ASYMPTOTIC OPTIMALITY OF GENERALIZED SEQUENTIAL LIKELIHOOD RATIO TESTS IN SOME CLASSICAL SEQUENTIAL TESTING PROBLEMS

BY

TZE LEUNG LAI

TECHNICAL REPORT NO. 7
DECEMBER 1988

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS 87-15614

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ASYMPTOTIC OPTIMALITY OF GENERALIZED SEQUENTIAL LIKELIHOOD RATIO TESTS IN SOME CLASSICAL SEQUENTIAL TESTING PROBLEMS

BY

TZE LEUNG LAI

TECHNICAL REPORT NO. 7
DECEMBER 1988

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS 87-15614

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ASYMPTOTIC OPTIMALITY OF GENERALIZED SEQUENTIAL LIKELIHOOD RATIO TESTS IN SOME CLASSICAL SEQUENTIAL TESTING PROBLEMS

Tze Leung Lai
Department of Statistics
Stanford University
Stanford, CA 94305

Abstract

Several classical optimality problems in sequential tests of composite hypotheses are reviewed and discussed herein. Asymptotic approximations are used to make the associated optimal stopping problems more tractable and to develop asymptotic lower bounds for the expected sample sizes of sequential tests subject to error probability constraints. Such asymptotic analysis of the different sequential testing problems leads to a unified asymptotic theory that gives certain generalized sequential likelihood ratio tests as the asymptotically optimal solutions to these problems.
1. INTRODUCTION

Let $X_1, X_2, \ldots$ be i.i.d. random variables with a common distribution $P$. Wald’s sequential probability ratio test (SPRT) of the null hypothesis $H : P = P_0$ versus $K : P = P_1$ stops sampling at stage

$$N = \inf\{n \geq 1 : R_n \geq A \text{ or } R_n \leq B\} \quad (\inf \phi = \infty),$$

(1.1)

where $A > 1 > B > 0$ are stopping bounds and

$$R_n = \prod_{i=1}^{n} \{p_i(X_i)/p_0(X_i)\}$$

(1.2)

is the likelihood ratio, $p_i$ being the density of $P_i$ with respect to some common dominating measure $\nu$ ($i = 0, 1$). When stopping occurs, $H$ or $K$ is accepted according as $R_N \leq B$ or $R_N \geq A$. The choice of the stopping bounds is dictated by the error probabilities $\alpha = P_0\{R_N \geq A\}$ and $\beta = P_1\{R_N \leq B\}$. This simple test was shown by Wald and Wolfowitz (1948) to be the optimal solution of testing $H$ versus $K$, in the sense that the SPRT minimizes both $E_0(T)$ and $E_1(T)$ among all tests whose sample size $T$ has a finite expectation under both $H$ and $K$, and whose error probabilities satisfy

$$P_0\{\text{Reject } H\} \leq \alpha \quad \text{and} \quad P_1\{\text{Reject } K\} \leq \beta.$$ 

(1.3)

To prove the optimality of the SPRT, a Lagrange–multiplier–type approach to handle the error constraints (1.3) leads to the auxiliary Bayes problem of minimizing

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

The Bayes solution can be decomposed into two components. The first component is the optimal terminal decision rule $\delta^*$, which turns out to be the rule that accepts the $H_i$ with the smaller posterior loss $\ell^*_n$, when stopping occurs at stage $n$. The second component of the Bayes test is the stopping rule $T^*$, which can be studied by the theory of optimal stopping for the stochastic sequence $cn + \ell^*_n$. In the present setting of a two–point prior distribution (simple null $P_0$ versus simple alternative $P_1$) and i.i.d. observations $X_1, X_2, \ldots$ with distribution $P_i$ given $i = 0$ or $i = 1$, the optimal stopping problem is of stationary Markov type (cf. Chow, Robbins and Siegmund, 1970), and the optimal stopping rule turns out to be of the simple form (1.1). Thus, the Bayes solution is an SPRT and the Wald–Wolfowitz theorem then follows by varying the parameters $p, c, w_0, w_1$ (cf. Lehmann, 1959).

Note the close analogy between Wald’s SPRT and the classical Neyman–Pearson fixed sample size test of the simple null hypothesis $H$ versus the simple alternative $K$, subject to the Type I error constraint $P_0\{\text{Reject } H\} \leq \alpha$. Both tests involve the likelihood ratios (1.2) and are solutions to natural optimization problems. While the Neyman–Pearson optimization criterion is to minimize the Type II error $P_1\{\text{Reject } K\}$ for the given sample size and Type I error bound, the Wald–Wolfowitz criterion is to minimize both $E_0T$ and $E_1T$ under the Type I and Type II error constraints (1.3). While the extension of the Neyman–Pearson theory to composite hypotheses is by no means an easy task, it is even much more difficult to extend the Wald–Wolfowitz theory beyond the setting of testing sequentially simple hypotheses with i.i.d. observations.
For fixed sample size tests, a first step to extend the Neyman–Pearson theory from simple to composite hypotheses is to consider one-sided composite hypotheses of the form $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ in the case of parametric families with monotone likelihood ratio in a real parameter $\theta$. In this case, the rejection region of the level-$\alpha$ Neyman–Pearson test of $H : \theta = \theta_0$ versus $K : \theta = \theta_1(> \theta_0)$ can be expressed as $\{T(X_1, \ldots, X_n) \geq c\}$, where the statistic $T(X_1, \ldots, X_n)$ does not involve $\theta_1$, and $c$ is so chosen that $P_{\theta_0}\{T(X_1, \ldots, X_n) \geq c\} = \alpha$, with possibly extraneous randomization on the event $\{T(X_1, \ldots, X_n) = c\}$. Hence this Neyman–Pearson test does not depend on the alternative $\theta_1$; moreover, by monotonicity, $P_{\theta}\{T(X_1, \ldots, X_n) \geq c\} \leq P_{\theta_0}\{T(X_1, \ldots, X_n) \geq c\} = \alpha$ for $\theta \leq \theta_0$ and therefore the test has level $\alpha$ for testing the composite hypotheses $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. Thus, the ability to reduce the composite hypotheses $H_0$ versus $H_1$ to the problem of simple hypotheses $H$ versus $K$ gives an optimal solution (in the sense of uniformly most powerful level-$\alpha$ tests) in this case (cf. Lehmann, 1959).

In the sequential setting, however, we cannot reduce the optimality considerations for one-sided composite hypotheses to those for simple hypotheses even in the presence of monotone likelihood ratio. For example, let $X_1, X_2, \ldots$ be i.i.d. normal random variables with mean $\theta$ and variance 1. To test $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_1(> \theta_0)$ subject to the error constraints

$$P_{\theta}\{\text{Reject } H_0\} \leq \alpha \quad \text{for } \theta \leq \theta_0, \quad P_{\theta}\{\text{Reject } H_1\} \leq \beta \quad \text{for } \theta \geq \theta_1,$$

one can use the SPRT of $H : \theta = \theta_0$ versus $K : \theta = \theta_1$ with Type I and Type II error probabilities $\alpha$ and $\beta$, and the monotone likelihood ratio structure implies that the SPRT also satisfies the error constraints (1.4) for the composite hypotheses problem. However, while this SPRT has minimum expected sample size at $\theta = \theta_0$ and at $\theta = \theta_1$ by the Wald–Wolfowitz theorem, its expected sample size may be unsatisfactory at other parameter points. In particular, for $\alpha = \beta$, its maximum expected sample size occurs at $\theta^* = (\theta_0 + \theta_1)/2$ and can be considerably larger than the optimal fixed sample size. This led Bechhofer (1960) to consider the minimax problem of finding a sequential test which minimizes $\sup_\theta E_\theta T$ over all tests that satisfy (1.4). In the case $\alpha = \beta$, the minimax problem can be reduced to a Bayes problem involving only three parameters $\theta_0, \theta_1$ and $\theta^*$. While the corresponding optimal stopping problem is Markovian, it is nonstationary and does not have a simple closed-form solution. However, as shown by Lai (1973), the complicated optimal solution can be well approximated by a simple stopping rule of form: Stop sampling as soon as $|\Sigma_i (X_i - \theta^*)| + (\theta_1 - \theta_0)n/4 \geq c$. This simple approximation is asymptotically optimal as $\alpha(= \beta) \to 0$, and Lorden (1976) has extended the result to general parametric families, where the corresponding asymptotically optimal procedure is a certain generalized sequential likelihood ratio test, which he calls a 2-SPRT. This and other developments in the minimax problem, often called the "Kiefer–Weiss problem", will be reviewed in Section 2.

Instead of the frequentist minimax approach to the preceding problem of testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_1$ for the mean of a normal distribution, or more generally, for the natural parameter of an exponential family, one may adopt a Bayesian approach, putting a prior distribution $G$ on $\theta$ and introducing a cost $c$ for each observation together with a loss $\ell(\theta)$, at $\theta$, of accepting the incorrect hypothesis. However, the optimal stopping problems associated with Bayes sequential tests do not give simple solutions beyond the two-point priors considered by Wald and Wolfowitz. A major thrust in the literature on Bayes sequential tests of $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_1$ is,
therefore, to develop simple approximations to the Bayes tests and their operating characteristics. Section 3 reviews the work of Schwarz (1962) and subsequent authors who developed asymptotic approximations by keeping \( \theta_0 \) and \( \theta_1 (> \theta_0) \) and letting \( c \to 0 \).

Another important development in the asymptotic theory of Bayes sequential tests of one-sided composite hypotheses is Chernoff’s (1961, 1965a,b, 1972) work on testing sequentially \( H'_0 : \theta < 0 \) versus \( H'_1 : \theta > 0 \) for the mean \( \theta \) of a normal distribution. Instead of assuming an indifference zone as in Schwarz’s theory, Chernoff’s theory assumes a loss of \(|\theta|\) for the wrong decision. His results and methods are reviewed in Section 3, and they are markedly different from those in Schwarz’s theory. Letting \( \theta_1 \) approach \( \theta_0 \) in Schwarz’s theory does not give Chernoff’s result, and this unsatisfactory situation of having two very different asymptotic approximations depending on whether one assumes a fixed indifference zone was recently resolved by Lai (1988a), who improved and refined the asymptotic approximation of Schwarz and obtained a Chernoff-type continuation region as a limiting case of this refined approximation when \( \theta_1 - \theta_0 \to 0 \). This unified asymptotic theory of Bayes tests of \( H_0 : \theta \leq \theta_0 \) versus \( H_1 : \theta \geq \theta_1 \) and of \( H'_0 : \theta < \theta_0 \) versus \( H'_1 : \theta > \theta_0 \), for the natural parameter \( \theta \) of an exponential family, leads to a class of generalized sequential likelihood ratio tests with time-varying stopping boundaries that will be reviewed in Section 4.

An important idea in the theory of optimal fixed sample size tests is to make use of invariance to reduce composite hypotheses to simple ones (cf. Lehmann, 1959). The same idea enables one to generalize Wald’s SPRT to an invariant SPRT, in which \( R_n \) in (1.1) now takes the form

\[
R_n = \frac{p_{1n}(T_n)}{p_{0n}(T_n)},
\]

(1.5)

where \( T_n \) is a maximal invariant with respect to a group \( G \) of transformations that leaves the problem invariant and \( p_n \) is the density of the maximal invariant under \( H_i \) \((i = 0, 1)\). This idea has led to the sequential \( t \)-test, the sequential \( F \)-test, the sequential \( T^2 \)-test, etc. (cf. Ghosh, 1970). However, since the proof of the Wald–Wolfowitz theorem on the optimum character of the SPRT based on i.i.d. observations depends very heavily on the fact that \( \{\log R_n, n \geq 1\} \) in (1.2) is a random walk, the argument breaks down when \( R_n \) is given by (1.5) instead, in which case \( \{\log R_n\} \) is no longer a random walk. Lai (1981) gave a general asymptotic analogue of the optimum character of Wald’s SPRT when the likelihood ratio \( R_n \) is based on dependent observations. Central to this development are asymptotic extensions of Wald’s (1945) lower bounds for the expected sample size of a sequential test. These and other related results will be reviewed in Section 5.

In their pioneering work, Wald and Wolfowitz (1948) laid the foundations of optimal stopping theory and successfully applied the theory to establish the optimum character of the SPRT for testing simple hypotheses. Although optimal stopping theory typically leads to computationally complicated and analytically intractable procedures in the case of composite hypotheses, we show in this review that for several classical sequential testing problems involving composite hypotheses, asymptotic approximations can be used to make the associated optimal stopping problems more tractable or to develop asymptotic lower bounds for the expected sample sizes of sequential tests. Such asymptotic analysis of these different sequential testing problems leads to a unified asymptotic theory that gives certain generalized sequential likelihood ratio tests as the asymptotically optimal solutions.
2. THE KIEFER–WEISS PROBLEM AND 2–SPRT

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

\[
f_\theta(x) = \exp\{\theta x - \psi(\theta)\}.
\]

Let \( \theta_1 > \theta_0 \) and consider the class \( T(\alpha, \beta) \) of all sequential tests \( (T, \delta) \), based on the \( X_i \), of \( H_0 : \theta \leq \theta_0 \) versus \( H_1 : \theta \geq \theta_1 \) subject to the error constraints (1.4), where \( T \) denotes the stopping rule and \( \delta \) denotes the terminal decision rule. The so-called "Kiefer–Weiss problem" is to minimize \( \sup_\theta E_\theta T \) over this class. Actually Kiefer and Weiss (1957) considered the following simpler problem:

**Minimize** \( E_\theta T \) **at a fixed parameter** \( \theta \) **subject to**

\[
P_{\theta_0}\{ \text{Reject } H_0 \} \leq \alpha \text{ and } P_{\theta_1}\{ \text{Reject } H_1 \} \leq \beta.
\]

They proved certain structure theorems about the optimal solution of (2.2) which they showed to be a generalized sequential probability ratio test (i.e., one whose stopping rule is of the form (1.2) but with the constant boundaries \( A, B \) in (1.1) replaced by time-varying boundaries \( A_n, B_n \)). Later Weiss (1962) showed that the minimax problem of minimizing \( \sup_\theta E_\theta T \), raised by Bechhofer (1960), can be reduced to the formulation (2.2) in symmetric cases involving normal and binomial distributions. He used this reduction to study the minimax problem.

In particular, for the case of normal \( N(\theta, 1) \) random variables \( X_i \) with \( \theta_0 = -\theta_1 \) and \( \alpha = \beta < \frac{1}{2} \), it follows from this reduction and the Kiefer–Weiss (1957) characterization of (2.2) that the minimax problem is equivalent to the Bayes problem of minimizing

\[
(1-p)E_0 T + (p/2)P_{\theta_1}\{ \text{Reject } H_0 \} + (p/2)P_{\theta_1}\{ \text{Reject } H_1 \},
\]

for some \( 0 < p < 1 \). Let \( S_n = X_1 + \ldots + X_n \). Since the Bayes terminal decision rule rejects \( H_0 \) or \( H_1 \) according as \( S_T > 0 \) of \( S_T < 0 \), the Bayes problem of minimizing (2.3) is further reduced to the optimal stopping problem of finding the stopping rule \( T \) that minimizes

\[
E_0\{ cT + \exp(-\theta_1^2 T/2 + \theta_1 |S_T|) \}, \quad \text{where} \quad c = 2(1-p)/p,
\]

cf. Lai (1973). The optimal stopping problem is Markovian but nonstationary, and the optimal stopping rule is of the form

\[
T^* = \min\{n \leq M : |S_n| \geq b_n \} \quad \text{with} \quad b_M = 0,
\]

where \( M \) is the smallest positive integer \( \geq -2\theta_1^{-2}\log c \). For every given value of \( c \), the stopping boundary \( b_n \) and the value function

\[
u_n(x) = \inf_{T \in \mathbb{N}} E_0[cT + \exp(-\theta_1^2 T/2 + \theta_1 |S_T|)|S_n = x]
\]

can be determined by backward induction on \( n \). Lai (1973) found simple upper and lower bounds for \( b_n \) and \( u_n \) and used these bounds to derive the following asymptotic shape of the optimal continuation region:

\[
\{(\log c)^{-1}(\theta_1^2 n, \theta_1 S) : n < M, |S| < b_n \} \rightarrow \{(t, w) : 0 \leq t < 2, |w| < 1 - t/2 \},
\]
the convergence being uniform in $\theta_1 \in C$ for any compact subset $C$ of $(0, \infty)$. These bounds developed for the discrete-time stopping problem can also be applied to give bounds on the partial derivatives of the value function in the corresponding continuous-time optimal stopping problem that replaces $\{S_n,n \geq 1\}$ in (2.4) by a standard Wiener process $\{W(t), t \geq 0\}$, and thereby to develop asymptotic approximations for the optimal stopping boundary in the continuous-time problem as well, cf. Lai (1973, Lemma 1 and Theorems 6 and 7). In particular, the asymptotic shape (2.5) for the optimal continuation region holds for the continuous-time problem as well. Note that the limiting continuation region in (2.5) is triangular and has a pair of symmetric line segments as its boundary. Such triangular continuation regions were first proposed by Anderson (1960) in his modification of the SPRT for testing the drift of a Wiener process to reduce the maximum expected sample size.

Lorden (1980) subsequently studied the problem (2.2) for the general exponential family (2.1), which can again be reduced to an optimal stopping problem of the form

$$\inf_{T} E\{T + \min[u \Pi_{t=1}^{T}(f_{\theta_1}(X_t)/f_{\theta}(X_t)), v \Pi_{t=1}^{T}(f_{\theta_1}(X_t)/f_{\theta}(X_t))\}], \quad (2.6)$$

for some $u \geq 0, v \geq 0$. Letting $U_0 = u, V_0 = v, U_n = U_{n-1} f_{\theta_1}(X_n)/f_{\theta}(X_n), V_n = V_{n-1} f_{\theta_1}(X_n)/f_{\theta}(X_n)$, Lorden observed that the optimal stopping problem in terms of the two-dimensional stochastic sequence $\{(U_n, V_n), n \geq 1\}$ is stationary Markov, although it is nonstationary as an optimal stopping problem in $\{X_n, n \geq 1\}$. A nonstationary Markovian stopping problem in $\{X_n\}$ means that the optimal stopping boundary is typically of the form $g(X_n, n) = 0$, where $g$ is a function of both the time variable $n$ and the space variable $X_n$, while a stationary problem means that the optimal stopping boundary is of the form $g(X_n) = 0$ with $g$ depending the the space variable $X_n$. Thus, a nonstationary problem in $X_n$ becomes stationary in terms of the two-dimensional space-time process $(X_n, n)$. The “state variable” $(U_n, V_n)$ that makes the problem stationary can be regarded as a transformation of the space-time variable $(X_n, n)$. Lorden developed bounds for the optimal continuation region parametrized by the two-dimensional state variable $(u, v)$. Assuming that $\theta_0 < \theta < \theta_1$ and choosing $a, b$ such that

$$a(\theta - \theta_0) + b(\theta - \theta_1) = 0, \quad a(\psi(\theta) - \psi(\theta_0)) + b(\psi(\theta) - \psi(\theta_1)) = -1,$$

he applied the transformation $t = a \log u + b \log v, s = \log(v/u)$ to establish the asymptotic behavior of the transformed optimal stopping boundary as $t \to \infty$, noting that $u V_n/v U_n$ reduces to the likelihood ratio $R_n$ defined in (1.2). This asymptotic behavior of the optimal stopping boundary agrees with that of the stopping boundary of an approximating 2-SPRT described below, extending Lai's (1973) results for the optimal stopping problem in the normal case.

Earlier Lorden (1976) extended Anderson's (1960) triangular stopping boundaries for testing the drift of a Wiener process to a class of simple sequential tests, called 2-SPRTs, for the problem of testing two simple hypotheses $H_0 : f = f_0$ versus $H_1 : f = f_1$ based on i.i.d. observations with a common density $f$ (with respect to some measure $\nu$) so that the expected sample size is minimized under $f = p$ (a third given density), at least asymptotically as the error probabilities $\alpha$ and $\beta$ tend to 0. The stopping rule of a 2-SPRT is of the form

$$M = M(A, B) = \inf \{n : \Pi_{i=1}^{n} f_0(X_i)/p(X_i) \leq A \text{ or } \Pi_{i=1}^{n} f_1(X_i)/p(X_i) \leq B\}, \quad (2.7)$$
and the test rejects \( H_0 \) if \( \Pi_1^M f_0(X_i)/f(X_i) \leq A \) and rejects \( H_1 \) if \( \Pi_1^M f_1(X_i)/p(X_i) \leq B \), where \( 0 \leq A, B \leq 1 \), are so chosen that the Type I and Type II error probabilities of the test are \( \alpha \) and \( \beta \), respectively. Let \( n(A, B) \) denote the infimum of \( E_{\theta} N \) over all sequential tests whose Type I and Type II error probabilities are bounded by \( \alpha \) and \( \beta \), respectively. Under the assumption that \( E_{\theta} \{ \log^2[f_0(X_1)/p(X_1)] + \log^2[f_1(X_1)/p(X_1)] \} < \infty \), Lorden (1976) showed that

\[
E_{\theta} M(A, B) - n(A, B) \to 0 \quad \text{as} \quad \min(A, B) \to 0. \tag{2.8}
\]

While Lorden (1976, 1980) focussed on the modified problem (2.2) instead of the minimax Kiefer–Weiss problem, Huffman (1983) subsequently extended his results to show that suitably chosen 2-SPRTs indeed provide asymptotically optimal solutions to the minimax sequential testing problem of \( H_0 : \theta \leq \theta_0 \) versus \( H_1 : \theta \geq \theta_1 \) in a general exponential family and without assuming \( \alpha = \beta \), in the sense of a weaker conclusion than (2.8). First let

\[
I(\theta, \lambda) = E_\theta \log \{ f_\theta(X_1)/f_\lambda(X_1) \} = (\theta - \lambda) \psi'(\theta) - (\psi(\theta) - \psi(\lambda)) \tag{2.9}
\]
denote the Kullback–Leibler information number, and define \( \theta^* \in (\theta_0, \theta_1) \) and \( n^* \) by

\[
|\log \alpha|/I(\theta^*, \theta_0) = |\log \beta|/I(\theta^*, \theta_1) = n^*. \tag{2.10}
\]

Let \( a_i^* = (\theta^* - \theta_i)/I(\theta^*, \theta_i) \) for \( i = 0, 1 \) and define \( r^* \) by \( \Phi(r^*) = -a_1^*/(a_0^* - a_1^*) \), where \( \Phi \) is the standard normal distribution function. Letting

\[
\tilde{\theta} = \theta^* + r^*[n^* \psi''(\theta^*)]^{-\frac{1}{2}},
\]

define the 2-SPRT with stopping rule

\[
\tilde{M} = \inf \{ n : \Pi_1^n f_{\theta_0}(X_i)/f_{\tilde{\theta}}(X_i) \leq \alpha \quad \text{or} \quad \Pi_1^n f_{\theta_1}(X_i)/f_{\tilde{\theta}}(X_i) \leq \beta \}. \tag{2.10}
\]

Noting that the 2-SPRT belongs to the class \( T(\alpha, \beta) \) of tests satisfying the error constraints (1.4), Huffman (1983) showed that as \( \alpha \to 0 \) and \( \beta \to 0 \) such that \( \log \alpha \sim \log \beta \),

\[
(\sup_\theta E_\theta \tilde{M})/\inf_\theta \{ \sup_\theta E_\theta T : (T, \delta) \in T(\alpha, \beta) \} = 1 + o(|\log \alpha|^{-\frac{1}{2}}). \tag{2.11}
\]

He also carried out extensive computer calculations for the case of testing the unknown mean of an exponential distribution to compare the proposed 2-SPRT with the minimax test. His numerical results show that the 2-SPRT comes within 2% of minimizing the maximum expected sample size over a broad range of error probability and parameter values. Earlier, Lorden (1976) carried out similar calculations for the problem of testing the mean of a normal distribution in the symmetric case \( \alpha = \beta \) (so that the minimax problem reduces to the problem (2.2)) and found that the 2-SPRT comes within 1% of minimizing the maximum expected sample over a broad range of parameter values.

Summarizing, 2-SPRTs provide excellent approximate solutions to the problem (2.2) of minimizing the expected sample size, at a given parameter value \( \theta' \), of a sequential test of \( H_0 : \theta \leq \theta_0 \).
versus \( H_1 : \theta \geq \theta_1 \) subject to the error constraints (1.4). They are easy to use (cf. Lorden, 1976) and can be interpreted as generalized sequential likelihood ratio tests for the problem (2.2). As will be discussed in Section 4, this also suggests the form of generalized sequential likelihood ratio tests for more general testing problems. Moreover, a suitable choice of \( \theta^* \) in the 2–SPRT also provides an asymptotically optimal solution of the minimax problem of minimizing \( \sup_{\theta} E_{\theta} T \) subject to the error constraints (1.4).

3. ASYMPTOTIC THEORIES OF CHERNOFF AND SCHWARZ FOR BAYES SEQUENTIAL TESTS OF ONE–SIDED HYPOTHESES

Consider the Bayes problem of testing \( H_0 : \theta \leq \theta_0 \) versus \( H_1 : \theta \geq \theta_1 \) for the parameter \( \theta \) of the exponential family (2.1), with a prior distribution \( G \) on \( \theta \), a loss \( \ell(\theta) \) at \( \theta \) of accepting the incorrect hypothesis and a cost \( c \) per observation. The Bayes risk of a sequential test \((T, \delta)\) with stopping rule \( T \) and terminal decision rule \( \delta \) is given by

\[
\mathcal{R}(T, \delta) = c \int E_{\theta} T dG + \int_{\theta \leq \theta_0} \ell(\theta) P_{\theta} \{\delta \text{ accepts } H_1\} dG \\
+ \int_{\theta \geq \theta_1} \ell(\theta) P_{\theta} \{\delta \text{ accepts } H_0\} dG.
\]  

(3.1)

The support of \( G \) is assumed to be contained in the natural parameter space \( \Theta = \{\theta : \int e^{\theta x} d\nu(x) < \infty\} \) and to satisfy \( \int \ell(\theta) dG(\theta) < \infty \).

A well known asymptotic solution to the Bayes problem of minimizing \( \mathcal{R}(T, \delta) \) is due to Schwarz (1962). Let \( B(c) \) denote the continuation region of the Bayes rule (which continues sampling at stage \( n + 1 \) iff \( (n, S_n) \in B(c) \)). 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's asymptotic theory leads to the following limiting continuation region of the Bayes rule: As \( c \to 0, \)

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

(3.2)

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

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

or equivalently, to stop sampling at stage

\[
N_c = \inf\{n \geq 1 : \max \{\Pi_{i=1}^{n} f_{\hat{\theta}_n}(X_i) / \Pi_{i=1}^{n} f_{\theta_0}(X_i), \Pi_{i=1}^{n} f_{\hat{\theta}_n}(X_i) / \Pi_{i=1}^{n} f_{\theta_1}(X_i)\} \geq c^{-1}\},
\]

(3.4)

where \( \hat{\theta}_n \) is the maximum likelihood estimate of \( \theta \). The terminal decision rule \( \delta^* \) is to accept \( H_1 \) (or \( H_0 \)) if \( \Pi_{i=1}^{N_c} f_{\hat{\theta}_1}(X_i) > (\leq) \Pi_{i=1}^{N_c} f_{\theta_0}(X_i) \).
There are two main steps in Schwarz's derivation of the above asymptotic approximation to the Bayes rule. The first step involves upper and lower bounds for the Bayes continuation region $B(c)$. Let
\[ L(n, x) = \frac{\min_{i=0,1} \ell(\theta_i) \exp\{\theta x - n\psi(\theta)\} dG(\theta)}{\int \exp\{\theta x - n\psi(\theta)\} dG(\theta)} \tag{3.5} \]
be the so-called "stopping risk," which is the posterior loss due to the wrong decision of the Bayes test if stopping occurs at stage $n$ and $S_n = x$. Let $\mathcal{R}(c) = \{(n, x) : L(n, x) \geq c\}$. Schwarz showed that for sufficiently small $c > 0$,
\[ \mathcal{R}(c) \supset B(c) \supset \mathcal{R}(3\Delta^{-1}c|\log c|), \tag{3.6} \]
where $\Delta = \psi(\theta_0) + \psi(\theta_1) - 2\psi((\theta_0 + \theta_1)/2)$. The next step is to apply Laplace's asymptotic method to evaluate the integrals in (3.5), leading to the asymptotic approximation
\[ \log L(n, S_n) \sim \min_{i=0,1} \{\theta_i S_n - n\psi(\theta_i)\} - \{\hat{\theta}_n S_n - n\psi(\hat{\theta}_n)\}. \tag{3.7} \]
Combining (3.7) and (3.6) gives the "asymptotic shape" (3.3) for the Bayes continuation region $B(c)$.

Let $J(\theta) = \max\{I(\theta, \theta_0), I(\theta, \theta_1)\}$, where $I(\theta, \lambda)$ is the Kullback–Leibler information number (2.9). Wong (1968) showed that for fixed $\theta_0$ and $\theta_1$, as $c \to 0$,
\[ E_\theta N_c \sim |\log c| / J(\theta) \quad \text{for every} \quad \theta, \]
\[ \sup_{\delta \leq \delta_0} P_\theta\{\delta^* \text{ accepts } H_1\} = o(c|\log c|) = \sup_{\theta \geq \theta_1} P_\theta\{\delta^* \text{ accepts } H_0\}, \]
and the Bayes risk of the test $(N_c, \delta^*)$ satisfies
\[ r(N_c, \delta^*) \sim c|\log c| \int d\pi(\theta)/J(\theta) \sim \inf_{(T, \delta)} r(T, \delta). \tag{3.8} \]
Hence the test $(N_c, \delta^*)$ is asymptotically Bayes both in the sense of the asymptotic shape (3.2) and in the sense of the asymptotic risk (3.8) as $c \to 0$, for fixed $\theta_0 < \theta_1$.

A different asymptotic theory in Bayes sequential tests was developed by Chernoff (1961, 1965 a,b) in the context of testing $H_0 : \theta < 0$ versus $H_1 : \theta > 0$ for the mean $\theta$ of a normal distribution with unit variance. Instead of assuming an indifference zone $(\theta_0, \theta_1)$ and a general loss function $\ell(\theta) > 0$ for $\theta \notin (\theta_0, \theta_1)$ as in Schwarz's theory, Chernoff's theory considers the special loss function $\ell(\theta) = |\theta|$ for $\theta \neq 0$ and assumes a normal prior distribution $G$ with mean 0 and variance $\sigma^2$. The Bayes terminal decision rule accepts $H_0$ (or $H_1$) according as $S_n \leq 0$ (or $S_n > 0$) when stopping occurs at stage $n$. Thus, the Bayes problem reduces to the optimal stopping problem of finding an optimal stopping rule to minimize
\[ r(T) = c \int_0^\infty E_\theta T dG + \int_0^- |\theta| P_\theta\{S_T > 0\} dG + \int_0^\infty \theta P_\theta\{S_T \leq 0\} dG. \tag{3.9} \]
To study the optimal stopping rule, Chernoff (1961) introduced the normalization
\[ t = c^{2/3}(n + \sigma^{-2}), \quad w = c^{1/3} S_n, \tag{3.10} \]
which is different from Schwarz’s normalization \( t = n/|\log c| \) and \( w = S_n/|\log c| \). With the normalization (3.10) for the problem, Chernoff obtained 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

\[
f(t) = \left\{ 3t[\log t^{-1} - (\log 8\pi)/3 + o(1)] \right\}^{1/2} \quad \text{as} \quad t \to 0, \tag{3.11}
\]

\[
f(t) \sim (4t)^{-1} \quad \text{as} \quad t \to \infty, \tag{3.12}
\]

(cf. Chernoff, 1965a, Breakwell and Chernoff, 1964). It is interesting to compare this with the boundary \( |w| = [(2t)^{1/2} - \Delta t]^+ \) in Schwarz’s limiting region (3.3) (with the different normalization \( t = n/|\log c|, \ w = S_n/|\log c| \)) in the normal case with \( \theta_1 = \Delta = -\theta_0 \).

There is, therefore, a big difference in the asymptotic solutions as \( c \to 0 \) between Schwarz’s theory, which assumes a fixed indifference zone \((\theta_0, \theta_1)\) and Chernoff’s theory, which assumes (instead of an indifference zone) the special loss function \(|\theta|\) 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 theory appears to be considerably simpler than Chernoff’s theory, and is derived in the general context of the exponential family instead of Chernoff’s special setting of the normal distribution.

A basic problem with Schwarz’s theory is whether the generalized sequential likelihood ratio test with stopping rule (3.4) actually provides an adequate approximation to the Bayes rule. To obtain a more refined approximation, Fushimi (1967) and Schwarz (1969) replaced the asymptotic approximation (3.7) for \( \log L(n, S_n) \) by a second-order asymptotic expansion. However, it is not clear that these modifications provide actual improvement of the Bayes risks. A more basic question in connection with possible improvements of Schwarz’s asymptotic theory, considered by Kiefer and Sacks (1963) and by Lorden (1967), is to find out which (if any) of Schwarz’s bounds \( \mathcal{R}(c) \) and \( \mathcal{R}(3\Delta^{-1}|\log c|) \) for the Bayes continuation region in (3.6) is of the right order of magnitude as \( c \to 0 \). Let \( T_c \) denote the stopping rule with continuation region \( \mathcal{R}(c) \) and \( \delta_G \) denote the terminal decision rule that chooses the \( \Theta_i \) which gives the smaller posterior risk with respect to the prior distribution \( G \). Let \( (T^*, \delta_G) \) be the Bayes rule that minimizes the Bayes risk (3.1). Kiefer and Sacks (1963) showed that \( r(T_c, \delta_G) \sim r(T^*, \delta_G) \) if \( G \) has compact support. Lorden (1967) later refined this result as

\[
r(T_{Qc}, \delta_G) - r(T^*, \delta_G) \leq cM(Q) \quad \text{for every} \quad Q > 0 \quad \text{and} \quad c > 0, \tag{3.13}
\]

where \( M(Q) \) is a constant depending only on \( Q \).

Another issue concerning the adequacy of Schwarz’s approximation (3.4) to the Bayes rule, noted by Lai (1988a), is the lack of uniformity in the indifference zone parameters \( \theta_0, \theta_1 \) in the convergence behavior (3.2) of the normalized Bayes continuation region. Compare this with the asymptotic shape result (2.5) for the Kiefer-Weiss problem, which is uniform in the indifference zone parameters \( \pm \theta_1 \) over compact sets of \( \theta_1 \) values. Note also that in the normalization for (2.5), \( \theta_1 \) appears in \((\theta_1^2 \sigma, \theta_1 S)\), in addition to the scaling factor \(|\log c|^{-1}\) for both coordinates. To see how the lack of uniformity in the indifference zone parameters may lead to difficulties in applying
Schwarz's asymptotic shape (3.2), consider the simple case where \( \theta_0 = -\theta_1 < 0 \) and the \( X_i \) are normal with mean \( \theta \) and variance 1, so that \( J(\theta) = (|\theta| + \theta_1)^2/2 \). Although \( c \) may be very small (say \( c = 10^{-20} \)) and \( \theta_1 \) appears to be much larger so that regarding it as "fixed" seems justified (say \( \theta_1 = 0.1 \)), \( |\log c| \) may turn out to be smaller than \( 1/J(\theta) \) for \( |\theta| \leq \theta_1 \). Since the asymptotic formula for the Bayes risk (3.8) involves \( |\log c|/J(\theta) \), if one uses the "c \( \rightarrow 0 \)" approximation in this case, it seems more reasonable not to consider \( \theta_1 \) as "fixed" but to also let \( \theta_1 \rightarrow 0 \), as pointed out by Lai (1988a).

4. GENERALIZED SEQUENTIAL LIKELIHOOD RATIO TESTS
AND A UNIFIED ASYMPTOTIC THEORY FOR BAYES
SEQUENTIAL TESTS OF ONE-SIDED HYPOTHESES

It seems somewhat artificial that there should be two different kinds of asymptotic approximations to Bayes tests of one-sided hypotheses, depending on whether there is an indifference zone or not. A more natural asymptotic theory should have the property that the approximation in the absence of an indifference zone can be obtained as the limit of the approximations with shrinking indifference zones. To see the feasibility of this unified approach, Lai (1988a) first studied the problem of testing sequentially \( H_0 : \mu \leq -\gamma \) versus \( H_1 : \mu \geq \gamma \) (with an indifference zone \((-\gamma, \gamma))\) and \( H'_0 : \mu < 0 \) versus \( H'_1 : \mu > 0 \) (without an indifference zone) for the drift coefficient \( \mu \) of a Wiener process \( \{w(t), t \geq 0\} \), assuming the \( 0-1 \) loss, a flat prior on \( \mu \), and a cost of \( t \) for observing the process for a period of length \( t \). Note that the Bayes terminal decision rule for either problem accepts the null or the alternative hypothesis according as \( w(t) < 0 \) or \( w(t) > 0 \) when stopping occurs at time \( t \).

For the problem of testing \( H'_0 : \mu < 0 \) versus \( H'_1 : \mu > 0 \), the posterior loss \( L_0(t, w) \) at time \( t \) if \( w(t) = w \) is observed and we decide to stop and accept \( H'_0 \) or \( H'_1 \) according as \( w < 0 \) or \( w > 0 \) is given by

\[
L_0(t, w) = t + \Phi(-|w| t^{-\frac{1}{2}}),
\]

where \( \Phi \) is the standard normal distribution function. For the problem of testing \( H_0 : \mu \leq -\gamma \) versus \( H_1 : \mu \geq \gamma \), the corresponding posterior loss is given by

\[
L_\gamma(t, w) = t + \Phi(-|w| t^{-\frac{1}{2}} - \gamma t^{\frac{1}{2}}).
\]

Thus, (3.1) can be regarded as a special case of (4.2) with \( \gamma = 0 \). For \( \gamma \geq 0 \), introduce the transformation

\[
s = 1/t, \quad y = w/t, \quad v_\gamma(s, y) = L_\gamma(t, w).
\]

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

\[
\frac{1}{2} \frac{\partial^2 u}{\partial y^2} = \frac{\partial u}{\partial s} \quad \text{for} \quad |y| < y_\gamma(s),
\]

\[
u(s, y) = v_\gamma(s, y) \quad \text{and} \quad \partial u/\partial y = \partial v_\gamma/\partial y \quad \text{for} \quad y = y_\gamma(s).
\]

10
Since \( s = 1/t \) and \( Y(s) = w(t)/t \), the optimal stopping rule can be written in the form

\[
\tau_\gamma = \inf \{ t > 0 : |w(t)| \geq h_\gamma(t) \}, \quad \text{where} \quad h_\gamma(t) = ty_\gamma(1/t).
\]
(4.5)

An asymptotic analysis of the free boundary problem (4.4) gives the following results on the asymptotic behavior of \( h_\gamma(t) \), which is positive for \( t > 0 \), as \( t \to 0 \) and as \( t \to \infty \).

**LEMMA 1** (Lai, 1988a). (i) For fixed \( 0 \leq \gamma < \infty \), as \( t \to 0 \),

\[
h_\gamma(t) = \left\{ 2t [\log t^{-1} + \frac{1}{2} \log \log t^{-1} - \frac{1}{2} \log 4\pi + o(1)] \right\}^{1/2}.
\]
(4.6)

(ii) Let \( \gamma \to \infty \). Then the asymptotic expansion (4.6) still holds as \( t \to 0 \) such that \( t = o((\gamma^2 \log \gamma^2)^{-1}) \).

(iii) For fixed \( 0 \leq \gamma < \infty \), as \( t \to \infty \),

\[
h_\gamma(t) \sim \frac{1}{4} \left( \frac{2}{\pi} \right)^{1/2} t^{-1/2} \exp\left\{ -\frac{1}{2} \gamma^2 t \right\}.
\]
(4.7)

(iv) Let \( \gamma \to \infty \). Then \( \sup_{t \geq \lambda \gamma^{-2} \log \gamma^2} h_\gamma(t) \to 0 \) for every \( \lambda > 2 \). Moreover, for every fixed \( 0 < \rho < 2 \), as \( t \to 0 \) such that \( t \leq \rho \gamma^{-2} \log \gamma^2 \),

\[
h_\gamma(t) = -\gamma t + \left\{ 2t [\log t^{-1} + \frac{1}{2} \log \log t^{-1} + O(1)] \right\}^{1/2}.
\]
(4.8)

Lemma 1 shows that the optimal boundary for the problem of testing \( H_0 : \mu \leq -\gamma \) versus \( H_1 : \mu \geq \gamma \) for fixed \( \gamma > 0 \) has the same asymptotic behavior (4.6) for small \( t \) as that for the problem of testing \( H_0' : \mu < 0 \) versus \( H_1' : \mu > 0 \). Moreover, even when \( \gamma \to \infty \) (large indifference zones), (4.6) still holds for \( t = o((\gamma^2 \log \gamma^2)^{-1}) \) and a modified form (4.8) also holds for \( t \leq (2-\epsilon)\gamma^{-2} \log \gamma^2 \). Thus, in the case of \( 0 - 1 \) loss, there is a unified asymptotic theory as \( t \to 0 \) for continuous-time Bayes sequential tests of \( H_0' \) versus \( H_1' \) and of \( H_0 \) versus \( H_1 \).

Using the numerical methods developed by Chernoff and Petkau (1986) to compute the optimal stopping boundaries \( h_\gamma \) for a variety of values of \( \gamma \) and combining the numerical results with the above asymptotic results, Lai (1988a) suggested the following simple approximations to \( h_\gamma \). First, for \( \gamma = 0 \), define the approximation

\[
h_0^*(t) = \frac{1}{4} \left( \frac{2}{\pi} \right)^{1/2} (t^{-1/2} - 5t^{-5/2}/48\pi) \quad \text{if} \quad t \geq 0.8,
\]

\[
= \exp(-0.69t - 1) \quad \text{if} \quad 0.1 \leq t < 0.8,
\]

\[
= 0.39 - 0.015t^{-\frac{3}{2}} \quad \text{if} \quad 0.01 \leq t < 0.1,
\]

\[
= \{4[2 \log t^{-1} + \log \log t^{-1} - \log 4\pi - 3\exp(-0.016t^{-\frac{3}{2}})]\} \quad \text{if} \quad t < 0.01.
\]

For \( 0 < \gamma \leq 20 \), use the approximation

\[
h_\gamma^*(t) = h_0^*(t) \exp(-\frac{1}{2} \gamma^2 t) \quad \text{if} \quad t \geq 1,
\]

\[
= h_0^*(t) \exp(-\frac{1}{2} \gamma^2 t^{1.125}) \quad \text{if} \quad 0 < t < 1.
\]
For $\gamma > 20$, use the approximation

$$
h^*_\gamma(t) = \lfloor (2 \log t^{-1} + \log \log t^{-1} - \log 4\pi)^{\frac{1}{2}} - \gamma t \rfloor^+.
$$

Lai (1988a) next considered the problems of testing (i) $H_0^' : \theta < 0$ versus $H_1^' : \theta > 0$, and

(ii) $H_0 : \theta \leq -\Delta$ versus $H_1 : \theta \geq \Delta$ for the mean $\theta$ of i.i.d. normal random variables $X_1, X_2, \ldots$.

Define

$$
t = cn, \quad w(t) = c^{\frac{1}{3}} S_n, \quad \mu = c^{-\frac{1}{2}} \theta, \quad \gamma = c^{-\frac{1}{2}} \Delta.
$$

(4.9)

Since $c^{\frac{1}{3}} \theta n = \mu t$, $w(t)$ is a Wiener process with drift coefficient $\mu$ and with $t$ 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 $G$ on $\theta$ such that $G$ has a positive continuous density $G'$, the density function $\pi_c$ of $\mu = c^{-\frac{1}{2}} \theta$ is given by

$$
\pi_c(x) = c^{\frac{1}{4}} G'(c^{\frac{1}{3}} x) \sim c^{\frac{1}{3}} G'(0) \quad \text{as} \quad c \to 0,
$$

and thereby the family of probability measures with densities $\pi_c$ converges to Lebesgue measure (flat prior). This suggests using the flat-prior continuous-time Bayes stopping boundaries $h_0$ and $h_\gamma$, or their approximations $h^0_0, h^*\gamma$, as approximations to the Bayes tests of $H_0^'$ versus $H_1^'$ and of $H_0$ versus $H_1$, respectively.

Lai (1988a) then extended these approximations from the normal distribution to the exponential family (2.1), under the assumption that $\theta$ is known to lie in an open interval $A$ with end-points $-\infty < a_1 < a_2 \leq \infty$ such that

$$
\inf_{a_1 - \eta < \theta < a_2 + \eta} \psi''(\theta) > 0, \quad \sup_{a_1 - \eta < \theta < a_2 + \eta} \psi''(\theta) < \infty, \quad \text{and} \quad \psi''
$$

(4.10)

is uniformly continuous on $(a_1 - \eta, a_2 + \eta)$ for some $\eta > 0$.

Since $\theta$ is known to lie in $A$, the maximum likelihood estimate $\hat{\theta}_n$ is obtained from the sample mean $\overline{X}_n = n^{-1} S_n$ by

$$
\hat{\theta}_n = (\psi')^{-1}(\overline{X}_n) \quad \text{if} \quad \psi'(a_1) < \overline{X}_n < \psi'(a_2),
$$

(4.11)

$$
= a_1 \quad \text{if} \quad \overline{X}_n \leq \psi'(a_1),
$$

$$
= a_2 \quad \text{if} \quad \overline{X}_n \geq \psi'(a_2).
$$

Of primary importance in the extension is the Kullback–Leibler information number $I(\theta, \lambda)$ in (2.9).

Note that in the case of a normal distribution with mean $\theta$ and variance 1, $I(\theta, \lambda) = (\theta - \lambda)^2 / 2$, and $\hat{\theta}_n = \overline{X}_n$ (when $A = (-\infty, \infty)$). Moreover, letting $g_\gamma(t) = (h_\gamma(t) + \gamma t)^2 / 2t$ for $\gamma \geq 0$, it follows from (4.9) that

$$
|w(t)| \geq h_\gamma(t) \iff |w(t)| + \gamma t)^2 / 2t \geq g_\gamma(t)
$$

$$
\iff (|S_n| + \Delta n)^2 / 2n \geq g_\gamma(cn)
$$

$$
\iff \max\{I(\hat{\theta}_n, \theta_0), I(\hat{\theta}_n, \theta_1)\} \geq n^{-1} g_\gamma(cn)
$$

where $\theta_1 = \Delta = -\theta_0$ in the case $H_0 : \theta \leq -\Delta$ versus $H_1 : \theta \geq \Delta$, and $\theta_1 = 0 = -\theta_0$ in the case $H : \theta < 0$ versus $K : \theta > 0$. Hence, the aforementioned approximation to the Bayes stopping rule for $H_0$ versus $H_1$ can be expressed in the form

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

(4.12)
with \( g = g_\gamma \), while the stopping rule for \( H_0' \) versus \( H_1' \) can be expressed as
\[
T_c = \inf\{ n \geq 1 : I(\hat{\theta}_n, \theta_0) \geq n^{-1} g_0(cn) \}. \tag{4.13}
\]

Lai (1988a) showed that the stopping rules (4.12) and (4.13) are asymptotically optimal, as \( c \to 0 \), for testing \( H_0 : \theta \leq \theta_0 \) versus \( H_1 : \theta \geq \theta_1 \) and for testing \( H : \theta < \theta_0 \) versus \( K : \theta > \theta_0 \), respectively, for the parameter \( \theta \) of the exponential family (2.1), with respect to the \( 0-1 \) loss, cost \( c \) per observation, and a general class of prior distributions on \( A \). By (4.11) and (2.9),
\[
nI(\hat{\theta}_n, \theta) = \log \left\{ \prod_{i=1}^{n} f_{\theta_i}(X_i) / \prod_{i=1}^{n} f_{\theta}(X_i) \right\}, \quad \text{if } X_n \in \psi'(A). \tag{4.14}
\]

Therefore, Schwarz's stopping rule (3.4) for testing \( H_0 : \theta \leq \theta_0 \) versus \( H_1 : \theta \geq \theta_1 \) is essentially equivalent to
\[
\inf\{ n \geq 1 : \max[I(\hat{\theta}_n, \theta_0), I(\hat{\theta}_0, \theta_0)] \geq n^{-1} |\log c| \}. \tag{4.15}
\]

Thus, the stopping rule (4.12) simply modifies Schwarz's rule by using \( g(cn) \) in place of the factor \( |\log c| \) in (4.15). Moreover, letting \( \theta_1 = \theta_0 \) and \( g = g_0 \) in (4.12) leads to the stopping rule (4.13) for testing \( H_0' : \theta < \theta_0 \) versus \( H_1' : \theta > \theta_0 \). The stopping rules (4.12) and (4.13) therefore provide a unified treatment of testing \( H_0 \) versus \( H_1 \) (with an indifference zone) and of testing \( H_0' \) versus \( H_1' \) (without an indifference zone).

As we have indicated in Section 3, a difficulty with Schwarz's approximation (4.15) for the Bayes rule is that the convergence behavior (3.2) of the normalized Bayes continuation region is not uniform in \( \theta_0 \) and \( \theta_1 \) over compact sets. Investigation of the special case of normal random variables by analyzing the associated continuous-time optimal stopping problem suggests that \( |\log c| \) in (4.15) should in fact be of the form \( g(cn) \) instead, with \( g(t) \sim \log t^{-1} \) as \( t \to 0 \). Important insights into why such modification is able to redress the lack of uniformity in the indifference zone parameters in Schwarz's asymptotic shape approximation, not only for normal distributions but also for general exponential families, are provided by the following asymptotic results on boundary crossing probabilities and expected crossing times that are uniform in \( \theta \).

**LEMMA 2** (Lai, 1988 a,b). Let \( X_1, X_2, \ldots \) be i.i.d. random variables having common density (2.1) with respect to some nondegenerate measure \( \nu \). Let \( A = (a_1, a_2) \), with \(-\infty \leq a_1 < a_2 \leq \infty \), be a subinterval of the natural parameter space \( \Theta \) such that (4.10) holds. Let \( g \) be a nonnegative function on \((0, \infty)\) such that \( \sup_{t \geq a} g(t) / t < \infty \) for all \( a > 0 \) and
\[
g(t) \sim \log t^{-1} \quad \text{and} \quad g(t) \geq \log t^{-1} + \xi \log \log t^{-1} + \mathcal{O}(1) \quad \text{as} \quad t \to 0, \tag{4.16}
\]
for some real number \( \xi \). For \( c > 0 \) and \( \theta_0, \theta_1 \in A \) with \( \theta_0 < \theta_1 \), define \( N(g, c) \) by (4.12) and let
\[
T(g, c) = \inf\{ n \geq 1 : I(\hat{\theta}_n, \theta_0) \geq n^{-1} g(cn) \},
\]
where \( \hat{\theta}_n \) is defined in (4.11). Let \( d_c < D_c \) be positive numbers satisfying
\[
d_c \to 0 \quad \text{and} \quad d_c / e^{\xi} \to \infty, \quad D_c \to \infty \quad \text{and} \quad D_c = o(|\log c|^{\frac{1}{2}}), \quad \text{as} \quad c \to 0. \tag{4.17}
\]

(i) As \( c \to 0 \), \( E_{\theta} T(g, c) \sim \{ \log(c^{-1}(\theta - \theta_0)^2) \} / I(\theta, \theta_0) \) uniformly in \( \theta \in A \) and \( \theta_0 \in A \) with \( d_c \leq |\theta - \theta_0| \leq D_c \). Moreover,
\[
P_{\theta} \{ \hat{\theta}_{T(\theta, c)} > \theta_0 \} I_{\{ \theta < \theta_0 \}} + P_{\theta} \{ \hat{\theta}_{T(\theta, c)} > \theta_0 \} I_{\{ \theta > \theta_0 \}} = O\left( c(\theta - \theta_0)^{-2} \{ \log(c^{-1}(\theta - \theta_0)^2) \}^{-\xi^{-\frac{1}{2}}} \right).
\]
uniformly in \( \theta \in \mathcal{A} \) and \( \theta_0 \in \mathcal{A} \) with \( |\theta - \theta_0| \geq d_c \).

(ii) The function \( J(\theta) = \max\{I(\theta, \theta_0), I(\theta, \theta_1)\} \) is maximized at \( \theta^* \), where \( \theta_0 < \theta^* < \theta_1 \) is defined by \( I(\theta^*, \theta_0) = I(\theta^*, \theta_1) \). Let \( n_c = \inf\{n : nJ(\theta^*) \geq g(cn)\} \). Then \( N(g, c) \leq n_c \) and as \( c \to 0 \),

\[
 n_c \sim \frac{\log(c^{-1}(\theta_1 - \theta_0)^2))}{J(\theta^*)} \text{ uniformly in } \theta_0, \theta_1 \in \mathcal{A} \text{ with } d_c \leq \theta_1 - \theta_0 \leq D_c, \\
 \mathbb{E}_\theta N(g, c) \sim \frac{\log(c^{-1}J(\theta))}{J(\theta)} \text{ uniformly in } \theta, \theta_0, \theta_1 \in \mathcal{A} \text{ with } d_c^2 \leq J(\theta) \leq D_c^2.
\]

\[
P_{\theta}\{\hat{\theta}_{N(g, c)} \geq \theta^*\} = O\left(\frac{c(\theta - \theta_0)^{-2}}{\log(c^{-1}(\theta_1 - \theta_0)^2))}^{-\frac{\xi-\frac{1}{2}}{2}}\right) \text{ uniformly in } \theta, \theta_0, \theta_1 \in \mathcal{A} \text{ with } \theta_0 - \theta \geq d_c \text{ and } \theta_1 - \theta_0 \geq d_c,
\]

\[
P_{\theta}\{\hat{\theta}_{N(g, c)} \geq \theta^*\} = O\left(\frac{c(\theta_1 - \theta_0)^{-2}}{\log(c^{-1}(\theta_1 - \theta_0)^2))}^{-\frac{\xi+3/2}{2}}\right) \text{ uniformly in } \theta_0, \theta_1 \in \mathcal{A} \text{ with } \theta_1 - \theta_0 \geq d_c.
\]

Similar results also hold for \( P_{\theta}\{\hat{\theta}_{N(g, c)} \leq \theta^*\} \) with \( \theta \geq \theta_1 \).

Note that in Lemma 2(i), the probability is of a smaller order of magnitude than \( cE_\theta T(g, c) \) if \( \xi > -3/2 \), while in Lemma 2(ii), the probabilities are of a smaller order of magnitude than \( cE_\theta N(g, c) \) if \( \xi > -1/2 \). This suggests that the asymptotic behavior as \( t \to 0 \) of the boudaries \( g(t) = (h_\gamma(t)+\gamma t)^2/2t \), with \( \gamma = (\theta_1-\theta_0)/2c^\frac{1}{2} \), given in Lemma 1 indeed provides an asymptotically optimal balance between the expected sample size and the probability of correct decision (cf. Lai, 1988b, for further discussion of this point and for more precise results on the probabilities). Making use of Lemma 2 together with weak convergence arguments and Hoeffding's (1960) lower bounds for the expected sample size of a sequential test, Lai (1988a) proved the following two results on the asymptotic optimality of the stopping rules (4.13) and (4.12) for the problems of testing \( H_0^1 : \theta < \theta_0 \) versus \( H_1^1 : \theta > \theta_0 \) and \( H_0 : \theta < \theta_0 \) versus \( H_1 : \theta \geq \theta_0 \) for the parameter \( \theta \) of the exponential family (2.1).

**THEOREM 1.** Let \( G \) be a prior distribution on \( \mathcal{A} \) such that \( G \) has a positive continuous density \( G' \) in some neighborhood of \( \theta_0(\in \mathcal{A}) \). For a sequential test \((T, \delta)\) of \( H_0^1 : \theta < \theta_0 \) versus \( H_1^1 : \theta > \theta_0 \) define the Bayes risk

\[
r(T, \delta) = c \int E_\theta TdG + \int_{\theta<\theta_0} P_{\theta}\{\delta \text{ accepts } H_1^1\}dG + \int_{\theta>\theta_0} P_{\theta}\{\delta \text{ accepts } H_0^1\}dG,
\]

with respect to the prior distribution \( G \), cost \( c \) per observation and the \( 0-1 \) loss function. Define \( T_c \) by (4.13) and let \( \delta^* \) denote the terminal decision rule that accepts \( H_0^1 \) or \( H_1^1 \) according as \( \hat{\theta}_{T_c} < \theta_0 \) or \( \hat{\theta}_{T_c} > \theta_0 \). Then as \( c \to 0 \),

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

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

where \( w(t), t \geq 0, \) denotes the Wiener process with drift coefficient \( \mu \) under \( P(\cdot|\mu) \) and \( \tau_0 = \inf\{t > 0 : |w(t)| \geq h_0(t)\} \).
THEOREM 2. Let $G$ be a prior distribution on $A$ and let $\tau(T, \delta)$ be the Bayes risk $(3.1)$ of a test of $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_1$, with cost $c$ per observation and loss $\ell(\theta)$ for the wrong decision such that $\int \ell(\theta) dG(\theta) < \infty$ and

$$\ell(\theta) \geq \Delta \text{ for all } \theta \notin (\theta_0, \theta_1) \text{ and some } \Delta > 0. \quad (4.18)$$

Let $g$ be a nonnegative function on $(0, \infty)$ such that $(4.16)$ holds for some $\xi > -1/2$. Define the stopping rule $N(g, c)$ by (4.12), and let $\delta^*$ be the terminal decision rule that accepts $H_0$ or $H_1$ according as $\hat{\theta}(g, c) \leq \theta^* \text{ or } \hat{\theta}(g, c) \geq \theta^*$, where $\theta^* \in (\theta_0, \theta_1)$ is such that $I(\theta^*, \theta_0) = I(\theta^*, \theta_1)$.

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

$$G([x, y]) \leq \rho(y - x) \text{ for all } x, y \in [\theta_0 - \epsilon, \theta_0] \cup [\theta_1, \theta_1 + \epsilon] \text{ with } x < y.$$ 

Then for fixed $\theta_0$ and $\theta_1$, as $c \to 0$,

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

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

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

$$r(N(g, c), \delta^*) \sim (8G'(\theta_0)/\psi''(\theta_0)c(\theta_1 - \theta_0)^{-1}\log[(\theta_1 - \theta_0)^2/c] \sim \inf_{(T, \delta)} r(T, \delta).$$

(iii) Suppose that $\ell(\theta) \to 1$ as $(\theta - \theta_0)I_{\{\theta < \theta_0\}} + (\theta - \theta_1)I_{\{\theta > \theta_1\}} \to 0$. Let $0 \leq \gamma < \infty$ and define $h_\gamma, \tau_\gamma$ as in (4.5). Let $g_\gamma(t) = (h_\gamma(t) + \gamma t)^2/2t$. Then $g_\gamma$ satisfies condition (4.16) with $\xi = 1/2$. Assume that $G$ has a positive continuous density $G'$ in some neighborhood of $\theta_0$. Then as $c \to 0$ and $\theta_1 \to \theta_0$ such that $(\theta_1 - \theta_0)/2c^{1/2} \to \gamma$,

$$\inf_{(T, \delta)} r(T, \delta) \sim r(N(g_\gamma, c), \delta^*) \sim \frac{c^{1/2}G'(\theta_0)}{\psi''(\theta_0)} \left\{ \int_{-\infty}^{\infty} E(\tau_\gamma|\mu)d\mu + \int_{-\infty}^{-\gamma} P[w(\tau_\gamma) > 0|\mu]d\mu + \int_{\gamma}^{\infty} P[w(\tau_\gamma) < 0|\mu]d\mu \right\}.$$

The simulation studies of the risk functions of the tests in Theorem 1 and 2 reported in Lai (1988a) show that the stopping rules (4.12) and (4.13) not only provide approximate Bayes solutions with respect to a large class of priors but also have nearly optimal frequentist properties. Lai (1988c) extends Theorem 1 dealing with the case of no indiﬁerence zone from the $0 - 1$ loss to general loss functions $\ell(\theta)$ such that

$$\ell(\theta) \sim \lambda|\theta - \theta_0|^p \text{ as } \theta \to \theta_0, \text{ for some } \lambda > 0 \text{ and } p > -1. \quad (4.19)$$

The $0 - 1$ loss is a special case of (4.19) with $p = 0$, while the loss function $|\theta - \theta_0|$ in Chernoff's theory reviewed in Section 3 is a special case of (4.19) with $p = 1$. The stopping rule $T_c$ in (4.13) for the $0 - 1$ loss case is now generalized to the form

$$T_c^* = \inf\{n \geq 1 : I(\hat{\theta}_n, \theta_0) \geq n^{-1} B(c^{2/(p+2)}n)\} \quad (4.20)$$
for the loss function (4.19), cf. Lai (1988c). In view of (4.14), the tests \((T_\varepsilon, \delta^*)\) and \((T'_\varepsilon, \delta^*)\) can be regarded as generalized sequential likelihood ratio tests of \(H'_0: \theta < \theta_0\) versus \(H'_1: \theta > \theta_0\), while the test \((N(g, c), \delta^*)\) and Schwarz's test \((N_\varepsilon, \delta^*)\) are generalized sequential likelihood ratio tests of \(H_0: \theta \leq \theta_0\) versus \(H_1: \theta \geq \theta_0\). Our approximations to the Bayes rules use time-varying stopping boundaries of the form \(g(cn)\) or \(B(c^{2/(r+3)n})\) for the generalized log likelihood ratios instead of constant boundaries (such as \(|\log c|\) used by Schwarz).

Making use of Lemma 2(ii) together with Hoeffding's (1960) lower bounds for the expected sample size of a sequential test, Lai (1988a) also established the following asymptotically optimal frequentist properties of the generalized sequential likelihood ratio test \((N(g, c), \delta^*)\).

**THEOREM 3.** Let \(g\) be a nonnegative function on \((0, \infty)\) such that (4.16) holds for some \(\xi\), and define the stopping rule \(N(g, c)\) by (4.12). Let \(\alpha = P_{\theta_0}\{\hat{\theta}_{N(G, c)} > \theta^*\}\), \(\beta = P_{\theta_1}\{\hat{\theta}_{N(G, c)} \leq \theta^*\}\), where \(\theta^* \in (\theta_0, \theta_1)\) is such that \(I(\theta^*, \theta_0) = I(\theta^*, \theta_1)\). Let \(T(\alpha, \beta)\) denote the class of all sequential tests of \(H_0: \theta \leq \theta_0\) versus \(H_1: \theta \geq \theta_1\) that satisfy the error constraints (1.4).

(i) For fixed \(\theta_0\) and \(\theta_1\), as \(c \to 0\), \(\log \alpha \sim \log \beta \sim \log c\), and for every bounded subset \(B\) of \(A\),

\[
E_\theta N(g, c) \sim |\log c|/J(\theta) \sim \inf_{(T, \delta) \in T(\alpha, \beta)} E_\theta T \text{ uniformly in } \theta \in B. \tag{4.21}
\]

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

\[
\log \alpha \sim \log \beta \sim \log(c/d^2), \text{ where } d = \theta_1 - \theta_0, \tag{4.22}
\]

\[
\sup_\delta E_\theta N(g, c) \sim 8d^{-2}(\log c^{-1}d^2)/\psi'(\theta_0) \sim \inf_{(T, \delta) \in T(\alpha, \beta)} \sup_\theta E_\theta T. \tag{4.23}
\]

Moreover, for every distribution function \(G\) on \(A\) having a positive continuous derivative \(G'\) in some neighborhood of \(\theta_0\),

\[
\int E_\theta N(g, c)dG(\theta) \sim (8G'(\theta_0)/\psi''(\theta_0))d^{-1}\log(d^2/c) \sim \inf_{(T, \delta) \in T(\alpha, \beta)} \int E_\theta TdG(\theta). \tag{4.24}
\]

Theorem 3 reveals an interesting connection between the generalized sequential likelihood ratio test \((N(g, c), \delta^*)\) and the 2-SPRTs discussed in Section 2. As shown in Section 2, the 2-SPRT that stops sampling at stage

\[
M_\theta = \inf \left\{ n: \sum_{i=1}^{n} \log[f_\theta(X_i)/f_{\theta_0}(X_i)] \geq b \text{ or } \sum_{i=1}^{n} \log[f_\theta(X_i)/f_{\theta_1}(X_i)] \geq b \right\} \tag{4.25}
\]

is asymptotically optimal as \(b \to \infty\), in the sense of (2.8) for the problem of minimizing the expected sample size at a given \(\theta\) among all sequential tests with the same (or smaller) error probabilities at \(\theta_0\) and \(\theta_1\). Since the true value of \(\theta\) is typically unknown, it is natural to try replacing \(\theta\) in (4.25) by its maximum likelihood estimator \(\hat{\theta}_n\). The accuracy of \(\hat{\theta}_n\) as an estimate of \(\theta\) varies with \(n\), and the stopping rule \(N(g, c)\) takes this into account by adjusting the constant boundary \(b\) in (4.25) with the simple time-varying boundary \(g(cn)\), cf. (4.12) and (4.14). Thus, the generalized sequential likelihood ratio test \((N(g, c), \delta^*)\) can be viewed as an adaptive 2-SPRT, with the value of \(\theta\) in (4.25) being chosen adaptively and with a corresponding adjustment of the
stopping boundary $b$. Theorem 3(i) shows that this idea still leads to a first-order asymptotically optimal solution in ignorance of $\theta$ for fixed $\theta_0$ and $\theta_1$, although the conclusion (4.21) is weaker than the higher-order asymptotically optimum character (2.8) for the 2-SPRT that assumes knowledge of $\theta$. Moreover, even as $\theta_1 \to \theta_0$, Theorem 3(ii) shows that the test $\{(N(g,c),\delta^*)\}$ 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 (1.4).

5. ASYMPTOTIC OPTIMALITY OF AN INVARIANT SPRT AND OF SPRT’S BASED ON DEPENDENT OBSERVATIONS

For the SPRT (1.1) with stopping bounds $A > 1 > B > 0$, Wald (1945) showed that the error probabilities $\alpha = P_0 \{R_N \geq A\}$ and $\beta = P_1 \{R_N \leq B\}$ are related to $A$ and $B$ by the inequalities

$$\alpha \leq A^{-1}(1 - \beta), \quad \beta \leq B(1 - \alpha). \tag{5.1}$$

and that equalities would hold in (5.1) if there is no overshoot. Ignoring overshoots, Wald used (5.1) to give approximate determinations of $A$ and $B$. These arguments and the bounds (5.1) do not depend on the fact that the likelihood ratio $R_n$ is of the special form (1.2) which assumes the successive observations to be i.i.d. Consequently, most of the early works on sequential tests of composite hypotheses are concerned with the case where the composite hypotheses $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$ can be reduced to simple ones by the principle of invariance so that invariant SPRTs can be applied. If $G$ is a group of transformations leaving the problem invariant, then the distribution of a maximal invariant depends on $P$ only through its orbit. Therefore, if $\mathcal{P}_0$ and $\mathcal{P}_1$ form two distinct orbits and only invariant sequential tests are considered, then the hypotheses become simple (cf. Chapter 4 of Ghosh, 1970). Hence in analogy with Wald’s SPRT, we again stop with $N$ given by (1.1) but with $R_n$ now defined by

$$R_n = p_{1n}(T_n)/p_{0n}(T_n), \tag{5.2}$$

where the random vector $T_n$ is a maximal invariant with respect to $G$ based on the first $n$ observations and $p_{in}$ is the density of this maximal invariant under $H_i$ ($i = 0, 1$). Classical examples of these tests, called invariant SPRTs, are the sequential $t$-test, the sequential $T_2$-test, the sequential $F$-test, and the Savage–Sethuraman sequential rank test.

For these invariant SPRTs, $\{\log R_n, n \geq 1\}$ is no longer a random walk and the arguments in the proof of the Wald–Wolfowitz theorem on the optimum character of Wald’s SPRT cannot be extended to them. The corresponding optimal stopping problems associated with invariant sequential tests of $H_0 : P \in \mathcal{P}_0$ versus $H_1 : P \in \mathcal{P}_1$ are typically quite intractable. To study the long standing problem concerning what kind of optimum properties, if any, these invariant SPRTs may have (cf. Ghosh, 1970), Lai (1981) developed asymptotic extensions of Wald’s (1945) lower bounds on the expected sample size of a sequential test based on i.i.d. observations subject to the error constraints (1.3). Wald’s lower bounds are attained by his approximations for the expected sample sizes (and error probabilities) of the SPRT, ignoring overshoots (cf. Wald, 1945, page 157). This suggests that a similar approach may establish the asymptotic optimality of invariant SPRTs.
However, although Wald's bounds (5.1) still hold for invariant SPRTs, his approximations for the expected sample size of an SPRT make use of the random walk structure of \( \{\log R_n\} \) and do not generalize to the expected sample size of an invariant SPRT. To analyze the expected sample sizes, Lai (1975, 1981) first made use of a representation of \( \log R_n \) as a random walk plus a remainder term that satisfies an \( r \)-quick strong law with \( r \geq 1 \), and later also applied nonlinear renewal theory to develop asymptotic expansions of the expected sample sizes up to the \( o(1) \) term. (A sequence of random variables \( Y_n \) is said to converge to 0 \( r \)-quickly if \( EL^r \epsilon < \infty \) for every \( \epsilon > 0 \), where \( L_\epsilon = \sup \{ n \geq 1 : |Y_n| \geq \epsilon \} \), cf. Lai, 1976, Section 6. This strengthens the notion of a.s. convergence of \( Y_n \) to 0, which can be stated as \( P\{L_\epsilon < \infty \} = 1 \) for every \( \epsilon > 0 \).) Making use of this representation of \( \log R_n \) and a nonlinear renewal theorem of Lai and Siegmund (1979), Lai (1981) obtained the following asymptotic generalization of Wald's lower bounds.

**THEOREM 4.** Let \( Z_1, Z_2, \ldots \) be a sequence of random variables on a \( \sigma \)-finite measure space \((\Omega, \mathcal{F}, Q)\). Let \( P_i, i = 0, 1 \), be two probability measures on \((\Omega, \mathcal{F})\) such that under \( P_i \), \( (Z_1, \ldots, Z_n) \) has joint density \( p_{in} \) with respect to the restriction of \( Q \) to the \( \sigma \)-field \( \mathcal{F}_n \) generated by \( Z_1, \ldots, Z_n \). Define the likelihood ratio

\[
R_n = p_{1n}(Z_1, \ldots, Z_n)/p_{0n}(Z_1, \ldots, Z_n). \tag{5.3}
\]

Let \( T(\alpha, \beta) \) denote the class of all sequential tests \((T, \delta)\) of \( H : P = P_0 \) versus \( K : P = P_1 \) based on the sequence \( \{Z_n\} \) satisfying the error constraints (1.3). Let \( \mathcal{G}_1 \subset \mathcal{G}_2 \subset \ldots \) be a sequence of sub-\( \sigma \)-fields of \( \mathcal{F} \) such that \( \mathcal{F}_n \subset \mathcal{G}_n \) for every \( n \). Suppose that under \( P_i(i = 0, 1) \), \( \log R_n \) has a representation of the form

\[
\log R_n = \sum_{j=1}^{n} Y_j + \xi_n, \tag{5.4}
\]

where \( Y_1, Y_2, \ldots \) are i.i.d. with \( E_i Y^2_1 < \infty \), \( Y_n \) is \( \mathcal{G}_n \)-measurable and \( Y_{n+1} \) is independent of \( \mathcal{G}_n \) for every \( n \geq 1 \), and

\[
E_0 Y_1 = \lambda_0 < 0, \quad E_1 Y_1 = \lambda_1 > 0, \tag{5.5}
\]

\( \xi_n \) converges in distribution to some random variable \( \xi \). \tag{5.6}

(The random variables \( Y_n, \xi_n \), and \( \xi \) may depend on \( i \in \{0, 1\} \).) Assume for \( i = 0, 1 \) that there exist positive constants \( \delta \leq 1, \Delta_n, \rho \) and events \( A_n \) (which may also depend on \( i \)) such that

\[
P_i\left( \bigcup_{n \leq k \leq n+n^\rho \delta} A_k \right) = o(n^{-1}) \quad (\bar{A} = \text{complement of } A), \tag{5.7}
\]

\[
\{\xi_n | I(\bigcap_{n \leq k \leq n+n^\rho \delta} A_k), n \geq n_0 \} \text{ is uniformly integrable under } P_i \text{ for some } n_0, \tag{5.8}
\]

\[
\lim_{n \to \infty} \Delta_n = 0 \quad \text{and} \quad P_i\left\{ \max_{n \leq k \leq n+n^\rho \delta} |\xi_k - \xi_n| > \Delta_n \right\} = o(n^{-1}), \tag{5.9}
\]

\[
P_i\{\max_{j \leq n} (\log R_j) > \lambda_1 n + cn^\delta\} + P_i\{\log R_n < \lambda_1 n - cn^\delta\} = o(n^{-1}), \tag{5.10}
\]

\[
P_0\{\min_{j \leq n} (\log R_j) < \lambda_0 n - cn^\delta\} + P_0\{\log R_n > \lambda_0 n + cn^\delta\} = o(n^{-1})
\]
for all $c > 0$. Then as $\alpha + \beta \to 0$ such that $\alpha \log \beta + \beta \log \alpha \to 0$,

$$
\inf_{(T, \delta) \in T(\alpha, \beta)} E_0 T \geq |\lambda_0|^{-1} \{ |\log \beta| + E_0 \xi \} + o(1),
$$

(5.11)

$$
\inf_{(T, \delta) \in T(\alpha, \beta)} E_1 T \geq \lambda_1^{-1} \{ |\log \alpha| - E_1 \xi \} + o(1).
$$

(5.12)

**REMARK.** When $Z_1, Z_2, \ldots$ are i.i.d., Wald (1945) obtained the lower bounds

$$
|\lambda_0| E_0 T \geq (1 - \alpha) \log((1 - \alpha)/\beta) + \alpha \log(\alpha/(1 - \beta)), \\
\lambda_1 E_1 T \geq (1 - \beta) \log((1 - \beta)/\alpha) + \beta \log(\beta/(1 - \alpha)),
$$

(5.13)

for the expected sample size $E_i T$ of an arbitrary test $(T, \delta) \in T(\alpha, \beta)$ with $E_i T < \infty$, where $\lambda_i = E_i \{ \log[p_1(Z_i)/p_0(Z_i)] \}$. Ignoring overshoots, Wald’s SPRT with boundaries $A, B$ given by equalities in (5.1) attain the lower bounds in (5.13). Moreover, as $\alpha + \beta \to 0$ such that $\alpha \log \beta + \beta \log \alpha \to 0$, these lower bounds reduce to (5.11) and (5.12) with $\xi = 0$. In fact, in this i.i.d. setting, $\xi_n = 0$ and (5.6)–(5.9) are obviously satisfied (with $A_n = \Omega, Y_n = \log[p_1(Z_n)/p_0(Z_n)]$) and (5.10) holds with $1/2 < \delta \leq 1$ if $E_i |Y_i|^2/\delta < \infty$.

By showing that the sequential $t$–test, sequential $T^2$–test and sequential $F$–test attain the asymptotic lower bounds given in Theorem 4, Lai (1981) established the asymptotic optimality of these invariant SPRTs whose stopping rules are of the form (1.1) but with $R_n$ given by (5.2), in the sense that

$$
E_i N - \inf_{(T, \delta) \in T(\alpha, \beta)} E_i T = O(1) \text{ for } i = 0, 1,
$$

(5.14)

as $\alpha + \beta \to 0$ such that $\alpha \log \beta + \beta \log \alpha \to 0$. The $O(1)$ term in (5.14) is due to the overshoot $\log(R_N/B)$ or $\log(R_N/A)$ in the invariant SPRT. Note in this connection that Wald’s lower bounds (5.13) are not attained by the SPRT with error probabilities $\alpha, \beta$ when overshoots are present, so the optimum character of the SPRT needs deeper arguments involving optimal stopping theory for its proof. Replacing the assumption (5.4) on $\log R_n$ by the simple stability assumption

$$
n^{-1} \log R_n \to \lambda_i \text{ r-quickly under } P_i \text{ (i = 0, 1)},
$$

(5.15)

Lai (1981) also established the following first–order asymptotically optimum property of SPRTs based on dependent observations $Z_1, Z_2, \ldots$:

$$
E_i N^r / \inf_{(T, \delta) \in T(\alpha, \beta)} E_i T^r \to 1 \text{ for } i = 0, 1,
$$

(5.16)

as $\alpha + \beta \to 0$, and applied this result to the Savage–Sethuraman (1966) two–sample rank–order SPRT for Lehmann alternatives and to SPRTs of hypotheses concerning the mean level of an autoregressive Gaussian sequence.

**REFERENCES**


