The Annals of Statistics
1981, Vol. 9, No. 2, 318–333

ASYMPTOTIC OPTIMALITY OF INVARIANT SEQUENTIAL PROBABILITY RATIO TESTS[1]

BY TZE LEUNG LAI

Columbia University

It is well known that Wald's SPRT for testing simple hypotheses based on i.i.d. observations minimizes the expected sample size both under the null and under the alternative hypotheses among all tests with the same or smaller error probabilities and with finite expected sample sizes under the two hypotheses. In this paper it is shown that this optimum property can be extended, at least asymptotically as the error probabilities tend to 0, to invariant SPRTs like the sequential t-test, the Savage-Sethuraman sequential rank-order test, etc. In fact, not only do these invariant SPRTs asymptotically minimize the expected sample size, but they also asymptotically minimize all the moments of the sample size distribution among all invariant tests with the same or smaller error probabilities. Modifications of these invariant SPRTs to asymptotically minimize the moments of the sample size at an intermediate parameter are also considered.

1. Introduction. Let $Z_1, Z_2, \dots$ be i.i.d. random variables with a common distribution P. Wald's sequential probability ratio test (SPRT) of testing the null hypothesis $H: P = P_0$ versus $K: P = P_1$ stops sampling at stage

$$\tau = \inf\{n \geq 1 : R_n \geq A \quad \text{or} \quad R_n \leq B\}, \qquad (\inf \phi = \infty), \tag{1.1}$$

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

$$R_n = \textstyle\prod_{j=1}^{n} \{p_1(Z_j)/p_0(Z_j)\} \tag{1.2}$$

is the likelihood ratio, p_i being the density of P_i with respect to some common dominating measure $Q (i \neq 0, 1)$. When stopping occurs, H or K is accepted according as $R_\tau \leq B$ or $R_\tau \geq A$. The choice of the stopping bounds is dictated by the error probabilities $\alpha = P_0[R_\tau \geq A]$ and $\beta = P_1[R_\tau \leq B]$. Wald [14] has shown that α and β are related to A and B by the following inequalities:

$$\alpha \leq A^{-1}(1 - \beta), \qquad \beta \leq B(1 - \alpha), \tag{1.3}$$

and that equalities would hold in (1.3) if there is no overshoot. Ignoring overshoots, (1.3) gives approximate determinations of A and B in terms of the error probabilities, and the ease of defining the test given the error probabilities is one of the attractive properties of Wald's SPRT.

The above idea of Wald has been generalized in the literature to obtain sequential tests of composite hypotheses $H_0: P \in \mathscr{P}_0$ versus $H_1: P \in \mathscr{P}_1$ when these composite hypotheses can be reduced to simple ones by the principle of invariance. If G is a group leaving the problem invariant, then the distribution of a maximal invariant depends on P only through its orbit. Therefore, if $\mathscr{P}_0$ and $\mathscr{P}_1$ form two distinct orbits and only invariant sequential tests are considered, then the hypotheses become simple (cf. Chapter 6 of [4]). Hence in analogy with Wald's SPRT, we can again stop at stage τ given by (1.1) but with R_n now defined by

$$R_n = p_{1n}(\mathbf{T}_n)/p_{0n}(\mathbf{T}_n), \tag{1.4}$$

Received November 1978; revised June 1979.

[1] Research supported by the National Science Foundation under grant NSF-MCS-78-09179.

AMS 1970 *subject classifications.* Primary 62L10; secondary 62F05, 62M10, 60F15.

Key words and phrases. Invariant SPRT, Wald-Wolfowitz theorem, asymptotic optimality, r-quick convergence, Wald's lower bounds for the expected sample size.

where the random vector $\mathbf{T}_n$ 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 $H_i (i = 0, 1)$. This extension of Wald's SPRT still preserves the simple inequalities (1.3) (which are approximate equalities ignoring overshoots), and therefore the stopping bounds A and B in (1.1) can again be approximately determined in terms of the error probabilities when R_n in (1.1) is the likelihood ratio of the maximal invariant as defined by (1.4) (cf. [4], page 89) Classical examples of these tests, called invariant sequential probability ratio tests (cf. [4]), are the sequential t-test, the sequential T^2-test, the sequential F-test, and the Savage-Sethuraman sequential rank-order test. These examples will be considered in Sections 2 and 3.

A remarkable property of Wald's SPRT for testing H versus K with error probabilities α and β is that it minimizes both E_0N and E_1N (N being the sample size) among all tests (sequential or not) for which

$$P_0[\text{rejecting } H] \leq \alpha, \qquad P_1\,[\text{rejecting } K] \leq \beta, \tag{1.5}$$

and for which E_0N and E_1N are both finite. This optimum property of Wald's SPRT, first established by Wald and Wolfowitz [15], is a justification of its use at least for testing simple hypotheses. The proof of this optimum property depends heavily on the fact that since $Z_1, Z_2, \ldots$ are i.i.d., $\log R_n$ (where R_n is defined by (1.2)) is a sum of i.i.d. random variables. The argument breaks down when R_n is the likelihood ratio of a maximal invariant as defined by (1.4), in which case $\{\log R_n\}$ is no longer a random walk. Therefore, the Wald-Wolfowitz theorem on the optimum property of Wald's SPRT does not extend to cover invariant SPRTs, and it has been an open problem as to what kind of optimum properties, if any, these invariant SPRTs have (cf. [4], pages 146–147 and 292, and [17]). Unlike Wald's SPRT which simply involves the first exit time of a random walk from a finite interval, the stopping time τ of an invariant SPRT is very difficult to analyze. In fact a major part of the literature on invariant SPRTs has been concerned with the problem of termination with probability 1 and with the stronger property of exponential boundedness of τ (see [17] for a survey of the present status of the subject). Only recently have 110asymptotic approximations to the moments of τ been obtained (cf. [3], [6], [7], [8], [10], asymptotic approximations to the moments of τ been obtained (cf. [3], [6], [7], [8], [10], [12]).

In Section 2 below we shall obtain a first-order asymptotic analogue of the Wald-Wolfowitz theorem for a large class of invariant SPRTs including all the classical examples as special cases. More specifically, we shall show that such an invariant SPRT for testing H_0 versus H_1 with error probabilities α and β asymptotically (a $\alpha + \beta \to 0$) minimizes all the moments of the sample size distribution (both under H_0 and under H_1) for invariant tests of H_0 versus H_1 with error probabilities not exceeding α and β. Our asymptotic analogue of the Wald-Wolfowitz theorem in fact extends well beyond invariant SPRTs. As an illustration of its wide applicability, we shall apply it in Section 2 to study the SPRT of $H: \theta = \theta_0$ versus $K: \theta = \theta_1$ when the observations $Z_1, Z_2, \ldots$ form a finite-order autoregressive stationary Gaussian sequence with unknown mean level θ.

Let $Z_1, Z_2, \ldots$ be i.i.d. $N(\theta, 1)$ random variables. As is well known, although Wald's SPRT for testing $H: \theta = \theta_0$ versus $K: \theta = \theta_1$ has optimum expected sample size at these parameter values, its expected sample size becomes substantially larger when θ is around the midpoint $\frac{1}{2}\,(\theta_0 + \theta_1)$, and there exists a simple modification of the SPRT, due to Anderson [1], which has a much smaller expected sample size than that of the SPRT at $\frac{1}{2}$ $(\theta_0 + \theta_1)$. We shall show in Section 3 that a similar phenomenon also holds for the sequential t-test, the sequential F-test, the sequential T^2-test, and the SPRT for the mean level of an autoregressive Gaussian sequence. Moreover, we shall examine the more general problem of asymptotically minimizing the moments of the sample size at an intermediate parameter without the i.i.d. structure.

In Section 4, we shall study the problem of higher-order asymptotic optimality; and it will be shown that ignoring overshoots, invariant SPRTs such as the sequential t-test in fact asymptotically minimize the expected sample size up to the $o(1)$ term (both under H_0 and under H_1) among all invariant tests with the same or smaller error probabilities. This result is obtained by developing full asymptotic expansions of lower bounds for the expected sample size of invariant sequential tests and extending the classical lower bounds of Wald [14] from the i.i.d. case to a much more general setting.

2. An asymptotic analogue of the Wald-Wolfowitz theorem with applications to invariant SPRTs. Throughout the sequel we shall use the following notation. Let $X_1, X_2, \ldots$ be a sequence of random variables defined on the same underlying measurable space $(\Omega, \mathscr{F})$. Let $\mathscr{F}_n$ be the sub-σ-field of $\mathscr{F}$ generated by $X_1, \ldots, X_n$. Let Q be a σ-finite measure on $(\Omega, \mathscr{F})$ and let Q_n be the restriction of Q to $\mathscr{F}_n$. Let P_i, $i = 0, 1$, be two probability measures on $(\Omega, \mathscr{F})$ such that under P_i, $(X_1, \ldots, X_n)$ has a joint density $p_{in}(x_1, \ldots, x_n)$ with respect to Q_n for every $n \geq 1$. A (nonrandomized) test (either sequential or fixed sample size) of $H_0: P = P_0$ versus $H_1: P = P_1$ based on the sequence $\{X_n\}$ can be characterized by

(i) a stopping rule N relative to $\{\mathscr{F}_n\}$, i.e., a positive extended integer valued random variable N such that $\{N = n\} \in \mathscr{F}_n$ for all n, and

(ii) a terminal decision rule $d(= d(X_1, \ldots, X_N))$ which accepts either H_0 or H_1 upon stopping.

The test will be denoted by (N, d). We shall consider the class $\mathscr{T}(\alpha, \beta)$ of tests (N, d) which satisfy prescribed bounds $0 < \alpha, \beta < ½$ on the error probabilities, i.e.,

$$P_0[(N, d) \text{ rejects } H_0] \leq \alpha, \qquad P_1[(N, d) \text{ rejects } H_1] \leq \beta. \tag{2.1}$$

We shall let $S(A, B)$ denote the SPRT of H_0 versus H_1 based on the sequence $\{X_n\}$ and having stopping bounds $A > B > 0$, i.e., $S(A, B)$ stops sampling at stage τ given by (1.1) but with R_n now defined by

$$R_n = p_{1n}(X_1, \ldots, X_n)/p_{0n}(X_1, \ldots, X_n), \tag{2.2}$$

and accepts H_0 or H_1 when stopping occurs according as $R_\tau \leq B$ or $R_\tau \geq A$.

If α and β are the error probabilities of $S(A, B)$, then the inequalities in (1.3) hold (cf. [4], page 89). Regarding (1.3) as approximate equalities by neglecting overshoots leads to the approximations

$$A \approx (1 - \beta)/\alpha, \qquad B \approx \beta/(1 - \alpha). \tag{2.3}$$

We note that the approximate solutions for A and B in (2.3) are asymptotic to α^{-1} and β respectively as $\alpha + \beta \to 0$. Moreover, (1.3) implies that $S(\alpha^{-1}, \beta)$ satisfies the error constraints (2.1).

In the special case where $X_1, X_2, \ldots$ are i.i.d., $\{\log R_n\}$ is a random walk and the Wald-Wolfowitz theorem on the optimality of the SPRT $S(A, B)$ in this i.i.d. case depends heavily on the random walk structure. As $\{\log R_n\}$ is no longer a random walk when $X_1, X_2, \ldots$ are not i.i.d., our method is to replace the random walk structure by the property that $\{\log R_n\}$ is asymptotically stable in the following sense. First note that when $X_1, X_2, \ldots$ are i.i.d., by the strong law of large numbers,

$$n^{-1} \log R_n \to \lambda_i \text{ a.s. } [P_i], \tag{2.4}$$

where $\lambda_i = E_i\{\log(p_1(X_i)/p_0(X_i))\}$ so that $\lambda_0 < 0$ and $\lambda_1 > 0$ if $P_i[p_0(X_1) \neq p_1(X_i)] > 0$ for $i = 0, 1$. It turns out that for the applications which we shall consider, this stability property still holds (with constants $\lambda_0 < 0$ and $\lambda_1 > 0$ defined differently) although $X_1, X_2, \ldots$ are no longer i.i.d. The stability of $\{\log R_n\}$ is sufficient to imply the asymptotic optimality of the SPRT in the sense of the following theorem.

THEOREM 1. *Defining R_n as the likelihood ratio in* (2.2), *assume that there exist finite constants $\lambda_0 < 0$ and $\lambda_1 > 0$ such that* (2.4) *holds for $i = 0, 1$.*

(*i*) *For $0 < \alpha, \beta < 1$, let $\mathscr{T}(\alpha, \beta)$ be the class of all* (*nonrandomized*) *tests* (*sequential or fixed sample size*) *of $H_0: P = P_0$ versus $H_1: P = P_1$ based on the sequence $\{X_n\}$ and satisfying* (2.1). *Then for every $0 < \delta < 1$, as $\alpha + \beta \to 0$,*

$$\inf_{(N,d) \in \mathscr{T}(\alpha,\beta)} P_0[N > \delta |\log \beta| / |\lambda_0|] \to 1, \tag{2.5a}$$

and

$$\inf_{(N,d) \in \mathscr{T}(\alpha,\beta)} P_1[N > \delta |\log \alpha| / \lambda_1] \to 1. \tag{2.5b}$$

(ii) *For $0 < \alpha, \beta < 1$, let $A_{\alpha,\beta} > B_{\alpha,\beta} > 0$ be so chosen that* (2.1) *holds for $S(A_{\alpha,\beta}, B_{\alpha,\beta})$ and*

$$\log A_{\alpha,\beta} \sim \log \alpha^{-1}, \qquad \log B_{\alpha,\beta} \sim \log \beta \qquad \textit{as} \quad \alpha + \beta \to 0. \tag{2.6a}$$

Let $\tau_{\alpha,\beta}$ denote the stopping rule of $S(A_{\alpha,\beta}, B_{\alpha,\beta})$. Then as $\alpha + \beta \to 0$,

$$\tau_{\alpha,\beta} / |\log \beta| \to |\lambda_0|^{-1} \text{ a.s. } [P_0], \tag{2.6b}$$

and

$$\tau_{\alpha,\beta} / |\log \alpha| \to \lambda_1^{-1} \text{ a.s. } [P_1]. \tag{2.6c}$$

Consequently, for every $0 < \delta < 1$, as $\alpha + \beta \to 0$,

$$\inf_{(N,d) \in \mathscr{T}(\alpha,\beta)} P_i[N > \delta \tau_{\alpha,\beta}] \to 1 \qquad \textit{for} \quad i = 0, 1. \tag{2.7}$$

Since the optimality criterion in the Wald-Wolfowitz theorem is about the minimization of the expected sample size both under H_0 and under H_1, it is natural to ask whether the asymptotic optimality of $S(A_{\alpha,\beta}, B_{\alpha,\beta})$ in the sense of (2.7) above implies the asymptotic minimization of moments of the sample size. While the almost sure convergence in the stability property (2.4) does not even guarantee the finiteness of moments of $\tau_{\alpha,\beta}$, it turns out that in our applications the almost sure convergence in (2.4) can in fact be strengthened into the notion of r-quick convergence which provides a useful tool in our argument. For $r > 0$, a sequence $\{Y_n\}$ of random variables is said to *converge r-quickly* to a constant λ if $E(L_a)^r < \infty$ for all $a > 0$, where $L_a = \sup\{n \geq 1 : |Y_n - \lambda| \geq a\}$ ($\sup \phi = 0$) (cf. [9]). Note that $Y_n \to \lambda$ a.s. iff $P[L_a < \infty] = 1$ for all $a > 0$. An r-quick theory for random walks and for various statistics has been developd in [6], [7], and [9]. In particular, if $X_1, X_2, \ldots$ are i.i.d. and $\lambda_i = E_i\{\log(p_1(X_1)/p_0(X_1))\}$ is finite, then $E_i|\log(p_1(X_1)/p_0(X_1))|^{r+1} < \infty$ is both necessary and sufficient for the r-quick convergence of $n^{-1} \log R_n$ to λ_i under P_i (cf. [9]). Under this stronger mode of convergence of $n^{-1} \log R_n$, we obtain from Theorem 1 the following

COROLLARY 1. *With the same notations and assumptions as in Theorem 1, suppose that* (2.4) *is strengthened into*

$$n^{-1} \log R_n \to \lambda_i \qquad \textit{r-quickly under} \quad P_i \qquad i = 0, 1 \tag{2.8}$$

for some positive constant r. Then for all $0 < B < A$, the sample size of $S(A, B)$ has a finite rth moment under P_0 and under P_1. Moreover, as $\alpha + \beta \to 0$,

$$\begin{aligned} &\inf_{(N,d) \in \mathscr{T}(\alpha,\beta)} E_0 N^r \sim |\lambda_0|^{-r} |\log \beta|^r \sim E_0 \tau^r_{\alpha,\beta}, \qquad \textit{and} \\ &\inf_{(N,d) \in \mathscr{T}(\alpha,\beta)} E_1 N^r \sim \lambda_1^{-r} |\log \alpha|^r \sim E_1 \tau^r_{\alpha,\beta}. \end{aligned} \tag{2.9}$$

The proof of Theorem 1 and Corollary 1 will be given in Section 3 as a special case of a more general result. In the rest of this section, we shall apply Corollary 1 to obtain the asymptotic analogue of the Wald-Wolfowitz theorem for various invariant SPRTs. Throughout the sequel we shall use the notation $\bar{a}_n$ for the arithmetic mean of n numbers $a_1, \ldots, a_n$.

EXAMPLE 1. (*Sequential t-test*). Let $Z, Z_1, Z_2, \ldots$ be i.i.d. normal $N(\mu, \sigma^2)$ random variables. The two hypotheses here are $H_i: \mu/\sigma = \gamma_i\,(i = 0, 1)$, where γ_0, γ_1 are given distinct numbers. A maximal invariant of $(Z_1, \ldots, Z_n)$ with respect to the group of scale changes $Z_j \to cZ_j\,(c > 0)$ is $(X_1, \ldots, X_n)$, where $X_j = Z_j/|Z_1|$. Hence any invariant test of H_0 versus H_1 is based on the sequence $\{X_n\}$ (cf. [4], page 301). The likelihood ratio of $(X_1, \ldots, X_n)$ is of the form

$$R_n = U_n(\gamma_1)/U_n(\gamma_0), \qquad \text{where}$$

$$U_n(\gamma) = \int_0^\infty u^{-1} \exp[n\, f(u, T_n, \gamma)]\, du, \; T_n = \bar{Z}_n/\{n^{-1} \textstyle\sum_1^n Z_j^2\}^{1/2}, \tag{2.10}$$

$$f(u, y, \gamma) = -\tfrac{1}{2} u^2 + \gamma y u + \log u - \tfrac{1}{2} \gamma^2$$

(cf. [16], page 1866). Define

$$h(\gamma, u) = -\tfrac{1}{2}\gamma^2 + \tfrac{1}{4}\{\gamma^2 u^2 + \gamma u(\gamma^2 u^2 + 4)^{1/2}\} + \log\{\gamma u + (\gamma^2 u^2 + 4)^{1/2}\}, \tag{2.11}$$

$$\Psi(u) = h(\gamma_1, u) - h(\gamma_0, u).$$

As shown in [6], page 587, there exists a constant C for which

$$|\log R_n - n\Psi(T_n)| \le C, \qquad n = 1, 2, \ldots. \tag{2.12}$$

For $i = 0, 1$, since $E_i|Z|^s < \infty$ for all $s > 0$, it follows that under H_i, $T_n \to (E_i Z)/(E_i Z^2)^{1/2} = \gamma_i/(1 + \gamma_i^2)^{1/2}$ r-quickly for all $r > 0$ (cf. [9]), and therefore by (2.12),

$$n^{-1} \log R_n \to \Psi(\gamma_i/(1 + \gamma_i^2)^{1/2}) = \lambda_i \quad r\text{-quickly} \tag{2.13}$$

for all $r > 0$. For fixed $-1 < u < 1$, the function $h_u(\gamma) = h(\gamma, u)$ (as defined in (2.4)) has its maximum at $\gamma = u/(1 - u^2)^{1/2}$. This implies that $\Psi(\gamma_0/(1 + \gamma_0^2)^{1/2}) < 0$ and $\Psi(\gamma_1/(1 + \gamma_1^2)^{1/2}) > 0$, i.e., $\lambda_0 < 0$ and $\lambda_1 > 0$. Hence the conditions of Corollary 1 are satisfied; and by Corollary 1, the sequential t-test $S(A_{\alpha,\beta}, B_{\alpha,\beta})$ is asymptotically (as $\alpha + \beta \to 0$) optimal within the class $\mathscr{T}(\alpha, \beta)$ of invariant sequential tests in the sense that (2.9) holds for all $r > 0$.

EXAMPLE 2. (*Sequential F-test*). Let $\mathbf{Z}, \mathbf{Z}_1\, \mathbf{Z}_2, \ldots$ be i.i.d. k-dimensional random vectors such that $\mathbf{Z}$ has the multivariate normal $\mathbf{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ distribution. Let μ_l and $Z_{nl}(l = 1, \ldots, k)$ denote the components of $\boldsymbol{\mu}$ and $\mathbf{Z}_n$ respectively. The parametric model assumes that for some given $s < k$, $\mu_{s+1} = \ldots = \mu_k = 0$. For some given $1 \le q \le s$, let $\theta = (\sum_{l=1}^q \mu_l^2)/(k\sigma^2)$, and the hypotheses are $H_0: \theta = \theta_0$ and $H_1: \theta = \theta_1$, where θ_0, θ_1 are distinct nonnegative constants. Let

$$\begin{aligned} \nu &= 0 && \text{if } \theta_0 > 0 \text{ and } \theta_1 > 0, \\ &= \tfrac{1}{2}(1 - q) && \text{if } \theta_1 > 0 \text{ and } \theta_0 = 0, \\ &= \tfrac{1}{2}(q - 1) && \text{if } \theta_0 > 0 \text{ and } \theta_1 = 0. \end{aligned} \tag{2.14}$$

For $\theta, x \ge 0$, define

$$H(\theta, x) = -\tfrac{1}{2}\theta + \tfrac{1}{4}\{\theta x + (\theta x)^{1/2}(\theta x + 4)^{1/2}\} + \log\{(\theta x)^{1/2} + (\theta x + 4)^{1/2}\}, \tag{2.15}$$

$$G(x) = k\{H(\theta_1, x) - H(\theta_0, x)\}.$$

Note that $H(\theta, x) = h(\theta^{1/2}, x^{1/2})$, where h is defined in (2.11). Sufficiency and invariance under some group G of transformations reduce the data to the sequence $\{X_n\}$, where

$$X_n = (\textstyle\sum_{l=1}^{q} \bar{Z}_{nl}^2)/[n^{-1} \sum_{j=1}^{n}\{\sum_{l=1}^{q} Z_{jl}^2 + \sum_{l=q+1}^{s}(Z_{jl} - \bar{Z}_{nl})^2 + \sum_{l=s+1}^{k} Z_{jl}^2\}] \tag{2.16}$$

(cf. [17]), and as has been shown in [6], pages 593–595, the likelihood ratio $R_n = p_{1n}(X_n)/p_{0n}(X_n)$ has the following approximation:

$$|\log R_n - \{n\, G(X_n) + \nu \log n\}| \le c, \qquad n = 1, 2, \dots, \tag{2.17}$$

where c is a positive constant and ν is as defined in (2.14). Under $H_i (i = 0, 1)$, since $X_n \to \theta_i/(1 + \theta_i)$ r-quickly for all $r > 0$, it follows from (2.17) that

$$n^{-1} \log R_n \to G(\theta_i/(1 + \theta_i)) \quad r\text{-quickly} \tag{2.18}$$

for all $r > 0$. Since $H(\theta, x) = h(\theta^{1/2}, x^{1/2})$, where h is as defined in Example 1, the same argument as in Example 1 shows that $G(\theta_0/(1 + \theta_0)) < 0$ and $G(\theta_1/(1 + \theta_1)) > 0$. Hence by Corollary 1, the sequential F-test $S(A_{\alpha,\beta}, B_{\alpha,\beta})$ is asymptotically (as $\alpha + \beta \to 0$) optimal within the class $\mathscr{T}(\alpha, \beta)$ of invariant sequential tests in the sense that (2.9) holds for all $r > 0$.

EXAMPLE 3. (*Sequential T^2-test*). Let $\mathbf{Z}, \mathbf{Z}_1, \mathbf{Z}_2, \dots$ be i.i.d. $\mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ k-dimensional random vectors with $\boldsymbol{\Sigma}$ nonsingular. Letting $\theta^2 = \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$, the two hypotheses are $H_i: \theta = \theta_i$ $(i = 0, 1)$, where θ_0, θ_1 are distinct nonnegative constants. Sufficiency and invariance under the general linear group $GL(k)$ of linear transformations $\mathbf{Z} \to \mathbf{CZ}$ ($\mathbf{C}$ nonsingular) reduce the data to the sequence $\{X_n\}$, where

$$X_n = V_n/(1 + V_n), \qquad V_n = \bar{\mathbf{Z}}_n' \mathbf{S}_n^{-1} \bar{\mathbf{Z}}_n, \tag{2.19}$$

and $\bar{\mathbf{Z}}_n$ and $\mathbf{S}_n$ are the sample mean vector and the sample covariance matrix (at stage n) respectively. Hence any invariant test of H_0 versus H_1 is based on the sequence $\{X_n\}$. Defining ν and G as in (2.14) and (2.15), the likelihood ratio $R_n = p_{1n}(V_n)/p_{0n}(V_n)$ in the present case also satisfies the approximation (2.17) (see (48) and (55) of [6]). Since $X_n \to \theta_i/(1 + \theta_i)$ r-quickly for all $r > 0$ under $H_i (i = 0, 1)$, it follows from Corollary 1 that the sequential T^2-test $S(A_{\alpha,\beta}, B_{\alpha,\beta})$ is asymptotically (as $\alpha + \beta \to 0$) optimal within the class $\mathscr{T}(\alpha, \beta)$ in the sense that (2.9) holds for all $r > 0$.

EXAMPLE 4. (*Sequential rank-order test for Lehmann alternatives*). Suppose that $Y_1, Y_2, \dots$ are i.i.d. with a continuous distribution function F, and are independent of $Z_1, Z_2, \dots$ which are i.i.d. with a continuous distribution function G. Here F and G are unknown, and the hypotheses are $H_0: G = F$ and $H_1: G = F^A$, where $0 < A \neq 1$ is a known constant. Let $F_n(x) = n^{-1} \sum_1^n I_{[Y_j \le x]}$, $G_n(x) = n^{-1} \sum_1^n I_{[Z_j \le x]}$. At stage n, a maximal invariant with respect to the group of transformations $(Y_j, Z_j) \to (\psi(Y_j), \psi(Z_j))$ (ψ being an arbitrary continuous increasing function) is the vector of ranks of $Y_1, \dots, Y_n$ among $Y_1, \dots, Y_n, Z_1, \dots, Z_n$, and the likelihood ratio of this maximal invariant is

$$R_n = A^n((2n)!)/\{n^{2n} \textstyle\prod_{j=1}^{n} [F_n(Y_j) + AG_n(Y_j)][F_n(Z_j) + AG_n(Z_j)]\} \tag{2.20}$$

(cf. [13]). Define

$$S(A, F, G) = \log 4A - 2 - \int \log(F(x) + AG(x))(dF(x) + dG(x)). \tag{2.21}$$

As shown in [13], $S(A, F, F) = \lambda_0(A) < 0$ and $S(A, F, F^A) = \lambda_1(A) > 0$ for all $0 < A \neq 1$. Moreover, by Lemma 1 of [13], $n^{-1} \log R_n \to \lambda_i(A)$ r-quickly for every $r > 0$ under H_i $(i = 0, 1)$. Hence by Corollary 1, the Savage-Sethuraman test $S(A_{\alpha,\beta}, B_{\alpha,\beta})$ is asymptotically (as $\alpha + \beta \to 0$) optimal within the class $\mathscr{T}(\alpha, \beta)$ of sequential rank tests in the sense that (2.9) holds for all $r > 0$.

EXAMPLE 5. (*Test for the mean level of a Gaussian sequence*). Let $\{X_n\}$ be a pth order autoregressive stationary Gaussian sequence with unknown mean level θ. Thus $X_n = Y_n + \theta$, where

$$Y_n = \beta_1 Y_{n-1} + \cdots + \beta_p Y_{n-p} + Z_n, \qquad n > p, \tag{2.22}$$

and $Z_{p+1}, Z_{p+2}, \ldots$ are i.i.d. $N(0, \sigma^2)$ random variables which are independent of $Y_1, \ldots, Y_p$ with joint density $g(y_1, \ldots, y_p)$. We shall assume that $\sigma, \beta_1, \ldots, \beta_p$ and g are known, and without loss of generality, we shall take $\sigma = 1$. The two hypotheses here are $H_i: \theta = \theta_i (i = 0, 1)$, and the likelihood ratio R_n with $n > p$ is of the form

$$\begin{aligned}\log R_n = {} & (1 - \beta_1 - \cdots - \beta_p)(\theta_1 - \theta_0) \textstyle\sum_{j=p+1}^n (X_j - \beta_1 X_{j-1} - \cdots - \beta_p X_{j-p}) \\ & - \tfrac{1}{2}(n-p)(1 - \beta_1 - \cdots - \beta_p)^2(\theta_1^2 - \theta_0^2) \\ & + \log\{g(X_1 - \theta_1, \ldots, X_p - \theta_1)/g(X_1 - \theta_0, \ldots, X_p - \theta_0)\}\end{aligned} \tag{2.23}$$

(cf. [2], page 184). Letting $S_n = \sum_1^n Y_j$, S_n is $N(0, \sigma_n^2)$, where $\sigma_n^2 = \text{Var } S_n = 0(n)$ since $\text{Cov}(Y_1, Y_n)$ converges to 0 exponentially fast (cf. [2], page 175), and consequently

$$n^{-1} S_n \to 0 \qquad r\text{-quickly for every} \quad r > 0. \tag{2.24}$$

Since $X_n = Y_n + \theta$, it follows from (2.23) and (2.24) that under $H_i (i = 0, 1)$,

$$n^{-1} \log R_n \to (-1)^{i+1}(1 - \beta_1 - \cdots - \beta_p)^2(\theta_1 - \theta_0)^2/2 \quad r\text{-quickly} \tag{2.25}$$

for every $r > 0$. Hence by Corollary 1, the SPRT $S(A_{\alpha\beta}, B_{\alpha,\beta})$ is asymptotically (as $\alpha + \beta \to 0$) optimal within the class $\mathscr{T}(\alpha, \beta)$ of sequential tests based on $\{X_n\}$ in the sense that (2.9) holds for all $r > 0$.

3. Sequential tests which are asymptotically optimal at an intermediate parameter.

For the sequential t-test $S(A_{\alpha,\beta}, B_{\alpha,\beta})$ of Example 1, the rth moment of the stopping rule $\tau_{\alpha,\beta}$ at the parameter $\mu/\sigma = \gamma$ is finite for all $r > 0$ and is given by

$$\begin{aligned} E(\tau_{\alpha,\beta}^r \mid \gamma) & \sim |\log \alpha|^r / \Psi^r(\gamma/(1+\gamma^2)^{1/2}) \quad \text{if} \quad \Psi(\gamma/(1+\gamma^2)^{1/2}) > 0, \\ & \sim |\log \beta|^r / |\Psi(\gamma/(1+\gamma^2)^{1/2})|^r \quad \text{if} \quad \Psi(\gamma/(1+\gamma^2)^{1/2}) < 0, \end{aligned} \tag{3.1}$$

where Ψ is as defined in (2.11) and $E(\cdot \mid \gamma)$ denotes expectation with respect to the probability measure under which $Z_1, Z_2, \ldots$ are i.i.d. $N(\mu, \sigma^2)$ with $\mu/\sigma = \gamma$ (cf. [6]). Since $\Psi'(u) \neq 0$ for all u, the equation $\Psi(\gamma/(1 + \gamma^2)^{1/2}) = 0$ has a unique root γ^*, and γ^* lies between γ_0 and γ_1. By Theorem 3 of [8], as $\alpha + \beta \to 0$ such that $|\log \alpha| \sim |\log \beta|$,

$$E(\tau_{\alpha,\beta}^r \mid \gamma^*) \sim \lambda_r |\log \alpha|^{2r}, \tag{3.2}$$

where λ_r is a positive constant depending on r. Thus, although the sequential t-test asymptotically minimizes the rth moment of the sample size at γ_0 and γ_1, it has an inordinately large sample size at the least favorable parameter γ^*. From the minimax point of view, it is therefore of interest to minimize the rth moment of the sample size at the parameter γ^*. The following theorem and its corollary deal with the general problem of asymptotically minimizing the rth moment of the sample size under a probability measure P which need not equal P_0 or P_1.

THEOREM 2. *For $0 < \alpha, \beta < 1$, let $\mathscr{T}(\alpha, \beta)$ be the class of tests of H_0 versus H_1 as defined in Theorem 1, and for $i = 0, 1$, let $p_{in}(x_1, \ldots, x_n)$ denote the joint density (under H_i) of $(X_1, \ldots, X_n)$ with respect to Q_n. Let P be a probability measure on $(\Omega, \mathscr{F})$ such that* *under P*, $(X_1, \ldots, X_n)$ *has a joint density* $p_n(x_1, \ldots, x_n)$ *with respect to* Q_n *for every* $n \geq 1$. *Define*

$$R_n^{(i)} = p_n(X_1, \ldots, X_n)/p_{in}(X_1, \ldots, X_n), \qquad i = 0, 1. \tag{3.3}$$

Assume that there exist finite constants η_0 *and* η_1 *such that*

$$\eta_0 \geq 0, \qquad \eta_1 \geq 0, \qquad \max\{\eta_0, \eta_1\} > 0, \tag{3.4}$$

and

$$n^{-1} \log R_n^{(i)} \to \eta_i \text{ a.s. } [P], \qquad i = 0, 1. \tag{3.5}$$

(*i*) *For every* $0 < \delta < 1$, *as* $\alpha + \beta \to 0$.

$$\inf_{(N,d) \in \mathscr{T}(\alpha,\beta)} P[N > \delta \min\{|\log \alpha|/\eta_0, |\log \beta|/\eta_1\}] \to 1 \tag{3.6}$$

(*where* $a/0$ *is defined as* ∞ *for* $a > 0$).

(ii) *For* $0 < \alpha, \beta < 1$, *let* $C_{\alpha,\beta}$ *and* $D_{\alpha,\beta}$ *be positive constants such that*

$$\log C_{\alpha,\beta} \sim |\log \alpha|, \log D_{\alpha,\beta} \sim |\log \beta| \quad as \quad \alpha + \beta \to 0. \tag{3.7}$$

Define

$$T_{\alpha,\beta} = \inf\{n \geq 1 : R_n^{(0)} \geq C_{\alpha,\beta} \quad or \quad R_n^{(1)} \geq D_{\alpha,\beta}\} \quad (\inf \phi = \infty). \tag{3.8}$$

Let $(T_{\alpha,\beta}, d^*)$ *be the test which stops sampling at stage* $T_{\alpha,\beta}$ *and rejects* H_0 iff $R_{T_{\alpha,\beta}}^{(0)} \geq C_{\alpha,\beta}$. *Then as* $\alpha + \beta \to 0$,

$$\frac{T_{\alpha,\beta}}{\min\{|\log \alpha|/\eta_0, \ |\log \beta|/\eta_1\}} \to 1 \quad \text{a.s. } [P], \tag{3.9}$$

and consequently, for every $0 < \delta < 1$,

$$\inf_{(N,d) \in \mathscr{T}(\alpha,\beta)} P[N > \delta \, T_{\alpha,\beta}] \to 1. \tag{3.10}$$

Moreover, the error probabilities of the test $(T_{\alpha,\beta}, d^*)$ *satisfy*

$$\begin{aligned} P_0[(T_{\alpha,\beta}, d^*) \text{ rejects } H_0] &\leq C_{\alpha,\beta}^{-1} \, P[(T_{\alpha,\beta}, d^*) \text{ rejects } H_0], \\ P_1[(T_{\alpha,\beta}, d^*) \text{ rejects } H_1] &\leq D_{\alpha,\beta}^{-1} \, P[(T_{\alpha,\beta}, d^*) \text{ rejects } H_1]. \end{aligned} \tag{3.11}$$

PROOF. Let $l_n^{(i)} = \log R_n^{(i)}$. Let $0 < \delta < 1$ and $\tilde{\delta} > 1$ such that $\delta\tilde{\delta} < 1$. Let m be the greatest integer $\leq \delta \min\{|\log \alpha|/\eta_0, |\log \beta|/\eta_1\}$. Then for $(N, d) \in \mathscr{T}(\alpha, \beta)$,

$$\begin{aligned} \alpha &= \int_{\{N<\infty,(N,d) \text{ rejects } H_0\}} \exp(-\, l_N^{(0)}) \, dP \\ &\geq \int_{\{N \leq m, l_N^{(0)} \leq \tilde{\delta}\eta_0 m, (N,d) \text{ rejects } H_0\}} \exp(-\, l_N^{(0)}) \, dP \\ &\geq \exp(-\, \tilde{\delta}\eta_0 m) P[N \leq m, \, l_N^{(0)} \leq \tilde{\delta}\eta_0 m, \, (N, d) \text{ rejects } H_0]. \end{aligned} \tag{3.12}$$

Since $\tilde{\delta}\eta_0 m \leq \delta\tilde{\delta}\,|\log \alpha|$, it follows from (3.12) that

$$\begin{aligned} P[N \leq m, \, (N, d) \text{ rejects } H_0] &\leq \alpha^{1-\delta\tilde{\delta}} + P[N \leq m, \, l_N^{(0)} > \tilde{\delta}\eta_0 m] \\ &\leq \alpha^{1-\delta\tilde{\delta}} + P[\max_{j \leq m} l_j^{(0)} > \tilde{\delta}\eta_0 m]. \end{aligned} \tag{3.13}$$

Using a similar argument, we also obtain that

$$P[N \le m,\ (N, d) \text{ rejects } H_1] \le \beta^{1-\delta\tilde{\delta}} + P[\max_{j\le m} l_j^{(1)} > \tilde{\delta}\eta_1 m]. \tag{3.14}$$

From (3.13) and (3.14), it follows that

$$\begin{aligned}\sup_{(N,d)\in\mathscr{T}(\alpha,\beta)} P[N \le m] \le \alpha^{1-\delta\tilde{\delta}} + \beta^{1-\delta\tilde{\delta}} + P[\max_{j\le m} l_j^{(0)} > \tilde{\delta}\eta_0 m] \\ + P[\max_{j\le m} l_j^{(1)} > \tilde{\delta}\eta_1 m].\end{aligned} \tag{3.15}$$

Since $j^{-1}l_j^{(i)} \to \eta_i$ a.s. $[P]$ for $i = 0, 1$ and $\tilde{\delta} > 1$, (3.6) follows from (3.15).

The a.s. asymptotic behavior (3.9) of $T_{\alpha,\beta}$ follows easily from (3.5) and (3.7). From (3.6) and (3.9), (3.10) follows immediately. The bounds in (3.11) for the error probabilities of $(T_{\alpha,\beta}, d^*)$ can be proved by essentially the same standard argument used for Wald's SPRT. □

COROLLARY 2. *With the same notations and assumptions as in Theorem 2, suppose that* (3.5) *is strengthened into*

$$n^{-1}\log R_n^{(i)} \to \eta_i \quad r\text{-quickly under } P \qquad i = 0, 1, \tag{3.16}$$

for some positive constant r. Then $ET_{\alpha,\beta}^r < \infty$, *and as* $\alpha + \beta \to 0$,

$$\inf_{(N,d)\in\mathscr{T}(\alpha,\beta)} EN^r \sim ET_{\alpha,\beta}^r \sim (\min\{|\log\alpha|/\eta_0, |\log\beta|/\eta_1\})^r, \tag{3.17}$$

where E denotes expectation with respect to P.

PROOF. Take $0 < \delta < 1$. It follows from (3.6) that as $\alpha + \beta \to 0$,

$$\inf_{(N,d)\in\mathscr{T}(\alpha,\beta)} EN^r \ge \delta^r(\min\{|\log\alpha|/\eta_0, |\log\beta|/\eta_1\})^r(1 + o(1)). \tag{3.18}$$

First consider the case where $\eta_0 > 0$ and $\eta_1 > 0$. Let $0 < a < \min\{\eta_0, \eta_1\}$ and define $L = \sup\{n \ge 1 : \max_{i=0,1}|n^{-1}l_n^{(i)} - \eta_i| > a\}$ ($\sup\phi = 0$), where $l_n^{(i)} = \log R_n^{(i)}$. On the event $\{T_{\alpha,\beta} - 1 > L\}$,

$$\begin{aligned}(\eta_i - a)(T_{\alpha,\beta} - 1) < l_{T_{\alpha,\beta}-1}^{(i)} &< \log C_{\alpha,\beta}, && i = 0, \\ &< \log D_{\alpha,\beta}, && i = 1.\end{aligned} \tag{3.19}$$

Since $T_{\alpha,\beta} \le L + 1$ on the complementary event, we obtain from (3.19) that

$$T_{\alpha,\beta} \le L + 1 + \min\{(\eta_0 - a)^{-1}\log C_{\alpha,\beta}, (\eta_1 - a)^{-1}\log D_{\alpha,\beta}\}. \tag{3.20}$$

Since $EL^r < \infty$ by (3.16), (3.20) implies that $ET_{\alpha,\beta}^r < \infty$ and that

$$ET_{\alpha,\beta}^r \sim (\min\{|\log\alpha|/\eta_0, |\log\beta|/\eta_1\})^r, \tag{3.21}$$

in view of (3.7), (3.9), and the dominated convergence theorem. The validity of (3.21) in the case where $\eta_0 = 0 < \eta_1$, say, can be proved by applying a similar argument to $\tilde{T}_{\alpha,\beta} = \inf\{n : R_n^{(1)} \ge D_{\alpha,\beta}\}$ $(\ge T_{\alpha,\beta})$ and noting that (3.21) reduces to $ET_{\alpha,\beta}^r \sim (|\log\beta|/\eta_1)^r$ in this case.

From (3.18) (with $\delta \uparrow 1$), (3.21), and Remark (i) below, the desired conclusion (3.17) follows immediately. □

REMARKS. (i) In view of the bounds in (3.11) for the error probabilities of $(T_{\alpha,\beta}, d^*)$, if we take $C_{\alpha,\beta} = \alpha^{-1}$ and $D_{\alpha,\beta} = \beta^{-1}$, then (2.1) holds for $(T_{\alpha,\beta}, d^*)$ which therefore belongs to $\mathscr{T}(\alpha, \beta)$; moreover, (3.7) obviously holds.

(ii) Setting $P = P_0$ in Theorem 2(i), we obtain (2.5a) of Theorem 1(i), while (2.5b) follows from Theorem 2(i) with $P = P_1$. The proof of Theorem 1(ii) and Corollary 1 is exactly analogous to that of Theorem 2(ii) and Corollary 2.

(iii) When $X, X_1, X_2, \ldots$ are i.i.d., tests of the form $(T_{\alpha,\beta}, d^*)$ were first proposed by Anderson [1] in the case where X is normal under P_0, P_1, and P, and were recently extended by Lorden [11] to the case where $E\{\log(p(X)/p_0(X))\}^2 + E\{\log(p(X)/p_1(X))\}^2 < \infty$, P_0, P_1, and P being mutually distinct. Note that these assumptions of Lorden in the i.i.d. case imply that the assumption (3.16) of Corollary 2 is satisfied with $r = 1$ (cf. [9]).

For the problem in Example 1 of testing $H_0: \gamma = \gamma_0$ versus $H_1: \gamma = \gamma_1$, where $\gamma = \mu/\sigma$ (μ and σ^2 being the mean and variance of the i.i.d. normal observations $Z_1, Z_2, \ldots$), we note that the assumption (3.16) of Corollary 2 is satisfied for all $r > 0$ at any parameter γ (not necessarily equal to γ_0 or γ_1). To see this, let $R_n^{(i)} = U_n(\gamma)/U_n(\gamma_i)$, $i = 0, 1$, where $U_n(\cdot)$ is as defined in (2.10). Replacing (γ_0, γ_1) in the argument leading to (2.13) in Example 1 by (γ_i, γ), we obtain that (3.16) holds with

$$\eta_i = \eta_i(\gamma) = h(\gamma, \gamma/(1+\gamma^2)^{1/2}) - h(\gamma_i, \gamma/(1+\gamma^2)^{1/2}) \qquad (>0 \text{ for } \gamma \neq \gamma_i), \tag{3.22}$$

where h is as defined in (2.11). Hence letting $\eta(\gamma) = \max\{\eta_0(\gamma), \eta_1(\gamma)\}$, it follows from Corollary 2 that as $\alpha + \beta \to 0$ such that $|\log \alpha| \sim |\log \beta|$,

$$\inf_{(N,d) \in \mathscr{T}(\alpha,\beta)} E(N^r \mid \gamma) \geq (1 + o(1)) |\log \alpha|^r / \eta^r(\gamma) \tag{3.23}$$

for all $r > 0$. Moreover, the lower bound in (3.23) is asymptotically attained by the test $(T_{\alpha,\beta}(\gamma), d^*)$ where

$$T_{\alpha,\beta}(\gamma) = \inf\{n \geq 1: U_n(\gamma)/U_n(\gamma_0) \geq \alpha^{-1} \quad \text{or} \quad U_n(\gamma)/U_n(\gamma_1) \geq \beta^{-1}\}. \tag{3.24}$$

The minimum of $\eta(\gamma)$ occurs at the root γ^* of the equation $\Psi(\gamma/(1+\gamma^2)^{1/2}) = 0$, where Ψ is as defined in (2.11). Hence in view of (3.23) with $\gamma = \gamma^*$ and Lemma 1 below, the test $(T_{\alpha,\beta}(\gamma^*), d^*)$ is asymptotically minimax within the class $\mathscr{T}(\alpha, \beta)$ in the sense that for all $r > 0$

$$\begin{aligned} \inf_{(N,d) \in \mathscr{T}(\alpha,\beta)} \sup_\gamma E(N^r \mid \gamma) &\sim E(T^r_{\alpha,\beta}(\gamma^*) \mid \gamma^*) \\ &\sim \sup_\gamma E(T^r_{\alpha,\beta}(\gamma^*) \mid \gamma) \text{ as } \alpha + \beta \to 0 \text{ such that } |\log \alpha| \sim |\log \beta|. \end{aligned} \tag{3.25}$$

LEMMA 1. *Let $Z_1, Z_2, \ldots$ be i.i.d. $N(\mu, \sigma^2)$ with $\mu/\sigma = \gamma$. Define $U_n(\cdot)$ as in (2.10), h as in (2.11), and let $\gamma^*, \gamma_0, \gamma_1, T_{\alpha,\beta}(\gamma^*)$ be the same as in (3.25). Define $\Psi_i^*(u) = h(\gamma^*, \mu) - h(\gamma_i, u)$, $i = 0, 1$. For $a > 0$, let*

$$L_{a,\gamma} = \sup\{n \geq 1: \max_{i=0,1} |n^{-1} \log(U_n(\gamma^*)/U_n(\gamma_i)) - \Psi_i^*(\gamma/(1+\gamma^2)^{1/2})| > a\} \qquad (\sup \phi = 0).$$

Then for all $r > 0$ and $a > 0$, $\sup_\gamma E(L^r_{a,\gamma} \mid \gamma) < \infty$, and consequently,

$$\sup_\gamma E(T^r_{\alpha,\beta}(\gamma^*) \mid \gamma) \sim E(T^r_{\alpha,\beta}(\gamma^*) \mid \gamma^*) \text{ as } \alpha + \beta \to 0 \text{ such that } |\log \alpha| \sim |\log \beta|. \tag{3.26}$$

PROOF. Let $Y_1, Y_2, \ldots$ be i.i.d. $N(0, 1)$. Writing $Z_i = \sigma Y_i + \mu$, we can regard all the random variables (including $L_{a,\gamma}$ for all γ) as being defined on the same probability space and generated by the sequence $\{Y_n\}$. As in (2.12), there exists a constant c for which

$$|\log(U_n(\gamma^*)/U_n(\gamma_i)) - n\Psi_i^*(T_n)| \leq c, \qquad n = 1, 2, \ldots, i = 0, 1, \tag{3.27}$$

where

$$(3.28)\qquad T_n = \frac{\bar{Z}_n}{(n^{-1}\sum_1^n Z_i^2)^{1/2}} = \frac{\bar{Y}_n + \gamma}{\{n^{-1}\sum_1^n(Y_i - \bar{Y}_n)^2 + (\bar{Y}_n + \gamma)^2\}^{1/2}}.$$

The functions $\Psi_i^*(u)$, $i = 0, 1$, are both uniformly continuous for $|u| \le 1$, while the function $\phi(x, y) = y/(x + y^2)^{1/2}$ is uniformly continuous for $2 \ge x \ge \frac{1}{2}$ and $-\infty < y < \infty$. Hence in view of (3.27), (3.28), and the fact that $|T_n| \le 1$, for every $a > 0$ there exists $0 < b = b(a) < \frac{1}{2}$ such that

$$(3.29)\qquad \sup_\gamma L_{a,\gamma} \le L_b = \sup\{n \ge 1 : |\bar{Y}_n| > b \quad \text{or} \quad |n^{-1}\sum_1^n(Y_i - \bar{Y}_n)^2 - 1| > b\}.$$

We note that the function $w(\gamma) = \max_{i=0,1}\Psi_i^*(\gamma/(1 + \gamma^2)^{1/2})$ has its minimum at γ^*, and that $w(\gamma^*) = \eta(\gamma^*)$, where $\eta(\gamma) = \max_{i=0,1}\eta_i(\gamma)$ and $\eta_i(\gamma)$ is as defined in (3.22). Let $T_{\alpha,\beta}(\gamma^* \mid \gamma)$ denote the stopping rule $T_{\alpha,\beta}(\gamma^*) = \inf\{n : U_n(\gamma^*)/U_n(\gamma_0) \ge \alpha^{-1}$ or $U_n(\gamma^*)/U_n(\gamma_1) \ge \beta^{-1}\}$ when the underlying sequence $\{Z_n = \sigma Y_n + \mu\}$ satisfies $\mu/\sigma = \gamma$. Let $0 < a < w(\gamma^*)$. Since $(w(\gamma) - a)^{-1} \le (w(\gamma^*) - a)^{-1}$ for all γ, it can be shown by the same argument as in (3.19)–(3.20) that

$$(3.30)\qquad \sup_\gamma T_{\alpha,\beta}(\gamma^* \mid \gamma) \le \sup_\gamma L_{a,\gamma} + 1 + \{w(\gamma^*) - a\}^{-1}\max\{|\log \alpha|, |\log \beta|\}.$$

Since $E(T_{\alpha,\beta}^r(\gamma^*) \mid \gamma^*) \sim \{|\log \alpha|/w(\gamma^*)\}^r$ as $\alpha + \beta \to 0$ such that $|\log \alpha| \sim |\log \beta|$, and since $EL_b^r < \infty$ for all $b > 0$ (cf. [9]), the desired conclusion (3.26) follows from (3.29) and (3.30) (where a can be arbitrarily small). □

For the k-variate normal models in Examples 2 and 3, we again have a parametric family $\{P(\theta)\}$ of distributions (where $\theta = \sum_{j=1}^q \mu_j^2)/(k\sigma^2)$ in Example 2 and $\theta^2 = \mu' \sum^{-1} \mu$ in Example 3) for an invariantly sufficient sequence $\{X_n\}$, and it can similarly be shown that the assumption (3.16) of Corollary 2 is satisfied for all $r > 0$ by $P = P(\theta)$ for every $\theta \ge 0$. Hence Corollary 2 is applicable and gives asymptotically minimax tests (within the class $\mathscr{T}(\alpha, \beta)$) of $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. Likewise, it can be shown that Corollary 2 is also applicable to the problem of testing the mean level of the autoregressive Gaussian sequence of Example 5.

4. Higher-order asymptotic optimality and extensions of Wald's lower bounds for the expected sample size. When $X_1, X_2, \ldots$ are i.i.d., Wald [14] obtained the lower bounds

$$(4.1)\qquad \begin{aligned} |\lambda_0| E_0 N &\ge (1 - \alpha)\log((1 - \alpha)/\beta) + \alpha \log(\alpha/(1 - \beta)), \\ \lambda_1 E_1 N &\ge (1 - \beta)\log((1 - \beta)/\alpha) + \beta \log(\beta/(1 - \alpha)), \end{aligned}$$

for the expected sample size $E_i N$ of an arbitrary test $(N, d) \in \mathscr{T}^*(\alpha, \beta)$, *where* $\mathscr{T}^*(\alpha, \beta) = \{(N, d) \in \mathscr{T}(\alpha, \beta) : P_i[N < \infty] = 1 \text{ for } i = 0, 1\}$ and $\lambda_i = E_i\{\log(p_1(X_1)/p_0(X_1))\}$. Ignoring overshoots, Wald's SPRT with boundaries, A, B given by equalities in (1.3) attain the lower bounds in (4.1) (cf. [14], page 157). As $\alpha + \beta \to 0$ such that $\alpha \log \beta + \beta \log \alpha \to 0$, these lower bounds reduce to

$$(4.2)\qquad \begin{aligned} \inf_{(N,d)\in \mathscr{T}^*(\alpha,\beta)} E_0 N &\ge |\lambda_0|^{-1}|\log \beta| + o(1), \\ \inf_{(N,d)\in \mathscr{T}^*(\alpha,\beta)} E_1 N &\ge \lambda_1^{-1}|\log \alpha| + o(1). \end{aligned}$$

While Wald's lower bounds (4.1) depend very heavily on the i.i.d. structure of $\{X_n\}$, the following theorem shows that asymptotic expansions similar to (4.2) actually hold in a much more general setting. As an application, we shall show that invariant SPRTs like the sequential t-test are not only first-order asymptotically optimal in the sense of Corollary 1, but they also attain (like Wald's SPRT) the Wald-type lower bounds up to the $o(1)$ term when the overshoots are neglected.

THEOREM 3. *With the same notation as in Theorem* 1, *let* $\mathscr{G}_1 \subset \mathscr{G}_2 \subset \dots$ *be a sequence of sub-σ-fields of* $\mathscr{F}$ *such that* $\mathscr{F}_n \subset \mathscr{G}_n$ *for every n. Suppose that under* P_i $(i = 0, 1)$, $\log R_n$ *has a representation of the form*

$$\log R_n = \sum_{j=1}^{n} Y_j + \xi_n, \tag{4.3}$$

where $Y_1, Y_2, \dots$ *are* i.i.d. *with* $E_i Y_1^2 < \infty$, Y_n *is* $\mathscr{G}_n$*-measurable and* Y_{n+1} *is independent of* $\mathscr{G}_n$ *for every* $n \geq 1$, *and*

$$E_0 Y_1 = \lambda_0 < 0, \qquad E_1 Y_1 = \lambda_1 > 0, \tag{4.4}$$

$$\xi_n \text{ converges in distribution to some random variable } \xi. \tag{4.5}$$

(*The random variables* Y_n, ξ_n, *and* ξ *may depend on* $i \in \{0, 1\}$.) *Assume for* $i = 0, 1$ *that there exist constants* $\Delta_n > 0$, $\rho > 0$, $0 < \delta \leq 1$, *and events* A_n (*which may also depend on* i) *such that*

$$P_i(\cup_{n \leq k \leq n+\rho n^\delta} \tilde{A}_k) = o(n^{-1}) \qquad (\tilde{A} = \text{complement of } A), \tag{4.6}$$

$$\{ |\xi_n| I(\cap_{n \leq k \leq n+\rho n^\delta} A_k),\ n \geq n_0 \} \text{ is uniformly integrable under } P_i, \tag{4.7}$$
$$\text{for some } n_0 \qquad (I(A) = \text{indicator functon of } A),$$

$$\lim_{n\to\infty} \Delta_n = 0 \qquad \text{and} \quad P_i[\max_{n \leq k \leq n+\rho n^\delta} |\xi_k - \xi_n| > \Delta_n] = o(n^{-1}), \tag{4.8}$$

$$P_1[\max_{j \leq n}(\log R_j) > \lambda_1 n + c n^\delta] + P_1[\log R_n < \lambda_1 n - c n^\delta] = o(n^{-1}), \tag{4.9}$$
$$P_0[\min_{j \leq n}(\log R_j) < \lambda_0 n - c n^\delta] + P_0[\log R_n > \lambda_0 n + c n^\delta] = o(n^{-1})$$
$$\text{for all} \quad c > 0.$$

Then as $\alpha + \beta \to 0$ *such that* $\alpha \log \beta + \beta \log \alpha \to 0$,

$$\inf_{(N,d) \in \mathscr{T}(\alpha,\beta)} E_0 N \geq |\lambda_0|^{-1}\{|\log \beta| + E_0 \xi\} + o(1), \tag{4.10a}$$

$$\inf_{(N,d) \in \mathscr{T}(\alpha,\beta)} E_1 N \geq \lambda_1^{-1}\{|\log \alpha| - E_1 \xi\} + o(1). \tag{4.10b}$$

REMARKS. (i) In view of (4.5) and (4.6), $\xi_n I(\cap_{n \leq k \leq n+\rho n^\delta} A_k)$ converges in distribution to ξ under P_i. Hence by (4.7) and Fatou's lemma, $E_i|\xi| < \infty$ for $i = 0, 1$.

(ii) In the case where $\xi_n = 0$, the conditions (4.5)–(4.8) are obviously satisfied with $A_k = \Omega$, and (4.9) holds with $1/2 < \delta \leq 1$ if $E_i|Y_1|^{2/\delta} < \infty$ (cf. [9]). Moreover, since $\xi = 0$, the bounds (4.10a) and (4.10b) are identical with the Wald bounds in (4.2).

PROOF OF THEOREM 3. We shall only prove (4.10b), as the proof of (4.10a) is similar. Let $0 < c < \min\{1, \rho/3\}$ and define

$$a = \lambda_1^{-1}|\log \alpha|, \qquad \underline{n} = [a - c a^\delta], \qquad \bar{n} = \underline{n} + [\rho \underline{n}^\delta], \qquad l_n = \log R_n, \tag{4.11}$$

where $[x]$ denotes the greatest integer $\leq x$. We first show that as $\alpha + \beta \to 0$,

$$\sup_{(N,d) \in \mathscr{T}(\alpha,\beta)} P_1[N \leq \underline{n},\ (N, d) \text{ rejects } H_0] = o(a^{-1}). \tag{4.12}$$

Using a similar argument as in (3.12), we obtain that for $(N, d) \in \mathscr{T}(\alpha, \beta)$

$$P_1[N \leq \underline{n},\ (N, d) \text{ rejects } H_0,\ l_N \leq \lambda_1 \underline{n} + \tfrac{1}{2}\lambda_1 c \underline{n}^\delta] \tag{4.13}$$
$$\leq \alpha \exp(\lambda_1 \underline{n} + \tfrac{1}{2}\lambda_1 c \underline{n}^\delta) = \exp\{-(\tfrac{1}{2}\lambda_1 c + o(1)) a^\delta\}, \qquad \text{by (4.11)}.$$

Since $P_1[N \leq \underline{n},\ l_N > \lambda_1 \underline{n} + \frac{1}{2}\lambda_1 c \underline{n}^\delta] \leq P_1[\max_{j \leq \underline{n}} l_j > \lambda_1 \underline{n} + \frac{1}{2}\lambda_1 c \underline{n}^\delta] = o(a^{-1})$ by (4.9), (4.12) follows from (4.13).

Define

(4.14) $$A = (\cap_{k=\underline{n}}^{\bar{n}} A_k) \cap \{\max_{\underline{n} \le k \le \bar{n}} |\xi_k - \xi_{\underline{n}}| \le \Delta_n\} \cap \{l_{\bar{n}} \ge |\log \alpha| + \lambda_1 c a^{\delta}\}.$$

Since $\lambda_1 \bar{n} - \lambda_1 c \bar{n}^{\delta} = |\log \alpha| + \{\lambda_1(\rho - 2c) + o(1)\} a^{\delta}$ and $\rho - 2c > c$, it follows from (4.6), (4.8), and (4.9) that as $\alpha + \beta \to 0$,

(4.15) $$P(\tilde{A}) = o(a^{-1}).$$

Define $\Omega_{N,d} = \{N > \underline{n}, (N, d) \text{ rejects } H_0\}$. By (4.12),

(4.16) $$\sup_{(N,d) \in \mathcal{T}(\alpha,\beta)} P_1(\tilde{\Omega}_{N,d}) = o(a^{-1})$$

as $\alpha + \beta \to 0$ such that $\beta \log \alpha \to 0$ (and therefore $\beta = o(a^{-1})$). On $A \cap \Omega_{N,d}$, $|\xi_{N \wedge \bar{n}} - \xi_{\underline{n}}| \le \Delta_{\underline{n}}$, where $N \wedge \bar{n}$ denotes $\min\{N, \bar{n}\}$. Therefore, letting $S_n = \sum_{j=1}^{n} (Y_j - \lambda_1)$, we obtain from (4.3) that

(4.17) $$\int_{A \cap \Omega_{N,d}} l_{N \wedge \bar{n}}\, dP_1 \le \lambda_1 E_1 N + \int_{A \cap \Omega_{N,d}} (S_{N \wedge \bar{n}} + \xi_{\underline{n}})\, dP_1 + \Delta_{\underline{n}}.$$

From (4.5), (4.7), (4.15), and (4.16), it follows that as $\alpha + \beta \to 0$ such that $\beta \log \alpha \to 0$,

(4.18) $$\sup_{(N,d) \in \mathcal{T}(\alpha,\beta)} \left| \int_{A \cap \Omega_{N,d}} \xi_{\underline{n}}\, dP_1 - E_1 \xi \right| \to 0.$$

Note that $(S_n, \mathcal{G}_n, n \ge 1)$ is a martingale and that N is also a stopping time relative to $\{\mathcal{G}_n\}$ since $\mathcal{F}_n \subset \mathcal{G}_n$. Therefore $E_1 S_{N \wedge \bar{n}} = 0$ by Wald's lemma and$E_1 S_{N \wedge \bar{n}}^2 = \sigma^2 E(N \wedge \bar{n})$, where $\sigma^2 = E_1(Y_1 - \lambda_1)^2$. It then follows that

(4.19) $$\left| \int_{A \cap \Omega_{N,d}} S_{N_{\wedge \bar{n}}}\, dP_1 \right| = \left| \int_{\tilde{A} \cup \tilde{\Omega}_{N,d}} S_{N \wedge \bar{n}}\, dP_1 \right| \le \sigma\{E_1(N \wedge \bar{n})\}^{1/2} \{P_1(\tilde{A} \cup \tilde{\Omega}_{N,})\}^{1/2},$$

by the Schwarz inequality. From (4.15), (4.16), and (4.19), we obtain that as $\alpha + \beta \to 0$ such that $\beta \log \alpha \to 0$

(4.20) $$\sup_{(N,d) \in \mathcal{T}(\alpha,\beta)} \left| \int_{A \cap \Omega_{N,d}} S_{N \wedge \bar{n}}\, dP_1 \right| \to 0.$$

On the event A, $1/R_{\bar{n}} = \exp(-l_{\bar{n}}) \le \alpha \exp(-\lambda_1 c a^{\delta})$, and therefore for $(N, d) \in \mathcal{T}(\alpha, \beta)$,

(4.21) $$P_0(\Omega_{N,d}) + \int_{A \cap \Omega_{N,d} \cap \{N > \bar{n}\}} (1/R_{\bar{n}})\, dP_1 \le \alpha\{1 + \exp(-\lambda_1 c a^{\delta})\}.$$

Hence by Lemma 2 below,

(4.22) $$\int_{A \cap \Omega_{N,d}} l_{N \wedge \bar{n}}\, dP_1 \ge P_1(A \cap \Omega_{N,d}) \log\left\{ \frac{P_1(A \cap \Omega_{N,d})}{\alpha(1 + \exp(-\lambda_1 c a^{\delta}))} \right\}.$$

Since $\inf_{(N,d) \in \mathcal{T}(\alpha,\beta)} P_1(A \cap \Omega_{N,d}) = 1 + o(a^{-1})$ by (4.15) and (4.16), the desired conclusion (4.10b) follows from (4.17), (4.18), (4.20), and (4.22). □

LEMMA 2. *With the same notation as in Theorem 1, let N be a stopping rule relative to $\{\mathcal{F}_n\}$. Then for $A \in \mathcal{F}$, $B \in \mathcal{F}_N$* (i.e., $B \cap \{N = n\} \in \mathcal{F}_n$ *for all* n), *and* $m = 1, 2, \ldots,$

$$\int_{A \cap B} \log R_{N \wedge m}\, dP_1 \ge P_1(A \cap B) \log\left\{ \frac{P_1(A \cap B)}{P_0(B) + \int_{A \cap B \cap \{N > m\}} R_m^{-1}\, dP_1} \right\}.$$

PROOF. By Jensen's inequality,

$$(4.23)\qquad E_1[\log R_{N\wedge m}\,|\,A\cap B] \ge \log\{1/E_1[R_{N\wedge m}^{-1}\,|\,A\cap B]\}.$$

We note that

$$(4.24)\qquad \begin{aligned}\int_{A\cap B} R_{N\wedge m}^{-1}\,dP_1 &\le \int_{B\cap\{N\le m\}} R_N^{-1}\,dP_1 + \int_{A\cap B\cap\{N>m\}} R_m^{-1}\,dP_1\\ &= P_0(B\cap\{N\le m\}) + \int_{A\cap B\cap\{N>m\}} R_m^{-1}\,dP_1, \qquad \text{since}\quad B\in\mathscr{F}_N.\end{aligned}$$

From (4.23) and (4.24), the desired conclusion follows. □

We now apply Theorem 3 to the problem of testing $H_0:\gamma = \gamma_0$ versus $H_1:\gamma=\gamma_1$ of Example 1, where $\gamma = \mu/\sigma$ and μ, σ^2 are the mean and variance of the i.i.d. observations Z_1, $Z_2, \ldots$. Defining R_n, $U_n(\cdot)$, T_n, f as in (2.10) and applying Laplace's asymptotic formula (cf. Theorem 4.1 of [16]) to the integral $\int_0^\infty u^{-1}\exp[nf(u, T_n, \gamma)]\,du$, we obtain that

$$(4.25)\qquad \log R_n = n\Psi(T_n) + \tfrac12\log\{g(\gamma_0 T_n)/g(\gamma_1 T_n)\} + \theta_n(T_n),$$

where Ψ is as defined in (2.11) and

$$(4.26)\qquad g(u) = \tfrac14\{u + (u^2+4)^{1/2}\}^2 + 1,$$

$$(4.27)\qquad \theta_n(x)\to 0 \quad \text{uniformly in}\quad |x|\le 1.$$

Since $|T_n|\le 1$ and Ψ is continuous, it follows from (4.25) that there exist $C, D>0$ such that

$$(4.28)\qquad |\log R_n| \le C + Dn, \qquad n = 1, 2, \ldots.$$

Since the distribution of $\log R_n$ is scale invariant, we shall assume throughout the sequel that $\sigma = 1$ so that $\mu=\gamma$. Let

$$(4.29)\qquad t = \gamma/(1+\gamma;2)^{1/2}, \qquad v = 1+\gamma^2(=EZ_1^2).$$

Set $\psi(x,y) = \Psi(xy^{-1/2})$. Noting that $|T_n|\le 1$ and $T_n = \psi(\bar Z_n, n^{-1}\sum_1^n Z_j^2)$, we obtain by (4.25) and Taylor's expansion on ψ that

$$(4.30)\qquad \begin{aligned}\log R_n &= n\Psi(t) + (\textstyle\sum_1^n Z_j - n\gamma)\,\psi_x + (\sum_1^n Z_j^2 - nv)\,\psi_y\\ &\quad + \tfrac12 n^{-1}(\textstyle\sum_1^n Z_j - n\gamma)^2\psi_{xx} + n^{-1}(\sum_1^n Z_j - n\gamma)(\sum_1^n Z_j^2 - nv)\,\psi_{xy}\\ &\quad + \tfrac12 n^{-1}(\textstyle\sum_1^n Z_j^2 - nv)^2\,\psi_{yy} + \tfrac12\log\{g(\gamma_0 t)/g(\gamma_1 t)\} + r_n,\end{aligned}$$

where ψ_x, ψ_y, ψ_{xx}, ψ_{xy}, ψ_{yy} denote the partial derivatives of $\psi(x,y)$ evaluated at $x=\gamma$, $y=v$ (i.e., $\psi_x = (\partial/\partial x)\psi(\gamma, v) = v^{-1/2}\,\Psi'(t)$, etc.), and

$$(4.31)\qquad |r_n| \le K\{|T_n - t| + n|\bar Z_n - \gamma|^3 + n|n^{-1}\textstyle\sum_1^n Z_j^2 - v|^3\} + c_n,$$

K and c_n being nonrandom constants such that $\lim_{n\to\infty} c_n = 0$ (since $\log g(\gamma_i u)$ has a bounded derivative for $|u|\le 1$, $i=0,1$, and the third-order partial derivatives of $\psi(x,y)$ are bounded for $|xy^{-1/2}|\le 1$).

Let $Y_j = \Psi(t) + (Z_j-\gamma)\psi_x + (Z_j^2 - v)\psi_y$, and

$$(4.32)\qquad \begin{aligned}\xi_n^{(1)} &= n^{-1}(\textstyle\sum_1^n Z_j - n\gamma)^2, \qquad \xi_n^{(2)} = n^{-1}(\sum_1^n Z_j^2 - nv)^2,\\ \xi_n^{(3)} &= 2n^{-1}(\textstyle\sum_1^n Z_j - n\gamma)(\sum_1^n Z_j^2 - nv) = n^{-1}\{\sum_1^n Z_j\\ &\quad - n\gamma + \textstyle\sum_1^n Z_j^2 - nv\}^2 - \xi_n^{(1)} - \xi_n^{(2)},\\ \xi_n &= \tfrac12\{\xi_n^{(1)}\psi_{xx} + \xi_n^{(2)}\psi_{yy} + \xi_n^{(3)}\psi_{xy} + \log(g(\gamma_0 t)/g(\gamma_1 t))\} + r_n.\end{aligned}$$

Let $A_n = \{|T_n - t| + n|\bar Z_n - \gamma|^3 + n|n^{-1}\sum_1^n Z_j^2 - v|^3 < (\log n)^{-2}\}$. Then

$$P_i(\tilde{A}_n) = o(n^{-p}) \quad \text{for all} \quad p > 0 \qquad i = 0, 1. \tag{4.33}$$

Take any $\rho > 0$ and $\frac{1}{2} < \delta < 1$. Then (4.33) implies (4.6). Since it follows from (4.31) that

$$|r_n| \le K(\log n)^{-2} + c_n \quad \text{on} \quad A_n, \tag{4.34}$$

we obtain by applying Proposition 1 of [10] to $\xi_n^{(1)}$, $\xi_n^{(2)}$, $\xi_n^{(3)}$ that conditions (4.5) and (4.7) are satisfied by ξ_n. The random variable ξ in (4.5) is given by

$$\xi = \tfrac{1}{2} W^2\psi_{xx} + W(W^2 + 2\gamma W - 1)\psi_{xy} + \tfrac{1}{2} (W^2 + 2\gamma W - 1)^2\psi_{yy} + \tfrac{1}{2} \log\{g(\gamma_0 t)/g(\gamma_1 t)\}, \tag{4.35}$$

where W denotes generically the $N(0, 1)$ random variable. By an argument similar to that in the proof of Proposition 1 of [10], page 69, it can be shown that (4.8) holds with $\Delta_n = (\log n)^{-2}$ for each of $\xi_n^{(1)}$, $\xi_n^{(2)}$, and $\xi_n^{(3)}$. Therefore, in view of (4.33) and (4.34), (4.8) also holds with $\Delta_n = (\log n)^{-1} + 2 \sup_{k\ge n} c_k$.

Let $\lambda_i = \Psi(\gamma_i/(1 + \gamma_i^2)^{1/2}) = E_i Y_1$. Then as shown in Example 1, $\lambda_0 < 0$ and $\lambda_1 > 0$. Hence (4.4) holds. Let $0 < \theta < 1$ such that $D\,\theta < \min\{|\lambda_0|, \lambda_1\}$, where D is given in (4.28). Then by (4.28), for all large n,

$$\begin{aligned} P_1[\max_{j\le n}(\log R_j) > \lambda_1 n + cn^{\delta}] &\le P_1[\max_{\theta n\le j\le n} |\log R_j - \lambda_1 j| > cn^{\delta}], \\ P_0[\min_{j\le n}(\log R_j) < \lambda_0 n - cn^{\delta}] &\le P_0[\max_{\theta n\le j\le n} |\log R_j - \lambda_0 j| > cn^{\delta}]. \end{aligned} \tag{4.36}$$

Since $\delta > \frac{1}{2}$, it follows from (4.30), (4.33), and (4.34) that for $i = 0, 1$,

$$P_i[\max_{\theta n\le m\le n} |\log R_m - \lambda_i m| > cn^{\delta}] = o(n^{-p}) \qquad \text{for all} \quad c > 0 \quad \text{and} \quad p > 0. \tag{4.37}$$

In view of (4.36) and (4.37), the condition (4.9) is satisfied.

Since the assumptions of Theorem 3 are satisfied, it follows from Theorem 3 that the expected sample size of any invariant sequential test of $H_0: \gamma = \gamma_0$ versus $H_1: \gamma = \gamma_1$ satisfies the lower bounds (4.10a) and (4.10b) with ξ given by (4.35). For the sequential t-test $S(A, B)$ with stopping rule τ, it can be shown by using Theorem 3 of [10] that as $A \to \infty$ and $B \to 0$ such that $\log A/|\log B|$ is bounded away from 0 and ∞,

$$E_0\tau = |\lambda_0|^{-1}\{|\log B| + E_0\xi + C_0\} + o(1), \tag{4.38a}$$

$$E_1\tau = \lambda_1^{-1} \{\log A - E_1\xi + C_1\} + o(1), \tag{4.38b}$$

where $C_i > 0$ is the mean of the limiting distribution under H_i of the overshoot $\log (B/R_\tau)$ for $i = 0$ and $\log (R_\tau/A)$ for $i = 1$. Hence, ignoring overshoots, the sequential t-test $S(A, B)$ with the boundaries A and B given by equalities in (1.3) attain the lower bounds in (4.10a) and (4.10b) as $\alpha + \beta \to 0$ such that $|\log \alpha|/|\log \beta|$ is bounded away from 0 and ∞.

By an argument similar to the preceding, it can be shown that Theorem 3 is also applicable to the parametric models of Examples 2 and 3. Thus, ignoring overshoots, the sequential F-test and sequential T^2-test with boundaries given by equalities in (1.3) again attain the lower bounds given by Theorem 3 up to the $o(1)$ term. Since the overshoots are actually not negligible, our results in this paper indicate that the sequential t-test is asymptotically optimal up to the $O(1)$ term within the class $\mathcal{T}(\alpha, \beta)$ of invariant sequential tests. This is also true for the sequential F-test or sequential T^2-test.

It is natural to ask whether these invariant SPRTs are in fact asymptotically optimal up to the $o(1)$ term within the class of invariant sequential tests. Another interesting problem is related to asymptotic expansions for lower bounds of the expected sample size at an intermediate parameter and extensions along the lines of Theorem 3 of Hoeffding's lower bounds [5] in the i.i.d. case. These and other related problems require deeper techniques and will be treated elsewhere.

REFERENCES

[1] ANDERSON, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size. *Ann. Math. Statist.* **31** 165–197.
[2] ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.
[3] BERK, R. H. (1973). Some asymptotic aspects of sequential analysis. *Ann. Statist.* **1** 1126–1138.
[4] GHOSH, B. K. (1970). *Sequential Tests of Statistical Hypotheses.* Addison-Wesley, Reading.
[5] HOEFFDING, W. (1960). Lower bounds for the expected sample size and the average risk of a sequential procedure. *Ann. Math. Statist.* **31** 352–368.
[6] LAI, T. L. (1975a). Termination, moments and exponential boundedness of the stopping rule for certain invariant sequential probability ratio tests. *Ann. Statist.* **3** 581–598.
[7] LAI, T. L. (1975b). On Chernoff-Savage statistics and sequential rank tests. *Ann. Statist.* **3** 825–845.
[8] LAI, T. L. 1975c). A note on first exist times with applications to sequential analysis. *Ann. Statis.* **3** 999–1005.
[9] LAI, T. L. (1976). On r-quick convergence and a conjecture of Strassen. *Ann. Probability* **4** 612–627.
[10] LAI, T. L. AND SIEGMUND, D. (1979). A nonlinear renewal theory with applications to sequential analysis II. *Ann. Statist.* **7** 60–76.
[11] LORDEN, G. (1976). 2-SPRT's and the modified Kiefer-Weiss problem of minimizing the expected sample size. *Ann. Statist.* **4** 281–291.
[12] SACKS, J. (1965). A note on the sequential t-test. *Ann. Math. Statist.* **36** 1867–1869.
[13] SAVAGE, I. R. AND SETHURAMAN, J. (1966). Stopping time of a rank-order sequential probability ratio test based on Lehmann alternatives. *Ann. Math. Statist.* **37** 1154–1160. Correction **38** 1309.
[14] WALD, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Statist.* **16** 117–1186.
[15] WALD, A. AND WOLFOWITZ, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* **19** 326–339.
[16] WIJSMAN, R. A. (1971). Exponentially bounded stopping time of sequential probability ratio tests for composite hypotheses. *Ann. Math. Statist.* **42** 1859–1869.
[17] WIJSMAN, R. A. (1978). Stopping time of invariant sequential probability ratio tests. In *Developments in Statistics* (P. R. Krishnaiah ed.) **2** 235–314.

DEPARTMENT OF MATHEMATICAL STATISTICS
COLUMBIA UNIVERSITY
MATHEMATICS BUILDING
NEW YORK, NEW YORK 10027