SEQUENTIAL ANALYSIS
Vol. 23, No. 4, pp. 585–602, 2004

Sequential Probability Ratio Tests Based on Improper Priors

J. K. Ghosh,[1,*] Sumitra Purkayastha,[2] and Tapas Samanta[3]

[1]Indian Statistical Institute, Kolkata, India and Department of Statistics, Purdue University, West Lafayette, Indiana, USA
[2]Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Kolkata, India
[3]Applied Statistics Unit, Indian Statistical Institute, Kolkata, India

ABSTRACT

One way of handling composite hypotheses by SPRT's is to integrate out nuisance parameters under both the null and alternative hypotheses and then form an SPRT. If the prior is improper, Wald inequalities will not hold in general. We provide sufficient conditions and then verify that the usual sequential $|t|$-test satisfies these conditions but Wald's $|t|$-test does not satisfy these conditions and in fact cannot satisfy a strengthened form of these inequalities in general. Nonetheless, simulations show Wald's inequalities for A, B, corresponding to usual small α, β, will usually hold.

*Correspondence: J. K. Ghosh, Indian Statistical Institute, 203 B.T. Road, Kolkata 700 108, India and Department of Statistics, Purdue University, West Lafayette, IN 47907, USA; E-mail: ghosh@stat.purdue.edu and jayanta@isical.ac.in.

Key Words: Improper priors; Sequential probability ratio test; Sequential t-test; Wald inequalities.

AMS 2000 Subject Classifications: 62L10; 62L99.

1. INTRODUCTION

Sequential Probability Ratio Tests (SPRT's) for simple and composite hypotheses were first introduced by Wald[14,15]. For composite hypotheses, they are based either on integration of the nuisance parameters with respect to priors (Wald[15]), or elimination of nuisance parameters via invariance and sufficiency (Hall et al.[10]). In many cases, the priors, conditional on the hypotheses $H_i, i = 0, 1$, are improper, making the Wald inequalities somewhat doubtful in general. We explore this question in detail, specially in the context of sequential two-sided t-test. The corresponding ratio of integrated likelihoods depends on $|t|$, and hence we call this test the sequential $|t|$-test.

We first prove a general result that ensures the validity of Wald inequalities, and then examine the following special case. Suppose X_i's are independent and identically distributed (i.i.d.) as $N(\mu, \sigma^2)$, the null hypothesis is $H_0 : \mu = 0$ and the alternative is $H_1 : \mu/\sigma = \pm\Delta$ for some specified Δ. Slightly extending Hoel[11], we consider the family of priors under H_i:

$$\pi(\sigma) = \frac{1}{\sigma^\gamma}, \quad 0 < \sigma < \infty \, (0 \leq \gamma \leq 1),$$

$\mu = 0$ under H_0, and μ/σ given σ, equals $\Delta, -\Delta$ with probability half under H_1. They all lead to sequential $|t|$-tests in the sense that the likelihood ratio (defined in (2.1) and (2.4) below) at nth stage is a monotone function of the corresponding $|t|$ statistic. In the rest of the paper, unless stated otherwise, by "likelihood ratio" we shall mean "ratio of integrated likelihoods".

If $\gamma = 1$, we get the usual sequential $|t|$-test, often called the WAGR sequential test, vide Barnard[1], Rushton[13], Ghosh[6]. The validity of this test is proved rigorously in Barnard[2] via convergence arguments. Its validity is also proved in Hall et al.[10]. It is possible to show in this case that the ratio of integrated likelihoods at nth stage is actually the likelihood ratio of $|t|$ with $df = n - 1$, the numerator involving an appropriate non-centrality parameter. Details are given in Sec. 2. We show our general theorem applies in this case.

If $\gamma = 0$, we get the sequential $|t|$-test proposed by Wald. One of the reasons for writing this note is to settle whether the famous Wald[14,15] inequalities hold for Wald's test. We show that our general theorem does not hold. In fact, the likelihood ratios with approximating proper priors do not converge to the limiting likelihood ratio of Wald under the sequence of approximating priors. This explains why a limiting argument cannot hold. In principle, Wald inequalities can still hold. We prove a general result on likelihood ratios and show that a certain strengthened version of Wald inequalities (see (2.11)) which is an essential step in Wald's argument cannot hold.

These arguments are quite general and apply to all tests with

$$\pi(\sigma) = \frac{1}{\sigma^\gamma}, \quad 0 < \gamma < 1.$$

Incidentally, our results make it clear that Hoel's[11] concern about these tests was justified. They also show why a rigorous proof or at least numerical support from simulations is needed in Statistics in spite of it being an applied discipline, see in this context the interesting discussion in Barnard[1,2]. In view of these facts, the claim in Ghosh[7] is not true for Wald's test; the claim is true for the usual test with $\gamma = 1$. It is pointed out in Ghosh[7] that the Wald inequalities only lead to approximations for error probabilities and hence Hoel's[11] argument does not really apply. This point remains true, the so-called (approximate) double minimax property defined in Hoel[11] holds for the WAGR test, and any other test for which the Wald approximations hold. In this connection, it is worth pointing out that the prior associated with the usual sequential $|t|$-test avoids the marginalization paradox (vide Dawid et al.[5]) but the other priors may not.

It is our finding that notwithstanding these theoretical discussions for Wald's test, Wald's inequalities hold for all the cases where we have undertaken simulations so far. This may be because the boundaries of this test are close to the boundaries of the test with $\gamma = 1$ and Wald inequalities are quite conservative for the latter. Results of some simulations are presented.

It is worth mentioning that all these tests terminate with probability one, at least under H_0 and H_1. This can be verified by employing Theorem 2.5 of Berk[3]. Also all of them have monotonically decreasing error of second kind as $|\mu|/\sigma$ increases, vide Ghosh[7,8].

Wald was one of the greatest statisticians of the twentieth century. He packed in a period of ten to fifteen years what would have been for lesser people the work of a lifetime. Had his life not been cut short by the tragic accident in India, he would have himself settled the questions relating to use of improper priors. Incidentally, for a fascinating description of Wald's visit to the Indian Statistical Institute and the problems caused by improper priors in another area see the autobiographical account of D. Basu in Ghosh et al.[9].

2. RESULTS

Let $X_1, X_2, \ldots$ be i.i.d. observations from a distribution with density $f(x, \theta)$. We want to test the hypothesis $H_0 : \theta \in \Theta_0$ against the alternative $H_1 : \theta \in \Theta_1$. Let $\pi_j(\theta)$ be improper noninformative prior density of θ under H_j and Π_j be the corresponding measure, $j = 0, 1$.

Let

$$\lambda_n = \frac{p_{1n}}{p_{0n}}, \quad n \geq 1, \tag{2.1}$$

where

$$p_{jn} \equiv p_{jn}(X_1, \ldots, X_n) \stackrel{def}{=} \int_{\Theta_j} \prod_{i=1}^{n} f(X_i, \theta)\pi_j(\theta)d\theta, \quad j = 0, 1, \; n \geq 1.$$

Note that p_{1n} and p_{0n} are not proper probability density functions. Consider Wald's SPRT with boundaries $B < A$ and likelihood ratio $\lambda_n(X_1, \ldots, X_n)$ at stage $n \geq 1$. Let N be the stopping time, i.e., $N \stackrel{def}{=} \min\{n \geq 1 : \lambda_n \notin (B, A)\}$.

We consider a sequence of sets $\Theta_{jM} \uparrow \Theta_j$ as $M \to \infty$ such that $\Pi_j(\Theta_{jM}) < \infty$ for all M, $j = 0, 1$, and the proper likelihood ratios based on the approximating proper priors $\pi_j(\theta)/\Pi_j(\Theta_{jM})$ on Θ_{jM}:

$$\lambda_{nM} = \frac{p_{1nM}}{p_{0nM}}, \quad n \geq 1, \tag{2.2}$$

where

$$p_{jnM} \equiv p_{jnM}(X_1, \ldots, X_n) \stackrel{def}{=} \frac{\int_{\Theta_{jM}} \prod_{i=1}^{n} f(X_i, \theta)\pi_j(\theta)d\theta}{\Pi_j(\Theta_{jM})},$$
$$j = 0, 1, \quad n \geq 1, \quad M \geq 1.$$

The functions $p_{jnM}, j = 0, 1$, are proper probability density functions. Let $P_{jnM}, j = 0, 1$, be the corresponding probability distributions.

We also assume that $\Pi_0(\Theta_{0M}) = \Pi_1(\Theta_{1M})$ for all M so that

$$\lambda_{nM} = \frac{\int_{\Theta_{1M}} \prod_{i=1}^{n} f(X_i, \theta)\pi_1(\theta)d\theta}{\int_{\Theta_{0M}} \prod_{i=1}^{n} f(X_i, \theta)\pi_0(\theta)d\theta}.$$

The basic idea is as follows. Since p_{1n} and p_{0n} are not proper probability density functions, we cannot verify directly whether the Wald inequalities hold for the ratio λ_n. We, therefore, try to approximate λ_n by the proper likelihood ratios λ_{nM} which are based on the approximating proper priors $\pi_j(\theta)/\Pi_j(\Theta_{jM})$ on Θ_{jM}. The Wald inequalities will hold with these likelihood ratios λ_{nM}. Thus, when λ_{nM} approximates λ_n well, the Wald inequalities for λ_n are also expected to hold. In other words, approximation of λ_n by λ_{nM} is the key issue. This issue is explored in detail.

We state below (in Theorem 2.1) a version of the Wald inequalities.

Let P_{0M} and P_{1M} be the probabilities on $\mathbb{R}^\infty$ defined by the finite dimensional distributions $\{P_{0nM} : n \geq 1\}$ and $\{P_{1nM} : n \geq 1\}$, respectively. We assume that

(A) for any $n \geq 2$, the distribution of $(\lambda_2, \ldots, \lambda_n)$ under $P_{jnM}(j = 0, 1)$ converges weakly to some continuous $(n-1)$-dimensional distribution as $M \to \infty$.

This collection of finite dimensional limiting distributions defines a unique probability distribution, denoted P_{H_j}, on $\mathbb{R}^\infty$ which is the limit of the distribution of $(\lambda_2, \lambda_3, \ldots)$ under P_{jM} as $M \to \infty$. Note that P_{H_j} is a probability on the space of $(\lambda_2, \lambda_3, \ldots)$ and the event that H_0 is accepted (or rejected) can be described in terms of $(\lambda_2, \lambda_3, \ldots)$. We define the event H_0 is accepted (rejected) as $\{N < \infty, H_0 \text{ is accepted (rejected)}\}$. If $N < \infty$ with probability one under both hypotheses, then we can drop "$N < \infty$".

Theorem 2.1. *Suppose that for each $n \geq 2$, $\lambda_{nM} \xrightarrow{P_{jnM}} \lambda_n$ as $M \to \infty$, i.e., for every $\epsilon > 0$, $P_{jnM}(|\lambda_{nM} - \lambda_n| > \epsilon) \longrightarrow 0$ as $M \to \infty$, for $j = 0, 1$, and that Assumption* (A) *above holds. Let $P_{H_j}, j = 0, 1$, be the limiting probabilities as defined above. Then for Wald's* SPRT *with boundaries A and B and likelihood ratios $\{\lambda_n : n \geq 2\}$, we have*

(a) $P_{H_1}(H_0$ *is accepted*$) \leq B\, P_{H_0}(H_0$ *is accepted*$)$, *and*
(b) $P_{H_1}(H_0$ *is rejected*$) \geq A\, P_{H_0}(H_0$ *is rejected*$)$.

Proof. We prove part (a) only, the proof of part (b) being similar.

Note that the event H_0 is accepted can be written as a countable disjoint union as follows:

$$\{H_0 \text{ is accepted}\} = \bigcup_{n\geq 2}\{B < \lambda_2 < A, \ldots, B < \lambda_{n-1} < A, \lambda_n \leq B\}.$$

Under the assumptions of the theorem,

$$\begin{aligned} &P_{jnM}(B < \lambda_2 < A, \ldots, B < \lambda_{n-1} < A, \lambda_n \leq B) \\ &\quad - P_{jnM}(B < \lambda_{2M} < A, \ldots, B < \lambda_{n-1,M} < A, \lambda_{n,M} \leq B) \longrightarrow 0. \\ &\qquad \text{as } M \to \infty 0. \end{aligned} \tag{2.3}$$

The proof of (2.3) follows from the definition of weak convergence (see, for example, Theorem 25.4 of Billingsley[4]).

Since $\lambda_{nM} = p_{1nM}/p_{0nM}$ is a proper likelihood ratio,

$$\begin{aligned} &P_{1nM}(B < \lambda_{2M} < A, \ldots, B < \lambda_{n-1,M} < A, \lambda_{n,M} \leq B) \\ &\quad \leq BP_{0nM}(B < \lambda_{2M} < A, \ldots, B < \lambda_{n-1,M} < A, \lambda_{n,M} \leq B). \end{aligned}$$

Therefore, from (2.3),

$$\begin{aligned} &\sum_{n=2}^{\infty} \lim_{M\to\infty} P_{1nM}(B < \lambda_2 < A, \ldots, B < \lambda_{n-1} < A, \lambda_n \leq B) \\ &\quad \leq B \sum_{n=2}^{\infty} \lim_{M\to\infty} P_{0nM}(B < \lambda_2 < A, \ldots, B < \lambda_{n-1} < A, \lambda_n \leq B). \end{aligned}$$

This completes the proof of Theorem 2.1. □

We now consider the special case where the observations X_i's are random samples from a $N(\mu, \sigma^2)$ population with μ and σ^2 both unknown. The null hypothesis to be tested and the alternative are respectively

$$H_0 : \mu = 0 \quad \text{and} \quad H_1 : \mu/\sigma = \pm\Delta$$

for some specified Δ. We consider the family of priors

$$\pi(\sigma) = \frac{1}{\sigma^\gamma}, \quad 0 < \sigma < \infty \, (0 \leq \gamma \leq 1)$$

under both H_0 and H_1, $\mu = 0$ under H_0, and μ/σ given σ, equals Δ and $-\Delta$ with probability half under H_1. Note that $\pi(\sigma)$ is an improper prior for σ. We consider approximating proper priors $\pi(\cdot)$ restricted to (a_M, b_M), where a_M, b_M are positive numbers, $a_M \to 0$ and $b_M \to \infty$ as $M \to \infty$. When $0 \leq \gamma < 1$, we may, however, take $a_M = 0$.

With $p_{jn}, p_{jnM} (j = 0, 1)$ as defined above, we then have

$$\begin{aligned}
p_{0n} &= \int_0^\infty f(X_1, \ldots, X_n; 0, \sigma)\frac{1}{\sigma^\gamma}d\sigma, \\
p_{1n} &= \frac{1}{2}\left[\int_0^\infty f(X_1, \ldots, X_n; \Delta\sigma, \sigma)\frac{1}{\sigma^\gamma}d\sigma \right.\\
&\qquad \left. + \int_0^\infty f(X_1, \ldots, X_n; -\Delta\sigma, \sigma)\frac{1}{\sigma^\gamma}d\sigma\right], \\
p_{0nM} &= C_M \int_{a_M}^{b_M} f(X_1, \ldots, X_n; 0, \sigma)\frac{1}{\sigma^\gamma}d\sigma, \\
p_{1nM} &= \frac{C_M}{2}\left[\int_{a_M}^{b_M} f(X_1, \ldots, X_n; \Delta\sigma, \sigma)\frac{1}{\sigma^\gamma}d\sigma \right.\\
&\qquad \left. + \int_{a_M}^{b_M} f(X_1, \ldots, X_n; -\Delta\sigma, \sigma)\frac{1}{\sigma^\gamma}d\sigma\right],
\end{aligned} \tag{2.4}$$

where

$$f(X_1, \ldots, X_n; \mu, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - \mu)^2\right)$$

and

$$C_M^{-1} = \int_{a_M}^{b_M} \frac{1}{\sigma^\gamma}d\sigma.$$

Notice that for any constant δ,

$$\int_0^\infty f(X_1, \ldots, X_n; \delta\sigma, \sigma)\frac{1}{\sigma^\gamma}d\sigma = \frac{1}{(\sqrt{2\pi})^n s^{n+\gamma-1}}\int_0^\infty g_\gamma(u, t; \delta)du, \tag{2.5}$$

where

$$(n-1)s^2 = \sum_{i=1}^n (X_i - \overline{X})^2, \quad \overline{X} = \frac{1}{n}\sum_{i=1}^n X_i, \quad t = \frac{\sqrt{n}\overline{X}}{s},$$

and

$$g_\gamma(u,t;\delta) \stackrel{def}{=} u^{n+\gamma-2} \exp\left[-\frac{1}{2}\{(n-1)u^2 + (tu - \delta\sqrt{n})^2\}\right], \quad u > 0. \tag{2.6}$$

Recall now that the expression for the density function of non-central t-distribution $T(p,\theta)$ with degrees of freedom p and non-centrality parameter θ is given by

$$f(t;p,\theta) \stackrel{def}{=} \frac{p^{p/2}}{\sqrt{2\pi}2^{\frac{p}{2}-1}\Gamma(\frac{p}{2})} \int_0^\infty u^p \exp\left\{-\frac{1}{2}\left[pu^2 + (tu-\theta)^2\right]\right\} du,$$
$$t \in \mathbb{R}.$$

Hence, when $\gamma = 1$, except for the leading constant the expression appearing on the right hand side of (2.5) coincides with $f(t; n-1, \delta\sqrt{n})$. Thus when $\gamma = 1$,

$$\lambda_n = \frac{f(t; n-1, \Delta\sqrt{n}) + f(t; n-1, -\Delta\sqrt{n})}{2f(t; n-1, 0)}$$
$$= \frac{f(t; n-1, \Delta\sqrt{n}) + f(-t; n-1, \Delta\sqrt{n})}{2f(t; n-1, 0)},$$

and hence

$$\lambda_n = \frac{\text{density of } |T(n-1, \Delta\sqrt{n})| \text{ evaluated at } |t|}{\text{density of } |T(n-1, 0)| \text{ evaluated at } |t|}.$$

We first consider the case with $\gamma = 1$.

Proposition 2.1. *If $\gamma = 1$, then for any fixed $n \geq 2, \lambda_{nM} \xrightarrow{P_{jnM}} \lambda_n$ as $M \to \infty$, $j = 0, 1$, i.e., for every $\epsilon > 0, P_{jnM}(|\lambda_{nM} - \lambda_n| > \epsilon) \longrightarrow 0$ as $M \to \infty$.*

To prove this proposition, we need the following lemma.

Lemma 2.1. *Let s be the sample standard deviation defined as $(n-1)s^2 = \sum_{i=1}^n (X_i - \overline{X})^2$ with $\overline{X} \stackrel{def}{=} \frac{1}{n}\sum_{i=1}^n X_i$. Then $s/b_M \xrightarrow{P_{jnM}} 0$ and $a_M/s \xrightarrow{P_{jnM}} 0$ as $M \to \infty, j = 0, 1$.*

Proof. All the probability statements in the proof refer to the joint distribution of $(X_1, \ldots, X_n)$ and (μ, σ).

We choose $K > 0$ such that $Pr(s/\sigma < K)$ can be made sufficiently close to 1. This is possible as distribution of s/σ, given (μ, σ), is free of

(μ, σ). Also, for any $\epsilon > 0$,

$$Pr(\sigma/b_M < \epsilon) = C_M \int_{a_M}^{\epsilon b_M} \frac{1}{\sigma} d\sigma = \frac{\log(\epsilon b_M) - \log(a_M)}{\log b_M - \log a_M} \to 1$$

as $M \to \infty$. Therefore, $s/b_M = \frac{\sigma}{b_M} \cdot \frac{s}{\sigma} \xrightarrow{P_{jnM}} 0$ as $M \to \infty$.

To prove the second part of the lemma, we choose $K' > 0$ such that $Pr(\sigma/s < K')$ is sufficiently close to 1. Also, for any $\epsilon > 0$,

$$Pr(a_M/\sigma < \epsilon) = \frac{\log(b_M) - \log(a_M/\epsilon)}{\log b_M - \log a_M} \to 1$$

as $M \to \infty$. Then $a_M/s = \frac{\sigma}{s} \cdot \frac{a_M}{\sigma} \xrightarrow{P_{jnM}} 0$ as $M \to \infty$. □

We also need the following result. We omit the proof.

Lemma 2.2. *Let $U, V, U_m, V_m, m \geq 1$, be random variables defined on a measurable space $(\Omega, \mathscr{A})$ and P_m, $m \geq 1$, be a sequence of probability measures on $(\Omega, \mathscr{A})$ such that $U_m \xrightarrow{P_m} U$ and $V_m \xrightarrow{P_m} V$ as $m \to \infty$. If $\{U_m\}$ and $\{V_m\}$ are stochastically bounded and $\{V_m\}$ is stochastically bounded away from zero under $\{P_m\}$, then*

$$\left|\frac{U_m}{V_m} - \frac{U}{V}\right| \xrightarrow{P_m} 0 \quad \text{as } m \to \infty.$$

Proof of Proposition 2.1. It is easy to verify that for $0 \leq a < b \leq \infty$, and any constant δ,

$$\int_a^b f(X_1, \ldots, X_n; \delta\sigma, \sigma) \frac{1}{\sigma} d\sigma = \frac{1}{(\sqrt{2\pi} s)^n} \int_{s/b}^{s/a} g_1(u, t; \delta) du$$

(see (2.5) and (2.6)). From (2.1), (2.2), and (2.4), we have

$$\lambda_{nM} = \frac{\int_{s/b_M}^{s/a_M} g_1(u, t; \Delta) du + \int_{s/b_M}^{s/a_M} g_1(u, t; -\Delta) du}{2 \int_{s/b_M}^{s/a_M} g_1(u, t; 0) du}.$$

and

$$\lambda_n = \frac{\int_0^\infty g_1(u, t; \Delta) du + \int_0^\infty g_1(u, t; -\Delta) du}{2 \int_0^\infty g_1(u, t; 0) du}.$$

Then

$$|\text{numerator of } \lambda_{nM} - \text{numerator of } \lambda_n|$$
$$\leq 2\int_0^{s/b_M} u^{n-1}\exp\left[-\frac{1}{2}(n-1)u^2\right]du$$
$$+2\int_{s/a_M}^{\infty} u^{n-1}\exp\left[-\frac{1}{2}(n-1)u^2\right]du, \quad (2.7)$$

which converges to zero under P_{jnM} probability ($j = 0, 1$) as $M \to \infty$ in view of Lemma 2.1. The difference between the denominators of λ_{nM} and λ_n also has the same bound as (2.7) and hence converges to zero under P_{jnM} probability.

Now note that we may choose $K > 0$ such that $P_{jnM}(|t| < K)$ can be made sufficiently close to 1. Then by Lemma 2.1,

$$\text{denominator of } \lambda_{nM} = 2\int_{s/b_M}^{s/a_M} g_1(u, t; 0)du > 2\int_1^2 g_1(u, K; 0)du$$

with P_{jnM}-probability sufficiently close to 1. Therefore, the denominator of λ_{nM} is stochastically bounded away from zero. Also, both numerator and denominator of λ_{nM} or λ_n are bounded by

$$2\int_0^{\infty} u^{n-1}\exp\left[-\frac{1}{2}(n-1)u^2\right]du$$

which is finite. The result now follows by an application of Lemma 2.2. □

We now note that Assumption (A) of Theorem 2.1 holds as distribution of $(\lambda_2, \ldots, \lambda_n)$ under P_{jnM} does not depend on $M (j = 0, 1)$. Thus the Wald inequalities as stated in Theorem 2.1 hold.

We now consider the case $0 \leq \gamma < 1$. The case $\gamma = 0$ corresponds to the test proposed by Wald[15]. In this case, λ_{nM} does not converge (in probability) to λ_n as $M \to \infty$. In what follows, we show this only under P_{0nM}-probability. Here we take $a_M = 0 \; \forall M$ and λ_{nM} and λ_n can be written as

$$\lambda_{nM} = F(s/b_M, t), \quad \lambda_n = F(0, t), \quad (2.8)$$

where

$$F(w,t) = \frac{\int_w^\infty g_\gamma(u,t;\Delta)du + \int_w^\infty g_\gamma(u,t;-\Delta)du}{2\int_w^\infty g_\gamma(u,t;0)du}, \quad w > 0,$$

and $g_\gamma(u,t;\delta)$ is defined in (2.6). All the probability statements below refer to the joint distribution of $(X_1,\ldots,X_n)$ and σ, where σ has density $\propto 1/\sigma^\gamma$, $0 < \sigma < b_M$.

Lemma 2.3. *The distribution of $(s/b_M, t)$ is free of M.*

Proof. We write $(\frac{s}{b_M}, t)$ as $(\frac{s}{\sigma}\cdot\frac{\sigma}{b_M}, t)$. Note that $(s/\sigma, t)$ is a function of $Y_i := X_i/\sigma, i = 1,\ldots,n$, which are i.i.d. $N(0,1)$ variables, given σ. Therefore, unconditional distribution of $(s/\sigma, t)$ is free of M and $(s/\sigma, t)$ is stochastically independent of σ/b_M. Also, σ/b_M has a distribution free of M which is given by

$$Pr(\sigma/b_M \le v) = \frac{1-\gamma}{b_M^{1-\gamma}}\int_0^{vb_M}\frac{1}{\sigma^\gamma}d\sigma = v^{1-\gamma}, \quad 0 \le v \le 1.$$

This proves the result. □

Proposition 2.2. *λ_{nM} does not converge in probability (under both P_{0nM} and P_{1nM}) to λ_n as $M \to \infty$.*

Proof. In what follows, we show this only under P_{0nM}-probability.

Suppose, if possible,

$$\lambda_{nM} \xrightarrow{P_{0nM}} \lambda_n \quad \text{as } M \to \infty. \tag{2.9}$$

Then by Lemma 2.3, writing $s/b_M = \eta$, we have from (2.8)

$$F(\eta,t) = F(0,t) \text{ with probability } 1,$$

i.e., with a slight abuse of notation,

$$F(\eta,t) = F(0,t) \quad \text{a.e.} \quad (\eta,t), \tag{2.10}$$

under the joint distribution of (η,t).

One can find the joint density of (η,t) as indicated in the proof of Lemma 2.3 and check that it is supported on the whole of $\mathbb{R}^+ \times \mathbb{R}$.

Therefore, using continuity of $F(\eta, t)$ as a function of (η, t), we have from (2.10)

$$F(\eta, t) = F(0, t), \quad 0 < \eta < \infty, \quad -\infty < t < \infty,$$

i.e.,

$$\int_{\eta}^{\infty} [g_\gamma(u, t; \Delta) + g_\gamma(u, t; -\Delta)]\, du$$
$$= 2 \int_{\eta}^{\infty} g_\gamma(u, t; 0) du \times F(0, t) \quad \forall\ (\eta, t).$$

Differentiating both sides of this identity with respect to η and simplifying, we have

$$\exp[\eta t \Delta \sqrt{n}] + \exp[-\eta t \Delta \sqrt{n}] = 2F(0, t) \exp[\Delta^2 n/2] \quad \forall\ (\eta, t).$$

This cannot be true as the right-hand side is free of η.

Hence, (2.9) does not hold. □

We have proved that the assumption

$$\lambda_{nM} \xrightarrow{P_{jnM}} \lambda_n \quad \text{as } M \to \infty, \quad n \geq 2,$$

of Theorem 2.1 does not hold when $\gamma < 1$. It is natural to ask whether the Wald inequalities can still hold even though Theorem 2.1 cannot be applied. Interpreting the Wald inequalities in the stronger sense of (2.11) below, we now show the answer is no.

For $0 \leq \gamma \leq 1$, let us denote the likelihood ratio (2.1) by $\lambda_{n,\gamma}$. Let P_i denote probability under H_i, and Q_i be the distribution of $\lambda_{n,\gamma}$ under $P_i, i = 0, 1$.

Wald's inequalities are based on the inequalities of the form

$$B P_0(E \cap \{B \leq \lambda_{n,\gamma} \leq A\}) \leq P_1(E \cap \{B \leq \lambda_{n,\gamma} \leq A\})$$
$$\leq A P_0(E \cap \{B \leq \lambda_{n,\gamma} \leq A\}) \qquad (2.11)$$

for any $-\infty \leq B < A \leq \infty$ and $n \geq 1$, where E is an event expressible in terms of $X_1, \ldots, X_n$.

Suppose that the following weaker version of (2.11) holds: for any $B < A$,

$$B P_0(B \le \lambda_{n,\gamma} \le A) \le P_1(B \le \lambda_{n,\gamma} \le A) \le A P_0(B \le \lambda_{n,\gamma} \le A). \tag{2.12}$$

We show below that (2.12) cannot hold for $0 \le \gamma < 1$.
First, we show that

$$\frac{dQ_1}{dQ_0}(x) = x.$$

Here Q_0, Q_1 are measures on $(0, \infty)$. We define a martingale sequence $\{Z_m, \mathscr{F}_m\}_{m \ge 1}$ on $(0, \infty)$ as follows. Let

$$\Omega = (0, \infty) = \bigcup_{r=0}^{\infty} (r, r+1].$$

Let $\mathscr{F}_m (m \ge 1)$ be the σ-field generated by the partition of Ω into dyadic intervals $(r + k2^{-m}, r + (k+1)2^{-m}]$, $0 \le k < 2^m$, $r \ge 0$ and Z_m, $m \ge 1$, are $\mathscr{F}_m$-measurable random variables defined as

$$Z_m(x) = \frac{Q_1(r + k2^{-m}, r + (k+1)2^{-m}]}{Q_0(r + k2^{-m}, r + (k+1)2^{-m}]} \quad \text{if } x \in (r + k2^{-m}, r + (k+1)2^{-m}]. \tag{2.13}$$

Then Z_m is the Radon–Nikodym derivative of Q_1 with respect to Q_0 on $\mathscr{F}_m$ and the sequence $\{Z_m : m \ge 1\}$ is martingale relative to the σ-fields $\{\mathscr{F}_m : m \ge 1\}$ (see, for example, Billingsley[4], Sec. 35).

By Theorem 35.8 of Billingsley[4], with probability one

$$Z_m \longrightarrow \frac{dQ_1}{dQ_0}.$$

For each $x > 0$ and $m \ge 1$, we choose dyadic rationals $a_m(x) = r + k2^{-m}$ and $b_m(x) = r + (k+1)2^{-m}$ such that $a_m(x) < x \le b_m(x)$. Then, if (2.12) holds, it follows from (2.13) that for each $x > 0$,

$$a_m(x) \le Z_m(x) \le b_m(x)$$

implying that $Z_m(x) \to x$ as $m \to \infty$. Thus

$$\frac{dQ_1}{dQ_0}(x) = x.$$

Now for $B < A$,

$$\begin{aligned} P_1(B \le \lambda_{n,\gamma} \le A) &= \int_{\{B \le \lambda_{n,\gamma} \le A\}} dP_1 = \int_{\{B \le x \le A\}} dQ_1(x) \\ &= \int_{\{B \le x \le A\}} x\, dQ_0(x) = \int_{\{B \le \lambda_{n,\gamma} \le A\}} \lambda_{n,\gamma}\, dP_0. \end{aligned}$$

For any $b < a$,

$$\{b \le |t| \le a\} = \{B \le \lambda_{n,\gamma} \le A\}$$

for some B, A as $\lambda_{n,\gamma}$ is a strictly increasing function of $|t|$. This property of $\lambda_{n,\gamma}$ follows from the MLR property of non-central $|t|$, which is a special case of non-central F (Lehmann[12]). Therefore,

$$P_1(\{b \le |t| \le a\}) = \int_{\{b \le |t| \le a\}} \lambda_{n,\gamma}\, dP_0,$$

and hence

$$P_1(\{b \le |t| \le a\}) = \int_{\{b \le |t| \le a\}} \lambda_{n,\gamma}\, dP_0 = \int_{\{b \le |t| \le a\}} \lambda_{n,1}\, dP_0.$$

Then by a monotone class argument

$$\int_S \lambda_{n,\gamma}\, dP_0 = \int_S \lambda_{n,1}\, dP_0 \quad \forall \text{ Borel subsets } S \text{ of } (0, \infty).$$

This implies

$$\lambda_{n,\gamma} = \lambda_{n,1} \quad \text{a.s.} \tag{2.14}$$

with respect to the distribution of $|t|$ under P_0. However, by using property of confluent hypergeometric function it is possible to show the following (Hoel[11]): for $0 \leq \gamma < 1$,

$$\lambda_{n,\gamma}(t) < \lambda_{n,1}(t), \quad t > 0. \tag{2.15}$$

Hence, (2.14) cannot hold. This proves in turn that (2.12) cannot hold.

3. SIMULATIONS

For an SPRT with boundaries $A > B$ and probabilities of errors of first and second kind α, β, the following inequalities hold:

$$\beta \leq B(1-\alpha), \quad A\alpha \leq (1-\beta)$$

(vide (a) and (b) of Theorem 2.1). These are referred to as Wald inequalities. Wald gave an argument that in many cases these inequalities become approximate equalities. Then given α, β, one can find the boundaries approximately. These inequalities and approximations play a fundamental role in Wald's[15] book.

We investigate through simulation whether Wald's inequalities hold for Wald's sequential $|t|$-test, and also demonstrate validity of these inequalities for the WAGR sequential test, theoretical support of which has already been provided (see Theorem 2.1).

Consider i.i.d. observations X_i's from a $N(\mu, \sigma^2)$ population with μ and σ^2 both unknown. The null hypothesis to be tested and the alternative are respectively

$$H_0 : \mu = 0 \quad \text{and} \quad H_1 : \mu/\sigma = \pm\Delta$$

for some specified Δ. We consider the family of priors $\pi(\sigma) = 1/\sigma^\gamma$, $0 < \sigma < \infty$ $(0 \leq \gamma \leq 1)$ under both H_0 and H_1, $\mu = 0$ under H_0, and μ/σ given σ equals Δ and $-\Delta$ with probability half under H_1. The expression for the "likelihood ratio" appears in (2.8).

We present in the table below $P_{H_1}(H_0$ is accepted), $P_{H_0}(H_0$ is accepted), $B\,P_{H_0}(H_0$ is accepted), $P_{H_1}(H_0$ is rejected), $P_{H_0}(H_0$ is rejected), and $A\,P_{H_0}(H_0$ is rejected), for varying choices of the target error probabilities $\alpha^\star, \beta^\star$ with $\alpha^\star = \beta^\star$, and Δ. We denote the common value of $\alpha^\star$ and $\beta^\star$ by τ. Also, we report our findings both for $\gamma = 1$ and $\gamma = 0$, the former corresponding to the WAGR sequential test and the latter to Wald's sequential $|t|$-test. The column indices appear in the first row.

It is easy to verify in this situation that

$$\begin{aligned}
P_{H_0}(H_0 \text{ is accepted}) &= P_{0,1}(\lambda_N \leq B),\\
P_{H_0}(H_0 \text{ is rejected}) &= P_{0,1}(\lambda_N \geq A),\\
P_{H_1}(H_0 \text{ is accepted}) &= \frac{1}{2}[P_{\Delta,1}(\lambda_N \leq B) + P_{-\Delta,1}(\lambda_N \leq B)]\\
&= P_{\Delta,1}(\lambda_N \leq B),\\
P_{H_1}(H_0 \text{ is rejected}) &= \frac{1}{2}[P_{\Delta,1}(\lambda_N \geq A) + P_{-\Delta,1}(\lambda_N \geq A)]\\
&= P_{\Delta,1}(\lambda_N \geq A),
\end{aligned} \tag{3.1}$$

where $P_{\theta,1}(C)$ denotes the probability of the event C when X_i's are i.i.d. observations from a $N(\theta, 1)$ population.

The set of equalities in (3.1) enables us to calculate easily the quantities reported in columns 6–11 of the table. In all the calculations below, $A = (1-\beta^\star)/\alpha^\star = (1-\tau)/\tau B = \beta^\star/(1-\alpha^\star) = \tau/(1-\tau)$. Also, we denote the events $\{H_0$ is accepted$\}$ and $\{H_0$ is rejected$\}$ by D and E, respectively. Hence, $P_{H_0}(E)$ and $P_{H_1}(D)$ are the error probabilities. They are reported in columns 10 and 6 of the table, respectively.

The simulations have been done by two different programmes for calculating the integrals appearing in λ_n. One of these methods calculates them by a recursive method, and the other employs numerical integration. The results match substantially. We report the results obtained by the first method. A few words about this method are in order.

Calculation of the integrals appearing in λ_n can be shown to boil down to calculation of integrals of the form

$$\int_0^\infty x^m \exp\left(-\frac{(x-\theta)^2}{2\tau^2}\right)dx =: I_m, \quad m = 0, 1, 2, \ldots. \tag{3.2}$$

It can be proved that

$$\begin{aligned}
I_0 &= \sqrt{2\pi}\tau\Phi\left(\frac{\theta}{\tau}\right),\\
I_1 &= \tau^2 \exp\left(-\frac{\theta^2}{2\tau^2}\right) + \theta I_0,\\
I_{m+2} &= \theta I_{m+1} + (m+1)\tau^2 I_m, \quad m \geq 0,
\end{aligned} \tag{3.3}$$

where Φ is the cdf of the standard normal distribution. These, i.e., the equalities in (3.3), can be used to evaluate I_m for every m.

Finally, the number of simulations for each entry is 25,000.

1	2	3	4	5	6	7	8	9	10	11
γ	Δ	τ	A	B	$P_{H_1}(D)$	$P_{H_0}(D)$	$BP_{H_0}(D)$	$P_{H_1}(E)$	$P_{H_0}(E)$	$AP_{H_0}(E)$
1	0.5	0.05	19	1/19	0.0408	0.9581	0.0504	0.9592	0.0419	0.7957
		0.01	99	1/99	0.0087	0.9924	0.0100	0.9913	0.0076	0.7488
0	0.5	0.05	19	1/19	0.0432	0.9627	0.0507	0.9568	0.0373	0.7091
		0.01	99	1/99	0.0094	0.9939	0.0100	0.9906	0.0061	0.6059

For each row, we need to compare entries of columns 6 and 8, and also of columns 9 and 11 (cf. Theorem 2.1). It is seen that Wald's inequalities hold for all the cases we have considered, for both the WAGR sequential test and Wald's sequential $|t|$-test.

Similar studies were also made in other cases, namely, for $\Delta = 0.1, 0.5, 1, 2$ and small, moderate (=0.5) and large (close to 1) values of B (=$1/A$). In all cases Wald's inequalities hold for Wald's $|t|$-test ($\gamma = 0$). So it remains open whether there are any α, β, Δ for which Wald's inequalities are violated. Incidentally, using (2.15), it can be shown that one of the two Wald inequalities, namely, part (b) of Theorem 2.1 will hold for Wald's $|t|$-test.

Finally, we report a simulation indicating violation of the strengthened version (2.12) for the above test.

In (2.12), we take $n = 4, \gamma = 0, B = 0.98, A = 1.02, \Delta = 0.5$, and obtain the first, second and third terms of the inequalities as 0.0268, 0.0320 and 0.0279, respectively, based on 100,000 simulations.

ACKNOWLEDGMENTS

We thank Mr. Nilanjan Mukhopadhyay for his help in computation. We are grateful also to two referees for their careful reading of the earlier draft of the paper.

REFERENCES

1. Barnard, G.A. The frequency justification of certain sequential tests. Biometrika **1952**, *39*, 144–150.
2. Barnard, G.A. The frequency justification of certain sequential tests – addendum. Biometrika **1953**, *40*, 468–469.
3. Berk, R.H. Stopping times of SPRT's based on exchangeable models. Ann. Math. Statist. **1970**, *41*, 979–990.
4. Billingsley, P. *Probability and Measure*, 2nd Ed.; Wiley: New York, 1986.
5. Dawid, A.P.; Stone, M.; Zidek, J.V. Marginalization paradoxes in Bayesian and structural inference (with discussion and a reply by the authors). J. Roy. Statist. Soc. Ser. B **1973**, *35*, 189–233.
6. Ghosh, B.K. *Sequential Tests of Statistical Hypotheses*; Addison-Wesley Publishing Co.: Reading, MA, 1970.
7. Ghosh, J.K. On some properties of sequential *t*-test. Cal. Statist. Assoc. Bull. **1960**, *9*, 77–86.
8. Ghosh, J.K. On the monotonicity of the OC of a class of sequential probability ratio tests. Cal. Statist. Assoc. Bull. **1960**, *9*, 139–144.
9. Ghosh, J.K.; Mitra, S.K.; Parthasarathy, K.R., Eds. *Glimpses of India's Statistical Heritage*; Wiley Eastern Limited: New Delhi, 1992.
10. Hall, W.J.; Wijsman, R.A.; Ghosh, J.K. The relationship between sufficiency and invariance with applications in sequential analysis. Ann. Math. Statist. **1965**, *36*, 575–614. [Correction in Ann. Statist. **1995**, *23*, 705].
11. Hoel, P.G. On a property of the sequential *t*-test. Skand. Aktuarietidskr **1954**, *37*, 19–22.
12. Lehmann, E.L. *Testing Statistical Hypotheses*, 2nd Ed.; Wiley: New York, 1986.
13. Rushton, S. On a sequential *t*-test. Biometrika **1950**, *37*, 326–333.
14. Wald, A. Sequential tests of statistical hypotheses. Ann. Math. Statist. **1945**, *16*, 117–186.
15. Wald, A. *Sequential Analysis*; Wiley: New York, 1947.

Received June 2003
Revised November 2003
Accepted January 2004
Recommended by Masafumi Akahira