The Annals of Mathematical Statistics
1971, Vol. 42, No. 6, 1859–1869

EXPONENTIALLY BOUNDED STOPPING TIME OF SEQUENTIAL PROBABILITY RATIO TESTS FOR COMPOSITE HYPOTHESES[1]

BY R. A. WIJSMAN

University of Illinois at Urbana-Champaign

Let N be the stopping variable of a SPRT for testing one composite hypothesis against another, based on i.i.d. observations $Z_1, Z_2, \cdots$ with common distribution P. P need not belong to the model. N is termed *exponentially bounded* if for every choice of stopping bounds there exists $c < \infty$ and $\rho < 1$ such that $P\{N > n\} < c\rho^n$; if this does not hold P is called *obstructive*. The main theorem presents sufficient conditions, both on the model and on P, for N to be exponentially bounded. Under weaker conditions the theorem proves $P\{N < \infty\} = 1$. Two applications of the theorem are given: 1. In the problem of testing $\sigma = \sigma_1$ against $\sigma = \sigma_0$ in a normal population with unknown mean it is proved that N is exponentially bounded for every P except if $P\{Z_1 = \zeta \pm a\} = \frac{1}{2}$ (ζ arbitrary and a^2 a given function of σ_1 and σ_2) in which case P is obstructive. 2. In the sequential t-test it is proved that N is exponentially bounded for every P for which ${Z_1}^2$ has finite mgf and is not a member of a certain family of two-point distributions.

1. Introduction. Let $Z_1, Z_2, \cdots$ be i.i.d. random variables with common distribution P. The joint distribution of the Z's will also be denoted P. Stein [4] showed that the stopping time N (= random sample size) of Wald's [5] sequential probability ratio test (SPRT) for testing one simple hypothesis against another is *exponentially bounded*, i.e. satisfies, for some $c < \infty$, $0 < \rho < 1$:

$$P\{N > n\} < c\rho^n, \qquad n = 1, 2, \cdots \tag{1.1}$$

for every P except for those P under which the log probability ratio is degenerate at 0. The reason Wald's SPRT can be treated with such relative ease is that $\{L_n, n = 1, 2, \cdots\}$ is a random walk, where L_n is the log probability ratio at the nth stage.

If the hypotheses to be tested are composite, Wald [5] suggested a reduction to simple hypotheses by means of weight functions and it is then possible to define a SPRT in terms of these simple hypotheses. A special case of this method is the use of an invariance reduction, provided there exists a group of invariance transformations that reduces both composite hypotheses to simple ones. The resulting test will be called an *invariant* SPRT. Our applications of the main theorem in this paper will in fact be exclusively to invariant SPRT's, but it should be kept in mind that the theorem could also be applied to weight function SPRT's that are not obtained by an invariance reduction.

Received November 23, 1970.

[1] Research supported by the National Science Foundation under grant GP 23835.

AMS 1970 *subject classification*: 6245.

Key words and phrases. Stopping time, exponentially bounded, invariant SPRT, obstructive, composite hypotheses.

Whereas in Wald's SPRT for testing a simple hypothesis against a simple alternative L_n is a random walk, this is no longer true for a weight function SPRT and consequently it is much harder to try to establish (1.1). The general results that have been obtained are still very incomplete in that they require conditions on P that are almost certainly unnecessary while, at the same time, not demonstrating for which P's (1.1) fails ([7] gives a more thorough discussion of this, as well as a list of references). In particular, for the validity of (1.1) in a large class of parametric problems it has been necessary to assume the existence of finite moment generating functions (mgf) of certain functions of Z_1. On the other hand, in [3] two examples are given in which (1.1) is demonstrated to hold for all P except possibly for a class of P's exhibiting a certain degeneracy, thereby showing that at least in those examples the existence of moment generating functions is irrelevant.

In the following we shall call P *obstructive* if (1.1) is not satisfied for all stopping bounds. The main theorem in the present paper (Theorem 2.1) establishes (1.1) under conditions which are admittedly again restrictive. On the other hand, thanks to this theorem it is possible to present an example where it is possible to prove that (1.1) holds for every P except for P in a small class of obstructive two-point distributions. This will be done in Section 3.

The theorem is also applied to the sequential t-test (in Section 4). It is shown that if $Z_1{}^2$ has finite mgf and P is not in a certain small class of two-point distributions then (1.1) holds. This constitutes a strengthening of certain results in [6] since it is no longer necessary to cope with so-called "exceptional" P's (not to be confused with obstructive P's: the exceptional P's are defined in terms of certain moments and form a much larger class than the obstructive P's). For the exceptional P's (1.1) could not be established in [6], only a weaker property. Thus, the results in Section 4 are essentially the same as obtained by Berk [1] following a different method. Our method provides slightly more information about a subclass of all two-point distributions that may contain the obstructive P's.

Throughout this paper R_n denotes the probability ratio at the nth stage. For invariant SPRT's, under some additional conditions, a more or less explicit expression for R_n is given in [7] (2.1) but will not be used until Section 4. We define $L_n = \log R_n$, L_n being slightly more convenient to work with. The stopping time N of an invariant SPRT is then defined in the usual way:

$$N = \text{smallest } n \geqq 1 \text{ such that } l_1 < L_n < l_2 \text{ is violated.} \tag{1.2}$$

where $-\infty < l_1 < l_2 < \infty$ are the chosen stopping bounds for $\{L_n\}$.

We shall assume that there exists a function s from the range of Z_1 into Euclidean k-space, E^k, for some $k \geqq 1$, such that, with the notation

$$X_i = s(Z_i), \qquad i = 1, 2, \cdots \tag{1.3}$$

$$\bar{X}_n = (1/n) \textstyle\sum_{i=1}^n X_i, \tag{1.4}$$

R_n (and therefore L_n) is a function of $Z_1, \cdots, Z_n$ only through n and $\bar{X}_n$. The domain of applicability includes therefore exponential models but not such nonparametric tests as, for example, sequential rank tests.

2. The main theorem. Theorem 2.1 below consists of two parts, the second part proving (1.1) and the first proving termination with probability one, i.e.

$$P(N < \infty) = 1, \tag{2.1}$$

under weaker conditions on P. Even though we are mainly interested in (1.1), most of the labor that goes into proving (1.1) can also be used for (2.1) so that we obtain the latter at no extra cost. The method of proof utilizes an idea used by Stein in [4].

In the assumptions below the existence will be assumed of a certain real-valued function Φ on E^k possessing certain properties. There is a natural candidate for Φ, as is clear from [7, Section 2], but in this section it is immaterial where Φ comes from. We shall make the following two sets of assumptions, the first used to prove (2.1) and both to prove (1.1). (Notation: vectors will be understood to be column vectors and prime denotes transposition.)

ASSUMPTION A. (i) $E_P X_1 = \xi$ exists and is finite; (ii) there exists a neighborhood V of ξ and a real-valued continuous function Φ on V and a finite constant B_1 such that

$$|L_n - n\Phi(\bar{X}_n)| < B_1 \quad \text{if} \quad \bar{X}_n \in V, \qquad n = 1, 2, \cdots; \tag{2.2}$$

(iii) Φ has continuous first partial derivatives on V; let grad Φ be the vector of first partials and $\Delta =$ grad Φ evaluated at ξ, then

$$P\{\Delta'(X_1 - \xi) = 0\} < 1. \tag{2.3}$$

ASSUMPTION B. For all components X_{1j} $(j = 1, \cdots, k)$ of X_1 we have $E_P \exp[tX_{1j}] < \infty$ for t in some neighborhood about 0.

Clearly, Assumption B implies Assumption A(i).

THEOREM 2.1. *Let* $-\infty < l_1 < l_2 < \infty$ *be arbitrary and* N *defined by* (1.2). *Then* (2.1) *is true if Assumption* A *is satisfied and* (1.1) *is true if Assumptions* A *and* B *are satisfied. Assumption* A(iii) *is not needed if* $\Phi(\xi) \neq 0$.

PROOF. Put $B = B_1 + \max(|l_1|, |l_2|)$, and

$$\Phi_n = n\Phi(\bar{X}_n). \tag{2.4}$$

Comparison with (1.2) and using (2.2) shows that if, for some n, $\bar{X}_n \in V$ and $|\Phi_n| \geqq B$ then $N \leqq n$. We distinguish two cases: $\Phi(\xi) \neq 0$ and $\Phi(\xi) = 0$. In case 1 suppose $\Phi(\xi) > 0$ (the case $\Phi(\xi) < 0$ is treated entirely analogously). Choose $0 < \varepsilon < \Phi(\xi)$ and reduce V, if necessary, to ensure $\Phi > \varepsilon$ on V (this can be done since Φ is continuous at ξ). For any $n > B/\varepsilon$, if $\bar{X}_n \in V$ then $n\Phi(\bar{X}_n) > B$ so that $N \leqq n$. Let $N' =$ first $n > B/\varepsilon$ such that $\bar{X}_n \in V$. By the above we have $N \leqq N'$.

By Assumption A(i), $\bar{X}_n \to \xi$ a.e. P so that $\bar{X}_n \in V$ eventually a.e. P. That is, $P(N' < \infty) = 1$ and so *a fortiori* $P(N < \infty) = 1$. If also Assumption B holds, then it follows essentially from [3] Theorem 1 (see also [6, Section 3]) that

$$P\{\bar{X}_n \notin V\} < c_1 \rho_1^{\,n} \tag{2.5}$$

for some $c_1 < \infty$, $0 < \rho_1 < 1$. Since $n > B/\varepsilon$ and $\bar{X}_n \in V$ imply $N \leqq n$ it follows that $P\{N > n\} < c_1 \rho_1^{\,n}$ for all $n > B/\varepsilon$, and (1.1) follows.

Now consider case 2: $\Phi(\xi) = 0$. By making a translation in E^k we may assume $\xi = 0$, so $\Phi(0) = 0$, and $\Delta = \operatorname{grad} \Phi$ evaluated at 0. By Assumption A(iii) $\Delta' X_1$ is not degenerate at 0. Choose any $\delta > 0$; then there exists a positive integer r and $\varepsilon > 0$ such that $P\{|\Delta'(X_1 + \cdots + X_r)| \geqq 2B + 2\delta\} > 2\varepsilon$. Denote $S_n = X_{n+1} + \cdots + X_{n+r}$, $n = 0, 1, \cdots$, so that the above inequality may be written $P\{|\Delta' S_0| \geqq 2B + 2\delta\} > 2\varepsilon$. Then there exists A such that $P\{|\Delta' S_0| \geqq 2B + 2\delta,\ \|S_0\| \leqq A\} > \varepsilon$, which is equivalent to

$$P\{|\Delta' S_0| < 2B + 2\delta \quad \text{or} \quad \|S_0\| > A\} < 1 - \varepsilon. \tag{2.6}$$

According to (2.4) we can write

$$\Phi_{n+r} - \Phi_n = (n+r)[\Phi(\bar{X}_{n+r}) - \Phi(\bar{X}_n)] + r\Phi(\bar{X}_n). \tag{2.7}$$

Put $u = \operatorname{grad} \Phi - \Delta$ so that $u(0) = 0$ and u is continuous on V. Provided $\bar{X}_n, \bar{X}_{n+r} \in V$ we can then write

$$\Phi(\bar{X}_n) = (\Delta + u_1)' \bar{X}_n, \tag{2.8}$$

$$\Phi(\bar{X}_{n+r}) - \Phi(\bar{X}_n) = (\Delta + u_2)'(\bar{X}_{n+r} - \bar{X}_n) \tag{2.9}$$

in which the random variables u_1 and u_2 (the dependency on n has been suppressed in the notation) are the values of u at intermediate points: $u_1 = u(\alpha_1 \bar{X}_n)$, $u_2 = u(\alpha_2 \bar{X}_n + (1-\alpha_2)\bar{X}_{n+r})$, $0 \leqq \alpha_1, \alpha_2 \leqq 1$, α_1 depending on $\bar{X}_n$, α_2 on both $\bar{X}_n$ and $\bar{X}_{n+r}$. Multiply (2.9) on both sides by $n+r$ and observe that on the right-hand side $(n+r)(\bar{X}_{n+r} - \bar{X}_n) = S_n - r\bar{X}_n$. Then substitute together with (2.8) into the right-hand side of (2.7):

$$\Phi_{n+r} - \Phi_n = (\Delta + u_2)' S_n + r(u_1 - u_2)' \bar{X}_n. \tag{2.10}$$

We may choose V convex and so small that if $x_1, x_2, x_3, x_4 \in E^k$ the following implication holds:

$$[x_1, x_2, x_3 \in V, \|x_4\| \leqq A] \Rightarrow [r|(u(x_1) - u(x_2))' x_3| < \delta, |u(x_2)' x_4| < \delta]. \tag{2.11}$$

If $\bar{X}_n, \bar{X}_{n+r} \in V$ then so are u_1 and u_2. Then by (2.11):

$$[\bar{X}_n \in V, \bar{X}_{n+r} \in V, \|S_n\| \leqq A] \Rightarrow [r|(u_1 - u_2)' \bar{X}_n| < \delta, |u_2' S_n| < \delta] \tag{2.12}$$

so that, using (2.10),

$$\begin{aligned}[\bar{X}_n \in V, \bar{X}_{n+r} \in V, \|S_n\| \leqq A, |\Phi_{n+r} - \Phi_n| < 2B] \\ \Rightarrow [\bar{X}_n \in V, \bar{X}_{n+r} \in V, \|S_n\| \leqq A, |\Delta' S_n| < 2B + 2\delta].\end{aligned} \tag{2.13}$$

From (2.13) follows

$$(2.14)\quad [\bar{X}_n \in V, \bar{X}_{n+r} \in V, |\Phi_{n+r} - \Phi_n| < 2B] \Rightarrow [\|S_n\| > A \quad \text{or} \quad |\Delta' S_n| < 2B + 2\delta] = C_n, \qquad \text{say.}$$

Since the S_n, $n = 0, 1, \cdots$, are equidistributed, we have from (2.6)

$$(2.15)\qquad PC_n < 1 - \varepsilon, \qquad n = 0, 1, \cdots$$

By Assumption A(i) $\bar{X}_n \to \xi$ a.e. P, so that $\bar{X}_n \in V$ eventually, with P-probability 1. This implies that given any $\varepsilon_1 > 0$ there exists an integer n_0 such that $PD < \varepsilon_1$, where D is the complement of $\{\bar{X}_n \in V, n \geqq n_0\}$. For the event that the test never terminates we have now the following string of inclusions:

$$\begin{aligned}
\{N = \infty\} &\subset D \cup \{\bar{X}_n \in V, |\Phi_n| < B, n \geqq n_0\}, \\
(2.16)\qquad &\subset D \cup \{\bar{X}_n \in V, |\Phi_{n+r} - \Phi_n| < 2B, n \geqq n_0\}, \\
&\subset D \cup \{\bar{X}_n \in V, |\Phi_{n+r} - \Phi_n| < 2B, n = n_0 + ir, i = 0, 1, \cdots\}, \\
&\subset D \cup \textstyle\bigcap_{i=0}^{\infty} C_{n_0+ir},
\end{aligned}$$

where in the last inclusion we have used (2.14). Now the C_{n_0+ir}, $i = 0, 1, \cdots$, are independent, and $PC_{n_0+ir} < 1 - \varepsilon$ by (2.15). Therefore, $P \bigcap_{i=0}^{\infty} C_{n_0+ir} = 0$ so that by the last inclusion in (2.16) $P(N = \infty) \leqq PD < \varepsilon_1$. Since ε_1 was arbitrary, $P(N = \infty) = 0$, thereby proving the first part of the theorem in case 2.

For the proof of the second part of the theorem in case 2 we use Assumption B to obtain (2.5), as in case 1. We have then

$$(2.17)\quad P\{N > (r+1)n\} \leqq \textstyle\sum_{i=0}^{n} P\{\bar{X}_{n+ir} \notin V\} + P\{\bar{X}_{n+ir} \in V, |\Phi_{n+ir}| < B, i = 0, \cdots, n\}.$$

Using (2.5), the sum on the right-hand side of (2.17) is bounded by $c_1(1-\rho_1{}^r)^{-1}\rho_1{}^n = c_2\rho_1{}^n$, say. The remaining term on the right-hand side of (2.17) is $\leqq P\{\bar{X}_n \in V, \bar{X}_{n+ir} \in V, |\Phi_{n+ir} - \Phi_{n+(i-1)r}| < 2B, i = 1, \cdots, n\}$. This, in turn, using (2.14) and (2.15), is $\leqq P \bigcap_{i=0}^{n-1} C_{n+ir} < (1-\varepsilon)^n$. Therefore, $P\{N > (r+1)n\} < c_2\rho_1{}^n + (1-\varepsilon)^n$. This establishes (1.1) for n running through integral multiples of r. To establish (1.1) for all n is then a trivial matter and follows along the lines of [4].

3. Complete characterization of distributions P for which N is exponentially bounded in a special example. In Example 1 of [7] a certain class of obstructive P's was exhibited. In this section we shall complement the result in [7] by showing that the obstructive P's in that example are the only ones.

Let $Z_1, Z_2, \cdots$ be i.i.d. normal with mean ζ, variance σ^2, both unknown. We want to test $H_1 : \sigma = \sigma_1$, against $H_2 : \sigma = \sigma_2$, where the σ_j are given and distinct (thus,

ζ is a nuisance parameter). Under the transformations $Z_i \to Z_i + b$ $(i = 1, 2, \cdots)$, $\zeta \to \zeta + b$, $\sigma \to \sigma$, $-\infty < b < \infty$, the problem is invariant, and it was shown in [7] (4.1) that

$$(3.1) \qquad L_n = ((2\sigma_1^2)^{-1} - (2\sigma_2^2)^{-1}) \textstyle\sum_{i=1}^n (Z_i - \bar{Z}_n)^2 + (n-1)\log(\sigma_1/\sigma_2)$$

in which $\bar{Z}_n = (1/n)\sum_1^n Z_i$. For the purpose of exponential boundedness we may multiply L_n by any nonzero constant. We may pretend then that

$$(3.2) \qquad L_n = \textstyle\sum_{i=1}^n (Z_i - \bar{Z}_n)^2 - (n-1)a^2$$

in which

$$(3.3) \qquad a^2 = (\log \sigma_2 - \log \sigma_1)/((2\sigma_1^2)^{-1} - (2\sigma_2^2)^{-1}).$$

Now let the actual common distribution of the Z_i be P and try to establish (1.1), with N defined in (1.2).

Case 1, *P unbounded*. We shall prove that (1.1) holds for every $-\infty < l_1 < l_2 < \infty$. This will be accomplished by showing—following the Stein method [4]—that there exists $\varepsilon > 0$ such that for all n

$$(3.4) \qquad P\{L_{n+1} > l_2 \mid Z_1, \cdots, Z_n, l_1 < L_n < l_2\} \geqq \varepsilon.$$

An elementary computation, using (3.2), shows that

$$(3.5) \qquad L_{n+1} - L_n = \frac{n}{n+1}(Z_{n+1} - \bar{Z}_n)^2 - a^2.$$

Now put $d = l_2 - l_1$, then given $l_1 < L_n < l_2$ the event $L_{n+1} > l_2$ is implied by $L_{n+1} - L_n > d$, i.e., using (3.5), $(Z_{n+1} - \bar{Z}_n)^2 > ((n+1)/n)(a^2 + d)$. This, in turn, is implied by $|Z_{n+1} - \bar{Z}_n| > c$, where $c^2 = 2(a^2 + d)$. Therefore, the left-hand side of (3.4) is $\geqq$

$$(3.6) \qquad P\{|Z_{n+1} - \bar{Z}_n| > c \mid Z_1, \cdots, Z_n, l_1 < L_n < l_2\}.$$

It suffices therefore to find a positive lower bound for (3.6). Since Z_{n+1} is independent of $(Z_1, \cdots, Z_n)$, the value of (3.6) depends on the conditioning only through $\bar{Z}_n$. Furthermore, $P\{|Z_{n+1} - \bar{Z}_n| > c \mid \bar{Z}_n = z\} = P\{|Z_{n+1} - z| > c\} = P\{|Z_1 - z| > c\}$. We shall show that there exists $\varepsilon > 0$ such that

$$(3.7) \qquad P\{|Z_1 - z| > c\} \geqq \varepsilon \quad \text{for all} \quad -\infty < z < \infty.$$

Since P is unbounded, there is a number z_0 such that $P\{Z_1 < z_0 - c\} = \varepsilon_1 > 0$ and $P\{Z_1 > z_0 + c\} = \varepsilon_2 > 0$. Since $P\{Z_1 < z - c\}$ is non-decreasing in z, $P\{|Z_1 - z| > c\} \geqq \varepsilon_1$ if $z \geqq z_0$. Similarly, $P\{|Z_1 - z| > c\} \geqq \varepsilon_2$ if $z \leqq z_0$. Then taking $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$ gives (3.7).

Case 2, *P bounded*. In this case Theorem (2.1) will be used. The function s in (1.3) will be chosen: $s(z) = (z^2, z)'$, $-\infty < z < \infty$ (thus $k = 2$), so that $X_i = (Z_i^2, Z_i)'$. Now choose the function Φ on E^2 as follows:

$$\Phi(x_1, x_2) = x_1 - x_2^2 - a^2 \tag{3.8}$$

with a^2 given in (3.3). We compute

$$n\Phi(\bar{X}_n) = \sum_{i=1}^{n} (Z_i - \bar{Z}_n)^2 - na^2 \tag{3.9}$$

and comparing this with (3.2) we see that Assumption A(ii) is satisfied, with $B_1 = a^2$, no matter what neighborhood V is chosen. The boundedness of P guarantees the validity of Assumption B. For convenience we shall write $E_P Z_1 = \zeta$, $E_P Z_1^2 = \sigma^2 + \zeta^2$. Then $\xi = E_P X_1 = (\sigma^2 + \zeta^2, \zeta)'$. From the form of L_n given by (3.2) it is obvious that all distributions P obtained from a single one by translation produce the same stochastic behavior of $\{L_n\}$. It suffices, therefore, to assume $\zeta = 0$ so that $\xi = (\sigma^2, 0)'$. Substituting ξ for x into (3.8) we get $\Phi(\xi) = \sigma^2 - a^2$ so that $\Phi(\xi) \neq 0$ provided $\sigma \neq a$. For any such P we can therefore conclude, by Theorem 2.1, that (1.1) holds.

Now suppose P is such that $\sigma = a$, so that $\Phi(\xi) = 0$. In order to conclude (1.1) we now also need Assumption A(iii). From (3.8) we compute grad $\Phi = (1, -2x_2)'$ so that $\Delta = (1, 0)'$ and $\Delta'(X_1 - \xi) = Z_1^2 - \sigma^2 = Z_1^2 - a^2$. Hence if $P(Z_1^2 - a^2 = 0) < 1$ Assumption A(iii) is satisfied and we can conclude (1.1). On the other hand, if $P(Z_1^2 - a^2 = 0) = 1$ then in order that ζ be equal to 0 we must have $P(Z_1 = \pm a) = \frac{1}{2}$. In this case it was shown in [7, Section 4] (for $a = 1$, but the extension to arbitrary a is trivial) that (1.1) fails for sufficiently wide stopping bounds. On the other hand, (2.1) is still valid.

Summarizing, in the present example N is exponentially bounded for every P except if

$$P\{Z_1 = \zeta \pm a\} = \tfrac{1}{2} \quad \text{for some} \quad -\infty < \zeta < \infty \tag{3.10}$$

with a given by (3.3). The distributions defined by (3.10) are obstructive, but for every such P we still have $P(N < \infty) = 1$.

4. Another application: the sequential *t*-test. Let $Z_1, Z_2, \cdots$ be i.i.d. normal with mean ζ and variance σ^2, both unknown. Put $\gamma = \zeta/\sigma$ and test $\gamma = \gamma_1$ against $\gamma = \gamma_2$ where γ_1 and γ_2 are any two distinct finite numbers. The problem is invariant under the transformations $Z_i \to cZ_i$ $(i = 1, 2, \cdots)$, $\zeta \to c\zeta$, $\sigma \to c\sigma$, $c > 0$. Thus, the group G of invariance transformations consists of the positive reals c under multiplication. As right invariant (= left invariant) measure on G we shall take $\nu_G(dg) = dc/c$. Put $\theta = (\zeta, \sigma)$, then the orbit of θ under G is $G\theta = \{g\theta : g \in G\} = \{(c\zeta, c\sigma) : c > 0\}$. Taking, in particular, $\theta_j = (\gamma_j, 1)$, $j = 1, 2$, the two orbits $G\theta_j$ are the two composite hypotheses that are to be tested. From [7, (2.1)] we take the representation for the probability ratio at the nth stage

$$R_n = J_n(\theta_2)/J_n(\theta_1) \tag{4.1}$$

in which

$$J_n(\theta) = \int \prod_{i=1}^n p_{g\theta}(Z_i)\nu_G(dg) \tag{4.2}$$

and p_θ is the density of Z_1 with respect to Lebesgue measure

$$p_\theta(z) = (2\pi)^{-\frac{1}{2}}\sigma^{-1}\exp[-(2\sigma^2)^{-1}(z-\zeta)^2]. \tag{4.3}$$

It is seen that (4.2) depends on the Z_i only through $\sum_1^n Z_i^2$ and $\sum_1^n Z_i$. Thus, in (1.3) we shall take s as in Section 3: $s(z) = (z^2, z)'$ so that $X_i = (Z_1^2, Z_i)'$. After setting $g\theta = (\gamma c, c)'$, $\nu_G(dg) = dc/c$ and making a change of variable of integration: $c = 1/t$, we can write (4.2) as

$$J_n(\theta) = \int_0^\infty \exp[n\psi(t, \bar{X}_n; \gamma)]t^{-1}\,dt \tag{4.4}$$

in which

$$\psi(t, x; \gamma) = -\tfrac{1}{2}x_1 t^2 + \gamma x_2 t + \log t - \tfrac{1}{2}\gamma^2 - \tfrac{1}{2}\log(2\pi) \tag{4.5}$$

and $x = (x_1, x_2)$. For fixed γ, $x_1 > 0$, x_2, ψ as a function of t has a unique maximum. Put

$$\varphi(x; \gamma) = \max_{t>0}\psi(t, x; \gamma) \tag{4.6}$$

and

$$\Phi(x) = \varphi(x; \gamma_2) - \varphi(x; \gamma_1). \tag{4.7}$$

As explained in [7, Section 2] we may expect R_n to behave asymptotically as $\exp[n\Phi(\bar{X}_n)]$ so that Φ defined in (4.7) is the natural candidate to try for the application of Theorem 2.1. The basis for this phenomenon is the application of Laplace's method (see e.g. [2]) which gives the result

$$(2\pi)^{-\frac{1}{2}}n^{\frac{1}{2}}(x_1 + t_m^{-2})^{\frac{1}{2}}t_m\exp[-n\varphi(x;\gamma)] \cdot \int_0^\infty \exp[n\psi(t, x; \gamma)]t^{-1}\,dt \to 1 \quad \text{as} \quad n \to \infty, \tag{4.8}$$

where $t_m = t_m(x, \gamma)$ is the value of t that maximizes $\psi(t, x; \gamma)$. However, there is a difficulty in applying (4.8) to (4.4) which stems from the fact that on the right-hand side of (4.4) there is a random $\bar{X}_n$ rather than a fixed x, and there is a priori no guarantee that in (4.8) the convergence is uniform in x. On the other hand, in order to satisfy Assumption A(ii) we only need the convergence in (4.8) to be uniform for $x \in V$, where V is any neighborhood of ξ. Thus, what is needed is the following:

THEOREM 4.1. (uniform Laplace). *Let*

$$J(x, n) = \int_{-\infty}^\infty \exp(n\psi(t, x))h(t)\,dt, \qquad x \in V, \tag{4.9}$$

with V some set, and assume that the following conditions are fulfilled: There is an interval $T = [t_1, t_2]$ with $-\infty < t_1 < t_2 < \infty$, and there are finite numbers B_1, B_2 and a nonnegative integer n_0 such that for all $x \in V$:

(i) $\psi(t, x) > B_1$, $t \in T$; *h continuous and* > 0 *on* T; $h \geqq 0$ *everywhere*;

(ii) $\int_{-\infty}^{\infty} \exp[n_0\psi(t,x)]h(t)\,dt < B_2$;

(iii) $\psi(\cdot\,, x)$ *attains a maximum at* $t_m = t_m(x) \in (t_1, t_2)$; *put* $\varphi(x) = \psi(t_m, x)$;

(iv) *given any* $\delta > 0$ *there exists* $a(\delta) > 0$ *such that* $\psi(t, x) < \varphi(x) - a(\delta)$ *if* $|t - t_m| > \delta$;

(v) *there exists* $b = b(x) > 0$ *such that given any* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that* $|\psi(t, x) - \varphi(x) + (b/2)(t - t_m)^2| \leqq \varepsilon(t - t_m)^2$ *if* $|t - t_m| \leqq \delta$.

Then $(2\pi)^{-\frac{1}{2}}(nb(x))^{\frac{1}{2}}[h(t_m(x))]^{-1} \exp(-n\varphi(x))$, $J(x, n) \to 1$ *as* $n \to \infty$, *uniformly in* $x \in V$.

The proof is essentially the same as for the standard theorem when there is no extra variable x (see e.g. [2]) and will be omitted. The theorem will be applied with $\psi(t, x) = \psi(t, x; \gamma)$ given in (4.5), with $\gamma = \gamma_1$ or γ_2, $h(t) = 0$ or t^{-1} according as $t \leqq 0$ or > 0, V any compact subset of $\{x = (x_1, x_2): x_1 > 0\}$, $0 < t_1 < t_2 < \infty$ chosen suitably so that $t_1 < t_m(x) < t_2$ for all $x \in V$, and n_0 any positive integer, e.g. 1. It is obvious from (4.5) that for fixed x, $\psi \to -\infty$ as $t \to 0$ or $\to \infty$, so there must be at least one maximum t_m, and $\partial\psi/\partial t \equiv -x_1 t + \gamma x_2 + t^{-1} = 0$ at $t = t_m$. The latter equation has exactly one positive root:

$$t_m = x_1^{-\frac{1}{2}}\alpha(\gamma x_2 x_1^{-\frac{1}{2}}) \tag{4.10}$$

in which

$$\alpha(u) = \tfrac{1}{2}[u + (u^2 + 4)^{\frac{1}{2}}], \qquad -\infty < u < \infty, \tag{4.11}$$

and condition (iii) of Theorem 4.1 is therefore fulfilled. To check (iv) and (v) observe that $\partial^2\psi/\partial t^2 = -x_1 - t^{-2}$ which is continuous on T (verifying (v) with $b = x_1 + t_m^{-2}$) and $< -c$ for some $c > 0$ for all $x \in V$ (verifying (iv)).

Next we shall investigate Assumptions A and B of Section 2. Since $X_1 = (Z_1^2, Z_1)'$, in order to satisfy Assumption B it suffices to require of P that $E_P \exp[tZ_1^2] < \infty$ for t in a neighborhood of 0. The function Φ in Assumption A(ii) is taken to be the Φ defined in (4.7). The inequality (2.2) is verified by using the conclusion of Theorem 4.1, applied once for $\gamma = \gamma_1$, once for $\gamma = \gamma_2$, and observing that $[h(t_m(x; \gamma))]^{-1} = t_m(x; \gamma) \in T$ if $x \in V$ so that $|\log t_m(x; \gamma_2) - \log t_m(x; \gamma_1)| \leqq \log(t_2/t_1)$, while $\log b(x; \gamma_2) - \log b(x; \gamma_1)$ is similarly bounded. It remains to verify Assumption A(iii). Substitution of (4.10) into (4.5) yields for (4.6):

$$\varphi(x; \gamma) = \beta(\gamma x_2 x_1^{-\frac{1}{2}}) - \tfrac{1}{2}\log x_1 - \tfrac{1}{2}\gamma^2 - \tfrac{1}{2}\log(2\pi) - \tfrac{1}{2} \tag{4.12}$$

in which

$$\beta(u) = \tfrac{1}{2}u\alpha(u) + \log\alpha(u). \tag{4.13}$$

Substitution into (4.7) then gives

$$\Phi(x) = \beta(\gamma_2 x_2 x_1^{-\frac{1}{2}}) - \beta(\gamma_1 x_2 x_1^{-\frac{1}{2}}) - \tfrac{1}{2}\gamma_2^2 + \tfrac{1}{2}\gamma_1^2. \tag{4.14}$$

From this we compute (using $\beta'(u) = \alpha(u)$)

$$\frac{\partial\Phi}{\partial x_1} = -\tfrac{1}{2}x_2 x_1^{-\frac{3}{2}}(g_2(x)-g_1(x)), \tag{4.15}$$

$$\frac{\partial\Phi}{\partial x_2} = x_1^{-\frac{1}{2}}(g_2(x)-g_1(x))$$

in which

$$g_j(x) = \gamma_j\alpha(\gamma_j x_2 x_1^{-\frac{1}{2}}), \qquad j = 1,2. \tag{4.16}$$

From (4.15) and (4.16) it is seen that grad Φ is continuous. If $x_2 = 0$, $g_2(x)-g_1(x) = \gamma_2-\gamma_1 \neq 0$. If $x_2 \neq 0$, $g_2(x)-g_1(x)$ is also $\neq 0$ because the function $u\alpha(u)$ is strictly increasing. We shall determine now for which P's (2.3) is not satisfied, i.e., for which P's

$$\Delta'(X_1-\xi) = 0 \quad \text{with } P\text{-probability} \quad 1. \tag{4.17}$$

Writing $E_P Z_1 = \zeta$, $E_P Z_1^2 = \sigma^2+\zeta^2$ so that $\xi' = (\sigma^2+\zeta^2, \zeta)$, and observing that in (4.15) $g_2(x)-g_1(x) \neq 0$, we compute from (4.15) that Δ' is proportional to $(-\zeta/(\sigma^2+\zeta^2), 2)$. After substitution into (4.17) we have with P-probability 1:

$$\zeta Z_1^2 - 2(\sigma^2+\zeta^2)Z_1 + \zeta(\sigma^2+\zeta^2) = 0. \tag{4.18}$$

If $\zeta = 0$, (4.18) has as its only solution $Z_1 = 0$. But if $P(Z_1 = 0) = 1$ then also $P(X_1 = 0) = 1$ so that $\bar{X}_n = 0$ for all n with P-probability 1, and after consulting (4.5) we see that the integral on the right hand side of (4.4) does not converge for any n. In that case the sequential t-test is undefined and we shall therefore exclude the possibility $P(Z_1 = 0) = 1$ from consideration. (Note that for all other one-point distributions the sequence $\{R_n\}$ is well-defined by (4.1) even though with probability 1 the t-ratio at each n is ∞.) If $\zeta \neq 0$, (4.18) has exactly two distinct solutions and the probabilities in these two points can be determined from the equation $E_P Z_1 = \zeta$. The result is

$$P\{Z_1 = (\sigma^2+\zeta^2)^{\frac{1}{2}}\zeta^{-1}((\sigma^2+\zeta^2)^{\frac{1}{2}} \pm \sigma)\} = \tfrac{1}{2}[1 \mp \sigma(\sigma^2+\zeta^2)^{-\frac{1}{2}}] \qquad \sigma > 0, \zeta \neq 0. \tag{4.19}$$

In summary, we have shown that in the sequential t-test, when Z_1 has distribution P, the stopping time N is exponentially bounded if Z_1^2 has a finite mgf and if P is not one of the two-point distributions defined by (4.19). If we do not require Z_1^2 to have a finite mgf but only a finite expectation, then $P(N < \infty) = 1$ if P does not satisfy (4.19). Berk [1], by a different method, obtained the same conclusions except that his method necessitates the exclusion of a family of two-point distributions different from (4.19) (and the family depends on (γ_1, γ_2)). (Assumption 2.4 (c) in [1] at first seems to imply that all two-point distributions have to be excluded. However, after the publication of [1] it was noticed by Berk that the proofs of various theorems only use Assumption 2.4 (c) with $\theta = \theta_p$, $\theta' = \theta_q$. For the normal distribution the assumption then is equivalent to the exclusion of a certain two-parameter subfamily of all two-point distributions.)

REMARKS. 1. Under the assumption that ${Z_1}^2$ has finite mgf a distribution P is not obstructive unless both $\Phi(\xi) = 0$ and (4.17) holds. It can be shown from (4.14) that the equation $\Phi(\xi) = 0$ has exactly one solution for ζ/σ, its value depending on γ_1 and γ_2, and $\zeta/\sigma = 0$ iff ${\gamma_1}^2 = {\gamma_2}^2$. With this value of ζ/σ (provided it is $\neq 0$) (4.19) defines a one-parameter family of distributions obtained from a single one by scale transformations. This same family is also obtained as the intersection of (4.19) and Berk's two-parameter family of distributions mentioned above.

2. In the sequential t-test it is not known whether there are any obstructive distributions at all. In the light of the results of Section 3 and of [7, Examples 2 and 3] one may conjecture that the distributions mentioned in Remark 1 are indeed obstructive and that they are the only ones (the latter would be proved if it were shown that any unbounded P is not obstructive, as in Section 3).

3. Other classical invariant SPRT's, such as the sequential F-test, etc., are amenable to the same treatment as afforded the sequential t-test in this section. However, to this end it is necessary to extend Theorem 4.1 and permit t to be vector-valued as well as cope with functions h that are not > 0 at t_m but instead behave as a product of powers of some of the components of $t - t_m$.

Acknowledgment. I am indebted to R. H. Berk for valuable discussion and comments.

REFERENCES

[1] BERK, R. H. (1970). Stopping times of SPRTS based on exchangeable models. *Ann. Math. Statist.* **41** 979–990.

[2] DE BRUIJN, N. G. (1958). *Asymptotic Methods in Analysis.* North-Holland Publishing Co., Amsterdam.

[3] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Statist.* **23** 493–507.

[4] STEIN, C. (1946). A note on cumulative sums. *Ann. Math. Statist.* **17** 498–499.

[5] WALD, A. (1947). *Sequential Analysis.* Wiley, New York.

[6] WIJSMAN, R. A. (1968). Bounds on the sample size distribution for a class of invariant sequential probability ratio tests. *Ann. Math. Statist.* **39** 1048–1956.

[7] WIJSMAN, R. A. (1970). Examples of exponentially bounded stopping time of invariant sequential probability ratio tests when the model may be false. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* To appear.