The Annals of Statistics
1983, Vol. 11, No. 1, 306–316

AN EFFICIENT APPROXIMATE SOLUTION TO THE KIEFER-WEISS PROBLEM

BY MICHAEL D. HUFFMAN

Montana State University

The problem is to decide on the basis of repeated independent observations whether θ_0 or θ_1 is the true value of the parameter θ of a Koopman-Darmois family of densities, where the error probabilities are at most α_0 and α_1. An explicit method is derived for determining a combination of one-sided SPRT's, known, as a 2-SPRT, which minimizes the maximum expected sample size to within $o((\log \alpha_0^{-1})^{1/2})$ as α_0 and α_1 go to 0, subject to the condition that $0 < C_1 < \log \alpha_0/\log \alpha_1 < C_2 < \infty$ for fixed but arbitrary constants C_1 and C_2. For the case of testing the mean of an exponential density, extensive computer calculations comparing the proposed 2-SPRT with optimal procedures show that the 2-SPRT comes within 2% of minimizing the maximum expected sample size over a broad range of error probability and parameter values.

1. Introduction. Based on a sequence of independent random variables $X_1, X_2, \cdots$ with common density f_θ with respect to some σ-finite measure μ, the problem is to test $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ where $\theta_0 < \theta_1$, with error probabilities at most α_0 and α_1. A test T consists of the pair (N, D), where N is an extended stopping rule with respect to the sequence $\mathscr{F}_n$ of smallest σ-algebras with respect to which $X_1, \cdots, X_n$ are measurable, and D is an $\mathscr{F}_N$-measurable decision rule specifying which hypothesis is to be accepted once sampling has stopped.

The performance of any sequential test is judged on the basis of its error probabilities and expected sample sizes. Wald and Wolfowitz (1948) established the property that the sequential probability ratio test (SPRT) (Wald, 1947) minimizes both $E_{\theta_0}N$ and $E_{\theta_1}N$ among all tests—sequential or not—with equal or smaller error probabilities. ($E_\theta N$ denotes the expected value of N when θ is the true parameter value). Even though the SPRT has this remarkable optimality property, its performance is unsatisfactory for values of θ between θ_0 and θ_1. In some cases $E_\theta N$ is larger than the number of observations required by a fixed sample size test with the same error probabilities. Much work has been directed toward finding procedures which reduce the sample size of the SPRT for these parameter values.

Let $\mathscr{T}(\alpha_0, \alpha_1)$ denote the class of all tests (N, D) which have error probabilities at most α_0 and α_1, and define

$$n(\alpha_0, \alpha_1) = \inf\{\sup_\theta E_\theta N \mid (N, D) \in \mathscr{T}(\alpha_0, \alpha_1)\}.$$

The problem of finding a procedure (N', D') which minimizes the maximum expected sample size subject to the error probability constraints α_0 and α_1—that is, so that $\sup_\theta E_\theta N' = n(\alpha_0, \alpha_1)$—is known as the Kiefer-Weiss problem. No optimal results (in the sense of the optimality property of the SPRT) have been found for this problem. Kiefer and Weiss (1957) proved structure theorems about tests which minimize $E_\phi N$ for a fixed $\theta = \phi$, which is called the modified Kiefer-Weiss problem. Weiss (1962) showed that the Kiefer-Weiss problem reduces to the modified problem in symmetric cases involving normal and binomial distributions, while Lai (1973) investigated the Wiener process case. More recently Lorden (1980) characterized the basic structure of optimal tests for the modified problem, with particularly informative results for the Koopman-Darmois families.

A test (N, D) with error probabilities α_0 and α_1 is customarily judged by its efficiency,

Received April 1981; revised August 1982.

AMS 1980 *subject classification.* Primary 62L10; secondary 62F03.

Key words and phrases. Asymptotic efficiency, sequential probability ratio test, 2-SPRT.

which in the context of the Kiefer-Weiss problem is

$$\frac{n(\alpha_0, \alpha_1)}{\sup_\theta E_\theta N}. \tag{1.1}$$

A procedure is said to be asymptotically efficient if (1.1) tends to 1 as α_0 and α_1 tend to 0, and for such tests the rate of approach of (1.1) to 1 is of interest. In particular, finding fairly simple procedures which are not only asymptotically efficient, but have efficiencies close to 1 for practical values of α_0 and α_1, has been an important problem.

Anderson (1960) studied a class of easily constructable procedures for the symmetric case of testing the mean drift of a Wiener process. In a general context Lorden (1976) studied a subclass of Anderson's procedures related to SPRT's and called 2-SPRT's defined as follows. Given $\theta_0 < \theta < \theta_1$ and $0 < A_0, A_1 \le 1$, the stopping rule $M(\theta, A_0, A_1)$ is the smallest n (or ∞ if there is no n) such that

$$\frac{f_{\theta_i n}}{f_{\theta_n}} \le A_i \tag{1.2}$$

for either $i = 0$ or 1, where $f_{\phi n} = f_\phi(x_1) f_\phi(x_2) \cdots f_\phi(x_n)$. The decision rule D rejects θ_0 if (1.2) holds only for $i = 0$, and rejects θ_1 if it holds only for $i = 1$. If (1.2) is true for both values of i, then any fixed rule can be used for deciding between θ_0 and θ_1. A useful alternative way to write the stopping rule is $M(\theta, A_0, A_1) = \min(M_0(\theta, A_0), M_1(\theta, A_1))$ where $M_i(\theta, A_i)$ is the smallest n such that (1.2) holds.

As Lorden pointed out, the method which Wald used to derive upper bounds for the error probabilities of an SPRT is applicable to the 2-SPRT and yields $\alpha_i < A_i P_\theta$ (reject θ_i), i = 0, 1 so that setting $A_i = \alpha_i$ in (1.2) ensures error probabilities of at most α_0 and α_1. The main theorem in Lorden (1976) states that if α_0 and α_1 are the true error probabilities of the 2-SPRT $(M(\theta, A_0, A_1), D)$ then

$$E_\theta M(\theta, A_0, A_1) = \inf\{E_\theta N \mid (N, D) \in \mathscr{T}(\alpha_0, \alpha_1)\} + o(1)$$

as $\alpha_0, \alpha_1 \to 0$ where θ is fixed. Thus, for any fixed θ, the 2-SPRT provides an asymptotic solution to the modified Kiefer-Weiss problem. In the symmetric normal case, where θ is the mean and $\alpha_0 = \alpha_1$, say the Kiefer-Weiss problem reduces to the modified problem for $\phi = (\theta_0 + \theta_1)/2$ (Weiss, 1962), where only procedures symmetric about ϕ need be considered. So in this case, the 2-SPRT gives an approximate solution to the Kiefer-Weiss problem. Lorden's numerical results indicate that over a wide range of values of α, θ_0 and θ_1, the 2-SPRT has an efficiency of more than 99.2%.

Setting $A_i = \alpha_i$ so that the 2-SPRT's are in $\mathscr{T}(\alpha_0, \alpha_1)$, Lorden's theorem suggests that if $\tilde{\theta}$ can be found so that $E_\theta M(\tilde{\theta}, \alpha_0, \alpha_1)$ is nearly maximized at $\theta = \tilde{\theta}$, then the resulting 2-SPRT will be an approximate solution to the Kiefer-Weiss problem. The following section describes an explicit method for determining $\tilde{\theta}$ as a function of α_0 and α_1, in the context of the Koopman-Darmois family of densities, and states the main result, whose proof is contained in Section 4. Section 3 describes the results of computer calculations comparing the 2-SPRT with actual Kiefer-Weiss solutions in the case of an exponential density. A method of computing the latter was developed, incorporating the technique of backward induction for computing modified Kiefer-Weiss solutions. In addition the expected sample sizes under θ_0 and θ_1 and the maximum expected sample size of the 2-SPRT are compared with those of the SPRT having the same error probabilities.

2. Formulation of the 2-SPRT and statement of the main result. $X_1, X_2, \cdots$ are assumed to be independent and identically distributed with one of the Koopman-Darmois densities $f_\theta(x) = \exp\{\theta x - b(\theta)\}$, $\underline{\theta} < \theta < \bar{\theta}$ with respect to a non-degenerate σ-finite measure μ. The function $b(\theta)$ is necessarily convex and infinitely differential on $(\underline{\theta}, \bar{\theta})$, and its first two derivatives satisfy $b'(\theta) = E_\theta(X)$ and $b''(\theta) = \mathrm{Var}_\theta(X) \equiv \sigma^2(\theta)$ (Koopman, 1936). Furthermore a simple calculation shows that the Kullback-Leibler information numbers $I(\theta, \phi) = E_\theta \log\{f_\theta(X)/f_\phi(X)\}$ are given by $I(\theta, \phi) = (\theta - \phi)b'(\theta) - \{b(\theta) - b(\phi)\}$.

For the discussion in this section it suffices to choose $A_i = \alpha_i$ for $i = 0, 1$ which ensures that all the 2-SPRT's under consideration are in $\mathscr{T}(\alpha_0, \alpha_1)$. It will also be assumed that there are fixed but arbitrary constants C_1 and C_2 such that

$$0 < C_1 < \frac{\log \alpha_0}{\log \alpha_1} < C_2 < \infty. \tag{2.1}$$

Let $S_n = X_1 + X_2 + \cdots + X_n$ for $n = 1, 2, \cdots$, and define the log-likelihood ratios $\ell_i(\theta, n) = (\theta - \theta_i)S_n - n\{b(\theta) - b(\theta_i)\}$, $i = 0, 1$. Then (1.2) is equivalent to $\ell_i(\theta, n) > \log \alpha_i^{-1}$, and the 2-SPRT can be described graphically in the plane of n and S_n. There are two lines given by

$$(\theta - \theta_i)S_n - n\{b(\theta) - b(\theta_i)\} = \log \alpha_i^{-1}.$$

Defining $I_i(\theta) = I(\theta, \theta_i)$ and $a_i(\theta) = (\theta - \theta_i)/I_i(\theta)$ these lines intersect at $(n(\theta), v(\theta))$ where

$$n(\theta) = \left\{\frac{\log \alpha_i^{-1}}{I_1(\theta)} a_0(\theta) - \frac{\log \alpha_0^{-1}}{I_0(\theta)} a_1(\theta)\right\}\{a_0(\theta) - a_1(\theta)\}^{-1} \tag{2.2}$$

and

$$v(\theta) = n(\theta)b'(\theta) + \left\{\frac{\log \alpha_0^{-1}}{I_0(\theta)} - \frac{\log \alpha_1^{-1}}{I_1(\theta)}\right\}\{a_0(\theta) - a_1(\theta)\}^{-1}. \tag{2.3}$$

By virtue of the fact that $b(\theta)$ is convex, $I_0(\theta)$ is strictly increasing, $I_1(\theta)$ is strictly decreasing and both are positive on (θ_0, θ_1). Thus for any θ in (θ_0, θ_1), $a_1(\theta) < 0 < a_0(\theta)$, which implies that $n(\theta)$ is positive. Therefore, the two lines are converging, sampling is stopped as soon as the sequence $(1, S_1)$, $(2, S_2)$, $\cdots$ leaves the triangular region bounded by the lines, and the decision depends on which line is crossed.

The key to the theorem is to choose $\tilde{\theta}$ so that the supremum over θ of $E_\theta M(\tilde{\theta}, \alpha_0, \alpha_1)$ is attained at $\tilde{\theta}$, at least to within $o((n(\tilde{\theta}))^{1/2})$. To determine how to choose, $\tilde{\theta}$, first define θ^* so that

$$\frac{\log \alpha_0^{-1}}{I_0(\theta^*)} = \frac{\log \alpha_1^{-1}}{I_1(\theta^*)} \tag{2.4}$$

and let n^* be the common value of the two sides (which by (2.2) equals $n(\theta^*)$). Let a_i^*, I_i^*, and σ^* denote the values of $a_i(\theta)$, $I_i(\theta)$ and $\sigma(\theta)$ for $\theta = \theta^*$. Also let $M^* = \min(M_0^*, M_1^*)$ represent $M(\theta^*, \alpha_0, \alpha_1)$.

In the (n, S_n) plane, relation (2.3) shows that the line determined by the points $(n, E_{\theta^*}S_n) = (n, nb'(\theta^*))$ for $n = 1, 2, \cdots$ passes through the vertex $(n^*, v(\theta^*))$. So under θ^* the points (n, S_n) will tend to drift toward the vertex. In general however, $E_\theta M^*$ is not maximized at $\theta = \theta^*$. This is because for $n < n^*$ one of the boundaries will be closer to the line $(n, nb'(\theta^*))$ so that when θ^* is the true parameter value the fluctuations in S_n will cause the 2-SPRT to end too early by going over the closer boundary.

More precisely, essentially the same argument that will be used to show (2.9) of the theorem can be extended to show that for $\theta = \theta^* + c(n^*)^{-1/2}$ (where c is restricted to any bounded interval)

$$E_\theta M^* = n^* - \sigma^*(n^*)^{1/2}E(\max_{i=0,1}\{a_i^*(Z + \sigma^* c)\}) + o((n^*)^{1/2}), \tag{2.5}$$

where the expectation on the right-hand side is with respect to the standard normal variable Z. Choosing θ to maximize $E_\theta M^*$ to within $o((n^*)^{1/2})$ is then equivalent to finding c to minimize the expectation on the right-hand side. With $\Phi(x)$ and $\phi(x)$ as the standard normal distribution and density functions, respectively, straightforward calculation shows that the expectation equals

$$\sigma^* c a_1^* + (a_0^* - a_1^*)\{\sigma^* c \Phi(\sigma^* c) + \phi(\sigma^* c)\}.$$

Differentiation with respect to c shows that the minimum value occurs at $c = r^*/\sigma^*$ where

$$\Phi(r^*) = \frac{a_1^*}{(a_1^* - a_0^*)}. \tag{2.6}$$

In addition the value of the expectation at $c = r^*/\sigma^*$ is given by $(a_0^* - a_1^*)\phi(r^*)$.

Define $\tilde{\theta}$ and $\tilde{r}$ by

$$\tilde{\theta} = \theta^* + \frac{r^*}{\sigma^*(n^*)^{1/2}} \tag{2.7}$$

and

$$\Phi(\tilde{r}) = \frac{a_1(\tilde{\theta})}{a_1(\tilde{\theta}) - a_0(\tilde{\theta})}. \tag{2.8}$$

As above for M^*, let $\tilde{M} = M(\tilde{\theta}, \alpha_0, \alpha_1) = \min(\tilde{M}_0, \tilde{M}_1)$ and define $\tilde{n}$, $\tilde{I}_i$, $\tilde{\sigma}$ and $\tilde{a}_i$ accordingly. Then it turns out that the analog of (2.5) using $\tilde{M}$ in place of M^* is extremized by the same choice of c, as stated in the following theorem.

THEOREM. *Let $\tilde{\theta}$ and $\tilde{r}$ be determined by relations* (2.4), (2.6), (2.7) *and* (2.8). *If* (2.1) *is satisfied then as α_0 and $\alpha_1 \to 0$,*

$$\sup_\theta E_\theta \tilde{M} = \tilde{n} - \tilde{\sigma}(\tilde{a}_0 - \tilde{a}_1)\phi(\tilde{r})\tilde{n}^{1/2} + o(\tilde{n}^{1/2}) \tag{2.9}$$

and

$$n(\alpha_0, \alpha_1) = \tilde{n} - \tilde{\sigma}(\tilde{a}_0 - \tilde{a}_1)\phi(\tilde{r})\tilde{n}^{1/2} + o(\tilde{n}^{1/2}), \tag{2.10}$$

where $\phi(\cdot)$ is the standard normal density function. Thus

$$\frac{n(\alpha_0, \alpha_1)}{\sup_\theta E_\theta M} = 1 - o((\log \alpha_0^{-1})^{-1/2}). \tag{2.11}$$

Before giving the proof of the theorem in Section 4, the next section will describe the results of extensive computer calculations comparing the 2-SPRT with Kiefer-Weiss solutions and with the SPRT for the case of testing the mean of an exponential density.

3. Comparison of 2-SPRT's with Kiefer-Weiss solutions. The actual error probabilities of the 2-SPRT using $\tilde{M}$ can be evaluated asymptotically using the relations

$$P_{\theta_i}(\text{reject } \theta_i) = P_{\tilde{\theta}}(\text{reject } \theta_i)A_i E_{\tilde{\theta}}[\exp\{\log \alpha_i^{-1} - \ell_i(\tilde{\theta}, \tilde{M})\} \mid \text{reject } \theta_i],$$

for $i = 0, 1$. Using relation (4.5) of Section 4, the limit distribution of $T_m = S_m - mb'(\tilde{\theta})$ where $m = [\tilde{n} - \tilde{n}^{1/2} \log \tilde{n}]$ and arguments similar to those in Section 4.2, $P_{\tilde{\theta}}(\text{reject } \theta_0)$ is asymptotically $P(Z > -\tilde{r}) = \tilde{a}_1/(\tilde{a}_1 - \tilde{a}_0)$. Since $\ell_0(\tilde{\theta}, M) - \log \alpha_0^{-1}$ is the excess over the boundary when θ_0 is rejected, Theorem 5 of Lorden (1977) shows for the nonlattice case that

$$P_{\theta_0}(\text{reject } \theta_0) \sim \frac{\tilde{a}_1}{\tilde{a}_1 - \tilde{a}_0} A_0 \frac{L(\tilde{\theta}, \theta_0)}{I_0(\tilde{\theta})}$$

where $L(\tilde{\theta}, \theta_0)$ is defined by Lorden. A similar expression holds for the other error probability.

In practice it seems advisable to use this information in defining the test, so that the error probabilities attained by the 2-SPRT will be closer to those desired. The following formulation, used for the calculations presented here, is recommended for practical use. Define

$$A_0(\theta) = \frac{a_1(\theta) - a_0(\theta)}{a_1(\theta)} \alpha_0$$

and

$$A_1(\theta) = \frac{a_0(\theta) - a_1(\theta)}{a_0(\theta)} \alpha_1.$$

Choose ϕ^* to satisfy

$$\frac{\log\{A_0(\phi^*)\}^{-1}}{I_0(\phi^*)} = \frac{\log\{A_1(\phi^*)\}^{-1}}{I_1(\phi^*)}$$

and let $n(\phi^*)$ be the common value of the two sides. Let

$$\tilde{\phi} = \phi^* + \frac{r(\phi^*)}{\sigma(\phi^*)\{n(\phi^*)\}^{1/2}}$$

and use the 2-SPRT $\tilde{N} = M(\tilde{\phi}, A_0(\tilde{\phi}), A_1(\tilde{\phi}))$. The factors $L(\theta, \theta_0)/I_0(\theta)$ and $L(\theta, \theta_1)/I_1(\theta)$, which are corrections for the excess over the boundary, could be used in the definitions of $A_0(\theta)$ and $A_1(\theta)$. However, since these factors generally will be close to 1, and the computation of the L numbers is quite involved, their use is not recommended in practice. Also, the theorem now holds for $\tilde{N}$, with α_0 and α_1 replaced by $A_0(\tilde{\phi})$ and $A_1(\tilde{\phi})$ and $\tilde{n}$, $\tilde{\sigma}$, $\tilde{a}_i$ and $\tilde{r}$ determined by $\tilde{\phi}$, with the proof as contained in Section 4 going through nearly unchanged.

Calculations were carried out comparing the above 2-SPRT with Kiefer-Weiss solutions in the case of testing the parameter θ of the exponential density $f_\theta(x) = \theta\exp(-\theta x)$, $\theta > 0$, $x > 0$. In testing $\theta = \theta_0$ against $\theta = \theta_1$ it can be assumed that $\theta_0 = 1$, since that can always be achieved by scaling the X's. Desired values of α_0 and α_1 were used to define $\tilde{\phi}$ and $\tilde{N}$. $E_{\tilde{\phi}}N$, sup $E_\theta\tilde{N}$ and the actual error probabilities α_0' and α_1' of the 2-SPRT were computed. Then, as is described in Section 6 and Remark 1 of Lorden (1980), the boundaries of the Kiefer-Weiss solution with error probabilities α_0' and α_1' were calculated along with its operating characteristics. This provided the values of $n(\alpha_0', \alpha_1')$ used to compute the efficiency $n(\alpha_0', \alpha_1')/\sup_\theta E_\theta\tilde{N}$ of the 2-SPRT.

Figure 1 pictures both the 2-SPRT and Kiefer-Weiss boundaries attained by this process

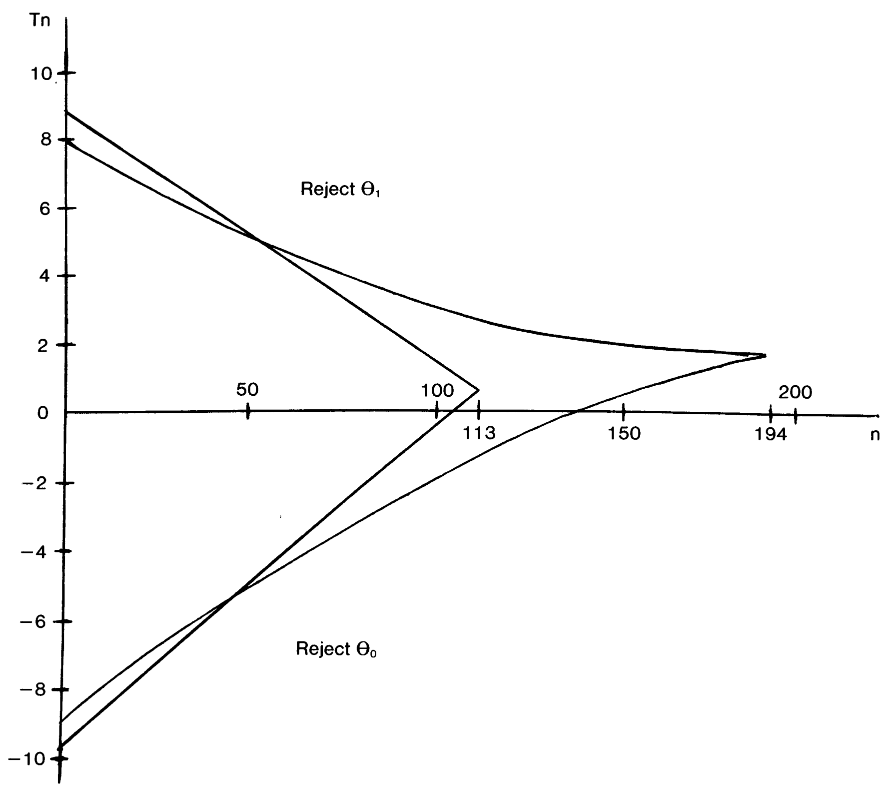

FIG. 1. 2-SPRT *and Kiefer-Weiss boundaries for testing* $\theta_0 = 1$ *against* $\theta_1 = 1.5$ *with* $\alpha_0' = 4.5\%$ *and* $\alpha_1' = 4.4\%$.

TABLE 1
Error probabilities and efficiencies, $\theta_0 = 1$, $\theta_1 = 2$

α_0	α_1	α_0'	α_1'	$n(\alpha_0', \alpha_1')$	$\sup_\theta E_\theta \tilde{N}$	% efficiency	$E_{\tilde{\phi}} \tilde{N}$
10	5	9.5	3.3	14.84	14.95	99.2	14.88
5	5	4.1	4.1	18.96	19.08	99.3	19.00
5	1	5.1	0.6	25.65	25.83	99.3	25.76
1	5	0.7	5.8	26.98	27.24	99.1	27.11
.1	5	0.06	8.2	37.02	37.60	98.4	37.35

α_i = *desired error probabilities (in %).*
α_i' = *actual error probabilities attained by* 2-SPRT *(in %).*

TABLE 2
Comparison of expected sample sizes of 2-SPRT *and* SPRT N,
$\theta_0 = 1$, $\theta_1 = 2$

α_0'	α_1'	$E_{\theta_0} \tilde{N}$	$E_{\theta_0} N$	$E_{\theta_1} \tilde{N}$	$E_{\theta_1} N$	$\sup_\theta E_\theta \tilde{N}$	$\sup_\theta E_\theta N$
9.5	3.3	10.60	9.93	12.33	11.15	14.95	16.48
4.1	4.1	11.47	10.36	16.27	15.08	19.08	21.43
5.1	0.6	17.30	16.08	18.21	15.22	25.83	31.85
0.7	5.8	12.33	10.07	24.38	23.13	27.24	31.41

for testing $\theta_0 = 1$ against $\theta_1 = 1.5$ with desired error probabilities of $\alpha_0 = \alpha_1 = 0.05$. The straight-line boundaries are those of the 2-SPRT, defined by $\tilde{\phi} = 1.25$, which had actual error probabilities of $\alpha_0' = 0.045$ and $\alpha_1' = 0.044$. The curved boundaries are those of the corresponding Kiefer-Weiss solution. For convenience these were drawn in the (n, T_n) plane where $T_n = S_n - E_{\tilde{\phi}} S_n = S_n - 0.8n$. For this case, $\sup_\theta E_\theta N = 51.72$ whereas $n(\alpha_0, \alpha_1) = 51.39$, resulting in an efficiency of 99.3%. A typical feature illustrated in Figure 1 is that the maximum possible number of observations with the 2-SPRT is much smaller than that of the Kiefer-Weiss solution. The truncation point of the 2-SPRT for this case is at 113 observations, while that of the Kiefer-Weiss solution is at 194 observations.

The most extensive calculations were carried out for the case $\theta_1 = 2$ and are recorded in Table 1. The 2-SPRT is seen to have an efficiency of over 98% for a broad range of desired error probabilities, with both the efficiency and the closeness of the actual error probabilities to the desired ones decreasing as the ratio of α_0 to α_1 becomes extreme. The last column records the values of $E_{\tilde{\phi}} \tilde{N}$, which are in general within 0.5% of $\sup_\theta E_\theta \tilde{N}$ indicating that $\tilde{\phi}$ indeed nearly maximizes $E_\theta \tilde{N}$. Lorden (1976) indicated that in the symmetric normal case the observed efficiencies depended on the desired error probabilities, but that over a broad range they depended hardly at all on the parameter values. To confirm this for the exponential density, two cases were computed for $\theta_1 = 1.5$. As stated earlier, the $\alpha_0 = \alpha_1 = 0.05$ case resulted in 99.3% efficiency. The case $\alpha_0 = 0.1$, $\alpha_1 = 0.05$ attained $\alpha_0' = 0.11$ and $\alpha_1' = 0.035$ with an efficiency of 99.2%. Both of these efficiencies agree exactly with the corresponding cases for $\theta_1 = 2$.

In addition to the characteristics already mentioned, $E_{\theta_0} \tilde{N}$ and $E_{\theta_1} \tilde{N}$ were computed. For the exponential case, Dvoretsky, Kiefer and Wolfowitz (1950) provide exact formulas for the operating characteristic and expected sample sizes of an SPRT. Typical results comparing $E_{\theta_0} \tilde{N}$, $E_{\theta_1} \tilde{N}$ and $\sup_\theta E_\theta \tilde{N}$ with the corresponding quantities for the SPRT are recorded in Table 2. For these cases the 2-SPRT requires on the average between 5% and 25% more observations than the SPRT if the null or alternative hypothesis is true, while the SPRT requires between 10% and 25% more observations than the 2-SPRT at the maximum expected sample size. So here the 2-SPRT will take between 1 and 3 more observations at θ_0 or θ_1 but save up to 6 observations for θ between θ_0 and θ_1.

4. Proof of the main result. The proof of the theorem consists of establishing relations (4.1) – (4.3) below. (2.9) follows immediately from

$$E_{\tilde\theta}\tilde M \geq \tilde n - \tilde\sigma(\tilde a_0 - \tilde a_1)\phi(\tilde r)\tilde n^{1/2} - o(\tilde n^{1/2}) \tag{4.1}$$

and

$$\sup_\theta E_\theta \tilde M \leq \tilde n - \tilde\sigma(\tilde a_0 - \tilde a_1)\phi(\tilde r)\tilde n^{1/2} + o(\tilde n^{1/2}), \tag{4.2}$$

while (2.10) follows from these relations together with

$$\inf_{\mathscr{T}(\alpha_0,\alpha_1)} E_{\tilde\theta} N \geq E_{\tilde\theta}\tilde M - O(1). \tag{4.3}$$

It will be assumed in the remainder of this section that α_0 and α_1 are small enough so that θ^* and $\tilde\theta$ are in a closed subinterval $[\phi_0, \phi_1]$ of (θ_0, θ_1). This assures that a_i^*, $\tilde a_i$ and the information numbers are bounded away from 0 and ∞.

Before continuing to the proofs of (4.1) – (4.3) given in the following subsections, some relationships concerning the boundaries of the 2-SPRT will be given. Let $T_n = S_n - nb'(\tilde\theta)$ for $n = 1, 2, \cdots$ and let $\tilde s = v(\tilde\theta) - \tilde n b'(\tilde\theta)$. In the (n, T_n) plane, the boundaries of the continuation region for $\tilde M$ are lines with slope $-1/\tilde a_i$, intercept $\log \alpha_i^{-1}/(\tilde\theta - \theta_i)$, passing through the vertex $(\tilde n, \tilde s)$. So the boundaries are given by

$$U_i(n) = \tilde s + (\tilde n - n)/\tilde a_i = \log \alpha_i^{-1}/(\tilde\theta - \tilde\theta_i) - n/\tilde a_i.$$

(Sampling is stopped as soon as either $T_n \geq U_0(n)$ or $T_n \leq U_1(n)$, the decision depending on which inequality holds.) Solving for $\log \alpha_i^{-1}$ yields

$$\log \alpha_i^{-1} = (\tilde n + \tilde a_i \tilde s)\tilde I_i. \tag{4.4}$$

The final relationship is given by

$$\frac{\tilde s}{\tilde\sigma(\tilde n)^{1/2}} = -\tilde r + o(1). \tag{4.5}$$

To show (4.5) it suffices to establish the same relationship with σ^*, n^* and r^* in place of $\tilde\sigma$, $\tilde n$ and $\tilde r$. This in turn follows in a straightforward manner by using the definitions of $\tilde s$ and n^* and Taylor expansions for the information numbers involved.

4.1. *Proof of* (4.3). As pointed out, $\tilde a_i$, $\tilde I_i$ and $\tilde\sigma$ are bounded away from 0 and ∞, so that the following lemma yields (4.3).

LEMMA 4.1. *If* $\theta \in (\theta_0, \theta_1)$, *then*

$$E_\theta M(\theta, \alpha_0, \alpha_1) - \inf_{\mathscr{T}(\alpha_0,\alpha_1)} E_\theta N \leq \textstyle\sum_{i=0}^1 \left\{ a_i^2(\theta)\sigma^2(\theta) + \dfrac{\log 2}{I_i(\theta)} \right\}. \tag{4.6}$$

PROOF. Let M and M_i denote $M(\theta, \alpha_0, \alpha_1)$ and $M_i(\theta, \alpha_i)$ respectively. Let (N, D) be any test in $\mathscr{T}(\alpha_0, \alpha_1)$, and let $\{D = i\}$ be the event that θ_i is rejected by that test. As in the proof of Theorem 1 in Lorden (1972), define $N_i = \min(M_i, N\{D = i\})$ where $N\{D = i\} = N$ if $D = i$ and ∞ otherwise. Clearly for all θ

$$M - N \leq \textstyle\sum_{i=0}^1 (M_i - N_i). \tag{4.7}$$

By Wald's equation $I_i(\theta)E_\theta M_i = \log \alpha_i^{-1} + \delta$, where $\delta = E_\theta\{\ell_i(\theta, M_i) - \log \alpha_i^{-1}\}$ is the expected excess over the boundary $\log \alpha_i^{-1}$. By Theorem 1 of Lorden (1970),

$$\delta \leq \frac{\mathrm{Var}_\theta \ell_i(\theta, 1)}{I_i(\theta)} = \frac{(\theta - \theta_i)^2\sigma^2(\theta)}{I_i(\theta)}$$

which implies

$$E_\theta M_i \leq \frac{\log \alpha_i^{-1}}{I_i(\theta)} + a_i^2(\theta)\sigma^2(\theta). \tag{4.8}$$

An argument similar to the proof of Wald's lower bounds on the expected sample size of a sequential test (Wald, 1947, page 197) shows $I_i(\theta)\cdot E_\theta N_i \geq -\log P_{\theta_i}(N_i < \infty)$. Combining this with the inequality $P_{\theta_i}(N_i < \infty) \leq 2\alpha_i$ yields

$$E_\theta N_i > \frac{\log \alpha_i^{-1} - \log 2}{I_i(\theta)}. \tag{4.9}$$

Taking expectations in (4.7) and using (4.8) and (4.9) shows that (4.6) is true.

4.2 *Proof of* (4.1). Let $m = [\tilde{n} - \tilde{n}^{1/2}\log \tilde{n}]$. The derivation of (4.1) relies mainly on two facts. First, with overwhelming probability under $\tilde{\theta}$, the test requires at least m observations; and second, once the m observations are taken, the behavior of the remainder of the test is sufficiently predictable by the value of T_m. More specifically, the first claim is given by

$$P_{\tilde{\theta}}(\tilde{M} \leq m) \leq O(\tilde{n}^{-1}) \tag{4.10}$$

and will be proved in Lemma 4.4 at the end of this section, while the second is given by the following lemma.

LEMMA 4.2. *On the event* $\{\tilde{M} > m\}$

$$E_{\tilde{\theta}}(\tilde{M} \mid X_1, \cdots, X_m) \geq \tilde{n} - \max_{i=0,1}\{\tilde{a}_i(T_m - \tilde{s})\} - o(\tilde{n}^{1/2}). \tag{4.11}$$

PROOF. At time m the log-likelihood ratios have values $\ell_i(\tilde{\theta}, m) = (\tilde{a}_i T_m + m)\tilde{I}_i$ for $i = 0, 1$. If $\tilde{M} > m$ then based on observations $Y_1, Y_2, \cdots$ where $Y_k = X_{m+k}$, let N_i be the first n such that $\ell_i(\tilde{\theta}, n) \geq K_i = \log \alpha_i^{-1} - (\tilde{a}_i T_m + m)\tilde{I}_i$. Then on $\{\tilde{M} > m\}$,

$$E_{\tilde{\theta}}(\tilde{M} \mid X_1, \cdots, X_m) = m + E_{\tilde{\theta}}\{\min(N_0, N_1)\}. \tag{4.12}$$

By Lemma 4.3 below there is a constant D such that

$$E_{\tilde{\theta}}\{\min(N_0, N_1)\} \geq \min_{i=0,1}\{K_i/\tilde{I}_i\} - D(\min_{i=0,1}\{K_i/\tilde{I}_i\})^{1/2}. \tag{4.13}$$

Substituting for $\log \alpha_i^{-1}$ according to (4.4) shows

$$\min_{i=0,1}\{K_i/I_i\} = \tilde{n} - m - \max_{i=0,1}\{\tilde{a}_i(T_m - \tilde{s})\} \tag{4.14}$$

which is at most $\tilde{n}^{1/2} \log \tilde{n}$. Using (4.13) and (4.14) in (4.12) yields relation (4.11).

The following lemma establishes (4.13) by giving a general lower bound on $E_\theta M(\theta, A_0, A_1)$.

LEMMA 4.3. *Let* $D = \frac{1}{2} \max_{[\phi_0,\phi_1]}[\{a_0(\theta) - a_1(\theta)\}\sigma(\theta)]$ *and* $K = \min_{i=0,1}\{\log A_i^{-1}/I_i(\theta)\}$. *Then for any* θ *in* $[\phi_0, \phi_i]$

$$E_\theta M(\theta, A_0, A_1) \geq K - D\,K^{1/2}. \tag{4.15}$$

PROOF. As in the proof of inequality (1.4) in Hoeffding (1960), define $Y_{in} = \ell_i(\theta, n)/I_i(\theta)$ for $i = 0, 1$ and $n = 1, 2, \cdots$, and $Y_n = Y_{0,n} - Y_{1,n}$. Clearly $K \leq \max(Y_{0,M}, Y_{1,M}) = \frac{1}{2}(Y_{0,M} + Y_{1,M}) + \frac{1}{2}\,|Y_M|$. Therefore, since $E_\theta Y_{i,M} = E_\theta M$ by Wald's equation,

$$K \leq E_\theta M + \tfrac{1}{2}E_\theta |Y_M|. \tag{4.16}$$

Since $E_\theta Y_1 = 0$, Wald's second moment equation and a simple computation show

$$E_\theta(Y_M^2) = (E_\theta M)\mathrm{Var}_\theta Y_1 = \{a_0(\theta) - a_1(\theta)\}^2\sigma^2(\theta)E_\theta M. \tag{4.17}$$

Combining (4.16) and (4.17) with the inequality $E_\theta |Y_M| \leq (E_\theta Y^2)^{1/2}$ and the definition of D leads to $K \leq E_\theta M + D(E_\theta M)^{1/2}$ from which (4.15) follows easily, proving the lemma.

Lemma 4.2 and the estimate of $P_{\tilde{\theta}}(\tilde{M} \leq m)$ in (4.10) give

$$E_{\tilde{\theta}} M \geq \tilde{n} - E_{\tilde{\theta}}(\max_{i=0,1}\{\tilde{a}_i(T_m - \tilde{s})\}) + o(\tilde{n}^{1/2}).$$

The expectation on the right-hand side can be written

$$\tilde{\sigma}m^{1/2}\int_0^\infty P_{\tilde{\theta}}[\max_{i=0,1}\{\tilde{a}_i(T_m-\tilde{s})/\tilde{\sigma}m^{1/2}\}>t]\,dt.$$

The integrand is the sum of the probabilities of the inequality holding for $i = 0$ and for $i = 1$. In the case $i = 0$, for example, this equals

$$P_{\tilde{\theta}}\left(\frac{T_m}{\tilde{\sigma}m^{1/2}}>\frac{t}{\tilde{a}_0}+\frac{s}{\tilde{\sigma}m^{1/2}}\right)=P\left(Z>\frac{t}{\tilde{a}_0}+\frac{s}{\tilde{\sigma}m^{1/2}}\right)+O(m^{-1/2}),$$

where Z is standard normal, by virtue of the Berry-Esseen theorem (Feller, 1971, page 542), and the fact that $\mathrm{Var}_{\tilde{\theta}}T_i$ and $E_{\tilde{\theta}}(|T_1|^3)$ are bounded away from 0 and ∞, respectively. Since $\tilde{s}/\tilde{\sigma}\,\tilde{n}^{1/2}$—and hence $\tilde{s}/\tilde{\sigma}\,m^{1/2}$—tends to $-\tilde{r}$, the first term on the right-hand side converges to $P\{\tilde{a}_0(Z+\tilde{r})>t\}$. Together with a similar result for $i = 1$, this implies that the above integrand converges pointwise to $P(\max_{i=0,1}\{\tilde{a}_i(Z+\tilde{r})\}>t)$. Chebyshev's inequality shows that the integrand is also bounded above by a function that goes to 0 like t^{-2} as $t\to\infty$. Therefore, by the dominated convergence theorem

$$E_{\tilde{\theta}}\tilde{M}>\tilde{n}-\tilde{\sigma}\tilde{n}^{1/2}E(\max_{i=0,1}\{\tilde{a}_i(Z+\tilde{r})\})+o(\tilde{n}^{1/2}).$$

Evaluation of the expectation on the right now yields (4.1).

To complete the proof of (4.1), then, it remains only to establish (4.10), which is contained in the following lemma (also to be used in the proof of (4.3)).

LEMMA 4.4. *Let B be a positive constant. Then $P_\theta(\tilde{M}<m)\le O(\tilde{n}^{-1})$ uniformly for $|\theta-\tilde{\theta}|<B\tilde{n}^{-1/2}$.*

PROOF. It suffices to show

$$P_\theta\{T_k\ge U_0(k)\}=O(\tilde{n}^{-2}),\quad k=1,\cdots,m, \tag{4.18}$$

uniformly in k and $|\theta-\tilde{\theta}|\le B\tilde{n}^{-1/2}$. Combined with a similar bound for $P_\theta\{T_k\le U_1(k)\}$, the lemma follows by summing over $k = 1, \cdots, m$.

The boundedness of $\tilde{a}_0$ together with (4.5) and the equations for the boundary $U_0(n)$ imply there is a positive constant c such that $U_0(k)\ge c\tilde{n}^{1/2}\log\tilde{n}\equiv\gamma$ for all $k\le m$. Now for any $t\ge 0$, $P_\theta(T_k\ge\gamma)=P_\theta[\exp\{t(T_k-\gamma)\}\ge 1]$, so that Chebyshev's inequality gives

$$\begin{aligned}P_\theta(T_k\ge\gamma)&\le\exp(-\gamma t)[E_\theta\{\exp(tT_1)\}]^k\\&=\exp[-\gamma t+k\{b(t+\theta)-b(\theta)-tb'(\tilde{\theta})\}].\end{aligned}$$

Taylor expansions of $b(t+\theta)$ about θ and $b'(\theta)$ about $\tilde{\theta}$ yield positive constants q and q' such that

$$P_\theta(T_k\ge\gamma)\le\exp(-\gamma t+ktq'B\tilde{n}^{-1/2}+qkt^2).$$

Replacing k by n on the right-hand side and setting $t=2/(c\tilde{n}^{1/2})$ establishes (4.18).

4.3. *Proof of* (4.2). The proof of (4.2) is divided into two parts; the first is to show that there is a $B > 0$ such that

$$\sup_{\{|\theta-\tilde{\theta}|>B\tilde{n}^{-1/2}\}}E_\theta M\le\tilde{n}-\tilde{\sigma}(\tilde{a}_0-\tilde{a}_1)\phi(\tilde{r})\tilde{n}^{1/2}+o(\tilde{n}^{1/2}). \tag{4.19}$$

Only the case $\theta>\tilde{\theta}$ will be considered, as the case $\theta<\tilde{\theta}$ is similar.

The argument for (4.19) in the case $\theta>\tilde{\theta}$ involves comparing $\tilde{M}$ with $\tilde{M}_0$. Theorem 1 of Lorden (1970) shows that the expected excess of $\tilde{M}_0$ over $\log\alpha_0^{-1}$ is at most

$$\frac{(\theta-\theta_0)^2\sigma^2(\theta)}{E_\theta\ell_0(\tilde{\theta},1)}. \tag{4.20}$$

For $B' > 0$ (to be chosen below) there is a $B > 0$ such that $\theta > \tilde{\theta} + B\tilde{n}^{-1/2}$ implies $E_\theta \ell_0(\tilde{\theta}, 1) \geq \tilde{I}_0 + B'\tilde{n}^{-1/2}$. Thus for $\theta > \tilde{\theta} + B\tilde{n}^{-1/2}$, (4.20) is bounded, so that by Wald's equation

$$E_\theta \tilde{M} \leq E_\theta \tilde{M}_0 \leq \frac{\log \alpha_0^{-1}}{\tilde{I}_0 + B'\,\tilde{n}^{-1/2}} + O(1).$$

Replacing $\log \alpha_0^{-1}$ according to (4.4) and using (4.5) yields

$$E_\theta \tilde{M} \leq \tilde{n} - (B'\tilde{I}_0^{-1} + \tilde{a}_0 \tilde{r}\tilde{\sigma})\tilde{n}^{1/2} + o(\tilde{n}^{1/2}).$$

Choose B' sufficiently large that (4.19) follows.

For the remainder of this section, let J denote the interval of values given by $|\theta - \tilde{\theta}| < B\tilde{n}^{-1/2}$. The second part of the proof of (4.2) parallels Lemma 4.2 by establishing a bound on the conditional expectation of $\tilde{M}$ given continuation past m.

LEMMA 4.5. *On the event* $\{\tilde{M} > m\}$

$$E_\theta(\tilde{M} \mid X_1, \cdots, X_m) \leq \tilde{n} - \max_{i=0,1}\{\tilde{a}_i(T_m - \tilde{s})\} + o(\tilde{n}^{1/2}) \tag{4.21}$$

uniformly for θ *in* J.

PROOF. Assume $T_m \geq \tilde{s}$; the proof for $T_m < \tilde{s}$ is similar. Define N_0 and K_0 as in the proof of Lemma 4.2. Under θ the expected excess of N_0 over K_0 is bounded above by (4.20), and is thus bounded uniformly on J. Hence

$$E_\theta(\tilde{M} \mid X_1, \cdots, X_m) \leq m + \frac{\{\tilde{n} - \tilde{a}_0(T_m - \tilde{s}) - m\}I_0(\tilde{\theta})}{E_\theta \ell_0(\tilde{\theta}, 1)} + O(1) \tag{4.22}$$

uniformly for θ in J. Since the ratio of $I_0(\tilde{\theta})$ to $E_\theta \ell_0(\tilde{\theta}, 1)$ is uniformly $1 + O(\tilde{n}^{1/2})$ and $\tilde{n} - \tilde{a}_0(T_m - \tilde{s}) - m \leq \tilde{n} - m \leq O(\tilde{n}^{1/2}\log \tilde{n})$, (4.22) implies that on $\{\tilde{M} > m\}$

$$E_\theta(\tilde{M} \mid X_1, \cdots, X_m) \leq \tilde{n} - \tilde{a}_0(T_m - \tilde{s}) + O(\log \tilde{n})$$

uniformly for θ in J, which yields (4.21) for the case $T_m \geq \tilde{s}$, proving the lemma.

From (4.21)

$$E_\theta \tilde{M} \leq \tilde{n} - \tilde{\sigma}\tilde{n}^{1/2} E_\theta[\max_{i=0,1}\{\tilde{a}_i(T_m - \tilde{s})/\tilde{\sigma}\tilde{n}^{1/2}\} 1\{\tilde{M} > m\}] + o(\tilde{n}^{1/2}),$$

where $1\{\cdot\}$ denotes the indicator function and the inequality holds uniformly for θ in J. To complete the proof of (4.2) it will suffice to show

$$E_\theta[\max_{i=0,1}\{\tilde{a}_i(T_m - \tilde{s})/\tilde{\sigma}\tilde{n}^{1/2}\} 1\{\tilde{M} > m\}] \geq \inf_r E[\max_{i=0,1}\{\tilde{a}_i(Z + r)\}] - o(1) \tag{4.23}$$

uniformly on J, since the right-hand side is at least $(\tilde{a}_0 - \tilde{a}_1)\phi(\tilde{r}) + o(1)$. To prove (4.23) note that by arguing as in the paragraph following the proof of Lemma 4.3, for $t > 0$.

$$P_\theta[\max_{i=0,1}\{\tilde{a}_i(T_m - \tilde{s})/\tilde{\sigma}\tilde{n}^{1/2}\} > t] = P[\max_{i=0,1}\{\tilde{a}_i(Z + m^{1/2}\tilde{\sigma}^{-1}E_\theta T_1 + \tilde{r})\} > t] + o(1)$$

uniformly in J.

Since $P_\theta(\tilde{M} \leq m) \to 0$ uniformly by Lemma 4.4, the last relation yields for fixed $L > 0$

$$\begin{aligned} &\int_0^L P_\theta[\max_{i=0,1}\{\tilde{a}_i(T_m - \tilde{s})/\tilde{\sigma}\tilde{n}^{1/2}\} 1\{\tilde{M} > m\} > t]\, dt \\ &\qquad = \int_0^L P_\theta[\max_{i=0,1}\{\tilde{a}_i(Z + m^{1/2}\tilde{\sigma}^{-1}E_\theta T_1 + \tilde{r})\} > t]\, dt + o(1) \end{aligned} \tag{4.24}$$

uniformly for θ in J. Because $\tilde{r} + m^{1/2}\tilde{\sigma}^{-1}E_\theta T_1 = \tilde{r} + m^{1/2}\tilde{\sigma}^{-1}\{b'(\theta) - b'(\tilde{\theta})\}$ is bounded for θ in J, there is a Q such that the integral on the right-hand side of (4.24) is at least

$$\inf_{|r|\le Q}\int_0^L P[\max_{i=0,1}\{\tilde{a}_i(Z+r)\}>t]\,dt$$

$$\ge \inf_{|r|\le Q}\int_0^\infty P[\max_{i=0,1}\{\tilde{a}_i(Z+r)\}>t]\,dt-\int_L^\infty g(t)\,dt, \tag{4.25}$$

where $g(\cdot)$ is an integrable function which can be chosen to dominate the integrands (since the range of r is bounded).

(4.24) and (4.25) establish (4.23) to within the last term in (4.25), which can be made arbitrarily small by choosing L large. Thus, (4.23) follows and the proof of (4.2) and, hence, the theorem is complete.

Acknowledgment. The author would like to thank his thesis advisor, Dr. Gary Lorden, for suggesting this problem and for his invaluable guidance in directing this research.

REFERENCES

ANDERSON, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size. *Ann. Math. Statist.* **31** 165–197.

DVORETSKY, A., KIEFER, J. and WOLFOWITZ, J. (1950). Sequential decision problems for processes with continuous time parameter. Testing hypotheses. *Ann. Math. Statist.* **24** 254–264.

FELLER, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. 2. Wiley, New York.

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

KIEFER, J. and WEISS, L. (1957). Some properties of generalized sequential probability ratio tests. *Ann. Math. Statist.* **28** 57–75.

KOOPMAN, B. O. (1936). On distributions admitting a sufficient statistic. *Trans. Amer. Math. Soc.* **39** 399–409.

LAI, T. L. (1973). Optimal stopping and sequential tests which minimize the maximum expected sample size. *Ann. Statist.* **1** 659–673.

LORDEN, G. (1976). 2-SPRT's and the modified Kiefer-Weiss problem of minimizing an expected sample size. *Ann. Statist.* **4** 281–291.

LORDEN, G. (1970). On excess over the boundary. *Ann. Math. Statist.* **41** 520–527.

LORDEN, G. (1972). Likelihood ratio tests for sequential k-decision problems. *Ann. Math. Statist.* **43** 1412–1427.

LORDEN, G. (1977). Nearly-optimal sequential tests for finitely many parameter values. *Ann. Statist.* **5** 1–21.

LORDEN, G. (1980). Structure of sequential tests minimizing an expected sample size. *Z. Wahrsch. verw. Gebiete* **51** 291–302.

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

WALD, A. and WOLFOWITZ, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* **19** 326–339.

WEISS, L. (1962). On sequential tests which minimize the maximum expected sample size. *J. Amer. Statist. Assoc.* **57** 551–566.

DEPARTMENT OF MATHEMATICAL SCIENCES
MONTANA STATE UNIVERSITY
BOZEMAN, MONTANA 59717