The Annals of Statistics
1976, Vol. 4, No. 2, 281-291

2-SPRT'S AND THE MODIFIED KIEFER-WEISS PROBLEM OF MINIMIZING AN EXPECTED SAMPLE SIZE

BY GARY LORDEN[1]

California Institute of Technology

A simple combination of one-sided sequential probability ratio tests, called a 2-SPRT, is shown to approximately minimize the expected sample size at a given point θ_0 among all tests with error probabilities controlled at two other points, θ_1 and θ_2. In the symmetric normal and binomial testing problems, this result applies directly to the Kiefer-Weiss problem of minimizing the maximum over θ of the expected sample size.

Extensive computer calculations for the normal case indicate that 2-SPRT's have efficiencies greater than 99% regardless of the size of the error probabilities. Accurate approximations to the error probabilities and expected sample sizes of these tests are given.

1. Introduction. A substantial part of the development of sequential analysis has been directed toward improving the performance of the sequential probability ratio test (SPRT) by reducing the expected sample size for parameter values between the hypotheses. Kiefer and Weiss [9] proved structure theorems about optimal tests for several formulations, including the problem of minimizing the expected sample size, $E_{\theta_0}N$, at a point θ_0, subject to error probability bounds, α and β, at two other points, θ_1 and θ_2. Bechhofer [2] pointed out the desirability of minimizing the maximum expected sample size over all possible θ, and Weiss [17] showed how that problem, the so-called Kiefer-Weiss problem, reduces to the previous formulation in symmetric cases involving normal and binomial distributions. Weiss used this reduction to obtain properties of the solution of the Kiefer-Weiss problem in these cases. Recently, Lai [10] investigated the Wiener process case.

The present paper is concerned with the problem of minimizing $E_{\theta_0}N$, i.e., the modified Kiefer-Weiss problem which was shown in [9] to be equivalent to the Bayes problem of minimizing a weighted average of $E_{\theta_0}N$ and the two error probabilities. For the class of parametric families in [9], which includes the Koopman-Darmois families, the solutions to this problem have bounded sample size. Hence, the backwards iterative procedure for solving n-stage Bayes problems ([3], [14], [16]) is applicable. As Weiss [17] and Lai [10] point out, heavy computational work is required to carry out this algorithm, rendering it unsuitable for routine use.

Anderson [1] showed, however, that in the Wiener process case with $\alpha = \beta$, remarkably high efficiencies are attainable by using a pair of converging straight

---

Received June 1974; revised July 1975.

[1] Research supported by the National Science Foundation under Grant GP-37819.

AMS 1970 *subject classifications.* Primary 62L10; Secondary 62F20.

Key words and phrases. Sequential probability ratio test, Bayes solution, asymptotic optimality.

lines as boundaries. He derived formulas for the exact operating characteristics of tests in this class and chose boundary slopes to optimize the results, which were then compared with a highly accurate lower bound of Hoeffding [7].

In the general context of Koopman–Darmois families, a single straight-line boundary in the plane of n and S_n (the cumulative sum sufficient statistic) corresponds to a one-sided SPRT. Thus, Anderson's approach can be viewed as performing simultaneously a pair of one-sided SPRT's, one for the (possible) rejection of θ_1 and the other for rejection of θ_2. The tests of the present paper, called 2-SPRT's, are particularly natural choices within that class of tests. A one-sided SPRT of θ_0 vs. θ_1 is used to reject θ_1, while another one-sided SPRT of θ_0 vs. θ_2 is used to reject θ_2, with a fixed (but arbitrary) rule for choosing between the two in case both SPRT's stop at the same time.

An important feature of 2-SPRT's is the applicability of Wald's simple error probability bounds [13]. These bounds are particularly useful in symmetric cases, where they can be improved by a factor of two. Generalizations of Wald's approach are considered in [5] and [11].

The results of extensive computer calculations for the symmetric normal case are given in Section 3. Over the broad range of parameter values studied, the 2-SPRT minimizes $E_{\theta_0}N$, and hence the maximum $E_\theta N$, within 1 % for all sample sizes and error probabilities. These results suggest that most of the difference between the expected sample sizes of Anderson's tests for the Wiener process and Hoeffding's lower bound is due to the latter's underestimation of the true minimum. The high efficiencies attained by the 2-SPRT's suggest, however, that Anderson's different choice of boundary slopes to minimize $E_{\theta_0}N$ offers no significant savings over the 2-SPRT's. The latter tests are especially appealing in the symmetric normal case because of the availability of highly accurate approximations to their operating characteristics based on Wald's approach to estimating SPRT error probabilities. These approximations are given in Section 3.

2. Approximate optimality of the 2-SPRT. The modified Kiefer–Weiss problem is formulated as follows. Observations $X_1, X_2, \cdots$ are random variables on a sample space $(\Omega, \mathscr{F})$ on which the true probability measure is one of three measures, F, G, and P. Under each of these, the X's are independent and identically distributed. Probability densities (i.e., Radon–Nikodym derivatives) of $(X_1, \cdots, X_n)$ are taken with respect to a suitable σ-finite measure (e.g., the one induced by $(F + G + P)/3$) and are denoted by f_n, g_n, and p_n, respectively. Write f_n/p_n, g_n/p_n, etc., for the usual likelihood ratios $f_1(X_1) \cdots f_1(X_n)/p_1(X_1) \cdots p_1(X_n)$, etc., and define $f_0 \equiv g_0 \equiv p_0 \equiv 1$. Let E denote expectation under P. A test, $(N, \hat{N})$, is a stopping rule N together with a terminal decision rule $\hat{N}$.

To allow $N = 0$ and ∞, it is required that N be an extended stopping variable with respect to $\mathscr{F}_0 = \{\emptyset, \Omega\}$, $\mathscr{F}_1, \mathscr{F}_2, \cdots$, the sigma-fields generated by the X's. The terminal decision $\hat{N}$ is defined on $\{N < \infty\}$, takes the values F and G, and satisfies $\{N = n, \hat{N} = F\} \in \mathscr{F}_n$ for $n = 0, 1, \cdots$. The framework is easily extended to allow randomized stopping variables and decisions, but the problem is essentially unchanged (see, for example, Theorem 5.3 of [3]). Without loss of generality, only tests satisfying $P(N < \infty) = 1$ are considered. However, $F(N < \infty)$ and $G(N < \infty)$ may be less than one. It is assumed in the sequel that F, G, and P are mutually absolutely continuous and distinct, and that

$$E \log^2 \left(\frac{p_1}{f_1}\right) \quad \text{and} \quad E \log^2 \left(\frac{p_1}{g_1}\right) \quad \text{are finite.} \tag{1}$$

The distinctness of F, G, and P insures that the information numbers

$$I(P, F) = E \log \frac{p_1}{f_1} \quad \text{and} \quad I(P, G) = E \log \frac{p_1}{g_1}$$

are positive.

The error probabilities of $(N, \hat{N})$ are expressible in the form

$$F(\hat{N} = G) = E \frac{f_N}{p_N} \mathbf{1}\{\hat{N} = G\} \quad \text{and} \quad G(\hat{N} = F) = E \frac{g_N}{p_N} \mathbf{1}\{\hat{N} = F\}, \tag{2}$$

where $\mathbf{1}\{\ \}$ denotes the indicator function ($= 1$ if the event occurs, 0 if it doesn't). These relations are proved by essentially the same argument Wald used to derive error probability approximations for SPRT's [13] (see the equality case of Lemma 3 of [12]).

Given $0 \leqq A, B \leqq 1$, not both zero, define a 2-SPRT by the stopping time $M(A, B) =$ smallest $n \geqq 0$ (or ∞ if there is no n) such that

$$\frac{f_n}{p_n} \leqq A \quad \text{or} \quad \frac{g_n}{p_n} \leqq B, \tag{3}$$

with the following requirement for $\hat{M}$. Choose $\hat{M} = F$ if the first inequality in (3) is not satisfied and $\hat{M} = G$ if the second inequality is not satisfied. For Theorem 1 it is immaterial what rule is used for choosing between F and G when both inequalities in (3) are satisfied. (In Section 3 for the symmetric normal case, $\hat{M} = F$ was chosen whenever $f_M > g_M$.) Note that $EM(A, B) < \infty$, since $I(P, F)$ and $I(P, G)$ are positive.

Applying (2) and (3) with the requirement for $\hat{M}$,

$$F(\hat{M} = G) \leqq AP(\hat{M} = G) \quad \text{and} \quad G(\hat{M} = F) \leqq BP(\hat{M} = F). \tag{4}$$

Thus, the sum of the error probabilities is $\leqq \max(A, B)$. In symmetric cases, $P(\hat{M} = G) = P(\hat{M} = F) = \frac{1}{2}$, which makes the bounds in (4) more useful.

THEOREM 1. *Let $\alpha(A, B)$ and $\beta(A, B)$ denote the error probabilities of the* 2-SPRT $(M(A, B), \hat{M})$. *Let $n(A, B)$ denote the infimum of EN over all tests satisfying $\alpha \leqq \alpha(A, B)$ and $\beta \leqq \beta(A, B)$. Under assumption* (1), *if* $A,B > 0$,

$$EM(A, B) - n(A, B) \to 0 \qquad \text{as } \min(A, B) \to 0.$$

PROOF. Following the pattern of Wald and Wolfowitz ([15]), define the integrated risk of $(N, \hat{N})$ for given $u, v \geqq 0$ as

$$r(N, \hat{N}) = EN + uF(\hat{N} = G) + vG(\hat{N} = F) . \tag{5}$$

By (2), this can be written

$$\begin{aligned} r(N, \hat{N}) &= E(N + U_N \mathbf{1}\{\hat{N} = G\} + V_N \mathbf{1}\{\hat{N} = F\}) \\ &\geqq E(N + \min (U_N, V_N)) , \end{aligned} \tag{6}$$

where

$$U_n = u \frac{f_n}{p_n} \quad \text{and} \quad V_n = v \frac{g_n}{p_n} \qquad n = 0, 1, \cdots .$$

Define the Bayes risk, $R(u, v)$, as the minimum integrated risk over all $(N, \hat{N})$. Since choosing $\hat{N} = F$ or $\hat{N} = G$ without sampling achieves integrated risk v or u, respectively, evidently

$$R(u, v) \leqq \min (u, v) , \tag{7}$$

and sampling is called for only if strict inequality holds. It is well known ([14], [16]) that the Bayes risk is attained and that a test attaining it, $(\underline{N}, \underline{\hat{N}}) = (\underline{N}(u, v), \underline{\hat{N}}(u, v))$, can be defined as follows. The stopping time, $\underline{N}$, is the smallest $n \geqq 0$ (or ∞ if there is no n) such that

$$R(U_n, V_n) = \min (U_n, V_n) , \tag{8}$$

so that further sampling is not called for. (Multiplying both sides of (7) by $p_n/(p_n + uf_n + vg_n)$ gives the usual characterization, "*a posteriori* risk equals stopping risk.") The terminal decision, $\underline{\hat{N}}$, chooses G if and only if $U_{\underline{N}} = \min (U_{\underline{N}}, V_{\underline{N}})$, so that $R(u, v) = E(\underline{N} + \min (U_{\underline{N}}, V_{\underline{N}}))$.

Theorems 4.2 and 5.2 of [3] suffice to establish that

$$E[\underline{N} + \min (U_{\underline{N}}, V_{\underline{N}}) \mid \mathscr{F}_n] = n + R(U_n, V_n) \qquad \text{on} \quad \{\underline{N} \geqq n\} ,$$

i.e., the conditional expected risk of $(\underline{N}, \underline{\hat{N}})$ upon reaching (U_n, V_n) after n observations equals the minimal (Bayes) risk from that point forward plus the cost, n, of the observations already taken. Thus if another stopping time, M, satisfies $M \leqq \underline{N}$, then

$$E[\underline{N} + \min (U_{\underline{N}}, V_{\underline{N}}) \mid \mathscr{F}_n] = M + R(U_M, V_M) \qquad \text{on} \quad \{M = n\} .$$

Integrating and summing over n yields

$$R(u, v) = EM + ER(U_M, V_M) \qquad \text{if} \quad M \leqq \underline{N}(u, v) . \tag{9}$$

Let

$$\bar{U} = \sup \{u \mid u = R(u, v) \ \text{for some} \ v\}$$

and

$$\bar{V} = \sup \{v \mid v = R(u, v) \ \text{for some} \ u\} .$$

It is easy to see that $R(1, 1) = 1$, so $\bar{U}, \bar{V} \geqq 1$. Furthermore, $\bar{U}$ is finite, since $EM(\frac{1}{2}, 0)$ is finite and for all v

$$R(u, v) \leqq EM(\tfrac{1}{2}, 0) + \tfrac{1}{2}u < u \qquad \text{if} \quad u > 2EM(\tfrac{1}{2}, 0) ,$$

using the bounds from (4), $\alpha(\frac{1}{2}, 0) \leqq \frac{1}{2}$ and $\beta(\frac{1}{2}, 0) = 0$. Also, $\bar{U}$ is by definition the limit from below of a sequence $\{u_k\}$ such that $u_k = R(u_k, v_k)$ for some v_k. Since $R(\cdot, \cdot)$ is clearly nondecreasing in each variable,

$$u_k \leqq R(\bar{U}, v_k) \leqq \lim_{v\to\infty} R(\bar{U}, v) \leqq \bar{U},$$

using (7).

Letting k go to infinity it follows that

$$\bar{U} = \lim_{v\to\infty} R(\bar{U}, v). \tag{10}$$

By similar reasoning, $\bar{V}$ is finite and

$$\bar{V} = \lim_{u\to\infty} R(u, \bar{V}). \tag{11}$$

Note also that

$$R(u, v) - u = \min_{(N,\hat{N})} \{EN + u(F(\hat{N} = G) - 1) + vG(\hat{N} = F)\} \downarrow \text{in } u, \tag{12}$$

not necessarily strictly.

Now, for A and B between 0 and 1, let

$$u = \frac{\bar{U}}{A} \quad \text{and} \quad v = \frac{\bar{V}}{B},$$

so that, by (3), $M = M(A, B)$ stops the first time

$$U_n \leqq \bar{U} \quad \text{or} \quad V_n \leqq \bar{V}. \tag{13}$$

For the remainder of the proof, u and v are chosen in this way and M denotes $M(A, B)$.

By the definition of $\bar{U}$ and $\bar{V}$, (8) implies (13), so that $\underline{N}(u, v) \geqq M$. Therefore, (9) applies and, using (6), it follows that

$$\begin{aligned} r(M, \hat{M}) - R(u, v) = E\{&(U_M - R(U_M, V_M))\mathbf{1}\{\hat{M} = G\} \\ &+ (V_M - R(U_M, V_M))\mathbf{1}\{\hat{M} = F\}\}. \end{aligned} \tag{14}$$

Recall that $\hat{M}$ is required to be F if the first inequality in (3) is not satisfied at time M. Hence, $U_M \leqq \bar{U}$ is a necessary condition for $\hat{M} = G$ and, thus, on $\{\hat{M} = G\}$ the following relations hold:

$$\begin{aligned} &U_M \leqq \bar{U}, \\ &V_M \geqq \max(U_M - \bar{U}, V_M - \bar{V}, 0) = Y_M, \quad \text{say}, \end{aligned}$$

and, by (12) and the monotonicity of $R(\bar{U}, \cdot)$,

$$U_M - R(U_M, V_M) \leqq \bar{U} - R(\bar{U}, V_M) \leqq \bar{U} - R(\bar{U}, Y_M). \tag{15}$$

Similarly, on $\{\hat{M} = F\}$

$$V_M - R(U_M, V_M) \leqq \bar{V} - R(Y_M, \bar{V}). \tag{16}$$

Using the estimates (15) and (16) in (14),

$$0 \leqq r(M, \hat{M}) - R(u, v) \leqq E[\max(\bar{U} - R(\bar{U}, Y_M), \bar{V} - R(Y_M, \bar{V}))], \tag{17}$$

the first inequality by definition of $R(u, v)$. As an immediate consequence of Lemma 2 below,

$$\max(U_M, V_M) \to \infty \quad \text{in probability} \quad (P) \tag{18}$$

as $\min(A, B) \to 0$. It follows at once that $Y_M \to \infty$ and, by (10) and (11), that the bracketed quantity in (17) goes to zero in probability. Therefore, by the bounded convergence theorem,

$$r(M, \hat{M}) - R(u, v) \to 0 \qquad \text{as} \quad \min(A, B) \to 0 \,. \tag{19}$$

For fixed A and B, consider tests $(N, \hat{N})$ such that $F(\hat{N} = G) \leqq \alpha(A, B)$ and $G(\hat{N} = F) \leqq \beta(A, B)$. By (5), evidently

$$r(M, \hat{M}) - EM \geqq r(N, \hat{N}) - EN \,.$$

Since $r(N, \hat{N})$ is at least the minimal value, $R(u, v)$,

$$r(M, \hat{M}) - EM \geqq R(u, v) - EN \,.$$

This relation holds for all tests included in the definition of the infimum, $n(A, B)$. Therefore, the relation holds with $n(A, B)$ in place of EN, and the theorem follows upon letting $\min(A, B) \to 0$ and invoking (19).

Since $U_M = \bar{U} f_M / p_M A$ and $V_M = \bar{V} g_M / p_M B$, it is sufficient for (18) to prove the following lemma.

LEMMA 1. *Under assumption* (1), *if* $M = M(A, B)$ *and* $A, B > 0$, *then*

$$\max\left(\frac{f_M}{p_M A}, \frac{g_M}{p_M B}\right) \to \infty \quad \textit{in probability} \quad (P) \tag{20}$$

as $\min(A, B) \to 0$.

REMARK. If $I(P, F)/I(P, G)$ is not a limit point of $\log A^{-1}/\log B^{-1}$, (20) is an easy consequence of the strong law of large numbers, assuming only that the information numbers are finite, without invoking (1). Theorem 1 in this case essentially amounts to a reaffirmation of the optimality property of the one-sided SPRT, since, with P-probability approaching one, $M =$ the stopping time for P vs. F (P vs. G) when $I(P, F)/I(P, G) - \log A^{-1}/\log B^{-1}$ is positive (resp. negative).

PROOF. Consider the special case where $B \to 0$ while A remains bounded below by a positive number, δ. Then $M \leqq \tilde{M} =$ smallest $n \geqq 0$ such that $f_n/p_n \leqq \delta$. Hence, since (1) implies $P(g_n = 0) = 0$,

$$\frac{g_M}{p_M B} \geqq \frac{1}{B} \min\left(\frac{g_0}{p_0}, \ldots, \frac{g_{\tilde{M}}}{p_{\tilde{M}}}\right) \to \infty \tag{21}$$

in probability (P) as $B \to 0$. By similar reasoning, (20) holds in the special case where B remains bounded below by a positive number.

It remains to prove (20) in the case where both $A \to 0$ and $B \to 0$. By standard reasoning about subsequences, this case can be combined with those of the preceding paragraph to establish (20). Define

$$S_n = \log \frac{p_n}{f_n} \quad \text{and} \quad T_n = \log \frac{p_n}{g_n}, \qquad n = 0, 1, \cdots, \tag{22}$$

and observe that M is the smallest $n \geqq 0$ (or ∞ if there is no n) such that $S_n \geqq \log A^{-1}$ or $T_n \geqq \log B^{-1}$. S_n and T_n are each cumulative sums of independent and identically distributed random variables with positive means $I(P, F)$ and $I(P, G)$, respectively. Letting M_1 and M_2 denote the first time $S_n \geqq \log A^{-1}$ and the first time $T_n \geqq \log B^{-1}$, respectively, it is clear that $M = \min(M_1, M_2)$. Now, it is well known that

$$\frac{M_1}{\log A^{-1}/I(P, F)} \to 1 \quad \text{and} \quad \frac{M_2}{\log B^{-1}/I(P, G)} \to 1 \quad \text{in probability} \quad (P)$$

as $A, B \to 0$. (See [4], page 127 for an elementary proof of an even stronger a.s. result.)

It follows routinely that

$$\frac{M}{C} = \frac{\min(M_1, M_2)}{C} \to 1 \quad \text{in probability} \quad (P) \tag{23}$$

where $C = \min(\log A^{-1}/I(P, F), \log B^{-1}/I(P, G)) \to \infty$ as $A, B \to 0$.

Now, let $r = I(P, F)/I(P, G)$, so that $E(S_1 - rT_1) = 0$ and let $\operatorname{Var}(S_1 - rT_1) = \sigma^2$. The variance σ^2 cannot be zero, for this would imply $S_1 = rT_1$, which yields $f_1 = (g_1/p_1)^r p_1$. Integrating this relation leads to $E(g_1/p_1)^r = 1 = 1^r = (E(g_1/p_1))^r$, which can hold only if $r = 1$ or g_1/p_1 is constant ($= 1$), both of which are contrary to the assumed distinctness of F, G, and P. By (1) and the Central Limit Theorem for randomly stopped sums ([4], page 197; [18]), $C^{-\frac{1}{2}}(S_M - rT_M) \to \sigma Z$ in distribution, where Z is standard normal. The convergence of the distribution functions is uniform because σZ has a continuous distribution function. Therefore, the probability that $S_M - rT_M$ falls in an interval $J = J(A, B)$, say, goes to zero if the length of J is less than $C^{\frac{1}{4}}$ (say). Obviously, this fact carries over to $S_M - \log A^{-1} - r(T_M - \log B^{-1})$. Hence

$$C^{-\frac{1}{4}}|S_M - \log A^{-1} - r(T_M - \log B^{-1})| \to \infty \quad \text{in probability} \quad (P)$$

and, therefore,

$$C^{-\frac{1}{4}} \max(|S_M - \log A^{-1}|, |T_M - \log B^{-1}|) \to \infty \quad \text{in probability} \quad (P). \tag{24}$$

Now, without the absolute value signs in (24), the result would be convergence to *zero*, because the indicated maximum would then be either the excess of S_{M_1} over the boundary $\log A^{-1}$ or the excess of T_{M_2} over the boundary $\log B^{-1}$, both of which have proper limit distributions by virtue of the finite variances ([6], page 355). Evidently, then, (24) holds only because

$$C^{-\frac{1}{4}} \max(\log A^{-1} - S_M, \log B^{-1} - T_M) \to \infty \quad \text{in probability} \quad (P),$$

from which (20) follows immediately, proving the lemma.

REMARK. In the case of Koopman–Darmois families (e.g., normal, gamma, binomial) with P between F and G, there exist positive numbers λ and μ such that $\lambda \log (f_1/p_1) + \mu \log (g_1/p_1) = -1$. This leads to the relation

$$\lambda \log \frac{f_n}{p_n A} + \mu \log \frac{g_n}{p_n B} = -n + \lambda \log A^{-1} + \mu \log B^{-1}, \tag{25}$$

which simplifies the proof of the lemma. If $\max (f_M/p_M A, g_M/p_M B) < y$, say, then the relation yields $M > m = [\lambda \log A^{-1} + \mu \log B^{-1} - (\lambda + \mu) \log y]$ (integer part). Since M hasn't stopped by time m, both terms on the left-hand side of (25) are positive at $n = m$, whence

$$0 < \lambda \log \frac{f_m}{p_m A} < -m + \lambda \log A^{-1} + \mu \log B^{-1} < (\lambda + \mu) \log y + 1,$$

which places $\log (f_m/p_m)$ in an interval of length $(1 + \mu/\lambda) \log y + \lambda^{-1}$. As $\min (A, B) \to 0$, $m \to \infty$ and an application of the ordinary Central Limit Theorem shows that the probability of this event goes to zero for every y. The Wiener process case can also be handled in this way and the proof of Theorem 1 goes through essentially unchanged.

3. Symmetric normal case. The symmetric case of the Kiefer–Weiss problem for testing a normal mean is the case where the error probability bounds, α and β, at given values θ_1, θ_2 of the mean are equal. As Weiss showed ([17]), the problem reduces in this case to minimizing the expected sample size at $\theta_0 = (\theta_1 + \theta_2)/2$. Furthermore, the class of all solutions to this problem for different choices of α coincides with the class of solutions to the Bayes problem with equal weights on θ_1 and θ_2, corresponding to $u = v$ in Section 2. A version of the standard iterative scheme ([10], page 662) was used to compute a sequence of Bayes solutions and their operating characteristics for each of several choices of the parameter $\delta = (\theta_2 - \theta_1)/2\sigma$ (σ = the known standard deviation). Operating characteristics of 2-SPRT's were also computed. All computations were performed in double precision on an XDS Sigma 5 computer and the results were checked and confirmed to have a relative accuracy of 10^{-7}.

The fact that δ characterizes the problem can be seen as follows. The distributions P, F, and G are normal with variance σ^2 and means θ_0, $\theta_0 - \delta\sigma$ and $\theta_0 + \delta\sigma$, respectively. Standardizing the observations in the form $Y_i = (X_i - \theta_0)/\sigma$ transforms P, F and G, into normal distributions with variance one and means 0, $-\delta$, and δ. (It is also instructive to consider the problem as one of observing at times $t = \delta^2, 2\delta^2, 3\delta^2, \cdots$ a Wiener process $X(t)$ with variance one per unit time and mean 0, $-t$, or t.) The values $\delta = .1, .2, .3, \cdots, 1.0$ were used in the computations. The efficiencies of the 2-SPRT's were computed as the ratio $E_{\theta_0} N / E_{\theta_0} M$, where N is an optimal test (Bayes solution) and M is a 2-SPRT having the same error probability. Since the efficiencies vary insignificantly as a function of δ, only the values $\delta = .1, .2$, and $.4$ are used for illustration in the following table.

TABLE 1

Expected (and maximum possible) sample sizes of optimal tests and 2-SPRT'*s*

		$\alpha = .10$	$\alpha = .05$	$\alpha = .01$
$\delta = .4$	Optimal	7.35 (30)	12.27 (40)	25.38 (62)
	2-SPRT	7.40 (18)	12.35 (26)	25.49 (46)
$\delta = .2$	Optimal	28.62 (153)	48.29 (194)	100.76 (281)
	2-SPRT	28.83 (75)	48.61 (110)	101.19 (190)
$\delta = .1$	Optimal	113.70 (749)	192.42 (913)	402.25 (1263)
	2-SPRT	114.57 (311)	193.66 (449)	404.00 (771)
	Efficiency†	99.25%	99.36%	99.57%

† The efficiencies were obtained not from the rounded-off EN's shown, but from more precise expressions. The same efficiencies apply to all three cases, $\delta = .1, .2, .4$, with the exception that for $\delta = .4$ actual efficiencies are .01% higher for $\alpha = .10$ and .05.

Although error probabilities larger than 10% are infrequently acceptable in practice, it is interesting to note how the efficiency of the 2-SPRT varies with α. It turns out that the minimum efficiency is attained for an α of approximately .17 and that as α is increased past this level the efficiency actually increases. The results in Table 2 were obtained for $\delta = .1$, but differ very little from those for other values of δ.

TABLE 2

Efficiency of the 2-SPRT

α (in %)	40	25	17	10	7	4	2	1	$\frac{1}{2}$	$\frac{1}{4}$
% efficiency	99.68	-*.25	—.19	—.25	—.30	—.40	—.49	—.57	—.63	—.67

* The digits to the left of the decimal point are 99 in all cases.

The value 99.19% shown for $\alpha = .17$ is the lowest obtained in the computer calculations, which covered the range from $\delta = .1$ to 1.0 and from sample size one to the size needed for $\alpha = \frac{1}{4}\%$.

For the problem of testing $\theta = \theta_0 - \delta\sigma$ against $\theta = \theta_0 + \delta\sigma$ based on independent normal observations $X_1, X_2, \cdots$ with mean θ and variance σ^2, the symmetric 2-SPRT $M(A, A)$ stops sampling as soon as

$$|Y_1 + \cdots + Y_n| \geqq \delta^{-1} \log A^{-1} - \tfrac{1}{2}n\delta ,$$

where the Y_i's are the standardized observations $(X_i - \theta_0)/\sigma$. Two questions of practical importance are the following. How should A be chosen to achieve a desired error probability, α? What will be the resulting expected sample size when θ_0 is true?

The answer to the first question is available to a high degree of precision as a result of the computer calculations performed for δ in the range .1 to 1.0. It turns out that the ratio of the true error probability, $\alpha(A)$, to A is constant to within 1 part in 4 million for $\delta = .1$ and α between 10% and .1%. Similar results were obtained for $\delta = .2$, .3, and .4. For $\delta = 1.0$ the constancy holds only within 1 part in 200. By fitting a quadratic to the values of these ratios for the given values of δ, the following formula for A was obtained:

$$A = \frac{\alpha}{.4996 - .28645\delta + .0696\delta^2}.$$

For α between 10% and .1%, the choice of A given by this formula results in an actual α agreeing with the desired α to within 1 part in 5000 for δ between .1 and .5. For larger δ (up to 1.0) the accuracy falls off, reaching 1 part in 100 for $\delta = 1$. This should still be sufficient for practical purpose, e.g., a desired α of 5% would result in an actual α between 4.95% and 5.05%.

The answer to the second question, concerning expected sample sizes, is available to a lesser degree of precision. The lower bound of Hoeffding [7] is

$$E_{\theta_0} N \geqq 2\delta^{-2}(1 - \log(2\alpha) - [1 - 2\log(2\alpha)]^{\frac{1}{2}}) = H(\alpha, \delta).$$

This lower bound is remarkably sharp over a broad range of values of α. For values of α between 5% and .1%, the ratios of $E_{\theta_0} M$ for the 2-SPRT to $H(\alpha, \delta)$ are approximately as follows.

δ	.1	.2	.3	.4	.5	.6	.7	.8
$E_{\theta_0} M/H(\alpha, \delta)$	1.039	1.042	1.046	1.052	1.06	1.07	1.09	1.10

A useful approximation to $E_{\theta_0} M$ is obtained for given α and δ by multiplying $H(\alpha, \delta)$ by the ratio indicated for δ (or an interpolated value). The relative accuracy of this approximation is at worst 1 part in 100 for $\delta \leqq .5$, falling off to 1 part in 20 for $\delta = .8$. The imprecision in the actual α resulting from the A chosen above has a relatively small effect on the accuracy of the $E_{\theta_0} M$ approximation. Of course, the expected sample sizes become quite small for larger δ, e.g., 3.5 for $\alpha = 5\%$ when $\delta = .8$, so that a relative accuracy of 1 part in 20 may be sufficient. The expected sample sizes for $\alpha = .05$ and .01 in Table 1 are all within 0.4% of the approximations based on the tabulated ratios to $H(\alpha, \delta)$.

Acknowledgments. The author is indebted to F. R. Maiocco of the Jet Propulsion Laboratory for his very generous help with the computer calculations. The author also wishes to thank David Siegmund, Tze Lai, and Michael Klass for helpful discussions.

REFERENCES

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

[2] BECHHOFFER, R. E. (1960). A note on the limiting relative efficiency of the Wald sequential probability ratio test. *J. Amer. Statist. Assoc.* **55** 660-663.

[3] CHOW, Y. S., ROBBINS, H. and SIEGMUND, D. (1971). *Great Expectations: The Theory of Optimal Stopping.* Houghton Mifflin, Boston.

[4] CHUNG, K. L. (1968). *A Course in Probability Theory.* Harcourt, Brace, and World, New York.

[5] FABIAN, V. (1974). Note on Anderson's sequential procedures with triangular boundary. *Ann. Statist.* **2** 170-176.

[6] FELLER, W. (1966). *An Introduction to Probability Theory and its Applications* **2**. Wiley, New York.

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

[8] KIEFER, J. and SACKS, J. (1963). Asymptotically optimum sequential inference and design. *Ann. Math. Statist.* **34** 705–750.

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

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

[11] LAWING, W. and DAVID, H. T. (1966). Likelihood ratio computations of operating characteristics. *Ann. Math. Statist.* **37** 1704–1716.

[12] SIMONS, G. (1967). Lower bounds for average sample number of sequential multihypothesis tests. *Ann. Math. Statist.* **38** 1343–1364.

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

[14] WALD, A. (1950). *Statistical Decision Functions.* Wiley, New York.

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

[16] WALD, A. and WOLFOWITZ, (1950). Bayes solutions of sequential problems. *Ann. Math. Statist.* **21** 82–99.

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

[18] WITTENBERG, H. (1964). Limiting distributions of random sums of independent random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **1** 7–18.

ALFRED P. SLOAN LABORATORY OF
MATHEMATICS AND PHYSICS
DEPARTMENT OF MATHEMATICS
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125