OPTIMUM PROPERTIES AND ADMISSIBILITY OF SEQUENTIAL TESTS[1]

By D. L. Burkholder and R. A. Wijsman

University of Illinois

1. Introduction and summary. In Wald's sequential probability ratio test (SPRT) [7], with stopping bounds B, $A (B < A)$, for testing a simple hypothesis H_1 against a simple alternative H_2, sampling continues until the first time the probability ratio is either less than or equal to B, or greater than or equal to A. In the former case H_1 is accepted, in the latter case H_2 is accepted. This test is known to have a certain optimum property (OP) first conjectured by Wald [6], [7] Section A7, later proved by Wald and Wolfowitz [9], [10] and partly by Arrow, Blackwell and Girshick [1] (see also Wolfowitz [11]). This OP can be expressed in words rather roughly as follows: Among all sequential tests whose error probabilities do not exceed those of the SPRT under consideration, the latter has the smallest expected sample size under both distributions. However, the validity of the OP has been demonstrated only under two conditions. The first condition is that $B \leqq 1 \leqq A$. The second condition is that only sequential tests with finite expected sample sizes are considered. The question arises then whether these conditions are necessary. The purpose of this paper is partly to show that the second condition is not necessary, but the first one is. The superfluousness of the second condition is shown for a rather general cost function in the following form: If a test has OP among all tests with finite expected sample sizes, then it has OP among all tests. If $B \leqq 1 \leqq A$ does not hold, then the SPRT is not admissible, in a sense which will be made precise in Section 2. This will be demonstrated in a manner which at the same time exhibits a test that not only improves upon the SPRT, but which itself possesses OP. However, if only tests are considered which take at least one observation, then there are no restrictions on A and B (other than $B \leqq A$) for a SPRT to have OP. A summary of the main results appeared in [2]. Some of the results were also obtained by Ghosh [3] under slightly less general conditions.

The methods employed in this paper yield some additional results that are of interest. Part of the Wald and Wolfowitz OP proof [9] consists in showing the existence of a loss function that makes a given SPRT Bayes for a given *a priori* distribution of the two hypotheses. We give a simpler existence proof in Section 5, as well as proving a lemma that, although slightly weaker, achieves the same aim. In Section 6 a certain continuity property of the error probabilities of a SPRT as functions of the stopping bounds is shown. If the lower bound is fixed, the error probabilities are left or right continuous functions of the upper stopping bound depending on whether the latter is a stopping point for the probability ratio or not. Similar statements hold for the error probabilities as functions of the lower stopping bound.

Received June 2, 1961; revised July 2, 1962.

[1] Work supported by the National Science Foundation, grant NSF G-9104 and G-11382

2. Definitions and main theorems. Throughout this paper we will employ the following abbreviations: PR = probability ratio, SPRT = sequential probability ratio test, OP = optimum property.

Let P_1 and P_2 be two probability measures over some measurable space, and let $X_1, X_2, \cdots$, be random variables on this space such that with respect to P_i, $i = 1, 2$, $X_1, X_2, \cdots$ are independent and identically distributed. Throughout, i will always run over 1,2. The X_j can be considered as independent observations on a random variable X whose distribution is determined either by P_1 or by P_2. We exclude the case that the distributions of X induced by P_1 and P_2 are the same. Let H_i be the hypothesis that the true distribution is P_i. We shall denote the PR at the nth stage of sampling by Y_n, i.e.,

$$Y_n = p_2(X_1) \cdots p_2(X_n)/p_1(X_1) \cdots p_1(X_n)$$

($n = 1, 2, \cdots$), where p_i is the common density of the X's under H_i with respect to any sigma-finite measure that dominates P_1 and P_2. For any (possibly randomized) sequential test T let N be the random number of observations, and let $\alpha_i(T)$ and $\nu_i(T)$ be the error probabilities and expected sample sizes, respectively, i.e.,

$$\alpha_i(T) = P_i(H_i \text{ is rejected} \mid T) \tag{2.1}$$

$$\nu_i(T) = E_i(N \mid T). \tag{2.2}$$

In Theorem 2.1 we shall define the ν_i a little more generally, as follows. Suppose the cost of n observations is $c_i(n)$ if H_i is true, with

$$0 = c_i(0) \leqq c_i(1) \leqq \cdots < c_i(\infty) = \infty \tag{2.3}$$

and $c_i(n) \to \infty$ as $n \to \infty$. Then we define

$$\nu_i(T) = E_i(c_i(N) \mid T). \tag{2.4}$$

It is seen that (2.2) is a special case of (2.4), with $c_i(n) = n$.

For some results it is essential to allow randomized tests. We shall be mostly interested in the type of randomization where, in advance, a random device selects one test from among a finite collection of tests. If this collection consists of the tests $T_1, \cdots, T_k$, and the random device selects T_m with probability π_m, $m = 1, \cdots, k$, then the resulting randomized test will be denoted symbolically by

$$T = \pi_1 T_1 + \cdots + \pi_k T_k. \tag{2.5}$$

We shall say that T is a mixture of $T_1, \cdots, T_k$.

Given a cost function $c_i(n)$, admissibility and inadmissibility of tests is defined in terms of the α_i and ν_i as follows:

DEFINITION 2.1. *A test T will be called inadmissible relative to a cost function $c_i(n)$ if there is a test T^* such that $\alpha_i(T^*) \leqq \alpha_i(T)$, $\nu_i(T^*) \leqq \nu_i(T)$, with strict inequality in at least one of the four inequalities.*

We shall define two kinds of optimum property, termed OP I and OP II, respectively, the latter being somewhat stronger than the former.

DEFINITION 2.2. *T^* is said to have optimum property* I (OP I) *relative to a cost function $c_i(n)$ if $\nu_i(T^*) < \infty$, and if for each T satisfying $\nu_i(T) < \infty$ we have*

$$(2.6)\quad \alpha_i(T) \leqq \alpha_i(T^*), \qquad i = 1, 2, \quad \text{implies} \quad \nu_i(T) \geqq \nu_i(T^*), \qquad i = 1, 2.$$

Definition 2.2 gives the usual form of the optimum property as introduced by Wald [6], [7] Section A7. It should be remarked at this point that Definition 2.2 would allow at least one strict inequality in (2.6) on the left of the implication sign, and at the same time equality for both i on the right, in which case T is obviously a better test than T^*. However, this can actually not happen, as will follow from Theorem 2.1 below, so that Definition 2.2 is a sensible one. It also follows immediately by examining the proof of OP I, [9] Section 7, that in (2.6) on the right at least one of the inequalities must be strict if at least one of the inequalities on the left is strict. However, Theorem 2.1 gives more, namely that both inequalities on the right in (2.6) are strict if at least one inequality on the left is. OP II makes this as part of its definition, and, in addition, drops the restriction to finite ν_i.

DEFINITION 2.3. *T^* is said to have optimum property* II (OP II) *relative to a cost function $c_i(n)$ if, for each T* (2.6) *holds, and if for each T satisfying the left side of the implication* (2.6) *with at least one strict inequality, the right side is satisfied with both inequalities strict.*

Since to all appearances Definition 2.3 is stronger than Definition 2.2, one would be inclined to think that there could be a test T^* having OP I but not OP II. More specifically, there could conceivably be a test T such that $\alpha_i(T) < \alpha_i(T^*)$, $\nu_1(T) = \infty$ and $\nu_2(T) < \nu_2(T^*)$. That this cannot be the case after all follows from

THEOREM 2.1. *Relative to a cost function $c_i(n)$, if T^* has* OP I, *it has* OP II.

The proof is given in Section 3. Thus, if T^* is a SPRT with $B \leqq 1 \leqq A$, so that T^* has OP I relative to $c_i(n) = n$, it follows from Theorem 2.1 that T^* also has OP II relative to $c_i(n) = n$. Therefore, the usual restriction of comparing T^* with tests T having finite expected sample sizes is unnecessary. In other words, T^* has OP among all sequential tests. From now on, if a test has OP I, and hence also OP II, by Theorem 2.1, we shall simply say that it has OP.

In the remainder of this section it is assumed that $c_i(n) = n$, i.e., (2.2) is taken as the definition of the ν_i. Since a SPRT with $B \leqq 1 \leqq A$ has OP, it is certainly admissible. What can one say about a SPRT with $1 < B$ or with $A < 1$? It turns out that such a test is not even admissible, let alone that it has OP. This will be shown by exhibiting another test with the same α_i but smaller ν_i. It will be necessary to consider SPRT's of slightly different form than the usual. Usually, the acceptance interval for H_1 is taken as the closed interval $I_1 = [0, B]$, and the acceptance interval for H_2 as $I_2 = [A, \infty]$. There is, however, no reason why we could not let I_1 be open on the right and/or I_2 open on the left. We shall, therefore, consider all four combinations. As a matter of notation, if

$I_1 = [0, B)$, we shall denote the lower stopping bound by $B-$. Likewise, $I_2 = (A, \infty]$ will be indicated by $A+$. Thus, we consider the four types of SPRT's denoted by $T(B, A)$, $T(B-, A)$, $T(B, A+)$ and $T(B-, A+)$. All four are defined if $B < A$; the last three are also defined if $B = A$. Instead of $\alpha_i(T(B, A))$, $\nu_i(T(B, A))$, we shall write $\alpha_i(B, A)$, $\nu_i(B, A)$; and similarly for the other kinds of tests. Furthermore, it is convenient to introduce the quantities β_i defined by

$$\beta_1 = \alpha_1/(1 - \alpha_2), \qquad \beta_2 = \alpha_2/(1 - \alpha_1). \tag{2.7}$$

Since the α_i depend on the test T, the notation $\beta_i(T)$, $\beta_i(B, A)$, etc., is self-explanatory. From a notational point of view it is fortunate that a quantity like $\alpha_i(B, A+)$ turns out to be the same as lim $\alpha_i(B, A')$ when $A' \downarrow A$. This will be shown in Section 6 after Lemma 6.2. Concerning the difference between, say, $T(B, A)$ and $T(B, A+)$, it goes without saying that these tests are really the same if the PR cannot ever assume the value A. Even if this value can be assumed, but only with probability 0 for both i (e.g., if the PR is continuously distributed under both H_i), the two tests are still the same with probability 1. This leads to

DEFINITION 2.4. *Two tests will be called* equivalent *if they differ on a set of sample sequences of probability* 0 *under both* P_i.

With these preliminaries, we are able to state the main theorems concerning inadmissible tests. We repeat that, throughout, the cost function is taken to be $c_i(n) = n$, i.e., the ν_i are defined by (2.2).

THEOREM 2.2. *If* $B > 1$ *the* SPRT $T(B, A)$ *is inadmissible unless it is equivalent to the* SPRT $T(1, A)$. *Similarly, if* $A < 1$ $T(B, A)$ *is inadmissible unless it is equivalent to* $T(B, 1)$. *The same statements are true if* B *is replaced by* $B-$, *and/or* A *by* $A+$.

Theorem 2.2 is really a corollary of the next theorem. The latter gives a more or less explicit description of the test which improves upon the inadmissible test, but cannot be improved upon itself. For simplicity, the theorem will be stated only for the case $T(B, A)$, with $B > 1$, but analogous statements are true for the other cases mentioned in Theorem 2.2. First we dispose of some trivial special cases. If $B > 1$ but $T(B, A)$ is equivalent to $T(1, A)$, the former is not only admissible, but has OP, since the latter does. In the same way, if $A < 1$, $T(B, A)$ has OP if it is equivalent to $T(B, 1)$. Now let $Y = p_2(X)/p_1(X)$ be the PR of one observation and suppose $B > 1$ and $P_2(Y > B) = 0$, so that also $P_1(Y > B) = 0$. Then for any $A > B$ (including $A = B+$) we have $\alpha_1(B, A) = 0$, $\alpha_2(B, A) = 1$, $\nu_i(B, A) = 1$. Clearly, we do better by taking the test T_1 which accepts H_1 without taking any observation, because then $\alpha_1(T_1) = 0$, $\alpha_2(T_1) = 1$, $\nu_i(T_1) = 0$.

THEOREM 2.3. *Let* $1 < B < A$ *and* $P_2(Y > B) > 0$, *where* $Y = p_2(X)/p_1(X)$. *Let* $T = T(B, A)$ *and suppose that* T *is not equivalent to* $T(1, A)$. *Then there exists* $A' > 1$ *and a mixture* T' *of* $T(1, A')$ *and* $T(1, A'+)$ *such that* $\beta_1(T') = \beta_1(T)$ *and* $\alpha_2(T') < \alpha_2(T)$. *Let* λ *be defined by*

$$\lambda = [1 - \alpha_2(T)]/[1 - \alpha_2(T')]. \tag{2.8}$$

so that $0 < \lambda < 1$. *Let* T_1 *be the test that accepts* H_1 *without taking any observation. Then the mixture* $T^* = (1 - \lambda)T_1 + \lambda T'$ *satisfies* $\alpha_i(T^*) = \alpha_i(T)$, $\nu_i(T^*) \leqq \nu_i(T)$, *with at least one of the inequalities strict. Moreover,* T^* *has* OP.

The proof of this theorem is given in Section 6.

The test T^* in Theorem 2.3 uses T_1 with positive probability $1 - \lambda$. There may be occasions where it is objectionable to use a test which takes no observation with positive probability.[2] For instance, suppose that P_1 and P_2 are only two members of a family of distributions indexed by a real parameter θ, and suppose that it is desired that the power function of a test should approach 0 as $\theta \to -\infty$, and 1 as $\theta \to \infty$. Then any test that takes no observation, with positive probability, is ruled out. The test T^*, for instance, has its power function bounded above by $\lambda < 1$. Now, if we restrict ourselves to sequential tests, possibly randomized, that take at least one observation, then it turns out that every SPRT has OP, whether $B \leqq 1 \leqq A$ or not. This is stated, as well as the ordinary OP statement, in Corollary 2.2, which follows from Theorem 2.4 below. Before stating this theorem we shall introduce the notion of *extended* SPRT.

Let T_i be the test that accepts H_i without taking any observation. It is sometimes convenient to consider each T_i also as a SPRT. We can do this by considering the stage before any observation has been taken as the *zero*th stage of sampling, denoting the PR at this stage by Y_0, and putting $Y_0 = 1$. If $B > 1$, then $Y_0 < B$ so that we decide to stop and accept H_1, which is test T_1. Similarly, if $A < 1$ we obtain T_2. If $B = 1 < A$ we have the option to stop and accept H_1, or take an observation, or randomize between these two possibilities. Similarly if $B < 1 = A$. If $B = 1 = A$ we may randomize between three decisions: accept H_1, accept H_2, or take an observation. The situation is the same at the nth stage of sampling, $n \geqq 1$, except that now the randomization probabilities may also depend (measurably) on the whole past history, i.e., on $X_1, \cdots, X_n$. This kind of a test is strictly speaking not necessarily a SPRT, since it may depend on the X's in a more complicated way than only through the sequence of probability ratios. Tests of this kind are implicit in [1], [8] and [10], since they arise naturally as Bayes procedures. However, in its application to the SPRT these references do not explicitly state that, if the PR equals A or B, the decision rule may depend on the past history. This was pointed out by Hoeffding ([4], p. 359). We shall call such a generalization of the usual SPRT an *extended* SPRT, and for future reference give here the formal

DEFINITION 2.5. *A test will be called an extended* SPRT *with stopping bounds* A *and* B *if, for* $n = 0, 1, \cdots$, H_1 *is accepted if* $Y_n < B$, H_2 *accepted if* $Y_n > A$, *sampling continues if* $B < Y_n < A$, *and a possibly randomized rule depending on* $X_1, \cdots, X_n$ *is adopted to decide between accepting* H_1 *and continuing sampling if* $Y_n = B$, *between accepting* H_2 *and continuing sampling if* $Y_n = A$; *where* $Y_0 = 1$, *and for* $n \geqq 1$ Y_n *is the* PR *at the* nth *stage of sampling.*

[2] This was pointed out to us by Colin R. Blyth.

REMARKS. 2.1 The above definition also makes sense if $B = A$, except that whenever PR $= B = A$ the decision rule has to decide between the three alternatives: accepting H_1, accepting H_2, and continuing sampling. 2.2. Any SPRT $T(B, A)$ is an extended SPRT provided $B \leqq 1 \leqq A$. More generally, any mixture of $T(B, A)$, $T(B-, A)$, $T(B, A+)$ and $T(B-, A+)$ is an extended SPRT provided $B \leqq 1 \leqq A$. For instance, if $A > 1$, $\pi T(1, A) + (1 - \pi)T(1-, A)$ is an extended SPRT with stopping bounds 1 and A and the following rule concerning what to do if PR $= 1$ or A: at any sampling stage accept H_2 if PR $= A$; at the *zero*th stage (when PR $= 1$) continue sampling; with probability π accept H_1 at every stage $n \geqq 1$ if PR $= 1$, and with probability $1 - \pi$ continue sampling at every stage $n \geqq 1$ if PR $= 1$.

If, in the following, we say that a test T agrees with another test T' from the mth observation on, we mean that the decision rules of T and T' are the same for all sampling stages $n \geqq m$.

We can state now, in Theorem 2.4 below, a slight generalization of the usual optimum property statement of SPRT's. A proof of the theorem is sketched in Section 5.

THEOREM 2.4. *For any integer $m \geqq 0$, if a test T takes at least m observations and agrees with an extended* SPRT *from the* mth *observation on, then T has* OP *among all tests that take at least m observations.*

By taking $m = 0$ in Theorem 2.4 we have

COROLLARY 2.1. *Every extended* SPRT *has* OP.

By taking $m = 1$ in Theorem 2.4 we get

COROLLARY 2.2. *Every* SPRT *has* OP *among all tests that take at least one observation, and any* SPRT *with $B \leqq 1 \leqq A$ has* OP.

The last part of Corollary 2.2 follows from Remark 2.2 and Corollary 2.1. It is, of course, the usual OP statement of SPRT's. As an application of Corollary 2.1 it also follows that the test T_i that accepts H_i without taking any observation has OP, which is of course trivially true.

3. Proof of Theorem 2.1. We shall use repeatedly the fact that if $T = (1 - \lambda)T_1 + \lambda T_2$, then $\alpha_i(T) = (1 - \lambda)\alpha_i(T_1) + \lambda\alpha_i(T_2)$, and $\nu_i(T) = (1 - \lambda)\nu_i(T_1) + \lambda\nu_i(T_2)$. The proof proceeds in four stages, (A), (B), (C) and (D). Stage (A) is a preliminary, (B) and (C) together establish (2.6), and (D) establishes the fact that in (2.6) there are two strict inequalities on the right if there is at least one on the left. Throughout, T^* has OP I and T satisfies the left hand side of implication (2.6).

(A) *If for either $j = 1$ or $j = 2$ we have $\alpha_j(T) < \alpha_j(T^*)$ and $P_j(N < \infty \mid T) = 1$, then the right hand side of* (2.6) *is true.*

For each positive integer k, let T_k be the test agreeing with T at stages $0, 1, 2, \cdots, k - 1$, and such that if the kth stage is arrived at, the sampling stops and H_j is rejected. Then

$$\alpha_{3-j}(T_k) \leqq \alpha_{3-j}(T) \leqq \alpha_{3-j}(T^*), \qquad k = 1, 2, \cdots \tag{3.1}$$

and

$$(3.2)\quad \alpha_j(T_k) = \sum_{n=0}^{k-1} P_j(H_j \text{ is rejected}, N = n \mid T) + P_j(N \geqq k \mid T).$$

Since, by the assumption made in (A), $P_j(N \geqq k \mid T) \to 0$ as $k \to \infty$, the right hand side of (3.2) converges to $\alpha_j(T)$, which is $< \alpha_j(T^*)$ by assumption. Hence, for some k,

$$(3.3)\qquad \alpha_j(T_k) < \alpha_j(T^*).$$

Clearly, $\nu_i(T_k) < \infty$. Since by (3.1) and (3.3) we have $\alpha_i(T_k) \leqq \alpha_i(T^*)$, and since T^* has OP I, we conclude $\nu_i(T_k) \geqq \nu_i(T^*)$. The observation $\nu_i(T) \geqq \nu_i(T_k)$ completes the proof of stage (A).

(B) *If* $\alpha_i(T^*) > 0$ *then the right hand side of* (2.6) *is true.*

Suppose (B) were not true, then either $\nu_1(T) < \nu_1(T^*)$, or $\nu_2(T) < \nu_2(T^*)$. Suppose the latter, the other case being similar. Then $\nu_2(T) < \infty$, so, *a fortiori*, $P_2(N < \infty \mid T) = 1$. Let T' be a fixed sample size test such that $\alpha_i(T') < \alpha_i(T^*)$, and let $T_\lambda = (1 - \lambda)T + \lambda T'$, $0 < \lambda < 1$. For any such λ we have $P_2(N < \infty \mid T_\lambda) = 1$, and $\alpha_i(T_\lambda) < \alpha_i(T^*)$ since $\alpha_i(T) \leqq \alpha_i(T^*)$ and $\alpha_i(T') < \alpha_i(T^*)$. Furthermore, λ can be chosen so small that $\nu_2(T_\lambda) < \nu_2(T^*)$. This contradicts (A), and therefore establishes the truth of (B).

(C) *If* $\alpha_1(T^*) = 0$ *or* $\alpha_2(T^*) = 0$, *then the right hand side of* (2.6) *is true.*

Suppose $\alpha_1(T^*) = 0$, the other case being similar. We also assume $\alpha_2(T^*) < 1$, the case $\alpha_2(T^*) = 1$ being trivial. Thus,

$$(3.4)\qquad 0 = \sum_{n=0}^{\infty} P_1(H_2 \text{ accepted}, N = n \mid T^*)$$

and

$$(3.5)\qquad 0 < 1 - \alpha_2(T^*) = \sum_{n=0}^{\infty} P_2(H_2 \text{ accepted}, N = n \mid T^*),$$

using the fact that $P_2(N < \infty \mid T^*) = 1$. Since $P_i(H_2 \text{ accepted}, N = 0 \mid T^*)$ does not depend on i, and therefore, by (3.4), is 0, it follows from (3.5) that for some positive n we have $P_2(H_2 \text{ accepted}, N = n \mid T^*) > 0$. Thus there is a set F_2 such that $0 < P_2(X \varepsilon F_2) = \sup P_2(X \varepsilon G)$, $P_1(X \varepsilon F_2) = 0$, where the supremum is taken over all sets G satisfying $P_2(X \varepsilon G) > 0$, $P_1(X \varepsilon G) = 0$, and where X is a random variable with the same distribution as the X_j. Let F_1 be the empty set if there is no G such that $P_1(X \varepsilon G) > 0$, $P_2(X \varepsilon G) = 0$. Otherwise, let F_1 be defined similarly to F_2.

Let T' be the test which stops and accepts H_i as soon as an observation falls in F_i, but otherwise is the same as T. Then $\alpha_i(T') \leqq \alpha_i(T) \leqq \alpha_i(T^*)$, $\nu_i(T') \leqq \nu_i(T)$, and

$$(3.6)\qquad \begin{aligned} P_2(N < \infty \mid T') &= 1 - \lim_{n\to\infty} P_2(N > n \mid T') \\ &\geqq 1 - \lim_{n\to\infty} [P_2(X \not\varepsilon F_2)]^n = 1. \end{aligned}$$

We now show that

$$\nu_i(T^*) \leqq \nu_i(T'), \tag{3.7}$$

implying (2.6).

Case C1: $\alpha_2(T^*) = 0$. Then all three tests, T^*, T and T', have their error probabilities equal to 0. Clearly, if a test has both error probabilities equal to 0, then it is equivalent to a test which, at the nth stage, stops and decides H_i only if the observation falls in F_i. Moreover, of all such tests, the one which stops as soon as an observation falls in F_1 or F_2 minimizes the expected sample sizes. The test T' is of this nature. Thus, $\nu_i(T') \leqq \nu_i(T^*)$, so that $\nu_i(T') < \infty$. Since T^* has OP I and $\alpha_i(T') \leqq \alpha_i(T^*)$, (3.7) follows (actually with equality for both i).

Case C2: $\alpha_2(T^*) > 0$. Let T_k be the fixed sample size test which takes k observations and is such that if one of the k observations falls in F_2, then H_2 is accepted, otherwise H_1 is accepted. Then $\alpha_1(T_k) = 0$, $\alpha_2(T_k) = [P_2(X \not\in F_2)]^k$, $k = 1, 2, \cdots$. We can choose k so large that $\alpha_2(T_k) < \alpha_2(T^*)$. Let $T_\lambda = (1 - \lambda)T_k + \lambda T'$, $0 < \lambda < 1$, then $\alpha_1(T_\lambda) = 0$, $\alpha_2(T_\lambda) < \alpha_2(T^*)$, and $P_2(N < \infty \mid T_\lambda) = 1$ using (3.6). By (A) we have then $\nu_i(T^*) \leqq \nu_i(T_\lambda) = (1 - \lambda)c_i(k) + \lambda\nu_i(T')$. Letting $\lambda \to 1$ gives (3.7).

(D) Suppose one of the inequalities on the left in (2.6) is strict. Assume $\alpha_2(T) < \alpha_2(T^*)$, the other case being similar. Let $T_\lambda = (1 - \lambda)T + \lambda T_1$, $0 < \lambda < 1$, where T_1 accepts H_1 without taking any observation, so that $\alpha_1(T_1) = 0$, $\alpha_2(T_1) = 1$, $\nu_i(T_1) = 0$. Then λ can be chosen so small that $\alpha_2(T_\lambda) < \alpha_2(T^*)$. For any λ we have $\alpha_1(T_\lambda) = (1 - \lambda)\alpha_1(T) \leqq \alpha_1(T^*)$. By (B) and (C) we conclude then $\nu_i(T^*) \leqq \nu_i(T_\lambda)$. But $\nu_i(T_\lambda) = (1 - \lambda)\nu_i(T) < \nu_i(T)$, so that $\nu_i(T^*) < \nu_i(T)$. This concludes the proof of Theorem 2.1.

4. Characterization of Bayes tests as extended SPRT's. In this section a summary will be given of some of the main results in [1] and [10] as applied to the two decision problem. A geometric interpretation (Fig. 1) of these results is given and will be helpful in establishing the existence lemmas of Section 5.

Throughout this section we shall assume that a loss is associated with a wrong decision. Let $W = (W_1, W_2)$, $0 < W_i < \infty$, where W_i is the loss if H_i is true and the wrong hypothesis is accepted. In this section W will be kept fixed. The cost of each observation will be taken to be 1 unit, i.e., the cost function in Section 2 is taken as $c_i(n) = n$. We shall consider all possible tests, allowing randomization at each stage of sampling between stopping and continuing sampling, and, if it is decided to stop, between accepting H_1 and accepting H_2. For any test T let $r_i(T)$ be the expected cost if H_i is true:

$$r_i(T) = W_i\alpha_i(T) + \nu_i(T) \tag{4.1}$$

with the α_i, ν_i defined by (2.1) and (2.2). The point $r(T) = (r_1(T), r_2(T))$ will be called the *risk point* of T. Let S be the set of all risk points $r(T)$, S_0 the set of $r(T)$ with T restricted to tests that do not take any observations, and S_1

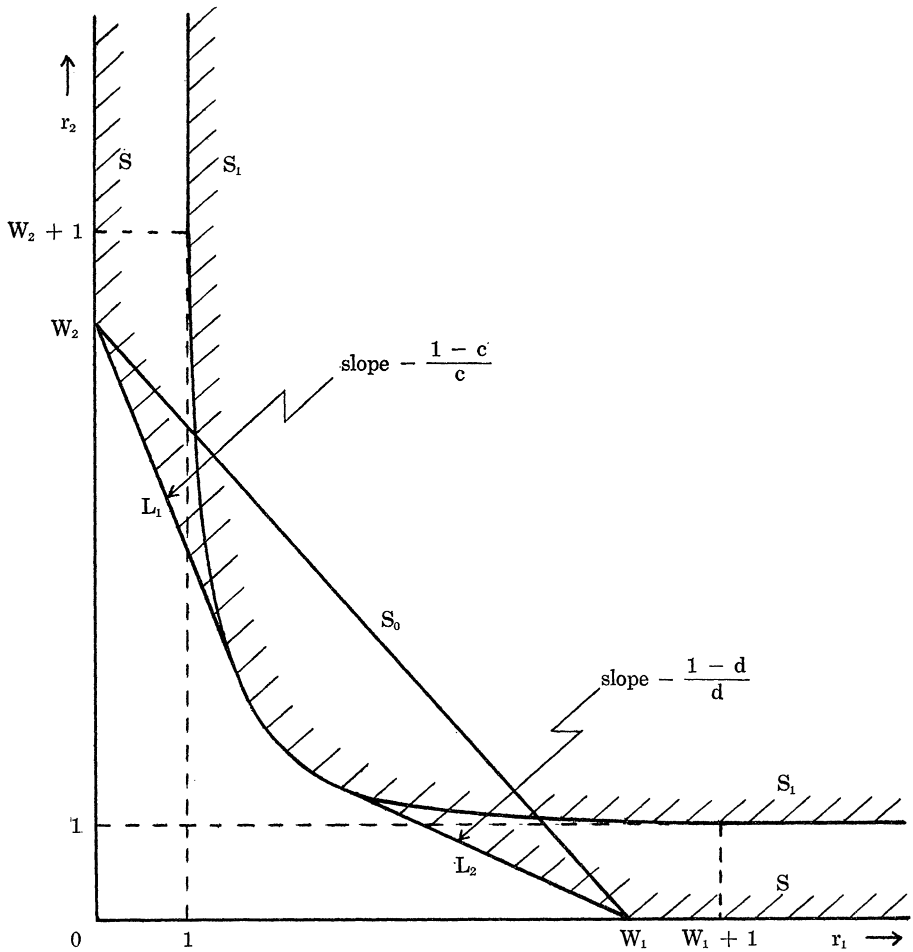

FIG. 1

the set of $r(T)$ with T restricted to tests that take at least one observation. These three sets are indicated in Figure 1 for a typical case. It is easy to show that all three sets are convex and that S_0 is the line segment connecting $(0, W_2)$ and $(W_1, 0)$. A test that takes at least one observation has obviously $\nu_i(T) \geqq 1$, so that, by (4.1), $r_i(T) \geqq 1$. That means that S_1 is contained in the portion of the plane where $r_i \geqq 1$. Consequently, if the W_i are sufficiently small, e.g., if one of the W_i is $\leqq 1$, the sets S_0 and S_1 will be disjoint. This would necessitate exceptions of a trivial nature to be appended to statements that follow later in this section. In order to avoid these trivialities, it will be assumed in this section that W is such that S_0 and S_1 have at least one point in common. It follows from monotonicity lemmas in Section 5 that if S_0 and S_1 are disjoint it is possible to make them intersect by increasing the W_i, while keeping W_2/W_1 constant.

Let an *a priori* distribution over H_1 and H_2 be given by $g = (g_1, g_2)$, $g_i \geqq 0$, $g_1 + g_2 = 1$, where g_i is the probability that H_i is true. The overall expected cost is then

$$r(T, g) = \sum g_i r_i(T) \tag{4.2}$$

with the $r_i(T)$ given by (4.1). A test T that minimizes the overall expected cost (4.2) is called a Bayes test with respect to the *a priori* distribution g, or, simply Bayes (g). Wald and Wolfowitz [10] and Arrow, Blackwell and Girshick [1] proved that the Bayes (g) tests can be described in the following way. First the *a posteriori* distribution is introduced. The joint density of $X_1, \cdots, X_n$ at $(x_1, \cdots, x_n)$ is $p_i(x_1) \cdots p_i(x_n)$, but for short we shall write x^n for $(x_1, \cdots, x_n)$ and $p_{in}(x^n)$ for $p_i(x_1) \cdots p_i(x_n)$. After observing $x_1, \cdots, x_n$ the *a posteriori* distribution $g(x^n)$ is defined by its components

$$g_i(x^n) = \frac{g_i\, p_{in}(x^n)}{\sum_{j=1}^{2} g_j\, p_{jn}(x^n)}. \tag{4.3}$$

A characterization of Bayes (g) tests can now be made as follows. There are two critical numbers, which we shall denote c and d,[3] $0 < c \leqq d < 1$, independent of g (the dependence on W will be taken up in Section 5). At the *zero*th stage of sampling, if $g_2 < c$ no observation is taken and H_1 is accepted, if $g_2 > d$ no observation is taken and H_2 is accepted, if $c < g_2 < d$ an observation is taken. If $g_2 = c$ one may randomize between taking an observation and accepting H_1 without taking any observation; similarly if $g_2 = d$. At the nth stage of sampling, $n = 1, 2, \cdots$, the decision rule is the same as at the *zero*th stage, except that $g_2(x^n)$ has to be substituted for g_2. Using the definition (4.3), with $i = 2$, for $g_2(x^n)$, the above rule is recognized to define the class of extended SPRT's (Definition 2.5) with stopping bounds given by

$$B = (g_1/g_2)[c/(1 - c)], \qquad A = (g_1/g_2)[d/(1 - d)]. \tag{4.4}$$

There is an immediate geometrical interpretation of c and d in Figure 1. The two lines marked L_1 and L_2 play a special role. L_2 is the supporting line of S through $(W_1, 0)$ of minimum slope (maximum negative slope); L_1 is defined analogously. We then recognize the slope of L_1 to be $-(1 - c)/c$, since $g_2 < c$ is equivalent to $g_1/g_2 > (1 - c)/c$, and if g_1/g_2 is larger than the negative of the slope of L_1, then clearly $\sum g_i r_i$ is minimized only by the risk point $(0, W_2)$, i.e., $\sum g_i r_i(T)$ is minimized only by the test T_1 that accepts H_1 without taking an observation. Similarly, the slope of L_2 is $-(1 - d)/d$. We have $c = d$ if and only if the slopes of L_1 and L_2 are equal, and this happens if and only if S_0 does not contain interior points of S_1.

The geometric interpretation of c and d leads to some useful expressions and inequalities. The slope $-(1 - c)/c$ of L_1 is the smallest slope (largest negative

[3] They are denoted h', h'' in [8] and c', c'' in [10].

slope) that any line connecting a point $r \varepsilon S$ with $(0, W_2)$ can have. Therefore, $(1 - c)/c = \sup (W_2 - r_2)/r_1$, where the supremum is taken over all $r \varepsilon S$ (actually the supremum is a maximum). It suffices to take the sup over $r \varepsilon S_0$ and over $r \varepsilon S_1$, and then take the largest of these two numbers. The sup over $r \varepsilon S_0$ is simply W_2/W_1. Thus we have

$$\frac{1-c}{c} = \max\left(\frac{W_2}{W_1}, \sup_{r \varepsilon S_1} \frac{W_2 - r_2}{r_1}\right). \tag{4.5}$$

(In (4.5) we have not assumed, as we did in the rest of Section 4, that S_0 and S_1 intersect, the reason being that we shall need (4.5) in Section 5.) An analogous expression for d is obtained by replacing in (4.5) c by $1 - d$ and interchanging the subscripts 1 and 2. This enables to turn any true statement about c into one about d. In the remainder of the paper this procedure, i.e., interchanging c and $1 - d$ and simultaneously interchanging subscripts 1 and 2, will be termed "by symmetry." From (4.5) and by symmetry we have the inequalities

$$0 < (1 - d)/d \leqq W_2/W_1 \leqq (1 - c)/c < \infty. \tag{4.6}$$

If S_0 contains interior points of S_1, the middle inequalities in (4.6) are both strict; otherwise both are equalities. In particular, these equalities will occur if one of the W_i is $\leqq 1$.

Till now we have considered Bayes tests among all tests. One can also restrict the class of tests and ask for Bayes tests within the restricted class. In particular, let C_m be the class of tests that take at least m observations. How can the Bayes tests within C_m be characterized? The answer is essentially contained in [1] and [10], and is stated in the following lemma for future reference. The notation *Bayes* (g, W) is employed in the anticipation of Section 5 where W is no longer held fixed.

LEMMA 4.1. *For $m \geqq 0$, let C_m be the class of tests that take at least m observations. A test in C_m is Bayes (g, W) among all tests in C_m if and only if it agrees with an extended* SPRT *from the* mth *observation on, with the stopping bounds A and B given by* (4.4).

For $m = 0$ Lemma 4.1 reduces to the characterization of the Bayes (g, W) tests among all tests.

5. Existence and uniqueness lemmas. In this section we shall consider the W_i as variables, and we shall sometimes write $c(W_1, W_2)$, $d(W_1, W_2)$, or $c(W)$, $d(W)$, to express the dependence of c, d on the W_i. In this notation we rewrite the Equations (4.4):

$$B = \frac{g_1}{g_2}\frac{c(W)}{1 - c(W)}, \qquad A = \frac{g_2}{g_2}\frac{d(W)}{1 - d(W)}. \tag{5.1}$$

Wald and Wolfowitz [9] proved that, given A, B and g, with $0 < B < A < \infty$, and both $g_i > 0$, there exists W satisfying (5.1). Since their proof is somewhat involved, we propose to give a simpler proof in this section. First we restate the proposition in an equivalent form as

LEMMA 5.1. *Given* c_0, d_0, *with* $0 < c_0 \leqq d_0 < 1$, *the pair of equations*

$$c(W_1, W_2) = c_0, \qquad d(W_1, W_2) = d_0 \tag{5.2}$$

has a solution for the W_i, *with* $1 \leqq W_i < \infty$.

Parallel to the existence question is the question of uniqueness, of some interest in itself although not used in the OP proof. Concerning uniqueness we have

LEMMA 5.2. *If* $0 < c_0 < d_0 < 1$ *and the Equations* (5.2) *have a solution, this solution is unique.*

The proofs of Lemmas 5.1 and 5.2 will be given a little later in this section. A lemma similar to Lemma 5.1, serving the same purpose, was proved in an ingenious way by L. LeCam (in [5] Chapter 3, Lemma 6), who also obtained the uniqueness result.

Lemmas 4.1 and 5.1 are the two crucial lemmas from which the OP of SPRT's follows at once. Since we have stated the OP in slightly more general form than usual in Theorem 2.4, and also to show how Lemmas 5.1 and 5.3 are used, we shall give below the few simple steps of the proof, following [9], Section 7.

PROOF OF THEOREM 2.4. Let C_m be the class of all tests that take at least m observations, so that the test T in Theorem 2.4 is a member of C_m. Let T' be any other test in C_m with $\alpha_i(T') \leqq \alpha_i(T)$ and $\nu_i(T') < \infty$. Denote $\alpha_i(T) = \alpha_i$, $\nu_i(T) = \nu_i$, $\alpha_i(T') = \alpha_i'$, $\nu_i(T') = \nu_i'$. Let the stopping bounds of the extended SPRT, as mentioned in the theorem, be A and B, $B \leqq A$. Let any g be chosen, with both $g_i > 0$. With these values of A, B, g_1, g_2, the Equations (5.1) have a solution for $W = (W_1, W_2)$, by Lemma 5.1, with $1 \leqq W_i < \infty$. With this W and the given g, T is Bayes (g, W) in C_m, using Lemma 4.1. Hence $\sum g_i(W_i\alpha_i + \nu_i) \leqq \sum g_i(W_i\alpha_i' + \nu_i')$, or $\sum g_i(\nu_i' - \nu_i) \geqq \sum g_iW_i(\alpha_i - \alpha_i') \geqq 0$. Since this holds for any g with both $g_i > 0$, we must have $\nu_i' - \nu_i \geqq 0$, and Theorem 2.4 is proved.

From the proof of Theorem 2.4 it is clear that instead of using Lemma 5.1, we could get by with the somewhat weaker but easier to prove

LEMMA 5.3. *Given* A, B *and* ϵ, *with* $0 < B \leqq A < \infty$ *and* $\epsilon > 0$, *there exists* (i) W *and* g, *with* $g_1 < \epsilon$, *satisfying* (5.1); *and* (ii) W' *and* g', *with* $g_2' < \epsilon$, *satisfying* (5.1).

Before proving Lemmas 5.1 through 5.3, we need a few simple properties of the functions $c(W)$ and $d(W)$. In (4.5) the r_i are functions of W and T, as given by (4.1). We rewrite

$$\frac{1 - c(W)}{c(W)} = \max\left(\frac{W_2}{W_1}, \sup_{T \varepsilon C_1} f(T, W)\right) \tag{5.3}$$

in which

$$f(T, W) = \frac{W_2(1 - \alpha_2(T)) - \nu_2(T)}{W_1\alpha_1(T) + \nu_1(T)}. \tag{5.4}$$

In (5.3) the supremum is taken over all tests that take at least one observation.

In view of the expression (5.4) for f we may restrict the supremum to all $T \varepsilon C_1$ with $\alpha_2(T) < 1$ and $\nu_i(T) \leqq B_i$, for some suitable B_i (e.g., $B_i = W_i$ would suffice.) Denoting this restricted class of tests by C_1^*, we see from (5.4) that $f(T, W)$ is continuous as a function of W, uniformly in $T \varepsilon C_1^*$. From this it follows immediately that the right hand side of (5.3) is continuous in W. Thus, we conclude that c is a continuous function of W. The same is then true for $d(W)$, by symmetry (see Section 4, after (4.5)).

In (5.3) on the right hand side, the supremum is attained by some $T \varepsilon C_1^*$, say by T^*. We can write (5.3) then as

$$(5.5) \qquad [1 - c(W)/c(W)] = \max\,(W_2/W_1, f(T^*, W)).$$

The dependence of T^* on W is suppressed in the notation.

There are some obvious monotonicity properties of the functions c and d. In (5.5) if T^* is kept fixed, but W changed to W' in such a way that W_2/W_1 increases and $f(T^*, W)$ increases, then

$$\frac{1 - c(W^1)}{c(W')} \geqq \max\left(\frac{W_2'}{W_1'},\ f(T^*, W')\right) > \max\left(\frac{W_2}{W_1},\ f(T^*, W)\right) = \frac{1 - c(W)}{c(W)}$$

so that $(1 - c)/c$ increases. Consulting (5.4) we see that $f(T^*, W)$ is increased if W_1 is fixed and W_2 increased (remember $\alpha_2(T^*) < 1$), and $f(T^*, W)$ is not decreased if W_2 is fixed and W_1 decreased (if $\alpha_1(T^*) > 0$, f is actually increased). Furthermore, by (5.5), if W_1 is fixed and $W_2 \to \infty$, $c \to 0$. All this proves

LEMMA 5.4. *If W_1 is fixed, c is a strictly decreasing function of W_2, and $c \to 0$ as $W_2 \to \infty$. If W_2 is fixed, c is a non-decreasing function of W_1. Similar statements for d follow by interchanging c and $1 - d$, W_1 and W_2.*

Somewhat less obvious is the following

LEMMA 5.5. *If $W_i \to \infty$, with W_2/W_1 fixed, then $c \to 0$, $d \to 1$.*

PROOF. It is sufficient to show $c \to 0$ since $d \to 1$ then follows by symmetry. Let $W_2/W_1 = \tau$. It is sufficient to take the sequence $(n^2, \tau n^2)$ of W's and show that $c \to 0$ as $n \to \infty$. Using (4.5) it is sufficient to show that there is a sequence $\{T_n\}$ of tests for which the quantity $(W_2 - r_2)/r_1 \to \infty$ as $n \to \infty$. For each $n \geqq 1$ let T_n' be the test which takes exactly n observations and which minimizes $\max\,(\alpha_1, \alpha_2)$. Write $\alpha_i(n)$ for $\alpha_i(T_n')$. It is well-known that $\alpha_i(n) \to 0$ as $n \to \infty$. We have for T_n':

$$\frac{W_2 - r_2}{r_1} = \frac{\tau n^2(1 - \alpha_2(n)) - n}{n^2\alpha_1(n) + n}$$

which $\to \infty$ as $n \to \infty$. This proves the lemma.

LEMMA 5.6. *Let W be such that $c(W) < d(W)$. Let W be changed to $W + \Delta W$, with $\Delta W_i \geqq 0$ and at least one strict inequality. Then c decreases if $\Delta(W_2/W_1) \geqq 0$ and d increases if $\Delta(W_2/W_1) \leqq 0$.*

PROOF. It is sufficient to prove the first part of the conclusion of the lemma, the second part then following by symmetry. The condition $\Delta(W_2/W_1) \geqq 0$ together with the conditions on the ΔW_i imply $\Delta W_2 > 0$. Since $c < d$ for the given W, the two middle inequalities in (4.6) are strict so that on the right hand side of (5.5) $f(T^*, W)$ is the larger of the two quantities of which the maximum is taken. Using (5.4), keeping T^* fixed, it is elementary to show that $f(T^*, W)$ increases if $\Delta(W_2/W_1) \geqq 0$ and $\Delta W_2 > 0$, which establishes the lemma.

PROOF OF LEMMA 5.3. We need only prove part (i), part (ii) then following by symmetry. Choose $\tau > 0$ in such a way that

$$\tau A < \epsilon/(1 - \epsilon) \tag{5.6}$$

In the following W_2/W_1 will be kept fixed and equal to τ, i.e., W is of the form $(w, \tau w)$. The quantities c and d will then be continuous function of w; we will sometimes write $c(w)$, $d(w)$. The quantity $d(1 - c)/c(1 - d)$ is a continuous function of w which equals 1 if $w = 1$ (since then $c = d$) and which $\rightarrow \infty$ as $w \rightarrow \infty$, by Lemma 5.5. Since $A/B \geqq 1$, there is a value of w, say w_0, with $w_0 \geqq 1$, such that

$$\frac{A}{B} = \frac{d(w_0)}{1 - d(w_0)} \frac{1 - c(w_0)}{c(w_0)}. \tag{5.7}$$

Finally choose g such that

$$\frac{g_1}{g_2} = A \frac{1 - d(w_0)}{d(w_0)}. \tag{5.8}$$

By (4.6) we have

$$[1 - d(w_0)]/d(w_0) \leqq \tau \tag{5.9}$$

and (5.6), (5.8) and (5.9) imply $g_1/g_2 < \epsilon/(1 - \epsilon)$, or $g_1 < \epsilon$. The equations (5.7) and (5.8) are equivalent to the equations (5.1), which establishes (i) of the lemma, with the chosen g and with $W_1 = w_0$, $W_2 = \tau w_0$.

PROOF OF LEMMA 5.2. Suppose W and $W' = W + \Delta W$ are two solutions of (5.2), with given $c_0 < d_0$. Thus, when passing from W to W', c and d do not change. In that case, however, the ΔW_i cannot be of the same sign without both being 0, by Lemma 5.6, and the ΔW_i cannot be of opposite sign without both being 0, by Lemma 5.4. Thus, $W = W'$.

PROOF OF LEMMA 5.1. First it should be observed that it is possible to increase both W_i in such a way that the image $(c(W), d(W))$ keeps its first coordinate c fixed. For, increasing W_1 with W_2 fixed does not decrease c; an increase of c can be annihilated by increasing W_2 while leaving W_1 fixed, since, again by Lemma 5.4, c decreases and $\rightarrow 0$ as $W_2 \rightarrow \infty$. The strict decrease of c when W_2 increases while W_1 is fixed also shows that there is only one value of W_2 which restores the original value of c. This makes d a function of W_1, given the value of c. We shall write $d = d(W_1; c)$. By Lemma 5.6 this function is strictly increasing whenever $d > c$. It is also easy to see that the function is continuous: If W_1 is changed by an amount ΔW_1, keeping W_2 fixed, and then W_2 changed, keeping W_1 fixed, to restore the original value of c, d first changes by Δd_1, then by Δd_2, producing an overall change $\Delta d = \Delta d_1 + \Delta d_2$. Now Δd_1 has the same sign as ΔW_1, and the same is true of Δd, by the monotonicity of

$d(W_1; c)$, whereas Δd_2 has the opposite sign from ΔW_1. As a result, $|\Delta d| \leqq |\Delta d_1|$ which $\to 0$ as $\Delta W_1 \to 0$, since d is a continuous function of W_1 with W_2 fixed.

Suppose the W_i are increased in such a way that c remains constant >0. Then it must be true that $W_2/W_1 \to 0$, otherwise c could not remain constant, using Lemmas 5.5 and 5.4. It follows then that $d(W_1; c) \to 1$ as $W_1 \to \infty$, using Lemmas 5.5 and 5.4. Let now $W^* = (W_1^*, W_2^*)$, with $W_1^* = 1$, $W_2^* = (1 - c_0)/c_0$, then $c(W^*) = d(W^*) = c_0$, using (4.6) (observe that the two middle inequalities are equalities due to $W_1^* = 1$). We have now that $d(W_1; c_0)$ equals c_0 when $W_1 = W_1^* = 1$, and $\to 1$ as $W_1 \to \infty$. Since the function $d(W_1; c_0)$ is continuous and $c_0 \leqq d_0 < 1$ there is a value of $W_1 \geqq 1$ for which $d(W_1; c_0) = d_0$. That the corresponding W_2 is also $\geqq 1$ follows by symmetry and the uniqueness Lemma 5.2. This concludes the proof of Lemma 5.1.

6. Proof of Theorem 2.3. We precede the proof by a few lemmas. As before, the probability ratio of the first n observations will be denoted by Y_n; instead of Y_1 we shall simply write Y.

LEMMA 6.1. *Let* $0 < u' < u$, $P_2(Y > u) > 0$, $P_2(u' < Y \leqq u) > 0$, *and let* β_1 *be defined by* (2.7), *then* $\beta_1(u', u'+) > \beta_1(u, u+)$.

PROOF. Denote $\alpha_i(u, u+) = \alpha_i$, $\alpha_i(u', u'+) = \alpha_i + \Delta\alpha_i$, then

$$\alpha_1 = P_1(Y > u), \qquad 1 - \alpha_2 = P_2(Y > u), \tag{6.1}$$

$$\Delta\alpha_1 = P_1(u' < Y \leqq u), \qquad \Delta(1 - \alpha_2) = P_2(u' < Y \leqq u). \tag{6.2}$$

Consider the inequalities

$$\frac{1}{u'} > \frac{P_1(u' < Y \leqq u)}{P_2(u' < Y \leqq u)} \geqq \frac{1}{u} > \frac{P_1(Y > u)}{P_2(Y > u)}. \tag{6.3}$$

Substituting (6.1) and (6.2) into the second and fourth members of (6.3) gives $\Delta\alpha_1/\Delta(1 - \alpha_2) > \alpha_1/(1 - \alpha_2)$, which leads immediately to the conclusion of the lemma.

LEMMA 6.2. *For fixed* B, *the* $\alpha_i(B, A)$ *and* $\alpha_i(B-, A)$ *are left continuous, the* $\alpha_i(B, A+)$ *and* $\alpha_i(B-, A+)$ *right continuous functions of* A. *For fixed* A, *the* $\alpha_i(B, A)$ *and* $\alpha_i(B, A+)$ *are right continuous, the* $\alpha_i(B-, A)$ *and* $\alpha_i(B-, A+)$ *left continuous functions of* B.

PROOF. We need only show the continuity properties of $\alpha_1(B, A)$ and $\alpha_1(B, A+)$ for fixed B as a function of A. We shall drop the subscript 1 on P_1, and instead of the variable A we shall write x. Let $N(x)$ be the random number of observations determined by the SPRT $T(B, x)$, $B < x$. The probability ratio at stopping is then $Y_{N(x)}$. We shall show that $P(Y_{N(x)} \geqq x)$ is left continuous. In the first place, $EN(x) < \infty$, or $\sum_{n \geqq 1} P(N(x) \geqq n) < \infty$. Then, given $\epsilon > 0$, there exists k such that $\sum_{n \geqq k+1} P(N(x) \geqq n) < \epsilon$. Take y so that $B < y < x$, then

$$0 \leqq P(Y_{N(y)} \geqq y) - P(Y_{N(x)} \geqq x)$$
$$\leqq \sum_{n=1}^{\infty} P(B < Y_m < y, m = 1, \cdots, n - 1, y \leqq Y_n < x).$$

The last series we split into the sum from $n = 1$ to k, and a remaining series from $n = k + 1$ to ∞. The sum from $n = 1$ to k is bounded by $\sum_{n=1}^{k} P(Y \leqq Y_n < x)$ which $\downarrow 0$ as $y \uparrow x$. The series from $n = k + 1$ to ∞ is bounded by $\sum_{n \geqq k+1} P(B < Y_m < y, m = 1, \cdots, n - 1) \leqq \sum_{n \geqq k+1} P(N(x) \geqq n) < \epsilon$. This proves the left continuity of $\alpha_1(B, x)$. To show the right continuity of $\alpha_1(B, x+)$, let $M(x)$ be the random number of observations determined by $T(B, x+)$. Let $z > x$, then $P(Y_{M(x)} > z) \leqq P(Y_{M(z)} > z)$, so that $P(Y_{M(x)} > x) - P(Y_{M(z)} > z) \leqq P\{Y_{M(x)} > x\} - P\{Y_{M(x)} > z\} = P(x < Y_{M(x)} \leqq z) \downarrow 0$ as $z \downarrow x$. This proves the right continuity of $P(Y_{M(x)} > x) = \alpha_1(B, x+)$, and the proof of the lemma is complete.

We can verify now the assertion made in Section 2 that $\alpha_i(B, A+) = \lim \alpha_i(B, A')$ as $A' \downarrow A$, and similar statements with $B-$. Verifying this for $\alpha_1(B, A+)$ we have $\alpha_1(B, A'+) \leqq \alpha_1(B, A') \leqq \alpha_1(B, A+)$. If we let now $A' \downarrow A$, the two extreme members of the resulting inequalities are equal because of the right continuity of $\alpha_1(B, A'+)$ as a function of A'.

LEMMA 6.3. *For any B for which $P_2(Y > B) > 0$ let β_1 be defined by* (2.7). *Then, as a function of x, $\beta_1(B, x)$ is non-increasing, continuous from the left, and $\beta_1(B, x) \to 0$ as $x \to \infty$. More specifically, if $x' > x$ and $T(B, x)$ is not equivalent to $T(B, x')$, then $\beta_1(B, x) > \beta_1(B, x')$.*

PROOF. The monotonicity of $\beta_1(B, x)$ was proved in [12], Corollary 2. The left continuity follows from Lemma 6.2. Finally,

$$\beta_1(B, x) = \frac{\alpha_1(B, x)}{1 - \alpha_2(B, x)} = \frac{P_1(Y_N \geqq x)}{P_2(Y_N \geqq x)} \leqq \frac{1}{x}.$$

PROOF OF THEOREM 2.3. Let $T = T(B, A)$ be as in the hypothesis of Theorem 2.3. We shall first show the existence of the test T' mentioned in the theorem. According to Lemma 6.3, $\beta_1(B, B+) \geqq \beta_1(B, A)$, and according to Lemma 6.1 $\beta_1(1, 1+) \geqq \beta_1(B, B+)$. Moreover, equalities in both of these inequalities can occur only if $T(B, A)$ is equivalent to $T(B, B+)$, and $T(B, B+)$ is equivalent to $T(1, 1+)$. However, in that case $T(B, A)$ would be equivalent to $T(1, A)$. Since this is excluded, by the hypothesis of the theorem, one of the inequalities must be strict, and we have $\beta_1(1, 1+) > \beta_1(T)$. Observing that $P_2(Y > 1) > 0$, we can apply Lemma 6.3 to the function $\beta_1(1, x)$. There is then a value A' for x, $A' > 1$, such that $\beta_1(1, A') \geqq \beta_1(T) \geqq \beta_1(1, A'+)$. It follows that there exists π, $0 \leqq \pi \leqq 1$, such that $T' = (1 - \pi)T(1, A') + \pi T(1, A'+)$ has $\beta_1(T') = \beta_1(T)$. Of course, if $\beta_1(1, x)$ is continuous in A' we can simply take $\pi = 0$.

It should be noticed that $\beta_1(T') = \beta_1(T)$ implies that in the α-plane the point $\alpha(T')$ lies on the line through $(0, 1)$ and $\alpha(T)$. Note also that $(0, 1) = \alpha(T_1)$. It will be shown now that $\alpha_2(T') < \alpha_2(T)$, or, in other words, that $\alpha(T')$ does not lie on the segment between $(0, 1)$ and $\alpha(T)$. Suppose first $A' \leqq A$. Then to pass from T to T' the lower stopping bound is decreased and the upper one not increased. Since the two tests are not equivalent, by the hypothesis of the theorem, α_2 decreases. If, on the other hand, $A' > A$, to pass from T to T'

the lower stopping bound is decreased, the upper one increased. Since the two tests are not equivalent, $\nu_i(T') > \nu_i(T)$. If we had $\alpha_2(T') \geqq \alpha_2(T)$, i.e., $\alpha(T')$ lying on the segment between $(0, 1)$ and $\alpha(T)$, there would be a mixture T'' of T_1 and T such that $\alpha(T'') = \alpha(T')$, namely $T'' = (1 - \pi')T_1 + \pi' T$ with $\pi' = (1 - \alpha_2(T'))/(1 - \alpha_2(T))$. Since $\nu_i(T'') = \pi' \nu_i(T) \leqq \nu_i(T) < \nu_i(T')$, the test T'' has the same α_i as T' but has smaller ν_i. This is impossible since T' has OP.

Let T^* be as in the conclusion of the theorem. With λ defined by (2.8) and remembering $\beta_1(T) = \beta_1(T')$, we compute easily $\alpha_1(T^*) = \lambda\alpha_1(T') = \alpha_1(T)$, and $\alpha_2(T^*) = 1 - \lambda + \lambda\alpha_2(T') = \alpha_2(T)$, so that $\alpha_i(T^*) = \alpha_i(T)$ as asserted in the conclusion of the theorem. To prove the remaining assertions note that T^* is an extended SPRT (Definition 2.5) so that T^* has OP by Corollary 2.1. Since $\alpha_i(T) \leqq \alpha_i(T^*)$ (actually equal) it follows that $\nu_i(T) \geqq \nu_i(T^*)$. If these latter inequalities were both equalities, T would be Bayes (g, W) for any (g, W) for which T^* is Bayes. Now take any g with $g_i > 0$, substitute in Equations (5.1) $B = 1$, $A = A'$, and solve for W (the existence of a solution guaranteed by Lemma 5.1). With this g and W, if a test is Bayes (g, W), it is equivalent to some extended SPRT with stopping bounds 1 and A'. T^* satisfies this but T does not, so that at least one of the inequalities $\nu_i(T) \geqq \nu_i(T^*)$ must be strict. This concludes the proof of the theorem.

REFERENCES

[1] Arrow, K. J., Blackwell, D. and Girshick, M. A. (1949). Bayes and minimax solutions of sequential decision problems. *Econometrica* **17** 213–244.

[2] Burkholder, D. L. and Wijsman, R. A. (1960). Optimum properties and admissibility of sequential tests. *Ann. Math. Statist.* **31** 232 (abstract).

[3] Ghosh, J. K. (1961). On the optimality of probability ratio tests in sequential and multiple sampling. *Calcutta Statist. Assoc. Bull.* **10** 73–92.

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

[5] Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

[6] Wald, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Statist.* **16** 117–186.

[7] Wald, A. (1947). *Sequential Analysis*. Wiley, New York.

[8] Wald, A. (1950). *Statistical Decision Functions*. Wiley, New York.

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

[10] Wald, A. and Wolfowitz, J. (1950). Bayes solutions of sequential decision problems. *Ann. Math. Statist.* **21** 82–99.

[11] Wolfowitz, J. (1955). Review of *Theory of Games and Statistical Decisions* by David Blackwell and M. A. Girshick. *Bull. Amer. Math. Soc.* **61** 247–253.

[12] Wijsman, R. A. (1960). A monotonicity property of the sequential probability ratio test. *Ann. Math. Statist.* **31** 677–684.