OPTIMUM CHARACTER OF THE SEQUENTIAL PROBABILITY RATIO TEST

A. WALD AND J. WOLFOWITZ

Columbia University

1. Summary. Let S_0 be any sequential probability ratio test for deciding between two simple alternatives H_0 and H_1, and S_1 another test for the same purpose. We define $(i, j = 0, 1)$:

$\alpha_i(S_j)$ = probability, under S_j, of rejecting H_i when it is true;

$E_i^j(n)$ = expected number of observations to reach a decision under test S_j when the hypothesis H_i is true. (It is assumed that $E_i^1(n)$ exists.)

In this paper it is proved that, if

$$\alpha_i(S_1) \leq \alpha_i(S_0) \qquad (i = 0,1),$$

it follows that

$$E_i^0(n) \leq E_i^1(n) \qquad (i = 0, 1).$$

This means that of all tests with the same power the sequential probability ratio test requires on the average fewest observations. This result had been conjectured earlier ([1], [2]).

2. Introduction. Let $p_i(x)$, $i = 0, 1$, denote two different probability density functions or (discrete) probability functions. (Throughout this paper the index i will always take the values 0, 1). Let X be a chance variable whose distribution can only be either $p_0(x)$ or $p_1(x)$, but is otherwise unknown. It is required to decide between the hypotheses H_0, H_1, where H_i states that $p_i(x)$ is the distribution of X, on the basis of n independent observations $x_1, \cdots, x_n$ on X, where n is a chance variable defined (finite) on almost every infinite sequence

$$\omega = x_1, x_2, \cdots$$

i.e., n is finite with probability one according to both $p_0(x)$ and $p_1(x)$. The definition of $n(\omega)$ together with the rule for deciding on H_0 or H_1 constitute a sequential test.

A sequential probability ratio test is defined with the aid of two positive numbers, $A^* > 1$, $B^* < 1$, as follows: Write for brevity

$$p_{ij} = \prod_{k=1}^{j} p_i(x_k).$$

Then $n = j$ if

$$\frac{p_{1j}}{p_{0j}} \geq A^* \qquad \text{or} \qquad \leq B^*$$

and

$$B^* < \frac{p_{1k}}{p_{0k}} < A^*, \qquad k < j.$$

If

$$\frac{p_{1n}}{p_{0n}} \geq A^*, \qquad \text{the hypothesis } H_1 \text{ is accepted,}$$

if

$$\frac{p_{1n}}{p_{0n}} \leq B^* \text{ the hypothesis } H_0 \text{ is accepted.}$$

In this paper we limit consideration to sequential tests for which $E_i(n)$ exists, where $E_i(n)$ is the expected value of n when H_i is true (i.e., when $p_i(x)$ is the distribution of X). It has been proved in [3] that all sequential probability ratio tests belong to this class. The purpose of the paper is to prove the result stated in the first section. Throughout the proof we shall find it convenient to assume that there is an a priori probability g_i that H_i is true ($g_0 + g_1 = 1$; we shall write $g = (g_0, g_1)$). We are aware of the fact that many statisticians believe that in most problems of practical importance either no a priori probability distribution exists, or that even where it exists the statistical decision must be made in ignorance of it; in fact we share this view. Our introduction of the a priori probability distribution is a purely technical device for achieving the proof which has no bearing on statistical methodology, and the reader will verify that this is so. We shall always assume below that $g_0 \neq 0, 1$.

Let W_0, W_1, c be given positive numbers. We define

$$R = g_0(W_0\alpha_0 + cE_0(n)) + g_1(W_1\alpha_1 + cE_1(n)),$$

and call R the average risk associated with a test S and a given g (obviously R is a function of both). We shall say that H_i is accepted when the decision is made that $p_i(x)$ is the distribution of X. We shall say that H_0 is rejected when H_1 is accepted, and vice versa. The reader may find it helpful to regard W_i as a weight which measures the loss caused by rejecting H_i when it is true, c as the cost of a single observation, and R as the average loss associated with a given g and a test S. For mathematical purposes these are simply quantities which we manipulate in the course of the proof.

3. Role of the probability ratio. Let g, $W = (W_0, W_1)$, and c be fixed. Let S be a given sequential test, with $R(S)$ the associated risk and $n(\omega, S)$ the associated "sample size" function. Let $\psi(x_1, \cdots, x_n)$ be the "decision" function; this is a function which takes only the values 0 and 1, and such that, when $x_1, \cdots, x_n$ is the sample point, the hypothesis with index $\psi(x_1, \cdots, x_n)$ is *rejected*. Define the following decision function $\varphi(x_1, \cdots, x_n)$: $\varphi = 0$ when

$$\lambda = \frac{W_1 g_1 p_{1n}}{W_0 g_0 p_{0n}}$$

is greater than 1, and $\varphi = 1$ when $\lambda < 1$. When $\lambda = 1$, φ may be 0 or 1 at pleasure.

It must be remembered that an actual decision function is a single-valued function of $(x_1, \cdots, x_n)$. We note, however, that

a) the relevant properties of a test are not affected by changing the test on a set T of points ω whose probability is zero according to both H_0 and H_1, i.e., changing the definition on T of n and/or of the decision function, leaves α_0, α_1, $E_0(n)$ and $E_1(n)$ unaltered. In particular, the average risk R remains unchanged.

b) the set of points for which $p_{0n} = p_{1n} = 0$ and λ is indeterminate, has probability zero according to both H_0 and H_1.

In view of the above we decide arbitrarily, in all sequential tests which we shall henceforth consider, to define $n = j$, and $\psi = 0$, whenever $p_{0j} = p_{1j} = 0$, and $n \neq 1, \cdots, (j-1)$. By this arbitrary action $R(S)$ will not be changed.

Let now

$$L_{in} = \frac{W_i g_i p_{in}}{g_0 p_{0n} + g_1 p_{1n}};$$

$$L_n = cn + \min(L_{0n}, L_{1n}).$$

We have

$$EL_{\psi n} = \Sigma g_i W_i \alpha_i$$

where the operator E denotes the expected value with respect to the joint distribution of H_i and $(x_1, \cdots, x_n)$, i.e., E is the operator $g_0E_0 + g_1E_1$. If now the event $\{\psi(S) \neq \varphi \text{ and } \lambda \neq 1\}$ has positive probability according to either H_0 or H_1, we would have, for $n = n(\omega, S)$,

$$EL_{\varphi n} < EL_{\psi n}.$$

Hence, if the decision function ψ connected with the test S were replaced by the decision function φ, R would be decreased. Since our object throughout this proof will be to make R as small as possible, we shall confine ourselves henceforth, except when the contrary is explicitly stated, to tests for which φ is the decision function. This will be assumed even if not explicitly stated.

The function φ has not yet been uniquely defined when $\lambda = 1$. A definition convenient for later purposes will be given in the next section. R is the same for all definitions.

We thus have that φ is a function only of λ, or, what comes to the same thing when W is fixed, of $r_n = \frac{p_{1n}}{p_{0n}}$. Define

$$r_j = \frac{p_{1j}}{p_{0j}}, \qquad j = 1, 2, \cdots.$$

We shall now prove

LEMMA 1. *Let g, W, and c be fixed. There exists a sequential test S^* for which the average risk is a minimum. Its sample size function $n(\omega, S^*)$ can be defined by means of a properly chosen subset K of the non-negative half-line as follows: For any ω consider the associated sequence*

$$r_1, r_2, \cdots$$

and let j be the smallest integer for which $r_j \in K$. Then $n = j$. The function n may be undefined on a set of points ω whose probability according to H_0 and H_1 is zero.

Let $a = (a_1, \cdots, a_d)$ be any point in some finite d-dimensional Euclidean space, provided only that $p_{0d}(a)$ and $p_{1d}(a)$ are not both zero. Let $b = \dfrac{p_{1d}(a)}{p_{0d}(a)}$ and let $l(a) = cd + \min(L_{0d}, L_{1d})$. Let D be any sequential test whatever for which $n(\omega, D) > d$ for any ω whose first d coördinates are the same as those of a, and for which $E(n \mid a, D) < \infty$, where $E(n \mid a, D)$ is the conditional expected value of n according to the test D under the condition that the first d coördinates of ω are the same as those of a. For brevity let G represent the set of points ω which fulfill this last condition, i.e., that the first d coördinates of ω are the same as those of a. Finally, let $E(L_n \mid a, D)$ be the conditional expected value of L_n according to D under the condition that ω is in the set G. We know that $\min(L_{0d}, L_{1d})$ depends only on $r_d(a) = b$.

Write

$$\nu(a) = \sup_D [l(a) - E(L_n \mid a, D)].$$

Let $a_0 = (a_{01}, \cdots, a_{0k})$ be any point such that

$$\frac{p_{1d}(a)}{p_{0d}(a)} = \frac{p_{1k}(a_0)}{p_{0k}(a_0)}.$$

Let D_0 be any sequential test whatever for which $n(\omega, D_0) > k$ for any ω whose first k coordinates are the same as those of a_0, and for which $E(n \mid a_0, D_0) < \infty$ Let

$$\nu(a_0) = \sup_{D_0} [l(a_0) - E(L_n \mid a_0, D_0)].$$

We shall prove that $\nu(a) = \nu(a_0)$. Thus we shall be justified in writing

$$\gamma(b) = \nu(a) = \nu(a_0).$$

Suppose, therefore that $\nu(a) > \nu(a_0)$. Let D_1 be a test of the type D such that

$$l(a) - E(L_n \mid a, D_1) > \frac{\nu(a) + \nu(a_0)}{2}.$$

We now partially define another sequential test D_{10} of the type D_0 as follows: Let

$$\bar{a} = a_1, \cdots, a_d, y_1, \cdots, y_t,$$

be any sequence such that $n(\bar{a}, D_1) = d + t$. Then for the sequence

$$\bar{a}_0 = a_{01}, \cdots, a_{0k}, y_1, \cdots, y_t,$$

let $n(\bar{a}_0, D_{10}) = k + t$. The decision function ψ_0 associated with D_{10} will be partially defined as follows:

$$\psi_0(\bar{a}_0) = \varphi(\bar{a}).$$

(The reader will observe that it may happen that $\psi_0(\bar{a}_0) \neq \varphi(\bar{a}_0)$). Since $r_d(a) = r_k(a_0)$ it follows that

$$l(a) - E(L_n \mid a, D_1) = l(a_0) - E(L_n \mid a_0, D_{10}) > \frac{\nu(a) + \nu(a_0)}{2} > \nu(a_0),$$

in violation of the definition of $\nu(a_0)$. A similar contradiction is obtained if $\nu(a) < \nu(a_0)$. Hence $\nu(a) = \nu(a_0)$ as was stated above.

We define K to consist of all numbers b which are such that there exist points a with $r_d(a) = b$, and for which $\gamma(b) \leq 0$. We shall now prove that the test S^* defined in the statement of the lemma is such that $R(S^*)$ is a minimum. Recall that the average risk is the expected value of L_n. Let S be any other test. Let $a^* = (\overset{*}{a_1}, \cdots, \overset{*}{a_{d^*}})$ be any sequence such that either $n(a^*, S^*) = d^*$, or $n(a^*, S) = d^*$, but $n(a^*, S^*) \neq n(a^*, S)$. We exclude the trivial case that the probability of the occurrence of such a sequence, under both H_0 and H_1, is zero. Let $r_{d^*}(a^*) = b^*$. The sequence a^* may be one of three types:

1) $\gamma(b^*) < 0$. Hence $b^* \in K$, $n(a^*, S) > d^*$. It is more advantageous, from the point of view of diminishing the average risk, to terminate the sequential process at once, since $E(L_n \mid a^*, S) > l(a^*)$.

2) $\gamma(b^*) = 0$. Hence $b^* \in K$, $n(a^*, S) > d^*$. If $l(a^*) - E(L_n \mid a^*, S) = 0$, i.e., the supremum is actually attained by S, then, as far as the average risk is concerned, it makes no difference whether the sequential process is terminated with a^* or continued according to S. If, however, $l(a^*) - E(L_n \mid a^*, S) < 0$, it is clearly disadvantageous to proceed according to S. It is impossible that $l(a^*) - E(L_n \mid a^*, S) > 0$, since $\gamma(b^*) = 0$.

3) $\gamma(b^*) > 0$. Hence $b^* \notin K$, $n(a^*, S) = d^*$. Clearly it is more advantageous from the point of view of diminishing the average risk not to terminate the sequential process, but to continue with at least one more observation. After one more observation we are either in case 1 or 2, where it is advantageous to terminate the sequential process, or again in case 3, where it is advantageous to take yet another observation.

We conclude that $R(S^*)$ is a minimum, as was to be proved.

4. A fundamental lemma. Consider the complement of K with respect to the non-negative half-line, and from it delete all points b' for which there exists no point a in some d-dimensional Euclidean space such that $r_d(a) = b'$. The point 1 is never to be considered as of the type of b', i.e., 1 is never to be deleted. Designate the resulting set by $\bar{K}$.

Our proof of the theorem to which this paper is devoted hinges on the following lemma:

LEMMA 2. *Let W, g, c be fixed, and $\bar{K}$ be as defined above. There exist two positive numbers A and B, with $B \leq \frac{W_0 g_0}{W_1 g_1} \leq A$, such that*

a) if $b \in K$, then either $b \geq A$ or $b \leq B$

b) if $b \in \bar{K}$, $B \leq b \leq A$.

Two remarks may be made before proceeding with the proof:

1) We may now complete the definition of φ for tests of the type of S^*. The reader will recall that φ was not uniquely defined when $\lambda = 1$, i.e., when $r_n = \frac{W_0 g_0}{W_1 g_1}$. Lemma 2 shows that it is necessary to define $\varphi(\lambda)$ only when $\lambda = \frac{W_0 g_0}{W_1 g_1} \in K$ and λ is therefore either A or B. We will define $\varphi\left(\frac{W_0 g_0}{W_1 g_1}\right)$ as 0 or 1, according as $\frac{W_0 g_0}{W_1 g_1}$ is A or B, and $A \neq B$. This is simply a convenient definition which will give uniqueness. When $A = B = \frac{W_0 g_0}{W_1 g_1} \in K$, the situation is completely trivial, and we may take $\varphi = 0$ arbitrarily.

2) If $1 \in K$ the above lemma shows that the average risk is minimized (for fixed W, g, c, of course) by taking no observations at all. We have $\varphi = 0$ or 1 according as $1 \geq A$ or $1 \leq B$.

PROOF OF THE LEMMA: Let $h > \frac{W_0 g_0}{W_1 g_1}$ be a point in $\bar{K}$. We will prove that any point h' such that $\frac{W_0 g_0}{W_1 g_1} \leq h' < h$, and such that there exists a point a' in some d'-dimensional Euclidean space for which $r_{d'}(a') = h'$, is also in $\bar{K}$. In a similar way it can be shown that, if $h_0 < \frac{W_0 g_0}{W_1 g_1}$ is any point in $\bar{K}$, any point h_0' such that $h_0 < h_0' \leq \frac{W_0 g_0}{W_1 g_1}$, and such that there exists a point a_0' in some d''-dimensional Euclidean space for which $r_{d''}(a_0') = h_0'$, is also in $\bar{K}$. This will prove the lemma.

Let therefore h and h' be as above. Let S^* be the sequential test based on K, with the decision function φ. Let a be a point in d-space such that $r_d(a) = h$. Since $h \in \bar{K}$ we have $\gamma(h) > 0$.

We now wish to define partially another sequential test $\bar{S}$, with a decision function which may be different from φ, as follows: Let a' be defined as above. Write

$$a = (a_1, \cdots, a_d)$$
$$a' = (a_1', \cdots, a_{d'}').$$

Let

$$\bar{a} = a_1, \cdots, a_d, y_1, \cdots, y_t$$

be any sequence such that $n(\bar{a}, S^*) = d + t$. Then for the sequence

$$\bar{a}' = a_1', \cdots, a_{d'}', y_1, \cdots, y_t$$

let $n(\bar{a}', \bar{S}) = d' + t$. The decision function ψ associated with $\bar{S}$ will be partially defined as follows:

$$\psi(\bar{a}') = \varphi(\bar{a}).$$

Clearly

$$E_i(n \mid a, S^*) - d = E_i(n \mid a', \bar{S}) - d' \qquad (i = 0, 1) \tag{4.1}$$

and

$$E_i(\varphi \mid a, S^*) = E_i(\psi \mid a', \bar{S}) \qquad (i = 0, 1). \tag{4.2}$$

Furthermore, we have

$$\begin{aligned} &l(a) - E(L_n \mid a, S^*) \\ &\quad = \frac{g_0}{g_0 + g_1 h} \{W_0 + cd - cE_0(n \mid a, S^*) - W_0[1 - E_0(\varphi \mid a, S^*)]\} \\ &\qquad + \frac{g_1 h}{g_0 + g_1 h} \{cd - cE_1(n \mid a, S^*) - W_1 E_1(\varphi \mid a, S^*)\}. \end{aligned} \tag{4.3}$$

Since $\gamma(h) > 0$, and since

$$cd - cE_1(n \mid a, S^*) - W_1 E_1(\varphi \mid a, S^*) < 0, \tag{4.4}$$

we must have

$$W_0 + cd - cE_0(n \mid a, S^*) - W_0[1 - E_0(\varphi \mid a, S^*)] > 0. \tag{4.5}$$

From $h' < h$ it follows that

$$\frac{g_0}{g_0 + g_1 h'} > \frac{g_0}{g_0 + g_1 h}, \qquad \text{and} \qquad \frac{g_1 h'}{g_0 + g_1 h'} < \frac{g_1 h}{g_0 + g_1 h}. \tag{4.6}$$

Relations (4.1), (4.2), (4.4), (4.5) and (4.6) imply that the value of the right hand member of (4.3) is increased by replacing φ, h, a, S^* and d by ψ, h', a', $\bar{S}$, and d', respectively. This proves our lemma.

If there are values which r_j cannot assume the pair B, A might not be unique. For convenience we shall define A and B uniquely in the manner described below. We will always adhere to this definition thereafter.

We shall first define $\gamma(h)$ for all positive h in a manner consistent with the previous definition, which defined $\gamma(h)$ only for those values of h which could be assumed by r_j. Let h be any positive number and $D(h)$ be any sequential test with the following properties:

(4.7) there exists a set $Q(h)$ of positive numbers such that $n = j$ if and only if the j-th member of the sequence

$$hr_1, hr_2, hr_3, \cdots$$

is the first element of the sequence to be in $Q(h)$

$$E_i(n \mid D(h)) < \infty \qquad (i = 0, 1). \tag{4.8}$$

We define, for $h \geq \frac{W_0 g_0}{W_1 g_1}$,

$$\gamma(h \mid D(h)) = \frac{g_0}{g_0 + g_1 h} \{W_0 E_0(\varphi \mid D(h)) - cE_0(n \mid D(h))\} \tag{4.9}$$

$$+ \frac{g_1 h}{g_0 + g_1 h} \{-W_1 E_1(\varphi \mid D(h)) - cE_1(n \mid D(h))\},$$

$$\gamma(h) = \sup_{D(h)} \gamma(h \mid D(h)) \tag{4.10}$$

with a corresponding definition for $h \leq \frac{W_0 g_0}{W_1 g_1}$. Thus $\gamma(h)$ is defined for all positive h. This definition coincides with the previous definition whenever the latter is applicable. It is true that the supremum operation in (4.10) is limited to tests which depend only on the probability ratio, as (4.7) implies, but the argument of Lemma 1 shows that this limitation does not diminish the supremum. (It might appear that, for $h = \frac{W_0 g_0}{W_1 g_1}$, $\gamma(h)$ is not uniquely defined. We shall shortly see that this is not the case.)

The quantity $\gamma(h)$ depends, of course, on g_0 and g_1. To put this in evidence, we shall also write $\gamma(h, g_0, g_1)$. One can easily verify that

$$\gamma(h, g_0, g_1) = \gamma\left(1, \frac{g_0}{g_0 + g_1 h}, \frac{g_1 h}{g_0 + g_1 h}\right).$$

More generally, for any positive values h and h', we have $\gamma(h, g_0, g_1) = \gamma(h', \bar{g}_0, \bar{g}_1)$, where $\bar{g}_0$ and $\bar{g}_1$ are suitable functions of g_0, g_1, h, and h'. Thus, if h is not an admissible value of the probability ratio and h' is any admissible value, we can interpret the value of $\gamma(h, g_0, g_1)$ as the value of γ corresponding to h' and some properly chosen a priori probabilities $\bar{g}_0$ and $\bar{g}_1$.

We now define A as the greatest lower bound of all points $h \geq \frac{W_0 g_0}{W_1 g_1}$ for which $\gamma(h) \leq 0$. We define B as the least upper bound of all points $h \leq \frac{W_0 g_0}{W_1 g_1}$ for which $\gamma(h) \leq 0$. If $\gamma(h) \leq 0$ for all h the above definition implies $A = B = \frac{W_0 g_0}{W_1 g_1}$.

The argument of Lemma 2 shows that $\gamma(h)$ is monotonically increasing in the interval $\left(B, \frac{W_0 g_0}{W_1 g_1}\right)$, and that $\gamma(h)$ is monotonically decreasing in the interval $\left(\frac{W_0 g_0}{W_1 g_1}, A\right)$.

We shall now define a sequential test $S^*(h)$ for every positive h. The decision function of $S^*(h)$ will be φ, and $n = j$ if and only if the j-th member of the sequence

$$\gamma(hr_1), \quad \gamma(hr_2), \quad \gamma(hr_3), \cdots$$

is the first element to be ≤ 0. We see that

$$\gamma(h) = \gamma(h \mid S^*(h)) \tag{4.11}$$

for *all* h. Incidentally, this proves that $\gamma(h)$ was uniquely defined at $h = \frac{W_0 g_0}{W_1 g_1}$.

We shall now prove

LEMMA 3. *The function $\gamma(h)$ has the following properties:*

a) It is continuous for all h.

b) $\gamma(A) = \gamma(B) = 0$

c) $\gamma(h) < 0$ for $h > A$ or $< B$.

Only a) and c) require proof, since b) is a trivial consequence of a) and the definition of A and B.

Let h be any point except $\frac{W_0 g_0}{W_1 g_1}$, and let z be any point in a neighborhood of h. Within a neighborhood of h both $E_0(n \mid S^*(z))$ and $E_1(n \mid S^*(z))$ are bounded. Let Δ be an arbitrarily given, positive number. Let h' and h'' be any two points in a sufficiently small neighborhood of h, to be described shortly. We proceed as in the argument of Lemma 2, with the present h' corresponding to h of Lemma 2, the present h'' corresponding to h' of Lemma 2, and with $S^*(h')$ corresponding to S^* of Lemma 2. Since $\frac{g_0}{g_0 + g_1 z}$ and $\frac{g_1 z}{g_0 + g_1 z}$ are continuous functions of z, and since $E_0(n \mid S^*(z))$ and $E_1(n \mid S^*(z))$ are bounded functions of z, we conclude that, when the neighborhood of h is sufficiently small,

$$\gamma(h'') \geq \gamma(h') - \Delta.$$

Reversing the roles of h' and h'' we obtain that in this neighborhood

$$\gamma(h') \geq \gamma(h'') - \Delta,$$

and conclude that

$$| \gamma(h') - \gamma(h'') | \leq \Delta.$$

Since Δ was arbitrary, this implies the continuity of $\gamma(h)$ everywhere, except perhaps at $h = \frac{W_0 g_0}{W_1 g_1}$.

To deal with the point $h = \frac{W_0 g_0}{W_1 g_1}$, proceed as follows: Using the above argument and the definition (4.9), (4.10), we prove that $\gamma(h)$ is continuous on the right at $h = \frac{W_0 g_0}{W_1 g_1}$. Using, at the point $h = \frac{W_0 g_0}{W_1 g_1}$, the definition of $\gamma(h \mid D(h))$ for $h \leq \frac{W_0 g_0}{W_1 g_1}$ i.e.,

$$\gamma(h \mid D(h)) = \frac{g_0}{g_0 + g_1 h} \{-W_0 E_0(1 - \varphi \mid D(h)) - cE_0(n \mid D(h))\} \tag{4.12}$$
$$+ \frac{g_1 h}{g_0 + g_1 h} \{W_1 E_1(1 - \varphi \mid D(h)) - cE_1(n \mid D(h))\},$$

(4.10) and (4.11), we prove that $\gamma(h)$ is continuous on the left at $h = \frac{W_0 g_0}{W_1 g_1}$. This proves a).

To prove c), we proceed as follows: Suppose for $h_0 > A$ we had $\gamma(h_0) = 0$. Since

$$\{ -W_1 E_1(\varphi \mid S^*(h_0)) - cE_1(n \mid S^*(h_0))\} < 0,$$

we would have that

$$\{W_0 E_0(\varphi \mid S^*(h_0)) - cE_0(n \mid S^*(h_0))\} > 0.$$

An argument like that of Lemma 2 would then show that $\gamma(h) > 0$ for $\frac{W_0 g_0}{W_1 g_1} < h < h_0$. This, however, is impossible, because it is a violation of the definition of A.

In a similar way we prove that if $h < B$, $\gamma(h) < 0$. This proves c) and with it the lemma.

5. The behavior of A and B. LEMMA 4. *Let g and c be fixed. Then A and B are continuous functions of W_0 and W_1.*

PROOF: It will be sufficient to prove that A is continuous, the proof for B being similar. Suppose $A > B$. Let h_1 and h_2 be such that

a) $B < h_1 < A < h_2$;

b) $h_2 - h_1 < \Delta$ for an arbitrary positive Δ.

We write $\gamma(h)$ temporarily as $\gamma(h, W_0, W_1)$ in order to exhibit the dependence on W_0 and W_1. Then

$$\gamma(h_1, W_0, W_1) > 0;$$
$$\gamma(h_2, W_0, W_1) < 0.$$

It follows from (4.9) that $\gamma(h \mid D(h))$ is continuous in W_0, W_1, uniformly in $D(h)$. Hence $\gamma(h, W_0, W_1) = \sup_{D(h)} \gamma(h \mid D(h))$ is also continuous in W_0, W_1. Hence, for ΔW_0 and ΔW_1 sufficiently small,

$$\gamma(h_1, W_0 + \Delta W_0, W_1 + \Delta W_1) > 0;$$
$$\gamma(h_2, W_0 + \Delta W_0, W_1 + \Delta W_1) < 0.$$

Therefore

$$h_1 \leq A(W_0 + \Delta W_0, W_1 + \Delta W_1) \leq h_2,$$

which proves continuity, since Δ was arbitrary.

If $\frac{W_0 g_0}{W_1 g_1} = A = B$, we take $h_1 < \frac{W_0 g_0}{W_1 g_1} < h_2$, $h_2 - h_1 < \Delta$, and by a similar argument show that

$$\gamma(h_1, W_0 + \Delta W_0, W_1 + \Delta W_1) < 0;$$
$$\gamma(h_2, W_0 + \Delta W_0, W_1 + \Delta W_1) < 0.$$

Thus

$$h_1 \leq B(W_0 + \Delta W_0, W_1 + \Delta W_1) \leq A(W_0 + \Delta W_0, W_1 + \Delta W_1) \leq h_2.$$

This proves the lemma.

LEMMA 5. *Let g, c, and W_1 be fixed. A is strictly monotonic in W_0. As W_0 approaches 0, A approaches 0; as W_0 approaches $+\infty$, A also approaches $+\infty$.*

PROOF: Since $A \geq \frac{W_0 g_0}{W_1 g_1}$, $A \to +\infty$ as $W_0 \to +\infty$. If $W_0 < c$ no reduction in average risk could compensate for taking even a single observation, no matter what the value of h. Hence $\gamma(h) \leq 0$ for all h when $W_0 < c$, so that $A = B$. Since $B \leq \frac{W_0 g_0}{W_1 g_1}$, $B \to 0$ as $W_0 \to 0$. Hence $A \to 0$ as $W_0 \to 0$. It is evident from (4.9) that $\gamma(h \mid D(h))$ is non-decreasing with increasing W_0 (everything else fixed). Hence also

$$\gamma(h) = \sup_{D(h)} \gamma(h \mid D(h)),$$

is non-decreasing with increasing W_0, for fixed $h > \frac{W_0 g_0}{W_1 g_1}$ and fixed W_1. For a positive Δ sufficiently small and for any h such that $A \leq h < A + \Delta$, we have that

$$E_0 (\varphi \mid S^*(h)) > 0.$$

Hence, for such h, $\gamma(h, W_0, W_1)$ is strictly monotonically increasing with increasing W_0. Therefore A is (strictly) monotonically increasing with increasing W_0.

We now define the function $W_0(W_1, \delta)$ of the two positive arguments W_1, δ so that

$$A(W_0(W_1, \delta), W_1) = \delta.$$

By Lemma 5 such a function exists and is single-valued.

6. Properties of the function $W_0(W_1, \delta)$. LEMMA 6. *$W_0(W_1, \delta)$ is continuous in W_1.*

PROOF: Let

$$\lim_{N \to \infty} W_{1N} = W_1,$$

and suppose that the sequence $\{W_0(W_{1N}, \delta)\}$ did not converge. Suppose W_0' and W_0'' were two distinct limit points of this sequence. From the continuity of A (Lemma 4) it would follow that

$$A(W_0', W_1) = A(W_0'', W_1)$$

This, however, violates Lemma 5. The only remaining possibility to be considered is that

$$\lim_{N\to\infty} W_0(W_{1N}, \delta) = \infty.$$

If that were the case, then, since $A \geq \dfrac{W_0 g_0}{W_1 g_1}$, it would follow that $A \to \infty$, in violation of the fact that $A \equiv \delta$.

LEMMA 7. *We have, for fixed* δ,

$$\lim_{W_1\to 0} W_0(W_1) = 0;$$

$$\lim_{W_1\to\infty} W_0(W_1) = \infty.$$

PROOF: If, for small W_1, $W_0(W_1)$ were bounded below by a positive number, then, since $A \geq \dfrac{g_0 W_0(W_1, \delta)}{W_1 g_1}$, we could make A arbitrarily large by taking W_1 sufficiently small, in violation of the fact that $A \equiv \delta$. To prove the second half of the lemma, assume that $W_0(W_1)$ is bounded above as $W_1 \to \infty$. Then $B\left(\leq \dfrac{W_0 g_0}{W_1 g_1}\right)$ will approach zero as $W_1 \to \infty$. Let h be fixed so that $B < h < \delta$. Consider the totality of points ω for which there exists an integer $n^*(\omega)$ such that:

$$hr_{n^*} \leq B;$$

$$B < hr_j < \delta, \qquad j < n^*.$$

The conditional expected value of n^* in this totality, when H_0 is true, may be made arbitrarily large by making B sufficiently small. Hence, when W_1 is sufficiently large, for fixed but arbitrary $h < \delta$, the optimum procedure from the point of minimizing the average risk is to reject H_0 at once without taking any more observations. This, however, contradicts the fact that $h < \delta$, and proves the lemma.

LEMMA 8. *We have, for fixed* $\delta > 1$,

$$\lim_{W_1\to 0} B(W_0(W_1, \delta), W_1) = \delta;$$

$$\lim_{W_1\to\infty} B(W_0(W_1, \delta), W_1) = 0.$$

PROOF: By Lemma 7,

$$\lim_{W_1\to 0} W_0(W_1) = 0.$$

When, for fixed c, both W_0 and W_1 are small enough, then, no matter what the value of h, $\gamma(h) < 0$. Hence $A = B$, which proves the first half of the lemma.

Let now $\{W_{1N}\}$ be a sequence such that $\lim W_{1N} = \infty$. Let $\delta > 1$. For the sake of brevity we write $B(W_{1N})$ instead of

$$B(W_0(W_{1N}\delta), W_{1N}).$$

Suppose that, for sufficiently large N, $B(W_{1N})$ is bounded below by a positive number. Hence, for sufficiently large N, the probability of rejecting H_1 when it it is true is bounded below by a positive number. Moreover, since $B \leq \frac{W_0 g_0}{W_1 g_1} \leq A$, it follows that, for N sufficiently large, $\frac{W_{0N}g_0}{W_{1N}g_1}$ is bounded above and below by positive constants. Thus, for large N the average risk of the test defined by $B(W_{1N})$, δ, is greater than ug_1W_{1N}, where u is a positive constant which does not depend on N. Moreover, from the definition of $B(W_{1N})$, this risk is a minimum.

Let ϵ be a positive number such that $\epsilon\left(\frac{W_{0N}g_0}{W_{1N}g_1} + 1\right) < \frac{u}{2}$ for all N sufficiently large. Let V_1, V_2, with $0 < V_1 < 1 < V_2$, be two constants such that, for the sequential probability ratio test determined by them, both α_0 and α_1 are $< \epsilon$. Of course E_0n and E_1n are finite and determined by the test. For this test the average risk is less than

$$\begin{aligned}
&\epsilon(g_0 W_{0N} + g_1 W_{1N}) + cg_0 E_0 n + cg_1 E_1 n \\
&< \frac{u}{2} g_1 W_{1N} + cg_0 E_0 n + cg_1 E_1 n \\
&< \frac{3u}{4} g_1 W_{1N},
\end{aligned}$$

for W_{1N} large enough. This however contradicts the fact that the minimum risk is $> ug_1W_{1N}$, and proves the lemma.

7. Proof of the theorem. Let a given sequential probability ratio test S_0 be defined by B^*, A^*; $B^* < 1 < A^*$. Let $\alpha_i(S_0)$ be the probability, according to S_0, of rejecting H_i when it is true. Let c be fixed.

By Lemma 4, B is a continuous function of W_0 and W_1. Let $\delta = A^*$ in Lemma 8. Then there exists a pair $\overline{W}_0$, $\overline{W}_1$, with $\overline{W}_0 = W_0(\overline{W}_1, A^*)$, such that

$$A(\overline{W}_0, \overline{W}_1) = A^*;$$

$$B(\overline{W}_0, \overline{W}_1) = B^*.$$

Hence the average risk

$$\sum_i g_i [\overline{W}_i \alpha_i(S_0) + cE_i^0(n)],$$

corresponding to the sequential test S_0 is a minimum.

Now let S_1 be any other test for deciding between H_0 and H_1 and such that

$$\alpha_i(S_1) \leq \alpha_i(S_0), \quad \text{and} \quad E_i^1(n) \text{ exists } (i = 1, 2).$$

Then

$$\sum_i g_i\,[\overline{W}_i\,\alpha_i(S_0) + cE_i^0(n)] \leq \sum_i g_i\,[\overline{W}_i\,\alpha_i(S_1) + cE_i^1(n)].$$

Since $\alpha_i(S_1) \leq \alpha_i(S_0)$, we have

$$\sum_i g_i\,E_i^0(n) \leq \sum_i g_i\,E_i^1(n).$$

Now g_0, g_1 were arbitrarily chosen (subject, of course, to the obvious restrictions). Hence it must be that

$$E_i^0(n) \leq E_i^1(n).$$

This, however, is the desired result.

REFERENCES

1] A. Wald, "Sequential tests of statistical hypotheses", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 117–186.

[2] A. Wald, *Sequential Analysis*, John Wiley and Sons, Inc., New York, 1947.

[3] Charles Stein, "A note on cumulative sums", *Annals of Math. Stat.*, Vol. 17 (1946), pp. 498–499.