LOWER BOUNDS FOR THE EXPECTED SAMPLE SIZE AND THE AVERAGE RISK OF A SEQUENTIAL PROCEDURE

by

Wassily Hoeffding

University of North Carolina

This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-261. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Institute of Statistics
Mimeograph Series No. 227
June, 1959

LOWER BOUNDS FOR THE EXPECTED SAMPLE SIZE AND THE AVERAGE RISK OF A SEQUENTIAL PROCEDURE[1]

By Wassily Hoeffding
University of North Carolina

Summary. Sections 1-6 are concerned with lower bounds for the expected sample size $E_o(N)$ of an arbitrary sequential test whose error probabilities at two parameter points θ_1 and θ_2 do not exceed given numbers α_1 and α_2, where $E_o(N)$ is evaluated at a third parameter point θ_o. The bounds in (1.3) and (1.4) are shown to be attainable or nearly attainable in certain cases where θ_o lies between θ_1 and θ_2. In section 7 lower bounds for the average risk of a general sequential procedure are obtained.

1. Introduction and main results. Let $X_1, X_2, \ldots$ be a sequence of independent random variables having a common probability density f with respect to a σ-finite measure μ. One of two decisions, d_1 and d_2, is to be made. Let f_1 and f_2 be two probability densities such that decision d_2 (d_1) is considered as wrong if $f = f_1$ (f_2). We shall consider sequential tests (decision rules) for making decision d_1 or d_2, such that the probability of a wrong decision does not exceed a positive number α_i when $f = f_i$ ($i=1,2$). Let N denote the (random) number of observations required by such a test. This paper is mainly concerned with lower

1. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-261. Reproduction in whole or in part is permitted for any purpose of the United States Government.
Part of this work was done while the author was a visiting professor at Stanford University.

bounds for $E_0(N)$, the expected sample size when $f = f_0$, where f_0 is in general different from f_1 and f_2.

The background of this problem is as follows. Suppose that f depends on a real parameter θ and f_i corresponds to the value θ_i, where $\theta_1 < \theta_2$. Suppose further that decision d_1 or d_2 is preferred according as $\theta \leq \theta_1$ or $\theta \geq \theta_2$, and that neither decision is strongly preferred if $\theta_1 < \theta < \theta_2$. If we require that the probability of a wrong decision does not exceed α_1 (α_2) if $\theta \leq \theta_1$ ($\theta \geq \theta_2$), the condition of the preceding paragraph will be satisfied. (In many important cases a test which satisfies the latter condition also satisfies the former.) It is known [14] that Wald's sequential probability ratio (SPR) test for testing θ_1 against θ_2, with error probabilities equal to α_1 and α_2, minimizes the expected sample size at these two parameter values. In typical cases its expected sample size is largest when θ is between θ_1 and θ_2 (that is, when neither decision is strongly preferred), and in general there exist tests whose expected sample size at these intermediate θ values is smaller than that of the SPR test. (A special case in which a SPR test minimizes the maximum expected sample size will be discussed in section 4.) In principle it is possible to construct a test which minimizes the expected sample size at an arbitrary θ value or minimizes the maximum expected sample size. Kiefer and Weiss [7] have proved important qualitative properties of such tests. However, the actual construction of a test having this property, as well as the evaluation of its expected sample size and its error probabilities, meets with difficulties which have not been overcome so far (except for a few special cases). Therefore attempts have been made to find a test which, without actually minimizing the maximum expected sample size, comes close to this goal, or at least substantially improves upon the performance of known tests. I mention in particular the work of Donnelly [5] and Anderson [1] who, independently of each other, considered a class of tests such that, if θ is the mean of a normal distribution, the boundaries for the cumulative sums are not parallel lines, as in the SPR test, but converging straight lines. (Anderson also considered truncated tests of this type.) The performance of these and other tests can, to some extent, be judged by comparing, at any parameter point θ, the expected sample size of the test with the smallest expected sample size attainable by any test having the same error probabilities at θ_1 and θ_2. In the ignorance of the minimum expected sample size, the comparison may be made with a lower bound for this minimum. If the discrepancy is small, both the test (as judged by this criterion) and the bound cannot be greatly improved. Our main concern will be with bounds which are best when θ is between θ_1 and θ_2.

We admit arbitrary (in general, randomized) sequential tests which terminate with probability one under each of f_o, f_1 and f_2. We also assume with no loss of generality that $E_o(N) < \infty$. To exclude trivialities we suppose that $\alpha_1 + \alpha_2 < 1$.

The first lower bound for the expected sample size was given by Wald ([10], p. 156) who proved for the case $f_o = f_1$ that

$$(1.1)\quad E_1(N) \geq \frac{\alpha_1 \log \frac{\alpha_1}{1-\alpha_2} + (1-\alpha_1) \log \frac{1-\alpha_1}{\alpha_2}}{\int f_1 [\log (f_1/f_2)] d\mu}$$

and an analogous inequality for $f_o = f_2$. (Wald's proof assumes a non-randomized test, but this restriction is easy to remove.) Both the numerator and the denominator in (1.1) are positive (since $\log x > 1-x^{-1}$ for $x > 0$ unless $x=1$); the integral in the denominator can be equal to $+\infty$, in which case the lower bound has the trivial value 0. The sign of equality in (1.1) can be attained with a SPR test in the case where the ratio f_1/f_2 takes on the two values C and $1/C$ only, provided that the values α_1 and α_2 can be achieved as error probabilities in this test. In certain other cases the sign of equality can be nearly attained with a SPR test.

An extension of (1.1) to the case of an arbitrary f_o has been given by the author [6]:

$$(1.2)\quad E_o(N) \geq \sup_{o<c<1} \frac{-\log [\alpha_1{}^c (1-\alpha_2)^{1-c} + (1-\alpha_1)^c \alpha_2{}^{1-c}]}{c \int f_o \, (\log \frac{f_o}{f_1}) d\mu + (1-c) \int f_o \, (\log \frac{f_o}{f_2}) d\mu}$$

For $f_o = f_1$ and $c \to 1$, (1.2) reduces to (1.1). This bound is likely to be close when f_o is close to f_1 or f_2.

In this paper two new inequalities will be proved,

$$(1.3)\quad E_o(N) \geq \frac{1 - \alpha_1 - \alpha_2}{1 - \int \min (f_o, f_1, f_2) \, d\mu}$$

and

$$(1.4)\quad E_o(N) \geq \frac{\{[(\tau/4)^2 - \zeta \log(\alpha_1+\alpha_2)]^{1/2} - \tau/4\}^2}{\zeta^2},$$

where

$$(1.5) \qquad \zeta = \max(\zeta_1, \zeta_2), \quad \zeta_i = \int f_o (\log \frac{f_o}{f_i}) d\mu, \ i=1,2,$$

and

$$(1.6) \qquad \tau^2 = \int (\log \frac{f_2}{f_1} - \zeta_1 + \zeta_2)^2 f_o \, d\mu \, .$$

Note that $\zeta_i \geq 0$, and $\zeta_i > 0$ is f_o and f_i are densities of different distributions.

In the proof of (1.4) it will be assumed that, in addition to the existence of the integrals in (1.5) and (1.6),

$$(1.7) \qquad f_o(x) = 0 \text{ implies } \min [f_1(x), f_2(x)] = 0,$$

except perhaps on a set of probability zero under f_o, and that the equation

$$(1.8) \qquad E_o (\sum_{j=1}^{N} Y_j)^2 = \tau^2 E_o(N),$$

is satisfied, where

$$(1.9) \qquad Y_j = \log \frac{f_2(X_j)}{f_1(X_j)} - \zeta_1 + \zeta_2 .$$

Concerning the last assumption we note that $\zeta_1 - \zeta_2 = \int f_o [\log (f_2/f_1)] d\mu$, so that $E_o(Y_j) = 0$ and, by (1.6), $E(Y_j^2) = \tau^2$. Equation (1.8) has been proved by Wald [9] and Wolfowitz [15] under certain conditions; see also Seitz and Winkelbauer [8]. It certainly holds if N is bounded or if $Y_1 + \ldots + Y_m$ is bounded for $m < N$. It is clear that if condition (1.8) is satisfied for a test which minimizes

$E_o(N)$, then inequality (1.4) is true also for any other test. In particular this is true under the assumptions of Theorem 4 of Kiefer and Weiss [7], which imply that if a test which minimizes $E_o(N)$, then N is bounded.

Inequalities (1.3) and (1.4) will be proved and discussed in the following sections. Here we mention only the conditions for the attainment or near-attainment of equlaity. In inequality (1.3) the sign of equality holds under certain conditions which are typified by the following two cases. In the first case the densities f_i are arbitrary except that $f_o(x) \geq \min [f_1(x), f_2(x)]$, but α_1 and α_2 are restricted to values which are attainable with a test which requires at most one observation, x_1, and decision $d_1(d_2)$ is made if $f_1(x_1) - f_2(x_1) > 0$ (<0). In the second case the f_i are rectangular densities on intervals of common length such that the mean of f_o is between the means of f_1 and f_2. Then equality in (1.3) is attained with a version of the SPR test for arbitrary values α_1 and α_2.

In (1.4) strict equality is not attainable except in trivial cases. If f_o, f_1 and f_2 are normal probability densities with variance 1 and respective means $\theta = 0$, $-\delta$ and δ, then for $\alpha_1 = \alpha_2 = \alpha < \frac{1}{2}$ equality in (1.4) is nearly attained with a fixed sample size test if α is very small and with a SPR test if α is sufficiently large. For $\alpha = 0.05$ and $\alpha = 0.01$, the expected sample size at $\theta=0$ of a test considered by Anderson [1] comes remarkably close to the lower bound in (1.4).

In section 7 lower bounds for the average risk of a general sequential procedure are derived. Inequality (1.3) can be obtained from one of these bounds.

2. Some lemmas. A randomized sequential test for deciding between d_1 and d_2, based on the sequence X_1, X_2, ..., can be characterized by two sequences of random variables, ψ_o, ψ_1, ψ_2, ... and ϕ_o, ϕ_1, ϕ_2, ... such that $\psi_n \geq 0$, $\psi_o + \psi_1 + \psi_2 + \dots \leq 1$, $0 \leq \phi_n \leq 1$, and both ψ_n and ϕ_n are functions of $X_1, \dots, X_n$ only; ψ_o and ϕ_o are constants. Here ψ_n denotes the probability of $N = n$, under the condition that the values $X_1, \dots, X_n$ have been observed, where N is the number of observations taken before making a terminal decision, and ϕ_n $(1-\phi_n)$ is the probability of making decision d_2 (d_1) under the condition that N=n and the values $X_1, \dots, X_n$ have been observed. A test defined in this way will be denoted by $\{\psi_n, \phi_n\}$. The sequence $\{\psi_n\}$ will be referred to as the stopping rule and $\{\phi_n\}$ as the terminal decision rule of the test. It will be assumed that $N < \infty$, that is, $\psi_o+\psi_1+\dots=1$, with probability one when the common probability density f of the independent random variables X_1, X_2, ... is any one of the functions f_o, f_1, f_2. We note that the probability of making decision d_2 when $f=f_i$ equals

$$E_i(\phi_N) = E_i \left(\sum_{n \geq 0} \psi_n \phi_n\right).$$

The probability density $\prod_{j=1}^{n} f_i(x_j)$ with respect to the product measure μ^n $(n \geq 1)$ will be written $f_{i,n}$ for short. It will be convenient to define

$$\frac{f_{i,n}}{f_{j,n}} = 1 \quad \text{if } n = 0,$$

in accordance with the convention that an "empty" product is equal to 1.

Similarly, the empty sum $\Sigma_{j=1}^{n}$ with n=0, is defined to be 0. The notation $\{\phi_n^*\}$ will serve to denote any terminal decision rule such that for n = 1,2,...

$$(2.1) \qquad \phi_n^* = \begin{cases} 1 & \text{if} \quad f_{1n} < f_{2n} \\ 0 & \text{if} \quad f_{1n} > f_{2n} \end{cases} .$$

The following lemmas will be needed. Lemmas 1, 3, 4, 5 and 6 will be used in the proof of inequality (1.3), Lemmas 1, 2 and 4 in the proof of (1.4). Most of the lemmas are known. The simple proofs of all but Lemma 4 are included for convenience.

Lemma 1. If $\{\psi_n, \phi_n\}$ is an arbitrary sequential test,

$$(2.2) \qquad E_1\ (\phi_N) + E_2\ (1-\phi_N) \geq E_1\ (\phi_N^*) + E_2\ (1-\phi_N^*).$$

Lemma 2. If $f_o(x) = 0$ implies $\min\ [f_1(k),\ f_2(x)] = 0$ except perhaps on a set of probability zero under f_o, then for any stopping rule $\{\psi_n\}$,

$$(2.3) \qquad E_1(\phi_N^*) + E_2\ (1-\phi_N^*) = E_o\ [\min\ (f_{1N}, f_{2N})/f_{oN}].$$

Lemma 3. For any stopping rule $\{\psi_n\}$,

$$(2.4) \qquad E_1(\phi_N^*) + E_2\ (1-\phi_N^*) \geq E_o\ [\min\ (f_{oN}, f_{1N}, f_{2N})/f_{oN}] ,$$

where the sign of equality holds if

$$(2.5) \qquad f_o(x) \geq \min\ [f_1(x),\ f_2(x)]$$

except perhaps on a set of probability zero under f_o.

To prove these three lemmas we note that for any test $\{\psi_n, \phi_n\}$

$$E_1(\phi_N) + E_2(1-\phi_N) = \psi_o + \sum_{n\geq 1} \int \psi_n [\phi_n f_{1,n} + (1-\phi_n)f_{2n}] d\mu^n$$

(2.6)

$$\geq \psi_o + \sum_{n\geq 1} \int \psi_n \min (f_{1,n}, f_{2,n}) d\mu^n$$

with equality for $\phi_n = \phi_n^*$, $n=1,2,\ldots$ This proves Lemma 1.

If the condition of Lemma 2 is satisfied, we can write $\min (f_{1,n}, f_{2,n}) = [\min (f_{1,n}, f_{2,n})/f_{o,n}] f_{o,n}$ in the integrand in (2.6), which implies Lemma 2.

Finally, using (2.6),

$$E_1(\phi_N^*) + E_2 (1-\phi_N^*) \geq \psi_o + \sum_{n\geq 1} \int \psi_n \min (f_{o,n}, f_{1,n}, f_{2,n}) d\mu^n .$$

Upon dividing and multiplying by $f_{o,n}$ in the integrand we obtain inequality (2.4). The condition for equality in Lemma 3 is easy to verify.

<u>Lemma 4</u>. If $E_o(N) < \infty$ and $t(x)$ is a real-valued function such that $E_o [t(x_1)]$ exists, then

$$\text{(2.7)} \qquad E_o \left[\sum_{j=1}^{N} t(X_j)\right] = E_o [t(X_1)] E_o(N).$$

Equation (2.7) is originally due to Wald [11] and has been proved under the present assumptions, except for the trivial extension to randomized tests, by Blackwell [3]; see also Wolfowitz [15].

Lemma 5. If $a_j \geq 0$, $b_j \geq 0$, $c_j \geq 0$, $j=1, \ldots, n$, then

$$\min \left(\prod_{j=1}^{n} a_j, \prod_{i=1}^{n} b_j, \prod_{i=1}^{n} c_j \right) \geq \prod_{j=1}^{n} \min (a_i, b_i, c_j). \tag{2.8}$$

The proof is by induction.

Lemma 6. If $0 \leq d_j \leq 1$, $j = 1, \ldots, n$, then

$$\sum_{j=1}^{n} (1-d_j) \geq 1 - \prod_{j=1}^{n} d_j. \tag{2.9}$$

The sign of equality is attained if and only if all but at most one of $d_1, \ldots, d_n$ are equal to 1.

Lemma 6, with the condition for equality, follows from the identity

$$\sum_{j=1}^{n} (1-d_j) - 1 + \prod_{j=1}^{n} d_j = \sum_{m=2}^{n} (1-d_m) \left(1-\prod_{j=1}^{m-1} d_i\right).$$

Lemma 7. If U is a random variable, $E(e^U) \geq e^{E(U)}$ whenever the expectations exist. The sign of equality holds if and only if U is equal to a constant with probability one.

Proof. Let $V = U - E(U)$. By Taylor's formula, $e^V \geq 1+V$, with equality only if $V=0$. Hence $E(e^V) \geq 1$, and the lemma follows.

3. Proof of inequality (1.3). By Lemma 1, for any test which satisfies $E_1(\phi_N) \leq \alpha_1$ and $E_2(1-\phi_N) \leq \alpha_2$,

$$\alpha_1 + \alpha_2 \geq E_1(\phi_N^*) + E_2(1-\phi_N^*). \tag{3.1}$$

By Lemma 3,

$$E_1(\phi_N^*) + E_2(1-\phi_N^*) \geq E_o [\min (f_{oN}, f_{1N}, f_{2N}) / f_{oN}]. \tag{3.2}$$

By Lemma 5,

$$(3.3)\qquad \min(f_{o,n}, f_{1,n}, f_{2,n}) \geq \prod_{j=1}^{n} \min[f_o(x_i), f_1(x_i), f_2(x_i)].$$

Hence if we write

$$r(x) = \min[f_o(x), f_1(x), f_2(x)]/f_o(x),$$

we have

$$(3.4)\qquad E_o[\min(f_{o,N}, f_{1,N}, f_{2,N})/f_{oN}] \geq E_o[\prod_{j=1}^{N} r(X_j)].$$

Note that $0 \leq r(x) \leq 1$. If we apply Lemma 6 and then Lemma 4, we obtain

$$(3.5)\qquad \begin{aligned} E_o[\prod_{1=1}^{N} r(X_j)] &\geq E_o\{1-\sum_{i=1}^{N}[1-r(X_j)]\} \\ &= 1 - E_o(N)\, E_o[1-r(X_1)] \\ &= 1 - E_o(N)[1 - \int \min(f_o, f_1, f_2)d\mu] \end{aligned}$$

Inequality (1.3) now follows from (3.1), (3.2), (3.4) and (3.5).

4. Discussion of inequality (1.3). The sign of equality in (1.3) holds if and only if it holds in each of the four inequalities used in the proof. Equality in (3.1) is attained if $\phi_n = \phi_n^*$ and

$$(4.1)\qquad E_1(\phi_N^*) = \alpha_1,\quad E_2(1-\phi_N^*) = \alpha_2.$$

In (3.2) it is attained, by Lemma 3, if

$$(4.2)\qquad f_o(x) \geq \min[\bar{f}_1(x), f_2(x)].$$

By Lemma 6, equality in (3.5) holds if and only if, for each $n \geq 1$, $N = n$ implies that all but at most one of $r(X_1), \ldots, r(X_n)$ are equal to 1, with probability one under f_o. This is the case if

$$(4.3) \qquad f_o(X_j) = f_1(X_j) = f_2(X_j) \text{ for } 1 \leq j \leq N-1.$$

If this condition is satisfied, equality also holds in (3.4). Thus conditions (4.1) (4.2) and (4.3) are sufficient for the attainment of equality in (1.3).

Condition (4.2) is satisfied for many common one-parameter families of distributions when θ_o is between θ_1 and θ_2. Under this condition, (1.3) reduces to

$$(4.4) \qquad E_o(N) \geq \frac{1-\alpha_1-\alpha_2}{1-\int \min(f_1, f_2) d\mu} = \frac{1-\alpha_1-\alpha_2}{\frac{1}{2}\int |f_1-f_2| \, d\mu}$$

Condition (4.3) is satisfied, and equality holds in (4.4), in the following two cases.

The first case is where the densities are arbitrary, subject only to (4.2), but the values α_1 and α_2 are such that they can be attained as error probabilities with a test which requires at most one observation ($N \leq 1$), and if an observation x is taken, decision d_1 (d_2) is made if $f_1(x)-f_2(x)$ is > 0 (< 0).

The second case in which equality in (4.4) is attained is where, in addition to (4.2), the set $\mathcal{E} = \{x \mid f_o(x)=f_1(x)=f_2(x)\}$ has a positive probability. Let $\mathcal{E}_o \subseteq \mathcal{E}$ and let the complement of $\mathcal{E}_o$ be subdivided into two disjoint sets $\mathcal{E}_1$ and $\mathcal{E}_2$ such that $f_1(x)-f_2(x)>0$ if

$x \epsilon C_1$ and ≤ 0 if $x \epsilon C_2$. Let N be the least n such that $x_n \notin C_o$. Decision d_i is made if $x_N \epsilon C_i$, i=1,2. (Instead, suitable randomized decisions can be made when $x_n \epsilon C$.) Then it can be readily verified that equality holds in (4.4), with $E_o(N) = (1-p_o)^{-1}$, $\alpha_1 = p_{12}(1-p_o)^{-1}$, $\alpha_2 = p_{21}(1-p_o)^{-1}$, where p_o is the probability of C_o (under any f_i) and p_{ij} is the probability of C_j under f_i. In the particular case where μ is linear Lebesgue measure, $f_i(x) = g(x-\Theta_i)$, $g(x) = 1/L$, $-L/2 \leq x \leq L/2$, $g(x) = 0$ otherwise, $0 < \Theta_2 - \Theta_1 < L$, and $\Theta_1 \leq \Theta_o \leq \Theta_2$, we have $C = [\Theta_2 - L/2,\ \Theta_1 + L/2]$. Let $\Theta_2 - L/2 \leq c \leq d \leq \Theta_1 + L/2$, $C_o = (c,d)$, $C_1 = (-\infty, c]$, $C_2 = [d, +\infty)$. Then with the test just described, perhaps preceded by a randomized decision as to whether to take at least one observation, any error probabilities can be attained ($\alpha_1 \geq 0$, $\alpha_2 \geq 0$, $\alpha_1 + \alpha_2 \leq 1$). Moreover, the maximum with respect to all real Θ of the expected sample size of this test when the density is $g(x-\Theta)$ is attained when Θ is between Θ_1 and Θ_2. Hence the test minimizes the maximum expected sample size. It should be noted that the present test is a modified version of the SPR test as defined in Wald [12], p. 120. It differs from the latter only in this respect: If the probability ratio after n observations equals one of the two numbers A and B (in Wald's notation; in our case A=B=1), the stopping decision and the terminal decision may depend on the position of the sample point in the corresponding sets, instead of being randomized decisions.

It is of interest to note that the bound in (1.3) is always positive whereas the bounds in (1.1) and (1.2) take on the trivial value 0 if the integrals in their denominators are equal to $+\infty$.

However, in most of the more common cases the bounds (1.1) and (1.2) (as well as (1.4)) are better than (1.3). For instance, if f_o, f_1, and f_2 are normal distributions with a common variance and respective means 0, $-\delta$ and δ, the bound in (1.3) is of the order δ^{-1}, but those in (1.2) and (1.4) are proportional to δ^{-2} and hence better than (1.3) if δ is small.

There is an interesting similarity between Wald's inequality (1.1) and inequality (1.3) (or (4.4)) for $f_o=f_1$. If α_1 and α_2 denote the actual error probabilities, both inequalities are of the form

$$E_1(N) \geq \frac{D(f_1^*,f_2^*)}{D(f_1,f_2)}, \tag{4.5}$$

where D is the measure of discrepancy between two distributions which appears in the denominators of (1.1) and (4.4), and f_i^* denotes the distribution on the two points d_1,d_2 of the decision space such that the probability assigned to d_j is the probability of making decision d_j when $f=f_i$; more precisely, f_i^* is the probability density with respect to a measure μ^* such that $\mu^*(d_1)=\mu^*(d_2)=1$ and $1-f_1^*(d_1)=f_1^*(d_2)=\alpha_1$, $1-f_2^*(d_2)=f_2^*(d_1)=\alpha_2$.

It will be seen in section 7 that inequality (1.3) can be deduced from a lower bound for the average risk of a general sequential procedure. However, the direct proof given in section 3 makes it easier to determine the conditions for equality.

5. Proof of inequality (1.4). We assume that the integrals ζ_1, ζ_2 and τ^2 in (1.5) and (1.6) exist and that the conditions (1.7) and (1.8) are satisfied. Let, for i=1,2,

$$Z_{i,n} = \sum_{j=1}^{n} \left(\log \frac{f_o(X_j)}{f_i(X_j)} - \zeta_i\right) \tag{5.1}$$

and let

$$Z_n = Z_{1,n} - Z_{2,n} = \sum_{j=1}^{n} Y_j, \tag{5.2}$$

where Y_j is defined in (1.9). Then

$$\frac{f_{i,n}}{f_{o,n}} = e^{-Z_{i,n}-\zeta_i n} .$$

Hence, by Lemma 2,

$$\begin{aligned} E_1(\phi_N^*) + E_2(1-\phi_N^*) &= E_o\left[\min\left(\frac{f_{1,N}}{f_{o,N}}, \frac{f_{2,N}}{f_{o,N}}\right)\right] \\ &= E_o\left[e^{-\max(Z_{1,N}+\zeta_1 N, Z_{2,n}+\zeta_2 N)}\right] \\ &\geq E_o\left[e^{-\max(Z_{1,N},Z_{2,N})-\zeta N}\right], \end{aligned} \tag{5.3}$$

where $\zeta=\max(\zeta_1,\zeta_2)$. By Lemma 7,

$$E_o\left[e^{-\max(Z_{1,N},Z_{2,N})-\zeta N}\right] \geq e^{-E_o\left[\max(Z_{1,N},Z_{2,N})\right]-\zeta E_o(N)} \tag{5.4}$$

Since $2\max(Z_{1,N},Z_{2,N}) = Z_{1,N} + Z_{2,N} + |Z_{1,N} - Z_{2,N}|$, $Z_{1,N}-Z_{2,N}=Z_N$, and, by Lemma 4, $E_o(Z_{1,N}) = E_o(Z_{2,N})=0$, we have

$$E_o\left[\max(Z_{1,N},Z_{2,N})\right] = \tfrac{1}{2}E_o(|Z_N|). \tag{5.5}$$

Also

$$E_0(|Z_N|) \leq [E_0(Z_N^2)]^{1/2} = \tau[E_0(N)]^{1/2}, \tag{5.6}$$

where we have used equation (1.8).

Thus if (ψ_n, ϕ_n) is any test such that $E_1(\phi_N) \leq \alpha_1$, $E_2(1-\phi_N) \leq \alpha_2$, and equation (1.8) is satisfied, it follows from Lemma 1 and the relations (5.3), (5.4), (5.5) and (5.6) that

$$\log(\alpha_1+\alpha_2) \geq -\frac{\tau}{2}[E_0(N)]^{1/2} - \zeta E_0(N).$$

Solving this inequality for $E_0(N)$, we obtain (1.4).

6. Discussion of inequality (1.4). Inequality (1.4) has been obtained by combining the four inequalities (3.1), (5.3), (5.4) and (5.6). Equality in (3.1) is always attainable for suitable α_1 and α_2, and in (5.3) it holds if $\zeta_1=\zeta_2$ $(=\zeta)$. In (5.4) the sign of equality holds if and only if $\max(Z_{1,N}, Z_{2,N}) + \zeta N$ is constant with probability one (see Lemma 7), and in (5.6) it holds if and only if $|Z_N|$, that is $|Z_{1,N} - Z_{2,N}|$, is constant with probability one, both probabilites evaluated under f_0. The last two conditions cannot be satisfied simultaneously except in trivial cases.

To obtain an idea of how close the bound in (1.4) can come to the minimum attainable value of $E_0(N)$, we shall consider the following special case. Let f_i be the normal probability density with variance 1 and mean θ_i, where $\theta_0 = 0$, $\theta_1=-\delta$ and $\theta_2=\delta>0$. Then $\zeta_1=\zeta_2=\delta^2/2$, $\tau=2\delta$, and inquality (1.4) becomes

$$E_0(N) \geq \delta^{-2} \{[1-2 \log (2\alpha)]^{1/2} - 1\}^2 \tag{6.1}$$

where $2\alpha = \alpha_1+\alpha_2$. This bound will be compared with the values of $E_0(N)$ for a fixed sample size test, Wald's SPR test, and a test considered by Anderson, with error probabilities $\alpha_1=\alpha_2=\alpha(>\frac{1}{2})$ in each case.

Let $S_n=X_1+\ldots+X_n$. For a fixed sample size test such that decision d_1 or d_2 is made according as $S_n < 0$ or $S_n > 0$, the error probabilities at $\Theta=-\delta$ and $\Theta=\delta$ are both equal to $\Phi(-\delta n^{1/2})$, where

$$\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-y^2/2} dy.$$

Hence $E_0(N)$ is the least n such that $\Phi(-\delta n^{1/2}) \leq \alpha$. If $\lambda=\lambda(\alpha)$ is defined by $\Phi(-\lambda) = \alpha$, we have

$$E_0(N) = \delta^{-2}\lambda^2, \tag{6.2}$$

exactly or with a good approximation. If $\alpha \longrightarrow 0$, then $\lambda \longrightarrow \infty$ and

$$\alpha = \Phi(-\lambda) = (2\pi)^{-1/2} \lambda^{-1} e^{-\lambda^2/2} (1+0(\lambda^{-2}))$$

Hence

$$\lambda^2 = -2 \log \alpha + 0 [\log (-2 \log \alpha)].$$

The factor of δ^{-2} in inequality (6.1) is

$$\{[1-2 \log (2\alpha)]^{1/2}-1\}^2 = -2 \log \alpha + 0 [(-2 \log \alpha)^{1/2}].$$

Thus if α is small enough, the bound in (6.1) is nearly attained with a fixed sample size test, although the asymptotic approach is extremely slow. It follows that the fixed sample size test nearly minimizes the expected sample size at $\theta = 0$ when α is (very) small.

Now consider the SPR test which stops as soon as $2\delta |S_n| > \log A$ (>0). Then $(\log A)^2 \leq 4\delta^2 E_o(S_N^2) = 4\delta^2 E_o(N)$ by (1.8), and $A \leq \frac{1-\alpha}{\alpha}$. These inequalities are close approximations for α fixed and δ small enough (Wald [10]). With this approximation,

$$E_o(N) = \delta^{-2}\left(\frac{1}{2} \log \frac{1-\alpha}{\alpha}\right)^2. \tag{6.3}$$

Put $\alpha = (1-\epsilon)/2$ and let $\epsilon \longrightarrow 0$. Then

$$\left(\frac{1}{2} \log \frac{1-\alpha}{\alpha}\right)^2 = \epsilon^2 + \frac{2}{3}\epsilon^4 + \frac{23}{45}\epsilon^6 + \dots$$

and

$$\{[1-2 \log (2\alpha)]^{1/2}-1\}^2 = \epsilon^2 + \frac{2}{3}\epsilon^4 - \frac{1}{6}\epsilon^5 + \dots$$

Thus if α is close to its upper bound $\frac{1}{2}$, and δ is small enough, the lower bound in (6.1) is nearly attained with a SPR test. Hence the SPR test nearly minimizes $E_o(N)$ in this case. Table 1 shows that even for $\alpha=0.2$ the expected sample size exceeds the lower bound by only 3%. (The lower bound in (1.2) with $c=\frac{1}{2}$ also approaches $E_o(N)$ for the SPR test as $\alpha \longrightarrow \frac{1}{2}$. However, inequality (6.1) is better than (1.2), as applied to the present case, for all values of α.)

For α values not close to 0 or $\frac{1}{2}$ we compare the bound in (6.1) with the expected sample size of a test considered by Anderson [1]. This test stops as soon as $|S_n| \geq c + dn$, where $d < 0 < c$. Anderson approximated the sequence $\{S_n\}$ by a Wiener process so that his values for the expected stopping time, $E_0(\tau)$, when the mean of the process is 0 are approximations to $E_0(N)$. He chose the constants c and d so as to minimize $E_0(\tau)$ subject to prescribed error probabilities $\alpha_1=\alpha_2=\alpha$ at $\theta = \pm\ \delta$, for δ=0.1 and α=0.01 and 0.05. Anderson's values are given in Table 1. The expected sample sizes exceed the lower bounds only by 3.6% and 2.8%, respectively. This shows that both Anderson's test (as judged by the expected sample size at θ=0) and inequality (6.1) cannot be greatly improved in these cases.

TABLE 1

Values of $E_0(N)$ and of the lower bound in (6.1) for δ = 0.1.

α =	0.0001	0.001	0.01	0.05	0.1	0.2	0.3
Fixed sample size	1383	955	541.2	270.6	164.3	70.8	27.5
SPR test	2121	1193	527.9	216.7	120.7	48.0	17.9
Anderson's test	--	--	402.2	192.2	--	--	--
Lower bound (6.1)	1054	710	388.3	187.0	111.1	46.6	17.8

To conclude this section, it will be shown that for each of the two sequential tests here considered the expected sample size attains its maximum when the mean θ of the normal distribution is 0. In conjunction with the preceding results this implies that each of these tests (as well as the fixed sample size test) comes close to minimizing the maximum expected sample size for certain α values.

Both tests are such that sampling is stopped as soon as $|S_n| \geq c_n$, where c_1, c_2, ... are nonnegative constants. The expected value of N at θ is the sum of the probabilities $P[N>n|\theta]$. We can write

$$P[N > n \mid \theta] = \int_A f(y-\theta z)dy,$$

where $y = (y_1, \ldots, y_n)$, $z=(1,1, \ldots, 1)$, f is the probability density of n independent normal random variables with mean 0 and variance 1, and $A = \{y \mid |y_1+\ldots+y_m| < c_m, m=1,\ldots,n\}$. The set A is convex, and $y\epsilon A$ implies $-y\epsilon A$. It follows from a theorem of Anderson [2] that $P[N > n \mid \theta]$ attains its maximum at $\theta=0$ (and is monotone for $\theta < 0$ and $\theta > 0$). Thus the same conclusion is true for the expected value of N.

7. <u>Lower bounds for the average risk</u> In this section a sequence of increasingly better lower bounds for the average risk of a general sequential procedure will be derived. Under certain conditions these bounds converge to the minimum average risk. They are similar to the bounds given by Blackwell and Girshick [4] and will be obtained as a consequence of results of Wald and Wolfowitz [13] which are also contained in Wald's book [12]. In slight extension of the assumptions in [13] and [4], the cost per observation will be allowed to depend on the parameter; due to this assumption the bounds can be used to obtain lower bounds for the expected sample size (see section 8).

The random variables X_1, X_2, ... are assumed to be independent with a common probability density f_θ with respect to a σ-finite measure μ, where the parameter θ is contained in a space Θ. To simplify the exposition, the assumptions of [13], section 2, will be made (with some obvious changes in notation), with two exceptions stated below. In particular, μ is Lebesgue or counting measure on the real Borel sets (this is not essential); the loss function W on $\Theta \times D$ is nonnegative and bounded; the terminal decision space D is compact in the sense of the convergence $\sup_\theta |W(\theta,d_i)-W(\theta,d_o)| \longrightarrow 0$; the a priori distributions ξ are the probability measures on a fixed Borel field of subsets of Θ. The cost of m observations is assumed to be $c(\theta)m$, where $c(\theta)$ is nonnegative, bounded and measurable on the given Borel field of subsets of Θ. (In [13], $c(\theta)$ is a constant.) In addition, we assume that the function $\inf_{\theta \in \Theta} f_\theta$ is Borel measurable. The class Δ consists of all sequential decision functions δ which satisfy the needed measurability conditions as specified in [13]. For the other measurability assumptions we also refer to [13].

Denote by $r(\theta,\delta)$ the risk (expected loss plus expected cost) when the decision function δ is used and the parameter is θ. For any a priori distribution ξ over Ω let $r(\xi,\delta) = \int r(\theta,\delta)d\xi$. Let $\rho(\xi)$ denote the infimum of the average risk $r(\xi,\delta)$ for $\delta\epsilon\Delta$. Let

$$c(\xi) = \int c(\theta)d\xi, \quad \rho_o(\xi) = \inf_{d\epsilon D}\int W(\theta,d)d\xi, \quad f_\xi(y) = \int f_\theta(y)d\xi,$$

and let ξ_y denote the distribution over Ω defined by $d\xi_y = f_\theta(y)d\xi/f_\xi(y)$. Then the function $\rho(\xi)$ satisfies the equation

$$(7.1) \quad \rho(\xi) = \min\left[\rho_o(\xi), \int \rho(\xi_y)f_\xi(y)d\mu(y) + c(\xi)\right] .$$

This is a straightforward extension of Theorem 3.2 of [13] .

For $n \geq 0$ let $\rho_n(\xi)$ denote the infimum of $r(\xi,\delta)$ for $\delta\epsilon\Delta_n$, the class of all decision functions in Δ which terminate after at most n observations. (This is consistent with the definition of $\rho_o(\xi)$ above.) By direct extension of Theorem 3.1 of [13] we have

$$(7.2) \quad \rho_n(\xi) = \min\left[\rho_o(\xi), \int\rho_{n-1}(\xi_y)f_\xi(y)d\mu(y) + c(\xi)\right], \qquad n = 1, 2, \ldots$$

Clearly $\rho_o(\xi) \geq \rho_1(\xi) \geq \rho_2(\xi) \geq \cdots \geq \rho(\xi)$. In [13] it is shown that if $c(\theta) = c > 0$, then $\lim \rho_n(\xi) = \rho(\xi)$.

Blackwell and Girshick ([4], pp. 255-256) have given lower bounds for $\rho(\xi)$ which with the present cost function can be defined as follows. Let $r_o^*(\xi) = 0$ and define recursively for $n = 1, 2, \ldots$

$$(7.3) \qquad r_n^*(\xi) = \min\left[\rho_o(\xi), \int r_{n-1}^*(\xi_y)f_\xi(y)d\mu(y) + c(\xi)\right] .$$

Then $r_o^*(\xi) \leq r_1^*(\xi) \leq r_2^*(\xi) \leq \cdots \leq \rho(\xi)$, and if $c(\theta) = c > 0$, then $\lim r_n^*(\xi) = \rho(\xi)$ [4].

It will now be shown that the lower bounds (7.3) can be improved with the help of an inequality of Wald and Wolfowitz [13]. Sufficient conditions for the convergence of these lower bounds and of the upper bounds $\rho_n(\xi)$ to $\rho(\xi)$ when $c(\Theta)$ is not constant will also be given.

Let

$$\lambda = 1 - \int \inf_{\Theta \in \Omega} f_\Theta(y) d\mu(y) \quad . \tag{7.4}$$

Excluding the trivial case where all distributions f_Θ are identical, we have $0 < \lambda \leq 1$. Now define

$$\rho_o'(\xi) = \min [\rho_o(\xi), \lambda^{-1} c(\xi)] \tag{7.5}$$

and recursively for $n = 1, 2, \ldots$

$$\rho_n'(\xi) = \min [\rho_o(\xi), \int \rho_{n-1}'(\xi_y) f_\xi(y) d\mu(y) + c(\xi)] \quad . \tag{7.6}$$

We shall write $f_{\Theta,n}$ for $\prod_{j=1}^n f_\Theta(x_j)$, $f_{\xi,n}$ for $\int f_{\Theta,n} d\xi$ and $\xi^{(n)}$ for the a posteriori distribution over Ω after n observations $x_1, \ldots, x_n$, so that $d\xi^{(n)} = f_{\Theta,n} d\xi / f_{\xi,n}$.

<u>Theorem 1</u>. We have

$$\rho_o'(\xi) \leq \rho_1'(\xi) \leq \rho_2'(\xi) \leq \cdots \leq \rho(\xi) \quad . \tag{7.7}$$

In order that

$$\lim_{n \to \infty} \rho_n'(\xi) = \lim_{n \to \infty} \rho_n(\xi) = \rho(\xi) \quad , \tag{7.8}$$

it is sufficient that either

$$\lim_{n \to \infty} \int \rho_o(\xi^{(n)}) f_{\xi,n} d\mu^{(n)} = 0 \tag{7.9}$$

or

$$(7.10) \qquad \xi\left\{ c(\Theta) > 0 \right\} = 1.$$

<u>Remark 1.</u> If $\lambda = 1$, then $\rho'_{n-1}(\xi) = r^*_n(\xi)$, so that the two sequences of bounds are equivalent. We always have $\rho'_{n-1}(\xi) \geq r^*_n(\xi)$.

<u>Remark 2.</u> The integral in (7.9) is the risk of the (fixed sample size) Bayes procedure based on n observations when $c(\Theta) \equiv 0$. Thus condition (7.9) is satisfied for all ξ if the maximum expected loss of some decision rule based on n observations tends to 0 as $n \longrightarrow \infty$. An upper bound for the ingegral in (7.9) (which, in turn, is an upper bound for $\rho_n(\xi) - \rho'_n(\xi)$) for the case of finite Γ is given in Theorem 2 below.

<u>Remark 3.</u> In section 8 it will be shown that the inequality $\rho(\xi) \geq \rho'_o(\xi)$ implies inequality (1.3). The discussion in section 4 shows that equality in $\rho(\xi) \geq \rho'_o(\xi)$ is attained in special cases.

<u>Proof of Theorem 1.</u> Since $\rho(\xi) = \inf_\delta \int r(\Theta,\delta)d\xi$ and

$$\int \rho(\xi_y) f_\xi(y) d\mu(y) = \int \inf_\delta \left[\int r(\Theta,\delta) f_\Theta(y) d\xi(\Theta) \right] d\mu(y) \qquad ,$$

we have

$$(7.11) \qquad \int \rho(\xi_y) f_\xi(y) d\mu(y) \geq \rho(\xi) \int \inf_\Theta f_\Theta(y) d\mu(y) = (1-\lambda)\rho(\xi) \qquad .$$

(This is essentially equivalent to inequality (3.22) of [13].) Hence, by (7.1), if $\rho(\xi) < \rho_o(\xi)$, then $\rho(\xi) \geq \lambda^{-1}c(\xi)$. Therefore $\rho(\xi) \geq \rho'_o(\xi)$. It now follows from (7.1) and (7.6) by induction that $\rho(\xi) \geq \rho'_n(\xi)$ for all $n \geq 0$.

To complete the proof of (7.7) we now show that

$$(7.12) \qquad \rho'_n(\xi) \geq \rho'_{n-1}(\xi) \qquad , \qquad n = 1, 2, \ldots$$

It can be seen in a similar way as in the proof of (7.11) that

$$\int \rho_o'(\xi_y) f_\xi(y) d\mu(y) \geq (1 - \lambda)\rho_o'(\xi) \quad .$$

Hence, by (7.6) with n = 1,

$$\rho_1'(\xi) \geq \min [\rho_o(\xi), \quad (1 - \lambda)\rho_o'(\xi) + c(\xi)] \quad .$$

It is readily shown that the right side of this inequality is equal to $\rho_o'(\xi)$. Thus (7.12) is proved for n = 1. For n = 2, 3, ... the result follows by induction from (7.6).

To prove the remaining part of the theorem, we first observe that $\rho_n'(\xi)$ (just as $r_n^*(\xi)$; see [4]) can be interpreted as the minimum average risk in a modified decision problem. Let D' denote the original terminal decision space D, augmented by a terminal decision $d_o \notin D$. Let the loss function be $W(\Theta,d)$ if $d \neq d_o$, but $\lambda^{-1}c(\Theta)$ if $d = d_o$. The cost function is that of the original problem. Let Δ_n' denote the class of all sequential decision functions (subject to measurability assumptions analogous to those in [13]) which terminate after at most n (≥ 0) observations, such that decision d_o is allowed only after the n-th observation has been taken. If $r'(\Theta,\delta)$ denotes the risk function in the modified problem, it can be seen that the minimum of $r'(\xi,\delta)$ for δ in Δ_n' is equal to $\rho_n'(\xi)$ as defined by (7.5) and (7.6).

Since $\rho_n'(\xi) \leq \rho(\xi) \leq \rho_n(\xi)$, (7.8) will be proved if we show that

$$(7.13) \qquad \lim_{n \to \infty} [\rho_n(\xi) - \rho_n'(\xi)] = 0 \quad .$$

For a fixed a priori distribution ξ, let δ_n' be a Bayes decision function in Δ_n', so that $\rho_n'(\xi) = r'(\xi,\delta_n')$. Let δ_n be the decision function in Δ_n which is identical with δ_n' before the n-th observation is taken and makes the optimal terminal decision after the n-th observation. Denote by $\psi_n' = \psi_n'(x_1, \ldots, x_{n-1})$

the probability that the sample size N' required by procedure δ_n' is equal to n, given that the first n-1 observations are $x_1, \ldots, x_{n-1}$. Then

$$\rho_n(\xi) - \rho_n'(\xi) \le r(\xi,\delta_n) - r'(\xi,\delta_n') = \int \psi_n' [\rho_o(\xi^{(n)}) - \rho_o'(\xi^{(n)})] f_{\xi,n} d\mu^n$$

Therefore

$$(7.14) \qquad \rho_n(\xi) - \rho_n'(\xi) \le \int \psi_n' \rho_o(\xi^{(n)}) f_{\xi,n} d\mu^n \quad .$$

It follows immediately that condition (7.9) is sufficient for (7.13) and hence for (7.8). Also, if $\overline{W}$ is an upper bound for $W(\Theta,d)$ and hence for $\rho_o(\xi)$, (7.14) implies

$$(7.15) \qquad \rho_n(\xi) - \rho_n'(\xi) \le \overline{W} \int \psi_n' f_{\xi,n} d\mu^n = \overline{W} P_\xi(N' = n) \quad .$$

Now

$$\begin{aligned} \overline{W} \ge \rho_n'(\xi) &= r'(\xi,\delta_n') \ge \int c(\Theta) E_\Theta(N') d\xi \\ &\ge \int c(\Theta) n P_\Theta(N'=n) d\xi \ge n^{1/2} \int_{\{c(\Theta) \ge n^{-1/2}\}} P_\Theta(N' = n) d\xi \\ &= n^{1/2} [P_\xi(N'=n) - \int_{\{c(\Theta) < n^{-1/2}\}} P_\Theta(N'=n) d\xi] \\ &\ge n^{1/2} [P_\xi(N' = n) - \xi \{ c(\Theta) < n^{-1/2} \}] \quad . \end{aligned}$$

Thus

$$(7.16) \qquad P_\xi(N' = n) \le n^{-1/2} \overline{W} + \xi \{ c(\Theta) < n^{-1/2} \} \quad .$$

Letting $n \longrightarrow \infty$, it follows from (7.15) and (7.16) that condition (7.10) is sufficient for (7.8). This completes the proof of the theorem.

The section is concluded by a theorem which shows that if Ω is finite, then under a natural assumption on the loss function the difference $\rho_n(\xi) - \rho_n'(\xi)$ converges to 0 uniformly in ξ at an exponential rate.

Theorem 2. If Ω consist of k points, $\theta = 1, 2, \ldots, k$, say, and if for each $\theta \in \Omega$ there is a $d_\theta \in D$ such that $W(\theta, d_\theta) = 0$, then

$$(7.17) \qquad \rho_n(\xi) - \rho_n'(\xi) \leq \overline{W}(k-1)\gamma^n \quad ,$$

where $\overline{W}$ is an upper bound for $W(\theta, d)$ and

$$\gamma = \max_{i \neq j} \int (f_i f_j)^{1/2} d\mu \quad .$$

We remark that $\gamma < 1$ if it is understood that no two of the functions $f_1, \ldots, f_k$ are densities of the same distribution. The theorem exhibits a particularly simple bound for $\rho_n(\xi) - \rho_n'(\xi)$; closer bounds are contained in the proof.

To prove the theorem we note that by (7.14)

$$(7.18) \qquad \rho_n(\xi) - \rho_n'(\xi) \leq \int \rho_o(\xi^{(n)}) f_{\xi,n} d\mu^n \quad .$$

Let ξ assign probability g_i to the point $\theta = i$. Then

$$\int \rho_o(\xi^{(n)}) f_{\xi,n} d\mu^n = \int \inf_d \sum_{i=1}^{k} g_i W(i,d) f_{i,n} d\mu^n$$

$$\leq \int \min_{j=1,\ldots,k} \sum_{i=1}^{k} g_i W(i,d_j) f_{i,n} d\mu^n$$

$$\leq \int \sum_{j=1}^{k} \phi_j \sum_{i=1}^{k} g_i W(i,d_j) f_{i,n} d\mu^n,$$

where $\phi_1, \ldots, \phi_k$ are arbitrary nonnegative measurable functions of $x_1, \ldots, x_n$ such that $\phi_1 + \ldots + \phi_k = 1$. Hence, recalling that $W(i, d_i) = 0$,

$$\int \rho_o(\xi^{(n)}) f_{\xi,n} d\mu^n \leq \overline{W} \int \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j\neq i}}^{k} \phi_j g_i f_{i,n} d\mu^n$$

$$= \overline{W} \sum_{i=1}^{k} g_i \int (1 - \phi_i) f_{i,n} d\mu^n \quad .$$

Let, in particular, $\phi_i = \prod_{j=1}^{k} \phi_{ij}$, where $\phi_{ii} = 1$ and for $i < j$, $1 - \phi_{ji} = \phi_{ij} = 1$ if $f_{i,n} > f_{j,n}$ and $= 0$ otherwise. The conditions $\phi_i \geq 0$ and $\phi_1 + \ldots + \phi_k = 1$ are satisfied. (Note that $\phi_i = 1$ if $f_{i,n} = \max_j f_{j,n}$.) By Lemma 6 of section 2, $1 - \phi_i \leq \Sigma_{j=1}^{k}(1 - \phi_{ij})$, where the term with $j = i$ is zero. Hence

$$\int (1 - \phi_i) f_{i,n} d\mu^n \leq \sum_{\substack{j=1 \\ j\neq i}}^{k} \int (1 - \phi_{ij}) f_{i,n} d\mu^n \quad .$$

Now if $i \neq j$,

$$\int (1 - \phi_{ij}) f_{i,n} d\mu^n \leq \int \min(f_{i,n}, f_{j,n}) d\mu^n$$

$$\leq \int (f_{i,n}, f_{j,n})^{1/2} d\mu^n = \left[\int (f_i f_j)^{1/2} d\mu\right]^n \leq \gamma^n.$$

Hence $\int \rho_o(\xi^{(n)}) f_{\xi,n} d\mu^n \leq \overline{W}\, \Sigma_{i=1}^{k}\, g_i (k-1)\gamma^n = \overline{W}(k-1)\gamma^n$, and the theorem follows from (7.18).

8. Further lower bounds for the expected sample size.

The lower bounds for $\rho(\xi)$ in section 7 can be used to obtain lower bounds for the expected sample size at a specified parameter point Θ_o in terms of upper bounds on the expected loss or (by choosing a suitable loss function) in terms of upper or lower bounds on the probabilities of various decisions at selected parameter points. For this purpose one chooses the cost function so that $c(\Theta) = 0$ for $\Theta \neq \Theta_o$ and $c(\Theta_o) > 0$. The explicit result will be stated only for a two-decision problem; extensions to problems involving more than two decisions will be obvious.

Let Ω consist of the three points 0, 1, 2, and let there be two decisions, d_1 and d_2. Put $W(1,d_2) = W(2,d_1) = 1$, $W(i,d_j) = 0$ otherwise, $c(0)=1$, $c(1) = c(2) = 0$. Let ξ assign probability g_i to the point i ($i = 0, 1, 2$). Then $\rho_o(\xi) = \min(g_1,g_2)$, $c(\xi) = g_o$, and, with $\delta = \{ \psi_n, \phi_n \}$,

$$r(\xi,\delta) = g_o E_o(N) + g_1 E_1(\phi_N) + g_2 E_2(1 - \phi_N) \quad .$$

For any $n \geq 0$, $r(\xi,\delta) \geq \rho_n'(\xi)$. Hence if $E_1(\phi_N) \leq \alpha_1$ and $E_2(1 - \phi_N) \leq \alpha_2$,

$$(8.1) \qquad E_o(N) \geq \sup_\xi \frac{\rho_n'(\xi) - g_1\alpha_1 - g_2\alpha_2}{g_o} \quad , \quad n = 0, 1, 2, \ldots$$

This gives a sequence of increasingly better lower bounds for $E_o(N)$. In particular, $\rho_o'(\xi) = \min(g_1,g_2,\lambda^{-1}g_0)$, where $\lambda = 1 - \int \min(f_o,f_1,f_2)d\mu$. The ratio in (8.1) with $n = 0$ is maximized by letting $g_1 = g_2 = \lambda^{-1}g_o$, and the resulting inequality is equivalent to (1.3).

References

[1] T. W. Anderson, "A modification of sequential analysis to reduce the sample size," Ann. Math. Stat.

[2] T. W. Anderson, "The integral of a symmetric unimodel function," Proc. Amer. Math. Soc., Vol. 6 (1955), pp. 170-176.

[3] David Blackwell, "On an equation of Wald," Ann. Math. Stat., Vol. 17 (1946), pp. 84-87.

[4] David Blackwell and M. A. Girshick, Theory of Games and Statistical Decisions, John Wiley and Sons, Inc., New York 1954.

[5] T. G. Donnelly, "A family of sequential tests," Ph.D. dissertation, University of North Carolina, 1957.

[6] Wassily Hoeffding, "A lower bound for the average sample number of a sequential test," Ann. Math. Stat., Vol. 24 (1953), pp. 127-130.

[7] J. Kiefer and Lionel Weiss, "Some properties of generalized sequential probability ratio tests," Ann. Math. Stat., Vol. 28 (1957), pp. 57-74.

[8] J. Seitz and K. Winkelbauer, Remark concerning a paper of Kolmogorov and Prohorov," Czechoslovak Math. J., Vol. 3 (78) (1953), pp. 89-91 (Russian with English summary.)

[9] Abraham Wald, "Differentiation under the expectation sign in the fundamental identity of sequential analysis," Ann. Math. Stat. Vol. 17 (1946), 493-497.

[10] Abraham Wald, Sequential Analysis, John Wiley and Sons, Inc., New York, 1947.

[11] Abraham Wald, "Sequential tests of statistical hypotheses," Ann. Math. Stat., Vol. 16 (1945), 117-186.

[12] Abraham Wald, Statistical Decision Functions, John Wiley and Sons, Inc., New York 1950.

[13] A. Wald and J. Wolfowitz, "Bayes solutions of sequential decision problems," Ann. Math. Stat., Vol. 21 (1950), pp. 82-99.

[14] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," Ann. Math. Stat., Vol. 19 (1948), pp. 326-339.

[15] J. Wolfowitz, "The efficiency of sequential estimates and Wald's equation for sequential processes", Ann. Math. Stat., Vol. 18 (1947), pp. 215-230.