SEQUENTIAL TESTS OF STATISTICAL HYPOTHESES

By A. Wald

Columbia University

Table of Contents

A. Introduction

By a sequential test of a statistical hypothesis is meant any statistical test procedure which gives a specific rule, at any stage of the experiment (at the n-th trial for each integral value of n), for making one of the following three decisions: (1) to accept the hypothesis being tested (null hypothesis), (2) to reject the null hypothesis, (3) to continue the experiment by making an additional observation. Thus, such a test procedure is carried out sequentially. On the basis of the first trial, one of the three decisions mentioned above is made. If the first or the second decision is made, the process is terminated. If the third decision is made, a second trial is performed. Again on the basis of the first two trials one of the three decisions is made and if the third decision is reached a third trial is performed, etc. This process is continued until either the first or the second decision is made.

An essential feature of the sequential test, as distinguished from the current test procedure, is that the number of observations required by the sequential test is not predetermined, but is a random variable due to the fact that at any stage of the experiment the decision of terminating the process depends on the results of the observations previously made. The current test procedure may be considered a limiting case of a sequential test in the following sense: For any positive integer n less than some fixed positive integer N, the third decision is always taken at the n-th trial irrespective of the results of these first n trials. At the N-th trial either the first or the second decision is taken. Which decision is taken will depend, of course, on the results of the N trials.

In a sequential test, as well as in the current test procedure, we may commit two kinds of errors. We may reject the null hypothesis when it is true (error of the first kind), or we may accept the null hypothesis when some alternative hypothesis is true (error of the second kind). Suppose that we wish to test the null hypothesis H_0 against a single alternative hypothesis H_1, and that we want the test procedure to be such that the probability of making an error of the first kind (rejecting H_0 when H_0 is true) does not exceed a preassigned value α, and the probability of making an error of the second kind (accepting H_0 when H_1 is true) does not exceed a preassigned value β. Using the current test procedure, i.e., a most powerful test for testing H_0 against H_1 in the sense of the Neyman-Pearson theory, the minimum number of observations required by the test can be determined as follows: For any given number N of observations a most powerful test is considered for which the probability of an error of the first kind is equal to α. Let $\beta(N)$ denote the probability of an error of the second kind for this test procedure. Then the minimum number of observations is equal to the smallest positive integer N for which $\beta(N) \leq \beta$.

In this paper a particular test procedure, called the sequential probability ratio test, is devised and shown to have certain optimum properties (see section 4.7). The sequential probability ratio test in general requires an expected number of observations considerably smaller than the fixed number of observations needed by the current most powerful test which controls the errors of the first

and second kinds to exactly the same extent (has the same α and β) as the sequential test. The sequential probability ratio test frequently results in a saving of about 50% in the number of observations as compared with the current most powerful test. Another surprising feature of the sequential probability ratio test is that the test can be carried out without determining any probability distributions whatsoever. In the current procedure the test can be carried out only if the probability distribution of the statistic on which the test is based is known. This is not necessary in the application of the sequential probability ratio test, and only simple algebraic operations are needed for carrying it out. Distribution problems arise in connection with the sequential probability ratio test only if we want to make statements about the probability distribution of the number of observations required by the test.

This paper consists of two parts. Part I deals with the theory of sequential tests for testing a simple hypothesis against a single alternative. In Part II a theory of sequential tests for testing simple or composite hypotheses against infinite sets of alternatives is outlined. The extension of the probability ratio test to the case of testing a simple hypothesis against a set of one-sided alternatives is straight forward and does not present any difficulty. Applications to testing the means of binomial and normal distributions, as well as to testing double dichotomies are given. The theory of sequential tests of hypotheses with no restrictions on the possible values of the unknown parameters is, however, not as simple. There are several unsolved problems in this case and it is hoped that the general ideas outlined in Part II will stimulate further research.

Sections 5.2, 5.3 and 5.4 in Part II deal with the applications of the sequential probability ratio test to binomial distributions, double dichotomies and normal distributions. These sections are nearly self-contained and can be understood without reading the rest of the paper. Thus, readers who are primarily interested in these special cases of the sequential probability ratio test rather than in the general theory, may profitably read only the above mentioned sections For the benefit of readers who lack a sufficient background in the mathematical theory of statistics the exposition in sections 5.2, 5.3 and 5.4 is kept on a fairly elementary level.

It should be pointed sut that whenever the number of observations on which the test is based is for some reason determined in advance, for instance, if certain data are available from past history and no additional data can be obtained, then the current most powerful test procedure is preferable. The superiority of the sequential probability ratio test is due to the fact that it requires a smaller expected number of observations than the current most powerful test. This feature of the sequential probability ratio test is, however, of no value if the number of observations is for some reason determined in advance.

B. Historical Note

To the best of the author's knowledge the first idea of a sequential test, i.e., a test where the number of observations is not predetermined but is dependent

on the outcome of the observations, goes back to H. F. Dodge and H. G. Romig who proposed a double sampling inspection procedure [1]. In this double sampling scheme the decision whether a second sample should be drawn or not depends on the outcome of the observations in the first sample. The reason for introducing a double sampling method was, of course, the recognition of the fact that double sampling results in a reduction of the amount of inspection as compared with "single" sampling.

The double sampling method does not fully take advantage of sequential analysis, since it does not allow for more than two samples. A multiple sampling scheme for the particular case of testing the mean of a binomial distribution was proposed and discussed by Walter Bartky [2]. His procedure is closely related to the test which results from the application of the sequential probability ratio test to testing the mean of a binomial distribution. Bartky clearly recognized the fact that multiple sampling results in a considerable reduction of the average amount of inspection.

The idea of chain experiments discussed briefly by Harold Hotelling [3] is also somewhat related to our notion of sequential analysis. An interesting example of such a chain of experiments is the series of sample censuses of area of jute in Bengal carried out under the direction of P. C. Mahalanobis [6]. The successive preliminary censuses, steadily increasing in size, were primarily designed to obtain some information as to the parameters to be estimated so that an efficient design could be set up for the final sampling of the whole immense jute area in the province.

In March 1943, the problem of sequential analysis arose in the Statistical Research Group, Columbia University,[1] in connection with a specific question posed by Captain G. L. Schuyler of the Bureau of Ordnance, Navy Department. It was pointed out by Milton Friedman and W. Allen Wallis that the mere notion of sequential analysis could slightly improve the efficiency of some current most powerful tests. This can be seen as follows: Suppose that N is the planned number of trials and W_N is a most powerful critical region based on N observations. If it happens that on the basis of the first n trials ($n < N$) it is already certain that the completed set of N trials must lead to a rejection of the null hypothesis, we can terminate the experiment at the n-th trial and thus save some observations. For instance, if W_N is defined by the inequality $x_1^2 + \ldots + x_N^2 \geq c$, and if for some $n < N$ we find that $x_1^2 + \ldots + x_n^2 \geq c$, we can terminate the process at this stage. Realization of this naturally led Friedman and Wallis to the conjecture that modifications of current tests may exist which take advantage of sequential procedure and effect substantial improvements. More specifically, Friedman and Wallis conjectured that a sequential test may exist that controls the errors of the first and second kinds to exactly the same extent as the current

[1] The Statistical Research Group operates under a contract with the Office of Scientific Research and Development and is directed by the Applied Mathematics Panel of the National Defense Research Committee.

most powerful test, and at the same time requires an expected number of observations substantially smaller than the number of observations required by the current most powerful test.[2]

It was at this stage that the problem was called to the attention of the author of the present paper. Since infinitely many sequential test procedures exist, the first and basic problem was, of course, to find the particular sequential test procedure which is most efficient, i.e., which effects the greatest possible saving in the expected number of observations as compared with any other (sequential or non-sequential) test. In April, 1943 the author devised such a test, called the sequential probability ratio test, which for all practical purposes is most efficient when used for testing a simple hypothesis H_0 against a single alternative H_1.

Because of the substantial savings in the expected number of observations effected by the sequential probability ratio test, and because of the simplicity of this test procedure in practical applications, the National Defense Research Committee considered these developments sufficiently useful for the war effort to make it desirable to keep the results out of the reach of the enemy, at least for a certain period of time. The author was, therefore, requested to submit his findings in a restricted report [7] which was dated September, 1943.[3] In this report the sequential probability ratio test is devised and its mathematical theory is developed. In July 1944 a second report [8] was issued by the Statistical Research Group which gives an elementary non-mathematical exposition of the applications of the sequential probability ratio test, together with charts, tables and computational simplifications to facilitate applications.

Independently of the developments here, G. A. Barnard [9] recognized the merits of a sequential method of testing, i.e., the possibility of a saving in the number of observations as compared with the current most powerful test. He also devised an interesting sequential test for testing double dichotomies, which differs from the one obtained by applying the sequential probability ratio test.

Some further developments in the theory of the sequential probability ratio test took place in 1944. Extending the methods used in [7], C. M. Stockman [10] found the operating characteristic curve of the sequential probability ratio test applied to a binomial distribution. Independently of Stockman, Milton Friedman and George W. Brown (independently of each other) obtained the same result which can be extended to the normal distribution and a few other specific distributions, but is not applicable to more general distributions. The general operating characteristic curve for any sequential probability ratio test was derived by the author [11]. A few months later the author developed a general theory of cumulative sums [4] which gives not only the operating char-

[2] Bartky's multiple sampling scheme [2] for testing the mean of a binomial distribution provides, of course, an example of such a sequential test (see, for example, the remarks on p. 377 in [2]). Bartky's results were not known to us at that time, since they were published nearly a year later.

[3] The material was recently released making the present publication possible.

acteristic curve for any sequential probability ratio test but also the characteristic function of the number of observations required by the test.

The theory of the sequential probability ratio test as given in the present paper differs considerably from the exposition given in [7], since the new developments in [4] have been taken into account. However, some tables and a few sections of the original report [7] are included in the present paper without any substantial changes.

Part I. Sequential Test of a Simple Hypothesis Against a Single Alternative

1. The Current Test Procedure

Let X be a random variable. In what follows in this and the subsequent sections it will be assumed that the random variable X has either a continuous probability density function or a discrete distribution. Accordingly, by the probability distribution $f(x)$ of a random variable X we shall mean either the probability density function of X or the probability that $X = x$, depending upon whether X is a continuous or a discrete variable. Let the hypothesis H_0 to be tested (null hypothesis) be the statement that the distribution of X is $f_0(x)$. Suppose that H_0 is to be tested against the single alternative hypothesis H_1 that the distribution of X is given by $f_1(x)$.

According to the Neyman-Pearson theory of testing hypotheses a most powerful critical region W_N for testing H_0 against H_1 on the basis of N independent observations $x_1, \cdots, x_N$ on X is given by the set of all sample points $(x_1, \cdots, x_N)$ for which the inequality

$$\frac{f_1(x_1)f_1(x_2) \cdots f_1(x_N)}{f_0(x_1)f_0(x_2) \cdots f_0(x_N)} \geq k \tag{1.1}$$

is fulfilled. The quantity k on the right hand side of (1.1) is a constant and is chosen so that the size of the critical region, i.e., the probability of an error of the first kind should have the required value α.

For a fixed sample size N the probability β of an error of the second kind is a single valued function of α, say $\beta_N(\alpha)$, if a most powerful critical region is used. Thus, if in addition to fixing the value of α it is required that the probability of an error of the second kind should have a preassigned value β, or at least it should not exceed a preassigned value β, we are no longer free to choose the sample size N. The minimum number of observations required by the test satisfying these conditions is equal to the smallest integral value of N for which $\beta_N(\alpha) \leq \beta$.

Thus, the current most powerful test procedure for testing H_0 against H_1 can be briefly stated as follows: We choose as critical region the region defined by (1.1) where the constant k is determined so that the probability of an error of the first kind should have a preassigned value α and N is equal to the smallest integer for which the probability of an error of the second kind does not exceed a preassigned value β.

2. The Sequential Test Procedure: General Definitions

2.1. *Notion of a sequential test.* In current tests of hypotheses the number of observations is treated as a constant for any particular problem. In sequential tests the number of observations is no longer a constant, but a random variable. In what follows the symbol n is used for the number of observations required by a sequential test and the symbol N is used when the number of observations is treated as a constant.

Sequential tests can be described as follows: For each positive integer m the m-dimensional sample space M_m is subdivided into three mutually exclusive parts R_m^0, R_m^1 and R_m. After the first observation x_1 has been drawn H_0 is accepted if x_1 lies in R_1^0, H_0 is rejected (i.e., H_1 is accepted) if x_1 lies in R_1^1, or a second observation is drawn if x_1 lies in R_1. If the third decision is reached and a second observation x_2 drawn, H_0 is accepted, H_1 is accepted, or a third observation is drawn according as the point (x_1, x_2) lies in R_2^0, R_2^1 or in R_2. If (x_1, x_2) lies in R_2, a third observation x_3 is drawn and one of the three decisions is made according as (x_1, x_2, x_3) lies in R_3^0, R_3^1 or in R_3, etc. This process is stopped when, and only when, either the first decision or the second decision is reached. Let n be the number of observations at which the process is terminated. Then n is a random variable, since the value of n depends on the outcome of the observations. (It will be seen later that the probability is one that the sequential process will be terminated at some finite stage.)

We shall denote by $E_0(n)$ the expected value of n if H_0 is true and by $E_1(n)$ the expected value of n if H_1 is true. These expected values, of course, depend on the sequential test used. In order to put this dependence in evidence, we shall occasionally use the symbols $E_0(n \mid S)$ and $E_1(n \mid S)$ to denote the values $E_0(n)$ and $E_1(n)$, respectively, when the sequential test S is applied.

2.2. *Efficiency of a sequential test.* As in the current test procedure, errors of two kinds may be committed in sequential analysis. We may reject H_0 when it is true (error of the first kind), or we may accept H_0 when H_1 is true (error of the second kind). With any sequential test there will be associated two numbers α and β between 0 and 1 such that if H_0 is true the probability is α that we shall commit an error of the first kind and if H_1 is true, the probability is β that we shall commit an error of the second kind. We shall say that two sequential tests S and S' are of equal strength if the values α and β associated with S are equal to the corresponding values α' and β' associated with S'. If $\alpha < \alpha'$ and $\beta \leq \beta'$, or if $\alpha \leq \alpha'$ and $\beta < \beta'$, we shall say that S is stronger than S' (S' is weaker than S). If $\alpha > \alpha'$ and $\beta < \beta'$, or if $\alpha < \alpha'$ and $\beta > \beta'$, we shall say that the strength of S is not comparable with that of S'.

Restricting ourselves to sequential tests of a given strength, we want to make the number of observations necessary for reaching a final decision as small as possible. If S and S' are two sequential tests of equal strength we shall say that S' is better than S if either $E_0(n \mid S') < E_0(n \mid S)$ and $E_1(n \mid S') \leq E_1(n \mid S)$, or $E_0(n \mid S') \leq E_0(n \mid S)$ and $E_1(n \mid S') < E_1(n \mid S)$. A sequential test will be said to be an admissible test if no better test of equal strength exists.

If a sequential test S satisfies both inequalities $E_0(n \mid S) \leq E_0(n \mid S')$ and $E_1(n \mid S) \leq E_1(n \mid S')$ for any sequential test S' of strength equal to that of S, then the test S can be considered to be a best sequential test. That such tests exist, i.e., that it is possible to minimize $E_0(n)$ and $E_1(n)$ simultaneously, is not proved here; but it is shown later (section 4.7) that for the so called sequential probability ratio test defined in section 3.1 both $E_0(n)$ and $E_1(n)$ are very nearly minimized.[4] Thus, for all practical purposes the sequential probability ratio test can be considered best.

Since it is unknown that a sequential test always exists for which both $E_0(n)$ and $E_1(n)$ are exactly minimized, we need a substitute definition of an optimum test. Several substitute definitions are possible. We could, for example, require that the test be admissible and the maximum of the two values $E_0(n)$ and $E_1(n)$ be minimized, or that the mean $\frac{E_0(n) + E_1(n)}{2}$, or some other weighted average be minimized. All these definitions are equivalent if a sequential test exists for which both $E_0(n)$ and $E_1(n)$ are minimized; but if they cannot be minimized simultaneously the definitions differ. Which of them is chosen is of no significance for the purpose of this paper, since for the sequential probability ratio test proposed later both expected values $E_0(n)$ and $E_1(n)$ are, if not exactly, very nearly minimized. If we had a priori knowledge as to how frequently H_0 and how frequently H_1 will be true in the long run, it would be most reasonable to minimize a weighted average (weighted by the frequencies of H_0 and H_1, respectively) of $E_0(n)$ and $E_1(n)$. However, when such knowledge is absent, as is usually the case in practical applications, it is perhaps more reasonable to minimize the maximum of $E_0(n)$ and $E_1(n)$ than to minimize some weighted average of $E_0(n)$ and $E_1(n)$. Hence the following definition is introduced.

A sequential test S is said to be an optimum test if S is admissible and Max $[E_0(n \mid S), E_1(n \mid S)] \leq$ Max $[E_0(n \mid S'), E_1(n \mid S')]$ for all sequential tests S' of strength equal to that of S.

By the efficiency of a sequential test S is meant the value of the ratio[5]

$$\frac{\text{Max } [E_0(n \mid S^*), E_1(n \mid S^*)]}{\text{Max } [E_0(n \mid S), E_1(n \mid S)]}$$

where S^* is an optimum sequential test of strength equal to that of S.

2.3. *Efficiency of the current procedure, viewed as a particular case of a sequential test.* The current test procedure can be considered as a particular case of a sequential test. In fact, let N be the size of the sample used in the current procedure and let W_N be the critical region on which the test is based. Then the

[4] The author conjectures that $E_0(n)$ and $E_1(n)$ are exactly minimized for the sequential probability ratio test, but he did not succeed in proving this, except for a special class of problems (see section 4.7).

[5] The existence of an optimum sequential test is not essential for the definition of efficiency, since Max $[E_0(n \mid S^*), E_1(n \mid S^*)]$ could be replaced by the greatest lower bound of Max $[E(_0n \mid S'), E_1(n \mid S')]$ with respect to all sequential tests S' of strength equal to that of S.

current procedure can be considered as a sequential test defined as follows: For all $m < N$, the regions R_m^0, R_m^1 are the empty subsets of the m-dimensional sample space M_m, and $R_m = M_m$. For $m = N$, R_N^1 is equal to W_N, R_N^0 is equal to the complement $\overline{W}_N$ of W_N and R_N is the empty set. Thus, for the current procedure we have $E_0(n) = E_1(n) = N$.

It will be seen later that the efficiency of the current test based on the most powerful critical region is rather low. Frequently it is below $\frac{1}{2}$. In other words, an optimum sequential test can attain the same α and β as the current most powerful test on the basis of an expected number of observations much smaller than the fixed number of observations needed for the current most powerful test.

In the next section we shall propose a simple sequential test procedure, called the sequential probability ratio test, which for all practical purposes can be considered an optimum sequential test. It will be seen that these sequential tests usually lead to average savings of about 50% in the number of trials as compared with the current most powerful test.

3. Sequential Probability Ratio Test

3.1. *Definition of the sequential probability ratio test.* We have seen in section 2.1 that the sequential test procedure is defined by subdividing the m-dimensional sample space M_m ($m = 1, 2, \cdots$, ad inf.) into three mutually exclusive parts R_m^0, R_m^1 and R_m. The sequential process is terminated at the smallest value n of m for which the sample point lies either in R_n^0 or in R_n^1. If the sample point lies in R_n^0 we accept H_0 and if it lies in R_n^1 we accept H_1.

An indication as to the proper choice of the regions R_m^0, R_m^1 and R_m can be obtained from the following considerations: Suppose that before the sample is drawn there exists an a priori probability that H_0 is true and the value of this probability is known. Denote this a priori probability by g_0. Then the a priori probability that H_1 is true is given by $g_1 = 1 - g_0$, since it is assumed that the hypotheses H_0 and H_1 exhaust all possibilities. After a number of observations have been made we gain additional information which will affect the probability that H_i ($i = 0, 1$) is true. Let g_{0m} be the a posteriori probability that H_0 is true and g_{1m} the a posteriori probability that H_1 is true after m observations have been made. Then according to the well known formula of Bayes we have

$$g_{0m} = \frac{g_0 p_{0m}(x_1, \cdots, x_m)}{g_0 p_{0m}(x_1, \cdots, x_m) + g_1 p_{1m}(x_1, \cdots, x_m)} \tag{3.1}$$

and

$$g_{1m} = \frac{g_1 p_{1m}(x_1, \cdots, x_m)}{g_0 p_{0m}(x_1, \cdots, x_m) + g_1 p_{1m}(x_1, \cdots, x_m)} \tag{3.2}$$

where $p_{im}(x_1, \cdots, x_m)$ denotes the probability density in the m-dimensisnal sample space calculated under the hypothesis H_i ($i = 0, 1$).[6] As an abbreviation for $p_{im}(x_1, \cdots, x_m)$ we shall use simply p_{im}.

[6] If the probability distribution is discrete $p_{im}(x_1, \cdots, x_m)$ denotes the probability that the sample point $(x_1, \cdots, x_m)$ will be obtained.

Let d_0 and d_1 be two positive numbers less than 1 and greater than $\frac{1}{2}$. Suppose that we want to construct a sequential test such that the conditional probability of a correct decision under the condition that H_0 is accepted is greater than or equal to d_0, and the conditional probability of a correct decision under the condition that H_1 is accepted is greater than or equal to d_1.[7] Then the following sequential process seems reasonable: At each stage calculate g_{0m} and g_{1m}. If $g_{1m} \geq d_1$, accept H_1. If $g_{0m} \geq d_0$, accept H_0. If $g_{1m} < d_1$ and $g_{0m} < d_0$, draw an additional observation. R_m^0 in this sequential process is thus defined by the inequality $g_{0m} \geq d_0$, R_m^1 by the inequality $g_{1m} \geq d_1$, and R_m by the simultaneous inequalities $g_{1m} < d_1$ and $g_{0m} < d_0$. It is necessary that the sets R_m^0, R_m^1 and R_m be mutually exclusive and exhaustive. For this it suffices that the inequalities

$$g_{1m} = \frac{g_1 p_{1m}}{g_0 p_{0m} + g_1 p_{1m}} \geq d_1 \tag{3.3}$$

and

$$g_{0m} = \frac{g_0 p_{0m}}{g_0 p_{0m} + g_1 p_{1m}} \geq d_0 \tag{3.4}$$

be not fulfilled simultaneously. To show that (3.3) and (3.4) are incompatible, we shall assume that they are simultaneously fulfilled and derive a contradiction from this assumption. The two inequalities sum to

$$g_{1m} + g_{0m} \geq d_1 + d_0 . \tag{3.5}$$

Since $g_{0m} + g_{1m} = 1$, we have

$$1 \geq d_1 + d_0$$

which is impossible, since by assumption $d_i > \frac{1}{2}$ $(i = 0, 1)$. Hence it is proved that the sets R_m^0, R_m^1 and R_m are mutually exclusive and exhaustive.

The inequalities (3.3) and (3.4) are equivalent to the following inequalities, respectively:

$$\frac{p_{1m}}{p_{0m}} \geq \frac{g_0}{g_1} \frac{d_1}{1 - d_1} \tag{3.6}$$

and

$$\frac{p_{1m}}{p_{0m}} \leq \frac{g_0}{g_1} \frac{1 - d_0}{d_0} . \tag{3.7}$$

The constants on the right hand sides of (3.6) and (3.7) do not depend on m.

If an a priori probability of H_0 does not exist, or if it is unknown, the inequalities (3.6) and (3.7) suggest the use of the following sequential test: At each stage

[7] The restriction $d_0 > 1/2$ and $d_1 > 1/2$ are imposed because otherwise it might happen that the hypothesis with the smaller a posteriori probability will be accepted.

calculate p_{1m}/p_{0m}. If $p_{1m} = p_{0m} = 0$, the value of the ratio p_{1m}/p_{0m} is defined to be equal to 1. Accept H_1 if

$$\frac{p_{1m}}{p_{0m}} \geq A. \tag{3.8}$$

Accept H_0 if

$$\frac{p_{1m}}{p_{0m}} \leq B. \tag{3.9}$$

Take an additional observation if

$$B < \frac{p_{1m}}{p_{0m}} < A. \tag{3.10}$$

Thus, the number n of observations required by the test is the smallest integral value of m for which either (3.8) or (3.9) holds. The constants A and B are chosen so that $0 < B < A$ and the sequential test has the desired value α of the probability of an error of the first kind and the desired value β of the probability of an error of the second kind. We shall call the test procedure defined by (3.8), (3.9) and (3.10), a sequential probability ratio test.

The sequential test procedure given by (3.8), (3.9) and (3.10) has been justified here merely on an intuitive basis. Section 4.7, however, shows that for this sequential test the expected values $E_0(n)$ and $E_1(n)$ are very nearly minimized.[8] Thus, for practical purposes this test can be considered an optimum test.

3.2. *Fundamental relations among the quantities α, β, A and B.* In this section the quantities α, β, A and B will be related by certain inequalities which are of basic importance for the sequential analysis.

Let $\{x_m\}(m = 1, 2, \cdots, \text{ad inf.})$ be an infinite sequence of observations. The set of all possible infinite sequences $\{x_m\}$ is called the infinite dimensional sample space. It will be denoted by M_∞. Any particular infinite sequence $\{x_m\}$ is called a point of M_∞. For any set of n given real numbers $a_1, \cdots, a_n$ we shall denote by $C(a_1, \cdots, a_n)$ the subset of M_∞ which consists of all points (infinite sequences) $\{x_m\}$ $(m = 1, 2, \cdots, \text{ad inf.})$ for which $x_1 = a_1, \cdots, x_n = a_n$. For any values $a_1, \cdots, a_n$ the set $C(a_1, \cdots, a_n)$ will be called a cylindric point of order n. A subset S of M_∞ will be called a cylindric point, if there exists a positive integer n for which S is a cylindric point of order n. Thus, a cylindric point may be a cylindric point of order 1, or of order 2, etc. A cylindric point $C(a_1, \cdots, a_n)$ will be said to be of type 1 if

$$\frac{p_{1n}}{p_{0n}} = \frac{f_1(a_1)f_1(a_2)\cdots f_1(a_n)}{f_0(a_1)f_0(a_2)\cdots f_0(a_n)} \geq A$$

[8] It seems likely to the author that $E_0(n)$ and $E_1(n)$ are exactly minimized for the sequential probability ratio test. However, he did not succeed in proving it, except for a special class of problems (see section 4.7).

and

$$B < \frac{p_{1m}}{p_{0m}} = \frac{f_1(a_1) \cdots f_1(a_m)}{f_0(a_1) \cdots f_0(a_m)} < A \quad (m = 1, \cdots, n-1)$$

A cylindric point $C(a_1, \cdots, a_n)$ will be said to be of type 0 if

$$\frac{p_{1n}}{p_{0n}} = \frac{f_1(a_1) \cdots f_1(a_n)}{f_0(a_1) \cdots f_0(a_n)} \leq B$$

and

$$B < \frac{p_{1m}}{p_{0m}} = \frac{f_1(a_1) \cdots f_1(a_m)}{f_0(a_1) \cdots f_0(a_m)} < A \quad (m = 1, \cdots, n-1).$$

Thus, if a sample $(x_1, \cdots, x_n)$ is observed for which $C(x_1, \cdots, x_n)$ is a cylindric point of type i, the sequential test defined by (3.8), (3.9) and (3.10) leads to the acceptance of H_i $(i = 0, 1)$.

Let Q_i be the sum of all cylindric points of type i $(i = 0, 1)$. For any subset M of M_∞ we shall denote by $P_i(M)$ the probability of M calculated under the assumption that H_i is true $(i = 0, 1)$. Now we shall prove that

$$P_i(Q_0 + Q_1) = 1 \qquad (i = 0, 1) \tag{3.11}$$

This equation means that the probability is equal to one that the sequential process will eventually terminate. To prove (3.11) we shall denote the variate $\log \frac{f_1(x_i)}{f_0(x_i)}$ by z_i and $z_1 + \cdots + z_m$ by Z_m $(i, m = 1, 2, \cdots, \text{ad inf.})$. Furthermore, denote by n the smallest integer for which either $Z_n \geq \log A$ or $Z_n \leq \log B$. If no such finite integer n exists we shall say that $n = \infty$. Clearly, n is the number of observations required by the sequential test and (3.11) is proved if we show that the probability that $n = \infty$ is zero. But the latter statement was proved by the author elsewhere (see Lemma 1 in [4]). Hence equation (3.11) is proved.

With the help of (3.11) we shall be able to derive some important inequalities satisfied by the quantities α, β, A and B. Since for each sample $(x_1, \cdots, x_n)$ for which $C(x_1, \cdots, x_n)$ is an element of Q_1 the inequality $p_{1n}/p_{0n} \geq A$ holds, we see that

$$P_1(Q_1) \geq AP_0(Q_1) \tag{3.12}$$

Similarly, for each sample $(x_1, \cdots, x_n)$ for which $C(x_1, \cdots, x_n)$ is a point of Q_0 the inequality $p_{1n}/p_{0n} \leq B$ holds. Hence

$$P_1(Q_0) \leq BP_0(Q_0). \tag{3.13}$$

But $P_0(Q_1)$ is the probability of committing an error of the first kind and $P_1(Q_0)$ is the probability of making an error of the second kind. Thus, we have

$$P_0(Q_1) = \alpha; \qquad P_1(Q_0) = \beta. \tag{3.14}$$

Since Q_0 and Q_1 are disjoint, it follows from (3.11) that

$$P_0(Q_0) = 1 - \alpha; \qquad P_1(Q_1) = 1 - \beta. \tag{3.15}$$

From the relations (3.12)–(3.15) we obtain the important inequalities

$$1 - \beta \geq A\,\alpha \tag{3.16}$$

and

$$\beta \leq B\,(1 - \alpha). \tag{3.17}$$

These inequalities can be written as

$$\frac{\alpha}{1 - \beta} \leq \frac{1}{A} \tag{3.18}$$

and

$$\frac{\beta}{1 - \alpha} \leq B. \tag{3.19}$$

The above inequalities are of great value in practical applications, since they supply upper limits for α and β when A and B are given. For instance, it follows immediately from (3.18) and (3.19), and the fact that $0 < \alpha < 1, 0 < \beta < 1$ that

$$\alpha \leq \frac{1}{A} \tag{3.20}$$

and

$$\beta \leq B. \tag{3.21}$$

A pair of values α and β can be represented by a point in the plane with the coordinates α and β. It is of interest to determine the set of all points (α, β) which satisfy the inequalities (3.18) and (3.19) for given values of A and B. Consider the straight lines L_1 and L_2 in the plane given by the equations

$$A\alpha = 1 - \beta \tag{3.22}$$

and

$$\beta = B(1 - \alpha), \tag{3.23}$$

respectively. The line L_1 intersects the abscissa axis at $\alpha = \frac{1}{A}$ and the ordinate axis at $\beta = 1$. The line L_2 intersects the abscissa axis at $\alpha = 1$ and the ordinate axis at $\beta = B$. The set of all points (α, β) which satisfy the inequalities (3.18) and (3.19) is the interior and the boundary of the quadrilateral determined by the lines L_1, L_2 and the coordinate axes. This set is represented by the shaded area in figure 1.

The fundamental inequalities (3.18) and (3.19) were derived under the assumption that $x_1, x_2, \cdots$, ad inf. are independent observations on the same random

variable X. The independence of the observations is, however, not necessary for the validity of (3.18) and (3.19). In fact, the independence of the observations was used merely to show the validity of (3.11). But (3.11) can be shown to hold also for dependent observations under very general conditions. Hence, if H_i states that the joint distribution of $x_1, x_2, \cdots, x_m$ is given by the joint probability density function $p_{im}(x_1, \cdots, x_m)$[9] $(i = 0, 1; m = 1, 2, \cdots$, ad inf.) and if (3.11) holds, then for the sequential test of H_0 against H_1, as defined by (3.8), (3.9) and (3.10), the inequalities (3.18) and (3.19) remain valid. For instance, let λ_0 and λ_1 be two different positive values <1 and let $H_i(i = 0, 1)$ be the hypothesis that the joint probability density function of $x_1, \cdots, x_m$ is given by

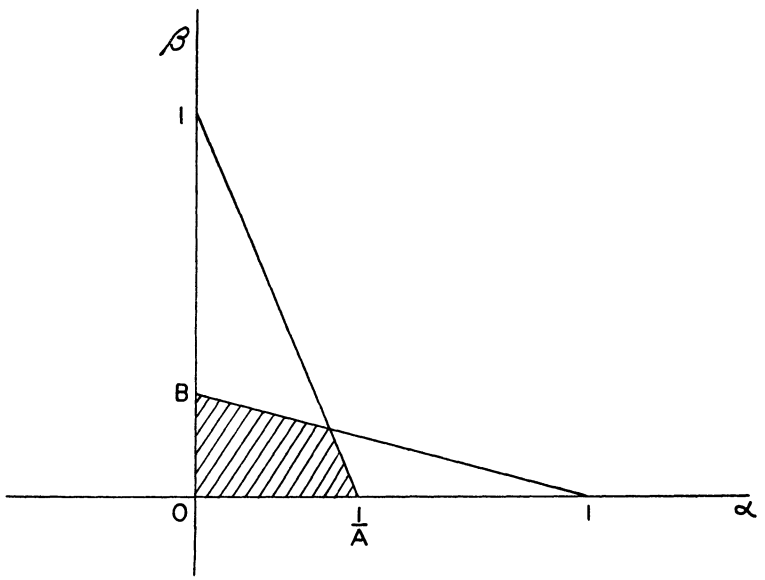

FIG. 1

$$p_{im}(x_1, \cdots, x_m) = \frac{1}{(2\pi)^{m/2}} e^{-\frac{1}{2}x_1^2 - \frac{1}{2}\sum_{j=2}^{m}(x_j - \lambda_i x_{j-1})^2} \qquad (i = 0, 1)$$

i.e., that x_1 and $(x_j - \lambda_i x_{j-1})(j = 2, 3, \cdots$, ad inf.) are normally and independently distributed with zero means and unit variances, then the inequalities (3.18) and (3.19) will hold for the sequential test defined by (3.8), (3.9) and (3.10).

3.3. *Determination of the values A and B in practice.* Suppose that we wish to have a sequential test such that the probability of an error of the first kind is equal to α and the probability of an error of the second kind is equal to β. De-

[9] Of course, for any positive integers m and m' with $m < m'$ the marginal distribution of $x_1, \cdots, x_m$ determined on the basis of the joint distribution $P_{im'}(x_1, \cdots, x_{m'})$ must be equal to $P_{im}(x_1, \cdots, x_m)$.

note by $A(\alpha, \beta)$ and $B(\alpha, \beta)$ the values of A and B for which the probabilities of the errors of the first and second kinds will take the desired values α and β. The exact determination of the values $A(\alpha, \beta)$ and $B(\alpha, \beta)$ is rather laborious, as will be seen in Section 3.4. The inequalities at our disposal, however, permit the problem to be solved satisfactorily for practical purposes. From (3.18) and (3.19) it follows that

$$A(\alpha, \beta) \leq \frac{1 - \beta}{\alpha} \tag{3.24}$$

and

$$B(\alpha, \beta) \geq \frac{\beta}{1 - \alpha}. \tag{3.25}$$

Suppose we put $A = \frac{1 - \beta}{\alpha} = a(\alpha, \beta)$ (say), and $B = \frac{\beta}{1 - \alpha} = b(\alpha, \beta)$ (say). Then A is greater than or equal to the exact value $A(\alpha, \beta)$, and B is less than or equal to the exact value $B(\alpha, \beta)$. This procedure, of course, changes the probabilities of errors of the first and second kind. If we were to use the exact value of B and a value of A which is greater than the exact value, then evidently we would lower the value of α, but slightly increase the value of β. Similarly, if we were to use the exact value of A and a value of B which is below the exact value, then we would lower the value of β, but slightly increase the value of α. Thus, it is not clear what will be the resulting effect on α and β if a value of A is used which is higher than the exact value, and a value of B is used which is lower than the exact value. Denote by α' and β' the resulting probabilities of errors of the first and second kind, respectively, if we put $A = \frac{1 - \beta}{\alpha}$ and $B = \frac{\beta}{1 - \alpha}$.

We now derive inequalities satisfied by the quantities α', β', α and β. Substituting $a(\alpha, \beta)$ for A, $b(\alpha, \beta)$ for B, α' for α and β' for β we obtain from (3.18) and (3.19)

$$\frac{\alpha'}{1 - \beta'} \leq \frac{1}{a(\alpha, \beta)} = \frac{\alpha}{1 - \beta} \tag{3.26}$$

and

$$\frac{\beta'}{1 - \alpha'} \leq b(\alpha, \beta) = \frac{\beta}{1 - \alpha}. \tag{3.27}$$

From these inequalities it follows that

$$\alpha' \leq \frac{\alpha}{1 - \beta} \tag{3.28}$$

and

$$\beta' \leq \frac{\beta}{1 - \alpha}. \tag{3.29}$$

Multiplying (3.26) by $(1 - \beta)(1 - \beta')$ and (3.27) by $(1 - \alpha)(1 - \alpha')$ and adding the two resulting inequalities, we have

$$\alpha' + \beta' \leq \alpha + \beta. \tag{3.30}$$

Thus, we see that at least one of the inequalities $\alpha' \leq \alpha$ and $\beta' \leq \beta$ must hold. In other words, by using $a(\alpha, \beta)$ and $b(\alpha, \beta)$ instead of $A(\alpha, \beta)$ and $B(\alpha, \beta)$, respectively, at most one of the probabilities α and β may be increased.

If α and β are small (say less than .05), as they frequently will be in practical applications, $\frac{\alpha}{1 - \beta}$ and $\frac{\beta}{1 - \alpha}$ are nearly equal to α and β, respectively. Thus, we see from (3.28) and (3.29) that the quantity by which α' can possibly exceed α, or β' can exceed β, must be small. Section 3.4 contains further inequalities which show that the amount by which $\alpha'(\beta')$ can possibly exceed $\alpha(\beta)$ is indeed extremely small. Thus, for all practical purposes $\alpha' \leq \alpha$ and $\beta' \leq \beta$.

If $f_1(x)$ (the distribution under the alternative hypothesis) is sufficiently near $f_0(x)$ (the distribution under the null hypothesis), $A(\alpha, \beta)$ and $B(\alpha, \beta)$ will be nearly equal to $\frac{1 - \beta}{\alpha}$ and $\frac{\beta}{1 - \alpha}$, respectively; and consequently α' and β' are also very nearly equal to α and β respectively. The reason that (3.18) and (3.19) and therefore also (3.24) and (3.25), are inequalities instead of equalities is that the sequential process may terminate with $\frac{p_{1n}}{p_{0n}} > A$ or $\frac{p_{1n}}{p_{0n}} < B$. If at the final stage $\frac{P_{1n}}{P_{0n}}$ were exactly equal to A or B, then $A(\alpha, \beta)$ and $B(\alpha, \beta)$ would be exactly $\frac{1 - \beta}{\alpha}$ and $\frac{\beta}{1 - \alpha}$, respectively. If $f_1(x)$ is near $f_0(x)$, it is almost certain that the value of $\frac{P_{1n}}{P_{0n}}$ is changed only slightly by one additional observation. Thus, at the final stage $\frac{P_{1n}}{P_{0n}}$ will be only slightly above A, or slightly below B and consequently $A(\alpha, \beta)$ and $B(\alpha, \beta)$ will be nearly equal to $\frac{1 - \beta}{\alpha}$ and $\frac{\beta}{1 - \alpha}$, respectively. If fractional observations were possible, that is to say, if the number of observations were a continuous variable, $\frac{P_{1m}}{P_{0m}}$ would also be a continuous function of m and consequently $A(\alpha, \beta)$ and $B(\alpha, \beta)$ would be exactly equal to $\frac{1 - \beta}{\alpha}$ and $\frac{\beta}{1 - \alpha}$, respectively. Thus, we have inequalities in (3.24) and (3.25) instead of equalities merely on account of the fact that the number m of observations is discontinuous, i.e., m can take only integral values.

Hence for all practical purposes the following procedure can be adopted: *To construct a sequential test such that the probability of an error of the first kind does not exceed α and the probability of an error of the second kind does not exceed β, put*

$A = \frac{1-\beta}{\alpha}$ *and* $B = \frac{\beta}{1-\alpha}$ *and carry out the sequential test as defined by the inequalities* (3.8), (3.9) *and* (3.10).

In most practical cases the calculation of the exact values $A(\alpha, \beta)$ and $B(\alpha, \beta)$ will be of little interest for the following reasons: When $A = a(\alpha, \beta) = \frac{1-\beta}{\alpha}$ and $B = b(\alpha, \beta) = \frac{\beta}{1-\alpha}$, the probability α' of an error of the first kind cannot exceed α and the probability β' of an error of the second kind cannot exceed β, except by a very small quantity which can be neglected for practical purposes. Thus, for all practical purposes the use of $a(\alpha, \beta)$ and $b(\alpha, \beta)$ instead of $A(\alpha, \beta)$ and $B(\alpha, \beta)$ will not decrease the strength of the sequential test. The only possible disadvantage from the substitution is that it may increase the expected number of trials necessary for a decision. Since the discrepancy between $A(\alpha, \beta)$ and $B(\alpha, \beta)$ on the one hand and $a(\alpha, \beta)$ and $b(\alpha, \beta)$ on the other, arises only from the discontinuity of the number m of observations, it is clear that the increase in the expected number of trials caused by the use of $a(\alpha, \beta)$ and $b(\alpha, \beta)$ will be slight. This slight increase, however, cannot be considered entirely a loss for the following reason: if $a(\alpha, \beta) > A(\alpha, \beta)$ or $b(\alpha, \beta) < B(\alpha, \beta)$, then we can sharpen the inequality (3.30) to $\alpha' + \beta' < \alpha + \beta$. Hence by using $a(\alpha, \beta)$ and $b(\alpha, \beta)$ we gain in strength.

The fact that for practical purposes we may put $A = a(\alpha, \beta)$ and $B = b(\alpha, \beta)$ brings out a surprising feature of the sequential test as compared with current tests. While current tests cannot be carried out without finding the probability distribution of the statistic on which the test is based, there are no distribution problems in connection with sequential tests. In fact, $a(\alpha, \beta)$ and $b(\alpha, \beta)$ depend on α and β only, and the ratio $\frac{p_{1m}}{p_{0m}}$ can be calculated from the data of the problem without solving any distribution problems. Distribution problems arise in connection with the sequential process only if it is desired to find the probability distribution of the number of trials necessary for reaching a final decision. (This subject is discussed later.) But this is of secondary importance as long as we know that the sequential test on the average leads to a saving in the number of trials.

3.4. *Probability of accepting* H_0 *(or* H_1*) when some third hypothesis* H *is true.* In Section 3.2 we were concerned with the probability that the sequential probability ratio test will lead to the acceptance of H_0 (or H_1) when H_0 or H_1 is true. Since in Part II we shall admit an infinite set of alternatives, and since this is the practically important case, it is of interest to study the probability of accepting H_0 (or H_1) when any third hypothesis H, not necessarily equal to H_0 or H_1, is true. Let H be the hypothesis that the distribution of X is given by $f(x)$. If $f(x)$ is equal to $f_0(x)$ or $f_1(x)$ we have the special case discussed in Section 3.2. In what follows in this and the subsequent sections any probability relationship

will be stated on the assumption that H is true, unless a statement to the contrary is explicitly made. Denote by γ the probability that the sequential probability ratio test will lead to the acceptance of H_1.[10] Clearly, if $H = H_0$, then $\gamma = \alpha$ and if $H = H_1$, then $\gamma = 1 - \beta$.

The probability γ can readily be derived on the basis of the general theory of cumulative sums given in [4]. Denote $\log \frac{f_1(x_i)}{f_0(x_i)}$ by z_i. Then $\{z_i\}$ $(i = 2, \cdots,$ ad inf.) is a sequence of independent random variables each having the same distribution. Denote by Z_j the sum of the first j elements of the sequence $\{z_i\}$ i.e.,

$$Z_j = z_1 + \cdots + z_j \qquad (j = 1, 2, \cdots, \text{ad inf.}) \tag{3.31}$$

For any relation R we shall denote by $P(R)$ the probability that R holds. For any random variable Y the symbol EY will denote the expected value of Y. Let n be the smallest positive integer for which either $Z_n \geq \log A$ or $Z_n \leq \log B$ holds. If $\log B < Z_m < \log A$ holds for $m = 1, 2, \cdots$, ad inf., we shall say that $n = \infty$. Obviously, n is the number of observations required by the sequential probability ratio test. As we have seen in Section 3.3, in practice we shall put $A = a(\alpha, \beta) = \frac{1 - \beta}{\alpha}$ and $B = b(\alpha, \beta) = \frac{\beta}{1 - \alpha}$. Since B must be less than A, we shall consider only values α and β for which $\frac{1 - \beta}{\alpha} > \frac{\beta}{1 - \alpha}$. This inequality is equivalent to $\alpha + \beta < 1$, which in turn implies that $B < 1$ and $A > 1$. Thus, in all that follows it will be assumed that $A > 1$ and $B < 1$. We shall also assume that the variance of z_i is not zero.

According to Lemma 1 in [4] the relation $P(n = \infty) = 0$ holds. Hence, the probability is equal to one that the sequential process will eventually terminate. This implies that the probability of accepting H_0 is equal to $1 - \gamma$.

Let z be a random variable whose distribution is equal to the common distribution of the variates z_i $(i = 1, 2, \cdots,$ ad inf.). Denote by $\varphi(t)$ the moment generating function of z, i.e.,

$$\varphi(t) = Ee^{zt}.$$

It was shown in [4] that under very mild restrictions on the distribution of z there exists exactly one real value h such that $h \neq 0$ and $\varphi(h) = 1$. Furthermore, it was shown in [4] (see equation (16) in [4]) that

$$Ee^{Z_n h} = 1. \tag{3.32}$$

Let E^* be the conditional expected value of $e^{Z_n h}$ under the restriction that H_0 is accepted, i.e., that $Z_n \leq \log B$, and let E^{**} be the conditional expected value of $e^{Z_n h}$ under the restriction that H_1 is accepted, i.e., that $Z_n \geq \log A$. Then we obtain from (3.32)

$$(1 - \gamma)E^* + \gamma E^{**} = 1 \tag{3.33}$$

[10] The probability that H_0 will be accepted is equal to $1 - \gamma$, as will be seen later.

Solving for γ we obtain

$$\gamma = \frac{1 - E^*}{E^{**} - E^*}. \tag{3.34}$$

If both the absolute value of Ez and the variance of z are small, which will be the case when $f_1(x)$ is near $f_0(x)$, E^* and E^{**} will be nearly equal to B^h and A^h, respectively. Hence, in this case a good approximation to γ is given by the expression

$$\bar{\gamma} = \frac{1 - B^h}{A^h - B^h}. \tag{3.35}$$

It is easy to verify that $h = 1$ if $H = H_0$, and $h = -1$ if $H = H_1$. The difference $\bar{\gamma} - \gamma$ approaches zero if both the mean and the variance of z converge to zero.

To judge the goodness of the approximation given by $\bar{\gamma}$, it is desirable to derive lower and upper limits for γ. Such limits for γ can be obtained by deriving lower and upper limits for E^* and E^{**}. First we consider the case when $h > 0$. Let ζ be a real variable restricted to values >1, and let ρ be a positive variable restricted to values <1. For any random variable Y and any relationship R we shall denote by $E(Y \mid R)$ the conditional expected value of Y under the restriction that R holds. It was shown in [4] that the following inequalities hold:[11]

$$B^h \left\{ \text{g.l.b.}_{\zeta} \zeta E\left(e^{hz} \mid e^{hz} \leq \frac{1}{\zeta}\right)\right\} \leq E^* \leq B^h \qquad (h > 0) \tag{3.36}$$

and

$$A^h \leq E^{**} \leq A^h \left\{ \text{l.u.b.}_{\rho} \rho E\left(e^{hz} \mid e^{hz} \geq \frac{1}{\rho}\right)\right\} \qquad (h > 0). \tag{3.37}$$

The symbol g.l.b._{ζ} stands for the greatest lower bound with respect to ζ, and the symbol l.u.b._{ρ} stands for least upper bound with respect to ρ. Putting

$$\text{g.l.b.}_{\zeta} \zeta E\left(e^{hz} \mid e^{hz} \leq \frac{1}{\zeta}\right) = \eta \tag{3.38}$$

and

$$\text{l.u.b.}_{\rho} \rho E\left(e^{hz} \mid e^{hz} \geq \frac{1}{\rho}\right) = \delta, \tag{3.39}$$

the inequalities (3.36) and (3.37) can be written as

$$B^h \eta \leq E^* \leq B^h \qquad (h > 0) \tag{3.40}$$

[11] See relations (23) and (26) in [4]. The notation used here is somewhat different from that in [4].

and

$$A^h \leq E^{**} \leq A^h\delta \cdot \qquad (h > 0). \tag{3.41}$$

Since $B < 1$ and $A > 1$, we see that $E^* < 1$ and $E^{**} > 1$ if $h > 0$. From this and the relations (3.34), (3.40) and (3.41) it follows easily that

$$\frac{1 - B^h}{\delta A^h - B^h} \leq \gamma \leq \frac{1 - \eta B^h}{A^h - \eta B^h} \qquad (h > 0) \tag{3.42}$$

If $h < 0$, limits for γ can be obtained as follows: Let $z' = -z$, $A' = \frac{1}{B}$, $B' = \frac{1}{A}$. Then $h' = -h > 0$ and $\gamma' = 1 - \gamma$. Thus, according to (3.42) we have

$$\frac{1 - (B')^{h'}}{\delta'(A')^{h'} - (B')^{h'}} \leq \gamma' \leq \frac{1 - \eta'(B')^{h'}}{(A')^{h'} - \eta'(B')^{h'}} \tag{3.43}$$

where δ' and η' are equal to the expressions we obtain from (3.38) and (3.39), respectively, by substituting h' for h and z' for z. Since η and δ depend only on the product $hz = h'z'$, we see that $\delta' = \delta$ and $\eta' = \eta$. Hence, we obtain from (3.43)

$$\frac{1 - A^h}{\delta B^h - A^h} \leq 1 - \gamma \leq \frac{1 - \eta A^h}{B^h - \eta A^h} \qquad (h < 0) \tag{3.44}$$

where δ and η are given by (3.38) and (3.39), respectively.

In Section 3.5 we shall calculate the value of η and δ for binomial and normal distributions. If the limits of γ, as given in (3.42) and (3.44), are too far apart, it may be desirable to determine the exact value of γ, or at least to find a closer approximation to γ than that given in (3.35). A solution of this problem is given in [4] (see section 7 of that paper). There the exact value of γ is derived when z can take only a finite number of integral multiples of a constant d. If z does not have this property, arbitrarily fine approximation to the value of γ can be obtained, since the distribution of z can be approximated to any desired degree by a discrete distribution of the type mentioned before if the constant d is chosen sufficiently small. The results obtained in [4] can be stated as follows: There is no loss of generality in assuming that $d = 1$, since the quantity d can be chosen as the unit of measurement. Thus, we shall assume that z takes only a finite number of integral values. Let g_1 and g_2 be two positive integers such that $P(z = -g_1)$ and $P(z = g_2)$ are positive and z can take only integral values $\geq -g_1$ and $\leq g_2$. Denote $P(z = i)$ by h_i. Then the moment generating function of z is given by

$$\varphi(t) = \sum_{i=-g_1}^{g_2} h_i e^{it}.$$

Put $u = e^t$ and let $u_1, \cdots u_g$ be the $g = g_1 + g_2$ roots of the equation of g-th degree

$$\sum_{i=-g_1}^{g_2} h_i u^i = 1. \tag{3.45}$$

Denote by $[a]$ the smallest integer $\geq \log A$, and by $[b]$ the largest integer $\leq \log B$. Then Z_n can take only the values

$$(3.46) \quad [b] - g_1 + 1, [b] - g_1 + 2, \cdots, [b], [a], [a] + 1, \cdots, [a] + g_2 - 1.$$

Denote the g different integers in (3.46) by $c_1, \cdots, c_g$, respectively. Let Δ be the determinant value of the matrix $|| u_i^{c_j} ||$ $(i, j = 1, \cdots, g)$ and let Δ_j be the determinant we obtain from Δ by substituting 1 for the elements in the j-th column. Then, if $\Delta \neq 0$, the probability that $Z_n = c_j$ is given by

$$(3.47) \qquad P(Z_n = c_j) = \frac{\Delta_j}{\Delta}.$$

Hence

$$(3.48) \qquad \gamma = P(Z_n \geq [a]) = \sum_j \frac{\Delta_j}{\Delta}$$

where the summation is to be taken over all vaues of j for which $c_j \geq [a]$.

3.5. *Calculation of δ and η for binomial and normal distributions.* Let X be a random variable which can take only the values 0 and 1. Let the probability that $X = 1$ be p_i if H_i is true $(i = 0, 1)$, and p if H is true. Denote $1 - p$ by q and $1 - p_i$ by q_i $(i = 0, 1)$. Then $f_i(1) = p_i$; $f_i(0) = q_i$, $f(1) = p$ and $f(0) = q$. It can be assumed without loss of generality that $p_1 > p_0$. The moment generating function of $z = \log \frac{f_1(x)}{f_0(x)}$ is given by

$$\varphi(t) = E\left(\frac{f_1(x)}{f_0(x)}\right)^t = p\left(\frac{p_1}{p_0}\right)^t + q\left(\frac{q_1}{q_0}\right)^t.$$

Let $h \neq 0$ be the value of t for which $\varphi(h) = 1$, i.e.,

$$p\left(\frac{p_1}{p_0}\right)^h + q\left(\frac{q_1}{q_0}\right)^h = 1.$$

First we consider the case when $h > 0$. It is clear that $e^{zh} = \left(\frac{f_1(x)}{f_0(x)}\right)^h > 1$ implies that $x = 1$. Hence $e^{zh} > 1$ implies that $e^{zh} = \left(\frac{f_1(1)}{f_0(1)}\right)^h = \left(\frac{p_1}{p_0}\right)^h$. From this and the definition of δ given in (3.39) it follows that

$$(3.49) \qquad \delta = \left(\frac{p_1}{p_0}\right)^h \qquad (h > 0).$$

Similarly, the inequality $e^{zh} < 1$ implies that $e^{zh} = \left(\frac{q_1}{q_0}\right)^h$. From this and the definition of η given in (3.38) it follows that

$$(3.50) \qquad \eta = \left(\frac{q_1}{q_0}\right)^h \qquad (h > 0).$$

If $h < 0$, it can be shown in a similar way that

$$\delta = \left(\frac{q_1}{q_0}\right)^h \qquad (h < 0) \tag{3.51}$$

and

$$\eta = \left(\frac{p_1}{p_0}\right)^h \qquad (h < 0). \tag{3.52}$$

Now we shall calculate the values of δ and η if X is normally distributed. Let

$$f_i(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta_i)^2} \qquad (i = 0, 1) \tag{3.53}$$

and

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}. \tag{3.54}$$

We can assume without loss of generality that $\theta_0 = -\Delta$ and $\theta_1 = \Delta$ where $\Delta > 0$, since this can always be achieved by a translation. Then

$$z = \log \frac{f_1(x)}{f_0(x)} = 2\Delta x. \tag{3.55}$$

The moment generating function of z is given by

$$\varphi(t) = e^{2\Delta\theta t + 2\Delta^2 t^2}. \tag{3.56}$$

Hence

$$h = -\frac{\theta}{\Delta}. \tag{3.57}$$

Substituting this value of h in (3.38) and (3.39) we obtain

$$\delta = \text{l.u.b.}_{\rho}\, \rho E\left(e^{-2\theta x} \mid e^{-2\theta x} \geq \frac{1}{\rho}\right) \tag{3.58}$$

and

$$\eta = \text{g.l.b.}_{\zeta}\, \zeta E\left(e^{-2\theta x} \mid e^{-2\theta x} \leq \frac{1}{\zeta}\right). \tag{3.59}$$

For any relation R let $P^*(R)$ denote the probability that the relation R holds calculated under the assumption that the distribution of x is normal with mean θ and variance unity. Furthermore, let $P^{**}(R)$ denote the probability that R holds if the distribution of x is normal with mean $-\theta$ and variance unity. Since $e^{-2\theta x}$ is equal to the ratio of the normal probability density function with mean $-\theta$ and variance unity to the normal probability density function with mean θ and variance unity, we see that

$$E\left(e^{-2\theta x} \mid e^{-2\theta x} \geq \frac{1}{\rho}\right) = \frac{P^{**}\left(e^{-2\theta x} \geq \frac{1}{\rho}\right)}{P^{*}\left(e^{-2\theta x} \geq \frac{1}{\rho}\right)}, \tag{3.60}$$

and

$$E\left(e^{-2\theta x} \mid e^{-2\theta x} \leq \frac{1}{\zeta}\right) = \frac{P^{**}\left(e^{-2\theta x} \leq \frac{1}{\zeta}\right)}{P^{*}\left(e^{-2\theta x} \leq \frac{1}{\zeta}\right)}. \tag{3.61}$$

It can easily be verified that the right hand side expressions in (3.60) and (3.61) have the same values for $\theta = \lambda$ as for $\theta = -\lambda$. Thus, also δ and η have the same values for $\theta = \lambda$ as for $\theta = -\lambda$. It will be, therefore, sufficient to compute δ and η for negative values of θ. Let $\theta = -\lambda$ where $\lambda > 0$. First we show that $\eta = \frac{1}{\delta}$. Clearly

$$\frac{\zeta P^{**}\left(e^{2\lambda x} \leq \frac{1}{\zeta}\right)}{P^{*}\left(e^{2\lambda x} \leq \frac{1}{\zeta}\right)} = \frac{\zeta P^{**}(e^{-2\lambda x} \geq \zeta)}{P^{*}(e^{-2\lambda x} \geq \zeta)} \qquad (1 \leq \zeta < \infty). \tag{3.62}$$

Putting $\zeta = \frac{1}{\rho}$ $(0 < \rho \leq 1)$ in (3.62) gives

$$\frac{\zeta P^{**}\left(e^{2\lambda x} \leq \frac{1}{\zeta}\right)}{P^{*}\left(e^{2\lambda x} \leq \frac{1}{\zeta}\right)} = \frac{P^{**}\left(e^{-2\lambda x} \geq \frac{1}{\rho}\right)}{\rho P^{*}\left(e^{-2\lambda x} \geq \frac{1}{\rho}\right)}. \tag{3.63}$$

Hence

$$\eta = \underset{\zeta}{\text{g.l.b.}} \left\{\frac{\zeta P^{**}\left(e^{2\lambda x} \leq \frac{1}{\zeta}\right)}{P^{*}\left(e^{2\lambda x} \leq \frac{1}{\zeta}\right)}\right\} = \frac{1}{\underset{\rho}{\text{l.u.b.}} \left\{\frac{\rho P^{*}\left(e^{-2\lambda x} \geq \frac{1}{\rho}\right)}{P^{**}\left(e^{-2\lambda x} \geq \frac{1}{\rho}\right)}\right\}}. \tag{3.64}$$

Because of the symmetry of the normal distribution, it is easily seen that

$$\underset{\rho}{\text{l.u.b.}} \left\{\frac{\rho P^{*}\left(e^{-2\lambda x} \geq \frac{1}{\rho}\right)}{P^{**}\left(e^{-2\lambda x} \geq \frac{1}{\rho}\right)}\right\} = \underset{\rho}{\text{l.u.b.}} \left\{\frac{\rho P^{**}\left(e^{2\lambda x} \geq \frac{1}{\rho}\right)}{P^{*}\left(e^{2\lambda x} \geq \frac{1}{\rho}\right)}\right\} = \delta.$$

Hence

$$\eta = \frac{1}{\delta}. \tag{3.65}$$

Now we shall calculate the value of δ. Denote $\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt$ by $G(x)$. Then

$$P^{**}\left(e^{2\lambda x} \geq \frac{1}{\rho}\right) = P^{**}\left(2\lambda x \geq \log\frac{1}{\rho}\right)$$

$$= P^{**}\left(x \geq \frac{1}{2\lambda}\log\frac{1}{\rho}\right) = G\left(2\lambda \log\frac{1}{\rho} - \lambda\right).$$

Similarly

$$P^{*}\left(e^{2\lambda x} \geq \frac{1}{\rho}\right) = P^{*}\left(x \geq \frac{1}{2\lambda}\log\frac{1}{\rho}\right) = G\left(\frac{1}{2\lambda}\log\frac{1}{\rho} + \lambda\right).$$

Denote $\frac{1}{2\lambda}\log\frac{1}{\rho}$ by u. Since ρ can vary from 0 to 1, u can take any value from 0 to ∞. Since $\rho = e^{-2\lambda u}$, we have

$$\delta = \underset{\rho}{\text{l.u.b.}}\left\{\frac{\rho P^{**}\left(e^{2\lambda x} \geq \frac{1}{\rho}\right)}{P^{*}\left(e^{2\lambda x} \geq \frac{1}{\rho}\right)}\right\} = \underset{u}{\text{l.u.b.}}\left\{e^{-2\lambda u}\frac{G(u-\lambda)}{G(u+\lambda)}\right\} \quad (0 \leq u < \infty). \tag{3.66}$$

We shall prove that

$$e^{-2u\lambda}\frac{G(u-\lambda)}{G(u+\lambda)} = \chi(u) \qquad \text{(say)} \tag{3.67}$$

is a monotonically decreasing function of u and consequently the maximum is at $u = 0$. For this purpose it suffices to show that the derivative of $\log \chi(u)$ is never positive. Now

$$\log \chi(u) = \log G(u-\lambda) - \log G(u+\lambda) - 2\lambda u. \tag{3.68}$$

Denote $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ by $\Phi(x)$. Since $\frac{d}{du}G(u) = -\Phi(u)$ it follows from (3.68) that

$$\frac{d}{du}\log\chi(u) = -\frac{\Phi(u-\lambda)}{G(u-\lambda)} + \frac{\Phi(u+\lambda)}{G(u+\lambda)} - 2\lambda. \tag{3.69}$$

It follows from the mean value theorem that the right hand side of (3.69) is never positive if $\frac{d}{du}\left(\frac{\Phi(u)}{G(u)}\right)$ is equal to or less than 1 for all values of u. Thus, we need merely to show that

$$\frac{d}{du}\left(\frac{\Phi(u)}{G(u)}\right) = \frac{\Phi'(u)G(u) - G'(u)\Phi(u)}{G^2(u)} \tag{3.70}$$

$$= \frac{\Phi'(u)G(u) + \Phi^2(u)}{G^2(u)} = \frac{\Phi^2(u)}{G^2(u)} - u\frac{\Phi(u)}{G(u)} \leq 1.$$

Denote $\frac{\Phi(u)}{G(u)}$ by y. The roots of the equation $y^2 - uy - 1 = 0$ are

$$y = \frac{u \pm \sqrt{u^2 + 4}}{2}.$$

Hence the inequality $y^2 - uy - 1 \leq 0$ holds if and only if

$$\frac{u - \sqrt{u^2 + 4}}{2} \leq y \leq \frac{u + \sqrt{u^2 + 4}}{2}.$$

Since y cannot be negative, this inequality is equivalent to

$$\frac{\Phi(u)}{G(u)} = y \leq \frac{u + \sqrt{u^2 + 4}}{2}. \tag{3.71}$$

Thus we have merely to prove (3.71). We shall show that (3.71) holds for all real values of u. Birnbaum has shown [5] that for $u > 0$

$$\frac{\sqrt{u^2 + 4} - u}{2}\Phi(u) \leq G(u). \tag{3.72}$$

Hence

$$\frac{\Phi(u)}{G(u)} \leq \frac{2}{\sqrt{u^2 + 4} - u} = \frac{\sqrt{u^2 + 4} + u}{2} \qquad (u > 0) \tag{3.73}$$

which proves (3.71) for $u > 0$. Now we prove (3.71) for $u < 0$. Let $u = -v$ where $v > 0$. Then it follows from (3.73) that

$$\frac{\Phi(v)}{G(v)} \leq \frac{2}{\sqrt{4 + v^2} - v}. \tag{3.74}$$

Taking reciprocals, we obtain from (3.74)

$$\frac{G(v)}{\Phi(v)} \geq \frac{\sqrt{4 + v^2} - v}{2}. \tag{3.75}$$

Since

$$\frac{G(u)}{\Phi(u)} \geq \frac{G(v) + 2v\Phi(v)}{\Phi(v)} = \frac{G(v)}{\Phi(v)} + 2v$$

we obtain from (3.75)

$$\frac{G(u)}{\Phi(u)} \geq \frac{\sqrt{v^2 + 4} + 3v}{2} \geq \frac{\sqrt{v^2 + 4} + v}{2}. \tag{3.76}$$

Taking reciprocals, we obtain

$$\frac{\Phi(u)}{G(u)} \leq \frac{2}{\sqrt{v^2+4}+v} = \frac{\sqrt{v^2+4}-v}{2} = \frac{\sqrt{u^2+4}+u}{2}.$$

Hence (3.71) is proved for all values of u and consequently δ is equal to the value of the expression (3.67) if we substitute 0 for u. Thus,

$$\delta = \frac{G(-\lambda)}{G(\lambda)}. \tag{3.77}$$

4. The Number of Observations Required by the Sequential Probability Ratio Test

4.1. *Expected number of observations necessary for reaching a decision.* As before, let

$$z = \log\frac{f_1(x)}{f_0(x)}, \qquad z_i = \log\frac{f_1(x_i)}{f_0(x_i)} \qquad (i = 1, 2, \cdots, \text{ad inf.})$$

and let n be the number of observations required by the sequential test, i.e., n is the smallest integer for which $Z_n = z_1 + \cdots + z_n$ is either $\geq \log A$ or $\leq \log B$. To determine the expected value $E(n)$ of n under any hypothesis H we shall consider a fixed positive integer N. The sum $Z_N = z_1 + \cdots + z_N$ can be split in two parts as follows

$$Z_N = Z_n + Z'_n \tag{4.1}$$

where $Z'_n = z_{n+1} + \cdots + z_N$ if $n \leq N$ and $Z'_n = Z_N - Z_n$ if $n > N$. Taking expected values on both sides of (4.1) we obtain

$$NEz = EZ_n + EZ'_n. \tag{4.2}$$

Since the probability that $n > N$ converges to zero as $N \to \infty$, and since $|Z'_n| < 2(\log A + |\log B|)$ if $n > N$, it can be seen that

$$\lim_{N=\infty} [EZ'_n - E(N-n)Ez] = 0. \tag{4.3}$$

From (4.2) and (4.3) it follows that

$$EZ_n = EnEz. \tag{4.4}$$

Hence

$$En = \frac{EZ_n}{Ez} \tag{4.5}$$

Let E^*Z_n be the conditional expected value of Z_n under the restriction that the sequential analysis leads to the acceptance of H_0, i.e. that $Z_n \leq \log B$. Similarly, let $E^{**}Z_n$ be the conditional expected value of Z_n under the restriction that H_1 is accepted, i.e., that $Z_n \geq \log A$. Since γ is the probability that $Z_n \geq \log A$, we have

$$EZ_n = (1-\gamma)E^*Z_n + \gamma E^{**}Z_n. \tag{4.6}$$

From (4.5) and (4.6) we obtain

$$En = \frac{(1 - \gamma)E^*Z_n + \gamma E^{**}Z_n}{Ez}. \tag{4.7}$$

The exact value of EZ_n, and therefore also the exact value of En, can be computed if z can take only integral multiples of a constant d, since in this case the exact probability distribution of Z_n was obtained (see equation (3.47)). If z does not satisfy the above restriction, it is still possible to obtain arbitrarily fine approximations to the value of EZ_n, since the distribution of z can be approximated to any desired degree by a discrete distribution of the type mentioned above if the constant d is chosen sufficiently small.

If both $|\, Ez \,|$ and the standard deviation of z are small, E^*Z_n is very nearly equal to $\log B$ and $E^{**}Z_n$ is very nearly equal to $\log A$. Hence in this case we can write

$$En \sim \frac{(1 - \gamma) \log B + \gamma \log A}{Ez}. \tag{4.8}$$

To judge the goodness of the approximation given in (4.8) we shall derive lower and upper limits for En by deriving lower and upper limits for E^*Z_n and $E^{**}Z_n$. Let r be a non-negative variable and let

$$\xi = \operatorname*{Max}_r E(z - r \,|\, z \geq r) \qquad (r \geq 0) \tag{4.9}$$

and

$$\xi' = \operatorname*{Min}_r E(z + r \,|\, z + r \leq 0). \qquad (r \geq 0) \tag{4.10}$$

It is easy to see that

$$\log A \leq E^{**}Z_n \leq \log A + \xi \tag{4.11}$$

and

$$\log B + \xi' \leq E^*Zn \leq \log B. \tag{4.12}$$

We obtain from (4.7), (4.11) and (4.12)

$$\frac{(1 - \gamma)(\log B + \xi') + \gamma \log A}{Ez} \leq En \leq \frac{(1 - \gamma) \log B + \gamma(\log A + \xi)}{Ez} \tag{4.13}$$

if $Ez > 0$

and

$$\frac{(1 - \gamma) \log B + \gamma(\log A + \xi)}{Ez} \leq En \leq \frac{(1 - \gamma)(\log B + \xi') + \gamma \log A}{Ez} \tag{4.14}$$

if $Ez < 0$.

4.2. *Calculation of the quantities ξ and ξ' for binomial and normal distributions.* Let X be a random variable which can take only the values 0 and 1. Let the probability that $X = 1$ be p_i if H_i is true ($i = 0, 1$), and p if H is true. Denote

$1 - p$ by q and $1 - p_i$ by q_i $(i = 0, 1)$. Then $f_i(1) = p_i$, $f_i(0) = q_i$, $f(1) = p$ and $f(0) = q$. It can be assumed without loss of generality that $p_1 > p_0$. It is clear that $\log \frac{f_1(x)}{f_0(x)} > 0$ implies that $x = 1$ and consequently $\log \frac{f_1(x)}{f_0(x)} = \log \frac{f_1(1)}{f_0(1)} = \log \frac{p_1}{p_0}$. Hence

$$\xi = \operatorname*{Max}_r E(z - r \mid z \geq r) = \log \frac{p_1}{p_0}. \tag{4.15}$$

Since $\log \frac{f_1(x)}{f_0(x)} \leq 0$ implies that $x = 0$, we have

$$\xi' = \operatorname*{Min}_r E(z + r \mid z + r \leq 0) = \log \frac{q_1}{q_0}. \tag{4.16}$$

Now we shall calculate the values ξ and ξ' if X is normally distributed. Let

$$f_i(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\theta_i)^2/2} \qquad (i = 0, 1)\ (\theta_1 > \theta_0)$$

and

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}.$$

We may assume without loss of generality that $\theta_0 = -\Delta$ and $\theta_1 = \Delta$ where $\Delta > 0$, since this can always be achieved by a translation. Then

$$z = \log \frac{f_1(x)}{f_0(x)} = 2\Delta x. \tag{4.17}$$

Denote $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \cdot x^2}$ by $\Phi(x)$ and $\frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{1}{2} \cdot t^2}\, dt$ by $G(x)$. Let $t = x - \theta$. Then $z = 2\Delta(t + \theta)$ and

$$\begin{aligned} E(z - r \mid z - r \geq 0) &= 2\Delta E\left(t + \theta - \frac{r}{2\Delta} \,\middle|\, t + \theta - \frac{r}{2\Delta} \geq 0\right) \\ &= \frac{2\Delta}{G(t_0)} \int_{t_0}^\infty (t - t_0)\Phi(t)\, dt = \frac{2\Delta}{G(t_0)} [-t_0 G(t_0) + \Phi(t_0)] \end{aligned} \tag{4.18}$$

where

$$t_0 = \frac{r}{2\Delta} - \theta. \tag{4.19}$$

In section 3.5 (see equation (3.70)) it was proved that $\frac{\Phi(t_0)}{G(t_0)} - t_0$ is a monotonically decreasing function of t_0. Hence the maximum of $E(z - r \mid z - r \geq 0)$ is reached for $r = 0$ and consequently

$$\xi = \frac{2\Delta}{G(-\theta)} [\theta G(-\theta) + \Phi(-\theta)] = 2\Delta\left[\theta + \frac{\Phi(-\theta)}{G(-\theta)}\right]. \tag{4.20}$$

Now we shall calculate ξ'. We have

$$\xi' = \operatorname*{Min}_r E(z + r \mid z + r \leq 0) = -\operatorname*{Max}_r E(-z - r \mid -z - r \geq 0)$$

$$= -2\Delta \operatorname*{Max}_r E\left(-x - \frac{r}{2\Delta} \,\middle|\, -x - \frac{r}{2\Delta} \geq 0\right). \tag{4.21}$$

Let $t = -x + \theta$ and $t_0 = \frac{r}{2\Delta} + \theta$. Then

$$E\left(-x - \frac{r}{2\Delta} \,\middle|\, -x - \frac{r}{2\Delta} \geq 0\right) = E(t - t_0 \mid t - t_0 \geq 0)$$

$$= \frac{1}{G(t_0)} \int_{t_0}^{\infty} (t - t_0)\Phi(t)\, dt = \frac{\Phi(t_0)}{G(t_0)} - t_0 . \tag{4.22}$$

Since this is a monotonically decreasing function of t_0, we have

$$\operatorname*{Max}_r E\left(-x - \frac{r}{2\Delta} \,\middle|\, -x - \frac{r}{2\Delta} \geq 0\right) = \frac{\Phi(\theta)}{G(\theta)} - \theta. \tag{4.23}$$

From (4.21) and (4.23) we obtain

$$\xi' = -2\Delta\left[\frac{\Phi(\theta)}{G(\theta)} - \theta\right]. \tag{4.24}$$

4.3. *Saving in the number of observations as compared with the current test procedure.* We consider the case of a normally distributed variate, such that

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta_0)^2}$$

and

$$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta_1)^2} \qquad (\theta_1 \neq \theta_0).$$

Denote by $n(\alpha, \beta)$ the minimum number of observations necessary in the current most powerful test for the probabilities of errors of the first and second kinds to be α and β, respectively, or less.

We shall calculate the number of observations required by the most powerful test. It can be assumed without loss of generality that $\theta_0 \leq \theta_1$. According to the current most powerful test procedure the hypothesis H_0 is accepted if $\bar{x} \leq d$ and the hypothesis H_1 is accepted if $\bar{x} > d$, where $\bar{x}$ is the arithmetic mean of the observations and d is a properly chosen constant. The probability of an error of the first kind is given by $G[\sqrt{n}(d - \theta_0)]$ and the probability of an error of the second kind is given by $1 - G[\sqrt{n}(d - \theta_1)]$ where $G(t) = \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-x^2/2}\, dx$. To equate these probabilities to α and β, respectively, the quantities d and n must satisfy

$$G[\sqrt{n}(d - \theta_0)] = \alpha \tag{4.25}$$

and

$$1 - G[\sqrt{n}(d - \theta_1)] = \beta. \tag{4.26}$$

Denote by λ_0 and λ_1 the values for which $G(\lambda_0) = \alpha$ and $G(\lambda_1) = 1 - \beta$. Then we have

$$\sqrt{n}(d - \theta_0) = \lambda_0 \tag{4.27}$$

and

$$\sqrt{n}(d - \theta_1) = \lambda_1. \tag{4.28}$$

Subtracting (4.27) from (4.28) we obtain

$$\sqrt{n}(\theta_0 - \theta_1) = \lambda_1 - \lambda_0. \tag{4.29}$$

From (4.29)

$$n = n(\alpha, \beta) = \frac{(\lambda_1 - \lambda_0)^2}{(\theta_0 - \theta_1)^2}. \tag{4.30}$$

If the expression on the right hand side of (4.30) is not an integer, $n(\alpha, \beta)$ is the smallest integer in excess.

In the sequential probability ratio test we put $A = a(\alpha, \beta) = \frac{1 - \beta}{\alpha}$ and $B = b(\alpha, \beta) = \frac{\beta}{1 - \alpha}$. Then the probability of an error of the first (second) kind cannot exceed $\alpha(\beta)$ except by a negligible amount. Let $A(\alpha, \beta)$ and $B(\alpha, \beta)$ be the values of A and B for which the probabilities of errors of the first and second kinds become exactly equal to α and β, respectively. It has been shown in Section 3.2 that $A(\alpha, \beta) \leq a(\alpha, \beta)$ and $B(\alpha, \beta) \geq b(\alpha, \beta)$. Thus, the expected values $E_1(n)$ and $E_0(n)$ are only increased by putting $A = a\ (\alpha, \beta)$ and $B = b\ (\alpha, \beta)$ instead of $A = A\ (\alpha, \beta)$ and $B = B\ (\alpha, \beta)$.

Consider the case where $|\ \theta_1 - \theta_0\ |$ is small so that the quantities ξ and ξ' can be neglected. Thus, we shall use the approximation (4.8). Since $\gamma = \alpha$ if $H = H_0$ and $\gamma = 1 - \beta$ if $H = H_1$, we obtain from (4.8)

$$E_1(n) = \frac{a^*}{E_1(z)} - \beta \frac{a^* + |\ b^*\ |}{E_1(z)} \tag{4.31}$$

and

$$E_0(n) = \frac{-b^*}{E_0(-z)} - \alpha \frac{-b^* + a^*}{E_0(-z)} \tag{4.32}$$

where $a^* = \log a(\alpha, \beta) = \log \frac{1 - \beta}{\alpha}$ and $b^* = \log b(\alpha, \beta) = \log \frac{\beta}{1 - \alpha}$. Since

$$E_1(z) = \tfrac{1}{2}(\theta_0 - \theta_1)^2 \tag{4.33}$$

and

$$E_0(-z) = \tfrac{1}{2}(\theta_0 - \theta_1)^2, \tag{4.34}$$

it follows from (4.30), (4.31) and (4.32) that $\frac{E_1(n)}{n(\alpha, \beta)}$ and $\frac{E_0(n)}{n(\alpha, \beta)}$ are independent of the parameters θ_0 and θ_1.

TABLE 1

Average percentage saving of sequential analysis, as compared with current most powerful test for testing mean of a normally distributed variate

A. When alternative hypothesis is true:

β \ α	.01	.02	.03	.04	.05
.01	58	60	61	62	63
.02	54	56	57	58	59
.03	51	53	54	55	55
.04	49	50	51	52	53
.05	47	49	50	50	51

B. When null hypothesis is true:

β \ α	.01	.02	.03	.04	.05
.01	58	54	51	49	47
.02	60	56	53	50	49
.03	61	57	54	51	50
.04	62	58	55	52	50
.05	63	59	55	53	51

The average saving of the sequential analysis as compared with the current method is $100\left(1 - \frac{E_1(n)}{n(\alpha, \beta)}\right)$ per cent if H_1 is true, and $100\left(1 - \frac{E_0(n)}{n(\alpha, \beta)}\right)$ per cent if H_0 is true. In Table 1 the expression $100\left(1 - \frac{E_1(n)}{n(\alpha, \beta)}\right)$ is shown in Panel A, and the expression $100\left(1 - \frac{E_0(n)}{n(\alpha, \beta)}\right)$ in Panel B, for several values of α and β. Because of the symmetry of the normal distribution, Panel B is obtained from Panel A simply by interchanging α and β.

As can be seen from the table, for the range of α and β from .01 to .05 (the range most frequently employed), the sequential process leads to an average saving of at least 47 per cent in the necessary number of observations as compared with the current procedure. The true saving is slightly greater than shown in the table, since $E_i(n)$ calculated under the condition that $A = a\ (\alpha, \beta)$ and $B = b\ (\alpha, \beta)$ is greater than $E_i(n)$ calculated under the condition that $A = A\ (\alpha, \beta)$ and $B = B\ (\alpha, \beta)$.

4.4. *The characteristic function, the moments and the distribution of the number of observations necessary for reaching a decision.* It was shown in [4] (see equation (15) in [4]) that the following fundamental identity holds

$$E\{e^{Z_n t}[\varphi(t)]^{-n}\} = 1 \qquad (\varphi(t) = Ee^{zt}) \tag{4.35}$$

for all points t of the complex plane for which $\varphi(t)$ exists and $|\varphi(t)| \geq 1$. The symbol n denotes the number of observations required by the sequential test, i.e., n is the smallest positive integer for which Z_n is either $\geq \log A$ or $\leq \log B$, and $\varphi(t)$ denotes the moment generating function of z.

On the basis of the identity (4.35) the exact characteristic function of n is derived in section 7 of [4] in the case when z can take only integral multiples of a constant. If the number of different values which Z_n can take is large, the calculation of the exact characteristic function is cumbersome, because a large number of simultaneous linear equations have to be solved. However, if $|Ez|$ and σ_z are small so that $|Z_n - \log A|$ (when $Z_n \geq \log A$) and $|Z_n - \log B|$ (when $Z_n \leq \log B$) can be neglected, the calculation of the characteristic function is much simpler, as was shown in [4]. We shall briefly state the results obtained in [4]. Let h be the real value $\neq 0$ for which $\varphi(h) = 1$. Furthermore let $t = t_1(\tau)$ and $t = t_2(\tau)$ be the roots of the equation in t

$$-\log \varphi(t) = \tau$$

such that $\lim_{\tau=0} t_1(\tau) = 0$ and $\lim_{\tau=0} t_2(\tau) = h$. Finally, let $\psi_1(\tau)$ the characteristic function of the conditional distribution of n under the restriction that $Z_n \geq \log A$, and $\psi_2(\tau)$ the characteristic function of the conditional distribution of n under the restriction that $Z_n \leq \log B$. Then, if $|Z_n - \log A|$ (when $Z_n \geq \log A$) and $|Z_n - \log B|$ (when $Z_n \leq \log B$) can be neglected, $\psi_1(\tau)$ and $\psi_2(\tau)$ are the solutions of the linear equations

$$\gamma\psi_1(\tau)A^{t_1(\tau)} + (1 - \gamma)\psi_2(\tau)B^{t_1(\tau)} = 1 \tag{4.36}$$

and

$$\gamma\psi_1(\tau)A^{t_2(\tau)} + (1 - \gamma)\psi_2(\tau)B^{t_2(\tau)} = 1 \tag{4.37}$$

where

$$\gamma = P(Z_n \geq \log A) = \frac{1 - B^h}{A^h - B^h}.$$

The characteristic function of the unconditional distribution of n is

$$\psi(\tau) = \gamma\psi_1(\tau) + (1 - \gamma)\psi_2(\tau). \tag{4.38}$$

As an illustration we shall determine $\psi_1(\tau)$, $\psi_2(\tau)$ and $\psi(\tau)$ when z has a normal distribution. Then we have

$$-\log \varphi(t) = -(Ez)t - \frac{\sigma_z^2}{2}t^2 = \tau.$$

Hence

$$h = -\frac{2Ez}{\sigma_z^2} \tag{4.39}$$

$$t_1(\tau) = \frac{1}{\sigma_z^2}\left(-Ez + \sqrt{(Ez)^2 - 2\sigma_z^2\tau}\right).$$

$$t_2(\tau) = \frac{1}{\sigma_z^2}\left(-Ez - \sqrt{(Ez)^2 - 2\sigma_z^2\tau}\right). \tag{4.40}$$

From (4.36), (4.37) and (4.38) we obtain

$$\gamma\psi_1(\tau) = \frac{B^{\sigma_2} - B^{\sigma_1}}{A^{\sigma_1}B^{\sigma_2} - A^{\sigma_2}B^{\sigma_1}}, \tag{4.41}$$

$$(1-\gamma)\psi_2(\tau) = \frac{A^{\sigma_1} - A^{\sigma_2}}{A^{\sigma_1}B^{\sigma_2} - A^{\sigma_2}B^{\sigma_1}} \tag{4.42}$$

and

$$\psi(\tau) = \frac{A^{\sigma_1} + B^{\sigma_2} - A^{\sigma_2} - B^{\sigma_1}}{A^{\sigma_1}B^{\sigma_2} - A^{\sigma_2}B^{\sigma_1}} \tag{4.43}$$

where

$$g_1 = \frac{1}{\sigma_z^2}\left(-Ez + \sqrt{(Ez)^2 - 2\sigma_z^2\tau}\right) \tag{4.44}$$

and

$$g_2 = \frac{1}{\sigma_z^2}\left(-Ez - \sqrt{(Ez)^2 - 2\sigma_z^2\tau}\right). \tag{4.45}$$

For any positive integer r the r-th moment of n i.e., $E(n^r)$ is equal to the r-th derivative of $\psi(\tau)$ taken at $\tau = 0$. Let $E^*(n^r)$ be the conditional expected value of n^r under the restriction that $Z_n \leq \log B$, and let $E^{**}(n^r)$ be the conditional expected value of n^r under the restriction that $Z_n \geq \log A$. Then

$$E^*(n^r) = \frac{d^r\psi_2(\tau)}{d\tau^r}\bigg|_{\tau=0} \quad \text{and} \quad E^{**}(n^r) = \frac{d^r\psi_1(\tau)}{d\tau^r}\bigg|_{\tau=0}. \tag{4.46}$$

It may be of interest to note that $\dfrac{d^r\psi_k(\tau)}{d\tau^r}\bigg|_{\tau=0}$ $(k = 1, 2)$ and therefore also the moments of n can be obtained from the identity (4.35) directly by successive differentiation. In fact, the identity (4.35) can be written as (neglecting the excess of Z_n over the boundaries $\log A$ and $\log B$)

$$\gamma A^t\psi_1[-\log\varphi(t)] + (1-\gamma)B^t\psi_2[-\log\varphi(t)] = 1. \tag{4.47}$$

Taking the first r derivatives of (4.47) with respect to t at $t = 0$ and $t = h$ we obtain a system of $2r$ linear equations in the $2r$ unknowns $\frac{d^j\psi_k(\tau)}{d\tau^j}\Big|_{\tau=0}$ $(k = 1, 2; j = 1, \cdots, r)$ from which these unknowns can be determined. For example, $\frac{d\psi_k(\tau)}{d\tau}\Big|_{\tau=0}$ $(k = 1, 2)$ can be determined as follows: Taking the first derivative of (4.47) with respect to t and denoting $\frac{d^r\psi_k(\tau)}{d\tau^r}$ by $\psi_k^{(r)}(\tau)$ we obtain

$$\begin{aligned}\gamma(\log A)A^t\psi_1[-\log\varphi(t)] - \gamma A^t\frac{\varphi'(t)}{\varphi(t)}\psi_1^{(1)}[-\log\varphi(t)] \\ + (1-\gamma)(\log B)B^t\psi_2[-\log\varphi(t)] \\ - (1-\gamma)B^t\frac{\varphi'(t)}{\varphi(t)}\psi_2^{(1)}[-\log\varphi(t)] = 0.\end{aligned} \tag{4.48}$$

Putting $t = 0$ and $t = h$ we obtain the equations

$$\gamma\log A - \gamma\frac{\varphi'(0)}{\varphi(0)}\psi_1^{(1)}(0) + (1-\gamma)\log B - (1-\gamma)\frac{\varphi'(0)}{\varphi(0)}\psi_2^{(1)}(0) = 0 \tag{4.49}$$

and

$$\begin{aligned}\gamma(\log A)A^h - \gamma A^h\frac{\varphi'(h)}{\varphi(h)}\psi_1^{(1)}(0) \\ + (1-\gamma)(\log B)B^h - (1-\gamma)B^h\frac{\varphi'(h)}{\varphi(h)}\psi_2^{(1)}(0) = 0\end{aligned} \tag{4.50}$$

from which $\psi_1^{(1)}(0)$ and $\psi_2^{(1)}(0)$ can be determined.

The distribution of n can be obtained by inverting the characteristic function of $\psi(\tau)$. This was done in [4] (neglecting the excess of Z_n over $\log A$ and $\log B$) in the case when z is normally distributed. The results obtained in [4] can be briefly stated as follows: If $B = 0$, or if $B > 0$ and $A = \infty$, the distribution of n is a simple elementary function. If $B = 0$ and $Ez > 0$, the distribution of $m = \frac{1}{2\sigma_z^2}(Ez)^2n$ is given by

$$F(m)\,dm = \frac{c}{2\Gamma(\frac{1}{2})m^{\frac{1}{2}}}e^{-c^2/4m-m+c}\,dm \qquad (0 \le m < \infty) \tag{4.51}$$

where

$$c = \frac{1}{\sigma_z^2}(Ez)\log A. \tag{4.52}$$

If $B > 0$, $A = \infty$ and $Ez < 0$ the distribution of $m = \frac{1}{2\sigma_z^2}(Ez)^2n$ is given by the expression we obtain from (4.51) if we substitute $\frac{1}{\sigma_z^2}(Ez)\log B$ for c.

If $B > 0$ and $A < \infty$, the distribution of m is given by an infinite series where each term is of the form (4.51) (see equation (76) in [4]).

Since m is a discrete variable, it may seem paradoxical that we obtained a probability density function for m. However, the explanation lies in the fact that we neglected the excess of Z_n over $\log A$ and $\log B$ which is zero only in the limiting case when Ez and σ_z approach zero.

The distribution of m given in (4.51) can be used as a good approximation to the exact distribution of m even if $B > 0$, provided that the probability that $Z_n \geq \log A$ is nearly equal to 1.

It was pointed out in [4] that if $| Ez |$ and σ_z are sufficiently small, the distribution of n determined under the assumption that z is normally distributed will be a good approximation to the exact distribution of n even if z is not normally distributed.

4.5. *Lower limit of the probability that the sequential process will terminate with a number of trials less than or equal to a given number.* Let $P_i(n_0)$ be the probability that the sequential process will terminate at a value $n \leq n_0$, calculated under H_i $(i = 0, 1)$. Let

$$\overline{P}_0(n_0) = P_0\left[\sum_{\alpha=1}^{n_0} z_\alpha \leq \log B\right] \tag{4.53}$$

and

$$\overline{P}_1(n_0) = P_1\left[\sum_{\alpha=1}^{n_0} z_\alpha \geq \log A\right]. \tag{4.54}$$

It is clear that

$$\overline{P}_i(n_0) \leq P_i(n_0) \qquad (i = 0, 1). \tag{4.55}$$

For calculating $\overline{P}_i(n_0)$ we shall assume that n_0 is sufficiently large so that $\sum_{\alpha=1}^{n_0} z_\alpha$ can be regarded as normally distributed. Let $G(\lambda)$ be defined by

$$G(\lambda) = \frac{1}{\sqrt{2\pi}} \int_\lambda^\infty e^{-\frac{1}{2}t^2}\, dt. \tag{4.56}$$

Furthermore, let

$$\lambda_1(n_0) = \frac{\log A - n_0 E_1(z)}{\sqrt{n_0}\,\sigma_1(z)} \tag{4.57}$$

and

$$\lambda_0(n_0) = \frac{\log B - n_0 E_0(z)}{\sqrt{n_0}\,\sigma_0(z)} \tag{4.58}$$

where $\sigma_i(z)$ is the standard deviation of z under H_i. Then

$$\overline{P}_1(n_0) = G[\lambda_1(n_0)] \tag{4.59}$$

and

$$\overline{P}_0(n_0) = 1 - G[\lambda_0(n_0)]. \tag{4.60}$$

Hence we have the inequalities

$$P_1(n_0) \geq G[\lambda_1(n_0)] \tag{4.61}$$

and

$$P_0(n_0) \geq 1 - G[\lambda_0(n_0)]. \tag{4.62}$$

Putting $\log A = \log \frac{1-\beta}{\alpha}$ and $\log B = \log \frac{\beta}{1-\alpha}$, Table 2 shows the values of $\overline{P}_1(n_0)$ and $\overline{P}_0(n_0)$ corresponding to different pairs (α, β) and different values of n_0. In these calculations it has been assumed that the distribution under H_0 is a normal distribution with mean zero and unit variance, and the distribution under H_1 is a normal distribution with mean θ and unit variance. For each pair (α, β) the value of θ was determined so that the number of observations required by the current most powerful test of strength (α, β) is equal to 1000.

TABLE 2

Lower bound of the probability that a sequential analysis will terminate within various numbers of trials, when the most powerful current test requires exactly 1000 trials*

Number of trials	$\alpha = .01$ and $\beta = .01$		$\alpha = .01$ and $\beta = .05$		$\alpha = .05$ and $\beta = .05$	
	Alternative hypothesis true	Null hypothesis true	Alternative hypothesis true	Null hypothesis true	Alternative hypothesis true	Null hypothesis true
1000	.910	.910	.799	.891	.773	.773
1200	.950	.950	.871	.932	.837	.837
1400	.972	.972	.916	.957	.883	.883
1600	.985	.985	.946	.972	.915	.915
1800	.991	.991	.965	.982	.938	.938
2000	.995	.995	.977	.989	.955	.955
2200	.997	.997	.985	.993	.967	.967
2400	.999	.999	.990	.995	.976	.976
2600	.999	.999	.994	.997	.982	.982
2800	1.00	1.00	.996	.998	.987	.987
3000	1.00	1.00	.997	.999	.990	.990

* The probabilities given are lower bounds for the true probabilities. They relate to a test of the mean of a normally distributed variate, the difference between the null and alternative hypothesis being adjusted for each pair of values of α and β so that the number of trials required under the most powerful current test is exactly 1000.

4.6. *Truncated sequential analysis.* In some applications a definite upper bound for the number of observations may be desirable. Thus, a certain integer n_0 is chosen so that if the sequential process does not lead to a final decision for $n \leq n_0$, a new rule is given for the acceptance or rejection of H_0 at the stage $n = n_0$.

A simple and reasonable rule for the acceptance or rejection of H_0 at the stage $n = n_0$ can be given as follows: If $\sum_{\alpha=1}^{n_0} z_\alpha \leq 0$ we accept H_0 and if $\sum_{\alpha=1}^{n_0} z_\alpha > 0$

we accept H_1. By thus truncating the sequential process we change, however, the probabilities of errors of the first and second kinds. Let α and β be the probabilities of errors of the first and second kinds, respectively, if the sequential test is not truncated. Let $\alpha(n_0)$ and $\beta(n_0)$ be the probabilities of errors of the first and second kinds if the test is truncated at $n = n_0$. We shall derive upper bounds for $\alpha(n_0)$ and $\beta(n_0)$.

First we shall derive an upper bound for $\alpha(n_0)$. Let $\rho_0(n_0)$ be the probability (under the null hypothesis) that the following three conditions are simultaneously fulfilled:

(i) $\log B < \sum_{\alpha=1}^{n} z_\alpha < \log A \qquad \text{for } n = 1, \cdots, n_0 - 1$

(ii) $0 < \sum_{\alpha=1}^{n_0} z_\alpha < \log A$

(iii) continuing the sequential process beyond n_0, it terminates with the acceptance of H_0.

It is clear that

$$\alpha(n_0) \leq \alpha + \rho_0(n_0). \tag{4.63}$$

Let $\bar{\rho}_0(n_0)$ be the probability (under the null hypothesis) that $0 < \sum_{\alpha=1}^{n_0} z_\alpha < \log A$. Then obviously

$$\rho_0(n_0) \leq \bar{\rho}_0(n_0)$$

and consequently

$$\alpha(n_0) \leq \alpha + \bar{\rho}_0(n_0). \tag{4.64}$$

Let $\rho_1(n_0)$ be the probability under the alternative hypothesis that the following three conditions are simultaneously fulfilled:

(i) $\log B < \sum_{\alpha=1}^{n} z_\alpha < \log A \qquad \text{for } n = 1, \cdots, n_0 - 1$

(ii) $\log B < \sum_{\alpha=1}^{n_0} z_\alpha \leq 0$

(iii) continuing the sequential process beyond n_0, it terminates with the acceptance of H_1.

It is clear that

$$\beta(n_0) \leq \beta + \rho_1(n_0). \tag{4.65}$$

Let $\bar{\rho}_1(n_0)$ be the probability (under the alternative hypothesis) that $\log B < \sum_{\alpha=1}^{n_0} z_\alpha \leq 0$. Then $\rho_1(n_0) \leq \bar{\rho}_1(n_0)$ and consequently

$$\beta(n_0) \leq \beta + \bar{\rho}_1(n_0). \tag{4.66}$$

Let

$$\nu_1 = \frac{-n_0 E_0(z)}{\sqrt{n_0}\,\sigma_0(z)}$$

$$\nu_2 = \frac{\log A - n_0 E_0(z)}{\sqrt{n_0}\,\sigma_0(z)}, \qquad \nu_3 = \frac{-n_0 E_1(z)}{\sqrt{n_0}\,\sigma_1(z)}, \qquad \nu_4 = \frac{\log B - n_0 E_1(z)}{\sqrt{n_0}\,\sigma_1(z)}$$

where $\sigma_i(z)$ is the standard deviation of z under H_i $(i = 0, 1)$. Then

$$\bar{\rho}_0(n_0) = G(\nu_1) - G(\nu_2) \tag{4.67}$$

and

$$\bar{\rho}_1(n_0) = G(\nu_4) - G(\nu_3). \tag{4.68}$$

From (4.64), (4.66), (4.67) and (4.68) we obtain

$$\alpha(n_0) \leq \alpha + G(\nu_1) - G(\nu_2) \tag{4.69}$$

and

$$\beta(n_0) \leq \beta + G(\nu_4) - G(\nu_3). \tag{4.70}$$

The upper bounds given in (4.69) and (4.70) may considerably exceed $\alpha(n_0)$ and $\beta(n_0)$, respectively. It would be desirable to find closer limits.

Table 3 shows the values of the upper bounds of $\alpha(n_0)$ and $\beta(n_0)$ given by formulas (4.69) and (4.70) corresponding to different pairs (α, β) and different values of n_0. In these calculations we have put $\log A = \log \frac{1-\beta}{\alpha}$, $\log B = \log \frac{\beta}{1-\alpha}$ and assumed that the distribution under H_0 is a normal distribution with mean zero and unit variance, and the distribution under H_1 is a normal distribution with mean θ and unit variance. For each pair (α, β) the value of θ has been determined so that the number of observations required by the current most powerful test of strength (α, β) is equal to 1000.

It seems to the author that the upper limits given in (4.69) and (4.70) are considerably above the true $\alpha(n_0)$ and $\beta(n_0)$ respectively, when n_0 is not much higher than the value of n needed for the current most powerful test.

4.7. *Efficiency of the sequential probability ratio test.* Let S be any sequential test for which the probability of an error of the first kind is α, the probability of an error of the second kind is β and the probability that the test procedure will eventually terminate is one. Let S' be the sequential probability ratio test whose strength is equal to that of S. We shall prove that the sequential probability ratio test is an optimum test, i.e., that $E_i(n \mid S) \geq E_i(n \mid S')$ $(i = 0, 1)$, if for S' the excess of Z_n over $\log A$ and $\log B$ can be neglected. This excess is exactly zero if z can take only the values d and $-d$ and if $\log A$ and $\log B$ are integral multiples of d. In any other case the excess will not be identically zero. However, if $|Ez|$ and σ_z are sufficiently small, the excess of Z_n over $\log A$ and $\log B$ is negligible.

For any random variable u we shall denote by $E_i^*(u \mid S)$ the conditional expected value of u under the hypothesis H_i $(i = 0, 1)$ and under the restriction

that H_0 is accepted. Similarly, let $E_i^{**}(u \mid S)$ be the conditional expected value of u under the hypothesis H_i $(i = 0, 1)$ and under the restriction that H_1 is accepted. In the notations for these expected values the symbol S stands for

TABLE 3

Effect on risks of error of truncating a sequential analysis at a predetermined number of trials*

Number of trials	$\alpha = .01$ and $\beta = .01$		$\alpha = .01$ and $\beta = .05$		$\alpha = .05$ and $\beta = .05$	
	Upper bound of effective α	Upper bound of effective β	Upper bound of effective α	Upper bound of effective β	Upper bound of effective α	Upper bound of effective β
1000	.020	.020	.033	.070	.095	.095
1200	.015	.015	.024	.063	.082	.082
1400	.013	.013	.019	.058	.072	.072
1600	.012	.012	.016	.055	.066	.066
1800	.011	.011	.014	.053	.062	.062
2000	.010	.010	.012	.052	.058	.058
2200	.010	.010	.012	.051	.056	.056
2400	.010	.010	.011	.051	.055	.055
2600	.010	.010	.011	.051	.053	.053
2800	.010	.010	.010	.050	.053	.053
3000	.010	.010	.010	.050	.052	.052

* If the sequential analysis is based on the values α and β shown, but a decision is made at n_0 trials even when the normal sequential criteria would require a continuation of the process, the realized values of α and β will not exceed the tabular entries. The table relates to a test of the mean of a normally distributed variate, the difference between the null and alternative hypotheses being adjusted for each pair (α,β) so that the number of trials required by the current test is 1000.

the sequential test used. Denote by $Q_i(S)$ the totality of all samples for which the test S leads to the acceptance of H_i. Then we have

$$E_0^*\left(\frac{p_{1n}}{p_{0n}} \mid S\right) = \frac{P_1[Q_0(S)]}{P_0[Q_0(S)]} = \frac{\beta}{1-\alpha} \tag{4.71}$$

$$E_0^{**}\left(\frac{p_{1n}}{p_{0n}} \mid S\right) = \frac{P_1[Q_1(S)]}{P_0[Q_1(S)]} = \frac{1-\beta}{\alpha} \tag{4.72}$$

$$E_1^*\left(\frac{p_{0n}}{p_{1n}} \mid S\right) = \frac{P_0[Q_0(S)]}{P_1[Q_0(S)]} = \frac{1-\alpha}{\beta} \tag{4.73}$$

and

$$E_1^{**}\left(\frac{p_{0n}}{p_{1n}} \mid S\right) = \frac{P_0[Q_1(S)]}{P_1[Q_1(S)]} = \frac{\alpha}{1-\beta}. \tag{4.74}$$

To prove the efficiency of the sequential probability ratio test, we shall first derive two lemmas.

LEMMA 1. *For any random variable u the inequality*

$$e^{Eu} \leq Ee^{u} \tag{4.75}$$

holds.

PROOF: Inequality (4.75) can be written as

$$1 \leq Ee^{u'} \tag{4.76}$$

where $u' = u - Eu$. Lemma 1 is proved if we show that (4.76) holds for any random variable u' with zero mean. Expanding $e^{u'}$ in a Taylor series around $u' = 0$, we obtain

$$e^{u'} = 1 + u' + \tfrac{1}{2}u'^2 e^{\xi(u')} \quad \text{where } 0 \leq \xi(u') \leq u'. \tag{4.77}$$

Hence

$$Ee^{u'} = 1 + \tfrac{1}{2}E[u'^2 e^{\xi(u')}] \geq 1 \tag{4.78}$$

and Lemma 1 is proved.

LEMMA 2. *Let S be a sequential test such that there exists a finite integer N with the property that the number n of observations required for the test is $\leq N$. Then*

$$E_i(n \mid S) = \frac{E_i\left(\log \frac{p_{1n}}{p_{0n}} \middle| S\right)}{E_i(z)} \qquad (i = 0, 1). \tag{4.79}$$

The proof is omitted, since it is essentially the same as that of equation (4.5) for the sequential probability ratio test.

On the basis of Lemmas 1 and 2 we shall be able to derive the following theorem.

THEOREM. *Let S be any sequential test for which the probability of an error of the first kind is α, the probability of an error of the second kind is β and the probability that the test procedure will eventually terminate is equal to one. Then*

$$E_0(n \mid S) \geq \frac{1}{E_0(z)}\left[(1 - \alpha)\log\frac{\beta}{1 - \alpha} + \alpha \log \frac{1 - \beta}{\alpha}\right] \tag{4.80}$$

and

$$E_1(n \mid S) \geq \frac{1}{E_1(z)}\left[\beta \log\frac{\beta}{1 - \alpha} + (1 - \beta)\log \frac{1 - \beta}{\alpha}\right] \tag{4.81}$$

PROOF: First we shall prove the theorem in the case when there exists a finite integer N such that n never exceeds N. According to Lemma 2 we have

$$\begin{aligned} E_0(n \mid S) &= \frac{1}{E_0(z)} E_0\left(\log \frac{p_{1n}}{p_{0n}} \middle| S\right) \\ &= \frac{1}{E_0(z)}\left[(1 - \alpha)E_0^*\left(\log \frac{p_{1n}}{p_{0n}} \middle| S\right) + \alpha E_0^{**}\left(\log \frac{p_{1n}}{p_{0n}} \middle| S\right)\right] \end{aligned} \tag{4.82}$$

and

$$E_1(n \mid S) = \frac{1}{E_1(z)} E_1\left(\log \frac{p_{1n}}{p_{0n}} \middle| S\right) = \frac{1}{E_1(z)}\left[\beta E_1^*\left(\log \frac{p_{1n}}{p_{0n}} \middle| S\right) + (1 - \beta) E_1^{**}\left(\log \frac{p_{1n}}{p_{0n}} \middle| S\right)\right]. \tag{4.83}$$

From equations (4.71)–(4.74) and Lemma 1 we obtain the inequalities

$$E_0^*\left(\log \frac{p_{1n}}{p_{0n}} \middle| S\right) \leq \log \frac{\beta}{1 - \alpha} \tag{4.84}$$

$$E_0^{**}\left(\log \frac{p_{1n}}{p_{0n}} \middle| S\right) \leq \log \frac{1 - \beta}{\alpha} \tag{4.85}$$

$$E_1^*\left(\log \frac{p_{0n}}{p_{1n}} \middle| S\right) = -E_1^*\left(\log \frac{p_{1n}}{p_{0n}} \middle| S\right) \leq \log \frac{1 - \alpha}{\beta} \tag{4.86}$$

and

$$E_1^{**}\left(\log \frac{p_{0n}}{p_{1n}} \middle| S\right) = -E_1^{**}\left(\log \frac{p_{1n}}{p_{0n}} \middle| S\right) \leq \log \frac{\alpha}{1 - \beta}. \tag{4.87}$$

Since $E_0(z) < 0$, (4.80) follows from (4.82), (4.84) and (4.85). Similarly, since $E_1(z) > 0$, (4.81) follows from (4.83), (4.86) and (4.87). This proves the theorem when there exists a finite integer N such that $n \leq N$.

To prove the theorem for any sequential test S of strength (α, β), for any positive integer N let S_N be the sequential test we obtain by truncating S at the N-th observation if no decision is reached before the N-th observation. Let (α_N, β_N) be the strength of S_N. Then we have

$$E_0(n \mid S) \geq E_0(n \mid S_N) \geq \frac{1}{E_0(z)}\left[(1 - \alpha_N) \log \frac{\beta_N}{1 - \alpha_N} + \alpha_N \log \frac{1 - \beta_N}{\alpha_N}\right] \tag{4.88}$$

and

$$E_1(n \mid S) \geq E_1(n \mid S_N) \geq \frac{1}{E_1(z)}\left[\beta_N \log \frac{\beta_N}{1 - \alpha_N} + (1 - \beta_N) \log \frac{1 - \beta_N}{\alpha_N}\right. \tag{4.89}$$

Since $\lim_{N=\infty} \alpha_N = \alpha$ and $\lim_{N=\infty} \beta_N = \beta$, inequalities (4.80) and (4.81) follow from (4.88) and (4.89). Hence the proof of the theorem is completed.

If for the sequential probability ratio test S' the excess of the cumulative sum Z_n over the boundaries $\log A$ and $\log B$ is zero, $E_0(n \mid S')$ is exactly equal to the right hand side member of (4.80) and $E_1(n \mid S')$ is exactly equal to the right hand side member of (4.81). Hence, in this case S' is exactly an optimum test. If both $\mid Ez \mid$ and σ_z are small, also the expected value of the excess over the boundaries will be small and, therefore, $E_0(n \mid S')$ and $E_1(n \mid S')$ will be only slightly larger than the right hand members of (4.80) and (4.81), respectively. Thus, in such a case the sequential probability ratio test is, if not exactly, very nearly an optimum test.[12]

[12] The author conjectures that the sequential probability ratio test is exactly an optimum test even if the excess of Z_n over the boundaries is not zero. However, he did not succeed in proving this.

PART II. SEQUENTIAL TEST OF A SIMPLE OR COMPOSITE HYPOTHESIS AGAINST A SET OF ALTERNATIVES

In Part I we have dealt with the problem of testing a simple hypothesis H_0 against a single alternative H_1. Here we shall consider the problem of testing a simple or composite hypothesis against a set of infinitely many alternatives. By a simple hypothesis we mean a hypothesis which specifies uniquely the probability distribution of the random variable x under consideration. A hypothesis is called composite, if it is not simple.

5. Test of a Simple Hypothesis Against One-sided Alternatives

5.1. *General remarks.* Let $f(x, \theta)$ be the probability density function of a random variable X, where θ is an unknown parameter. Suppose that it is required to test the simple hypothesis that $\theta = \theta_0$ and that the alternative values of θ are restricted to values $\theta > \theta_0$. Assume that it is desired to have a sequential test such that the probability of an error of the first kind is equal to a given α.

The probability of an error of the second kind is no longer a single value, but is a function of the true value of θ. If $f(x, \theta)$ is a continuous function of x and θ, the probability of an error of the second kind will be arbitrarily near $1 - \alpha$ if the true value of θ is sufficiently near θ_0. Hence, if α is small, the probability of an error of the second kind is necessarily large when the true value of θ is very near θ_0. In most practical applications we do not care if the probability of an error of the second kind is high when the true value of θ is very near θ_0, since in this case the error committed by accepting θ_0 is usually of very little importance. However, there will be a value $\theta_1 > \theta_0$ such that we wish the probability of an error of the second kind to be less than or equal to a given small positive value β whenever the true value of θ is greater than or equal to θ_1.

In this case we can proceed as follows: Consider the single alternative hypothesis H_1 that $\theta = \theta_1$. Construct a sequential test for testing $\theta = \theta_0$ against the single alternative H_1 such that the probability of an error of the first kind is α and the probability of an error of the second kind, i.e., the probability of accepting θ_0 when θ_1 is true, is β. If this sequential test has the further property that the probability of an error of the second kind is less than or equal to β whenever the true value of θ is greater than θ_1, then this sequential test provides a satisfactory solution of the problem of testing the hypothesis that $\theta = \theta_0$ against the set of alternatives $\theta > \theta_0$.

In most of the important cases occurring in practice, such as when X has a normal, binomial, or Poisson distribution, etc., the sequential probability ratio test for testing the hypothesis that $\theta = \theta_0$ against a single alternative θ_1 ($\theta_1 > \theta_0$) satisfies the condition that the probability of an error of the second kind is a monotonically decreasing function of θ in the domain $\theta > \theta_0$. Thus, in all these cases the sequential probability ratio test for testing the hypothesis that $\theta = \theta_0$ against a properly chosen alternative θ_1 provides a satisfactory solution of our problem.

The case in which the alternative values of θ are restricted to values less than θ_0 is entirely analogous to that in which the alternatives are restricted to values greater than θ_0, and need not be discussed separately.

It should be pointed out that the test procedure for testing $\theta = \theta_0$ against alternatives $\theta > \theta_0$, as described in this section, is also suitable for testing the composite hypothesis that $\theta \leq \theta_0$, provided that the probability of rejecting the null hypothesis is $\leq \alpha$ whenever the true value of θ is $\leq \theta_0$. This condition is fulfilled, for instance, when X has a normal, binomial or Poisson distribution.

5.2. *Application to binomial distributions.* 5.2.1. *Statement of the problem.* The case of a binomial distribution arises when the result of a single observation is a classification into one of two categories. For example, this is the situation in acceptance inspection of manufactured products, if each unit inspected is classified into one of the two categories, non-defective and defective. Let p denote the probability that an item belongs to a given category. The value of p is usually unknown. We shall deal here with the problem of testing the hypothesis that p does not exceed a given value p' against the alternative possibility that $p > p'$.

Since acceptance inspection of manufactured products is perhaps the most important and widest field of application of such a test procedure, we shall, in continuing the discussion, use the terminology of acceptance inspection. This, of course, does not mean that the test procedure is not applicable to other cases. Suppose that a lot containing a large number of units is submitted for sampling inspection. Let p denote the proportion of defective units contained in the lot. The probability that a unit drawn at random from the lot will be defective is equal to p. If m units are drawn at random from the lot, the probability that there will be d defectives among them is given by[13]

$$\frac{m!}{d!(m-d)!}\, p^d(1-p)^{m-d} \qquad (d = 0, 1, \cdots, m). \tag{5.1}$$

The probability distribution as given in (5.1) is called a binomial distribution.

The purpose of sampling inspection is to decide whether the lot should be accepted or rejected. It is clear that for high values of p we want to reject the lot and for low values of p we want to accept the lot. Thus, it will be possible to specify a particular value of p, say p', so that if $p \leq p'$ we wish to accept the lot, and if $p > p'$ we wish to reject the lot. Thus, our problem is to devise a proper sampling inspection plan for testing the hypothesis that $p \leq p'$.

5.2.2. *Tolerated risks for making a wrong decision.* No sampling inspection plan can guarantee that the correct decision will always be made, i.e., that the lot will always be accepted when $p \leq p'$ and the lot will always be rejected when $p > p'$, unless the lot is inspected completely. A complete inspection is usually

[13] Formula (5.1) is exact only if the lot contains infinitely many units. While the lot is always finite in practice, we shall assume that m is small as compared with the lot size so that formula (5.1) can be used.

rather uneconomical and one is willing to take some risk of making a wrong decision if this permits a reduction in the amount of inspection. Hence, recommendations as to the proper choice of a sampling inspection plan can be made only after the risks that can be tolerated have been stated.

If p is equal to the marginal value p', we may say that it is indifferent to us whether the lot is accepted or rejected. If $p < p'$ we prefer acceptance and this preference is the stronger the smaller p. Similarly, if $p > p'$ we prefer rejection of the lot and this preference increases as p increases. Thus, it will be possible to select a value $p_0 < p'$ and a value $p_1 > p'$ such that the error is considered serious only if we accept the lot when $p \geq p_1$, or we reject the lot when $p \leq p_0$.

After the two values p_0 and p_1 have been selected the risks that we are willing to tolerate may reasonably be stated as follows: a sampling inspection plan is required such that the probability of rejecting the lot is less than or equal to a preassigned value α whenever $p \leq p_0$, and the probability of accepting the lot is less than or equal to a preassigned value β whenever $p \geq p_1$. Thus, the tolerated risks are characterized by the four quantities p_0, p_1, α and β. The proper sampling plan can be determined after these four quantities have been chosen.

5.2.3. *The sequential probability ratio test corresponding to the quantities* p_0, p_1, α *and* β. Let H_0 be the hypothesis that $p = p_0$ and H_1 the hypothesis that $p = p_1$. Consider the sequential probability ratio test T for testing H_0 against H_1 for which α is the probability of accepting H_1 when H_0 is true (error of the first kind) and β is the probability of accepting H_0 when H_1 is true (error of the second kind). This probability ratio test will satisfy all our requirements, since for this test the probability of accepting the lot (accepting H_0) is $\leq \beta$ whenever $p \geq p_1$ and the probability of rejecting the lot (accepting H_1) is $\leq \alpha$ whenever $p \leq p_0$.

According to formulas (3.8), (3.9), (3.10) and section 3.3 the sequential test T is given as follows: At each stage of the inspection, at the m-th observation for each integral value of m, calculate the quantity

$$\frac{p_{1m}}{p_{0m}} = \frac{p_1^{d_m}(1 - p_1)^{m-d_m}}{p_0^{d_m}(1 - p_0)^{m-d_m}} \qquad (m = 1, 2, \cdots) \tag{5.2}$$

where d_m denotes the number of defectives found in the first m units inspected. Reject the lot (accept H_1) if

$$\frac{p_{1m}}{p_{0m}} \geq \frac{1 - \beta}{\alpha}. \tag{5.3}$$

Accept the lot if

$$\frac{p_{1m}}{p_{0m}} \leq \frac{\beta}{1 - \alpha}. \tag{5.4}$$

Take an additional observation if[14]

$$\frac{\beta}{1-\alpha} < \frac{p_{1m}}{p_{0m}} < \frac{1-\beta}{\alpha}. \tag{5.5}$$

For the purpose of practical computations it is useful to rewrite the inequalities (5.3), (5.4) and (5.5) in a somewhat different form. Taking the logarithms of both sides of the inequalities (5.3), (5.4) and (5.5) one can easily verify that these inequalities are equivalent to

$$d_m \geq \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{p_1}{p_0} - \log \frac{1-p_1}{1-p_0}} + m \frac{\log \frac{1-p_0}{1-p_1}}{\log \frac{p_1}{p_0} - \log \frac{1-p_1}{1-p_0}} \tag{5.6}$$

$$d_m \leq \frac{\log \frac{\beta}{1-\alpha}}{\log \frac{p_1}{p_0} - \log \frac{1-p_1}{1-p_0}} + m \frac{\log \frac{1-p_0}{1-p_1}}{\log \frac{p_1}{p_0} - \log \frac{1-p_1}{1-p_0}} \tag{5.7}$$

and

$$\begin{aligned} &\frac{\log \frac{\beta}{1-\alpha}}{\log \frac{p_1}{p_0} - \log \frac{1-p_1}{1-p_0}} + m \frac{\log \frac{1-p_0}{1-p_1}}{\log \frac{p_1}{p_0} - \log \frac{1-p_1}{1-p_0}} \\ &< d_m < \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{p_1}{p_0} - \log \frac{1-p_1}{1-p_0}} + m \frac{\log \frac{1-p_0}{1-p_1}}{\log \frac{p_1}{p_0} - \log \frac{1-p_1}{1-p_0}}. \end{aligned} \tag{5.8}$$

Using the inequalities (5.6), (5.7) and (5.8) the test procedure can easily be carried out as follows: For each m we compute the acceptance number

$$A_m = \frac{\log \frac{\beta}{1-\alpha}}{\log \frac{p_1}{p_0} - \log \frac{1-p_1}{1-p_0}} + m \frac{\log \frac{1-p_0}{1-p_1}}{\log \frac{p_1}{p_0} - \log \frac{1-p_1}{1-p_0}} \tag{5.9}$$

and the rejection number

$$R_m = \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{p_1}{p_0} - \log \frac{1-p_1}{1-p_0}} + m \frac{\log \frac{1-p_0}{1-p_1}}{\log \frac{p_1}{p_0} - \log \frac{1-p_1}{1-p_0}}. \tag{5.10}$$

[14] There is a slight approximation involved in the formulas (5.3), (5.4) and (5.5). For details see section 3.3.

These acceptance numbers A_m and rejection numbers R_m are best tabulated before inspection starts. Inspection is continued as long as $A_m < d_m < R_m$. At the first time when d_m does not lie between the acceptance and rejection numbers, the sampling inspection is terminated. The lot is accepted if $d_m \leq A_m$ and the lot is rejected if $d_m \geq R_m$.

The test procedure can also be carried out graphically as indicated in Figure 2. The number m of observations made is measured along the abscissa axis. Since A_m is a linear function of m, the points (m, A_m) will lie on a straight line L_0. Similarly, the points (m, R_m) will lie on a straight line L_1. We draw the lines L_0 and L_1 and the points (m, d_m) are plotted as inspection goes on. At the first time when the point (m, d_m) does not lie between the lines L_0 and L_1 inspection

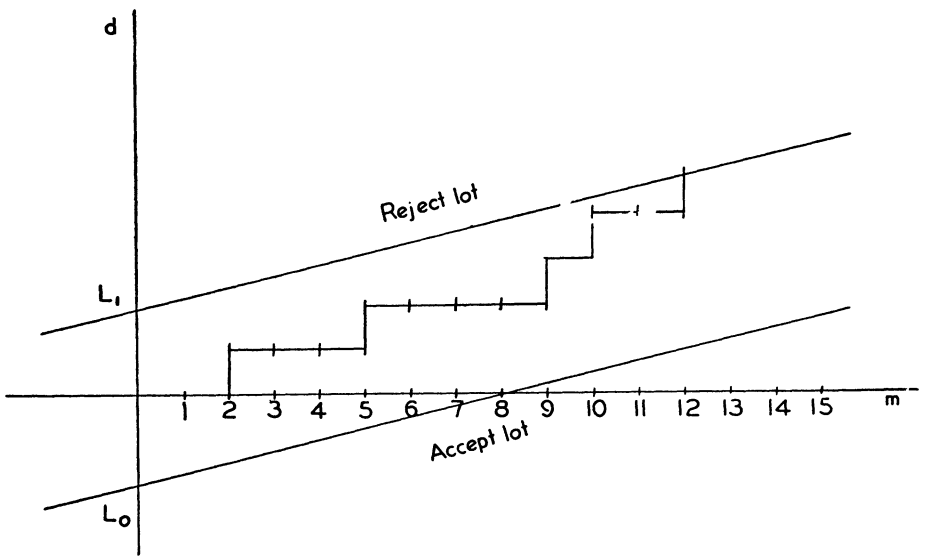

FIG. 2

is terminated. The lot is rejected if the point (m, d_m) lies on L_1 or above, and the lot is accepted if the point (m, d_m) lies on L_0 or below.

5.2.4. *The operating characteristic curve of the test.* As mentioned in section 5.2.3 the test procedure defined by the inequalities (5.6), (5.7) and (5.8) will satisfy the requirement that the probability of accepting the lot is $\leq \beta$ whenever $p \geq p_1$ and the probability of rejecting the lot is $\leq \alpha$ whenever $p \leq p_0$. Although this already describes the essential features of the test procedure, it may be desirable to know the probability L_p of accepting the lot for any possible value p of the proportion of defectives in the lot. Clearly, L_p will be a function of p and can be plotted as shown in Figure 3. The curve L_p is called the operating characteristic curve. The range of p is, of course, from 0 to 1. $L_p = 1$ for $p = 0$ and $L_p = 0$ for $p = 1$. The value of L_p decreases as p increases. We already know that $L_{p_0} = 1 - \alpha$ and $L_{p_1} = \beta$. Now we shall give a method

for computing the value of L_p for any p. If p_1 is not far from p_0, which will usually be the case in practice, a good approximation to L_p is given by (see equation 3.35)

$$L_p \sim 1 - \frac{1 - \left(\frac{\beta}{1-\alpha}\right)^h}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h} = \frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h}. \tag{5.11}$$

where h is equal to the non-zero root of the equation

$$p\left(\frac{p_1}{p_0}\right)^h + (1-p)\left(\frac{1-p_1}{1-p_0}\right)^h = 1. \tag{5.12}$$

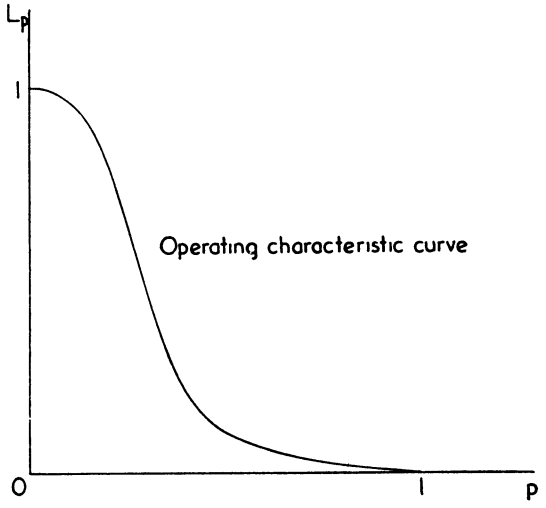

FIG. 3

To plot the operating characteristic curve, it is not necessary to solve (5.12) with respect to h. Instead we can proceed as follows: From (5.12) we express p as a function of h, i.e.,

$$p = \frac{1 - \left(\frac{1-p_1}{1-p_0}\right)^h}{\left(\frac{p_1}{p_0}\right)^h - \left(\frac{1-p_1}{1-p_0}\right)^h}. \tag{5.13}$$

For any given value h we compute the value of p from (5.13) and the value of L_p from (5.11). The point (p, L_p) obtained in this way will be a point of the operating characteristic curve. Doing this for various values of h we can obtain a sufficient number of points on the operating characteristic curve so that the curve can be drawn.

5.2.5. *The average amount of inspection required by the test.* Denote by $E_p(n)$ the expected value of the number of observations required by the test. Clearly, $E_p(n)$ is a function of p. According to (4.8) a good approximation to the value of $E_p(n)$ is given by

$$E_p(n) \sim \frac{L_p \log \frac{\beta}{1-\alpha} + (1 - L_p) \log \frac{1-\beta}{\alpha}}{p \log \frac{p_1}{p_0} + (1 - p) \log \frac{1 - p_1}{1 - p_0}} \tag{5.14}$$

where L_p is given by (5.11). Plotting $E_p(n)$ as a function of p, the curve obtained will, in general, be of the type shown in Fig. 4. The maximum will ordinarily be reached between p_0 and p_1. Furthermore, the curve will, in general, be increasing as p increases from 0 to p_0, and decreasing as p increases from p_1 to 1.

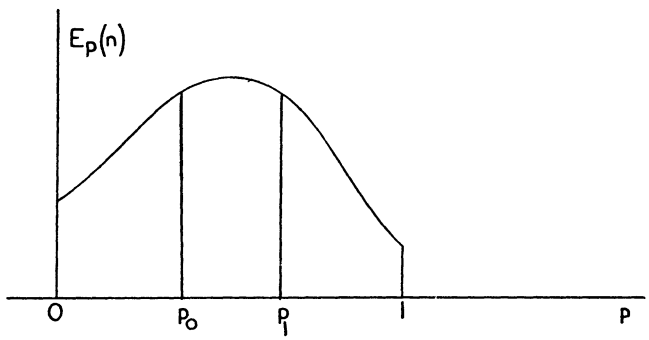

FIG. 4

5.3. *Sequential analysis of double dichotomies.* 5.3.1. *Formulation of the problem.* Suppose that we want to compare the effectiveness of two production processes where the effectiveness of a production process is measured in terms of the proportion of effective units in the sequence produced. We shall say that a unit is effective if it has a certain desirable property, for example, if it withstands a certain strain. Let p_1 be the proportion of effectives if process 1 is used, and p_2 the proportion of effectives if process 2 is used. In other words, p_1 is the probability that a unit produced will be effective if process 1 is used, and p_2 is the probability that a unit produced will be effective if process 2 is used. Suppose that the manufacturer does not know the values of p_1 and p_2, and that process 1 is in operation. If $p_1 \geq p_2$, then the manufacturer wants to retain process 1. However, if $p_1 < p_2$, especially if p_1 is substantially smaller than p_2, the manufacturer would like to replace process 1 by process 2. Thus, we are interested in testing the hypothesis that $p_1 \geq p_2$ against the alternative that $p_1 < p_2$.

A more general formulation of the problem can be given as follows: Consider two binomial distributions. Let p_1 be the probability of a success in a single

trial according to the first binomial distribution, and let p_2 be the probability of a success in a single trial according to the second binomial distribution. We shall use the symbol 1 for success and the symbol 0 for failure. Suppose that the probabilities p_1 and p_2 are unknown. We consider the problem of testing the hypothesis that $p_1 \geq p_2$ on the basis of a sample consisting of N_1 observations from the first binomial distribution and N_2 observations from the second binomial population. Since in many experiments the case $N_1 = N_2$ is mainly of interest, and since this case (as we shall see later) makes an exact and simplified mathematical treatment of the problem possible, we shall assume in what follows that $N_1 = N_2 = N$ (say).

Thus, on the basis of the outcome of the two series of N independent trials we have to decide whether the hypothesis $p_1 \geq p_2$ should be accepted or rejected.

5.3.2. *The classical method.* The classical solution of the problem for large N is given as follows: Let S_1 be the number of successes in the first set of N trials (drawn from the first binomial population), and let S_2 be the number of successes in the second set of N trials (drawn from the second binomial population). Denote $\frac{S_1 + S_2}{2N}$ by $\bar{p}$ and $1 - \bar{p}$ by $\bar{q}$. Then for large N the expression

$$\frac{S_2 - S_1}{\sqrt{2N\bar{p}\bar{q}}} \tag{5.15}$$

is normally distributed with zero mean and unit variance if $p_1 = p_2$. Suppose that the level of significance we wish to choose is α. Let λ_α be the value for which the probability that a normal variate with zero mean and unit variance will exceed λ_α is equal to α. (For example, if $\alpha = .05$, $\lambda_\alpha = 1.64$). Thus, if $p_1 = p_2$, the probability that the expression (5.15) will exceed λ_α is equal to α. If $p_1 > p_2$, the probability that the expression (5.15) will exceed λ_α is less than α. According to the classical method the hypothesis that $p_1 \geq p_2$ is rejected if the observed value of (5.15) exceeds λ_α. This method involves an approximation. The distribution of the expression (5.15) is not exactly normal even for large N. For small N this method cannot be used, since the distribution of (5.15) is far from normal. For small N, R. A. Fisher has proposed an exact method which, however, involves cumbersome calculations. In section 5.3.3. we shall suggest another method which is exact (does not involve any approximations) and is simple to apply as far as computations are concerned. The latter method has the further advantage of being suitable for sequential analysis to which existing methods are not readily adaptable.

5.3.3. *An exact method.* Let $a_1, \cdots, a_N$ be the results in the first set of N trials, and $b_1, \cdots, b_N$ the results in the second set of N trials. These results are arranged in the order observed. Consider the sequence of N pairs

$$(a_1, b_1), \cdots, (a_N, b_N). \tag{5.16}$$

Let t_1 be the number of pairs (1, 0) and t_2 the number of pairs (0, 1) in this sequence. We consider only the pairs (0, 1) and (1, 0) and base the test on them.

Let a be the outcome of an observation from the first population, and b the outsome of an observation from the second population. The probability that $(a, b) = (1, 0)$ is equal to $p_1(1 - p_2)$, and the probability that $(a, b) = (0, 1)$ is equal to $(1 - p_1)p_2$. Hence, knowing that (a, b) is equal to one of the pairs $(0, 1)$ and $(1, 0)$, the (conditional) probability that it is equal to $(0, 1)$ is given by

$$p = \frac{(1 - p_1)p_2}{p_1(1 - p_2) + p_2(1 - p_1)}, \tag{5.17}$$

and the (conditional) probability that it is equal to $(1, 0)$ is given by

$$1 - p = \frac{p_1(1 - p_2)}{p_1(1 - p_2) + (1 - p_1)p_2}. \tag{5.18}$$

Hence, considering only the pairs $(1, 0)$ and $(0, 1)$ the variate t_2 is distributed like the number of successes in a sequence of $t = t_1 + t_2$ independent trials, the probability of a success in a single trial being equal to p. One can easily verify that $p = \frac{1}{2}$ if $p_1 = p_2$, $p < \frac{1}{2}$ if $p_1 > p_2$ and $p > \frac{1}{2}$ if $p_1 < p_2$. Thus, the hypothesis to be tested, i.e., the hypothesis that $p_1 \geq p_2$, is equivalent to the hypothesis that $p \leq \frac{1}{2}$. Thus, we can test the hypothesis that $p_1 \geq p_2$ by testing the hypothesis that $p \leq \frac{1}{2}$ on the basis of the observed value of t_2. Since the distribution of t_2 is the same as the distribution of the number of successes in $t = t_1 + t_2$ independent trials (t is treated as a constant and the probability of a success in a single trial is equal to p), the test procedure can be carried out in the usual manner. If we want a level of significance α, a critical value T is chosen so that for $p = \frac{1}{2}$ the probability that $t_2 \geq T$ is equal to α. The hypothesis that $p \leq \frac{1}{2}$ is rejected if and only if the observed t_2 is greater than or equal to the critical value T. The value of T can be obtained from a table of the binomial distribution. If t is large, t_2 is nearly normally distributed and the critical value T can be obtained from a table of the normal distribution.

This procedure thus provides a simple test of the hypothesis that $p_1 \geq p_2$. The question arises whether the efficiency of this method is as high as that of the classical method. It would seem that the method suggested here cannot be a most efficient procedure, since the values of t_1 and t_2 depend on the order of the elements in the sequences $(a_1, \cdots, a_N)$ and $(b_1, \cdots, b_N)$, and there is no particular reason to arrange them in the order observed. However, it has been shown in [7] that the loss in efficiency as compared with the classical method is negligible if the number N of trials is large.[15]

It should be pointed out that the procedure for testing the hypothesis that $p_1 \geq p_2$ can be used also for testing the hypothesis that $p_1 = p_2$ if the alternative hypotheses are restricted to $p_2 > p_1$.

In addition to simplicity and exactness the present method seems superior to the classical one in the following respect: Suppose that (contrary to the original

[15] The author believes that the loss in efficiency is slight even when N is small, although no exact investigation of this case has been made.

assumption) the probability of a success varies from trial to trial. Denote by $p_1^{(i)}$ the probability of success in the i-th trial of the first set, and by $p_2^{(i)}$ the probability of success in the i-th trial in the second set ($i = 1, \cdots, N$). Assume that that the probabilities $p_1^{(i)}$ and $p_2^{(i)}$ are entirely unknown and we wish to test the hypothesis that $p_1^{(i)} - p_2^{(i)} = \cdots = p_1^{(N)} - p_2^{(N)} = 0$. In this case the classical method is not applicable, but the present method provides a correct procedure. Such a situation may arise, for instance, if we want to test the hypothesis that the probability of a success (hitting the target) is the same for two different guns. In the course of the experiments the probability of a hit may change due to external conditions such as wind, disposition of the gunner, etc. However, these external conditions are likely to affect both guns equally if the trials are made alternately (or approximately alternately), so that if the two guns are equally good we have $p_1^{(i)} = p_2^{(i)}$ ($i = 1, \cdots, N$).

5.3.4. *Sequential test of the hypothesis that $p_1 \geq p_2$.* In order to devise a proper sequential test for testing the hypothesis that $p_1 \geq p_2$, we have to state first what risks of making wrong decisions we are willing to tolerate. The efficiency of the production process 1 may be measured by the ratio of effectives to ineffectives produced, i.e., by $k_1 = \frac{p_1}{1 - p_1}$. Production process 1 may be regarded the more efficient the larger the value of k_1. Similarly, the efficiency of production process 2 may be measured by $k_2 = \frac{p_2}{1 - p_2}$. The relative superiority of production process 2 over the process 1 can then reasonably be measured by the ratio of k_2 to k_1 i.e., by

$$u = \frac{k_2}{k_1} \frac{p_2(1 - p_1)}{p_1(1 - p_2)} \tag{5.19}$$

If $u = 1$, the two processes are equally good. If $u > 1$, process 2 is superior to process 1, and if $u < 1$, process 1 is superior to process 2. Thus, the manufacturer will, in general, be able to select two values of u, u_0 and u_1 say ($u_0 < u_1$) such that the rejection of process 1 in favor of process 2 is considered an error of practical importance whenever the true value of $u \leq u_0$, and the maintainance of process 1 is considered an error of practical importance whenever $u \geq u_1$. If u lies between u_0 and u_1, the manufacturer does not care particularly which decision is taken.

Clearly, we will always have $u_0 < u_1$. If the transition from production process 1 to process 2 involves some cost or other inconveniences, it seems reasonable to put $u_0 = 1$ (or u_0 may even be slightly greater than one). This choice of u_0 really means that we consider the rejection of process 1 a serious error whenever this process is not inferior to process 2. On the other hand, if the transition from process 1 to process 2 does not involve any inconveniences, the rejection of process 1 in favor of 2 cannot be a serious error when the two processes are equally efficient, i.e., when $u = 1$. Thus, in such a case, it seems reasonable to choose u_0 somewhat below 1.

After the quantities u_0 and u_1 have been chosen the risks that we are willing to tolerate may reasonably be expressed in the following form: The probability of rejecting process 1 should not exceed a preassigned value α whenever $u \leq u_0$, and the probability of maintaining process 1 should not exceed a preassigned value β whenever $u \geq u_1$.

Thus, the risks that we are willing to tolerate are characterized by the four quantities u_0, u_1, α and β. After these four quantities have been chosen, a proper sequential test can be carried out as follows: The (conditional) probability that we obtain a pair (0, 1), as given in (5.17), can be expressed as a function of u. In fact

$$(5.20) \qquad p = \frac{(1 - p_1)p_2}{p_1(1 - p_2) + p_2(1 - p_1)} = \frac{\frac{(1 - p_1)p_2}{p_1(1 - p_2)}}{1 + \frac{p_2(1 - p_1)}{p_1(1 - p_2)}} = \frac{u}{1 + u}.$$

Let H_0 denote the hypothesis that $p = \frac{u_0}{1 + u_0}$, and H_1 the hypothesis that $p = \frac{u_1}{1 + u_1}$. A proper sequential test satisfying our requirements concerning tolerated risks is the sequential probability ratio test of H_0 against H_1. The acceptance and rejection numbers for this sequential test can be obtained from (5.9) and (5.10) by substituting $\frac{u_0}{1 + u_0}$ for p_0 $\frac{u_1}{1 + u_1}$ for p_1 and $t = t_1 + t_2$ for m.

Thus, for each value of t the acceptance number is given by

$$(5.21) \qquad A_t = \frac{\log \frac{\beta}{1 - \alpha}}{\log u_1 - \log u_0} + t \frac{\log \frac{1 + u_1}{1 + u_0}}{\log u_1 - \log u_0}$$

and the rejection number is given by

$$(5.22) \qquad R_t = \frac{\log \frac{1 - \beta}{\alpha}}{\log u_1 - \log u_0} + t \frac{\log \frac{1 + u_1}{1 + u_0}}{\log u_1 - \log u_0}$$

These acceptance numbers A_t and rejection numbers $R_t (t = 1, 2, \cdots)$ are best tabulated before experimentation starts. The sequential test is then carried out as follows: The observations are taken in pairs where each pair consists of an observation from the first process and an observation from the second process. We continue taking pairs as long as $A_t < t_2 < R_t$. At the first time when t_2 does not lie between the acceptance and rejection numbers, experimentation is terminated. Process 1 is maintained if at this final stage $t_2 \leq A_t$, and process 1 is rejected in favor of 2 if $t_2 \geq R_t$.

The test procedure can also be carried out graphically as shown in Figure 5. The total number m of pairs (0, 1) and (1, 0) is measured along the horizontal axis. The points (t, A_t) will lie on a straight line L_0, since A_t is a linear function of t. The points (t, R_t) will lie on a parallel line L_1. We draw the lines L_0 and

L_1 and plot the points (t, t_2) as experimentation goes on. At the first time when the point (t, t_2) is not within the lines L_0 and L_1 experimentation is terminated. Process 1 is maintained if at the final stage the point (t, t_2) lies on L_0 or below, and process 1 is rejected if the point (t, t_2) lies on L_1 or above.

5.3.5. *The operating characteristic curve of the test.* For any value u of the ratio $\frac{k_2}{k_1}$ we shall denote by L_u the probability of maintaining process 1. Clearly, L_u is a function of u. This function L_u is called the operating characteristic curve of the test. The operating characteristic curve can be determined from the equations (5.11) and (5.13) by substituting $\frac{u_1}{1+u_1}$ for p_1 and $\frac{u_0}{1+u_0}$ for p_0.

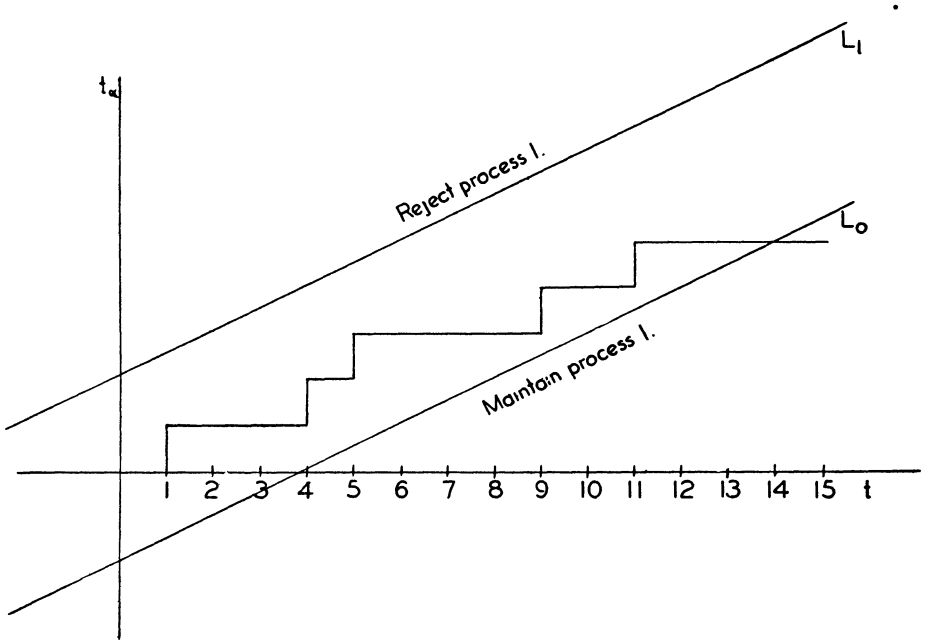

FIG. 5

These equations are:

$$L_u \sim \frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h} \tag{5.23}$$

and

$$\frac{u}{1+u} = \frac{1 - \left(\frac{1+u_0}{1+u_1}\right)^h}{\left(\frac{u_1(1+u_0)}{u_0(1+u_1)}\right)^h - \left(\frac{1+u_0}{1+u_1}\right)^h}. \tag{5.24}$$

For any given value h we compute the values of u and L_u from these equations. The point (u, L_u) obtained in this way will be a point of the operating characteristic curve. Calculating the points (u, L_u) for a sufficiently large number of values of h we can draw the operating characteristic curve.

5.3.6. *The average amount of inspection required by the test.* For any value u of the ratio $\frac{k_2}{k_1}$ denote by $E_u(t)$ the expected value of the total number of pairs (0, 1) and (1, 0) required by the test. The value of $E_u(t)$ can be obtained from (5.14) by substituting $E_u(t)$ for $E_p(n)$, L_u for L_p, $\frac{u_1}{1+u_1}$ for p_1 and $\frac{u_0}{1+u_0}$ for p_0. Thus

$$E_u(t) \sim \frac{L_u \log \frac{\beta}{1-\alpha} + (1 - L_u) \log \frac{1-\beta}{\alpha}}{\frac{u}{1+u} \log \frac{u_1(1+u_0)}{u_0(1+u_1)} - \frac{1}{1+u} \log \frac{1+u_0}{1+u_1}}. \tag{5.25}$$

To compute the expected value of the total number of pairs (including also the pairs (0, 0) and (1, 1)), we merely have to divide the right side expression in (5.25) by $p_1(1 - p_2) + p_2(1 - p_1)$.

In the rare event that no decision is yet reached at a number of pairs equal to three times the expected value, we can truncate the test at that stage without seriously affecting the probabilities of making a wrong decision (see section 4.6 in Part I).

5.3.7. *Observations made in groups of r.* In applications it may happen that at each stage in the sequential process instead of drawing a single observation we draw r observations from each of the binomial distributions. Hence, instead of a single pair, we have two sets of r observations. If the order of observations in each such set of r is recorded, we can establish the number of pairs (0, 1) and the number of pairs (1, 0) for each pair of sets of r observations. In such a case the test can be carried out as described in section 5.3.4, since after each pair of sets of r observations we can compute t and t_2. The only effect of taking the observations in groups of r is that more observations will generally be necessary (approximately enough to fill out a group) and thereby the probability of making an incorrect decision will be made somewhat smaller. However, if the order of observations in such groups of r is not recorded, the difficulty arises that we are not able to determine the values of t and t_2 needed for the test procedure. It has been shown in [7] that in such a case we may replace t and t_2 by certain estimates of t and t_2 without affecting seriously the probability of making an incorrect decision. The estimates of t_1 and t_2 (and thereby also an estimate of $t = t_1 + t_2$) are obtained as follows: Let r_1 be the number of successes in the group of r observations drawn from the first binomial distribution, and let r_2 be the number of successes in the group of r observations drawn from the second binomial distribution. Then for this pair of groups of r observations, we estimate the number

of pairs (1, 0) to be $r_1 - \frac{r_1 r_2}{r}$ and the number of pairs (0, 1) to be $r_2 - \frac{r_1 r_2}{r}$. Thus, an estimate of t_1 is obtained by summing $r_1 - \frac{r_1 r_2}{r}$ over all pairs of groups observed, and that of t_2 is obtained by summing $r_2 - \frac{r_1 r_2}{r}$ over all pairs of groups observed.

5.4. *Application to testing the mean of a normal distribution with known standard deviation.* 5.4.1. *Formulation of the problem.* Suppose that a measurable quantity x is normally distributed with unknown mean θ and known standard deviation σ. For example, x may be some measurable quality characteristic of a unit of a certain product where x is normally distributed with a known standard deviation in the population of all units. The problem we shall consider here is to test the hypothesis that the unknown mean θ is less than a specified value θ'. This problem arises frequently, for example, in quality control. Suppose that the quality of the product is considered the better the higher the mean value of x. Thus, there will be a value θ' such that the product is considered sub-standard if $\theta < \theta'$ and the product is considered to meet specifications if $\theta \geq \theta'$. Since θ is unknown, we are usually interested in testing the hypothesis that $\theta < \theta'$, i.e., that the product is sub-standard.

Since quality control is an important field of application for such test procedures, the discussion will be continued in the terminology of quality control. This, of course, should not be interpreted as a restriction upon the general validity and applicability of the test procedure. The problem treated in section 5.4 can now be stated as follows: Let x be a measurable quality characteristic of a unit of a certain product. The variable x is supposed to be normally distributed with known standard deviation in the population of all units produced. The problem is to devise a sampling plan for testing the hypothesis that the product is sub-standard. The product is said to be sub-standard, if the mean θ of x is less than a given specified value θ'.

5.4.2. *Tolerated risks for making a wrong decision.* No sampling plan can guarantee that the correct decision will always be made, i.e., that the product will be declared sub-standard if and only if $\theta < \theta'$. The larger the amount of inspection, the smaller we can make the risks for making a wrong decision. If inspection is costly, or destructive, we are willing to tolerate some risks of making wrong decisions in order to reduce the necessary amount of inspection. Thus, a proper sampling plan can be recommended only after the risks that can be tolerated have been stated.

If the quality of the product is exactly on the margin, i.e., if $\theta = \theta'$, then it will make little difference whether the product is classified as sub-standard or not. However, if θ is considerably smaller than θ', then the acceptance of the hypothesis that the product meets specifications (rejection of the hypothesis that the product is sub-standard) will usually be considered as a serious error.

Similarly, if θ is much larger than θ', the acceptance of the hypothesis that the product is sub-standard will generally be considered as a serious error. Thus, the manufacturer will, in general, be able to select two values of θ, θ_0 and θ_1 say ($\theta_0 < \theta'$ and $\theta_1 > \theta'$) such that the classification of the product as satisfactory (meeting specifications) is considered an error of practical importance whenever $\theta \leq \theta_0$, and the classification of the product as sub-standard is considered an error of practical importance whenever $\theta \geq \theta_1$. If θ lies between θ_0 and θ_1, a wrong classification of the product will not be viewed as a serious error, since in this case θ is near the marginal value θ'.

After the two values θ_0 and θ_1 have been selected, the risks that we are willing to tolerate can be stated in the following form: A sampling plan is required such that the probability of classifying the product as satisfactory is less than or equal to a preassigned quantity α whenever $\theta \leq \theta_0$, and such that the probability of classifying the product as sub-standard is less than or equal to a preassigned quantity β whenever $\theta \geq \theta_1$. Thus, the tolerated risks are characterized by the four quantities θ_0, θ_1, α and β. A proper sampling plan can be devised after these four quantities have been selected.

5.4.3. *A sequential test of the hypothesis that $\theta < \theta'$ (the product is sub-standard).* Let H_0 be the hypothesis that $\theta = \theta_0$ and let H_1 be the hypothesis that $\theta = \theta_1$. Let T be the sequential probability ratio test for testing H_0 against H_1 such that α is the probability of accepting H_1 when H_0 is true and β is the probability of accepting H_0 when H_1 is true. This sequential test will satisfy all our requirements, since for this test the probability of accepting H_0 (declaring the product as sub-standard) is $\leq \beta$ whenever $\theta \geq \theta_1$, and the probability of accepting H_1 (declaring the product as satisfactory) is $\leq \alpha$ whenever $\theta \leq \theta_0$.

The sequential test T is given as follows: Denote the successive observations on x by $x_1, x_2, \cdots$, etc. Accept the hypothesis that the product is satisfactory at the m-th observation if

$$\log \frac{e^{-(1/2\sigma^2)\sum_{\alpha=1}^{m}(x_\alpha-\theta_1)^2}}{e^{-(1/2\sigma^2)\sum_{\alpha=1}^{m}(x_\alpha-\theta_0)^2}} \geq \log \frac{1-\beta}{\alpha}. \tag{5.26}$$

Accept the hypothesis that the product is sub-standard if

$$\log \frac{e^{-(1/2\sigma^2)\sum_{\alpha=1}^{m}(x_\alpha-\theta_1)^2}}{e^{-(1/2\sigma^2)\sum_{\alpha=1}^{m}(x_\alpha-\theta_0)^2}} \leq \log \frac{\beta}{1-\alpha} \tag{5.27}$$

Take an additional observation if

$$\log \frac{\beta}{1-\alpha} < \log \frac{e^{-(1/2\sigma^2)\sum_{1}^{m}(x_\alpha-\theta_1)^2}}{e^{-(1/2\sigma^2)\sum_{1}^{m}(x_\alpha-\theta_0)^2}} < \log \frac{1-\beta}{\alpha}. \tag{5.28}$$

The inequalities (5.26), (5.27) and (5.28) are equivalent to

$$\sum_{\alpha=1}^{m} x_\alpha \geq \frac{\sigma^2}{\theta_1 - \theta_0} \log \frac{1-\beta}{\alpha} + m \frac{\theta_0 + \theta_1}{2} \tag{5.29}$$

$$\sum_{\alpha=1}^{m} x_\alpha \leq \frac{\sigma^2}{\theta_1 - \theta_0} \log \frac{\beta}{1-\alpha} + m \frac{\theta_0 + \theta_1}{2} \tag{5.30}$$

and

$$\frac{\sigma^2}{\theta_1 - \theta_0} \log \frac{\beta}{1-\alpha} + m \frac{\theta_0 + \theta_1}{2} < \sum_{\alpha=1}^{m} x_\alpha < \frac{\sigma^2}{\theta_1 - \theta_0} \log \frac{1-\beta}{\alpha} + m \frac{\theta_0 + \theta_1}{2}, \tag{5.31}$$

respectively.

Using the inequalities (5.29), (5.30) and (5.31) the test procedure can easily be carried out as follows: For each m compute the acceptance number

$$A_m = \frac{\sigma^2}{\theta_1 - \theta_0} \log \frac{\beta}{1-\alpha} + m \frac{\theta_0 + \theta_1}{2} \tag{5.32}$$

and the rejection number

$$R_m = \frac{\sigma^2}{\theta_1 - \theta_0} \log \frac{1-\beta}{\alpha} + m \frac{\theta_0 + \theta_1}{2}. \tag{5.33}$$

These acceptance numbers A_m and rejection numbers R_m are best tabulated before inspection starts. Inspection is continued as long as $A_m < \sum_{\alpha=1}^{m} x_\alpha < R_m$. At the first time that $\sum_{\alpha=1}^{m} x_\alpha$ does not lie between A_m and R_m, inspection is terminated. If at this final stage $\sum_{\alpha=1}^{m} x_\alpha \leq A_m$, the hypothesis that the product is sub-standard is accepted, and if $\sum_{\alpha=1}^{m} x_\alpha \geq R_m$, the hypothesis that the product is sub-standard is rejected.

The test procedure can also be carried out graphically as shown in Figure 6. The number m of observations is measured along the horizontal axis. The points (m, A_m) will lie in a straight line L_0 and the points (m, R_m) will lie on a parallel line L_1. We draw the parallel lines L_0 and L_1 and plot the points $\left(m, \sum_{\alpha=1}^{m} x_\alpha\right)$ as inspection goes on. At the first time when the point $\left(m, \sum_{\alpha=1}^{m} x_\alpha\right)$ does not lie between the lines L_0 and L_1 inspection is terminated. The hypothesis that the product is sub-standard is rejected if the point $\left(m, \sum_{\alpha=1}^{m} x_\alpha\right)$ lies on L_1 or above. The hypothesis in question is accepted if the point $\left(m, \sum_{\alpha=1}^{m} x_\alpha\right)$ lies on L_0 or below.

5.4.4. *The operating characteristic curve of the test.* For any value θ denote by L_θ the probability that the hypothesis that the product is sub-standard is accepted. Obviously, L_θ will be a function of θ and is called the operating characteristic curve of the test. The shape of the operating characteristic curve will, in general, be of the type shown in Figure 7. L_θ approaches 1 as $\theta \to -\infty$ and L_θ approaches zero as $\theta \to \infty$. Furthermore, L_θ is a decreasing function of θ. We already know the values of L_θ for $\theta = \theta_0$ and $\theta = \theta_1$. Now we shall give a method for computing the value of L_θ for any θ. If $\frac{\theta_1 - \theta_0}{\sigma}$ is fairly small,

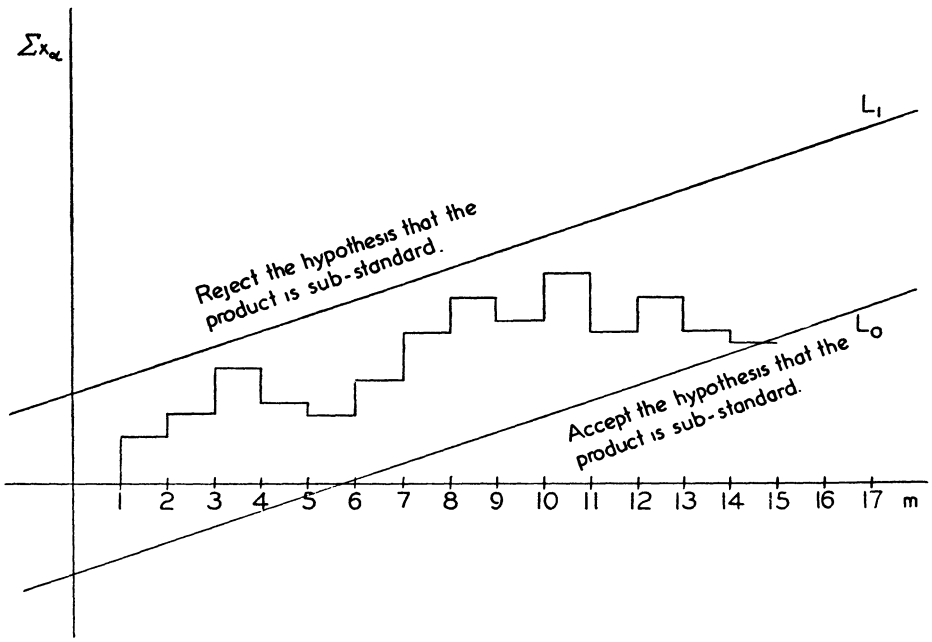

FIG. 6

which will usually be the case in practice, a good approximation to L_θ is given by (see equation 3.35)

$$L_\theta \sim 1 - \frac{1 - \left(\frac{\beta}{1-\alpha}\right)^h}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h} = \frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h} \tag{5.34}$$

where the constant h is determined as follows: First we compute the characteristic function $\varphi(t)$ of the variate

$$z = \log \frac{e^{-\frac{1}{2\sigma^2}(x-\theta_1)^2}}{e^{-\frac{1}{2\sigma^2}(x-\theta_0)^2}} = \frac{1}{2\sigma^2}[2(\theta_1 - \theta_0)x + \theta_0^2 - \theta_1^2]. \tag{5.35}$$

Thus, z is normally distributed with mean $= \frac{\theta_0^2 - \theta_1^2}{2\sigma^2} + \frac{(\theta_1 - \theta_0)\theta}{\sigma^2}$ and variance $= \frac{(\theta_1 - \theta_0)^2}{\sigma^2}$. Consequently, $\varphi(t)$ is given by

$$\varphi(t) = e^{\left[\frac{\theta_0^2-\theta_1^2}{2\sigma^2}+\frac{(\theta_1-\theta_0)\theta}{\sigma^2}\right]t+\frac{(\theta_1-\theta_0)^2}{2\sigma^2}t^2}. \tag{5.36}$$

The value h is the non-zero real root of the equation $\varphi(t) = 1$. Hence

$$h = \frac{(\theta_1^2 - \theta_0^2) - 2(\theta_1 - \theta_0)\theta}{(\theta_1 - \theta_0)^2} = \frac{\theta_1 + \theta_0 - 2\theta}{\theta_1 - \theta_0} \tag{5.37}$$

The operating characteristic curve can be computed from (5.34) substituting the right hand side member of (5.37) for h.

5.4.5. *The average amount of inspection required by the test.* Let $E_\theta(n)$ denote the expected value of the number of observations required by the test when θ

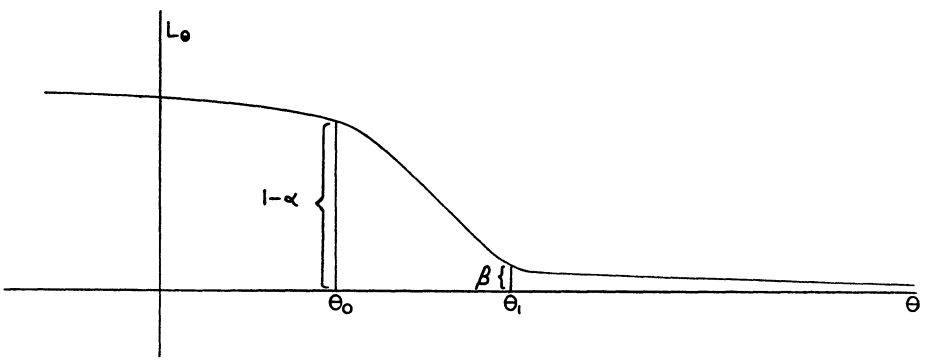

FIG. 7

is the true mean of x. According to (4.8) a good approximation to the value of $E_\theta(n)$ is given by

$$E_\theta(n) \sim 2\sigma^2 \frac{L_\theta \log \frac{\beta}{1-\alpha} + (1 - L_\theta) \log \frac{1-\beta}{\alpha}}{\theta_0^2 - \theta_1^2 + 2(\theta_1 - \theta_0)\theta}$$

where L_θ is given by (5.34).

In the rare event that the number of observations reaches three times the expected value before the test is terminated, we can truncate the test at this stage without seriously affecting the probabilities of making a wrong decision. (See section 4.6 in Part I).

6. Outline of a General Theory of Sequential Tests of Hypotheses when No Restrictions Are Imposed on the Alternative Values of the Unknown Parameters

6.1. *Sequential test of a simple hypothesis with no restrictions on the alternative values of the unknown parameters.* Consider the following general case. Let

$X_1, \cdots, X_p$ be a set of p random variables and let $f(x_1, \cdots, x_p, \theta_1, \cdots, \theta_k)$ be the joint probability density function of these random variables involving k unknown parameters $\theta_1, \cdots, \theta_k$. Suppose that we wish to test the hypothesis H_0 that $\theta_1 = \theta_1^0, \cdots, \theta_k = \theta_k^0$, where $\theta_1^0, \cdots, \theta_k^0$ are some given specified values. Denote the set of all *a priori* possible parameter points by Ω. Assume that Ω contains at least a finite k-dimensional sphere with the center $(\theta_1^0, \cdots, \theta_k^0)$. Let Ω^* be the set of all possible alternative parameter points; i.e., Ω^* is the whole parameter space Ω with the exception of the point $\theta^0 = (\theta_1^0, \cdots, \theta_k^0)$. For any statistical procedure for testing H_0, the probability of an error of the first kind, will have a definite value, but the probability of an error of the second kind will depend on the true alternative; i.e., it will be a single valued function $\beta(\theta)$ defined over all points θ of Ω^*. Let $w(\theta)$ be some non-negative function, called weight function, such that $\int_{\Omega^*} w(\theta)\, d\theta = 1$. Suppose that we wish to construct a sequential test such that the probability of an error of the first kind is equal to a given α and that the weighted average $\int_{\Omega^*} w(\theta)\beta(\theta)\, d(\theta)$ of the probabilities of errors of the second kind is equal to some given positive value β.

This problem can easily be solved as follows: Let p_{0n} be equal to the product $\prod_{\alpha=1}^{n} f(x_{1\alpha}, \cdots, x_{p\alpha}, \theta_1^0, \cdots, \theta_k^0)$ where $x_{i\alpha}$ denotes the αth observation on x_i $(i = 1, \cdots, p; \alpha = 1, \cdots, n)$. Furthermore, let p_{1n} be defined by

$$p_{1n} = \int_{\Omega^*} w(\theta) \left[\prod_{\alpha=1}^{n} f(x_{1\alpha}, \ldots, x_{p\alpha}, \theta_1, \ldots, \theta_k) \right] d\theta. \tag{6.1}$$

The expression p_{1n} can be interpreted as the probability density in the sample space of n observations on the variates $x_1, \cdots, x_p$, if we assume that the parameter point θ in Ω^* has a probability distribution given by the density function $w(\theta)\, d\theta$.

We shall denote by H_1 the hypothesis that the probability density function in the sample space of n observations on $X_1, \cdots, X_p$ is given by p_{1n} defined in equation (6.1). The problem of testing H_0 against the single alternative H_1 is not exactly of the type discussed in Part I, since p_{1n} given in (6.1) cannot be represented, in general, as a product of n factors where the αth factor depends only on the observations $x_{1\alpha}, \cdots, x_{p\alpha}$. However, it was pointed out in section 3.2 that the fundamental inequalities derived in Section 3.2 remain valid also when p_{1n} is given by an expression of the type (6.1). Thus, we can use the sequential probability ratio test for testing H_0 against the single alternative H_1. We reject H_0 if

$$\frac{p_{1n}}{p_{0n}} \geq A, \tag{6.2}$$

we accept H_0 if

$$\frac{p_{1n}}{p_{0n}} \leq B, \tag{6.3}$$

and we make an additional observation if

$$B < \frac{p_{1n}}{p_{0n}} < A. \tag{6.4}$$

The expression p_{1n} is given by (6.1) and the constants A and B are chosen so that the probability of accepting H_1 when H_0 is true is α, and the probability of accepting H_0 when H_1 is true is β. Thus, for practical purposes we may put $A = \frac{1-\beta}{\alpha}$ and $B = \frac{\beta}{1-\alpha}$.

Using the sequential process defined by the inequalities (6.2), (6.3), and (6.4) we obviously have

$$\int_{\Omega^*} w(\theta)\,\beta(\theta)\,d\theta = \beta \tag{6.5}$$

where for each point θ in Ω^*, $\beta(\theta)$ denotes the probability of accepting H_0 under the assumption that θ is the true parameter point.

Thus, the sequential test given by (6.2), (6.3), and (6.4) provides a satisfactory solution of the problem if we want a test procedure such that the probability of an error of the first kind is α and the weighted average $\int_{\Omega^*} w(\theta)\beta(\theta)\,d\theta$ of the probabilities of errors of the second kind is β. Practical problems, however, do not always take this form. Many instances require a test procedure such that $\beta(\theta)$ should be less than or equal to a given positive value β for all parameter points θ whose "distance" (defined in some sense) from θ^0 is greater than or equal to some given positive value d_0. The "distance" of two parameter points θ^1 and θ^2 may be defined by some function $\delta(\theta^1, \theta^2)$ which is equal to zero if $\theta^1 = \theta^2$ and is greater than zero if $\theta^1 \neq \theta^2$. Furthermore, for any three points θ^1, θ^2, θ^3 we have $\delta(\theta^1, \theta^2) = \delta(\theta^2, \theta^1)$ and $\delta(\theta^1, \theta^2) + \delta(\theta^3, \theta^2) \geq \delta(\theta^1, \theta^3)$. The distance function will, in general, be chosen according to practical needs and mathematical convenience.

Given the distance function $\delta(\theta^1, \theta^2)$ and given the requirements that the probability of an error of the first kind be α and the probability of an error of the second kind should not exceed β whenever the distance of the true parameter point from θ^0 is greater than or equal to d_0, the aim is, of course, to construct a sequential test which satisfies these requirements with a minimum expected number of observations.

While an exact solution of this problem has not yet been found, the following approach seems reasonable: Let Ω_0 be the set of all parameter points θ for which $\delta(\theta^0, \theta) \geq d_0$. We restrict ourselves to the class C_S of sequential tests based on the ratio $\frac{p_{1n}}{p_{0n}}$ where

$$p_{0n} = \prod_{\alpha=1}^{n} f(x_{1\alpha}, \cdots, x_{p\alpha}, \theta_1^0, \cdots, \theta_k^0), \tag{6.6}$$

$$p_{1n} = \int_{\Omega_0} w(\theta) \prod_{\alpha=1}^{n} f(x_{1\alpha}, \cdots, x_{p\alpha}, \theta_1, \cdots, \theta_k)\,d\theta \tag{6.7}$$

and $w(\theta)$ may be any non-negative function of θ, called weight function, for which

$$\int_{\Omega_0} w(\theta)d\theta = 1. \tag{6.8}$$

For carrying out the sequential test two constants A and B are chosen. The hypothesis H_0 is accepted if $\frac{p_{1n}}{p_{0n}} \leq B$, H_0 is rejected if $\frac{p_{1n}}{p_{0n}} \geq A$, and an additional observation is made if $B < \frac{p_{1n}}{p_{0n}} < A$. The restriction to the class C_S of sequential tests is suggested by the fact that we are led to these tests if it is required that some weighted average of the probabilities of errors of the second kind be equal to a given value β.

Accepting the restriction that the sequential test should be a member of the class C_S, we still need a principle for choosing the weight function $w(\theta)$. It is clear that the maximum of $\beta(\theta)$ in Ω_0 depends on the quantities A, B, and the weight function $w(\theta)$. Denote this maximum value by $\beta_{\text{Max}}[A, B, w(\theta)]$. Since it is desirable to make $\beta_{\text{Max}}[A, B, w(\theta)]$ as small as possible, it is proposed to determine $w(\theta)$ so that the expression $\beta_{\text{Max}}[A, B, w(\theta)]$ becomes a minimum with respect to $w(\theta)$. Since for given values A and B the value of the weighted average $\int_{\Omega_0} w(\theta)\beta(\theta)\, d\theta$ is practically independent of $w(\theta)$ (it is nearly equal to $\frac{B(A-1)}{A-B}$), minimizing $\beta_{\text{Max}}[A, B, w(\theta)]$ is practically equivalent to minimizing the difference $\beta_{\text{Max}}[A, B, w(\theta)] - \int_{\Omega_0} w(\theta)\beta(\theta)\, d\theta$. For convenience we determine $w(\theta)$ so that $\beta_{\text{Max}}[A, B, w(\theta)] - \int_{\Omega_0} w(\theta)\beta(\theta)\, d\theta$ becomes a minimum. For this weight function the maximum of $\beta(\theta)$ in Ω_0 will depend only on A and B. Denote this value by $\beta(A, B)$. Finally we determine the values A and B so that $\beta(A, B) = \beta$ and the probability of an error of the first kind becomes α.

The determination of $w(\theta)$ is a problem in the calculus of variations. In some important cases, however, the solution can be obtained by the following simple procedure: Let $S(d)$ be the set of all parameter points θ for which $\delta(\theta^0, \theta) = d$. Let $v(\theta)$ be a non-negative weight function defined over the surface $S(d_0)$ so that the surface integral $\int_{S(d_0)} v(\theta)\, d\omega = 1$ (where $d\omega$ denotes the infinitesimal surface element). Consider the following sequential procedure: Reject H_0 if

$$\frac{\int_{S(d_0)} v(\theta) \left[\prod_\alpha f(x_{1\alpha}, \cdots, x_{p\alpha}, \theta_1, \cdots, \theta_k)\right] d\omega}{\prod_\alpha f(x_{1\alpha}, \cdots, x_{p\alpha}, \theta_1^0, \cdots, \theta_k^0)} \tag{6.9}$$

is greater than or equal to A, accept H_0 if (6.9) is less than or equal to B, and make an additional observation if the value of (6.9) lies between A and B. The

constants A and B are so chosen that the probability of an error of the first kind is α and $\int_{S(d_0)} \beta(\theta)v(\theta)\,d\omega = \beta$. In many statistical problems it is possible to find a weight function $v(\theta)$ such that for a conveniently chosen distance function $\delta(\theta^1, \theta^2)$ the probability $\beta(\theta)$ of an error of the second kind becomes constant on the surface $S(d)$ for any value d, and, furthermore, $\beta(\theta)$ decreases with increasing d. For such a weight function $v(\theta)$, the sequential test based on (6.9), will provide a solution of the problem. In fact, the weight function $v(\theta)$ over the surface $S(d_0)$ can be considered a limiting case of a weight function $w(\theta)$ defined in Ω_0 which takes the value zero for any θ whose distance from θ^0 is greater than $d_0 + \Delta$ with Δ approaching zero in the limit. For the weight function $v(\theta)$ the maximum of $\beta(\theta)$ in Ω_0 is equal to the weighted integral of $\beta(\theta)$. Thus, for this weight function the difference between the maximum of $\beta(\theta)$ and the weighted integral of $\beta(\theta)$ is minimized.

We shall illustrate this procedure by a simple example. Let $X_1, \cdots, X_k$ be k normally and independently distributed variates with unit variances. The mean values $\theta_1, \cdots, \theta_k$ are unknown. Suppose that it is required to test the hypothesis H_0 that $\theta_1 = \cdots = \theta_k = 0$. Assume that the distance of two points θ^1 and θ^2 is equal to

$$+\sqrt{(\theta_1^1 - \theta_1^2)^2 + \cdots + (\theta_k^1 - \theta_k^2)^2}.$$

Then $S(d)$ is a sphere with center at the origin and radius d. Let $v(\theta)$ be constant on $S(d_0)$ and equal to the reciprocal of the area of $S(d_0)$. We shall show that for this weight function $v(\theta)$, $\beta(\theta)$ is constant on the sphere $S(d)$ and is monotonically decreasing with increasing d. For this purpose we prove first that (6.9) is a monotonically increasing function of $\bar{x}_1^2 + \cdots + \bar{x}_k^2$ where $\bar{x}_i$ is the arithmetic mean of the observations on x_i. In fact, the expression (6.9) becomes

$$\frac{c_k \dfrac{1}{(2\pi)^{kn/2}} \displaystyle\int_{S(d_0)} \exp\left[-\frac{1}{2}\sum_{i=1}^{k}\sum_{\alpha=1}^{n}(x_{i\alpha} - \theta_i)^2\right] d\omega}{\dfrac{1}{(2\pi)^{kn/2}} \exp\left[-\tfrac{1}{2}\Sigma\Sigma x_{i\alpha}^2\right]} \tag{6.10}$$

$$= c_k \exp\left[-\tfrac{1}{2} nd_0^2\right] \int_{S(d_0)} \exp\left[n\Sigma\bar{x}_i\theta_i\right] d\omega$$

where c_k is the reciprocal of the area of $S(d_0)$ and $\bar{x}_i$ is the arithmetic mean of the n observations $x_{i\alpha}$ $(\alpha = 1, \cdots, n)$. Let r_x denote $\left|\sqrt{\sum_i \bar{x}_i^2}\right|$ and let $\alpha(\theta)$ $(0 \leq \alpha \leq \pi)$ denote the angle between the vector $(\bar{x}_1, \cdots, \bar{x}_k)$ and the vector $(\theta_1, \cdots, \theta_k)$. Then (6.10) can be written

$$c_k \exp\left[-\tfrac{1}{2} nd_0^2\right] \int_{S(d_0)} \exp\left(nr_x d_0 \cos\left[\alpha(\theta)\right]\right) d\omega. \tag{6.11}$$

Because of the symmetry of the sphere, the value of (6.11) will not be changed if we substitute $\gamma(\theta)$ for $\alpha(\theta)$ where $\gamma(\theta)$ $(0 \leq \gamma(\theta) \leq \pi)$ denotes the angle

between the vector θ and an arbitrarily chosen fixed vector u. From this it follows that the value of (6.11) depends only on r_x.

Now we shall show that (6.11) is a strictly increasing function of r_x. For this purpose we have merely to show that

$$I(r_x) = \int_{S(d_0)} \exp\,(nr_x d_0 \cos\,[\gamma(\theta)])d\omega \tag{6.12}$$

is a strictly increasing function of r_x. We have

$$\frac{dI(r_x)}{dr_x} = \int_{S(d_0)} nd_0 \cos\,[\gamma(\theta)] \exp\,(nr_x d_0 \cos\,[\gamma(\theta)])d\omega. \tag{6.13}$$

Denote by ω_1 the subset of $S(d_0)$ in which $0 \leq \gamma(\theta) \leq \frac{\pi}{2}$, and by ω_2 the subset in which $\frac{\pi}{2} \leq \gamma(\theta) \leq \pi$. Because of the symmetry of the sphere we have

$$\begin{aligned}\int_{\omega_2} nd_0 \cos\,[\gamma(\theta)] \exp\,(nr_x d_0 \cos\,[\gamma(\theta)])\,d\omega \\ = \int_{\omega_1} nd_0 \cos\,[\pi - \gamma(\theta)] \exp\,(nr_x d_0 \cos\,[\pi - \gamma(\theta)])\,d\omega \\ = -\int_{\omega_1} nd_0 \cos\,[\gamma(\theta)] \exp\,(-nr_x d_0 \cos\,[\gamma(\theta)])\,d\omega.\end{aligned}$$

Hence

$$\begin{aligned}\frac{dI(r_x)}{dr_x} = nd_0 \int_{\omega_1} \cos\,[\gamma(\theta)] \\ \{\exp\,(nd_0 r_x \cos\,[\gamma(\theta)]) - \exp\,(-nd_0 r_x \cos\,[\gamma(\theta)])\}\,d\omega\end{aligned} \tag{6.14}$$

The right hand side of (6.14) is positive. Hence, we have proved that expression (6.11) (or (6.10)) is a strictly increasing function of r_x.

To show that $\beta(\theta)$ is constant on $S(d)$ and is monotonically decreasing with increasing d, let $y_1, \cdots, y_k$ be an orthogonal linear transformation of $x_1, \cdots, x_k$ so that $E(y_1) = \sqrt{\theta_1^2 + \cdots + \theta_k^2}$, $E(y_i) = 0$ $(i = 2, \cdots, k)$. Since $\bar{y}_1^2 + \cdots + \bar{y}_k^2 = \bar{x}_1^2 + \cdots + \bar{x}_k^2$ and since (6.11) depends only on $\bar{x}_1^2 + \cdots + \bar{x}_k^2$, it is seen that the sequence of expression (6.11) formed for any sequence of integers n has a joint distribution which depends only on $\sqrt{\theta_1^2 + \cdots + \theta_k^2}$. Hence $\beta(\theta)$ is constant on any sphere with center at the origin. Since (6.11) is a strictly increasing function of r_x, it can be shown that $\beta(\theta)$ is a monotonically decreasing function of $\sqrt{\theta_1^2 + \cdots + \theta_k^2}$. Hence, we can test the hypothesis H_0 by the sequential process based on (6.10).

If $k = 1$—that is, if we test the mean value of a single normal variate—the

sphere $S(d)$ is a 0-dimensional sphere consisting of the two points $\theta_1 = +d$ and $\theta_1 = -d$ and expression (6.10) reduces to

$$\frac{\frac{1}{2}\frac{1}{(2\pi)^{n/2}}\{\exp[-\tfrac{1}{2}\Sigma_\alpha(x_\alpha - d_0)^2] + \exp[-\tfrac{1}{2}\Sigma_\alpha(x_\alpha + d_0)^2]\}}{\frac{1}{(2\pi)^{n/2}}\exp[-\tfrac{1}{2}\Sigma x_\alpha^2]} = \tfrac{1}{2}\exp[-\tfrac{1}{2}nd_0^2]\{\exp[n\bar{x}d_0] + \exp[-n\bar{x}d_0]\}. \tag{6.15}$$

6.2. *Sequential test of a composite hypothesis.* We shall give only a brief outline of the principles on which a sequential test of a composite hypothesis can be based, since they are analogous to those for a simple hypothesis. Let $X_1, \cdots, X_p$ be a set of p random variables and let $f(x_1, \cdots, x_p, \theta_1, \cdots, \theta_k)$ be the joint probability density function of these variables involving k unknown parameters $\theta_1, \cdots, \theta_k$. Denote the set of all possible parameter points $\theta = (\theta_1, \cdots, \theta_k)$ by Ω. Suppose that we wish to test the hypothesis H_0 that the true parameter point θ is contained in the subset ω of Ω. Let $\bar{\omega}$ be the set of all points of Ω which are not contained in ω. Furthermore, let $w_0(\theta)$ and $w_1(\theta)$ be two non-negative functions of θ, called weight functions, such that

$$\int_\omega w_0(\theta)d\theta = 1 \text{ and } \int_{\bar{\omega}} w_1(\theta)d\theta = 1. \tag{6.16}$$

If ω is a surface in the space Ω then the integral over ω is meant to be the surface integral over ω.

In testing a composite hypothesis the probability of an error of the first kind need not necessarily be the same for all points θ in ω. It will, in general, be a function $\alpha(\theta)$ of the true point θ in ω. Similarly the probability of an error of the second kind is a function $\beta(\theta)$ of θ defined for all points in $\bar{\omega}$. Suppose that we wish to construct a sequential test such that the weighted average $\int_\omega w(\theta)\alpha(\theta)\,d\theta$ of the probabilities of errors of the first kind is a given value α, and the weighted average $\int_{\bar{\omega}} w(\theta)\beta(\theta)\,d\theta$ of the probabilities of errors of the second kind is a given value β. Then the following sequential test can be used: Denote by H_0^* the hypothesis that the probability density in the sample space of n observations on $X_1, \cdots, X_p$ is given by

$$p_{0n} = \int_\omega w_0(\theta)[\prod_\alpha f(x_{1\alpha}, \cdots, x_{p\alpha}, \theta_1, \cdots, \theta_k)]\,d\theta \tag{6.17}$$

and by H_1^* the hypothesis that the density in the sample space is given by

$$p_{1n} = \int_{\bar{\omega}} w_1(\theta)[\prod_\alpha f(x_{1\alpha}, \cdots, x_{p\alpha}, \theta_1, \cdots, \theta_k)]\,d\theta. \tag{6.18}$$

The sequential probability ratio test for testing H_0^* against the single alternative H_1^* provides a solution of our problem. If the constants A and B in this sequen-

tial test are chosen so that the probability is α that we reject H_0^* when H_0^* is true, and the probability is β that we accept H_0^* when H_1^* is true, then for this sequential test we have

$$\int_{\omega} w_0(\theta)\alpha(\theta)\, d\theta = \alpha$$

and

$$\int_{\bar{\omega}} w_1(\theta)\beta(\theta)\, d\theta = \beta.$$

This can be proved in the same way as the corresponding statement in the case of a simple hypothesis.

Frequently we may require a sequential test procedure such that the least upper bound of $\alpha(\theta)$ in ω is equal to a given α and $\beta(\theta)$ is less than or equal to a given β for all points θ whose "distance" (defined in some sense) from ω is greater than or equal to a given positive value d_0. The "distance" of a parameter point θ from ω may be defined by some function $\delta(\theta, \omega)$ which is positive if θ is not in ω and is zero if θ is in ω. The distance function will be chosen in general according to practical needs and mathematical convenience. For reasons similar to those discussed in the case of a simple hypothesis, an appropriate sequential test procedure with the desired properties can be found as follows: Let $\bar{\omega}(d)$ be the set of all points θ for which $\delta(\theta, \omega) \geq d$. Let, furthermore, $w_0(\theta)$ and $w_1(\theta)$ be two weight functions such that

$$\int_{\omega} w_0(\theta)\, d\theta = \int_{\bar{\omega}(d_0)} w_1(\theta)\, d\theta = 1. \tag{6.19}$$

Denote by H_0^* the hypothesis that the probability density in the sample space of n observations on $X_1, \cdots, X_p$ is given by

$$p_{0n} = \int_{\omega} w_0(\theta)\left[\prod_{\alpha=1}^{n} f(x_{1\alpha}, \cdots, x_{p\alpha}, \theta)\right] d\theta \qquad (n = 1, 2, \cdots) \tag{6.20}$$

and by H_1^* the hypothesis that the probability density in the sample space of n observations on $X_1, \cdots, X_p$ is given by

$$p_{1n} = \int_{\bar{\omega}(d_0)} w_1(\theta)\left[\prod_{\alpha=1}^{n} f(x_{1\alpha}, \cdots, x_{p\alpha}, \theta)\right] d\theta. \quad (n = 1, 2, \cdots) \tag{6.21}$$

Consider the sequential probability ratio test for testing the simple hypothesis H_0^* against the single alternative H_1^*. For any θ in ω let $\alpha(\theta)$ be the probability of accepting H_1^* when θ is true, and for any θ in $\bar{\omega}$ let $\beta(\theta)$ be the probability of accepting H_0^* when θ is true. It is clear that $\alpha(\theta)$ and $\beta(\theta)$ depend on the constants A and B used in the sequential process and on the weight functions $w_0(\theta)$ and $w_1(\theta)$. For given A, B, $w_0(\theta)$ and $w_1(\theta)$ let $\beta[A, B, w_0(\theta), w_1(\theta)]$ be the least upper bound of $\beta(\theta)$ in $\bar{\omega}(d_0)$ and let $\alpha[A, B, w_0(\theta), w_1(\theta)]$ be the least upper bound of $\alpha(\theta)$ in ω. Consider the difference

$$\Delta\alpha[A, B, w_0(\theta), w_1(\theta)] = \alpha[A, B, w_0(\theta), w_1(\theta)] - \int_{\omega} w_0(\theta)\alpha(\theta)\, d\theta$$

and

$$\Delta\beta[A, B, w_0(\theta), w_1(\theta)] = \beta[A, B, w_0(\theta), w_1(\theta)] - \int_{\bar{\omega}(d_0)} w_1(\theta)\beta(\theta)\, d\theta.$$

Determine $w_0(\theta)$ and $w_1(\theta)$ so that Max $[\Delta\alpha, \Delta\beta]$ is a minimum. For these weight functions the least upper bound of $\alpha(\theta)$ in ω and the least upper bound of $\beta(\theta)$ in $\bar{\omega}(d_0)$ will be functions of A and B only. Finally, we determine A and B so that the least upper bound of $\alpha(\theta)$ in ω becomes α, and the least upper bound of $\beta(\theta)$ in $\bar{\omega}(d_0)$ becomes β.

The determination of $w_0(\theta)$ and $w_1(\theta)$ involves the solution of problems in the calculus of variations. However, in some important cases the solution of the problem can easily be derived, since weight functions $w_0(\theta)$ and $w_1(\theta)$ can be found for which $\Delta\alpha = \Delta\beta = 0$. Such a situation is given, for instance, in the following case: Let $S(d)$ be the set of all points θ for which $\delta(\theta, \omega) = d$. Suppose that we can find two weight functions $v_0(\theta)$ and $v_1(\theta)$ such that $\int_\omega v_0(\theta)\, d\theta = \int_{S(d_0)} v_1(\theta)\, dS = 1$ (dS denotes the infinitesimal surface element of $S(d_0)$) and the sequential probability ratio test based on

$$\frac{\int_{S(d_0)} v_1(\theta)[\prod_\alpha f(x_{1\alpha}, \cdots, x_{p\alpha}, \theta)]\, dS}{\int_\omega v_0(\theta)[\prod_\alpha f(x_{1\alpha}, \cdots, x_{p\alpha}, \theta)]\, d\theta}$$

has the following properties: (1) $\alpha(\theta)$ is constant in ω; (2) $\beta(\theta)$ is constant on $S(d)$ for any $d \geq d_0$; (3) $\beta(\theta)$ is strictly decreasing with increasing d in the domain $d \geq d_0$. Then for these weight functions we evidently have $\Delta\alpha = \Delta\beta = 0$.

Let us illustrate this by a simple example. Let X be a normally distributed variate with unknown mean μ and unknown variance σ^2. Suppose that we want to test the hypothesis H_0 that $\mu = 0$ and that the distance of the point (μ, σ) from the set ω is defined by $\left|\frac{\mu}{\sigma}\right|$.

The set $S(d_0)$ then consists of all points (μ, σ) for which $\mu = +d_0\sigma$ or $\mu = -d_0\sigma$. The set ω consists of all points $(0, \sigma)$ where σ can take any arbitrary positive value. Let r be a positive value. We define the weight functions $v_{0r}(\sigma)$ and $v_{1r}(\sigma)$ as follows: $v_{0r}(\sigma) = \frac{1}{r}$ if $0 \leq \sigma \leq r$ and equals zero for all other values of σ. The weight function $v_{1r}(\sigma)$ is equal to $\frac{1}{2r}$ if $0 \leq \sigma \leq r$ and $\mu = \pm d_0\sigma$ and equal to zero otherwise.

Then

$$p_{1n} = \int_{S(d_0)} v_{1r}(\sigma)\, \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[-\frac{1}{2}\frac{\Sigma(x_\alpha - \mu)^2}{\sigma^2}\right] d\sigma$$

$$= \frac{1}{(2\pi)^{n/2}} \frac{1}{2r} \int_0^r \left\{\frac{1}{\sigma^n} \exp\left[-\frac{1}{2}\frac{\Sigma(x_\alpha - d_0\sigma)^2}{\sigma^2}\right] + \frac{1}{\sigma^n} \exp\left[-\frac{1}{2}\frac{\Sigma(x_\alpha + d_0\sigma)^2}{\sigma^2}\right]\right\} d\sigma \tag{6.22}$$

and

$$p_{0n} = \frac{1}{(2\pi)^{n/2}} \frac{1}{r} \left\{\int_0^r \frac{1}{\sigma^n} \exp\left[-\frac{1}{2}\frac{\Sigma x_\alpha^2}{\sigma^2}\right] d\sigma\right\}. \tag{6.23}$$

Hence

$$\frac{p_{1n}}{p_{0n}} = \frac{\frac{1}{2}\int_0^r \frac{1}{\sigma^n} \exp\left[-\frac{1}{2}\frac{\Sigma(x_\alpha - d_0\sigma)^2}{\sigma^2}\right] d\sigma}{\int_0^r \frac{1}{\sigma^n} \exp\left[-\frac{1}{2}\frac{\Sigma x_\alpha^2}{\sigma^2}\right] d\sigma} + \frac{\frac{1}{2}\int_0^r \frac{1}{\sigma^n} \exp\left[-\frac{1}{2}\frac{\Sigma(x_\alpha + d_0\sigma)^2}{\sigma^2}\right] d\sigma}{\int_0^r \frac{1}{\sigma^n} \exp\left[-\frac{1}{2}\frac{\Sigma x_\alpha^2}{\sigma^2}\right] d\sigma}. \tag{6.24}$$

We consider the limiting case when $r \to \infty$. Then

$$\frac{p_{1n}}{p_{0n}} = \frac{\frac{1}{2}\int_0^\infty \frac{1}{\sigma^n} \exp\left[-\frac{1}{2}\frac{\Sigma(x_\alpha - d_0\sigma)^2}{\sigma^2}\right] d\sigma}{\int_0^\infty \frac{1}{\sigma^n} \exp\left[-\frac{1}{2}\frac{\Sigma x_\alpha^2}{\sigma^2}\right] d\sigma} + \frac{\frac{1}{2}\int_0^\infty \frac{1}{\sigma^n} \exp\left[-\frac{1}{2}\frac{\Sigma(x_\alpha + d_0\sigma)^2}{\sigma^2}\right] d\sigma}{\int_0^\infty \frac{1}{\sigma^n} \exp\left[-\frac{1}{2}\frac{\Sigma x_\alpha^2}{\sigma^2}\right] d\sigma}. \tag{6.25}$$

The sequential test based on the ratio (6.25) provides a solution of the problem if it can be shown to have the following three properties: (1) $\alpha(\theta)$ is constant in ω; (2) $\beta(\theta)$ is only a function of $\left|\frac{\mu}{\sigma}\right|$; (3) $\beta(\theta)$ is monotonically decreasing with increasing $\left|\frac{\mu}{\sigma}\right|$. Denote $\frac{\sum_{\alpha=1}^n x_\alpha}{n}$ by $\bar{x}$ and $\sum_{\alpha=1}^n (x_\alpha - \bar{x})^2$ by S^2. Since the distribution of $\left|\frac{\bar{x}}{S}\right|$ depends only on $\left|\frac{\mu}{\sigma}\right|$, the first two properties are proved if we

show that the ratio (6.25) is a single valued function of $\left|\frac{\bar{x}}{S}\right|$.

First we show that the numerator of the ratio (6.25) is a homogenous function of $(x_1, \cdots, x_n)$ of degree $-(n-1)$. In fact, making the transformation $\sigma = \lambda t$ we obtain

$$\int_0^\infty \left\{\frac{1}{\sigma^n}\exp\left[-\frac{1}{2}\frac{\Sigma(\lambda x_\alpha - d_0\sigma)^2}{\sigma^2}\right] + \frac{1}{\sigma^n}\exp\left[-\frac{1}{2}\frac{\Sigma(\lambda x_\alpha + d_0\sigma)^2}{\sigma^2}\right]\right\} d\sigma$$

$$= \int_0^\infty \left\{\frac{1}{(\lambda t)^n}\exp\left[-\frac{1}{2}\frac{\Sigma(x_\alpha - d_0 t)^2}{t^2}\right] + \frac{1}{(\lambda t)^n}\exp\left[-\frac{1}{2}\frac{\Sigma(x_\alpha + d_0 t)^2}{t^2}\right]\right\} d(\lambda t)$$

$$= \frac{1}{\lambda^{n-1}}\int_0^\infty \left\{\frac{1}{t^n}\exp\left[-\frac{1}{2}\frac{\Sigma(x_\alpha - d_0 t)^2}{t^2}\right] + \frac{1}{t^n}\exp\left[-\frac{1}{2}\frac{\Sigma(x_\alpha + d_0 t)^2}{t^2}\right]\right\} dt.$$

This proves that the numerator of (6.25) is a homogenous function of $-(n-1)$ degree. Similarly, it can be shown that the denominator of (6.25) is also a homogenous function of degree $-(n-1)$. Thus the ratio (6.25) is a homogenous function of zero degree in the variables $x_1, \cdots, x_n$.

It can be seen that (6.25) is a function of the two expressions Σx_α^2 and Σx_α only; i.e.,

$$\frac{p_{1n}}{p_{0n}} = \phi(\Sigma x_\alpha^2, \Sigma x_\alpha). \tag{6.26}$$

Let $v = \left|\sqrt{\Sigma x_\alpha^2}\right|$. Since (6.26) is a homogenous function of zero degree, its value is not changed by substituting $\frac{x_\alpha}{v}$ for x_α. Hence,

$$\frac{p_{1n}}{p_{0n}} = \phi\left[\sum_\alpha\left(\frac{x_\alpha}{v}\right)^2, \sum\frac{x_\alpha}{v}\right] = \phi\left[1, \frac{n\bar{x}}{v}\right]. \tag{6.27}$$

Since $\phi(\Sigma x_\alpha^2, -\Sigma x_\alpha) = \phi(\Sigma x_\alpha^2, \Sigma x_\alpha)$, we see that

$$\frac{p_{1n}}{p_{0n}} = \psi\left[\frac{\bar{x}^2}{v^2}\right].$$

Since $\frac{(\bar{x})^2}{v^2}$ is a single valued function of $\left|\frac{\bar{x}}{S}\right|$, we have proved that $\frac{p_{1n}}{p_{0n}}$ is a single valued function of $\left|\frac{\bar{x}}{S}\right|$.

In order to prove property (3) of the sequential test based on the ratio (6.25), we have merely to show that (6.25) is a strictly increasing function of $\left|\frac{\bar{x}}{S}\right|$. Since $\frac{\bar{x}^2}{v^2}$ is a strictly increasing function of $\left|\frac{\bar{x}}{S}\right|$, we have only to show that (6.25) is a strictly increasing function of $\frac{\bar{x}^2}{v^2}$. The latter statement is obviously proved if we show that (6.25) increases with increasing value $|\bar{x}|$ while keeping

v fixed. For fixed value of v the denominator of (6.25) is constant. Thus, we have merely to show that the numerator of (6.25) increases with increasing $|\bar{x}|$ while keeping v fixed. This follows easily from the fact that

$$\exp\left[\frac{(\Sigma x_\alpha)\, d_0}{\sigma}\right] + \exp\left[\frac{-(\Sigma x_\alpha)\, d_0}{\sigma}\right]$$

is a strictly increasing function of $|\bar{x}|$.

REFERENCES

[1] H. F. Dodge and H. G. Romig, "A method of sampling inspection," *The Bell System Tech. Jour.*, Vol. 8 (1929), pp. 613–631.

[2] Walter Bartky, "Multiple sampling with constant probability", *Annals of Math. Stat.*, Vol. 14 (1943), pp. 363–377.

[3] Harold Hotelling, "Experimental determination of the maximum of a function", *Annals of Math. Stat.*, Vol. 12 (1941).

[4] Abraham Wald, "On cumulative sums of random variables", *Annals of Math. Stat.*, Vol. 15 (Sept. 1944).

[5] Z. W. Birnbaum, "An inequality for Mill's ratio", *Annals of Math. Stat.*, Vol. 13 (1942).

[6] P. C. Mahalanobis, "A sample survey of the acreage under jute in Bengal, with discussion on planning of experiments," Proc. 2nd Ind. Stat. Conf., Calcutta, Statistical Publishing Soc. (1940).

[7] Abraham Wald, *Sequential Analysis of Statistical Data: Theory*. A report submitted by the Statistical Research Group, Columbia University to the Applied Mathematics Panel, National Defense Research Committee, Sept. 1943.

[8] Harold Freeman *Sequential Analysis of Statistical Data: Applications*. A Report submitted by the Statistical Research Group, Columbia University to the Applied Mathematics Panel, National Defense Research Committee, July 1944.

[9] G. A. Barnard, M.A., *Economy in Sampling with Reference to Engineering Experimentation*, (British) Ministry of Supply, Advisory Service on Statistical Method and Quality Control, Technical Report, Series 'R', No. Q.C./R/7 Part 1.

[10] C. M. Stockman, *A Method of Obtaining an Approximation for the Operating Characteristic of a Wald Sequential Probability Ratio Test Applied to a Binomial Distribution*, (British) Ministry of Supply, Advisory Service on Statistical Method and Quality Control, Technical Report, Series 'R' No. Q.C./R/19.

[11] Abraham Wald, *A General Method of Deriving the Operating Characteristics of any Sequential Probability Ratio Test*. A Memorandum submitted to the Statistical Research Group, Columbia University, April 1944.