CONTRIBUTIONS TO THE THEORY OF SEQUENTIAL ANALYSIS. I

By M. A. Girshick

United States Department of Agriculture

PART I Applications of Sequential Analysis to the Ranking of Two Populations With Respect to a Single Parameter.

1. Summary. Given two populations π_1 and π_2 each characterized by a distribution density $f(x, \theta)$ which is assumed to be known, except for the value of the parameter θ. It is desired to test the composite hypothesis $\theta_1 < \theta_2$ against the alternative hypothesis $\theta_1 > \theta_2$ where θ_i is the value of the parameter in the distribution density of π_i, $(i = 1, 2)$.

The criterion proposed for testing this hypothesis is based on the sequential probability ratio and consists of the following:

Choose two positive constants a and b and two values of θ, say θ_1^0 and θ_2^0. Take pairs of observations $x_{1\alpha}$ from π_1 and $x_{2\alpha}$ from π_2, $(\alpha = 1, 2, \dots)$, in sequence and compute $Z_j = \sum_{\alpha=1}^{j} z_\alpha$ where

$$z_\alpha = \log\left[\frac{f(x_{2\alpha}, \theta_1^0)f(x_{1\alpha}, \theta_2^0)}{f(x_{2\alpha}, \theta_2^0)f(x_{1\alpha}, \theta_1^0)}\right].$$

The hypothesis tested is accepted or rejected depending on whether $Z_n \geq a$ or $Z_n \leq -b$ where n is the smallest integer j for which either one of these relationships is satisfied.

The boundaries a and b are partly given in terms of the desired risks of making an erroneous decision. The values θ_1^0 and θ_2^0 define the magnitude of the difference between the values of θ in π_1 and in π_2 which is considered worth detecting. It is shown that the power of this test is constant on a curve $h(\theta_1, \theta_2) = $ constant. If $E\left(\log\dfrac{f(x, \theta_2^0)}{f(x, \theta_1^0)}\right)$ is a monotonic function of θ, then the test is unbiased in the sense that all points (θ_1, θ_2) which lie on the curve $h(\theta_1, \theta_2) = $ constant are such that either every $\theta_1 < \theta_2$ or every $\theta_1 > \theta_2$. For a large class of known distributions the quantity h is shown to be an appropriate measure of the difference between θ_1 and θ_2 and the test procedure for this class of distributions is simple and intuitively sensible.

For the case of the binomial, the exact power of this test as well as the distribution of n is given.

1.1 General discussion. Consider two processes (populations) π_1 and π_2 each yielding a measurable quantity x whose distribution density $f(x, \theta)$ is assumed to be known except for the value of the parameter θ. On the basis of a random sample obtained from each, it is desired to choose that process which yields the smaller (or larger) θ. That is, it is desired to devise a test which will result in a high probability of accepting π_1 if the θ characterizing its distribution density is smaller (or larger) than the θ in π_2, a high probability of rejecting π_1 (i.e. accepting π_2) when the opposite is true, and approximately equal probability of making one or the other decision if the value of θ in π_1 is the same as in π_2.

As an illustration of the type of problem here considered, let us assume that a manufacturer is faced with a choice between two competing processes of production, each process yielding an unknown fraction defective p and each entailing about the same operating cost. Based on the evidence of a random sample selected from each, the manufacturer wishes to choose that process which yields the smaller fraction defective. If the fractions defective in the two processes differ by a significant amount, he will want a test which guarantees a high probability of making a correct decision. If, however, the fraction defective in the two processes are of approximately the same magnitude, it will be a matter of indifference to him which decision is reached.

The solution given in this paper to the above problem is based on Wald's sequential probability ratio test [1]. The resulting procedure not only requires on the average, fewer observations for the same protection than any other test (which is always the case with sequential tests of this type) but is also direct and simple when applied to a large class of distributions commonly met in practice.

1.2 Derivation of the sequential test when the existence of a priori probabilities is assumed. The choice of the probability ratio as a method of discriminating between the two processes is suggested by considerations of a priori probabilities. Let us assume that each process may have either θ_1^0 or θ_2^0 as the value of a parameter θ in its distribution density and that the value θ_1^0 is more desirable than θ_2^0. Let us further assume that there exists an a priori probability g_1 that a process will have θ_1^0 as a parameter and an a priori probability $g_2 = 1 - g_1$ that it will have θ_2^0 as a parameter. Let the likelihood for n observations $x_{11}, x_{12}, \cdots, x_{1n}$ drawn from π_1 be designated by $p(x_{11}, z_{12}, \cdots, x_{1n}, \theta_1^0)$ when θ_1^0 is the parameter in π_1, and by $p(x_{11}, x_{12}, \cdots, x_{1n}, \theta_2^0)$ when θ_2^0 is the parameter in π_1. Let the likelihoods $p(x_{21}, x_{22}, \cdots, x_{2n}, \theta_1^0)$ and $p(x_{21}, x_{22}, \cdots, x_{2n}, \theta_2^0)$ be similarly defined for n observations $x_{21}, x_{22}, \cdots, x_{2n}$ drawn from π_2. Then

$$p(x_{i1}, x_{i2}, \cdots, x_{in}, \theta_j^0) = \prod_{\alpha=1}^{n} f(x_{i\alpha}, \theta_j^0), \qquad i, j = 1, 2. \tag{1.201}$$

Let β_{ij}, $(i, j = 1, 2)$, be the a posteriori probability that having obtained $x_{i\alpha}$, $(\alpha = 1, 2, \cdots, n)$, that process π_i has θ_j^0 as a parameter in its distribution density. Then

$$\beta_{i1} = \frac{g_1 p(x_{i1}, x_{i2}, \cdots, x_{in}, \theta_1^0)}{g_1 p(x_{i1}, \cdots, x_{in}, \theta_1^0) + g_2 p(x_{i1}, \cdots, x_{in}, \theta_2^0)} \tag{1.202}$$

and

$$\beta_{i2} = \frac{g_2 p(x_{i1}, \cdots, x_{in}, \theta_2^0)}{g_1 p(x_{i1}, \cdots, x_{in}, \theta_1^0) + g_2 p(x_{i1}, \cdots, x_{in}, \theta_2^0)} \tag{1.203}$$

for $i = 1, 2$.

In order to decide whether the hypothesis that θ_1^0 belongs to the distribution density of π_1 is more tenable than the hypothesis that it belongs to the distribution of π_2, it is only necessary to compare β_{11} with β_{21}. But if β_{11} is equal to or greater than β_{21}, the ratio β_{11}/β_{12} must be equal to or greater than β_{21}/β_{22} and conversely. For assume that $\beta_{11} \geq \beta_{21}$. Subtracting $\beta_{11}\beta_{21}$ from each side of the inequality we get $\beta_{11}(1 - \beta_{21}) \geq \beta_{21}(1 - \beta_{11})$. But since $1 - \beta_{21} = \beta_{22}$ and $1 - \beta_{11} = \beta_{21}$, we see that $\beta_{11}/\beta_{12} \geq \beta_{21}/\beta_{22}$. Conversely, let $\beta_{11}/\beta_{12} \geq \beta_{21}/\beta_{22}$. Then $\beta_{11}(1 - \beta_{21}) \geq \beta_{21}(1 - \beta_{11})$, or $\beta_{11} \geq \beta_{21}$.

From the above it would appear that a sensible sequential procedure for deciding whether θ_1^0 is more likely to belong to π_1 than to π_2 is as follows: Select two positive quantities A and B with $A > 1$ and $B < 1$. Take a pair of observations $(x_{1\alpha}, x_{2\alpha})$, $(\alpha = 1, 2, \cdots)$, at a time, one from each process. At each step (i.e., for each sample size n) compute the ratio $\lambda = \dfrac{\beta_{21}}{\beta_{22}} \Big/ \dfrac{\beta_{11}}{\beta_{12}}$. If at any stage $\lambda \leq B$, terminate the sampling and accept the hypothesis that θ_1^0 is a parameter in the distribution density of π_1. On the other hand, if at any stage $\lambda \geq A$, terminate sampling and accept the hypothesis that θ_1^0 is a parameter of the distribution density in π_2. If neither holds, that is if $B < \lambda < A$, then take another pair of observations, consisting of one from each process. Continue this procedure until one or the other decision is reached.[1]

The interesting point here is that the decision function λ is independent of g_1 and g_2. In fact, it is easily seen from equations (1.202) and (1.203) that

$$\lambda = \frac{p(x_{21}, x_{22}, \cdots, x_{2n}, \theta_1^0)p(x_{11}, x_{12}, \cdots, x_{1n}, \theta_2^0)}{p(x_{21}, x_{22}, \cdots, x_{2n}, \theta_2^0)p(x_{11}, x_{12}, \cdots, x_{1n}, \theta_1^0)}. \tag{1.204}$$

1.3 The proposed sequential test as a special case of a sequential probability ratio test. If we examine the expression given in (1.204) we see that it is a ratio of two likelihoods. The numerator of the ratio is the likelihood of the $2n$ observations under the hypothesis that θ_2^0 is a parameter in π_1 and θ_1^0 is a praameter in π_2; the denominator is the likelihood of the $2n$ observations under the hypothesis that θ_1^0 is a parameter in π_1 and θ_2^0 is a parameter in π_2. Thus, the proposed sequential test is equivalent to a sequential probability ratio test (see [1]) for testing the simple hypothesis that θ_1^0 belongs to π_1 and θ_2^0 belongs to π_2 against the alternative hypothesis that θ_2^0 belongs to π_1 and θ_1^0 belongs to π_2. We can, therefore, apply the theory of sequential analysis developed by A. Wald ([1] and [2]) to this problem.

While the test is posed in terms of a simple hypothesis, the solution, as will be shown later, is in fact a solution to a composite hypothesis. In order to bring this out more clearly we shall rederive a few of the results which have already been obtained by A. Wald. This will be done in sections 1.4, 1.5, and 1.6.

[1] That a decision will be reached eventually can be asserted with probability one if the variance of the variate z_α (defined by (1.301)) below is different from zero (or if it is zero, the value of z_α is different from zero). See [2], Lemma 1. As we shall see later, if, in fact, both processes have either θ_1^0 or θ_2^0 as parameters, then the above sequential procedure will result in the acceptance of either process with approximately equal probability.

In what follows we shall speak of the hypothesis (θ_1, θ_2) to mean the hypothesis that θ_1 is the value of the parameter in the distribution density of π_1 and θ_2 is the value of the parameter in the distribution density of π_2. The hypothesis (θ_1^0, θ_2^0) will represent a specific hypothesis which we may wish to test and will be used to define the decision function (the probability ratio) of the sequential test.

Let us fix $A > 1$ and $B < 1$ and set

$$z_\alpha = \log\left[\frac{f(x_{2\alpha}, \theta_1^0)f(x_{1\alpha}, \theta_2^0)}{f(x_{2\alpha}, \theta_2^0)f(x_{1\alpha}, \theta_1^0)}\right] \tag{1.301}$$

where $x_{1\alpha}$ is the αth observation from π_1, $x_{2\alpha}$ is the αth observation from π_2 and (θ_1^0, θ_2^0) is the particular hypothesis to be tested against the alternative hypothesis (θ_2^0, θ_1^0). Let $a = \log A$ and $-b = \log B$. Then a and b are positive. Since the observations from π_1 and π_2 are assumed to be independent, $\log \lambda = \sum_{\alpha=1}^{n} z_\alpha$. Hence the proposed sequential test can be carried out in the following manner. Draw one pair of observations at a time, one from π_1 and one from π_2. Let $z_1, z_2, \cdots$ be the values of z_α obtained from the first, second, etc. trial. Let $Z_n = z_1 + z_2 + \cdots + z_n$, $(n = 1, 2, \cdots)$. Continue sampling as long as $-b < Z_n < a$. Whenever $Z_n \geq a$, $(n = 1, 2, 3, \cdots)$, terminate sampling and accept π_2 (or π_1). Whenever $Z_n \leq -b$, $(n = 1, 2, 3, \cdots)$, terminate sampling and accept π_1 (or π_2).

1.3a. *Basic assumptions.* In this section and throughout this paper, we shall be dealing with sequential tests involving, as above, a decision function $Z_n = z_1 + z_2 + \cdots + z_n$, $(n = 1, 2, \cdots$, ad inf.), where the z_α's are independently distributed random variables having a common distribution function. Let z denote a random variable whose distribution is the same as the common distribution of z_α, $(\alpha = 1, 2, \cdots$, ad inf.). It will be assumed, even if not explicitly stated, that the distribution of z satisfies the following conditions.

CONDITION I. *Both the expected value Ez of z and the variance of z exist and are unequal to zero.*

CONDITION II. *There exists a positive δ such that $P(e^z > 1 + \delta) > 0$ and $P(e^z < 1 - \delta) > 0$.*

CONDITION III. *For any real value h, the expected value $Ee^{hz} = g(h)$ exists.*

CONDITION IV. *The first two derivatives of the function $g(h)$ exist and may be obtained by differentiating under the integral sign.*

1.3b. *Fundamental properties of sequential tests.* Let z be defined as in 1.3a. Then under the assumption that the distribution of z satisfies the conditions specified, Wald [2] has proved the following:

LEMMA I. *The probability that a decision is reached in a finite number of steps is unity.*

LEMMA II. *There exists one and only one real value $h \neq 0$ such that the expected value $Ee^{hz} = 1$.*

FUNDAMENTAL IDENTITY: *The fundamental identity $Ee^{Z_n t}[\phi(t)]^{-n} = 1$ holds for all points in the complex plane for which $|\phi(t)| \geq 1$ where $\phi(t) = Ee^{tz}$.*

Let $w = \log \frac{f(x, \theta_1^0)}{f(x, \theta_2^0)}$ and let the distribution density of x be $f(x, \theta)$. Let θ_1 and θ_2 be any two values of θ which may be distinct from θ_1^0 and θ_2^0. Then it can easily be verified that if w satisfies the conditions specified in section 1.3a under the hypothesis $\theta = \theta_1$ as well as the hypothesis $\theta = \theta_2$, and if moreover the expected values of w under these two hypotheses are not equal, then $z = \log \frac{f(x_2, \theta_1^0)f(x_1, \theta_2^0)}{f(x_2, \theta_2^0)f(x_1, \theta_1^0)}$ will also satisfy these conditions when the joint distribution density of x_1 and x_2 (x_1 representing the measurable characteristic in π_1 and x_2 in π_2) is either $f(x_1, \theta_1) f(x_2, \theta_2)$ or $f(x_1, \theta_2) f(x_2, \theta_1)$.

In what follows, we shall assume that the distribution of w satisfies the required restrictions for the θ_1 and θ_2 under consideration and that the expectation of w under the hypothesis $\theta = \theta_1$ is unequal to the expectation of w under the hypothesis $\theta = \theta_2$. Consequently, we shall assume that Lemmas I and II and the Fundamental Identity hold for all the sequential tests we shall consider.

1.4 The power of the proposed test. Let x_1 be an observation from π_1 and x_2 an observation from π_2. Let

$$z = \log \frac{f(x_2, \theta_1^0)f(x_1, \theta_2^0)}{f(x_2, \theta_2^0)f(x_1, \theta_1^0)} \tag{1.401}$$

where θ_1^0 and θ_2^0 are specified parameters in the probability density of π_1 and π_2 respectively. Furthermore, let $\phi(t \mid \theta_1, \theta_2) = E(e^{tz} \mid \theta_1, \theta_2)$ be the moment generating function of z under the hypothesis (θ_1, θ_2). Then

$$E(e^{tz} \mid \theta_1, \theta_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{f(x_2, \theta_1^0)f(x_1, \theta_2^0)}{f(x_2, \theta_2^0)f(x_1, \theta_1^0)}\right]^t f(x_1, \theta_1)f(x_2, \theta_2)\, dx_1\, dx_2. \tag{1.402}$$

By Lemma II there exists one and only one real number $h \neq 0$ such that $E(e^{hz} \mid \theta_1, \theta_2) = 1$. Let $L_h = P(Z_n \leq -b \mid \theta_1, \theta_2)$ be the probability that the sequential test terminates and $Z_n \leq -b$ under the hypothesis (θ_1, θ_2). Then by Lemma I, $1 - L_h = P(Z_n \geq a \mid \theta_1, \theta_2)$. For any random variable u considered under the hypothesis (θ_1, θ_2), let the symbol $E_b(u)$ stand for the expected value of u under the restriction that $Z_n \leq -b$ and $E_a(u)$ stand for the expected value of u under the restriction that $Z_n \geq a$. In terms of the above definitions, the Fundamental Identity can be expressed as follows:

$$L_h E_b e^{tZ_n}[\phi(t \mid \theta_1, \theta_2)]^{-n} + (1 - L_h)E_a e^{tZ_n}[\phi(t \mid \theta_1, \theta_2)]^{-n} = 1. \tag{1.403}$$

Setting $t = h$ in (1.403) we get

$$L_h E_b e^{hZ_n} + (1 - L_h)E_a e^{hZ_n} = 1. \tag{1.404}$$

Following Wald [2], we define a two valued random variable $\bar{Z}_n$ in this manner: $\bar{Z}_n = a$ if $Z_n \geq a$ and $\bar{Z}_n = -b$ if $Z_n \leq -b$. Let $\bar{Z}_n - Z_n = \epsilon$. Then ϵ is also a random variable. In what follows, we shall substitute 0 for ϵ. The error committed in neglecting ϵ is small when θ_1^0 is close to θ_2^0. As we shall indicate later, the quantity ϵ can, in fact, be neglected without error in the special case where $f(x, \theta)$ is the binomial distribution.

Substituting $\bar{Z}_n$ for Z_n in (1.404) we get

$$L_h e^{-hb} + (1 - L_h)e^{ha} = 1. \tag{1.405}$$

Solving for L_h we get[2]

$$L_h = \frac{1 - e^{ha}}{e^{-hb} - e^{ha}} = \frac{e^{h(a+b)} - e^{hb}}{e^{h(a+b)} - 1}. \tag{1.406}$$

As we shall see later, $h = 0$ when $\theta_1 = \theta_2$. But when $h = 0$, L_h in (1.406) is indeterminate. However, it can be easily seen that

$$\lim_{h\to 0} L_h = \frac{a}{a + b}. \tag{1.407}$$

It follows from (1.406) that the power of the test is constant for all θ_1 and θ_2 which give the same root $t = h$. The quantity h is thus fundamental in this test, and as we shall see later, is an appropriate measure of the difference between θ_1 and θ_2 for a large class of distributions.

1.5 Method of determining the sequential test. Let z be defined as in (1.401) and let $\phi_1(t) = E(e^{tz} \mid \theta_1^0, \theta_2^0)$ be the moment generating function of z under the hypothesis (θ_1^0, θ_2^0), and let $\phi_2(t) = E(e^{tz} \mid \theta_2^0, \theta_1^0)$ be the moment generating function of z under the hypothesis (θ_2^0, θ_1^0). Furthermore, let $\alpha = P(\bar{Z}_n = a \mid \theta_1^0, \theta_2^0)$ and $\beta = P(\bar{Z}_n = -b \mid \theta_2^0, \theta_1^0)$. Then by Lemma I, $1 - \alpha = P(\bar{Z}_n = -b \mid \theta_1^0, \theta_2^0)$ and $1 - \beta = P(\bar{Z}_n = a \mid \theta_2^0, \theta_1^0)$. Now, applying Wald's Fundamental Identity we have,

$$(1 - \alpha)e^{-tb}E_{1b}[\phi_1(t)]^{-n} + \alpha e^{ta}E_{1a}[\phi_1(t)]^{-n} = 1, \tag{1.501}$$

$$\beta e^{-tb}E_{2b}[\phi_2(t)]^{-n} + (1 - \beta)e^{ta}E_{2a}[\phi_2(t)]^{-n} = 1, \tag{1.502}$$

where the symbol E_{1a} stands for the conditional expectation knowing that $\bar{Z}_n = a$ and E_{1b} stand for the conditional expectation knowing that $\bar{Z}_n = -b$; with both expectations taken under the hypothesis (θ_1^0, θ_2^0). The symbols E_{2a} and E_{2b} are similarly defined but under the hypothesis (θ_2^0, θ_1^0). Setting $t = 1$ in (1.501) and $t = -1$ in (1.502), we get, in view of Corollary 2, Theorem 2 below,

$$(1 - \alpha)e^{-b} + \alpha e^{a} = 1, \tag{1.503}$$

$$\beta e^{b} + (1 - \beta)e^{-a} = 1. \tag{1.504}$$

[2] In what follows, L_h will always stand for the probability that a sequential test will terminate with $Z_n \leq -b$. In any given problem, the interpretation of the event $Z_n \leq -b$ will be clear from the context.

Now $a = \log A$ and $-b = \log B$. Hence, equations (1.503) and (1.504) become

$$(1.505) \qquad (1 - \alpha) B + \alpha A = 1,$$

$$(1.506) \qquad \frac{\beta}{B} + \frac{1 - \beta}{A} = 1,$$

or

$$(1.507) \qquad A = \frac{1 - \beta}{\alpha} \quad \text{and} \quad a = \log \frac{1 - \beta}{\alpha},$$

$$(1.508) \qquad B = \frac{\beta}{1 - \alpha} \quad \text{and} \quad b = \log \frac{1 - \alpha}{\beta}.$$

From (1.507) and (1.508) we see that the sequential test is completely determined by the function z, which, in turn, is defined by θ_1^0 and θ_2^0, and by the probabilities of making a decision for the two hypotheses (θ_1^0, θ_2^0) and (θ_2^0, θ_1^0).

Once z is defined in terms of a specific (θ_1^0, θ_2^0), the probability that $Z_n \leq -b$ will be equal to $1 - \alpha$ and the probability that $Z_n \geq a$ will be α (if we neglect the fact that $| Z_n |$, at a decision point, might exceed a or b) for the totality of hypotheses (θ_1, θ_2) for which the moment generating function $\phi(t \mid \theta_1, \theta_2) = 1$ when $t = 1$. A similar statement can be made for the corresponding hypotheses (θ_2, θ_1) for which the moment generating function will equal unity when $t = -1$. Hence, we see that while the test is defined by specifying two points (θ_1^0, θ_2^0) and (θ_2^0, θ_1^0) in the parameter space, the pre-assigned risks α and β of making the correct decision will be approximately constant on the set of points for which the moment generating function equals unity when $t = 1$ and when $t = -1$, respectively. This set of points usually will constitute a smooth curve.

If $\theta_1 = \theta_2$, $L_0 = \dfrac{a}{a + b}$ (by 1.407). Hence, the probability of accepting π_1 will be close to $\frac{1}{2}$ if a is close to b, and will equal $\frac{1}{2}$ if $a = b$. But from (1.507) and (1.508) we see that $a = b$ if $\alpha = \beta$. Thus, if we construct a test which will give a probability of rejecting π_1 when (θ_1^0, θ_2^0) is true equal to the probability of accepting π_1 when (θ_2^0, θ_1^0) is true, we shall be accepting π_1 and π_2 with equal frequency when in fact $\theta_1 = \theta_2$.

1.6 The average number of pairs of observations required to reach a decision. Let $E(n \mid \theta_1, \theta_2)$ be the expected number of pairs of observations required to reach a decision under the hypothesis (θ_1, θ_2). We shall show that

$$(1.601) \qquad E(n \mid \theta_1, \theta_2) = \frac{a(1 - L_h) - bL_h}{Ez}.$$

PROOF: Differentiating the Fundamental Identity,

$$(1.602) \qquad Ee^{tZ_n}[\phi(t)]^{-n} = 1,$$

with respect to t, we get[3]

$$E\{Z_n e^{tZ_n}[\phi(t)]^{-n} - n e^{tZ_n}\phi'(t)[\phi(t)]^{-n-1}\} = 0. \tag{1.603}$$

Setting $t = 0$, we get

$$EZ_n - \phi'(0)E(n \mid \theta_1, \theta_2) = 0. \tag{1.604}$$

But

$$E\bar{Z}_n = a(1 - L_h) - bL_h \tag{1.605}$$

and

$$\phi'(0) = Ez. \tag{1.606}$$

Hence, solving for $E(n \mid \theta_1, \theta_2)$ in (1.604) and substituting from (1.605) and (1.606) we get

$$E(n \mid \theta_1, \theta_2) = \frac{a(1 - L_h) - bL_h}{Ez}. \tag{1.607}$$

While L_h is approximately constant for all values of (θ_1, θ_2) for which the moment generating function equals unity for $t = h$ the expected value of n given by (1.607) will depend on the particular hypothesis (θ_1, θ_2). This follows from the fact that Ez is not necessarily constant for the same set of points (θ_1, θ_2) for which L_h is constant.

1.7 Some general properties of the proposed test.

THEOREM 1. *Let* $z = \log \frac{f(x_2, \theta_1^0)f(x_1, \theta_2^0)}{f(x_2, \theta_2^0)f(x_1, \theta_1^0)}$ *where* x_1 *is an observation from* π_1 *and* x_2 *from* π_2. *Then if* $F(z)$ *is the distribution density of* z *under the hypothesis* (θ_1, θ_2), $F(-z)$ *is the distribution density of* z *under the hypothesis* (θ_2, θ_1).

PROOF: Let t be a real number and let $\psi_1(t) = E(e^{itz} \mid \theta_1, \theta_2)$ be the characteristic function of z under the hypothesis (θ_1, θ_2). Then

$$\psi_1(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\frac{f(x_2, \theta_1^0)f(x_1, \theta_2^0)}{f(x_2, \theta_2^0)f(x_1, \theta_1^0)}\right]^{it} f(x_1, \theta_1)f(x_2, \theta_2)\, dx_1\, dx_2. \tag{1.701}$$

Now let $\psi_2(t) = E(e^{-itz} \mid \theta_2, \theta_1)$ be the characteristic function of $-z$ under the hypothesis (θ_2, θ_1). Then

$$\psi_2(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\frac{f(x_2, \theta_1^0)f(x_1, \theta_2^0)}{f(x_2, \theta_2^0)f(x_1, \theta_1^0)}\right]^{-it} f(x_1, \theta_2)f(x_2, \theta_1)\, dx_1\, dx_2. \tag{1.702}$$

Interchanging the variables of integration in (1.702) we see that $\psi_1(t) = \psi_2(t)$. Consequently, the distribution of z under the hypothesis (θ_1, θ_2) is the same as

---

[3] This assumes that the Fundamental Identity can be differentiated with respect to t. The results that follow can be derived without any reference to the Fundamental Identity. See Wald [1], page 142.

the distribution of $-z$ under the hypothesis (θ_2, θ_1). This theorem in conjunction with the fact that $E(z \mid \theta_1, \theta_2) \neq 0$ when $\theta_1 \neq \theta_2$ shows that the decision function z discriminates in a real sense between the two alternative hypotheses (θ_1, θ_2) and (θ_2, θ_1).

THEOREM 2. *Let $E(e^{tz} \mid \theta_1, \theta_2)$ be the moment generating function of z under the hypothesis (θ_1, θ_2) and let $E(e^{tz} \mid \theta_2, \theta_1)$ be the moment generating function of z under the hypothesis (θ_2, θ_1). Then, if $t = h$ is a root of the equation $E(e^{tz} \mid \theta_1, \theta_2) = 1$, then $t = -h$ is a root of the equation $E(e^{tz} \mid \theta_2, \theta_1) = 1$.*

PROOF: The same as Theorem 1. As we have seen in Section 1.4, the power of the proposed sequential test (neglecting ϵ) depends only on h. This theorem shows that if the probability of accepting π_1 is large under the hypothesis (θ_1, θ_2), it will be small under the hypothesis (θ_2, θ_1), and conversely.

COROLLARY 1. *The only value of t for which $E(e^{tz} \mid \theta, \theta) = 1$ is $t = 0$.* This follows from Theorem 2.

COROLLARY 2. *The values of t for which $E(e^{tz} \mid \theta_1^0, \theta_2^0) = 1$ and $E(e^{tz} \mid \theta_2^0, \theta_1^0) = 1$ are $t = 1$ and $t = -1$ respectively.* This can be seen by expressing $E(e^{tz} \mid \theta_1^0, \theta_2^0)$ as a double integral and setting $t = 1$.

THEOREM 3. *Let ω be the totality of points (θ_1, θ_2) in the parameter space for which $\theta_1 < \theta_2$. Then a necessary and sufficient condition that the values of h (for which $E(e^{hz} \mid \theta_1, \theta_2) = 1$) be of the same sign for all points in ω is that*

$$Ew \mid \theta = \int_{-\infty}^{\infty} \log \frac{f(x, \theta_2^0)}{f(x, \theta_1^0)} f(x, \theta)\, dx \tag{1.703}$$

be a monotonic function of θ.

To prove this theorem we need the following lemma.

LEMMA 1. *Let $g(x, \theta)$ be the distribution density of x and $\psi(t)$ its moment generating function. Let h be the real non-zero value of t for which $\psi(t) = 1$. Then the sign of h is opposite in sign to Ex (the expected value of x) if $Ex \neq 0$.*

PROOF: For any random variable u, Wald [1] has shown that the inequality

$$Eu \leq \log Ee^u \tag{1.704}$$

holds.

Setting $u = tx$, where t is a constant, we get

$$tEx \leq \log Ee^{tx} = \log \psi(t). \tag{1.705}$$

Setting $t = h$ in (1.705) we get $hEx \leq 0$. This proves the lemma.

Now let $E(z \mid \theta_1, \theta_2)$ be the expected value of z under the hypotheses (θ_1, θ_2) where (θ_1, θ_2) belongs to ω. Then

$$\begin{aligned} E(z \mid \theta_1, \theta_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \log \frac{f(x_2, \theta_1^0) f(x_1, \theta_2^0)}{f(x_2, \theta_2^0) f(x_1, \theta_1^0)} f(x_1, \theta_1) f(x_2, \theta_2)\, dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \log \frac{f(x, \theta_2^0)}{f(x, \theta_1^0)} f(x, \theta_1)\, dx \\ &\quad - \int_{-\infty}^{\infty} \log \frac{f(x, \theta_2^0)}{f(x, \theta_1^0)} f(x, \theta_2)\, dx = Ew \mid \theta_1 - Ew \mid \theta_2. \end{aligned} \tag{1.706}$$

From (1.706) we see that if $Ew \mid \theta$ is monotonic in θ, $E(z \mid \theta_1, \theta_2)$ will have a constant sign for all points (θ_1, θ_2) in ω and hence by Lemma 1, h will have a constant sign. Conversely, if h is of constant sign for all (θ_1, θ_2) in ω, so will $E(z \mid \theta_1, \theta_2)$ be. Consequently, by (1.706) $Ew \mid \theta$ must be monotonic.

COROLLARY 1. *Let $Ew \mid \theta$ be a monotonic function of θ and let ω_h, $(h \neq 0)$, be the totality of points (θ_1, θ_2) in the parameter space for which the power of the sequential test is constant. Then the coordinatss of the points (θ_1, θ_2) in ω_h are such that either every $\theta_1 < \theta_2$ or every $\theta_1 > \theta_2$.*

PROOF: By assumption all points in ω_h have the same power. Since L_h in (1.406) is a strictly increasing function of h, the points in ω_h must yield the same h. However, if we assume that ω_h contains a point (θ_1', θ_2') with $\theta_1' < \theta_2'$ and a point (θ_1'', θ_2'') with $\theta_1'' > \theta_2''$, the sign of $E(z \mid \theta_1', \theta_2')$ by (1.706) will be opposite to the sign of $E(z \mid \theta_1'', \theta_2'')$. Hence, the value of h yielded by (θ_1', θ_2') is opposite in sign to that yielded by (θ_1'', θ_2''), which contradicts the assumption that both points yield the same h.

Theorem 3 and Corollary 1 show that if $Ew \mid \theta$ is monotonic in θ, the proposed sequential test is unbiased in the sense that all points (θ_1, θ_2) that lie on the curve $h = $ constant (and hence have the same power) will have the property that either the inequality $\theta_1 < \theta_2$ holds or the inequality $\theta_1 > \theta_2$ holds. The equality sign will hold if and only if $h = 0$.

1.8 The proposed test applied to distributions which admit sufficient statistics. Let $f(x, \theta)$ admit a sufficient estimate of θ. Then it is well known that $f(x, \theta)$ can be written in the form[4]

$$f(x, \theta) = e^{u(x)v(\theta)+r(x)+w(\theta)}. \tag{1.801}$$

Setting $z = \log \frac{f(x_2, \theta_1^0)f(x_1, \theta_2^0)}{f(x_2, \theta_2^0)f(x_1, \theta_1^0)}$, we see that for this class of distributions the decision function assumes the simple form:

$$z = [u(x_2) - u(x_1)][v(\theta_1^0) - v(\theta_2^0)]. \tag{1.802}$$

Let $a^* = \frac{a}{v(\theta_1^0) - v(\theta_2^0)}$ and $b^* = \frac{b}{v(\theta_1^0) - v(\theta_2^0)}$. Then the decision function becomes

$$z^* = u(x_2) - u(x_1). \tag{1.803}$$

We shall now show that, for this class of distributions, the power of the sequential test is a function of $v(\theta_1) - v(\theta_2)$. To prove this, it is only necessary to show that $E(e^{tz^*} \mid \theta_1, \theta_2)$ equals unity for $t = v(\theta_1) - v(\theta_2)$. Now

$$\begin{aligned} E(e^{tz^*} \mid \theta_1, \theta_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t[u(x_2)-u(x_1)]} f(x_1, \theta_1)f(x_2, \theta_2)\, dx_1\, dx_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{u(x_1)[v(\theta_1)-t]+u(x_2)[t+v(\theta_2)]+r(x_1)+r(x_2)+w(\theta_1)+w(\theta_2)}\, dx_1\, dx_2. \end{aligned} \tag{1.804}$$

If we set $t = v(\theta_1) - v(\theta_2)$ in (1.804), we see that the statement is proved.

[4] See, for example, [3].

Let $En \mid h$ be the average number of pairs of observations required to reach a decision when $v(\theta_1) - v(\theta_2) = h$. Then by formula (1.607) we have

$$(1.805) \qquad E(n \mid h) = \frac{a^*(1 - L_h) - b^* L_h}{E[u(x_2) - u(x_1)]} = \frac{(1 - L_h) \log A + L_h \log B}{h_0 E[u(x_2) - u(x_1)]}.$$

Since the expected value of $u(x)$ will not necessarily equal $v(\theta)$, the average number of pairs of observations required to reach a decision will depend not only on $v(\theta_1) - v(\theta_2)$ but also on the particular hypothesis (θ_1, θ_2) considered.

Since the power of the test for this class of distributions depends on $v(\theta_1) - v(\theta_2)$, it will be constant for all θ_1 and θ_2 which lie on the curve defined by $v(\theta_1) - v(\theta_2) =$ constant. In particular, if the sequential test is defined with risks α and β, the probability of accepting π_1 (or π_2) will be approximately α for all hypotheses (θ_1, θ_2) which lie on the curve defined by $v(\theta_1) - v(\theta_2) = v(\theta_1^0) - v(\theta_2^0) = h_0$ and the probability of accepting π_2 (or π_1) will be approximately β for all hypotheses (θ_2, θ_1) which lie on the curve defined by $v(\theta_2) - v(\theta_1) = h_0$. Now, the decision function z as well as the boundaries a^* and b^* will be identical for all sequential tests provided they are defined by the same risks α and β and the parameters θ_1 and θ_2 which determine the decision function all lie on the curve $v(\theta_1) - v(\theta_2) = h_0$. Since Wald [1] has proved that the sequential probability ratio test minimizes $E(n)$, the expected number of observations required to reach a decision, when the hypothesis tested is true as well as when the alternative hypothesis is true, it must follow that in the case under consideration $E(n)$ is minimized for *all* hypotheses (θ_1, θ_2) which lie either on the curve defined by $v(\theta_1) - v(\theta_2) = h_0$ or on the curve defined by $v(\theta_2) - v(\theta_1) = h_0$. If $v(\theta)$ is a monotonic function of θ, then the test is unbiased (i.e. all points (θ_1, θ_2) which lie on the curve $v(\theta_1) - v(\theta_2) =$ constant will have the property that either every $\theta_1 < \theta_2$ or every $\theta_1 > \theta_2$).

For this type of distribution, the importance of the difference between θ_1 and θ_2 may be measured by $v(\theta_1) - v(\theta_2)$. We shall now show that the function $v(\theta_1) - v(\theta_2)$ is an appropriate measure of the difference between these parameters for a wide class of distributions which often occur in practice.

1.9 The proposed test applied to known distributions.

1.9a. *The problem of discriminating between means when the variances are known.* Let $f(x, \mu)$ be a normal distribution function with unknown mean μ and known variance σ^2 which we shall assume, without loss of generality, to be unity. Let x_1 be an observation from π_1 and x_2 an observation from π_2. Let the distribution density of x_1 be designated by $f(x_1, \mu_1)$ and that of x_2 by $f(x_2 \cdot \mu_2)$. The problem is to decide which process has the larger μ.

Since $f(x, \mu)$ is a normal distribution, it is given by

$$(1.901) \qquad f(x, \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}.$$

Hence $f(x, \mu)$ is of the form considered in Section 1.8 with $u(x) = x$ and $v(\mu) = \mu$. Therefore, the decision function is given by

$$(1.902) \qquad z^* = x_2 - x_1$$

and the power of the test depends on $h = \mu_1 - \mu_2$ and is given by (1.406) with a and b replaced by a^* and b^*, respectively.

The sequential test is performed in the following manner: We take a pair of observations, one from π_1 and one from π_2, in sequence. If at any stage $\sum_{\alpha=1}^{n} (x_{2\alpha} - x_{1\alpha}) \leq -b^*$, we accept the hypothesis that π_1 has the larger mean. If, however, at any stage $\sum_{\alpha=1}^{m} (x_{2\alpha} - x_{1\alpha}) \geq a^*$, we accept the hypothesis that π_2 has the larger mean. If neither holds, we continue sampling. According to section 1.8 $a^* = \dfrac{\log A}{\mu_1 - \mu_2}$ and $-b^* = \dfrac{\log B}{\mu_1 - \mu_2}$, where $\mu_1 - \mu_2$ is assumed to be positive.

In order to determine a sequential test, we must fix a^* and b^*. That is, we must fix the quantities $\mu_1 - \mu_2$, A, and B. This can be accomplished by deciding: (1) the smallest difference between the means of the two processes which is considered worth detecting. This determines $h_0 = \mu_1^0 - \mu_2^0$, which we shall assume to be positive: (2) the maximum probability α of rejecting the hypothesis that π_1 has the larger mean when in fact μ_1 in π_1 differs from μ_2 in π_2 by as much as h_0; and (3) the maximum probability β of accepting the hypothesis that π_1 has the larger mean when in fact the difference between μ_1 and μ_2 is as large as h_0 negatively.[5] When α and β are fixed, A and B are determined by equations (1.507) and (1.508).

1.9b. *The problem of discriminating between variances when the means are known.* Let us assume that the distribution of x_1 in π_1 and x_2 in π_2 are normal with known means but unknown variances. We are required to choose that process which has the smaller variance. Without any loss of generality we shall suppose that the means of x_1 and x_2 are zero. Since $f(x, \sigma)$ is normal, it is given by

$$\text{(1.903)} \qquad \frac{1}{\sqrt{2\pi}\sigma} e^{-(x^2/2\sigma^2)} = e^{-(x^2/2\sigma^2) - \log \sigma\sqrt{2\pi}}$$

which is of the form considered in Section 1.8 with $u(x) = x^2$ and $V(\sigma) = \dfrac{1}{2\sigma^2}$. Hence the decision function z^* is given by

$$\text{(1.904)} \qquad z^* = x_2^2 - x_1^2$$

and the power of the test depends on $h = \frac{1}{2}(\sigma_2^{-2} - \sigma_1^{-2})$ and is given by (1.406) with a and b replaced by a^* and b^*, respectively. The sequential test is performed in the following manner: We take one pair of observations at a time, one from π_1 and one from π_2. We continue sampling as long as $\sum_{\alpha=1}^{n} (x_{2\alpha}^2 - x_{1\alpha}^2)$ lies between $-b^*$ and a^*. Whenever $\sum_{\alpha=1}^{n}(x_{2\alpha}^2 - x_{1\alpha}^2) \geq a^*$, we conclude that $\sigma_2^2 > \sigma_1^2$.

---

[5] The power curve defined by (1.406) is a monotonic function of $h = \mu_1 - \mu_2$. Hence the probability of rejecting the hypothesis that π_1 has the larger mean is $\leq \alpha$ whenever $\mu_1 - \mu_2 \geq h_0$. Thus α is in fact the *maximum* risk of making an erroneous decision. A similar statement can be made concerning the risk β.

Whenever $\sum_{\alpha=1}^{n} (x_{2\alpha}^2 - x_{1\alpha}^2) \leq -b^*$, we conclude that $\sigma_2^2 < \sigma_1^2$. The quantities a^* and b^* are defined by

$$a^* = \frac{\log A}{\frac{1}{2}[(\sigma_2^0)^{-2} - (\sigma_1^0)^{-2}}$$

and

$$-b^* = \frac{\log B}{\frac{1}{2}[(\sigma_2^0)^{-2} - (\sigma_1^0)^{-2}]}.$$

Thus a^* and b^* are defined by a specific value of $\sigma_2^{-2} - \sigma_1^{-2}$ and A and B. If we take $(\sigma_2^0)^{-2} - (\sigma_1^0)^{-2}$ as negative, then $A = \frac{\beta}{1-\alpha}$ and $B = \frac{1-\beta}{\alpha}$ where α = probability of concluding that $\sigma_1^2 < \sigma_2^2$ when in fact $\sigma_2^{-2} - \sigma_1^{-2} = -[(\sigma_2^0)^{-2} - (\sigma_1^0)^{-2}]$ and β is the probability of concluding $\sigma_1^2 < \sigma_2^2$ when in fact $\sigma_2^{-2} - \sigma_1^{-2} = [(\sigma_2^0)^{-2} - (\sigma_1^0)^{-2}]$.

1.9c. *The problem of discriminating between variances when the means are unknown.* Let the measured characteristics in π_1 and π_2 be assumed to be normally distributed with unknown means and unknown variances. We desire to choose, on the basis of a sequential test, that process which has the smaller variance no matter what the means are. This will be accomplished by reducing the problem to that treated in Section 1.9b.

Let $x_{11}, x_{12}, x_{13}, \cdots$ be the successive observations from π_1 and $x_{21}, x_{22}, x_{23}, \cdots$ the successive observations from π_2. Consider the transformation

$$\begin{aligned}
y_{11} &= \frac{1}{\sqrt{2}} x_{11} - \frac{1}{\sqrt{2}} x_{12}, \\
y_{12} &= \frac{1}{\sqrt{2.3}} x_{11} + \frac{1}{\sqrt{2.3}} x_{12} - \frac{2}{\sqrt{2.3}} x_{13}, \\
&\vdots \\
y_{1(n-1)} &= \frac{1}{\sqrt{n(n-1)}} x_{11} + \frac{1}{\sqrt{n(n-1)}} x_{12} \cdots - \frac{n-1}{\sqrt{n(n-1)}} x_{1n}, \\
&\cdots\cdots\cdots\cdots
\end{aligned}$$

with $y_{21}, y_{22}, \cdots y_{2(n-1)} \cdots$ similarly defined in terms of $x_{21}, x_{22}, \cdots x_{2n} \cdots$. It is obvious that this transformation can be applied sequentially. Moreover, it is easy to show that

(1) The expected values of the y's are zero.

(2) The variances of the y's are the same as the variances of the x's.

(3) The y's are normally and independently distributed.

Hence we can apply the sequential test developed in Section 1.9b to the y's without any alterations. The decision function Z_n^* will be given by

$$Z_n^* = \sum_{\alpha=1}^{n} (y_{2\alpha}^2 - y_{1\alpha}^2). \tag{1.905}$$

But it can be easily shown that

$$\sum_{\alpha=1}^{n} y_{2\alpha}^2 = \sum_{\alpha=1}^{n+1} (x_{2\alpha} - \bar{x}_2)^2$$

and

$$\sum_{\alpha=1}^{n} y_{1\alpha}^2 = \sum_{\alpha=1}^{n+1} (x_{1\alpha} - \bar{x}_1)^2$$

where $\bar{x}_1$ and $\bar{x}_2$ are the arithmetic means of the observations in π_1 and π_2 respectively. Hence (1.905) is equivalent to

$$(1.906) \qquad Z_n^* = \sum_{\alpha=1}^{n+1} (x_{2\alpha} - \bar{x}_2)^2 - \sum_{\alpha=1}^{n+1} (x_{1\alpha} - \bar{x}_1)^2 .$$

Thus, to perform this sequential test, the population means need not be known. The only difference between the tests considered in 1.9b and 1.9c is that 1.9c requires one additional pair of observations.[6]

1.9d. *The problem of discriminating between means when the variates have a Poisson distribution.* Let the distribution of x_1 in π_1 be given by $\frac{e^{-m_1} m_1^{x_1}}{x_1!}$ and the distribution of x_2 in π_2 be given by $\frac{e^{-m_2} m_2^{x_2}}{x_2!}$ where x_1 and x_2 each take on the values 0, 1, 2, $\cdots$. It is desired to test the hypothesis that the mean in π_1 is smaller than the mean in π_2 against the alternative that the reverse is true. Since the Poisson distribution can be written as

$$(1.907) \qquad f(x, m) = \frac{1}{x!} e^{x \log m - m},$$

it is of the form considered in Section 1.8 with $u(x) = x$ and $v(m) = \log m$. Hence the decision function z^* is given by

$$z^* = x_2 - x_1$$

and the power of the test depends on $h = \log \frac{m_1}{m_2}$. The sequential test is performed in the following manner: We take one observation from π_1 and one from π_2 in succession. If at any stage $\sum_{\alpha=1}^{n} (x_{2\alpha} - x_{1\alpha}) \leq -b^*$, we conclude that m_2 is smaller than m_1. If $\sum_{\alpha=1}^{n} (x_{2\alpha} - x_{1\alpha}) \geq a^*$, we conclude that m_1 is smaller than m_2. If neither holds, we take another pair of observations. This process is

[6] The method employed here was discovered independently by Charles Stein and the author as a solution to a different sequential problem.

continued until one or the other decision is reached. The quantities a^* and b^* are given by

$$a^* = \frac{\log \dfrac{\beta}{1-\alpha}}{\log u_0} \tag{1.908}$$

$$b^* = \frac{\log \dfrac{1-\beta}{\alpha}}{\log u_0} \tag{1.909}$$

where $u_0 = m_1^0/m_2^0$ which is assumed to be less than one, α is the desired probability of concluding that m_2 is smaller than m_1 when in fact $m_1^0/m_2^0 = u_0 < 1$, and β is the probability of concluding that m_1 is smaller than m_2 when in fact $m_1^0/m_2^0 = 1/u_0$. The power curve is given by

$$L_u = \frac{u^{a^*+b^*} - u^{b^*}}{u^{a^*+b^*} - 1}, \tag{1.910}$$

where $u = m_1/m_2$.

1.9e. *Double dichotomies.*[7] We are given two processes π_1 and π_2, one yielding a fraction defective p_1 and the other p_2. We shall assume that p_1 and p_2 are unknown. We desire to choose on the basis of a sample that process which gives the smaller fraction defective. That is, we wish to devise a test which gives a high probability of accepting π_1 if $p_1 < p_2$ and a high probability of accepting π_2 if $p_2 < p_1$. If $p_1 = p_2$, we might be more or less indifferent as to which process we select.

Before we can answer this question, we must decide: (a) the minimum difference between the two processes which we consider worth detecting; and (b) if the two processes differ at least by the amount specified in (a), the minimum probability with which we desire to make the correct decision.

In the proposed test, the decision function is given by $z^* = x_2 - x_1$ where x_i, $(i = 1, 2)$, takes on the values 0 or 1, depending on whether the ith process yields a nondefective or defective item. The difference between the two processes is measured by[8] $u = \frac{p_1}{1-p_1} / \frac{p_2}{1-p_2}$ (the ratio of the odds). It can easily be seen that when $u < 1$, $p_1 < p_2$ and when $u > 1$, $p_1 > p_2$. If $u = 1$, $p_1 = p_2$. Let u_0 represent a quantity less than 1. Furthermore, let α be the probability of accepting π_2 when in fact the point (p_1, p_2) lies on the curve $\frac{p_1 q_2}{q_1 p_2} = u_0$; and β be the probability of accepting π_1 when in fact the true point (p_1, p_2) lies on the

[7] For a solution of a more general problem in double dichotomies using a different approach, see [1], section 5.32 and [4] section 3.

[8] This follows from the fact that the binomial distribution can be written as $f(x, p) = e^{x \log(p/q) + \log q}$ where x takes on the values 0 or 1. Hence the distribution is of the form considered in section 1.8 with $v(p) = \log p/q$, $w(p) = \log q$, and $z^* = x_2 - x_1$.

curve $\frac{p_2 q_1}{q_2 p_1} = u_0$. Once u_0, α and β are chosen, we compute

$$(1)\ a^* = \frac{\log \frac{\beta}{1-\alpha}}{\log u_0}$$

and

$$(2)\ -b^* = \frac{\log \frac{1-\beta}{\alpha}}{\log u_0}.$$

We then proceed as follows: We take one item from each process in sequence and cumulate the number of defective d_1 in process π_1 and d_2 in process π_2. Whenever $d_2 - d_1 \leq -b^*$, we choose process π_2. Whenever $d_2 - d_1 \geq a^*$, we choose process π_1. Whenever $d_2 - d_1$ lies between a^* and $-b^*$, we take another pair of observations, one from each process. This procedure is continued until one or the other decision is reached.

1.9e1. *The exact value of the power function for double dichotomies.* Since $d_2 - d_1$ changes at most in steps of one unit, it must follow that whenever a decision is reached at a^*, the difference between a^* and $d_2 - d_1$ is either zero (if a^* is an integer), or the difference between a^* and $d_2 - d_1$ is constant for all values of n. A similar argument holds for b^*. This permits us to compute the power function without any approximations. Let $\bar{a}$ be the next positive integer larger than a^* if a^* is not an integer, and $\bar{a} = a^*$ if a^* is an integer. Let $\bar{b}$ be the next positive integer larger than b^* if b^* is not an integer, and $\bar{b} = b^*$ if b^* is an integer. Then we see that the equation (1.406) for the power curve can be given without any approximations by the formula

$$(1.9101) \qquad L_u = (u^{\bar{a}+\bar{b}} - u^{\bar{b}})/(u^{\bar{a}+\bar{b}} - 1)$$

1.9e2. *The exact average sample number for double dichotomies.* Let $Z_n = d_2 - d_1$ and let the point (p_1, p_2) be on some curve $\frac{p_1 q_2}{p_2 q_1} = u$. Let $E(n \mid p_1, p_2)$ be the expected number of pairs of observations required before a decision is reached. Let L_u = probability of reaching $-\bar{b}$ (i.e., L_u is the probability that π_2 is accepted). Then $1 - L_u$ is the probability of reaching $\bar{a}$ (i.e., $1 - L_u$ is the probability that π_1 is accepted). Then by Wald's Fundamental Identity we have[9]

$$(1.911) \qquad EZ_n = EzE(n \mid p_1, p_2).$$

Now, $Ez = p_2 - p_1$, and $EZ_n = -L_u\bar{b} + (1 - L_u)\bar{a}$. Hence

$$(1.912) \qquad E(n \mid p_1, p_2) = \frac{L_u(\bar{a} + \bar{b}) - \bar{a}}{p_2 - p_1}.$$

[9] For a derivation of formula (1.911) which does not depend on the Fundamental Identity, see Wald [1], page 142.

It will be noted that while L_u depends only on $u = \frac{p_1q_2}{p_2q_1}$, $E(n \mid p_1, p_2)$ depends not only on the ratio of the odds but also on the difference between the two fraction defectives.

1.9e3. *The distribution of n for double dichotomies.* In this section we shall be concerned with the probability of reaching a decision with exactly n pairs of observations.

Let a and b be two positive integers and let the sequential test be defined by the decision function $Z_n^* = \sum_{\alpha=1}^{n} z_\alpha$ where z_α takes on the values -1, 0, and 1 with probabilities P_1, P_2, and P_3, respectively. In terms of double dichotomies, $Z_n^* = d_2 - d_1$ where d_2 and d_1 are the cumulative number of defectives obtained sequentially from π_1 and π_2, respectively, and $P_1 = p_1q_2$, $P_2 = p_1p_2 + q_1q_2$, $P_3 = p_2q_1$, where p_1 is the fraction defective yielded by π_1 and p_2 the fraction defective yielded by π_2.

By the Fundamental Identity we have for any t in the complex plane for which $|\phi(t)| \geq 1$,

$$L_u e^{-tb} E_1[\phi(t)]^{-n} + (1 - L_u) e^{ta} E_2[\phi(t)]^{-n} = 1 \tag{1.913}$$

where L_u is the probability that $Z_n^* = -b$ when p_1 and p_2 are such that $\frac{P_1}{P_3} = u$, E_1 and E_2 are the appropriate conditional expectations, and

$$\phi(t) = P_1e^{-t} + P_2 + P_3e^t. \tag{1.914}$$

If we examine Wald's proof of Lemma II [2], we see that $\phi(t) \geq 1$ for all real values of t which lie outside the open interval $(0, h)$ where h is the root of the equation $\phi(t) = 1$. Hence, it must follow that the Fundamental Identity (1.913) must also hold for all real values of t with the possible exception of the open interval $(0, h)$. This fact will be used in the subsequent discussion.

We shall first obtain the distribution of n when $a = \infty$. From equation (1.910) we see that when a approaches ∞, L_u approaches 1 for $u \geq 1$ and u^b for $u < 1$. We shall assume that $u \geq 1$. Then for t negative and $a = \infty$, the Fundamental Identity (1.913) becomes

$$e^{-tb}E[\phi(t)]^{-n} = 1 \tag{1.915}$$

or

$$E[\phi(t)]^{-n} = e^{tb}. \tag{1.916}$$

Now for all $u > 1$, $P_1 > P_3$, and hence $Ez = P_3 - P_1$ is negative. Since the real roots of $\phi(t) = 1$ are opposite in sign to Ez, it must follow that (1.916) holds for all t in the interval $(-\infty, 0)$. Now set $e^t = x$. Then (1.916) can be written as

$$E\left(P_1 \frac{1}{x} + P_2 + P_3x\right)^{-n} = x^b \tag{1.917}$$

and (1.917) is valid for all x in the interval $0 \leq x \leq 1$.

Now set

$$P_1 \frac{1}{x} + P_2 + P_3 x = \frac{1}{\tau}. \tag{1.918}$$

Then for any specified value of τ there will be two values of x, say $x_1(\tau)$ and $x_2(\tau)$. As τ approaches 0, one of these values of x will approach zero and the other infinity. Let $x_1(\tau)$ be the value of x in (1.918) which approaches zero as τ approaches zero. Substituting (1.918) in (1.917) we get

$$E\tau^n = [x(\tau)]^b. \tag{1.919}$$

But $E\tau^n$ is the generating function of n. Hence if we could expand $E\tau^n$ as a power series in τ, then the probability $Z_n^* = -b$ in exactly n steps would be given by the coefficient of τ^n. We are thus led to consider the expansion of $[x(\tau)]^b$ in a power series in τ.

We multiply (1.918) by τx and get

$$x = \tau(P_3 x^2 + P_2 x + P_1). \tag{1.920}$$

Then since $x_1(\tau)$ approaches 0 as τ approaches 0, we can expand $[x_1(\tau)]^b$ by Lagrange formula,[10] and get

$$[x_1(\tau)]^b = \sum \frac{b^m}{m!} \frac{d^{m-1}}{d\xi^{n-1}} [\xi^{b-1}(P_1 + P_2 \xi + P_3 \xi^2)^m]_{\xi=0} \tag{1.921}$$

where the expansion is valid for $x_1(\tau)$ sufficiently close to zero. Hence, if $P_n(b)$ is the probability that exactly n pairs of observations are required to reach a decision, then

$$P_n(b) = \frac{b}{n!} \frac{d^{n-1}}{d\xi^{n-1}} [\xi^{b-1}(P_1 + P_2 \xi + P_3 \xi^2)^n]_{\xi=0}. \tag{1.922}$$

Now

$$\begin{aligned} &\frac{d^{n-1}}{d\xi^{n-1}} [\xi^{b-1}(P_1 + P_2 \xi + P_3 \xi^2)^n]_{\xi=0} \\ &= \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} P_3^i \sum_{j=0}^{n-1} \frac{(n-i)!}{j!(n-i-j)!} P_1^j P_2^{n-i-j} \frac{d^{n-1}}{d\xi^{n-1}} \xi^{n+i-j+b-1}]_{\xi=0}. \end{aligned} \tag{1.923}$$

But

$$\left. \frac{d^{n-1}}{d\xi^{n-1}} \xi^{n+i-j+b-1} \right]_{\xi=0} = 0 \tag{1.924}$$

unless $n = n + i - j + b$, i.e., $j = i + b$, in which case

$$\left. \frac{d^{n-1}}{d\xi^{n-1}} \xi^{n+i-j+b-1} \right|_{\xi=0} = (n-1)! \tag{1.925}$$

[10] See, for example, *Mathematical Analysis*, Vol. 1 (paragraph 189), by Goursat-Hedrick.

Also, since the subscript j ranges from 0 to $n - i$, it must follow that $j \leq n - i$. Hence, $i + b \leq n - i$, or $i \leq \frac{n - b}{2}$. Substituting (1.924) and (1.925) into (1.923) and simplifying, we get for $P_n(b)$

$$P_n(b) = b \sum_{i=0}^{m} \frac{(n-1)!\, P_1^{i+b} P_2^{n-2i-b} P_3^i}{i!(i+b)!(n-2i-b)!} \tag{1.926}$$

where $m = \frac{n-b}{2}$ when $n - b$ is even and $m = \frac{n-b-1}{2}$ when $n - b$ is odd.

We shall now obtain the distribution of n when a is finite.

As before, let $x_1(\tau)$ and $x_2(\tau)$ be the roots of the equation (1.918). Then from (1.913) we have

$$L_u[x_1(\tau)]^{-b} E_1\tau^n + (1 - L_u)\,[x_1(\tau)]^a E_2\tau^n = 1, \tag{1.927}$$

$$L_u[x_2(\tau)]^{-b} E_1\tau^n + (1 - L_u)\,[x_2(\tau)]^a E_2\tau^n = 1. \tag{1.928}$$

Solving for $E_1\tau^n$ and $E_2\tau^n$ from (1.927) and (1.928) we get

$$L_u E_1 \tau^n = \frac{[x_1(\tau)x_2(\tau)]^b[x_2(\tau)^a - x_1(\tau)^a]}{x_2(\tau)^{a+b} - x_1(\tau)^{a+b}} \tag{1.929}$$

$$(1 - L_u)E_2\tau^n = \frac{x_2(\tau)^b - x_1(\tau)^b}{x_2(\tau)^{a+b} - x_1(\tau)^{a+b}}. \tag{1.930}$$

We shall first obtain the probability $Q_n(b)$ that $Z_n^* = -b$. This is given by the coefficient of τ^n in the expansion of $L_u E_1\tau^n$ in a power series in τ. From (1.918) we see that $x_1(\tau)x_2(\tau) = \frac{P_1}{P_3}$. Hence we can write (1.929) as

$$L_u E_1 \tau^n = \frac{x_1(\tau)^b - \left(\frac{P_3}{P_1}\right)^a x_1(\tau)^{b+2a}}{1 - \left(\frac{P_3}{P_1}\right)^{b+a} x_1(\tau)^{2b+2a}}. \tag{1.931}$$

Applying Lagrange formula, we get for $Q_n(b)$

$$Q_n(b) = \frac{1}{n!} \frac{d^{n-1}}{d\xi^{n-1}} [(P_1 + P_2\xi + P_3\xi^2)^n f'(\xi)]_{\xi=0} \tag{1.932}$$

where

$$f(\xi) = \frac{\xi^b - \left(\frac{P_3}{P_1}\right)^a \xi^{b+2a}}{1 - \left(\frac{P_3}{P_1}\right)^{b+a} \xi^{2b+2a}}. \tag{1.933}$$

But $f(\xi)$ can be expanded in a power series in ξ,

$$(1.934)\qquad f(\xi) = \sum_{k=0}^{\infty} \left(\frac{P_3}{P_1}\right)^{kb+ka} \xi^{(2k+1)b+2ka} - \left(\frac{P_3}{P_1}\right)^{kb+(k+1)a} \xi^{(2k+1)b+(2k+2)a}$$

Hence

$$(1.935)\qquad \begin{aligned} Q_n(b) = \frac{1}{n!} \sum_{k=0}^{\infty} [(2k+1)b + 2ka] \left(\frac{P_3}{P_1}\right)^{kb+ka} \\ \cdot \frac{d^{n-1}}{d\xi^{n-1}} [\xi^{(2k+1)b+2ka-1} (P_1 + P_2\xi + P_3\xi^2)^n]_{\xi=0} \\ - \frac{1}{n!} \sum_{k=0}^{\infty} [(2k+1)b + (2k+2)a] \left(\frac{P_3}{P_1}\right)^{kb+(k+1)a} \\ \cdot \frac{d^{n-1}}{d\xi^{n-1}} [\xi^{(2k+1)b+(2k+2)a-1} (P_1 + P_2\xi + P_3\xi^2)^n]_{\xi=0} . \end{aligned}$$

Comparing (1.935) with (1.922) we see that

$$(1.936)\qquad Q_n(b) = P_n(b) - \left(\frac{P_3}{P_1}\right)^{a} P_n(b+2a) + \left(\frac{P_3}{P_1}\right)^{b+a} P_n(3b+2a) - \cdots ,$$

the terms in the series being alternately of the form

$$\left(\frac{P_3}{P_1}\right)^{kb+ka} P_n[(2k+1)b + 2ka] \quad \text{and}$$

$$- \left(\frac{P_3}{P_1}\right)^{kb+(k+1)b} P_n[(2k+1)b + (2k+2)a], \quad \textit{for } k = 0, 1, \cdots .$$

The series stops by itself as soon as the argument of P_n becomes greater than n

If we compare (1.930) with (1.929), we see that the probability that $Z_n^* = a$ with exactly n pairs of observations is given by (1.936) with a and b interchanged and the result multiplied by $(P_3/P_1)^a$.

It is to be noted that the problem of double dichotomies is similar to the following problem in games of chance. Two players A and B, possessing a and b dollars, respectively, are playing a game of chance which admits a draw. The stake is one dollar per game. The probability that A will win one dollar is P_1, the probability that B will win one dollar is P_3 and the probability of a draw is P_2. In terms of this game, L_u given by (1.910) is the probability that B will be ruined in the long run, and $Q_n(b)$ in (1.936) is the probability that B will be ruined in exactly n games.

For a discussion of games of chance which do not permit a draw, see *Introduction to Mathematical Probability*, Chapter VIII, by J. V. Uspensky. The development presented above is in some respects similar to that given in Uspensky's book. In Part II, we shall give a different and more general approach to the problem of deriving the distribution of n for sequential tests in which the variate takes on a finite number of integral values.

REFERENCES

[1] ABRAHAM WALD, "Sequential tests of statistical hypotheses," *Annals of Math. Stat.*, Vol. 16 (1945).
[2] ABRAHAM WALD, "On cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 15 (1944).
[3] B. O. KOOPMAN, "On distributions which admit a sufficient statistic," *Trans. Am. Math. Soc.*, Vol. 39 (1936).
[4] STATISTICAL RESEARCH GROUP, Columbia University, "Sequential analysis of statistical data, applications," *Columbia Univ. Press*, 1945.