CONTRIBUTIONS TO THE THEORY OF SEQUENTIAL ANALYSIS, II, III

By M. A. Girshick

United States Department of Agriculture

Summary. This is a continuation of a paper Part I of which was published in the June, 1946 issue of the *Annals of Mathematical Statistics*. The present paper is divided into two parts, Parts II and III, which are summarized as follows:

Part II. The Exact Power Curve and the Distribution of n for Sequential Tests Where z Takes on a Finite Number of Integral Values.

Consider a sequential test defined by a decision function $Z_n = \sum_{\alpha=1}^{n} z_\alpha$ with boundaries $-b$ and a where a and b are positive integers and z_α is the αth observation of a variate z which takes on a finite number of integral values ranging from the negative integer $-r$ to the positive integer m with respective probabilities $p_{-r}, \cdots, p_m$. Let $\xi_{ai} = P[Z_n = (a + i)]$, $(i = 1, 2, \cdots, m - 1)$, and $\xi_{bj} = P[Z_n = -(b + j)]$, $(j = 1, 2, \cdots, r - 1)$. Furthermore, let A be a square matrix of $a + b - 1$ rows and columns with elements defined by: $a_{ii} = 1 - p_0$ for all i; $a_{i,i+k} = -p_k$ for $k = 1, 2, \cdots, m$; $a_{i,i-j} = -p_{-j}$ for $j = 1, 2, \cdots, r$; and $a_{ij} = 0$ otherwise.

It is proved that

$$\text{(i)} \quad \xi_{bj} = \sum_{i=0}^{r-j-1} p_{i-r} A_{r-j-i,b}, \qquad (j = 0, 1, \cdots, r - 1)$$

$$\text{(ii)} \quad \xi_{aj} = \sum_{i=0}^{m-j-1} p_{i+j+1} A_{a+b-i-1,b}, \qquad (j = 0, 1, \cdots, m - 1),$$

where A_{kb} is the element of the kth row and bth column in A^{-1}. Let $E_{aj}\tau^n$ and $E_{bj}\tau^n$ be the conditional generating function of n under the restriction that $Z_n = (a + j)$ and $Z_n = -(b + j)$ respectively. Then $\xi_{bj}E_{bj}\tau^n$ is obtained by substituting τp_j for each p_j occurring in equation (i) and $\xi_{aj}E_{aj}\tau^n$ is obtained by substituting τp_j for each p_j occurring in equation (ii). The probability that $Z_n = a + j$ in exactly n steps is given by the coefficient of τ^n in the expansion of $\xi_{aj}E_{aj}\tau^n$ in a power series in τ. The probability that $Z_n = -(b + j)$ in exactly n steps is similarly obtained.

This method is applied to the derivation of the exact power function and the distribution of n for the sequential binomial probability ratio test.

Part III. On Conjugate Distributions.

Consider a random variable X with a distribution density $f(x, \theta)$ which satisfies certain specified conditions. Let θ_1 and θ_2 be two values of θ and let $z = \log (f(x, \theta_2)/f(x, \theta_1))$. For any hypothesis $\theta = \theta'$, let $\varphi(t \mid \theta')$ be the moment

generating function of z and h the non-zero value of t for which $\varphi(t \mid \theta') = 1$. We set $F(x) = e^{hz}f(x, \theta')$. Then f and F are conjugate distributions. If $F = f(x, \theta'')$, then θ' and θ'' are defined as conjugate pairs.

A method is given for obtaining the totality of conjugate pairs for the general class of distributions which admit a sufficient statistic. It is then shown that the power of the sequential probability ratio test based on such distributions is given explicitly in terms of these pairs. It is proven that within the approximation obtained by neglecting the excess of $| Z_n |$ over a and b at a decision point the following relationship holds:

$$P_b(n \mid F) = e^{-hb}P_b(n \mid f)$$
$$P_a(n \mid F) = e^{ha}P_a(n \mid f)$$

where $P_b(n \mid g)$ and $P_a(n \mid g)$ stand for the probability that $Z_n \geq a$ and $Z_n \leq -b$ respectively in exactly n steps under the hypothesis g.

II. The Exact Power Curve and the Distribution of n for Sequential Tests Where z Takes on a Finite Number of Integral Values

2.1. General discussions. Let a sequential test be defined by a decision function $Z_n = \sum_{\alpha=1}^{n} z_\alpha$ with boundaries $-b$ and a where a and b are positive and z_α is the αth observation of a variate z which takes on a finite number of integral values, $-r, r + 1, \cdots, -1, 0, 1, 2, \cdots, m$. Let $P(z = i) = p_i$ where $P(z = i)$ stands for the probability that z takes on the value i. We shall assume without any loss of generality that a and b are integers.

When the sequential test terminates with $Z_n \geq a$, the possible values that Z_n can take on are: $a, a + 1, \cdots, a + m - 1$. Similarly, when the sequential test terminates with $Z_n \leq -b$, the possible values which Z_n can take on are: $-b, -(b + 1), \cdots, -(b + r - 1)$. Let $\xi_{ai} = P[Z_n = (a + i)], i = 0, 1, \cdots, m - 1$, and $\xi_{bi} = P[Z_n = -(b + i)], i = 0, 1, \cdots, r - 1$.

For any variate u, let the symbol $E_{bi}(u)$ stand for the expected value of u under the restriction that $Z_n = -(b + i)$, and the symbol $E_{ai}(u)$ stand for the expected value of u under the restriction that $Z_n = a + i$. Let $\phi(t)$ be the generating function of z. Then

$$\phi(t) = \sum_{i=-r}^{m} p_i t^i. \tag{2.101}$$

In terms of the generating function, the Fundamental Identity (see section 2.32 in [6]) can be written as

$$\sum_{i=0}^{r-1} \xi_{bi} t^{-(b+i)} E_{bi}[\varphi(t)]^{-n} + \sum_{i=0}^{m-1} \xi_{ai} t^{a+i} E_{ai}[\varphi(t)]^{-n} = 1. \tag{2.102}$$

It follows from (2.102) that for all values of t for which

$$\phi(t) = \sum_{i=-r}^{m} p_i t^i = 1, \tag{2.103}$$

$$(2.104) \qquad \psi(t) = \sum_{i=0}^{r-1} \xi_{bi} t^{-(b+i)} + \sum_{i=0}^{m-1} \xi_{ai} t^{a+i} = 1$$

where $\psi(t)$ is the generating function of Z_n.

In the paper "The cumulative sums of random variables" [2] Wald has given the following method for obtaining the probabilities ξ_{ai} and ξ_{bi}. Let $t_1, t_2, \cdots, t_{r+m}$ be the $r + m$ roots of (2.103). Substituting these in (2.104) we get $r + m$ linear equations in the $r + m$ unknowns, ξ_{ai} and ξ_{bi}. Thus, if the determinant of these equations is different from zero, the unknowns can be solved in terms of the roots of (2.103). In a similar manner, the characteristic function of n under the restriction that $Z_n = i$ can also be obtained.

The above method has two disadvantages. First, it involves solving for all the roots of a polynomial which will often be of a high degree and second, it involves solving a set of linear equations with coefficients which are powers of complex numbers.

The method outlined below is in many respects much simpler. It requires only the evaluation of one column of the inverse of a matrix of $a + b - 1$ rows and columns. The elements of the matrix are given explicitly and are either 0, 1 or p_i. This permits obtaining general solutions for special classes of sequential tests.

2.2. Derivation of the exact power functions. We multiply $\phi(t) - 1$ by t^r and $\psi(t) - 1$ by t^{b+r-1} and obtain two polynomials.

$$(2.201) \qquad f(t) = \sum_{j=0}^{m+r} (p_{j-r} - \delta_{jr}) t^j$$

and

$$(2.202) \qquad g(t) = \sum_{j=0}^{r-1} \xi_{bj} t^{r-j-1} - t^{b+r-1} + \sum_{j=0}^{m-1} \xi_{aj} t^{a+b+r+j-1}$$

where $\delta_{ik} = 1$ when $i = k$ and zero otherwise.

By the Fundamental Identity, every root of $f(t)$ is also a root of $g(t)$. Since $f(t)$ is of degree $m + r$ and $g(t)$ is of degree $a + b + m + r - 2$, it must follow that $g(t)$ equals $f(t)$ times a polynomial of degree $a + b - 2$.[1] That is,

$$(2.203) \qquad g(t) = f(t) \sum_{i=0}^{a+b-2} c_i t^i$$

where the c's are undetermined constants. Substituting from (2.201) in (2.203) we obtain

$$(2.204) \qquad g(t) = \sum_{j=0}^{a+b+m+r-2} Q_j t^j$$

[1] It is assumed here that $f(t)$ has no multiple roots. The author conjectures that this is true for the polynomial under consideration for all values of p.

where

$$Q_j = \sum_{i=0}^{j} (p_{i-r} - \delta_{ir})c_{j-i}. \tag{2.205}$$

Comparing the coefficients of (2.204) with those of (2.202) and taking into account the fact that $p_k = 0$ when $k > m$ and $c_k = 0$ when $k > a + b - 2$, we get

$$\xi_{bj} = \sum_{i=0}^{r-j-1} p_{i-r}c_{r-j-i-1}, \qquad (j = 0, 1, \cdots, r-1), \tag{2.206}$$

and

$$\xi_{aj} = \sum_{i=0}^{m-j-1} p_{i+j+1}c_{a+b-i-2}, \qquad (j = 0, 1, \cdots, m-1). \tag{2.207}$$

Thus, if the c's (we require only the first r and the last m) are determined, the probabilities ξ_{ai} and ξ_{bi} are also determined from (2.206) and (2.207). But, if we examine the structure of $g(t)$ in (2.202) we see that the coefficients of all the powers of t from r to $(a + b + r - 2)$ inclusive are zero except for the coefficient of t^{b+r-1} which is equal to -1. Consequently, if in (2.204) we set $Q_j = -\delta_{j,b+r-1}$, for all $j = r, r + 1, \cdots, a + b + r - 2$, we shall have the required number of equations to solve for the $a + b - 1$ unknown c's.

In view of (2.205) these equations can be written as

$$\sum_{i=0}^{j} (\delta_{ir} - p_{i-r})c_{j-i} = \delta_{j,b+r-1}, \qquad (j = r, \cdots, a + b + r - 2). \tag{2.208}$$

Changing the range of the subscript j, we get

$$\sum_{i=0}^{j+r-1} (\delta_{ir} - p_{i-r})c_{j+r-i-1} = \delta_{jb}, \qquad (j = 1, 2, \cdots, a + b - 1), \tag{2.209}$$

with the understanding that $p_k = 0$ when $k > m$ and $c_k = 0$ when $k > a + b - 2$.

Let A be the matrix of the equations in (2.209). Then A is of the following form. The elements in the main diagonal are $(1 - p_0)$. In the diagonals to the right of and parallel to the main diagonal, the elements are $-p_{-1}, -p_{-2}, \cdots, -p_{-r}, 0, \cdots, 0$ successively; in the diagonals to the left of and parallel to the main diagonal, the elements are $-p_1, -p_2, \cdots, -p_m, 0, \cdots, 0$ successively. Assume that the determinant of A is different from zero[2] and let A^{-1} be the inverse of A. Let the elements of A^{-1} be designated by A_{ij}, $(i, j = 1, 2, \cdots, a + b - 1)$. Then, in view of (2.209) we get

$$c_j = A_{j+1,b}, \qquad (j = 0, 1, 2, \cdots, a + b - 2). \tag{2.210}$$

Finally, from (2.206) and (2.207), we have,

$$\xi_{bj} = \sum_{i=0}^{r-j-1} p_{i-r}A_{r-j-i,b}, \qquad (j = 0, 1, 2, \cdots, r-1), \tag{2.211}$$

[2] P. L. Hsu has submitted a simple proof to the author that A is non-singular.

and

$$\xi_{aj} = \sum_{i=0}^{m-j-1} p_{i+j+1} A_{a+b-i-1,b}, \qquad (j = 0, 1, 2, \cdots, m-1) \tag{2.212}$$

where, as before, it is understood that $p_k = 0$ when $k > m$ and $A_{kb} = 0$ when, $k > a + b - 1$.

From (2.211) and (2.212) we can obtain the probability that $Z_n \leq -b$ and the probability that $Z_n \geq a$ since these are given by

$$\sum_{j=0}^{r-1} \xi_{bj} \quad \text{and} \quad \sum_{j=0}^{m-1} \xi_{aj} \left(= 1 - \sum_{j=0}^{r-1} \xi_{bj}\right)$$

respectively. We can also obtain En, the average number of steps required to reach a decision. For, if we differentiate (2.102) with respect to t and set $t = 1$, we get

$$E(n) = \frac{EZ_n}{Ez} = \frac{\sum_{i=0}^{m-1} \xi_{ai}(a+i) - \sum_{i=0}^{r-1} \xi_{bi}(b+i)}{\sum_{i=-r}^{m} ip_i} \tag{2.213}$$

2.3. Derivation of the probability that the sequential test will terminate in exactly n steps. Let $\phi(t)$ be the generating function of z and $\psi(t, \tau)$ the joint generating function of Z_n and n. Then

$$\phi(t) = \sum_{i=-r}^{m} p_i t^i \tag{2.301}$$

and

$$\psi(t, \tau) = \sum_{i=0}^{r-1} \xi_{bi} t^{-(b+i)} E_{bi} \tau^n + \sum_{i=0}^{m-1} \xi_{ai} t^{a+i} E_{ai} \tau^n. \tag{2.302}$$

Furthermore, let $\phi_1(t, \tau) = \tau\phi(t) - 1$ and $\psi_1(t, \tau) = \psi(t, \tau) - 1$. In terms of these functions, the Fundamental Identity can be stated as follows: For a fixed τ, every root of $\phi_1(t, \tau)$ is also a root of $\psi_1(t, \tau)$. Let $f(t, \tau) = t^r\phi_1(t, \tau)$ and $g(t, \tau) = t^{b+r-1}\psi_1(t, \tau)$. Then

$$f(t, \tau) = \sum_{j=0}^{m+r} (\tau p_{j-r} - \delta_{jr}) t^j \tag{2.303}$$

and

$$g(t, \tau) = \sum_{j=0}^{r-1} (\xi_{bj} E_{bj} \tau^n) t^{r-j-1} - t^{b+r-1} + \sum_{j=0}^{m-1} (\xi_{aj} E_{aj} \tau^n) t^{a+b+r+j-1}. \tag{2.304}$$

Since for a fixed τ, every root of $f(t, \tau)$ is a root of $g(t, \tau)$, and since $f(t, \tau)$ is a polynomial in t of degree $m + r$ and $g(t, \tau)$ is a polynomial in t of degree $a + b + m - 2$, it must follow that[3]

[3] See footnote 1, section 2.2.

$$g(t, \tau) = f(t, \tau) \sum_{i=0}^{a+b-2} d_i t^i. \tag{2.305}$$

The rest of the argument is identical with that of section 2.2 except that the unknowns in this case are $\xi_{bj}E_{bj}\tau^n$ and $\xi_{aj}E_{aj}\tau^n$ and are given by

$$\xi_{bj} E_{bj} \tau^n = \sum_{i=0}^{r-j-1} \tau p_{i-r} d_{r-j-i-1}, \qquad (j = 0, 1, \cdots, r-1), \tag{2.306}$$

and

$$\xi_{aj} E_{aj} \tau^n = \sum_{i=0}^{m-j-1} \tau p_{i+j+1} d_{a+b-i-2}, \qquad (j = 0, 1, \cdots, m-1), \tag{2.307}$$

(see (2.206), and (2.207)) where the d's are obtained by solving the linear equations:

$$\sum_{i=0}^{j+r-1} (\delta_{i-r} - \tau p_{i-r}) d_{j+r-i-1} = \delta_{jb}, \qquad (j = 1, 2, \cdots, a+b-1), \tag{2.308}$$

(see (2.209)). Thus, we see that the solution for $\xi_{bi}E_{bi}\tau^n$ is obtainable from the solution given in 2.2 for ξ_{bi} by substituting τp_i for every p_i appearing in the expression (2.211). Similarly, the solution for $\xi_{ai}E_{ai}\tau^n$ is obtainable from the solution given for ξ_{ai} by substituting τp_i for every p_i appearing in the expression (2.212).

Let $p(Z_n = k \mid n)$ stand for the probability that $Z_n = k$ in exactly n steps and let $p_{ai}(n) = p[Z_n = (a+i) \mid n]$ and $p_{bi}(n) = p[Z_n = -(b+i) \mid n]$. Then $p_{ai}(n)$ and $p_{bi}(n)$ are given by the coefficient of τ^n in the expansion of $\xi_{ai}E_{ai}\tau^n$ and $\xi_{bi}E_{bi}\tau^n$ respectively in a power series in τ. That the expansions are valid can be seen from the following considerations: If we examine the solutions given for $\xi_{ai}E_{ai}\tau^n$ $(i = 0, 1, \cdots, m-1)$, and $\xi_{bi}E_{bi}\tau^n$ $(i = 0, 1, \cdots, r-1)$, we see that each is a ratio of two polynomials in τ, the polynomial in the denominator is, in each case, the determinant of the linear equations (2.308). Now, it is easy to see that this determinant eqals 1 when $\tau = 0$. Hence the expansions are valid in a neighborhood of $\tau = 0$.[4]

Let $p_{an} = p[Z_n \geq a \mid n]$ and $p_{bn} = p[Z_n \leq -b \mid n]$; then

$$p_{an} = \sum_{i=0}^{m-1} p_{ai}(n) \tag{2.309}$$

and

$$p_{bn} = \sum_{i=0}^{r-1} p_{bi}(n). \tag{2.310}$$

We have also:

$$\sum_{m_1}^{\infty} p_{an} = \sum_{i=0}^{m-1} \xi_{ai} = p(Z_n \geq a) \tag{2.311}$$

---

[4] It can be seen from (2.303) that for a fixed $\tau, f(t, \tau) = 0$ implies that $\varphi(t) = 1/\tau$. Hence if $\tau \leq 1$, $\varphi(\tau) \geq 1$. Thus, the Fundamental Identity is valid in the neighborhood of $\tau = 0$.

and

$$\sum_{m_2}^{\infty} p_{bn} = \sum_{i=0}^{r-1} \xi_{bi} = p(Z_n \leq -b) \tag{2.312}$$

where m_1 is the smallest integer greater than or equal to a/m, and m_2 is the smallest integer greater than or equal to b/r.

2.4. Application of the method to the binomial distribution. We shall consider the binomial in terms of acceptance inspection although the results are general.

Let a sequential acceptance inspection plan be defined by p_1, p_2, α and β where p_1 is the fraction defective which can be tolerated in the lot, p_2 is the fraction defective which cannot be tolerated, α is the maximum probability that the lot will be rejected when the fraction defective is p_1 or less and β is the maximum probability that the lot will be accepted when the faction defective is p_2 or greater. Then the sequential criterion is given by two parallel lines ([1] and [3]).

$$d_1 = -h_1 + sn \tag{2.401}$$

$$d_2 = h_2 + sn \tag{2.402}$$

where

$$h_1 = \frac{\log \dfrac{1-\alpha}{\beta}}{\log \dfrac{p_2(1-p_1)}{p_1(1-p_2)}} \tag{2.403}$$

$$h_2 = \frac{\log \dfrac{1-\beta}{\alpha}}{\log \dfrac{p_2(1-p_1)}{p_1(1-p_2)}} \tag{2.404}$$

$$s = \frac{\log \dfrac{1-p_1}{1-p_2}}{\log \dfrac{p_2\,(1-p_1)}{p_1\,(1-p_2)}} \tag{2.405}$$

and n is the number of observations taken sequentially. We assume that $\alpha + \beta < 1$ and $p_1 < p_2$. Then h_1 and h_2 are positive and s lies between 0 and 1.

The sequential procedure is as follows: Items are examined one at a time in sequence. If at any stage, the cumulative number of defectives found in the sample thus far taken is less than or equal to d_1 given by (2.401), the lot is accepted; if the cumulative number of defectives is greater than d_2 given by (2.402), the lot is rejected; if neither holds then another observation is taken and the process continued.

It is easy to show that the sequential test described above is equivalent to the following: A variate z takes on the values $-s$ and $1 - s$ with respective probabilities q and p. A sequential test is defined by the two boundaries $-h_1$ and h_2 and by the decision function $Z_n = \sum_{\alpha=1}^{n} z_\alpha$ where z_α is the αth observation on z. The sequential test terminates if and only if $Z_n \leq -h_1$ or $Z_n \geq h_2$.

As was mentioned above, s lies between 0 and 1.[5] We shall derive the exact power and the distribution of n for this sequential test by assuming that $s = u/v$ where u and v are integers and $u < v$. This restriction is not serious since every value of s can be approximated to any degree of accuracy by a rational fraction; and, moreover, when the sequential test is applied in practice, s is always taken as rational.

Suppose $s = u/v$. Then the sequential test is equivalent to a test in which the variate z takes on the values $-u$ and $v - u$ with probabilities q and p, respectively, and the boundaries are given by $-h_1v$ and h_2v. Let b be the smallest integer greater than or equal to h_1v and a be the smallest integer greater than or equal to h_2v. Then, since u and v are integers, there is no loss in generality in assuming that the boundaries are $-b$ and a. We shall also assume that u and v are prime to each other (i.e. the fraction u/v is reduced to lowest terms) so that the interval $(-b, a)$ is the shortest possible for this test.

The above discussion shows that a sequential test based on the binomial can be considered as a special case of the class of tests treated in this section. Since z takes only on two values, the linear equations (2.209) assume the simple form:

$$-pC_{j+u-v-1} + C_{j-1} - qC_{j+u-1} = \delta_{bj}, \qquad (j = 1, 2, \cdots, a + b - 1) \tag{2.401}$$

where $C_k = 0$ when k is negative or greater than $a + b - 2$. In terms of the C's, the ξ_{bj} and E_{aj} are given by

$$\xi_{bj} = qC_{u-j-1}, \qquad (j = 0, 1, \cdots, u - 1), \tag{2.402}$$

$$\xi_{aj} = qC_{a+b+u-v+j-1}, \quad (j = 0, 1, \cdots, v - u - 1). \tag{2.403}$$

The conditional generating functions of n are obtained by solving (2.401) with τp substituted for p and τq substituted for q.

Since the first $v - u$ and the last u equations in (2.401) contain only two terms and all the other equations contain only three terms, the C's can be obtained without too much difficulty by direct substitution provided $a + b$ is not very large. When $a + b$ is sizeable, a general solution is called for. So far, the author has been able to obtain this only for the case $u = 1$. This special case also has been considered by Walter Bartky [4].

Setting $u = 1$ in (2.401) we get

$$-pC_{j-v} + C_{j-1} - qC_j = \delta_{bj}, \qquad (j = 1, 2, \cdots, a + b - 1), \tag{2.407}$$

where $C_k = 0$ when k is negative or greater than $a + b - 2$.

[5] In fact, it follows from Theorem 1, section 3.2 below that $p_1 \leq s \leq p_2$.

Consider a general set of equations of the form (2.407) with the subscripts ranging from 1 to an arbitrary integer k. Let the determinant of these equations be designated by Δ_k. Then by direct expansion it can be shown that Δ_k satisfies the difference equation.

$$(2.408) \qquad \Delta_k = \Delta_{k-1} - pq^{v-1}\Delta_{k-v}$$

with the initial conditions

$$(2.409) \qquad \Delta_i = 1, \qquad i = 1, 2, \cdots, v-1; \qquad \Delta_v = 1 - pq^{v-1}.$$

The difference equation (2.408) can be solved by well known methods. We set

$$(2.410) \qquad \phi(x) = \sum_{j=1}^{\infty} \Delta_j x^{j-1}$$

and then multiply each side of (2.410) by $1 - x + pq^{v-1}x^v$. This yields

$$(2.411) \qquad (1 - x + pq^{v-1}x^v)\phi(x) = \sum_{j=1}^{\infty} [\Delta_j - \Delta_{j-1} + pq^{v-1}\Delta_{j-v}]x^{j-1}.$$

But by (2.408) and (2.409) we find that the right-hand side of (2.411) equals $1 - pq^{v-1}x^{v-1}$. Therefore,

$$(2.412) \qquad \phi(x) = \frac{1 - pq^{v-1}x^{v-1}}{1 - x + pq^{v-1}x^v}$$

If we expand (2.412) in a power series in x, the coefficient of x^k will be Δ_{k+1}. This expansion can be performed readily and we get:

$$(2.413) \qquad \Delta_{k+1} = \sum_{j=0}^{m_1} (-1)^j C_j^{k-j(v-1)}(pq^{v-1})^j - \sum_{j=0}^{m_2} (-1)^j C_j^{k-v-j(v-1)+1}(pq^{v-1})^{j+1}$$

where m_1 stands for the largest integer less than or equal to,k/v m_2 stands for the largest integer less than or equal to $k - v + 1/v$ and $C_t^r = r!/t!(r-t)!$.

Let us define $\Delta_0 = 1$ and $\Delta_k = 0$ when $k < 0$. Then, in terms of the extended definition of Δ_k, C_j is given by

$$(2.414) \qquad C_j = \frac{\Delta_j \Delta_{a-1} - \Delta_{j-b}\Delta_{a+b-1}}{q^{j-b+1}\Delta_{a+b-1}}$$

for $j = 0, 1, \cdots, a + b - 2$. To prove this, we substitute in the left-hand member of (2.407) the expression for C_k given in (2.414) and get

$$(2.415) \qquad \frac{\Delta_{a+b-1}(\Delta_{j-b} - \Delta_{j-b-1} + pq^{v-1}\Delta_{j-v-b}) - \Delta_{a-1}(\Delta_j \Delta_{j-1} + pq^{v-1}\Delta_{j-v})}{q^{j-b}\Delta_{a+b-1}}.$$

But in view of (2.408), (2.409) and the extended definition of Δ_k, the expression in (2.415) vanishes for all $j \neq b$. When $j = b$, the expression equals 1. Hence, it follows that (2.414) is the desired solution.

Let $L_p = p[Z_n \leq -b]$. Then L_p, when plotted against p, gives the operating characteristic curve for this sequential test. But $L_p = qC_0$. Hence, we have

(2.416) $$L_p = q^b \frac{\Delta_{a-1}}{\Delta_{a+b-1}}.$$

As a final remark, we wish to point out that the solution to the sequential problem presented in this section, where taken in conjunction with Wald's solution, is of mathematical interest, since it relates each element of the inverse of a square matrix (designated by A in this section) with the roots of a polynomial $f(t)$ given by (2.201).

III. Conjugate Distributions

3.1. General discussion. Consider a random variable X with a distribution density $f(x, \theta)$.[6] Let θ_1 and θ_2 be two specified values of θ and let

(3.101) $$z = \log \frac{f(x, \theta_2)}{f(x, \theta_1)}.$$

For any hypothesis $\theta = \theta'$, let $\phi(t \mid \theta')$ be the moment generating function of z. That is,

(3.102) $$\phi(t \mid \theta') = \int_{-\infty}^{\infty} e^{tz} f(x, \theta') \, dx.$$

Let h be the real non-zero value of t for which $\phi(t \mid \theta') = 1$[7] and let

(3.103) $$F(x) = e^{hz} f(x, \theta').$$

Then $F(x)$ is a distribution density. Following Wald [5], we shall call $F(x)$ and $f(x, \theta')$ conjugate distributions.

The distribution density $F(x)$ depends on θ_1, θ_2, and θ'. In some instances $F(x)$ will be a member of the class of distributions $f(x, \theta)$. This is the case, for example, when z is a discrete variate. It is the case also if $\theta' = \theta_1$. For then $h = 1$ and $F(x) = f(x, \theta_2)$. If $F(x)$ belongs to the class of distributions $f(x, \theta)$, we shall designate $F(x)$ by $f(x, \theta'')$ and call θ' and θ'' a conjugate pair.

3.2. Conjugate pairs and the power curve for sequential probability ratio tests in which the underlying distributions admit a sufficient statistic. Let $f(x, \theta)$ admit a sufficient statistic and let a sequential test be defined in terms of the probability ratio z given by (3.101) for some specified hypothesis θ_1 and alternative hypothesis θ_2 with $\theta_1 < \theta_2$. Let the boundaries be given by $-b$ and a where a and b are positive. Since $f(x, \theta)$ admits a sufficient statistic, it can be written in the form

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

The probability ratio z is then given by the simple expression

(3.202) $$z = u(x)[v(\theta_2) - v(\theta_1)] + w(\theta_2) - w(\theta_1).$$

---

[6] If X is discrete, then $f(x, \theta)$ stands for the probability that $X = x$ when θ is true.

[7] See section 2.31 and Lemma II, section 2.32 in [6].

Let

$$b^* = \frac{b}{v(\theta_2) - v(\theta_1)} \tag{3.203}$$

$$a^* = \frac{a}{v(\theta_2) - v(\theta_1)} \tag{3.204}$$

$$s = \frac{w(\theta_1) - w(\theta_2)}{v(\theta_2) - v(\theta_1)}. \tag{3.205}$$

In terms of b^*, a^* and s, the sequential criterion is defined by two parallel lines[8]

$$A_n = -b^* + sn \tag{3.206}$$

$$R_n = a^* + sn \tag{3.207}$$

and the decision functions $\sum_{\alpha=1}^{n} u(x_\alpha)$. The hypothesis $\theta = \theta_1$ is accepted whenever $\sum_{\alpha=1}^{n} u(x_\alpha) \leq A_n$ and rejected whenever $\sum_{\alpha=1}^{n} u(x_\alpha) \geq R_n$. If $A_n < \sum_{\alpha=1}^{n} u(x_\alpha) < R_n$, another observation is taken. This process is continued until one or the other decision is reached.

In what follows, we shall restrict ourselves to the general class of functions $f(x, \theta)$ for which the differentiations under the integral sign indicated below are permissible and $v(\theta)$ is a monotonic function of θ.

Consider the function

$$\psi(\theta) = sv(\theta) + w(\theta). \tag{3.208}$$

We shall show that $\psi(\theta) =$ constant has exactly two roots in θ. To this end, we prove the following theorems.

Theorem 1. *Let $Eu(x) \mid \theta$ be the expected value of $u(x)$ under the assumption that θ is true. Then there exists a value of $\theta = \theta_0$ such that (a) $Eu(x) \mid \theta_0 = s$; (b) $\theta_1 \leq \theta_0 \leq \theta_2$ and $Eu(x) \mid \theta_1 \leq s \leq Eu(x) \mid \theta_2$ if $v(\theta)$ is an increasing function of θ, and the inequalities are reversed if $v(\theta)$ is a decreasing function of θ.*

Proof: Assume that $v(\theta)$ is an increasing function of θ. Let $z^* = u(x) - s$ and let $\phi(t) \mid \theta$ be the moment generating function of z^* under the hypothesis that θ is true. Then, it is easy to see that $\phi(h \mid \theta_1) = 1$ and $\phi(-h \mid \theta_2) = 1$ where $h = v(\theta_2) - v(\theta_1)$. Since h is positive, it follows by Lemma 1, section 2.6 of [6], that $Ez^* \mid \theta_1 < 0$ and $Ez^* \mid \theta_2 > 0$. Therefore, $Eu(x) \mid \theta_1 < s$ and $Eu(x) \mid \theta_2 > s$. Moreover, as we shall see in the proof of Theorem 2 below, $Eu(x) \mid \theta$ is assumed to be a continuous function of θ and proved to be mono-

[8] It is here assumed that $v(\theta_2) - v(\theta_1) > 0$. If this is not the case, then a^* and b^* have to be interchanged.

tonically increasing. Hence it must follow that there exists a $\theta = \theta_0$ such that $Eu(x) \mid \theta_0 = s$ and $\theta_1 \leq \theta_0 \leq \theta_2$. This proves the theorem in case $v(\theta)$ is monotonically increasing. However, the argument is identically the same in case $v(\theta)$ is monotonically decreasing.

THEOREM 2. *Let $\psi(\theta)$ be defined as in* (3.208). *Then $\psi(\theta)$ is a monotonically increasing function of θ in the interval $\theta < \theta_0$; assumes a maximum at $\theta = \theta_0$; and is a monotonically decreasing function of θ in the interval $\theta > \theta_0$.*

PROOF. If we differentiate twice the identity

$$\int_{-\infty}^{\infty} e^{u(x)v(\theta)+r(x)+w(\theta)}\, dx = 1 \tag{3.209}$$

with respect to θ we get

$$v'(\theta)Eu(x) \mid \theta + w'(\theta) = 0 \tag{3.210}$$

and

$$v''(\theta)Eu(x) \mid \theta + w''(\theta) = [v'(\theta)]^2 \sigma^2_{u(x)} \tag{3.211}$$

where $\sigma^2_{u(x)}$ is the variance of $u(x)$. Also, if we differentiate under the integral sign the function $Eu(x) \mid \theta$ with respect to θ, we get

$$\frac{dEu(x) \mid \theta}{d\theta} = v'(\theta)\sigma^2_{u(x)}. \tag{3.212}$$

Now by hypothesis, $v(\theta)$ is monotonic in θ. Hence from (3.212) we see that $Eu(x) \mid \theta$ is also monotonic. Moreover, if $v(\theta)$ is an increasing function of θ, so is $Eu(x) \mid \theta$, and conversely. Let us assume that $v(\theta)$ increases with θ. Then for all $\theta < \theta_0$, $Eu(x) \mid \theta < s$ and for all $\theta > \theta_0$, $Eu(x) \mid \theta > s$. Consequently, we have

$$\psi'(\theta) > v'(\theta)Eu(x) \mid \theta + w'(\theta) \tag{3.213}$$

for all $\theta < \theta_0$ and

$$\psi'(\theta) < v'(\theta)Eu(x) \mid \theta + w'(\theta) \tag{3.214}$$

for all $\theta > \theta_0$. But by (3.210) the right-hand side of these inequalities is equal to zero for all θ. Hence $\psi'(\theta) > 0$ for $\theta < \theta_0$ and $\psi'(\theta) < 0$ for $\theta > \theta_0$. The same argument holds when $v(\theta)$ is a decreasing function of θ. Now let $\theta = \theta_0$. Then by (3.210), we see that $\psi'(\theta_0) = 0$. Hence, $\psi(\theta)$ is a maximum at $\theta = \theta_0$. This proves the theorem.

Let c be any constant $< \psi(\theta_0)$ within the domain of $\psi(\theta)$. Then by Theorem 2, the equation $\psi(\theta) = c$ has two roots in θ. Let these roots be designated by θ' and θ''. We now prove the following theorem.

THEOREM 3. *Let z^* and $\phi(t \mid \theta)$ be defined as above. Then* (*a*) *$\phi(t \mid \theta') = 1$ for $t = v(\theta'') - v(\theta')$*; (*b*) *$\phi(t \mid \theta'') = 1$ for $t = v(\theta') - v(\theta'')$*; *and* (*c*) *$\theta'$ and θ'' from a conjugate pair with respect to z^*.*

PROOF: By definition

$$\phi(t \mid \theta') = \int_{-\infty}^{\infty} e^{u(x)[v(\theta')+t)]+r(x)+w(\theta')-ts}\, dx. \tag{3.215}$$

Now let $t = v(\theta'') - v(\theta') = h$. Then, in view of the fact that $\psi(\theta') = \psi(\theta'')$, we get

$$\phi(h \mid \theta') = \int_{-\infty}^{\infty} e^{u(x)v(\theta'')+r(x)+w(\theta'')}\, dx = 1. \tag{3.216}$$

In a similar manner, it can be shown that $\phi(-h \mid \theta'') = 1$. Moreover, the same argument also shows that $f(x, \theta'') = e^{hz^*}f(x, \theta')$. This proves the theorem.

Turning now to the sequential test defined by (3.206) and (3.207), we see that it is equivalent to a test with the decision function $Z_n^* = \sum_{\alpha=1}^{n} z_\alpha^*$ and the two boundaries $-b^*$ and a^*. Let L_θ be the probability that the sequential test will terminate and $Z_n^* \leq -b^*$ (i.e. the hypothesis θ_1 is accepted) when θ is true. Then (neglecting the fact that at a decision point Z_n^* might exceed a^* or fall short of $-b_1^*$), $L_{\theta'}$ and $L_{\theta''}$ are given by (see for example (2.406) in [6]).

$$L_{\theta'} = \frac{e^{(a^*+b^*)h} - e^{b^*h}}{e^{(a^*+b^*)h} - 1} \tag{3.217}$$

and

$$L_{\theta''} = \frac{e^{-h(a^*+b^*)} - e^{-hb^*}}{e^{-h(a^*+b^*)} - 1} = e^{-b^*h} L_{\theta'} \tag{3.218}$$

where $h = v(\theta'') - v(\theta')$. Thus, we see that the two roots of the equation $\psi(\theta) = c$ determine two points on the power curve for the sequential test. By assigning various values to c we obtain as many pairs of points as desired.

The above results show that for the class of distributions under consideration, the real non-zero roots of $\varphi(t \mid \theta) = 1$ are obtainable from the roots of $\psi(\theta) =$ constant. Since $\psi(\theta)$ is completely defined by the form of the distribution $f(x, \theta)$, the power curve of the sequential test can be obtained without a knowledge of the moment generating function of z^*. This might be advantageous in some cases.

3.3. The distribution of n under conjugate hypotheses. Let $P_b(n \mid g)$ stand for the probability that a sequential test will terminate with $Z_n \leq -b$ in exactly n steps when the distribution density of x is g. Let $P_a(n \mid g)$ be similarly defined.

THEOREM 1. *If we neglect the excess of Z_n over a and $-b$ at a decision point,*

$$P_b(n \mid F) = e^{-hb}P_b(n \mid f) \tag{3.301}$$

$$P_a(n \mid F) = e^{ha}P_a(n \mid f) \tag{3.302}$$

where f and F are conjugate distributions as defined in (3.103) *and h is the non-zero* *real value of t for which $\phi(t \mid f)$, the characteristic function of $z = \log f(x, \theta_2)/f(x, \theta_1)$ underthe hypothesis f, equals* 1.

PROOF: Since, by definition, $F = e^{zh}f$, it follows that $\psi(t - h \mid F) = \phi(t \mid f)$ where $\psi(t \mid F)$ is the characteristic function of z under the hypothesis F. Let

$$\phi(t \mid f) = e^{-\tau} \tag{3.303}$$

where τ is a pure imaginary. Furthermore, let $t_1(\tau)$ and $t_2(\tau)$ be the roots of (3.303) such that $\lim_{\tau \to 0} t_1(\tau) = 0$ and $\lim_{\tau \to 0} t_2(\tau) = h$ (see [2], page 289). Then $t_1(\tau) - h$, and $t_2(\tau) - h$ will be the corresponding roots of

$$\psi(t \mid F) = e^{-\tau}. \tag{3.304}$$

Now by the Fundamental Identity we have

$$L_f e^{-bt_1(\tau)} E_{bf} e^{\tau n} + (1 - L_f) e^{at_1(\tau)} E_{af} e^{\tau n} = 1 \tag{3.305}$$

$$L_f e^{-bt_2(\tau)} E_{bf} e^{\tau n} + (1 - L_f) e^{at_2(\tau)} E_{af} e^{\tau n} = 1 \tag{3.306}$$

and

$$L_F e^{-b[t_1(\tau)-h]} E_{bF} e^{\tau n} + (1 - L_F) e^{a[t_1(\tau)-h]} E_{aF} e^{\tau n} = 1 \tag{3.307}$$

$$L_F e^{-b[t_2(\tau)-h]} E_{bF} e^{\tau n} + (1 - L_F) e^{a[t_2(\tau)-h]} E_{aF} e^{\tau n} = 1 \tag{3.308}$$

where $L_f = P[Z_n \leq -b \mid f]$, E_{bf} stands for the expected value of $e^{\tau n}$ under the hypothesis f and the restriction $Z_n \leq -b$; E_{af} stands for the expected value of $e^{\tau n}$ under the hypothesis f and the restriction $Z_n \geq a$; and the symbols L_F, E_{bF} and E_{aF} are similarly defined.

By comparing equations (3.305) and (3.306) with (3.307) and (3.308) we see that

$$L_F E_{bF} e^{\tau n} = e^{-hb} L_f E_{bf} e^{\tau n} \tag{3.309}$$

and

$$(1 - L_F) E_{aF} e^{\tau n} = e^{ha}(1 - L_f) E_{af} e^{\tau n}. \tag{3.310}$$

Since the above relationships hold for the characteristic functions of n, they must also hold for the distribution of n. This proves the theorem.

If we set $\tau = 0$ in (3.309) and (3.310) we also get

$$L_F = e^{-hb} L_f \tag{3.311}$$

and

$$1 - L_F = e^{ha}(1 - L_f). \tag{3.312}$$

In view of (3.311) and (3.312) we see from (3.309) and (3.310) that

$$E_{bF} e^{\tau n} = E_{bf} e^{\tau n} \tag{3.313}$$

and

$$E_{aF}e^{\tau n} = E_{af}e^{\tau n}. \tag{3.314}$$

From (3.313) and (3.314) we obtain the following rather surprising theorem.

THEOREM 4. *Except for the approximation indicated in Theorem 1, the conditional distribution of n under the restriction that $Z_n \leq -b$ as well as the restriction that $Z_n \geq a$ is identical for the two hypotheses F and f.*

The above theorems are of particular interest when F is a member of the class of distributions f. In any given sequential test the results of Theorem 1 can be used to facilitate the computation of the probabilities of making a decision. Furthermore, the results of Theorem 4 show that the conditional distribution of n throws no light on the parameter θ involved in the distribution of z. This follows since the conditional distribution of n is identical for the conjugate pair θ' and θ'', and, in any practical problem, θ' and θ'' will represent opposing hypotheses.

We shall now establish exact relationships of the type considered above when the variate z takes on a finite number of integral values.

Let z take on the values $-r, -r+1, \cdots, -1, 0, 1, 2, \cdots, m$ with $P(z = i) = p_i$. Furthermore, let $P_i = e^{hi}p_i$ where h is the real non-zero root of

$$\sum_{i=-r}^{m} p_i e^{ti} = 1. \tag{3.315}$$

Then the probabilities P_i and p_i are conjugate. We set $e^t = u$ and define $\phi(u \mid \theta)$ to be the generating function of z under the hypothesis $p(z = i) = \theta_i$. Then

$$\phi(u \mid p) = \sum_{i=-r}^{m} p_i u^i \tag{3.316}$$

and

$$\phi(u \mid P) = \sum_{i=-r}^{m} P_i u^i = \sum_{i=-r}^{m} p_i(e^h u)^i. \tag{3.317}$$

Consider a sequential test defined by two boundaries $-b$ and a and a decision function $Z_n = \sum_{\alpha=1}^{n} z_\alpha$. Let ξ_{bi}^{θ} and ξ_{ai}^{θ} stand for the probabilities that $Z_n = -(b+i)$ and $Z_n = a + i$ respectively under the hypothesis that $\theta_i = P(z = i)$. Furthermore, let $P_{bi}(n \mid \theta)$ and $P_{ai}(n \mid \theta)$ stand for the probabilities that $Z_n = -(b+i)$ and $Z_n = (a+i)$ respectively in exactly n steps, under the hypothesis $\theta_i = P(z = i)$. Also, let the symbols E_{bi}^{θ} and E_{ai}^{θ} stand for conditional expectations under the hypothesis $\theta_i = P(z = i)$ and under the restriction that $Z_n = -(b+i)$ and $Z_n = a + i$ respectively.

Since z takes on a finite number of integral values, the Fundamental Identity for the two conjugate hypotheses, p and P can be written as:

$$\sum_{i=0}^{r-1} \xi_{bi}^{p} u^{-(b+i)} E_{bi}^{p}[\phi(u \mid p)]^{-n} + \sum_{i=0}^{m-1} \xi_{ai}^{p} u^{a+i} E_{ai}^{p}[\phi(u \mid p)]^{-n} = 1 \tag{3.318}$$

and

$$\sum_{i=0}^{r-1} \xi_{bi}^{P} u^{-(b+i)} E_{bi}^{P}[\phi(u \mid P)]^{-n} + \sum_{i=0}^{m-1} \xi_{ai}^{P} u^{a+i} E_{ai}^{P}[\phi(u \mid P)]^{-n} = 1. \tag{3.319}$$

For any real number τ let $u_1(\tau), u_2(\tau), \cdots, u_{r+m}(\tau)$ be the $r + m$ roots of the equation:

$$\phi(u \mid p) = \sum_{i=-r}^{m} p_i u^i = \frac{1}{\tau}. \tag{3.320}$$

Then, in view of (3.317) the corresponding roots of

$$\phi(u \mid P) = \sum_{i=-r}^{m} P_i u^i = \frac{1}{\tau} \tag{3.321}$$

are given by $u_1(\tau)e^{-h}, u_2(\tau)e^{-h}, \cdots, u_{r+m}(\tau)e^{-h}$. Substituting these roots in (3.318) and (3.319) successively, we get

$$\sum_{i=0}^{r-1} \xi_{bi}^{p} u_j(\tau)^{-(b+i)} E_{bi}^{p} \tau^n + \sum_{i=0}^{m-1} \xi_{ai}^{p} u_j(\tau)^{a+i} E_{ai}^{p} \tau^n = 1 \tag{3.322}$$

and

$$\sum_{i=0}^{r-1} \xi_{bi}^{P}[u_j(\tau)e^{-h}]^{-(b+i)} E_{bi}^{P} \tau^n + \sum_{i=0}^{m-1} \xi_{ai}^{P}[u_j(\tau)e^{-h}]^{a+i} E_{ai}^{P} \tau^n = 1 \tag{3.323}$$

for $j = 1, 2, \cdots, r + m$. Since the roots $u_j(\tau)$ are assumed to be known, the unknowns in (3.322) and (3.323) can be solved in terms of these roots provided the determinant of the equations is different from zero. But in section 2, we have indirectly shown that for a sufficiently small τ, the determinant is different from zero. Thus, assuming that the solution has been obtained we see from (3.322) and (3.323) that

$$\xi_{bi}^{P} E_{bi}^{P} \tau^n = e^{-h(b+i)} \xi_{bi}^{p} E_{bi}^{p} \tau^n \tag{3.324}$$

and

$$\xi_{ai}^{P} E_{ai}^{P} \tau^n = e^{h(a+i)} \xi_{ai}^{p} E_{ai}^{p} \tau^n. \tag{3.325}$$

Setting $\tau = 1$, we get

$$\xi_{bi}^{P} = e^{-h(b+i)} \xi_{bi}^{p} \tag{3.326}$$

and

$$\xi_{ai}^{P} = e^{h(a+i)} \xi_{ai}^{p}. \tag{3.327}$$

Moreover, if we expand the expressions in (3.324) and (3.325) in a power series in τ (which by section 2 is permissible), and compare coefficients of τ^n we get

$$P_{bi}(n \mid P) = e^{-h(b+i)} P_{bi}(n \mid p) \tag{3.328}$$

and

$$P_{ai}(n \mid P) = e^{h(a+i)} P_{ai}(n \mid p). \tag{3.329}$$

REFERENCES

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

[2] Abraham Wald, "On cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 283–296.

[3] Statistical Research Group, Columbia University, *Sequential Analysis of Statistical Data, Applications*, Columbia Univ. Press, 1945.

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

[5] Abraham Wald, "Some generalizations of the theory of cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 287–293.

[6] M. A. Girshick, "Contributions to the theory of sequential analysis, I," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 123–143.