FORMULAS FOR THE GROUP SEQUENTIAL SAMPLING OF ATTRIBUTES

BY HOWARD L. JONES

Illinois Bell Telephone Company

1. Summary. When an infinite lot consisting of defective and nondefective items is investigated by means of a group sequential sampling plan, the use of matrices and vectors is helpful in determining the probabilities of various combinations of the two classes of items and in computing unbiased estimates of the lot fraction defective. For a sequential plan of the Bartky [1] type, the infinite summation of such vectors leads to an exact, explicit formula for the average number of items inspected,

$$\begin{aligned} \bar{n}_p = p^{-1} \{L_p[G(h_1 + h_2 - 1) - (h_1 + h_2)] \\ - G(h_2 - 1) + h_1 + h_2 - [h_1]\}, \end{aligned} \tag{1.1}$$

where p is the fraction defective in the lot, L_p is the probability of arriving at a decision to accept the lot, h_1 and h_2 are parameters of the plan as defined by the Statistical Research Group [3], $G(i)$ is defined by Bartky's equation (36), and $[h_1]$ is the largest integer equal to or less than h_1.

In approximating L_p, or in finding the parameters of a sequential plan with specified risks, the formulas proposed by Wald [2] and the Statistical Research Group can be improved by adding an adjustment,

$$a = \tfrac{1}{3}(1 - 2s), \tag{1.2}$$

to the value of h_2 wherever it occurs. Their formula for approximating $\bar{n}_p$ can be improved by adding the adjustment

$$cq = aq/(1 - s) \tag{1.3}$$

wherever h_2 occurs, provided that the value of L_p which appears in this formula is arrived at by employing adjustment (1.2).

2. Introduction. The sampling plans considered here are among those used in acceptance sampling where the purpose of the plan is to provide objective criteria for deciding whether the fraction defective in a lot of infinite size is excessive or not. Inspection of randomly selected items from the lot continues as long as

$$ns - h_1 < d < ns + h_2, \tag{2.1}$$

where n is the cumulative number of items inspected, d is the cumulative number of defective items found, and s, h_1, and h_2 are positive numbers that are chosen to give the sampling plan certain desired properties and that may be regarded as parameters defining a particular plan of this type. When (2.1) is no longer true, inspection ceases and the indicated decision is recorded. If

$d \leq ns - h_1$, the decision is favorable, and the lot is accepted as not having an excessive number of defective items. If $d \geq ns + h_2$, the decision is unfavorable, and the lot is rejected. The optimum properties of a sampling plan of this type, with respect to the amount of inspection necessary before a decision is reached, have been discussed by Wald and Wolfowitz [6].

If $1/s$, h_1/s, h_2/s, and $h_1 + h_2$ are all positive integers, the plan is equivalent to a group plan like that outlined in Table 1, where an initial group of size $v_0 \geq 0$ is selected, followed if necessary by additional groups of size $v = 1/s > v_0$. The sampling plan for this case has the important practical advantage that reference to a chart or a table (like Table 1) is necessary only once for each group inspected.

For group plans of this type, Bartky [1] derived an exact formula for the probability of acceptance, here denoted by L_p, and for the average or expected value of the total number of groups that have been selected and wholly or partly inspected when inspection ceases and the appropriate decision is reached. These formulas were obtained by summing vectors, one vector for each value of n for which $ns - h_1$ is an integer, with an element for each of the $h_1 + h_2 - 1$ integral values of d satisfying (2.1), and with each element equal to the joint probability (a) that n or more items will be inspected before reaching a decision, and (b) that exactly d defective items will be found among the first n items inspected. By a slightly different approach, Girshick [5] derived a formula for L_p that is equivalent to Bartky's. His results indicate that this formula holds also when $h_1 + h_2$ is not an integer. For the still more general case where s is any rational number between 0 and 1, Pólya [7] described a method of computing L_p and $\bar{n}_p$ by solving difference equations, $\bar{n}_p$ being used here to denote the average or expected number of items inspected until a decision is reached to accept or reject the lot. His results, however, are stated in terms of polynomials for which explicit formulas are not given except for an illustrative example. Walker [9] obtained an exact, explicit formula for L_p for the case where s is rational, and found the mean and variance of the number of items inspected in terms of functions not stated explicitly. Wald [2] gave formulas for approximating L_p and $\bar{n}_p$ for the still more general case where s, h_1, and h_2 are not necessarily rational. As pointed out by Mrs. Robinson [8], the errors in Wald's approximations are sometimes of considerable size.

For plans of the type outlined in Table 1, the approach taken here is to define a vector for every integer $n \geq 0$, including values of n where $ns - h_1$ is not an integer as well as the values discussed by Bartky. The method of computing these vectors is described in Section 3 following. As indicated in Section 4, such vectors are useful in arriving at unbiased estimates of the lot fraction defective by the general method suggested by Girshick, Mosteller, and Savage [4]. They also simplify the summing of probabilities (or its description, at any rate), and facilitate the use of some results already obtained by Bartky. These results are briefly reviewed in Sections 5 and 6, and are used in Section 7 to derive an exact, explicit formula for $\bar{n}_p$. An approach similar to this can obviously be followed in connection with double sampling, truncated sequential sampling, and other plans for sampling attributes. Matrices with more than two dimensions may be employed where there are more than two attributes.

Bartky suggested methods of approximating L_p and the average number of groups selected. These methods can also be extended to the approximation of $\bar{n}_p$. Such approximations are frequently much closer to the exact values than the approximations proposed by Wald. By adding certain adjustments to the parameter h_2 in Wald's formulas, however, it is possible to obtain fairly simple approximations that are not greatly different from those resulting from the application of Bartky's suggestions. These various approximations are discussed in Section 8, and comparisons between them and the exact values for illustrative examples are shown in the accompanying tables. The nature of the errors in

TABLE 1

A group sequential plan of the Bartky type

Total number of groups selected		Total number of items selected	Upon finding d defective items ($d = 0, 1, \cdots$)—		
Initial group	Additional groups		accept if d is equal to or less than—	continue inspection if d is equal to one of the numbers—	reject if d is equal to or greater than—
1	0	v_0	$v_0s - h_1$	$v_0s - h_1 + 1, \cdots, v_0s + h_2 - 1$	$v_0s + h_2$
1	1	$v_0 + v$	$v_0s - h_1 + 1$	$v_0s - h_1 + 2, \cdots, v_0s + h_2$	$v_0s + h_2 + 1$
1	2	$v_0 + 2v$	$v_0s - h_1 + 2$	$v_0s - h_1 + 3, \cdots, v_0s + h_2 + 1$	$v_0s + h_2 + 2$
1	3	$v_0 + 3v$	$v_0s - h_1 + 3$	$v_0s - h_1 + 4, \cdots, v_0s + h_2 + 2$	$v_0s + h_2 + 3$
...	...	...	.....	....	.....

NOTE: $1/s$, h_1/s, h_2/s, and $h_1 + h_2$ are positive integers; $h_1 + h_2 \geq 2$; $v = 1/s \geq 2$; $0 \leq v_0 = (h_1 - [h_1])/s < v$, where $[h_1]$ is the largest integer $\leq h_1$.

Wald's approximations is outlined in Section 9, and suggestions are offered for improving his procedure for choosing the value of h_2 when designing a sequential sampling plan.

The notation used here largely follows that introduced by the Statistical Research Group, Columbia University [3], where applicable. Elsewhere, Bartky's symbols have been used extensively. In making close comparisons with his article [1], however, it should be noted that the definitions of some of his symbols have been changed slightly in order to simplify the summarization of his results. Some of this simplification is made possible by restricting our discussion to sampling plans where $h_1 > 0$, whereas Bartky also considered plans where $h_1 \leq 0$.

3. The probability that inspection will continue. Let n, s, h_1, and h_2 have the same meaning as in inequality (2.1); and let d_1 be the smallest integer greater than $ns - h_1$. For $i = 2, 3, \cdots, k$, let

$$d_i = d_1 + i - 1, \tag{3.1}$$

where k is the smallest integer equal to or greater than $h_1 + h_2$. A vector $P(n)$ can now be defined for every integer $n \geq 0$ by letting $P_i(n)$ denote the probability that inspection will continue on account of finding exactly d_i defective items among the first n items inspected. In other words, let $P_i(n)$ represent the joint probability (a) that n or more items will be inspected before arriving at a decision to accept or reject the lot, and (b) that exactly d_i defective items will be found among the first n items inspected, with the special condition imposed that

$$P_i(n) = 0 \text{ if } d_i \geq ns + h_2 . \tag{3.2}$$

If we let p equal the fraction defective in the lot, and q equal $1 - p$, we can now write

$$P(n + 1) = J(n)P(n), \qquad n = 0, 1, 2, \cdots , \tag{3.3}$$

where $P(0)$ is the vector with elements

$$P_i(0) = \begin{cases} 1 & \text{if } i \text{ is the smallest integer} \geq h_1 , \\ 0 & \text{otherwise,} \end{cases} \tag{3.4}$$

for $i = 1, 2, \cdots , k$, and where $J(n)$ is one of several $k \times k$ matrices with elements equal to p, q, or 0. In particular, if $0 < s \leq \frac{1}{2}$, then

$$J(n) = \begin{cases} A & \text{if} \quad (n + 1)s - h_1 < d_1 , \quad \text{and} \quad d_k < (n + 1)s + h_2 , \\ B & \text{if} \quad (n + 1)s - h_1 \geq d_1 , \\ C & \text{if} \quad (n + 1)s - h_1 < d_1 , \quad \text{and} \quad d_k \geq (n + 1)s + h_2 , \end{cases} \tag{3.5}$$

where the matrices A, B, and C have the respective elements

$$A_{ij} = \begin{cases} q & \text{if} \quad j = i, \\ p & \text{if} \quad j = i - 1, \\ 0 & \text{otherwise,} \end{cases} \tag{3.6}$$

$$B_{ij} = \begin{cases} p & \text{if} \quad j = i < k, \\ q & \text{if} \quad j = i + 1, \\ 0 & \text{otherwise,} \end{cases} \tag{3.7}$$

$$C_{ij} = \begin{cases} q & \text{if} \quad j = i < k, \\ p & \text{if} \quad j = i - 1 < k - 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.8}$$

Examples of several successive vectors for an illustrative example are shown in Table 2. (If $\frac{1}{2} < s < 1$, the same approach may be followed after substituting $1 - s$ for s and interchanging h_1 and h_2, and p and q, with corresponding changes in the interpretation of the results.)

For a group sampling plan of the type indicated in Table 1, where an initial group is selected of size v_0 such that $0 \leq v_0 < 1/s$, followed by one or more additional groups, if necessary, of size $v = 1/s$, let c represent the number of groups of size v that have been completely inspected when the nth item is inspected. Also, for $n \geq v_0$, let

$$t = n - (v_0 + cv), \qquad c = 0, 1, 2, \cdots, \tag{3.9}$$

so that wherever defined, t has one of the values $0, 1, \cdots, v - 1$, representing the number of items inspected in the $(c + 1)$th group of size v. Then it can be

TABLE 2

Some values of $P(n)$ for the illustrative case where $s = .3$, $h_1 = .7$, $h_2 = 1.5$

n	$ns - h_1$	d_1	d_2	d_3	$ns + h_2$	$P_1(n)$	$P_2(n)$	$P_3(n)$
0	$-.7$	0	1	2	1.5	1	0	0
1	$-.4$	0	1	2	1.8	q	p	0
2	$-.1$	0	1	2	2.1	q^2	$2pq$	p^2
3	.2	1	2	3	2.4	$3pq^2$	$3p^2q$	0
4	.5	1	2	3	2.7	$3pq^3$	$6p^2q^2$	0

NOTE: $P(1) = CP(0)$; $P(2) = AP(1)$; $P(3) = BP(2)$; $P(4) = CP(3)$.

verified that

$$P(n) = \begin{cases} A^t M^c \mathrm{P}(0) & \text{if} \quad v_0 = 0, \\ A^n P(0) & \text{if} \quad 0 \leq n < v_0 \geq 1, \\ A^t M^c B A^{v_0 - 1} P(0) & \text{if} \quad n \geq v_0 \geq 1, \end{cases} \tag{3.10}$$

where the superscripts indicate repeated multiplication,

$$M = BA^{v-1}, \tag{3.11}$$

and $A^0 = M^0 = I$, the identity matrix. Let C_a^b be defined for any integer a and any nonnegative integer b by the relationship

$$C_a^b = \begin{cases} b!/[a!(b - a)!] & \text{if} \quad 0 \leq a \leq b, \\ 0 & \text{otherwise}, \end{cases} \tag{3.12}$$

where $0! = \Gamma(1) = 1$. Then M has the elements

$$M_{ij} = \begin{cases} C_{i-j+1}^v p^{i-j+1} q^{v-i+j-1} & \text{if} \quad i = 1, 2, \cdots, k - 1, \\ 0 & \text{if} \quad i = k, \end{cases} \tag{3.13}$$

being equivalent to M_{ij} in Bartky's [1] equation (7) except for an additional row and column.

4. Estimating the fraction defective. Suppose a decision to accept or reject a particular lot has been reached after inspecting n_p items and finding d_p defective items, and that it is desired to obtain an unbiased estimate, $\hat{p}$, of the fraction defective. Following the method of Girshick, Mosteller, and Savage [4], we may write

$$\hat{p} = K^*(n_p, d_p)/K(n_p, d_p), \tag{4.1}$$

where $K(n_p, d_p)$ is the number of ways of selecting d_p defective items and $n_p - d_p$ nondefective items without arriving at a decision to accept or reject the lot before inspecting the n_pth item, and $K^*(n_p, d_p)$ is the number of such ways when the first item selected is defective. But since a decision has finally been reached after finding d_p defective items among the n_p items inspected, there must have been d_r defective items among the first n_r items inspected, where $n_r = n_p - 1$, and $d_r = d_p$ or $d_p - 1$ according to whether the lot has been accepted or rejected. Conversely, after finding d_r defective items among the first n_r items inspected, there was just one way of finding d_p defective items among the n_p items inspected. It follows that

$$\hat{p} = K^*(n_r, d_r)/K(n_r, d_r). \tag{4.2}$$

Let $P_r(n_r)$ denote the probability of finding exactly d_r defective items among the first n_r items inspected; and let $P_r^*(n_r)$ denote the conditional probability of finding exactly d_r defective items among the first n_r items when the first item inspected is defective. Also, let d_1 be the smallest integer greater than $n_r s - h_1$, and r be the integer determined by substituting r for i and d_r for d_i in (3.1). Then $P_r(n_r)$ is the rth element in the vector $P(n_r)$ as defined in Section 3. Similarly, $P_r^*(n_r)$ is the rth element in the vector $P^*(n_r)$ defined by the equations

$$P^*(1) = J^*(0)P(0), \tag{4.3}$$

$$P^*(n + 1) = J(n)P^*(n), \tag{4.4}$$

where $J^*(0)$ is the matrix obtained by substituting 1 for p and 0 for q in $J(0)$ as defined in that section. Moreover,

$$P_r(n_r) = K(n_r, d_r)p^{d_r}q^{n_r - d_r}, \tag{4.5}$$

and

$$P_r^*(n_r) = K^*(n_r, d_r)p^{d_r - 1}q^{n_r - d_r}. \tag{4.6}$$

Hence,

$$\hat{p} = pP_r^*(n_r)/P_r(n_r), \tag{4.7}$$

where the right-hand side of this equation is determined for any arbitrary value of p such that $0 < p < 1$.

For a sampling plan like that shown in Table 1, a convenient way of computing $P_r^*(n_r)$ and $P_r(n_r)$ is by first finding the operator $O(n_r)$ that is equivalent to the product of the matrices shown on the right-hand side of (3.10) for $n = n_r$, and then employing the relationships

$$P(n_r) = O(n_r)P(0), \tag{4.8}$$

$$P^*(n_r) = O(n_r)P^*(0). \tag{4.9}$$

The use of (4.9), however, will necessitate finding elements of a vector $P^*(0)$ satisfying equations (4.3) and (4.4) simultaneously for $n = 0$. It can be verified that these two equations are satisfied if $P^*(0)$ has the following set of elements (not unique unless $J(0) = A$):

$$(4.10) \qquad P_i^*(0) = \begin{cases} 0 & \text{if} \quad i \le u, \\ (-1)^{i-u+1} p^{i-u-1} q^{-i+u} & \text{if} \quad i > u, \end{cases}$$

where u is the smallest integer greater than $h_1 - s$.

An alternative method of computing $\hat{p}$ is to find vectors $K(n_r)$ and $K^*(n_r)$ by substituting 1 for every p and q in the vectors $P(0)$ and $P^*(0)$ and in each matrix $J(n)$, and then performing the operations analogous to those indicated in (3.10), (4.8), and (4.9). The rth elements in these vectors are equal to $K(n_r, d_r)$ and $K^*(n_r, d_r)$, respectively; whence

$$(4.11) \qquad \hat{p} = K_r^*(n_r)/K_r(n_r).$$

5. The probability of acceptance. For iv an integer, let

$$(5.1) \qquad g(i) = \sum_{d<i} (-1)^d \, C_d^{(i-d)v+d-1} p^d q^{-(i-d)v-d}$$

for $d = 0, 1, \cdots$, where $C_d^{(i-d)v+d-1}$ is defined by (3.12). Also, for a sampling plan of the type shown in Table 1, let the vector V be defined by

$$(5.2) \qquad V = \sum_{c=0}^{\infty} P(v_0 + cv), \qquad c = 0, 1, 2, \cdots;$$

that is, let V be the infinite sum of those vectors in equation (3.10) for which $t = 0$. Then from Bartky's results [1], the elements of V are

$$(5.3) \qquad V_i = \begin{cases} g(i)g(h_2)/g(h_1 + h_2) & \text{if} \quad i \le h_1, \\ g(i)g(h_2)/g(h_1 + h_2) - g(i - h_1) & \text{if} \quad i > h_1, \end{cases}$$

for $i = 1, 2, \cdots, k$, where $k = h_1 + h_2$; and the probability of accepting the lot is

$$(5.4) \qquad L_p = g(h_2)/g(h_1 + h_2) = \begin{cases} q^v V_1 & \text{if} \quad v_0 = 0, \\ q^{v_0} P_1(0) + q^v V_1 & \text{if} \quad v_0 > 0. \end{cases}$$

In evaluating $g(i)$ for small i, it is convenient to use

$$(5.5) \qquad g(i) = q^{-vi} \Delta_{vi-1}$$

in conjunction with Girshick's [5] difference equation

$$(5.6) \qquad \Delta_m = \Delta_{m-1} - pq^{v-1} \Delta_{m-v}, \qquad m \ge v,$$

with the initial conditions

$$(5.7) \qquad \Delta_m = 1 \qquad \text{if } m = 0, 1, \cdots, v - 1.$$

6. The average number of items selected. For the sampling plan indicated in Table 1, suppose the items are selected in groups of size v after the initial group of size v_0, the rth group being selected after the $(r - 1)$th group has been completely inspected. The total number of items selected, when a decision has been reached to accept or reject the lot, is therefore either v_0, or the sum of v_0 and some multiple of v. Let this number be denoted by the product vr_p, so that r_p is the sum of a nonnegative integer and the fraction v_0/v, where $0 \leq v_0/v < 1$. From Bartky's [1] equation (14), the average or expected value of r_p is given by

$$\bar{r}_p = v_0/v + \sum_{i=1}^{k} V_i, \qquad k = h_1 + h_2, \tag{6.1}$$

where V_i is defined by (5.3). Also, if we let

$$G(i) = \sum_{d<i} g(i - d), \qquad d = 0, 1, \cdots, \tag{6.2}$$

where $g(i - d)$ is defined by (5.1), then

$$\sum_{i=1}^{k} V_i = L_p G(h_1 + h_2 - 1) - G(h_2 - 1), \tag{6.3}$$

where L_p is the probability of acceptance, as in equation (5.4). Since $1/s$ and h_1/s are integers for the type of plan considered here,

$$v_0/v = h_1 - [h_1], \tag{6.4}$$

where $[h_1]$ is the largest integer $\leq h_1$. It follows that

$$\bar{r}_p = L_p G(h_1 + h_2 - 1) - G(h_2 - 1) + h_1 - [h_1], \tag{6.5}$$

and that the average number of items selected is $v\bar{r}_p$.

7. The average amount of inspection. If inspection ceases immediately when inequality (2.1) is no longer true, the average or expected number of items inspected for the sampling plan indicated in Table 1 may be derived as outlined in the following paragraphs.

Let S_i denote the ith element in the infinite sum of vectors

$$S = \sum_{n=0}^{\infty} P(n). \tag{7.1}$$

Employing (3.10) and (5.2), we write

$$S = \sum_{n=0}^{v_0-1} A^n P(0) + \sum_{t=0}^{v-1} A^t V = (I - A)^{-1}[(I - A^{v_0})P(0) + (I - A^v)V]. \tag{7.2}$$

The elements of $(I - A^{v_0})P(0)$ are 0 if $v_0 = 0$; otherwise they are $(1 - q^{v_0})P_i(0)$ for $i = 1$, and $P_i(0) - P_{i-1}(v_0)$ for $i = 2, 3, \cdots, k$. Similarly, the elements of $(I - A^v)V$ are $(1 - q^v)V_1$ for $i = 1$, and $V_i - V_{i-1} + P_{i-1}(v_0)$ for $i = 2, 3, \cdots, k$. The elements of $(I - A)^{-1}$ are equivalent to p^{-1} for $j \leq i$, and 0 for $j > i$. The elements of S are therefore

$$
(7.3) \qquad S_i = \begin{cases} p^{-1}\left\{(1 - q^v)V_1 + \sum_{j=2}^{i} [V_j - V_{j-1} + P_{j-1}(0)]\right\} & \text{if} \quad v_0 = 0, \\ p^{-1}\left\{1 - q^{v_0})P_1(0) + (1 - q^v)V_1 \right. \\ \qquad \left. + \sum_{j=2}^{i} [V_j - V_{j-1} + P_j(0)]\right\} & \text{if} \quad v_0 > 0, \end{cases}
$$

where the summation is taken to be 0 for $i = 1$. This equation may be reduced to

$$
(7.4) \qquad S_i = \begin{cases} p^{-1}(V_i - L_p) & \text{if} \quad i \leq h_1, \\ p^{-1}(V_i - L_p + 1) & \text{if} \quad i > h_1. \end{cases}
$$

To find the average amount of inspection, we can now employ

$$
(7.5) \qquad \bar{n}_p = \sum_{i=1}^{k} S_i,
$$

the derivation of which is similar to that used by Bartky [1] to obtain his equation (14). It follows that

$$
(7.6) \qquad \begin{aligned} \bar{n}_p &= p^{-1}\{\bar{r}_p + h_2 - (h_1 + h_2)L_p\} \\ &= p^{-1}\{L_p[G(h_1 + h_2 - 1) - (h_1 + h_2)] - G(h_2 - 1) + h_1 + h_2 - [h_1]\}, \end{aligned}
$$

where $\bar{r}_p$ and $G(i)$ are defined by equations (6.5) and (6.2).

8. Approximation formulas. Several formulas have been proposed for approximating the probability of acceptance and the average amount of inspection for sequential sampling plans. Such formulas are convenient not only for the type of plan shown in Table 1 where s, h_1, and h_2 are rational numbers, but particularly for plans where any or all of these parameters may not be rational and approximation by step-by-step evaluation of terms like those in the last three columns of Table 2 would be long and tedious. They are also useful in designing plans that are to have certain specified properties, as outlined in Section 9.

The following formulas were proposed by Bartky [1]:

$$
(8.1) \quad g(i) \sim \begin{cases} (2vi + 2v/3 - 4/3)(v - 1)^{-1} & \text{if} \quad vp = 1, \\ (1 - vp)^{-1} + [q - (v - 1)px]^{-1}x^{-i+1/v} & \text{if} \quad vp \neq 1, \end{cases}
$$

$$
(8.2) \quad G(i) \sim \begin{cases} (vi^2 + 5vi/3 + v/18 - 4i/3 \\ \quad - 1/18 - v^{-1}/9)(v - 1)^{-1} & \text{if} \quad vp = 1, \\ i(1 - vp)^{-1} - \tfrac{1}{2}v(v - 1)p^2(1 - vp)^{-2} \\ \quad + [q - (v - 1)px]^{-1}(1 - x)^{-1}x^{-i+1/v} & \text{if} \quad vp \neq 1, \end{cases}
$$

where $v = 1/s$ and x is the real positive root of

$$
(8.3) \qquad (px + q)^v = x
$$

that is unequal to 1. In practical problems, we may first compute p for selected real positive values of the dummy variable x by employing

$$p = (x^s - 1)(x - 1)^{-1}, \tag{8.4}$$

which is equivalent to (8.3), and then find the corresponding values of L_p, the probability of acceptance, and $\bar{n}_p$, the average amount of inspection, from equa-

TABLE 3

Computation of L_p, the probability of acceptance, to three decimal places for illustrative examples

x	10	5	2	1	.5	.2	.1
Example 1: $s = .04$, $h_1 = h_2 = 1$							
Exact	.963	.911	.759	.577	.380	.182	.096
Formula (8.1)	.954	.899	.746	.566	.373	.180	.095
Formula (8.5)	.966	.915	.765	.582	.383	.183	.097
Formula (8.6)*	.909	.833	.667	.500	.333	.167	.091
Formula (8.8)	.955	.900	.747	.566	.373	.180	.096
Example 2: $s = .04$, $h_1 = 2$, $h_2 = 1$							
Exact	.959	.893	.674	.403	.169	.036	.010
Formula (8.1)	.950	.881	.662	.395	.166	.035	.009
Formula (8.5)	.962	.898	.681	.408	.170	.036	.010
Formula (8.6)*	.901	.806	.571	.333	.143	.032	.009
Formula (8.8)	.951	.882	.663	.395	.166	.035	.010
Example 3: $s = .04$, $h_1 = 1$, $h_2 = 2$							
Exact	.996	.981	.888	.698	.444	.196	.100
Formula (8.1)	.995	.980	.887	.698	.444	.196	.100
Formula (8.5)	.996	.981	.891	.701	.445	.196	.100
Formula (8.6)*	.991	.968	.857	.667	.429	.194	.099
Formula (8.8)	.996	.980	.888	.698	.444	.196	.100

* Values shown for Formula (8.6) are taken from [3].

tions (5.4) and (7.6) in conjunction with (8.1) and (8.2). Comparison of the results with exact values for illustrative examples are shown in Tables 3 and 4 on the lines headed "Formula (8.1)" and "Formula (8.2)". In this connection, it should be observed that Bartky recommended these approximations only for cases corresponding to those where $h_2 \geq 3$. The examples illustrated here were deliberately chosen to show comparisons under more adverse conditions.

If v is large and p small, so that the probability of finding exactly d defective items in a group of size v approximates the corresponding term in the Poisson

TABLE 4

Computation of $\bar{n}_p$, the average amount of inspection, to one decimal place for illustrative examples

x	10	5	2	1	.5	.2	.1
Example 1:							
$s = .04$, $h_1 = h_2 = 1$							
Exact	31.2	33.9	36.6	36.2	32.7	25.9	21.2
Formula (8.2)	30.6	33.0	35.3	34.6	31.1	24.3	19.6
Formula (8.5)	29.5	32.4	35.6	35.4	32.3	25.7	21.0
Formula (8.7)	27.9	28.5	28.0	26.0	22.7	17.6	14.2
Formula (8.9)	30.6	32.9	34.9	34.2	30.1	23.2	18.5
Example 2:							
$s = .04$, $h_1 = 2$, $h_2 = 1$							
Exact	63.6	70.4	77.0	71.2	54.7	34.1	24.6
Formula (8.2)	62.8	69.0	74.9	68.9	52.7	32.4	24.0
Formula (8.5)	60.0	67.2	74.7	69.7	54.0	33.7	24.4
Formula (8.7)	58.2	60.7	60.1	52.1	38.9	23.8	16.8
Formula (8.9)	62.8	68.9	74.3	68.2	51.4	31.1	21.8
Example 3:							
$s = .04$, $h_1 = 1$, $h_2 = 2$							
Exact	33.7	40.1	53.1	60.6	58.0	44.7	35.4
Formula (8.2)	33.6	40.0	53.0	60.5	57.8	44.5	35.0
Formula (8.5)	31.6	38.1	51.1	58.9	56.7	43.9	34.7
Formula (8.7)	32.2	38.6	48.1	52.1	48.6	37.4	29.5
Formula (8.9)	33.6	40.0	52.8	60.2	57.0	43.5	33.9

series, then $g(i)$ is approximately equal to its limiting value as $p \to 0$ and $v \to \infty$, and we may write

$$g(i) \sim \sum_{d<i} (d!)^{-1} [(d - i)vp]^d e^{(i-d)vp}, \qquad d = 0, 1, \cdots, \tag{8.5}$$

where $0! = \Gamma(1) = 1$. The limiting value of $G(i)$ can be found by combining (8.5) with (6.2). These limiting values of $g(i)$ and $G(i)$, for $i = 1, 2, \cdots, 5$ and for selected values of x as defined above, are given in Bartky's Table II. The resulting approximations to L_p and $\bar{n}_p$ for illustrative examples are shown in our Tables 3 and 4 on the line headed "Formula (8.5)".

Wald's [2] formulas for approximating L_p and $\bar{n}_p$ have been transformed by the Statistical Research Group [3] into the following:

$$L_p \sim \begin{cases} h_2/(h_1 + h_2) & \text{if } p = s, \\ (x^{h_1+h_2} - x^{h_1})/(x^{h_1+h_2} - 1) & \text{if } p \neq s, \end{cases} \tag{8.6}$$

$$\bar{n}_p \sim \begin{cases} h_1h_2/[s(1 - s)] & \text{if } p = s, \\ [L_p(h_1 + h_2) - h_2]/(s - p) & \text{if } p \neq s, \end{cases} \tag{8.7}$$

where x has the same meaning as in (8.3) and (8.4) above. The results obtained by applying these formulas to the examples shown in Tables 3 and 4 are indicated on the lines headed "Formula (8.6)" and "Formula (8.7)". These results are evidently less satisfactory than those obtained from Bartky's formulas, particularly for small values of h_2. For other comparisons, see Mrs. Robinson's note [8].

For reasons outlined in Section 9 following, Wald's formulas can be improved by adding appropriate adjustments to the parameter h_2, so that the approximations are computed from the following relationships:

$$L_p \sim \begin{cases} (h_2 + a)/(h_1 + h_2 + a) & \text{if } p = s, \\ (x^{h_1+h_2+a} - x^{h_1})/(x^{h_1+h_2+a} - 1) & \text{if } p \neq s, \end{cases} \tag{8.8}$$

$$\bar{n}_p \sim \begin{cases} h_1(h_2 + b)/[s(1 - s)] & \text{if } p = s, \\ [L_p(h_1 + h_2 + cq) - (h_2 + cq)]/(s - p) & \text{if } p \neq s, \end{cases} \tag{8.9}$$

where

$$a = \tfrac{1}{3}(1 - 2s), \tag{8.10}$$

$$b = a[1 + s(h_1 + h_2 + a)^{-1}], \tag{8.11}$$

$$c = a/(1 - s). \tag{8.12}$$

For the illustrative examples shown in Tables 3 and 4, the application of these formulas leads to the approximate values shown on the lines headed "Formula (8.8)" and "Formula (8.9)". The derivation of these semiempirical formulas is discussed in Section 9.

Where sample items are selected as outlined in Section 6, the average number of items selected can be approximated by combining the foregoing approximations with the equation

$$\bar{r}_p = (h_1 + h_2)L_p - h_2 + p\bar{n}_p, \tag{8.13}$$

obtained from (7.6), and then dividing by s.

9. Some comments on Wald's approximations. Wald's [2] formulas for sequential sampling were developed to provide an objective criterion for deciding which of two alternative hypotheses, H_1 and H_2, concerning the population sampled is the correct one. In general, a sample statistic, X_n, is observed or computed for $n = 1, 2, \cdots$, successively, where n is the cumulative number of observations. Let $f_1(X_n)$ and $f_2(X_n)$ denote the relative probabilities that X_n will be found in n sample observations when H_1 and H_2, respectively, are true. Also, let α denote the risk we are willing to run of making the wrong decision when H_1 is true, and β the corresponding risk when H_2 is true. Then for a Wald sequential plan, sampling continues as long as the likelihood ratio satisfies the inequality

$$\frac{\beta}{1 - \alpha} < \frac{f_2(X_n)}{f_1(X_n)} < \frac{1 - \beta}{\alpha}. \tag{9.1}$$

When inequality (9.1) no longer holds, however, sampling ceases and we conclude that H_1 is true if the second member of (9.1) is equal to or less than the first member, or that H_2 is true if the second member is equal to or greater than the third member. Wald's choice of this type of plan was based on his conjecture, later proved [6], that it would minimize the average number of observations for the given risks when either H_1 or H_2 is true.

If the population consists of an infinite lot of items that can each be classified as defective or nondefective, if the alternative hypotheses state that the lot fraction defective $p = p_1$ and p_2, respectively, and if $X_n = d$ (the cumulative number of defective items in the first n sample items observed), then inequality (9.1) becomes

$$\frac{\beta}{1 - \alpha} < \frac{p_2^d q_2^{n-d}}{p_1^d q_1^{n-d}} < \frac{1 - \beta}{\alpha}. \tag{9.2}$$

This is equivalent to inequality (2.1), where

$$s = \frac{\log [q_1/q_2]}{\log [(p_2 q_1)/(p_1 q_2)]}, \tag{9.3}$$

$$h_1 = \frac{\log [(1 - \alpha)/\beta]}{\log [(p_2 q_1)/(p_1 q_2)]}, \tag{9.4}$$

and

$$h_2 = \frac{\log [(1 - \beta)/\alpha]}{\log [(p_2 q_1)/(p_1 q_2)]}, \tag{9.5}$$

"log" denoting logarithm to any convenient base. Formulas (9.3), (9.4), and (9.5), or their equivalents, were therefore proposed by Wald and by the Statistical Research Group [3] for use in designing a sampling plan with parameters s, h_1 and h_2 such that the operating characteristic curve representing the functional relationship between p, the fraction defective, and L_p, the probability of deciding that H_1 is true, will pass through the specified points $(p_1, 1 - \alpha)$ and (p_2, β). In practice, for $p_1 < p_2$, a decision that H_1 is true means that the lot being sampled is accepted as satisfactory, while a decision that H_2 is true means that the lot is rejected.

In a situation where the specified points, $(p_1, 1 - \alpha)$ and (p_2, β), and the values of s, h_1, and h_2 computed from formulas (9.3), (9.4), and (9.5) are such that the first two members of (2.1) are exactly equal whenever a decision is reached to accept the lot, and the second and third members are exactly equal whenever a decision is reached to reject the lot, the operating characteristic curve will actually pass through the specified points. Moreover, if we set

$$(p_2 q_1)/(p_1 q_2) = 1/x, \tag{9.6}$$

the elimination of p_1 and q_1 from (9.3) and (9.6), together, will yield an equation in p_2 equivalent to (8.4); while the elimination of p_1, q_1, and α from (9.4), (9.5), and (9.6) will yield a value of β equal to the right-hand side of (8.6). In other words, for the situation just described, equations (9.3) to (9.6) can be used to derive the exact formula for the operating characteristic curve in terms of x. As Pólya [7] pointed out, the necessary and sufficient conditions for this situation are that $s = \frac{1}{2}$ and that both $2h_1$ and $2h_2$ be positive integers.

In other situations, when a decision to accept the lot is reached, the first and second members of (2.1) will be equal if $1/s$ and h_1/s are integers, and in no case will the first member exceed the second member by as much as s. But when decisions to reject are reached, the second and third members can not always be equal if $s < \frac{1}{2}$, and the difference may be almost as large as $1 - s$. It follows that when s is small, most of the difficulty with Wald's procedure for choosing the parameters of the sampling plan lies in the formula for computing h_2. This difficulty can be largely surmounted by subtracting an adjustment, a, from the right-hand side of (9.5). This implies a similar adjustment of (8.6) by adding the correction a to each value of h_2, as shown in (8.8). The value of a here proposed is that shown in (8.10), which was chosen so that, for $p = s$, (8.8) would agree with the approximations to L_p resulting from the combination of (5.4) and (8.1). The suggested formula for h_2, therefore, is

$$h_2 = \frac{\log [(1 - \beta)/\alpha]}{\log [(p_2 q_1)/(p_1 q_2)]} - \tfrac{1}{3}(1 - 2s). \tag{9.7}$$

Comparison between actual and specified risks for illustrative sampling plans resulting from the use of formulas (9.5) and (9.7), in conjunction with (9.3) and (9.4), is shown in Table 5.

To investigate formula (8.7), consider a situation where the points $(p_1, 1 - \alpha)$ and (p_2, β) and the resulting values of s, h_1, and h_2 lead to a sampling plan such that a decision to accept or to reject the lot can be reached only when the last item in some group has been inspected. In that case, the relationship between n_p, the number of items inspected when a decision is reached, and r_p, the equivalent number of groups of size $1/s$ selected, is necessarily $r_p = sn_p$. Taking expected values and combining with (7.6) to eliminate $\bar{r}_p$ leads to (8.7), which is therefore exact for the situation just described. This situation, as Pólya noted, is precisely the same as the one where (8.6) is exact; that is, the necessary and sufficient conditions for (8.7) to be exact are that $s = \frac{1}{2}$ and that both $2h_1$ and $2h_2$ be integers.

In other situations where $1/s$ and h_1/s are integers, the relationship $r_p = sn_p$ will hold if the sampling inspection leads to a decision to accept the lot. But for $s < \frac{1}{2}$, a decision to reject the lot may be reached before the r_pth group is

TABLE 5

Comparison between actual and specified risks, for different methods of computing h_2 leading to the same sampling plan, where $p_1 = .010720$ and $p_2 = .097766$

	Example 1	Example 2	Example 3
Specified risks:			
Formula (9.5)			
α	.090909	.099099	.009009
β	.090909	.009009	.099099
Formula (9.7)			
α	.044638	.048886	.004444
β	.095577	.009511	.099556
Parameters of resulting plan:			
s	.04	.04	.04
h_1	1.00	2.00	1.00
h_2	1.00	1.00	2.00
Actual risks:			
α	.037	.041	.0044
β	.096	.0096	.0996

completely inspected. In other words, for all plans of the type shown in Table 1, the relationship between r_p and n_p for every decision is

$$r_p - 1 < sn_p \leq r_p . \tag{9.8}$$

This means that

$$\bar{n}_p = [L_p(h_1 + h_2) - h_2 - f_p]/(s - p), \tag{9.9}$$

where f_p is some function of p that satisfies the conditions

$$0 \leq f_p < 1 \tag{9.10}$$

for all values of p, and $f_p \to 0$ as $p \to 0$ or 1. The formula

$$f_p \sim cq(1 - L_p), \tag{9.11}$$

where c is defined by (8.12), results in approximations to $\bar{n}_p$ that satisfy these conditions, and that are about the same as the approximations resulting from the combination of (7.6) and (8.2) for values of p in the neighborhood of s. Rewriting (9.9) and (9.11) as indicated in (8.9) shows that most of the difficulty in formula (8.7) is associated with the parameter h_2.

Formulas (8.8) and (8.9) are equivalent to Wald's formulas when $s = \frac{1}{2}$, and are therefore exact for this value of s when $2h_1$ and $2h_2$ are integers. Again like Wald's formulas, the approximations to L_p and $\bar{n}_p$ approach the exact values as $p \to 0$ or 1. Investigation of these formulas and their first derivatives for $p = s$ indicates increasingly close agreement with the approximations computed from (8.1) and (8.2) as h_2 increases in size.

REFERENCES

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

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

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

[4] M. A. Girshick, F. Mosteller, and L. J. Savage, "Unbiased estimates for certain binomial sampling problems with applications," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 13–23.

[5] M. A. Girshick, "Contributions to the theory of sequential analysis, II, III," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 282–298.

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

[7] G. Pólya, "Exact formulas in the sequential analysis of attributes," *Univ. California Publ. Math.*, N. S., Vol. 1 (1948), pp. 229–240.

[8] J. Robinson, "A note on exact sequential analysis," *Univ. California Publ. Math.*, N. S., Vol. 1 (1948), pp. 241–246.

[9] A. M. Walker, "Note on sequential sampling formulae for a binomial population," *Journal Roy. Stat. Soc., Ser. B*, Vol. 12 (1950), pp. 301–307.