On Sequential Tests of the Binomial Distribution

ARVID CORNELIUSSEN AND DANIEL W. LADD
The MITRE Corporation, Bedford, Massachusetts

In this paper, we consider the detailed characteristics of the standard sequential test of a simple hypothesis versus a single alternative for the case of the binomial distribution. A method is developed which permits evaluation of the associated probabilities of acceptance and/or rejection of the null hypothesis at each stage. From this information, the quantitative features of the test (Operating Characteristic and test termination probability) are completely determined. Curves are presented to facilitate assessment of the sequential procedure in practical cases. A sequential test is compared with a fixed sample size test having approximately the same Operating Characteristic. Wald's approximation for the average sample number is compared with exact values.

1. Introduction

Consider an experiment which can result in success with probability $p$ and failure with probability $1 - p$. The problem to be considered is that of testing the (null) hypothesis,

$$H_0 : p = p_0,$$

against the alternative

$$H_1 : p = p_1 < p_0$$

A simple test based on the observation of a fixed number of independent samples, $N$, can be developed as follows: Let $\alpha$ be the magnitude of the Type I error (probability that $H_1$ will be rejected when true) that can be tolerated, and choose $k$ to most nearly satisfy

$$L(p_0) = E(N, k, p_0) = \sum_{i=0}^{N-k} \binom{N}{i} p_0^i (1 - p_0)^{N-i} \approx 1 - \alpha$$

$H_0$ will be accepted if there are at least $k$ out of $N$ successes and rejected ($H_1$ accepted) otherwise.

The function $L(p)$ is the Operating Characteristic (O.C.) of the test, the probability of accepting $H_0$ when the true parameter is $p$. In particular, $L(p_1) = \beta$, the probability of accepting $H_0$ when $H_1$ is true (Type II error).

The work of Wald [1] has indicated that some economy in the average number of samples required can be effected by the use of sequential methods. However, it is well known, e.g. Wetherill [2, pg. 21], that the distribution of the sample size typically has a large variance. Since a general, complete, and precise descrip-
tion of the distributional properties of test sample size is not achievable by the methods of Wald, this topic has received attention.

It is possible to obtain an exact formulation of the properties of the sequential test procedure in cases where the number of possible outcomes of an experiment (or sampling) at any stage is finite.

One example of such a case is the treatment of the binomial distribution as developed by Burman [3]. By formulating an integer approximation to the sequential procedure, he derived and solved difference equations which result in expressions for the O.C. curve, the average sample size, and the variance of the sample size. His formulas are, however, difficult to evaluate numerically.

Polya [4] has studied the case where the slope, \( s \), of the Wald procedure is a rational number. For this case, he has developed recursion methods which permit a complete analysis of the problem. Robinson [5] has performed calculations by Polya's formulae and compared the results with Wald's approximations for eight selected cases. However, even when \( s \) is rational, the computational effort is substantial in many cases of interest.

Lindley and Barnett [6] develop an exact analysis for a sequential sampling plan significantly different from that of Wald.

An approach of greater similarity to that developed in this paper is that taken by Aroian [7], in his treatment of the truncated exponential distribution. By developing recursion formulas for probabilities on a grid network, he obtains exact expressions for the O.C. curve and the average time to test termination.

In a recent paper [8], the same author has described a direct method for determining the significant quantitative properties of sequential tests. For a Wald sequential test of a simple hypothesis, the procedure involves the computation of successive convolutions. He has applied this method to truncated sequential tests of the mean of a normal distribution and to a binomial test. For the latter case, he obtains results similar to those in [9].

The literature on sequential analysis is quite extensive, as may be seen from the bibliographies of Johnson [10] and Jackson [11].

In the procedure developed below, we are concerned with non-truncated sequential tests of the binomial distribution, and our objective is to obtain the exact distribution of sample numbers (test stages) required to terminate the test, together with the identification of acceptance and rejection stages. From the distribution, we obtain the mean (ASN), the O.C. function, percentiles, and various other parameters which provide an assessment of specific proposed test designs. This is accomplished by development of a recursion algorithm which can be readily applied to a large class of problems characterized by the binomial distribution, and which requires only modest computing capability.

Numerical results based on the algorithm are presented, offering comparison with the equivalent fixed sample test and Wald's approximate expression for the average sample number [1]. Numerical results to aid in the design of sequential tests are tabulated.

2. Basic Theory of the Method

Following Wald [1], consider a sequential test procedure in which \( x_1, x_2, \ldots, x_n, \ldots \) each represent a success or a failure in successive outcomes of
sequential trials. If $x_i$ is a success, we set $\log \left( \frac{p_{1,i}}{p_0,i} \right) = \log \frac{p_1}{p_0} = -S_D$, while if $x_i$ is a failure, we set $\log \left( \frac{p_{1,i}}{p_0,i} \right) = \log \left( \frac{(1 - p_1)/(1 - p_0)}{1 - p_0} \right) = S_U$. For the $m$th trial, we set $\log \left( \frac{p_{1,m-1}}{p_0,m-1} \right) = \log \left( \frac{p_1}{p_0} \right) + S_D$ for a failure or $\log \left( \frac{p_{1,m-1}}{p_0,m-1} \right) = -S_D$ for a success.

A graphical representation of a hypothetical test sequence is depicted in Figure 1, where the sample size, $m$, is the abscissa, and $\log \left( \frac{p_{1,m}}{p_0,m} \right)$ is the ordinate (for convenience we abbreviate $p_{1,m}$ to $p_{1,m}$, etc.). Horizontal lines are shown at $B_U = \log \left( \frac{(1 - \beta)/(\alpha)}{1 - \alpha} \right)$ and $B_D = \log \left( \frac{\beta}{1 - \alpha} \right)$. A particular sample sequence can be represented in the test diagram by a step function with discontinuities of either $S_D$ or $-S_D$ at each $m$. The first time that this discontinuous line intersects the lines $B_U$ or $B_D$, the test is terminated with rejection or acceptance of the null hypothesis respectively.

Figure 1 shows a test diagram for the parameter values $p_0 = 0.7$, $\alpha = 0.048$, $p_1 = 0.4$, and $\beta = 0.127$. These particular values of $\alpha$ and $\beta$ were selected to coincide with those of a fixed sample size test with $N = 20$, in particular $E(20, 11, 0.7) = 0.952 = 1 - \alpha$ and $E(20, 11, 0.4) = 0.127 = \beta$. 
Suppose the test has not terminated prior to step m. Then, the value of 
\[ \log \frac{p_{1,m}}{p_{0,m}} \] at step m must assume one of the values \( L_{n_1,n_2}^{(m)} \), where
\[ L_{n_1,n_2}^{(m)} = n_1 S_U - n_2 S_D, \] (1)
where \( n_1 \) and \( n_2 \) are such that
\[ n_1 + n_2 = m; \quad n_1 \geq 0; \quad n_2 \geq 0 \] (2)
and
\[ B_D - S_D < L_{n_1,n_2}^{(m)} < B_U + S_U. \]

In Figure 1, these possible states are illustrated for \( m = 9 \) by the circled points; \( L_{6,3}^{(8)}; L_{4,5}^{(8)}; L_{6,3}^{(9)}; \) and \( L_{7,7}^{(9)} \).

We can associate with each state \( L_{n_1,n_2}^{(m)} \), a probability \( p_{n_1,n_2}^{(m)} \), that \( \log \frac{p_{1,m}}{p_{0,m}} = L_{n_1,n_2}^{(m)} \) after \( m \) samples have been processed. This probability is the sum of the probability associated with each sample sequence for which \( B_j < \log \frac{p_{1,m}}{p_{0,m}} < B_{j+1} \) for \( j = 1, 2, \ldots, m - 1 \) and \( \log \frac{p_{1,m}}{p_{0,m}} = L_{n_1,n_2}^{(m)} \). It is easy to show that
\[ p_{n_1,n_2}^{(m)} = \varepsilon_1(1 - p)p_{n_1-1,n_2}^{(m-1)} + \varepsilon_2 p_{n_1,n_2-1}^{(m-1)} \]
where
\[ \begin{align*}
\varepsilon_1 &= \begin{cases} 1: & L_{n_1-1,n_2}^{(m-1)} > B_D \\ 0: & L_{n_1-1,n_2}^{(m-1)} \leq B_D \end{cases} \\
\varepsilon_2 &= \begin{cases} 1: & L_{n_1,n_2-1}^{(m-1)} < B_U \\ 0: & L_{n_1,n_2-1}^{(m-1)} \geq B_U \end{cases}
\end{align*} \]

Starting from the relation \( p_0^{(1)} = 1 - p \) and \( p_{1,0}^{(1)} = p \), we can, by recursion, obtain \( p_{n_1,n_2}^{(m)} \) for any \( m \) and for \( n_1 \) and \( n_2 \) satisfying equation (2).

If for a given \( m \),
\[ IP \left[ \frac{mS_U - B_U}{S_U + S_D} \right] > \frac{(m - 1)S_U - B_U}{S_U + S_D} \] (4a)
where \( IP \) represents the integer part, there is a rejection at the \( m \)th stage with
\[ n_1 = m - IP \left[ \frac{mS_U - B_U}{S_U + S_D} \right] \] (4b)
For a proof of equation (4), see the Appendix. If
\[ IP \left[ \frac{mS_D + B_D}{S_U + S_D} \right] > \frac{(m - 1)S_D + B_D}{S_U + S_D} \] (5a)
there is an acceptance at the \( m \)th stage with
\[ n_1 = IP \left[ \frac{mS_D + B_D}{S_U + S_D} \right] \] (5b)

The sum over all \( m \), satisfying (5a), of the probabilities \( p_{n_1,n_2}^{(m)} \), where the values of \( n_1 \) for each such \( m \) are given by (5b) is the Operating Characteristic. The probability of termination at stage \( m \) is the sum over all \( n_1 \) satisfying (4) or (5) of the probabilities \( p_{n_1,n_2}^{(m)} \); the aggregate of these termination prob-
abilities constitutes the probability distribution function of the test termination time and the ASN is the expected value of this random variable.

Although it is possible to proceed further analytically, equations (1) (5) constitute a suitable algorithm for a computer solution.

3. RESULTS AND DISCUSSION

The algorithm described above has been implemented to yield distributions of the sample size. The run terminates as soon as the cumulative probability of termination exceeds 0.999.

An objective of this study was to compare sequential tests against standard tests with fixed sample sizes. The case selected for comparison is that depicted in Figure 1 (p₀ = 0.7, α = .048, p₁ = 0.4, β = .127), corresponding to a fixed sample size with N = 20 and k = 11.

The basic information obtained is depicted in the plots of Cumulative Termination Probability, shown as Figure 2. An alternative means of depicting this information is by means of Termination Probability Contours, shown in Figure 3. This curve is obtained directly from the corresponding Figure 2 and shows the number of samples required for 50%, 80%, 95%, and 99.9% probability of termination as a function of the true (unknown) probability, p. Because of the approximate symmetry in the Termination Contours, it suffices to show only values of p ≤ 0.55 in Figure 2.

In Figure 4, we have shown both the O.C. curve and the probability of terminating before the corresponding fixed sample size case. The crosses on Figure 4 show the corresponding points on the O.C. curve for the fixed sample size test.
It is clear from Figures 2, 3, and 4 that for a range of values of \( p \) in the vicinity of \( p = \frac{1}{2}(p_0 + p_1) \), the sequential test will not be preferable to the fixed sample size test, unless a suitable rule for termination of such cases is employed ([1], Section 3.8).

Figure 5 shows the discrete probability function for the number of samples required to terminate the test. The data is shown for \( p = \mu_1 = 0.4 \).

There are a number of factors that determine the form of the termination probability distribution. At a given stage, there is a probability of acceptance, of rejection, of neither, or of both. For example, in the case shown, trials 5, 6, 8, 10, 12, 14, 16, 17, 19, \( \ldots \) may result in rejection, trials 4, 6, 9, 11, 13, 15, \( \ldots \) may result in acceptance, trials 6, 26, 35, \( \ldots \) may result in either, and trials 7, 27, 36, \( \ldots \) can result in neither.

For neighboring stages, the probabilities associated with rejection are considerably larger than those associated with acceptance. For the case under consideration, it is seen from equations (4a) and (5a) that two successive rejection trials can have at most one non-rejection trial between them and also that two successive acceptance trials can have either one or two non-acceptance trials between them.

It is also of interest to assess the accuracy of Wald's approximate expression for the ASN ([1], page 99). In Figure 6 we show the mean sample number as computed by the program, and also Wald's approximate expression.

Figure 6 indicates that Wald's approximation may be quite misleading in some cases—It predicts a maximum ASN approximately 20% less than the exact calculation (there is some evidence that the approximation is better in large sample cases). However, in view of the very large spread in sample number,
we do not believe the ASN function to be a particularly useful measure of the effort required to conduct a sequential test. The Termination Probability Contours appear to be a more useful guide to the amount of sampling required.

4. Summary Data For Test Design

Table I presents pertinent summary data for a group of parameter values \((p_0, p_1, \alpha, \beta)\) which were chosen to have a fairly general range of applicability. For each set of parameters, we tabulate \(L(p)\) (the O.C. function), the average sample number (as computed by the program), and the sample number for which there is a 95 percent probability of terminating the test.

Because of the symmetry of the problem when \(\alpha = \beta\) it is redundant to tabulate the quantities for \(p_0 < 0.5\). Thus, in designing tests for such cases \((p_0 < 0.5)\), one enters the appropriate table with \(p' = 1 - p\) and employs the relations

\[
N_{30}(p') = N_{30}(p) \quad \text{ASN}(p') = \text{ASN}(p) \quad L(p') = 1 - L(p)
\]

![Figure 4—Test Characteristics](image)
A method has been developed for determining the exact statistical properties of the binomial sequential test. This theory has been applied to a comparison of a sequential test with an equivalent fixed sample test and with Wald's approximate expression for the average sample number. In addition, the statistical properties of sequential tests for a number of parameters of general interest for test design, are tabulated.

**Appendix—the Rejection States**

A sequential test, as defined in Section 2, terminates at the $m$th sample with rejection of the null hypothesis if integers $n_1$ and $n_2$ are such that the following inequality is satisfied:

$$B_U \leq L_{n_1, n_2}^{(m)} < B_U + S_U .$$

Writing $L_{n_1, n_2}^{(m)}$ in the form
ON SEQUENTIAL TESTS OF THE BINOMIAL DISTRIBUTION

\[ L_{n_1, n_2}^{(m)} = (m - n_s)S_D - n_sS_o, \]

the inequality \( A-1 \) can be written as an inequality for \( n_s \) in the form:

\[ \frac{(m - 1)S_D - B_D}{S_D + S_o} < n_s \leq \frac{mS_D - B_D}{S_D + S_o} \quad A-2 \]

To select a non-negative integer \( n_s \) satisfying \( A-2 \), it is necessary that the upper limit be non-negative. That is, there can be rejection states only for

\[ m \geq \frac{B_D}{S_D} \quad A-3 \]

Each value of \( m \) defines an interval \( I_m \) on the real line defined by equation \( A-2 \). Each such interval is of length \( S_D/(S_D + S_o) < 1 \). Thus, a given interval can contain at most one integer. It is clear that if an integer lies within the interval \( I_m \),

\[ IP\left( \frac{mS_D - B_D}{S_D + S_o} \right) > \frac{(m - 1)S_D - B_D}{S_D + S_o} \quad A-4 \]

Figure 6—Comparison of Average Sample Number
where \( \lfloor p \rfloor \) denotes integer part. It also follows that the integer \( n_2 \) is given by

\[
n_2 = \lfloor p \rfloor \left[ \frac{mS_u - Bu}{S_u + S_o} \right]
\]

with

\[
n_1 = m - \lfloor p \rfloor \left[ \frac{mS_u - Bu}{S_u + S_o} \right]
\]

The intervals \( I_k, k \geq B_u/S_o, k = k_1, k_1 + 1, k_1 + 2, \ldots \) contain all points \( X \) on the real line for which

\[
\frac{(k_1 - 1)S_u - Bu}{S_u + S_o} < X < \frac{(k_1 + 1)S_u - Bu}{S_u + S_o}
\]
and each such point $X$ lies in only one interval since the intervals are disjoint. Suppose that $X$ lies in $I_a$. As $X$ increases, it passes into $I_{a+1}$, then $I_{a+2}$ and so on. For the $m$th interval, let

$$IP\left[\frac{mS_U - B_U}{S_U + S_D}\right] = n_m$$

(The quantity $n_m$ may or may not lie in $I_a$—in the former case, $n_m = n_a$ and there is a rejection at state $m$). There are $n_m + 1$ non-negative integers $n_k \leq n_m$ and each of these integers lies in an interval $I_k$ for $k \leq m$. Thus there are

$$IP\left[\frac{mS_U - B_U}{S_U + S_D}\right] + 1$$

rejection states for $m$ or less samples.

| Test Summary Data (Continued) |

| $p_0 = 0.5, p_1 = 0.8$ | $p_0 = 0.8, p_1 = 0.7$ |

| $\alpha = 0.05$ | $\alpha = 0.10$ | $\alpha = 0.05$ | $\alpha = 0.10$ |

| $p$ | Lp N | ASN | N 95 | Lp N | ASN | N 95 | Lp N | ASN | N 95 |

| 0.0 | 0.00 | 5.0 | 5 | 0.00 | 4.0 | 4 | 0.00 | 4.0 | 4 |
| 0.1 | 0.00 | 5.5 | 7 | 0.00 | 4.4 | 6 | 0.00 | 4.4 | 6 |
| 0.2 | 0.00 | 6.3 | 9 | 0.00 | 5.0 | 7 | 0.00 | 5.0 | 7 |
| 0.3 | 0.00 | 7.3 | 12 | 0.00 | 5.8 | 10 | 0.00 | 5.8 | 10 |
| 0.4 | 0.00 | 8.3 | 18 | 0.00 | 7.0 | 13 | 0.00 | 7.0 | 13 |
| 0.5 | 0.00 | 11.2 | 20 | 0.00 | 8.9 | 17 | 0.00 | 8.9 | 17 |
| 0.6 | 0.00 | 15.5 | 29 | 0.00 | 12.2 | 20 | 0.00 | 12.2 | 20 |
| 0.7 | 0.00 | 19.8 | 35 | 0.00 | 19.8 | 35 | 0.00 | 19.8 | 35 |
| 0.8 | 0.00 | 27.3 | 45 | 0.00 | 28.2 | 45 | 0.00 | 28.2 | 45 |
| 0.9 | 0.00 | 35.1 | 54 | 0.00 | 35.1 | 54 | 0.00 | 35.1 | 54 |
| 1.0 | 0.00 | 45.1 | 78 | 0.00 | 45.1 | 78 | 0.00 | 45.1 | 78 |

| $p_0 = 0.7, p_1 = 0.6$ | $p_0 = 0.6, p_1 = 0.5$ |

| $\alpha = 0.05$ | $\alpha = 0.10$ | $\alpha = 0.05$ | $\alpha = 0.10$ |

| $p$ | Lp N | ASN | N 95 | Lp N | ASN | N 95 | Lp N | ASN | N 95 |

| 0.0 | 0.00 | 11.0 | 11 | 0.00 | 8.0 | 8 | 0.00 | 8.0 | 8 |
| 0.1 | 0.00 | 12.7 | 17 | 0.00 | 9.8 | 12 | 0.00 | 9.8 | 12 |
| 0.2 | 0.00 | 15.3 | 21 | 0.00 | 11.7 | 17 | 0.00 | 11.7 | 17 |
| 0.3 | 0.00 | 19.7 | 31 | 0.00 | 14.9 | 25 | 0.00 | 14.9 | 25 |
| 0.4 | 0.00 | 22.6 | 43 | 0.00 | 19.8 | 37 | 0.00 | 19.8 | 37 |
| 0.5 | 0.00 | 28.2 | 74 | 0.00 | 26.0 | 71 | 0.00 | 26.0 | 71 |
| 0.6 | 0.00 | 37.9 | 111 | 0.00 | 30.8 | 94 | 0.00 | 30.8 | 94 |
| 0.7 | 0.00 | 45.7 | 129 | 0.00 | 37.7 | 104 | 0.00 | 37.7 | 104 |
| 0.8 | 0.00 | 48.1 | 184 | 0.00 | 37.7 | 104 | 0.00 | 37.7 | 104 |
| 0.9 | 0.00 | 54.3 | 249 | 0.00 | 41.9 | 121 | 0.00 | 41.9 | 121 |
| 1.0 | 1.00 | 60.1 | 318 | 1.00 | 46.9 | 148 | 1.00 | 46.9 | 148 |
Corresponding results for acceptance terminations (equations 5a and 5b) can be proved in a similar fashion. The total number of termination states for m or less samples is given by:

\[ IP \left[ \frac{mS_U - B_U}{S_U + S_D} \right] + IP \left[ \frac{mS_U + B_D}{S_U + S_D} \right] + 2. \]

**REFERENCES**


