SEQUENTIAL TESTING OF SAMPLE SIZE

BY

DAVID G. HOEL

TECHNICAL REPORT NO. 17
MAY 31, 1967

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT USPHS-5T1 GM 25-09

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SEQUENTIAL TESTING OF SAMPLE SIZE

by

David G. Hoel

TECHNICAL REPORT NO. 17

May 31, 1967

PREPARED UNDER THE AUSPICES

OF

PUBLIC HEALTH SERVICE GRANT USPHS-5T1 GM 25-09*

*This research was also partially supported by U.S. Army Research Office Grant No. DA-ARO-D-31-124-G670 to the University of North Carolina - Chapel Hill.

Reproduction in Whole or in Part is Permitted for any Purpose of the United States Government

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
SEQUENTIAL TESTING OF SAMPLE SIZE

David G. Hoel

In life testing observations $x_1, \ldots, x_r$ are made in time until a predetermined number $r$ have been obtained from a total sample of $n$. By means of sequentially observing the $x_i$'s, methods are proposed for estimating and testing the unknown sample size $n$. Their properties are given and also a conjectured approximation to the a.s.n. of an s.p.r.t. is discussed.
1. INTRODUCTION

In certain "life-testing" problems, observations $x_1, \ldots, x_r$ are made in time until a predetermined number $r$ have been obtained from a total sample of $n$. Assuming that the sampling is random, we then have available the first $r$ order statistics from a sample of size $n$. From this information an inference is made concerning the distribution from which the observations have been taken. In this paper we will instead be concerned with the situation in which the population distribution is known but the sample size is unknown. An application might be where an unknown number of individuals in a population have been exposed at the same point in time to radiation or some other health problem. Then, assuming knowledge of the distribution of time from exposure until detection of exposure, one would be interested in estimating the total number of individuals affected.

Johnson [1962] has considered the nonsequential aspects of this problem giving the properties of the fixed sample size test based upon the likelihood ratio for discriminating between two values of $n$. He also discusses the maximum likelihood estimate of $n$.

In the following, we will both consider the properties of the sequential probability ratio test, s.p.r.t., for discriminating between two values of $n$ and give a sequential estimate of $n$. One reason for looking at this problem other than for its own sake, is that although the observations are neither independent nor identically distributed, the properties of this s.p.r.t. can be obtained exactly without resorting to
quadrature and are independent of the distribution from which the sample was drawn. This then provides a model to which various approximations to the properties of an s.p.r.t. may be applied. In particular we shall consider in section 3 a conjectured approximation to the average sample number, a.s.n., given by Bate [1955].

2. A SEQUENTIAL TEST OF SAMPLE SIZE

Suppose we have a random sample of size \( n \) from an absolutely continuous distribution \( G \), with probability density function \( g \). Let \( x_1, \ldots, x_n \) represent the \( n \) observations after they have been ordered such that \( x_1 \leq \ldots \leq x_n \). Then if \( f(x_1, \ldots, x_k|n) \) denotes the joint density of \( x_1, \ldots, x_k \), we have

\[
f(x_1, \ldots, x_k|n) = \frac{n!}{(n-k)!} \left[ \prod_{i=1}^{k} g(x_i) \right] \left[ 1-G(x_k) \right]^{n-k},
\]

for \( k = 1, \ldots, n \). Now suppose that the \( x \)'s are sequentially observed and that we wish to choose between two possible values of \( n \) by means of an s.p.r.t. discriminating between

\[
H_0: \ n = n_0 \quad \text{and} \quad H_1: \ n = n_1
\]

where \( n_0 < n_1 \). If we define

\[
L_k = \frac{f(x_1, \ldots, x_k|n_1)}{f(x_1, \ldots, x_k|n_0)} = \frac{n_1!(n_1-k)!}{n_0!(n_0-k)!} \left[ 1-G(x_k) \right]^{n_1-n_0},
\]

for \( k = 1, \ldots, n_0 \), and
\[ L_{n_0+1} = \begin{cases} A & \text{if } n > n_0 \\ B & \text{if } n = n_0 \end{cases} , \]

where \( 0 < B < l < A \), it then follows that Wald's s.p.r.t. of (1) is to continue sampling if \( B < L_k < A \) and to stop if either of these inequalities is violated, at which time \( H_0 \) is accepted if \( L_k \leq B \) while \( H_1 \) is accepted if \( L_k > A \). If sampling proceeds to the \( n_0 + 1 \) stage and \( X_{n_0+1} \) does not exist, then \( H_0 \) is true. Thus the procedure will be logical as long as we reject \( H_0 \) if \( X_{n_0+1} \) is observed and accept \( H_0 \) if \( X_{n_0+1} \) does not exist. In this respect there is considerable freedom in defining \( L_{n_0+1} \). It should also be noted that at each stage the sequential procedure depends solely on the most recently observed variable and the stage index.

For purposes of determining the properties of the test, we shall restrict our attention to those values of \( n \) which belong to the interval \([n_0, n_1]\). Suppose we are able to obtain

\[ p_n(j) = \mathbb{P}(L_j \leq B \text{ and } B < L_i < A \text{ for } i = 1, \ldots, j-1|n) \]

\[ q_n(j) = \mathbb{P}(L_j \geq A \text{ and } B < L_i < A \text{ for } i = 1, \ldots, j-1|n) \]

for \( j = 1, \ldots, n_0 \); then the properties of the test are known, since

\[ \mathbb{P}([\text{accept } H_0|n]) = \begin{cases} \sum_{j=1}^{n_0} p_n(j) & \text{if } n > n_0 \\ 1 - \sum_{j=1}^{n_0} q_n(j) & \text{if } n = n_0 \end{cases} \]
\[ P(N = j | n) = \begin{cases} \frac{p_n(j) + q_n(j)}{n_0} & \text{if } j \leq n_0 \\ 1 - \sum_{i=1}^{n_0} (p_n(i) + q_n(i)) & \text{if } j = n_0 + 1 \\ 0 & \text{if } j > n_0 + 1, \end{cases} \]

where $N$ denotes the sample size of the s.p.r.t.

Next define

\[ b_i = \max \left\{ 0, 1 - \left[ \frac{n_0!(n_1-k)!B}{n_1!(n_0-k)!} \right] \frac{1}{n_1-n_0} \right\} \]

\[ a_i = \max \left\{ 0, 1 - \left[ \frac{n_0!(n_1-k)!A}{n_1!(n_0-k)!} \right] \frac{1}{n_1-n_0} \right\} \]

and observe that $1 \geq b_i \geq a_i \geq 0$, $b_{i+1} \geq b_i$ and $a_{i+1} \geq a_i$. Also let

\[ B_i(t) = \int_{a_i}^{b_i} \int_{a_{i-1}}^{b_{i-1}} \cdots \int_{a_1}^{b_1} c(t_1, \ldots, t_i) dt_1 \cdots dt_i \]

for $t \geq a_i$ where

\[ c(t_1, \ldots, t_i) = \begin{cases} 1 & \text{if } 0 \leq t_1 \leq \ldots \leq t_i \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

and notice that $B_i(t)$ is a polynomial in $t$ of degree at most $i$. Now if we set $u_k = G(x_k)$, then

\[ p_n(j) = P(b_j \leq u_j \text{ and } a_i < u_i < b_i \text{ for } i = 1, \ldots, j-1 | n). \]
Since $u_k$ is the $k^{th}$ order statistic of a sample of size $n$ from a $[0,1]$ uniform distribution, we have

$$p_n(j) = \frac{n!}{(n-j)!} \int_{b_j}^{1} (1-t)^{n-j} B_{j-1} \left( \min(b_{j-1}, t) \right) dt$$

$$= \frac{n!}{(n-j+1)!} (1-b_j)^{n-j+1} B_{j-1} \left( b_{j-1} \right),$$

where $B_0(t) = 1$. Similarly, we have

$$q_n(j) = P[a_j > u_j \text{ and } a_i < u_i < b_i \text{ for } i = 1, \ldots, j-1 | n]$$

$$= \frac{n!}{(n-j)!} \int_{a_{j-1}}^{a_j} (1-t)^{n-j} B_{j-1} \left( \min(b_{j-1}, t) \right) dt$$

$$= \begin{cases} 
\frac{n!}{(n-j)!} \int_{a_{j-1}}^{a_j} (1-t)^{n-j} B_{j-1}(t) dt & \text{for } b_{j-1} \geq a_j \\
\frac{n!}{(n-j)!} \int_{a_{j-1}}^{b_{j-1}} (1-t)^{n-j} B_{j-1}(t) dt + \frac{n!}{(n-j+1)!} \left((1-b_{j-1})^{n-j+1} - (1-a_j)^{n-j+1}\right) B_{j-1}(b_{j-1}) & \text{for } b_{j-1} < a_j.
\end{cases}$$

Now since $B_i(t)$ is a polynomial, the above expressions can easily be used to obtain $p_n(j)$ and $q_n(j)$. Although there does not appear to be a simple expression for $B_i(t)$, it can be obtained recursively for $a_i \leq t \leq b_i$ as follows:
\[
B_{i+1}(t) = \begin{cases} 
(t-a_{i+1})B_1(b_i) & \text{if } b_i \leq a_{i+1} \leq t \\
\int_{a_{i+1}}^{t} B_1(u)du & \text{if } a_{i+1} \leq t \leq b_i \\
\int_{a_{i+1}}^{b_i} B_1(u)du + (t-b_i)B_1(b_i) & \text{if } a_{i+1} \leq b_i \leq t.
\end{cases}
\]

In Table 1 a comparison is given with the fixed sample results to be found in Table 2 of Johnson [1962]. The values of \( A \) and \( B \) were varied so that the error probabilities agree exactly with those found by Johnson. Also it is observed that the typical savings of about one-third is found in comparing the a.s.n. with the sample size of the nonsequential test. Finally, the sequential test has the feature that although the stopping rule depends upon the distribution \( G \), the properties of the test do not. A computing algorithm and additional tables of these properties are given in section 5.

---

Insert Table 1
---

3. AN APPROXIMATION TO THE A.S.N.

Bhate [1955] has conjectured that for an s.p.r.t., the a.s.n. may be obtained approximately as the solution for \( m \) of

\[
\hat{\theta}(Z_m) = P[\text{accept } H_0] \log B + P[\text{accept } H_1] \log A
\]

(3)

where \( Z_m \) denotes the log likelihood ratio of the first \( m \) observations. For the case where \( Z_m \) is a sum of independent and identically distributed
### TABLE 1

<table>
<thead>
<tr>
<th>$n_0$</th>
<th>$n_1$</th>
<th>$\alpha = \beta$</th>
<th>$r$</th>
<th>$\delta_{n_0} (N)$</th>
<th>$\delta_{n_1} (N)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>15</td>
<td>0.294</td>
<td>5</td>
<td>3.2</td>
<td>3.7</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>0.185</td>
<td>5</td>
<td>3.0</td>
<td>3.8</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>0.123</td>
<td>5</td>
<td>2.8</td>
<td>3.8</td>
</tr>
<tr>
<td>15</td>
<td>20</td>
<td>0.360</td>
<td>5</td>
<td>3.3</td>
<td>3.7</td>
</tr>
<tr>
<td>15</td>
<td>25</td>
<td>0.266</td>
<td>5</td>
<td>3.1</td>
<td>3.8</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td>0.394</td>
<td>5</td>
<td>3.4</td>
<td>3.7</td>
</tr>
<tr>
<td>15</td>
<td>20</td>
<td>0.256</td>
<td>10</td>
<td>6.3</td>
<td>7.3</td>
</tr>
<tr>
<td>15</td>
<td>25</td>
<td>0.134</td>
<td>10</td>
<td>5.8</td>
<td>7.4</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td>0.326</td>
<td>10</td>
<td>6.4</td>
<td>7.1</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td>0.233</td>
<td>15</td>
<td>9.8</td>
<td>11.2</td>
</tr>
</tbody>
</table>

where $r$ is the fixed sample size and $\alpha, \beta$ are the corresponding type I, II errors, respectively.
random variables, (3) is simply Wald's formula for approximating the a.s.n. In the more general situation it appears to be a natural extension of Wald's formula.

In some cases of interest, for example, the sequential t-test, the calculation of the left-hand side of (3) is quite laborious. So, Ray [1956] has made a further conjecture. He suggests that if \( Z_m \) may be expressed as \( \log \ell(t_m) \), then the left-hand side of (3) be replaced by \( \log \ell[\hat{g}(t_m)] \).

Applying Bhat's conjecture to the test of section 2, we have that

\[
\hat{g}_n[\log L_k] = \log(n_1) - \log(n_0) + (n_1-n_0)\hat{g}_n[\log(1-u_k)].
\]

Since \( u_k \) is distributed as the \( k \)th order statistic from a uniform distribution, it follows that \( -\log(1-u_k) \) is distributed as the \( k \)th order statistic from an exponential distribution. Hence \( \hat{g}(-\log(1-u_k)) = \frac{1}{n-1} \sum_{i=1}^{k} \frac{1}{n-i+1} \) and therefore, \( \hat{g}_n(\log L_k) \) is equal to

\[
\log(n_1) - \log(n_0) - (n_1-n_0)\sum_{i=1}^{k} \frac{1}{n-i+1} + (n_1-n_0)\log(1-\hat{g}_n(u_k)).
\]

(4)

For Ray's approximation we replace \( \hat{g}_n[\log L_k] \) by

\[
\log(n_1) - \log(n_0) + (n_1-n_0)\log(1-\hat{g}_n(u_k)),
\]

which is equal to

\[
\log(n_1) - \log(n_0) + (n_1-n_0)\log(1-\frac{k}{n+1}).
\]

(5)

For purposes of comparing the expressions (4) and (5), it should be noted that

\[
-\log(1-\frac{k}{n}) > \sum_{i=1}^{k} \frac{1}{n-i+1} > -\log(1-\frac{k}{n+1}).
\]
Now by equating (4) and (5) to the previously determined values of
$P[\text{accept } H_0] \log B + P[\text{accept } H_1] \log A$ and solving for $k$, we arrive
at the conjectured values of the a.s.n.

Table 2 is typical of the cases which were considered. From the
values it appears that the approximation cannot be relied on for all $n$
Siskind [1964] considered both of these approximations in connection with
the two-sided sequential t-test. There he found the approximation to be
adequate except midway between his hypothesized values.

---------------
Insert Table 2
---------------

Bhate's conjecture consists of two approximations. The first is the
neglect of overshoot which is usually considered reasonable. In general,
however, when the a.s.n. is small, the overshoot might be quite large.
From Table 3, we observe that the approximation is good for the larger
values of the a.s.n. but poor for several of the small values. Thus,
for these small values, we might be justified in explaining away part of
the inaccuracy of Bhate's approximation by the neglect of overshoot.

---------------
Insert Table 3
---------------

The second approximation in Bhate's conjecture consists of setting
$\delta(N) = m$ where $m$ is the solution of $\delta(Z_m) = \delta(Z_N)$. Define $z_i = Z_i -$
$Z_{i-1}$, $z_1 = Z_1$ and restrict attention to the case where $z_1, z_2, \ldots$
are independent (this is, in effect, true for our test, since it can be
shown that $z_1, \ldots, z_n$ are independent). If we further assume$^1$ that

$$\delta(Z_N) = \sum_{i=1}^{\infty} \delta(z_i)P[N \geq 1],$$  \hspace{1cm} (6)

$^1$ $\delta(N) < \infty, \delta(z_i)$ uniformly bounded and independence are sufficient for (6).

For further details see Chow, Robbins & Teicher [1965].
<table>
<thead>
<tr>
<th>n</th>
<th>n_{0} = 10</th>
<th>n_{1} = 25</th>
<th>\alpha = \beta = .05</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>a.s.n.</td>
<td>Bhate</td>
<td>Ray</td>
</tr>
<tr>
<td>10</td>
<td>3.83</td>
<td>2.41</td>
<td>2.67</td>
</tr>
<tr>
<td>11</td>
<td>4.29</td>
<td>2.83*</td>
<td>3.02*</td>
</tr>
<tr>
<td>12</td>
<td>4.74</td>
<td>2.65*</td>
<td>3.23*</td>
</tr>
<tr>
<td>13</td>
<td>5.17</td>
<td>1.71*</td>
<td>3.13*</td>
</tr>
<tr>
<td>14</td>
<td>5.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>5.92</td>
<td>10.17</td>
<td>9.72</td>
</tr>
<tr>
<td>16</td>
<td>6.03</td>
<td>9.43</td>
<td>8.05</td>
</tr>
<tr>
<td>17</td>
<td>6.15</td>
<td>8.01</td>
<td>7.06</td>
</tr>
<tr>
<td>18</td>
<td>6.21</td>
<td>7.20</td>
<td>6.55</td>
</tr>
<tr>
<td>19</td>
<td>6.21</td>
<td>6.72</td>
<td>6.26</td>
</tr>
<tr>
<td>20</td>
<td>6.17</td>
<td>6.36</td>
<td>6.04</td>
</tr>
<tr>
<td>21</td>
<td>6.09</td>
<td>6.09</td>
<td>5.82</td>
</tr>
<tr>
<td>22</td>
<td>6.00</td>
<td>5.86</td>
<td>5.63</td>
</tr>
<tr>
<td>23</td>
<td>5.89</td>
<td>5.65</td>
<td>5.46</td>
</tr>
<tr>
<td>24</td>
<td>5.77</td>
<td>5.46</td>
<td>5.30</td>
</tr>
<tr>
<td>25</td>
<td>5.66</td>
<td>5.29</td>
<td>5.15</td>
</tr>
</tbody>
</table>

Note: For those values which are followed by an asterisk, there are two solutions—the one which is given and another which falls between 10 and 11.
\[ \alpha = \beta = .05 \]

<table>
<thead>
<tr>
<th>( n_0 )</th>
<th>( n_1 )</th>
<th>a.s.n.</th>
<th>Bhate</th>
<th>a.s.n.</th>
<th>Bhate</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10</td>
<td>3.76</td>
<td>3.04</td>
<td>4.84</td>
<td>4.68</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>2.64</td>
<td>1.39</td>
<td>3.86</td>
<td>3.44</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>2.11</td>
<td>0.77</td>
<td>3.21</td>
<td>2.77</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>7.57</td>
<td>7.32</td>
<td>9.12</td>
<td>9.22</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>5.13</td>
<td>4.06</td>
<td>7.06</td>
<td>6.89</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>3.83</td>
<td>2.41</td>
<td>5.66</td>
<td>5.29</td>
</tr>
<tr>
<td>15</td>
<td>20</td>
<td>11.78</td>
<td>11.97</td>
<td>13.62</td>
<td>13.94</td>
</tr>
<tr>
<td>15</td>
<td>25</td>
<td>8.10</td>
<td>7.37</td>
<td>10.56</td>
<td>10.67</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td>16.21</td>
<td>16.75</td>
<td>18.25</td>
<td>18.76</td>
</tr>
</tbody>
</table>
then Ehate's second approximation consists of finding an m such that either

\[
\sum_{i=1}^{m} \delta(z_i) < \sum_{i=1}^{\infty} \delta(z_i) P[N \geq i] < \sum_{i=1}^{m+1} \delta(z_i)
\]

or

\[
\sum_{i=1}^{m+1} \delta(z_i) < \sum_{i=1}^{\infty} \delta(z_i) P[N \geq i] < \sum_{i=1}^{m} \delta(z_i),
\]

and approximating the a.s.n. by a value, which is found by interpolation, between m and m + 1 (if \( \sum_{i=1}^{m} \delta(z_i) = \sum_{i=1}^{\infty} \delta(z_i) P[N \geq i] \) then the a.s.n. is approximated by m). Hopefully it will then be the case that

\[
m \leq \sum_{i=1}^{\infty} P[N \geq i] \leq m + 1.
\]

In general (7) does not imply (8) as can be seen by defining \( z_{n+1} \) of section 2, to be a sufficiently small constant.

It can also be shown for an s.p.r.t. that \( \delta_{H_0}(z_i) \leq 0 \) and \( \delta_{H_1}(z_i) \geq 0 \), but in general it is not necessarily true that \( \delta(z_i) \geq 0 \) for all i or \( \delta(z_i) \leq 0 \) for all i. Thus \( \sum_{i=1}^{n} \delta(z_i) \) need not be either increasing or decreasing in n. Hence there is no assurance that a unique m which satisfies (7) exists (this is illustrated in Table 2 by n = 11, ..., 14).

Now, if \( u \geq \delta(z_i) \geq \ell > 0 \), it then follows from (7) that

\[
m \ell \leq \sum_{i=1}^{m} \delta(z_i) \leq \sum_{i=1}^{\infty} \delta(z_i) P[N \geq i] \leq \sum_{i=1}^{m+1} \delta(z_i) \leq (m+1)u
\]

and

\[
\ell \sum_{i=1}^{\infty} P[N \geq i] \leq \sum_{i=1}^{\infty} \delta(z_i) P[N \geq i] \leq u \sum_{i=1}^{\infty} P[N \geq i].
\]
Hence we have

\[ \frac{u}{m} \leq \delta(N) \leq \frac{u}{2} (m+1) \]

and a similar expression for \( 0 > u \geq \delta(z_i) \geq \delta \). Therefore we may conclude that the second approximation in Bhat's conjecture will be reasonable, in the case of independence, if there is little relative variation among the values of \( \delta(z_i) \) and they are bounded away from zero. In general, however, it is clear that Bhat's conjecture may often lead to fallacious results when applied to cases midway between \( H_0 \) and \( H_1 \).

4. ESTIMATION

Johnson [1962] has shown that the maximum likelihood estimator \( \hat{n} \), given \( x_1, \ldots, x_r \) of \( n \) is the integer value lying between the limits

\[ ru_r^{-1} - 1 \quad \text{and} \quad ru_r^{-1}. \]

For purposes of calculating the properties of \( \hat{n} \), he approximated \( \hat{n} \) with the continuous variate \( ru_r^{-1} \) finding\(^1\)

\[ \delta(\hat{n}) = \frac{r}{r-1} n \]

\[ \text{Var}(\hat{n}) = \left( \frac{r}{r-1} \right)^2 n \left( \frac{n-r+1}{r-2} \right). \]

\(^1\) The denominator of expression (15) in Johnson's paper should read \( (r-1) \ldots (r-s) \) instead of \( r(r-1) \ldots (r-s+1) \).
It also easily follows that \((r-1)u_r^{-1}\) is an unbiased estimator of \(n\) with variance

\[
\frac{n(n-r+1)}{r-2}.
\]

In giving a sequential estimate of \(n\), it is reasonable to base the stopping rule on \(u_r\) since it is sufficient. Thus, we shall take for our rule--continue sampling until \(u_k \geq c\). Then we have

\[
P[N=k] = P[u_1 < c, \ldots, u_{k-1} < c, u_k \geq c],
\]

which is easily shown to be equal to

\[
\binom{n}{k-1}c^{k-1}(1-c)^{n-k+1}.
\]

Therefore \(N - 1\) is distributed as binomial \(B(n,c)\). Thus we will take as our sequential estimator \(\tilde{n}\) of \(n\) to be \(\frac{N-1}{c}\). It then follows that

\[
\delta(\tilde{n}) = n
\]

\[
\text{Var}(\tilde{n}) = n\left(\frac{1}{c} - 1\right)
\]

\(\text{a.s.n.} = nc + 1\).

In comparing \(\text{Var}(\tilde{n})\) with the variances of the two fixed sample estimators, we have by setting \(nc + 1 = r\) that

\[
\text{Var}(\tilde{n}) = n\left(\frac{n-r+1}{r-1}\right).
\]

Although this value is smaller than the previous two, in practice we would not be able to predetermine the a.s.n. without knowing \(n\).
5. TABLES

For computing purposes the following algorithm simplifies the determination of $B_i(t)$ for $a_i \leq t \leq b_i$. Let $B_i(k,j)$ be the coefficient of $t^{i-1}$ in $B_i(t)$ for $b_{k-1} < t \leq b_k$. Then, since $B_i(t) = t^{a_i}$, we have $B_i(1,1) = -a_i$ and $B_i(1,2) = 1$. Now

(a) if $b_i \leq a_{i+1}$ then

\[
B_{i+1}(k,*) = 0 \quad \text{for } k \leq i
\]

\[
B_{i+1}(i+1,1) = -a_{i+1} \sum_{p=1}^{i+1} B_i(i,p)b_{i+1}^p
\]

\[
B_{i+1}(i+1,2) = \sum_{p=1}^{i+1} B_i(i,p)b_{i+1}^p
\]

\[
B_{i+1}(i+1,j) = 0 \quad \text{for } j > 2
\]

(b) if $b_i \geq a_{i+1}$ and $k \leq i$ then

\[
B_{i+1}(k,j) = \frac{1}{j-1} B_i(k,j-1) \quad \text{for } j > 1
\]

\[
B_{i+1}(i+1,1) = -\sum_{p=1}^{i+1} \frac{1}{p} B_i(i_0,p)a_{i+1}^p
\]

\[
B_{i+1}(k,1) = \sum_{n=1}^{k-i_0-1} \frac{1}{p} B_i(i_0+n,p)(b_{i_0+n}^p - b_{i_0+n-1}^p)
\]

\[+ \sum_{p=1}^{i+1} \frac{1}{p} B_i(i_0,p)(b_{i_0}^p - a_{i+1}^p) - \sum_{p=1}^{i+1} \frac{1}{p} B_i(k,p)b_{k-1}^p, \text{ for } k \geq i_0,
\]

where $i_0$ is defined as the smallest $j$ such that $b_j \geq a_{i+1}$.

(c) if $b_i \geq a_{i+1}$ and $k = i + 1$ then
\[ B_{i+1}(i+1,j) = 0 \quad \text{for} \quad j > 2 \]

\[ B_{i+1}(i+1,1) = \sum_{p=1}^{i+2} B_{i+1}(i,p) b_{i+1}^{p-1} - \sum_{p=1}^{i+1} B_i(i,p) b_i^p \]

\[ B_{i+1}(i+1,2) = \sum_{p=1}^{i+1} B_i(i,p) b_i^{p-1} . \]

Applying the algorithm, the operating characteristic function and a.s.n. are given in the following tables for various pairs of \( n_0 \) and \( n_1 \). In each case the values of \( A \) and \( B \) which produce \( \alpha = \beta = .05 \) are given.
### TABLE 4

\[ n_0 = 5 \quad n_1 = 10 \quad B = 0.1898 \quad A = 7.881 \]

<table>
<thead>
<tr>
<th>n</th>
<th>O.C.</th>
<th>a.n.s.n.</th>
<th>n</th>
<th>O.C.</th>
<th>a.n.s.n.</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.950</td>
<td>3.763</td>
<td>8</td>
<td>0.175</td>
<td>4.915</td>
</tr>
<tr>
<td>6</td>
<td>0.526</td>
<td>4.387</td>
<td>9</td>
<td>0.094</td>
<td>4.918</td>
</tr>
<tr>
<td>7</td>
<td>0.313</td>
<td>4.762</td>
<td>10</td>
<td>0.050</td>
<td>4.839</td>
</tr>
</tbody>
</table>

### TABLE 5

\[ n_0 = 5 \quad n_1 = 15 \quad B = 0.2150 \quad A = 9.4216 \]

<table>
<thead>
<tr>
<th>n</th>
<th>O.C.</th>
<th>a.n.s.n.</th>
<th>n</th>
<th>O.C.</th>
<th>a.n.s.n.</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.950</td>
<td>2.643</td>
<td>11</td>
<td>0.197</td>
<td>4.057</td>
</tr>
<tr>
<td>6</td>
<td>0.795</td>
<td>3.077</td>
<td>12</td>
<td>0.140</td>
<td>4.041</td>
</tr>
<tr>
<td>7</td>
<td>0.647</td>
<td>3.453</td>
<td>13</td>
<td>0.100</td>
<td>3.995</td>
</tr>
<tr>
<td>8</td>
<td>0.502</td>
<td>3.738</td>
<td>14</td>
<td>0.071</td>
<td>3.931</td>
</tr>
<tr>
<td>9</td>
<td>0.375</td>
<td>3.924</td>
<td>15</td>
<td>0.050</td>
<td>3.858</td>
</tr>
<tr>
<td>10</td>
<td>0.274</td>
<td>4.024</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 6

\[ n_0 = 5 \quad n_1 = 20 \quad B = 0.2574 \quad A = 10.0055 \]

<table>
<thead>
<tr>
<th>n</th>
<th>O.C.</th>
<th>a.n.s.n.</th>
<th>n</th>
<th>O.C.</th>
<th>a.n.s.n.</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.950</td>
<td>2.110</td>
<td>13</td>
<td>0.253</td>
<td>3.455</td>
</tr>
<tr>
<td>6</td>
<td>0.875</td>
<td>2.398</td>
<td>14</td>
<td>0.201</td>
<td>3.456</td>
</tr>
<tr>
<td>7</td>
<td>0.786</td>
<td>2.673</td>
<td>15</td>
<td>0.160</td>
<td>3.437</td>
</tr>
<tr>
<td>8</td>
<td>0.684</td>
<td>2.916</td>
<td>16</td>
<td>0.127</td>
<td>3.404</td>
</tr>
<tr>
<td>9</td>
<td>0.579</td>
<td>3.114</td>
<td>17</td>
<td>0.100</td>
<td>3.363</td>
</tr>
<tr>
<td>10</td>
<td>0.481</td>
<td>3.264</td>
<td>18</td>
<td>0.079</td>
<td>3.315</td>
</tr>
<tr>
<td>11</td>
<td>0.392</td>
<td>3.367</td>
<td>19</td>
<td>0.063</td>
<td>3.265</td>
</tr>
<tr>
<td>12</td>
<td>0.316</td>
<td>3.428</td>
<td>20</td>
<td>0.050</td>
<td>3.214</td>
</tr>
</tbody>
</table>
### TABLE 7

\( n_0 = 10 \quad n_1 = 15 \quad B = 0.14918 \quad A = 9.304 \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>o.c.</th>
<th>a.s.n.</th>
<th>( n )</th>
<th>o.c.</th>
<th>a.s.n.</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.950</td>
<td>7.569</td>
<td>13</td>
<td>0.188</td>
<td>9.305</td>
</tr>
<tr>
<td>11</td>
<td>0.568</td>
<td>8.489</td>
<td>14</td>
<td>0.098</td>
<td>9.286</td>
</tr>
<tr>
<td>12</td>
<td>0.342</td>
<td>9.070</td>
<td>15</td>
<td>0.050</td>
<td>9.121</td>
</tr>
</tbody>
</table>

### TABLE 8

\( n_0 = 10 \quad n_1 = 20 \quad B = 0.14845 \quad A = 11.518 \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>o.c.</th>
<th>a.s.n.</th>
<th>( n )</th>
<th>o.c.</th>
<th>a.s.n.</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.950</td>
<td>5.126</td>
<td>16</td>
<td>0.218</td>
<td>7.532</td>
</tr>
<tr>
<td>11</td>
<td>0.837</td>
<td>5.813</td>
<td>17</td>
<td>0.152</td>
<td>7.487</td>
</tr>
<tr>
<td>12</td>
<td>0.704</td>
<td>6.442</td>
<td>18</td>
<td>0.105</td>
<td>7.376</td>
</tr>
<tr>
<td>13</td>
<td>0.358</td>
<td>6.946</td>
<td>19</td>
<td>0.073</td>
<td>7.226</td>
</tr>
<tr>
<td>14</td>
<td>0.421</td>
<td>7.291</td>
<td>20</td>
<td>0.050</td>
<td>7.056</td>
</tr>
<tr>
<td>15</td>
<td>0.307</td>
<td>7.478</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 9

\( n_0 = 10 \quad n_1 = 25 \quad B = 0.16529 \quad A = 11.735 \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>o.c.</th>
<th>a.s.n.</th>
<th>( n )</th>
<th>o.c.</th>
<th>a.s.n.</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.950</td>
<td>3.829</td>
<td>18</td>
<td>0.286</td>
<td>6.209</td>
</tr>
<tr>
<td>11</td>
<td>0.901</td>
<td>4.286</td>
<td>19</td>
<td>0.225</td>
<td>6.209</td>
</tr>
<tr>
<td>12</td>
<td>0.831</td>
<td>4.741</td>
<td>20</td>
<td>0.176</td>
<td>6.166</td>
</tr>
<tr>
<td>13</td>
<td>0.742</td>
<td>5.165</td>
<td>21</td>
<td>0.137</td>
<td>6.092</td>
</tr>
<tr>
<td>14</td>
<td>0.643</td>
<td>5.531</td>
<td>22</td>
<td>0.107</td>
<td>5.996</td>
</tr>
<tr>
<td>15</td>
<td>0.541</td>
<td>5.820</td>
<td>23</td>
<td>0.083</td>
<td>5.888</td>
</tr>
<tr>
<td>16</td>
<td>0.445</td>
<td>6.026</td>
<td>24</td>
<td>0.064</td>
<td>5.774</td>
</tr>
<tr>
<td>17</td>
<td>0.359</td>
<td>6.152</td>
<td>25</td>
<td>0.050</td>
<td>5.658</td>
</tr>
</tbody>
</table>
### TABLE 10

<table>
<thead>
<tr>
<th>$n_0$ = 15</th>
<th>$n_1$ = 20</th>
<th>$B$ =</th>
<th>$A$ =</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>o.c.</td>
<td>a.s.n.</td>
<td>n</td>
</tr>
<tr>
<td>15</td>
<td>.950</td>
<td>11.779</td>
<td>18</td>
</tr>
<tr>
<td>16</td>
<td>.584</td>
<td>12.891</td>
<td>19</td>
</tr>
<tr>
<td>17</td>
<td>.354</td>
<td>13.606</td>
<td>20</td>
</tr>
</tbody>
</table>

### TABLE 11

<table>
<thead>
<tr>
<th>$n_0$ = 15</th>
<th>$n_1$ = 25</th>
<th>$B$ =</th>
<th>$A$ =</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>o.c.</td>
<td>a.s.n.</td>
<td>n</td>
</tr>
<tr>
<td>15</td>
<td>.950</td>
<td>8.097</td>
<td>21</td>
</tr>
<tr>
<td>16</td>
<td>.852</td>
<td>8.990</td>
<td>22</td>
</tr>
<tr>
<td>17</td>
<td>.727</td>
<td>9.819</td>
<td>23</td>
</tr>
<tr>
<td>18</td>
<td>.583</td>
<td>10.497</td>
<td>24</td>
</tr>
<tr>
<td>19</td>
<td>.443</td>
<td>10.970</td>
<td>25</td>
</tr>
<tr>
<td>20</td>
<td>.324</td>
<td>11.230</td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 12

<table>
<thead>
<tr>
<th>$n_0$ = 20</th>
<th>$n_1$ = 25</th>
<th>$B$ =</th>
<th>$A$ =</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>o.c.</td>
<td>a.s.n.</td>
<td>n</td>
</tr>
<tr>
<td>21</td>
<td>.593</td>
<td>17.460</td>
<td>24</td>
</tr>
<tr>
<td>22</td>
<td>.361</td>
<td>18.270</td>
<td>25</td>
</tr>
</tbody>
</table>

19
ACKNOWLEDGMENTS

I gratefully acknowledge the suggestions and guidance given by Dr. N. L. Johnson while I was a student at the University of North Carolina. I would also like to thank Dr. Rupert Miller for reading the manuscript.

BIBLIOGRAPHY


Siskind, V. (1964) On certain suggested formulae applied to the sequential t-test, Biometrika, 51, 97-106.