SEQUENTIAL ESTIMATION OF THE MEAN
OF A NEGATIVE BINOMIAL DISTRIBUTION

BY

MICHAEL BINNS

TECHNICAL REPORT NO. 85
JULY 31, 1974

PREPARED UNDER CONTRACT
NO0014-67-A-0112-0078 (NR-042-993)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SEQUENTIAL ESTIMATION OF THE MEAN
OF A NEGATIVE BINOMIAL DISTRIBUTION

by

MICHAEL BINNS
Stanford University
Canada Department of Agriculture, Ottawa

TECHNICAL REPORT NO. 85
July 31, 1974

PREPARED UNDER CONTRACT
N00014-67-A-0112-0078 (NR-042-993)
FOR THE OFFICE OF NAVAL RESEARCH

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

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SEQUENTIAL ESTIMATION OF THE MEAN OF A NEGATIVE BINOMIAL DISTRIBUTION

by

Michael Binns

1. Introduction

Studies of biological populations (see, for example, Bliss and Fisher, 1953) have shown that in many instances the distribution of the number of individuals per sampling unit obeys the negative binomial law. This may be defined as (Anscombe, 1950)

\[
P(r) = \text{probability of exactly } r \text{ individuals in a sampling unit} = \frac{\Gamma(k+r)}{\Gamma(r+1)\Gamma(k)} \left(\frac{k}{\theta+k}\right)^k \left(\frac{\theta}{\theta+k}\right)^r
\]

where \(\theta\) is the mean of the distribution and \(k\), the exponent, reflects the patchiness of the distribution: the larger the value of \(k\), the less patchy the distribution; the limiting distribution as \(k\) tends to infinity is Poisson.

Bliss and Owen (1958) have shown that, at least for certain biological populations, variation in the sample values of \(k\) may be explained adequately by sampling variation only. Thus, familiarity with such a population allows a good estimate of \(k\) to be obtained and used (cf. Waters, 1955). If the value of \(k\) is assumed to be known, the method described here may be used to estimate \(\theta\) with a precision which amounts to having a nearly constant coefficient of variation.

Much work has been done on the sequential estimation of the mean
of Poisson and exponential distributions [Sandelius (1950), Weiler (1972), Starr and Woodroofe (1972)], but rather less on the corresponding problem for the negative binomial distribution [Gerrard and Cook (1972)]. The procedure proposed here is to obtain random samples 

\[ X_1 \]

from a negative binomial distribution with mean \( \theta \) (unknown) and exponent \( k \) (known), and to continue sampling as long as the number of samples \( n \) and the cumulative totals \( s_n = \sum_{i=1}^{n} X_i \) satisfy

\[ (s_n - a^2)(nk - 0.5 - a^2) < a^4 \]

for some fixed a (see Figure 1). Define \( t_l \) to be the value of \( n \) when sampling stops - i.e., the point \( (t_l, s_{t_l}) \) is on or beyond the stopping boundary - and \( (t, s_t) \) to be the point of intersection of the straight line joining \( (t_{l-1}, s_{t_{l-1}}) \) to \( (t_l, s_{t_l}) \) and the boundary. Then we estimate \( \tilde{\theta} \) by \( \tilde{\theta} = ks_{t_l}/(tk-0.5) \). This "adjustment for continuity" is necessary to obtain the results below; the distribution of an estimate based on \( (t_l, s_{t_l}) \) alone would depend, in a complicated way, on \( \theta \) and \( k \) (see Section 5).

It will be shown that \([\tilde{\theta}(1-c), \tilde{\theta}(1+c)]\) is a confidence interval for \( \theta \) with confidence probability \( 1-\alpha \). Since \( a \) may be regarded as a function of \( c \) and \( \alpha \) (see Section 3), a boundary may be defined by \( c \) and \( \alpha \); some typical values are shown in Table 1. Furthermore the distribution of \( Y = \tilde{\theta}/\theta \) has (approximately) the simple form

\[ \Pr(Y \leq y) \approx \Phi(a\sqrt{y} - a/\sqrt{y}) \]

where
\[ \phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}z^2} \, dz. \]

2. Derivation of the Distribution of \( Y = \tilde{\theta}/\theta \)

Let \( X_1, X_2, X_3, \ldots \) be random samples from a negative binomial distribution with unknown mean \( \theta \) and known exponent \( k \). At each step compute \( s_n = \sum_{i=1}^{n} X_i \) and continue sampling until \( nk > a^2 + 0.5 \) and

\[ s_n \geq a^2 + \frac{4a}{nk - 0.5a^2} = b(n), \text{ say.} \]

Denote by \( t_1 \) the value of \( n \) when sampling stops, and by \( (t, s_t) \) the point of intersection of the sample path and the stopping boundary (see Figure 1) \( (t_1 - 1 < t < t_1) \). Estimate \( \theta \) by \( \tilde{\theta} = ks_t/(tk - 0.5) \).

We shall show in Section 5 that although the distribution of \( Y_1 = \tilde{\theta}_1/\theta = ks_{t_1}/[\theta(t_1k - 0.5)] \) is not (2.1) below, the distribution of \( Y = \tilde{\theta}/\theta \) is close enough to (2.1) for any differences to be ignored in practice.

The proof of Theorem 1 is almost the same as the proof of Theorem 2.1 in Binns (1974b). It is given here again for completeness.

**Theorem 1.**

If \( k > 0 \) is finite and \( \theta > 0 \), the distribution of \( Y = \tilde{\theta}/\theta = ks_t/[\theta(tk - 0.5)] \) is given approximately by

\[ \text{pr}(Y \leq y) \approx \Phi(a\sqrt{y} - a/\sqrt{y}). \]

(2.1)
where

$$\phi(z) = \int_{-\infty}^{Z} \varphi(z)dz$$

and

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

Proof.

Since for finite $k$ the boundary $b(n)$ is decreasing and $s_n$ is non-decreasing,

$$\Pr(t > n) = \Pr(t_1 > n) = \Pr(s_n \leq I[b(n)])$$

where $I[x]$ denotes the largest integer less than $x$. Using an approximation to the negative binomial distribution (Binns, 1974d),

$$\Pr(t > n) \approx \phi \left( \frac{(I[b(n)] + 0.5)k - (nk-0.5)\theta}{\sqrt{(I[b(n)] + nk)k\theta}} \right).$$

Writing $d = b(n) - I[b(n)] - 0.5$ ($|d| \leq 0.5$), $\theta^* = kb(n)/(nk-0.5)$,

$$\Pr(t > n) \approx \phi \left( \frac{[b(n)-d]k - (nk-0.5)\theta}{\sqrt{[b(n)-d-0.5+nk]k\theta}} \right)$$

$$= \phi \left( \frac{[b(n)-d]k - (nk-0.5)\theta}{\sqrt{[b(n)(nk-0.5)/a^2-d]k\theta}} \right)$$

$$= \phi \left( \frac{a[\theta^*-d/(nk-0.5)]}{\sqrt{[\theta^*-kda^2/(nk-0.5)^2]}} \right)$$

4
which, if $a^2$ and $n$ are large, leads to the approximation

\[(2.3) \quad \Pr(t > n) \approx \Phi\left[ a \left( \frac{\theta^* - \theta}{\sqrt{n} \theta^*} \right) \right].\]

Since the boundary is decreasing, $t > n$ implies $kb(t)/(kt-0.5) < \theta^*_n$, and since by definition $(t,s_\theta)$ is on the boundary,

\[\tilde{\theta} = \frac{ks_\theta}{(tk-0.5)} = \frac{kb(t)/(tk-0.5)}{\leq \theta^*_n}.\]

Thus $\Pr(\tilde{\theta} \leq \theta^*_n)$ and $\Pr(t > n)$ are equal. Treating $n$ as though it were varying continuously, $\theta^*_n$ may be replaced by an arbitrary $\theta^*$ in (2.3) yielding the result (2.1).

There are two points worth noting in the above:

(i) The terms ignored to obtain (2.3) are increasing functions of $k$. The distribution (2.1) may therefore become a worse approximation as $k$ increases.

(ii) The replacement of $\theta^*_n$ by an arbitrary $\theta^*$ may worsen the approximation. The use of linear interpolation is intended to overcome this.

Monte Carlo methods were used to study the combined effects of all approximations involved (Section 5).

3. Properties of the Estimate

The distribution (2.1) of $Y = \tilde{\theta}/\theta$ has been investigated elsewhere (Binns, 1974b); the main results are as follows:

(i) Given $\alpha$ and $c$, $a$ may be calculated from the expression

\[(3.1) \quad \Phi[ca/\sqrt{(1-c)}] - \Phi[-ca/\sqrt{(1-c)}] = 1-\alpha.\]
Some typical values of $\theta$ are presented in Table 1.

(ii) Moments of $\tilde{\theta}$:

\[(3.2)\]
\[E(\tilde{\theta}) = \theta(2\theta^2+1)/2\theta^2 \]
\[\text{var}(\tilde{\theta}) = \theta^2(4\theta^2+5)/4\theta^4 . \]

(iii) Average sample size:

Since $(t, s_t)$ is on the boundary,

\[\tilde{\theta} = k s_t / (tk - 0.5) \]
\[= k s^2 / (tk - 0.5 - a^2) . \]

Therefore $\theta(tk - 0.5 - a^2)/ka^2$ has the same distribution as $Y^{-1}$ and so also of $Y$. In particular the moments are equal:

\[E[\theta(tk - 0.5 - a^2)/ka^2] = E(Y) = (2a^2+1)/2a^2 . \]

Therefore

\[(3.3)\]
\[E(t) = \frac{0.5 + a^2}{k} + \frac{(2a^2+1)}{2\theta} = (a^2+0.5)(\frac{1}{k} + \frac{1}{\theta}) . \]

4. Incorrect Values of $k$

It is useful to know what happens when the value of the exponent used in the sample path, $K$ say, is not exactly equal to the true value, $k$. Writing $K = k(1+5)$ in the right-hand side of (2.2),

6
\[ pr(\theta \leq \theta^*_n) = \Phi \left( \frac{[b(n) - d]k - (nk - 0.5)d}{\sqrt{k\theta [b(n)(nk - 0.5)^2/a^2 - d]}} \right) \]

\[ = \Phi \left( \frac{\theta^*_nK\theta - dK/(nk - 0.5)}{\sqrt{C\theta^*_n - K\theta^2/(nk - 0.5)(nk - 0.5)}} \right) \]

(after some simple manipulation), where

\[ C = \frac{K(nk - 0.5)}{k(nk - 0.5)} = 1 - \frac{0.5\theta}{nk - 0.5} \]

Therefore, for small \( \theta \), the terms ignored in the approximation of Section 2 are almost unchanged. If they are ignored, the effect of not knowing \( k \) precisely is seen to be to shift the distribution of \( \tilde{\theta} \) to the left or right accordingly as \( \theta > 0 \) or \( \theta < 0 \), but the size of the shift is negligible. It is reasonable to suppose, therefore, that the sequential procedure is not seriously affected if \( k \) is not known exactly.

If \( k \) is not known exactly, one may regard the procedure in terms of double sampling: since \( nK \) must be at least \( a^2 + 0.5 \), obtain \( (a^2 + 0.5)/K \) samples and use them to update the value of \( K \), which may then be used for the remainder of the procedure. Also, as Robbins (1959) has shown for the normal distribution with unknown mean and variance, one may update \( K \) after every sample. Although in this case there is no guarantee that the distribution of \( Y = \tilde{\theta}/\theta \) is given by (2.1), the fact that a small error in \( K \) has little effect suggests that the distribution would be close to (2.1).
5. Monte Carlo Simulation

The empirical distributions of \( Y_1 = \tilde{\theta}_1 / \theta \) and of \( Y = \tilde{\theta} / \theta \) (see Section 2) were compared with the distribution (2.1) for various values of \( \theta \) and \( k \). Since there was no reason to suppose that comparisons would be different for different values of \( a^2 \), only \( a^2 = 385 \) was used. A wide range of values of the ratio \( \theta / k \) was tested - comparisons for different \( \theta \) and \( k \) given a fixed ratio were not expected to vary.

The results, shown in Tables 2 and 2a, indicate that the distribution of \( Y_1 \) differs from the distribution (2.1) in an unpredictable way. The differences between (2.1) and the distribution of \( Y \), however, appear to be ignorable. Although it was noted at the end of Section 2 that the approximation may be worse for larger \( k \), there seems to be no evidence for this.

6. Comparison with Other Sequential Procedures

It is interesting to compare this estimation procedure with one recently proposed by Gerrard and Cook (1974). Their proposal is to take samples sequentially but, instead of counting all individuals in a sample, they note only if the sample contains any individuals at all. By sampling until a predetermined number, \( \tau \) say, of empty samples has been found, the probability \( \pi = [k/(\theta+k)]^k \) of an empty sample can be estimated by standard methods. If \( k \) is known, this can be inverted to give an estimate of \( \theta \). If the individuals are hard to find (for example, golden nematodes in a soil sample), this procedure has much to recommend it.
A convenient way of comparing the two procedures is to compare their Kullback-Leibler information numbers. These are defined for discrete distributions as 

\[
I(\theta_1, \theta_2) = \sum p_i(\theta_1) \log[p_i(\theta_1)/p_i(\theta_2)] .
\]

Therefore, the information number for a negative binomial distribution \[I(\text{NB}, \theta_1, \theta_2, k)\] is

\[
I(\text{NB}, \theta_1, \theta_2, k) = \sum_{i=0}^{\infty} \frac{\Gamma(i+k)}{\Gamma(i+1) \Gamma(k)} \left( \frac{k}{\theta_1 + k} \right)^k \left( \frac{\theta_1}{\theta_2} \right)^i \left( \frac{\theta_2}{k+\theta_1} \right)^{i+k} 
\]

\[
= (\theta_1 + k) \log\left( \frac{k+\theta_2}{k+\theta_1} \right) + \theta_1 \log\left( \frac{\theta_1}{\theta_2} \right),
\]

while for Gerrard and Cook's induced negative binomial

\[
I(\text{GC}, \theta_1, \theta_2, k) = k \left( \frac{k}{\theta_1 + k} \right)^k \log\left( \frac{\theta_2 + k}{\theta_1 + k} \right) + \left[ 1 - \left( \frac{k}{\theta_1 + k} \right)^k \right] \log\left[ \frac{1 - \left( \frac{k}{\theta_2 + k} \right)^k}{1 - \left( \frac{k}{\theta_1 + k} \right)^k} \right].
\]

The ratio \( R(\theta_1, \theta_2, k) = I(\text{NB}, \theta_1, \theta_2, k)/I(\text{GC}, \theta_1, \theta_2, k) \) is tabulated in Table 3 for various values of \( \theta_1, \theta_2, \) and \( k \). It is certainly true here, and may be true in general, that for given \( k \) and \( \theta_1 \), the ratio is an increasing or decreasing function of \( \theta_2 \) accordingly as \( k \) > or < 1. Certain values and limit values are of interest:
\[ R(\theta_1, \theta_2, l) = \theta_1 + 1 \]
\[ R(\theta_1, \theta_1, k) = \left( \frac{1}{k} + \frac{1}{\theta_1} \right) \left[ \left( \frac{\theta_1 + k}{k} \right)^k - 1 \right] \]
\[ R(\theta_1, 0, k) = \theta_1 \left[ 1 - \left( \frac{k}{k+\theta_1} \right)^k \right] \]
\[ R(\theta_1, \infty, k) = \left( \frac{k+\theta_1}{k} \right)^k \]
\[ R(\theta_1, \theta_2, \infty) = \frac{\theta_2 - \theta_1 + \theta_1 \log(\theta_1/\theta_2)}{e^{-\theta_1(\theta_2 - \theta_1) + (1-e^{-\theta_1})[\log((1-e^{-\theta_1})/(1-e^{-\theta_2}))]} \right] \]
\[ R(\theta_1, \infty, 0) = 1 \]
\[ R(0, \theta_2, k) = 1 \]

There are therefore some ranges of the parameters for which the method proposed here is much better, but also ranges where the efficiency is only slightly improved. Of course, \( R \) is never less than 1.

It is worth noting that for large \( \sigma^2 \), \( R(\theta_1, \theta_1, k) \) is also the ratio of expected sample sizes when the variances are constrained to be equal. \( R(\theta_1, \theta_2, \infty) \) is the ratio when the underlying distribution is Poisson.

Since for large \( k \) the two schemes are almost equivalent, Weiler's (1973) results for the Poisson distribution may be compared with the results here (letting \( k \) be large). Writing \( t \) as the sample size, for Weiler's scheme
\[ E(\theta t) = a^2 - 0.5 + 0.5 \theta \]
\[ \text{var}(\theta t) = a^2 - 0.5 + \theta^2/12 \]

and for the present scheme it may be shown that

\[ E(\theta t) = a^2 + 0.5 + O(1/k) \]
\[ \text{var}(\theta t) = a^2 + 5/4 + O(1/k) \]

The differences, negligible for large \( a^2 \), are due to approximations used in the derivations. See Section 2.

**Comments**

(i) It would have been possible to obtain stopping boundaries using Bayes methods and dynamic programming (the F-distribution would be a convenient prior). However, samples taken from biological populations and the estimates they may provide of abundance are generally collected together and used in some comparative study. In such a study it is important to have some idea of the precision of the data and to be able to control it beforehand. Also, leaving aside the general planning of the project, the (variable) cost of an individual sample should not affect its precision. The criterion used here (constant confidence probability for a certain form of interval) is reasonable in such a situation; in this case it also leads to a simple form for the distribution of the estimate which, as mentioned elsewhere (Binns, 1974b, Section 8) is approximately lognormal for large \( a^2 \).

(ii) It is possible to obtain stopping boundaries for other forms of confidence interval. (3.1) expresses the fact that when \( \theta \) is estimated
by \( \tilde{\theta} = k_{\alpha/2}(tk-0.5) \) and \( a^2(s_t^2 + tk-0.5) = st(tk-0.5) \), then
\( (\tilde{\theta} - c\tilde{\theta}, \tilde{\theta} + c\tilde{\theta}) \) is a \( 100(1-\alpha)\% \) confidence interval. Suppose that
an interval of the form \([\tilde{\theta} - \gamma f(\tilde{\theta}), \tilde{\theta} + \gamma f(\tilde{\theta})]\) is required, where \( f(x) \) is a
given function. Substitution of the above formula for \( a \) and
\( \tilde{\theta} = \gamma f(\tilde{\theta}) \) into the left-hand side of (3.1) and solution of the
equation will give a curve in the \((t,s_t)\) plane. If this curve is
decreasing in \( t \) (as is the curve in Figure 1), it may be used
without further analysis as a stopping boundary for the \( 100(1-\alpha)\% \)
confidence interval \( \tilde{\theta} \pm \gamma f(\tilde{\theta}) \). This was done, as an example, for
\( f(x) = \sqrt{(x-x_0^2/k)} \). As expected, the curves turned out to be \( n \neq \)
constant (Figure 2). Unfortunately it is no longer possible in
general to obtain a simple expression for the distribution of \( \tilde{\theta} \),
although it could be approximated by an obvious normal distribution.

(iii) The relationship between the negative and positive binomial
distributions may be used to suggest a sequential sampling procedure
for the positive binomial: let \( X_1, X_2, \ldots \) be independent positive
binomial, \( \Pr(X_i=1) = p = 1-\Pr(X_i=0) \), and \( s_n = \sum_{i=1}^{n} X_i \). Keep sampling
until the sample path of \( (n-s_n, s_n) \) reaches or crosses the boundary
\( na^2 = (n-s_n)s_n \). Let \( t \) be the value of \( n \) when sampling stops
(using linear interpolation to obtain a point on the boundary). Then
the odds ratio \( \rho = p/q = p/(1-p) \) may be estimated by \( \tilde{\rho} = s_t/(t-s_t) \)
and \( 100(1-\alpha)\% \) confidence interval \([\tilde{\rho}(1-c), \tilde{\rho}(1+c)]\). The numbers
\( \alpha, c, a^2 \) are as before. Also

\[ \Pr(Y = \tilde{\rho}/\rho \leq y) \approx \Phi(a\sqrt{y} - a/\sqrt{y}). \]
(iv) Anscombe (1948) proposed a variance stabilizing transformation for a negative binomial \(X\), namely \(W = \log(X+k/2)\) which has the approximate variance \(\Psi'(k)\) when \(\theta\) is large and \(k > l\) \([\Psi(u)\) is the second derivative of \(\log\Gamma(u)\)]. How does a fixed size sample using this transformation compare with the method proposed here? Since so many approximations are involved, it is difficult to be precise. However, we have that \(\hat{\theta} = ks_t/(tk-0.5)\) is approximately lognormal with variance \(a^{-2}\). For the fixed sample size, \(\Sigma X_i + nk/2\) is approximately lognormal with variance \(\Psi'(nk)\). Now [Abramowitz and Stegun (1972), page 260, formula 6.4.12]

\[
\Psi'(nk) \sim (nk)^{-1}.
\]

So, constraining the variances to be equal by letting \(nk = a^2\), the ratio of sample sizes (using formula 3.3 above) is

\[
r = n/E(t) = \frac{a^2\theta}{(a^2+0.5)(k+\theta)} = \frac{\theta}{(k+\theta)}.
\]

Although this appears to show that Anscombe's method is preferable, it should be noted that his derivation of the variance of \(W = \log(X+k/2)\) was made letting \((\theta+k)/\theta\) tend to 1. So \(r \neq 1\), and we are in effect treating the case where the sample path of the proposed method meets the boundary near the asymptote \(nk = a^2+0.5\).
Table 1. Values of $a^2$ corresponding to various $c$ and $\alpha$, and expected sample sizes (ESS) for the indicated values of $k$ and $\theta$ (formula 3.3).*

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$k$</th>
<th>$\theta$</th>
<th>$a^2$</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1</td>
<td>5</td>
<td>ESS</td>
<td>3192</td>
<td>804</td>
<td>361</td>
<td>206</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>ESS</td>
<td>2926</td>
<td>737</td>
<td>331</td>
<td>189</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>5</td>
<td>ESS</td>
<td>798</td>
<td>201</td>
<td>90</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>ESS</td>
<td>532</td>
<td>134</td>
<td>60</td>
<td>34</td>
</tr>
<tr>
<td>0.05</td>
<td>1</td>
<td>5</td>
<td>ESS</td>
<td>1537</td>
<td>385</td>
<td>174</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>ESS</td>
<td>1844</td>
<td>462</td>
<td>206</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>5</td>
<td>ESS</td>
<td>1691</td>
<td>424</td>
<td>189</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>ESS</td>
<td>461</td>
<td>116</td>
<td>52</td>
<td>29</td>
</tr>
</tbody>
</table>

*Expected sample size (ESS) is a symmetric function of $\theta$ and $k$. Therefore, for example, 3192 is also the ESS for $a^2 = 2660$, $\theta = 1$, $k = 5$. Also if for fixed $a$, both $k$ and $\theta$ are multiplied by a number $m$, say, the ESS is divided by $m$. 

14
Table 2. Monte Carlo frequency table for $Y$ and $Y$ (see Section 5).

\[
\theta^2 = 385 \text{ and various values of } \theta^1 \text{ and } k.
\]

<table>
<thead>
<tr>
<th>$\theta^1$</th>
<th>10</th>
<th>10</th>
<th>10</th>
<th>10</th>
<th>10</th>
<th>10</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>less than or equal to 0.90</td>
<td>77</td>
<td>64</td>
<td>70</td>
<td>68</td>
<td>73</td>
<td>56</td>
<td>59</td>
</tr>
<tr>
<td>0.90 to 0.92</td>
<td>126</td>
<td>121</td>
<td>121</td>
<td>122</td>
<td>131</td>
<td>124</td>
<td>128</td>
</tr>
<tr>
<td>0.92 to 0.94</td>
<td>246</td>
<td>246</td>
<td>245</td>
<td>264</td>
<td>257</td>
<td>243</td>
<td>259</td>
</tr>
<tr>
<td>0.94 to 0.96</td>
<td>397</td>
<td>375</td>
<td>397</td>
<td>374</td>
<td>383</td>
<td>394</td>
<td>413</td>
</tr>
<tr>
<td>0.96 to 0.98</td>
<td>537</td>
<td>507</td>
<td>535</td>
<td>535</td>
<td>570</td>
<td>539</td>
<td>547</td>
</tr>
<tr>
<td>0.98 to 1.00</td>
<td>616</td>
<td>583</td>
<td>610</td>
<td>653</td>
<td>643</td>
<td>635</td>
<td>613</td>
</tr>
<tr>
<td>1.00 to 1.02</td>
<td>605</td>
<td>638</td>
<td>587</td>
<td>595</td>
<td>616</td>
<td>620</td>
<td>580</td>
</tr>
<tr>
<td>1.02 to 1.04</td>
<td>512</td>
<td>517</td>
<td>511</td>
<td>527</td>
<td>483</td>
<td>471</td>
<td>521</td>
</tr>
<tr>
<td>1.04 to 1.06</td>
<td>377</td>
<td>384</td>
<td>384</td>
<td>385</td>
<td>377</td>
<td>364</td>
<td>344</td>
</tr>
<tr>
<td>1.06 to 1.08</td>
<td>244</td>
<td>289</td>
<td>261</td>
<td>219</td>
<td>231</td>
<td>292</td>
<td>261</td>
</tr>
<tr>
<td>1.08 to 1.10</td>
<td>139</td>
<td>152</td>
<td>153</td>
<td>148</td>
<td>131</td>
<td>123</td>
<td>152</td>
</tr>
<tr>
<td>over 1.10</td>
<td>123</td>
<td>124</td>
<td>126</td>
<td>110</td>
<td>105</td>
<td>139</td>
<td>123</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>18.6</td>
<td>4.2</td>
<td>11.4</td>
<td>10.2</td>
<td>23.9</td>
<td>12.3</td>
<td>32.8</td>
</tr>
</tbody>
</table>

a) Based on the distribution (2.1) in the text.
b) $\chi^2$ goodness of fit with 11 degrees of freedom.
Table 2a. Monte Carlo frequency table. Continuation of Table 2.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td><strong>expected values</strong></td>
<td><strong>Y</strong>&lt;sub&gt;1&lt;/sub&gt;</td>
<td><strong>Y</strong>&lt;sub&gt;1&lt;/sub&gt;</td>
<td><strong>Y</strong>&lt;sub&gt;1&lt;/sub&gt;</td>
<td><strong>Y</strong>&lt;sub&gt;1&lt;/sub&gt;</td>
<td><strong>Y</strong>&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
<tr>
<td>less than or equal to 0.90</td>
<td>77</td>
<td>83</td>
<td>86</td>
<td>92</td>
<td>93</td>
</tr>
<tr>
<td>0.90 to 0.92</td>
<td>126</td>
<td>140</td>
<td>132</td>
<td>128</td>
<td>126</td>
</tr>
<tr>
<td>0.92 to 0.94</td>
<td>246</td>
<td>179</td>
<td>212</td>
<td>249</td>
<td>268</td>
</tr>
<tr>
<td>0.94 to 0.96</td>
<td>397</td>
<td>403</td>
<td>386</td>
<td>392</td>
<td>406</td>
</tr>
<tr>
<td>0.96 to 0.98</td>
<td>537</td>
<td>591</td>
<td>565</td>
<td>561</td>
<td>536</td>
</tr>
<tr>
<td>0.98 to 1.00</td>
<td>616</td>
<td>552</td>
<td>587</td>
<td>619</td>
<td>616</td>
</tr>
<tr>
<td>1.00 to 1.02</td>
<td>605</td>
<td>647</td>
<td>626</td>
<td>577</td>
<td>583</td>
</tr>
<tr>
<td>1.02 to 1.04</td>
<td>512</td>
<td>472</td>
<td>497</td>
<td>520</td>
<td>524</td>
</tr>
<tr>
<td>1.04 to 1.06</td>
<td>377</td>
<td>407</td>
<td>375</td>
<td>381</td>
<td>375</td>
</tr>
<tr>
<td>1.06 to 1.08</td>
<td>244</td>
<td>253</td>
<td>254</td>
<td>239</td>
<td>231</td>
</tr>
<tr>
<td>1.08 to 1.10</td>
<td>139</td>
<td>138</td>
<td>148</td>
<td>125</td>
<td>117</td>
</tr>
<tr>
<td>over 1.10</td>
<td>123</td>
<td>135</td>
<td>132</td>
<td>117</td>
<td>125</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>42.4</td>
<td>12.0</td>
<td>7.4</td>
<td>10.8</td>
<td>46.8</td>
</tr>
</tbody>
</table>
Table 3. Values of $R(\theta_1, \theta_2, k) = I(NB, \theta_1, \theta_2, k)/I(SC, \theta_1, \theta_2, k)$

<table>
<thead>
<tr>
<th>$\theta_1$</th>
<th>$k$</th>
<th>.01</th>
<th>.1</th>
<th>1</th>
<th>10</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.01</td>
<td>3.40</td>
<td>2.67</td>
<td>2.13</td>
<td>1.83</td>
<td>1.65</td>
</tr>
<tr>
<td>0.1</td>
<td>1.47</td>
<td>1.44</td>
<td>1.36</td>
<td>1.29</td>
<td>1.24</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.10</td>
<td>1.10</td>
<td>1.10</td>
<td>1.10</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.06</td>
<td>1.06</td>
<td>1.06</td>
<td>1.08</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.05</td>
<td>1.05</td>
<td>1.06</td>
<td>1.08</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.01</td>
<td>13.70</td>
<td>8.04</td>
<td>4.77</td>
<td>3.32</td>
<td>2.63</td>
</tr>
<tr>
<td>0.1</td>
<td>4.30</td>
<td>3.79</td>
<td>2.98</td>
<td>2.39</td>
<td>2.05</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.64</td>
<td>1.66</td>
<td>1.75</td>
<td>2.13</td>
<td>2.43</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.60</td>
<td>1.62</td>
<td>1.72</td>
<td>2.21</td>
<td>2.58</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.01</td>
<td>73.40</td>
<td>33.40</td>
<td>14.00</td>
<td>7.16</td>
<td>4.61</td>
</tr>
<tr>
<td>0.1</td>
<td>21.90</td>
<td>16.80</td>
<td>9.92</td>
<td>5.92</td>
<td>4.14</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>11.00</td>
<td>11.00</td>
<td>11.00</td>
<td>11.00</td>
<td>11.00</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>12.00</td>
<td>13.80</td>
<td>23.10</td>
<td>204.60</td>
<td>706.30</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>12.70</td>
<td>15.20</td>
<td>29.60</td>
<td>1516.00</td>
<td>10019.00</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Stopping boundary for the procedure with an example of a sample path.

\( s_n = a^2 \)

\( nk = a^2 + 0.5 \)

\( (s_n - a^2)(nk - 0.5 - a^2) = a^4 \)
Figure 2. Stopping boundaries for 95% confidence intervals of the form
\[ \hat{\theta} \pm 0.1\sqrt{\hat{\theta}^2 + \frac{\hat{\theta}^2}{k}} \]
Acknowledgment

I would like to thank Professor Herman Chernoff for many useful discussions relating to this subject and for helping me clarify my thoughts.
References


References (Cont.)

Waters, W. E. (1955), "Sequential analysis of forest insect surveys," 
Forest Science, Vol. 1, page 68.

Weiler, H. (1972), "Inverse sampling of a Poisson distribution," Biometrics, 
Vol. 28, page 959.
Sequential Estimation of the Mean of a Negative Binomial Distribution

Michael Binns

Department of Statistics
Stanford University
Stanford, California

Office of Naval Research
Statistics & Probability Program
Code 436
Arlington, Virginia 22217

Approved for public release; distribution unlimited.

A sequential procedure is proposed to estimate the mean of a negative binomial distribution when the value of the exponent \( k \) is known. An approximation is obtained for the distribution of the estimate from which it may be shown that the precision amounts to having a predetermined coefficient of variation. The effect of imperfect information on \( k \) is investigated. Comparisons are made with other procedures including the variance-stabilizing logarithmic transformation.