A HYPOTHESIS TEST OF THE COVERAGE OF A SAMPLE FROM AN INFINITE POPULATION

Carlos M. Hernandez and Olga I. Codero-Brana

BU-1325-M March 1996
A HYPOTHESIS TEST OF THE COVERAGE OF A SAMPLE FROM AN INFINITE POPULATION

BY CARLOS M. HERNANDEZ AND OLGA I. CORDERO-BRAÑA *

ABSTRACT: We consider sampling from a population composed of $K$ classes or species in proportions $\theta_i$ where $K$ and $\theta_1, \ldots, \theta_K$ are unknown and the population size is infinite. By allowing $K$ not to be restricted to the number of species observed in the sample we obtain unrestricted maximum likelihood estimates of the $\theta_i$'s. Using these estimates a test of hypothesis on the coverage of a sample is given. The method does not require any of the usual assumptions, for example equal class sizes, but that all individuals have the same probability of being detected at any moment. Results from a Monte Carlo study suggest the practicality of the method in determining whether a desired coverage had been achieved.

KEY WORDS: coverage of a sample; unrestricted maximum likelihood estimator; singletons; number of classes.

*Carlos M. Hernandez is a graduate student at Cornell University, Biometrics Unit Ithaca, NY 14853-7801 and Olga I. Cordero-Braña is Assistant Professor, Department of Mathematics and Statistics, American University, Washington, DC 20016-8050. This research has been funded partially by a Conacyt Scholarship to Carlos Hernandez and by grant DEB-9253570 (Presidential Faculty Award to Carlos Castillo-Chavez).
1. INTRODUCTION

Consider sampling from a population composed of $K$ classes or "species" in proportions $\theta_i$ where $K$ and $\theta_1, \ldots, \theta_K$ are unknown. Two related problems associated with this situation are the estimation of $K$ and the coverage, $u$, of a sample on the basis of a sample of size $n$. In the latter, the aim is to make inferences on the proportion that accounts for the species found in the sample. That is, $u = \sum_{i=1}^{K} I_{(n_i > 0)} \theta_i$ where $n_i$ is the number of individuals from the $i$th class in the sample, $i = 1, \ldots, K$, and $I_A$ is the indicator function of the event $A$. However, these problems arise from different objectives. For example, the estimation of the number of species in a biological population provides information on the diversity of the whole population. The coverage, on the other hand, conveys information about the reach of the sample; this is very relevant for survey opinions, where it is desired to make inferences on the proportion of people whose opinions have been recorded. Note that the random vector $\mathbf{n} = (n_1, \ldots, n_K)^T$ is not observable; instead, we observe the vector $\mathbf{c} = (c_1, \ldots, c_n)^T$ where $c_j$ denotes the number of classes represented $j$ times in the sample. For example, $c_1$ is the number of singletons, $c_2$ is the number of twin pairs, and so on. Let $c = \sum_{j=1}^{n} c_j$. Under the assumption of equal class sizes, $u = c/K$ so that given a predictor $\hat{u}$ of $u$, $K$ is estimated by $\hat{K} = c/\hat{u}$. 
Bunge and Fitzpatrick (1993) provided a comprehensive review of the literature related to the estimation of the number of species and the coverage. Under the assumption of equal class sizes, Good (1953) and Good and Toulmin (1953) proposed the first estimator of \( u \) using the proportion of species in the sample represented by singletons. That is, \( \hat{u} = 1 - (c_1/n) \). Robbins (1968) used the proportion of singletons in the sample as an estimator of the proportion occupied by the uncovered species but only after taking an additional sample. The resulting estimator is unbiased. Engen (1975) used a negative binomial model and using a normal approximation to the ratio of two densities derived a confidence interval for \( u \). Starr (1979) considered generalizing Robbins (1968) procedure to an extended search of \( m \) individuals and derived a linear estimator for the expected value of the probability of finding a new species at the end of the \( n + m \) stage. Esty (1982) proposed an estimator of \( u \) by using a Poisson approximation to the number of classes represented at least twice in the population.

Here, we propose a simple and practical solution for applied researchers to decide if the desired coverage is reached. The only requirement for this procedure is that at any time all individuals have the same probability of being detected, namely \( 1/N \) where \( N \) is the unknown population size. Results from a moderate Monte Carlo study are presented.
2. THEORETICAL RESULTS

When assuming sampling from a multinomial distribution with vector of proportion parameters $\Theta$ one can get two types of maximum likelihood estimators. By restricting the dimension of $\Theta$ to equate the number of classes observed in the sample we obtain the classical maximum likelihood estimator, MLE. By allowing $K$ to vary freely, we get what we call the unrestricted maximum likelihood estimator, UMLE, of $\Theta$. Observe that the likelihood of the sample under the UMLE is always greater than or equal to the likelihood using the classical MLE since the optimization of the former is done over a larger set.

The following result considers the case where the sample is comprised of singletons only. Although a simple result it is essential to the proof of Proposition 2.2. Intuitively, if all individuals in the population are different then the probability that a sample of size $n$ contains singletons only is maximized.

Proposition 2.1 If a sample of size $n$ contains singletons only then the UMLE of $\Theta$ has dimension $N$, the population size, with $\theta_i = 1/N$ for $i = 1, \ldots, N$. In this case the coverage is minimized over all possible values.

Proof. In order to have a sample of size $n$ composed of singletons only it is necessary that every sampled individual be different from all the previous ones. Let $\theta_1$ be the population proportion of the first detected species, $\theta_2$
that of the second one, and so on. Note that \( \theta_j \), can be any of the \( \theta_i \), \( i = 1, \ldots, K \). The probability of getting \( n \) singletons is given by

\[
Pr(n \text{ singletons}) = (1 - \theta_1)(1 - \theta_1 - \theta_2) \cdots (1 - \theta_1 - \theta_2 - \cdots - \theta_{n-1}).
\]

This probability is maximized by letting \( \theta_j \) be the smallest possible; that is, by choosing \( \theta_j = 1/N \) for \( j = 1, \ldots, N \). But this requires \( K = N \), which implies that all individuals of the population must be different. To show that for this particular UMLE the coverage is minimized, note that after taking a sample of size \( n \), with singletons only, the minimum value for the proportion of covered species is \( n/N \) and for this population the coverage is effectively \( n/N \).

The next result deals with the case of a mixed sample of singletons and non-singletons species. To find the UMLE in this case it is necessary to have all non-singleton species present in the sample and the remaining species, observed singletons and unobserved ones, being singletons.

**Proposition 2.2** Suppose that a sample of size \( n \) contains \( r \) species, \( s \) of them being singletons and the remaining \( n-s \) individuals belonging to \( r-s \) different types. Let \( x_i \) denote the number of individuals of type \( i \) in the sample, \( i = 1, \ldots, r-s \). Then the UMLE of \( \Theta \) is given by \( \theta_i = x_i/n \) for the \( r-s \)
non-singleton species and the remaining individuals are all of different type
and have a total proportion of \( s/n \).

\textbf{Proof.} Consider the \( r \times 1 \) vector \( x = (x_1, \ldots, x_{r-s}, 1, \ldots, 1)^T \) with \( x_i \) being
the number of individuals of the \( i \)th non-singleton species, \( i = 1, \ldots, r-s \).

Note that the \( s \) singleton individuals can be considered indistinguishable, for
if we substitute any of them by another (imaginary) type, the structure of
\( x \) is not altered. This does not hold if we substitute a particular individual
from a non-singleton species. Let \( \theta_i \) denote the population proportion of
the \( i \)th non-singleton species detected in the sample, \( i = 1, \ldots, r-s \), and let
\( \phi_1, \ldots, \phi_{K-r+s} \) represent the proportions of the remaining \( K-r+s \) species. The
likelihood function of the sample, \( L(x|\theta, \phi) \), is given by

\[
L(x|\theta, \phi) = \frac{n!}{x_1! \cdots x_{r-s}! 1! \cdots 1!} \theta_1^{x_1} \cdots \theta_{r-s}^{x_{r-s}} \phi_1^0 \cdots \phi_{r-s}^0 \phi_{r-s+1}^1 \cdots \phi_{K-r+s}^0 + \\
\frac{n!}{x_1! \cdots x_{r-s}! 1! \cdots 1!} \theta_1^{x_1} \cdots \theta_{r-s}^{x_{r-s}} \phi_1^0 \phi_2^1 \cdots \phi_{r-s+1}^0 \phi_{r-s+2}^0 \cdots \phi_{K-r+s}^0 + \\
+ \cdots + \\
\frac{n!}{x_1! \cdots x_{r-s}!} \theta_1^{x_1} \cdots \theta_{r-s}^{x_{r-s}} \phi_1^0 \cdots \phi_{K-r+s}^0 \phi_{K-r+s+1}^1 \cdots \phi_{K-r+s}^1. \tag{1}
\]

Observe that there are \( (K-r+s)!/s!(K-r)! \) terms in the sum and that each
term represents a different way of getting \( s \) singletons out of the \( K-r+s \)
species. The set of the \( K \) classes in the population can be written as the
union of the set of non-singletons in the sample and the set of singletons in
the sample and the unobserved species. Let E be the set of non-singleton species in the sample and let Pr(E) denote its probability. Similarly, let E\(^c\) denote the complement of event E which consists of the remaining \(K-r+s\) species and let \(\Phi \equiv Pr(E^c) = \sum_{i=1}^{K-r+s} \phi_i\).

Define events A and B as follows:

\[
A = \{x_1 \text{ individuals of type } 1, \ldots, x_{r-s} \text{ individuals of type } r-s \},
\]
given that a sample of size \(n-s\) is taken from E\(^c\)

\[
B = \{s \text{ singletons given that a random sample of size } s \text{ is taken from } E^c\}.
\]

Then the likelihood function (1) can be written as

\[
L(x|\theta, \phi) = \frac{n!}{s!(n-s)!} [Pr(E)]^{n-s} [Pr(E^c)]^s Pr(A) Pr(B)
\]

\[
= \frac{n!}{s!(n-s)!} (1 - \Phi)^{n-s} \Phi^s Pr(A) Pr(B)
\]

(2)

But,

\[
Pr(A) = \frac{(n-s)!}{x_1! \cdots x_{r-s}!} \left( \frac{\theta_1}{\sum \theta_t} \right)^{x_1} \cdots \left( \frac{\theta_{r-s}}{\sum \theta_t} \right)^{x_{r-s}}
\]

\[
= \frac{(n-s)!}{x_1! \cdots x_{r-s}!} \theta_1^{x_1} \cdots \theta_{r-s}^{x_{r-s}} (1 - \Phi)^{-\sum_{i=1}^{r-s} x_i}
\]

\[
= \frac{(n-s)!}{x_1! \cdots x_{r-s}!} \theta_1^{x_1} \cdots \theta_{r-s}^{x_{r-s}} (1 - \Phi)^{-(n-s)}
\]

(3)

since the probability of event A reduces to a multinomial probability with \(r-s\) cells with proportions \(\theta_i/\sum \theta_t, \sum_{t=1}^{r-s} \theta_t = 1 - \Phi\), and \(\sum_{t=1}^{r-s} x_i = n-s\). Applying
(3) to (2) we obtain

\[ L(x|\theta, \phi) = \frac{n!}{x_1! \cdots x_{r-s-1}!} \theta_1^{x_1} \cdots \theta_{r-s-1}^{x_{r-s-1}} \phi^{x_{r-s}} \Pr(B) \]  

(4)

Observe that \( \Pr(B) \) depends on the relative proportion of the \( \phi_i \)'s with respect to \( \Phi \) and not on any of the \( \theta_i \)'s. Therefore we can maximize \( \Pr(B) \) independently of the rest of the expression in (4). But from Proposition 2.1, we know that \( \Pr(B) \) is maximized if each individual in the group \( E^c \) belongs to a different species and the maximum value occurs at \( \phi_i = 1/(N-n+s) \). Hence, maximizing (4) is equivalent to maximizing

\[ L(x|\theta, \phi) = \frac{n!}{x_1! \cdots x_{r-s-1}!} \theta_1^{x_1} \cdots \theta_{r-s-1}^{x_{r-s-1}} \phi^{x_{r-s}} \]  

(5)

But, \( \theta_1 + \cdots + \theta_{r-s} + \Phi = 1 \) and \( x_1 + \cdots + x_{r-s} + s = n \) so that the likelihood equation (5) represents the likelihood function of a multinomial sample for which the usual MLE of \( \Theta \) exists namely, \( \hat{\theta}_i = x_i/n \) and \( \hat{\phi} = s/n \).

As mentioned earlier, the UMLE maximizes over a larger set therefore the next result is stated without proof.

**Corollary 2.1** The likelihood function of \( x \) evaluated at the UMLE is greater or equal to the likelihood function evaluated at the MLE.
Corollary 2.2 If the proportion of undetected species is \( \Pi \), then the UMLE of \( \Pi \) is given by

\[
\hat{\Pi} = s \left( \frac{1}{n} - \frac{1}{N} \right).
\]

Proof. The UMLE of \( \Phi \) is \( s/n \), so by subtracting the proportion occupied by the \( s \) singletons detected in the sample the results follows.

With this approach, the use of singletons as statistics to estimate the proportion of undetected species appears very natural. In fact when \( N \to \infty \), the UMLE of \( \Phi \to s/n \) which is precisely Good’s estimator (1953). If there are \( r-s \) non-singletons in the sample, we can think of sampling from \( r-s+1 \) urns where each of the first \( r-s \) urns have been sampled at least once and thus completely covered. Taking \( s \) individuals from the last urn contributes \( s/n \) to the coverage of the sample, hence the proportion of singletons in the sample can be used to estimate and to test hypotheses on the proportion occupied by the last urn. The next section explains the methodology.

3. Hypothesis Test

From Corollary 2.2, with \( N \) big, a test on the size of \( \Pi \) can be approximated with a test on \( \Phi \). Let the random variable \( S \) be the number of singletons in the sample. To test the hypothesis \( H_0 : \Phi \geq \Phi_0 \) a first approximation is to calculate \( Pr(S \leq s|\Phi = \Phi_0) \) where \( S \) is binomial with parameters \( n \) and \( \Phi \).
Therefore, for a given $\alpha$ and $p_0$ the decision rule is:

Reject $H_0 : \Phi \geq p_0$ if

$$Pr(S \leq s \mid \Phi = p_0) = \sum_{i=0}^{s} \frac{n!}{i!(n-i)!} (1 - p_0)^{n-i} p_0^i < \alpha.$$  \hspace{1cm} (6)

Engen (1978) reported a "suspiciously small" probability when evaluating Eldridge statistics of fully inflected words in American newspaper English (Eldridge, 1911). In that work $n=43989$ words were sampled from which 2976 were singletons. Assuming $N$ is infinite, the UMLE of the proportion of uncovered species is equal to $2976/n=0.068$. Engen (1978, p. 67) wrote: "The initial (prior) probability that none of the words that did not appear in the sample should be observed is actually as small as $(1-0.068)^{43989} \approx 10^{-914}$. This number may at first glance look suspiciously small, but it is well known in probability theory that quite unlikely events do occur." The observed event is not unlikely at all. Nothing is wrong in the above calculation if we refer to the particular set of words that did not come up in the sample. Assuming $N$ is infinite, the UMLE of the proportion of uncovered species is equal to $2976/n=0.068$. To test the hypothesis $H_0 : \Phi \geq p_0$ we use (6). Calculating $Pr(S \leq 2976 \mid \Phi = p_0)$ gives that for $\alpha = .05$ we reject $H_0$ for values of $p_0$ bigger than 0.0697. That is, the coverage is at least 93.03%. For $\alpha = .01$, the coverage is at least 92.95%.
We can improve this approximation using the following argument. Assuming no prior knowledge of the composition of the target population, we can base our decision on having achieved a given coverage when some species are found repeatedly in the sample in a frequency beyond what we would expect if the proportion of undetected species is very large. In a sample of size one nothing can be stated on the proportion occupied by this species until we take a second sample whose outcome is a Bernoulli random variable with probability of success equal to the proportion of the previous species. In general, we can test on the proportion occupied by some particular species, $S_i$, based on the number of times we find it among the maximum possible $n-1$. The number of repeated samples of this particular species $S_i$ is then a binomial random variable with parameters $n-1$ and $p=\theta_i$.

According to Proposition 2.2, the singletons belong to a set of species with total proportion $\Phi$, with the characteristic that each species is represented by a single individual. This implies that their contribution to the coverage of the sample is negligible, $1/N$, and thus the coverage achieved is due mainly to the non-singleton species. Therefore, we can test on the proportion occupied by the non-singleton species found using the total number of “repeated” samples from this group. Each repetition corresponds to a “success” and thus “failures” corresponds to every singleton found.
To establish the independence and constant probability required in the assumptions for a binomial distribution we analyze the frequency of failures. If \( w \) non-singletons out of \( r \) species are found, the maximum number of failures possible is \( n-w \). The decision on the coverage achieved by the particular set of species chosen, the non-singletons, can be based on the number of failures, singletons, that have a constant probability \( \Phi \) of being detected at any one of \( n-w \) trials. Since \( w=r-s \), it follows that the distribution of \( S \), the number of singletons in the sample, is binomial with parameters \( n-r+s \) and \( p=\Phi \). The decision rule is then:

\[
\text{Reject } H_0 : \Phi \geq p_0 \text{ if } \Pr(S \leq s|\Phi = p_0) = \sum_{i=0}^{s} \frac{(n-r+s)!}{i!(n-r+s-i)!} (1 - p_0)^{n-r+s-i} p_0^i < \alpha. \tag{7}
\]

Applying the above rule on Eldridge’s data we have \( r=6001 \) so that we assume \( S \) has a binomial distribution with parameters \( n=40964 \) and \( p=\Phi \). For \( \alpha = .05 \) we reject \( H_0 \) for values of \( p_0 \) bigger than 0.0748 providing a coverage of at least 92.52%. For \( \alpha = .01 \), the coverage is at least 92.43%. With such a large sample no improvement, compared with the calculation using rule (6), is found. However, as suggested by the results from a moderate Monte Carlo study to be presented in the next section, when small sample sizes are considered the effectiveness of the method is noticeable.
4. MONTE CARLO SIMULATIONS

This section reports the results of a Monte Carlo study designed to test the hypothesis $H_0 : \Phi \geq p_0$ using the rejection rules (6) and (7). Six different populations were used with 15, 30, 50, 100, 200, and 500 classes. Figure 1 shows the distributions of classes for these populations. Observe that population 6 has 500 equally likely classes. The simulations performed on every population for every combination of $\alpha$ and $p_0$ were as follows. An initial sample of size 5 was taken. For fixed $\alpha$ and $p_0$, $H_0$ was tested using the rejection rule (7), and an additional randomly chosen individual is added until $H_0$ is rejected. This rule is referred as “Repeated Observations”. As a comparison, the rejection rule (6) is also evaluated. This method is referred as “Total”. During the simulations it is always possible to record the exact sample size at which the desired coverage was achieved. Thus, for each simulation, three sample sizes were recorded: the sample size when the intended coverage was achieved and the sample sizes when $H_0$ was rejected according to each method. Also, for every simulation, the true coverage is evaluated and it was noted whether the intended coverage was reached or not.

We considered $\alpha$ values of 0.05 and 0.01, and $p_0$ values of 0.05, 0.1, 0.2 and 0.3 for intended coverages of 95, 90, 80 and 70 percent respectively. This sampling procedure was repeated 10000 times on each population. Tables
1-4 summarize the results from these simulations. Given in the tables are average and standard deviation values of the coverage and the sample size at which the coverage was achieved. Also given is the proportion of the 10000 repetitions for which the desired coverage was achieved. This proportion is referred as “proportion of success”.

Evaluating the efficacy of a given stopping rule is a difficult task. Relevant measures are the coverage achieved, the frequency in which an intended coverage is achieved, frequency of success, and the sample size required. The task is especially difficult since it is possible to obtain a high proportion of successes with an insensitive stopping rule. As a rule of thumb, a good procedure should have a proportion of success slightly larger than 1-\(\alpha\), small sample size and a mean coverage at least as the desired one.

Examination of the tables shows the mean coverage of the 10000 simulations for each population is always greater than the intended one with both methods. However, the frequency of success is always closer to 1-\(\alpha\) for the “Repeated” method than for “Total”. Also, larger sample sizes are required for the “Repeated” procedure than for “Total”, but this is due to the fact that the latter method stops early, hence it has a smaller proportion of successes.
Fig. 1 Class Distribution

Population 1: 15 classes

Population 2: 30 classes

Population 3: 50 classes

Population 4: 100 classes

Population 5: 200 classes

* Horizontal labels are upper limits for each interval.
Table 1. Method "Repeated observations"

<table>
<thead>
<tr>
<th>Pop. 1</th>
<th>Pop. 2</th>
<th>Pop. 3</th>
<th>Pop. 4</th>
<th>Pop. 5</th>
<th>Pop. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.863 ± 0.078</td>
<td>0.847 ± 0.077</td>
<td>0.838 ± 0.063</td>
<td>0.827 ± 0.046</td>
<td>0.812 ± 0.034</td>
</tr>
<tr>
<td></td>
<td>32.76 ± 7.3</td>
<td>43.54 ± 9.8</td>
<td>77.36 ± 11.1</td>
<td>150.33 ± 15.5</td>
<td>276.41 ± 21.8</td>
</tr>
<tr>
<td>0.8</td>
<td>0.911 ± 0.062</td>
<td>0.898 ± 0.053</td>
<td>0.894 ± 0.045</td>
<td>0.884 ± 0.033</td>
<td>0.874 ± 0.025</td>
</tr>
<tr>
<td></td>
<td>46.03 ± 8.9</td>
<td>63.89 ± 13.2</td>
<td>106.94 ± 14.6</td>
<td>203.91 ± 19.1</td>
<td>381.66 ± 26.9</td>
</tr>
<tr>
<td>0.9</td>
<td>0.961 ± 0.037</td>
<td>0.949 ± 0.030</td>
<td>0.948 ± 0.025</td>
<td>0.942 ± 0.020</td>
<td>0.936 ± 0.015</td>
</tr>
<tr>
<td></td>
<td>71.65 ± 15.8</td>
<td>116.3 ± 23.5</td>
<td>171.66 ± 22.5</td>
<td>322.36 ± 29.6</td>
<td>605.66 ± 42.2</td>
</tr>
<tr>
<td>0.95</td>
<td>0.983 ± 0.021</td>
<td>0.973 ± 0.017</td>
<td>0.975 ± 0.015</td>
<td>0.971 ± 0.012</td>
<td>0.967 ± 0.009</td>
</tr>
<tr>
<td></td>
<td>111.36 ± 20.3</td>
<td>198.87 ± 39.4</td>
<td>264.63 ± 38.1</td>
<td>485.38 ± 47.9</td>
<td>912.64 ± 66.9</td>
</tr>
<tr>
<td>Mean coverage ± std</td>
<td>Mean sample size ± std</td>
<td>Proportion of success</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Method "Total"

<table>
<thead>
<tr>
<th>Pop. 1</th>
<th>Pop. 2</th>
<th>Pop. 3</th>
<th>Pop. 4</th>
<th>Pop. 5</th>
<th>Pop. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.824 ± 0.1</td>
<td>0.808 ± 0.097</td>
<td>0.782 ± 0.081</td>
<td>0.766 ± 0.06</td>
<td>0.749 ± 0.044</td>
</tr>
<tr>
<td></td>
<td>26.27 ± 7.5</td>
<td>34.56 ± 9.3</td>
<td>60.15 ± 11.1</td>
<td>116.29 ± 15.2</td>
<td>213 ± 21.5</td>
</tr>
<tr>
<td>0.8</td>
<td>0.89 ± 0.071</td>
<td>0.877 ± 0.065</td>
<td>0.863 ± 0.058</td>
<td>0.846 ± 0.043</td>
<td>0.834 ± 0.032</td>
</tr>
<tr>
<td></td>
<td>39.09 ± 9.0</td>
<td>53.28 ± 12.9</td>
<td>87.77 ± 14.5</td>
<td>165.59 ± 18.7</td>
<td>308.79 ± 26.7</td>
</tr>
<tr>
<td>0.9</td>
<td>0.952 ± 0.042</td>
<td>0.941 ± 0.035</td>
<td>0.936 ± 0.030</td>
<td>0.927 ± 0.025</td>
<td>0.919 ± 0.018</td>
</tr>
<tr>
<td></td>
<td>65.73 ± 12.2</td>
<td>102.86 ± 23.4</td>
<td>149.69 ± 22.1</td>
<td>278.95 ± 28.9</td>
<td>522.49 ± 41.2</td>
</tr>
<tr>
<td>0.95</td>
<td>0.979 ± 0.026</td>
<td>0.971 ± 0.019</td>
<td>0.970 ± 0.017</td>
<td>0.965 ± 0.014</td>
<td>0.961 ± 0.010</td>
</tr>
<tr>
<td></td>
<td>103.53 ± 20.6</td>
<td>183.8 ± 40.1</td>
<td>240.33 ± 36.5</td>
<td>438.98 ± 47.3</td>
<td>823.43 ± 65.7</td>
</tr>
<tr>
<td>Mean coverage ± std</td>
<td>Mean sample size ± std</td>
<td>Proportion of success</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 3. Method "Repeated Observations"

**Alpha = 0.01**

<table>
<thead>
<tr>
<th></th>
<th>Pop. 1</th>
<th>Pop. 2</th>
<th>Pop. 3</th>
<th>Pop. 4</th>
<th>Pop. 5</th>
<th>Pop. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>0.895 ± 0.067</td>
<td>0.877 ± 0.061</td>
<td>0.864 ± 0.049</td>
<td>0.843 ± 0.041</td>
<td>0.826 ± 0.032</td>
<td>0.82 ± 0.022</td>
</tr>
<tr>
<td></td>
<td>39.74 ± 7.2</td>
<td>52.27 ± 9.8</td>
<td>87.63 ± 10.7</td>
<td>162.61 ± 15.5</td>
<td>294.7 ± 21.8</td>
<td>858.79 ± 26.2</td>
</tr>
<tr>
<td>0.8</td>
<td>0.937 ± 0.045</td>
<td>0.918 ± 0.038</td>
<td>0.91 ± 0.037</td>
<td>0.896 ± 0.031</td>
<td>0.884 ± 0.023</td>
<td>0.882 ± 0.017</td>
</tr>
<tr>
<td></td>
<td>56.03 ± 8.5</td>
<td>76.39 ± 13.5</td>
<td>119.85 ± 14.2</td>
<td>220.78 ± 19.7</td>
<td>405.01 ± 27.7</td>
<td>1072.58 ± 28.9</td>
</tr>
<tr>
<td>0.9</td>
<td>0.972 ± 0.027</td>
<td>0.96 ± 0.02</td>
<td>0.958 ± 0.022</td>
<td>0.95 ± 0.018</td>
<td>0.942 ± 0.014</td>
<td>0.942 ± 0.012</td>
</tr>
<tr>
<td></td>
<td>69.84 ± 30.7</td>
<td>139.76 ± 23.6</td>
<td>192.92 ± 24.2</td>
<td>350.3 ± 31.2</td>
<td>642.82 ± 43.2</td>
<td>1429.5 ± 37.2</td>
</tr>
<tr>
<td>0.95</td>
<td>0.989 ± 0.014</td>
<td>0.98 ± 0.012</td>
<td>0.98 ± 0.012</td>
<td>0.975 ± 0.01</td>
<td>0.971 ± 0.008</td>
<td>0.9722 ± 0.008</td>
</tr>
<tr>
<td></td>
<td>136.62 ± 22.9</td>
<td>240.08 ± 42.6</td>
<td>302.26 ± 41.4</td>
<td>529.63 ± 50.3</td>
<td>972.54 ± 71.3</td>
<td>1793.5 ± 51.7</td>
</tr>
<tr>
<td></td>
<td>0.9822</td>
<td>0.9777</td>
<td>0.9752</td>
<td>0.977</td>
<td>0.9833</td>
<td>0.9938</td>
</tr>
</tbody>
</table>

Mean coverage ± std  
Mean sample size ± std  
Proportion of success

### Table 4. Method "Total"

**Alpha = 0.01**

<table>
<thead>
<tr>
<th></th>
<th>Pop. 1</th>
<th>Pop. 2</th>
<th>Pop. 3</th>
<th>Pop. 4</th>
<th>Pop. 5</th>
<th>Pop. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>0.869 ± 0.075</td>
<td>0.847 ± 0.076</td>
<td>0.82 ± 0.062</td>
<td>0.789 ± 0.053</td>
<td>0.767 ± 0.040</td>
<td>0.741 ± 0.027</td>
</tr>
<tr>
<td></td>
<td>33.02 ± 7.3</td>
<td>42.85 ± 9.8</td>
<td>69.54 ± 10.5</td>
<td>127.21 ± 15.2</td>
<td>228.75 ± 21.4</td>
<td>676.46 ± 28.2</td>
</tr>
<tr>
<td>0.8</td>
<td>0.924 ± 0.052</td>
<td>0.9 ± 0.047</td>
<td>0.883 ± 0.049</td>
<td>0.863 ± 0.038</td>
<td>0.848 ± 0.029</td>
<td>0.830 ± 0.022</td>
</tr>
<tr>
<td></td>
<td>48.69 ± 8.6</td>
<td>64.24 ± 13.1</td>
<td>99.67 ± 14.2</td>
<td>180.99 ± 19.1</td>
<td>330.46 ± 27.0</td>
<td>889.36 ± 30.0</td>
</tr>
<tr>
<td>0.9</td>
<td>0.961 ± 0.032</td>
<td>0.954 ± 0.024</td>
<td>0.948 ± 0.027</td>
<td>0.937 ± 0.021</td>
<td>0.927 ± 0.017</td>
<td>0.919 ± 0.014</td>
</tr>
<tr>
<td></td>
<td>77.47 ± 11.1</td>
<td>125.04 ± 23.4</td>
<td>169.72 ± 23.3</td>
<td>305.37 ± 30.2</td>
<td>557.26 ± 42.3</td>
<td>1260.3 ± 38.2</td>
</tr>
<tr>
<td>0.95</td>
<td>0.987 ± 0.016</td>
<td>0.979 ± 0.013</td>
<td>0.976 ± 0.014</td>
<td>0.970 ± 0.012</td>
<td>0.965 ± 0.019</td>
<td>0.962 ± 0.009</td>
</tr>
<tr>
<td></td>
<td>128.38 ± 22.7</td>
<td>224.4 ± 42.2</td>
<td>275.95 ± 40.3</td>
<td>481.94 ± 49.4</td>
<td>880.87 ± 69.7</td>
<td>1637.4 ± 49.9</td>
</tr>
<tr>
<td></td>
<td>0.9737</td>
<td>0.965</td>
<td>0.9515</td>
<td>0.9351</td>
<td>0.9317</td>
<td>0.9005</td>
</tr>
</tbody>
</table>

Mean coverage ± std  
Mean sample size ± std  
Proportion of success
REFERENCES


