A CONDITIONALLY UNBIASED ESTIMATOR FOR THE EQUAL-CATCHABILITY MODEL

I. B. J. Goudie* and J. Ashbridge
School of Mathematics and Statistics,
University of St Andrews, St Andrews, Scotland

ABSTRACT

In this paper we reconsider an estimator of population size previously advocated for use when sampling from a population subdivided into different types. We show that it may be usefully adopted in the simple equal-catchability model used in mark-recapture. Unlike the commonly used maximum likelihood estimator, this conditionally unbiased estimator is always finite-valued. Except in situations in which the data contain little relevant information, its performance, in terms of bias and precision, is seen to be at least as good as that of the maximum likelihood estimator. Two estimators of the standard deviation of the conditionally unbiased estimator are considered.

Key words: capture-recapture sampling; C-number; factorial series distribution; population size.

*Correspondence: I. B. J. Goudie, School of Mathematics and Statistics, University of St Andrews, St Andrews, KY16 9SS, Scotland; E-mail: ig@st-andrews.ac.uk.

1. INTRODUCTION

The standard equal-catchability mark-recapture model known as $M_0$ (Otis et al. (1)) is too simplistic to serve as a complete model in most problems of population size estimation. It is, however, often used to provide benchmark results against which those from more realistic models may be judged. It can also have a role in situations in which division of the population
into more homogeneous strata is deemed viable.

Under the model, a sample is obtained on each of $t$ occasions from a population of unknown size $N > 0$, in such a way that, independently of other individuals and of its previous capture history, each member of the population is captured on any given occasion with probability $p$. The probability $p$ is thus constant over time, common to all individuals in the population and is unaffected by whether or not the individual has previously been captured. On the first occasion on which a member of the population is captured it is tagged to distinguish it from other members of the population.

It is recognised that, in practical studies, it is relatively rare for complete populations to satisfy the strong assumptions of model $M_0$. Other models within the class discussed by Otis et al. (1) are likely to be more appropriate. Model $M_0$ can, nonetheless, have a role to play when time-dependence, behavioural effects and heterogeneity of capture probabilities are manifest only in weak forms. For instance, Lee and Chao (2) commend the use of $M_0$ in preference to the heterogeneity model, model $M_h$, in cases of weak heterogeneity where the coefficient of variation of the capture probabilities does not exceed 0.4.

The notion of dividing a population into more homogeneous strata has been used in post-enumeration surveys after recent U.S. and U.K. censuses (Brown et al. (3)). These surveys, designed to address the problem of under-enumeration in the census count, can be viewed as the second sample in a two-stage process. Although this indicates that the concept of sub-division can have important applications, it should be noted, however, that in these examples weaker assumptions sufficed. More precisely, the underlying model for these two-sample procedures was model $M_t$ of the Otis class, which permits time variation of detection probabilities: it was not appropriate to assume that the detection probability in the post-enumeration survey equalled that for the census itself.

Maximum likelihood estimates obtained from model $M_0$ are often cited for comparative purposes. These estimates, which can be evaluated using standard mark-recapture software, such as program MARK, can demonstrate, for instance, how much a population size estimate
is influenced by the heterogeneity of the capture probabilities between individuals. An example of this type is the comparative assessment of estimators for model $M_h$ given by Ashbridge and Goudie (4).

For estimating the population size $N$ under model $M_0$, we advocate in this paper the use of an alternative estimator, defined in section 3 below. Conditional on the event that the total number $Z$ of captures equals $z$, this alternative estimator is unbiased provided that $N \leq z$. In section 2 we show that, under the same conditioning, the distribution of the sufficient statistics belongs to the family of factorial series distributions, given by Berg (5). This conditional distribution arose as the unconditional distribution of the data in the sampling problem considered by Charalambides (6), and he proposed the estimator that we now discuss as an unbiased estimator for that more restricted context. Estimation of the standard deviation of the conditionally unbiased estimator is considered in section 6, and the relationship of this estimator to those of Pathak (7) and Chapman (8) is demonstrated in section 4. In section 5 the performance of the conditionally unbiased estimator is seen to compare favourably with that of the maximum likelihood estimator.

2. DISTRIBUTIONAL RESULTS

Notionally numbering the members of the population, define the $N \times t$ matrix $Y = (y_{jk})$ by $y_{jk} = 1$ if individual $j$ is captured in sample $k$, and $y_{jk} = 0$ otherwise. By independence,

$$pr(Y = y) = \prod_{j=1}^{N} \prod_{k=1}^{t} p^{y_{jk}} (1 - p)^{1-y_{jk}} = p^z (1 - p)^{Nt-z},$$

where the random variable $Z = y_{11} + \ldots + y_{Nt}$, which gives the total number of captures made over the $t$ samples, is clearly binomial with parameters $Nt$ and $p$. Hence, given $Z = z$, the conditional probability that any particular $z$ of the $Nt$ elements of the matrix $Y$ equal unity is given by the constant $z!/(Nt)_z$, where $(b)_z = b(b-1)\ldots(b-z+1)$ and $(b)_0 = 1$. These $z$ elements may thus be regarded as a random sample of the $Nt$ elements, selected without replacement. As shown below, given $Z = z$, the conditional probability function of
the number $X$ of different individuals observed is

$$p(x|z) = (N)_x C(z, x, t)/(Nt)_z \quad x = \min(1, z), \ldots, \min(N, z) \quad (1)$$

where $C(z, x, t)$ is a $C$-number. The $C$-numbers, which are discussed by Charalambides and Singh (9), are defined by $C(0, 0, t) = 1$, and by

$$C(z, x, t) = \frac{1}{x!} \sum_{v=0}^{x} (-1)^v \binom{x}{v} (t(x - v))_z = \frac{1}{x!} [\Delta^x (Nt)_z]_{N=0}$$

for positive integers $z$ and $x$, with $\Delta$ denoting the forward finite difference operator. It follows that, for $z = 0, \ldots, Nt$ and $x = \min(1, z), \ldots, \min(N, z)$, the joint probability function for $Z$ and $X$ is

$$p(z, x) = (N)_x C(z, x, t)p^x(1 - p)^{Nt - x}/z! \quad (2)$$

Note also that the distribution (1) has the form of a factorial series distribution, as defined by Berg (5), with series function $(Nt)_z$.

One method of obtaining the conditional distribution (1) is to note that the problem of sampling without replacement from the elements of $Y$ is equivalent to the scenario, considered by Charalambides (6), of sampling from a population consisting of $t$ individuals of each of $N$ types, with all $Nt$ individuals being equiprobable. He showed that, when a sample of known size $z$ is chosen without replacement from the population, the conditional probability that $x$ different types are observed is given by (1). Another probabilistically equivalent situation arose in the restricted occupancy problem considered by Fang (10), in which $z$ indistinguishable balls are randomly assigned to $N$ distinguishable urns, each of which is subdivided into $t$ distinguishable cells.

Alternatively the result (1) may be obtained directly by noting that the conditional probability of observing $x$ different individuals, given a total of $z$ captures, is the conditional probability that there are exactly $x$ rows of $Y$ having one or more of the $z$ non-zero elements. A total of $(Nt)_z/z!$ different possible samples of size $z$ can be drawn from the elements of $Y$, and, by
the inclusion-exclusion principle (Johnson et al. (11), p407), the number of these that assign
the \( z \) non-zero entries to \( x \) particular rows of \( Y \), so that each of these \( x \) rows has a non-zero
entry, is given by \((x!/z!)C(z, x, t)\). Since the \( x \) rows can be chosen in \((N)_x/x!\) different ways,
the conditional probability function (1) is obtained.

One way to record the data is in terms of capture histories. If \( \omega \) denotes a subset of the
integers \( \{1, 2, \ldots, t\} \), let \( u_\omega \) equal the number of members of the population which are caught
in all the samples at times \( k \) for which \( k \in \omega \), but are not caught in any other samples. The
probability function of \( \{u_\omega\} \) is multinomial, and the results of Darroch (12) imply that

\[
\text{pr}\{\{u_\omega\}\} = \left(\frac{N}{x}\right)^z p^z (1 - p)^{Nt - z} / \prod_{\omega} u_\omega!.
\]

Inspection of this probability function shows that \((X, Z)\) is sufficient for \((N, p)\).

Suppose that \( n_i \) denotes the number of individuals caught in sample \( i \) \((i = 1, \ldots, t)\). Under
model \( M_0 \), which is our primary concern, \( n_1, \ldots, n_t \) are independent binomial random
variables, each with parameters \( N \) and \( p \). The alternative model \( M_f \), in which the vector
\( n = (n_1, \ldots, n_t) \) of sample sizes is not random but pre-chosen, was considered by Berg
(5). For \( N > \max(n_i) \), he showed that, under model \( M_f \), the random variable \( X \) has
a factorial series distribution with series function \( A(N, n) = (N)_{n_1} \cdots (N)_{n_t} / (n_1! \cdots n_t!) \).

Explicitly, the probability function of \( X \), which we write as a conditional distribution prior
to reconsidering it below in the context of model \( M_0 \), is

\[
p(x|n) = (N)_x a(x, n) / A(N, n) \quad x = \max(n_i), \ldots, \min(N, z),
\]

where

\[
a(x, n) = \frac{1}{x!} [\Delta^x A(N, n)]_{N=0} = \frac{1}{x!} \sum_{v=0}^{x} (-1)^v \binom{x}{v} \prod_{i=1}^{t} \binom{x - v}{n_i}
\]

Hence, under model \( M_0 \), the joint probability function of \( X \) and \( n \), for \( n_i = 0, \ldots, N \)
\((i = 1, \ldots, t)\) and \( x = \max(n_i), \ldots, \min(N, z) \), is given by

\[
p(x, n) = (N)_x a(x, n)p^x (1 - p)^{Nt - z}.
\]
3. A CONDITIONALLY UNBIASED ESTIMATOR

Consider the estimator $\tilde{N}_U(z, x)$ defined by

$$
\tilde{N}_U(z, x) = \begin{cases} 
  x + \{C(z, x-1, t)/C(z, x, t)\} & z = 1, 2, \ldots; x = 1, \ldots, z; \\
  0 & z = 0, x = 0.
\end{cases} \tag{5}
$$

Since the conditional distribution (1) is a factorial series distribution, the results of Berg (5) show that, provided $N \leq z$, this is the unique unbiased estimator of $N$ with respect to the conditional distribution of the data given $Z = z$. This estimator is identical to one proposed by Charalambides (6) for the problem, discussed above, of sampling from a population consisting of $Nt$ equiprobable individuals, classified into $N$ equiprobable types. Changed into our notation, he expressed the estimator in the form

$$
\tilde{N}_U(z, x) = (z/t) + [C(z+1, x, t)/\{tC(z, x, t)\}] \quad z = 1, 2, \ldots; x = 1, \ldots, z. \tag{6}
$$

The equivalence of the two forms is easily seen using the standard triangular recurrence relation for the $C$-numbers (Charalambides and Singh (9)), which states that, for positive integers $z$ and $x$,

$$
C(z, x, t) = (tx - z + 1)C(z - 1, x, t) + tC(z - 1, x - 1, t), \tag{7}
$$

where $C(0, 0, t) = 1$, $C(z, 0, t) = 0$ for $z = 1, 2 \ldots$, and $C(0, x, t) = 0$ for $x = 1, 2 \ldots$.

The major difference between the sampling problem discussed by Charalambides (6) and the inference problem for model $M_0$ is that in the former the sample size $z$ is pre-chosen. As far as we are aware, the estimator $\tilde{N}_U$ has not previously been employed in situations where $Z$ is a random variable, as is the case for model $M_0$. Moreover, for model $M_0$, the probability that the condition $N \leq z$ holds is less than unity, and hence the unbiasedness of $\tilde{N}_U$ with respect to the conditional distribution given $Z$ does not imply that $\tilde{N}_U$ is unconditionally unbiased.
As the true population size $N$ is integer valued, it is appropriate to round the value of $\tilde{N}_U$ and define the conditionally unbiased estimator by $\hat{N}_U(z, x) = [\tilde{N}_U(z, x) + 0.5]$, where the square brackets denote the integer part.

The conditionally unbiased estimator has a theoretical advantage over the maximum likelihood estimator in being a closed form estimator. With modern computing facilities, however, this is much less important than it would once have been. Any such advantage is further dissipated by the relative difficulty of computing the $C$-numbers, except when their arguments are small. The $C$-numbers grow rapidly in size, and recursive ways of computing the joint probability function (2) and the conditionally unbiased estimator (5) are helpful. It follows from the triangular recurrence relation (7) that

$$p(z, x) = p\{(tx - z + 1)p(z - 1, x) + (N - x + 1) t p(z - 1, x - 1)\}/\{(1 - p)z\}.$$ 

Moreover a recursive relationship, again of triangular form, can be obtained for the estimator $\tilde{N}_U$, by generalising Property 5 of Berg (14). Explicitly, for $x = 2, 3, \ldots$, an expression for $\tilde{N}_U(z, x)$ is given by the relationship

$$\{\tilde{N}_U(z, x) - x\}/\{\tilde{N}_U(z - 1, x) - x\} = \{t\tilde{N}_U(z - 1, x - 1) - z + 1\}/\{t\tilde{N}_U(z - 1, x) - z + 1\},$$

which can be verified by using equations (5) and (6) to express the left and right hand sides respectively as ratios of $C$-numbers. Since $C(z, 0, t) = 0$ for $z = 1, 2 \ldots$, definition (5) provides one boundary condition, namely $\tilde{N}_U(z, 1) = 1$ for $z = 1, 2 \ldots$. To obtain a second, note that, as $C(z, x, t) = 0$ for $z < x$, the triangular recurrence relation (7) implies that $C(z, z, t) = t^z$ for $z = 1, 2 \ldots$. Then use (5) and (6) to show that $\tilde{N}_U(z, z)$ satisfies the recurrence relation

$$\tilde{N}_U(z, z) - \tilde{N}_U(z - 1, z - 1) = \{z(t - 1) + 1\}/t \quad z = 1, 2 \ldots ,$$

with $\tilde{N}_U(0, 0) = 0$. This then yields the second boundary condition

$$\tilde{N}_U(z, z) = z(tz - z + t + 1)/(2t) \quad z = 0, 1 \ldots .$$
This equation also shows that, unlike the maximum likelihood estimator, the estimator $\tilde{N}_U$ remains finite when $Z = X$.

It is of interest to note the connection between the estimator $\tilde{N}_U$ and the estimator of Harris (13) for the classical occupancy problem. For a population of fixed size $N$, envisage a sequence of experiments in which the number $t$ of sampling occasions increases but the capture probability $p$ decreases in such a way that $E(Z) = Ntp$ remains constant. The matrix $Y$, indicating which animals are caught in each sample, becomes increasingly sparse. The sampling occasions on which an animal is seen resemble those in the classical occupancy problem in the sense that just a single observation occurs which will be of one of $N$ types.

More precisely, Charalambides (6) noted that, as $t \to \infty$, $C(z, x, t)/t^z$ tends to the Stirling number $S(z, x)$ of the second kind, defined by

$$S(z, x) = \frac{1}{x!} \sum_{u=1}^{x} (-1)^{x-u} \binom{x}{u} u^z,$$

and hence, as $t \to \infty$, the conditional distribution (1) for $X$ given $Z = z$ tends to the classical occupancy distribution. A further use of the limit of $C(z, x, t)/t^z$ shows that, when $x > 0$, the estimator $\tilde{N}_U$ tends to the estimator $x + \{S(z, x - 1)/S(z, x)\}$ as $t \to \infty$. Application of the triangular recurrence relation $S(z + 1, x) = xS(z, x) + S(z, x - 1)$ for the Stirling numbers of the second kind implies that this limiting estimator equals $S(z + 1, x)/S(z, x)$. Apart from the difference that $Z$ here is a random variable rather than a pre-chosen constant, this is the estimator of Harris (13) for the classical occupancy problem.

4. RELATIONSHIP TO PATHAK’S ESTIMATOR

For the fixed sample sizes model $M_f$, Pathak (7) proposed an estimator which may be written as $N^*(x, n) = x + \{a(x - 1, n)/a(x, n)\}$, where $a(x, n)$ is defined by (3). Observing that $n_1 + \ldots + n_t = z$, denote the set of integers $N$ such that $N \in [\max(n_i), z]$ by $\Theta$. Pathak proved that, for $N \in \Theta$, the estimator $N^*$ is the minimum variance unbiased estimator of $N$ under model $M_f$. Reinterpreted in the context of model $M_0$, this result implies that, for
\( N \in \Theta \), the estimator \( N^* \) is unbiased under the conditional distribution of the data given \( n \).

Now define \( n_{z,x} \equiv \{ n : n_1 + \ldots + n_t = z, \; n_i \leq x \; (i = 1, \ldots, t) \} \). From (2) and (4), the conditional probability function of \( n \) given \( X \) and \( Z \) is

\[
p(n|x,z) = \frac{p(x,n)}{p(x,z)} = \frac{a(x,n)}{C(z,x,t)} \; \; \; n \in n_{z,x}.
\]

Use of the Rao-Blackwell Theorem suggests the estimator

\[
E(N^*|X,Z) = x + \left\{ \frac{z!}{C(z,x,t)} \sum a(x-1,n) \right\},
\]

where the summation is over \( n \in n_{z,x} \). This expression is equivalent to

\[
x + \frac{z!}{(x-1)! \; C(z,x,t)} \sum_{v=0}^{x-1} (-1)^v \binom{x-1}{v} \sum_{n \in n_{z,x}} \prod_{i=1}^t \binom{x-1-v}{n_i}
\]

in which the product of combinatorial terms gives the number of ways of achieving a total of \( z \) captures, involving a particular \( x-1-v \) animals, in such a way that \( n_i \) of these captures occur in sample \( i \). Summing this product over \( n_{z,x} \) gives (c.f. Feller (15), p58) the number of ways of achieving \( z \) captures from a possible total of \( t(x-1-v) \), which is \( \{ t(x-1-v) \}_{z/z!} \). It is thus readily confirmed that the estimator \( E(N^*|X,Z) \) is simply \( \tilde{N}_U \).

In general, \( x!a(x,n) \) gives the number of ways of drawing the samples of sizes \( n_1, \ldots, n_t \) in such a way that a particular \( x \) animals are all seen at least once. In the case \( t = 2 \), this equals the number of ways in which the \( x \) animals may be divided into a group of size \( z-x = n_1 + n_2 - x \) seen in both samples, a group of size \( n_1 - (z-x) = x - n_2 \) seen only in the first sample and a group of size \( x - n_1 \) seen only in the second. So, in this case, \( x!a(x,n) \) is simply the trinomial coefficient \( x! \{ (x-n_1)!(x-n_2)!(n_1+n_2-x)! \} \). It follows that, when \( t = 2 \), Pathak’s estimator can be written as

\[
N^* = x + \frac{(x-n_1)(x-n_2)}{n_1 + n_2 - x + 1} = \frac{n_1n_2 + x}{n_1 + n_2 - x + 1}.
\]

9
So, as noted by Berg (16), this estimator coincides with the variant of the Petersen estimator proposed by Chapman (8).

When \( t = 2 \), each of the \((x - n_1) + (x - n_2) = 2x - z\) animals that is seen exactly once is captured in the first sample with probability 0.5, independently of other animals. Hence the conditional distribution for \( X - n_1 \), given \( X = x \) and \( Z = z \), is binomial. The estimator \( \tilde{N}_U = (z^2 + z + 2x)/\{4(z - x + 1)\} \) can then be obtained as the expectation of \( N^* \) with respect to this distribution. Alternatively, to verify that this result agrees with (5), sum the above trinomial coefficients over all possible values of \( n_1 \), for fixed \( x \) and \( z \). This indicates that the number of ways in which a total of \( z \) captures can arise from the \( x \) animals, with each being caught at least once, is given by \( x! \cdot 2^{2x-z}/\{(z - x)! \cdot (2x - z)!\} \). The discussion in section 2 shows that this equals \((x!/z!)\cdot C(z, x, 2)\), thus yielding an expression for the required \( C \)-number. It follows that, when \( t = 2 \), the resulting estimator \( \tilde{N}_U \) may be viewed as the improvement to Chapman’s estimator provided by the Rao-Blackwell Theorem.

5. COMPARISON WITH THE MAXIMUM LIKELIHOOD ESTIMATOR

Here we compare the performance of the conditionally unbiased estimator \( \hat{N}_U \) with that of the maximum likelihood estimator \( \hat{N}_M \). Results are presented in Table 1 for situations in which \( t = 5, 10, 15 \) or 20 samples are taken from populations of sizes \( N = 50 \) and 100, when the capture probability \( p = 0.05, 0.1 \) or 0.2. The performance of each estimator is indicated by the statistics \( m \) and \( s \) which are respectively the mean and standard deviation conditional on the event \( E = \{Z > X\} \). Although this conditioning is not required for the estimator \( \hat{N}_U \), which is finite with probability one, it is necessary for the estimator \( \hat{N}_M \) if the statistics \( m \) and \( s \) are to be finite. The conditioning thus enables a precise comparison to be made. All the moments of estimators that are given in this paper are conditional on the event \( E \), but, for brevity and to avoid confusion with the more important conditioning on \( Z \), we will omit the word ‘conditional’ on each occasion. The conditional distributions, given the event \( E \), of the estimators can readily be obtained using the joint probability function (2) of \( Z \) and \( X \).
Table 1 indicates that the conditionally unbiased estimator $\hat{N}_U$ should be preferred to the maximum likelihood estimator $\hat{N}_M$ for the three larger sample sizes corresponding to $t = 10$, 15 or 20. For all such cells in the table, $\hat{N}_U$ is less biased, and only in two of these eighteen cases does $\hat{N}_M$ have a somewhat smaller standard deviation. In fact, to three significant figures, $\hat{N}_U$ is unbiased for those cells for which $t \geq 10$ and $p \geq 0.1$.

For the remaining case $t = 5$, the situation is less clear-cut. The cells corresponding to this case indicate that $\hat{N}_U$ has the smaller bias for $p \geq 0.1$, and consistently has the smaller standard deviation, but $\hat{N}_M$ is preferable in terms of bias when $p = 0.05$ and $N = 50$. For $t = 5$, $p = 0.05$ and $N = 100$, the biases of the estimators have opposite signs. Indeed it is worth noting that $\hat{N}_M$ shows appreciable positive bias in several cells corresponding to smaller values of $p$. For a different population size model, Goudie et al. (17) noted the danger, particularly if judgment is made on the basis of mean square error, of favouring estimators which display negative bias simply because it is often accompanied by greater precision. Nevertheless, in a conservation context, for instance, a positive bias would be undesirable, since over-estimation of population size is liable to have more serious consequences than under-estimation. Taking a more theoretical stand-point, observe also that, although Table 1 suggests that the bias of $\hat{N}_U$ decreases monotonically as the sample size $t$ increases, it is evident that $\hat{N}_M$ does not have this property.

The balance of the above argument indicates clearly that the conditionally unbiased estimator has a useful role to play. Moreover the arguments for using it have been somewhat understated. The conditioning on the event $E = \{Z > X\}$, which was required in order to make a precise comparison, serves to exacerbate the negative bias of $\hat{N}_U$ in those situations where bias is present. Furthermore the fact that $\hat{N}_M$ can be infinite with non-negligible probability is a further shortcoming of $\hat{N}_M$ as an estimator. This probability, given by $1 - P(E)$, is less than 0.0003 for all but four of the situations considered in Table 1. When $t = 5$ and $p = 0.05$, however, it equals 0.1018 for $N = 100$ and is as large as 0.3190 for $N = 50$. 

11
One estimator of the standard deviation of the conditionally unbiased estimator $\hat{N}_U$ can be obtained by taking the square root $\hat{S}_U(z, x)$ of the variance estimator $\hat{V}_U(z, x)$, given by

$$\hat{V}_U(z, x) = \{\tilde{N}_U(z, x) - x\}\{\tilde{N}_U(z, x) - \tilde{N}_U(z, x - 1)\}.$$ 

The motivation for this estimator comes from the results of Berg (5) on factorial series distributions. Application of his results shows that, for fixed $z \geq N$, $\hat{V}_U(z, x)$ is the unique unbiased estimator of the variance of the raw form $\tilde{N}_U$ of the conditionally unbiased estimator prior to rounding. Under the joint distribution of $Z$ and $X$, the condition that $z \geq N$ is not satisfied with probability one, and hence, unconditionally, the bias of $\hat{V}_U(z, x)$ can deviate from zero. The estimator $\hat{S}_U(z, x)$ is also biased: the extent of the bias, conditional on the event $E = \{Z > X\}$, is examined in Table 2.

Another estimator $\hat{S}_B(z, x)$ of the standard deviation of $\tilde{N}_U$ can be found using a parametric bootstrap. If the estimate of population size is $\tilde{N}_U(z, x)$, realisations are simulated from a population of size $\tilde{N}_U(z, x)$, with each realisation comprising $t$ samples, in which the constant capture probability is given by $z/\{t\tilde{N}_U(z, x)\}$. When it is desired to condition on the event $E$, realisations for which $Z = X$ can be discarded. For each simulated realisation of the sampling process, the value of the conditionally unbiased estimator is calculated. The standard deviation $\hat{S}_B(z, x)$ of the estimates from the resamples then provides an estimated standard deviation for the conditionally unbiased estimator.

In Table 2 we compare these estimators of the standard deviation in each of the situations previously considered in Table 1. The expected value of $\hat{S}_U(z, x)$ is given by $s_U$, while $s$ again denotes the true standard deviation of the conditionally unbiased estimator $\tilde{N}_U$. The performance of the parametric bootstrap was assessed by randomly selecting 500 values of $(Z, X)$ under its conditional distribution given $E$. For each $(Z, X)$, 500 realisations of the sampling process were generated in the manner described above. The value $\hat{s}_B$ shown in the table is the mean of the resulting 500 estimates of the standard deviation.

Table 2 indicates that the performance of both of these estimators of the standard deviation
of $\hat{N}_U$ is usually satisfactory. The bootstrap estimate $\hat{S}_B(z, x)$ is biased positively more often than negatively, and, when quantifying the precision of a point estimate, it is safer to err on the side of caution. A further advantage of using the resampling method to estimate the standard deviation $s$ occurs when an interval estimate of $N$ is desired, since this approach can readily be used to generate bootstrap confidence intervals directly (see, for example, Buckland and Garthwaite (18)).

In all but one of the cases shown the estimator $\hat{S}_U(z, x)$ displays negative bias, but that should not exclude it from consideration as the size of that bias is generally small. Moreover it should be noted that, particularly when $t$ and $p$ are both small, $\hat{S}_B(z, x)$ can show a larger negative bias than $\hat{S}_U(z, x)$. It is particularly important not to underestimate the standard deviation in this case in which the information from the data is weak, and it could therefore be argued that $\hat{S}_U(z, x)$ represents a safer choice than $\hat{S}_B(z, x)$.

7. DISCUSSION

The results presented in this paper suggest that the conditionally unbiased estimator $\hat{N}_U$ should prove helpful to practitioners in population size studies. The computations in section 5 indicated that, for the larger sample sizes corresponding to $t = 10, 15$ or 20, $\hat{N}_U$ performed at least as well as the maximum likelihood estimator. It was, in fact, only when $t$ and $p$ were both small that the latter could be considered preferable. In this case, however, the data contain only weak information about $N$, with the result that the true coefficient of variation of the estimator is relatively large. When this happens, it is important not to underestimate the standard deviation $s$, as its magnitude clearly gives a useful warning that the information obtained is low. Moreover the appropriate response in practice may well be to collect more data rather than resorting to use of the maximum likelihood estimator. The arguments for using the maximum likelihood estimator for model $M_0$ in preference to the conditionally unbiased estimator would therefore appear to be few.
REFERENCES


Table 1. The mean $m$ and standard deviation $s$ of the maximum likelihood estimator and the conditionally unbiased estimator.

<table>
<thead>
<tr>
<th>$p$</th>
<th>$N$</th>
<th>$\hat{N}_M$</th>
<th>$\hat{N}_U$</th>
<th>$t$</th>
<th>$5$</th>
<th>$10$</th>
<th>$15$</th>
<th>$20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>50</td>
<td>50.5</td>
<td>49.9</td>
<td>5</td>
<td>9.05</td>
<td>3.00</td>
<td>1.48</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>99.6</td>
<td>50.0</td>
<td>10</td>
<td>2.98</td>
<td>1.50</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50.6</td>
<td>50.0</td>
<td>15</td>
<td>4.17</td>
<td>2.08</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>57.7</td>
<td>50.0</td>
<td>20</td>
<td>4.86</td>
<td>2.08</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>50</td>
<td>57.7</td>
<td>48.2</td>
<td>5</td>
<td>31.8</td>
<td>18.1</td>
<td>11.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50.6</td>
<td>50.0</td>
<td>10</td>
<td>9.29</td>
<td>8.54</td>
<td>8.64</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>49.8</td>
<td>50.0</td>
<td>15</td>
<td>4.97</td>
<td>4.86</td>
<td>4.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>49.6</td>
<td>50.0</td>
<td>20</td>
<td>3.24</td>
<td>3.21</td>
<td>3.00</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>50</td>
<td>43.3</td>
<td>30.1</td>
<td>5</td>
<td>24.4</td>
<td>13.4</td>
<td>11.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>57.7</td>
<td>48.6</td>
<td>10</td>
<td>31.7</td>
<td>18.3</td>
<td>12.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>52.3</td>
<td>49.9</td>
<td>15</td>
<td>16.3</td>
<td>12.7</td>
<td>11.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50.6</td>
<td>50.0</td>
<td>20</td>
<td>9.37</td>
<td>8.64</td>
<td>8.64</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>108</td>
<td>102</td>
<td>5</td>
<td>66.4</td>
<td>39.2</td>
<td>18.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>101</td>
<td>10</td>
<td>39.2</td>
<td>39.2</td>
<td>18.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80.6</td>
<td>80.6</td>
<td>15</td>
<td>29.1</td>
<td>29.1</td>
<td>17.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80.6</td>
<td>80.6</td>
<td>20</td>
<td>29.1</td>
<td>29.1</td>
<td>17.2</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. The mean $s_U$ of the estimator $\hat{S}_U(z, x)$ and the estimated mean $\hat{s}_B$ of the estimator $\hat{S}_B(z, x)$ compared with the true standard deviation $s$.

<table>
<thead>
<tr>
<th>$p$</th>
<th>$s$</th>
<th>$s_U$</th>
<th>$\hat{s}_B$</th>
<th>$s$</th>
<th>$s_U$</th>
<th>$\hat{s}_B$</th>
<th>$s$</th>
<th>$s_U$</th>
<th>$\hat{s}_B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>8.30</td>
<td>7.74</td>
<td>8.86</td>
<td>18.1</td>
<td>16.5</td>
<td>15.7</td>
<td>13.4</td>
<td>13.2</td>
<td>9.63</td>
</tr>
<tr>
<td>0.1</td>
<td>8.54</td>
<td>7.99</td>
<td>10.3</td>
<td>8.64</td>
<td>8.11</td>
<td>9.03</td>
<td>18.3</td>
<td>16.5</td>
<td>16.3</td>
</tr>
<tr>
<td>0.05</td>
<td>8.64</td>
<td>8.11</td>
<td>9.03</td>
<td>33.2</td>
<td>35.2</td>
<td>27.3</td>
<td>29.1</td>
<td>25.8</td>
<td>29.0</td>
</tr>
</tbody>
</table>