

Population size estimation based on estimating functions for closed capture–recapture models[☆]

Anne Chao^{a,*}, Paul S.F. Yip^b, Shen-Ming Lee^c, Wenten Chu^a

^a*Institute of Statistics, National Tsing Hua University, Hsin-Chu 30043, Taiwan*

^b*Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong*

^c*Department of Statistics, Feng-Chia University, Tai-Chung 40724, Taiwan*

Received 19 June 1998; received in revised form 15 April 2000; accepted 5 May 2000

Abstract

This paper explores a unified approach to inference on the size of a closed population using capture–recapture data where time, behavior or heterogeneity may affect the capture probabilities. The unifying idea is that of an optimal estimating function incorporating the notion of sample coverage. The proposed procedure generalizes the martingale estimating function approach (Lloyd, 1994, *Biometrika*, 81, 305–315) and also extends the sample coverage procedure (Lee and Chao, 1994, *Biometrics*, 50, 88–97). Results of a simulation study are reported to investigate the behavior of the proposed procedure. An example concerning captures of deer mice is given to illustrate the method. © 2001 Elsevier Science B.V. All rights reserved.

MSC: primary 62P10; secondary 62G05

Keywords: Behavioral response; Heterogeneity; Mark-recapture; Sample coverage; Time variation

1. Introduction

In biological and ecological sciences, population size estimation is a classical problem and has been extensively discussed in the literature. The pioneering work of Darroch (1958) provided a mathematical foundation for the theory. Seber (1982, 1986, 1992) and

[☆] This research was supported by the National Science Council of Taiwan under Contracts NSC-87-2118-M007-010 and NSC-88-2118-M007-009.

* Corresponding author.

E-mail address: chao@stat.nthu.edu.tw (A. Chao).

Nomenclature

t	number of trapping samples
N	population size
p_{ij}	capture probability of the i th animal in the j th trapping sample
p_i	first-capture probability of the i th individual animal; p_1, p_2, \dots, p_N have mean $\bar{p} = \sum_{i=1}^N p_i/N$ and coefficient of variation (CV) $\gamma = [\sum_{i=1}^N (p_i - \bar{p})^2/N]^{1/2}/\bar{p}$
e_j	multiplicative time effect of the j th trapping sample
ϕ	proportional constant relating recapture probabilities to initial-capture probabilities
C_k	sample coverage of the first k samples, $k = 1, 2, \dots, t$, $C_0 \equiv 0$
X_{ij}	=1 if the i th animal is caught in the j th sample, =0 otherwise
n_k	number of animals captured in the k th sample
u_k	number of unmarked animals captured in the k th sample
m_k	number of marked animals captured in the k th sample, $m_1 = 0$
f_{jk}	number of animals captured exactly j times in samples $1, 2, \dots, k$, $f_{10} \equiv 0$
M_k	number of distinct animals captured just before the k th sample, $M_1 = 0$
Z_ω	number of animals with capture history ω , where ω is a non-empty subset of $\{1, 2, \dots, t\}$, i.e., the number of animals caught in each of the samples indexed by elements of ω , and in no others.

Schwarz and Seber (1999) presented comprehensive reviews on the topic. We focus on population size estimation for the hierarchy of closed capture–recapture models proposed by Pollock (e.g., Pollock, 1991) and Otis et al. (1978). The three basic models considered are (a) model \mathcal{M}_t , which allows capture probabilities to vary by time; (b) model \mathcal{M}_b , which allows behavioral responses to capture; and (c) model \mathcal{M}_h , which allows heterogeneous capture probabilities. Various combinations of these three types of unequal capture probabilities (i.e., models \mathcal{M}_{tb} , \mathcal{M}_{th} , \mathcal{M}_{bh} and \mathcal{M}_{tbh}) are also considered. The special model \mathcal{M}_0 , in which no variation exists, will not be discussed due to its limited practical use.

As noted in the reviews of Seber (1982, 1986, 1992) and Schwarz and Seber (1999), estimators for various models may be found in the literature. These estimators rely on many different approaches: maximum likelihood, jackknife method, log-linear models, Bayesian methods, mixture models, regression methods, sample coverage procedures, martingale estimating functions and an estimator averaging concept. The most general model (model \mathcal{M}_{tbh}) has been considered only conceptually useful and too complex to be applied to practical situations because too many parameters are involved. Nevertheless, it has been selected as the most likely model for estimating squirrel populations

(White et al., 1982, p. 149) and in the example discussed below. If the heterogeneity effects can be characterized by a limited number of parameters using random-effects or similar types of models, then the analysis of model \mathcal{M}_{tbh} becomes possible when data are not sparse. Lee's (1996) conditional sample coverage approach (given the capture data of the first sample) represents the first attempt to deal with the model. He successfully applied the method to the squirrel data of White et al. (1982).

Our interest in this general model was motivated by the deer mouse data presented in Otis et al. (1978, p. 93) which we re-examine in Section 3. The model selection procedure in Otis et al. (1978) shows that the most likely model is model \mathcal{M}_{tbh} and model \mathcal{M}_{b} is the next most likely model. The application of Lee's (1996) approach to this data set yields an estimate with large standard error due to the small amount of information in the first sample. We were thus motivated to extend an approach using estimating functions, previously applied to homogeneous models by Chao et al. (2000), to heterogeneous models. This procedure also generalizes the martingale estimating function approach (Lloyd, 1994) and extends the sample coverage procedure (Lee and Chao, 1994) in which the relative time effects are assumed to be known constants. Recall that an estimating function is a vector function of the parameters and the data, and its roots yield the parameter estimates. An unbiased estimating function is such that the expectation of the function is zero for all parameters. The consistency of the resulting estimators under some regularity conditions relies on the unbiasedness of the estimating functions. Let θ denote the row vector of parameters involved in the model and \mathbf{g}_k be an unbiased estimating function (e.g., a martingale difference) associated with sample k , where $k = 1, 2, \dots, t$. The estimating functions considered in this paper are linear combinations of $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_t$, where the weight associated with \mathbf{g}_k is allowed to be a function of the parameters and the past up to time $k - 1$. Let E_c , var_c , cov_c and Cov_c denote, respectively, the expectation, variance, covariance and variance-covariance matrix conditional on the past up to time $k - 1$. The optimal estimating equation is given by (see Liang and Zeger, 1995 for necessary background.)

$$\mathbf{g} = \sum_{k=1}^t D_k^T V_k^{-1} \mathbf{g}_k = \mathbf{0}, \quad (1.1)$$

where D_k^T denotes the transpose of D_k , $D_k = E_c[\partial \mathbf{g}_k / \partial \theta]$ and $V_k = \text{Cov}_c(\mathbf{g}_k)$. Here the "optimality" is in the sense that the corresponding root θ of $\mathbf{g} = \mathbf{0}$ has minimum asymptotic variance in the class of weighted linear functions of $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_t$. The quantities $D_k^T V_k^{-1}$ play the role of the optimal weights.

Our unified approach is presented in Section 2 along with the resulting estimators of population size for each model. In Section 3, we re-examine the deer mouse data. Section 4 reports some limited simulation results that illustrate the general performance of our procedure and compare it with other approaches. A summary and discussion are given in Section 5. Details of all the derivations and the related asymptotic properties have been omitted and readers are referred to the appendices of a technical report by the authors (Chao et al., 1998).

2. Models and estimators

We first list the models and assumptions below:

Model	Assumption	Restriction in model \mathcal{M}_{tbh}
\mathcal{M}_{tbh}	$p_{ij} = \begin{cases} p_i e_j & \text{until first capture} \\ \phi p_i e_j & \text{for any recapture} \end{cases}$	
\mathcal{M}_{bh}	$p_{ij} = \begin{cases} p_i & \text{until first capture} \\ \phi p_i & \text{for any recapture} \end{cases}$	Set $e_j = 1$ in model \mathcal{M}_{tbh}
\mathcal{M}_{tb}	$p_{ij} = \begin{cases} e_j & \text{until first capture} \\ \phi e_j & \text{for any recapture} \end{cases}$	Set $p_i = 1$ in model \mathcal{M}_{tbh}
\mathcal{M}_{th}	$p_{ij} = p_i e_j$	Set $\phi = 1$ in model \mathcal{M}_{tbh}
\mathcal{M}_{h}	$p_{ij} = p_i$	Set $e_j = 1, \phi = 1$ in model \mathcal{M}_{tbh}
\mathcal{M}_{b}	$p_{ij} = \begin{cases} p & \text{until first capture} \\ \phi p & \text{for any recapture} \end{cases}$	Set $p_i = p, e_j = 1$ in model \mathcal{M}_{tbh}
\mathcal{M}_{t}	$p_{ij} = e_j$	Set $p_i = 1, \phi = 1$ in model \mathcal{M}_{tbh}

Thus the parameters $\{e_1, e_2, \dots, e_t\}$, $\{p_1, p_2, \dots, p_N\}$ and ϕ are used, respectively, to model the time effects, individual heterogeneity and the behavioral response to capture. A multiplicative form of the capture probabilities is assumed for the time and heterogeneity effects, and the recapture probabilities are proportional to the initial capture probabilities. The individual probabilities $\{p_1, p_2, \dots, p_N\}$ could be treated as either fixed or random effects. In a fixed-effects model, it is impractical to estimate the individual p_i 's. To reduce the number of parameters and to remove the non-identification caused by the numerous parameters, we further assume in a fixed-effects model that all the heterogeneity effects are characterized by the mean $\bar{p} = \sum_{i=1}^N p_i / N$ and the coefficient of variation (CV) $\gamma = [\sum_{i=1}^N (p_i - \bar{p})^2 / N]^{1/2} / \bar{p}$. That is, for models with heterogeneity, only the mean and CV of the probabilities $\{p_1, p_2, \dots, p_N\}$ need to be identified. (In a random-effects model, this assumption is equivalent to assuming that $\{p_1, p_2, \dots, p_N\}$ is a random sample from a parametric family with two parameters.) Our derivation focuses on the fixed-effects approach, unless otherwise specified. Both approaches lead to the same estimators.

The animals are assumed to act independently. The capture histories can be expressed in a matrix (X_{ij}) , where $X_{ij} = I$ [the i th animal is caught in the j th sample], and $I[\cdot]$ is the usual indicator function.

To motivate and provide a background to our approach to heterogeneity models, in Sections 2.1 and 2.2, we discuss the widely used homogeneous models \mathcal{M}_{t} and \mathcal{M}_{b} . In Section 2.3, we give the unified estimating approach for the most general model \mathcal{M}_{tbh} . Other models (\mathcal{M}_{tb} , \mathcal{M}_{bh} , \mathcal{M}_{th} and \mathcal{M}_{h}) will be treated as special cases of model \mathcal{M}_{tbh} . Some estimation formulae are outlined in Section 2.4 and variance estimation is discussed in Section 2.5.

2.1. Model \mathcal{M}_t

Model \mathcal{M}_t assumes that the capture probability for all animals in the k th sample is e_k for $k = 1, 2, \dots, t$. Given the capture histories over the first $k - 1$ samples, u_k is a binomial random variable with parameters $(N - M_k)$ and e_k , and m_k is a binomial random variable with parameters M_k and e_k . For each sample k , construct the following unbiased estimating functions ($m_1 = 0$, $M_1 = 0$):

$$\mathbf{g}_k = \begin{bmatrix} g_{1k} \\ g_{2k} \end{bmatrix} = \begin{bmatrix} u_k - (N - M_k)e_k \\ m_k - M_k e_k \end{bmatrix}.$$

For model \mathcal{M}_t , we have $\boldsymbol{\theta} = (N, e_1, e_2, \dots, e_t) \equiv (N, \mathbf{e})$. The matrix D_k and V_k defined in (1.1) become

$$D_k = E_c \begin{bmatrix} \frac{\partial g_{1k}}{\partial N} & \frac{\partial g_{1k}}{\partial \mathbf{e}} \\ \frac{\partial g_{2k}}{\partial N} & \frac{\partial g_{2k}}{\partial \mathbf{e}} \end{bmatrix}, \quad V_k = \begin{bmatrix} \text{var}_c(g_{1k}) & \text{cov}_c(g_{1k}, g_{2k}) \\ \text{cov}_c(g_{1k}, g_{2k}) & \text{var}_c(g_{2k}) \end{bmatrix}. \quad (2.1)$$

Note that D_k is a matrix with dimension $2 \times (t + 1)$. Substituting $\text{var}_c(g_{1k}) = (N - M_k)e_k(1 - e_k)$, $\text{var}_c(g_{2k}) = M_k e_k(1 - e_k)$ and $\text{cov}_c(g_{1k}, g_{2k}) = 0$ (u_k and m_k are conditionally independent) into the above, we obtain the following $t + 1$ equations for \mathbf{g} :

Equation for N :

$$\sum_{k=1}^t [(N - M_k)(1 - e_k)]^{-1} [u_k - (N - M_k)e_k] = 0, \quad (2.2a)$$

Equation for e_k :

$$n_k - Ne_k = 0, \quad k = 1, 2, \dots, t. \quad (2.2b)$$

From (2.2b), we have $e_k = n_k/N$ and Eq. (2.2a) is then equivalent to

$$\sum_{k=1}^t [(N - M_k)(N - n_k)]^{-1} [M_k u_k - (N - M_k)m_k] = 0.$$

The above equation was first derived by Yip (1991), who expressed $M_k u_k - (N - M_k)m_k$ as a martingale difference and $[(N - M_k)(N - n_k)]^{-1}$ as the optimal weight associated with that martingale difference.

2.2. Model \mathcal{M}_b

For model \mathcal{M}_b , the two unbiased estimating functions are $g_{1k} = u_k - (N - M_k)p$ and $g_{2k} = m_k - M_k p$. Therefore, equations in (1.1) are

Equation for N :

$$\sum_{k=1}^t [(N - M_k)(1 - p)]^{-1} [u_k - (N - M_k)p] = 0, \quad (2.3a)$$

Equation for p :

$$\sum_{k=1}^t [p(1 - p)]^{-1} [u_k - (N - M_k)p] = 0, \quad (2.3b)$$

Equation for ϕ :

$$\sum_{k=1}^t [\phi p(1 - \phi p)]^{-1} [m_k - M_k \phi p] = 0. \quad (2.3c)$$

The constants ϕp , $1 - \phi p$, p and $1 - p$ in the denominators of the above equations can be removed without affecting the solution. The third equation is used to estimate the recapture probability only. The first two equations, which are identical to those in Lloyd (1994), are jointly used to solve for N and p . See Lloyd (1994) for details and the efficiency comparison.

2.3. Model \mathcal{M}_{tbh} and other special cases

We first define the sample coverage. The sample coverage C_k is defined as the proportion of the total heterogeneous probabilities that is associated with the captured animals in samples 1 to k . That is, C_k can be expressed as

$$C_k = \sum_{i=1}^N p_i I \left[\sum_{j=1}^k X_{ij} > 0 \right] / \sum_{i=1}^N p_i. \quad (2.4)$$

Note that for models without heterogeneity (i.e., $p_1 = p_2 = \dots = p_N$), C_k reduces to M_{k+1}/N . It can be shown under heterogeneous models that for $k \geq 1$, ($C_0 \equiv 0$)

$$E_c(u_k) = (1 - C_{k-1}) \left(\sum_{i=1}^N p_i \right) e_k = (N - NC_{k-1}) \bar{p} e_k,$$

$$E_c(m_k) = C_{k-1} \left(\sum_{i=1}^N p_i \right) \phi e_k = (NC_{k-1}) \phi \bar{p} e_k.$$

If NC_{k-1} is treated as a function of N , we can only estimate the confounded multiplicative effect of $N \bar{p} e_k$ and would not be able to separately estimate the parameter N and $\bar{p} e_k$. Noting that in models without heterogeneity, $NC_{k-1} = M_k$, in heterogeneous models we were motivated to express the quantity NC_{k-1} as a sum of M_k and other parameters (e.g., CV, the behavioral effect and/or the relative time effects). Therefore, we define $\alpha_k = \bar{p} e_k$, $M_1^* = 0$, $M_k^* = NC_{k-1}$ for $k \geq 2$, and construct unbiased estimating functions for each sample k :

$$\mathbf{g}_k = \begin{bmatrix} g_{1k} \\ g_{2k} \end{bmatrix} = \begin{bmatrix} u_k - (N - M_k^*) \alpha_k \\ m_k - M_k^* \phi \alpha_k \end{bmatrix}. \quad (2.5)$$

Given the capture data of the first $k - 1$ samples, we have

$$\text{cov}_c(g_{1k}, g_{2k}) = 0,$$

$$\text{var}_c(g_{1k}) = (1 - C_{k-1}) \left(\sum_{i=1}^N p_i \right) e_k - (1 - C'_{k-1}) \left(\sum_{i=1}^N p_i^2 \right) e_k^2,$$

$$\text{var}_c(g_{2k}) = C_{k-1} \left(\sum_{i=1}^N p_i \right) \phi e_k - C'_{k-1} \left(\sum_{i=1}^N p_i^2 \right) \phi^2 e_k^2,$$

where

$$C'_k = \sum_{i=1}^N p_i^2 I \left[\sum_{j=1}^k X_{ij} > 0 \right] / \sum_{i=1}^N p_i^2.$$

As indicated in Liang and Zeger (1995), even if the weight function is mis-specified, the consistency of the solution based on unbiased estimating functions remains valid. Since both variances are used only as the weight function in (1.1) and generally have little impact on the inference, they can be simplified to render the problem analytically tractable. Here we suppose $C_k \approx C'_k$ for all k so that

$$\text{var}_c(g_{1k}) = (N - NC_{k-1})\alpha_k[1 - (1 + \gamma^2)\alpha_k], \quad (2.6a)$$

$$\text{var}_c(g_{2k}) = NC_{k-1}\phi\alpha_k[1 - (1 + \gamma^2)\phi\alpha_k]. \quad (2.6b)$$

For the moment we assume that the M_k^* and CV are known. Based on (1.1), (2.5) and (2.6), the optimal estimating equations of (1.1) are the following:

Equation for N :

$$\sum_{k=1}^t \frac{u_k - (N - M_k^*)\alpha_k}{(1 - C_{k-1})[1 - (1 + \gamma^2)\alpha_k]} = 0, \quad (2.7a)$$

Equation for ϕ :

$$\sum_{k=1}^t \frac{m_k - M_k^*\phi\alpha_k}{[1 - (1 + \gamma^2)\phi\alpha_k]} = 0, \quad (2.7b)$$

Equation for α_k :

$$\frac{u_k - (N - M_k^*)\alpha_k}{[1 - (1 + \gamma^2)\alpha_k]} + \frac{m_k - M_k^*\phi\alpha_k}{[1 - (1 + \gamma^2)\phi\alpha_k]} = 0, \quad k = 1, 2, \dots, t. \quad (2.7c)$$

The constants $1/N$, $1/\phi$ and $1/\alpha_k$ are, respectively, dropped in (2.7a)–(2.7c) without affecting the solution. From the equation for α_k , we can express α_k in terms of ϕ and N :

$$\alpha_k = \alpha_k(\phi, N) = \frac{A_k - [A_k^2 - 4N\phi n_k(1 + \gamma^2)]^{1/2}}{2N\phi(1 + \gamma^2)}, \quad (2.7c')$$

where

$$A_k = A_k(\phi, N) = N + \phi n_k(1 + \gamma^2) + (\phi - 1)[NC_{k-1} - (1 + \gamma^2)m_k]. \quad (2.8)$$

For models with time-varying effects, (2.7c) implies that (2.7a) and (2.7b) are equivalent to

Equation for N :

$$\sum_{k=1}^t \frac{M_k^*(\phi u_k + m_k) - Nm_k}{(1 - C_{k-1})[1 + (\phi - 1)C_{k-1} - \phi(1 + \gamma^2)\alpha_k]} = 0, \quad (2.9a)$$

Equation for ϕ :

$$\sum_{k=1}^t \frac{M_k^*(\phi u_k + m_k) - Nm_k}{[1 + (\phi - 1)C_{k-1} - \phi(1 + \gamma^2)\alpha_k]} = 0. \quad (2.9b)$$

Now we can treat other models as special cases of model \mathcal{M}_{tb} :

- Model \mathcal{M}_{tb} : substitute $M_k^* = M_k$, $\bar{p} = 1$, $\alpha_k = e_k$, $\gamma^2 = 0$ into (2.7a)–(2.7c), the system of equations reduce to

Equation for N :

$$\sum_{k=1}^t \frac{u_k - (N - M_k)e_k}{(N - M_k)(1 - e_k)} = 0, \quad (2.10a)$$

Equation for ϕ :

$$\sum_{k=1}^t \frac{m_k - M_k \phi e_k}{(1 - \phi e_k)} = 0, \quad (2.10b)$$

Equation for e_k :

$$\frac{u_k - (N - M_k)e_k}{(1 - e_k)} + \frac{m_k - M_k \phi e_k}{(1 - \phi e_k)} = 0, \quad k = 1, 2, \dots, t. \quad (2.10c)$$

Eq. (2.10c) or (2.7c') reduces to

$$e_k = e_k(\phi, N) = \frac{A_k - (A_k^2 - 4N\phi n_k)^{1/2}}{2N\phi}, \quad (2.10c')$$

where $A_k = A_k(\phi, N) = N + \phi n_k + (\phi - 1)(M_k - m_k)$. Equations (2.9a) and (2.9b) then become

Equation for N :

$$\sum_{k=1}^t \frac{M_k(\phi u_k + m_k) - Nm_k}{(N - M_k)[N + (\phi - 1)M_k - N\phi e_k]} = 0, \quad (2.11a)$$

Equation for ϕ :

$$\sum_{k=1}^t \frac{M_k(\phi u_k + m_k) - Nm_k}{[N + (\phi - 1)M_k - N\phi e_k]} = 0, \quad (2.11b)$$

where $e_k = e_k(N, \phi)$ is given in (2.10c'). The above estimating equations (2.11a) and (2.11b) are identical to those derived in Lloyd (1994), but he used $e_k = u_k/(N - M_k)$ which is different from the optimal choice of (2.10c').

For the special case of model \mathcal{M}_t , it follows from (2.10c') that $e_k = n_k/N$ as in (2.2b), and observe that (2.10a) is identical to (2.2a). For the special model of model \mathcal{M}_b , we let $e_1 = e_2 = \dots = e_t \equiv p$ in the derivation of \mathcal{M}_{tb} . The resulting estimating equation is equivalent to summing the t equations in (2.10c) into one equation for p . It is clear that the resulting equations are equivalent to (2.3).

For models without heterogeneity (\mathcal{M}_t , \mathcal{M}_b and \mathcal{M}_{tb}), the usual maximum-likelihood estimator exists under the assumption that the recapture probabilities bear a constant relationship to the initial probabilities in model \mathcal{M}_{tb} . Lloyd (1994) established the asymptotic equivalence of the proposed estimator and the MLE for models \mathcal{M}_t and \mathcal{M}_b when the population size is large. Chao et al. (2000) extends Lloyd's result to

model \mathcal{M}_{tb} . For finite sizes, simulation results in Chao et al. (2000) have shown that both estimates are generally comparable. An obvious advantage of our approach is that it can be extended to heterogenous populations where the MLE usually does not exist.

- Model \mathcal{M}_{bh} : let $\alpha_1 = \alpha_2 = \dots = \alpha_t \equiv \bar{p}$ in the derivation of model \mathcal{M}_{tbh} . The resulting equation for \bar{p} is equivalent to that by summing over k in (2.7c). This yields Equation for N :

$$\sum_{k=1}^t \frac{u_k - (N - M_k^*)\bar{p}}{(1 - C_{k-1})} = 0, \quad (2.12a)$$

Equation for ϕ :

$$\sum_{k=1}^t (m_k - M_k^* \phi \bar{p}) = 0, \quad (2.12b)$$

Equation for \bar{p} :

$$\sum_{k=1}^t \left\{ \frac{u_k - (N - M_k^*)\bar{p}}{[1 - (1 + \gamma^2)\bar{p}]} + \frac{m_k - M_k^* \phi \bar{p}}{[1 - (1 + \gamma^2)\phi \bar{p}]} \right\} = 0. \quad (2.12c)$$

The constants $[1 - (1 + \gamma^2)\bar{p}]$ and $[1 - (1 + \gamma^2)\phi \bar{p}]$ have been removed in the denominators of (2.12a) and (2.12b), respectively, without affecting the solution. From (2.12c), we obtain

$$\bar{p} = \bar{p}(\phi, N) = \frac{A - [A^2 - 4tN\phi(\sum_{k=1}^t n_k)(1 + \gamma^2)]^{1/2}}{2tN\phi(1 + \gamma^2)}, \quad (2.12c')$$

where

$$A = A(\phi, N) = tN + \phi(\sum_{k=1}^t n_k)(1 + \gamma^2) + (\phi - 1) \sum_{k=1}^t [NC_{k-1} - (1 + \gamma^2)m_k].$$

- Model \mathcal{M}_{th} : remove the equation for ϕ and substitute $\phi=1$ into (2.7c). The resulting equations are:

Equation for N :

$$\sum_{k=1}^t \frac{u_k - (N - M_k^*)\alpha_k}{(1 - C_{k-1})[1 - (1 + \gamma^2)\alpha_k]} = 0, \quad (2.13a)$$

Equation for α_k :

$$\alpha_k = n_k/N, \quad k = 1, 2, \dots, t. \quad (2.13b)$$

Substituting $\alpha_k = n_k/N$ into (2.13a) yields

Equation for N :

$$\sum_{k=1}^t \frac{M_k^* n_k - N m_k}{(1 - C_{k-1})[N - (1 + \gamma^2)n_k]} = 0,$$

which is equivalent to (2.9a) with $\phi = 1$.

- Model \mathcal{M}_h : either take the equations for model \mathcal{M}_{bh} , omitting Eq. (2.12b) for ϕ , or use those for model \mathcal{M}_{th} , setting $\alpha_1 = \alpha_2 = \dots = \alpha_t = \bar{p}$ and summing Eqs. (2.13b) into a single equation. We have

Equation for N :

$$\sum_{k=1}^t \frac{u_k - (N - M_k^*)\bar{p}}{(1 - C_{k-1})} = 0, \quad (2.14a)$$

Equation for \bar{p} :

$$\bar{p} = \sum_{k=1}^t n_k / (tN). \quad (2.14b)$$

Substituting (2.14b) into (2.14a), we obtain an explicit estimator for N in terms of the other quantities involved.

2.4. Estimation of M_k^* , C_{k-1} and CV for heterogeneous models

For models without heterogeneity, we have $M_k^* = M_k$, $C_{k-1} = M_k/N$ and $CV = 0$. Thus the estimating equations are those described above. For models with heterogeneity, however, the three quantities (M_k^* , C_{k-1} and CV) need to be estimated or expressed as functions of N and ϕ . In this section, we only list the resulting formulae without giving the proofs.

- Model \mathcal{M}_h :

$$\hat{C}_{k-1} = 1 - f_{1k} / \sum_{i=1}^k n_i,$$

$$\hat{M}_k^* = M_k + f_{1,k-1} \hat{\gamma}_h^2,$$

$$\hat{\gamma}_h^2 = \max \left\{ \frac{\hat{N}_0 t \sum_{k=1}^t k(k-1) f_{kt}}{(t-1)(\sum_{k=1}^t k f_{kt})^2} - 1, 0 \right\}$$

$$\text{where } \hat{N}_0 = M_{t+1} / \left[1 - f_{1t} / \left(\sum_{k=1}^t k f_{kt} \right) \right].$$

- Model \mathcal{M}_{th} :

$$\hat{C}_{k-1} = 1 - u_k / n_k,$$

$$\hat{M}_k^* = M_k + f_{1,k-1} \hat{\gamma}_{th}^2,$$

$$\hat{\gamma}_{th}^2 = \max \left\{ \frac{\hat{N}_0 \sum_{k=1}^t k(k-1) f_{kt}}{(2 \sum_{j < k} n_j n_k)} - 1, 0 \right\}$$

$$\text{where } \hat{N}_0 = M_{t+1} / \left[1 - f_{1t} / \left(\sum_{k=1}^t k f_{kt} \right) \right].$$

- Model \mathcal{M}_{bh} :

$$\hat{C}_{k-1} = \hat{C}_{k-1}(\phi) = 1 - u_k/(u_k + m_k/\phi),$$

$$\hat{M}_k^* = M_k + (k-1)u_{k-1}\hat{\gamma}_{bh}^2,$$

$$\hat{\gamma}_{bh}^2 = \hat{\gamma}_{bh}^2(\phi) = \max \left\{ \frac{\hat{N}_h t \sum_{k=1}^t [k(k-1)f_{kt} + 2(\phi-1)(k-1)f_{kt}]}{(t-1)[\sum_{k=1}^t (m_k + \phi u_k)]^2} - 1, 0 \right\},$$

where \hat{N}_h is a simple estimator valid under model \mathcal{M}_h , i.e., $\hat{N}_h = [M_{t+1} + f_{1t}\hat{\gamma}_h^2]/[1 - f_{1t}/(\sum_{k=1}^t k f_{kt})]$.

- Model \mathcal{M}_{tbh} :

$$\hat{C}_{k-1} = \hat{C}_{k-1}(\phi) = 1 - u_k/(u_k + m_k/\phi),$$

$$\hat{M}_k^* = M_k + \left[\sum_{j=1}^{k-1} \hat{\rho}_{j,k-1} \right] u_{k-1} \hat{\gamma}_{tbh}^2,$$

$$\hat{\gamma}_{tbh}^2 = \hat{\gamma}_{tbh}^2(\phi) = \max \left\{ \frac{\hat{N}_{bh} \sum_{k=1}^t [k(k-1)f_{kt} + 2(\phi-1)(k-1)f_{kt}]}{[\sum_{k=1}^t (m_k + \phi u_k)]^2 - \sum_{k=1}^t (m_k + \phi u_k)^2} - 1, 0 \right\},$$

where \hat{N}_{bh} is a simple estimator valid under model \mathcal{M}_{bh} . Here $\rho_{j,k-1} = e_j/e_{k-1}$ denotes the unknown relative time-effect of sample j to sample $k-1$. A convenient estimator of $\rho_{j,k-1} = e_j/e_{k-1}$ as a function of ϕ is

$$\hat{\rho}_{j,k-1} = \hat{\rho}_{j,k-1}(\phi) = (u_j + m_j/\phi)/(u_{k-1} + m_{k-1}/\phi).$$

Remark 1 (*Computational algorithm*). We use model \mathcal{M}_{tbh} as an illustration. Given an initial value of (N, ϕ) , we can compute numerical estimates of \hat{C}_{k-1} , \hat{M}_k^* , $\hat{\gamma}_{tbh}$ as well as α_k as that given in (2.7c'). The numerical “optimal weights”, i.e., the denominators in (2.7a) and (2.7b) or (2.9), are subsequently calculated. Eqs. (2.7a), (2.7b) or (2.9) then become linear equations in (N, ϕ) . Thus a second value of (N, ϕ) can easily be obtained. The iterative steps continue until convergence. If the resulting estimate is less than the number of distinct animals caught (that is, M_{t+1}), then the estimate is replaced by M_{t+1} .

Remark 2 (*Motivation for the formula of \hat{M}_k^**). The basic motivation for using $\hat{M}_k^* = M_k + f_{1,k-1}\hat{\gamma}_h^2$ for models \mathcal{M}_h and \mathcal{M}_{th} is the following: If there exists a constant v such that $\{vp_1, vp_2, \dots, vp_N\}$ is a random sample from a gamma distribution, then it can be shown that for a known CV, we have $N^{-1}\{M_k^* - (M_k + f_{1,k-1}\gamma^2)\} \rightarrow 0$ with probability 1 for all $k=2, \dots, t$ as N becomes large. If an estimated CV is used, then bias may be introduced. Similarly, for model \mathcal{M}_{bh} we can also verify under the same condition that $N^{-1}\{M_k^* - [M_k + (k-1)u_{k-1}\gamma^2]\} \rightarrow 0$ with probability 1. For model \mathcal{M}_{tbh} , the conclusion holds only when all e_j 's are approximately equal. For unequal time effects, an additional source of bias is induced.

Remark 3 (*The estimation of CV*). All the CV estimators given above are calculable from data or in terms of the parameter ϕ which can be found in the iterative steps. We can consider alternative estimators which are also functions of the parameter N . For example, in the most widely used heterogeneous model \mathcal{M}_h , the CV estimator given above is derived based on the following equality:

$$\gamma^2 = \frac{Nt \sum_{k=1}^t k(k-1)E(f_{kt})}{(t-1)[E(\sum_{k=1}^t kf_{kt})]^2} - 1.$$

We could regard N in the above as an unknown parameter in the estimation procedure and obtain the resulting estimate by iterations, but this would generally result in increased variance.

2.5. Variance estimators and confidence intervals

The bootstrap procedure (Efron and Tibshirani, 1993) has been applied to capture–recapture sampling by Buckland (1984), Buckland and Garthwaite (1991) and Norris and Pollock (1996). We adopt a non-parametric bootstrap to estimate the variance of our population size estimator. The capture histories $\{Z_\omega | \omega \in \Omega\}$, where Ω denotes the collection of all subsets of $\{1, 2, \dots, t\}$ and z_\emptyset (\emptyset is an empty set) denotes the missing cell, are approximately distributed as a multinomial distribution; see Darroch et al. (1993).

Suppose that we want to obtain a bootstrap variance for any estimator \hat{N} on the basis of given observable capture histories $\{Z_\omega | \omega \in \Omega \setminus \emptyset\}$. A bootstrap sample $\{Z_\omega^* | \omega \in \Omega\}$ is generated from a multinomial distribution with cell total \hat{N} and cell probabilities Z_ω/\hat{N} for any observable cell and $1 - \sum Z_\omega/\hat{N}$ for the missing cell. Then based on the generated observed data $\{Z_\omega^* | \omega \in \Omega \setminus \emptyset\}$, a bootstrap estimate can be obtained. After B replications, the bootstrap variance estimator of \hat{N} is simply the sample variance of those B bootstrap estimates. A log-transformation suggested by Burnham (Chao, 1987) can then be used to construct an associated confidence interval based on the estimated bootstrap variance. The performance of the bootstrap variance estimator and the confidence interval will be discussed in Section 4.

3. Deer mouse example

The deer mouse data were originally collected by S. Hoffman and are described and analyzed in Otis et al. (1978, p. 93). Detailed capture information was given as an example in program CAPTURE (Rexstad and Burnham, 1991). Trapping was conducted on five days and 110 distinct mice were caught. The numbers of captures for the five samples are $(n_1, n_2, \dots, n_5) = (37, 54, 58, 65, 69)$ and of new captures are $(u_1, u_2, \dots, u_5) = (37, 31, 9, 21, 12)$. Hence we have $(M_2, M_3, \dots, M_6) = (37, 68, 77, 98, 110)$. The frequency counts in the five samples are $(f_{15}, f_{25}, \dots, f_{55}) = (34, 20, 28, 15, 13)$.

Table 1
Various estimates for the deer mouse data

Model	Estimator	\hat{N}	s.e.(\hat{N})	90% C. I. of N	$\hat{\phi}$	$\hat{\gamma}$
\mathcal{M}_b	MLE	142	16.4	(124, 181)	2.38	—
	(2.3)	140	21.7	(120, 197)	2.36	—
\mathcal{M}_{tb}	(2.10)	152	32.9	(125, 231)	2.87	—
\mathcal{M}_h	Jackknife	156	16.3	(136, 191)	—	—
	(2.14)	119	5.0	(114, 131)	—	0.51
\mathcal{M}_{th}	Sample coverage I	136	9.2	(125, 155)	—	0.53
	Sample coverage II	129	8.0	(119, 147)	—	0.48
	(2.13)	118	5.8	(113, 133)	—	0.53
\mathcal{M}_{bh}	Generalized removal	142	16.4	(124, 181)	2.38	—
	Jackknife	158	15.3	(139, 190)	—	—
	(2.12)	125	11.8	(115, 157)	1.43	0.44
\mathcal{M}_{tbh}	(2.7)	123	38.7	(111,267)	1.03	0.52

Otis et al. (1978) concluded that for these data behavior is the strongest factor affecting capture probabilities. The model selection procedure in Otis et al. (1978) shows that the most likely model is model \mathcal{M}_{tbh} and model \mathcal{M}_b is the next most likely model. All the models with behavior effects and/or heterogeneity as well as the associated estimates are presented in Table 1. The MLE for model \mathcal{M}_b and the generalized removal estimator for model \mathcal{M}_{bh} (Otis et al., 1978) yield the same estimate of 142 with an asymptotic s.e. of 16.4. Two jackknife estimators are given: the interpolated jackknife estimator (Burnham and Overton, 1978) for model \mathcal{M}_h , and the one proposed for model \mathcal{M}_{bh} in Pollock and Otto (1983). The two estimates of population size based on the sample coverage approach proposed in Lee and Chao (1994) for model \mathcal{M}_h are close to those for model \mathcal{M}_{th} . Hence only results for model \mathcal{M}_{th} are shown. There are, of course, other approaches (e.g., log-linear models, mixture models and the Bayesian approach among others) that can be used to estimate population size under heterogeneous models. We remark that in the log-linear model approach (Agresti, 1994; Cormack, 1989; Darroch et al., 1993), quasi-symmetric models are usually fitted to the observable capture history data and extrapolated to the missing cell. However, the fitting was not stable and the estimate was not comparable to others probably because of too many small cells in this data set.

The estimated s.e.'s in Table 1 were, except for the MLE and generalized removal estimator, obtained by generating 2000 bootstrap replications which subsequently were used to construct 90% confidence intervals based on the estimated variance. A log-transformation was applied to all estimation methods to obtain confidence intervals where the lower limit is always greater than the number of distinct mice caught in the experiment (that is, 110).

If model \mathcal{M}_{tbh} is assumed, Lee's (1996) conditional sample coverage approach produces an estimate of 277 with a large estimated standard error of 107 as there is insufficient conditional information. Solving the system of equations (2.7) shows that our proposed estimator yields an estimate of 123 with an estimated bootstrap s.e. of 38.7. The confidence interval associated with our estimate under model \mathcal{M}_{tbh} is unavoidably longer than others in Table 1 because of the estimation of more parameters.

It is interesting to note that the estimates for the behavioral effect and heterogeneity effect vary greatly between the models. If only the behavioral effect is allowed in the model, then there is strong evidence for a trap-happy response. However, if heterogeneity and time effects are also considered in the model, the relatively large value of the CV (0.52) shows strong heterogeneity and the behavioral effect becomes less evident ($\hat{\phi} = 1.03$). Since the most likely model suggested by Otis et al. (1978) is model \mathcal{M}_{tbh} and these data include a relatively high proportion of captures and recaptures to estimate all effects, the estimate associated with the most general model \mathcal{M}_{tbh} merits consideration. A program written in C language which calculates all the proposed estimators may be obtained from the first author upon request and will be available soon on the website <http://www.stat.nthu.edu.tw/~chao/>.

4. Simulation study

A simulation study was carried out to investigate the performance of the proposed estimators. For models without heterogeneity, Chao et al. (2000) have shown that the proposed estimator is asymptotically equivalent to the MLE. Therefore, we focus on heterogeneous models here. We concentrate on the two models \mathcal{M}_h and \mathcal{M}_{th} as there are more estimators available for our comparisons. The results are given in Tables 2 and 3, respectively. A more extensive simulation study for the most general model will be reported elsewhere.

The true population size (N) was fixed to be 400 and $t=5$ was chosen. The capture probabilities were randomly sampled from a beta distribution. The 10 beta distributions are arranged in increasing order of the value of the CV which varies from 0.2 to 0.7. The multiplicative time effects were selected as $(e_1, \dots, e_5) = (0.9, 0.8, 0.65, 0.8, 0.9)$ for model \mathcal{M}_{th} . Before the generation of each set of data, a set of 400 random variants was first generated from a specified beta (α, β) distribution as the capture probabilities $\{p_1, p_2, \dots, p_N\}$. Then they were kept fixed through the simulation. Their averages and values of the CV are also given in the two tables. For each trial, 200 data sets were generated. For each randomly generated data set, the following estimators were also calculated:

\hat{N}_t : the MLE under model \mathcal{M}_t ; see Seber (1982, p. 131). The estimator \hat{N}_t satisfies

$$1 - M_{t+1} = \prod_{i=1}^t (1 - n_i/N).$$

\hat{N}_{jk} : the widely used interpolated jackknife estimator proposed by Burnham and Overton (1978). This estimator is a linear function of the capture frequencies.

Table 2
Comparison of various estimators for model \mathcal{M}_h^a

Trial	Estimator	Average estimate	Sample s.e.	Average bootstrap s.e.	Sample RMSE	Bootstrap C.I. coverage (%)
Beta (10, 10)	\hat{N}_t	392	4.9	3.9	9.6	70.5
$\bar{p} = 0.492$	\hat{N}_{jk}	428	21.7	20.4	35.2	11.0
$CV = 0.223$	\hat{N}_{sc}	399	7.3	6.9	7.3	94.0
$M_{t+1} = 381$	\hat{N}_h	403	6.5	6.7	6.9 ^b	91.0
Beta (10, 20)	\hat{N}_t	383	9.2	8.7	19.0	64.5
$\bar{p} = 0.341$	\hat{N}_{jk}	441	13.4	18.7	43.2	12.5
$CV = 0.262$	\hat{N}_{sc}	403	15.1	14.9	15.4 ^b	93.0
$M_{t+1} = 341$	\hat{N}_h	407	14.2	14.8	16.0	92.5
Beta (5, 5)	\hat{N}_t	384	5.0	3.5	17.1	6.5
$\bar{p} = 0.499$	\hat{N}_{jk}	429	12.3	16.4	31.5	26.5
$CV = 0.306$	\hat{N}_{sc}	397	7.7	7.5	8.3	93.5
$M_{t+1} = 374$	\hat{N}_h	398	6.9	6.9	7.2 ^b	95.0
Beta (5, 8)	\hat{N}_t	375	8.2	6.6	26.4	17.5
$\bar{p} = 0.381$	\hat{N}_{jk}	434	11.8	16.0	35.9	23.0
$CV = 0.340$	\hat{N}_{sc}	397	12.7	12.7	13.0	94.5
$M_{t+1} = 348$	\hat{N}_h	400	11.7	12.0	11.7 ^b	96.5
Beta (4, 8)	\hat{N}_t	367	9.8	8.4	34.7	12.0
$\bar{p} = 0.333$	\hat{N}_{jk}	430	19.3	21.8	35.7	70.5
$CV = 0.385$	\hat{N}_{sc}	396	16.4	16.0	17.0	93.5
$M_{t+1} = 328$	\hat{N}_h	399	14.7	15.0	14.8 ^b	95.5
Beta (3, 5)	\hat{N}_t	358	8.0	6.6	42.6	0
$\bar{p} = 0.362$	\hat{N}_{jk}	422	17.5	19.5	27.8	80.0
$CV = 0.434$	\hat{N}_{sc}	388	14.9	13.9	19.0	88.0
$M_{t+1} = 332$	\hat{N}_h	389	12.8	12.7	17.0 ^b	90.5
Beta (3, 10)	\hat{N}_t	335	15.4	13.4	66.6	3.0
$\bar{p} = 0.230$	\hat{N}_{jk}	432	38.4	35.6	50.1	72.5
$CV = 0.518$	\hat{N}_{sc}	383	26.5	25.6	31.3	89.5
$M_{t+1} = 268$	\hat{N}_h	384	23.4	24.3	28.3 ^b	90.5
Beta (1, 1)	\hat{N}_t	339	6.3	2.4	61.3	0
$\bar{p} = 0.492$	\hat{N}_{jk}	394	14.2	16.6	15.3 ^b	95.0
$CV = 0.577$	\hat{N}_{sc}	368	10.0	9.0	34.0	19.5
$M_{t+1} = 335$	\hat{N}_h	361	8.6	7.4	40.3	2.5
Beta (2, 10)	\hat{N}_t	316	19.6	19.6	86.7	7.5
$\bar{p} = 0.165$	\hat{N}_{jk}	411	29.4	32.1	31.3 ^b	94.0
$CV = 0.573$	\hat{N}_{sc}	367	32.4	35.1	46.2	88.0
$M_{t+1} = 218$	\hat{N}_h	369	29.3	35.9	42.8	94.5
Beta (0.5, 0.5)	\hat{N}_t	300	5.4	1.3	99.9	0
$\bar{p} = 0.507$	\hat{N}_{jk}	351	20.1	18.3	53.3 ^b	48.5
$CV = 0.713$	\hat{N}_{sc}	322	8.5	7.1	78.7	0
$M_{t+1} = 300$	\hat{N}_h	312	6.9	5.2	88.2	0

^a $N=400,200$ simulation runs, 500 bootstrap replications, \hat{N}_t : the MLE under model \mathcal{M}_t ; \hat{N}_{jk} : the interpolated jackknife estimator; \hat{N}_{sc} : an estimator based on sample coverage approach; \hat{N}_h : the proposed estimator under model \mathcal{M}_h , see Eq. (2.14).

^bDenotes the smallest RMSE.

Table 3

Comparison of various estimators for model \mathcal{M}_{th}^a

Trial	Estimator	Average estimate	Sample s.e.	Average bootstrap s.e.	Sample RMSE	Bootstrap C.I. coverage (%)
Beta (10, 10)	\hat{N}_t	390	7.0	6.5	12.3	78.5
$\bar{p} = 0.492$	\hat{N}_{jk}	446	10.5	15.7	46.8	6.0
$CV = 0.223$	\hat{N}_{sc}	403	11.0	11.3	11.5	91.5
$M_{t+1} = 362$	\hat{N}_{th}	406	8.6	10.4	10.4 ^a	90.5
Beta (10, 20)	\hat{N}_t	383	13.5	12.7	21.8 ^b	76.0
$\bar{p} = 0.341$	\hat{N}_{jk}	452	33.5	32.0	62.1	55.5
$CV = 0.262$	\hat{N}_{sc}	412	21.7	21.6	24.9	88.5
$M_{t+1} = 314$	\hat{N}_{th}	416	18.9	19.8	24.5	86.5
Beta (5, 5)	\hat{N}_t	380	6.9	5.9	21.7	21.0
$\bar{p} = 0.499$	\hat{N}_{jk}	436	10.8	15.2	37.5	15.5
$CV = 0.306$	\hat{N}_{sc}	398	11.7	11.2	11.8	95.0
$M_{t+1} = 357$	\hat{N}_{th}	399	9.9	10.1	9.9 ^b	96.0
Beta (5, 8)	\hat{N}_t	372	11.0	9.9	29.9	36.0
$\bar{p} = 0.381$	\hat{N}_{jk}	436	22.9	24.9	42.5	65.0
$CV = 0.340$	\hat{N}_{sc}	401	18.0	17.8	18.0	92.5
$M_{t+1} = 324$	\hat{N}_{th}	403	15.8	16.0	16.1 ^b	94.0
Beta (4, 8)	\hat{N}_t	361	11.9	11.7	40.9	20.5
$\bar{p} = 0.333$	\hat{N}_{jk}	436	33.3	32.2	48.8	75.0
$CV = 0.385$	\hat{N}_{sc}	398	21.4	21.7	21.5	95.0
$M_{t+1} = 300$	\hat{N}_{th}	396	18.3	18.9	18.7 ^b	96.0
Beta (3, 5)	\hat{N}_t	356	11.7	9.9	45.6	6.5
$\bar{p} = 0.362$	\hat{N}_{jk}	427	28.6	28.1	39.1	83.5
$CV = 0.434$	\hat{N}_{sc}	393	19.1	19.4	20.2	94.5
$M_{t+1} = 308$	\hat{N}_{th}	392	17.3	16.9	18.9 ^b	95.0
Beta (3, 10)	\hat{N}_t	334	18.5	18.2	68.4	15.5
$\bar{p} = 0.230$	\hat{N}_{jk}	437	32.7	33.8	49.7	70.0
$CV = 0.518$	\hat{N}_{sc}	389	32.6	33.7	34.5	95.5
$M_{t+1} = 241$	\hat{N}_{th}	385	28.5	29.1	32.1 ^b	92.5
Beta (1, 1)	\hat{N}_t	332	7.1	4.1	68.8	0
$\bar{p} = 0.492$	\hat{N}_{jk}	397	21.8	20.5	22.1 ^b	91.5
$CV = 0.577$	\hat{N}_{sc}	362	12.3	11.2	40.1	19.5
$M_{t+1} = 310$	\hat{N}_{th}	359	10.5	9.5	42.5	9.5
Beta (2, 10)	\hat{N}_t	309	23.1	25.7	93.8	17.5
$\bar{p} = 0.165$	\hat{N}_{jk}	390	25.6	29.0	27.4 ^b	98.0
$CV = 0.573$	\hat{N}_{sc}	376	45.9	46.7	51.8	92.0
$M_{t+1} = 191$	\hat{N}_{th}	367	33.8	39.4	47.2	92.5
Beta (0.5, 0.5)	\hat{N}_t	292	6.0	2.4	108.6	0
$\bar{p} = 0.507$	\hat{N}_{jk}	341	16.6	18.1	61.6 ^b	32.0
$CV = 0.713$	\hat{N}_{sc}	314	9.5	8.1	86.3	0
$M_{t+1} = 288$	\hat{N}_{th}	307	8.0	6.3	93.3	0

^a $N=400$, 200 simulation runs, 500 bootstrap replications, \hat{N}_t : the MLE under model \mathcal{M}_t ; \hat{N}_{jk} : the interpolated jackknife estimator; \hat{N}_{sc} : an estimator based on sample coverage approach; \hat{N}_{th} : the proposed estimator under model \mathcal{M}_{th} , see Eq. (2.13).

^bDenotes the smallest RMSE.

- \hat{N}_{sc} : a population size estimator based on the sample coverage method proposed by Chao et al. (1992) and Lee and Chao (1994). This estimator takes the form of $\hat{N}_{sc} = (M_{t+1} + f_{1t}\hat{\gamma}^2)/\hat{C}_t$, where $\hat{C}_t = 1 - [f_{1t} - f_{2t}/(t-1)]/\sum_{k=1}^t k f_{kt}$ and the CV estimator (i.e., $\hat{\gamma}$) under models \mathcal{M}_h and \mathcal{M}_{th} is provided in Lee and Chao (1994). Here the estimator \hat{C}_t is an average of the two previous estimators derived in Lee and Chao (1994).
- \hat{N}_h : the proposed estimator under model \mathcal{M}_h ; see Eq. (2.14).
- \hat{N}_{th} : the proposed estimator under model \mathcal{M}_{th} ; see Eq. (2.13).

The two estimators \hat{N}_h and \hat{N}_{th} behave very similarly. Therefore, we only report one of them in each table. Iterative procedures, which converged in all the data sets that we generated, were required to find \hat{N}_{th} . Bootstrap standard error estimates and 95% confidence intervals (using a log-transformation) were calculated using 500 replications for each data set. The resulting 200 estimates and their bootstrap estimated standard errors were averaged to give the results “average estimate” and “average bootstrap s.e.”. The sample standard deviation as well as the sample root-mean-squared error (RMSE) were also obtained based on these 200 estimates. The percentage of data sets in which 95% confidence intervals cover the true value is shown in the last column of each table. We also list the average value of M_{t+1} (number of distinct animals captured) for each trial.

As indicated in many previous studies, the estimator \hat{N}_t without considering heterogeneity consistently underestimates and the magnitude of the bias increases as the CV becomes large. Although it has small RMSE when the CV is relatively small, the coverage probabilities of the associated confidence intervals are much lower than the nominal level. When the CV is large, the estimated coverage probabilities are 0 in some cases, i.e., none of the intervals cover the true parameter.

The jackknife estimator has been found to be the most robust method for estimating heterogeneous populations (Otis et al., 1978; Pollock and Otto, 1983). For the case of five trapping samples and beta-type heterogeneity, it generally yields positive biases when a relative large proportion of animals is caught as in the first few cases examined. On the other hand, it is biased downwards when there are un-trappable animals, e.g., beta (0.5, 0.5). As noted in Chao et al. (1992), the jackknife works relatively well when the capture proportion is low as in the case of beta (2, 10). In the last three trials, the jackknife has the smallest RMSE when other estimators perform unsatisfactorily. See the discussion below.

Both tables show that our proposed estimators and the estimator using the sample coverage method generally exhibit comparable biases. Our estimators are more precise since they have a smaller s.e. in all cases. The proposed estimator \hat{N}_h or \hat{N}_{th} usually is preferable in terms of RMSE except for the last three trials.

For both estimators, the bootstrap estimate of the s.e.’s is quite close to the sample s.e. The 95% confidence intervals associated with \hat{N}_h or \hat{N}_{th} cover the true parameter about 87–97% of the time except for the cases of the uniform (0,1) and beta (0.5, 0.5) trials. These are the two cases where both our point and interval estimators fail

to work. In such cases, our estimators exhibit negative bias and the largest coverage probability for the 95% confidence intervals can only reach 10%. Note that the density function of a beta (α, β) , $0 < \alpha, \beta < 1$ and of the uniform $(0,1)$ does not vanish near 0, which means that there are animals with very small capture probabilities. That portion of the population is almost un-catchable in the data. We feel in these extreme cases that the data cannot contain sufficient information to estimate the population size well. Consequently, what we can estimate is essentially the number of “catchable” or “trappable” animals rather than the size of the whole population. Therefore, especially in the case of beta $(0.5, 0.5)$ where the density goes up to infinity near 0, no methods seem to work properly, although the jackknife yields the smallest RMSE.

We remark that a relatively large amount of data is needed for our approach. If catchability is low as in the case of beta $(2, 10)$, where the density does vanish near 0, large variation is unavoidable, although the associated confidence interval covers the parameter at the anticipated level. Therefore, when there is enough capture information and the number of un-catchable animals is negligible, our approach can provide more precise and reliable estimates for the beta-type heterogeneity.

5. Summary and discussion

For a sequence of closed models we have developed a unified approach for estimating population size based on estimating functions. An optimal property is that the resulting estimator has the smallest asymptotic variance in the class of weighted combinations of the capture data of all trapping samples. Our approach produces estimators for all models including the most complex one with time-varying, behavior response and heterogeneity effects.

Stanley and Burnham (1998) recently modified the model averaging approach (Buckland et al. 1997) and proposed an estimator averaging over all competing estimators. Therefore, having estimators for all possible models in a unified way as we have proposed would be also useful in the applications of the estimator averaging method.

As indicated in Section 2, our approach sequentially considers the information from each sample with a weight depending on the previous samples. Therefore, the resulting estimators do depend on the temporal order of the samples. This is not unreasonable because capture–recapture experiments for animals are usually conducted in sequential order. For example, the Schnabel’s estimator (Seber, 1982, p. 139) varies with the order in which the samples are taken. Moreover, models with behavioral effects (\mathcal{M}_b , \mathcal{M}_{bh} and \mathcal{M}_{tbh}) implicitly consider temporal order in the models as well. In the simplest model \mathcal{M}_b (or equivalently, the removal model), the MLE of the first-capture probability is $M_{t+1}/(tN - \sum_{k=1}^t M_k)$ where $\sum_{k=1}^t M_k = \sum_{i=1}^{t-1} (t-i)u_i$ clearly depends on the order of the samples. Thus the MLE of population size is not invariant under a permutation of the samples. Seber and Whale (1970) derived a necessary and sufficient condition for the existence of the MLE. The condition is reduced to $u_1 > u_2$ for two

samples and $u_1 > u_3$ for three samples. Apparently, these conditions also depend on the order. Therefore, it seems intuitively justifiable that our estimators depend on the order because of the sequential use of the optimal estimating function. The capture–recapture models have also been applied to human populations, e.g., the ascertainment data in epidemiology or the multiple-system data in the census undercount problem. There is usually no temporal order for human lists in those applications, and any valid estimator in such cases is then required to be independent of the way that the lists are ordered.

For models without heterogeneity, our estimator is shown to be asymptotically equivalent to the MLE in Chao et al. (2000). This paper extends the methodology to the heterogeneous populations for which no MLE exists. For heterogeneous models with/without time-varying effects, the simulation results have shown that our proposed estimators work if the number of un-catchable animals is insubstantial. As shown in the deer mouse data in Section 3, the most general model with time-varying, behavior response and heterogeneity can be applied in practice when there is enough capture and recapture information to allow the stable estimation of time, heterogeneity and behavior effects. Further comparisons of the proposed and known approaches using simulation studies are required. Model selection is also a worthwhile problem.

Acknowledgements

We would like to thank Dr. Richard Huggins and reviewers for their constructive suggestions and comments as well as generous editorial help which substantially improved the readability of this paper.

References

- Agresti, A., 1994. Simple capture-recapture models permitting unequal catchability and variable sampling effort. *Biometrics* 50, 494–500.
- Buckland, S.T., 1984. Monte Carlo confidence interval. *Biometrics* 40, 811–817.
- Buckland, S.T., Burnham, K.P., Augustin, N.H., 1997. Model selection: an integral part of inference. *Biometrics* 53, 603–618.
- Buckland, S.T., Garthwaite, P.H., 1991. Quantifying precision of mark-recapture estimates using the bootstrap and related methods. *Biometrics* 47, 255–268.
- Burnham, K.P., Overton, W.S., 1978. Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* 65, 625–633.
- Chao, A., 1987. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43, 783–791.
- Chao, A., Chu, W., Hsu, C.-H., 2000. Capture-recapture when time and behavioral response affect capture probabilities. *Biometrics* 56, 427–433.
- Chao, A., Lee, S.-M., Jeng, S.-L., 1992. Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics* 48, 201–216.
- Chao, A., Yip, P., Lee, S.-M., Chu, W., 1998. Population size estimation based on estimating functions for closed capture-recapture models. Technical Report, Institute of Statistics, National Tsing Hua University, Taiwan.
- Cormack, R.M., 1989. Log-linear models for capture-recapture. *Biometrics* 45, 395–413.

- Darroch, J.N., 1958. The multiple recapture census. I: estimation of a closed population. *Biometrika* 45, 343–359.
- Darroch, J.N., Fienberg, S.E., Glonek, G.F.V., Junker, B.W., 1993. A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. *J. Amer. Statist. Assoc.* 88, 1137–1148.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman & Hall, London.
- Lee, S.-M., 1996. Estimating population size for capture-recapture data when capture probabilities vary by time, behavior and individual animal. *Comm. Statist. B-Simulation Comput.* 25, 431–457.
- Lee, S.-M., Chao, A., 1994. Estimating population size via sample coverage for closed capture-recapture models. *Biometrics* 50, 88–97.
- Liang, K.-Y., Zeger, S.L., 1995. Inference based on estimating functions in the presence of nuisance parameters (with discussion). *Statist. Sci.* 10, 158–199.
- Lloyd, C.J., 1994. Efficiency of martingale methods in recapture studies. *Biometrika* 81, 305–315.
- Norris, J.L., Pollock, K.H., 1996. Nonparametric MLE under two closed capture-recapture models with heterogeneity. *Biometrics* 52, 639–649.
- Otis, D.L., Burnham, K.P., White, G.C., Anderson, D.R., 1978. Statistical inference from capture data on closed animal populations, *Wildlife Monographs*, Vol. 62, pp. 1–135.
- Pollock, K.H., 1991. Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future. *J. Amer. Statist. Assoc.* 86, 225–238.
- Pollock, K.H., Otto, M.C., 1983. Robust estimation of population size in closed animal populations from capture-recapture experiments. *Biometrics* 39, 1035–1049.
- Rexstad, E., Burnham, K.P., 1991. User's Guide for Interactive Program CAPTURE. Colorado Cooperative Fish and Wildlife Research Unit, Fort Collins.
- Schwarz, C.J., Seber, G.A.F., 1999. A review of estimating animal abundance III. *Statist. Sci.* 14, 427–456.
- Seber, G.A.F., 1982. *The Estimation of Animal Abundance and Related Parameters*, 2nd Edition. Griffin, London.
- Seber, G.A.F., 1986. A review of estimating animal abundance. *Biometrics* 42, 267–292.
- Seber, G.A.F., 1992. A review of estimating animal abundance II. *Internat. Statist. Rev.* 60, 129–166.
- Seber, G.A.F., Whaley, J.F., 1970. The removal method for two and three samples. *Biometrics* 26, 393–400.
- Stanley, T.R., Burnham, K.P., 1998. Estimator selection for closed-population capture-recapture. *J. Agric. Biol. Environ. Statist.* 3, 131–150.
- White, G.C., Anderson, D.R., Burnham, K.P., Otis, D.L., 1982. Capture-recapture and removal methods for sampling closed populations. Publication LA-8787-NERP. Los Alamos National Laboratory, Los Alamos, New Mexico.
- Yip, P., 1991. A martingale estimating equation for a capture-recapture experiment. *Biometrics* 47, 1081–1088.