

Capture–Recapture When Time and Behavioral Response Affect Capture Probabilities

Anne Chao,* Wenten Chu, and Chiu-Hsieh Hsu

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

*email: chao@stat.nthu.edu.tw

SUMMARY. We consider a capture–recapture model in which capture probabilities vary with time and with behavioral response. Two inference procedures are developed under the assumption that recapture probabilities bear a constant relationship to initial capture probabilities. These two procedures are the maximum likelihood method (both unconditional and conditional types are discussed) and an approach based on optimal estimating functions. The population size estimators derived from the two procedures are shown to be asymptotically equivalent when population size is large enough. The performance and relative merits of various population size estimators for finite cases are discussed. The bootstrap method is suggested for constructing a variance estimator and confidence interval. An example of the deer mouse analyzed in Otis et al. (1978, *Wildlife Monographs* **62**, 93) is given for illustration.

KEY WORDS: Bootstrap; Capture–recapture; Maximum likelihood; Maximum quasi-likelihood; Optimal estimating equation; Population size.

1. Introduction

We focus on inference for a closed capture–recapture model in which both behavioral response to capture and variation among sampling times exist. The model originally proposed by Pollock is usually referred to as model \mathcal{M}_{tb} in the literature, where the subscript t denotes time variation and b denotes behavioral response. (See Otis et al. [1978] for a review of related models.) The analysis of capture–recapture data can be performed using the comprehensive computer program CAPTURE (Otis et al., 1978; White et al., 1982; Rexstad and Burnham, 1991), which is readily available from Gary White's website (<http://www.cnr.colostate.edu/~gwhite/software.html>).

As indicated in Otis et al. (1978, p. 38), model \mathcal{M}_{tb} is not identifiable without further restrictions on the parameters. Section 2 briefly reviews the problem of nonidentifiability. There is relatively little reported research for this model. Burnham, in an unpublished manuscript, was the first to derive the maximum likelihood estimator (MLE) of population size under the assumption that recapture probability for any sample is a power function of initial capture probability (see Rexstad and Burnham, 1991, p. 13). This MLE has been implemented in the program CAPTURE.

Despite the nonidentifiability, model \mathcal{M}_{tb} has been selected as the most likely model for estimating some biological populations. Pollock et al. (1990, pp. 15–17) gave some examples for quail data sets that were originally presented in O'Brien et al. (1985). The model selection procedure provided in CAPTURE chose model \mathcal{M}_{tb} for three out of the eight quail data sets.

Otis et al. (1978, p. 93) provided another interesting data set on deer mice, and it will be used in Section 3 for illustration. The data were collected by S. Hoffman in 1974 and are given as an example with the program CAPTURE. The live-trapping experiment was conducted for five consecutive mornings. A total of 110 distinct mice were caught out of 283 captures. The data detail and analysis will be given in Section 3. Otis et al. (1978, p. 94) concluded that behavior is the strongest factor affecting capture probabilities and that time is the next most significant factor. However, the iterative steps for Burnham's estimation procedure when applied to the deer mouse example fail to obtain an estimator. It seems that a problem with Burnham's estimator is the possible failure of convergence even for nonsparse data. We were thus motivated to find the MLE under other conditions for model \mathcal{M}_{tb} .

Huggins (1991) and Stanley and Burnham (1998, 1999) considered a logistic relationship between first-capture probabilities and recapture probabilities. Lloyd (1994) developed a martingale method under the assumption that recapture probabilities and first-capture probabilities are related by a constant proportional parameter, with the constant independent of the animal. Using a model selection and goodness-of-fit test, we show in Section 3 that the special type of model \mathcal{M}_{tb} proposed by Lloyd (1994) is an adequate model for the deer mouse data. The purpose of this article is to provide ways to deal with this model. Our estimators would also be useful for applications where an estimator averaging approach (Stanley and Burnham, 1998) is adopted.

Under Lloyd's assumption, we derive in Section 2.1 the MLE of population size and its variance estimation. Two types of MLEs (unconditional and conditional) are discussed.

In Section 2.2, we present the maximum quasi-likelihood estimator (MQLE) on the basis of optimal estimating equations. Sanathanan (1972) verified that the two types of MLEs are asymptotically equivalent. In this article, we show that the MLEs and the MQLE are also asymptotically equivalent. Results of a simulation study to assess the relative merits of various estimators for finite sizes are reported in Section 4. As will be shown in Section 4, the iterative procedures for both MLEs and the MQLE converged for most trials in our simulation if a sufficiently large portion of animals was caught.

2. Models and Estimators

2.1 Maximum Likelihood Estimator

Assume that the population size is N and that there are t trapping samples. Let p_{ij} be the capture probability of the i th animal in the j th trapping sample. Model \mathcal{M}_{tb} assumes that the capture probabilities vary with time and with behavioral response, i.e., it assumes that $p_{ij} = p_j$ for a first capture and $p_{ij} = c_j$ for a recapture. There are a total of $2t$ parameters, i.e., $(N, p_1, p_2, \dots, p_t, c_2, c_3, \dots, c_t)$. Let n_j be the number of animals captured in the j th sample, u_j and m_j denote, respectively, the number of unmarked and marked animals in the j th sample, and $m. = \sum_{j=1}^t m_j$ denote the sum of m_j 's. Let $M_j = u_1 + u_2 + \dots + u_{j-1}$ be the number of marked animals in the population prior to the j th sample. Hence, M_{t+1} is the number of distinct animals captured in the experiment.

As shown in Otis et al. (1978, p. 38), the minimal sufficient statistic is $(u_1, u_2, \dots, u_t, m_2, m_3, \dots, m_t)$, which is of dimension $2t - 1$. Therefore, the model is nonidentifiable without any restriction. In this article, we assume that the recapture probability c_j is equal to a constant multiple of the initial capture probability p_j , i.e., $c_j/p_j = \phi$ for all $j = 2, \dots, t$ so that the parameters reduce to $(N, \phi, p_1, p_2, \dots, p_t)$, which has the dimension $t + 2$. This suffices to make the model identifiable if there are at least three trapping samples. The model with this assumption will be called simply model \mathcal{M}_{tb} hereafter.

Let X_ω be the number of animals with capture history ω , where ω denotes a nonempty subset of $\{1, 2, \dots, t\}$. For example, X_{124} is the number of animals captured in the first, second, and the fourth samples but not in others. Let P_ω denote the probability of the history ω and $Q = \prod_{j=1}^t (1 - p_j)$ be the probability of not being captured in the experiment. Then the distribution of the set of all possible capture histories $\{X_\omega\}$ is a multinomial and can be written as $P[\{X_\omega\}] = N![(N - M_{t+1})! \prod_\omega X_\omega!]^{-1} [\prod_\omega P_\omega^{X_\omega}] Q^{N - M_{t+1}}$. Letting $\mathbf{p} = (p_1, p_2, \dots, p_t)$, the likelihood function under the above assumption is given by (Otis et al., 1978, p. 111)

$$L = L(N, \phi, \mathbf{p}) \propto N(N - 1) \dots (N - M_{t+1} + 1) p_1^{u_1} (1 - p_1)^{N - u_1} \phi^{m.} \times \prod_{j=2}^t p_j^{n_j} (1 - p_j)^{N - M_{j+1}} (1 - \phi p_j)^{M_j - m_j}. \quad (2.1)$$

As shown in Sanathanan (1972), this likelihood can be factored as $L = L_1(N, Q)L_2(\phi, \mathbf{p})$, where L_1 denotes a binomial likelihood and L_2 denotes a multinomial likelihood. Here $L_1(N, Q) = N![(N - M_{t+1})!M_{t+1}]^{-1}(1 - Q)^{M_{t+1}} \times Q^{N - M_{t+1}}$ (the distribution of M_{t+1}) and $L_2(\phi, \mathbf{p}) = [M_{t+1}!/\prod_\omega X_\omega!]\{\prod_\omega [P_\omega/(1 - Q)]^{X_\omega}\}$ (the conditional dis-

tribution of $\{X_\omega\}$ given M_{t+1}). Note that only N and Q are involved in L_1 and N is not involved in L_2 . There are two types of MLEs (Sanathanan, 1972),

- (1) Unconditional MLE (UMLE): This is the usual MLE obtained by maximizing the full likelihood (2.1) simultaneously with respect to (N, ϕ, \mathbf{p}) .
- (2) Conditional MLE (CMLE): The CMLE of (ϕ, \mathbf{p}) is first computed by maximizing the conditional likelihood L_2 and obtaining an MLE \hat{Q} of Q ; the CMLE of N is then determined by maximizing $L_1(N, \hat{Q})$. The second part of maximization is equivalent to estimating the size in a binomial case for a given trial probability $1 - \hat{Q}$. Thus, the CMLE is the integer part of $M_{t+1}/(1 - \hat{Q})$ (cf., Dahiya, 1981). For easy interpretation, the nearly exact solution $M_{t+1}/(1 - \hat{Q})$ instead of its integer part (error is less than unity) will be used and referred to as the CMLE throughout this article.

In deriving the UMLE, the parameter N can be treated as either an integer or a real number. If N is treated as an integer, maximization is possible (Otis et al., 1978; Dahiya, 1981; Lindsay and Roeder, 1987) but the numerical manipulations become less tractable. Since the likelihood (2.1) is meaningful for any real N , we treat N as a real number in this article. For the UMLE, we thus can take derivatives with respect to N , ϕ , and p_j , respectively, and the UMLE satisfies the following system of $t + 2$ equations:

$$\frac{\partial \log L}{\partial N} = \sum_{j=1}^{M_{t+1}} (N - j + 1)^{-1} + \sum_{j=1}^t \log(1 - p_j) = 0, \quad (2.2a)$$

$$\frac{\partial \log L}{\partial \phi} = \frac{m.}{\phi} - \sum_{j=2}^t \frac{(M_j - m_j)p_j}{1 - \phi p_j} = 0, \quad (2.2b)$$

$$\frac{\partial \log L}{\partial p_j} = \frac{n_j}{p_j} - \frac{N - M_{j+1}}{1 - p_j} - \frac{(M_j - m_j)\phi}{1 - \phi p_j} = 0, \quad j = 1, 2, \dots, t. \quad (2.2c)$$

The solution is denoted by $(\hat{N}_{UMLE}, \hat{\phi}_{UMLE}, \hat{\mathbf{p}}_{UMLE})$. The matrix of the second derivatives is evaluated in a technical report by the authors (Chao, Chu, and Hsu, 1998). From (2.2c), we can express p_j as a function of N and ϕ , i.e., we have, for $j = 2, \dots, t$,

$$p_j = p_j(N, \phi) = \frac{1}{2N\phi} \left\{ [N + n_j + (\phi - 1)M_{j+1}] - \{ [N + n_j + (\phi - 1)M_{j+1}]^2 - 4N\phi n_j \}^{1/2} \right\} \quad (2.3)$$

and $p_1 = n_1/N$. In (2.3), the smaller root is chosen because it attains the maximum of the likelihood function. (The larger root attains the corresponding minimum.) After substituting (2.3) and $p_1 = n_1/N$ into (2.2), the UMLE $(\hat{N}_{UMLE}, \hat{\phi}_{UMLE})$ under model \mathcal{M}_{tb} is the solution of the equations (2.2a) and (2.2b), where $p_j = p_j(N, \phi)$ is given in (2.3). A numerical technique such as the Newton-Raphson method is required to obtain the solution. A drawback of the UMLE is that it is not scale invariant. This was first noted by Schnute (1983)

for the submodel \mathcal{M}_b . Examples are given in the Discussion for model \mathcal{M}_b and in Section 3.

To find the CMLE, we first maximize L_2 with respect to (ϕ, \mathbf{p}) . Note that $\partial \log L / \partial p_j = [(N - M_{t+1})/Q - M_{t+1}/(1 - Q)] \partial Q / \partial p_j + \partial \log L_2 / \partial p_j$. Under the restriction $N = M_{t+1}/(1 - Q)$, it is readily seen that $(N - M_{t+1})/Q - M_{t+1}/(1 - Q) = 0$. Hence, both $\partial \log L_2 / \partial \phi = \partial \log L / \partial \phi$ and $\partial \log L_2 / \partial p_j = \partial \log L / \partial p_j$ ($j = 1, 2, \dots, t$) are valid. This implies that the CMLE $(\hat{N}_{\text{CMLE}}, \hat{\phi}_{\text{CMLE}})$ satisfies the following system of equations:

$$N = M_{t+1}/(1 - Q) = M_{t+1} / \left[1 - \prod_{j=1}^t (1 - p_j) \right],$$

or equivalently,

$$1 - \frac{M_{t+1}}{N} = \prod_{j=1}^t (1 - p_j), \tag{2.4a}$$

$$\frac{\partial \log L_2}{\partial \phi} = \frac{m}{\phi} - \sum_{j=2}^t \frac{(M_j - m_j)p_j}{1 - \phi p_j} = 0, \tag{2.4b}$$

where $p_j = p_j(N, \phi)$ is given in (2.3).

It is interesting to notice that (2.4a) is also the equation gotten by taking a first finite difference of the likelihood function with respect to N . The two systems of equations corresponding to the two MLEs differ only in (2.2a) and (2.4a). These two equations are approximately the same when N is large because $\sum_{j=1}^{M_{t+1}} (N - j + 1)^{-1} \approx \int_{N - M_{t+1}}^N x^{-1} dx = \log N - \log(N - M_{t+1})$. Sanathanan (1972) proved that the CMLE is not less than the UMLE and that the two MLEs are asymptotically equivalent when N is large. For finite values of N , the difference between them might be large (see the deer mouse example in Section 3).

When $\phi = 1$, model \mathcal{M}_{tb} reduces to the special case that the capture probabilities only vary with time, i.e., model \mathcal{M}_t . Under this special model, it follows from (2.3) that $p_j(N) = n_j/N$. The usual UMLE by treating N as a real value satisfies the equation $\sum_{j=1}^{M_{t+1}} (N - j + 1)^{-1} + \sum_{j=1}^t \log(1 - n_j/N) = 0$. The CMLE satisfies the equation $1 - M_{t+1}/N = \prod_{j=1}^t (1 - n_j/N)$. Here the CMLE is derived in the sense of conditioning on M_{t+1} in the estimation. Note that the CMLE is also an MLE conditional on the sizes $\{n_1, n_2, \dots, n_t\}$ (Darroch, 1958; Seber, 1982, p. 131).

For the special model \mathcal{M}_b , which allows behavioral response to capture, we can let $p_1 = p_2 = \dots = p_t = p$ and $c_2 = \dots = c_t = c$ in the likelihood (2.1). Define $M = \sum_{j=1}^t M_j$. The equation for the UMLE reduces to $\sum_{j=1}^{M_{t+1}} (N - j + 1)^{-1} + t \log[1 - M_{t+1}/(tN - M)] = 0$, whereas the CMLE satisfies $1 - M_{t+1}/N = [1 - M_{t+1}/(tN - M)]^t$. The latter estimating equation was also used by Zippin (1956). This model is also referred to as a removal model in the literature. Note that both MLEs depend only on the first capture data $\{u_1, u_2, \dots, u_t\}$ because M_{t+1} and M are functions of $\{u_1, u_2, \dots, u_t\}$. Schnute (1983) found that, for data $\{u_1, u_2, u_3\} = \{90, 60, 40\}$, the UMLE is $\hat{N}_{\text{UMLE}} = 265.2$, whereas for data $\{u_1, u_2, u_3\} = \{9, 6, 4\}$, the UMLE becomes $\hat{N}_{\text{UMLE}} = 23.2$, not 1/10 of 265.2. This indicates

that the UMLE may not be scaled accordingly when data are scaled by a factor. The practical interpretation is the following: If each fish is a unit, then the data set of $\{90, 60, 40\}$ implies that the initial population size is 265 fishes. However, if fishes are caught and counted in schools of 10, then the same data set of $\{9, 6, 4\}$ (in a unit of 10) implies 23.2 units (or 232 fishes). We can theoretically verify that the CMLE (and the MQLE discussed below) are scale invariant. For the above data sets, the CMLE (and the MQLE) is 270 for the former data and 27 for the latter data.

Under the regularity conditions specified in Sanathanan (1972), we can conclude that both MLEs are asymptotically normal and consistent if $N \rightarrow \infty$. The consistency here means that the estimator divided by N converges to one with probability one. Under our model, the asymptotic variance for both types of MLEs can be shown to be

$$N\phi \sum_{k=2}^t Q_{k-1}^2 A_k / \sum_{i=2}^{t-1} \sum_{j=i+1}^t A_i A_j (Q_{i-1} - Q_{j-1})^2, \tag{2.5}$$

where $A_k = [(1 - Q_{k-1})\phi p_k] / \{Q_{k-1}[q_k + (\theta - 1)Q_{k-1}]\}$, $q_k = 1 - p_k$, $\theta = 1/\phi$, and $Q_k = q_1 q_2 \dots q_k$. Replacing N , ϕ , p_k , and q_k in (2.5) by their estimates, we then obtain variance estimators based on the asymptotic variance formula. Simulation results have suggested that the asymptotic variance estimator generally works well when there are enough data, whereas it overestimates if the mean capture rate is relatively low. (See Section 4 for details.) 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 nonparametric bootstrap to obtain an alternative variance estimator. Let Ω denote the collection of all subsets of $\{1, 2, \dots, t\}$ and X_\emptyset (\emptyset denotes an empty set) denote the missing cell. The capture histories $\{X_\omega \mid \omega \in \Omega\}$, where X_ω is defined earlier in this section, are distributed as a multinomial distribution.

Suppose that we want to obtain a bootstrap variance for any estimator \hat{N} on the basis of given observable capture histories $\{X_\omega \mid \omega \in \Omega, \omega \neq \emptyset\}$. A bootstrap sample $\{X_\omega^* \mid \omega \in \Omega\}$ is generated from a multinomial distribution with parameter \hat{N} and cell probabilities X_ω/\hat{N} for any observable cell and $1 - \sum X_\omega/\hat{N}$ for the missing cell. Then, based on the generated observed data $\{X_\omega^* \mid \omega \in \Omega, \omega \neq \emptyset\}$, 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. The percentile method can then be used to construct an associated confidence interval based on the same B bootstrap estimates (see Efron and Tibshirani, 1993, Chapter 13). The performance of the bootstrap variance estimator and the confidence interval will be discussed in Section 4.

2.2 Maximum Quasi-Likelihood Estimator

We now present the optimal estimating function approach. An estimating function is a function of the parameter and data. An estimator of the parameter can be obtained as its root. An unbiased estimating function satisfies the expectation that the function is zero for all parameters. Let E_c , var_c , and cov_c

denote the expectation, variance, and covariance by holding the first $j - 1$ samples fixed. Given the capture history of the first $j - 1$ samples, u_j is a binomial random variable with parameters $(N - M_j)$ and p_j , and m_j is a binomial with parameters M_j and ϕp_j . Thus, for each sample j , we can construct the following unbiased estimating function ($m_1 = 0, M_1 = 0$):

$$g_j = \begin{bmatrix} g_{1j} \\ g_{2j} \end{bmatrix} = \begin{bmatrix} u_j - (N - M_j)p_j \\ m_j - M_j\phi p_j \end{bmatrix}. \tag{2.6}$$

We consider a linear combination of g_1, g_2, \dots, g_t into a single estimating function, and the weight associated with g_j is allowed to be a function of the parameters and the history of the previous $j - 1$ samples. The optimal estimating equation is given by (see Godambe and Heyde [1987] or Liang and Zeger [1995] for necessary background) $g = \sum_{j=1}^t D_j^T V_j^{-1} g_j = \mathbf{0}$, where D_j^T denotes the transpose of D_j and

$$D_j = E_c \begin{bmatrix} \frac{\partial g_{1j}}{\partial N} & \frac{\partial g_{1j}}{\partial \phi} & \frac{\partial g_{1j}}{\partial \mathbf{p}} \\ \frac{\partial g_{2j}}{\partial N} & \frac{\partial g_{2j}}{\partial \phi} & \frac{\partial g_{2j}}{\partial \mathbf{p}} \end{bmatrix}, \tag{2.7}$$

$$V_j = \begin{bmatrix} \text{var}_c(g_{1j}) & \text{cov}_c(g_{1j}, g_{2j}) \\ \text{cov}_c(g_{1j}, g_{2j}) & \text{var}_c(g_{2j}) \end{bmatrix},$$

and $\mathbf{p} = (p_1, p_2, \dots, p_t)$. Note that D_j is a matrix with dimension $2 \times (t+2)$. Substituting $\text{var}_c(g_{1j}) = (N - M_j)p_j(1 - p_j)$, $\text{var}_c(g_{2j}) = M_j\phi p_j(1 - \phi p_j)$, and $\text{cov}_c(g_{1j}, g_{2j}) = 0$ (u_k and m_k are conditionally independent) into (2.7), we obtain the following $t + 2$ equations for g :

$$\sum_{j=1}^t [(N - M_j)(1 - p_j)]^{-1} [u_j - (N - M_j)p_j] = 0, \tag{2.8a}$$

$$\sum_{j=2}^t [\phi(1 - \phi p_j)]^{-1} [m_j - M_j\phi p_j] = 0, \tag{2.8b}$$

$$\frac{u_j - (N - M_j)p_j}{1 - p_j} + \frac{m_j - M_j\phi p_j}{1 - \phi p_j} = 0, \tag{2.8c}$$

$j = 1, 2, \dots, t.$

The solution is designated as $(\hat{N}_{MQLE}, \hat{\phi}_{MQLE}, \hat{\mathbf{p}}_{MQLE})$. Note that a constant $1/p_j$ is dropped in (2.8c) without affecting the solution. These types of optimal equations are termed quasi-likelihood equations in Godambe and Heyde (1987). Thus, the solution of (2.8) will be referred to as the maximum quasi-likelihood estimator (MQLE).

Based on (2.8c), we can express p_j as a function of N and ϕ . The resulting function can be shown to be exactly the same as that in (2.3). It follows from (2.8c) that equations (2.8a) and (2.8b) are equivalent to

$$\sum_{j=2}^t \{(N - M_j)[(M_j(\phi - 1) + N(1 - \phi p_j))]\}^{-1} \times [\phi M_j u_j - (N - M_j)m_j] = 0, \tag{2.9a}$$

$$\sum_{j=2}^t [(M_j(\phi - 1) + N(1 - \phi p_j))^{-1} \times [\phi M_j u_j - (N - M_j)m_j] = 0, \tag{2.9b}$$

where $p_j = p_j(N, \phi)$ is given in (2.3). The above estimating equations are identical to those derived in Lloyd (1994), who suggested the use of $p_j = u_j/(N - M_j)$ in solving the equations. Our expression for p_j , given in (2.3), which uses both first-capture and recapture data, is more complex, but it is derived from a standpoint of optimal estimating functions.

For the special case of model \mathcal{M}_t , we let $\phi = 1$ in the derivation. Thus, (2.3) and (2.9a) imply that $p_j = n_j/N$ and the MQLE satisfies $\sum_{j=2}^t \{(N - M_j)(N - n_j)\}^{-1} [M_j u_j - (N - M_j)m_j] = 0$. The above result was first derived by Yip (1991). For the special case of model \mathcal{M}_b , we have $p_1 = p_2 = \dots = p_t = p$ and $c_2 = \dots = c_t = c$, and the unbiased estimating functions become $g_{1j} = u_j - (N - M_j)p$ and $g_{2j} = m_j - M_j c$. Therefore, the resulting MQLEs are the same as those derived in Lloyd (1994).

Under the conditions $\phi > 0$ and $0 < p_k < 1$ for $k = 1, 2, \dots, t$, the asymptotic normality and consistency of the MQLE as $N \rightarrow \infty$ is proved in a previously cited technical report by the authors. The asymptotic variance-covariance matrix of the solution $(\hat{N}_{MQLE}, \hat{\phi}_{MQLE}, \hat{\mathbf{p}}_{MQLE})$ is approximately the inverse of the matrix $E(\sum_{j=1}^t D_j^T V_j^{-1} D_j)$, where D_j and V_j are defined in (2.7). The asymptotic variances of the MQLE and the MLEs under model \mathcal{M}_{tb} turn out to be identical. Thus, an estimated variance estimator can be computed using (2.5). The bootstrap method proposed for the MLEs can be similarly applied to the MQLE as well.

3. Real Data Example

The deer mouse data considered in this section were originally collected by S. Hoffman and analyzed in Otis et al. (1978, p. 93). These data are given as an illustrative example with the program CAPTURE (Rexstad and Burnham, 1991). There were five trapping occasions; 110 distinct mice were caught out of 283 captures. The numbers of captures for the five capture occasions were $(n_1, \dots, n_5) = (37, 54, 58, 65, 69)$, and the numbers of unmarked caught were $(u_1, \dots, u_5) = (37, 31, 9, 21, 12)$. Hence, we have $(M_2, \dots, M_6) = (37, 68, 77, 98, 110)$.

If model \mathcal{M}_t is assumed, the UMLE of N is 112.6, which is only slightly higher than the number of distinct mice captured, 110. Assuming the underlying model is model \mathcal{M}_b , Otis et al. (1978) obtained the first-capture probability estimate of 0.26 and recapture probability estimate of 0.62. Thus, the animals showed a trap-happy behavioral response to capture. The resulting UMLE of the population size under model \mathcal{M}_b is 142 with an estimated SE of 16.4, which is used to construct an approximate 95% confidence interval of (109, 175).

Otis et al. (1978, p. 93) commented that neither model \mathcal{M}_t nor model \mathcal{M}_b fits the data using a goodness-of-fit (GOF) test. If we consider both time variation and behavioral response by adopting model \mathcal{M}_{tb} , the program CAPTURE fails to obtain an estimator under the assumption that recapture probabilities are a power function of initial capture probabilities. Assuming a constant relation between recapture probabilities and initial capture probabilities, we have $\hat{N}_{UMLE} = 161.1$, $\hat{\phi}_{UMLE} = 3.19$, and $\hat{\mathbf{p}}_{UMLE} = (0.23, 0.21, 0.21, 0.19, 0.18)$ based on (2.2). We first test the hypothesis of model \mathcal{M}_b versus model \mathcal{M}_{tb} using a likelihood ratio test (LRT); the test has an approximately chi-square distribution with 4 d.f. The hypothesis is rejected since the

chi-squared-based LRT yields a value of 66.38. Similarly, the LRT is 22.86 for testing the hypothesis of model \mathcal{M}_t versus model \mathcal{M}_{tb} . Hence, model \mathcal{M}_{tb} provides significant improvement over both model \mathcal{M}_t and model \mathcal{M}_b in maximizing the likelihood.

The usual Pearson's type of GOF test cannot be applied to these data because a large-scale pooling would be involved. Stanley and Burnham (1999) suggested a GOF test for model \mathcal{M}_{tb} based on the statistic $\chi_t^2 - \chi_{t/tb}^2$, where χ_t^2 denotes a GOF test for model \mathcal{M}_t and $\chi_{t/tb}^2$ denotes the chi-square-based LRT between model \mathcal{M}_t and model \mathcal{M}_{tb} . A widely used GOF test for model \mathcal{M}_t is Leslie's test (Leslie, 1958). Otis et al. (1978, p. 92) obtained the Leslie's GOF test of 84.12 with 66 d.f. Hence, a GOF test value for model \mathcal{M}_{tb} is $84.12 - 22.86 = 61.26$. Based on the chi-squared distribution with 65 d.f., the P -value of this GOF test is 0.61. Therefore, there is insufficient evidence to reject model \mathcal{M}_{tb} . The two MLEs and the MQLE are given below. The bootstrap SE for each estimator was based on 10,000 replications. The same 10,000 bootstrap estimates yielded a 95% confidence interval (CI) using the percentile method. The asymptotic standard errors were computed from equation (2.5).

Method	\hat{N}	Bootstrap SE	Asymptotic SE	95% Interval	$\hat{\phi}$
UMLE	161.1	42.79	41.72	(121, 283)	3.19
CMLE	173.7	45.25	55.69	(124, 289)	3.63
MQLE	152.0	29.94	32.87	(119, 229)	2.87

This case study provides an example that the UMLE and the CMLE may produce quite different results. Except for the CMLE, the asymptotic SE estimate for each of the other two estimators is close to the corresponding bootstrap SE estimate. All three values of $\hat{\phi}$ show strong evidence for a trap-happy case.

For each animal, if we add four additional animals with the same capture history into the data, then all summary statistics become five times the original data. For these expanded data sets, we have $\hat{N}_{CLME} = 868.5 = 5 \times 173.7$, $\hat{N}_{MQLE} = 760 = 5 \times 152$, which indicates that the CMLE and MQLE change accordingly when the data are scaled by a factor. However, $\hat{N}_{UMLE} = 854.4$, which is not five times 161.1. Hence, a drawback of the UMLE is that its estimates depend on the choice of the scale.

A computer program, written in the C language, that calculates all the 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 Results

We have carried out a limited simulation study to examine the performance of the proposed estimation procedures. Three values of N (100, 200, 400), two values of ϕ (1.5, 0.8), and $t = 5$ were selected. Here we only report the results for $N = 400$. Other results for $N = 100$ and 200 are given in a technical report by the authors (Chao, Chu, and Hsu, 1998).

The eight trials of model \mathcal{M}_{tb} are given below. The p_j 's in these eight trials were selected so that the expected number

of animals captured, $E(M_{t+1})$, increases from 200 to 375 in approximate increments of 25.

- Trial 1: $(p_1, p_2, p_3, p_4, p_5) = (0.17, 0.15, 0.12, 0.1, 0.12)$,
 $E(M_{t+1}) = 203$;
- Trial 2: $(p_1, p_2, p_3, p_4, p_5) = (0.12, 0.16, 0.20, 0.12, 0.16)$,
 $E(M_{t+1}) = 225$;
- Trial 3: $(p_1, p_2, p_3, p_4, p_5) = (0.2, 0.17, 0.25, 0.15, 0.12)$,
 $E(M_{t+1}) = 251$;
- Trial 4: $(p_1, p_2, p_3, p_4, p_5) = (0.20, 0.20, 0.15, 0.24, 0.24)$,
 $E(M_{t+1}) = 274$;
- Trial 5: $(p_1, p_2, p_3, p_4, p_5) = (0.30, 0.25, 0.15, 0.20, 0.30)$,
 $E(M_{t+1}) = 300$;
- Trial 6: $(p_1, p_2, p_3, p_4, p_5) = (0.23, 0.28, 0.35, 0.31, 0.25)$,
 $E(M_{t+1}) = 325$;
- Trial 7: $(p_1, p_2, p_3, p_4, p_5) = (0.38, 0.22, 0.34, 0.3, 0.44)$,
 $E(M_{t+1}) = 350$;
- Trial 8: $(p_1, p_2, p_3, p_4, p_5) = (0.45, 0.4, 0.4, 0.4, 0.45)$,
 $E(M_{t+1}) = 374$.

Table 1 presents the simulation comparison for the trap-happy cases only. The relative merits of various estimators are generally consistent for the trap-shy cases. For each fixed (p_1, p_2, \dots, p_5) and ϕ , 1000 sets of capture-recapture data were simulated. We mainly focused on the three population size estimators under model \mathcal{M}_{tb} : UMLE (\hat{N}_{UMLE}), CMLE (\hat{N}_{CMLE}), and MQLE (\hat{N}_{MQLE}). For each estimator, we computed (1) the point estimate, (2) the corresponding asymptotic SE estimate (see (2.5)), (3) the bootstrap SE estimate using 500 bootstrap replications for each data set, and (4) the 95% confidence interval using the percentile method based on the same set of bootstrap replications in (3).

We excluded those data sets for which the iterative steps did not converge for any estimator in our simulation. The procedure continued until 1000 data sets had been generated for which the iterative steps did converge for all estimators. Here divergence means that either iterations increase without a limit (the upper bound was set to be $7M_{t+1}$) or iterations bounce around without reaching a stable value in 1000 steps.

For the 1000 generated data sets, estimates and their asymptotic standard error estimates as well as bootstrap SE estimates were averaged to give the results average estimate, average asymptotic SE, and average bootstrap SE in Table 1. The sample standard error (with the heading of sample SE in the tables) and sample root mean squared error (with the heading of sample RMSE) were also obtained based on the resulting 1000 estimates. We also list the average number of distinct captured animals (M_{t+1}) for each trial. The percentage of 1000 simulated data sets in which the 95% confidence intervals covered the true value (400) was recorded. The coverage probabilities are listed under the heading bootstrap CI coverage. Based on Table 1 and other unreported results, we summarize below the performance of the various estimators.

The three estimators (UMLE, CMLE, and MQLE) are generally comparable. The CMLE and the MQLE exhibit

Table 1
*Comparison of various estimators for trap-happy cases ($\phi = 1.5$);
 $N = 400, 1000$ simulation runs, 500 bootstrap replications*

Trial (M_{t+1} in parentheses)	Estimator	Average estimate ^a	Sample SE	Average asymptotic SE	Average bootstrap SE	Sample RMSE	Bootstrap CI coverage (%) ^b
1 (203)	\hat{N}_{UMLE}	305	58.0	102.2	74.7	111.5	94
	\hat{N}_{CMLE}	344	84.3	164.5	78.1	101.2	96
	\hat{N}_{MQLE}	337	79.4	151.5	72.3	101.6	91
2 (226)	\hat{N}_{UMLE}	354	77.6	113.5	88.7	90.3	94
	\hat{N}_{CMLE}	387	102.9	159.8	93.2	103.7	97
	\hat{N}_{MQLE}	381	97.5	149.8	87.4	99.4	95
3 (252)	\hat{N}_{UMLE}	382	76.5	100.0	82.3	78.7	92
	\hat{N}_{CMLE}	407	98.7	132.4	86.0	99.0	95
	\hat{N}_{MQLE}	402	95.2	125.5	80.5	95.2	93
4 (275)	\hat{N}_{UMLE}	390	69.8	83.4	82.9	70.5	94
	\hat{N}_{CMLE}	411	86.7	106.0	87.8	87.4	96
	\hat{N}_{MQLE}	406	83.0	100.7	80.8	83.2	96
5 (300)	\hat{N}_{UMLE}	400	63.9	68.1	67.0	63.9	93
	\hat{N}_{CMLE}	417	77.9	84.6	70.6	79.9	96
	\hat{N}_{MQLE}	413	74.2	80.4	63.7	75.4	96
6 (326)	\hat{N}_{UMLE}	403	38.1	35.0	41.9	38.2	92
	\hat{N}_{CMLE}	409	41.3	38.1	45.0	42.4	92
	\hat{N}_{MQLE}	407	40.4	37.1	42.5	41.0	93
7 (351)	\hat{N}_{UMLE}	401	29.3	26.0	31.7	29.3	93
	\hat{N}_{CMLE}	406	32.7	29.0	34.9	33.3	94
	\hat{N}_{MQLE}	405	31.9	28.3	30.5	32.3	94
8 (375)	\hat{N}_{UMLE}	400	11.6	11.2	12.4	11.6	96
	\hat{N}_{CMLE}	402	12.2	11.9	13.3	12.4	95
	\hat{N}_{MQLE}	401	12.0	11.7	12.8	12.1	97

^a The SE of any average estimate is approximately the sample SE divided by $(1000)^{1/2}$.

^b The SE of any coverage probability is approximately $(p \times (1 - p)/1000)^{1/2}$, where p denotes the estimated coverage.

similar behavior. (Thus, we treat them as one group of estimators.) The value of the CMLE is consistently higher than that of the UMLE (a theoretical property proved by Sanathanan [1972]); the MQLE, on average, is somewhere in between (a numerical finding).

The UMLE generally increases to the true population size as the capture proportion tends to one. It has the lowest variance and RMSE, but it exhibits large negative bias when the mean capture rate is low, especially in the trap-shy cases (unreported). Both the CMLE and the MQLE first increase with the capture fraction, cross the true parameter, and then decrease to approach the population size. Hence, the direction of bias depends on the capture proportion. When the capture proportion is relatively low, as in trials 1–4, the CMLE and MQLE are preferable to the UMLE in terms of bias, although the former two estimators have larger variation.

The asymptotic variance formula given in (2.5) yields an adequate point estimator only for trials 6–8. When there is insufficient capture information, as in other trials, the asymptotic

estimated SE severely overestimates, especially for the CMLE. We generally recommend the use of the bootstrap method because the bootstrap SE is closer to the sample SE in most cases. The confidence intervals using the percentile method for each estimator perform reasonably well as regards coverage probability.

For trials 5–8, in which there were enough captures, convergence occurred for almost all data sets. When population size was increased, the divergence rate decreased as more data became available. The iterative steps did diverge for some trials. In the trap-happy cases, the divergence rates based on 1000 simulation runs for the UMLE for trials 1–4 were 25, 14, 7, and 4% and the corresponding rates for both the CMLE and the MQLE were about 33, 19, 9, and 6%.

In summary, we have developed for a special type of model \mathcal{M}_{tb} the unconditional MLE (UMLE), the conditional MLE (CMLE), and the maximum quasi-likelihood estimator (MQLE). The three estimators are asymptotically equivalent for large population sizes and are generally comparable for fi-

nite sizes. The CMLE and the MQLE behave quite similarly. These two estimators are scale invariant, but the UMLE is not. When there is a sufficient number of captures, the UMLE is preferable to the other two estimators, but it yields large negative bias otherwise. When the mean catchability is relatively low, the CMLE and the MQLE are each less biased than the UMLE. The bootstrap method (Buckland, 1984; Buckland and Garthwaite, 1991; Norris and Pollock, 1996) for constructing variance estimators and confidence intervals is recommended for practical use.

ACKNOWLEDGEMENTS

The authors thank the associate editor and two referees for carefully reading an earlier version of this article and for providing thoughtful comments. This research was supported by the National Science Council of Taiwan under contracts 85-2121-M007-004 and 86-2115-M007-033.

RÉSUMÉ

Nous étudions un modèle de capture-recapture dans lequel les probabilités de capture varient en fonction du temps et de la réponse comportementale. Nous développons deux procédures inférentielles, sous l'hypothèse que le rapport entre les probabilités de recapture et les probabilités de capture initiale est constant. Ces deux procédures sont la méthode du maximum de vraisemblance (conditionnelle ou inconditionnelle), et une méthode basée sur des fonctions d'estimation optimales. Nous montrons que les estimateurs de l'effectif de la population obtenus par les deux procédures sont asymptotiquement équivalents quand la population est assez grande. Nous comparons les performances et les qualités de divers estimateurs d'effectifs de population pour des cas finis. Nous suggérons l'utilisation du bootstrap pour construire un estimateur de la variance et un intervalle de confiance. Nous traitons l'exemple de la population de souris *Peromyscus* analysée par Otis et al. (1978, *Wildlife Monographs* **62**, p. 93).

REFERENCES

- Buckland, S. T. (1984). Monte Carlo confidence intervals. *Biometrics* **40**, 811–817.
- Buckland, S. T. and Garthwaite, P. H. (1991). Quantifying precision of mark-recapture estimates using the bootstrap and related methods. *Biometrics* **47**, 255–268.
- Chao, A., Chu, W., and Hsu, C.-H. (1998). *Capture-recapture when time and behavioral response affect capture probabilities*. Technical Report, Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan.
- Dahiya, R. C. (1981). An improved method of estimating an integer-parameter by maximum likelihood. *The American Statistician* **35**, 34–37.
- Darroch, J. N. (1958). The multiple recapture census. I: Estimation of a closed population. *Biometrika* **45**, 343–359.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Godambe, V. P. and Heyde, C. C. (1987). Quasi-likelihood and optimal estimation. *International Statistical Review* **55**, 231–244.
- Huggins, R. M. (1991). Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics* **47**, 725–732.
- Leslie, P. H. (1958). Statistical appendix. *Journal of Animal Ecology* **27**, 84–86.
- Liang, K.-Y. and Zeger, S. L. (1995). Inference based on estimating functions in the presence of nuisance parameters (with discussion). *Statistical Science* **10**, 158–199.
- Lindsay, B. G. and Roeder, K. (1987). A unified treatment of integer parameter models. *Journal of the American Statistical Association* **82**, 758–764.
- Lloyd, C. J. (1994). Efficiency of martingale methods in recapture studies. *Biometrika* **81**, 305–315.
- Norris, J. L. and Pollock, K. H. (1996). Including model uncertainty in estimating variances in multiple capture studies. *Environmental and Ecological Statistics* **3**, 233–244.
- O'Brien, T. G., Pollock, K. H., Davidson, W. R., and Kellogg, F. E. (1985). A comparison of capture-recapture with capture-removal for quail populations. *Journal of Wildlife Management* **49**, 1062–1065.
- Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs* **62**, 1–135.
- Pollock, K. H., Nichols, J. D., Brownie, C., and Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife Monographs* **107**, 1–97.
- Rexstad, E. and Burnham, K. P. (1991). *User's Guide for Interactive Program CAPTURE*. Fort Collins: Colorado Cooperative Fish and Wildlife Research Unit.
- Sanathanan, L. (1972). Estimating the size of a multinomial population. *The Annals of Mathematical Statistics* **43**, 142–152.
- Schnute, J. (1983). A new approach to estimating populations by the removal method. *Canadian Journal of Fisheries and Aquatic Sciences* **40**, 2153–2169.
- Seber, G. A. F. (1982). *The Estimation of Animal Abundance*, 2nd edition. London: Griffin.
- Stanley, T. R. and Burnham, K. P. (1998). Estimator selection for closed-population capture-recapture. *Journal of Agricultural, Biological, and Environmental Statistics* **3**, 131–150.
- Stanley, T. R. and Burnham, K. P. (1999). A goodness-of-fit test for capture-recapture model M_t under closure. *Biometrics* **55**, 366–375.
- White, G. C., Anderson, D. R., Burnham, K. P., and Otis, D. L. (1982). *Capture-recapture and removal methods for sampling closed populations*. Report LA-8787-NERP, Los Alamos National Laboratory, Los Alamos, New Mexico.
- Yip, P. (1991). A martingale estimating equation for a capture-recapture experiment in discrete time. *Biometrics* **47**, 1081–1088.
- Zippin, C. (1956). An evaluation of the removal method of estimating animal populations. *Biometrics* **12**, 163–189.

Received December 1997. Revised September 1999.

Accepted September 1999.