Environmental and Ecological Statistics **6,** 313–334 (1999)

An estimating function approach to the inference of catch-effort models

ANNE CHAO[1] and SHU-HUI CHANG[2]

[1]*Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043*
[2]*Department of Public Health, National Taiwan University, Taipei, Taiwan 10018*

Received May 1997; Revised February 1999

A class of catch-effort models, which allows for heterogeneous removal probabilities, is proposed for closed populations. The model includes three types of removal probabilities: multiplicative, Poisson and logistic. The usual removal and generalized removal models then become special cases. The equivalence of the proposed model and a special type of capture-recapture model is discussed. A unified estimating function approach is used to estimate the initial population size. For the homogeneous model, the resulting population size estimator based on optimal estimating functions is asymptotically equivalent to the maximum likelihood estimator. One advantage for our approach is that it can be extended to handle the heterogeneous populations in which the maximum likelihood estimators do not exist. The bootstrap method is applied to construct variance estimators and confidence intervals. We illustrate the method by two real data examples. Results of a simulation study investigating the performance of the proposed estimation procedure are presented.

Keywords: capture-recapture, heterogeneity, population size, removal model, sample coverage

1. Introduction

The capture-recapture and catch-effort models have been widely applied to estimate the population size of a closed population. In a typical capture-recapture experiment, we record and attach a unique tag to every unmarked animal, record any recapture, and return all animals to the population on each trapping sample. Thus both capture and recapture information is obtained. In a catch-effort model, the captured animals are removed from the population and there is no recapture information in the experiment. As will be discussed, there is a close relationship between the two models. The analysis of some capture-recapture models, which use only the first-capture data to estimate population size, is exactly the same as that of the catch-effort models.

This paper focuses on the catch-effort models. Catch-effort methods are mainly used for fish, small mammals and insect populations. Usually, the effort is measured in line, trap or fishing effort per unit time. The most common applications are killtrapping of mammals, electrofishing in ponds and trawling in oceans. The technique removes samples sequentially from the population and uses the decline in catch size and the known efforts to estimate initial population size. See Seber (1982, Chapter 7 and Section 12.9), Schnute

1352-8505 © 1999 Kluwer Academic Publishers

(1983) and Pollock (1981, 1991) for a general review on this topic. Table 1 lists all estimators and models that will be discussed in this paper.

The constant effort model is usually referred to as the removal model (see Otis *et al.* (1978)) and has been studied extensively in the literature. In the simplest removal model, all the individuals of the population have the same removal probability (homogeneous assumption), and this probability remains constant over samples. This is the model most often encountered in the literature. For example, Moran (1951) and Zippin (1956, 1958) derived the maximum likelihood estimator (MLE) of the initial population size and its asymptotic variance. Seber and Whale (1970) discussed the special cases of two or three samples and the effect of variable catchability on the usual MLE. Schnute (1983) presented two models that relax the assumption of constant catchability over samples.

Table 1. Various models and their estimators.

Model	*Estimator*
Homogeneous Constant Effort Model:	
Removal Model: $p_{ij} = \lambda$	MLE (Moran (1951); Zippin (1956, 1958); Seber and Whale (1971); Schnute (1983))
Heterogeneous Constant Effort Model:	
Generalized Removal Model: $p_{ij} = \lambda_i$	Generalized removal (Otis *et al.* (1978)) Sample coverage (Lee and Chao (1994)); see Equation (2.23) Jackknife (Pollock and Otto (1983)) Proposed MQLE; see Equation (2.22)
Homogeneous Variable Effort Model:	
(1) Multiplicative: $p_{ij} = \lambda e_j$	Regression (DeLury (1958); Ricker (1958); Leslie (1958)) MLE (Gould and Pollock (1997)); see Equation (2.8) Sample coverage (Lee and Chao (1994)); see Equation (2.13) Proposed MQLE; see Equation (2.5)
(2) Poisson: $p_{ij} = 1 - \exp(-\lambda e_j)$	Regression (DeLury (1958); Ricker (1958); Leslie (1958)) MLE (Gould and Pollock (1997); Seber (1982)); see Equation (2.8) Proposed MQLE; see Equation (2.5)
(3) Logistic: $p_{ij} = \exp(\lambda + \beta e_j)/[1 + \exp(\lambda + \beta e_j)]$	MLE (Gould and Pollock (1997)); see Equation (2.8) Proposed MQLE; see Equations (2.5) and (2.6)
Heterogeneous Variable Effort Model:	
(1) Multiplicative: $p_{ij} = \lambda_i e_j$	Sample coverage (Lee and Chao (1994)); see Equation (2.23) Proposed MQLE; see Equation (2.22)
(2) Poisson: $p_{ij} = 1 - \exp(-\lambda_i e_j)$	Proposed MQLE ; see Equation (2.22)
(3) Logistic: $p_{ij} = \exp(\lambda_i + \beta e_j)/[1 + \exp(\lambda_i + \beta e_j)]$	No estimator

Pollock, in his 1974 Ph.D. thesis and subsequent papers (e.g., Pollock (1991)) proposed a sequence of models for capture-recapture experiments. Three basic models are (a) model $\boldsymbol{M}_t$, which allows capture probabilities to vary by time; (b) model $\boldsymbol{M}_b$, which allows behavioral responses to capture; and (c) model $\boldsymbol{M}_h$, which allows heterogeneous animal capture probabilities. Various combinations of these three types of capture probabilities (i.e., models $\boldsymbol{M}_{tb}$, $\boldsymbol{M}_{th}$, $\boldsymbol{M}_{bh}$, $\boldsymbol{M}_{tbh}$) are also considered.

We now discuss the connection between removal models and capture-recapture models. In a capture-recapture model $\boldsymbol{M}_b$, it is assumed that all first-captures (i.e., unmarked animals) have one probability of capture, and all recaptures (i.e., marked animals) have another probability of capture. As shown by Otis *et al.* (1978, Appendix D), the recapture information is used only in the estimation of the probability of recapture. Only the first-capture information on each trapping occasion is used in terms of estimation of population size. The estimation procedure for model $\boldsymbol{M}_b$ is exactly the same as that of the removal model. Therefore, from the standpoint of inference, removal by marking is statistically equivalent to permanent removal. In model $\boldsymbol{M}_b$, the number of first-captures (unmarked) can be conceptually regarded as ''removed by being marked'' because any recapture of these animals does not contain any information about population size.

A generalized removal model relaxes the assumption of the homogeneous catchability among individuals. It is referred to as a heterogeneous constant effort model; see Table 1. A heterogeneous model assumes that each animal has its own unique catchability, which remains constant over samples. For this model, Otis *et al.* (1978) proposed a generalized removal method, which fits successively more general models to the data until an acceptable fit is found. They also mentioned that the generalized removal estimator performs poorly if it is applied to variable catch-effort sampling situations (Otis *et al.* (1978, p. 46)). Pollock and Otto (1983) proposed a jackknife estimator, which was shown by simulation to be an improvement over the generalized removal estimator. For a heterogeneous capture-recapture model, if initial and recapture probabilities are independent, then the recapture information contains no information about the population size. Thus the generalized removal model is statistically equivalent to a capture-recapture model $\boldsymbol{M}_{bh}$. Norris and Pollock (1995) obtained a non-parametric MLE of initial population size for model $\boldsymbol{M}_{bh}$. Their approach uses both capture and recapture data and thus cannot be applied to the catch-effort models. Gove *et al.* (1995) discussed some interesting effects of the violation of the equal catchability on statistical inferences.

When efforts vary from sample to sample, the commonly used estimators are regression type estimators. For example, Seber (1982, Section 7.1) discussed three types of regression estimators proposed respectively by Leslie (Leslie and Davis (1939)), Ricker (1958) and De Lury (1947). However, these regression approaches basically assume homogeneity, that is, all animals have the same removal probability for each sample. As also indicated by Gould and Pollock (1997), the assumptions and approximations used in the regression approaches are not practically valid. Moreover, Ricker's regression estimate varies with the scale of the efforts; see Section 3.2 for an example. The (conditional) MLE under homogeneous Poisson and logistic type models for variable effort case was discussed in Seber (1982) and Gould and Pollock (1997). Using simulations, Gould and Pollock (1997) found that the MLE is generally preferable to the regression estimators in terms of accuracy and precision.

Using the concept of sample coverage, Lee and Chao (1994) proposed an estimation procedure specifically for capture-recapture models. Since there is an equivalence relationship between catch-effort and capture-recapture models, some estimators that use first-captures only in Chao and Lee (1994) can be directly applied to the catch-effort model; see Table 1 and Section 2.

As indicated by Pollock (1991, page 231) ‘‘Currently, for closed populations there is no catch per unit effort model that allows for heterogeneity. This appears to be an obvious deficiency because there is a heterogeneity model for the equal effort case’’. In Section 2 of this paper, we present a class of variable catch-effort models with heterogeneous removal probabilities. This model bears a close relationship to some special types of capture-recapture models with behavioral response. The proposed models include three types of removal probabilities: multiplicative, Poisson and logistic. The usual removal and generalized removal models then become special cases. A unified nonparametric approach is developed using optimal estimating functions. Nonparametric estimators of the initial population size as well as its variance estimator are obtained. In Section 3, two real data sets are given for illustration. Section 4 reports a simulation study to show the relative merits of the proposed estimators and the previous approaches.

2. Models and estimators

Assume that the initial individuals are indexed by 1, 2, …, N and there are t removed samples. Let p_{ij} be the unknown removal probability of the ith individual in the jth trapping sample. Individuals are assumed to act independently. We also assume that the units of effort are independent, e.g., traps do not compete with each other. The data consist of a sequence of removals and the corresponding efforts: $\{(u_1, e_1), (u_2, e_2), \ldots, (u_t, e_t)\}$, where u_j denotes the removal and e_j denotes the effort of sample j. Our purpose is to estimate the initial population size N. It will be shown that the proposed estimators are independent of the scale of effort. Here u_k/e_k is the catch-per-unit-effort, usually referred to as CPUE in the fisheries literature. We propose the following three types of models for the heterogeneous model:

(1) Multiplicative model: The model assumes that

$$p_{ij} = \lambda_i e_j, \tag{2.1}$$

where λ_i is the unknown individual removal (or capture) probability for unit of effort and e_j is the known effort in the jth sample. Here, λ_i and e_j are defined only up to a multiplicative constant. Since $\lambda_i e_j$ may be greater than 1, a suitable modification is $p_{ij} = \min(\lambda_i e_j, 1)$

(2) Poisson model: The model assumes that

$$p_{ij} = 1 - \exp(-\lambda_i e_j). \tag{2.2}$$

It is derived under the assumption that sampling of the ith individual is a Poisson process with respect to effort with parameter λ_i. Parameter λ_i is referred to as a Poisson catchability coefficient in Seber (1982, p. 296), but Seber assumes that all individuals have the same coefficients.

(3) Logistic model: This model assumes that

$$p_{ij} = \exp(\lambda_i + \beta e_j)/[1 + \exp(\lambda_i + \beta e_j)], \tag{2.3}$$

i.e., there is logistic relationship between the catchabilities and the efforts.

In the foregoing models, if all efforts are equal, then without loss of generality, we may take $e_j = 1$ and this case is called a constant (or equal) effort model; otherwise, it is called a variable (or unequal) effort model. A homogeneous population means that all λ_i's are equal; otherwise, the population is called heterogeneous.

As previously stated, the constant effort model with homogeneous removal probability is statistically equivalent to the capture-recapture model $\boldsymbol{M}_b$ ($\boldsymbol{M}_{bh}$ for heterogeneous case) if initial and recapture probabilities in the latter model are independent. Under the same condition, our proposed variable effort model with homogeneous removal probability is equivalent to a capture-recapture model $\boldsymbol{M}_{tb}$ ($\boldsymbol{M}_{tbh}$ for heterogeneous case) with time effects being efforts. Therefore, any method only using unmarked information for model $\boldsymbol{M}_{tb}$ ($\boldsymbol{M}_{tbh}$) can be applied to the homogeneous (heterogeneous) variable effort models.

2.1 *Homogeneous models*

We first deal with the homogeneous case that all λ_i's in Equations (2.1)–(2.3) are equal. Let $\lambda_1 = \lambda_2 = \cdots = \lambda_N \equiv \lambda$ and let $\boldsymbol{\theta}$ denote the parameter vector involved in the model. Under the homogeneity assumption, there are only two parameters, i.e., $\boldsymbol{\theta} = (N, \lambda)^T$ (where the superscript T denotes the transpose of a matrix) in a multiplicative and a Poisson model, and there are three parameters, i.e., $\boldsymbol{\theta} = (N, \lambda, \beta)^T$ in a logistic model. Therefore, traditional inference including maximum likelihood can be applied. Homogeneity implies that the probabilities p_{ij} are the same for all i. To simplify our notation, let $p_{ij} \equiv p_j$. Define $\boldsymbol{X}_j = (x_{1j}, x_{2j}, \ldots, x_{Nj})$ as the state vector just after sample j, where $x_{ij} = 0$ if the ith animal remains uncaptured in samples 1 to j, and $x_{ij} = 1$ if the ith animal has been removed in samples 1 to j. Just after the $(k-1)$th sample, there are totally $D_{k-1} = \sum_{m=1}^{k-1} u_m = \sum_{i=1}^{N} x_{i,k-1}$ removals and there are $N - D_{k-1} = \sum_{i=1}^{N}(1 - x_{i,k-1})$ animals left in the population, $k = 1, 2, \ldots, t, D_0 \equiv 0$ and D_t is the number of total removals. Therefore, we have

$$E(u_k | \boldsymbol{X}_{k-1}) = (N - D_{k-1})p_k.$$

Thus for each sample k, we can construct an unbiased estimating function for the parameter (N, $\boldsymbol{\theta}$):

$$g_k = u_k - (N - D_{k-1})p_k,$$

for $k = 1, 2, \ldots, t$. To combine these t estimating functions, we consider a weight associated with g_k to be a function of $\boldsymbol{X}_{k-1}$. The optimal unbiased estimating equation with corresponding root that has minimum asymptotic variance is given by (e.g., see Liang and Zeger (1995)):

$$\boldsymbol{g} = \sum_{k=1}^{t} \boldsymbol{\Phi}_k V_k^{-1} g_k, \tag{2.4}$$

where $\mathbf{\Phi}_k = E\{(\partial g_k/\partial\theta)|\, \boldsymbol{X}_{k-1}\}$ and $V_k = \mathrm{var}(g_k|\, \boldsymbol{X}_{k-1})$. Notice that $\mathbf{\Phi}_k$ (and thus $\boldsymbol{g}$) is a column vector of dimension 2×1 in multiplicative and Poisson models, that is,

$$\mathbf{\Phi}_k = \begin{bmatrix} E\{(\partial g_k/\partial N)|\, \boldsymbol{X}_{k-1}\} \\ E\{(\partial g_k/\partial \lambda)|\, \boldsymbol{X}_{k-1}\} \end{bmatrix}.$$

Substituting the partial derivatives and the variance $V_k = (N - D_{k-1})p_k(1-p_k)$ into Equation (2.4), we obtain the following estimating equations:

$$\sum_{k=1}^{t} [(N - D_{k-1})(1-p_k)]^{-1}[u_k - (N - D_{k-1})p_k] = 0, \tag{2.5a}$$

$$\sum_{k=1}^{t} \left(\frac{\partial p_k}{\partial \lambda}\right)[p_k(1-p_k)]^{-1}[u_k - (N - D_{k-1})p_k] = 0. \tag{2.5b}$$

Letting $p_k = \lambda e_k$ or $p_k = 1 - \exp(-\lambda e_k)$ in Equation (2.5), we then have the estimating equations for multiplicative and Poisson models. Numerical iterative procedures are required to obtain the solution. Similarly, $\mathbf{\Phi}_k$ is a 3×1 matrix in a logistic model, where $p_k = \exp(\lambda + \beta e_k)/[1 + \exp(\lambda + \beta e_k)]$. In addition to Equation (2.5a) and (2.5b), the system of equations for a logistic model consists of one more equation:

$$\sum_{k=1}^{t} \left(\frac{\partial p_k}{\partial \beta}\right)[p_k(1-p_k)]^{-1}[u_k - (N - D_{k-1})p_k] = 0. \tag{2.6}$$

These types of optimal equations such as Equations (2.5) and (2.6) are termed quasi-likelihood equations in Godambe and Heyde (1987). Therefore, the solution of Equations (2.5) and (2.6) will be referred to as the maximum quasi-likelihood estimator (MQLE). The solution of (N, λ) for Equations (2.5) and (2.6) will be designated as $\hat{\boldsymbol{\theta}}^0_{MQLE} = (\hat{N}^0_{MQLE}, \hat{\lambda}^0_{MQLE})$, where the superscript ‘‘0’’ denotes no heterogeneity. Note that when the scale of the efforts is changed from e_k to be_k, for some constant $b>0$, the solution of λ becomes $b^{-1}\hat{\lambda}^0_{MQLE}$ but the solution of N remains unchanged. On the other hand, when the catch data u_k is changed to bu_k, it is easy to see that our estimate is changed accordingly to $b\hat{N}^0_{MQLE}$.

It follows from the theory of estimating function theory that the variance-covariance matrix takes the form (see McCullagh and Nelder (1989))

$$\mathrm{var}(\hat{\theta}^0_{MQLE}) \approx \left[E\sum_{k=1}^{t} \mathbf{\Phi}_k V_k^{-1} \mathbf{\Phi}_k^T\right]^{-1}.$$

This implies after some calculations that the variance of $\hat{N}^0_{MQLE}$ can be expressed as

$$\mathrm{var}(\hat{N}^0_{MQLE}) = N\sum_{k=1}^{t} \frac{Q_{k-1}}{p_k(1-p_k)}\left(\frac{\partial p_k}{\partial \boldsymbol{\lambda}}\right)^2 \left[\frac{1-Q_t}{Q_t}\sum_{i=1}^{t}\frac{Q_{i-1}}{p_i(1-p_i)}\left(\frac{\partial p_i}{\partial \boldsymbol{\lambda}}\right)^2 - \left(\sum_{i=1}^{t}\frac{1}{1-p_i}\frac{\partial p_i}{\partial \boldsymbol{\lambda}}\right)^2\right]^{-1}, \tag{2.7}$$

where $Q_k = \prod_{j=1}^{k}(1-p_j)$. The unbiasedness of $\boldsymbol{g}$ in Equation (2.4) implies the consistency of $\hat{N}^0_{MQLE}$ in the sense that $\hat{N}^0_{MQLE}/N$ converges to 1 in probability when N is large enough.

To obtain the MLE, we write down the multinomial likelihood function of $\boldsymbol{\theta} = (N, \lambda)$:

$$L(N,\lambda) = \frac{N!}{(N-D_t)!\prod_{k=1}^{t} u_k!} p_1^{u_1}[p_2(1-p_1)]^{u_2}\ldots\left[p_t\prod_{j=1}^{t-1}(1-p_j)\right]^{u_t}\left[\prod_{j=1}^{t}(1-p_j)\right]^{N-D_t}$$

Let $Q = \prod_{j=1}^{t}(1-p_j)$. We can further write $L(N,\lambda) = L_1(N,\lambda)L_2(\lambda)$, where

$$L_1(N,\lambda) = \frac{N!}{(N-D_t)!D_t!}(1-Q)^{D_t}Q^{N-D_t},$$

$$L_2(\lambda) = \frac{D_t!}{\prod_{k=1}^{t} u_k!}[p_1/(1-Q)]^{u_1}[p_2(1-p_1)/(1-Q)]^{u_2}\ldots\left[p_t\prod_{j=1}^{t-1}(1-p_j)/(1-Q)\right]^{u_t}.$$

Note that $L_1(N,\lambda)$ is a binomial likelihood and $L_2(\lambda)$ is a conditional multinomial likelihood given D_t. There are two types of maximum likelihood estimation procedures: unconditional and conditional (Sanathanan (1972, 1977); Bishop *et al*. (1975), Chapter 6). The unconditional MLE is obtained by maximizing $L(N,\lambda)$ simultaneously with respect to N and λ. Following the suggestion by Gould and Pollock (1997), we consider in this paper the conditional MLE. This conditional approach is to compute the MLE $\hat{\lambda}$ of λ by maximizing $L_2(\lambda)$ and then obtain an estimate $\hat{N}$ by maximizing $L_1(N,\hat{\lambda})$ over any integer $N, N \geq D_t$. The latter maximization yields that the conditional MLE for given Q is $[D_t/Q]$, where $[a]$ denotes the largest integer $\leq a$. See Dahiya (1981), Lindsay and Roeder (1987) for a unified treatment of integer parameter models. If we allow the resulting estimate to be a real number (i.e., D_t/Q) rather than the actual integer-value, then we can show that the conditional MLE satisfies the following system of equations:

$$1 - \frac{D_t}{N} = \prod_{k=1}^{t}(1-p_k), \tag{2.8a}$$

$$\sum_{k=1}^{t}\left(\frac{\partial p_k}{\partial\lambda}\right)[p_k(1-p_k)]^{-1}[u_k - (N-D_{k-1})p_k] = 0. \tag{2.8b}$$

The solution will be denoted by $(\hat{N}^0_{MLE}, \hat{\lambda}^0_{MLE})$. For a Poisson model, the resulting equations by substituting $p_k = 1-\exp(-\lambda e_k)$ into Eqaution (2.8) are identical to those given in Seber (1982, p. 297). The system of equations for MLE and MQLE are tabulated in Table 2 for comparison.

Under the regularity conditions specified in Sanathanan (1972), the resulting MLE for all three models is consistent. The variance of the MLE can be obtained by inverting the expected Fisher information matrix. That is, the asymptotic variance-covariance matrix of $(\hat{N}^0_{MLE}, \hat{\lambda}^0_{MLE})$ is the inverse of the expectation of the matrix

$$\begin{bmatrix} -\frac{\partial^2 \log L}{\partial N^2} & -\frac{\partial^2 \log L}{\partial\lambda\partial N} \\ -\frac{\partial^2 \log L}{\partial N\partial\lambda} & -\frac{\partial^2 \log L}{\partial\lambda^2} \end{bmatrix}.$$

After some direct manipulations, the MLE has exactly the same asymptotic variance as that given in Eqaution (2.7). Therefore, the MLE is asymptotically equivalent to the

Table 2. Computational formulas for various estimators from the data $\{(u_j,\ e_j), j = 1, 2, \ldots, t\}$,

$$D_k = \sum_{j=1}^{k} u_j, \quad A_k = \sum_{j=1}^{k} e_j / e_k.$$

Models or Estimators — *Equation Number in Text*

Homogeneous Model: $p_{ij} \equiv p_j$

MQLE ($\hat{N}^0_{MQLE}$): solution of the following system of equations

$$\sum_{k=1}^{t} [(N - D_{k-1})(1 - p_k)]^{-1} [u_k - (N - D_{k-1}) p_k] = 0, \tag{2.5a}$$

$$\sum_{k=1}^{t} \left(\frac{\partial p_k}{\partial \lambda}\right) [p_k(1 - p_k)]^{-1} [u_k - (N - D_{k-1}) p_k] = 0. \tag{2.5b}$$

Conditional MLE ($\hat{N}^0_{MLE}$): solution of the following system of equations

$$1 - \frac{D_t}{N} = \prod_{k=1}^{t} (1 - p_k), \tag{2.8a}$$

$$\sum_{k=1}^{t} \left(\frac{\partial p_k}{\partial \lambda}\right) [p_k(1 - p_k)]^{-1} [u_k - (N - D_{k-1}) p_k] = 0. \tag{2.8b}$$

$\hat{N}^0_{SC}$: see below.

Heterogeneous Model: $p_{ij} = \lambda_i e_j$

MQLE ($\hat{N}_{MQLE}$) and sample coverage estimator of N ($\hat{N}_{SC}$):

Step 1. For $k = 1, 2, \ldots, t - 1$, calculate the estimate of C_k

$$\hat{C}_k = 1 - \frac{u_{k+1}/e_{k+1}}{u_1/e_1}; \tag{2.12}$$

Step 2. Define

$$\tau = \max_{\hat{C}_k > 0\ (1 \le k \le t-1)} k$$

;

Step 3. Calculate $\hat{N}^0_{SC} = D_\tau / \hat{C}_\tau$; (2.13)

Step 4. Calculate CV estimator

$$\hat{\gamma}^2 = \max\{\hat{N}^0_{SC}\ [u_1 - u_2(e_1/e_2)]/u_1^2 - 1, \quad 0\}; \tag{2.18}$$

Step 5. The estimators are

$$\hat{N}_{MQLE} = \frac{\sum_{k=2}^{t} (e_k^2/u_k)[D_{k-1} + A_{k-1} u_{k-1} \hat{\gamma}^2]\ I\ [\hat{C}_{k-1} > 0]}{\sum_{k=2}^{t} (e_k^2/u_k) \hat{C}_{k-1} I\ [\hat{C}_{k-1} > 0]}, \tag{2.22}$$

$$\hat{N}_{SC} = \frac{\hat{D}^*_\tau}{\hat{C}_\tau} = \frac{D_\tau}{\hat{C}_\tau} + \frac{A_\tau\ u_\tau\ \hat{\gamma}^2}{\hat{C}_\tau}. \tag{2.23}$$

Note: An interactive computer program for calculating various estimators is available from the authors.

MQLE and both estimators have identical asymptotic distribution. Other estimators from the historical literature include the regression estimators and the sample coverage approach (see Equation (2.13)). See Table 1.

For the equal effort case (removal model), we let $e_k = 1$ in the three models and all models can be reparametrized as $p_{ij} \equiv \lambda$. In this case, we have $p_1 = p_2 = \cdots = p_t = \lambda$ in Equation (2.5) and we can drop the constant λ and 1-λ in the equations. The MQLE of (N, λ) for a homogeneous case can be simplified to the solution of the following equations:

$$\sum_{k=1}^{t}(N - D_{k-1})^{-1}[u_k - (N - D_{k-1})\lambda] = 0, \tag{2.9a}$$

$$\sum_{k=1}^{t}[u_k - (N - D_{k-1})\lambda] = 0. \tag{2.9b}$$

In this special case, the variance formula Equation (2.7) reduces to

$$\mathrm{var}(\hat{N}^0_{MQLE}) = N\frac{(1-\lambda)^t[1-(1-\lambda)^t]}{[1-(1-\lambda)^t]^2 - t^2\lambda^2(1-\lambda)^{t-1}},$$

which is the formula derived in Otis *et al.* (1978, p. 108). The commonly used (conditional) maximum likelihood estimator N under a removal model satisfies

$$1 - D_t/N = (1-\lambda)^t, \tag{2.10a}$$

$$\sum_{k=1}^{t}[u_k - (N - D_{k-1})\lambda] = 0, \tag{2.10b}$$

which is the same as those given in Moran (1951) and Zippin (1956).

To obtain a variance estimator for the MLE or MQLE, we could use the variance derived in (2.7) and replace all the parameters by their estimates. However, formula (2.7) is an asymptotic result and it is valid only for large population sizes. We suggest the use of a bootstrap procedure; see Buckland and Garthwaite (1991) for various applications of the bootstrap method. Under the homogeneous assumption, the distribution of $\{u_1, u_2, \ldots, u_t, u_0\}$ ($u_0 = N - D_t$ is the number of non-removals) follows an exact multinomial distribution. For notational simplicity, assume the estimator is denoted by $\hat{N}$. A bootstrap replication $\{u_0^*, u_1^*, \ldots, u_t^*, u_0^*\}$ is generated from a multinomial distribution with parameter $\hat{N}$ and cell probabilities $\{u_1/\hat{N}, \ldots,\ u_t/\hat{N},\ 1 - D_t/\hat{N}\}$. Then based on the bootstrap catch data, an estimate $\hat{N}^*$ is calculated. After B replications, the bootstrap variance of $\hat{N}$ is simply the sample variance of the B bootstrap estimates $\hat{N}^{*1}, \hat{N}^{*2}, \ldots,\ \hat{N}^{*B}$. The performance of bootstrap estimators will be investigated in Section 4. A confidence interval of N can be derived using a method presented in Chao (1989). An advantage of the interval is that the lower bound is always greater than the observed number of removals.

2.2 *Heterogeneous models*

The following discussion treats all λ_i 's in Equations (2.1)–(2.3) as fixed parameters. The relevant parameters are the mean $\bar{\lambda} = \sum_{i=1}^{N} \lambda_i/N$ and coefficient of variation (CV)

$\gamma = \sum_{i=1}^{N}[(\lambda_i - \bar{\lambda})^2/N]^{1/2}/\bar{\lambda}$. Here the CV is a measure of the degree of heterogeneity. The condition $\gamma = 0$ is equivalent to the homogeneous assumption, i.e., all animals have the same removal probabilities in any fixed sample. For a random-effects model, which treats $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ as a random sample from a distribution, all the following derivations are parallel and the same estimators yield. We will specifically deal with the multiplicative model, and the results are also approximately valid for a Poisson model. The computational algorithm is listed in Table 2. The necessary explanation and derivation are given below.

We first review the concept of sample coverage, since it will be used in our derivation of the optimal estimating functions for the heterogeneous cases. The basic motivation for a sample coverage approach is that the sample coverage (defined below) can be directly and well estimated even in the heterogeneous populations. The sample coverage of the first k samples, C_k, is defined as the probability-weighted fraction of the population that was removed in the first k samples, i.e.,

$$C_k = \frac{\sum_{i=1}^{N} \lambda_i \, x_{ik}}{\sum_{i=1}^{N} \lambda_i}, \tag{2.11}$$

where x_{ik} is a component of the state vector as defined in Section 2.1. Notice that if all λ_i 's are equal, then $C_k = D_k/N$. Hence a natural estimator of N under a model without heterogeneity is $D_k/\hat{C}_k$, where $\hat{C}_k$ is an ''estimator'' of C_k and

$$\hat{C}_k = 1 - \frac{u_{k+1}/e_{k+1}}{u_1/e_1}. \tag{2.12}$$

See the Appendix for a derivation of Equation (2.12). The expected value of $\hat{C}_k$ for all $k = 1, 2, \ldots, t-1$ is theoretically positive. Due to sampling variation, it might happen that, for some j, $\hat{C}_j$ is zero or negative. Therefore, we let τ be the largest k such that $\hat{C}_k > 0$ (that is, the largest k such that $u_{k+1}/e_{k+1} < u_1/e_1$), and the estimator of N based on the notion of sample coverage under the homogeneous model takes the form:

$$\hat{N}_{SC}^{0} = D_\tau/\hat{C}_\tau. \tag{2.13}$$

In most cases, $\tau = t - 1$.

We now present the optimal estimating equation for the heterogeneous case. All the derivation details are provided in the Appendix. Given the state vector of $\boldsymbol{X}_{k-1}$, we can easily express u_k as $u_k = \sum_{i=1}^{N}(x_{ik} - x_{i,k-1}) = \sum_{i=1}^{N} x_{ik}(1 - x_{i,k-1})$ because both x_{ik} and $x_{i,k-1}$ take values of 0 and 1 only. Therefore, we obtain the following conditional expectation:

$$\begin{aligned} E(u_k | \, \boldsymbol{X}_{k-1}) &= \sum_{i=1}^{N} \lambda_i e_k (1 - x_{i,k-1}) \\ &= (1 - C_{k-1}) \left(\sum_{i=1}^{N} \lambda_i \right) e_k = (N - NC_{k-1}) \bar{\lambda} e_k, \end{aligned} \tag{2.14}$$

where $C_0 \equiv 0$. In the heterogeneous case, NC_{k-1} is no longer equal to D_{k-1}. For notational simplicity, let $NC_{k-1} \equiv D_{k-1}^{*}$. The idea here is to find an estimator for the expected discrepancy between NC_{k-1} and D_{k-1}. Note that

$$E(D^*_{k-1}) = E(NC_{k-1}) = N\left[1 - \sum_{i=1}^{N} \lambda_i \prod_{j=1}^{k-1}(1-\lambda_i e_j) / \sum_{i=1}^{N}\lambda_i\right], \tag{2.15}$$

$$E(D_{k-1}) = N - \sum_{i=1}^{N}\prod_{j=1}^{k-1}(1-\lambda_i e_j). \tag{2.16}$$

If all e_j's are equal to 1, then we can derive that

$$E(D^*_{k-1}) - E(D_{k-1}) = (k-1)E(u_{k-1})\,\gamma^2 + \Delta_k, \tag{2.17a}$$

where Δ_k is the remainder term and γ is the CV defined before. In the Appendix, we show that if $\lambda_1, \lambda_2, \ldots, \lambda_N$ can be regarded as a random sample from a gamma distribution with density function $\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}/\Gamma(\alpha)$, then $\Delta_k/N \to 0$ for any fixed $k \geq 1$ as N is large enough. Because the gamma distribution can cover a wide range of types of heterogeneity, as shown by Fisher, Corbet and Williams (1943), this is our main motivation to ignore the remainder term in (2.17a). For the variable effort case, we can obtain a similar expansion

$$E(D^*_{k-1}) - E(D_{k-1}) = \left[\left(\sum_{m=1}^{k-1} e_m\right)/e_{k-1}\right] E(u_{k-1})\,\gamma^2 + \Delta^*_k. \tag{2.17b}$$

However, the remainder term Δ^*_k/N may deviate from zero unless all e_j's are equal. From this point of view, allocation of equal efforts is preferable to unequal effort. Let $A_{k-1} = [(\sum_{m=1}^{k-1} e_m)/e_{k-1}]$. If $NC_{k-1} \equiv D^*_{k-1}$ in Equation (2.14) is replaced by $D_{k-1} + A_{k-1}\, u_{k-1}\, \gamma^2$, then the CV (i.e., γ) is involved in the unbiased estimating function. We could regard the CV as one of the parameters and obtain the resulting estimating equations. Simulation results have suggested that the iterative steps failed to converge in many trials. Hence we adopt below the estimator of the CV from Lee and Chao (1994):

$$\hat{\gamma}^2 = \max\{\hat{N}^0_{SC}\,[u_1 - u_2(e_1/e_2)]/u_1^2 - 1, \quad 0\}, \tag{2.18}$$

where $\hat{N}^0_{sc}$ is derived in Equation (2.13). Now define $\hat{D}^*_0 = 0$ and for $k > 1$

$$\hat{D}^*_{k-1} = D_{k-1} + A_{k-1}\, u_{k-1}\, \hat{\gamma}^2,$$

and consider an approximate unbiased estimating function:

$$g_k = u_k - (N - \hat{D}^*_{k-1})\bar{\lambda} e_k. \tag{2.19}$$

The conditional variance can be shown to be

$$\mathrm{var}(g_k|\, \boldsymbol{X}_{k-1}) = \sum_{i=1}^{N} \lambda_i\, e_k(1-\lambda_i e_k)\,(1 - x_{i,k-1}) \approx (N - NC_{k-1})\bar{\lambda} e_k. \tag{2.20}$$

Substituting Equation (2.19), (2.20) into Equation (2.4) and using $\hat{D}^*_{k-1} \approx NC_{k-1} \approx N\hat{C}_{k-1}$, we have the estimator of $(N, \bar{\lambda})$ satisfy $(\hat{C}_0 \equiv 0)$

$$\sum_{k=1}^{t}(1-\hat{C}_{k-1})^{-1}[u_k - (N - \hat{D}^*_{k-1})\bar{\lambda} e_k] = 0, \tag{2.21a}$$

$$\sum_{k=1}^{t}[u_k - (N - \hat{D}^*_{k-1})\bar{\lambda} e_k] = 0. \tag{2.21b}$$

From (2.21b), we can express that $\bar{\lambda} = N^{-1}\sum_{k=1}^{t} u_k / \sum_{k=1}^{t}(1-\hat{C}_{k-1})e_k = u_1/(Ne_1)$ by use of Equation (2.12). Substituting this expression of $\bar{\lambda}$ into Equation (2.21a), we thus obtain an explicit solution of N:

$$\hat{N}_{MQLE} = \frac{\sum_{k=2}^{t}[e_k/(1-\hat{C}_{k-1})]\hat{D}^*_{k-1}}{\sum_{k=2}^{t}[e_k/(1-\hat{C}_{k-1})]\hat{C}_{k-1}} = \frac{\sum_{k=2}^{t}(e_k^2/u_k)[D_{k-1}+A_{k-1}u_{k-1}\hat{\gamma}^2]}{\sum_{k=2}^{t}(e_k^2/u_k)\hat{C}_{k-1}}.$$

Since the estimate $\hat{C}_{k-1}$ may be negative, a modified estimator is

$$\hat{N}_{MQLE} = \frac{\sum_{k=2}^{t}(e_k^2/u_k)[D_{k-1}+A_{k-1}u_{k-1}\hat{\gamma}^2]\ I\ [\hat{C}_{k-1}>0]}{\sum_{k=2}^{t}(e_k^2/u_k)\hat{C}_{k-1}I\ [\hat{C}_{k-1}>0]}, \tag{2.22}$$

where I[A] is an indicator function of the event A; that is, $I[A]=1$ if event A occurs and 0 otherwise. The estimator proposed in Lee and Chao (1994) has the form:

$$\hat{N}_{SC} = \frac{\hat{D}^*_\tau}{\hat{C}_\tau} = \frac{D_\tau}{\hat{C}_\tau} + \frac{A_\tau\ u_\tau\ \hat{\gamma}^2}{\hat{C}_\tau}, \tag{2.23}$$

where τ was defined before Equation (2.13), i.e., the largest k such that $\hat{C}_k>0$. Both Equations (2.22) and (2.23) theoretically show that estimators which do not incorporate CV estimates would underestimate the true size in heterogeneous populations (i.e., substitute $\hat{\gamma}^2=0$ in both formulas when the true $\gamma^2>0$). It is also clear that the proposed estimator Equation (2.22) is a weighted sum of the type of estimator Equation (2.23) with optimal weight being equal to $e_k^2/u_k \propto e_k/(1-\hat{C}_{k-1})$. That is, the weight is proportional to the effort and inversely proportional to $(1-\hat{C}_{k-1})$. If all the efforts are approximately equal, then we put more weight on the later samples because $E(C_{k-1})$ increases with k. Both $\hat{N}_{SC}$ and $\hat{N}_{MQLE}$ are obviously independent of the scale of efforts. Namely, no matter what scale for the efforts is recorded, it does not change the estimates. Also, when the catch data u_k is changed to bu_k, both estimates change accordingly. The estimators $\hat{N}_{SC}$ given in Equation (2.23) and $\hat{N}^0_{SC}$ given in Eqaution (2.13) will be referred to as ''sample coverage estimator'' (of the initial population size) hereafter if no confusion with $\hat{C}$ arises.

In the heterogeneous population, it is difficult to obtain an asymptotic variance. Although the exact distribution of $\{u_1, u_2, \ldots u_t, u_0\}$ ($u_0 = N - D_t$ is the number of non-removals) is unknown, it is approximately distributed as a multinomial as shown in Darroch *et al.* (1994). Therefore, a bootstrap variance estimator and confidence intervals can be constructed as in the homogeneity models. The performance of the bootstrap variance estimator will be examined in the simulation section.

3. Real data examples

3.1 *Fish removal data (constant effort model)*

We consider a removal data set for the fantail darter (*Etheostoma flabellare*) originally presented in Mahon (1980) and also discussed in White *et al.* (1982, p. 114). The numbers of removals for the seven samples are

$$180, 115, 94, 84, 75, 58, 60.$$

Table 3. Estimates and their standard errors.

Models or Estimators	*Fish Data*	*Lobster Data*
Homogeneous Model:		
Regression estimator:		
Leslie	858 (70)	120.5 (8.92)
Ricker	933 (68)	119.1 (6.72)
De Lury	856 (70)	115.2 (6.83)
MLE	904 (56)	125.5 (1.54)
MQLE	900 (53)	123.3 (1.48)
Sample coverage	909 (64)	130.2 (2.65)
Heterogeneous Model:		
Generalized removal	1025 (105)	—
Sample coverage	1339 (234)	192.0 (26.2)
Jackknife	1026 (50)	—
Proposed MQLE	1219 (168)	211.1 (26.6)

We choose this data set because the true $N = 1151$ is known; therefore, we can assess directly the performance of various estimates. All estimates are tabulated in Table 3.

If we incorrectly assume that all removal probabilities are equal, the unconditional MLE under a usual removal model is 900 (s.e. 48), as obtained by White *et al.* (1982). The conditional MLE using Equation (2.10) is 904 (bootstrap s.e. 56) and the MQLE using Equation (2.9) is 900 (s.e. 53). The usual regression estimators can also be applied to this equal effort case. The three regression estimates and their estimated s.e's are Leslie's estimate 858 (s.e. 70), Ricker's estimate 933 (s.e. 68) and DeLury's estimate 856 (s.e. 70). The estimated sample coverage for the first 6 samples is $\hat{C}_6 = 1 - u_7/u_1 = 66.7\%$ and the estimator without incorporating the estimation of CV (Equation (2.13)) is $\hat{N}^0_{SC} = 606/\hat{C}_6 = 909$ with an estimated s.e. of 64 based on 1000 bootstrap replications. Therefore, all estimates under the homogeneity assumption are very close.

However, the CV estimate is $\hat{\gamma} = \{\hat{N}^0_{SC}(u_1 - u_2)/u_1^2 - 1\}^{1/2} = 0.908$; see Equation (2.18). This large value of CV clearly indicates that the homogeneous assumption is not valid. The generalized removal estimate calculated by White *et al.* (1982) for these data is 1025 with an estimated s.e. of 105. The jackknife estimate proposed by Pollock and Otto (1983) is 1026 with an estimated s.e. of 50. The sample coverage approach using Equation (2.23) gives an estimate $\hat{N}_{SC} = 1339$ with a bootstrap estimated s.e. of 234 using 1000 replications.

We now apply the proposed approach to these data. Under the heterogeneous assumption, Equation (2.22) yields an estimate of $\hat{N}_{MQLE} = 1219$ (s.e. 168) and a 95% confidence interval of (975, 1657) based on a log-transformation as in Chao (1989). Comparing our estimator with the previously mentioned results, we find the proposed estimate is closest to the true $N = 1151$. More systematic comparisons will be given in Section 4.

3.2 *Lobster data (variable effort model)*

We consider a well-known set of 17-sample catch-effort lobster data originally given in De Lury (1947) and also discussed in Seber (1982, chapter 7). Here we identify ''pounds'' with ''number'' of individuals because the distribution of lobster size remained fairly constant during the sampling period (De Lury (1947)). The notation u_j and e_j in this paper correspond to Seber's n_j and f_j. All the estimates are given in Table 3.

Under the homogeneous assumption, the (conditional) MLE and MQLE for a multiplicative model are 125.5 and 123.3 (in 1000 pounds). Similar estimates were obtained under a Poisson model. The commonly used regression type estimators under a variable effort model are the following: Leslie's estimate is 120.5 (s.e. 8.92); Ricker's estimate is 119.1 (s.e. 6.72) and De Lury's estimate is 115.2 (s.e. 6.83). Thus the three regression methods produce similar results. We remark that if the effort is performed in a unit of 2000 traps so that each of the effort units becomes half the value of the original respective effort, then all estimates except for Ricker's estimator remain exactly the same. Ricker's estimate changes to 123. Similarly Ricker's estimate becomes 117 if the effort is changed to a unit of 500 traps. Although the differences in estimates are not significant in this example, this inconsistency is a drawback for Ricker's regression estimator. The sample coverage approach based on (2.13) under a homogeneous assumption is $\hat{N}^0_{SC} = 130.2$ (s.e. 2.65), which is also comparable to the regression results.

It follows from Eqaution (2.18) that the CV estimate is $\hat{\gamma} = 0.8$, which shows evidence of strong heterogeneity. Thus the estimator of N without incorporating the CV term is expected to be biased downwards. The sample coverage estimate of N using Equation (2.23) is $\hat{N}_{SC} = 192.0$ with an estimated s.e. of 26.2. Our proposed estimate given in Equation (2.22) yields an estimate of $\hat{N}_{MQLE} = 211.1$ (s.e. 26.6) and a 95% confidence interval of (168, 273). Our estimate and the sample coverage estimate are much higher than other estimates, which do not consider heterogeneity.

4. Simulation studies

We carried out a simulation study to investigate the behavior of the proposed estimators and to compare them with other estimates. The trials are described in Table 4. The true population size (N) for each trial was fixed to be 400 and $t = 5$ was selected. In most practical applications, the population sizes are likely to be very large. The reasons for selecting $N = 400$ are the following: Any of our proposed estimators (say, $\hat{N}$) would change accordingly to $b\hat{N}$ if the population size were changed to bN for some $b > 1$ (i.e., for each animal, add $b - 1$ additional animals with the same catchability into the population). Therefore, as far as the ''relative bias'' is concerned, our estimators are independent of the value of N, although the standard error does increase with N. Also, the larger the population size, the more the computing time needed in the simulation. Due to limited computing resources, $N = 400$ was chosen.

We focused on the multiplicative model $p_{ij} = \lambda_i e_j$ where $0 < \lambda_i e_j \leq 1$. A total of 16 trials were considered: 12 fixed-effects models (Trials 1–12) and four random-effects models (Trials 13–16). Three types of efforts were chosen:

Type A: $(e_1, e_2, e_3, e_4, e_5) = (1, 1, 1, 1, 1)$;

Type B: $(e_1, e_2, e_3, e_4, e_5) = (1.5, 1.5, 1, 0.5, 0.5)$;

Type C: $(e_1, e_2, e_3, e_4, e_5) = (1.0, 1.5, 0.5, 1.5, 0.5)$.

In the fixed-effects models, the population of 400 animals was divided into four different

Table 4. Description of the trials (in Trials 1–12, there are 100 animals with catchabilities $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ respectively, $\bar{\lambda}$ = average, CV = coefficient of variation).

Trial	*CV*	$\bar{\lambda}$	λ_1	λ_2	λ_3	λ_4
1	0	0.25	0.25	0.25	0.25	0.25
2	0	0.3	0.3	0.3	0.3	0.3
3	0	0.35	0.35	0.35	0.35	0.35
4	0.316	0.25	0.15	0.2	0.3	0.35
5	0.316	0.3	0.18	0.24	0.36	0.42
6	0.316	0.35	0.21	0.28	0.42	0.49
7	0.552	0.25	0.125	0.15	0.25	0.475
8	0.552	0.3	0.15	0.18	0.3	0.57
9	0.552	0.35	0.175	0.21	0.35	0.665
10	0.707	0.25	0.1	0.15	0.2	0.55
11	0.707	0.3	0.12	0.18	0.24	0.66
12	0.707	0.35	0.08	0.138	0.517	0.665
13	0.354	0.277	$\lambda_1, \lambda_2, \ldots, \lambda_N \propto$ gamma (8, 1)			
14	0.447	0.237	$\lambda_1, \lambda_2, \ldots, \lambda_N \propto$ gamma (5, 1)			
15	0.500	0.213	$\lambda_1, \lambda_2, \ldots, \lambda_N \propto$ gamma (4, 1)			
16	0.577	0.189	$\lambda_1, \lambda_2, \ldots, \lambda_N \propto$ gamma (3, 1)			

subpopulations of 100 animals each. That is, there were 100 animals with individual removal probability $\lambda_1, \lambda_2, \lambda_3$ and λ_4 respectively per unit of effort. The average of λ_i's is in the range of 0.25 to 0.35 and the CV ranges from 0 to 0.707. In the random-effects models, $\lambda_i = \lambda_i^*/[1.5 \times \max\{\lambda_1^*, \lambda_2^*, \ldots, \lambda_N^*\}]$ where $\{\lambda_1^*, \lambda_2^*, \ldots, \lambda_N^*\}$ were generated from a gamma density $\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}/\Gamma(\alpha)$. Here we properly scaled the gamma random variables so that $\lambda_i e_j \leq 1$. Four values of α ($\alpha = 3, 4, 5, 8$) and $\beta = 1$ were considered.

For each trial and each fixed type of efforts, 200 data sets were generated. Then for each generated data set, five estimates and their estimated standard errors were calculated. The five estimates included three estimators for homogeneous populations: $\hat{N}^0_{MLE}$ (Equation (2.8)), $\hat{N}^0_{MQLE}$ (Equation (2.5)), $\hat{N}^0_{SC}$ (Equation (2.13)), and two estimators for heterogeneous populations: $\hat{N}_{MQLE}$ (Equation (2.22)) and $\hat{N}_{SC}$ (Equation (2.23)). See Table 2 for a summary of the computational formulas. The regression estimates were not included because Gould and Pollock (1997) have concluded that the MLE is preferable to the regression estimators. The estimated standard errors were obtained by using 1000 bootstrap replications. Finally these 200 estimates and standard errors were averaged to yield the ''average estimate'', ''average bias'' and ''average estimated s.e.'' (see Table 5). The sample standard errors as well as the sample root mean squared errors (RMSE) were also computed. In Table 5, we list the simulation results for Type C of efforts. For the other types of efforts, the conclusions are generally consistent. The average number of total removals (D_t) for each trial is also tabulated in Table 5.

First notice from Table 5 that in all cases, the MLE $\hat{N}^0_{MLE}$ and the MQLE $\hat{N}^0_{MQLE}$ yield very similar results. For the homogeneous models (Trials 1–3, $CV = 0$), both estimators perform equally well and are superior to the sample coverage estimator $\hat{N}^0_{SC}$. The proposed estimator $\hat{N}^0_{MQLE}$ has slightly smaller bias and smaller RMSE than those of the MLE, but the differences are quite limited. Similar behavior was also found for $\hat{N}^0_{MQLE}$ in other types of efforts. Thus for homogeneous models, both $\hat{N}^0_{MLE}$ and our proposed estimator $\hat{N}^0_{MQLE}$

Table 5. Simulation results for variable effort $(e_1, e_2, e_3, e_4, e_5) = (1.0, 1.5, 0.5, 1.5, 0.5)$; 200 simulation runs, 1000 bootstrap replications for estimated s.e.

$\hat{N}^0_{MLE}$: MLE for homogeneous cases; see Equation (2.8).

$\hat{N}^0_{MQLE}$: MQLE for homogeneous cases; see Equation (2.5).

$\hat{N}^0_{SC}$: sample coverage estimator for homogeneous cases; see Equation (2.13).

$\hat{N}_{SC}$: sample coverage estimator for heterogeneous cases; see Equation (2.23).

$\hat{N}_{MQLE}$: MQLE for heterogeneous cases; see Equation (2.22).

Trial	*Estimator*	*Average Estimate*	*Average Bias*	*Average Estimated s.e.*	*Sample s.e.*	*Sample RMSE*
1	$\hat{N}^0_{MLE}$	408	8	33.6	30.1	31.1
$CV = 0$	$\hat{N}^0_{MQLE}$	406	6*	32.5	29.5	30.1*
$D_t = 310$	$\hat{N}^0_{SC}$	411	11	52.7	41.7	43.1
	$\hat{N}_{SC}$	436	36	86.4	57.6	67.9
	$\hat{N}_{MQLE}$	433	33	55.5	40.8	52.2
2	$\hat{N}^0_{MLE}$	401	1	19.2	18.6	18.6
$CV = 0$	$\hat{N}^0_{MQLE}$	400	0*	18.7	18.4	18.4*
$D_t = 339$	$\hat{N}^0_{SC}$	402	2	28.8	27.7	27.8
	$\hat{N}_{SC}$	423	23	47.8	45.0	50.7
	$\hat{N}_{MQLE}$	419	19	32.4	30.9	36.4
3	$\hat{N}^0_{MLE}$	401	1	12.5	12.4	12.4
$CV = 0$	$\hat{N}^0_{MQLE}$	400	0*	12.2	12.3	12.3*
$D_t = 360$	$\hat{N}^0_{SC}$	400	0*	18.4	17.4	17.4
	$\hat{N}_{SC}$	410	10	27.7	26.8	28.7
	$\hat{N}_{MQLE}$	410	10	20.7	19.5	22.1
4	$\hat{N}^0_{MLE}$	376	– 24	26.0	25.3	34.9*
$CV = 0.316$	$\hat{N}^0_{MQLE}$	374	– 26	25.2	24.8	35.7
$D_t = 299$	$\hat{N}^0_{SC}$	383	– 17	43.3	41.9	45.2
	$\hat{N}_{SC}$	412	12	73.7	65.8	66.8
	$\hat{N}_{MQLE}$	399	– 1*	44.1	42.7	42.7
5	$\hat{N}^0_{MLE}$	380	– 20	17.1	16.6	26.0*
$CV = 0.316$	$\hat{N}^0_{MQLE}$	379	– 21	16.5	16.6	26.9
$D_t = 326$	$\hat{N}^0_{SC}$	385	– 15	26.8	24.5	28.9
	$\hat{N}_{SC}$	406	6	43.8	40.0	40.4
	$\hat{N}_{MQLE}$	397	– 3*	29.4	27.6	27.7
6	$\hat{N}^0_{MLE}$	380	– 20	11.0	11.3	23.1
$CV = 0.316$	$\hat{N}^0_{MQLE}$	379	– 21	10.7	11.2	23.8
$D_t = 346$	$\hat{N}^0_{SC}$	384	– 16	17.5	17.2	23.7
	$\hat{N}_{SC}$	397	– 3*	27.2	25.2	25.4
	$\hat{N}_{MQLE}$	391	– 9	18.9	18.1	20.3*
7	$\hat{N}^0_{MLE}$	338	– 62	19.9	20.0	65.4
$CV = 0.552$	$\hat{N}^0_{MQLE}$	336	– 64	19.4	20.0	66.7
$D_t = 280$	$\hat{N}^0_{SC}$	345	– 55	32.7	29.3	62.5

Table 5. Continued.

Trial	*Estimator*	*Average Estimate*	*Average Bias*	*Average Estimated s.e.*	*Sample s.e.*	*Sample RMSE*
	$\hat{N}_{SC}$	379	– 21*	60.5	55.3	59.0
	$\hat{N}_{MQLE}$	361	– 39	36.1	34.9	52.2*
8	$\hat{N}^0_{MLE}$	347	– 53	13.8	14.7	55.3
$CV = 0.552$	$\hat{N}^0_{MQLE}$	345	– 55	13.4	14.4	56.5
$D_t = 305$	$\hat{N}^0_{SC}$	357	– 43	24.0	24.6	49.4
	$\hat{N}_{SC}$	386	– 14*	42.2	40.9	43.3
	$\hat{N}_{MQLE}$	367	– 33	25.7	24.6	41.2*
9	$\hat{N}^0_{MLE}$	352	– 48	9.9	11.2	49.3
$CV = 0.552$	$\hat{N}^0_{MQLE}$	351	– 49	9.6	11.2	50.5
$D_t = 323$	$\hat{N}^0_{SC}$	362	– 38	17.6	17.5	41.8
	$\hat{N}_{SC}$	389	– 11*	31.1	30.1	32.0*
	$\hat{N}_{MQLE}$	371	– 29	19.3	19.7	35.2
10	$\hat{N}^0_{MLE}$	311	– 89	16.7	16.1	90.5
$CV = 0.707$	$\hat{N}^0_{MQLE}$	310	– 90	16.1	15.7	91.8
$D_t = 264$	$\hat{N}^0_{SC}$	321	– 79	28.5	25.8	83.5
	$\hat{N}_{SC}$	358	– 42*	55.6	51.3	66.2*
	$\hat{N}_{MQLE}$	335	– 65	31.8	30.9	71.8
11	$\hat{N}^0_{MLE}$	321	– 79	12.0	12.8	80.1
$CV = 0.707$	$\hat{N}^0_{MQLE}$	319	– 81	11.5	12.5	81.8
$D_t = 287$	$\hat{N}^0_{SC}$	336	– 64	22.6	23.1	68.5
	$\hat{N}_{SC}$	373	– 27*	42.6	38.2	46.7*
	$\hat{N}_{MQLE}$	346	– 54	24.2	23.4	59.2
12	$\hat{N}^0_{MLE}$	299	– 101	5.5	8.1	101.0
$CV = 0.707$	$\hat{N}^0_{MQLE}$	298	– 102	5.3	8.1	102.2
$D_t = 287$	$\hat{N}^0_{SC}$	309	– 91	11.6	12.6	91.4
	$\hat{N}_{SC}$	324	– 76*	18.1	18.1	78.1*
	$\hat{N}_{MQLE}$	310	– 90	10.7	12.3	91.1
13	$\hat{N}^0_{MLE}$	373	– 27	19.8	21.4	34.4*
$CV = 0.354$	$\hat{N}^0_{MQLE}$	372	– 28	19.2	21.1	35.1
$D_t = 313$	$\hat{N}^0_{SC}$	377	– 23	30.5	27.6	35.7
	$\hat{N}_{SC}$	406	6	53.8	54.0	54.3
	$\hat{N}_{MQLE}$	396*	– 4*	34.9	36.5	36.8
14	$\hat{N}^0_{MLE}$	357	– 43	28.1	26.3	50.2
$CV = 0.447$	$\hat{N}^0_{MQLE}$	355	– 45	27.0	25.9	51.6
$D_t = 281$	$\hat{N}^0_{SC}$	364	– 36	46.9	38.4	52.4
	$\hat{N}_{SC}$	403	3*	84.8	76.4	76.5
	$\hat{N}_{MQLE}$	386	– 14	50.1	45.9	47.8*
15	$\hat{N}^0_{MLE}$	344	– 56	34.1	34.3	65.3
$CV = 0.500$	$\hat{N}^0_{MQLE}$	342	– 58	32.9	33.7	66.8
$D_t = 261$	$\hat{N}^0_{SC}$	352	– 48	58.6	49.6	69.3
	$\hat{N}_{SC}$	399	– 1*	109.2	109.2	109.2
	$\hat{N}_{MQLE}$	377	– 23	62.3	59.1	63.3*

Table 5. Continued.

Trial	*Estimator*	*Average Estimate*	*Average Bias*	*Average Estimated s.e.*	*Sample s.e.*	*Sample RMSE*
16	$\hat{N}^0_{MLE}$	330	− 70	41.9	38.0	80.1
$CV = 0.577$	$\hat{N}^0_{MQLE}$	327	− 73	42.0	37.5	81.9
$D_t = 239$	$\hat{N}^0_{SC}$	339	− 61	77.0	62.6	87.6
	$\hat{N}_{SC}$	389	− 11*	136.0	130.6	131.0
	$\hat{N}_{MQLE}$	364	− 36	77.4	66.0	75.4*

*Denotes the smallest bias or smallest RMSE.

are recommended for practical use. The two estimators $\hat{N}_{MQLE}$ and $\hat{N}_{SC}$ which incorporate the estimated CV are expected to have positive biases in homogeneous populations because the true CV is 0, whereas the estimates of it are slightly higher than 0.

As expected, when $CV > 0$, the three estimators $\hat{N}^0_{MLE}$, $\hat{N}^0_{MQLE}$ and $\hat{N}^0_{SC}$, without considering heterogeneity, consistently underestimate the true parameter, and the magnitude of the bias increases as the CV becomes larger. When the CV is relatively low ($CV = 0.316$), these estimators are still appropriate, but they become severely underestimated when the CV increases. Hence we only compare the two estimators $\hat{N}_{MQLE}$ and $\hat{N}_{SC}$ for Trials 4–16.

The relative merits of the proposed $\hat{N}_{MQLE}$ and the sample coverage estimator $\hat{N}_{SC}$ depend on the degree of heterogeneity. The sample coverage estimator generally yields a smaller bias but larger variation whereas the MQLE has a smaller variation but larger bias. Therefore our comparison criterion is based on the RMSE. For $CV = 0.316$, the proposed MQLE has a smaller RMSE than that of $\hat{N}_{SC}$; when $CV = 0.552$, both estimators are comparable; and in the case of $CV = 0.707$, the estimator based on sample coverage works better with respect to the RMSE. For the four random-effects models (Trials 13–16, $\alpha = 8, 5, 4, 3$ and the corresponding CVs are 0.354, 0.447, 0.500 and 0.577), the proposed estimator $\hat{N}_{MQLE}$ has a smaller RMSE than that of $\hat{N}_{SC}$.

Except for sparse data, the bootstrap standard error estimates are generally satisfactory compared with the sample standard errors. It is important to note that in our analysis we do need over 50% removal to produce reasonable estimates. If the removal rate is not sufficiently high, our estimators become unstable. Generally, sparse data provide relatively little information about the initial population size. The general guidelines about how large the removal rate should be are still unclear to us. However, as indicated by Gould and Pollock (1996, p. 896), it should depend on the population size and smaller populations require larger proportion of removals.

In summary, our proposed estimation procedure based on optimal estimation functions works equally well as the MLE for a homogeneous catch-effort model. For heterogeneous catch-effort models, our proposed estimator is generally more precise than the previous sample coverage estimator, although the latter is less biased. The proposed estimator is preferable in terms of RMSE if the degree of heterogeneity is relatively low or moderate. For highly heterogeneous populations, the sample coverage approach is recommended.

Appendix: Proof of Equations (2.12)–(2.17b) under a multiplicative model

Let the state vector $\boldsymbol{X}_j = (x_{1j}, x_{2j}, \ldots, x_{Nj})$ be defined as in Section 2.1. It is clear that

$$E(x_{ik}) = 1 - P(\text{the } i\text{th animal remains uncaptured in samples } 1-k)$$

$$= 1 - \prod_{j=1}^{k}(1-\lambda_i e_j). \quad \text{(A.1)}$$

Using the relationship $u_k = \sum_{i=1}^{N}(x_{ik} - x_{i,k-1})$, we have

$$E(u_k) = \sum_{i=1}^{N} \lambda_i e_k \left[\prod_{j=1}^{k-1}(1-\lambda_i e_j)\right]. \quad \text{(A.2)}$$

From the definition of sample coverage and Equation (A.1), the expected sample coverage can be expressed as

$$E(C_k) = \frac{\sum_{i=1}^{N}\lambda_i[1-\prod_{j=1}^{k}(1-\lambda_i e_j)]}{\sum_{i=1}^{N}\lambda_i} = 1 - \frac{E(u_{k+1})/e_{k+1}}{E(u_1)/e_1}.$$

Thus Equation (2.12) follows directly. Formulas (2.15) and (2.16) then follow from the definition of C_{k-1}, $D_{k-1} = \sum_{i=1}^{N} x_{i,k-1}$ and Equation (A.1).

Based on Equation (2.15) and (2.16), we obtain that

$$E(D^*_{k-1}) - E(D_{k-1}) = \sum_{i=1}^{N}\left[\prod_{j=1}^{k-1}(1-\lambda_i e_j)\right]\left(1-\frac{\lambda_i}{\bar{\lambda}}\right). \quad \text{(A.3)}$$

A simple expansion leads to Equation (2.17a) when all e_j's are equal to 1. Now we discuss the behavior of the remainder term of Δ_k in Equation (2.17a), where we have

$$\Delta_k = E(D^*_{k-1}) - E(D_{k-1}) \; - \; (k-1)\gamma^2 E(u_{k-1}).$$

If $\lambda_1, \lambda_2, \ldots, \lambda_N$ can be regarded as a random sample from a gamma distribution with density $\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}/\Gamma(\alpha)$, then

$$\frac{E(D_{k-1})}{N} = 1 - N^{-1}\sum_{i=1}^{N}(1-\lambda_i)^{k-1} \to 1 - E(1-\lambda)^{k-1} = 1 - \sum_{j=1}^{k-1}\binom{k-1}{j}(-1)^j E(\lambda^j),$$

$$\frac{E(D^*_{k-1})}{N} = 1 - \frac{\sum_{i=1}^{N}\lambda_i(1-\lambda_i)^{k-1}}{\sum_{i=1}^{N}\lambda_i} \to 1 - \frac{E[\lambda(1-\lambda)^{k-1}]}{E(\lambda)} = 1 - \sum_{j=1}^{k-1}\binom{k-1}{j}\frac{(-1)^j E(\lambda^{j+1})}{E(\lambda)},$$

$$\frac{E(u_{k-1})}{N} = N^{-1}\sum_{i=1}^{N}\lambda_i(1-\lambda_i)^{k-2} \to E[\lambda(1-\lambda)^{k-2}] = \sum_{j=1}^{k-2}\binom{k-2}{j}(-1)^j E(\lambda^{j+1}).$$

Using the above three formulas and substituting the moments $E(\lambda) = \alpha/\beta$, $\gamma^2 = 1/\alpha$ and $E(\lambda^k) = \Gamma(\alpha+k)/[\Gamma(\alpha)\beta^k]$, we can obtain that $\Delta_k/N \to 0$ for any fixed $k \geq 1$ as N is large enough. For the case of unequal efforts, a similar expansion of Equation (A.3) leads to

$$E(D^*_{k-1}) - E(D_{k-1}) = \gamma^2 \sum_{i=1}^{N} \left(\sum_{m=1}^{k-1} \lambda_i e_m \prod_{n \neq m}^{k-1} (1 - \lambda_i e_n) \right) + \Delta^*_k,$$

where Δ^*_k is the remainder term. Now based on the following approximation

$$\prod_{n \neq m}^{k-1} (1 - \lambda_i e_n) \approx \prod_{n=1}^{k-2} (1 - \lambda_i e_n),$$

we have

$$\sum_{m=1}^{k-1} \lambda_i e_m \prod_{n \neq m}^{k-1} (1 - \lambda_i e_n) \approx \sum_{m=1}^{k-1} \left(\frac{e_m}{e_{k-1}} \right) \left[\lambda_i e_{k-1} \prod_{n=1}^{k-2} (1 - \lambda_i e_n) \right]$$
$$= \sum_{m=1}^{k-1} \left(\frac{e_m}{e_{k-1}} \right) E(u_{k-1}).$$

Thus Equation (2.17b) follows.

Acknowledgements

The authors sincerely thank Dr Jon Schnute and two referees for their thoughtful suggestions and comments. Part of the research was supported by the National Science Council of Taiwan under contract NSC 86-2115-M007-033 and NSC 87-2118-M007-004.

References

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975) *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press, Cambridge, Massachusetts.

Buckland, S.T. and Garthwaite, P.H. (1991) Quantifying precision of mark-recapture estimates using bootstrap and related methods. *Biometrics*, **47**, 255–68.

Chao, A. (1989) Estimating population size for sparse data in capture-recapture experiments. *Biometrics*, **45**, 427–38.

Dahiya, R.C. (1981) An improved method of estimating an integer-parameter by maximum likelihood. *The American Statistician*, **35**, 34–7.

Darroch, J.N., Fienberg, S.E., Glonek, G.F.V., and Junker, B.W. (1993) A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. *Journal of the American Statistical Association*, **88**, 1137–48.

De Lury, D.B. (1947) On the estimation of biological populations. *Biometrics*, **3**, 145–67.

Fisher, R.A., Corbet, A.S., and Williams, C.B. (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, **12**, 42–58.

Gould, W.R. and Pollock, K.H. (1997) Catch-effort maximum likelihood estimation of important population parameters. *Canadian Journal of Fish and Aquatic Sciences*, **54**, 890–97.

Gove, J.H., Linder, E., and Tzilkowski, W.M. (1995) Bimodality of the combined removal and signs-of-activities estimator for sampling closed animal populations. *Environmental and Ecological Statistics*, **3**, 65–80.

Lee, S.-M. and Chao, A. (1994) Estimating population size via sample coverage for closed capture-recapture models. *Biometrics*, **50**, 88–97.

Leslie, P.H. and Davis, D.H.S. (1939) An attempt to determine the absolute number of rats on a given area. *Journal of Animal Ecology*, **8**, 94–113.

Liang, K.-Y. and Zeger, S.L. (1995) Inference based on estimating functions in the presence of nuisance parameters (with discussion). *Statistical Science*, **10**, 158–99.

Lindsay, B.G. and Roeder, K. (1987) A unified treatment of integer parameter models. *Journal of the American Statistical Association*, **82**, 758–64.

Mahon, R. (1980) Accuracy of catch-effort methods for estimating fish density and biomass in streams. *Environmental Biology of Fishes*, **5**, 343–63.

Moran, P.A.P. (1951) A mathematical theory of animal trapping. *Biometrika*, **38**, 307–11.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, Second Edition, Chapman and Hall, London.

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. (1981) Capture-recapture models: a review of current method, assumptions and experimental design. In *Estimating the Number of Terrestrial Birds, Studies in Avian Biology*, C.J. Ralph and J.M. Scott (eds), No. 6, pp. 426–35.

Pollock, K.H. (1991) Modeling capture, recapture and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present and future. *Journal of the American Statistical Association*, **86**, 225–38.

Pollock, K.H. and Otto, M.C. (1983) Robust estimation of population size in closed animal populations from capture-recapture experiments. *Biometrics*, **39**, 1035–49.

Ricker, W.E. (1958) Handbook of computations for biological statistics of fish populations. *Bulletin Fisheries Research Board of Canada*, **119**, 1–300.

Sanathanan, L. (1972) Estimating the size of a multinomial population. *The Annals of Mathematical Statistics*, **43**, 142–52.

Sanathanan, L. (1977) Estimating the size of a truncated sample. *Journal of the American Statistical Association*, **78**, 913–19.

Seber, G.A.F. (1982) *The Estimation of Animal Abundance*, 2nd edition. Griffin, London.

Seber, G.A.F. and Whale, J.F. (1970) The removal method for two and three samples. *Biometrics*, **26**, 393–400.

Schnute, J. (1983) A new approach to estimating populations by the removal method. *Canadian Journal of Fisheries and Aquatic Sciences*, **40**, 2153–69.

White, G.C., Anderson, D.R., Burnham, K.P., and Otis, D.L. (1982) Capture-recapture and removal methods for sampling closed populations. Los Alamos National Lab. Publication LA-8787-NERP.

Zippin, C. (1956) An evaluation of the removal method of estimating animal populations. *Biometrics*, **12**, 163–89.

Zippin, C. (1958) The removal method of population estimation. *Journal of Wildlife Management*, **22**, 82–90.

Biographical sketches

Anne Chao is Professor, Institute of Statistics, National Tsing Hua University, Taiwan. Her interest in ecological statistics began with analyzing birds' banding data in two estuaries of Taiwan. Since then, she has worked to enhance local people's understanding and appreciation of the country's dynamic treasure (estuaries and its bio-diversity). Specific methodological interests include species estimation, population size estimation and related biological samplings.

Shu-Hui Chang is Associate Professor, Department of Public Health, National Taiwan University. She received PhD degree in biostatistics from Johns Hopkins University in 1995. Her area of research interest is survival analysis and statistical issues in epidemiology.