Unified Maximum Likelihood Estimates for Closed Capture–Recapture Models Using Mixtures

Shirley Pledger

School of Mathematical and Computing Sciences, Victoria University of Wellington,
P.O. Box 600, Wellington, New Zealand
email: shirley.pledger@vuw.ac.nz

SUMMARY.   Agresti (1994, *Biometrics* **50**, 494–500) and Norris and Pollock (1996a, *Biometrics* **52**, 639–649) suggested using methods of finite mixtures to partition the animals in a closed capture–recapture experiment into two or more groups with relatively homogeneous capture probabilities. This enabled them to fit the models M_h, M_{bh} (Norris and Pollock), and M_{th} (Agresti) of Otis et al. (1978, *Wildlife Monographs* **62**, 1–135). In this article, finite mixture partitions of animals and/or samples are used to give a unified linear-logistic framework for fitting all eight models of Otis et al. by maximum likelihood. Likelihood ratio tests are available for model comparisons. For many data sets, a simple dichotomy of animals is enough to substantially correct for heterogeneity-induced bias in the estimation of population size, although there is the option of fitting more than two groups if the data warrant it.

KEY WORDS: Boundary estimation; Capture–recapture; Closed populations; Heterogeneity; Maximum likelihood; Mixture distribution; Multinomial model.

1. Introduction

There is an extensive literature on analyzing the eight closed-population capture–recapture models specified in Pollock (1974) and Otis et al. (1978). These models allow three possible factors to affect the probability of capture: time (probabilities vary from sample to sample), behavior (a trap response of animals to the first capture), and heterogeneity (different animals have intrinsically different capture probabilities).

Maximum likelihood (ML) has been used to fit some of the models (Fienberg, 1972; Otis et al., 1978; Cormack, 1989; Agresti, 1994; Norris and Pollock, 1995, 1996a), while other methods of model fitting in use include the jackknife (Burnham and Overton, 1978; Pollock and Otto, 1983), moment methods based on sample coverage (Chao, Lee, and Jeng, 1992; Lee and Chao, 1994; Lee, 1996), martingales (Yip, 1989, 1991; Lloyd and Yip, 1991; Lloyd, 1992; Yip and Fong, 1993), and Bayesian methods (Casteldine, 1981; Gazey and Staley, 1986; Bolfarine, Leite, and Rodrigues, 1992). Model selection has been done by various criteria, including discriminant analysis (Otis et al., 1978), likelihood ratio tests (LRTs) where available, an Akaike information criterion (AIC) (Burnham, White, and Anderson, 1995), or bootstrap-based LRTs (Norris and Pollock, 1996a). A discussion of these methods and model averaging is in Stanley and Burnham (1998).

A recurring theme has been methods of dealing with heterogeneity of capture among the animals, which causes negative bias in estimates of population size (N) in models assuming homogeneity. For the heterogeneous model M_h, estimators of N in wide use are Burnham and Overton's (1978) jackknife

and Chao et al.'s (1992) moment-based estimator using sample coverage.

Agresti (1994) and Norris and Pollock (1995, 1996a) use finite mixtures to allow for heterogeneity. In this article, the approach of Norris and Pollock is modified and extended to give maximum likelihood estimators (MLEs) for all eight models in Otis et al. (1978). The models are compared by likelihood ratio tests, using either Monte Carlo or nonstandard versions where appropriate. The pocket mouse (*Perognathus parvus*) data set of E. Larsen (Otis et al., 1978) is used to illustrate all the model fitting, and other examples are also given.

The desirability of a unified maximum-likelihood framework for fitting and comparing all the eight models of Otis et al. (1978) has long been recognized and is discussed in, e.g., Stanley and Burnham (1998). This article presents such a framework.

2. The Models

We assume a closed population of unknown size N is sampled k times, with independence between animals and across samples. At its first capture, an animal is permanently marked to make it recognizable on any subsequent captures. The capture of the ith animal at the jth sample is taken to be a Bernoulli trial with probability p_{ij} of capture. If D distinct animals are caught, the data form a $D \times k$ capture matrix with element $x_{ij} = 1$ if animal i is captured on occasion j; otherwise, $x_{ij} = 0$. Row i of the matrix is the capture history of animal i.

If there are N animals in the population, we may extend the capture matrix with $N - D$ rows of zeros to include the

uncaught animals. Using the multinomial assumptions of Otis et al. (1978) rather than the Poisson log-linear models of Fienberg (1972), Cormack (1989), and Agresti (1994), the likelihood for the extended capture matrix X is

$$L(N, p_{ij} \mid X) = \text{prob}(X \mid N, p_{ij})$$

$$= \frac{N!}{(\prod_h x_h!)(N-D)!} \prod_{i=1}^{N}$$

$$\times \prod_{j=1}^{k} \left\{ p_{ij}^{x_{ij}} (1 - p_{ij})^{1-x_{ij}} \right\}, \qquad (1)$$

where x_h is the number of observed animals with capture history h. In the model equations to follow, the term $\Pi_h\, x_h!$ is omitted, as it is not relevant to likelihood maximization or to model comparisons. The probabilities p_{ij} will be modeled as either fixed parameters or realizations of random variables.

2.1 *Model* M_{tbh}

The most general model of Otis et al. (1978) is M_{tbh}, which allows for time, behavior, and heterogeneity effects on the probability of capture. We start with model $M_{t \times b \times h_A}$, a fully interactive version of M_{tbh}. Assume there is a fixed number, A, of animal groups, within each of which there is homogeneity of capture. Which animals are in which group is not known; each animal is assumed to come independently from group a with probability π_a ($a = 1, \ldots, A$, $\Sigma\,\pi_a = 1$).

The probability of capture of animal i at sample j, p_{ij}, is assumed to be θ_{jba}, modeled with a linear logistic formulation,

$$\log\left(\frac{\theta_{jba}}{1 - \theta_{jba}}\right) = \mu + \tau_j + \beta_b + \eta_a + (\tau\beta)_{jb} + (\tau\eta)_{ja}$$

$$+ (\beta\eta)_{ba} + (\tau\beta\eta)_{jba}, \qquad (2)$$

where μ is a constant unknown parameter, τ_j (for time) is a fixed main effect depending on the sample j, β_b is a fixed behavior main effect (where $b = b_{ij} = 1$ if animal i was not caught before sample j; otherwise, $b = 2$), and η_a (for heterogeneity) is a (random) animal main effect, where $a = 1, \ldots, A$ with probabilities $\pi_1, \ldots, \pi_A$, respectively. The later terms are three two-way interactions and the three-way

interaction. No parameters $(\tau\beta)_{12}$ or $(\tau\beta\eta)_{12a}$ are needed, as no recaptures can occur in sample 1. To make the parameters identifiable, constraints are needed, e.g., either a corner-point parameterization or sum-to-zero constraints. Corner-point constraints could be defined by setting to zero any parameters with $j = 1$, $b = 1$, or $a = 1$. Also, with no parameters for recapture at sample 1, we set $(\tau\beta)_{22} = (\tau\beta\eta)_{22a} = 0$ for all $a = 1, \ldots, A$.

With these constraints, the numbers of independent parameters associated with the terms of logit(θ_{jba}) are μ, 1; τ_j, $k-1$; β_b, 1; η_a, $A-1$; $(\tau\beta)_{jb}$, $k-2$; $(\tau\eta)_{ja}$, $(A-1)(k-1)$; $(\beta\eta)_{ba}$, $A-1$; and $(\tau\beta\eta)_{jba}$, $(A-1)(k-2)$. These $2Ak - A$ independent parameters together with N and the $A-1$ independent π parameters for group membership give $2Ak$ independent parameters for model $M_{t \times b \times h_A}$.

The observed capture history for animal i has probability $\Sigma_{a=1}^{A} \pi_a \Pi_{j=1}^{k} \{\theta_{jba}^{x_{ij}}(1 - \theta_{jba})^{1-x_{ij}}\}$. Integrating equation (1) over each animal's independent group allocation gives likelihood

$$L(N, \{\pi\}, \{\theta\} \mid X)$$

$$= \frac{N!}{(N-D)!}$$

$$\times \prod_{i=1}^{N} \sum_{a=1}^{A} \left[\pi_a \prod_{j=1}^{k} \left\{ \left(\theta_{jba}\right)^{x_{ij}} \left(1 - \theta_{jba}\right)^{1-x_{ij}} \right\} \right]. \quad (3)$$

This general, fully interactive M_{tbh} model is labeled $M_{t \times b \times h_A}$. One submodel is the main effects model, M_{t+b+h}, in which only the constant and main effects terms of equation (2) are retained.

Model $M_{t \times b \times h_A}$ and its various submodels are specified in Table 1. Only hierarchical models are proposed: If an interaction term is present, so are all its associated main effects and lower order interactions. Table 1 omits a group of models between the fully interactive model $M_{t \times b \times h_A}$ and the main effects model M_{t+b+h_A}, these being the partially interactive models, which include the three main effects and some or all of the two-way interactions but not the three-way interaction.

Table 1

$M_{t \times b \times h_A}$ *and its submodels;* $\theta =$ *probability of capture and* N_{par} *is the number of independent parameters; there are also partially interactive models between* $M_{t \times b \times h_A}$ *and* M_{t+b+h_A}

Model	N_{par}	θ	Logit(θ)
$M_{t \times b \times h_A}$	$2Ak$	θ_{jba}	$\mu + \tau_j + \beta_b + \eta_a + (\tau\beta)_{jb} + (\tau\eta)_{ja} + (\beta\eta)_{ba} + (\tau\beta\eta)_{jba}$
M_{t+b+h_A}	$2A + k$	θ_{jba}	$\mu + \tau_j + \beta_b + \eta_a$
$M_{t \times b}$	$2k$	θ_{jb}	$\mu + \tau_j + \beta_b + (\tau\beta)_{jb}$
M_{t+b}	$k + 2$	θ_{jb}	$\mu + \tau_j + \beta_b$
$M_{t \times h_A}$	$Ak + A$	θ_{ja}	$\mu + \tau_j + \eta_a + (\tau\eta)_{ja}$
M_{t+h_A}	$2A + k - 1$	θ_{ja}	$\mu + \tau_j + \eta_a$
$M_{b \times h_A}$	$3A$	θ_{ba}	$\mu + \beta_b + \eta_a + (\beta\eta)_{ba}$
M_{b+h_A}	$2A + 1$	θ_{ba}	$\mu + \beta_b + \eta_a$
M_t	$k + 1$	θ_j	$\mu + \tau_j$
M_b	3	θ_b	$\mu + \beta_b$
M_{h_A}	$2A$	θ_a	$\mu + \eta_a$
M_0	2	θ	μ

The main effects models in this article are additive on the logit scale. An alternative approach, which can equally well be put into this framework, is to model capture probabilities with main effects additive on a log scale. These will be referred to as proportional models. Lee (1996) uses proportional main effects to model p_{ij} in M_{tbh} via conditional sample coverage. Lee and Chao (1994) also make the proportional assumption in a model of M_{tbh} in which catch-effort information is used to allow for the time effect. There may be problems in fitting the proportional models, i.e., a multiplier to allow for an extra factor may take a capture probability estimate above one. However, if capture probabilities are below 0.5, the proportional model and our logistic main effects model give similar estimates (Pledger, 1998).

Models are fitted by maximizing the log likelihood using the EM algorithm for unknown group membership. Details are available from the author. Profile likelihood intervals are given in this article. The advantages of these over traditional confidence intervals are discussed by Cormack (1992), Hirst (1994), and other authors.

2.2 Homogeneous Submodels

The submodels without heterogeneity, M_0, M_t, M_b, and $M_{t \times b}$, specified in Table 1 are reparameterizations of those discussed in Otis et al. (1978) and give the same ML estimates. The interactive model $M_{t \times b}$ cannot be fitted due to unidentifiable parameters. Otis et al. propose a model in which two parameters are arbitrarily set equal, but they comment that this is likely to be biologically unrealistic. Burnham's power relationship estimator, which is provided in the package CAPTURE, also makes a rather restrictive assumption, and a warning is given in Rexstad and Burnham (1991, p. 28). Our main effects model is feasible but is similar to Burnham's model and has the same potential problems.

2.3 Heterogeneous Submodels

Our model M_{h_A} is like the multinomial-based M_h of Norris and Pollock (1996a) but with one difference. Norris and Pollock condition on N, then fit a model with an unknown number of groups using nonparametric maximum likelihood estimation, NPMLE. There is a global maximum to the likelihood (Laird, 1978; Lindsay, 1983), and the number of groups is selected by finding this maximum. Then N is varied to find the overall maximum likelihood. In this article, a sequence of models, each with a fixed number of animal groups, is fitted without conditioning on N, and model comparisons are used to decide the number of animal groups. Often the two procedures lead to the same results, but they differ in that Norris and Pollock's procedure selects more animal groups because of a higher likelihood, while here, a significant increase of likelihood (decrease of residual deviance) is required before another animal group is added (see Section 3). Where different numbers of groups are chosen, Norris and Pollock's NPMLE has less bias, more parameters, and higher standard errors. With our sequential approach, as A increases, the residual deviances and $\hat{N}$ approach the NPMLE stable value, as noted by Aitkin (1996) when fitting such a sequence of models in the case of normal mixtures. Aitkin (1999) in Example 1 selects a two-component model, even though the NPMLE has three groups, since the two models give almost identical estimates. The models in

this article may have multimodalities in the likelihood function, as discussed in Section 4.

For M_{h_A}, the capture probability is θ_a, with

$$\log\left(\theta_a/(1-\theta_a)\right) = \mu + \eta_a,$$

and the likelihood is

$$
L(N, \{\pi\}, \{\theta\} \mid X)
$$
$$
= \frac{N!}{(N-D)!} \prod_{j=0}^{D} \left[\left\{ \sum_{a=1}^{A} \pi_a \theta_a^j (1-\theta_a)^{k-j} \right\}^{f_j} \right], \quad (4)
$$

where f_j is the number of animals with j captures, and so $f_0 = N - D$. With $2A$ independent parameters and k minimal sufficient statistics $f_1, \ldots, f_k$, we require $2A \leq k$ for model fitting to be possible. We fit a sequence of models with one, two, three, or more animal groups ($M_{h_1} = M_0$, M_{h_2}, M_{h_3}, etc.), comparing them via likelihood ratio tests.

Model M_{th} has interactive ($M_{t \times h_A}$) and main effects ($M_{t + h_A}$) versions and also a proportional version if the main effects are assumed additive on the log scale. The interactive model allows the different animal groups to have capture probabilities responding in different ways to the different samples, while the main effects and proportional models assume similar patterns of response through time. For example, lizards may all have low capture probabilities on dull days while exhibiting high heterogeneity of capture rates on bright days. The interactive model would be more suitable here. Agresti (1994) uses a main effects model with finite mixtures for a Poisson-based model of M_{th}. Chao et al. (1992), with their moment-based sample coverage method, use a proportional assumption for capture probability, i.e., $p_{ij} = p_i e_j$, where p_i is the capture probability for animal i and e_j is a multiplier for sample j.

Previous estimators of N for model M_{bh} are the generalized removal estimator of Otis et al. (1978), the jackknife and moment-based estimators of Pollock and Otto (1983), the sample coverage estimator of Lee and Chao (1994), and the mixture models of Norris and Pollock (1995, 1996a). Most of these are removal estimators, using data only up until the first capture for each animal, but Norris and Pollock (1995) use an interactive finite mixture model that includes both initial and subsequent capture information in the model. With interaction, the groups may respond differently to the first capture. For example, one group may have a high probability of capture throughout, while another may be initially timid but become emboldened after the first capture. The interactive model $M_{b \times h_A}$ specified in Table 1 matches Norris and Pollock's except that we fix the number of animal groups. The likelihood of equation (3) reduces to

$$
L(N, \{\pi\}, \{\theta\} \mid X)
$$
$$
= \frac{N!}{(N-D)!}
$$
$$
\times \left\{ \prod_{i=1}^{D} \sum_{a=1}^{A} \pi_a \theta_{1a} (1-\theta_{1a})^{w_i-1} \theta_{2a}^{r_i} (1-\theta_{2a})^{k-w_i-r_i} \right\}
$$
$$
\times \left\{ \sum_{a=1}^{A} \pi_a (1-\theta_{1a})^k \right\}^{N-D}, \quad (5)
$$

where w_i is the sample number at which animal i is first captured and r_i is its number of recaptures. The main effects model M_{b+h_A} has the same likelihood but omits the $b \times h$ interaction in the logistic formulation of the capture probability. An alternative would be the proportional model, with main effects additive on the log scale. The removal model, $M_{bh_A} \cdot r$ say, is used if either there are no data available on recaptures or it is thought undesirable to base the animal partition into groups on recaptures as well as first captures. This model, similar to that in Norris and Pollock (1996a), employs only data up to the first capture, and in equation (5), the terms in θ_{2a} are omitted.

2.4 *Models with Random Time Effects*

A major problem with the models presented so far is the large number of parameters assigned to the time factor. This spendthrift approach to the time effect parameters has also been noted by Jolly (1982) in the context of open population models. Another problem is that model $M_{t \times b}$ is unable to be fitted, as it has unidentifiable parameters. Both problems are solved if time (as well as heterogeneity) is modeled as a random effect.

The most general model is $M_{t_S \times b \times h_A}$, in which the samples are partitioned into S groups, $s = 1, \ldots, S$, and the animals are partitioned into A groups, $a = 1, \ldots, A$, a two-dimensional cross-partition of both rows and columns of the capture matrix. Each animal has probability π_a of belonging in animal group a, and each sample has probability λ_s of being in sample group s ($\Sigma \pi_a = 1$ and $\Sigma \lambda_s = 1$). For animal group a and sample group s, the probability of capture is modeled as

$$\log \left(\frac{\theta_{sba}}{1 - \theta_{sba}} \right) = \mu + \tau_s + \beta_b + \eta_a + (\tau\beta)_{sb} + (\tau\eta)_{sa}$$
$$+ (\beta\eta)_{ba} + (\tau\beta\eta)_{sba},$$

with b defined as before and with corner-point parameterization constraints of zero parameters if $s = 1$, $b = 1$, or $a = 1$. The likelihood is obtained by, first, integrating equation (1) over the different animals' group allocations, giving equation (3), and then integrating over the different samples' allocations into sample group. This gives likelihood

$$L(N, \{\pi\}, \{\lambda\}, \{\theta\} \mid X)$$

$$= \frac{N!}{(N-D)!} \sum_{s_k=1}^{S} \lambda_{s_k} \ldots \sum_{s_2=1}^{S} \lambda_{s_2} \sum_{s_1=1}^{S} \lambda_{s_1}$$

$$\times \prod_{i=1}^{N} \sum_{a=1}^{A} \left[\pi_a \prod_{j=1}^{k} \left\{ (\theta_{abs_j})^{x_{ij}} (1 - \theta_{abs_j})^{1-x_{ij}} \right\} \right], \quad (6)$$

where s_j is the allocation of sample j. The $2AS$ capture probability parameters, together with $A - 1$ independent π parameters, $S - 1$ independent λ parameters, and N, give a model with $2AS + A + S - 1$ parameters. If, as tends to happen, only two animal groups and two sample groups are needed, this means 11 parameters. By contrast, the model with fixed time effects and two animal groups has $4k$ parameters, e.g., 24 if $k = 6$.

The fully interactive model $M_{t_S \times b \times h_A}$ has new submodels $M_{t_S + b + h_A}$ (the main effects model), $M_{t_S \times b}$, $M_{t_S + b}$, $M_{t_S \times h_A}$, $M_{t_S + h_A}$, and M_{t_S}. There are also partially interactive models

between $M_{t_S \times b \times h_A}$ and $M_{t_S + b + h_A}$ and a proportional model alternative to each main effects model. All are feasible, and they usually have fewer parameters than their counterparts with fixed time effects. Of particular value is $M_{t_S \times b}$, as $M_{t \times b}$ was not feasible. There may well be an interaction between the behavior and time effects, with, e.g., more trap shyness on a bright day when the traps are clearly visible and less trap shyness on dull days when traps are not easily seen.

2.5 *Models with Covariates*

If relevant covariates are available for either the heterogeneity (e.g., animal sex, weight) or for the samples (e.g., catch-effort or weather variables), these are easily incorporated into this framework. For example, suppose there is heterogeneity and we have a useful search-effort covariate, say e_j, measured at the time of each sample. The capture probability could be modeled as

$$\log \left(\frac{\theta_{ja}}{1 - \theta_{ja}} \right) = \mu + \alpha e_j + \eta_a + \gamma_a e_j,$$

with constraints $\eta_1 = \gamma_1 = 0$. Here $\mu + \eta_a$ is the intercept for group a's logit capture probability versus e_j and $\alpha + \gamma_a$ is the slope. This model uses finite mixtures to model heterogeneity, while the effort variable allows for the time effect. The $\gamma_a e_j$ term would be omitted for a main effects (parallel lines) model. Multiple covariates may also be included, as in Pollock, Hines, and Nichols (1984) and Huggins (1989, 1991).

3. Model Comparison

A model selection procedure is now needed, and we must distinguish two situations depending on whether the regularity conditions for a likelihood ratio test (LRT) hold.

Between two nested models, standard likelihood ratio tests are valid provided no extra animal groups or sample groups are being introduced as we move from the submodel to the supermodel. For example, comparison of models M_{h_2} and M_{b+h_2} involves the testing of one parameter, $H_0 : \beta_2 = 0$. This is not a test at the edge of the parameter space, and so it is a regular LRT, in this case with null distribution χ_1^2. Other examples are given in Section 5.1.

However, if two nested models differ in the number of animal groups or sample groups (e.g., M_0 versus M_{h_2}), parameters at the boundary under H_0 mean that conditions for the standard LRT are not met. We then use a Monte Carlo LRT (parametric bootstrap LRT; cf., Manly, 1997). Populations are generated using the submodel with estimated parameters from the study, and the chosen test statistic is calculated for each population. This shows whether the observed test statistic is unusual in relation to these generated test statistics. Any nested pair of models may be compared in this way and also any pair of nonnested models by using the method of Williams (1970), as described in Brooks et al. (1997).

A faster alternative for any model comparison that introduces a two-group finite mixture is the nonstandard LRT given in Self and Liang (1987, Theorem 3, case 5). The null distribution of the LRT statistic is a 50:50 mixture of zeros and χ_1^2, which we label $0 : \chi_1^2$. Although conditions for the theorem are not met exactly due to a nuisance parameter disappearing under H_0, simulations confirmed that it is approximately correct for tests of M_0 versus M_{h_2} and of

M_b versus M_{b+h_2} (Pledger, 1998). A further nonstandard test from Self and Liang (1987, Theorem 3, case 6) allows a direct test of M_b versus $M_{b \times h_2}$, or M_b versus $M_{t_2 \times b}$ without going through the intermediate (possibly unrealistic) main effects model. This null distribution is a 50:50 mixture of χ_1^2 and χ_2^2 variates, labeled $\chi_1^2 : \chi_2^2$. For example, in testing M_b versus $M_{b \times h_2}$, the null hypothesis is H_0: $\eta_2 = 0$ and $(\beta \eta)_{22} = 0$. The test statistic is the sum of two variates, one log-likelihood ratio with a $0 : \chi_1^2$ distribution for M_b versus M_{b+h_2} and another with χ_1^2 null distribution for M_{b+h_2} versus $M_{b \times h_2}$. The sum is thus the 50:50 mixture of χ_1^2 and χ_2^2.

The addition of a third, fourth, or higher group is not covered by Self and Liang's (1987) theorems. Monte Carlo likelihood ratio tests are retained here, e.g., for M_{h_2} versus M_{h_3}.

4. Appraisals

An article with detailed appraisals of these models and comparisons with previously available models is in preparation. Some preliminary results from Pledger (1998, 1999) are given briefly here.

To investigate accuracy, bias, and precision of $\hat{N}$ for M_{h_A}, we used the same eight settings of parameters as in Norris and Pollock (1996a) to simulate M_{h_3} data, with results in Table 2. This indicates that, in this region of the sample space, going from M_0 to M_{h_2} corrects most of the bias in $\hat{N}$, with diminished returns in moving to M_{h_3}. For the fixed models, the root mean square error (RMSE) criterion selects model M_0 as best, followed by M_{h_2} and M_{h_3}, and for selection of the number of groups, a 5% LRT is preferable to the NPMLE of Norris and Pollock. With RMSE, high precision is outweighing the bias. However, the median absolute deviation (MAD) prefers more groups, and NPMLE is preferable to the 5% LRT, as this criterion gives more importance to bias reduction.

A check of sensitivity to misspecification of the distribution was done by simulating heterogeneous populations with a beta distribution of probability of capture and then analyzing the data using M_0, M_{h_2}, and M_{h_3}. (Model fitting using the beta mixing distribution often fails, as noted by Burnham and Overton [1978]. This is a common problem with infinite mixtures.) Parameter settings (k, N, and the mean and variance of capture probability) were again chosen to match Norris and Pollock's (1996a) choices. Details are in Pledger (1998, 1999); the results are close to those in Table 2, indicating that modeling the different animals' probability of

capture as a finite mixture distribution is acceptable even if the true distribution is an infinite mixture (in this case, beta binomial). This is in line with the comment in Lindsay (1995, Section 3.1) in comparing discrete and continuous mixture models: "My own preference is for the use of discrete distributions in cases of doubt.... A number of investigations have found that misspecifying the latent distribution has very little effect on bias, and minimal effect on standard errors beyond the necessary correction for overdispersion when mixing is present." The two-point support is often able to provide enough variability to model data from a distribution with more support points, even infinitely many.

The power to detect a time effect (M_t versus M_0) is higher than the power to detect a similar coefficient of variation of capture probabilities between animals (M_{h_2} versus M_0). This asymmetry is intrinsic to capture–recapture studies because of the oblong capture matrix, i.e., there is more information available per sample than per animal.

With a fixed number of animal groups, there may be multimodality in the likelihood function. Brooks et al. (1997) found it useful to obtain profile likelihoods plotted against the mixing probability (π_1) when detecting multimodality in mixture models with two components. Using this, we found in the 1600 simulated M_{h_3} populations only 23 cases where model M_{h_2} had multimodality. All of these were detected in our standard modeling routine, which checked the next model in the sequence, M_{h_3}, and compared it with the M_{h_2} fit. Details are in Pledger (1998, 1999), but overall it seems that multimodality of the M_{h_2} likelihood is fairly rare, even with data generated from more than two support points, and may be dealt with by a routine check of a higher model.

5. Examples

5.1 *Example 1: Pocket Mice*

E. Larsen (as reported in Otis et al., 1978) conducted a capture–recapture study on pocket mice (*Perognathus parvus*) for seven consecutive nights in Curlew Valley, Utah, capturing 55 distinct animals. Table 3 shows the results of fitting all the models to the pocket mouse data. Model $M_{t \times b \times h_2}$ is not considered as a valid option: with $\hat{N}$ at its lower bound of 55 and 4 of the 26 capture probability estimates (not shown) also at their boundaries, the data are not able to support this model. However, $M_{t_2 \times b \times h_2}$ is available as an alternative.

Using a forward selection procedure, from model M_0, the first step is to M_{h_2} ($p < 0.0001$ with a nonstandard $0 : \chi_1^2$ likelihood ratio test). The other options were M_b ($p = 0.0036$,

Table 2
Mean accuracy of $\hat{N}$ over the $2^3 = 8$ parameter settings of Norris and Pollock (1996a): $k = (10, 20)$, $N = (50, 100)$, three equiprobable groups with capture probabilities $(0.1, 0.2, 0.3)$ or $(0.1, 0.3, 0.5)$; 200 simulations per setting

Accuracy measure	Fixed model			Selected model	
	M_0	M_{h_2}	M_{h_3}	NPMLE	5% LRT
Mean bias	−5.28	−1.27	−0.36	−0.05	−1.62
Standard deviation	2.54	5.94	6.81	7.60	6.12
Root mean square error	5.91	6.21	6.85	7.64	6.39
Median absolute deviation	5.20	3.58	3.46	3.25	3.74

Table 3
Pocket mouse models and N estimates

Model	Number of parameters	Residual deviance[a]	$\hat{N}$
M_0	2	185.4	55.9
M_t	8	176.9	60.6
M_{t_2}	4	180.5	55.8
M_{t_3}	4	180.5	55.8
M_b	3	176.9	60.6
M_{h_2}	4	143.9	62.3
M_{h_3}	6	143.9	62.3
M_{t+b}	9	163.2	72.2
$M_{t \times b}$	NE	NE	NE
M_{t_2+b}	5	174.5	62.0
$M_{t_2 \times b}$	6	174.5	62.0
M_{t+h_2}	10	125.8	61.6
$M_{t \times h_2}$	16	117.9	61.6
$M_{t_2+h_2}$	6	139.3	62.2
$M_{t_2 \times h_2}$	7	137.7	62.1
M_{b+h_2}	5	142.5	69.7
$M_{b \times h_2}$	6	138.2	74.6
M_{t+b+h_2}	11	124.9	57.2
$M_{t \times b \times h_2}$	28	93.7	55.0
$M_{t_2+b+h_2}$	7	137.3	62.1
$M_{t_2 \times b \times h_2}$	11	136.7	62.3

[a] NE, not estimable.

a standard χ_1^2 test), M_{t_2} ($p = 0.0142$ with the nonstandard test on $0{:}\chi_1^2$), or M_t ($p = 0.0227$, a standard χ_6^2 test).

From M_{h_2}, there is no need to introduce another animal group, as the residual deviance for M_{h_3} is the same as for M_{h_2}. M_{h_2} was tested against all the M_{th} and M_{bh} models. We select a move to $M_{t_2+h_2}$ ($p = 0.0041$, $0{:}\chi_1^2$). Other options checked were $M_{t_2 \times h_2}$ ($p = 0.0202$, $\chi_1^2{:}\chi_2^2$), M_{t+h_2} ($p = 0.0060$, χ_6^2), $M_{t \times h_2}$ ($p = 0.0107$, χ_{12}^2), M_{b+h_2} ($p = 0.2394$, χ_1^2), or $M_{b \times h_2}$ ($p = 0.0578$, χ_2^2).

We stay with model $M_{t_2+h_2}$. Neither $M_{t_2 \times h_2}$ ($p = 0.2542$, χ_1^2) nor $M_{t_2+b+h_2}$ ($p = 0.4795$, χ_1^2) offered a substantially better fit to the data. Our chosen model $M_{t_2+h_2}$ has $\hat{N} = 62.2$ with a 95% profile likelihood interval of [56, 73]. (Assuming $N = 62$, 200 bootstrapped samples of 62 capture histories show a coverage of 94.5%.) The corresponding fixed-time model M_{t+h_2} is very similar, with $\hat{N} = 61.6$ and 95% profile likelihood interval [57, 73]. In both models, the grouping puts most mice in a low capture group, with a few having high probability of capture.

It is interesting that after heterogeneity is allowed for, the next most useful factor to include is the time effect. If heterogeneity is not allowed for, behavior appears more important than time. This kind of effect is common in stepwise regression. Our model choice differs from that of the CAPTURE package, which uses discriminant analysis to select model M_{tbh} or M_{bh}.

5.2 *Example 2: Cottontail Rabbits*

Edwards and Eberhardt (1967) did a capture–recapture study on a penned population of 135 wild cottontail rabbits, with $k = 18$ samples. Otis et al. (1978) reported high heterogeneity and low capture rates, making estimates unreliable.

Residual deviances are 375.59 for M_0, 361.62 for M_{h_2}, and 361.42 for M_{h_3}, with no change for M_{h_4}. M_{h_2} is chosen over M_0 ($p = 0.0001$ on $0{:}\chi_1^2$, with similar results from a Monte Carlo test), but the test statistic of $361.62 - 361.42 = 0.2$ is not large enough to choose M_{h_3} ($p = 0.09$ with a Monte Carlo test using 100 simulated populations).

The N estimates and 95% profile likelihood intervals are 96.3 [88, 107] for M_0, 135.5 [104, 347] for M_{h_2}, and 156.6 [105, 575720] for M_{h_3}. Using 200 bootstrapped samples of 135 capture histories, the 95% coverages of the intervals are 12.5% for M_0, 93.5% for M_{h_2}, and 93.0% for M_{h_3}, showing good coverage for the heterogeneous models.

The choice of model M_{h_2} is gratifying, as $\hat{N} = 135.5$, very close to the known value of 135. However, between M_{h_2} and M_{h_3}, $\hat{N}$ is sensitive to a small change in residual deviance; this sensitivity has not been found with other data sets. Also, the high upper bounds of the profile likelihood intervals for the heterogeneous models warn of extreme unreliability of the N estimates. Overall, it would be better to pool the samples, say, in groups of 3–6 samples with reasonably high capture probabilities.

5.3 *Example 3: Reid's Deer Mice*

Otis et al. (1978) discuss V. Reid's data set of deer mouse captures. There are 38 individuals caught over six trapping occasions, and the age group, sex, and weight of each animal is given. The CAPTURE model selection procedure chose M_b as the most suitable model. Huggins (1991) fitted his conditional likelihood models to these data (conditioning on the captured animals), incorporating heterogeneity by using age, sex, and weight covariates, as

$$\ln\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_0 + \beta_{\text{sex}} + \beta_{\text{age}} + \beta_{\text{wt}} \times \text{weight}(i)$$

where sex and age are included as factors each at two levels (the three semiadults being recorded as adults) and weight is included as a continuous covariate with coefficient β_{wt}. Huggins' method selects M_{bh}, using the AIC. This gives $\hat{N} = 47.1$.

Our finite mixture method, without using any covariates, chooses M_{b+h_2} (5% significance level), giving $\hat{N} = 45.6$. There is agreement of model choice and similarity of $\hat{N}$ using the two very different modeling methods. For situations in which useful heterogeneity covariates may not be available, finite mixtures may be able to provide a good alternative model.

6. Discussion

The option of modeling animals and/or samples as having random effects on the probability of capture has made available maximum likelihood estimates for a wide range of models. For unity, a linear logistic framework was chosen for the probabilities of capture. Either Monte Carlo or nonstandard likelihood ratio tests provide methods of model comparison to select the number of groups and the most appropriate model(s).

The finite-mixture linear-logistic models blend seamlessly with the covariate models in capture–recapture. Covariates, if available, are an alternative to finite mixtures for modeling

time effects, heterogeneity, or both. There is no barrier to a hybrid approach, modeling time with covariates and heterogeneity with finite mixtures, or vice versa.

These models could equally well have used the Norris and Pollock (1995, 1996a) NPMLE approach, in which the number of animal and/or sample groups is chosen by maximum likelihood within a model. The method proposed here, with different models for different numbers of groups, has the disadvantage of possible multimodality in the likelihood surface but the advantage of more parsimony in the number of parameters.

The behavior effect has been modeled at two levels, one for a first capture and the other for recapture. However, other choices could be specified. Pollock (1975) gives a general model in which capture probability depends on previous capture history, and Cormack (1989) suggests a limited memory model, in which the animal exhibits only temporary trap-response behavior.

Model selection by LRT often chooses only two groups, which can give an acceptable compromise in reducing bias in $\hat{N}$ while not inflating standard errors too much. This does not imply that we believe there really are two groups of animals; the grouping is essentially an artefact to allow for heterogeneity and thus correct for the bias in $\hat{N}$. This unreliability of grouping is implicit in Lindsay (1995, Section 1.4.2): "It is not uncommon that the goodness-of-fit of a mixture model to a data set does not change very much if we switch from a continuous latent distribution to a discrete one, or whether the discrete distribution has two components or four components." While estimates of N from finite mixture models are useful, estimates of π and θ parameters have high variance, as the likelihood surface is flat in those directions. The use of the π and θ estimates, e.g., to estimate a coefficient of variation of capture probabilities, is not recommended. (If, however, there is some reason to believe the animal groupings are valid, posterior probabilities of group membership may be found for each animal after estimating the parameters [Pledger, 1998]. These might be compared with unused covariates like sex to see how well they match. Similarly, posterior allocation of samples to sample groups could be compared with search effort or weather records.)

A desirable future tool would be a modified AIC to enable us to compare any two models. The traditional AIC is not correct for finite mixture models, as it relies on regularity conditions essentially the same as those for LRTs (Titterington, Smith, and Makov, 1985). This is being investigated.

These mixture models will be applied to open populations, modeling heterogeneity in capture rates and/or survival rates. They have immediate application to the robust design models of Pollock (1982) and Kendall, Pollock, and Brownie (1995), and their use in simple sampling scheme models like the Jolly–Seber is being investigated. The cross-partitions may have applications in latent class models and in random treatment and block models in experimental design.

Model uncertainty may be allowed for, as in Norris and Pollock (1996b), or by model averaging (Buckland, Burnham, and Augustin, 1997). Stanley and Burnham (1998) say: "We suspect that for the closed models of Otis et al. (1978), estimation and model selection need to be placed in a unified

likelihood framework.... In such a framework, information-theoretic approaches to model selection could be employed and could be used to supply data-specific weights for model averaging...." The results of this article could be applied in this way.

RÉSUMÉ

Agresti (1994, *Biometrics* **50**, 494–500) et Norris et Pollock (1996, *Biometrics* **52**, 639–649) ont suggéré d'utiliser des méthodes de mélanges finis pour partitionner les animaux, dans une expérience de capture-recapture au sein d'une population fermée, en deux groupes, ou plus, de probabilité de capture relativement homogènes. Cela leur a permis d'ajuster les modèles M_h, M_{bh} (Norris et Pollock) et M_{th} (Agresti) de Otis et al. (1978, *Wildlife Monographs* **62**, 1–135). Dans cet article, des partitions en mélange fini d'animaux et/ou d'échantillons sont utilisées pour donner un cadre unifié logistique-linéaire afin d'ajuster l'intégralité des huit modèles de Otis et al. par la méthode du maximum de vraisemblance. Des tests de rapport de vraisemblance sont disponibles pour la comparaison de modèles. Pour de nombreux jeux de données, une simple dichotomie des animaux est suffisante pour corriger de façon substantielle le biais introduit par l'hétérogénéité de capture sur l'estimateur de l'effectif de la population, quoiqu'on ait l'option d'ajuster plus de deux groupes si les données le permettent.

REFERENCES

Agresti, A. (1994). Simple capture–recapture models permitting unequal catchability and variable sampling effort. *Biometrics* **50**, 494–500.

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing* **6**, 251–262.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* **55**, 117–128.

Bolfarine, H., Leite, J. G., and Rodrigues, J. (1992). On the estimation of the size of a finite and closed population. *Biometrical Journal* **34**, 577–593.

Brooks, S. P., Morgan, B. J. T., Ridout, M. S., and Pack, S. E. (1997). Finite mixture models for proportions. *Biometrics* **53**, 1097–1115.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics* **53**, 603–618.

Burnham, K. P. and Overton, K. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**, 625–633.

Burnham, K. P., White, G. C., and Anderson, D. R. (1995). Model selection strategy in the analysis of capture–recapture data. *Biometrics* **51**, 888–898.

Casteldine, B. J. (1981). A Bayesian analysis of multiple capture–recapture sampling for a closed population. *Biometrika* **68**, 197–210.

Chao, A., Lee, S.-M., and Jeng, S.-L. (1992). Estimating population size for capture–recapture data when capture probabilities vary by time and individual animal. *Biometrics* **48**, 201–216.

Cormack, R. M. (1989). Log-linear models for capture–recapture. *Biometrics* **45**, 395–413.

Cormack, R. M. (1992). Interval estimation for mark–recapture studies of closed populations. *Biometrics* **48**, 567–576.

Edwards, W. R. and Eberhardt, L. L. (1967). Estimating cottontail abundance from live trapping data. *Journal of Wildlife Management* **31**, 87–96.

Fienberg, S. E. (1972). The multiple-recapture census for closed populations and incomplete contingency tables. *Biometrika* **59**, 591–603.

Gazey, W. J. and Staley, M. J. (1986). Population estimation from mark–recapture experiments using a sequential Bayes algorithm. *Ecology* **67**, 941–951.

Hirst, D. (1994). An improved removal method for estimating animal abundance. *Biometrics* **50**, 501–505.

Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika* **76**, 133–140.

Huggins, R. M. (1991). Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics* **47**, 725–732.

Jolly, G. M. (1982). Mark–recapture models with parameters constant in time. *Biometrics* **38**, 301–321.

Kendall, W. L., Pollock, K. H., and Brownie, C. (1995). A likelihood-based approach to capture–recapture estimation of demographic parameters under the robust design. *Biometrics* **51**, 293–308.

Laird, N. M. (1978). Nonparametric likelihood maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* **73**, 805–811.

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

Lee, S.-M. (1996). Estimating population size for capture–recapture data when capture probabilities vary by time, behavior and individual animal. *Communications in Statistics—Simulation* **25**, 431–457.

Lindsay, B. G. (1983). The geometry of mixture likelihoods: A general theory. *Annals of Statistics* **11**, 86–94.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications,* Volume 5, *NSF-CBMS Regional Conference Series in Probability and Statistics.* Hayward, California: Institute of Mathematical Statistics.

Lindsay, B. G. and Roeder, K. (1992). Residual diagnostics for mixture models. *Journal of the American Statistical Association* **87**, 785–794.

Lloyd, C. J. (1992). Testing capture homogeneity in a recapture model. *Biometrika* **79**, 555–561.

Lloyd, C. J. and Yip, P. (1991). A unification of capture–recapture experiments through martingale estimating functions. In *Estimating Functions,* V. P. Godambe (ed), 65–88. Oxford: Clarendon Press.

Manly, B. F. J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology,* 2nd edition. London: Chapman and Hall.

Norris, J. L. and Pollock, K. H. (1996a). Nonparametric MLE under two closed capture–recapture models with heterogeneity. *Biometrics* **52**, 639–649.

Norris, J. L., III, and Pollock, K. H. (1995). A capture–recapture model with heterogeneity and behavioural response. *Environmental and Ecological Statistics* **2**, 305–313.

Norris, J. L., III, and Pollock, K. H. (1996b). Including model uncertainty in estimating variances in multiple capture studies. *Journal of Ecological and Environmental Statistics* **3**, 235–244.

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.

Pledger, S. A. (1998). *Finite mixtures in closed capture–recapture models.* Research Report 98-18, School of Mathematical and Computing Sciences, Victoria University of Wellington, New Zealand.

Pledger, S. A. (1999). Finite mixtures in capture–recapture models. Ph.D. dissertation, Victoria University of Wellington, New Zealand.

Pollock, K. H. (1974). The assumption of equal catchability of animals in tag–recapture experiments. Ph.D. dissertation, Cornell University, Ithaca, New York.

Pollock, K. H. (1975). A K-sample tag–recapture model allowing for unequal survival and catchability. *Biometrika* **62**, 577–583.

Pollock, K. H. (1982). A capture–recapture design robust to unequal probability of capture. *Journal of Wildlife Management* **46**, 757–760.

Pollock, K. H. and Otto, M. C. (1983). Robust estimation of population size in closed animal populations from capture–recapture experiments. *Biometrics* **39**, 1035–1049.

Pollock, K. H., Hines, J. E., and Nichols, J. D. (1984). The use of auxiliary variables in capture–recapture and removal experiments. *Biometrics* **40**, 329–340.

Rexstad, E. and Burnham, K. P. (1991). *User's Guide for Interactive Program CAPTURE.* Fort Collins, Colorado: Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.

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.

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions.* New York: Wiley.

Williams, D. A. (1970). Discussion of Atkinson, A. C. (1970). A method of discriminating between models. *Journal of the Royal Statistical Society, Series B* **32**, 323–353.