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 $M_h$, $M_{th}$ (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 (Castakdine, 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 $i$th animal at the $j$th 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
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

\[
L(N, p_{ij} \mid X) = \text{prob}(X \mid N, p_{ij}) = \frac{N!}{(\prod_i x_i)! (N - D)!} \prod_{i=1}^{N} \left\{ \sum_{h} p_{ij} \left( 1 - p_{ij} \right)^{1-x_{ij}} \right\},
\]

(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 \( M_{tnbh} \), 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 \( \pi_a \) (\( a = 1, \ldots, A \)).

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

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

(2)

where \( \mu \) is a constant unknown parameter, \( \tau_j \) (for time) is a fixed main effect depending on the sample \( j \), \( \beta_b \) is a fixed behavior main effect (where \( b = b_{ij} = 1 \) if animal \( i \) was not caught before sample \( j \); otherwise, \( b = 2 \)), and \( \eta_a \) (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 \( \logit(\theta_{jba}) \) are \( \mu, 1; \tau_j, k - 1; \beta_b, 1; \eta_a, A - 1; (\tau \beta)_{jb}, k - 2; (\tau \eta)_{ja}, (A - 1)(k - 1); (\beta \eta)_{ba}, A - 1; \) and \( (\tau \beta \eta)_{jba}, (A - 1)(k - 2) \). These \( 2Ak - A \) independent parameters together with \( N \) and the \( A - 1 \) independent \( \pi \) parameters for group membership give \( 2Ak \) independent parameters for model \( M_{tbxhxhA} \).

The observed capture history for animal \( i \) has probability \( \pi_i = \frac{1}{\sum_a \pi_a^{(a-1)} \left\{ \sum_{j=1}^{k} \left( \theta_{jba} \right)^{x_{ij}} \right\} \} \). Integrating equation (1) over each animal's independent group allocation gives likelihood

\[
L(N, \{ \pi \}, \{ \theta \} \mid X) = \frac{N!}{(N - D)!} \prod_{i=1}^{N} \left\{ \sum_{a=1}^{A} \pi_a^{(a-1)} \left\{ \sum_{j=1}^{k} \left( \theta_{jba} \right)^{x_{ij}} \right\} \} \right\},
\]

(3)

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

Model \( M_{tbxhxhA} \) 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 \( M_{tbxhxhA} \) and the main effects model \( M_{tb+b+h} \), 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**

<table>
<thead>
<tr>
<th>Model</th>
<th>( N_{par} )</th>
<th>( \theta )</th>
<th>Logit(( \theta ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_{tbxhxhA} )</td>
<td>( 2Ak )</td>
<td>( \theta_{jba} )</td>
<td>( \mu + \tau_j + \beta_b + \eta_a + (\tau \beta)<em>{jb} + (\tau \eta)</em>{ja} + (\beta \eta)<em>{ba} + (\tau \beta \eta)</em>{jba} )</td>
</tr>
<tr>
<td>( M_{tb+b+h} )</td>
<td>( 2A + k )</td>
<td>( \theta_{jba} )</td>
<td>( \mu + \tau_j + \beta_b + \eta_a )</td>
</tr>
<tr>
<td>( M_{tb+b} )</td>
<td>( 2k )</td>
<td>( \theta_{jba} )</td>
<td>( \mu + \tau_j + \beta_b )</td>
</tr>
<tr>
<td>( M_{tb+h} )</td>
<td>( k + 2 )</td>
<td>( \theta_{jba} )</td>
<td>( \mu + \tau_j + \beta_b )</td>
</tr>
<tr>
<td>( M_{tbx} )</td>
<td>( Ak + A )</td>
<td>( \theta_{jba} )</td>
<td>( \mu + \tau_j + \eta_a + (\tau \eta)_{ja} )</td>
</tr>
<tr>
<td>( M_{tb+h} )</td>
<td>( 2A + k - 1 )</td>
<td>( \theta_{jba} )</td>
<td>( \mu + \tau_j + \eta_a )</td>
</tr>
<tr>
<td>( M_{tb} )</td>
<td>( 3A )</td>
<td>( \theta_{jba} )</td>
<td>( \mu + \beta_b + \eta_a + (\beta \eta)_{ba} )</td>
</tr>
<tr>
<td>( M_{tb} )</td>
<td>( 2A + 1 )</td>
<td>( \theta_{jba} )</td>
<td>( \mu + \beta_b + \eta_a )</td>
</tr>
<tr>
<td>( M_{tb} )</td>
<td>( k + 1 )</td>
<td>( \theta_{jba} )</td>
<td>( \mu + \tau_j )</td>
</tr>
<tr>
<td>( M_{tb} )</td>
<td>( 3 )</td>
<td>( \theta_{jba} )</td>
<td>( \mu + \beta_b )</td>
</tr>
<tr>
<td>( M_{tb} )</td>
<td>( 2A )</td>
<td>( \theta_{jba} )</td>
<td>( \mu + \eta_a )</td>
</tr>
<tr>
<td>( M_{tb} )</td>
<td>( 2 )</td>
<td>( \theta )</td>
<td>( \mu )</td>
</tr>
</tbody>
</table>
The main effects models in this article are additive on the log 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 $M_{i,th}$ via conditional sample coverage. Lee and Chao (1994) also make the proportional assumption in a model of $M_{i,th}$ 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, $M_0$, $M_1$, $M_h$, and $M_{1,th}$ specified in Table 1 are reparameterizations of those discussed in Otis et al. (1978) and give the same ML estimates. The interactive model $M_{1,th}$ cannot be fitted due to unidentified 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 $M_{h,h}$ 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 model, even though the NPMLE has three groups, since 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 $M_{h,h}$, the capture probability is $\theta_{a}$, with

$$\log \left( \frac{\theta_{a}}{1 - \theta_{a}} \right) = \mu + \eta_{a},$$

and the likelihood is

$$L(N, \{\pi\}, \{\theta\} | X) = \frac{N!}{(N-D)!} \prod_{j=0}^{D} \left\{ \sum_{a=1}^{A} \pi_{a} \theta_{a}^{j}(1 - \theta_{a})^{k-j} \right\} f_{j},$$

(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 $M_{i,th}$ has interactive ($M_{i,th,h}$) and main effects ($M_{i,th,h}$) 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 (1984) uses a main effects model with finite mixtures for a Poisson-based model of $M_{i,th}$. 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 $M_{i,th}$ 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 $M_{i,th,h}$ 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

$$L(N, \{\pi\}, \{\theta\} | X) = \frac{N!}{(N-D)!} \prod_{j=0}^{D} \left\{ \sum_{a=1}^{A} \pi_{a}\theta_{a}^{j}(1 - \theta_{a})^{k-j} \right\} f_{j},$$

(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 \( M_{b+h} \) 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, \( M_{b+h} \), say, is used if either there are no data available on recaptures or it is thought desirable to base the animal partition into groups on recaptures as well as first captures. This model, similar to that in Norris and Pollock (1996), employs only data up to the first capture, and in equation (5), the terms in \( \theta_{1a} \) 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 \( M_{t\times b\times h} \), 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-division of both rows and columns of the capture matrix. Each animal has probability \( \pi_a \) of belonging to animal group \( a \), and each sample has probability \( \lambda_s \) 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

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

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

\[
L(N, \{\pi\}, \{\lambda\}, \{\theta\} | X) = \frac{N!}{(N-D)!} \sum_{s_k=1}^S \lambda_{s_k} \sum_{s_2=1}^S \lambda_{s_2} \sum_{s_1=1}^S \lambda_{s_1} \times \prod_{i=1}^N \prod_{a=1}^A \prod_{j=1}^k \left( \{\theta_{sba}\}^{x_{ij}}(1 - \theta_{sba})^{1-x_{ij}} \right),
\]

where \( s_j \) is the allocation of sample \( j \). The 2AS capture probability parameters, together with \( A - 1 \) independent \( \pi \) parameters, \( S - 1 \) independent \( \lambda \) 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 \( M_{tS\times b\times h} \) has new submodels \( M_{tS\times b\times h} \) (the main effects model), \( M_{tS\times b}, M_{tS\times b}, M_{tS\times h}, M_{tS\times h}, \) and \( M_{tS} \). There are also partially interactive models between \( M_{tS\times b\times h} \) and \( M_{tS\times b\times h} \) 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 \( M_{tS\times b} \), 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

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

with constraints \( \eta_1 = \gamma_1 = 0 \). Here \( \mu + \eta_a \) is the intercept for group \( a \), \( \alpha \) 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 \( M_{b2} \) and \( M_{b+h2} \) 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 \( \chi^2 \). 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 \( M_{b2} \)), 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 \( \chi^2_1 \), which we label 0.5. 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 \( M_{b2} \) and of...
was done by simulating heterogeneous populations with a mixtures.) Parameter settings (\(k, N\), and the mean and variance of capture probability) were again chosen to match the data using Norris and Pollock’s NPMLE of Norris and Pollock's models, the root mean square error (RMSE) criterion selects the 5% LRT, as this criterion gives more importance to bias. However, the median absolute deviation of probability of capture and then analyzing the 5% LRT, as this criterion gives more importance to bias 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.

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 \(N\) for \(M_{h+1}\), we used the same eight settings of parameters as in Norris and Pollock (1996a) to simulate \(M_{h}\) data, with results in Table 2. This indicates that, in this region of the sample space, going from \(M_0\) to \(M_{h}\) corrects most of the bias in \(N\), with diminished returns in moving to \(M_{h+1}\). For the fixed models, the root mean square error (RMSE) criterion selects model \(M_0\) as best, followed by \(M_{h+1}\) and \(M_{h}\), 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 \(M_0\), \(M_{h}\), and \(M_{h+1}\). (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 (\(M_i\) versus \(M_0\)) is higher than the power to detect a similar coefficient of variation of capture probabilities between animals (\(M_{h+1}\) versus \(M_h\)). 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 (\(\rho_1\)) when detecting multimodality in mixture models with two components. Using this, we found in the 1600 simulated \(M_{h}\) populations only 23 cases where model \(M_h\) had multimodality. All of these were detected in our standard modeling routine, which checked the next model in the sequence, \(M_{h+1}\), and compared it with the \(M_h\) fit. Details are in Pledger (1998, 1999), but overall it seems that multimodality of the \(M_h\) 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 \(M_{1\times b\times h}\) is not considered as a valid option: with \(N\) 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, \(M_{2\times b\times h}\) is available as an alternative. Using a forward selection procedure, from model \(M_0\), the first step is to \(M_{h}\) (\(p < 0.0001\) with a nonstandard 0:1 likelihood ratio test). The other options were \(M_0\) (\(p = 0.0036, \ldots\)

---

**Table 2**

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

<table>
<thead>
<tr>
<th>Accuracy measure</th>
<th>Fixed model</th>
<th>Selected model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M_0)</td>
<td>(M_{h})</td>
</tr>
<tr>
<td>Mean bias</td>
<td>-5.28</td>
<td>-1.27</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>2.54</td>
<td>5.94</td>
</tr>
<tr>
<td>Root mean square error</td>
<td>5.91</td>
<td>6.21</td>
</tr>
<tr>
<td>Median absolute deviation</td>
<td>5.20</td>
<td>3.58</td>
</tr>
</tbody>
</table>
United Maximum Likelihood Estimates

Table 3
Pocket mouse models and N estimates

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of parameters</th>
<th>Residual deviancea</th>
<th>( \hat{N} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_0 )</td>
<td>2</td>
<td>185.4</td>
<td>55.9</td>
</tr>
<tr>
<td>( M_1 )</td>
<td>8</td>
<td>176.9</td>
<td>60.6</td>
</tr>
<tr>
<td>( M_2 )</td>
<td>4</td>
<td>180.5</td>
<td>55.8</td>
</tr>
<tr>
<td>( M_3 )</td>
<td>4</td>
<td>180.5</td>
<td>55.8</td>
</tr>
<tr>
<td>( M_b )</td>
<td>3</td>
<td>176.9</td>
<td>60.6</td>
</tr>
<tr>
<td>( M_{h2} )</td>
<td>4</td>
<td>143.9</td>
<td>62.3</td>
</tr>
<tr>
<td>( M_{b3} )</td>
<td>6</td>
<td>143.9</td>
<td>62.3</td>
</tr>
<tr>
<td>( M_{b+b} )</td>
<td>9</td>
<td>163.2</td>
<td>72.2</td>
</tr>
<tr>
<td>( M_{m+} )</td>
<td>NE</td>
<td>NE</td>
<td>NE</td>
</tr>
<tr>
<td>( M_{b+b} )</td>
<td>5</td>
<td>174.5</td>
<td>62.0</td>
</tr>
<tr>
<td>( M_{b+b} )</td>
<td>6</td>
<td>174.5</td>
<td>62.0</td>
</tr>
<tr>
<td>( M_{t+h2} )</td>
<td>10</td>
<td>125.8</td>
<td>61.6</td>
</tr>
<tr>
<td>( M_{t+h2} )</td>
<td>16</td>
<td>117.9</td>
<td>61.6</td>
</tr>
<tr>
<td>( M_{t+h2} )</td>
<td>6</td>
<td>139.3</td>
<td>62.2</td>
</tr>
<tr>
<td>( M_{t+h2} )</td>
<td>7</td>
<td>137.7</td>
<td>62.1</td>
</tr>
<tr>
<td>( M_{t+h2} )</td>
<td>5</td>
<td>142.5</td>
<td>69.7</td>
</tr>
<tr>
<td>( M_{t+b} )</td>
<td>6</td>
<td>138.2</td>
<td>74.6</td>
</tr>
<tr>
<td>( M_{t+b+h2} )</td>
<td>11</td>
<td>124.9</td>
<td>57.2</td>
</tr>
<tr>
<td>( M_{t+b+h2} )</td>
<td>28</td>
<td>93.7</td>
<td>55.0</td>
</tr>
<tr>
<td>( M_{t+b+h2} )</td>
<td>7</td>
<td>137.3</td>
<td>62.1</td>
</tr>
<tr>
<td>( M_{t+b+h2} )</td>
<td>11</td>
<td>136.7</td>
<td>62.3</td>
</tr>
</tbody>
</table>

a NE, not estimable.

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 \( M_{h2} \), and 361.42 for \( M_{h3} \), with no change for \( M_h \). \( M_{h2} \) is chosen over \( M_0 \) (\( p = 0.0001 \), 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 \( M_{h3} \) (\( p = 0.09 \) 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 \( M_{h2} \), and 156.6 [105, 575720] for \( M_{h3} \). Using 200 bootstrapped samples of 135 capture histories, the 95% coverages of the intervals are 12.5% for \( M_0 \), 93.5% for \( M_{h2} \), and 93.0% for \( M_{h3} \), showing good coverage for the heterogeneous models.

The choice of model \( M_{h2} \) is gratifying, as \( \hat{N} = 135.5 \), very close to the known value of 135. However, between \( M_{h2} \) and \( M_{h3} \), \( N \) 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 \( M_8 \) 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

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

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

Our finite mixture method, without using any covariates, chooses \( M_{t+h2} \) (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 $N$ 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 artifact to allow for heterogeneity and thus correct for the bias in $N$. 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 $\pi$ and $\theta$ parameters have high variance, as the likelihood surface is flat in those directions. The use of the $\pi$ and $\theta$ 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.

ACKNOWLEDGEMENTS

I thank Ken Pollock for early and continuing encouragement and helpful comments and Victoria University of Wellington for the research leave that enabled me to become temporarily trap shy about encounters with students. My thanks also to Richard Barker, Murray Efford, David Fletcher, Bryan Manly, and Gary White for valuable feedback at an informal seminar on this work. I am grateful to two referees and an associate editor for helpful comments and references.

RÉSUMÉ


REFERENCES


