Biometrics 61, 1010–1017
December 2005
DOI: 10.1111/j.1541-0420.2005.00372.x

Modeling Animals' Behavioral Response by Markov Chain Models for Capture–Recapture Experiments

Hsin-Chou Yang* **and Anne Chao**

Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043
*email: hsinchou@ibms.sinica.edu.tw

Summary. A bivariate Markov chain approach that includes both enduring (long-term) and ephemeral (short-term) behavioral effects in models for capture–recapture experiments is proposed. The capture history of each animal is modeled as a Markov chain with a bivariate state space with states determined by the capture status (capture/noncapture) and marking status (marked/unmarked). In this framework, a conditional-likelihood method is used to estimate the population size and the transition probabilities. The classical behavioral model that assumes only an enduring behavioral effect is included as a special case of the bivariate Markovian model. Another special case that assumes only an ephemeral behavioral effect reduces to a univariate Markov chain based on capture/noncapture status. The model with the ephemeral behavioral effect is extended to incorporate time effects; in this model, in contrast to extensions of the classical behavioral model, all parameters are identifiable. A data set is analyzed to illustrate the use of the Markovian models in interpreting animals' behavioral response. Simulation results are reported to examine the performance of the estimators.

Key words: Behavioral response; Ecological model; Mark-recapture; Markov chain; Population size.

1. Introduction

Multiple capture–recapture models are widely applied to estimate the population size and other parameters in closed populations. When identical trapping methods are used, empirical studies have provided evidence that mice, voles, small mammals, and rodents among others often exhibit a response to capture. A classical model that incorporates behavioral response to initial capture has been extensively discussed (Otis et al., 1978; Pollock, 1991) especially for animal studies. The model is referred to as model $\mathbf{M}_b$ in the literature, where the subscript b denotes behavioral response.

The classical model $\mathbf{M}_b$ assumes that animals exhibit an enduring behavioral response to the first capture. That is, after an animal has been captured once, the animal has a long memory of its first-capture experience and the effect lasts for the remainder of the experiment, leading to a higher (trap-happy) or lower (trap-shy) capture probability for all subsequent recaptures. Thus, all unmarked animals have one probability of capture and all marked animals have another probability of capture, and any animal's capture probability changes only once, at the time of its first capture.

For animals with "short-term" memory, the capture probability of an animal may be ephemeral and only depend on whether or not it was caught on the most recent occasion. This type of behavioral response model can be naturally modeled by a univariate Markov chain. Although Markov chain models have been adopted in many other biological applications (e.g., Solow, 2000), and there is a rich theory concerning the models, it seems these models have not been applied to model the behaviors of animals in capture–recapture experiments.

In this article, we consider a bivariate Markov chain approach that includes both enduring and ephemeral behavioral effects. The capture history of each animal is modeled as a Markov chain with state space determined by both capture status (capture/noncapture) and marking status (marked/unmarked). In the Markov approach, the capture probability for any marked animal can thus be altered more than once. This model includes both the classical model $\mathbf{M}_b$ and the univariate Markov chain model as special cases. Moreover, time effects and covariates can be incorporated in the transition probabilities to form a class of Markov chain models. A unified conditional-likelihood-based inference procedure (Huggins, 1989, 1991) is used to obtain estimators and the asymptotic variances of the population size and the transition probabilities.

Section 2 introduces the general framework of the bivariate Markov model. In Section 2.1, we deal with a Markov model with both enduring and ephemeral effects. Two submodels (enduring effect only and ephemeral effect only) are discussed in Sections 2.2 and 2.3, respectively. In Section 3.1, the extension to include time effects for the ephemeral effect model is presented. A Markov chain with logistic models for the transition probabilities that incorporate covariates is briefly described in Section 3.2. A data set is analyzed in Section 4 to illustrate the use of the Markov models in interpreting behavioral response. Simulation results are reported in Section 5 to examine the performance of the estimators. Concluding remarks and discussion are given in Section 6.

2. Markov Chain Approach

Consider a closed population of size N. A capture–recapture experiment is conducted over T sampling occasions (or samples). Denote the T occasions by the time points $t = 1, 2, \ldots, T$ and let the starting time of the experiment correspond to the initial time $t = 0$. Assume that all animals behave independently. Let M denote the total number of distinct animals that are caught in the experiment. Without loss of generality, we assume that those animals caught in the experiment are labeled from 1 to M and those not captured are labeled from $M + 1$ to N. The observed capture–recapture histories can be expressed as an $M \times T$ matrix $[X_{it}]$, where

$$X_{it} = I\,[\text{the } i\text{th animal is caught at time } t],$$

where $I[\cdot]$ denotes the usual indicator function. Let ω be a subset of $\{1, 2, \ldots, T\}$ and let Z_ω denote the number of animals captured in each of the samples indexed by elements of ω, and in no others. For animal i at time t, we define a state vector $(X_{it},\ Y_{it})$, where

$$Y_{it} = I\,[\text{the } i\text{th animal is marked at time } t].$$

So $Y_{it} = 1$ if animal i has been captured at least once in $(1, 2, \ldots, t)$. Then the state of an animal at time t is determined by the capture status (capture/noncapture at time t) and marking status (marked/unmarked by time t). Because $X_{it} = 1$ implies $Y_{it} = 1$, the state vector $(X_{it}, Y_{it}) = (1, 0)$ is not possible and there are only three states $\{(0, 0), (0, 1), (1, 1)\}$. For notational convenience, we let $a = (0, 0)$, $b = (0, 1)$, and $c = (1, 1)$.

The bivariate Markov chain model assumes that (X_{it}, Y_{it}) is a discrete-time stochastic process with state space $E = \{a, b, c\}$ for any animal with an initial state a, that is, $P\{(X_{i0}, Y_{i0}) = a\} = 1$. Let s_j denote the state vector at time j. The Markov chain model assumes for $t = 1, 2, \ldots, T$ that

$$\begin{aligned} P\{(X_{it}, Y_{it}) = s_t \mid (X_{i,t-1}, Y_{i,t-1}) = s_{t-1}, \ldots, (X_{i0}, Y_{i0}) = a\} \\ = P\{(X_{it}, Y_{it}) = s_t \mid (X_{i,t-1}, Y_{i,t-1}) = s_{t-1}\}. \end{aligned}$$

If animal i is in state a at time $t-1$, the probability of making a transition into state a at time t is defined as

$$P_{aa}(i,t) = P\{(X_{it}, Y_{it}) = a \mid (X_{i,t-1}, Y_{i,t-1}) = a\}. \quad (1)$$

The transition probabilities $P_{ab}(i, t)$, $P_{ba}(i, t), \ldots, P_{cc}(i, t)$ are similarly defined. It is readily seen that $P_{ab}(i, t) = P_{ba}(i, t) = P_{ca}(i, t) = 0$, $P_{aa}(i, t) + P_{ac}(i, t) = 1$, $P_{bb}(i, t) + P_{bc}(i, t) = 1$, and $P_{cb}(i, t) + P_{cc}(i, t) = 1$.

Define $n_{aa}(t)$ to be the number of transitions from state a at time $t-1$ to state a at time t in the *observed* capture history matrix. That is, for $t = 1, 2, \ldots, T$, we have

$$n_{aa}(t) = \sum_{i=1}^{M} I[(X_{i,t-1}, Y_{i,t-1}) = a, (X_{it}, Y_{it}) = a].$$

Similarly, define $n_{ac}(t)$, $n_{bb}(t)$, $n_{bc}(t)$, $n_{cb}(t)$, and $n_{cc}(t)$. We use the notational convention that when we sum over an index, the subscript corresponding to that index is replaced by a "+" sign. For example, $n_{a+}(t) = n_{aa}(t) + n_{ac}(t)$. It is obvious that $n_{bb}(1) = n_{bc}(1) = n_{cb}(1) = n_{cc}(1) = 0$, $n_{bb}(2) = n_{bc}(2) = 0$, $n_{++}(t) = M$ for $t \geq 1$.

In this article, we do not consider individual heterogeneity unless it can be explained by covariates. Thus, except for Section 3.2, we shall drop the index i and use (X_t, Y_t) instead of (X_{it}, Y_{it}) because all transition probabilities are independent of animal i.

2.1 *Model $\mathbf{MM}_2$(b): With Both Enduring and Ephemeral Behavioral Effects*

We first consider the case when the transition probabilities are time homogeneous. Denote these probabilities by P_{aa}, P_{ac}, P_{bb}, P_{bc}, P_{cb}, and $P_{cc}(P_{ab} = P_{ba} = P_{ca} = 0)$. This two-dimensional Markov chain model with both types of behavioral effects is denoted by $\mathbf{MM}_2$(b). The corresponding transition probability matrix is shown below.

State vector	$(X_{t+1},\ Y_{t+1})$ $= a = (0, 0)$	$b = (0, 1)$	$c = (1, 1)$
$(X_t,\ Y_t) = a = (0, 0)$	$1 - P_{ac}$	0	P_{ac}
$b = (0, 1)$	0	$1 - P_{bc}$	P_{bc}
$c = (1, 1)$	0	$1 - P_{cc}$	P_{cc}

Notice that $P_{bc} - P_{ac}$ measures an enduring behavioral effect as it is the difference in capture probability of a marked and an unmarked; and $P_{cc} - P_{bc}$ measures an ephemeral behavioral effect as this effect only concerns the transition immediately after a capture. If $P_{bc} = P_{ac}$, then there is no enduring effect and the model reduces to a model with an ephemeral effect only so that the effect of capture on the transition probabilities is limited to the next capture occasion. This model is called model $\mathbf{MM}_1$(b); see Section 2.2. If $P_{cc} = P_{bc}$, then the model reduces to a classical model $\mathbf{M}_\text{b}$ with an enduring effect only.

Let $n_{aa} = \sum_{t=1}^{T} n_{aa}(t)$ be the total number of transitions from state a to state a in the experiment and similarly define n_{ac}, n_{bb}, n_{bc}, n_{cb}, and n_{cc}. The parameters in the model are $\{N, P_{ac}, P_{bc}, P_{cc}\}$ and the minimum sufficient statistics are $\{n_{aa}, n_{ac}, n_{bb}, n_{bc}, n_{cb}, n_{cc}\}$. The statistic M is not included because $n_{ac} = M$. The likelihood under model $\mathbf{MM}_2$(b) can be expressed as

$$\begin{aligned} &L(N, P_{ac}, P_{bc}, P_{cc}) \\ &\quad = \frac{N!}{\left(\prod_{\omega} Z_\omega!\right)(N-M)!} \\ &\qquad \times \left[(1 - P_{ac})^{T(N-M)+n_{aa}} (P_{ac})^{n_{ac}} \right. \\ &\qquad \left. \times (1 - P_{bc})^{n_{bb}} (P_{bc})^{n_{bc}} \times (1 - P_{cc})^{n_{cb}} (P_{cc})^{n_{cc}}\right]. \quad (2) \end{aligned}$$

Let $Q = (P_{aa})^T = (1 - P_{ac})^T$ be the probability that an animal is not captured in the experiment and let $\boldsymbol{\eta} = \{P_{ac}, P_{bc}, P_{cc}\}$ be the set of transition probabilities. Observe that the likelihood function may be factorized as $L(N, \boldsymbol{\eta}) = L_b(N, \boldsymbol{\eta}) \times L_c(\boldsymbol{\eta})$ where

$$L_b(N, \boldsymbol{\eta}) = \frac{N!}{M!(N-M)!} Q^{N-M} (1-Q)^M \quad (3)$$

and

$$L_c(\boldsymbol{\eta}) = \frac{M!}{\left(\prod_{\omega} Z_\omega!\right)(1-Q)^M} \times \left\{(1-P_{ac})^{n_{aa}}(P_{ac})^{n_{ac}} \times (1-P_{bc})^{n_{bb}}(P_{bc})^{n_{bc}} \times (1-P_{cc})^{n_{cb}}(P_{cc})^{n_{cc}}\right\}. \tag{4}$$

Here $L_b(N, \boldsymbol{\eta})$ is a binomial likelihood with respect to the random variable M, and $L_c(\boldsymbol{\eta})$ is a conditional likelihood with respect to all observable counts. The latter is a multinomial probability, allocating the M animals to the observed cells. The usual unconditional MLE (UMLE) is obtained by maximizing $L(N, \boldsymbol{\eta})$ with respect to N and $\boldsymbol{\eta}$. For the conditional MLE (CMLE), $\hat{\boldsymbol{\eta}}$ is first computed by maximizing the conditional likelihood $L_c(\boldsymbol{\eta})$ and the CMLE of N is then obtained by maximizing $L_b(N, \hat{\boldsymbol{\eta}})$ with respect to N.

There are four reasons for adopting a conditional approach: (i) Both captured and noncaptured animals are considered in a full likelihood, but the covariates for the noncaptured animals are not observable. Thus, the usual likelihood method cannot be easily extended to models incorporating covariates. This difficulty can be avoided by using a likelihood conditional on the captured animals, so that the covariates of noncaptured animals are not needed; see Section 3.2 for the general models with covariates. (ii) For large population sizes, the two types of MLEs are asymptotically equivalent (Sanathanan, 1972) and lead to close estimates and variances, though for finite population sizes they might differ to some extent (Chao, Chu, and Hsu, 2000). (iii) The CMLE is also the Horvitz–Thompson-type estimator (Horvitz and Thompson, 1952). (iv) The CMLE is scale invariant whereas the UMLE is not (Chao et al., 2000).

It is shown in the Appendix that the CMLE $\hat{N}$ of N is the solution of the equation:

$$1 - \frac{M}{N} = \left(1 - \frac{n_{ac}}{T(N-M) + n_{a+}}\right)^T, \tag{5}$$

and the transition probabilities are estimated by

$$\hat{P}_{ac} = \frac{n_{ac}}{T(\hat{N}-M) + n_{a+}}, \quad \hat{P}_{bc} = \frac{n_{bc}}{n_{b+}}, \quad \hat{P}_{cc} = \frac{n_{cc}}{n_{c+}}. \tag{6}$$

Sanathanan (1972) showed that the UMLE and the CMLE have identical asymptotic variance, which can be obtained by inverting the expected Fisher information matrix. After some manipulations, the asymptotic variance is given by

$$\mathrm{Var}(\hat{N}) \cong N\left\{-1 + \frac{1}{Q} + \frac{T^2}{\sum_{t=1}^{T} P[(X_{t-1}, Y_{t-1}) = a]}\left(1 - \frac{1}{P_{aa}}\right)\right\}^{-1}. \tag{7}$$

Under model $\mathbf{MM}_2$(b), we have $\sum_{t=1}^{T} P[(X_{t-1}, Y_{t-1}) = a] = 1 + P_{aa} + P_{aa}^2 + \cdots + P_{aa}^{T-1} = (1-Q)/(1-P_{aa})$. This implies that

$$\mathrm{Var}(\hat{N}) \cong N\left\{-1 + \frac{1}{Q} + \frac{T^2(1-P_{aa})}{(1-Q)}\left(1 - \frac{1}{P_{aa}}\right)\right\}^{-1}. \tag{8}$$

A variance estimator can thus be obtained by substituting estimates for parameters in this variance formula. A confidence interval of N is constructed by using a log transformation (Chao, 1987) so that the lower bound is at least M.

2.2 *Model* $\mathbf{MM}_1$*(b): Ephemeral Behavioral Effect Only*

If we assume that there is no enduring effect, that is, $P_{ac} = P_{bc}$ in model $\mathbf{MM}_2$(b), then the transition probabilities only depend on whether an animal has been captured or not at each time. This ephemeral effect model becomes the univariate Markov chain model based on $\{X_t\}$ and is denoted by $\mathbf{MM}_1$(b). We can use a simpler notation p_{01} and p_{11} for transition probabilities. That is, $P_{ac} = P_{bc} \equiv p_{01}$ and $P_{cc} \equiv p_{11}$. Here, $p_{11} > p_{01}$ implies that a captured animal at the current time has a higher probability of being captured at the next time than a noncaptured one. This is the case of a Markov-type "trap-happy." Similarly, $p_{11} < p_{01}$ corresponds to a case of Markov-type "trap-shy."

Let $n_{jk} = \sum_{t=1}^{T} n_{jk}(t)$ denote the total number of transitions from state j to state k in the experiment, $j, k = 0$ or 1. Note that $n_{01} = n_{ac} + n_{bc}$, $n_{0+} = n_{a+} + n_{b+}$, $n_{11} = n_{cc}$. The likelihood function for model $\mathbf{MM}_1$(b) reduces to

$$L(N, p_{01}, p_{11}) = \frac{N!}{(N-M)!\prod_{\omega} Z_\omega!} \times (1-p_{01})^{T(N-M)+n_{00}} p_{01}^{n_{01}} (1-p_{11})^{n_{10}} p_{11}^{n_{11}}.$$

There are three parameters $\{N, p_{01}, p_{11}\}$. Because $n_{++} = TM$, the minimum sufficient statistics are $\{n_{00}, n_{01}, n_{10}, n_{11}\}$. A similar factorization of the likelihood as in model $\mathbf{MM}_2$(b) results in the CMLE of N, satisfying

$$1 - \frac{M}{N} = (1 - \hat{p}_{01})^T = \left(1 - \frac{n_{01}}{T(N-M) + n_{0+}}\right)^T. \tag{9}$$

The estimated transition probabilities are $\hat{p}_{01} = n_{01}/[T(\hat{N} - M) + n_{0+}]$ and $\hat{p}_{11} = n_{11}/n_{1+}$. A direct calculation leads to the following asymptotic formula:

$$\mathrm{Var}(\hat{N}) \approx N\left\{-1 + \frac{1}{Q} + \frac{T^2}{\sum_{t=1}^{T} P(X_{t-1} = 0)}\left[1 - \frac{1}{p_{00}}\right]\right\}^{-1}, \tag{10}$$

where Q reduces to $Q = (p_{00})^T$ and $P(X_{t-1} = 0) = [p_{10} + (p_{11} - p_{01})^{t-1} p_{01}]/(p_{01} + p_{10})$ by using a recursive relationship. The variance formula in (10) is in a similar form to that given in (7).

2.3 *Classical Model* $\mathbf{M}_b$*: Enduring Behavioral Effect Only*

In the special case $P_{bc} = P_{cc} \equiv r$ and $P_{ac} \equiv p$, the model $\mathbf{MM}_2$(b) reduces to the classical model $\mathbf{M}_\mathrm{b}$, where p denotes the capture probability for a first capture and r denotes the probability for a recapture. The likelihood in (2) is simplified to

$$L(N, p, r) = \frac{N!}{\left(\prod_{\omega} Z_\omega!\right)(N-M)!} \times \left[(1-p)^{T(N-M)+n_{aa}} p^{n_{ac}} \times (1-r)^{n_{bb}+n_{cb}} r^{n_{bc}+n_{cc}}\right]. \tag{11}$$

This likelihood can be shown to be identical to that given in Otis et al. (1978, p. 29). Comparing the likelihood functions in (2) and (11), we see that the estimation procedures of the population size for models $\mathbf{M}_\mathrm{b}$ and $\mathbf{MM}_2$(b) are the same. Because the classical behavioral model has been extensively discussed in the literature, further details are omitted here.

3. Extension to Incorporate Time Effects and/or Covariates

Otis et al. (1978, p. 38) indicated that the classical model $\mathbf{M}_\mathrm{tb}$ is nonidentifiable. The model $\mathbf{MM}_2$(tb), which extends the bivariate model $\mathbf{MM}_2$(b) to incorporate time effects, is not identifiable either. Nevertheless, the extended model with an ephemeral behavioral effect as well as time effects is identifiable, as is demonstrated in the following subsection.

3.1 *Model $\mathbf{MM}_1$(tb): With Time and Ephemeral Effect*

To incorporate time effects into the model $\mathbf{MM}_1$(b), first define $p_{11}(t) = P\{X_t = 1 \mid X_{t-1} = 1\}$ for $t = 1, 2, \ldots, T$ and similarly define $p_{01}(t)$. Let $\boldsymbol{\eta} = \{p_{01}(1), p_{01}(t), p_{11}(t); t = 2, \ldots, T\}$ denote the transition probabilities. The likelihood can be expressed as

$$L(N, \boldsymbol{\eta}) = \frac{N!}{(N-M)! \prod_\omega Z_\omega!} \times \prod_{t=1}^{T} \left\{[1 - p_{01}(t)]^{(N-M)+n_{00}(t)} [p_{01}(t)]^{n_{01}(t)} \times [1 - p_{11}(t)]^{n_{10}(t)} [p_{11}(t)]^{n_{11}(t)}\right\}.$$

There are $2T$ parameters in the model. With the constraints: $n_{00}(1) + n_{01}(1) = M$, $n_{+1}(t-1) = n_{1+}(t)$, and $n_{+0}(t-1) = n_{0+}(t)$, $t \geq 2$, there are $2T$ minimum sufficient statistics $\{n_{00}(1), n_{01}(1), n_{01}(t), n_{11}(t), t \geq 2\}$. Consequently, the Markov model with the ephemeral effect provides an identifiable model $\mathbf{MM}_1$(tb).

As in model $\mathbf{MM}_2$(b), we have the factorization $L(N, \boldsymbol{\eta}) = L_b(N, \boldsymbol{\eta}) \times L_c(\boldsymbol{\eta})$, where $L_b(N, \boldsymbol{\eta})$ is the same as that in equation (3) except for $Q = \prod_{t=1}^{T} p_{00}(t)$ and

$$L_c(\boldsymbol{\eta}) = \frac{M!}{\prod_\omega Z_\omega!} \frac{1}{(1-Q)^M} \times \prod_{t=1}^{T} \left\{[1 - p_{01}(t)]^{n_{00}(t)} [p_{01}(t)]^{n_{01}(t)} \times [1 - p_{11}(t)]^{n_{10}(t)} [p_{11}(t)]^{n_{11}(t)}\right\}.$$

Analogous to the derivation in the Appendix, it follows that the CMLE of N is the solution of the equation:

$$1 - \frac{M}{N} = \prod_{t=1}^{T} [1 - \hat{p}_{01}(t)] = \prod_{t=1}^{T} \left(1 - \frac{n_{01}(t)}{N - M + n_{0+}(t)}\right), \quad (12)$$

and the estimated transition probabilities are $\hat{p}_{01}(t) = n_{01}(t)/[\hat{N} - M + n_{0+}(t)]$ and $\hat{p}_{11}(t) = n_{11}(t)/n_{1+}(t)$. The asymptotic variance of the CMLE of N is given by

$$\mathrm{Var}(\hat{N}) \approx N \left\{-1 + \frac{1}{Q} + \sum_{t=1}^{T} \frac{1}{P(X_{t-1} = 0)} \left[1 - \frac{1}{p_{00}(t)}\right]\right\}^{-1}. \quad (13)$$

Using a recursive formula, we can compute $P(X_{t-1} = 0)$ and obtain a variance estimator.

If we assume that $p_{01}(t) = p_{11}(t) \equiv p(t)$ for all $t = 1, 2, \ldots, T$, then model $\mathbf{MM}_1$(tb) reduces to model $\mathbf{M}_\mathrm{t}$. If $p_{01}(t) = p_{01}$ and $p_{11}(t) = p_{11}$, then model $\mathbf{MM}_1$(tb) reduces to model $\mathbf{MM}_1$(b). If we further assume that $p_{01} = p_{11} \equiv p$, then model $\mathbf{MM}_1$(b) reduces to the classical model $\mathbf{M}_0$. See Otis et al. (1978) for estimation procedures for the classical models $\mathbf{M}_\mathrm{t}$ and $\mathbf{M}_0$.

3.2 *Model $\mathbf{MM}_1$(tbh): With Time, Behavioral, and Heterogeneity Effects*

In many cases, covariates can be used to explain the heterogeneity among individual capture probabilities. Suppose for each animal, there are s individual covariates. Denote the individual covariates for the ith animal by $\boldsymbol{W}_i' = (W_{i1}, W_{i2}, \ldots, W_{is})$ and let $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_s)$ denote the effects of these covariates. It is necessary to assume that the individual covariates are constant across the T capture occasions in the experiment, as they cannot be measured on any occasion where the individual is not captured. If heterogeneity is fully explained by an individual's covariates, then we can further extend the model $\mathbf{MM}_1$(tb) to a more general model with the heterogeneity effect being $\boldsymbol{\beta}'\boldsymbol{W}_i = \beta_1 W_{i1} + \beta_2 W_{i2} + \cdots + \beta_s W_{is}$. If the transition probabilities for the ith animal are denoted by $p_{01}(i, t)$ and $p_{11}(i, t)$, then a Markov chain with a logistic model for the transition probabilities denoted by model $\mathbf{MM}_1$(tbh) assumes

$$\begin{cases} logit\{p_{01}(i,t)\} = \mu + \alpha_t + \boldsymbol{\beta}'\boldsymbol{W}_i \\ logit\{p_{11}(i,t)\} = \mu + \lambda_t + \nu + \boldsymbol{\beta}'\boldsymbol{W}_i, \end{cases}$$

where μ denotes the baseline intercept, ν denotes the Markov-type behavioral effect, and $\{\alpha_1, \alpha_2, \ldots, \alpha_{T-1}\}$ and $\{\lambda_1, \lambda_2, \ldots, \lambda_{T-1}\}$ represent the time effects ($\alpha_T \equiv 0$, $\lambda_T \equiv 0$). The conditional-likelihood approach can be applied to make inferences; see Yang (2002) for details.

4. Example

The mouse (*Microtus pennsylvanicus*) data were first discussed in Nichols, Pollock, and Hines (1984). The original live-trapping experiment was conducted monthly from June to December, 1980. During each month, the capture–recapture procedure was repeated for 5 consecutive days. The detailed data are given in Williams, Nichols, and Conroy (2002, p. 525–528). We use the data collected in June. A total of 104 distinct mice were caught in the experiment. The numbers of captures for the five occasions were (63, 71, 74, 63, 63) and the numbers of new captures were (63, 19, 9, 7, 6); see Table 1. For each trapping occasion, the capture–recapture records and the transition data based on the capture status, $\{n_{00}(t), n_{01}(t), n_{10}(t), n_{11}(t); t = 1, \ldots, 5\}$, are shown in Table 1. The transition data based on both the capture status and marking status are given in Table 2.

In Table 3, we show the results for the classical model $\mathbf{M}_\mathrm{b}$, and three Markov chain models: $\mathbf{MM}_1$(b), $\mathbf{MM}_2$(b), and $\mathbf{MM}_1$(tb). Under each model, the CMLE of N, its estimated standard error (SE), the logarithm of the maximum conditional likelihood, the number of parameters, the estimated transition probabilities, and their SEs are shown. For comparison, the UMLE is also given in the same table. Although all the four models yield similar population size estimates (both

Table 1
Capture–recapture and transition records based on the capture status

Time	# Captures	Marked	Unmarked	$n_{00}(t)$	n_{01}(t)	n_{10}(t)	n_{11}(t)
$t = 1$	63	0	63	41	63	0	0
$t = 2$	71	52	19	22	19	11	52
$t = 3$	74	65	9	15	18	15	56
$t = 4$	63	56	7	13	17	28	46
$t = 5$	63	57	6	19	22	22	41

the UMLE and CMLE are 106 or 107) and similar estimated SEs, the interpretations of the animals' behavior are different.

The classical model $\mathbf{M}_b$ fits the data well based on the goodness-of-fit test using the counts $\{u_1, u_2, \ldots, u_5\}$, where u_j is the number of new captures at time j. The three Markov models $\mathbf{MM}_1$(b), $\mathbf{MM}_2$(b), and $\mathbf{MM}_1$(tb) also provide adequate fits to the transition counts. On the basis of the classical model $\mathbf{M}_b$, we perform the following three likelihood ratio tests (LRT) for two nested models based on the maximized conditional likelihood. Because this is a case of multiple comparison, the significance level for each test is fixed to be $1.67\% = (5\%)/3$ in order to control a Bonferroni-type overall level of 5%.

1. Classical $\mathbf{M}_b$ versus $\mathbf{MM}_2$(b): The chi-squared-based LRT with one degree of freedom yields a value of 6.14 and p value of 1.32%, which is significant. This indicates that the Markov chain model $\mathbf{MM}_2$(b) provides significant improvement over the classical behavioral model.
2. $\mathbf{MM}_2$(b) versus $\mathbf{MM}_1$(b): There is little difference between the maxima of the likelihoods. The two models are thus not statistically different and the simpler model $\mathbf{MM}_1$(b) is preferable. This can also be seen from the fitted model $\mathbf{MM}_2$(b), where $\hat{P}_{ac}$ is close to $\hat{P}_{bc}$, which indicates no enduring effect.
3. $\mathbf{MM}_1$(b) versus $\mathbf{MM}_1$(tb): The LRT for seven degrees of freedom gives a value of 13.76 and a p value of 5.56%, which is not significant under our multiple comparisons.

Based on the above multiple comparisons, we select the Markov chain model $\mathbf{MM}_1$(b). Under this model, the transition probabilities are estimated to be $\hat{p}_{01} = 0.53$ (SE 0.04), $\hat{p}_{11} = 0.72$ (SE 0.03), which indicates an ephemeral "trap-happy" case. If an animal has not been captured on the current occasion, then the probabilities for capture and noncapture on the next occasion are about the same (0.53 vs. 0.47). However, if captured, the probability of recapture on the next occasion becomes 0.72. These data provide an example that the Markov model with an ephemeral effect gives a better description of an animal's behavioral response.

Table 2
Transition records based on the capture status and marking status

Time	$n_{aa}(t)$	$n_{ac}(t)$	$n_{bb}(t)$	$n_{bc}(t)$	$n_{cb}(t)$	$n_{cc}(t)$
$t = 1$	41	63	0	0	0	0
$t = 2$	22	19	0	0	11	52
$t = 3$	13	9	2	9	15	56
$t = 4$	6	7	7	10	28	46
$t = 5$	0	6	19	16	22	41
Total	82	104	28	35	76	195

5. Simulation Studies

In order to examine the performances of the estimators, we carried out a simulation study for the Markov chain models without covariates. The results for the 12 representative trials described in Table 4 are reported. We focus six trials on the Markov chain model $\mathbf{MM}_1$(b) because it is the selected model for the data of our example. Because the traditional model $\mathbf{M}_{tb}$ is nonidentifiable, we think it is worthwhile to investigate the corresponding identifiable Markov chain model $\mathbf{MM}_1$(tb) in the other six trials. The transition probabilities were chosen so that the expected percentage of animals captured in the experiment is between 37.5% and 87.5%. The results are shown in Table 5. The reader is referred to Yang (2002) for more simulation trials.

In our simulations, the true population size N was fixed at 400 and the number of occasions T was fixed at 5. For each model and specified set of transition probabilities, 5000 data sets were generated. Then for each generated data set, we calculated the estimators from the two classical models $\mathbf{M}_t$ and $\mathbf{M}_b$, as well as the two Markov models $\mathbf{MM}_1$(b) and $\mathbf{MM}_1$(tb). The estimated SE for each estimator was computed by using the asymptotic variance discussed in Sections 2 and 3. The iterative method employed to compute the estimates diverged in only a negligible number of the generated data sets. The associated 95% confidence intervals were calculated using a log transformation as in Chao (1987). The resulting 5000 estimates and SE estimates were averaged and are reported as the "Average Estimate" and "Average Estimated SE" in Table 5. The sample SE and the sample root mean squared error (RMSE) were also computed and are reported as the "Sample SE" and "Sample RMSE." The proportion of trials in which the 95% confidence intervals covered the true parameter was also recorded and is tabulated in the last column of the same table. The findings are summarized below.

(1) The underlying model is $\mathbf{MM}_1$(tb) (Trials 1–6):

The estimators based on the two classical models ($\mathbf{M}_t$ and $\mathbf{M}_b$) and the Markov submodel $\mathbf{MM}_1$(b) are biased, and in some cases the biases are substantial. For example, in the trap-happy cases as in Trials 1–4, where $p_{11}(t) > p_{01}(t)$ for all t, the estimator under model $\mathbf{M}_t$ is negatively biased; it is positively biased in the trap-shy cases as in Trials 5 and 6, where $p_{11}(t) < p_{01}(t)$ for

Table 3
Estimation results for the mouse data

Model	#Parameters in L_c	Maximum $\log(L_c)$[a]	CMLE of N (SE)	UMLE of N (SE)	Estimated parameters (SE)
$\mathbf{M}_{\mathrm{b}}$	2	−332.66	107 (2.1)	106 (1.9)	$\hat{p} = 0.52(0.04)$, $\hat{r} = 0.69(0.03)$
$\mathbf{MM}_1$(b)	2	−329.69	106 (1.8)	106 (1.7)	$\hat{p}_{01} = 0.53(0.04)$, $\hat{p}_{11} = 0.72(0.03)$
$\mathbf{MM}_2$(b)	3	−329.59	107 (2.1)	106 (1.9)	$\hat{P}_{ac} = 0.52(0.04), \hat{P}_{bc} = 0.56(0.06),$ $\hat{P}_{cc} = 0.72(0.03)$
$\mathbf{MM}_1$(tb)	9	−322.81	107 (2.2)	106 (2.1)	$\hat{p}_{01}(t) = (0.59, 0.43, 0.50, 0.51, 0.50)$ SE = (0.05, 0.08, 0.09, 0.10, 0.08) $\hat{p}_{11}(t) = (-, 0.83, 0.79, 0.62, 0.65)$ SE = (−, 0.05, 0.05, 0.06, 0.06)

[a]Some constants shared by all models were dropped in the calculation.

all t. This conclusion is similar to that in the conventional models; that is, any estimator that wrongly ignores the behavioral response effect is biased downward in the trap-happy cases whereas it is biased upward in the trap-shy cases.

The classical behavioral model also yields severe bias, although there is no systematic finding about the direction of bias. In all trials, the two classical models have very large RMSE and the associated intervals do not perform satisfactorily as regards coverage probability (some of the estimated coverage probabilities are 0 or 1%).

For the Markov submodel $\mathbf{MM}_1$(b) without time effects, the direction of bias depends on the time effects. Although these three simpler models, $\mathbf{M}_{\mathrm{t}}$, $\mathbf{M}_{\mathrm{b}}$, and $\mathbf{MM}_1$(b) exhibit smaller variation (sample SEs are smaller) or smaller RMSE in the first two trials, the coverage probabilities of their associated 95% confidence intervals are much lower than the nominal level. Thus, none of these estimators are appropriate under model $\mathbf{MM}_1$(tb).

The proposed CMLE based on the correct model, $\mathbf{MM}_1$(tb), has the smallest bias in all trials. Although in the first two trials its RMSE is larger than that of a submodel, and it unavoidably has larger variation, the associated 95% confidence intervals perform well in maintaining the nominal coverage probability. When the fraction of animals caught in the experiment is over 40%, as in Trials 2–6, the asymptotic SE estimates produce satisfactory results compared with the sample SEs.

(2) The underlying model is $\mathbf{MM}_1$(b) (Trials 7–12):

The behavior of the estimator for the classical model $\mathbf{M}_{\mathrm{t}}$ is similar to that in Trials 1–6. That is, it underestimates the population size when $p_{11} > p_{01}$ as in Trials 7–10, and overestimates the population size when $p_{11} < p_{01}$ as in Trials 11 and 12. The estimator based on the classical behavioral model generally has the largest variance and its RMSE is larger than the estimators derived from the Markov chain models.

As expected, the estimator derived from the correct model performs best in terms of RMSE and interval coverage probability in Trials 7–12. The estimator based on a more general model, $\mathbf{MM}_1$(tb), yields comparable bias, RMSE, and coverage probability. Although the estimator under model $\mathbf{MM}_1$(tb) has slightly larger variance, the loss of efficiency is low in all cases. This is consistent with the findings of Lloyd (1994) for conventional models.

Table 4
Description of simulation trials

Trial	Model	Description of ($p_{01}(t)$, $p_{11}(t)$)	$E(M)$
1	$\mathbf{MM}_1$(tb)	$p_{01}(t) = (0.14, 0.02, 0.12, 0.14, 0.02)$ $p_{11}(t) = (-, 0.6, 0.4, 0.5, 0.6)$	150.0
2	$\mathbf{MM}_1$(tb)	$p_{01}(t) = (0.20, 0.05, 0.08, 0.15, 0.05)$ $p_{11}(t) = (-, 0.6, 0.4, 0.5, 0.6)$	174.2
3	$\mathbf{MM}_1$(tb)	$p_{01}(t) = (0.25, 0.05, 0.15, 0.15, 0.05)$ $p_{11}(t) = (-, 0.6, 0.4, 0.5, 0.6)$	204.4
4	$\mathbf{MM}_1$(tb)	$p_{01}(t) = (0.32, 0.12, 0.12, 0.16, 0.16)$ $p_{11}(t) = (-, 0.8, 0.6, 0.5, 0.5)$	251.4
5	$\mathbf{MM}_1$(tb)	$p_{01}(t) = (0.12, 0.32, 0.38, 0.25, 0.12)$ $p_{11}(t) = (-, 0.1, 0.15, 0.2, 0.1)$	302.1
6	$\mathbf{MM}_1$(tb)	$p_{01}(t) = (0.43, 0.4, 0.4, 0.2, 0.2)$ $p_{11}(t) = (-, 0.2, 0.2, 0.1, 0.1)$	347.5
7	$\mathbf{MM}_1$(b)	$p_{01} = 0.09$, $p_{11} = 0.20$	150.4
8	$\mathbf{MM}_1$(b)	$p_{01} = 0.11$, $p_{11} = 0.20$	176.6
9	$\mathbf{MM}_1$(b)	$p_{01} = 0.13$, $p_{11} = 0.20$	200.6
10	$\mathbf{MM}_1$(b)	$p_{01} = 0.18$, $p_{11} = 0.40$	251.7
11	$\mathbf{MM}_1$(b)	$p_{01} = 0.25$, $p_{11} = 0.15$	305.1
12	$\mathbf{MM}_1$(b)	$p_{01} = 0.34$, $p_{11} = 0.20$	349.9

In summary, the classical behavioral model yields severe bias or large RMSE when the true models are Markov models. The estimator derived under a general model $\mathbf{MM}_1$(tb) behaves well under the correct model and its submodel $\mathbf{MM}_1$(b); there is loss of efficiency under model $\mathbf{MM}_1$(b) but the loss is almost negligible. Under model $\mathbf{MM}_1$(tb), estimators ignoring either behavioral response or time effects may be seriously biased, although they exhibit smaller variance. Thus, the proposed Markov models not only provide new interpretation of animals' behavior but also offer a more satisfactory population size estimator when the behavioral effect lasts for only one occasion.

Table 5
Population size estimation for the trials given in Table 4 ($N = 400$, 5000 *simulations*)

True model	Assumed model	Average estimate	Sample SE	Average estimated SE	Sample RMSE	95% CI coverage (%)
Trial 1	$\mathbf{M}_t$	172	11.9	6.5	227.9	0
$\mathbf{MM}_1$(tb)	$\mathbf{M}_b$	228	38.1	45.8	176.6	22
	$\mathbf{MM}_1$(b)	370	77.9	85.7	83.3[a]	90
	$\mathbf{MM}_1$(tb)	413	94.7	113.5	95.6	97
Trial 2	$\mathbf{M}_t$	199	12.1	6.7	201.4	0
$\mathbf{MM}_1$(tb)	$\mathbf{M}_b$	218	21.8	19.8	182.8	1
	$\mathbf{MM}_1$(b)	346	57.4	55.5	79.0[a]	76
	$\mathbf{MM}_1$(tb)	414	79.3	82.4	80.5	95
Trial 3	$\mathbf{M}_t$	230	12.0	6.6	170.6	0
$\mathbf{MM}_1$(tb)	$\mathbf{M}_b$	236	15.7	12.3	164.6	0
	$\mathbf{MM}_1$(b)	351	43.7	40.4	65.8	73
	$\mathbf{MM}_1$(tb)	409	60.6	60.8	61.2[a]	95
Trial 4	$\mathbf{M}_t$	267	10.6	4.7	133.5	0
$\mathbf{MM}_1$(tb)	$\mathbf{M}_b$	287	16.6	12.8	113.8	1
	$\mathbf{MM}_1$(b)	349	26.6	23.9	57.8	52
	$\mathbf{MM}_1$(tb)	404	39.5	39.7	39.7[a]	95
Trial 5	$\mathbf{M}_t$	470	24.6	27.5	74.3	10
$\mathbf{MM}_1$(tb)	$\mathbf{M}_b$	430	32.4	43.2	44.4	94
	$\mathbf{MM}_1$(b)	449	27.2	28.4	56.1	46
	$\mathbf{MM}_1$(tb)	402	21.6	22.0	21.7[a]	95
Trial 6	$\mathbf{M}_t$	444	13.6	15.9	45.9	6
$\mathbf{MM}_1$(tb)	$\mathbf{M}_b$	361	7.6	5.0	40.1	0
	$\mathbf{MM}_1$(b)	406	11.7	12.0	13.2	90
	$\mathbf{MM}_1$(tb)	400	11.4	11.6	11.4[a]	95
Trial 7	$\mathbf{M}_t$	285	32.4	31.6	119.4	19
$\mathbf{MM}_1$(b)	$\mathbf{M}_b$	310	53.5	112.6	104.5	88
	$\mathbf{MM}_1$(b)	392	76.3	82.2	76.7[a]	94
	$\mathbf{MM}_1$(tb)	396	79.5	84.9	79.6	94
Trial 8	$\mathbf{M}_t$	319	31.3	30.3	86.8	38
$\mathbf{MM}_1$(b)	$\mathbf{M}_b$	356	62.6	115.6	76.4	92
	$\mathbf{MM}_1$(b)	398	59.5	63.9	59.5[a]	95
	$\mathbf{MM}_1$(tb)	400	61.9	65.5	61.9	95
Trial 9	$\mathbf{M}_t$	346	29.5	28.8	61.6	61
$\mathbf{MM}_1$(b)	$\mathbf{M}_b$	387	66.9	109.2	68.2	95
	$\mathbf{MM}_1$(b)	402	49.6	51.6	49.7[a]	95
	$\mathbf{MM}_1$(tb)	402	51.1	52.5	51.2	95
Trial 10	$\mathbf{M}_t$	313	15.2	12.1	88.1	0
$\mathbf{MM}_1$(b)	$\mathbf{M}_b$	413	64.6	75.2	65.9	96
	$\mathbf{MM}_1$(b)	404	34.4	34.6	34.6[a]	95
	$\mathbf{MM}_1$(tb)	403	35.0	35.2	35.2	95
Trial 11	$\mathbf{M}_t$	442	20.4	22.2	46.2	40
$\mathbf{MM}_1$(b)	$\mathbf{M}_b$	405	31.9	32.8	32.3	95
	$\mathbf{MM}_1$(b)	402	19.0	19.0	19.1[a]	95
	$\mathbf{MM}_1$(tb)	401	19.1	19.2	19.1[a]	95
Trial 12	$\mathbf{M}_t$	432	11.8	13.7	33.7	20
$\mathbf{MM}_1$(b)	$\mathbf{M}_b$	401	14.8	15.0	14.8	95
	$\mathbf{MM}_1$(b)	401	10.5	10.7	10.6[a]	95
	$\mathbf{MM}_1$(tb)	400	10.6	10.8	10.7	96

[a]Denotes the smallest RMSE.

6. Concluding Remarks and Discussion

Simulation results have shown that the classical behavioral model does not yield satisfactory estimates under models with an ephemeral behavioral effect. A bivariate Markov chain model and submodels have been proposed to include both enduring and ephemeral behavioral effects and to estimate population size. A unified likelihood approach is applicable to the class of Markov models. A real data example has shown that the Markov chain models provide significant improvement in the maximum likelihoods over the classical behavioral model.

A logistic-type Markov chain model accounting for the effects of an individual's covariates can also be applied.

In the Markov approach, we consider both the capture status (capture/noncapture) and marking status (marked/unmarked). We could further extend the current model by considering the number of occasions since the last capture instead of just the marking status. This would result in a model with a geometric-type drop off for behavioral effect. This may be a worthwhile topic for future research. Also, an extension to open populations is currently under investigation.

A GAUSS computer program which calculates the estimators proposed in this article will be available on the second author's website at `http://chao.stat.nthu.edu.tw/`.

Acknowledgements

Part of the material is based on the PhD thesis of the first author under the supervision of the second author. We would like to thank the associate editor and two referees for providing insightful suggestions and comments, which have motivated the extension of the univariate Markov chain model of the previous version to a bivariate approach. The extension has significantly improved the article.

References

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43,** 783–791.

Chao, A., Chu, W., and Hsu, C.-H. (2000). Capture-recapture when time and behavioral response affect capture probabilities. *Biometrics* **56,** 427–433.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47,** 663–685.

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.

Lloyd, C. J. (1994). Efficiency of martingale methods in recapture studies. *Biometrika* **81,** 305–315.

Nichols, J. D., Pollock, K. H., and Hines, J. E. (1984). The use of a robust capture-recapture design in small mammal population studies: A field example with *Microtus pennsylvanicus*. *Acta Theriologica* **29,** 357–365.

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. (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,** 225–238.

Sanathanan, L. (1972). Estimating the size of a multinomial population. *Annals of Mathematical Statistics* **43,** 143–152.

Solow, A. R. (2000). The effect of dependence on estimating sample coverage. *Environmetrics* **11,** 245–249.

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and Management of Animal Populations*. San Diego: Academic Press.

Yang, H.-C. (2002). The applications of Markov chain models and kernel smoothing in capture-recapture experiments, Ph.D. Thesis, National Tsing Hua University, Hsin-Chu, Taiwan.

Received March 2003. *Revised December* 2004.
Accepted January 2005.

Appendix

It follows from the likelihood given in (4) that the CMLE for $\boldsymbol{\eta} = \{P_{ac}, P_{bc}, P_{cc}\}$ satisfies the following equations:

$$\frac{\partial \log L_c(\boldsymbol{\eta})}{\partial P_{ac}} = -\frac{n_{aa}}{1-P_{ac}} + \frac{n_{ac}}{P_{ac}} - \frac{M}{1-Q}\frac{QT}{1-P_{ac}} = 0, \quad \text{(A.1)}$$

$$\frac{\partial \log L_c(\boldsymbol{\eta})}{\partial P_{bc}} = -\frac{n_{bb}}{1-P_{bc}} + \frac{n_{bc}}{P_{bc}} = 0, \quad \text{(A.2)}$$

$$\frac{\partial \log L_c(\boldsymbol{\eta})}{\partial P_{cc}} = -\frac{n_{cb}}{1-P_{cc}} + \frac{n_{cc}}{P_{cc}} = 0. \quad \text{(A.3)}$$

From (A.2) and (A.3), we have $\hat{P}_{bc} = n_{bc}/n_{b+}$ and $\hat{P}_{cc} = n_{cc}/n_{c+}$. After the substitution of $Q = (1 - P_{ac})^T$ into (A.1), the estimate $\hat{P}_{ac}$ could be numerically solved. A ratio method is then applied to the binomial likelihood $L_b(N, \hat{\boldsymbol{\eta}})$ given in (3) and the CMLE $\hat{N}$ for the integer-valued parameter N turns out to be $\hat{N} = [M/(1-\hat{Q})] = [M/(1-(1-\hat{P}_{ac})^T)]$, where $[x]$ denotes the largest integer $\leq x$. When N is treated as a real number, the estimation is equivalent to the following computational procedure: Substituting $M/(1-Q)$ and Q in (A.1) by N and $1 - M/N$, respectively, we obtain $\hat{P}_{ac} = n_{ac}/\{T(\hat{N} - M) + n_{a+}\}$ and the CMLE of N satisfies equation (5).