Review Article
Putting Markov Chains Back into Markov Chain Monte Carlo

Richard J. Barker and Matthew R. Schofield

Markov chain theory plays an important role in statistical inference both in the formulation of models for data and in the construction of efficient algorithms for inference. The use of Markov chains in modeling data has a long history, however the use of Markov chain theory in developing algorithms for statistical inference has only become popular recently. Using mark-recapture models as an illustration, we show how Markov chains can be used for developing demographic models and also in developing efficient algorithms for inference. We anticipate that a major area of future research involving mark-recapture data will be the development of hierarchical models that lead to better demographic models that account for all uncertainties in the analysis. A key issue is determining when the chains produced by Markov chain Monte Carlo sampling have converged.

1. Introduction

Markov chains, and related stochastic models, have long played an important role in helping ecologists understand population dynamics. In the main, this has been through the application of probability models to the problem of predicting the realized dynamics of plant and animal populations over time. In this context, the challenge is to construct models that are relatively simple, in terms of the numbers of parameters and the relationships between (and among) parameters and state variables, and yet that are able to provide a reasonable approximation to the sorts of dynamics typically observed in nature.

The development of these models has been driven by ecologists motivated by an interest in underlying theory, and ecologists with applications in mind. These applications are generally based on inference about the long-term dynamics of the study population. Examples include predicting the time to possible extinction for threatened populations or establishing safe levels of harvest for exploited populations.

Complementary to the problem of developing theoretical models that predict the temporal dynamics of populations has been the development of a body of theory for estimating population parameters. This estimation theory is the main emphasis in the remainder of this contribution. Almost universally, these models treat quantities such as population size at time t, N_t, or the numbers of individuals born between t and $t + \Delta t$, as fixed quantities to be predicted (here we use prediction in the sense of predicting the value of an unobserved realization of a random variable as well as future realizations). In contrast, the population models discussed above treat these same quantities as random variables whose behavior we are interested in describing.

A particularly important class of estimation models for population dynamics is the mark-recapture model [1]. Mark-recapture data comprises repeated measures on individual animals in the population obtained through samples of animals drawn from the population (usually) at discrete sample times $t_1, \ldots, t_k$. Because the capture process is imperfect, not all animals in the population at time t_j are captured, and a model is required to describe this process of repeated imperfect captures.

In the simplest case, the population is regarded as closed to births and deaths and so the population size at time t_j is the same for all k sample occasions. Studies of the dynamics of the population can then be based on repeated experiments generating a sequence of abundance estimates.

A more interesting class of models is the open population models in which individuals may enter (through birth or immigration) or leave (through death or emigration) during the interval (t_1, t_k). Here the challenge is to model the sequence of captures of animals in terms of parameters of demographic interest. These are usually survival probabilities S_j, and parameters that describe the birth process. Quantities such as N_j, the abundance of animals at the time of sample j, and B_j, the number of animals born between samples j and $j + 1$, are then predicted from the model.

Traditionally, inference for mark-recapture models has been based on maximum likelihood. Importantly, demographic models for $\{N_t\}$ have played no role in these estimation models. Inference about the $\{N_t\}$ process has instead been based on ad hoc methods for summarizing sequences of estimates such as $\{\hat{N}_t\}$. Recently there have been many developments that apply Bayesian inference methods to mark recapture models based on Markov chain Monte Carlo (MCMC). Here the utility of Markov chain theory appears in a fundamentally different context to that described above for population modeling; it arises as a tool for inference. An exciting feature of Bayesian inference methods based on MCMC is that fitting complicated hierarchical models has become feasible. Hierarchical models provide a link between the demographic models and estimation models in a way that should lead to better and more relevant inference. It is recent developments in Markov chain theory, in particular Gibbs sampling [2] and reversible jump Markov chain Monte Carlo [3], that allow this link to take place.
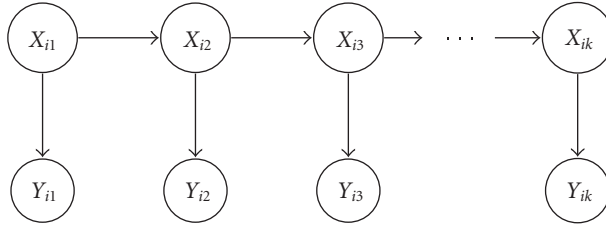
FIGURE 2.1. Directed acyclic graph for the hidden Markov model for mark-recapture data.

2. Open population capture-recapture models

Define v as the total number of animals alive on at least one of the sampling occasions t_j ($j = 1,\dots,k$). Let Y_{ij} be the indicator that animal i ($i = 1,\dots,v$) is caught in sample j, which takes place at time t_j. Any animal that is caught for the first time in a particular sample is marked with some form of unique tag and then released, with any recaptures noted. The data are of the form $y = \{y_{ij}\}$ for $i = 1,\dots,u.$ and $j = 1,\dots,k$, where u_j is the number of unmarked animals caught in sample j and $u. = \sum_j u_j$ is the total number of animals caught at least once.

For animal i, the set of values $y_i = \{y_{i1},\dots,y_{ik}\}$ is called the capture history and these provide censored information on the time of birth and death. We know that the time of birth occurred before the sample time corresponding to the first nonzero value of y_i and the time of death occurred sometime after the sample time corresponding to the last nonzero value of y_i.

2.1. Hidden Markov model. The problem with capture-recapture methods is that not all animals in the population are caught; this causes the censoring of the birth and death intervals. Define X_{ij} as the indicator of the event that animal i is alive and in the study population at time t_j. If $X = \{x_{ij}\}$, $i = 1,\dots,v$, $j = 1,\dots,k$, were observed for an animal population in which v animals were ever alive and available for capture on at least one of the k sample occasions, then inference about the parameters $\{S_j\}$ and $\{\beta_j\}$ would be straight-forward. It would also be straight-forward to calculate observed values of the random variables $N_j = \sum_i x_{ij}$.

Instead of observing x_{ij}, we observe y_{ij} with the joint distribution of the sequence of pairs $\{X_{ij}, Y_{ij}\}$ described by a hidden Markov model (Figure 2.1). Importantly, y does not include the history given by the vector of k zeros associated with animals that were available for capture at sometime during the experiment but never caught.

Conditional on the event that animal i is alive (and available for capture) on at least one of the capture occasions t_j, the sequence $(X_{i1},\dots,X_{ik})$ can then be modeled as a three-state Markov chain. $X_{ij} = 0$ corresponds to the event that animal i has not yet entered the population. $X_{ij} = 1$ corresponds to the event that animal i is in the population and alive, and $X_{ij} = 2$ corresponds to animal i being dead. The possible transitions are $0 \rightarrow 0$, $0 \rightarrow 1$,

$1 \to 1$, $1 \to 2$, and $2 \to 2$. Define:

β_0 as the probability that an animal is born before the start of the experiment, given that it is alive at some time during the experiment,

β_j as the probability that an animal is born between the times of sample j and $j+1$, given that it was available for capture at sometime during the experiment,

S_j as the probability that an animal alive at the time of sample j is still alive at the start of sample $j+1$.

Conditioning on the ν individuals that are alive and available for capture at sometime during the experiment means that $\beta_0 + \beta_1 + \ldots + \beta_{k-1} = 1$. If $\mathcal{A}$ is used to denote the event that animal i is alive and available for capture at sometime during the experiment, then

$$\Pr\left(X_{i1} = x_{i1} \mid \mathcal{A}\right) = \begin{cases} \beta_0, & x_{i1} = 1, \\ 1 - \beta_0, & x_{i1} = 0, \end{cases} \tag{2.1}$$

and for $j = 2,\ldots,k-1$,

$$\Pr\left(X_{ij} = x_{ij} \mid x_{ij-1}, \mathcal{A}\right) = \begin{cases} 1, & x_{ij} = 2, x_{ij-1} = 2, \\ 0, & x_{ij} = 1, x_{ij-1} = 2, \\ 0, & x_{ij} = 0, x_{ij-1} = 2, \\ 1 - S_j, & x_{ij} = 2, x_{ij-1} = 1, \\ S_j, & x_{ij} = 1, x_{ij-1} = 1, \\ 0, & x_{ij} = 0, x_{ij-1} = 1, \\ 0, & x_{ij} = 2, x_{ij-1} = 1, \\ \beta'_{j-1}, & x_{ij} = 1, x_{ij-1} = 0, \\ 1 - \beta'_{j-1}, & x_{ij} = 0, x_{ij-1} = 0, \end{cases} \tag{2.2}$$

where $\beta'_j = \beta_j / \sum_{h=j}^{k-1} \beta_h$. Note that we exclude from the study population individuals that are born and then die during (t_j, t_{j+1}). Clearly these individuals are invisible to the mark-recapture experiment.

2.2. Observed data likelihood.

A common assumption in mark recapture models is that $y_{ij}|x_{ij}$ is the outcome of a Bernoulli trial with probability p_j if $x_{ij} = 1$ or 0 otherwise. That is, only those animals alive and in the population at the time of sample j are at risk of being caught, which happens with probability p_j. The standard approach to fitting this model is to derive an observed data likelihood from the marginal distribution with pdf $[y \mid \nu, S, \beta, p]$ which is described by summing up across all possible sample paths for X that are compatible with the data (here we use the notation $[y|x]$ to denote the pdf of the distribution for the random variable Y, evaluated at y, conditional on X, evaluated at x).

Using the notation $y = \{y_{ij}\}$ ($i = 1,\ldots,u.$; $j = 1,\ldots,k$) and $u = (u_1,\ldots,u_k)'$ the likelihood can be expressed as

$$[y \mid \nu, S, \beta, p] = [u. \mid \nu][u \mid u., \beta, S][y \mid u, S, p], \tag{2.3}$$

where $[u.|v]$ describes a binomial distribution with index v and parameter π_0 and $[u|u.]$ a multinomial distribution with index $u.$ and parameter vector ξ.

Both π_0 and ξ are complicated functions of the parameters S and β that account for the censoring of the birth times in y. If we let ψ_j denote the probability that an animal is available for capture in sample j but has not yet been caught, then

$$\psi_{j+1} = \psi_j(1-p_j)S_j + \beta_j \tag{2.4}$$

for $j = 2,\ldots,k-1$, with $\psi_1 = \beta_0$. Such an animal is first caught in sample j with probability $\psi_j p_j$ hence

$$\pi_0 = 1 - \sum_{j=1}^{t}\psi_j p_j,$$
$$\xi_j = \frac{\psi_j p_j}{1-\pi_0}. \tag{2.5}$$

The term $[y \mid u, S, p]$ describes the celebrated Cormack-Jolly-Seber model (see [4–6]). If we index the first sample in which an animal i was caught by r_i and the last sample in which animal i was seen by l_i, then we can write

$$[y \mid u, S, p] = \prod_{i=1}^{u.}\prod_{j=r_i}^{l_i} p_j^{y_{ij}}(1-p_j)^{1-y_{ij}}\chi_{l_i}, \tag{2.6}$$

where the term χ_j, which accounts for the censoring of time of death, can be defined recursively as

$$\chi_j = \begin{cases} 1, & j = k, \\ 1 - S_j + S_j(1-p_{j+1})\chi_{j+1}, & j = 1,\ldots,k-1. \end{cases} \tag{2.7}$$

A nice feature of the observed data likelihood (2.3) is that it is straightforward to find maximum likelihood estimators. [7] showed that the parameter v is not identifiable, however the partial likelihood $[u \mid u., \beta, S][y \mid u, S, p]$ obtained by conditioning on $u.$ contains all practically useful information on the identifiable parameters in the model. Closed-form solutions to the ML equations exist for all identifiable parameters in this partial-likelihood.

The model can also be easily generalized to allow parameters to be individual-specific, by introducing covariates, or to allow captures to depend on the earlier capture history of the animal. The model (2.3) has also been extended by reparameterizing the model in terms of per capita birth rate f_j [7, 8] and an index to population growth rate [8]

$$\lambda_j = S_j + f_j$$
$$= \frac{E[N_{j+1} \mid N_j, v]}{E[N_j \mid v]}. \tag{2.8}$$

The complete model (2.3), which we refer to as the Crosbie-Manly-Arnason-Schwarz model, was first described in [9] building on earlier work of [10]. It differs from the well-known Jolly-Seber (JS) model ([5, 6]) in the way that captures of unmarked animals are modeled. In the JS model, the terms

$$[u. \mid v][u \mid u.] \tag{2.9}$$

are replaced by

$$[u \mid U] \tag{2.10}$$

and the elements of $U = (U_1, \dots, U_k)'$, where U_j is the number of unmarked animals in the population at the time of sample j, are treated as unknown parameters to be estimated. While historically important, this approach does not allow the reparameterizations in terms of f_j and λ_j described above as f_j and λ_j cannot be written as deterministic functions of the parameters S_j and β_j. Also, the U_j parameters are of little demographic interest although predictors of N_j and B_j exist.

2.3. Complete data likelihood.

An alternative to using the observed data likelihood is to describe the model in terms of the complete data likelihood (CDL) [11]. The CDL is the likelihood of *all* data, both missing and observed. The observed data are
   (i) the values y_{ij} for all $u.$ individuals that we observed at least once,
   (ii) the censored information about X that we obtain from y.
The missing data are
   (i) the unknown number of individuals that were available for capture but not caught (we include this by specifying v as a parameter),
   (ii) the realized but unknown values of X_{ij}, x_{ij}.
Including both the observed and missing data gives the CDL that we can express as

$$[x, y \mid S, \beta, p, v] = [y \mid p, x, v][x \mid S, \beta, v]. \tag{2.11}$$

The CDL has been naturally factored into a term that describes how the data were corrupted, $[y \mid p, x, v]$, and a term that models the underlying birth and death processes of interest, $[x \mid S, \beta, v]$. As for the observed data likelihood section, we model the corruption by assuming that an individual that was alive in sample j ($x_{ij} = 1$) was observed in sample j ($y_{ij} = 1$) with probability p_j. We model the birth and death process as in the Markov chain model for X described in Section 2.1.

Even though the CDL provides a natural approach to looking at the problem, we must still integrate over the missing data in order to obtain valid inference. Computational techniques such as Markov chain Monte Carlo (MCMC, discussed below) can be used that iteratively integrate out all missing data, allowing models to be specified in terms of the CDL. This means that in each iteration of the MCMC chain we need values for the missing quantities x and v (and all other parameters) that were obtained from the

posterior distribution of all unknowns. One such MCMC algorithm we can use is Gibbs sampling, described in the following section.

2.4. Markov chain Monte Carlo. The Gibbs sampler, also known as alternating conditional sampling, is a remarkable algorithm for efficiently constructing a Markov chain for complex joint probability distributions $[Z_1,\ldots,Z_k \mid \theta]$ by sampling from the full conditional distributions $[Z_i \mid \{Z_j\}_{j\neq i},\theta]$ of each component. The stationary distribution of this Markov chain has the target density $[Z_1,\ldots,Z_k \mid \theta]$. A particularly useful feature of the Gibbs sampler is that the Markov chain can be constructed even though the target joint probability density may only be known up to the normalizing constant. This has led to a resurgence of interest in Bayesian inference which, historically, has been held back by the need for high-dimension integration needed to normalize posterior probability densities.

Here, we outline the construction of a Gibbs sampler whose target density is the CDL described in the previous section. Once we have collected our mark-recapture data we have data y, that is known; and the unknown are any parameters and any unknown realized values of random variables of interest. For the Crosbie-Manly-Arnason-Schwarz model (2.3), the parameters are $p_1, p_2, \ldots, p_k$; $S_1, \ldots, S_{k-1}$, $\beta_0, \beta_1, \ldots, \beta_{k-1}$, and ν. In addition unknown realized random variables of interest (might) include $N_1, \ldots, N_k$, $B_0, B_1, \ldots, B_{k-1}$.

Starting with (2.3) and specifying independent beta $Be(\alpha_p, \gamma_p)$ prior distributions for the parameters p_j, we can show that the full conditional distribution for p_j is a beta $Be(n_j + \alpha_p, N_j - n_j + \gamma_p)$ for $j = 1, \ldots, k$, and where n_j is the total number of individuals caught at time of sample j.

If we specify independent beta $Be(\alpha_S, \gamma_S)$ prior distributions for the parameters S_j, we obtain beta $Be(N_j - D_j + \alpha_S, D_j + \gamma_S)$ full conditional distributions for $j = 1, \ldots, k-1$.

If we specify independent beta $Be(\alpha_{\beta'}, \gamma_{\beta'})$ prior distributions for the parameters β'_j, we obtain beta $Be(B_j + \alpha_{\beta'}, N - \sum_{h=0}^{j} B_h + \gamma_{\beta'})$ full conditional distributions for $j = 1, \ldots, k-2$ (note that $\beta'_{k-1} = 1$). If desired, we can transform the generated values of β'_j to any other birth parameter, such as β_j or η_j. For example, β_j, $j = 0, \ldots, k-1$, is obtained by taking

$$\beta_j = \begin{cases} \beta'_j, & j = 0, \\ \beta'_j \prod_{h=0}^{j-1} (1 - \beta'_h), & j = 1, \ldots, k-2. \end{cases} \tag{2.12}$$

We also need to calculate the full conditional probability of the missing values of x. This can be done a number of ways, but we choose to represent the information in x by matrices that give the interval censored times of birth and death, denoted by b and d, respectively. The value $b_{ij} = 1$ means individual i was born between sample j and $j+1$ with $b_{ij} = 0$ otherwise. The value $b_{i0} = 1$ means the individual was born before the study began. The value $d_{ij} = 1$ means that individual i died between sample j and $j+1$ with $d_{ij} = 0$ otherwise. The value $d_{ik} = 1$ means the individual was still alive at the end of the study. The assumptions about the underlying birth and death processes impose

restrictions on the values of b and d,

$$\sum_{j=0}^{k-1} b_{ij} = 1, \quad \sum_{j=1}^{k} d_{ij} = 1, \quad d_{ij} - b_{ij} \geq 0 \ \forall i, j. \tag{2.13}$$

The matrices b and d are censored. For the u. individuals that were ever observed we know that they were not born after first capture, $b_{ij} = 0$, $j = r_i, \ldots, k$, where r_i indexes the sample in which animal i first appeared. Likewise we know that they did not die before the last capture, $d_{ij} = 0$, $j = 1, \ldots, l_i$, where l_i indexes the sample in which animal i last appeared. For the $v - u$. individuals that were never observed we have no information about either b or d.

The full conditional distribution of the unknown values of b for individual i is a multinomial distribution with probability vector γ^b, where

$$\gamma_j^b = \frac{\zeta_j^b}{\sum_{h=0}^{\lambda_i^b - 1} \zeta_h^b},$$

$$\zeta_j^b = \beta_j \prod_{h=j}^{\lambda_i^b - 1} S_h (1 - p_h), \tag{2.14}$$

and λ_i^b is the time of first capture r_i for all individuals observed, and it is the last period in which the individual was alive (obtained from d) for individuals $i = u. + 1, \ldots, v$ that were not observed.

The full conditional distribution of the unknown values of d for individual i is a multinomial distribution with probability vector γ^d, where

$$\gamma_j^d = \frac{\zeta_j^d}{\sum_{h=\lambda_i^d}^{t} \zeta_h^d},$$

$$\zeta_j^d = \begin{cases} \displaystyle\prod_{h=\lambda_i^d}^{j-1} S_h \prod_{h=\lambda_i^d+1}^{j} (1 - p_h)(1 - S_j), & j < k, \\ \displaystyle\prod_{h=\lambda_i^d}^{j-1} S_h \prod_{h=\lambda_i^d+1}^{j} (1 - p_h), & j = k, \end{cases} \tag{2.15}$$

and λ_i^d is the time of last capture l_i for all individuals observed, and it is the first period in which the individual was alive (obtained from b) for individuals $i = u. + 1, \ldots, v$ that were not observed.

For the parameter v we specify a discrete uniform prior distribution $DU(0, \kappa_v)$. Obtaining a sample from the full conditional distribution for v has two problems:

(1) the full conditional distribution is only known to a proportionality constant,
(2) the value of the parameter changes the dimension of other unknowns in the model.

TABLE 2.1. Summary statistics for fitting the Crosbie-Manly-Arnason-Schwarz model to the meadow vole data. The statistic u_j is the number of individuals caught in sample j, R_j is the number of individuals released following sample j, q_j is the number of the R_j that were ever recaptured, and m_j is the number of marked animals caught in sample j.

Month	$u_j = R_j$	q_j	m_j
1	96	81	—
2	42	74	76
3	27	65	71
4	30	61	70
5	38	82	59
6	61	—	87

TABLE 2.2. Maximum likelihood estimates of identifiable parameters in the Crosbie-Manly-Arnason-Schwarz model fitted to the meadow vole data.

Parameter	Estimate	SE	95% C. I. lower	upper
S_1	0.875	0.041	0.770	0.936
S_2	0.659	0.049	0.559	0.747
S_3	0.681	0.050	0.576	0.770
S_4	0.619	0.050	0.518	0.710
p_2	0.905	0.040	0.792	0.960
p_3	0.855	0.047	0.737	0.925
p_4	0.934	0.036	0.816	0.979
p_5	0.909	0.039	0.800	0.961
f_2	0.220	0.056	0.130	0.347
f_3	0.253	0.062	0.152	0.391
f_4	0.378	0.078	0.242	0.538

To overcome the first problem we can use a sampling scheme, such as the Metropolis-Hastings algorithm, that allows us to sample from a distribution that we only know up to the proportionality constant. The second problem requires an extension of the Metropolis-Hastings algorithm called reversible jump Markov chain Monte Carlo [3] where there is dimension matching to ensure reversibility of the Markov chain. Details of the reversible-jump sampler are given in [11].

2.5. Example. We illustrate the use of the MCMC algorithm described in the previous section by fitting the model (2.3) to meadow vole (*Microtus pensylvannicus*) data collected at the Patuxent Wildlife Research Center, Laurel, Md, USA [12]. The meadow vole population was trapped at one-month intervals; untagged animals were tagged and released, tagged animals had their identity recorded and were then released. The model (2.3) can be fitted using sufficient statistics (Table 2.1) and closed-form solutions to the ML equations [7]. The ML estimates (Table 2.2) and the posterior summaries from fitting the model using the Gibbs sample algorithm (Table 2.3) of the previous section are

TABLE 2.3. Posterior distribution summary statistics for the Gibbs sampler posterior simulations and for the identifiable parameters in the Crosbie-Manly-Arnason-Schwarz model fitted to the meadow vole data. Included are predictions of the abundances N_j at the time of each sample and B_j, the number of individuals born in (t_j, t_{j+1}).

Parameters	Mean	Median	Quantile		
			0.025	0.5	0.975
S_1	0.874	0.042	0.788	0.876	0.953
S_2	0.659	0.049	0.563	0.659	0.755
S_3	0.681	0.051	0.580	0.682	0.779
S_4	0.616	0.050	0.517	0.616	0.711
f_2	0.222	0.058	0.123	0.218	0.348
f_3	0.257	0.064	0.147	0.252	0.397
f_4	0.377	0.079	0.241	0.370	0.549
p_2	0.886	0.043	0.791	0.891	0.956
p_3	0.839	0.048	0.733	0.843	0.921
p_4	0.913	0.041	0.819	0.919	0.975
p_5	0.891	0.041	0.799	0.896	0.958
Realized random variables					
N_2	132.4	5.1	125	132	145
N_3	116.3	5.1	108	116	128
N_4	108.8	3.9	103	108	118
N_5	108.1	3.7	103	107	117
B_2	28.7	4.1	21	29	37
B_3	29.3	3.4	22	29	36
B_4	41.0	3.6	35	41	49

in close agreement. For the Gibbs sampler we used beta $\mathrm{Be}(1,1)$ priors for S_i, p_i, and β'_i with the f_i parameters found deterministically as a function of the β_i and S_i parameters [7]. We used a discrete uniform $\mathrm{DU}(200000)$ prior for ν.

A nice feature of the MCMC approach is that it is relatively simple to obtain a posterior distribution for unobserved random variables of interest such as N_i and B_i, also reported in Table 2.3. Predictions for these in a likelihood analysis must be based on method of moment-type estimators as the N_i and B_i parameters do not explicitly appear in the likelihood function. Obtaining predictions using MCMC is straightforward with values sampled from the posterior predictive distribution. In practice, we simply use the set of b and d values obtained in the Markov chain and for each iteration compute the current value of the variable of interest. For example, [11] show that

$$N_j = \sum_{i=1}^{\nu} \left(\sum_{h=0}^{j-1} b_{ih} - \sum_{h=1}^{j-1} d_{ih} \right). \tag{2.16}$$

Progress of the Markov chains is instructive (Figure 2.2). Convergence of the Markov chains for the identifiable parameters $S_1, \ldots, S_{k-2}$; $p_2, \ldots, p_{k-1}$; $f_2, \ldots, f_{k-2}$ is rapid and the
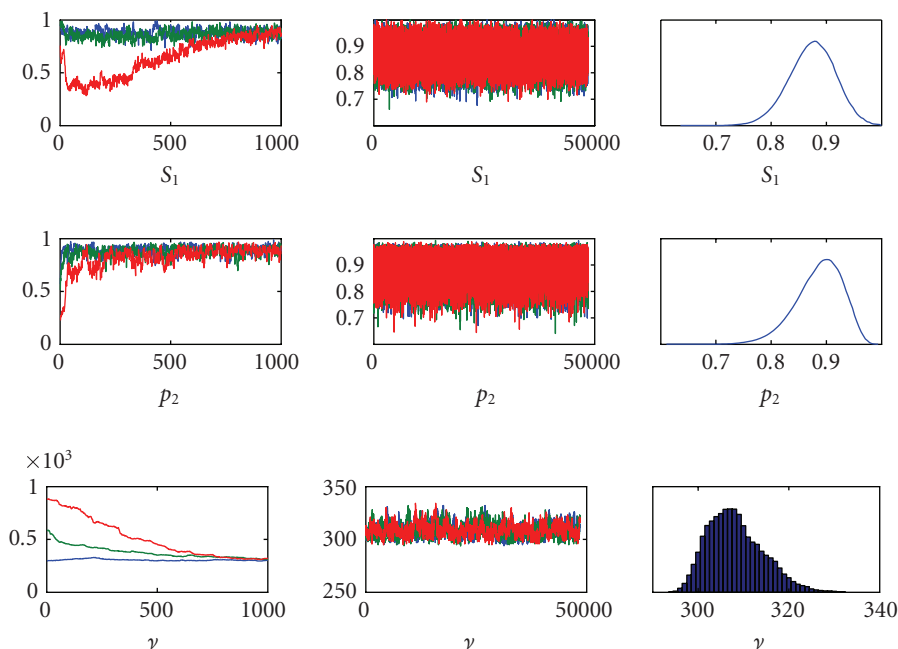
FIGURE 2.2. Gibbs sampler Markov chains for the first 1000 iterations (column of plots on left-hand side), for 50 000 iterations following a discarded burn-in of 10 000 iterations (middle column of plots) and posterior density plots (right-hand column of plots) for the identifiable parameters S_1 and p_1 and the nonidentifiable parameter ν, for the Gibbs sampler posterior simulations from fitting the Crosbie-Manly-Arnason-Schwarz model fitted to the meadow vole data.

chains appear well-mixed after a few thousand iterations. Even chains for nonidentifiable parameters such as ν appear well mixed after a few thousand iterations.

3. Discussion

Markov chain theory plays an important role in statistical inference both in the formulation of models for data and in the construction of efficient algorithms for inference. The ability to compactly represent stochastic processes that evolve over time has meant that Markov chain theory has played an important role in describing animal and plant population dynamics. The fact that each generation gives rise to the next makes the Markov property a natural starting point in developing population models.

The role of the Markov chain in statistical inference is a much more recent development. Although the Metropolis-Hastings algorithm first appeared in 1953 [13], with roots back to the Manhattan project, and the Gibbs sampler in 1984 [2] MCMC entered into widespread use in statistics only in the last 15 years. This followed [14] who illustrated the application of Gibbs sampling to a variety of statistical problems.

MCMC is particularly useful for fitting hierarchical models using Bayesian inference methods. A simple example of a hierarchical model is one where we have two components

to the model $[Y \mid \theta][\theta \mid \gamma]$. Here, conditional on parameters θ, the random variable Y has density $[Y \mid \theta]$. The parameters θ are themselves modeled as random variables and have distribution with pdf $[\theta \mid \gamma]$. In order to fit this model by a method such as maximum likelihood (ML), we must find the marginal likelihood: $\mathcal{L}(\theta \mid y) = \int_\theta [y \mid \theta][\theta \mid \gamma]d\theta$. In practice this integral is very difficult, if not practically impossible, in most cases of interest. The equivalent Bayesian problem is to find the posterior marginal density $[\gamma \mid y] \propto [Y \mid \theta][\theta \mid \gamma][\gamma]$ which involves essentially the same integral. However, MCMC allows us to approximate a sample from the required density but we must specify a prior distribution from γ. Thus, Bayesian inference methods based on MCMC allows inference for the model $[Y \mid \theta][\theta \mid \gamma]$ but the same problem may be difficult or impossible using standard ML theory.

In the context of mark-recapture modeling, the advent of MCMC has meant that it is now possible to fit complex hierarchical models. Starting with a model such as (2.3), it is relatively easy to add hierarchical structure. In the context of the hidden Markov model described in Figure 2.1, we mean that it is relatively straightforward to add in components that allow us to further model $[X \mid \beta, S, \nu]$. These issues have been explored by [7, 11]. [7] showed how the model (2.3) could be used to explore relationships among parameters and in particular considered a hierarchical model in which it was assumed that survival probabilities were related to per capita birth rates (functions of the β parameters). Such a model would make sense when there were common environmental influences on both survival probabilities and birth rates. Such an analysis is virtually impossible using ML methods. Similarly, [11] showed how the CDL discussed above could be used in hierarchical modeling, and in particular fit models in which there is feedback between the unknown random variables N_j and survival S_j or per capita birth rates f_j. The practical benefit of hierarchical modeling is that a much richer class of models is available for the ecologist to explore, and the advent of MCMC means that methods of inference are fully able to incorporate all different sources of uncertainty in the analysis.

A particularly important technical issue with MCMC is that of convergence. A typical problem in Markov chain theory is the determination of a stationary distribution given the transition kernel. In MCMC, the problem is reversed, and is one of constructing a transition kernel that has, as its stationary distribution, the target distribution of interest. Under mild conditions, the Metropolis-Hastings algorithm and the Gibbs sampler as a special case are methods of constructing transition kernels that have the required property. An important practical problem is the development of rules for helping determine when the chain has converged. Experience indicates that for some problems, convergence can be rapid and in others it can be slow. Rapid convergence appears to be an associated with likelihood functions that are well behaved. Methods for assessing convergence are ad hoc, and generally based on informal graphical assessment (Figure 2.2) or computation of simple statistics, such as the Brooks-Gelman-Rubin diagnostic statistic. There is surprisingly little underlying theory about the rate at which Markov chains constructed in MCMC should converge to the stationary distribution that can be applied to the practical problems of MCMC. Clearly, time to stationarity will be a property of the model. Again, observation indicates that problems for which the likelihood function are well behaved tend to converge rapidly. The work in [15, 16] is important in this regard.

References

[1] G. A. F. Seber, *The Estimation of Animal Abundance and Related Parameters*, Macmillan, New York, NY, USA, 2nd edition, 1982.

[2] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[3] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[4] R. M. Cormack, "Estimates of survival from the sighting of marked animals," *Biometrika*, vol. 51, pp. 429–438, 1964.

[5] G. M. Jolly, "Explicit estimates from capture-recapture data with both death and immigration-stochastic model," *Biometrika*, vol. 52, pp. 225–247, 1965.

[6] G. A. F. Seber, "A note on the multiple-recapture census," *Biometrika*, vol. 52, pp. 249–259, 1965.

[7] W. A. Link and R. J. Barker, "Modeling association among demographic parameters in analysis of open population capture-recapture data," *Biometrics*, vol. 61, no. 1, pp. 46–54, 2005.

[8] R. Pradel, "Utilization of capture-mark-recapture for the study of recruitment and population growth rate," *Biometrics*, vol. 52, no. 2, pp. 703–709, 1996.

[9] C. J. Schwarz and A. N. Arnason, "A general methodology for the analysis of capture-recapture experiments in open populations," *Biometrics*, vol. 52, no. 3, pp. 860–873, 1996.

[10] S. F. Crosbie and B. F. J. Manly, "Parsimonious modelling of capture-mark-recapture studies," *Biometrics*, vol. 41, no. 2, pp. 385–398, 1985.

[11] M. R. Schofield and R. J. Barker, "A unified capture-recapture framework," to appear in *Journal of Agricultural, Biological and Ecological Statistics*.

[12] B. K. Williams, J. D. Nichols, and M. J. Conroy, *Analysis and Management of Animal Populations*, Academic Press, London, UK, 2002.

[13] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[14] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398–409, 1990.

[15] J. J. Hunter, "Mixing times with applications to perturbed Markov chains," *Linear Algebra and Its Applications*, vol. 417, no. 1, pp. 108–123, 2006.

[16] J. J. Hunter, "Coupling and mixing times in a Markov chains," *Research Letters in Information and Mathematical Science*, vol. 11, pp. 1–22, 2007.

Richard J. Barker: Department of Mathematics and Statistics, University of Otago,
P.O. Box 56, Dunedin 9010, New Zealand
Email address: rbarker@maths.otago.ac.nz

Matthew R. Schofield: Department of Mathematics and Statistics, University of Otago,
P.O. Box 56, Dunedin 9010, New Zealand
Email address: mschofield@maths.otago.ac.nz