Hierarchical Capture-Recapture Models

Matthew R. Schofield

A thesis submitted for the degree of

Doctor of Philosophy in Statistics

at the University of Otago, Dunedin,

New Zealand

2007

Preface

A large proportion of the work in this thesis has been submitted for publication in various forms. I am lead author on Schofield and Barker (2008, 2007b,a) and co-author on Barker et al. (2008) and Barker and Schofield (2008).

Abstract

A defining feature of capture-recapture is missing data due to imperfect detection of individuals. The standard approach used to deal with the missing data is to integrate (or sum) over all the possible unknown values. The missing data is completely removed and the resulting likelihood is in terms of the observed data. The problem with this approach is that often biologically unnatural parameters are chosen to make the integration (summation) tractable. A related consequence is that latent variables of interest, such as the population size and the number of births are only available as derived quantities. As they are not explicitly in the model they are not available to be used in the model as covariates to describe population dynamics. Therefore, models including density dependence are unable to be examined using standard methods.

Instead of explicitly integrating out missing data, we choose to include it using data augmentation. Instead of being removed, the missing data is now present in the likelihood as if it were actually observed. This means that we are able to specify models in terms of the data we would like to have observed, instead of the data we actually did observe. Having the complete data allows us to separate the processes of demographic interest from the sampling process. The separation means that we can focus on specifying the model for the demographic processes without worrying about

the sampling model. Therefore, we no longer need to choose parameters in order to simplify the removal of missing data, but we are free to naturally write the model in terms of parameters that are of demographic interest. A consequence of this is that we are able write complex models in terms of a series of simpler conditional likelihood components. We show an example of this where we fit a CJS model that has an individual-specific time-varying covariate as well as live re-sightings and dead recoveries.

Data augmentation is naturally hierarchical, with parameters that are specified as random effects treated as any other latent variable and included into the likelihood. These hierarchical random effects models make it possible to explore stochastic relationships both (i) between parameters in the model, and (ii) between parameters and any covariates that are available.

Including all of the missing data means that latent variables of interest, including the population size and the number of births, can now be included and used in the model. We present an example where we use the population size (i) to allow us to parameterize birth in terms of the per-capita birth rates, and (ii) as a covariate for both the per-capita birth rate and the survival probabilities in a density dependent relationship.

Acknowledgements

First and foremost I want to express my gratitude to Prof. Richard Barker. He was very generous with his time and was always willing to help me. Any material of worth in this thesis is due to his influence. He is not only a supervisor, but also a friend.

I would have been unable to complete this Ph.D. without the scholarship provided from the Tertiary Education Commission. I am also thankful to the others who have provided travel funding throughout my Ph.D. These include the Division of Sciences and the Department of Mathematics of Statistics at the University of Otago, as well as the New Zealand Statistics Association.

My thanks goes to everybody in the Department of Mathematics and Statistics. The support staff, in particular, have been incredible. Irene, Leanne and Lenette were always more than willing to assist with any difficulty I faced.

I thank both Jason Rabbitt and Andrew Gormley for being more than hospitable in sharing an office with me.

Much of the credit of the thesis goes to my parents. Without their love and support I would not be here. Their love for each other, as well as for their children and the people around them is an example of how I want to live my life.

I cannot put into words the gratitude I have toward my wife, Bronwyn.

The list of sacrifices that she has made during this research would not fit on the pages of this thesis. I am sure that she had no idea of what she was in for when she said 'I do', but I have been thankful every day since that she did!

Perhaps the largest influence on this thesis has been from Jesus Christ. Without the sacrifice that he made 2000 years ago, I am sure that this thesis would not have been written.

> In him [Jesus Christ] lie hidden all the treasures of wisdom and knowledge. *Colossians 2:3*

Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Capture-Recapture Methods

Ecologists are interested in examining the dynamics of animal populations. In particular they want to describe demographic change in terms of survival probabilities, movement rates and fecundity and examine the effect of covariates on these processes. The problem is that we are unable to directly observe survival and fecundity rates and need some way of estimating these. One way to do this is with data from capture-recapture experiments. Capture-recapture experiments involve catching and marking a sample of animals from the population at some initial time t_1. The mark is in the form of a unique tag from which the animal can be identified in the future. The animals are then released back into the population. At $k-1$ subsequent times, $t_2, \ldots, t_k$ further samples are taken. Recaptures of previously marked animals are recorded, with any unmarked animals tagged.

The data from a capture-recapture study can be written in terms of a $u.$ by k capture matrix $\boldsymbol{X}^{obs}$, where $u.$ is the total number of individuals ever

observed. A typical capture history looks like

$$
\boldsymbol{X}^{obs} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}.
$$

The value $X_{ij}^{obs} = 1$ denotes a capture for individual i in sample j, with $X_{ij}^{obs} = 0$ otherwise. The superscript obs is used because $\boldsymbol{X}^{obs}$ only contains the capture histories for those individuals observed in one or more samples. It does not include the capture histories for the individuals in the population that were never caught.

In many capture-recapture experiments, covariates are collected when the individuals are caught. These covariates could be individual-specific, such as sex, length or weight, or they could be common to all individuals caught in that sample, for example, temperature.

The assumptions placed on animal behaviour both within and between the sampling occasions gives rise to a multitude of capture-recapture models (Otis et al. 1978, Pollock et al. 1990). For example, closed population models arise when the time period between capture occasions is short and it is assumed the population is closed to immigration, emigration, birth and death. Open population models arise when any of these assumptions is relaxed. It is also possible to have models that combine these assumptions. For example, the robust design assumes that there are series of capture occasions for which the population is closed interspersed with periods for which the population is assumed open (Pollock 1982).

The focus of this thesis is on population monitoring in ecology. However, the methods developed can be easily transferred to other areas where

capture-recapture models are applied, for example, in computer science (Briand et al. 1997) and in multiple-list studies in epidemiology (Hook and Regal 1995).

As much as possible we attempt to keep notation consistent. A full description of the notation used is given in appendix A.

1.1   Closed population models

The interest in closed population models is estimating the size of the population, denoted $\aleph$. In order to estimate $\aleph$, we write the likelihood in terms of capture probabilities

$$
\boldsymbol{p} = \begin{pmatrix} p_{11} & \cdots & p_{1k} \\ \vdots & & \vdots \\ p_{\aleph 1} & \cdots & p_{\aleph k} \end{pmatrix},
$$

where p_{ij} is the probability of individual i being caught in sample j. We use information from the individuals observed as well as assumptions about $\boldsymbol{p}$ to estimate the number of individuals with capture history $(0, \ldots, 0)$ that were available for capture but never seen. The standard likelihood is

$$
\mathcal{L}(\boldsymbol{p}, \aleph; X^{obs}) \propto \frac{\aleph!}{u.!(\aleph - u.)!} \prod_{i=1}^{u.} \prod_{j=1}^{k} p_{ij}^{X_{ij}^{obs}} (1 - p_{ij})^{1 - X_{ij}^{obs}} \prod_{i=u.+1}^{\aleph} \prod_{j=1}^{k} (1 - p_{ij})
$$

$$(1.1)$$

This is the most general likelihood with assumptions about $\boldsymbol{p}$ leading to different classes of model. Otis et al. (1978) give details about 8 classes of model: M_0, M_t, M_b, M_h, M_{tb}, M_{th}, M_{bh}, M_{tbh}.

1.1.1   Assumptions About The Capture Process

Model M_0

Model M_0 has one capture probability that is constant across individual and time,

$$p_{ij} = p, \; i = 1, \ldots, \aleph, \; j = 1, \ldots, k.$$

Model M_t

Model M_t has a different capture probability for each sample occasion which is constant across individual,

$$p_{ij} = p_j, \; i = 1, \ldots, \aleph, \; j = 1, \ldots, k.$$

Model M_b

Model M_b has two capture probabilities: p if an individual has never been caught and c if an individual has been caught before,

$$p_{ij} = \begin{cases} p & \text{if } i \text{ has never been caught} \\ c & \text{if } i \text{ has been caught before} \end{cases}, \; i = 1, \ldots, \aleph, \; j = 1, \ldots, k.$$

This model allows for the study to influence the behaviour of the animals. If the capture process is a bad experience for the animals, we get a trap-shy response with $p > c$. However, if the capture process is a good experience for the animals, we get a trap-happy response with $p < c$. Note that these responses can also be induced by the study design used. For example, if our sampling targets areas where animals were previously caught, we are likely to observe a trap-happy response.

Model M_h

Model M_h has a different capture probability for each individual in the

population which is constant through time,

$$p_{ij} = p_i, \ i = 1, \ldots, \aleph, \ j = 1, \ldots, k.$$

The historically popular approach to estimate $\aleph$ for model M_h is to use the non-parametric jackknife of Burnham and Overton (1978). Alternatives include using finite mixtures for p (Pledger 2000) or assuming the capture probabilities p are drawn from a probability density function with support on $[0, 1]$. Holzmann et al. (2006) showed that the population size $\aleph$ is identifiable in model M_h conditional on a model for p . However, Link (2003) showed that different probability density functions for p can lead to identical sufficient statistics with different values of $\aleph$.

Higher Order Models

The higher order models M_{tb}, M_{th}, M_{bh} and M_{tbh} are all extensions of models M_t, M_b and M_h. For example, M_{th} assumes that each individual has their own capture probability that changes through time. Models M_{tb}, M_{th}, M_{bh} and M_{tbh} are not identifiable without further assumptions (Otis et al. 1978).

Log-linear models can also be used to fit closed population models (Fienberg 1972, Cormack 1989, Agresti 1994). A virtue of log-linear models is that they provide flexible expressions for the capture probabilities with interest in the estimation of $\aleph$.

1.1.2 Covariates

Often there are covariates z that can improve estimation of $\aleph$ because they help to explain capture probabilities. When the covariate is time-varying and common across individuals, we can include a model where the capture

probabilities p are deterministically modeled in terms of the covariate z (Lebreton et al. 1992).

The more challenging case is when there are individual-specific covariates that are assumed to fully explain differences in capture probabilities between individuals. The problem with this model is that we do not know the covariates values for the individuals that we have never observed. Huggins (1989) solved this problem by using a likelihood that conditioned on $u.$, the number of individuals ever observed. He uses a method of moments estimator for $\aleph$ based on the Horvitz-Thompson estimator (Horvitz and Thompson 1952),

$$\hat{\aleph} = \sum_{i=1}^{u.} \frac{1}{\hat{p}_i(z_i)}$$

where z_i is the covariate value for individual i. This varies from the true Horvitz-Thompson estimator

$$\hat{\aleph} = \sum_{i=1}^{u.} \frac{1}{p_i(z_i)}$$

through replacing $p_i(z_i)$ by its estimate $\hat{p}_i(z_i)$. Instead of conditioning on $u.$, an alternate approach is to include $\aleph$ in the likelihood as a parameter and complete the data using data augmentation. This approach is investigated in more detail in section 3.1.1.

1.1.3 The Problem with Closed Population Modeling

As stated at the beginning of chapter 1, interest lies in examination of demographic change in the population. In this regard, there is little value in knowing $\aleph$. The interest is in investigating how $\aleph$ changes through time, in particular with respect to covariates. Open population models were introduced to investigate these changes in $\aleph$.

1.2 Open population models

1.2.1 Cormack-Jolly-Seber Model

The Cormack-Jolly-Seber (CJS) model (Cormack 1964, Jolly 1965, Seber 1965) was a breakthrough for capture-recapture modeling. Through conditioning on $\boldsymbol{u} = (u_1, \dots, u_k)$, where u_j is the number of unmarked individuals caught in sample j, a maximum likelihood framework was obtained in which the closed population assumption was relaxed by allowing death to occur during the study. We express the inclusion of mortality through the partially observed death covariate $\boldsymbol{d}$, where d_{ij} takes the value 1 if individual i died between sample j and $j + 1$ and it takes the value 0 otherwise. The value $d_{ik} = 1$ is taken to mean that an individual was still alive in the last sampling occasion, enforcing the constraint that $\sum_{j=1}^{k} d_{ij} = 1$, $i = 1, \dots, u..$ We know that an individual did not die before the time in which we last saw it (the sample of last capture $\mathcal{L}_i$) so the values d_{ij} for $j = 1, \dots, \mathcal{L}_i - 1$ are known to be 0. The remaining d_{ij} values are missing subject to the constraint $\sum_{j=1}^{k} d_{ij} = 1$. There are many possible covariates that could be specified to include the information on death that is contained in $\boldsymbol{d}$. For example, an indicator matrix $\boldsymbol{a}$ could be used, where $a_{ij} = 1$ if individual i was alive at time of sample j and $a_{ij} = 0$ otherwise. The choice of $\boldsymbol{d}$ over $\boldsymbol{a}$, or any other alternatives, was made for convenience.

To model the mortality, we introduce a vector of survival probabilities $\boldsymbol{S}$,

$$\boldsymbol{S} = (S_1, \dots, S_{k-1}),$$

where S_j is the probability of surviving from sample j to $j + 1$ conditional on being alive at the time of sample j. Even though it is possible to have parameters varying across individuals (Pledger et al. 2003), the CJS model

assumes both capture and survival probabilities do not vary between individuals but can vary through time. The vector of capture probabilities p is

$$p = (p_2, \ldots, p_k).$$

The likelihood for the CJS model is obtained by finding the joint probability of the capture histories for all individuals, conditional on first capture

$$\mathcal{L}(S, p; X^{obs}) \propto \prod_{i=1}^{u.} \left\{ \prod_{j=\mathfrak{F}_i}^{\mathfrak{L}_i - 1} S_j p_{j+1}^{X_{ij+1}} (1 - p_{j+1})^{1 - X_{ij+1}} \right\} \chi_{\mathfrak{L}_i},$$

where $\mathfrak{F}_i$ is the sample of first capture for individual i and χ_j is the probability of individual i never being observed after sample j. The value $\chi_{\mathfrak{L}_i}$ is required because it is necessary to integrate (or in this case sum) over the unknown values of d and remove them from the likelihood. The value $\chi_k = 1$ is known and a recursive relationship gives,

$$\chi_j = (1 - S_j) + S_j(1 - p_{j+1})\chi_{j+1}, \ j = 1, \ldots, k - 1. \qquad (1.2)$$

Closed form solutions for the CJS model are available and are given by Pollock et al. (1990) for the $2k - 3$ identifiable parameters:

$$p_2, \ldots, p_{k-1}, S_1, \ldots, S_{k-2}, p_k S_{k-1}.$$

Covariates

Often there are additional covariates collected that are assumed to affect the parameters in the model. These covariates may have been collected for one of two reasons:

1. To improve the estimation of the model. The covariates are expected to at least partially explain the variability in the parameters, but there is very little interest in the relationship. An example of this is when sampling effort is used as a covariate for capture probabilities.

2. To investigate biologically interesting relationships. In this case the effect of the covariates improves knowledge about the population dynamics of the species. An example of this is the effect of rainfall on survival rates. Lebreton et al. (1992) gave a deterministic framework for modeling the parameters in terms of known covariates. One such model could be

$$\text{logit}(S_j) = \gamma_0 + \gamma_1 z_j, \ j = 1, \ldots, k-1,$$

where the survival probabilities are a function of the covariate z_j. Even thought this was a large step forward, it is usually unsatisfactory. This is because we are assuming that the covariate fully explains the variability in the parameter.

A preferable approach is to hierarchically model the parameters using random effects (see section 4.1). One potential hierarchical model is that the parameters are themselves modeled as having a probability distribution function whose (hyper)parameters are explained by the covariates. Hierarchical models are more realistic as they allow relationships to be specified that only partially explain the variability in the parameter.

Individual-Specific Time-Varying Covariates

Individual-specific time-varying covariates are difficult to incorporate, as

they include missing covariate values. To overcome this, a model for the covariate must be specified and all missing values integrated out. When the covariate is categorical the common model used is the first order Markovian multi-state model. Here the "state" of individual i in sample j depends only on the state of i in $j - 1$ (Schwarz et al. 1993b). An extension to this is where the "state" of individual i in sample j depends on the state of i in $j - 1$ and $j - 2$ (Brownie et al. 1993). So long as the number of "states" is not too large, the summation over the missing covariates is computationally feasible, with the first order Markovian multi-state model able to be fitted in popular capture-recapture software such as MARK and M-SURGE (White and Burnham 1999, Choquet et al. 2004).

Including individual-specific time-varying continuous covariates is even more difficult. This is because it is harder to integrate over the missing covariate values than it is to sum over a discrete number of values. In general, analytic integration is impossible and numerical integration using quadrature is too computationally burdensome to be useful in practice. However, Bonner and Schwarz (2006) specified an extension to the CJS model where the missing covariate values were integrated out using Markov chain Monte Carlo within a Bayesian framework (see chapter 5). A more detailed discussion of this models is given in section 4.2.2 with an example in chapter 9.

1.2.2 The Jolly-Seber Model

To allow for birth, Jolly (1965) and Seber (1965) included a model for $u = (u_1, \ldots, u_k)$, the number of unmarked individuals caught in each sample. They assumed that u_j, the number of unmarked individuals caught in sample j, is binomially distributed with index U_j, the total number of unmarked

individuals in the population immediately before sample j,[1]

$$u_j \sim B(U_j, p_j), \; j = 1, \ldots, k,$$

with the values U_1 and U_k not identifiable. Including the information on first capture means that derived estimates[2] can be obtained for demographic summaries such as $\boldsymbol{N} = (N_1, \ldots, N_k)$, where N_j is the number of individuals alive in sample j (Pollock et al. 1990). Even though the information is available to estimate these quantities, we are unable to fully exploit this information. The unobserved random variable N_j is not available for direct modeling because it does not appear explicitly in the likelihood. This means we are unable to use N_j as a covariate in the model without adopting an errors-in-variables approach (Barker et al. 2002).

Unfortunately the "birth" parameters that are specified in the model, U_j, are not particularly useful to a biologist as they combine aspects of both the demographic processes of interest and the intensity of sampling. If intense sampling was undertaken in sample $j - 1$ then U_j will be lower than if very little sampling was undertaken in sample $j - 1$. This means that the parameter U_j is of little use for further hierarchical modeling.

1.2.3   The Crosbie-Manly-Arnason-Schwarz Model

Schwarz and Arnason (1996) followed the lead of Crosbie and Manly (1985) and modeled the first captures in terms of $\aleph$, the total number of individuals ever available for capture throughout the study. The Crosbie-Manly-Arnason-Schwarz (CMAS) model includes an unknown vector $\boldsymbol{B} = (B_0, \ldots, B_{k-1})$, where B_j is the number of individuals born between sample j and $j + 1$,

---

[1]Actually Jolly (1965) modeled r_j, the number of u_j subsequently released to take into account loss on capture.

[2]Note that only $N_2, \ldots, N_{k-1}$ are identifiable.

as the outcome of a multinomial distribution with index $\aleph$ and probabilities $\beta = (\beta_0, \ldots, \beta_{k-1})$,

$$\boldsymbol{B} \sim MN(\aleph, \beta),$$

with the constraint that $\sum_{h=0}^{k-1} \beta_h = 1$. While this is a step toward modeling in terms of biologically useful parameters, there are three problems:

1. The model is not identifiable, with $\aleph$, β and p_1 confounded.

2. The parameters β, while not sensitive to sampling intensity, are sensitive to aspects of the study design. To illustrate, consider a study with $k = 3$ and

$$\beta = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right).$$

   If the study were extended for a further s periods and birth occurred in at least one of those periods, then the values of β_0, β_1 and β_2 would all change. Therefore, as with the JS model there is little interest in the hierarchical modeling of β, as they are not natural demographic parameters.

3. Random variables of interest, such as N_j, are still not available for direct modeling and are only available as derived parameters.

1.2.4 Models with Natural Birth Parameters

Pradel (1996) and Link and Barker (2005) proposed models that parameterize birth in terms of per-capita birth rate index[3],

$$\eta'_j = \frac{E(B_j)}{E(N_j)}.$$

---

[3]The likelihoods of Pradel (1996) and Link and Barker (2005) are different but yield the same parameter estimates when there is no loss on capture. See Link and Barker (2005) for details about the difference if there is loss on capture.

It is referred to as an index because the expected abundance $E(N_j)$ is used instead of the actual value N_j. Both Pradel (1996) and Link and Barker (2005) construct the likelihood conditional on $u_.$ instead of $\aleph$, with Link and Barker (2005) justifying this by showing that the distribution $[u_.|\aleph]$ contains very little, if any, information about the identifiable parameters. Parameterizing in terms of per-capita birth rates has two advantages over the model of Schwarz and Arnason (1996):

1. The per-capita birth rate index is a more natural parameter for modeling. It describes the demographic change due to birth and does not depend on aspects of either study design or sampling intensity. For this reason Link and Barker (2005) included a hierarchical model where the per-capita birth rate index and survival probabilities were correlated. Including models of this sort allow biologists to investigate interesting relationships between birth and survival as described in section 4.1.

2. The model is specified in terms of identifiable parameters (Link and Barker 2005).

However, as with the Jolly-Seber and CMAS models, we are unable to include random variables of interest, such as N_j, directly in the likelihood to be used for modeling.

1.2.5   Dead Recoveries and Live Re-sightings

In many situations dead recoveries and/or live re-sightings occur outside of the sampling occasions. An example is in fisheries, where an angler might catch a fish and report the tag number. If the angler returns the fish after capture, the observation is a live re-sighting, otherwise it is a dead recovery. The recoveries and re-sightings may continue well after sampling has

finished. To account for this, we split the time after the end of sampling into periods of approximately the same length as the time between sampling periods. We denote the total number of re-sighting and recovery periods as k'.

The model of Burnham (1993) includes only the information from dead recoveries. An additional vector of parameters $r = (r_1, \ldots, r_{k'})$ is required in the model, where r_j is the probability that an individual that has died between sample j and $j + 1$ is reported dead.

The model of Barker (1997) allows live re-sightings as well as dead recoveries. As well as r, two additional vectors of parameters are required: $R = (R_1, \ldots, R_{k'})$ and $R' = (R'_1, \ldots, R'_{k'})$, where R_j is the probability an individual that survived until $j + 1$ is re-sighted alive between sample j and $j + 1$, and R'_j is the probability that an individual that has died between sample j and $j + 1$ is re-sighted alive without being found dead. Other parameterizations are also available. For example, one can parameterize in terms of R, $f = (f_1, \ldots, f_{k'})$ and $\nu = (\nu_1, \ldots, \nu_{k'})$, where f_j is the probability an individual is either re-sighted or recovered between j and $j + 1$ given that they were alive in sample j and ν_j is the probability that an individual is re-sighted alive given they were either re-sighted alive or recovered dead between j and $j + 1$. This parameterization can be used in fisheries where re-sightings and recoveries are obtained from anglers who catch fish and then decide whether to keep or return the fish. We use the R, f and ν parameterization in chapter 9.

The practical problem with having multiple sources of data is that it makes the likelihood, in particular, χ_j, more complex than for the standard CJS model, for example, see Barker et al. (2004). The benefit, however, is that so long as the standard regulatory conditions hold for the likelihood,

then the additional data will increase the amount of information about the parameters of interest, at least asymptotically (Barker and Kavalieris 2001).

Robust Design

Pollock (1982) suggests a study design for open population analysis that is robust to unequal probability of capture. This so-called robust design brings together the best elements of both closed population and open population estimation into one model. There are k primary sample periods in which the population is assumed to be open. Then, within each primary period there are a further ℓ secondary sampling periods in which the population is assumed closed[4]. A graphical representation is given in figure 1.1.

Primary
Periods

$$
\begin{array}{ccc}
1 & 2 & \cdots & k
\end{array}
$$

Secondary
Periods

$$
\begin{array}{ccccccccc}
1 & 2 \ldots \ell & & 1 & 2 \ldots \ell & & 1 & 2 \ldots \ell
\end{array}
$$
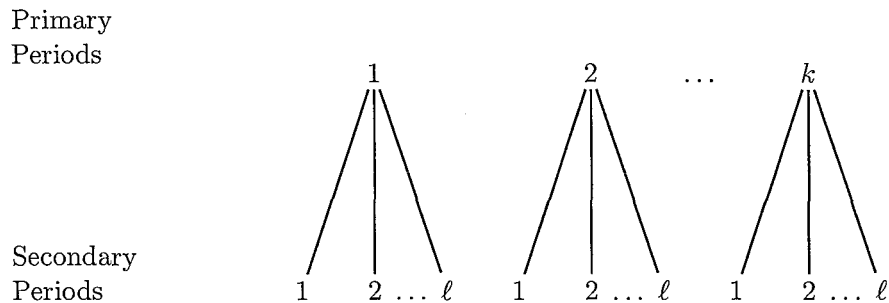
Figure 1.1: Graphical description of the robust design as given in Pollock et al. (1990).

Using the robust design allows us to choose closed population models such as model M_h for p. This relaxes the assumption that p is constant across individuals and makes the model robust to unequal capture probability. Many different versions of the robust design have been suggested and implemented, including combining it with other models such as the

---

[4]There need not be the same number of secondary periods in each primary periods.

multi-state model.

1.3   Missing Data in Capture-Recapture Models

A comparison of all the standard models described above reveals two common features:

1. Interest is in the flexible modeling of the demographic processes such as survival probabilities and birth rates.

2. Imperfect detection of individuals leads to missing data. All of the methods described above, with the exception of Bonner and Schwarz (2006), explicitly integrated out these missing components. Data augmentation (Tanner and Wong 1987) is an alternative method for incorporating missing data where all missing data is included directly into the likelihood. Having the missing data available leads naturally to hierarchical models that explore relationships between demographic parameters. Missing data methods, in particular data augmentation, are explored in the next chapter.

Chapter 2

Missing Data Methods

The process by which the data go missing is often referred to as the missingness mechanism. Anytime we have any missing data, we must determine whether the missingness mechanism for the data is ignorable. If the missingness mechanism is not ignorable then we need to explicitly model the process by which the data went missing.

To ascertain whether the missingness mechanism is ignorable we define y to be the matrix/vector of complete data, which is the data we would like to have observed, irrespective of whether we did or not. We also define an indicator $\mathcal{I}$, where

$$
\mathcal{I}_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is observed} \\ 0 & \text{if } y_{ij} \text{ is missing} \end{cases} \quad, \ \forall i, j.
$$

We separate the data y into an observed component y^{obs} and a missing component y^{mis} such that $y = (y^{obs}, y^{mis})$.

2.1   Ignorability

The missingness mechanism is non-ignorable if either:

1. The distribution of $\mathcal{I}$ depends on y^{mis},

$$[\mathcal{I}|y^{mis}, y^{obs}, \phi],$$

   where ϕ are the parameters modeling the missingness mechanism and
   $[x]$ denotes the probability density function for continuous x or prob-
   ability mass function for discrete x. Consider an example where a
   survey was conducted to investigate the relationship between age and
   income of respondents. Suppose that all respondents were first asked
   for their age[1]. Then all respondents were asked to specify their income
   if it was under $\$Z$. The missing income data is non-ignorable because
   the process by which it goes missing depends on the unknown value
   of the income.

2. The parameters used to model the data y (denoted as θ) are also used
   to model $\mathcal{I}$, that is $\phi = f(\theta)$. Note that if inference is conducted using
   Bayesian methods (see chapter 5), then the missingness mechanishm
   is also non-ignorable if ϕ and θ are separate but not distinct[2] (Gelman
   et al. 2004).

If the missingness mechanism does not depend on the missing data values
and ϕ and θ are separate and distinct, then the missingness mechanism is
said to be ignorable and $\mathcal{I}$ does not need to be included in the model.
Ignorable missing data can be further classified as either missing at random

---

[1]For the sake of the example we assume that all participants answered all questions
asked of them.

[2]Distinct parameters are independent in their prior distribution.

(MAR) or missing completely at random (MCAR). Assuming that ϕ and θ are separate and distinct, data are MAR if the missingness mechanism does not depend on the missing data values but does depend on the observed data values,

$$[\mathcal{I}|y^{obs}, \phi].$$

Gelman et al. (2004) uses the survey for income and age above to give an example of data MAR. As before, respondents were first asked for their age. Then only respondents under the age of Z were asked for their income. The missing income data is classified as MAR because the process by which it goes missing depends on observed data (the respondents age) and not the underlying value of the missing data (the respondents income).

Data are MCAR if the missingness mechanism does not depend on the data,

$$[\mathcal{I}|\phi].$$

In the survey for age and income, suppose we first ask respondents for their age. Then we flip a coin to decide whether or not to ask for their income. The missing income data is MCAR because the process by which the data go missing has nothing to do with either age or income.

To show that the missingness mechanism for data MAR or MCAR is ignorable, first consider the case where the missingness mechanism depends on the missing data

$$[\mathcal{I}|y^{mis}, y^{obs}, \phi].$$

The joint distribution of inclusion vector $\mathcal{I}$ and the observed data is

$$[y^{obs}, \mathcal{I}|\theta, \phi] = \int [y|\theta][\mathcal{I}|y, \phi] \, dy^{mis},$$

where θ are the parameters of interest. As we are unable to move the missingness mechanism out of the integral (as it involves y^{mis}) any inference about θ must include $[\mathcal{I}|y, \phi]$. However, if the missing data are MAR then,

$$[y^{obs}, \mathcal{I}|\theta, \phi] = [\mathcal{I}|y^{obs}, \phi] \int [y|\theta] \, dy^{mis}$$
$$= [\mathcal{I}|y^{obs}, \phi][y^{obs}|\theta].$$

Now any inference about θ need only include the term $[y^{obs}|\theta]$. Therefore, as far as inference about θ is concerned, the missingness mechanism is ignorable. The same process can be followed for data MCAR.

2.2   Data Augmentation and Latent Variables

Irrespective of whether the missingness mechanism is ignorable or not, the missing data itself needs to be specified and integrated out of the model. The standard approach is to explicitly integrate the missing data out of the model. However, there is an alternative approach called data augmentation where missing data is included in the model and integrated out as part of the model fitting procedure (Tanner and Wong 1987). Two of the commonly used algorithms for data augmentation are the EM algorithm and Markov chain Monte Carlo (MCMC) methods.

Data augmentation need not be restricted to the inclusion of missing data. Any latent/unobserved variables can be included, whether they are natural to the model or not. For example, slice sampling (see section 5.2.2) is a MCMC simulation technique where a latent variable is included for the sole purpose of invoking a Gibbs sampler in order to improve convergence of the algorithm.

2.2.1   Difficulties of Working With Data Augmentation

The key to data augmentation is specifying the latent variables in a way that
will ensure rapid convergence of the computational algorithm used to fit the
model. The problem is that the same model with a different expression
for the latent variables can converge at different rates (Meng and van Dyk
1997). Consider a hierarchical random effects model that can be written
either as,

$$x \sim N(z, \sigma^2)$$
$$z \sim N(\theta, \eta^2),$$

where z is latent, or as

$$x \sim N(\theta + \epsilon, \sigma^2)$$
$$\epsilon \sim N(0, \eta^2),$$

where ϵ is latent. Even though the two models are the same, the conver-
gence of the parameters in these two expressions can be drastically different
(Papaspiliopoulos et al. 2003, Neal and Roberts 2005). While there are the-
oretical reasons why one parameterization will work better than the other
for a given dataset,[3] often the best practical approach is a time consuming
one: chose one expression and if convergence is not satisfactory, try another
expression. For this reason, the development of efficient computational al-
gorithms that improve the convergence, irrespective of the parameterization
used, is an area of current research (van Dyk and Meng 2001).

---

[3] See Meng and van Dyk (1997), Papaspiliopoulos et al. (2003), Neal and Roberts
(2005) for details.

2.2.2 Complete Data Likelihood

We follow the notation of Gelman et al. (2004) and refer to the likelihood that must be used for inference as the observed data likelihood (ODL). This consists of the vector $\mathcal{I}$ and the observed data values[4],

$$\mathcal{L}_O(\theta, \phi; y^{obs}, \mathcal{I}) \propto [y^{obs}, \mathcal{I}|\theta, \phi].$$

Including the missing data through data augmentation means we have the complete data y available for modeling. Again we follow Gelman et al. (2004) and refer to this as the complete data likelihood (CDL),

$$\mathcal{L}_C(\theta, \phi; y, \mathcal{I}) \propto [y, \mathcal{I}|\theta, \phi].$$

A simple example that compares the difference between modeling using (i) the observed data likelihood directly and (ii) the complete data likelihood is given in appendix B.

Even though we are integrating out the missing data less efficiently, as we are integrating iteratively instead of explicitly, there can be a large advantage to working with the CDL. Often the CDL is in a form that is preferable to work with and has a natural factorization that the ODL does not. Essentially this is because we are modeling the data we would like to have observed instead of the data we have actually observed. This is the case when using data augmentation to obtain the complete data likelihood for the (open population) capture-recapture problem. The CDL presents the data that demographers would like to have in order to build hierarchical models to investigate demographic change. We are able to separate the nuisance capture component from the demographic processes of interest, such as birth

---

[4]Note that for ignorable missingness mechanisms we need not include $\mathcal{I}$ and ϕ.

and death. This allows the focus to be on building sensible and relevant scientifically driven models for examining relationships of interest instead of focusing on the complexities of obtaining the incomplete data.

Part II

Hierarchical

Capture-Recapture Models

Chapter 3

Capture-Recapture Using Data Augmentation

The first step in developing a framework for capture-recapture models using data augmentation is to incorporate standard models such as closed population, CJS, JS and CMAS models. Once these models have been developed, we can begin to investigate hierarchical extensions.

3.1 Closed Population Models

The complete data for the closed population capture-recapture model is the complete set of capture histories for every individual ever available for capture, X. The matrix X comprises the capture histories for all individuals available for capture X^{obs} and the capture histories for individuals we never saw X^{mis}.

$$X = \begin{pmatrix} X^{obs} \\ X^{mis} \end{pmatrix}.$$

The component X^{mis} is a matrix of zeros of unknown dimension. To overcome this, we specify $\aleph$, the total number of individuals ever available for capture as a parameter in the model, so that, conditional on $\aleph$ the matrix of capture histories X is fully specified. A separate variable $\mathcal{I}$ that specifies whether the data are observed or missing is not included in the model because $\mathcal{I} = f(X)$. Therefore, the missingness mechanism is already accounted for through the capture histories X.

The CDL for this problem is

$$\mathcal{L}_C(p, \aleph; X) \propto [X|p, \aleph]$$

where p is the matrix of capture probabilities. The joint distribution for X is the product of a series of Bernoulli distributions with a probability of capture p_{ij} for individual i in sample j,

$$[X|p, \aleph] \propto \frac{\aleph!}{(\aleph - u.)!} \prod_{i=1}^{\aleph} \prod_{j=1}^{k} p_{ij}^{X_{ij}} (1 - p_{ij})^{1-X_{ij}}. \tag{3.1}$$

The combinatorial term is included to account for the arbitrary ordering of individuals in the X matrix. In this case, the complete data likelihood in equation 3.1 is identical to the observed data likelihood used in equation 1.1. As described in section 1.1.1, models M_0, M_t, M_b, M_h, M_{tb}, M_{th}, M_{bh} and M_{tbh} are obtained by making assumptions about the capture probabilities p_{ij}.

A special case of model M_h is when the individual capture probabilities are assumed to be draws from a common probability distribution,

$$p_i \sim f(\theta_p), \quad i = 1, \ldots, \aleph.$$

The parameters $\boldsymbol{p}$ are no longer parameters, but random variables. We supplement the complete data by $\boldsymbol{p}$, which we include as latent variables using data augmentation. This means the CDL becomes

$$\mathcal{L}_C(\boldsymbol{\theta}_p, \aleph; \boldsymbol{X}, \boldsymbol{p}) \propto [\boldsymbol{X}, \boldsymbol{p}|\boldsymbol{\theta}_p, \aleph]$$
$$= [\boldsymbol{X}|\boldsymbol{p}, \aleph][\boldsymbol{p}|\boldsymbol{\theta}_p, \aleph].$$

The component $[\boldsymbol{X}|\boldsymbol{p}, \aleph]$ is specified in equation 3.1 with the constraint

$$p_{ij} = p_i, \quad i = 1, \ldots, \aleph, \ j = 1, \ldots, k.$$

The component $[\boldsymbol{p}|\boldsymbol{\theta}_p, \aleph]$ is the common distribution for $\boldsymbol{p}$,

$$[\boldsymbol{p}|\boldsymbol{\theta}_p, \aleph] = \prod_{i=1}^{\aleph}[p_i|\boldsymbol{\theta}_p],$$
$$[p_i|\boldsymbol{\theta}_p] = f(\boldsymbol{\theta}_p).$$

This model can also be represented using a directed acyclic graph (DAG). In a DAG, data and parameters are specified as nodes, with the directed arrows joining them specifying a relationship between the nodes. A single lined arrow denotes a stochastic conditional probability relationship and a double lined arrow denotes a deterministic relationship. The direction of the arrow determines the direction of relationship between the nodes. For example, in figure 3.1 the conditional probability of $\boldsymbol{X}$ depends on $\aleph$ and $\boldsymbol{p}$. DAGs are acyclic because starting at any one node we are unable to find a path that returns to that node. The DAG for model M_h is given in figure 3.1 with an example of fitting this model in section 7.1.
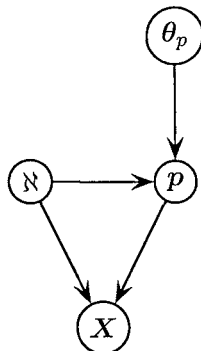
Figure 3.1: A DAG for a model M_h, where capture probabilities are from a common distribution with parameters θ_p.

3.1.1 Model M_h With Covariates

An extension to model M_h is when there are individual-specific covariates that explain the difference in the capture probabilities,

$$\text{logit}(p_i) = z\gamma,$$

where $\gamma = (\gamma_0, \gamma_1)'$ is a vector of parameters and

$$z = \begin{pmatrix} 1 & z_1 \\ 1 & z_2 \\ \vdots & \vdots \\ 1 & z_\aleph \end{pmatrix},$$

where z_i is the value of the covariate[1] for individual i. The values z_i, $i = 1, \ldots, u.$ are observed with z_i, $u. + 1, \ldots, \aleph$ unobserved. In order to fit this model Huggins (1989) conditions on $u.$ in order to obtain inference about $\aleph$, as described in section 1.1.2. This method imposes hidden assumptions

---

[1]This model can easily be extended to having more than one covariate.

about the covariate values for the unseen individuals. A better approach is to explicitly state the assumptions about the unknown covariate values through a model for z.

The complete data for this model consist of two components:

1. The complete set of capture histories for every individual ever available for capture, X.

2. The complete vector of covariate values z. As z_i, $u. + 1, \ldots, \aleph$ are unobserved we must include these using data augmentation. We model the covariate values z as being drawn from a common distribution with parameters θ_z,

$$z_i \sim g(\theta_z).$$

A separate variable $\mathcal{I}$ that specifies whether the covariate values were observed or missing is not included in the model because $\mathcal{I} = f(X)$. Therefore, the missingness mechanism is already accounted for through the capture histories X.

The CDL for this model is

$$\mathcal{L}_C(\theta_z, \gamma, \aleph; X, z) \propto [X, z | \theta_z, \gamma, \aleph]$$
$$= [X | z, \gamma, \aleph][z | \theta_z, \aleph].$$

The component $[X | z, \gamma, \aleph]$ is specified in equation 3.1 with

$$p_{ij} = \text{expit}(z\gamma), \quad i = 1, \ldots, \aleph, \ j = 1, \ldots, k.$$

The component $[z|\boldsymbol{\theta}_z, \aleph]$ is the common distribution for the covariate values,

$$[z|\boldsymbol{\theta}_z, \aleph] = \prod_{i=1}^{\aleph} [z_i|\boldsymbol{\theta}_z],$$

$$[z_i|\boldsymbol{\theta}_z] = g(\boldsymbol{\theta}_g).$$

The DAG for this model is given in figure 3.2 with an example of fitting this model in section 7.2.
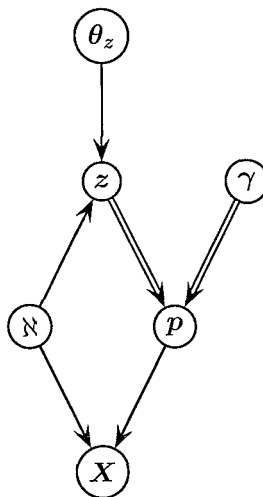
Figure 3.2: A DAG for a model M_h with individual specific covariates z. The covariates are modeled in terms of parameters $\boldsymbol{\theta}_z$.

In some circumstances z may be completely unknown, in which case we can model the entire z vector as a latent variable (Agresti 1994). For example, we could model the capture probabilities as a sample from a finite mixture distribution (Pledger 2000) with the unknown values of z_i denoting group assignment.

3.1.2 Splitting the capture history

In order to generalize this framework to open population studies, it is helpful to consider the capture history in two components: X_1, the capture history up to and including first capture, and X_2 the capture history subsequent to first capture. Note that X_1 and X_2 contain all the information about X.

For a closed population the likelihood contribution for each piece is,

$$[X_1|p, \aleph] = \frac{\aleph!}{(\aleph - u.)!} \prod_{i=1}^{\aleph} \prod_{j=1}^{\mathfrak{F}_i} p_{ij}^{X_{1ij}} (1 - p_{ij})^{1-X_{1ij}},$$

$$[X_2|X_1, p] = \prod_{i=1}^{u.} \prod_{j=\mathfrak{F}_i+1}^{k} p_{ij}^{X_{2ij}} (1 - p_{ij})^{1-X_{2ij}},$$

where $\mathfrak{F}_i$ is the sample of first capture and (for convenience) takes the value k if an individual was never caught.

This factorization is also convenient for closed population models with behavioural effects. For example, the likelihood for model M_b can be written as,

$$[X_1|p, \aleph] = \frac{\aleph!}{(\aleph - u.)!} \prod_{i=1}^{\aleph} \prod_{j=1}^{\mathfrak{F}_i} p^{X_{1ij}} (1 - p)^{1-X_{1ij}},$$

$$[X_2|X_1, c] = \prod_{i=1}^{u.} \prod_{j=\mathfrak{F}_i+1}^{k} c^{X_{2ij}} (1 - c)^{1-X_{2ij}}.$$

3.2 Open Population Models

The extension to open population models is achieved through the introduction of covariates, such as death, birth and movement that relax the closed population assumptions.

3.2.1   CJS

Introducing the interval censored time of death, d, defined in section 1.2.1 as a covariate using data augmentation provides an alternative description of the CJS model[2]. The traditional approach (in section 1.2.1) is to explicitly sum over all of the unknown covariate values to remove them from the model. This is done through the function χ_j specified in equation 1.2,

$$\chi_j = (1 - S_j) + S_j(1 - p_{j+1})\chi_{j+1},$$

where $\chi_k = 1$. Including the time of death using data augmentation means we no longer need to specify χ_j and explicitly integrate over the values. Instead the computational algorithm we choose, such as the EM algorithm or MCMC, integrates over the missing data as a part of the procedure.

In the same manner that we split the capture matrix X into X_1 and X_2 we also split the death matrix d into d_1, the information about mortality up to an including first capture and d_2 the information about mortality subsequent to first capture. As we are conditioning on first capture, we do not use d_1 for the CJS model.

To account for loss on capture we specify the vector ι. If $\iota_i = 0$ then individual i was lost on capture, with $\iota_i = 1$ otherwise. All individuals with $\iota_i = 0$ are known to have $d_{i \mathfrak{L}_i} = 1$ as they died at the time of last capture.

The complete data for the CJS consist of two components:

1. The capture histories subsequent to first capture X_2

2. The partially observed death covariate subsequent to first capture d_2

A separate indicator variable $\mathcal{I}$ that specifies whether the data in d_2 data is

---

[2]Note that this description is a special case of Dupuis (1995), who described the multistate model using data augmentation.

observed or missing is not included in the model because $\mathcal{I} = f(\boldsymbol{X_2})$. Even though we may only be interested in survival probabilities (and not capture probabilities) we must include a model for $\boldsymbol{X_2}$ because this component is modeling the non-ignorable missingness mechanism of $\boldsymbol{d_2}$.

The CDL for the CJS model is

$$
\begin{aligned}
\mathcal{L}_C(\boldsymbol{p}, \boldsymbol{S}; \boldsymbol{X_2}, \boldsymbol{d_2}, \boldsymbol{X_1}, \boldsymbol{\iota}) &\propto [\boldsymbol{X_2}, \boldsymbol{d_2}|\boldsymbol{p}, \boldsymbol{S}, \boldsymbol{X_1}, \boldsymbol{\iota}] \\
&= [\boldsymbol{X_2}|\boldsymbol{p}, \boldsymbol{d_2}, \boldsymbol{X_1}][\boldsymbol{d_2}|\boldsymbol{S}, \boldsymbol{X_1}, \boldsymbol{\iota}]
\end{aligned}
\tag{3.2}
$$

where $\boldsymbol{p} = (p_2, \ldots, p_k)$ and $\boldsymbol{S} = (S_1, \ldots, S_{k-1})$. We condition on $\boldsymbol{X_1}$ in order to condition on the first capture of each individual. We also condition on $\boldsymbol{\iota}$ because we do not want individuals lost on capture to contribute to mortality rates. If desired we could also model the loss on capture, however, in most cases this is a nuisance component of the model.

The survival component of equation 3.2 is further factorized as

$$
[\boldsymbol{d_2}|\boldsymbol{S}, \boldsymbol{X_1}, \boldsymbol{\iota}] = \prod_{i=1}^{N}[\boldsymbol{d_{2i:}}|\boldsymbol{S}, \boldsymbol{X_1}, \iota_i],
$$

$$
[\boldsymbol{d_{2i:}}|\boldsymbol{S}, \boldsymbol{X_1}, \iota_i] = MN(1, \boldsymbol{\xi}_i(\iota_i)),
$$

$$
\begin{aligned}
\boldsymbol{\xi}_i(\iota_i) = \Bigg(& (1 - S_{\mathfrak{F}_i}), S_{\mathfrak{F}_i}(1 - S_{\mathfrak{F}_i+1}), \ldots, \prod_{h=\mathfrak{F}_i}^{\mathfrak{L}_i-1} S_h(1 - S_{\mathfrak{L}_i})^{\iota_i}, \\
& \iota_i \prod_{h=\mathfrak{F}_i}^{\mathfrak{L}_i} S_h(1 - S_{\mathfrak{L}_i+1}), \ldots, \iota_i \prod_{h=\mathfrak{F}_i}^{k-2} S_h(1 - S_{k-1}), \iota_i \prod_{h=\mathfrak{F}_i}^{k-1} S_h \Bigg).
\end{aligned}
\tag{3.3}
$$

where $\boldsymbol{d_{2i:}}$ is the ith row of the matrix $\boldsymbol{d_2}$. Note that individuals lost on capture do not have a $(1 - S_j)$ term contributing to the likelihood. Throughout the rest of the model formulation in chapters 3 and 4 we ignore loss on capture for readability, noting that this formulation can incorporate loss on

capture unless otherwise specified.

The distribution for the capture matrix X_2 is largely the same as in model M_t for a closed population. The exception is that individuals are now only available for capture until the interval-censored time of death,

$$[X_2|d_2, p, X_1] \propto \prod_{i=1}^{u} \prod_{j=\mathfrak{F}_i+1}^{\mathfrak{D}_i} p_j^{X_{2ij}} (1-p_j)^{1-X_{2ij}}, \qquad (3.4)$$

where $\mathfrak{D}_i$ is the last sample in which individual i was alive.

The DAG for this model is specified in figure 3.3.
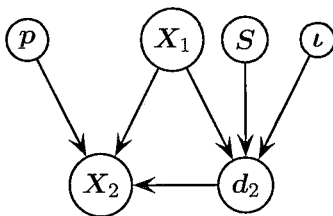
Figure 3.3: A DAG for the CJS model with loss on capture ι.

The factorization of the CDL in equation 3.2 shows a benefit of the data augmentation approach. We are able to separate the likelihood into a component for mortality and a component for capture conditional on death. This is not only the natural factorization of the data, but it also allows us to ignore the capture component when constructing the conditional likelihood for the mortality process. Another benefit of data augmentation is that we do not need to worry about how to average over the missing values of d_2 when constructing the conditional likelihood components.

3.2.2   The Jolly-Seber Model

Allowing for birth is an important extension to the CJS model. Jolly (1965) and Seber (1965) independently proposed to extend the CJS model through modeling the column sums of X_1,

$$u_j = \sum_{i=1}^{u.} X_{1ij}, \ j = 1, \dots, k$$

where u_j is the number of unmarked individuals observed in sample j.

The complete data for this problem is identical to the complete data for the CJS except now we also include the capture histories up to and including first capture, X_1. The CDL for the JS model is,

$$\mathcal{L}_C(p, S, U; X_2, d_2, X_1) \propto [X_1, X_2, d_2 | p, S, U]$$
$$\propto \underbrace{[X_1 | p, U]}_{\text{First Captures}} \underbrace{[X_2 | p, d_2, X_1][d_2 | S, X_1]}_{\text{CJS component}} \quad (3.5)$$

where $U = (U_1, \dots, U_k)$, with U_j being the total number of unmarked individuals in the population available for capture at sampling occasion j.

The component for first captures from equation 3.5 is

$$[X_1 | p, U] = \prod_{j=1}^{k} [u_j | p_j, U_j] \quad (3.6)$$

where

$$[u_j | p_j, U_j] = Bin(U_j, p_j), \ j = 1, \dots, k.$$

Even though both the index U_j and the probability p_j are unknown, the CJS component provides information that makes p_j identifiable for $j = 2, \dots, k - 1$, which in turn makes U_j, $j = 2, \dots, t - 1$ identifiable.

The DAG for this model is given in figure (figure 3.4).
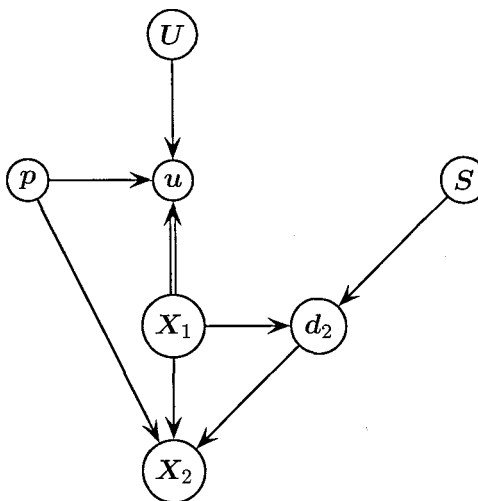
Figure 3.4: A DAG for the JS model, where $\boldsymbol{u} = (u_1, \ldots, u_k)$.

Values of demographic interest, such as N_j, the total population size in sample j, can be obtained from deterministic functions of the parameters,

$$N_j = U_j + M_j$$
$$M_j = \sum_{h=1}^{j-1} \left\{ u_h - \sum_{i=1}^{u.} d_{ih} \right\} \quad j = 1, \ldots, k \quad (3.7)$$

where M_j is the number of marked individuals in the population immediately before sample j. The DAG for this extension is in figure 3.5.

As mentioned in section 1.2.2, the values U_j are not particularly interesting to a biologist. They include both the sampling intensity as well as aspects of demographic interest, such as birth. The biologist has very little interest in further modeling of the U_j random variables. Therefore, models are required that parameterize in terms of more biologically interesting birth parameters. One such extension to the JS model is examined in section 4.3.1.
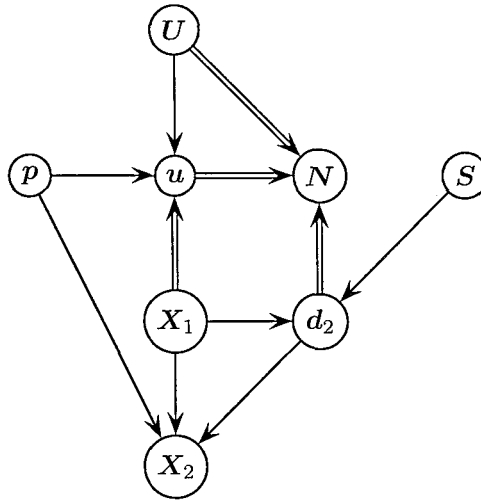
Figure 3.5: A DAG for the JS model including the population size in each sample $N = (N_1, \ldots, N_k)$.

3.2.3 CMAS

An alternative to including birth through modeling $u = (u_1, \ldots, u_k)$ was considered by Crosbie and Manly (1985) and Schwarz and Arnason (1996). They started with a closed population model which they relaxed in two ways:

1. They relaxed the assumption that all individuals remain available for capture until the end of the study. We express this through the introduction of a death covariate d as in the CJS.

2. They relaxed the assumption that all individuals are available for capture at the beginning of the study. We express this through the introduction of a birth covariate b. The birth covariate has value $b_{ij} = 1$ if individual i was born between sample j and $j + 1$ and $b_{ij} = 0$ otherwise. The value $b_{i0} = 1$ is taken to mean that an individual was

alive at the beginning of the study[3]. As with the death covariate the covariate b is partially known. We know that an individual was not born after the time in which we first observed it (the sample of first capture $\mathfrak{F}_i$ so the values b_{ij} for $j = \mathfrak{F}_i, \ldots, k - 1$ are known to be 0. The remaining b_{ij} values are missing (subject to the constraint that $\sum_{j=0}^{k-1} b_{ij} = 1$).

The complete data for the CMAS model consists of three components:

1. The capture histories $X = (X_1, X_2)$

2. The death covariate $d = (d_1, d_2)$

3. The birth covariate b

Separate variables $\mathcal{I}_b$ and $\mathcal{I}_d$ that specify whether the data in b and d are observed or missing are not included in the model because $\mathcal{I}_b = f(X)$ and $\mathcal{I}_d = f(X)$. This means that even though we may only be interested in birth rates and survival probabilities we must also include a model for X because this component is modeling the non-ignorable missingness mechanisms of b and d.

As with the closed population models we include $\aleph$, the total number of individuals ever available for capture, as a parameter in the model as we do not know the dimension of X.

The CDL for the CMAS model is,

$$
\begin{aligned}
\mathcal{L}_C(S, p, \beta, \aleph; X, b, d) &\propto [X, d, b | p, S, \beta, \aleph] \\
&= [X | p, d, b, \aleph][d | S, b, \aleph][b | \beta, \aleph],
\end{aligned}
\tag{3.8}
$$

where $\beta = (\beta_0, \ldots, \beta_{k-1})$ are the birth parameters with the constraint $\sum_j \beta_j = 1$. This factorization separates the joint likelihood into the natural

---

[3]This introduces the restriction that $\sum_{j=0}^{k-1} b_{ij} = 1$.

ordering of birth, then death conditional on birth, then capture conditional on birth and death.

The capture component is

$$[\boldsymbol{X}|\boldsymbol{p}, \boldsymbol{d}, \boldsymbol{b}, \aleph] = \prod_{i=1}^{\aleph} \prod_{j=\mathfrak{B}_i+1}^{\mathfrak{D}_i} p_j^{X_{ij}}(1 - p_j)^{1-X_{ij}}. \qquad (3.9)$$

This is identical to the representation of capture under model M_t except that individuals are only available for capture from the time of birth until the time of death.

The mortality component is

$$[\boldsymbol{d}|\boldsymbol{S}, \boldsymbol{b}, \aleph] = \prod_{i=1}^{\aleph}[\boldsymbol{d}_{i:}|\boldsymbol{b}, \boldsymbol{S}],$$

$$[\boldsymbol{d}_{i:}|\boldsymbol{b}, \boldsymbol{S}] = MN(1, \boldsymbol{\xi}_i),$$

$$\boldsymbol{\xi}_i = \left((1 - S_{\mathfrak{B}_i+1}), S_{\mathfrak{B}_i+1}(1 - S_{\mathfrak{B}_i+2}), \ldots, \prod_{h=\mathfrak{B}_i+1}^{k-2} S_h(1 - S_{k-1}), \prod_{h=\mathfrak{B}_i+1}^{k-1} S_h\right).$$

$$(3.10)$$

Note that we assume that an individual cannot die in the same period that it was born. Incorporating additional assumptions or different data can relax this assumption. For example, Crosbie and Manly (1985), Schwarz et al. (1993a) and Schwarz and Arnason (1996) introduce assumptions to allow individuals to die before they are available for capture. Here, we exclude such individuals from the definition of $\aleph$.

The birth component is

$$[\boldsymbol{b}|\boldsymbol{\beta}, \aleph] = \prod_{i=1}^{\aleph}[\boldsymbol{b}_{i:}|\boldsymbol{\beta}],$$

$$[\boldsymbol{b}_{i:}|\boldsymbol{\beta}] = MN(1, \boldsymbol{\beta}). \qquad (3.11)$$

This factorization allows us to focus on writing sensible and interesting models for birth and survival without worrying about (i) the model for the capture process or (ii) how to average across the missing b and d values.

The DAG for this model is in figure 3.6.
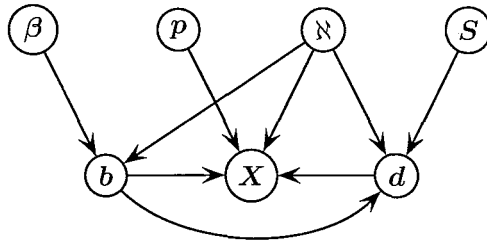
Figure 3.6: A DAG for the CMAS model.

As with the Jolly-Seber model we are able to obtain the population size in each sample deterministically from b, d_1 and d_2,

$$N_j = \sum_{i=1}^{\aleph} \left(\sum_{h=0}^{j-1} b_{ih} - \sum_{h=1}^{j-1} d_{ih} \right), \; j = 1, \ldots, k. \qquad (3.12)$$

The DAG for this extension is in figure 3.7.
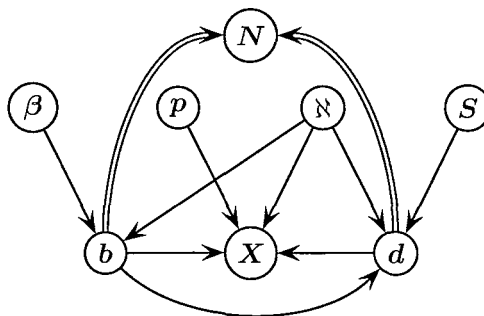
Figure 3.7: A DAG for the CMAS model including population size in each sample $\boldsymbol{N} = (N_1, \ldots, N_k)$.

An alternate factorization of the CDL in equation 3.8 is

$$\mathcal{L}_C(S, p, \beta, \aleph; X, b, d)$$

$$\propto \underbrace{[X_1|p, b, d_1, \aleph][d_1|S, b, \aleph][b|\beta, \aleph]}_{\text{First Capture}} \underbrace{[X_2|p, d_2, X_1][d_2|S, X_1]}_{\text{CJS component}} \quad (3.13)$$

This factorization specifies the CMAS model as a CJS model with additional components that model first capture through introduction of the birth covariate.

As mentioned in section 1.2.3, the CMAS model has several problems. As with the JS model one of those is that the birth parameters are not natural to a biologist, as they do not separate aspects of study design and the demographic process of interest. An extension of the CMAS to more natural birth parameters is examined in section 4.3.2.

Chapter 4

Extending the Hierarchy

In this chapter we extend the framework developed in the previous chapter. In particular we look on hierarchical extensions motivated by additional data, covariates or assumptions, with a focus on models that explore demographic relationships of interest.

4.1 Random effects models

In many situations we may wish to assume that parameters in the model are themselves drawn from a common probability distribution function. We refer to these as random effects models. There are three popular motivations for random effects models: (i) parsimony, (ii) prediction (iii) specifying relationship between parameters.

Consider a CJS model where the focus is on estimating survival probabilities with no potential covariates. Standard approaches give two alternatives for the survival probabilities (i) time specific and (ii) constant through time. Suppose the second model does not fit the data well, however, we feel that the first model is too general. Including a random effect achieves parsimony through assuming that the survival probabilities are time specific, but they

are sampled from some common probability distribution function. This effectively reduces the number of survival parameters from $k - 1$ to m, where

$$1 \leq m \leq k - 1.$$

Another example is where survival is assumed to depend on time-varying covariates. The standard approach of Lebreton et al. (1992) treats survival as a deterministic function of the covariate. However, even if the relationship is statistically significant it may not provide an adequate fit of the data. Random effects allow us to include additional variability in the regression relationship to account for the over-dispersion. More details about including covariates are given in section 4.2.

To show how random effects are used for prediction consider a CMAS model where the focus is in predicting population size forward δ years. As before, suppose that a constant survival rate is unrealistic and does not fit the data well. However, it is impossible to predict time specific survival probabilities δ years into the future without additional assumptions. One such assumption is that survival probabilities are sampled from a common probability distribution with location parameter μ_S and scale parameter τ_S. Including the survival probabilities as latent variables using data augmentation means that the CDL for the CJS model becomes

$$\mathcal{L}_C(\mu_S, \tau_S, \boldsymbol{p}; \boldsymbol{X}_2, \boldsymbol{X}_1, \boldsymbol{d}_2, \boldsymbol{S}) \propto \underbrace{[\boldsymbol{X}_2|\boldsymbol{p}, \boldsymbol{d}_2, \boldsymbol{X}_1][\boldsymbol{d}_2|\boldsymbol{S}, \boldsymbol{X}_1]}_{\text{CJS Component}} \underbrace{[\boldsymbol{S}|\mu_S, \tau_S]}_{\text{Random effect}} .$$

The terms $[\boldsymbol{X}_2|\boldsymbol{p}, \boldsymbol{d}_2, \boldsymbol{X}_1]$ and $[\boldsymbol{d}_2|\boldsymbol{S}, \boldsymbol{X}_1]$ are the usual CJS components specified in equations 3.4 and 3.3 with the term $[\boldsymbol{S}|\mu_S, \tau_S]$ specifying the random effect for $\boldsymbol{S}$. A common random effects model for survival is that the logit

of the survival probabilities is distributed as a normal distribution,

$$[\boldsymbol{S}|\mu_S, \tau_S] \propto \prod_{j=1}^{k-1} [S_j|\mu_S, \tau_S]$$

(4.1)

$$[S_j|\mu_S, \tau_S] = \text{logit} N(\mu_S, \tau_S)$$

The DAG for this model is given in figure 4.1. As usual, we choose our computational algorithm so that the latent variables $\boldsymbol{S}$ are integrated out of the model as part of the model fitting process.
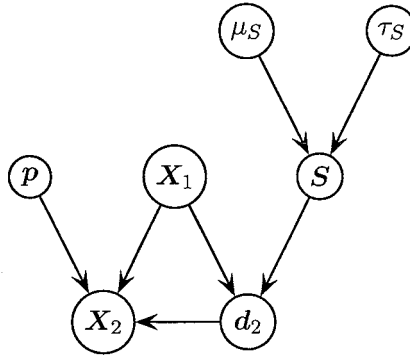
Figure 4.1: A DAG for the CJS model where the logit of the survival probabilities are assumed to be samples from a $N(\mu_S, \tau_S)$ random effects distribution.

More complex random effects distributions can be specified that often include relationships between parameters. Johnson and Hoeting (2003) specified the logit of the survival probabilities in a CJS model to have autoregressive errors. Link and Barker (2005) specified a multivariate random effects distribution for the per-capita birth rate index and the logit of survival in the model from section 1.2.4. They assumed that the survival probabilities and per-capita birth rate indices for each given sampling period

were correlated,

$$
\begin{pmatrix} \operatorname{logit}(S_j) \\ \log(\eta'_j) \end{pmatrix} \sim MVN\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right), \ j = 1, \ldots, k-1,
$$

where

$$
\boldsymbol{\mu} = \begin{pmatrix} \mu_S \\ \mu_\eta \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_S^2 & \rho\sigma_S\sigma_\eta \\ \rho\sigma_S\sigma_\eta & \sigma_\eta^2 \end{pmatrix}.
$$

If $\rho > 0$ then this suggests that survival is relatively high/low when fecundity is relatively high/low. This suggests that there will be "good" years in which individuals not only have higher survival probabilities but are also able to put effort into reproduction. Likewise, there will also be "bad" years when survival probabilities and fecundity are low.

4.2   Including Covariates

Often capture-recapture data is collected with additional covariates, z, that are assumed to affect the parameters in the model. The covariates z can either be fully observed or partially observed.

4.2.1   Fully Observed Covariates

Consider a CJS model where both the capture and survival probabilities are assumed to depend on some fully observed covariate z that varies through time but is constant between individuals. Lebreton et al. (1992) treats the parameter as a deterministic function of the covariate, for example,

$$
\begin{aligned}
\operatorname{logit}(S_j) &= \gamma_0 + \gamma_1 z_j, \ j = 1, \ldots, k-1, \\
\operatorname{logit}(p_j) &= \alpha_0 + \alpha_1 z_j, \ j = 2, \ldots, k.
\end{aligned}
\tag{4.2}
$$

The CDL for this model is

$$\mathcal{L}_C(\gamma, \alpha; X_2, X_1, d_2, z) \propto [X_2|\alpha, z, d_2, X_1][d_2|\gamma, z, X_1]$$

where $\gamma = (\gamma_0, \gamma_1)$ and $\alpha = (\alpha_0, \alpha_1)$. The component $[d_2|\gamma, z, X_1]$ is specified in equation 3.3 with

$$S_j = \text{expit}(\gamma_0 + \gamma_1 z_j), \ j = 1, \ldots, k-1.$$

The component $[X_2|\alpha, z, d_2, X_1]$ is specified in equation 3.4 with

$$p_j = \text{expit}(\alpha_0 + \alpha_1 z_j), \ j = 2, \ldots, k.$$

The DAG for this model is in figure 4.2.
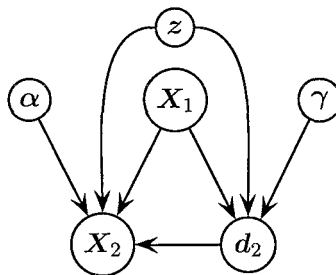
Figure 4.2: A DAG for the CJS model where the covariate z fully explains capture and survival probability.

The model specified in equation 4.2 assumes that all variability in S_j and p_j is explained by the covariate. As described in section 4.1, this is a problem if either the capture or survival probabilities are over-dispersed, as even if α_1 or γ_1 is significant, the model may not fit the data adequately. A more appropriate model accounts for any over-dispersion in p_j and S_j

by using data augmentation to include the deviations ϵ_{pj} and ϵ_{Sj}, which themselves are drawn from a specified distribution. The CDL for this model is

$$
\mathcal{L}_C(\gamma, \alpha, \tau_S, \tau_p; X_2, X_1, d_2, \epsilon_S, \epsilon_p, z)
$$
$$
\propto [X_2|\alpha, \epsilon_p, z, d_2, X_1][d_2|\gamma, \epsilon_S, z, X_1][\epsilon_p|\tau_p][\epsilon_S|\tau_S] \tag{4.3}
$$

where $\epsilon_S = (\epsilon_{S1}, \ldots, \epsilon_{Sk-1})$ and $\epsilon_p = (\epsilon_{p2}, \ldots, \epsilon_{pk})$. The component $[X_2|\alpha, \epsilon_p, z, d_2, X_1]$ is specified in 3.4 with

$$
p_j = \text{expit}(\alpha_0 + \alpha_1 z_j + \epsilon_{pj}), \ j = 2, \ldots, k.
$$

The component $[d_2|S, X_1]$ is specified in equation 3.3 with

$$
S_j = \text{expit}(\gamma_0 + \gamma_1 z_j + \epsilon_{Sj}), \ j = 1, \ldots, k-1.
$$

One set of possible random effects distributions is

$$
[\epsilon_S|\tau_S] \propto \prod_{j=1}^{k-1} [\epsilon_{Sj}|\tau_S],
$$
$$
[\epsilon_{Sj}|\tau_S] = N(0, \tau_S),
$$
$$
[\epsilon_p|\tau_p] \propto \prod_{j=2}^{k} [\epsilon_{pj}|\tau_p], \tag{4.4}
$$
$$
[\epsilon_{pj}|\tau_p] = N(0, \tau_p).
$$

The DAG for this model is given in figure 4.3.

4.2.2 Partially Observed Covariates

Covariates may not be fully observed for a number of reasons. We focus on situations where the covariate is individual-specific and time-varying.
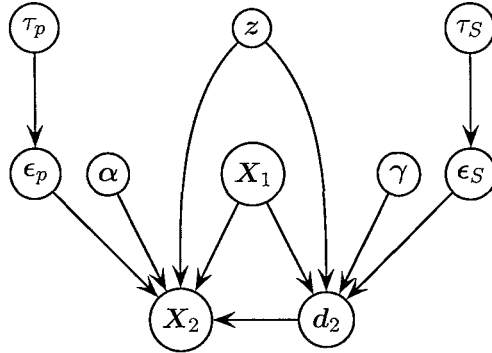
Figure 4.3: A DAG for the CJS model capture and survival probabilities are modeled as a random effect which depends on the fully observed covariate z.

This means that we are only able to observe the covariate value when the individual is caught, with the value missing otherwise.

The missingness mechanism for the covariates is not ignorable because whether the covariate is observed or not depends on the value of the missing covariate through the survival rates, birth rates and capture probabilities. However, as before, we do not need to separately include an indicator function $\mathcal{I}$ to model the missingness mechanism of z because $\mathcal{I} = f(\boldsymbol{X})$.

Even though we do not need to include $\mathcal{I}$, a model for the missing values of z is required. In most, if not all situations, we model both the observed and missing values of the covariate. To assist in the model formulation, we separate the covariate values z into a component z_1 that has information on z up to and including first capture for each individual and z_2 that has information on z after first capture. To include an individual-specific covariate z into a CJS model, we must (i) include the missing values of z into the model using data augmentation and (ii) specify a model for z_2 in terms of parameters $\boldsymbol{\Psi}$. A consequence of z being individual-specific is

that any parameters that depends on it, will also be individual-specific. To ensure identifiability, any random effect can only vary across time, with individual differences fully explained by z. The CDL for a CJS model with a time-varying individual-specific covariate z that affects capture and survival probabilities is

$$
\begin{aligned}
&\mathcal{L}_C(\gamma, \alpha, \tau_S, \tau_p, \Psi; X_2, X_1, d_2, \epsilon_S, \epsilon_p, z) \\
&\propto [X_2|\alpha, \epsilon_p, z, d_2, X_1][d_2|\gamma, \epsilon_S, z, X_1][\epsilon_p|\tau_p][\epsilon_S|\tau_S][z_2|\Psi, z_1]
\end{aligned}
\tag{4.5}
$$

The component $[X_2|\alpha, \epsilon_p, z, d_2, X_1]$ is the individual-specific analogue of equation 3.4,

$$
[X_2|d_2, p, X_1] \propto \prod_{i=1}^{u} \prod_{j=\mathfrak{F}_i+1}^{\mathfrak{D}_i} p_{ij}^{X_{2ij}} (1 - p_{ij})^{1-X_{2ij}},
\tag{4.6}
$$

with

$$
p_{ij} = \text{expit}(\alpha_0 + \alpha_1 z_{ij} + \epsilon_{pj}), \ \ i = 1, \ldots, u., \ \ j = 2, \ldots, k.
$$

The component $[d_2|\gamma, \epsilon_S, z, X_1]$ is the individual-specific analogue to equation 3.3,

$$
[d_2|S, X_1] = \prod_{i=1}^{\aleph} [d_{2i:}|S, X_1],
$$

$$
[d_{2i:}|S, X_1] = MN(1, \boldsymbol{\xi}_i),
$$

$$
\boldsymbol{\xi}_i = \left((1 - S_{i\mathfrak{F}_i}), S_{i\mathfrak{F}_i}(1 - S_{i\mathfrak{F}_i+1}), \ldots, \prod_{h=\mathfrak{F}_i}^{k-2} S_{ih}(1 - S_{ik-1}), \prod_{h=\mathfrak{F}_i}^{k-1} S_{ih} \right),
\tag{4.7}
$$

with

$$
S_{ij} = \text{expit}(\gamma_0 + \gamma_1 z_{ij} + \epsilon_{Sj}), \ \ i = 1, \ldots, u., \ \ j = 1, \ldots, k - 1.
$$

The components $[\epsilon_p|\tau_p]$ and $[\epsilon_S|\tau_S]$ are specified in equation 4.4. The additional component $[z_2|\Psi, z_1]$ models the unknown values of the covariate. In the next section we show how various models for the covariate z_2 give rise to different capture-recapture models. The DAG for this model is in figure 4.4.

The difference between the CDLs in equations 4.5 and 4.3 and the DAGs in figures 4.4 and 4.3 show there is very little difference between including fully observed and partially observed covariates. The basic model structure is the same with the only differences being that the parameters become individual-specific and an additional layer is required to model the missing covariate values.
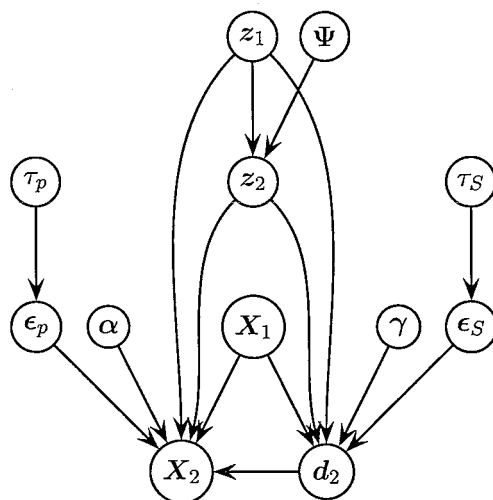
Figure 4.4: A DAG for the CJS model where capture and survival probabilities are modeled as a random effect which depends on the covariate z. The covariate values after first capture z_2 are only partially observed and modeled in terms of parameters Ψ.

Categorical Covariates

In some situations the partially observed covariate z is categorical. The multistate model is obtained when the model for categorical z_2 is assumed to be a first order Markov chain (Schwarz et al. 1993b, Dupuis 1995),

$$[z_2|\Psi, z_1] \propto \prod_{i=1}^{u.} \prod_{j=\mathfrak{F}_i+1}^{k} [z_{ij}|\Psi, z_{ij-1}],$$

where

$$[z_{ij+1}|\Psi, z_{ij}] \propto \prod_{h=1}^{n_z} \prod_{l=1}^{n_z} \Psi_{jkl}^{I(z_{ij+1}=l)I(z_{ij}=h)}, \tag{4.8}$$

where $\Psi_{jkl} = Pr(z_{ij+1} = l|z_{ij} = h)$, n_z is the number of possible "states" and $\sum_l \Psi_{jhl} = 1$, $\forall j, h$.

Dupuis (1995) also specified the multistate model in terms of the CDL and fitted the model using MCMC methods. Dupuis (1995) did not include the covariate d but treated death as an additional absorbing state of the covariate z. Although this changes the latent structure of the covariates z and d, the model is the same.

We are not forced into assuming that z_2 is a first order Markov chain. We can extend the model to a second or higher order Markov chain (Brownie et al. 1993), or we can model z_2 using any other model that is appropriate.

If we include first captures then we also need to specify the model for the covariate values from the time of birth until first capture. If we assume the covariate values are a first order Markov chain we need to model the covariate values in the first period after birth separately from subsequent periods,

$$[z_1|\pi, \Psi, b, \aleph] = [z_{i\mathfrak{B}_i+1}|\pi, \aleph] \prod_{j=\mathfrak{B}_i+2}^{\mathfrak{F}_i} [z_{ij}|\Psi, z_{ij-1}]$$

where π are parameters used to describe initial allocation into a state. A

possible model for the covariate in the period after birth is,

$$[z_{i\mathfrak{B}_i+1}|\pi, \aleph] = \prod_{h=1}^{n_z} \pi_{\mathfrak{B}_i+1h}^{I(z_{i\mathfrak{B}_i+1}=h)}, \ i = 1, \ldots, \aleph$$

where $\pi_{jh} = Pr(z_{ij} = h|b_{ij-1} = 1)$ with the constraint $\sum_h \pi_{jh} = 1, \ \forall j$ and $I(\cdot)$ is the indicator function.

The component $[z_{ij}|\Psi, z_{ij-1}], \ j = \mathfrak{B}_i + 2, \ldots, \mathfrak{F}_i$ is assumed to be the same as the model specified in equation 4.8.

Availability for capture

Availability is a commonly used partially observed categorical covariate for describing movement in and out of the study population (Jolly 1965, Seber 1965). We can express availability of capture through the covariate z, with $z_{ij} = 1$ if individual i is available for capture in sample j and $z_{ij} = 0$ otherwise.

Three common assumptions about movement are first order Markovian emigration, random emigration and permanent emigration (Barker 1997). First order Markovian emigration is when movement between the time of sample j and $j+1$ depends only on the covariate for individual i at time of sample j. The transition matrix Ψ_j for Markovian emigration is,

$$\Psi_j = \begin{bmatrix} 1 - F_j' & F_j' \\ 1 - F_j & F_j \end{bmatrix}, \ j = 1, \ldots, k-1$$

where

$$F_j = Pr(z_{ij+1} = 1|z_{ij} = 1)$$
$$F_j' = Pr(z_{ij+1} = 1|z_{ij} = 0)$$

Random emigration and permanent emigration are special cases of Markovian emigration. Under random emigration the movement probability does

not depend on the previous value of the covariate, that is, $F_j' = F_j$. Under permanent emigration, once an individual becomes unavailable for capture, it can never be available again, that is $F_j' = 0$.

Permanent Emigration

Permanent emigration is assumed in most models that include first captures, (Jolly 1965, Seber 1965, Schwarz and Arnason 1996, Link and Barker 2005). Averaging across the various combinations for b, d and z for each capture history reveals that the parameters S_j and F_j are confounded and there is not enough information to separately estimate birth, movement and survival parameters prior to the first capture. The confounding of S_j and F_j shows that we are unable to distinguish between those individual who emigrate (permanently) and those who die.

The standard approach is to not include the covariate z but combine (i) birth and immigration and (ii) death and emigration. This changes the meaning of both b and d. The value $b_{ij} = 1$ means that individual i was either born or immigrated into the population between sample j and $j+1$ with $b_{ij} = 0$ otherwise. The value $d_{ij} = 1$ means that individual i either died or emigrated out the population between sample j and $j+1$ with $d_{ij} = 0$ otherwise. As a result, the meanings of the parameters change to reflect the changes in the meaning of b and d. In general we parameterize in terms of $\phi_j = S_j F_j$, where $1 - \phi_j$ is the probability that an individual either died or emigrated out of the population between sample j and $j+1$.

First order Markovian Emigration

Without strong assumptions, first order Markovian emigration is not identifiable unless more complex study designs are used, such as the robust design, or models incorporating different types of re-encounter data. Even

with these designs, additional constraints about the time-specific covariate parameters, F_j and F_j' are required. One sufficient constraint is that movement parameters are fixed through time, $F_j = F$ and $F_j' = F'$ (Barker et al. 2004). A further problem is that there is not enough information to separate the birth parameters from the allocation probabilities π_{j1}. Possible solutions include gathering different information that can be used to separate the two parameters or assuming that all individuals are born unavailable (or available) for capture, $\pi_{j1} = 0$ (or $\pi_{j1} = 1$). The assumption $\pi_{j1} = 1$ is algebraically equivalent to combining individuals being born available for capture with immigrants becoming available for capture for the first time.

Random Emigration

Under random emigration F_j is confounded with p_{j+1} (Burnham 1991). The standard approach is to not include z but instead consider the identifiable parameter $p_{j+1}' = F_j p_{j+1}$, the joint probability of being available for capture and caught in sample $j + 1$. Including the first captures means that π_{j1} is also confounded with p_{j+1}. As with Markovian emigration, one possible solution is to assume all individuals are born unavailable (or available) for capture with $\pi_{j1} = 0$ (or $\pi_{j1} = 1$). Another possible solution is to assume that initial allocations are the same as subsequent movement probabilities, that is, $\pi_{j1} = F_j$ (Barker 1997). Under this assumption the algebraic structure for the model is identical to that of permanent emigration described above, with the only difference being the meaning of the parameters in the model.

Continuous covariates

The inclusion of individual-specific time-varying continuous covariates is very difficult using standard models because we must explicitly integrate out the missing values of continuous z, as mentioned in section 1.2.1. However, the CDL for continuous z is identical to the CDL for categorical z. The only differences in the model specification are in (i) the model specified for z and (ii) describing how the parameters depend on z. Bonner and Schwarz (2006) assumed that both survival and capture probabilities of meadow voles deterministically depended on body weight z. They used the CDL

$$\mathcal{L}_C(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\Psi}; \boldsymbol{X}_2, \boldsymbol{X}_1, \boldsymbol{d}_2, \boldsymbol{z})$$

$$\propto [\boldsymbol{X}_2 | \boldsymbol{\alpha}, \boldsymbol{d}_2, \boldsymbol{z}, \boldsymbol{X}_1][\boldsymbol{d}_2 | \boldsymbol{\gamma}, \boldsymbol{z}, \boldsymbol{X}_1][\boldsymbol{z}_2 | \boldsymbol{\Psi}, \boldsymbol{z}_1]$$

The capture component $[\boldsymbol{X}_2 | \boldsymbol{\alpha}, \boldsymbol{d}_2, \boldsymbol{z}, \boldsymbol{X}_1]$ is specified in equation 3.4 with

$$p_{ij} = \text{expit}(\alpha_0 + \alpha_1 z_{ij}), \quad j = 2, \ldots, k.$$

The survival component $[\boldsymbol{d}_2 | \boldsymbol{\gamma}, \boldsymbol{z}, \boldsymbol{X}_1]$ is specified in equation 3.3 with

$$S_{ij} = \text{expit}(\gamma_0 + \gamma_1 z_{ij}), \quad j = 1, \ldots, k - 1.$$

Bonner and Schwarz (2006) assumed that individual weight changed through time according to a Wiener process with drift,

$$[\boldsymbol{z}_2 | \boldsymbol{\Psi}, \boldsymbol{z}_1] \propto \prod_{i=1}^{u.} \prod_{j=\mathfrak{F}_i+1}^{k} [z_{ij} | \boldsymbol{\Psi}, z_{ij-1}]$$

$$[z_{ij+1} | \boldsymbol{\Psi}, z_{ij}] = N(z_{ij} + \Psi_{1j}, \Psi_2)$$

The parameter Ψ_{1j} is the mean weight change in the population between sample j and $j + 1$.

In section 6.4 we discuss fitting this model using WinBUGS and in chapter 9 we fit a capture-recapture model for rainbow trout where individual body length is a continuous covariate.

Uncertain Covariate Values

Often the covariates that we observe are uncertain. We do not wish to use the corrupted observed covariate z in our models, so we include the true covariate value z' into the model using data augmentation and specify:

1. A model describing the corruption of z from the true value z'.

2. A model for the true covariate values z'.

A necessary component of the model is either having the data or assumptions that make the modeling of the corruption process is identifiable. The multi-event model of Pradel (2005) is obtained when the covariate value is categorical.

4.3   Using latent variables

One of the advantages of the data augmentation approach is that the latent variables are available in the model to be used. They can be used in any way so long as the rules of conditional probability are obeyed. In this section we develop extensions to the capture-recapture model where the latent variables are used (i) to define more natural birth parameters and (ii) as a covariate.

4.3.1   Extension of JS Model to Include Birth Parameters

We can make use of latent variables to re-parameterize the JS model from section 3.2.2 to include biologically interesting birth parameters. The key step is recognizing that

$$U_j = U'_{j-1} + B_{j-1}, \quad j = 2, \ldots, k \tag{4.9}$$

where U'_{j-1} is the number of individuals that were unmarked immediately after sample $j-1$ and survived until j, and B_{j-1} is the number of individuals that were born between samples $j-1$ and j. The CDL for the standard Jolly-Seber model in equation 3.5 is extended to include hierarchical models for U'_j and B_j, factored in the natural order of events in time,

$$\mathcal{L}_C(p, S, \eta; X_2, d_2, X_1, U', B)$$

$$\propto [u_1|p_1, U_1] \prod_{j=2}^{k} [u_j|p_j, U'_{j-1}, B_{j-1}] \prod_{j=1}^{k-1} [U'_j|S_j, u_{1:j}, d_{2:(1:j-1)}, B_{j-1}, U'_{j-1}] \times$$

$$\prod_{j=1}^{k-1} [B_j|\eta_j, u_{1:j-1}, d_{2:(1:j-1)}, B_{j-1}, U'_{j-1}][X_2|p, d_2, X_1][d_2|S, X_1]$$

$$\tag{4.10}$$

where $U' = (U'_1, \ldots, U'_{k-1})$, $B = (B_1, \ldots, B_{k-1})$ and $\eta = (\eta_1, \ldots, \eta_{k-1})$.

The components $[u_1|p_1, U_1]$ and $[u_j|p_j, U'_{j-1}, B_{j-1}]$ are specified in equation 3.6,

$$[u_1|p_1, U_1] = Bin(U_1, p_1)$$

$$[u_j|p_j, U'_{j-1}, B_{j-1}] = Bin(U_j, p_j), \quad j = 2, \ldots, k$$

with U_j obtained from equation 4.9 for $j = 2, \ldots, k$.

We assume that all unmarked individuals in the population immediately

after the jth sampling period survive until sample $j+1$ with probability S_j. This gives,

$$[U'_j | S_j, \boldsymbol{u}_{1:j}, \boldsymbol{d}_{2:(1:j-1)}, B_{j-1}, U'_{j-1}] = Bin(N_j - M_j - u_j, S_j), \ j = 1, \ldots, k-1$$

(4.11)

where $B_0 = U_1$, $U'_0 = 0$ and

$$N_j = U_j + M_j$$
$$M_j = \sum_{h=1}^{j-1} \left\{ u_h - \sum_{i=1}^{u_\cdot} d_{2ih} \right\} \quad j = 1, \ldots, k. \qquad (4.12)$$

The value N_j is the number of individuals alive at time of sample j and M_j is the number of marked individuals in the population immediately before the jth sample.

We choose to model the number of births into the population between sample j and $j+1$ as a Poisson random variable. This gives,

$$[B_j | \eta_j, \boldsymbol{u}_{1:j-1}, \boldsymbol{d}_{2:(1:j-1)}, B_{j-1}, U'_{j-1}] = Pois(\eta_j N_j), \ j = 1, \ldots, k-1,$$

where
$$\eta_j = \frac{E(B_j)}{N_j}, \ j = 1, \ldots, k-1$$

is the per-capita birth rate with N_j given in equation 4.12. We choose to parameterize in terms of per-capita birth rates as these are natural birth parameters, The value η_j is the expected number of new individuals born between samples j and $j+1$ for every individual alive in sample j. They do not depend on any aspect of the study design[1] and can be used to model the changes in population dynamics. The per-capita birth rates η are similar

---

[1]Except for determining the relevant population.

to the per-capita birth rate indices η' considered by Pradel (1996) and Link and Barker (2005),

$$\eta'_j = \frac{E(B_j)}{E(N_j)}, \ j = 1, \ldots, k-1.$$

We refer to η'_j as an index because we use the expectation of N_j instead of its actual value.

The components $[\boldsymbol{X}_2|\boldsymbol{p}, \boldsymbol{d}_2, \boldsymbol{X}_1]$ and $[\boldsymbol{d}_2|\boldsymbol{S}, \boldsymbol{X}_1]$ are the CJS components for capture and mortality conditional on first capture and are specified in equations 3.4 and 3.3.

Using the Jolly-Seber model, the total number of individuals available for capture $\aleph$ is not a parameter in the model. However, it can be obtained using the deterministic relationship

$$\aleph = U_1 + \sum_{j=1}^{k-1} B_j.$$

The WinBUGS code used to fit this birth parameterization is given in section 6.6.

4.3.2   Extension of CMAS Model to Include Per-Capita Birth Rates

As with the Jolly-Seber model it is possible to re-parameterize the birth component in the CMAS model to include biologically interesting parameters such as the per-capita birth rate $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_{k-2})$. As the per-capita birth rate is not defined before the first sample we keep the parameter β_0 to describe birth before the study. This means the complete set of birth parameters is $(\beta_0, \eta_1, \ldots, \eta_{k-2})$.

In order to write the model in terms of the per-capita birth rate $\boldsymbol{\eta}$ we

must factorize the CDL in equation 3.8 carefully to ensure that we obey the rules of conditional probability. The CDL for the CMAS model is

$$
\mathcal{L}_C(\boldsymbol{S}, \boldsymbol{p}, \beta_0, \boldsymbol{\eta}, \aleph; \boldsymbol{X}, \boldsymbol{b}, \boldsymbol{d}) \propto [\boldsymbol{X}, \boldsymbol{b}, \boldsymbol{d} | \boldsymbol{p}, \beta_0, \boldsymbol{\eta}, \boldsymbol{S}, \aleph]
$$
$$
\propto [\boldsymbol{X} | \boldsymbol{p}, \boldsymbol{d}, \boldsymbol{b}, \aleph][\boldsymbol{b}, \boldsymbol{d} | \beta_0, \boldsymbol{\eta}, \boldsymbol{S}, \aleph]
$$

(4.13)

The capture component is specified in equation 3.9. The birth and death components are further factorized in terms of the natural order of events through time,

$$
[\boldsymbol{b}, \boldsymbol{d} | \beta_0, \boldsymbol{\eta}, \boldsymbol{S}, \aleph] \propto \prod_{i=1}^{\aleph} \left\{ [b_{i0} | \beta_0] \prod_{j=1}^{k-1} \left\{ [b_{ij} | \boldsymbol{b}_{:(0:j-1)}, \boldsymbol{d}_{:(1:j-1)}, \beta_0, \boldsymbol{\eta}_{1:j}, \aleph] \times \right. \right.
$$
$$
\left. \left. [d_{ij} | \boldsymbol{b}_{:(0:j)}, \boldsymbol{d}_{:(1:j-1)}, S_j] \right\} \right\}
$$

(4.14)

The death component is

$$
[d_{ij} | \boldsymbol{b}_{:(0:j)}, \boldsymbol{d}_{:(1:j-1)}, \boldsymbol{S}]
$$
$$
\propto \left\{ 1 - \left(\sum_{h=0}^{j-1} b_{ih} - \sum_{h=1}^{j-1} d_{ih} \right) (1 - S_j) \right\}^{1-d_{ij}}
$$
$$
\left\{ \left(\sum_{h=0}^{j-1} b_{ih} - \sum_{h=1}^{j-1} d_{ih} \right) (1 - S_j) \right\}^{d_{ij}} \quad j = 1, \dots, k-1
$$

(4.15)

$$
\propto \left\{ \sum_{h=1}^{j-1} d_{ih} \right\}^{1-d_{ij}} \left\{ 1 - \sum_{h=1}^{j-1} d_{ih} \right\}^{d_{ij}} \quad j = k,
$$

where the terms

$$
\left(\sum_{h=0}^{j-1} b_{ih} - \sum_{h=1}^{j-1} d_{ih} \right) \quad \text{and} \quad \sum_{h=1}^{j-1} d_{ih}
$$

enforce the constraints that individuals are born before they can die and

$\sum_{j=1}^{k} d_{ij} = 1$, $i = 1, \ldots, \aleph$. The death component assumes that, conditional on being alive at the time of sample j, the probability of death between j and $j+1$ is $1 - S_j$.

The birth component is

$$[b_{i0}|\beta_0] \propto \beta_0'^{b_{i0}}(1 - \beta_0')^{1-b_{i0}}$$

$$[b_{ij}|\boldsymbol{b}_{:(0:j-1)}, \boldsymbol{d}_{:(1:j-1)}, \beta_0, \boldsymbol{\eta}_{1:j}, \aleph] \qquad\qquad (4.16)$$

$$\propto \left\{ \left(1 - \sum_{h=0}^{j-1} b_{ih} \right) \beta_j' \right\}^{b_{ij}} (1 - \beta_j')^{\left(1 - \sum_{h=0}^{j} b_{ih}\right)}, \ j = 1, \ldots, k-1,$$

where

$$\beta_0' = \beta_0,$$

$$\beta_j' = \frac{\eta_j N_j}{\aleph \prod_{h=0}^{j-1}(1 - \beta_h')}, \ j = 1, \ldots, k-2,$$

$$\beta_{k-1}' = 1.$$

The constraint $\beta_{k-1}' = 1$ is equivalent to constraining $\sum_{h=0}^{k-1} \beta_h = 1$ and together with $(1 - \sum_{h=0}^{j-1} b_{ih})$ impose the constraint $\sum_{h=0}^{k-1} b_{ih} = 1$, $i = 1, \ldots, \aleph$. The birth component assumes that conditional on not previously being born, the probability of being born between sample j and sample $j+1$ is β_j'. As we are conditioning on $\boldsymbol{b}_{:(0:j-1)}$ and $\boldsymbol{d}_{:(1:j-1)}$ in the birth model for b_{ij} it is legitimate to use N_j as specified in equation 3.12 as a part of the model.

4.3.3 Density Dependence

As well as including N_j in the birth model, we can also use it as a covariate in a density dependence model. Density dependence assumes that the population size is an important predictor for both survival and fecundity of animal

populations. In particular, when the population size is high, competition
for resources is strong, reducing both survival probabilities and fecundity.
However, when the population size is low there is little competition for re-
sources, resulting in high survival probabilities and high fecundity. As noted
by Armstrong et al. (2005): "There is evidence for density dependence in
a wide range of species... but most studies can be challenged on statistical
grounds". The statistical problem is that when the interval-censored times
of birth and death are explicitly integrated out of the model in the standard
way, the population size is not explicitly included in the model. This means
that unless there is additional information on population size then the data
are used twice; once to estimate abundance and then again to use the abun-
dance estimate in a density dependent relationship. This is why Seber and
Schwarz (2002) state: "Tools to investigate the whole issue of density de-
pendence and dependence upon the actions of other individuals are not yet
readily available [for capture-recapture data]. Models that estimate abun-
dance (e.g., Jolly-Seber models) are available, but the feedback loop between
abundance and subsequent parameters has not yet been complete". We are
able to complete the feedback loop because we have $N = (N_1, \ldots, N_k)$ in
the model available to be used. We can specify density dependent models
for either the CMAS model or the Jolly-Seber model.

Density Dependence in the CMAS model

As we include the density dependence relationship through random effects
models for $S = (S_1, \ldots, S_{k-1})$ and $\eta = (\eta_1, \ldots, \eta_{k-2})$, we include these
values as latent variables using data augmentation. This means the CDL

for the CMAS becomes

$$\mathcal{L}_C(\beta_0, \gamma, \tau, p, \aleph; X, b, d, S, \eta) \propto [X, b, d, S, \eta | p, \beta_0, \gamma, \tau, \aleph]$$

$$\propto [X|p, d, b, \aleph][b, d, S, \eta | \beta_0, \gamma, \tau, \aleph],$$

where γ and τ are collections of parameters that model the density dependence. The component $[X|p, d, b, \aleph]$ is the standard CMAS capture distribution and is specified in equation 3.9. To include density dependence we factor the component $[b, d, S, \eta | \beta_0, \gamma, \tau, \aleph]$ in terms of the natural ordering of events in time,

$$[b, d, \eta, S | \beta_0, \gamma, \tau, \aleph] \propto \prod_{i=1}^{\aleph} \left\{ [b_{i0}|\beta_0] \times \right.$$
$$\prod_{j=1}^{k-1} \left\{ [b_{ij}|b_{:(0:j-1)}, d_{:(1:j-1)}, \beta_0, \eta_{1:j}, \aleph][d_{ij}|b_{:(0:j)}, d_{:(1:j-1)}, S_j] \right\} \right\} \times$$
$$\prod_{j=1}^{k-2} [\eta_j|b_{:(0:j-1)}, d_{:(1:j-1)}, \gamma, \tau, \aleph] \prod_{j=1}^{k-1} [S_j|b_{:(0:j-1)}, d_{:(1:j-1)}, \gamma, \tau, \aleph].$$

$$(4.17)$$

The factorization is similar to the one in section 4.3.2 with the components $[d_{ij}|b_{:(0:j)}, d_{:(1:j-1)}, S_j]$ and $[b_{ij}|b_{:(0:j-1)}, d_{:(1:j-1)}, \beta_0, \eta_{1:j}, \aleph]$ defined in equations 4.15 and 4.16. We have also factorized the survival and per-capita birth rates in time order. As both S_j and η_j depend on $b_{:(0:j-1)}$ and $d_{:(1:j-1)}$, we can legitimately model S_j and η_j in terms of N_j as given in equation 3.12. One potential model for the distributions for S_j and η_j is

$$[S_j|b_{:(0:j-1)}, d_{:(1:j-1)}, \gamma, \tau, \aleph] = \text{logit} N(\gamma_0 + \gamma_1 N_j, \tau_S), \ j = 1, \ldots, k-1,$$

$$[\eta_j|b_{:(0:j-1)}, d_{:(1:j-1)}, \gamma, \tau, \aleph] = LN(\gamma_2 + \gamma_3 N_j, \tau_\eta), \ j = 1, \ldots, k-2.$$

In section 8 we fit a density dependent model using the CMAS parameterization.

Density Dependence in the JS model

As we include the density dependence relationship through random effects models for $S = (S_1, \ldots, S_{k-1})$ and $\eta = (\eta_1, \ldots, \eta_{k-1})$, we include these values as latent variables using data augmentation. The CDL for the JS model in 4.10 is extended to include distributions for S_j and η_j which are factored in the natural order of events in time,

$$\mathcal{L}_C(p, \gamma, \tau; X_2, d_2, X_1, S, \eta, U', B)$$

$$\propto [u_1 | p_1, U_1] \prod_{j=2}^{k} [u_j | p_j, U'_{j-1}, B_{j-1}][X_2 | p, d_2, X_1][d_2 | S, X_1] \times$$

$$\prod_{j=1}^{k-1} \left\{ [U'_j | S_j, u_{1:j}, d_{2:(1:j-1)}, B_{j-1}, U'_{j-1}][S_j | u_{1:j-1}, d_{2:(1:j-1)}, U'_{j-1}, B_{j-1}] \times \right.$$

$$\left. [B_j | \eta_j, u_{1:j-1}, d_{2:(1:j-1)}, B_{j-1}, U'_{j-1}][\eta_j | u_{1:j-1}, d_{2:(1:j-1)}, U'_{j-1}, B_{j-1}] \right\},$$

where γ and τ are collections of parameters that model the density dependence. As both S_j and η_j depend on $u_{1:j-1}$, $d_{2:(1:j-1)}$, U'_{j-1} and B_{j-1}, we can legitimately model S_j and η_j in terms of N_j as given in equation 4.12. One potential model for the distributions for S_j and η_j is

$$[S_j | u_{1:j-1}, d_{2:(1:j-1)}, U'_{j-1}, B_{j-1}] = \text{logit} N(\gamma_0 + \gamma_1 N_j, \tau_S), \ j = 1, \ldots, k-1,$$

$$[\eta_j | u_{1:j-1}, d_{2:(1:j-1)}, U'_{j-1}, B_{j-1}] = LN(\gamma_2 + \gamma_3 N_j, \tau_\eta), \ j = 1, \ldots, k-1.$$

In section 6.6 we discuss fitting a density dependent model using the JS parameterization in WinBUGS.

4.4   Auxiliary Data

In many cases auxiliary data is available from other sources that can be included to improve estimation of demographic parameters in the model. There are many possible types of auxiliary data that may be available, but we will focus on information on live re-sightings and dead recoveries of the individuals between sampling periods. These observations may continue well after sampling has finished, so we split the time after the end of sampling into periods of approximately the same length as the time between sampling periods and denote the total number of re-sighting and recovery periods as k'. If $k' \geq k$, the death covariate d changes dimension to accommodate the additional information on mortality from the auxiliary information.

4.4.1   Recovery models

A common form of auxiliary information is the reporting of dead recoveries of the individuals outside of the sampling times (Burnham 1993). We include this data through the matrix Y_1, where $Y_{1ij} = 1$ denotes individual i being reported dead between sample j and $j+1$ and $Y_{1ij} = 0$ otherwise. Assuming that an individual died in the same period it was reported, the recoveries give us more information about the interval-censored times of death d. For all individuals that were ever recovered dead we know their time of death, so we have complete information about d for that individual. Including the information on death obtained from Y_1 means that X is no longer completely modeling the missingness mechanism for d. The missingness mechanism for d is only fully specified if both X and Y_1 are included in the model. The parameters required to include the recoveries are $r = (r_1, \ldots, r_{k'})$, where r_j is the probability that an individual who dies between sample j and $j + 1$

will be recovered dead. The CDL for the model with recoveries Y_1 is,

$$\mathcal{L}_C(S, p, r; X_2, X_1, d_2, Y_1) \propto [X_2, d_2, Y_1 | p, S, r, X_1]$$

$$\propto [X_2 | p, d_2, X_1][d_2 | S, X_1][Y_1 | r, d_2]. \tag{4.18}$$

The recovery component is assumed to be

$$[Y_1 | r, d_2] \propto \prod_{i=1}^{u.} r_{\mathfrak{D}_i}^{Y_{1i\mathfrak{D}_i}} (1 - r_{\mathfrak{D}_i})^{1 - Y_{1i\mathfrak{D}_i}}. \tag{4.19}$$

The distribution $[X_2 | p, d_2, X_1]$ is the capture component from the CJS model as specified in equation 3.4. The death component is

$$[d_2 | S, X_1] = \prod_{i=1}^{u.} [d_{2i:} | S, X_1],$$

$$[d_{2i:} | S, X_1] = MN(1, \boldsymbol{\xi}_i), \quad i = 1, \ldots, u.$$

Assuming that $k' \geq k$ we have

$$\boldsymbol{\xi}_i = \left((1 - S_{\mathfrak{F}_i}), S_{\mathfrak{F}_i}(1 - S_{\mathfrak{F}_i+1}), \ldots, \prod_{h=\mathfrak{F}_i}^{k'-1} S_h(1 - S_{k'}), \prod_{h=\mathfrak{F}_i}^{k'} S_h \right).$$

This is similar to the mortality component for the CJS model in equation 3.3 except there are k' survival periods instead of $k - 1$. If $k' < k$ the death component is specified in equation 3.3. The DAG for this model is in figure 4.5.

4.4.2 Joint re-sighting and recovery models

As well as information about dead recoveries, there may also be information from live re-sightings (Barker 1997). We include this through the matrix Y_2, where $Y_{2ij} = 1$ denotes individual i being re-sighted alive between sample j
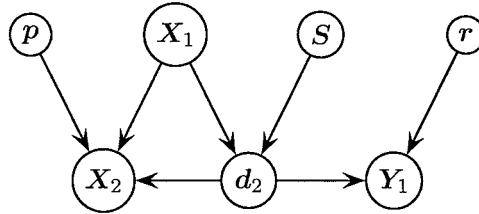
Figure 4.5: A DAG for the CJS model with auxiliary data on dead recoveries. This is the dead recovery model of Burnham (1993) with no movement.

and $j + 1$ and $Y_{2ij} = 0$ otherwise. Any re-sighting after the period of last capture provides information about the interval censored times of death d. Including the information on death obtained from Y_2 means that X and Y_1 are no longer sufficient to account for the missingness mechanism for d. The missingness mechanism for d is only fully specified when X, Y_1 and Y_2 are all included in the model. The parameters required to include the re-sightings are $R = (R_1, \ldots, R_{k'})$ and $R' = (R'_1, \ldots, R'_{k'})$, where R_j is the probability of an individual being re-sighted between sample j and $j + 1$ given that it is alive in both j and $j + 1$ and R'_j is the probability of an individual being re-sighted between samples j and $j + 1$ given that it died between j and $j + 1$ and was not recovered dead. The CDL for the model with recoveries Y_1 and re-sightings Y_2 is,

$$\mathcal{L}_C(S, p, r, R, R'; X_2, X_1, d_2, Y_1, Y_2)$$

$$\propto \underbrace{[X_2, d_2, Y_1 | p, S, r, X_1]}_{\text{Recovery Model from equation 4.18}} \times \underbrace{[Y_2 | R, R', d_2, Y_1, X_1]}_{\text{Re-sighting component}}.$$

We follow the model of Barker et al. (2004) and model Y_2 as,

$$[Y_2 | R, R', d_2, Y_1, X_1]$$

$$\propto \prod_{i=1}^{u.} \left\{ \left({R'_{\mathfrak{D}_i}}^{Y_{2i\mathfrak{D}_i}} (1 - R'_{\mathfrak{D}_i})^{1-Y_{2i\mathfrak{D}_i}} \right)^{1-Y_{1i\mathfrak{D}_i}} \prod_{j=\mathfrak{F}_i}^{\mathfrak{D}_i - 1} R_j^{Y_{2ij}} (1 - R_j)^{1-Y_{2ij}} \right\}. \tag{4.20}$$

The DAG for the CJS model with re-sightings and recoveries is in figure 4.6.
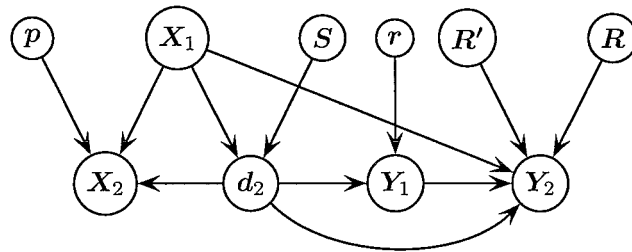
Figure 4.6: A DAG for the CJS model with auxiliary data on dead recoveries and live re-sightings.

An alternate parameterization used in Barker (1997) is in terms of $f = (f_1, \ldots, f_{k'})$, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_{k'})$ and $\boldsymbol{R} = (R_1, \ldots, R_{k'})$,

$$R_j = R_j$$
$$r_j = \frac{(1 - \nu_j) f_j}{1 - S_j} \qquad j = 1, \ldots, k' \tag{4.21}$$
$$R'_j = \frac{\frac{f_j - S_j R_j}{1 - S_j} - r_j}{1 - r_j}$$

where f_j is the probability an individual is either re-sighted or recovered between j and $j + 1$ given that they were alive in sample j and ν_j is the probability that an individual is re-sighting alive given they were either re-sighted alive or recovered dead between j and $j + 1$. The DAG for this

parameterization is in figure 4.7.
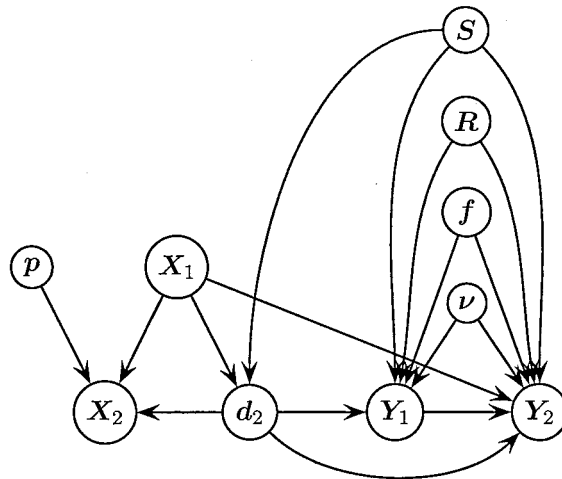
Figure 4.7: A DAG for the CJS model with auxiliary data on dead recoveries and live re-sightings using the f, ν and R parameterization.

We fit a joint re-sighting and recovery model using the f, ν and R parameterization in chapter 9.

4.5   Robust Design

The only change required to include the robust design is the constraining of parameters that allow individuals to enter and exit the population during the secondary sampling periods. Examples of appropriate constraints include $S_j = 1$ and $\eta_j = 0$. A nice result is that this means there are no major structural changes required to fit the robust design.

The importance of the robust design is best seen with complex models such as the CMAS model with density dependence on survival and per-capita birth rates. The same data is used to estimate all unknowns, including p, S, η and N with high sampling dependencies likely between parameters in

the model, making the model difficult to fit. This is not true with the robust design because the primary periods are used to estimate the demographic parameters S and η, while the secondary periods are used to estimate p and N. This should vastly improve the sampling correlation of the model.

4.6 Mother-of-all-models

The idea behind the mother-of-all-models is to have the likelihoods for commonly used capture-recapture models factorized into conditional likelihoods that can be called and combined on request to give a user specified model. Barker and White (2004) mapped out a conceptual plan for the mother-of-all-models that included the robust design and joint recapture, live resighting models. However they were unable to obtain a factorization that could easily include the multi-state model. Specifying the model in terms of the CDL overcomes these problems. As outlined above, closed population, CJS, JS, CMAS, multistate, continuous covariate, auxiliary data and robust design models can be obtained by selecting the appropriate conditional likelihood components and multiplying them together. At least in theory, this means that it is possible to develop software that allows for user-specific customization of the likelihood. Users would be able to select conditional likelihood components that are appropriate for the data they have collected, allowing the focus of modeling to be on choosing good scientific models for the appropriate components.

Part III

Model Fitting

Chapter 5

Inference

A Bayesian framework is adopted for inference. The Bayesian approach is preferred over other methods of inference because it has better machinery with which to fit hierarchical models that incorporate data augmentation.

5.1 Bayesian Methods

Adopting a Bayesian framework for statistical inference means that we express all of the uncertainty about parameters in terms of probability statements. In particular, we assign uncertainty in terms of a probability distribution function (pdf). Inference is based solely on the distribution of the parameters given the data, the so-called posterior probability distribution. The posterior distribution for parameters θ conditional on data y is found using

$$[\theta|y] = \frac{[y|\theta][\theta]}{\int [y|\theta][\theta]\ d\theta},\tag{5.1}$$

where $[\theta|y]$ is the posterior distribution, $[y|\theta]$ is proportional to the likelihood and $[\theta]$ is the joint prior distribution for the parameters.

The major criticism of Bayesian methods is the need to specify the prior

distribution $[\theta]$. The common argument used is that prior distributions make inference subjective because different people have different prior beliefs. Even though two people may have the same data y and use the same likelihood $[y|\theta]$ they may obtain a different posterior distribution because of the difference in their prior beliefs. While frequentists[1] point to the need for subjective priors as a fundamental flaw of the method, they must also recognize a certain subjectiveness in their own procedures. Even though maximum likelihood estimation is optimal asymptotically, there is no guarantee it will be for a finite sample. So how do we know which estimator to use for any given problem? An example of this subjectiveness is seen in linear calibration (also called inverse regression/prediction), see Osborne (1991) for a review of methods. Two main frequentist estimators have been proposed (i) the classical estimator[2] and (ii) the inverse estimator. Both estimators have virtues and downsides with no one estimator winning widespread approval. Interpreting the problem from a Bayesian perspective shows that the two different estimators are equivalent to choosing two different prior distributions for the missing predictor, that make very different assumptions (Hoadley 1970, Hunter and Lamboy 1981). One could argue that Bayesian inference has an advantage in that any assumptions and prior knowledge is explicitly stated and sensitivity to the prior distributions can be examined, whereas frequentist inference often relies on hidden assumptions that are sometimes very difficult to ascertain (as is the case for linear calibration). A good philosophy is that honesty should prevail in both the Bayesian and frequentist paradigm with full disclosure about methods and assumptions used. For Bayesian inference this means that the prior distributions should

[1]A frequentist is someone who considers probability in terms of long run frequencies of outcomes. Therefore, they reject the notion that probability can be used as a measure of uncertainty.

[2]Which is the MLE

always be disclosed.

Bayesian inference is conceptually very simple. Equation 5.1 is used to find the posterior distribution from which all inference is made. However, the integral in the denominator of 5.1 makes the calculation of the posterior distribution difficult in practice. For all but the very simplest of models, obtaining the marginal distribution of y has been prohibitive, especially when there are a large number of parameters.

The inability to obtain the posterior distribution, as well as questions about the foundations of Bayesian inference, led to a 200 year hiatus in the extensive use of the framework. However, widespread use of Bayesian methods has come with simulation techniques that make it possible to simulate from the required posterior distribution without the need to evaluate the marginal distribution of the data. These simulation techniques make it possible to sample from the proportional posterior distribution,

$$[\theta|y] \propto [y|\theta][\theta]. \tag{5.2}$$

This means we can avoid calculating the marginal likelihood

$$[y] = \int [y|\theta][\theta] \ d\theta.$$

However, even though we do not require the marginal likelihood to fit the model, it is still important, particularly when we are looking to perform model selection (see section 5.3.2).

The recent popularity of Bayesian methods has largely come because any missing data or latent variables can easily be included in the model using data augmentation. In the situation where we have missing data so that

$y = (y^{obs}, y^{mis})$, the required posterior distribution becomes

$$[\theta|y^{obs}] = \frac{\int [y|\theta][\theta] \, dy^{mis}}{\int\int [y|\theta][\theta] \, dy^{mis} \, d\theta}.$$

Latent variables, such as random effects are included in a similar way. Consider the latent variable η included as a random effect. The required posterior distribution becomes

$$[\theta|y^{obs}] = \frac{\int [y^{obs}|\eta][\eta|\theta][\theta] \, d\eta}{\int\int [y^{obs}|\eta][\eta|\theta][\theta] \, d\eta \, d\theta}.$$

Even though both of these calculations look prohibitive for practical inference, they are not. In practice, when using the simulation techniques that will be described in the next section, the only change that is required is that the missing data y^{mis} or latent variables η are treated like any other unknown and updated at every iteration.

5.2   Simulating from the Posterior Distribution

The methods below are not Bayesian per se. They are methods that allow us to draw a sample from any distribution that is specified up to its proportionality constant.

5.2.1   Rejection Sampling

Rejection sampling can be used to generate samples from a univariate distribution that has an unknown proportionality constant,

$$[x] \propto g(x),$$

where $g(x)$ is known. The technique requires an envelope function $f(x)$ that (i) integrates to 1, (ii) can be sampled from, and (iii) covers the proportional pdf $g(x)$ for all values in its support once appropriately scaled,

$$M \ f(x) > g(x), \ \forall x,$$

where M is the scaling factor that ensures coverage.

The algorithm consists of two steps:

1. Generate a sample $x^\star$ from the envelope function,

$$x^\star \sim f(x).$$

2. Accept the candidate with probability q,

$$q = \frac{g(x^\star)}{M \ f(x^\star)}.$$

If accepted: $x^\star \sim [x]$.

If rejected: return to step 1.

Example: Binomial Probability

Consider data y from a binomial distribution with index N and unknown probability p. If we were to specify a beta prior distribution with parameters α and β the posterior distribution for p is proportional to

$$[p|y, N] \propto p^{y+\alpha-1}(1-p)^{N-y+\beta-1} \tag{5.3}$$

The posterior distribution for p is known to be a beta distribution,

$$[p|y, N] = Beta(y + \alpha, N - y + \beta).$$

Alternatively, we could use equation 5.3 to generate samples from the posterior distribution for p using rejection sampling. We specify the envelope function to be $f(p)$

$$f(p) = Unif(0, 1)$$

with scaling factor

$$M = \tilde{p}^{y+\alpha-1}(1 - \tilde{p})^{N-y+\beta-1}$$

where

$$\tilde{p} = \frac{y + \alpha - 1}{N + \alpha + \beta - 2}$$

is the value of p that maximizes equation 5.3.

The algorithm proceeds as:

1. Generate the candidate value

$$p^\star \sim Unif(0, 1)$$

2. Accept $p^\star$ with probability q

$$q = \left(\frac{p^\star}{\tilde{p}}\right)^{y+\alpha-1} \left(\frac{1 - p^\star}{1 - \tilde{p}}\right)^{N-y+\beta-1}$$

If accepted: $p^\star$ is a sample from $[p|y]$

If rejected: return to step 1

For $y = 60$, $N = 100$, $\alpha = 1$ and $\beta = 1$, a kernel density smooth of $50,000$ values from the rejection sampler provides an excellent approximation to

the posterior distribution, as shown in figure 5.1.

Figure 5.1: Comparison between the true posterior distribution for p (red-line) and the posterior distribution estimated by a kernel density smooth of $50,000$ samples found using rejection sampling (blue line).

Different Envelopes

The efficiency of the algorithm can be improved by specifying an envelope distribution that closely approximates the posterior distribution, while still ensuring coverage. If the target distribution is log-concave, a piecewise exponential envelope function can give an excellent approximation to the posterior distribution, so long as enough pieces are chosen and they straddle the maximum. To further increase efficiency, the envelope can be adaptively improved at each iteration (Gilks and Wild 1992).

Problems with the rejection sampling

There are two major difficulties with rejection sampling:

- The extension to multivariate rejection sampling is very difficult due to the need to specify a multivariate envelope function. For practical applications, rejection sampling is only feasible in one dimension.

- It is very difficult to specify an envelope function for a non log-concave distribution with infinite support.

5.2.2   Markov Chain Monte Carlo Methods

Markov chain Monte Carlo (MCMC) is a technique where an ergodic[3] Markov chain is constructed that has as its limiting stationary distribution the posterior distribution of interest. The values generated using MCMC are only guaranteed to be a sample from the posterior distribution once the Markov chain has converged to its limiting distribution. Therefore, every time we use MCMC it is essential that we check that the chains appear to have converged, even though we can never be certain that they have in fact converged. One check of convergence is to see whether chains that are started at different points come together. This is called mixing. To ensure that the chains mix, multiple chains should be run with over-dispersed starting values[4]. A good diagnostic tool to determine convergence is simply by eye: do the chains look mixed? If we are unsure, the chains need to be run for longer. If required, there are also diagnostic tools available to help determine convergence (Brooks and Gelman 1998). The posterior sample should only be collected once we are sure the chains are well mixed.

---

[3]For more information about ergodicity see Cox and Miller (1965).

[4]Note that even though starting values should be over-dispersed they should also be reasonable. Once far enough into the tail of a distribution it can be so flat that algorithms may struggle to find a region of high probability.

There are many different MCMC algorithms (Lui 2001, Robert and Casella 2004) but we focus our attention on the Gibbs sampler, the Metropolis-Hastings algorithm and variants of these.

Gibbs Sampling

The Gibbs sampler (Geman and Geman 1984) is a technique used to sample from multivariate distributions. It is also known as alternative conditional sampling because the algorithm consists of sampling from the full conditional distribution of each unknown in turn. In situations where there is high sampling correlation between unknowns, blocks of unknowns can be updated in turn instead of one at a time.

Suppose we are wanting to sample from the distribution $[x]$, where $x = (x_1, x_2, x_3)$. If we are in iteration m, the standard Gibbs sampler alternates between the following full conditional distributions:

$$[x_1^{(m+1)} | x_2^{(m)}, x_3^{(m)}]$$
$$[x_2^{(m+1)} | x_1^{(m+1)}, x_3^{(m)}]$$
$$[x_3^{(m+1)} | x_1^{(m+1)}, x_2^{(m+1)}].$$

One advantage of Gibbs sampling is that we are able to use any sampling scheme to generate samples from the full conditional distributions. If the full conditional distribution is of known form, we can sample directly from that. However, if the full conditional distribution is not of known form then we can use rejection sampling, the Metropolis-Hastings algorithm or any other sampling scheme that we are smart enough to devise.

The Gibbs sampler is used for all models in this thesis. If the full conditional distributions are of known form we sample directly from these. Otherwise, we use another sampling scheme, most frequently the Metropolis-

Hastings algorithm to obtain the sample.

Slice Sampler

The slice sampler (Neal 2003, Higdon 1998, Damien et al. 1999) is a technique used to improve the convergence of the Markov chain. Interest is in sampling from the distribution $[x]$ which is not of known form and techniques that obtain independent samples from the required distribution, such as rejection sampling, cannot be easily used. The slice sampler uses data augmentation to introduce an auxiliary variable u that invokes a Gibbs sampler that has the distribution of interest as its limiting distribution. In each iteration we sample from:

1. The distribution for the auxiliary variable u conditional on the current value of the x:

$$[u^{(m+1)}|x^{(m)}] = Unif(0, f(x^{(m)})).$$

2. The distribution for x conditional on the current value of u:

$$[x^{(m+1)}|u^{(m+1)}] = Unif(a),$$

where a is the set of values of x such that $f(x) \geq u^{(m+1)}$.

The easiest way to understand the slice sampler is with an example[5]. We wish to sample from the distribution

$$[x] = \frac{1}{2}\exp\left(-\sqrt{x}\right), \ x > 0.$$

---

[5]The example is found in Robert and Casella (2004)

Assuming we are in iteration m, the first step is to generate a sample from $[u^{(m+1)}|x^{(m)}]$,

$$[u^{(m+1)}|x^{(m)}] = Unif\left(0, \frac{1}{2}\exp\left(-\sqrt{x^{(m)}}\right)\right).$$

The next step is to find the set of values a for which $f(x) \geq u^{(m+1)}$,

$$a = \left\{x : \frac{1}{2}\exp\left(-\sqrt{x}\right) > u^{(m+1)}\right\}$$
$$= \left\{x : -\sqrt{x} > \log\left(2u^{(m+1)}\right)\right\}$$
$$= \left\{x : -\left(\log\left(2u^{(m+1)}\right)\right)^2 < x < \left(\log\left(2u^{(m+1)}\right)\right)^2\right\}.$$

The constraint that $x > 0$ further restricts the set a,

$$a = \left\{x : 0 < x < \left(\log\left(2u^{(m+1)}\right)\right)^2\right\}.$$

Therefore, the full conditional distribution for $x^{(m+1)}$ is

$$[x^{(m+1)}|u^{(m+1)}] = Unif\left(0, \left(\log\left(2u^{(m+1)}\right)\right)^2\right).$$

Alternating between these two conditional distributions yields a Markov chain that has the distribution of interest $[x]$ as its limiting distribution.

Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm (Metropolis et al. 1953, Hastings 1970) is arguably the most important MCMC algorithm. The algorithm is very general and can be used for both univariate and multivariate distributions. We focus on the univariate case as we are making use of the Gibbs sampler for multivariate distributions.

Suppose that we are wanting samples from the distribution $[x]$ which has as unknown proportionality constant,

$$[x] \propto g(x),$$

where $g(x)$ is known. In iteration m, the Metropolis-Hastings algorithm proceeds as follows:

1. A candidate value of the parameter $x^\star$ is generated from a specified jumping distribution $J(x^\star|x^{(m)})$.

2. The candidate is accepted with probability q, where

$$q = \min\left(1, \frac{g(x^\star)J(x^{(m)}|x^\star)}{g(x^{(m)})J(x^\star|x^{(m)})}\right). \qquad (5.4)$$

   If accepted: $x^{(m+1)} = x^\star$

   If rejected: $x^{(m+1)} = x^{(m)}$.

3. Update the iteration counter to $m+1$ and return to step 1.

The sequence $(x_1, x_2, \ldots, x_M)$ forms a Markov chain that has as its stationary limiting distribution the distribution of interest $[x]$. In general we discard ℓ samples to ensure that the chain has converged to its limiting distribution and take $(x_{\ell+1}, \ldots, x_M)$ to be a sample from $[x]$.

In the Metropolis algorithm (Metropolis et al. 1953) the jumping distribution is symmetric, so $J(x^\star|x^{(m)}) = J(x^{(m)}|x^\star)$. These terms then cancel out of the acceptance calculation in equation 5.4.

As the jumping distribution is specified by the user it can be chosen to optimize the algorithm. If the approximate form of the posterior is known then an independent jumping distribution $J(x^\star)$ can be specified that is close to the posterior and has a high rate of acceptance. If the approximate form

of the posterior is not known we can specify a general jumping distribution that can be adapted during a training phase in order to draw an efficient sample. Consider a normal random walk jumping distribution,

$$J(x^\star | x^{(m)}) = N(x^{(m)}, \sigma_J^2)$$

where σ_J^2 is specified by the user. We can optimize the algorithm by choosing a good value of σ_J^2. If σ_J^2 is too large or too small the Markov chain is slow to mix. The optimal value for σ_J^2 is one that gives an acceptance rate of approximately 44% provided the target distribution $[x]$ is approximately normal (Gelman et al. 2004). This can be achieved through tuning σ_J during a training phase. We initialize σ_J to some value (say 0.001) and if the candidate value is accepted we then multiply σ_J by the value a (say $a = 1.1$). If the candidate value is rejected we then multiply σ_J by the value

$$b = a^{-\frac{44}{56}}$$

to achieve an approximate acceptance rate of 44%. Changing the value of b changes the expected acceptance rate. In general we set $b = a^{-1}$ in order to approximate 50% acceptance rates in the chains.

In situations where x is categorical a similar process can be used. We select a discrete uniform jumping distribution centered at the current value of the parameter. We give the current value the probability of 0, thus enforcing the candidate to be different from the current value. The first step of our proposal is generating a value A

$$A \sim Bern(0.5)$$

We then use this value in the discrete uniform jumping distribution

$$J(x^\star | x^{(m)}) = \begin{cases} DU(\theta^{(m)} - \kappa, \theta^{(m)} - 1) & \text{if } A = 0 \\ DU(\theta^{(m)} + 1, \theta^{(m)} + \kappa) & \text{if } A = 1 \end{cases}$$

The value that is adapted is κ. We initialize κ to some discrete positive value (say 1) and set up an adaptive scheme during the training phase. If the candidate value is accepted we add 1 to the value of κ and if the candidate value is rejected we then subtract 1 from the value of κ but ensure κ never goes below 1.

For both the continuous and categorical case, the samples obtained in the training phase must be discarded, even if it looks as if the Markov chain has mixed well. This is because we have no assurance that the Markov chain we have generated while changing the jumping distribution has the target distribution as its limiting distribution. Therefore, we run the training phase for ℓ iterations to obtain the parameters of the jumping distribution, which are then fixed for the subsequent samples.

Example: Binomial Probability

Consider the binomial problem from section 5.2.1. We are interested in sampling values from the posterior distribution for p when $y = 60$, $N = 100$, $\alpha = 1$ and $\beta = 1$. The proportional posterior for p is given in equation 5.3 as

$$[p|y, N] \propto p^{y+\alpha-1}(1-p)^{N-y+\beta-1}$$

We specify a random walk logit-normal jumping distribution for p,

$$J(p^\star | p^{(m)}) = \text{logit}N(\text{logit}(p^{(m)}), \sigma_J^2).$$

The acceptance probability for the candidate $p^\star$ in iteration m is,

$$q = \min\left(1, \left(\frac{p^\star}{p^{(m)}}\right)^{y+\alpha-1} \left(\frac{1-p^\star}{1-p^{(m)}}\right)^{N-y+\beta-1} \frac{p^\star(1-p^\star)}{p^{(m)}(1-p^{(m)})}\right),$$

where the term on the right is the ratio of jumping distributions. The parameter p was initialized to be 0.5, the value of a set to 1.02 and the value σ_J was initialized to 0.001. During a tuning phase of 5000 iterations the value of σ_J was optimized as described above. A kernel density smooth of the 50,000 samples subsequent to the tuning phase is in good agreement with the known posterior distribution, see figure 5.2. The value for σ_J after the tuning phase was 0.365470 and the 50,000 posterior samples had an acceptance rate of 53%, close to the desired rate of 50%. It is no surprise that the samples from rejection sampling (figure 5.1) give a better approximation to the true posterior than the samples from Metropolis-Hastings (figure 5.2). This is because the rejection algorithm draws independent samples from the posterior distribution, whereas Metropolis-Hastings generates correlated samples.

Reversible Jump MCMC

Reversible jump MCMC (Green 1995) is an extension of the Metropolis-Hastings algorithm. It is used when variables are being updated that define the dimension of other variables in the model. This dimension parameter will be referred to as the jumping parameter. An example is multi-model inference where a model indicator that determines the dimension of the parameter vector θ is included in the model as a parameter to be updated. The algorithm works by obtaining detailed balance so that lower/higher dimensions are not incorrectly favoured or penalized. Any two values of the
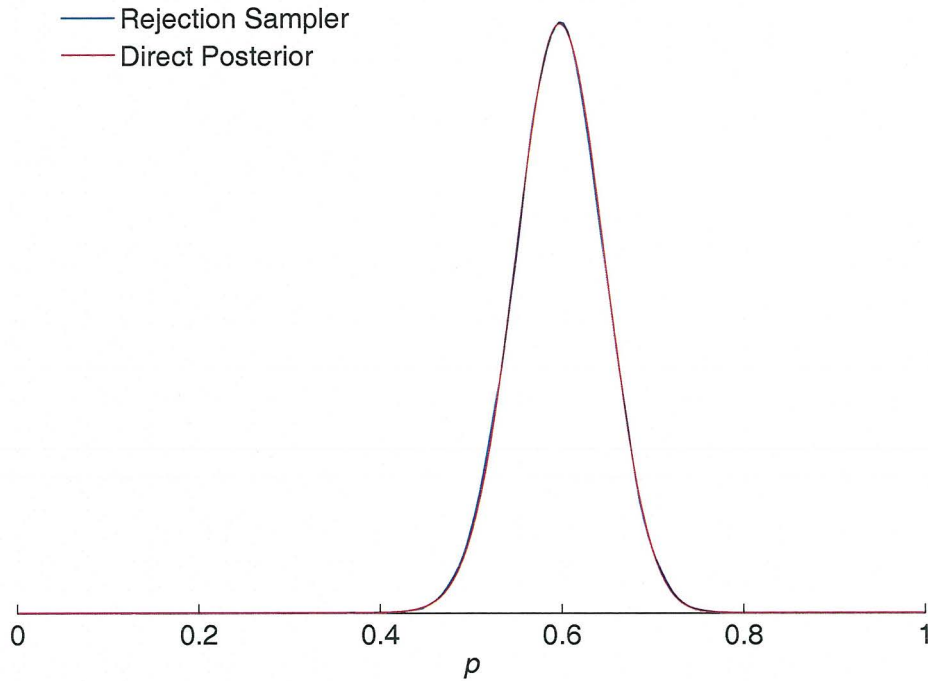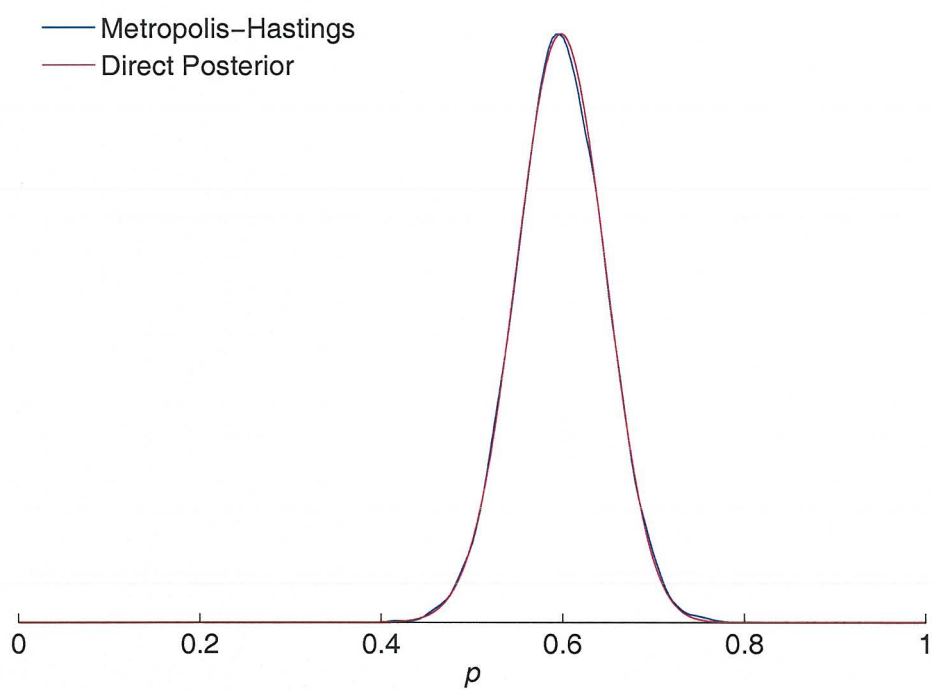
Figure 5.2: Comparison between the true posterior distribution for p (red-line) and the posterior distribution estimated by a kernel density smooth of $50,000$ samples found using the Metropolis-Hastings algorithm (blue line).

jumping parameter must have auxiliary parameters specified so that they
have the same dimension, with bijections to move from one value to another.

Consider the case where we want to sample from the distribution $[x, \zeta]$
which has an unknown proportionality constant,

$$[x, \zeta] \propto f(x, \zeta),$$

where $f(x, \zeta)$ is known. The jumping parameter is ζ and this determines
the dimension of x. In order to update ζ we propose a candidate $\zeta^\star$ from
$J(\zeta^\star | \zeta^{(m)})$. As the dimension of x changes we generate an augmenting vari-
able u from $H(u | \zeta^{(m)}, \zeta^\star, x)$, which together with the existing values $x^{(m)}$
and the bijection $(x^{(m)\star}, u^\star) = g_{\zeta^{(m)}, \zeta^\star}(x^{(m)}, u)$ give the candidate values
$x^{(m)\star}$. The candidate $\zeta^\star$ together with $x^{(m)\star}$ are accepted with probability
$q' = \min(1, q)$,

$$q = \frac{f(x^{(m)\star}, \zeta^\star) J(\zeta^{(m)} | \zeta^\star) H(u^\star | \zeta^\star, \zeta^{(m)}, x^{(m)\star})}{f(x^{(m)}, \zeta^{(m)}) J(\zeta^\star | \zeta^{(m)}) H(u | \zeta^{(m)}, \zeta^\star, x^{(m)})} \left| \frac{\nabla g_{\zeta^{(m)}, \zeta^\star}(x^{(m)}, u)}{\nabla (x^{(m)}, u)} \right| \quad (5.5)$$

where the component on the right is the determinant of the Jacobian of
the bijection. We then proceed as for the Metropolis-Hastings algorithm.
If the candidates are accepted then $(\zeta^{(m+1)}, x^{(m)}) = (\zeta^\star, x^{(m)\star})$, otherwise
$(\zeta^{(m+1)}, x^{(m)}) = (\zeta^{(m)}, x^{(m)})$. We keep the superscript m on x because we
are updating the jumping parameter ζ not x. The values of x get updated
in turn during other steps of the Gibbs sampler.

5.3 Model Checking and Selection

Two important questions when fitting models to data are:

   1. Does the model fit the data well?

2. How do you determine the best model?

The first question deals with ensuring that the model (or models) are sensible and that the data look like they could have come from the model. The second question deals with finding the best model from a set of candidate models.

5.3.1 Model Checking

A useful tool for model assessment is the posterior predictive distribution (Gelman et al. 2004). A set of replicate data is generated from the parameter values at each iteration. The replicate data is then compared to the observed data using an appropriate test statistic. If the model is good then the value of the test statistic for the replicate data should be similar to that from the observed data. However, if the model does not fit well then they will be quite different.

The key to this approach is the choice of the test statistic. There are various omnibus test statistics that have been used for capture-recapture type data, for example Brooks et al. (2000a) uses the Freeman-Tukey test statistic (Freeman and Tukey 1950, Bishop et al. 1975) and King and Brooks (2002) use the value of the likelihood as their test statistic. Omnibus tests are convenient as they available irrespective of the data being analyzed, however, this wide applicability means that they are less likely to discern lack of fit. For capture-recapture models the choice of adequate test statistics is a largely unexplored question. A promising, yet potentially time consuming approach would be to use a test statistic that is theoretically appropriate for the model being examined. For example, the test statistic used when examining a CJS model could be based on the results of Pollock et al. (1985).

5.3.2 Model Selection

Burnham and Anderson (1998) suggest that even before the data are collected much thought should go into into determining covariates likely to affect the system of interest, with only a small set of scientifically driven models considered. Once the data have been collected, Link and Barker (2006) suggest that Bayes factors are the quantity that should be used for multi-model-inference. The Bayes factor between model j and model h, BF_{jh}, is defined as the marginal likelihood of the data y under model j compared to the marginal likelihood of the data under model h,

$$BF_{jh} = \frac{[y|M_j]}{[y|M_h]}$$

where the marginal likelihood $[y|M_j]$ is evaluated as

$$[y|M_j] = \int [y|\theta, M_j][\theta|M_j] \, d\theta.$$

Bayes factors are combined with prior model weights to give posterior model weights that can be used for model weighting or model selection. For example, one can specify prior model weights so that the posterior model probabilities are the AIC model weights (Burnham and Anderson 1998, Link and Barker 2006). The problem is that Bayes factors are difficult to calculate, as they involve evaluation of the marginal probability of the data given the model (the marginal likelihood) by integrating the likelihood with respect to the prior. It is precisely this calculation that made practical Bayesian model fitting almost impossible except in relatively simple problems. The popularity of MCMC comes because it circumvents the need for the calculation of the marginal likelihood.

The calculation of Bayes factors is an area of active research in Bayesian

methods. Current approaches include estimation of the marginal likelihood from the MCMC output (Raftery et al. 2007) and methods that attempt to sample from the joint model and parameter space (Han and Carlin 2001). Of the second of these approaches, reversible-jump MCMC (RJMCMC) is perhaps the best known example among ecologists (Brooks et al. 2000b), and is currently the best general method for calculating Bayes factors, or equivalently posterior model weights given prior model weights.

In RJMCMC, the model is treated as another unknown to be estimated (see section 5.2.2). Posterior model weights are obtained which, combined with the prior model probabilities, give the Bayes factors. However, RJM-CMC is difficult to implement, due to the need to match parameter dimensions and specify efficient bijections between all candidate models. Unfortunately, no flexible and efficient RJMCMC algorithms are available in any commonly used statistical software.

For fixed effects models, the so-called Bayesian Information Criterion (BIC) can be used to find an approximation to marginal likelihood, with ratios of BICs giving the Bayes factors (Link and Barker 2006). The BIC is similar to AIC and easy to calculate,

$$BIC = -2[\boldsymbol{X}|\hat{\boldsymbol{\theta}}] + m \log(n),$$

where $[\boldsymbol{X}|\hat{\boldsymbol{\theta}}]$ is the likelihood at the posterior mode (or maximum likelihood estimate) of the parameters, m is the number of parameters in the model and n is the sample size. However, in many situations, the sample size n is not well defined and for random effects models the number of parameters m is unknown, making the criterion difficult to use (Raftery et al. 2007).

Random effects models can be used as an alternative to model averaging

and selection in many cases (Gelman et al. 2004). Consider the joint live re-sighting, dead recovery model of Barker (1997), with parameters S, R, R', r, p, F and F', where the effect of sex on survival is of interest. To complicate matters, it is thought that sex and time could effect the other (nuisance) parameters. For each parameter, we can specify (at least) 5 different models: (i) constant across time and sex, (ii) constant across time with a sex effect, (iii) varying through time with no sex effect, (iv) varying through time with an additive sex effect and (v) varying through time with a sex effect interaction. This gives a possible $5^7 = 78,125$ models we could examine. Instead of performing model selection on all, or a subset, of these nuisance parameters, we can specify a full model for each nuisance parameter, and obtain parsimony on nuisance parameters by modeling the parameters as random effects. This essentially removes the unneeded parameters from the model (as they are being sampled from the random effects distribution), so it saves us from the arduous task of performing model selection on the nuisance parameters. Conditional on having a random effects distribution for the nuisance parameters, we can then perform model selection across the 5 different models for S.

When using random effects, it is essential that distributional assumptions are appropriate. For example, the quantities being combined in the random effect could be mostly similar with the occasional spike. In order to specify an appropriate random effect, either:

- Suitable covariates need to be found that help to explain the variability, in particular, the variability associated with the spike.

- A heavy tailed distribution, such as a t-distribution, is specified so that the spike is reasonable under the model.

Chapter 6

Capture-Recapture Models

in WinBUGS

The most widely used program for Bayesian statistics is BUGS (Spiegel-halter et al. 2003), Bayes Using Gibbs Sampling. It is freely available in many forms including WinBUGS, for a windows operating system and versions for other operating systems, such as ClassicBUGS and LinBUGS. An open source version, OpenBUGS, is also available. BUGS is a general purpose statistical packages that can be used for a wide range of models and applications.

6.1   Program WinBUGS

As the name suggests, WinBUGS uses Gibbs sampling to generate samples from the posterior distribution. Models can be written in WinBUGS in two ways:

1. As a DAG in the doodleBUGS editor.

2. Writing code directly in the WinBUGS editor.

We only consider the second option as it offers the most flexibility. Nodes specified in WinBUGS are either stochastic or deterministic. The code

```
x ~ dbern(theta)
```

specifies that x is a stochastic node sampled from a Bernoulli distribution with parameter θ. The code

```
y <- sqrt(x)
```

specifies that y is a deterministic node that is the square root of the node x.

Several steps are required in order to run a model in WinBUGS:

1. The model is checked for syntax errors.

2. The data are loaded. The standard way to enter data is within in a `list(...)` statement as in program R.

3. Initial values are obtained for every unknown. These are either pre-specified or generated from the prior distribution. Pre-specified initial values are included within a separate `list(...)` statement.

4. The model is compiled into Pascal code.

5. The nodes to be monitored are selected.

6. The model is run.

6.1.1   Advantages of WinBUGS

There are many advantages to using WinBUGS:

- The full conditional distributions that are used in the Gibbs sampler are automatically calculated. This means that changes to the model, including hierarchical extensions, are easy to include and require virtually no programming effort.

- WinBUGS selects what it considers to be the best possible sampling algorithm for each full conditional distribution. If WinBUGS recognizes that the full conditional distribution is of known form then it will directly samples from the distribution. If direct sampling is not possible then WinBUGS selects a sampling algorithm from a range of possibilities, including rejection sampling, slice sampling and the Metropolis-Hastings algorithm.

- WinBUGS can be run as a script and can be called from other programs.

- WinBUGS has an inbuilt help file, including numerous examples.

6.1.2   Limits of WinBUGS

Unfortunately there are also several downsides to WinBUGS:

- There is not an explicit reversible jump step available in WinBUGS. Various tricks are required to include reversible jump steps in Win-BUGS. Durban and Elston (2005) show how inefficient data augmentation can be used to fit individual specific closed population capture-recapture models, such as model M_h in WinBUGS. These models require a reversible jump because the population size N determines the dimension of the capture history X. Link and Barker (2006) describe how to perform inefficient multi-model inference in WinBUGS. Even though these tricks are very helpful in certain situations, they can dramatically slow down the program, even to the point where models cannot be practically run.

- For large data sets WinBUGS can be very slow, especially for complex problems.

- It is very difficult to modify the algorithms to improve the rate of convergence.

- Tricks are required to use a probability distribution that is not hard-wired. One of these tricks is called the 'ones trick', where a value $w = 1$ is specified as data[1]. We then model w as a Bernoulli random variable with probability q, where we write q as the probability distribution function we wish to include. A similar trick is called the 'zeros trick', where $w = 0$ is specified as data, with w modeled as a Poisson random variable. Both the ones trick and the zeros trick are explained in full under the "Specifying a new sampling distribution" heading in the "Tricks: Advanced Use of the BUGS Language" section of the WinBUGS 1.4 help file (Spiegelhalter et al. 2003).

6.2   CJS

We examine the fitting of open population models in WinBUGS. The simplest open population model is the CJS model. Coding the CDL of the CJS in WinBUGS requires two steps:

1. Including and modeling the interval-censored time of death.

2. The modeling of capture conditional on time of death.

Implementing step 2 in WinBUGS is straight-forward. Step 1 is more difficult because the time of death is partially known through being right-censored. The complication arises because WinBUGS does not allow censored data for the multinomial distribution. Therefore, we must factor the mortality distribution for the CJS model $[d_2|S, X_1]$ defined in equation 3.3

---

[1]We can also specify a vector or matrix of ones if desired.

as

$$[d_2|S, X_1] = [d_2|d_{\min}, S][d_{\min}|S, X_1]$$

where $d_{\min}$ denotes the observed (right) censored time of death. The component $[d_2|d_{\min}, S]$ is a series of multinomial distributions from the time of last capture

$$[d_2|d_{\min}, S] = \prod_{i=1}^{u.} [d_{2i:}|d_{\min}, S]$$

$$[d_{2i:}|d_{\min}, S] = MN(1, \boldsymbol{\xi}_i)$$

where

$$\boldsymbol{\xi}_i = \left((1 - S_{\mathfrak{L}_i}), S_{\mathfrak{L}_i}(1 - S_{\mathfrak{L}_i+1}), \ldots, \prod_{h=\mathfrak{L}_i}^{k-2} S_h(1 - S_{k-1}), \prod_{h=\mathfrak{L}_i}^{k-1} S_h \right).$$

If an individual was lost on capture then $[d_{2i:}|d_{\min}, S, X_1] \propto 1$. The second term accounts for each individual surviving from first capture until last capture

$$[d_{\min}|S, X_1] \propto \prod_{i=1}^{u.} \prod_{j=\mathfrak{F}_i}^{\mathfrak{L}_i-1} S_j.$$

This factorization solves one problem but introduces another. WinBUGS does not allow a multinomial distribution to be specified for the individuals caught at the time of last capture as these individuals only have one possible outcome. There are two possible solutions: (i) writing the likelihood separately for individuals caught in the last sample, or (ii) including a $(k+1)$th column in d that has $Pr(d_{ik+1} = 0)$, $\forall i$. We prefer the second option as it allows us to have the same likelihood for all individuals.

An example of WinBUGS code that can be used to fit the CJS model $p(t)S(t)$ when imputing the times of death is given in appendix C.1. The key features of the model are:

- Line 8: Ensures that individuals are only available for capture until death.

- Line 36: Specifies $[d_2|d_{\min}, S, X_1]$.

- Line 47: Specifies $[d_{\min}|S, X_1]$ using the ones trick.

The data file requires several components:

- The capture histories X.

- The samples of first capture $\mathfrak{F} = (\mathfrak{F}_1, \ldots, \mathfrak{F}_{u.})$.

- The samples of last capture $\mathfrak{L} = (\mathfrak{L}_1, \ldots, \mathfrak{L}_{u.})$.

- The times of death d. Any missing values are including through specifying NA.

- The number of samples k

- The number of individuals caught u.

- The vector $wd = (1, \ldots, 1)$ for the ones trick.

An example of the data file needed is given in appendix C.2. The initial values for the stochastic nodes can either be specified in a way similar to the data or sampled from the prior distributions.

6.3   Jolly-Seber

The JS model extends the CJS model by including the first capture,

$$[u_j | p_j, U_j] = Bin(U_j, p_j), \ j = 1, \ldots, k.$$

The WinBUGS code that is required for the JS model is given in appendix C.3. The only additional data required is $\boldsymbol{u} = (u_1, \ldots, u_k)$.

6.4   Continuous Covariates

To include a continuous covariate into a CJS model requires two steps:

1. Including the covariate in the model for $\boldsymbol{p}$ and $\boldsymbol{S}$.

2. Specify a model for the covariate.

As an example, consider the continuous covariate model fitted by Bonner and Schwarz (2006). Individual body weight was assumed to be a predictor of both survival and probability of capture:

$$\begin{aligned} \text{logit}(S_{ij}) &= \gamma_1 + \gamma_2 z_{ij} \\ \text{logit}(p_{ij+1}) &= \gamma_3 + \gamma_4 z_{ij+1} \end{aligned} \qquad i = 1, \ldots, u., \ j = 1, \ldots, k-1$$

where z_{ij} is the standardized weight of individual i in sample j. As the weight was only obtained when each individual was caught, a model is required for the missing values. We model the weight as

$$z_{ij} \sim N(z_{ij-1} + \Delta_{j-1}, \tau), \ i = 1, \ldots, u., \ j = 1, \ldots, k.$$

where Δ_j is the mean increase in weight between j and $j+1$ and τ is the precision.

The WinBUGS code to include the continuous covariates is in appendix C.4. The only additional data that need to be specified for the continuous covariates is the matrix of covariates z. The missing values of z have the value NA and are updated as part of the Gibbs sampler.

6.5   Multi-state

If the covariate z is categorical the only differences from the continuous case is (i) the model for z and (ii) the model for how S and p depend on z. We fit a model where both survival and capture probability depend on z,

$$\begin{aligned} \mathrm{logit}(S_{ij}) &= \mathcal{S}_{hj} \\ \mathrm{logit}(p_{ij+1}) &= \pi_{lj} \end{aligned} \qquad i = 1, \ldots, u., \ j = 1, \ldots, k-1,$$

where $h = z_{ij}$, $l = z_{ij+1}$, $\mathcal{S}$ is a matrix of survival probabilities that are state and time specific and π is a matrix of capture probabilities that are state and time specific. The model for z is

$$Pr(z_{ij} = l | z_{ij-1} = h) = \Psi_{hl}, \ i = 1, \ldots, u., j = \mathfrak{F}_i + 1, \ldots, k,$$

where Ψ is a matrix of state specific transition probabilities that are constant through time. The WinBUGS code required to fit this model is in appendix C.5.

6.6   Density Dependence Using Jolly Seber Model

We make use of the density dependent specification of the JS model that is
developed in section 4.3.3. In particular we look at a model where

$$\text{logit}(S_j) \sim N(\beta_1 + \beta_2 \tilde{N}_j, \sigma_1^2), \quad j = 1, \ldots, k - 1$$

$$\log(\eta_j) \sim N(\beta_3 + \beta_4 \tilde{N}_j, \sigma_2^2), \quad j = 1, \ldots, k - 1,$$

$$\text{logit}(p_j) \sim N(\beta_5, \sigma_3^2), \quad j = 1, \ldots, k,$$

where $\tilde{N}_j$ is the value N_j that has been centered to reduce sampling corre-
lation between parameters.

In practice this model is slow to run in WinBUGS and can exhibit very
poor mixing in the Markov chain. As there is no easy way to modify the
sampling algorithms in WinBUGS there is little that can be done to improve
the convergence of the algorithm. As we are not estimating any individual-
specific parameters, one potential solution (at least for speed) is to re-write
the mortality component, d_2, in terms of sufficient statistics that require
data augmentation. We factorize the death component of the CJS defined
in equation 3.3 as

$$[d_2 | S, X_1] \propto \prod_{j=1}^{k-1} [D_j | m_j, R_j, T_j, M_j, S_j][T_j | S_j], \tag{6.1}$$

where D_j is the unknown number of individuals that die between samples
j and $j + 1$, m_j is the number of marked individuals caught in sample j,
R_j is the number of individuals caught in sample j that were subsequently
released, T_j is the number of individuals caught up to and including sample
j that were subsequently caught after j and M_j is the unknown number of
marked individual alive immediately before sample. The second component

of equation 6.1 models all individuals that are known to survive until period $j + 1$,

$$[T_j | S_j] \propto \prod_{h=1}^{k-1} S_h^{T_h}. \tag{6.2}$$

The first component of equation 6.1 is more difficult as it requires us to find the unknown values of M_j and D_j. As we know there were no marked individuals in first sample, $M_1 = 0$ with

$$M_{j+1} = M_j - m_j + R_j - D_j, \ j = 1, \ldots, k - 1.$$

The death component is then given as

$$[D_j | m_j, R_j, T_j, M_j, S_j] = Bin(M_j - m_j + R_j - T_j, 1 - S_j), \ j = 1, \ldots, k - 1.$$

The capture component for the CJS model defined in equation 3.4 must also change because we are now using the sufficient statistics,

$$[X_2 | p, d_2, X_1] = \prod_{j=2}^{k} [m_j | p_j, M_j] \tag{6.3}$$

$$[m_j | p_j, M_j] \propto p_j^{m_j} (1 - p_j)^{M_j - m_j}$$

A pleasant result is that writing the model in terms of the sufficient statistics not only improves the speed of the model, but also vastly improve mixing. The WinBUGS code for this model is given in appendix C.6. The data for this problem are the values $u = (u_1, \ldots, u_k)$, $m = (m_2, \ldots, m_k)$, $R = (R_1, \ldots, R_{k-1})$, $T = (T_1, \ldots, T_{k-1})$ as well as $wr = (1, \ldots, 1)$ and $wd = (1, \ldots, 1)$ which are used to include equations 6.2 and 6.3 into the model using the ones trick. An example of the data file needed is given in appendix C.7. Note that for this model, the initial values must be carefully specified

to ensure that all quantities in the model are in range.

Chapter 7

Capture-Recapture Models

Using Gibbs Sampling

Here we present how to fit a selection of standard capture-recapture models where any missing data is included using data augmentation. For the remainder of the thesis, all models presented are fitted using special purpose programs written in MATLAB using Gibbs sampling within a Bayesian framework. The Gibbs sampler requires a sample from the full conditional distribution of every unknown in every iteration. We directly sample from the full conditional distribution is we are able to, otherwise if the full conditional distribution is not of known form we use either the Metropolis-Hastings updater (MH) or use reversible jump Markov chain Monte Carlo (RJMCMC) as described in section 5.2.2. In general we do not explicitly include the iteration superscript on the unknowns. Unless otherwise specified we use the latest value for the unknown.

7.1 Model M_h

As specified in section 3.1, the CDL for model M_h with the capture probabilities drawn from a common distribution is

$$\mathcal{L}_C(\mu, \sigma^2, \aleph; \boldsymbol{X}, \boldsymbol{p}) \propto [\boldsymbol{X}|\boldsymbol{p}, \aleph][\boldsymbol{p}|\mu, \sigma^2, \aleph].$$

The DAG for this model is in 3.1. The capture component is

$$[\boldsymbol{X}|\boldsymbol{p}, \aleph] \propto \frac{\aleph!}{u.!(\aleph - u.)!} \prod_{i=1}^{\aleph} \prod_{j=1}^{k} p_i^{X_{ij}} (1 - p_i)^{1 - X_{ij}}. \tag{7.1}$$

The capture probabilities are specified to be samples from a beta distribution,

$$[\boldsymbol{p}|\mu, \sigma^2, \aleph] = \prod_{i=1}^{\aleph} [p_i|\mu, \sigma^2]$$

$$[p_i|\mu, \sigma^2] = Be(\alpha, \beta), \tag{7.2}$$

where we re-parameterize the beta distribution in terms of the mean μ and variance σ^2 to improve the sampling correlation,

$$\mu = \frac{\alpha}{\alpha + \beta}$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

The back transformation is

$$\alpha = \mu\psi$$

$$\beta = (1 - \mu)\psi \tag{7.3}$$

where

$$\psi = \frac{\mu(1 - \mu)}{\sigma^2} - 1.$$

The model specification is completed with prior distributions. The posterior distribution is

$$[\mu, \sigma^2, \boldsymbol{p}, \aleph | \boldsymbol{X}^{obs}] \propto [\boldsymbol{X} | \boldsymbol{p}, \aleph][\boldsymbol{p} | \mu, \sigma^2, \aleph][\sigma^2 | \mu][\mu][\aleph]$$

The prior distribution for σ^2 is conditional on μ in order to ensure that the values of μ and σ^2 are permissible, giving $\alpha > 0$ and $\beta > 0$. In particular, the prior distribution for σ^2 needs to ensure that $\sigma^2 < \mu(1 - \mu)$.

7.1.1   Full Conditional Distributions

A summary of the unknowns, their prior distributions and the updater used to sample from the full conditional distribution is given in table 7.1.

Unknown	Prior Distribution	Updater
μ	$Be(a, b)$	MH
σ^2	$U(0, \mu(1 - \mu))$	MH
$\aleph$	$DU(0, A)$	RJMCMC
p_i	–	Beta

Table 7.1: Table summarizing the Gibbs sampler for model M_h.

Full Conditional Distribution $[\mu | \cdot]$

The full conditional distribution for μ is

$$[\mu | \cdot] \propto [\boldsymbol{p} | \mu, \sigma^2, \aleph][\sigma^2 | \mu][\mu].$$

The prior distributions $[\mu]$ and $[\tau|\mu]$ are specified in table 7.1 and the distribution $[p|\mu,\sigma^2,\aleph]$ is specified in equation 7.2. This means

$$[\mu|\cdot] \propto \prod_{i=1}^{\aleph} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{\alpha-1} \mu^{a-1}(1-\mu)^{b-1} I\left(0 < \sigma^2 < \mu(1-\mu)\right)$$

where α and β are given in equation 7.3.

Full Conditional Distribution $[\sigma^2|\cdot]$

The full conditional distribution for σ^2 is

$$[\sigma^2|\cdot] \propto [p|\mu,\sigma^2,\aleph][\sigma^2|\mu].$$

The prior distribution $[\sigma^2]$ is specified in table 7.1 and the component $[p|\mu,\sigma^2,\aleph]$ is specified in equation 7.2. This gives

$$[\sigma^2|.] \propto \prod_{i=1}^{\aleph} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{\alpha-1} I\left(0 < \sigma^2 < \mu(1-\mu)\right)$$

where α and β are given in equation 7.3.

Full Conditional Distribution $[\aleph|\cdot]$

The full conditional distribution for $\aleph$ is

$$[\aleph|.] \propto [X|p,\aleph][p|\mu,\sigma^2,\aleph][\aleph].$$

The prior distribution $[\aleph]$ is specified in table 7.1, the distribution $[X|p,\aleph]$ is specified in equation 7.1 and the distribution $[p|\mu,\sigma^2,\aleph]$ is specified in equation 7.2. As $\aleph$ is discrete and defines the dimension of X and p it must be updated using RJMCMC (section 5.2.2). In iteration m we propose a candidate $\aleph^*$ from the jumping distribution $J(\aleph^*|\aleph^{(m)})$. Supposing that

$\aleph^\star > \aleph^{(m)}$ we must also generate two augmenting variables $\boldsymbol{u}_p$ and $\boldsymbol{u}_X$ so that both $\boldsymbol{p}^{(m)\star}$ and $\boldsymbol{X}^\star$ are of dimension $\aleph^\star$,

$$\boldsymbol{u}_p \sim H_1(\boldsymbol{u}|\aleph^{(m)}, \aleph^\star, \mu, \sigma^2)$$

$$\boldsymbol{u}_X \sim H_2(\boldsymbol{u}|\aleph^{(m)}, \aleph^\star)$$

We specify the bijection to be the identity,

$$\boldsymbol{p}^{(m)\star} = \begin{pmatrix} \boldsymbol{p}^{(m)} \\ \boldsymbol{u}_p \end{pmatrix}, \quad \boldsymbol{X}^\star = \begin{pmatrix} \boldsymbol{X}^{(m)} \\ \boldsymbol{u}_X \end{pmatrix}.$$

We choose $H_2(\cdot)$ to be a function yielding a $(\aleph^\star - \aleph^{(m)}) \times k$ matrix of zeros with probability 1 as we know that any additional individuals were not caught. We choose $H_1(\cdot)$ to be a function generating $(\aleph^\star - \aleph^{(m)})$ values from the $Be(\alpha, \beta)$ random effects distribution for $\boldsymbol{p}$. We then accept the candidate $\aleph^\star$ with probability $q' = \min(1, q)$, where

$$q = \frac{[\aleph^\star|.]J(\aleph^{(m)}|\aleph^\star)}{[\aleph^{(m)}|.]H_1(\boldsymbol{u}_p|\aleph^{(m)}, \aleph^\star, \mu, \sigma^2)J(\aleph^\star|\aleph^{(m)})}$$

$$= \frac{\aleph^\star(\aleph^{(m)} - u.)}{(\aleph^\star - u.)\aleph^{(m)}} \frac{\prod_{i=\aleph^{(m)}+1}^{\aleph^\star}\left(1 - p_i^{(m)\star}\right)^k J(\aleph^{(m)}|\aleph^\star)}{J(\aleph^\star|\aleph^{(m)})} I(u. \leq \aleph^\star \leq A).$$

All other terms cancel out of the expression because we were able to write the augmenting variable $\boldsymbol{u}_p$ in terms of the random effects distribution for $\boldsymbol{p}$. For the case when $\aleph^\star < \aleph^{(m)}$ we delete the last $(\aleph^{(m)} - \aleph^\star)$ rows of the $\boldsymbol{X}$ and $\boldsymbol{p}$ matrices and q becomes

$$q = \frac{\aleph^\star(\aleph^{(m)} - u.)}{(\aleph^\star - u.)\aleph^{(m)}} \frac{J(\aleph^{(m)}|\aleph^\star)}{\prod_{i=\aleph^\star+1}^{\aleph^{(m)}}\left(1 - p_i^{(m)}\right)^k J(\aleph^\star|\aleph^{(m)})} I(u. \leq \aleph^\star \leq A).$$

If accepted $\aleph^{(m+1)} = \aleph^\star$, $\boldsymbol{X}^{(m+1)} = \boldsymbol{X}^\star$ and $\boldsymbol{p}^{(m)} = \boldsymbol{p}^{(m)\star}$, otherwise we set $\aleph^{(m+1)} = \aleph^{(m)}$, $\boldsymbol{X}^{(m+1)} = \boldsymbol{X}^{(m)}$ and $\boldsymbol{p}^{(m)}$ remains unchanged.

Full Conditional Distribution $[p_i|\cdot]$

The full conditional distribution for p_i is

$$[p_i|\cdot] \propto [\boldsymbol{X}|\boldsymbol{p},\aleph][\boldsymbol{p}|\mu,\tau,\aleph], \ i = 1,\ldots,\aleph.$$

The components $[\boldsymbol{X}|\boldsymbol{p},\aleph]$ and $[\boldsymbol{p}|\mu,\tau,\aleph]$ are specified in equations 7.1 and 7.2. This gives

$$[p_i|\cdot] = Be(a',b'), \ i = 1,\ldots,\aleph,$$

where

$$a' = \sum_{j=1}^{k} X_{ij} + \alpha$$

$$b' = k - \sum_{j=1}^{k} X_{ij} + \beta,$$

where α and β are given in equation 7.3.

7.1.2 Example: Taxi Cabs in Edinburgh, Scotland

Carothers (1973) conducted a capture-recapture experiment in Edinburgh where any taxi cab that passed a preselected point in the city was recorded as a capture. Sampling occurred over $k = 10$ days with a different selection point on each day. The population was assumed closed with a known population size of $\aleph = 420$. We compare the results of the model fitted assuming the capture probabilities are drawn from a common beta distribution to the standard procedure in Otis et al. (1978) that uses the jackknife estimator (Burnham and Overton 1978).

The following prior distributions were used to fit the model:

$$\aleph \sim DU(0, 100000)$$

$$\mu \sim Be(1, 1)$$

$$\sigma^2 \sim U\left(0, \mu(1 - \mu)\right).$$

The Gibbs sampler was run on 3 independent chains started from different values. Each chain had an adaptive phase of $10,000$ iterations to optimize the Metropolis-Hastings and RJMCMC algorithm as specified in section 5.2.2. Each chain was then run for a further $100,000$ iterations. The chains mixed well and were combined to give a posterior sample of $300,000$ iterations.

There is a large difference in the results between the two models (Table 7.2). Not only does the median of $\aleph$ in the CDL differ from the MLE of $\aleph$ found using the jackknife, but the width of the corresponding confidence/credible interval also differs between the two models. The lower interval estimate of 418 taxi cabs in Otis et al. (1978) is higher than the median of the posterior distribution for $\aleph$ found using the CDL. The difference in the estimates is caused because the jackknife procedure yields a different model for the individual capture probabilities than was used in the CDL. Unfortunately, as model M_h is non-identifiable (Link 2003) as mentioned in section 1.1.1, there is no way to distinguish between these two different models. The estimated posterior densities for $\aleph$, μ and σ are given in figure 7.1.

Method	Estimate	2.5%	97.5%
Jackknife	471	418	524
CDL	408	358	525
True	420		

Table 7.2: Table comparing the estimates of N under model M_h for the taxi data of Carothers (1973) using (i) the jackknife (Burnham and Overton 1978) and (ii) the CDL assuming that capture probability are drawn from a beta distribution. The CDL estimate is the posterior median and the 2.5% and 97.5% quantiles are the equal sided credible interval values.
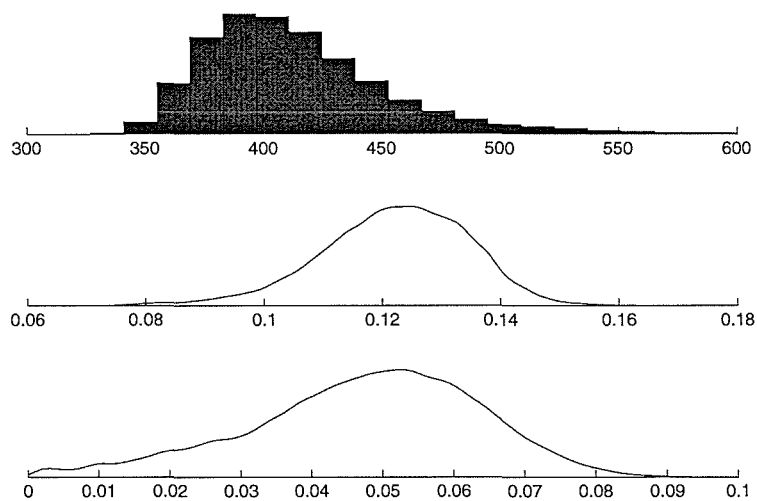
Figure 7.1: Estimates of the posterior distribution of N (top), μ (middle) and σ (bottom).

7.2   Model M_h with Individual-Specific Covariates

An extension to model M_h described in section 3.1.1 is where there are individual-specific covariates z that fully explain differences in individual capture probabilities. The CDL is

$$\mathcal{L}_C(\mu, \tau, \gamma, \aleph; X, z) \propto [X|z, \gamma, \aleph][z|\mu, \tau, \aleph].$$

The DAG for this model is in figure 3.2. We specify a logistic relationship between p and z,

$$\text{logit}(p_i) = \gamma_0 + \gamma_1 z_i,$$

so that the capture component $[X|z, \gamma, \aleph]$ is

$$[X|z, \gamma, \aleph] \propto \frac{\aleph!}{u.!(\aleph - u.)!} \prod_{i=1}^{\aleph} \prod_{j=1}^{k} \text{expit}(\gamma_0 + \gamma_1 z_i)^{X_{ij}} (1 - \text{expit}(\gamma_0 + \gamma_1 z_i))^{1 - X_{ij}}.$$

$$(7.4)$$

The covariate values are modeled as a sample from a normal distribution,

$$[z|\mu, \tau, \aleph] = \prod_{i=1}^{\aleph} [z_i|\mu, \tau]$$

$$(7.5)$$

$$[z_i|\mu, \tau] = N(\mu, \tau).$$

The model specification is completed with prior distributions. The posterior distribution is

$$[\gamma, \aleph, \mu, \tau | X^{obs}] \propto [X|z, \gamma, \aleph][z|\mu, \tau, \aleph] \prod_{h=1}^{2} [\gamma_h][\aleph][\mu][\tau].$$

7.2.1   Full Conditional Distributions

A summary of the unknowns, their prior distributions and the updater used to sample from the full conditional distribution is given in table 7.3.

Unknown	Prior Distribution	Updater
$\aleph$	$DU(0, A)$	RJMCMC
z_i^{mis}	–	MH
γ_j	$N(\mu_0, \tau_0)$	MH
τ	$Ga(a, b)$	Gamma
μ	$N(\mu_0, \tau_0)$	Normal

Table 7.3: Table summarizing the Gibbs sampler for model M_h with individual-specific covariates.

Full Conditional Distribution $[\aleph|\cdot]$

The full conditional distribution for $\aleph$ is

$$[\aleph|\cdot] \propto [\boldsymbol{X}|\boldsymbol{z}, \boldsymbol{\gamma}, \aleph][\boldsymbol{z}|\mu, \tau, \aleph][\aleph].$$

The prior distribution $[\aleph]$ is specified in table 7.3 and the distributions $[\boldsymbol{X}|\boldsymbol{z}, \boldsymbol{\gamma}, \aleph]$ and $[\boldsymbol{z}|\mu, \tau, \aleph]$ are specified in equations 7.4 and 7.5. As $\aleph$ is discrete and defines the dimension of $\boldsymbol{X}$ and $\boldsymbol{z}$ it must be updated using RJMCMC (section 5.2.2). In iteration m we propose a candidate $\aleph^\star$ from the jumping distribution $J(\aleph^\star|\aleph^{(m)})$. Supposing that $\aleph^\star > \aleph^{(m)}$ we must also generate two augmenting variables $\boldsymbol{u}_z$ and $\boldsymbol{u}_X$ so that both $\boldsymbol{z}^{(m)\star}$ and $\boldsymbol{X}^\star$ are of the dimension $\aleph^\star$,

$$\boldsymbol{u}_z \sim H_1(\boldsymbol{u}|\aleph^{(m)}, \aleph^\star, \mu, \tau)$$

$$\boldsymbol{u}_X \sim H_2(\boldsymbol{u}|\aleph^{(m)}, \aleph^\star)$$

We specify the bijection to be the identity,

$$\boldsymbol{z}^{(m)\star} = \begin{pmatrix} \boldsymbol{z}^{(m)} \\ \boldsymbol{u}_z \end{pmatrix}, \quad \boldsymbol{X}^\star = \begin{pmatrix} \boldsymbol{X}^{(m)} \\ \boldsymbol{u}_X \end{pmatrix}.$$

We choose $H_2(\cdot)$ to be a function yielding a $(\aleph^\star - \aleph^{(m)}) \times k$ matrix of zeros with probability 1 as we know that any additional individuals were not

caught. We choose $H_1(\cdot)$ to be a function generating $(\aleph^\star - \aleph^{(m)})$ values from the $N(\mu, \tau)$ random effects distribution for z. The candidate $\aleph^\star$ is accepted with probability $q' = \min(1, q)$, where

$$
q = \frac{[\aleph^\star|.] J(\aleph^{(m)}|\aleph^\star)}{[\aleph^{(m)}|.] H_1(u_z|\aleph^{(m)}, \aleph^\star, \mu, \tau) J(\aleph^\star|\aleph^{(m)})}
$$

$$
= \frac{\aleph^\star(\aleph^{(m)} - u.)}{(\aleph^\star - u.)\aleph^{(m)}} \frac{\prod_{i=\aleph^{(m)}+1}^{\aleph^\star} \left(1 - \text{expit}\left(\gamma_0 + \gamma_1 z_i^{(m)\star}\right)\right)^k J(\aleph^{(m)}|\aleph^\star)}{J(\aleph^\star|\aleph^{(m)})} I(u. \le \aleph^\star \le A).
$$

All other terms cancel out of the expression because we were able to write the augmenting variable u_z in terms of the random effects distribution for z. For the case when $\aleph^\star < \aleph^{(m)}$ we delete the last $(\aleph^{(m)} - \aleph^\star)$ rows of X and z with q becoming

$$
q = \frac{\aleph^\star(\aleph^{(m)} - u.)}{(\aleph^\star - u.)\aleph^{(m)}} \frac{J(\aleph^{(m)}|\aleph^\star)}{\prod_{i=\aleph^\star+1}^{\aleph^{(m)}} \left(1 - \text{expit}\left(\gamma_0 + \gamma_1 z_i^{(m)}\right)\right)^k J(\aleph^\star|\aleph^{(m)})} I(u. \le \aleph^\star \le A).
$$

If accepted $\aleph^{(m+1)} = \aleph^\star$, $X^{(m+1)} = X^\star$ and $z^{(m)} = z^{(m)\star}$, otherwise we set $\aleph^{(m+1)} = \aleph^{(m)}$, $X^{(m+1)} = X^{(m)}$ and $z^{(m)}$ remains unchanged.

Full Conditional Distribution $[z_i^{mis}|\cdot]$

The full conditional distribution for z_i^{mis}, $i = u. + 1, \ldots, \aleph$ is

$$
[z_i|\cdot] \propto [X|\gamma, z, \aleph][z|\mu, \tau, \aleph].
$$

The components $[X|\gamma, z, \aleph]$ and $[z|\mu, \tau, \aleph]$ are specified in equations 7.4 and 7.5. This gives

$$
[z_i^{mis}|\cdot] \propto \text{expit}(\gamma_0 + \gamma_1 z_i)^{\sum_{j=1}^k X_{ij}} (1 - \text{expit}(\gamma_0 + \gamma_1 z_i))^{k - \sum_{j=1}^k X_{ij}}
$$

$$
\times \exp\left(-\frac{\tau}{2}(z_i - \mu)^2\right), \quad i = u. + 1, \ldots, \aleph.
$$

Full Conditional Distribution $[\gamma_j|\cdot]$

The full conditional distribution for γ_j, $j = 0, 1$ is

$$[\gamma_j|\cdot] \propto [X|\gamma, z, \aleph][\gamma_j].$$

The prior distribution $[\gamma_j]$ is specified in table 7.3 and the distribution $[X|\gamma, z, \aleph]$ is specified in equation 7.4. This gives

$$[\gamma_j|\cdot] \propto \prod_{i=1}^{\aleph} \left(\text{expit}(\gamma_0 + \gamma_1 z_i)^{\sum_{j=1}^{k} X_{ij}} (1 - \text{expit}(\gamma_0 + \gamma_1 z_i))^{k - \sum_{j=1}^{k} X_{ij}} \right)$$

$$\times \exp\left(-\frac{\tau_0}{2}(\gamma_j - \mu_0)^2 \right), \ j = 0, 1.$$

Full Conditional Distribution $[\tau|\cdot]$

The full conditional distribution for τ is

$$[\tau|\cdot] \propto [z|\mu, \tau, \aleph][\tau].$$

The prior distribution $[\tau]$ is specified in table 7.3 and the distribution $[z|\mu, \tau, \aleph]$ is specified in equation 7.5. This gives

$$[\tau|\cdot] = Ga\left(\frac{\aleph}{2} + a, \frac{\sum_{i=1}^{\aleph}(z_i - \mu)^2}{2} + b \right)$$

Full Conditional Distribution $[\mu|\cdot]$

The full conditional distribution for μ is

$$[\mu|\cdot] \propto [z|\mu, \tau, \aleph][\mu].$$

The prior distribution $[\mu]$ is specified in table 7.3 and the component $[\boldsymbol{z}|\mu, \tau, \aleph]$ is specified in equation 7.5. This gives

$$[\mu|.] = N(a, b)$$

where b is a precision and

$$a = \frac{\tau \sum_{i=1}^{\aleph} z_i + \tau_0 \mu_0}{\aleph \tau + \tau_0}$$

$$b = \aleph \tau + \tau_0.$$

7.2.2   Example: Simulated Data

Data were simulated under the model for a $k = 10$ period study using:

$$\aleph = 100$$

$$\gamma = (-2, 1)$$

$$z_i \sim N(0, 1), \; i = 1, \dots, \aleph$$

The capture histories were stochastically simulated as

$$X_{ij} \sim Bern(\text{expit}(\gamma_0 + \gamma_1 z_i)), \; i = 1, \dots, \aleph, \; j = 1, \dots, k.$$

The resulting data set used for the analysis has $u. = 76$. As expected, the observed covariate values $\boldsymbol{z}^{obs}$ give a poor estimate of μ and σ because we are more likely to see individuals with high z_i values. The sample mean and standard deviation are 0.27 and 0.79 respectively.

We compare the results of the model that assumes the covariate values are drawn from a common normal distribution to the standard procedure in MARK in which we condition on the $u.$ individuals that were caught

(Huggins 1989).

The following prior distributions were used to fit the model:

$$\aleph \sim DU(0, 100000)$$

$$\gamma_j \sim N(0, 0.0001), \ j = 0, 1$$

$$\mu \sim N(0, 0.0001)$$

$$\tau \sim Ga(0.001, 0.001).$$

The Gibbs sampler was run on 3 independent chains started from different values. Each chain had an adaptive phase of $10,000$ iterations to optimize the Metropolis-Hastings and RJMCMC algorithms as specified in section 5.2.2. Each chain was then run for a further $100,000$ iterations. The chains mixed well and were combined to give a posterior sample of $300,000$ iterations.

There is very good agreement between the two methods used (Table 7.4). The only difference appears to be that the posterior distribution for $\aleph$ is skewed further to the right. It is reassuring to see that the estimates of μ and σ found including the missing individuals automatically adjust for the systematic biases that occur when estimating μ and σ from only the observed individuals. The point estimates for (μ, σ) have changed from $(0.27, 0.79)$ with only the observed individuals to $(0.00, 0.87)$ when including all individuals.

The posterior distribution for the population size $\aleph$ fitted using the CDL is in figure 7.2.

In general, caution should be used when placing non-informative priors on parameters that are on the logit scale, such as γ_0. A normal prior distribution with low precision on the $(-\infty, \infty)$ logit scale results is a bathtub shaped prior distribution on the original $[0, 1]$ scale with most mass near

	CDL Posterior			Conditional Likelihood		
Parameter	2.5%	Median	97.5%	2.5%	Mode	97.5%
$\aleph$	90	109	158	90	107	145
γ_0	-2.33	-1.98	-1.67	-2.25	-1.94	-1.62
γ_1	0.61	0.94	1.3	0.57	0.90	1.23
μ	-0.41	0.00	0.26			
σ	0.73	0.87	1.10			

Table 7.4: A comparison of parameter estimates for the model M_h with individual specific covariates using simulated data when (i) assuming the covariates are drawn from a common normal distribution and (ii) using the conditional likelihood approach of Huggins (1989).
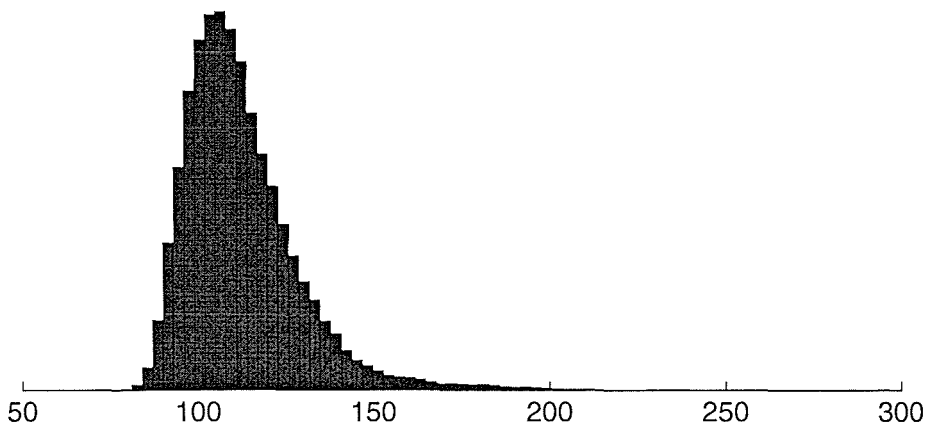
Figure 7.2: Posterior density estimate for the population size $\aleph$ for model M_h with individual-specific covariates that are assumed to be drawn from a common normal distribution. The data were simulated with a true population size of 100.

0 and 1. This prior is non-informative so long as the likelihood has very little mass near the boundaries. However, if there is reasonable mass in the likelihood close to either 0 or 1, then the prior can be unintentionally informative. One possible alternative is to use a logistic prior distribution with mean 0 and scale 1. This distribution on the $(-\infty, \infty)$ scale is equivalent to a beta distribution with $\alpha = 1$ and $\beta = 1$ on a $[0, 1]$ scale.

Potential improvements in the efficiency of the MCMC algorithm can be obtained from examining different link functions for p, such as the probit link. The use of different link functions may allow us to use more efficient sampling algorithms for some/all of the parameters in the model, improving the convergence of the Gibbs sampler. For example, in certain situations, data augmentation can be used to obtain known full conditional distributions for the parameters in probit regression (Gelman et al. 2004).

7.3 The CJS Model

The CDL for the time specific CJS model denoted $p(t)S(t)$ is specified in equation 3.2 as

$$\mathcal{L}_C(p, S; X_2, d_2, X_1) \propto [X_2|p, d_2, X_1][d_2|S, X_1],$$

where $[X_2|p, d_2, X_1]$ is specified in equation 3.4 and $[d_2|S, X_1]$ is specified in equation 3.3 (with no loss on capture so that $\iota_i = 0$, $\forall i$). The model specification is completed with prior distributions for the capture and survival probabilities. The posterior distribution is

$$[p, S, d_2^{mis}|X_2, d_2^{obs}, X_1] \propto [X_2|p, d_2, X_1][d_2|S, X_1] \prod_{h=2}^{k} [p_h] \prod_{h=1}^{k-1} [S_h].$$

7.3.1 Full Conditional Distributions

A summary of the unknowns, their prior distributions and the updater used to sample from the full conditional distribution is given in table 7.5.

Unknown	Prior Distribution	Updater
S_j	$Be(\alpha, \beta)$	Beta
p_j	$Be(\alpha, \beta)$	Beta
$d_{2i:}^{mis}$	–	Multinomial

Table 7.5: Table summarizing the Gibbs sampler for the CJS model $p(t)S(t)$.

Full Conditional Distribution $[S_j|\cdot]$

The full conditional distribution for S_j, $j = 1, \ldots, k-1$ is

$$[S_j|\cdot] \propto [d_2|S, X_1][S_j].$$

The prior distribution $[S_j]$ is specified in table 7.5 and the component $[d_2|S, X_1]$ is specified in equation 3.3. This gives

$$[S_j|\cdot] = Be(M_j^+ - D_j + \alpha, D_j + \gamma), \; j = 1, \ldots, k-1,$$

where D_j is the number of marked individuals that died between sample j and $j+1$ and M_j^+ is the number of marked individuals in the population immediately after sample j,

$$D_j = \sum_{i=1}^{u.} d_{2ij}$$

$$M_j^+ = \sum_{i=1}^{u.} \left\{ \sum_{h=1}^{j} X_{1ih} - \sum_{h=1}^{j-1} d_{2ih} \right\}.$$

Full Conditional Distribution $[p_j|\cdot]$

The full conditional distribution for p_j, $j = 2, \ldots, k$ is

$$[p_j|\cdot] \propto [X_2|p, d_2, X_1][p_j].$$

The prior distribution $[p_j]$ is specified in table 7.5 and the distribution $[X_2|p, d_2, X_1]$ is specified in equation 3.4. This gives

$$[p_j|\cdot] = Be(m_j + \alpha, M_j - m_j + \gamma), \ j = 2, \ldots, k,$$

where m_j is the number of marked individuals caught in sample j and M_j is the number of marked individuals in the population immediately before sample j,

$$m_j = \sum_{i=1}^{u.} X_{2ij}$$

$$M_j = \sum_{i=1}^{u.} \sum_{h=1}^{j-1} \{X_{1ih} - d_{2ih}\}.$$

Full Conditional Distribution $[d_{2i:}^{mis}|\cdot]$

The full conditional distribution for the missing values in the ith row of d_2 is

$$[d_{2i:}^{mis}|\cdot] \propto [X_2|p, d_2, X_1][d_2|S, X_1].$$

The components $[X_2|p, d_2, X_1]$ and $[d_2|S, X_1]$ are specified in equations 3.4 and 3.3. This gives a multinomial distribution,

$$[d_{2i:}^{mis}|\cdot] = MN(1, \zeta_i), \ i = 1, \ldots, u.$$

where $\zeta_i = (\zeta_{i\mathcal{L}_i}, \ldots, \zeta_{ik})$ and

$$\zeta_{ij} = \frac{\kappa_{ij}}{\sum_{h=\mathcal{L}_i}^{k} \kappa_{ih}}, \ j = \mathcal{L}_i, \ldots, k$$

$$\kappa_{ij} = \begin{cases} \prod_{h=\mathcal{L}_i}^{j-1} S_h(1 - p_{h+1})(1 - S_j), \ j = \mathcal{L}_i, \ldots, k-1 \\ \prod_{h=\mathcal{L}_i}^{j-1} S_h(1 - p_{h+1}), \ j = k \end{cases}$$

7.3.2 Example: European Dipper

We fit the CJS model to European dipper (*Cinclus cinclus*) data from Lebreton et al. (1992). The prior distributions used were

$$S_j \sim Be(1,1), \ j = 1, \ldots, k-1$$

$$p_j \sim Be(1,1), \ j = 2, \ldots, k.$$

The study consists of 7 periods with $u. = 294$ unique individuals observed. The Gibbs sampler was run on 3 independent chains started from different values. Each chain had a burn-in of $10,000$ followed by a further $100,000$ iterations. The chains mixed well and were combined to give a posterior sample of $300,000$ iterations.

The marginal posterior distributions of the identifiable parameters are summarised in table 7.6 and figure 7.3.

A nice feature of the CJS model is that the full conditional distributions are of known form. However, if we were to model either survival or capture probability in terms of covariates or include random effects, then the full conditional distributions would not be guaranteed to be of known form and sampling schemes such as the Metropolis-Hastings algorithm may become necessary.

Parameter	2.5%	Median	97.5%
S_1	0.4628	0.7226	0.9662
S_2	0.3187	0.4474	0.5965
S_3	0.3654	0.4800	0.6018
S_4	0.5095	0.6273	0.7450
S_5	0.4906	0.6019	0.7151
p_2	0.3962	0.6704	0.9054
p_3	0.6715	0.8822	0.9818
p_4	0.7321	0.8889	0.9730
p_5	0.7453	0.8830	0.9633
p_6	0.7826	0.9118	0.9787

Table 7.6: Posterior 2.5%, 50% and 97.5% quantile estimates for the survival and capture probabilities for the European dipper using a $S(t)p(t)$ CJS model.
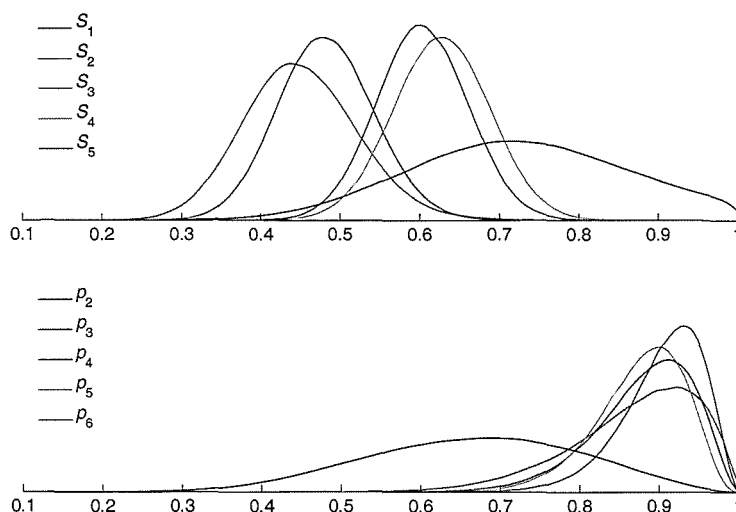
Figure 7.3: Estimates of the posterior distribution for the survival probabilities (top) and capture probabilities (bottom) for the European dipper using a $S(t)p(t)$ CJS model.

7.4   The CMAS Model

The CDL for the time specific CMAS model denoted $p(t)S(t)\beta(t)$ is specified in equation 3.2 as

$$\mathcal{L}_C(S, p, \beta, \aleph; X, b, d) \propto [X|p, d, b, \aleph][d|S, b, \aleph][b|\beta, \aleph]$$

where the three distributions are specified in equations 3.9, 3.10 and 3.11. The model specification is completed with prior distributions for all parameters. The posterior distribution is

$$[p, S, \beta, b^{mis}, d^{mis}\aleph|X^{obs}, b^{obs}, d^{obs}]$$

$$\propto [X|p, b, d, \aleph][d|S, b, \aleph][b|\beta, \aleph] \prod_{h=1}^{k} [p_h] \prod_{h=1}^{k-1} [S_h][\beta][\aleph].$$

Even though the model is specified in terms of β we are able to examine samples from the posterior distribution for $\eta = (\eta_1, \ldots, \eta_{k-1})$ through transforming the samples from the posterior distribution of β, b, d and $\aleph$,

$$\eta_j = \frac{\beta_j \aleph}{N_j}, \ j = 1, \ldots, k - 1.$$

For this reason, we also refer to this model as $p(t)S(t)\eta(t)$.

7.4.1   Full Conditional Distributions

A summary of the unknowns, their prior distributions and the updater used to sample from the full conditional distribution is given in table 7.7.

Unknown	Prior Distribution	Updater
$\aleph$	$DU(0, A)$	RJMCMC
p_j	$Be(\alpha, \gamma)$	Beta
S_j	$Be(\alpha, \gamma)$	Beta
β	$Dir(v)$	Dirichlet
$b_{i:}^{mis}$	–	Multinomial
$d_{i:}^{mis}$	–	Multinomial

Table 7.7: Table summarizing the Gibbs sampler for the CJS model $p(t)S(t)$.

Full Conditional Distribution $[\aleph|\cdot]$

The full conditional distribution for $\aleph$ is

$$[\aleph|\cdot] \propto [X|p, b, d, \aleph][d|S, b, \aleph][b|\beta, \aleph][\aleph].$$

The prior distribution $[\aleph]$ is specified in table 7.7 and the other components are specified in equations 3.9, 3.10 and 3.11. As $\aleph$ is discrete and defines the dimension of X, d and b it must be updated using RJMCMC (section 5.2.2). In iteration m we propose a candidate $\aleph^*$ from the jumping distribution $J(\aleph^*|\aleph^{(m)})$. Supposing that $\aleph^* > \aleph^{(m)}$ we must also generate three augmenting variables u_b, u_d and u_X so that both $b^{(m)*}$, $d^{(m)*}$ and X^* are of the dimension $\aleph^*$,

$$u_b \sim H_1(u|\aleph^{(m)}, \aleph^*, \beta)$$

$$u_d \sim H_2(u|\aleph^{(m)}, \aleph^*, S, u_b)$$

$$u_X \sim H_3(u|\aleph^{(m)}, \aleph^*)$$

We specify the bijection to be the identity,

$$b^{(m)*} = \begin{pmatrix} b^{(m)} \\ u_b \end{pmatrix}, \quad d^{(m)*} = \begin{pmatrix} d^{(m)} \\ u_d \end{pmatrix}, \quad X^* = \begin{pmatrix} X^{(m)} \\ u_X \end{pmatrix}.$$

We choose $H_3(\cdot)$ to be a function yielding a $(\aleph^\star - \aleph^{(m)}) \times k$ matrix of zeros with probability 1 as we know that any additional individuals were not caught. We choose $H_1(\cdot)$ to be a function generating $(\aleph^\star - \aleph^{(m)})$ vectors from the birth distribution $MN(1, \boldsymbol{\beta})$ specified in equation 3.11. We choose $H_2(\cdot)$ to be a function generating $(\aleph^\star - \aleph^{(m)})$ vectors according to the mortality distribution $MN(1, \boldsymbol{\xi}_i)$ specified in equation 3.10. We then accept the candidate $\aleph^\star$ with probability $q' = \min(1, q)$, where

$$
\begin{aligned}
q &= \frac{[\aleph^\star | \cdot] J(\aleph^{(m)} | \aleph^\star)}{[\aleph^{(m)} | \cdot] H_1(\boldsymbol{u}_b | \aleph^{(m)}, \aleph^\star, \boldsymbol{\beta}) H_2(\boldsymbol{u}_d | \aleph^{(m)}, \aleph^\star, \boldsymbol{S}, \boldsymbol{u}_b) J(\aleph^\star | \aleph^{(m)})} \\
&= \frac{\aleph^\star (\aleph^{(m)} - u.)}{(\aleph^\star - u.) \aleph^{(m)}} \frac{\prod_{i=\aleph^{(m)}+1}^{\aleph^\star} \prod_{j=\mathfrak{B}_i+1}^{\mathfrak{D}_i} (1 - p_j) J(\aleph^{(m)} | \aleph^\star)}{J(\aleph^\star | \aleph^{(m)})} I(u. \leq \aleph^\star \leq A).
\end{aligned}
$$

All other terms cancel out of the expression because we were able to write the augmenting variables $\boldsymbol{u}_b$ and $\boldsymbol{u}_d$ in terms of the conditional likelihood components of $\boldsymbol{b}$ and $\boldsymbol{d}$. For the case when $\aleph^\star < \aleph^{(m)}$ we delete the last $(\aleph^{(m)} - \aleph^\star)$ rows of the $\boldsymbol{X}$, $\boldsymbol{d}$ and $\boldsymbol{b}$ matrices and q becomes

$$
q = \frac{\aleph^\star (\aleph^{(m)} - u.)}{(\aleph^\star - u.) \aleph^{(m)}} \frac{J(\aleph^{(m)} | \aleph^\star)}{\prod_{i=\aleph^{(m)}+1}^{\aleph^\star} \prod_{j=\mathfrak{B}_i+1}^{\mathfrak{D}_i} (1 - p_j) J(\aleph^\star | \aleph^{(m)})} I(u. \leq \aleph^\star \leq A).
$$

If accepted $\aleph^{(m+1)} = \aleph^\star$, $\boldsymbol{X}^{(m+1)} = \boldsymbol{X}^\star$ and $\boldsymbol{z}^{(m)} = \boldsymbol{z}^{(m)\star}$, otherwise we set $\aleph^{(m+1)} = \aleph^{(m)}$, $\boldsymbol{X}^{(m+1)} = \boldsymbol{X}^{(m)}$ and $\boldsymbol{z}^{(m)}$ remains unchanged.

Full Conditional Distribution $[p_j | \cdot]$

The full conditional distribution for p_j, $j = 1, \ldots, k$ is

$$
[p_j | \cdot] \propto [\boldsymbol{X} | \boldsymbol{p}, \boldsymbol{b}, \boldsymbol{d}, \aleph][p_j].
$$

The prior distribution $[p_j]$ is specified in table 7.7 and component $[X|p, b, d, \aleph]$ is specified in equation 3.9. This gives

$$[p_j|\cdot] = Be(n_j + \alpha, N_j - n_j + \gamma), \ j = 1, \dots, k,$$

where n_j is the number of individuals caught in sample j and N_j is the number of individuals alive in sample j,

$$
\begin{aligned}
n_j &= \sum_{i=1}^{\aleph} X_{ij} \\
N_j &= \sum_{i=1}^{\aleph} \left(\sum_{h=0}^{j-1} b_{ih} - \sum_{h=1}^{j-1} d_{ih} \right).
\end{aligned}
\tag{7.6}
$$

Full Conditional Distribution $[S_j|\cdot]$

The full conditional distribution for S_j, $j = 1, \dots, k - 1$ is

$$[S_j|\cdot] \propto [d|S, b, \aleph][S_j]$$

The prior distribution $[S_j]$ is specified in table 7.7 and the component $[d|S, b, \aleph]$ is specified in equation 3.10. This gives

$$[S_j|\cdot] = Be(N_j - D_j + \alpha, D_j + \gamma), \ j = 1, \dots, k - 1,$$

where N_j is defined in equation 7.6 and D_j is the number of individuals that died between sample j and $j + 1$,

$$D_j = \sum_{i=1}^{\aleph} d_{ij}.$$

Full Conditional Distribution $[\beta|\cdot]$

The full conditional distribution for β is

$$[\beta|\cdot] \propto [b|\beta, \aleph][\beta]$$

The component $[b|\beta, \aleph]$ is given in equation 3.11 and the prior distribution $[\beta]$ is specified in table 7.7 where $v = (v_0, \ldots, v_{k-1})$. This gives

$$[\beta|\cdot] = Dir\left(B + v\right),$$

where $B = (B_0, \ldots, B_{k-1})$ and B_j is the number of individuals born between sample j and $j + 1$,

$$B_j = \sum_{i=1}^{\aleph} b_{ij}, \ j = 0, \ldots, k - 1.$$

Full Conditional Distribution $[b_{i:}^{mis}|\cdot]$

The full conditional distribution for the missing values in the ith row of b is

$$[b_{i:}^{mis}|\cdot] \propto [\boldsymbol{X}|p, b, d, \aleph][d|S, b, \aleph][b|\beta, \aleph]$$

The components are given in equations 3.9, 3.10 and 3.11. The gives a multinomial distribution,

$$[b_{i:}^{mis}|\cdot] = MN(1, \zeta_i), \ i = 1, \ldots, \aleph$$

where $\zeta_i = (\zeta_{i0}, \ldots, \zeta_{i\mathfrak{F}_i-1})$ and

$$\zeta_{ij} = \frac{\kappa_{ij}}{\sum_{h=0}^{\mathfrak{F}_i-1} \kappa_{ih}}, \ j = 0, \ldots, \mathfrak{F}_i - 1$$

$$\kappa_{ij} = \beta_j \prod_{h=j+1}^{\mathfrak{F}_i-1} S_h(1 - p_h), \ j = 0, \ldots, \mathfrak{F}_i - 1$$

Note that $\mathfrak{F}_i = \mathfrak{D}_i$ for all individuals never caught $i = u. + 1, \ldots, \aleph$.

Full Conditional Distribution $[d_{i:}^{mis}|\cdot]$

The full conditional distribution for the missing values in the ith row of d is,

$$[d_{i:}^{mis}|\cdot] \propto [\boldsymbol{X}|\boldsymbol{p}, \boldsymbol{b}, \boldsymbol{d}, \aleph][\boldsymbol{d}|\boldsymbol{S}, \boldsymbol{b}, \aleph]$$

The components are given in equations 3.9 and 3.10. This gives a multinomial distribution,

$$[\boldsymbol{d}_{i:}^{mis}|\cdot] = MN(1, \boldsymbol{\zeta}_i), \ i = 1, \ldots, u.$$

where $\boldsymbol{\zeta}_i = (\zeta_{i\mathfrak{L}_i}, \ldots, \zeta_{ik})$ and

$$\zeta_{ij} = \frac{\kappa_{ij}}{\sum_{h=\mathfrak{L}_i}^{k} \kappa_{ih}}, \ j = \mathfrak{L}_i, \ldots, k$$

$$\kappa_{ij} = \begin{cases} \displaystyle\prod_{h=\mathfrak{L}_i}^{j-1} S_h(1 - p_{h+1})(1 - S_j), \ j = \mathfrak{L}_i, \ldots, k - 1 \\ \displaystyle\prod_{h=\mathfrak{L}_i}^{j-1} S_h(1 - p_{h+1}), \ j = k \end{cases}$$

Note that $\mathfrak{L}_i = \mathfrak{B}_i + 1$ for all individuals never caught $i = u. + 1, \ldots, \aleph$.

7.4.2   Example: Meadow Vole

We fit the CMAS model to meadow vole (*Microtus pensylvannicus*) data collected at the Patuxent Wildlife Research Center, Laurel, Maryland (Williams et al. 2002). The prior distributions used are

$$S_j \sim Be(1,1), \; j = 1, \ldots, k-1$$

$$p_j \sim Be(1,1), \; j = 1, \ldots, k$$

$$\beta \sim Dir(\mathbf{1})$$

$$\aleph \sim DU(0, 100000)$$

where **1** is a vector of ones of length k. The study consists of 6 periods with $u. = 294$ unique individuals observed. The Gibbs sampler was run on 3 independent chains started from different values. Each chain had an adaptive phase of 10,000 iterations to optimize the RJMCMC algorithm as specified in section 5.2.2. Each chain was then run for a further 100,000 iterations. The chains mixed well and were combined to give a posterior sample of 300,000 iterations.

The marginal posterior distributions of the identifiable parameters are summarized in table 7.8 and figure 7.4. The posterior predictive distributions of population size N_j, birth B_j and death D_j are summarized in table 7.9. It is interesting to note that there is information about $\aleph$, even though it is not identifiable (Link and Barker 2005), as shown by its posterior distribution in figure 7.5.

Parameter	2.5%	Median	97.5%
S_1	0.7862	0.8752	0.9526
S_2	0.5632	0.6595	0.7544
S_3	0.5787	0.6809	0.7777
S_4	0.5174	0.6166	0.7114
p_2	0.7919	0.8918	0.9562
p_3	0.7343	0.8431	0.9218
p_4	0.8212	0.9201	0.9756
p_5	0.7975	0.8952	0.9579
η_2	0.1272	0.2182	0.331
η_3	0.1538	0.2537	0.375
η_4	0.2592	0.3779	0.5209

Table 7.8: Posterior 2.5%, 50% and 97.5% quantile estimates for the per-capita birth rates, survival and capture probabilities for the meadow vole using a $S(t)p(t)\eta(t)$ CMAS model.
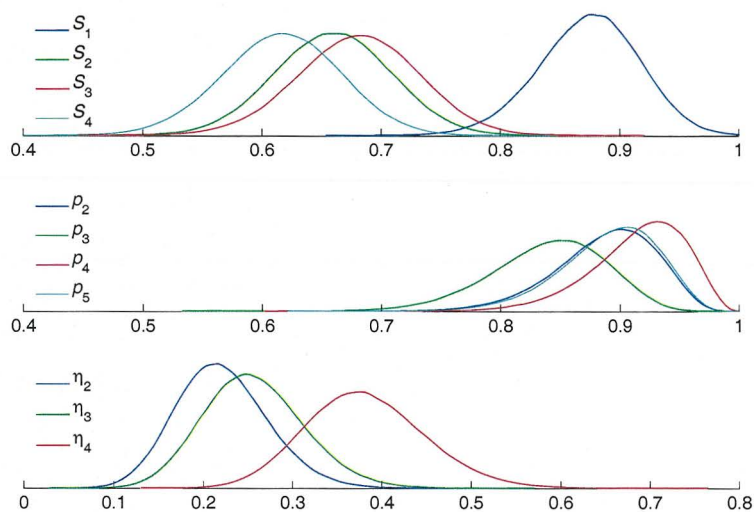
Figure 7.4: Estimates of the posterior distribution for the survival probabilities (top), per-capita birth rates (middle) and capture probabilities (bottom) for the meadow vole using a $S(t)p(t)\eta(t)$ CMAS model.

Quantity	2.5%	Median	97.5%
N_2	125	132	144
N_3	108	115	128
N_4	103	108	118
N_5	103	108	117
B_2	21	29	37
B_3	22	29	36
B_4	35	41	49
D_2	36	45	55
D_3	29	37	46
D_4	36	41	49

Table 7.9: Posterior predictive 2.5%, 50% and 97.5% quantile estimates for N_j, B_j and D_j for the meadow vole using a $S(t)p(t)\eta(t)$ CMAS model.
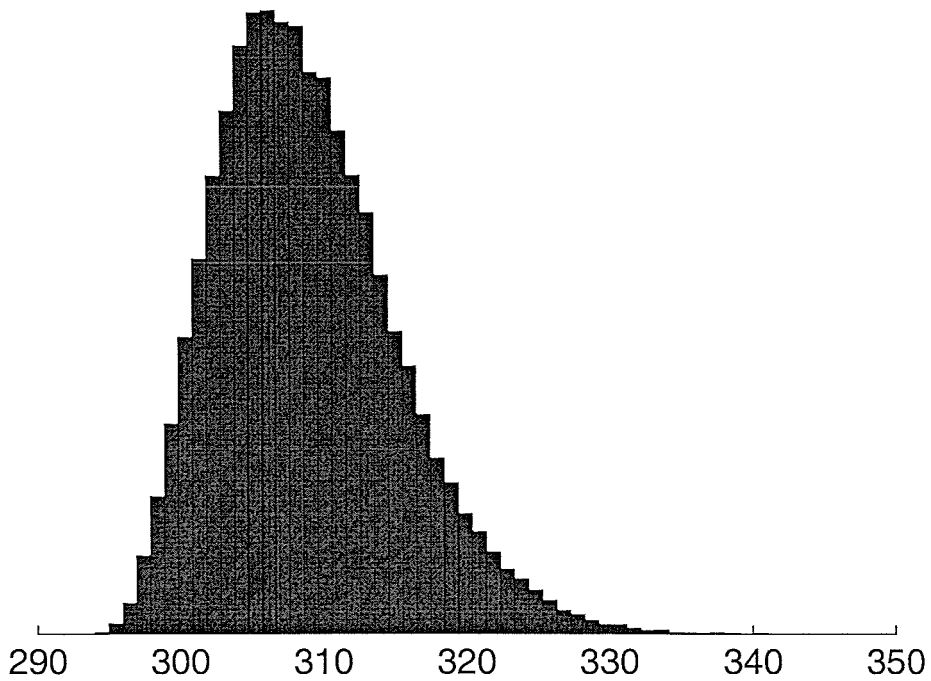
Figure 7.5: Estimate of the posterior distribution for the non-identifiable parameter $\aleph$ for the meadow vole.

Chapter 8

Density Dependent CMAS Model

Bishop et al. (1978) presents a capture-recapture study of the moth *Gonodontis bidentata* from Northwest England. There were $k = 17$ sampling occasions in which $u. = 689$ individuals were caught. The data were also analyzed by Crosbie and Manly (1985) and Link and Barker (2005). Link and Barker (2005) placed a multivariate random effects model on survival and per-capita birth rates as described in section 4.1. They found weak evidence of a positive correlation between $\text{logit}(S_j)$ and $\log(\eta_j')$, suggesting that survival probabilities and fecundity are positively related. Even though a model of this nature provides some insight into dynamics of the population, it leaves many questions unanswered. One such question is whether there are common covariate(s) that induce this positive correlation. An obvious choice for a covariate is the population size, giving a density dependent relationship. When population size is high and competition is strong, both survival and per-capita birth rates may be relatively low. However, when population is low with little competition, survival and per-capita birth rates

may be relatively high. Here we will fit a model to the *Gonodontis bidentata* dataset where both survival probability S_j and the per-capita birth rate η_j are modeled in terms of the latent variable N_j, the population size at time of sample j. We fit this model using the Gibbs sampler, generating a sample from the full conditional distribution of each unknown in every iteration.

8.1 Model

The factorization required for the density dependent CMAS model is described in section 4.3.3, in particular equation 4.17. The CDL is

$$
\mathcal{L}_C(\beta_0, \alpha, \gamma, \tau, p, \aleph; X, b, d, S, \eta) \propto [X|p, d, b, \aleph] \prod_{i=1}^{\aleph} \left\{ [b_{i0}|\beta_0] \times \right.
$$

$$
\left. \prod_{j=1}^{k-1} \left\{ [b_{ij}|b_{:(0:j-1)}, d_{:(1:j-1)}, \beta_0, \eta_{1:j}, \aleph][d_{ij}|b_{:(0:j)}, d_{:(1:j-1)}, S_j] \right\} \right\} \times
$$

$$
\prod_{j=1}^{k-2} [\eta_j|b_{:(0:j-1)}, d_{:(1:j-1)}, \gamma, \tau, \aleph] \prod_{j=1}^{k-1} [S_j|b_{:(0:j-1)}, d_{:(1:j-1)}, \gamma, \tau, \aleph].
$$

$$
(8.1)
$$

where the capture component $[X|p, d, b, \aleph]$ and the mortality component $[d_{ij}|b_{:(0:j)}, d_{:(1:j-1)}, S_j]$ are specified in equations 3.9 and 4.15. The models we use to include the density dependence are

$$
[S_j|b_{:(0:j-1)}, d_{:(1:j-1)}, \gamma, \tau_S, \aleph] = \text{logit} N(\gamma_0 + \gamma_1 \tilde{N}_j, \tau_S), \ j = 1, \ldots, k-1
$$

$$
[\eta_j|b_{:(0:j-1)}, d_{:(1:j-1)}, \alpha, \tau_\eta, \aleph] = LN(\alpha_0 + \alpha_1 \tilde{N}_j, \tau_\eta), \ j = 1, \ldots, k-2,
$$

$$
(8.2)
$$

where $\tilde{N}_j$ is N_j arbitrarily centered to reduce sampling correlation,

$$
\tilde{N}_j = \log(N_j) - 5.5.
$$

We include the error term in equation 8.2 to allow for potential over-dispersion in S and η. In order to reduce the number of computations performed at each iteration we re-parameterize the birth component by changing the variable from η_j to β_j using

$$\eta_j = \frac{\beta_j \aleph}{N_j}, \ j = 1, \ldots, k - 2. \tag{8.3}$$

The Jacobian required for this transformation is

$$\frac{\partial \eta_j}{\partial \beta_j} = \frac{\aleph}{N_j}.$$

The model specified for η_j in equation 8.2 is

$$[\eta_j | b_{:(0:j-1)}, d_{:(1:j-1)}, \alpha, \tau_\eta, \aleph] = \sqrt{\frac{\tau_\eta}{2\pi}} \exp\left(-\frac{\tau_\eta}{2}(\log(\eta_j) - \alpha_0 - \alpha_1 \tilde{N}_j)^2\right) \frac{1}{\eta_j}$$

Changing the variable to β_j gives

$$[\beta_j | b_{:(0:j-1)}, d_{:(1:j-1)}, \alpha, \tau_\eta, \aleph]$$
$$= \sqrt{\frac{\tau_\eta}{2\pi}} \exp\left(-\frac{\tau_\eta}{2}(\log(\beta_j) + \log(\aleph) - \log(N_j) - \alpha_0 - \alpha_1 \tilde{N}_j)^2\right) \frac{N_j}{\aleph \beta_j} \left|\frac{\partial \eta_j}{\partial \beta_j}\right|$$
$$= \sqrt{\frac{\tau_\eta}{2\pi}} \exp\left(-\frac{\tau_\eta}{2}(\log(\beta_j) + \log(\aleph) - \log(N_j) - \alpha_0 - \alpha_1 \tilde{N}_j)^2\right) \frac{1}{\beta_j},$$

$$\tag{8.4}$$

which is a log-normal distribution with mean $\alpha_0 + \alpha_1 \tilde{N}_j + \log(N_j) - \log(\aleph)$ and precision τ_η. The change of variable means that the birth component

of the CDL in equation 8.1 is now written in terms of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{k-1})$,

$$[b_{i0}|\beta_0] \propto {\beta_0'}^{b_{i0}} (1 - \beta_0')^{1-b_{i0}}$$

$$[b_{ij}|\boldsymbol{b}_{:(0:j-1)}, \boldsymbol{d}_{:(1:j-1)}, \beta_0, \beta_{1:j}]$$

$$\propto \left\{ \left(1 - \sum_{h=0}^{j-1} b_{ih}\right) \beta_j' \right\}^{b_{ij}} (1 - \beta_j')^{\left(1 - \sum_{h=0}^{j} b_{ih}\right)}, \ j = 1, \ldots, k-1$$

(8.5)

where

$$\beta_0' = \beta_0$$

$$\beta_j' = \frac{\beta_j}{\prod_{h=0}^{j-1}(1 - \beta_h')}, \ j = 1, \ldots, k-2$$

$$\beta_{k-1}' = 1$$

As with equation 4.16 the constraint $\beta_{k-1}' = 1$ is equivalent to $\sum_{h=0}^{k-1} \beta_h = 1$ and together with the $(1 - \sum_{h=0}^{j-1} b_{ih})$ term imposes the constraint $\sum_{h=0}^{k-1} b_{ih} = 1$.

The model specification is completed with prior distributions for all parameters. The posterior distribution is

$$[\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{p}, \boldsymbol{S}, \beta_0, \boldsymbol{\beta}, \boldsymbol{b}^{mis}, \boldsymbol{d}^{mis}, \aleph | \boldsymbol{X}^{obs}, \boldsymbol{d}^{obs}, \boldsymbol{b}^{obs}] \propto [\boldsymbol{X}|\boldsymbol{p}, \boldsymbol{d}, \boldsymbol{b}, \aleph] \times$$

$$\prod_{i=1}^{\aleph} \left\{ [b_{i0}|\beta_0] \prod_{j=1}^{k-1} \left\{ [b_{ij}|\boldsymbol{b}_{:(0:j-1)}, \boldsymbol{d}_{:(1:j-1)}, \beta_0, \beta_{1:j}] \times \right. \right.$$

$$\left. \left. [d_{ij}|\boldsymbol{b}_{:(0:j)}, \boldsymbol{d}_{:(1:j-1)}, S_j] \right\} \right\} \prod_{j=1}^{k-1} [S_j|\boldsymbol{b}_{:(0:j-1)}, \boldsymbol{d}_{:(1:j-1)}, \boldsymbol{\gamma}, \tau_S, \aleph] \times$$

$$\prod_{j=1}^{k-2} [\beta_j|\boldsymbol{b}_{:(0:j-1)}, \boldsymbol{d}_{:(1:j-1)}, \boldsymbol{\alpha}, \tau_\eta, \aleph] \prod_{h=0}^{1} [\alpha_h] \prod_{h=0}^{1} [\gamma_h][\tau_S][\tau_\eta] \prod_{h=1}^{k} [p_h][\beta_0][\aleph]$$

We denote this model $p(t)S(DD)\eta(DD)$.

8.2   Full Conditional Distributions

A summary of the unknowns, their prior distributions and the updater used

to sample from the full conditional distribution is given in table 8.1.

Unknown	Prior Distribution	Updater
$\aleph$	$DU(0, A)$	RJMCMC
p_j	$Be(a', b')$	Beta
α	$MVN(\mu_0, \Sigma_0)$	Multivariate Normal
γ	$MVN(\mu_0, \Sigma_0)$	Multivariate Normal
τ_η	$Ga(a', b')$	Gamma
τ_S	$Ga(a', b')$	Gamma
S_j	–	MH
β_0	$Be(a', b')$	MH
β_j	–	MH
b_i^{mis}	–	Multinomial
d_i^{mis}	–	Multinomial

Table 8.1: Table summarizing the Gibbs sampler for model the density
dependent CMAS model.

Full Conditional Distribution $[\aleph|\cdot]$

The full conditional distribution for $\aleph$ is

$$[\aleph|\cdot] \propto [X|p, d, b, \aleph] \prod_{i=1}^{\aleph} \left\{ [b_{i0}|\beta_0] \times \right.$$

$$\prod_{j=1}^{k-1} \left\{ [b_{ij}|b_{:(0:j-1)}, d_{:(1:j-1)}, \beta_0, \beta_{1:j}][d_{ij}|b_{:(0:j)}, d_{:(1:j-1)}, S_j] \right\} \bigg\} \times$$

$$\prod_{j=1}^{k-1} [S_j|b_{:(0:j-1)}, d_{:(1:j-1)}, \gamma, \tau_S, \aleph] \prod_{j=1}^{k-2} [\beta_j|b_{:(0:j-1)}, d_{:(1:j-1)}, \alpha, \tau_\eta, \aleph][\aleph].$$

The prior distribution $[\aleph]$ is specified in table 8.1 and the other components

are specified in equations 3.9, 8.5, 4.15, 8.2 and 8.4. As $\aleph$ is discrete and

defines the dimension of X, d and b it must be updated using RJMCMC

(section 5.2.2). In iteration m we propose a candidate $\aleph^*$ from the jumping

distribution $J(\aleph^*|\aleph^{(m)})$. Supposing that $\aleph^* > \aleph^{(m)}$ we must also generate

three augmenting variables u_b, u_d and u_X so that both $b^{(m)*}$, $d^{(m)*}$ and X^* are of the dimension $\aleph^*$,

$$u_b \sim H_1(u|\aleph^{(m)}, \aleph^*, \beta)$$

$$u_d \sim H_2(u|\aleph^{(m)}, \aleph^*, S, u_b)$$

$$u_X \sim H_3(u|\aleph^{(m)}, \aleph^*)$$

We specify the bijection to be the identity,

$$b^{(m)*} = \begin{pmatrix} b^{(m)} \\ u_b \end{pmatrix}, \qquad d^{(m)*} = \begin{pmatrix} d^{(m)} \\ u_d \end{pmatrix}, \qquad X^* = \begin{pmatrix} X^{(m)} \\ u_X \end{pmatrix}.$$

We choose $H_3(\cdot)$ to be a function yielding a $(\aleph^* - \aleph^{(m)}) \times k$ matrix of zeros with probability 1 as we know that any additional individuals were not caught. We choose $H_1(\cdot)$ to be a function generating $(\aleph^* - \aleph^{(m)})$ vectors from the multinomial birth distribution $MN(1, \beta)$ specified in equation 3.11. We choose $H_2(\cdot)$ to be a function generating $(\aleph^* - \aleph^{(m)})$ vectors according to the multinomial mortality distribution $MN(1, xi_i)$ specified in 3.10. We then accept the candidate $\aleph^*$ with probability $q' = \min(1, q)$, where

$$
\begin{aligned}
q &= \frac{[\aleph^*|\cdot] J(\aleph^{(m)}|\aleph^*)}{[\aleph^{(m)}|\cdot] H_1(u_b|\aleph^{(m)}, \aleph^*, \beta) H_2(u_d|\aleph^{(m)}, \aleph^*, S, u_b) J(\aleph^*|\aleph^{(m)})} \\
&= \frac{\aleph^*(\aleph^{(m)} - u.)}{(\aleph^* - u.)\aleph^{(m)}} \frac{\lambda_{DD} \prod_{i=\aleph^{(m)}+1}^{\aleph^*} \prod_{j=\mathfrak{B}_i+1}^{\mathfrak{D}_i}(1 - p_j) J(\aleph^{(m)}|\aleph^*)}{J(\aleph^*|\aleph^{(m)})} I(u. \leq \aleph^* \leq A),
\end{aligned}
$$

where

$$\lambda_{DD} = \lambda_1 \lambda_2$$

$$\lambda_1 = \frac{\exp\left(-\frac{\tau_\eta}{2} \sum_{j=1}^{k-2} \left(\log(\beta_j) + \log(\aleph^\star) - \log(N_j^\star) - \alpha_0 - \alpha_1 \tilde{N}_j^\star\right)^2\right)}{\exp\left(-\frac{\tau_\eta}{2} \sum_{j=1}^{k-2} \left(\log(\beta_j) + \log(\aleph^{(m)}) - \log(N_j^{(m)}) - \alpha_0 - \alpha_1 \tilde{N}_j^{(m)}\right)^2\right)}$$

$$\lambda_2 = \frac{\exp\left(-\frac{\tau_S}{2} \sum_{j=1}^{k-1} \left(\text{logit}(S_j) - \gamma_0 - \gamma_1 \tilde{N}_j^\star\right)^2\right)}{\exp\left(-\frac{\tau_S}{2} \sum_{j=1}^{k-1} \left(\text{logit}(S_j) - \gamma_0 - \gamma_1 \tilde{N}_j^{(m)}\right)^2\right)}$$

All other terms cancel out of the expression because we draw the augmenting variables u_b and u_d in terms of the multinomial expressions for b and d in the standard CMAS model. For the case when $\aleph^\star < \aleph^{(m)}$ we delete the last $(\aleph^{(m)} - \aleph^\star)$ rows of the X, d and b matrices and q becomes

$$q = \frac{\aleph^\star(\aleph^{(m)} - u.)}{(\aleph^\star - u.)\aleph^{(m)}} \frac{\lambda_{DD} J(\aleph^{(m)}|\aleph^\star)}{\prod_{i=\aleph^{(m)}+1}^{\aleph^\star} \prod_{j=\mathfrak{B}_i+1}^{\mathfrak{D}_i} (1 - p_j) J(\aleph^\star|\aleph^{(m)})} I(u. \leq \aleph^\star \leq A).$$

If accepted $\aleph^{(m+1)} = \aleph^\star$, $X^{(m+1)} = X^\star$ and $z^{(m)} = z^{(m)^\star}$, otherwise we set $\aleph^{(m+1)} = \aleph^{(m)}$, $X^{(m+1)} = X^{(m)}$ and $z^{(m)}$ remains unchanged.

Full Conditional Distribution $[p_j|\cdot]$

The full conditional distribution for p_j, $j = 1, \ldots, k$ is

$$[p_j|\cdot] \propto [X|p, b, d, \aleph][p_j]$$

The prior distribution $[p_j]$ is specified in table 8.1 and the distribution $[X|p, b, d, \aleph]$ is specified in equation 3.9. This gives

$$[p_j|\cdot] = Beta(n_j + a', N_j - n_j + b'), \quad j = 1, \ldots, k,$$

where n_j is the number of individuals caught in sample j and N_j is the number of individuals alive in sample j,

$$n_j = \sum_{i=1}^{\aleph} X_{ij}$$

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

Full Conditional Distribution $[\alpha|\cdot]$

The full conditional distribution for $\alpha = (\alpha_0, \alpha_1)$ is

$$[\alpha] \propto \prod_{h=1}^{k-2} [\beta_h | b_{:(0:h-1)}, d_{:(1:h-1)}, \alpha, \tau_\eta, \aleph][\alpha]$$

The component $[\beta_h | b_{:(0:h-1)}, d_{:(1:h-1)}, \alpha, \tau_\eta, \aleph]$ is specified in equation 8.4 and the prior distribution $[\alpha]$ is specified in table 8.1 where μ_0 is a 2×1 vector and Σ_0 is a 2×2 matrix. This gives

$$[\alpha|\cdot] = MVN \left(\tau_\eta \Sigma_\alpha A'_\alpha y_\alpha, \Sigma_\alpha \right),$$

where

$$\Sigma_\alpha = \left(\tau_\eta A'_\alpha A_\alpha + \Sigma_0^{-1} \right)^{-1},$$

$$A_\alpha = \begin{pmatrix} 1 & \tilde{N}_1 \\ \vdots & \vdots \\ 1 & \tilde{N}_{k-2} \end{pmatrix}, \quad y_\alpha = \begin{pmatrix} \log(\beta_1) + \log(\aleph) - \log(N_1) \\ \vdots \\ \log(\beta_{k-2}) + \log(\aleph) - \log(N_{k-2}) \end{pmatrix}.$$

Full Conditional Distribution $[\gamma|\cdot]$

The full conditional distribution for $\gamma = (\gamma_0, \gamma_1)$ is

$$[\gamma|\cdot] = \prod_{h=1}^{k-1} [S_h | b_{:(0:h-1)}, d_{:(1:h-1)}, \gamma, \tau_S, \aleph][\gamma]$$

The component $[S_h|b_{:(0:h-1)}, d_{:(1:h-1)}, \gamma, \tau_S, \aleph]$ is given in equation 8.2 and the prior distribution for γ is specified in table 8.1 where μ_0 is a 2×1 vector and Σ_0 is a 2×2 matrix. This gives

$$[\gamma|\cdot] = MVN\left(\tau_S \Sigma_\gamma A'_\gamma y_\gamma, \Sigma_\gamma\right),$$

where

$$\Sigma_\gamma = \left(\tau_S A'_\gamma A_\gamma + \Sigma_0^{-1}\right)^{-1},$$

$$A_\gamma = \begin{pmatrix} 1 & \tilde{N}_1 \\ \vdots & \vdots \\ 1 & \tilde{N}_{k-1} \end{pmatrix}, \quad y_\gamma = \begin{pmatrix} \text{logit}(S_1) \\ \vdots \\ \text{logit}(S_{k-1}) \end{pmatrix}.$$

Full Conditional Distribution $[\tau_\eta|\cdot]$

The full conditional distribution for τ_η is

$$[\tau_\eta|\cdot] \propto \prod_{h=1}^{k-2} [\beta_h|b_{:(0:h-1)}, d_{:(1:h-1)}, \alpha, \tau_\eta, \aleph][\tau_\eta]$$

The component $[\beta_h|b_{:(0:h-1)}, d_{:(1:h-1)}, \alpha, \tau_\eta, \aleph]$ is specified in equation 8.4 and the prior distribution $[\tau_\eta]$ is specified in table 8.1. This gives,

$$[\tau_\eta|\cdot] = Ga\left(a'', b''\right)$$

where

$$a'' = \frac{k-2}{2} + a'$$

$$b'' = \frac{\sum_{h=1}^{k-2} \left(\log(\beta_h) + \log(\aleph) - \log(N_h) - \alpha_0 - \alpha_1 \tilde{N}_h\right)^2}{2} + b'.$$

Full Conditional Distribution $[\tau_S | \cdot]$

The full conditional distribution for τ_S is

$$[\tau_S | \cdot] \propto \prod_{h=1}^{k-1} [S_h | b_{:(0:h-1)}, d_{:(1:h-1)}, \gamma, \tau_S][\tau_S]$$

The component $[S_h | b_{:(0:h-1)}, d_{:(1:h-1)}, \gamma, \tau_S]$ is specified in equation 8.2 and the prior distribution $[\tau_S]$ is specified in table 8.1. This gives

$$[\tau_S | \cdot] = Ga \left(\frac{k-1}{2} + a', \frac{\sum_{h=1}^{k-1} \left(\text{logit}(S_h) - \gamma_0 - \gamma_1 \tilde{N}_h \right)^2}{2} + b' \right).$$

Full Conditional Distribution $[S_j | \cdot]$

The full conditional distribution for S_j, $j = 1, \ldots, k-1$ is

$$[S_j | \cdot] \propto \prod_{i=1}^{\aleph} [d_{ij} | b_{:(0:j)}, d_{:(1:j-1)}, S_j][S_j | b_{:(0:j-1)}, d_{:(1:j-1)}, \gamma, \tau_S]$$

The components are specified in equations 4.15 and 8.2. This gives

$$[S_j | \cdot] \propto S_j^{N_j - D_j - 1} (1 - S_j)^{D_j - 1} \exp \left(-\frac{\tau_S}{2} (\text{logit}(S_j) - \gamma_0 - \gamma_1 \tilde{N}_j)^2 \right).$$

Full Conditional Distribution $[\beta_0 | \cdot]$

The full conditional distribution for β_0 is

$$[\beta_0 | \cdot] \propto \prod_{i=1}^{\aleph} \left\{ [b_{i0} | \beta_0] \prod_{j=1}^{k-1} [b_{ij} | b_{:(0:j-1)}, d_{:(1:j-1)}, \beta_0, \beta_{1:j}] \right\} [\beta_0]$$

The prior distribution $[\beta_0]$ is specified in table 8.1 and the other components are specified in equation 8.5. This gives

$$[\beta_0|\cdot] \propto \beta_0^{B_0+a'-1}(1-\beta_0)^{b'-1}\left(1-\sum_{h=0}^{k-2}\beta_h\right)^{B_{k-1}} I\left(1-\sum_{h=0}^{k-2}\beta_h > 0\right).$$

Full Conditional Distribution $[\beta_j|\cdot]$

The full conditional distribution for β_j, $j = 1,\ldots,k-2$ is

$$[\beta_j|\cdot] \propto \prod_{h=j}^{k-1}[b_{ih}|\boldsymbol{b}_{:(0:h-1)}, \boldsymbol{d}_{:(1:h-1)}, \beta_0, \beta_{1:h}][\beta_j|\boldsymbol{b}_{:(0:j-1)}, \boldsymbol{d}_{:(1:j-1)}, \boldsymbol{\alpha}, \tau_\eta, \aleph].$$

These components are given in equations 8.5 and 8.4. This gives

$$[\beta_j|\cdot] \propto \beta_j^{B_0-1}\left(1-\sum_{h=0}^{k-2}\beta_h\right)^{B_{k-1}} I\left(1-\sum_{h=0}^{k-2}\beta_h > 0\right) \times$$
$$\exp\left(-\frac{\tau_\eta}{2}(\log(\beta_j) + \log(\aleph) - \log(N_j) - \alpha_0 - \alpha_1\tilde{N}_j)^2\right).$$

Full Conditional Distribution $[b_i^{mis}|\cdot]$

The full conditional distribution for the missing values in the ith row of $\boldsymbol{b}$ is,

$$[b_{i:}^{mis}|\cdot] \propto [\boldsymbol{X}|\boldsymbol{p}, \boldsymbol{d}, \boldsymbol{b}, \aleph]\prod_{i=1}^{\aleph}\left\{[b_{i0}|\beta_0]\prod_{j=1}^{k-1}[b_{ij}|\boldsymbol{b}_{:(0:j-1)}, \boldsymbol{d}_{:(1:j-1)}, \beta_0, \beta_{1:j}]\times\right.$$
$$\left.[d_{ij}|\boldsymbol{b}_{:(0:j)}, \boldsymbol{d}_{:(1:j-1)}, S_j]\right\}\times$$
$$\prod_{j=1}^{k-1}[S_j|\boldsymbol{b}_{:(0:j-1)}, \boldsymbol{d}_{:(1:j-1)}, \gamma, \tau_S, \aleph]\prod_{j=1}^{k-2}[\beta_j|\boldsymbol{b}_{:(0:j-1)}, \boldsymbol{d}_{:(1:j-1)}, \boldsymbol{\alpha}, \tau_\eta, \aleph].$$

The components are specified in equations 3.9, 8.5, 4.15, 8.2 and 8.4. This gives a multinomial distribution,

$$[\boldsymbol{b}_{i:}^{mis}|\cdot] = MN(1, \boldsymbol{\zeta}_i), \quad i = 1, \ldots, \aleph.$$

where $\boldsymbol{\zeta}_i = (\zeta_{i0}, \ldots, \zeta_{i\mathfrak{F}_i-1})$ and

$$\zeta_{ij} = \frac{\kappa_{ij}}{\sum_{h=0}^{\mathfrak{F}_i-1} \kappa_{ih}}, \quad j = 0, \ldots, \mathfrak{F}_i - 1$$

$$\kappa_{ij} = \beta_j \prod_{h=j+1}^{\mathfrak{F}_i-1} S_h(1 - p_h) \times$$

$$\prod_{h=1}^{k-2} \exp\left(-\frac{\tau_\eta}{2}(\log(\beta_h) + \log(\aleph) - \log(N_h^\star) - \alpha_0 - \alpha_1 \tilde{N}_h^\star)^2\right) \times$$

$$\prod_{h=1}^{k-1} \exp\left(-\frac{\tau_S}{2}(\text{logit}(S_h) - \gamma_0 - \gamma_1 \tilde{N}_h^\star)^2\right), \quad j = 0, \ldots, \mathfrak{F}_i - 1.$$

The value $N_h^\star$ is the value of N_h when $b_{ij} = 1$. Note that $\mathfrak{F}_i = \mathfrak{D}_i$ for all individuals never caught $i = u. + 1, \ldots, \aleph$.

Full Conditional Distribution $[d_i^{mis}|\cdot]$

The full conditional distribution for the missing values in the ith row of $\boldsymbol{d}$ is,

$$[\boldsymbol{d}_{i:}^{mis}|\cdot] \propto [\boldsymbol{X}|\boldsymbol{p}, \boldsymbol{d}, \boldsymbol{b}, \aleph] \prod_{i=1}^{\aleph} \left\{ [b_{i0}|\beta_0] \prod_{j=1}^{k-1} [b_{ij}|\boldsymbol{b}_{:(0:j-1)}, \boldsymbol{d}_{:(1:j-1)}, \beta_0, \boldsymbol{\beta}_{1:j}] \times \right.$$

$$\left. [d_{ij}|\boldsymbol{b}_{:(0:j)}, \boldsymbol{d}_{:(1:j-1)}, S_j] \right\} \times$$

$$\prod_{j=1}^{k-1} [S_j|\boldsymbol{b}_{:(0:j-1)}, \boldsymbol{d}_{:(1:j-1)}, \boldsymbol{\gamma}, \tau_S, \aleph] \prod_{j=1}^{k-2} [\beta_j|\boldsymbol{b}_{:(0:j-1)}, \boldsymbol{d}_{:(1:j-1)}, \boldsymbol{\alpha}, \tau_\eta, \aleph].$$

The components are specified in equations 3.9, 8.5, 4.15, 8.2 and 8.4. This gives a multinomial distribution,

$$[d_{i:}^{mis}|\cdot] = MN(1, \zeta_i), \quad i = 1, \ldots, \aleph$$

where $\zeta_i = (\zeta_{i\mathfrak{L}_i}, \ldots, \zeta_{ik})$ and

$$\zeta_{ij} = \frac{\kappa_{ij}}{\sum_{h=\mathfrak{L}_i}^{k} \kappa_{ih}}, \quad j = \mathfrak{L}_i, \ldots, k$$

$$\kappa_{ij} = \begin{cases} \prod_{h=\mathfrak{L}_i}^{j-1} S_h(1-p_{h+1})(1-S_j) \times \\ \quad \prod_{h=1}^{k-2} \exp\left(-\frac{\tau_\eta}{2}(\log(\beta_h) + \log(\aleph) - \log(N_h^\star) - \alpha_0 - \alpha_1 \tilde{N}_h^\star)^2\right) \times \\ \quad \prod_{h=1}^{k-1} \exp\left(-\frac{\tau_S}{2}(\text{logit}(S_h) - \gamma_0 - \gamma_1 \tilde{N}_h^\star)^2\right), \quad j = \mathfrak{L}_i, \ldots, k-1 \\ \\ \prod_{h=\mathfrak{L}_i}^{j-1} S_h(1-p_{h+1}) \times \\ \quad \prod_{h=1}^{k-2} \exp\left(-\frac{\tau_\eta}{2}(\log(\beta_h) + \log(\aleph) - \log(N_h^\star) - \alpha_0 - \alpha_1 \tilde{N}_h^\star)^2\right) \times \\ \quad \prod_{h=1}^{k-1} \exp\left(-\frac{\tau_S}{2}(\text{logit}(S_h) - \gamma_0 - \gamma_1 \tilde{N}_h^\star)^2\right), \quad j = k. \end{cases}$$

The value $N_h^\star$ is the value of N_h when $d_{ij} = 1$. Note that $\mathfrak{L}_i - \mathfrak{B}_i + 1$ for all individuals never caught $i = u. + 1, \ldots, \aleph$.

8.3 Results

The prior distributions used were

$$\alpha \sim MVN\left((0,0)', 10000I_2\right)$$

$$\gamma \sim MVN\left((0,0)', 10000I_2\right)$$

$$\tau_\eta \sim Ga(0.001, 0.001)$$

$$\tau_S \sim Ga(0.001, 0.001)$$

$$\beta_0 \sim Be(1,1)$$

$$p_j \sim Be(1,1), \ j = 1, \ldots, k$$

$$\aleph \sim DU(0, 100000)$$

where I_n is the $n \times n$ identity matrix.

The Gibbs sampler was run on 3 independent chains started from different values. Each chain had an adaptive phase of $20,000$ iterations to optimize the Metropolis Hastings and RJMCMC algorithms as specified in section 5.2.2. As there was high autocorrelation in the values, each chain was run for a $100,000$ iteration burn-in followed by a further $300,000$ iterations. The chains mixed and were combined to give a posterior sample of $900,000$ iterations.

The posteriors of particular interest are the density dependent parameters α_1 and γ_1 (figure 8.3). The parameter α_1 has approximately 98% of mass below 0 suggesting that per-capita birth rates are negatively associated with population size. The parameter γ_1 has approximately 81% of the posterior mass below 0. This suggests a negative relationship between survival and population size, although it is not convincing. The evidence of density dependence raises the possibility that the correlation identified by Link and Barker (2005) was one induced by survival and birth rates both

being negatively related to population size. These results indicate that the population of *Gonodontis* is stable, at least in regard to fecundity: when the population size becomes large or small, birth rates adjust so that the population returns to somewhere near equilibrium.

Posterior summaries are available for the parameters in the density dependent relationship in table 8.2. To assess the difference between the density dependence model and standard models we compared the density dependent model to a standard CMAS $p(t)S(t)\eta(t)$ model as in section 7.4. Imposing the density dependent relationships results in shrinkage toward the predicted model (Figures 8.1 and 8.2). The shrinkage not only occurs for the estimates of survival probability and birth rates, but also for the demographic summaries predicted from the model, such as N_j. In most cases the posterior predictive distribution for N_j is tighter under the assumption of density dependence as shown in table 8.4.
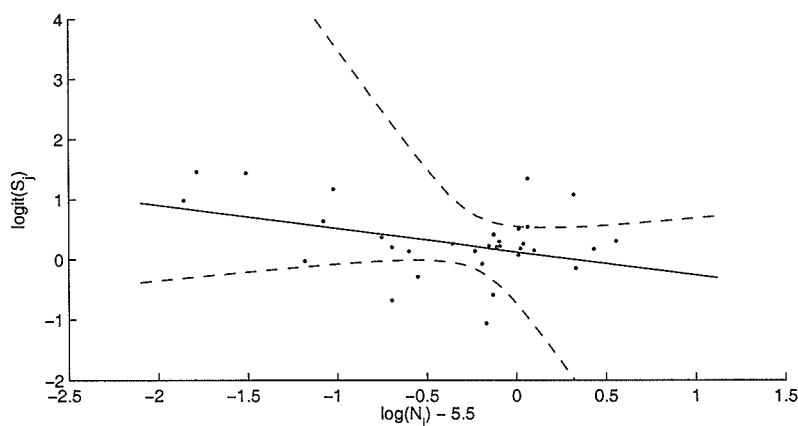
Figure 8.1: The posterior medians for logit(S_j) are plotted against the posterior medians for $\log(N_j) - 5.5$ for the $p(t)S(t)\eta(t)$ model (blue points) as well as for the $p(t)S(DD)\eta(DD)$ model (red points). The superimposed black lines are the median (solid line) and the 2.5% and 97.5% quantiles (dashed lines) of the estimated density dependent relationship.

Note that the changes in the posterior predictive distribution for N_j as
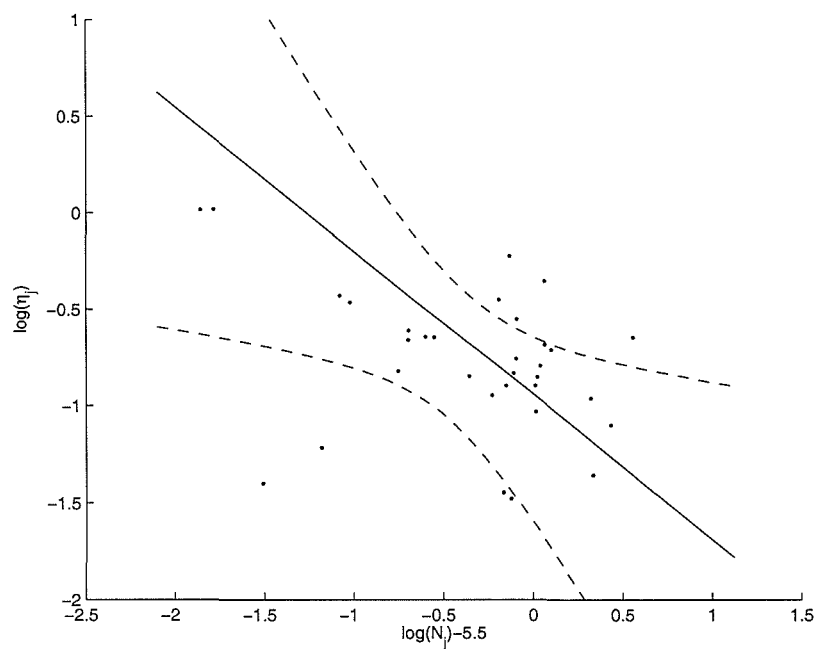
Figure 8.2: The posterior medians for $\log(\eta_j)$ are plotted against the posterior medians for $\log(N_j) - 5.5$ for the $p(t)S(t)\eta(t)$ model (blue points) as well as for the $p(t)S(DD)\eta(DD)$ (red points). The superimposed black lines are the median (solid line) and the 2.5% and 97.5% quantiles (dashed lines) of the estimated density dependent relationship.

well as sampling correlations among parameter estimates, are ignored when there is crude estimation of the density dependent relationship, for example, by regressing survival rate estimates on abundance estimates. Not accounting for the sampling process correctly could lead to incorrect inference for the study population. In order to compare the estimates we fitted a regression model to the posterior medians of S_j and η_j obtained from the $p(t)S(t)\eta(t)$ model,

$$\text{logit}(\tilde{S}_j) \sim N(\gamma_0^c + \gamma_1^c \tilde{N}_j, \tau_S^c), \ j = 1, \ldots, k-2,$$

$$\log(\tilde{\eta}_j) \sim N(\alpha_0^c + \alpha_1^c \tilde{N}_j, \tau_\eta^c), \ j = 2, \ldots, k-2,$$

where $\tilde{S}_j$ and $\tilde{\eta}_j$ are the posterior medians of S_j and η_j in the $p(t)S(t)\eta(t)$ model. The regression estimates are close to those found in the density dependence model $p(t)S(DD)\eta(DD)$ although the credible intervals are quite different (Table 8.3). The danger of the crude analysis is best seen in the posterior for α_1^c, with the 95% credible interval for the crude analysis incorrectly including 0.

Parameter	2.5%	50%	97.5%
γ_0	-0.58	0.13	0.58
γ_1	-3.09	-0.35	0.36
σ_S	0.03	0.29	1.14
α_0	-1.61	-0.97	-0.64
α_1	-1.60	-0.75	-0.09
σ_η	0.05	0.53	1.22

Table 8.2: Table of posterior 2.5%, 50% and 97.5% quantiles for γ_0, α_0, σ_S and σ_η.
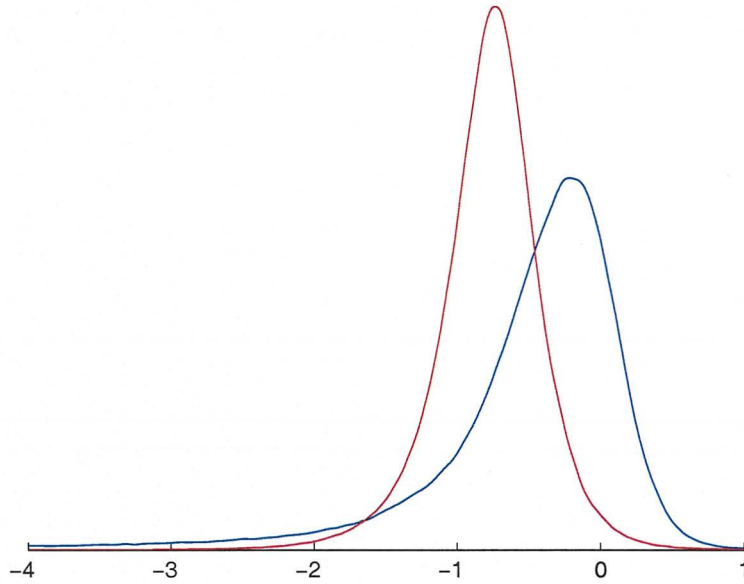
Figure 8.3: Posterior density estimates for γ_1, the effect of population size on survival probabilities (blue line) and α_1, the effect of population size on per-capita birth rates (red line) for *Gonodontis bidentata*.

Parameter	2.5%	50%	97.5%
γ_0^c	-0.22	0.26	0.73
γ_1^c	-1.06	-0.42	0.23
σ_S^c	0.57	0.81	1.27
α_0^c	-1.62	-1.12	-0.62
α_1^c	-1.67	-0.83	0.03
σ_η^c	0.60	0.85	1.37

Table 8.3: Table of posterior 2.5%, 50% and 97.5% quantiles for γ_0^c, α_0^c, σ_S^c and σ_η^c for the density dependence relationship found using the median values for survival and per-capita birth rates.

	$p(t)S(DD)\eta(DD)$			$p(t)S(t)\eta(t)$		
Quantity	2.5%	50%	97.5%	2.5%	50%	97.5%
N_2	60	83	133	61	88	155
N_3	85	120	194	92	141	225
N_4	89	129	207	83	122	199
N_5	51	103	197	38	54	95
N_6	129	201	347	132	215	385
N_7	153	233	377	150	261	455
N_8	145	208	304	133	216	366
N_9	148	219	334	146	261	503
N_{10}	192	282	500	252	427	698
N_{11}	160	255	457	180	377	755
N_{12}	177	258	405	203	337	536
N_{13}	186	253	384	218	341	517
N_{14}	154	218	320	150	248	425
N_{15}	129	189	279	116	207	354
N_{16}	78	160	254	49	75	162

Table 8.4: Comparison of posterior predictive distributions of N_j from the $p(t)S(DD)\eta(DD)$ model and the $p(t)S(t)\eta(t)$ model.

Chapter 9

CJS Model with Continuous Covariates and Auxiliary Data

A capture-recapture study on rainbow trout was conducted by Fish and Game New Zealand on the upper headwaters of the Rangitikei River, New Zealand as described in Barker et al. (2001). A total of $u. = 547$ rainbow trout were caught over $k = 13$ sampling periods. The trout were caught using natural bait fished upstream at a dead drift during a 2-week window each spring (between 7 and 31 October) and autumn (between 15 March and 15 April) from March 1993 until April 1999 (Barker et al. 2001). At time of capture, information was recorded about the sex and the length of each individual. In an attempt to improve the survival of the larger, breeding fish an upper size restriction of 550mm was imposed on the rainbow trout fishery during the study (October 1995), making it illegal to keep any fish over 550mm.

New recruits to the upper headwaters of the Rangitikei River are not

assumed to come through birth. Instead it is assumed that fish develop downstream and enter the upper headwaters after sexual maturity. Therefore, including the first captures in the model provides information about immigration, not fecundity. As understanding factors that influence immigration is not a focus of the study we look only at models that condition on first capture.

Additional information was available, with live re-sightings and dead recoveries collected from anglers. The information obtained for each reported fish varied, with length information included for some fish. All length information that was received from anglers was checked to ensure the measurement was correctly taken. Recoveries and re-sightings were recorded for $k' = k = 13$ periods.

The goal of the study was to determine whether the upper size limit (i) increased the return rate for fish over 550mm and (ii) increased the survival for the larger breeding fish. In an attempt to investigate these effects Barker et al. (2001) modeled the joint capture-recapture, re-sighting and recovery data. They used length at first capture to specify length effects for the period after first capture and disregarded all further length information[1]. The complete analysis requires length as a covariate for all survival probabilities, not just survival for the period after first capture. This requires length to be included as a continuous covariate with a model for the missing values.

In total there were 715 length measurements obtained from the $u. = 547$ fish. There was 1 fish with 5 length measurements, 4 with 4 measurements, 20 with 3 measurements, 113 with 2 measurements, 408 with 1 measurement and 1 fish with no length measurements.

The data comprise several components.

---

[1] Barker et al. (2001) also used an earlier version of the data.

Capture-Recapture Data

The observed capture-recapture histories are denoted by the matrix X^{obs}, where $X_{ij}^{obs} = 1$ means that individual i was captured in sample j and $X_{ij}^{obs} = 0$ otherwise, $i = 1, \ldots, u., \ j = 1, \ldots k$. As usual we separate X^{obs} into X_1 and X_2, where X_1 contains all information in X^{obs} up to and including first capture and X_2 contains all information in X^{obs} after first capture.

Angler Recoveries

The observed angler dead recoveries are denoted by the matrix Y_1. The value $Y_{1ij} = 1$ denotes individual i being recovered between sample j and $j + 1$, with $Y_{1ij} = 0$ otherwise, $i = 1, \ldots, u., \ j = 1, \ldots, k'$.

Angler Re-sightings

The observed angler live re-sightings are denoted by the matrix Y_2. The value $Y_{2ij} = 1$ denotes individual i being re-sighted between sample j and $j + 1$, with $Y_{2ij} = 0$ otherwise, $i = 1, \ldots, u., \ j = 1, \ldots, k'$.

Information on Death

The partially observed matrix d that denotes the interval censored time of death, with $d_{ij} = 1$ meaning that individual i died between sample j and $j + 1$ and $d_{ij} = 0$ otherwise, $i = 1, \ldots, u., \ j = 1, \ldots, k' + 1$. We separate d into d_1 and d_2, where d_1 contains all information in d up to and including first capture and d_2 contains all information in d after first capture. Note that the re-sightings and recoveries have extended the d matrix. This means that the value $d_{ik'+1} = 1$ denotes that the individual was still alive at the end of the study.

Loss on Capture

The information on loss on capture is denoted by $\iota = (\iota_1, \ldots, \iota_{u.})$, where $\iota_i = 0$ means that individual i was lost on capture, with $\iota_i = 1$ otherwise.

Availability for Capture

The matrix z is used to denote availability for capture, with $z_{ij} = 1$ meaning that individual i was available for capture in sample j with $z_{ij} = 0$ otherwise, $i = 1, \ldots, u.$, $j = 1, \ldots, k$. We separate z into z_1 and z_2, where z_1 contains all information in z up to and including first capture and z_2 contains all information in z after first capture.

Length Information from Capture-Recapture Model

The length information for individuals measured as part of the capture-recapture study is denoted by the matrix L_1. The value $L_{1i}(\delta_{1ih})$ denotes the recorded length of individual i in capture h, for $i = 1, \ldots, u.$, $h = 1, \ldots, \omega_{1i}$. The value δ_{1ih} is the time at capture h for individual i and ω_{1i} is the number of times i was captured with a length measurement taken.

Length Information from Recoveries

The length information for individuals measured when recovered is denoted L_2. The value $L_{2i}(\delta_{2ih})$ denotes the recorded length of individual i in recovery h, for $i = 1, \ldots, u.$, $h = 1, \ldots, \omega_{2i}$. The value δ_{2ih} is the time at recovery h for individual i and ω_{2i} is the number of recoveries for individual i with a length measurement taken. Note that it is possible for each individual to be only be recovered once in its lifetime, so ω_{2i} must either take the value 0 or 1.

Length Information from Re-sightings

The length information for individuals measured when re-sighted is denoted L_3. The value $L_{3i}(\delta_{3ih})$ denotes the recorded length of individual i in re-sighting h, for $i = 1, \ldots, u.$, $h = 1, \ldots, \omega_{3i}$. The value δ_{3ih} is the time at re-sighting h for individual i and ω_{3i} is the number of times i was re-sighted with a length measurement taken.

Information on Sex

The sex information that was recorded as part of the capture-recapture study is denoted $sex = (sex_1, \ldots, sex_{u.})$, with $sex_i = 1$ denoting a male with $sex_i = 0$ otherwise.

Information on Season

The information on whether sampling occurred in spring or autumn is denoted by the vector $sea = (sea_1, \ldots, sea_k)$, with $sea_j = 1$ denoting spring and $sea_j = 0$ otherwise.

Information on Regulation

The information on when the upper size limit was introduced is denoted $reg = (reg_1, \ldots, reg_k)$, with $reg_j = 1$ denoting that the upper size limit has been enforced, with $reg_j = 0$ otherwise.

Note the distinction between t_j, δ_{1ih}, δ_{2ih} and δ_{3ih}. The value t_j is the time at sampling occasion j and δ_{1ih}, δ_{2ih} and δ_{3ih} are the times at capture h, recovery h and re-sighting h for individual i respectively.

The information in sex is incomplete as not every individual has their sex assigned. We assume that the underlying sex of the individuals did not affect whether or not these observations were recorded. Therefore, the missingness mechanism is ignorable and the description for how the observations

went missing does not need to be included in the model.

The information on length is also incomplete. There are two reasons are length measurements go missing. The first reason is because we are not able to observed the fish in every sample. The reason why these observations go missing depends on the unobserved length values. The missingness mechanism for this process is fully specified by the capture-recapture process X_2 and the recovery and re-sighting processes Y_1 and Y_2. Therefore, we do not need to include an additional component that models how the data go missing. The second reason length observations are missing is because anglers (and in some cases fish and game officers) did not record length measurements when the fish was caught. We assume that this decision has nothing to do with the underlying measurement. Therefore, the missingness mechanism is ignorable and the description for how the observations went missing does not need to be included in the model. In principle this assumption can be relaxed if we believe that anglers were less likely to report the length of fish that were close to, or larger than, the legal limit (but still report the capture). However, this requires that at least some of the fish that are re-sighted with no measurement are subsequently caught and measured. There was not enough information in this dataset to investigate a model of this sort.

The joint model for this problem is complex, so we develop each of the conditional likelihood components in turn before combining them to give the CDL.

9.1   Length Model

The von Bertalanffy growth function (VBGF) is extensively used in fisheries to model the growth in length of fish (von Bertalanffy 1938). It assumes that

growth is monotonically increasing at a decreasing rate to an asymptote as shown in figure 9.1.
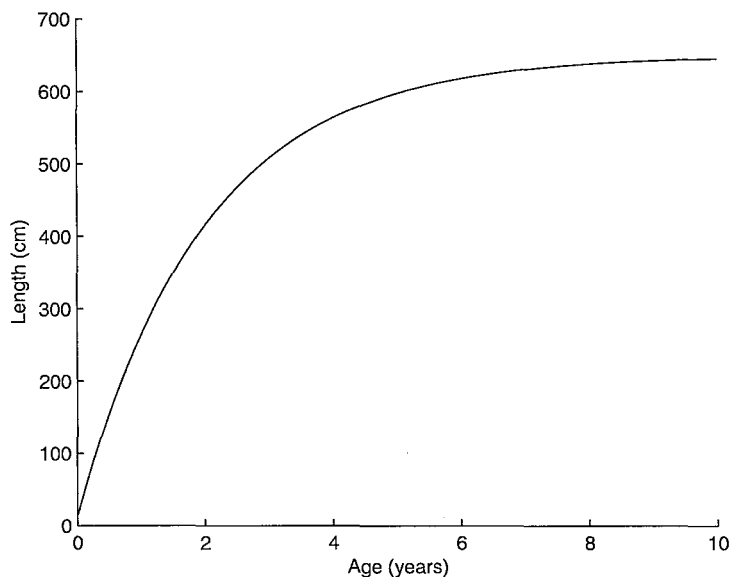
Figure 9.1: An example of the von Bertalanffy growth function.

9.1.1   Standard VBGF

The VBGF for the length of individual i at age t, denoted $\Lambda_{1i}(t)$ can be written as

$$\Lambda_{1i}(t) = L_{\infty i}(1 - \zeta_i \exp(-K_i t)), \ L_{\infty i} > 0, \ K_i > 0, \ 0 < \zeta_i < 1.$$

The parameters are $L_{\infty i}$, the asymptotic length, K_i a growth rate parameter that describes how quickly the growth rises to the asymptote and ζ_i is a parameter that allows for the size at birth to be larger than 0.

When the length data are obtained from capture-recapture studies the age of the individual is unknown, making the parameter ζ_i not identifiable.

The usual assumption that is incorporated to overcome this is that $\zeta_i = 1$ making the VBGF,

$$\Lambda_{1i}(t) = L_{\infty i}(1 - \exp(-K_i(A_i(\delta_{1i1}) + \Delta_i(t)))),$$

where $A_i(\delta_{1i1})$ is the (unknown) age from size 0 at first capture and $\Delta_i(t)$ is the time difference between t and first capture for individual i.

In order to estimate the growth parameters, Fabens (1965) conditioned on the observed length at first capture of each individual $L_i(\delta_{1i1})$ to give

$$D_i(t) = (L_{\infty i} - L_i(\delta_{1i1}))(1 - \exp(-K_i\Delta(t))),$$

where $D_i(t)$ is the expected difference between length at time t and length at first capture $L_i(\delta_{1i1})$, which is assumed fixed. Provided there is adequate data, least squares estimates can be found for $L_{\infty i}$ and K_i or restrictions of these parameters. Maller and deBoer (1988) showed that this procedure yields inconsistent estimates with biases that can potentially be very large. Francis (1988a) points out that a major source of the inconsistency is that the random variable $L_i(\delta_{1i1})$ is treated as known and overcame this through the re-parameterization described in Francis (1988b). James (1991) claimed to overcome the problem by modifying Fabens (1965) estimators to give distribution-free consistent estimators. Laslett et al. (2002) took a different approach and included the age at first capture in the model as a parameter to be estimated,

$$L_i(\delta_{1ih}) \sim N(\Lambda_{1i}(\delta_{1ih}), \sigma^2), \quad i = 1, \ldots, u., \quad h = 1, \ldots, \omega_{1i}$$

where $L_i(\delta_{1ih})$ is the observed length at time of capture h, σ^2 is the com-

bined measurement and model error[2] and ω_{1i} is the number of captures for individual i. In order to fit the model they assumed[3] that

1. $K_i = K$, $i = 1, \ldots, u$.

2. $L_{\infty i} \sim N(\mu_\infty, \tau_\infty)$, $i = 1, \ldots, u$.

3. $A_i(\delta_{1i1}) \sim LN(\mu_A, \tau_A)$, $i = 1, \ldots, u$.

Laslett et al. (2002) used maximum likelihood to fit the model, integrating out the random effects distribution for age. Note that if there is sufficient data it is also possible to extend the model of Laslett et al. (2002) to allow K_i to be a random effect, although within a maximum likelihood framework Eveson et al. (2004) noted that it "...may be too computer intensive to be feasible in practice". Fortunately, there is no such difficulty using Bayesian computational methods such as MCMC.

The standard VBGF has many statistical downsides. The parameter $L_{\infty i}$ is defined outside the range of the data as we never observe any individuals at infinite age. This is especially inappropriate when the individuals we catch are still growing rapidly. A related problem is that the parameters have high sampling correlation (Ratkowsky 1986, Francis 1988a,b) which can lead to unstable parameter estimates.

9.1.2   VBGF Mark II

Wang (2004) suggested a re-parameterization of the VBGF in terms of length

---

[2]Laslett et al. (2002) assumed that there was no error in the model. That is, if the individuals were accurately measured then they would fit exactly on the VBGF. This assumption means that Laslett et al. (2002) define σ^2 as measurement error.

[3]Note that as $L_i(\delta_{1ih}) > 0$ and $L_{\infty i} > 0$ it is preferable to consider positive distributions to enforce these constraints. For example,

$$L_i(\delta_{1ih}) \sim LN(\log(\Lambda_1(\delta_{1ih})), \tau), \quad i = 1, \ldots, u., \quad h = 1, \ldots, \omega_{1i}$$

assumes that the median size of the individual fish follow a VBGF.

at first capture instead of age at first capture,

$$\Lambda_{2i}(t) = \alpha_i + (L_{\infty i} - \alpha_i)(1 - \exp(-K_i \Delta_i(t))),$$

where α_i is the length at first capture,

$$\alpha_i = L_{\infty i}(1 - \zeta_i \exp(-K A_i(\delta_{1i1}))).$$

One virtue of using α_i is that we are able to directly check their values from the data. Furthermore, any additional information known about the first capture can be included to improve the model fit. We make use of this for the Rangitikei River fish because we expect the length distribution of the available fish to differ between the autumn and spring sampling periods. This difference is highlighted by an extraneous sampling period in the autumn of 1998 where many small fish were caught in the upper headwaters because of the high river temperatures in the summer of 1997/1998. It has been suggested that the temperature led to fish moving out of the middle reaches and into the cooler headwaters.

Even though this parameterization provides an improvement over the standard VBGF it still exhibits many of the statistical downsides mentioned in section 9.1.1, including the parameter $L_{\infty i}$ being out of range and the high sampling correlation between $L_{\infty i}$ and K_i.

9.1.3   VBGF Mark III

Many of the statistical problems revolve around the fact that $L_{\infty i}$ is defined outside the range of the data. To overcome this we re-parameterize in terms

of λ_i, the instantaneous growth rate at first capture,

$$\frac{\partial \alpha_i}{\partial A_i(\delta_{1i1})} = \lambda_i = K_i(L_{\infty i} - \alpha_i).$$

Using λ_i gives

$$\Lambda_{3i}(t) = \alpha_i + \frac{\lambda_i}{K_i}\left(1 - \exp(-K_i \Delta_i(t))\right).$$

The major advantage of this parameterization is that λ_i is well supported by the data, so will likely have lower sampling correlation with K_i.

9.1.4 Additional VBGF alternatives

There are an infinite number of possible re-parameterizations that could be examined for the VBGF. A particularly appealing model is described in Ratkowsky (1986), where the VBGF is parameterized predominantly in terms of features of the study,

$$\Lambda_{4i}(t) = \alpha_i + \frac{\Omega_i - \alpha_i}{1 - \exp\left(-K_i \Delta(t_{\Omega i})\right)}\left(1 - \exp(-K_i \Delta_i(t))\right)$$

where Ω_i is the length at last capture and $\Delta(t_{\Omega i})$ is the time difference between last capture and first capture for individual i. Ratkowsky (1986) concludes that this parameterization has low sampling correlations between parameters and converges well using non-linear least squares.

However, a necessary feature of the model is $\Delta(t_{\Omega i})$, which is not available for individuals caught less than two times. This means that the expected size of individuals caught less than twice is not defined at times other than first capture. As individuals caught once only are modeled entirely in terms of α_i, this is no problem for parameter estimation. However,

it makes prediction impossible for times other than the first capture for individuals caught less than twice. This means that the model is of little use as the very reason we are modeling length is to predict the missing values of the covariate for use in the capture-recapture model.

9.1.5   Model

As we are interested in predicting the missing length values, the VGBF model $\Lambda_{3i}(t_j)$ from section 9.1.3 is chosen as the preferred model for the length observations. For readability we suppress the subscript and denote this function as $\Lambda_i(t_j)$,

$$\Lambda_i(t) = \alpha_i + \frac{\lambda_i}{K_i}\left(1 - \exp(-K_i\Delta_i(t))\right). \tag{9.1}$$

We are able to extend the basic structure of the model to allow for the data collection being from both fish and game officers and anglers. We expect that the length observations of the anglers are from a separate distribution with a potential bias. Assuming independence between individuals and through time the model for length is

$$\left[L_1|\alpha,\lambda,K,\tau_1\right] = \prod_{i=1}^{u.}\prod_{j=1}^{\omega_{1i}}\left[L_{1i}(\delta_{1ij})|\alpha,\lambda,K,\tau_1\right],$$

$$\left[L_{1i}(\delta_{1ij})|\alpha,\lambda,K,\tau_1\right] = LN(\log(\Lambda_i(\delta_{1ij})),\tau_1),$$

$$\left[L_2|\alpha,\lambda,K,\mathcal{B},\tau_2\right] = \prod_{i=1}^{u.}\prod_{j=1}^{\omega_{2i}}\left[L_{2i}(\delta_{1ij})|\alpha,\lambda,K,\mathcal{B},\tau_2\right],$$

$$\left[L_{2i}(\delta_{1ij})|\alpha,\lambda,K,\mathcal{B},\tau_2\right] = LN(\log(\Lambda_i(\delta_{1ij})) + \mathcal{B},\tau_2),$$

$$\left[L_3|\alpha,\lambda,K,\mathcal{B},\tau_2\right] = \prod_{i=1}^{u.}\prod_{j=1}^{\omega_{3i}}\left[L_{3i}(\delta_{1ij})|\alpha,\lambda,K,\mathcal{B},\tau_2\right],$$

$$\left[L_{3i}(\delta_{1ij})|\alpha,\lambda,K,\mathcal{B},\tau_2\right] = LN(\log(\Lambda_i(\delta_{1ij})) + \mathcal{B},\tau_2),$$

$$\tag{9.2}$$

where $\mathcal{B}$ is the bias of the anglers, τ_1 is the precision for the fish and game officers and τ_2 is the precision for the anglers.

We assume that the instantaneous growth rates λ_i, $i = 1, \ldots, u.$ are drawn from a common log-normal distribution that depends on the sex of the individual. The conditional likelihood component for $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{u.})$ is

$$[\boldsymbol{\lambda}|\boldsymbol{\theta}_\lambda, \tau_\lambda, sex] = MVN(\boldsymbol{X}_\lambda \boldsymbol{\theta}_\lambda, \boldsymbol{\Sigma}_\lambda) \tag{9.3}$$

where

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{u.} \end{bmatrix}, \ \boldsymbol{X}_\lambda = \begin{bmatrix} 1 & sex_1 \\ 1 & sex_2 \\ \vdots & \vdots \\ 1 & sex_{u.} \end{bmatrix}, \ \boldsymbol{\theta}_\lambda = \begin{bmatrix} \theta_{\lambda 0} \\ \theta_{\lambda 1} \end{bmatrix}, \ \boldsymbol{\Sigma}_\lambda = \sigma_\lambda^2 \boldsymbol{I}_{u.}$$

where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. The common distribution for $\boldsymbol{\lambda}$ is essential because the value of λ_i is used to predict the missing values of the length covariate for times other than first capture. Without the common distribution, λ_i is informed by the data only for individuals with at least 2 captures.

We assume that the values α_i, $i = 1, \ldots, u.$ are drawn from a common log normal distribution that depends on the sex of the individual as well as the season of capture. We include a term in both the mean and variance for the autumn of 1998 that allows for the extraordinarily high summer river temperatures. The conditional likelihood component for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{u.})$ is

$$[\boldsymbol{\alpha}|\boldsymbol{\theta}_\alpha, \tau_\alpha, sex, sea] = MVN(\boldsymbol{X}_\alpha \boldsymbol{\theta}_\alpha, \boldsymbol{\Sigma}_\alpha) \tag{9.4}$$

where

$$
\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{u.} \end{bmatrix}, \; X_\alpha = \begin{bmatrix} 1 & sex_1 & sea_{\mathfrak{F}_1} & A98_1 \\ 1 & sex_2 & sea_{\mathfrak{F}_2} & A98_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & sex_{u.} & sea_{\mathfrak{F}_{u.}} & A98_{u.} \end{bmatrix}, \; \theta_\alpha = \begin{bmatrix} \theta_{\alpha 0} \\ \theta_{\alpha 1} \\ \theta_{\alpha 2} \\ \theta_{\alpha 3} \end{bmatrix},
$$

and

$$
\Sigma_\alpha = diag((\tau_{\alpha 1} + A98_1(\tau_{\alpha 2} - \tau_{\alpha 1})), \ldots, (\tau_{\alpha 1} + A98_{u.}(\tau_{\alpha 2} - \tau_{\alpha 1})))
$$

where $A98_i = 1$ if individual i was first caught in the autumn of 1998 and $A98_i = 0$ otherwise. Note that all off diagonals in Σ_α are zero.

To ensure identifiability, the parameter K_i is assumed to be the same for every individual in the population,

$$
K_i = K, \; i = 1, \ldots, u.
$$

9.1.6   Obtaining the Age at First Capture

The VBGF written in terms of length at first capture can be inverted to obtain the age at first capture,

$$
A_i(\delta_{1i1}) = \frac{\log\left(\frac{\lambda_i}{\lambda_i + K\alpha_i}\right)}{K}, \; i = 1, \ldots, u. \tag{9.5}
$$

$A_i(\delta_{1i1})$ is the age from size 0, not birth, which we are unable to estimate because we are using capture-recapture observations. However, we expect age from size 0 to be a good proxy for age from birth because trout are born small.

As most of our length information is from individuals larger than 450mm

the age estimate is highly model dependent.

9.2   Modeling Sex

We assume that the observed sex of the fish is the outcome of an independent Bernoulli trial,

$$[sex|\theta_g] = \prod_{i=1}^{u.}[sex_i|\theta_g]$$

$$[sex_i|\theta_g] = Bern(\theta_g),$$

(9.6)

where θ_g is the probability of being male.

9.3   Capture-Recapture Model

We fit a CJS model (section 3.2.1), including both the length covariate and availability for capture. Where possible we model the parameters in a similar manner to Barker et al. (2001).

We extend the model for the capture histories specified in equation 3.4 to include availability for capture,

$$[X_2|\theta_p, \epsilon_p, d_2, z, sea, \alpha, \lambda, K, X_1] = \prod_{i=1}^{u.} \prod_{j=\mathfrak{F}_i+1}^{\mathfrak{D}_i} \left\{ p_j^{X_{2ij}}(1-p_j)^{1-X_{2ij}} \right\}^{z_{ij}}.$$

(9.7)

The capture probabilities p_j are assumed to depend on season (autumn/spring) and vary through time according to a random effect,

$$\text{logit}(p_j) = \theta_{p0} + \theta_{p1}sea_j + \epsilon_{pj}, \; j = 2, \ldots, k,$$

where the model for $\epsilon_p = (\epsilon_{p2}, \ldots, \epsilon_{pk})$ is

$$[\epsilon_p | \tau_p] = \prod_{j=2}^{k} [\epsilon_{pj} | \tau_p]$$

$$[\epsilon_{pj} | \tau_p] = N(0, \tau_p).$$

(9.8)

We extend the survival model specified in equation 3.3 to allow survival probabilities to vary between individual and include the additional information from anglers,

$$[d_2 | \theta_S, \epsilon_S, \alpha, \lambda, K, \iota, X_1] = \prod_{i=1}^{N} MN(1, \xi_i(\iota_i)).$$

(9.9)

The probability vector $\xi_i(\iota_i)$ is

$$\xi_i(\iota_i) = \left((1 - S_{i\mathfrak{F}_i}), S_{i\mathfrak{F}_i}(1 - S_{i\mathfrak{F}_i+1}), \ldots, \prod_{h=\mathfrak{F}_i}^{\mathfrak{L}_i-1} S_{ih}(1 - S_{i\mathfrak{L}_i})^{\iota_i}, \right.$$
$$\left. \iota_i \prod_{h=\mathfrak{F}_i}^{\mathfrak{L}_i} S_{ih}(1 - S_{i\mathfrak{L}_i+1}), \ldots, \iota_i \prod_{h=\mathfrak{F}_i}^{k'-1} S_{ih}(1 - S_{ik'}), \iota_i \prod_{h=\mathfrak{F}_i}^{k'} S_{ih} \right).$$

The survival probabilities are assumed to vary through time according to a random effect. The mean of the distribution is assumed to depend on length and age with the length affect changing after the regulation,

$$\text{logit}(S_{ij}) = \theta_{S0} + \theta_{S1}\Lambda'_i(t_j) + \theta_{S2}reg_j + \theta_{S3}\Lambda'_i(t_j)reg_j + \theta_{S4}A_i(t_j) + \epsilon_{Sj},$$

$$i = 1, \ldots, u., \ j = \mathfrak{F}_i, \ldots, k',$$

where $A_i(t_j)$ is the age of individual i at time of sample j found using equation 9.5,

$$A_i(t_j) = A_i(\delta_{1i1}) + \Delta_i(t_j).$$

The value $\Lambda'_i(t_j)$ is the expected length of individual i at time t_j given by

equation 9.1 standardized with a mean of 559.4mm and standard deviation of 76.7mm[4]. We use the expected length as a proxy for true length, which we are not able to obtain because we cannot separate measurement and model error. The model for $\epsilon_S = (\epsilon_{S1}, \ldots, \epsilon_{Sk'})$ is

$$[\epsilon_S | \tau_S] = \prod_{j=1}^{k'} [\epsilon_{Sj} | \tau_S],$$

$$[\epsilon_{Sj} | \tau_S] = N(0, \tau_S).$$

(9.10)

9.4   Information from re-sighting and recoveries

As catching trout is illegal for anglers for a large majority of the winter season, we treat the re-sighting and recovery processes separately for summer and winter. During the winter season, the parameters R_j, R'_j and r_j are used and assumed constant through time, denoted R_{win}, R'_{win} and r_{win}. For the summer months we parameterize in terms of R_j, f_j and ν_j as described in section 4.4.2 as this is a natural parameterization for the process by which the data were collected[5]. When an angler catches a fish they must choose whether to release the fish back into the river (giving a re-sighting if reported) or kill the fish (giving a recovery if reported).

We extend the recovery model specified in equation 4.19 to allow for (i) parameters varying according the season and (ii) individuals not being able

---

[4]We choose these values to agree with the standardized values used by Barker et al. (2001).

[5]Barker et al. (2001) parameterized in terms of R, R' and r as this is the only parameterization available in program MARK (White and Burnham 1999).

to be recovered when lost on capture,

$$[\boldsymbol{Y}_1|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}]$$

$$\propto \prod_{i=1}^{u.} \left\{ \left(r_{i\mathfrak{D}_i}^{Y_{1i\mathfrak{D}_i}}(1 - r_{i\mathfrak{D}_i})^{1-Y_{1i\mathfrak{D}_i}} \right)^{sea_{\mathfrak{D}_i}} \left(r_{win}^{Y_{1i\mathfrak{D}_i}}(1 - r_{win})^{1-Y_{1i\mathfrak{D}_i}} \right)^{1-sea_{\mathfrak{D}_i}} \right\}^{\iota_i}$$

$$(9.11)$$

We extend the re-sighting model specified in equation 4.20 to again allow for (i) parameters varying according the season and (ii) individuals not being able to be re-sighted after they were lost on capture,

$$[\boldsymbol{Y}_2|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, \boldsymbol{X}_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, \boldsymbol{Y}_1]$$

$$\propto \prod_{i=1}^{u.} \left\{ \left(R_{i\mathfrak{D}_i}'^{Y_{2i\mathfrak{D}_i}}(1 - R'_{i\mathfrak{D}_i})^{1-Y_{2i\mathfrak{D}_i}} \right)^{sea_{\mathfrak{D}_i}} \left(R_{win}'^{Y_{2i\mathfrak{D}_i}}(1 - R'_{win})^{1-Y_{2i\mathfrak{D}_i}} \right)^{1-sea_{\mathfrak{D}_i}} \right\}^{\iota_i} \times$$

$$\prod_{j=\mathfrak{F}_i}^{\mathfrak{D}_i-1} \left(R_{ij}^{Y_{2ij}}(1 - R_{ij})^{1-Y_{2ij}} \right)^{sea_j} \left(R_{win}^{Y_{2ij}}(1 - R_{win})^{1-Y_{2ij}} \right)^{1-sea_j}$$

$$(9.12)$$

where the values of R'_{ij} and r_{ij} for the summer months are obtained from equation 4.21,

$$r_{ij} = \frac{(1 - \nu_{ij})f_{ij}}{1 - S_{ij}}$$

$$R'_{ij} = \frac{\frac{f_{ij} - S_{ij}R_{ij}}{1 - S_{ij}} - r_{ij}}{1 - r_{ij}}.$$

The parameters R_j and f_j are very similar and we model these in the same way. As anglers may target larger fish we allow these parameters to depend

on length as well as varying through time according to a random effect,

$$\text{logit}(R_{ij}) = \theta_{R0} + \theta_{R1}\Lambda_i'(t_j) + \epsilon_{Rj}, \; i = 1, \ldots, u., \; j = \{h : sea_h = 1\}$$

$$\text{logit}(f_{ij}) = \theta_{f0} + \theta_{f1}\Lambda_i'(t_j) + \epsilon_{fj}, \; i = 1, \ldots, u., \; j = \{h : sea_h = 1\}.$$

where $j = \{h : sea_h = 1\}$ is the set of all values of j such that $sea_j = 1$. To evaluate the effect of the upper size limit on release rate, we allow all individuals to have the same release rate before the regulation with fish larger than 550mm having a different release rate after the regulation,

$$\text{logit}(\nu_{ij}) = \theta_{\nu0} + \theta_{\nu1} I(\Lambda_i(t_j) > 550) reg_j + \epsilon_{\nu j},$$

$$i = 1, \ldots, u., \; j = \{h : sea_h = 1\}.$$

The model for ϵ_R is

$$[\epsilon_R | \tau_R] = \prod_{j : sea_j = 1} [\epsilon_{Rj} | \tau_R],$$

$$[\epsilon_{Rj} | \tau_R] = N(0, \tau_R). \tag{9.13}$$

where $\prod_{j : sea_j = 1}$ is the product over all values of j for which $sea_j = 1$. The model for ϵ_f is

$$[\epsilon_f | \tau_f] = \prod_{j : sea_j = 1} [\epsilon_{fj} | \tau_f],$$

$$[\epsilon_{fj} | \tau_f] = N(0, \tau_f). \tag{9.14}$$

The model for ϵ_ν is

$$[\epsilon_\nu | \tau_\nu] = \prod_{j : sea_j = 1} [\epsilon_{\nu j} | \tau_\nu],$$

$$[\epsilon_{\nu j} | \tau_\nu] = N(0, \tau_\nu). \tag{9.15}$$

9.5   Movement

We assume that the movement between available and unavailable for capture is first order Markovian and constant through time and across individuals,

$$
[z_2|\theta_z, z_1] = \prod_{i=1}^{u.} \prod_{j=\mathfrak{F}_i+1}^{k} [z_{ij}|z_{ij-1}, \theta_z]
$$

$$
[z_{ij}|z_{ij-1}, \theta_z] = \left(\theta_{z1}^{z_{ij}}(1-\theta_{z1})^{1-z_{ij}}\right)^{z_{ij-1}} \left(\theta_{z2}^{z_{ij}}(1-\theta_{z2})^{1-z_{ij}}\right)^{1-z_{ij-1}},
$$

(9.16)

where $\theta_z = (\theta_{z1}, \theta_{z2})$.

9.6   Posterior Distribution

The specification of the model is complete with prior distributions for all parameters. The posterior distribution is proportional to

$$
[X_2|\theta_p, \epsilon_p, d_2, z, sea, \alpha, \lambda, K, X_1][z_2|\theta_z, z_1][d_2|\theta_S, \epsilon_S, \alpha, \lambda, K, \iota, X_1] \times
$$

$$
[\epsilon_p|\tau_p][\epsilon_S|\tau_S]\prod_{h=0}^{4}[\theta_{Sh}]\prod_{h=0}^{1}[\theta_{ph}]\prod_{h=1}^{2}[\theta_{zh}][\tau_p][\tau_S] \times
$$

$$
[L_1|\alpha, \lambda, K, \tau_1][L_2|\alpha, \lambda, K, \mathcal{B}, \tau_2][L_3|\alpha, \lambda, K, \mathcal{B}, \tau_2][\alpha|\theta_\alpha, \tau_\alpha, sex, sea] \times
$$

$$
[\lambda|\theta_\lambda, \tau_\lambda, sex][sex|\theta_g][K][\mathcal{B}][\theta_g][\tau_1][\tau_2]\prod_{h=1}^{2}[\tau_{\alpha h}][\tau_\lambda][\theta_\alpha][\theta_\lambda] \times
$$

$$
[Y_1|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}] \times
$$

$$
[Y_2|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, Y_1] \times
$$

$$
[\epsilon_f|\tau_f][\epsilon_R|\tau_R][\epsilon_\nu|\tau_\nu]\prod_{h=0}^{1}\{[\theta_{fh}][\theta_{Rh}][\theta_{\nu h}]\}[\tau_f][\tau_R][\tau_\nu][r_{win}][R_{win}][R'_{win}]
$$

9.7   Full Conditional Distributions

9.7.1   Length Model

A summary of the unknowns in the length component, their prior distributions and the updater used to sample from the full conditional distribution is given in table 9.1.

Unknown	Prior Distribution	Updater
α_i	--	MH
λ_i	--	MH
K	$U(0, A)$	MH
$\boldsymbol{\theta}_\alpha$	$MVN(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$	Multivariate Normal
$\tau_{\alpha 1}$	$Ga(a, b)$	Gamma
$\tau_{\alpha 2}$	$Ga(a, b)$	Gamma
$\boldsymbol{\theta}_\lambda$	$MVN(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$	Multivariate Normal
τ_λ	$Ga(a, b)$	Gamma
τ_1	$Ga(a, b)$	Gamma
τ_2	$Ga(a, b)$	Gamma
$\mathcal{B}$	$N(\mu_0, \tau_0)$	Normal

Table 9.1: Table summarizing the Gibbs sampler for length component of the model for Rangitikei trout.

Full Conditional Distribution $[\alpha_i|\cdot]$

The full conditional distribution for α_i, $i = 1, \ldots, u.$ is

$$[\alpha_i|\cdot] \propto [L_1|\alpha, \lambda, K, \tau_1][L_2|\alpha, \lambda, K, \mathcal{B}, \tau_2][L_3|\alpha, \lambda, K, \mathcal{B}, \tau_2] \times$$

$$[\alpha|\boldsymbol{\theta}_\alpha, \tau_\alpha, sex, sea][d_2|\boldsymbol{\theta}_S, \epsilon_S, \alpha, \lambda, K, \iota, X_1] \times$$

$$[Y_1|\boldsymbol{\theta}_f, \boldsymbol{\theta}_R, \boldsymbol{\theta}_\nu, \boldsymbol{\theta}_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}] \times$$

$$[Y_2|\boldsymbol{\theta}_f, \boldsymbol{\theta}_R, \boldsymbol{\theta}_\nu, \boldsymbol{\theta}_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, Y_1]$$

These components are specified in equations 9.2, 9.4, 9.9, 9.11 and 9.12. This full conditional distribution is not of known form.

Full Conditional Distribution $[\lambda_i|\cdot]$

The full conditional distribution for λ_i, $i = 1, \ldots, u$. is

$$[\lambda_i|\cdot] \propto [L_1|\alpha, \lambda, K, \tau_1][L_2|\alpha, \lambda, K, \mathcal{B}, \tau_2][L_3|\alpha, \lambda, K, \mathcal{B}, \tau_2] \times$$

$$[\lambda|\theta_\lambda, \tau_\lambda, sex][d_2|\theta_S, \epsilon_S, \alpha, \lambda, K, \iota, X_1] \times$$

$$[Y_1|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}] \times$$

$$[Y_2|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, Y_1]$$

These components are specified in equations 9.2, 9.3, 9.9, 9.11 and 9.12. This full conditional distribution is not of known form.

Full Conditional Distribution $[K|\cdot]$

The full conditional distribution for K is

$$[K|\cdot] \propto [K][L_1|\alpha, \lambda, K, \tau_1][L_2|\alpha, \lambda, K, \mathcal{B}, \tau_2][L_3|\alpha, \lambda, K, \mathcal{B}, \tau_2] \times$$

$$[d_2|\theta_S, \epsilon_S, \alpha, \lambda, K, \iota, X_1] \times$$

$$[Y_1|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}] \times$$

$$[Y_2|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, Y_1]$$

The prior distribution $[K]$ is specified in table 9.1 and the other components are specified in equations 9.2, 9.9, 9.11 and 9.12. This full conditional distribution is not of known form.

Full Conditional Distribution $[\theta_\alpha|\cdot]$

The full conditional distribution for θ_α is

$$[\theta_\alpha|\cdot] \propto [\alpha|\theta_\alpha, \tau_\alpha, sex, sea][\theta_\alpha].$$

The prior distribution $[\boldsymbol{\theta}_\alpha]$ is specified in table 9.1 where $\boldsymbol{\mu}_0$ is a 4×1 vector and $\boldsymbol{\Sigma}_0$ is a 4×4 matrix. The distribution $[\boldsymbol{\alpha}|\boldsymbol{\theta}_\alpha, \tau_\alpha, sex, sea]$ is the multivariate normal distribution specified in equation 9.4. This gives a multivariate normal full conditional distribution for $\boldsymbol{\theta}_\alpha$,

$$[\boldsymbol{\theta}_\alpha|\cdot] = MVN(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$$

where

$$\boldsymbol{\mu}_n = \boldsymbol{\Sigma}_n \left(\boldsymbol{X}'_\alpha \boldsymbol{\Sigma}^{-1}_\alpha \log(\boldsymbol{\alpha}) + \boldsymbol{\Sigma}^{-1}_0 \boldsymbol{\mu}_0 \right)$$
$$\boldsymbol{\Sigma}^{-1}_n = \boldsymbol{X}'_\alpha \boldsymbol{\Sigma}^{-1}_\alpha \boldsymbol{X}_\alpha + \boldsymbol{\Sigma}^{-1}_0$$

Full Conditional Distribution $[\tau_{\alpha 1}|\cdot]$

The full conditional distribution for $\tau_{\alpha 1}$ is

$$[\tau_{\alpha 1}|\cdot] \propto [\boldsymbol{\alpha}|\boldsymbol{\theta}_\alpha, \tau_\alpha, sex, sea][\tau_{\alpha 1}]$$

The prior distribution $[\tau_{\alpha 1}]$ is specified in table 9.1 and the distribution $[\boldsymbol{\alpha}|\boldsymbol{\theta}_\alpha, \tau_\alpha, sex, sea]$ is the multivariate normal distribution specified in equation 9.4. This gives a gamma full conditional distribution for $\tau_{\alpha 1}$,

$$[\tau_{\alpha 1}|\cdot] = Ga(a', b')$$

where

$$a' = \frac{\sum_{i=1}^{u.} I(A98_i = 0)}{2} + a$$
$$b' = \frac{\sum_{i=1}^{u.} I(A98_i = 0)(\alpha_i - (\boldsymbol{X}_\alpha \boldsymbol{\theta}_\alpha)_i)^2}{2} + b.$$

where $(\boldsymbol{X}_\alpha \boldsymbol{\theta}_\alpha)_i$ is the ith element of the vector $\boldsymbol{X}_\alpha \boldsymbol{\theta}_\alpha$.

Full Conditional Distribution $[\tau_{\alpha 2}|\cdot]$

The full conditional distribution for $\tau_{\alpha 2}$ is

$$[\tau_{\alpha 2}|\cdot] \propto [\alpha|\boldsymbol{\theta}_\alpha, \tau_\alpha, sex, sea][\tau_{\alpha 2}]$$

The prior distribution $[\tau_{\alpha 2}]$ is specified in table 9.1 and the distribution $[\alpha|\boldsymbol{\theta}_\alpha, \tau_\alpha, sex, sea]$ is the multivariate normal distribution specified in equation 9.4. This gives a gamma full conditional distribution for $\tau_{\alpha 2}$,

$$[\tau_{\alpha 2}|\cdot] = Ga(a', b')$$

where

$$a' = \frac{\sum_{i=1}^{u.} I(A98_i = 1)}{2} + a$$
$$b' = \frac{\sum_{i=1}^{u.} I(A98_i = 1)(\alpha_i - (X_\alpha \boldsymbol{\theta}_\alpha)_i)^2}{2} + b$$

Full Conditional Distribution $[\boldsymbol{\theta}_\lambda|\cdot]$

The full conditional distribution for $\boldsymbol{\theta}_\lambda$ is

$$[\boldsymbol{\theta}_\lambda|\cdot] \propto [\boldsymbol{\lambda}|\boldsymbol{\theta}_\lambda, \tau_\lambda, sex][\boldsymbol{\theta}_\lambda].$$

The prior distribution $[\boldsymbol{\theta}_\lambda]$ is specified in table 9.1 where $\boldsymbol{\mu}_0$ is a 2×1 vector and Σ_0 is a 2×2 matrix. The distribution $[\boldsymbol{\lambda}|\boldsymbol{\theta}_\lambda, \tau_\lambda, sex]$ is the multivariate normal distribution specified in equation 9.3. This gives a multivariate normal full conditional distribution for $\boldsymbol{\theta}_\lambda$,

$$[\boldsymbol{\theta}_\lambda|\cdot] = MVN(\boldsymbol{\mu}_n, \Sigma_n)$$

where

$$\mu_n = \Sigma_n \left(X_\lambda' \Sigma_\lambda^{-1} \log(\lambda) + \Sigma_0^{-1} \mu_0 \right)$$

$$\Sigma_n^{-1} = X_\lambda' \Sigma_\lambda^{-1} X_\lambda + \Sigma_0^{-1}.$$

Full Conditional Distribution $[\tau_\lambda | \cdot]$

The full conditional distribution for τ_λ is

$$[\tau_\lambda | \cdot] \propto [\lambda | \theta_\lambda, \tau_\lambda, sex][\tau_\lambda]$$

The prior distribution $[\tau_\lambda]$ is specified in table 9.1 and the distribution $[\lambda | \theta_\lambda, \tau_\lambda, sex]$ is the multivariate normal distribution specified in equation 9.3. This gives a gamma full conditional distribution for τ_λ,

$$[\tau_\lambda | \cdot] = Ga(a', b')$$

where

$$a' = \frac{u.}{2} + a$$

$$b' = \frac{(\log(\lambda) - X_\lambda \theta_\lambda)' \left(\log(\lambda) - X_\lambda \theta_\lambda \right)}{2} + b$$

Full Conditional Distribution $[\tau_1 | \cdot]$

The full conditional distribution for τ_1 is

$$[\tau_1 | \cdot] \propto [L_1 | \alpha, \lambda, K, \tau_1][\tau_1]$$

The prior distribution $[\tau_1]$ is specified in table 9.1 and the distribution $[L_1 | \alpha, \lambda, K, \tau_1]$ is specified in equation 9.2. This gives a gamma full condi-

tional distribution for τ_1,

$$[\tau_1|\cdot] = Ga(a', b')$$

where

$$a' = \frac{\sum_{i=1}^{u.} \omega_{1i}}{2} + a$$

$$b' = \frac{\sum_{i=1}^{u.} \sum_{j=1}^{\omega_{1i}} \left\{ \log(L_{1i}(\delta_{1ij})) - \log(\Lambda_i(\delta_{1ij})) \right\}^2}{2} + b$$

Full Conditional Distribution $[\tau_2|\cdot]$

The full conditional distribution for τ_2 is

$$[\tau_2|\cdot] \propto [L_2|\alpha, \lambda, K, \mathcal{B}, \tau_2][L_3|\alpha, \lambda, K, \mathcal{B}, \tau_2][\tau_2]$$

The prior distribution $[\tau_2]$ is specified in table 9.1 and the other components are specified in equation 9.2. This gives a gamma full conditional distribution for τ_2,

$$[\tau_2|\cdot] = Ga(a', b')$$

where

$$a' = \frac{\sum_{i=1}^{u.} \sum_{j=2}^{3} \omega_{ji}}{2} + a$$

$$b' = \frac{\sum_{i=1}^{u.} \sum_{j=1}^{\omega_{2i}} \left\{ \log(L_{2i}(\delta_{2ij})) - \log(\Lambda_i(\delta_{2ij})) - \mathcal{B} \right\}^2}{2} +$$

$$+ \frac{\sum_{i=1}^{u.} \sum_{j=1}^{\omega_{3i}} \left\{ \log(L_{3i}(\delta_{3ij})) - \log(\Lambda_i(\delta_{3ij})) - \mathcal{B} \right\}^2}{2} + b$$

Full Conditional Distribution $[\mathcal{B}|\cdot]$

The full conditional distribution for the bias $\mathcal{B}$ is

$$[\mathcal{B}|\cdot] \propto [L_2|\alpha, \lambda, K, \mathcal{B}, \tau_2][L_3|\alpha, \lambda, K, \mathcal{B}, \tau_2][\mathcal{B}]$$

The prior distribution $[\mathcal{B}]$ is specified in table 9.1 and the other components are specified in equation 9.2. This results is a normal full conditional distribution for $\mathcal{B}$,

$$[\mathcal{B}|\cdot] = N(\mu_n, \tau_n)$$

where

$$\tau_n = \tau_2 \sum_{i=1}^{u.} \sum_{h=2}^{3} \omega_{hi} + \tau_0$$

$$\mu_n = \frac{a}{\tau_n}$$

and

$$a = \tau_2 \sum_{i=1}^{u.} \sum_{j=1}^{\omega_{2i}} (\log(L_{2i}(\delta_{2ij})) - \log(\Lambda_i(\delta_{2ij}))) +$$

$$\tau_2 \sum_{i=1}^{u.} \sum_{j=1}^{\omega_{3i}} (\log(L_{3i}(\delta_{3ij})) - \log(\Lambda_i(\delta_{3ij}))) + \mu_0\tau_0$$

9.7.2   Capture-Recapture Model

A summary of the unknowns in the capture-recapture component, their prior distributions and the updater used to sample from the full conditional distribution is given in table 9.2.

Unknown	Prior Distribution	Updater
$d_{2i:}^{mis}$	–	Multinomial
θ_{Sh}	$N(\mu_0, \tau_0)$	MH
ϵ_{Sh}	–	MH
τ_S	$Ga(a, b)$	Gamma
θ_{ph}	$N(\mu_0, \tau_0)$	MH
ϵ_{ph}	–	MH
τ_p	$Ga(a, b)$	Gamma

Table 9.2: Table summarizing the Gibbs sampler for capture-recapture component of the model for Rangitikei trout.

Full Conditional Distribution $[d_{2i:}^{mis}|\cdot]$

The full conditional distribution for the missing values in the ith row of d_2 is,

$$[d_{2i:}^{mis}|\cdot] \propto [X_2|\theta_p, \epsilon_p, d_2, z, sea, \alpha, \lambda, K, X_1][d_2|\theta_S, \epsilon_S, \alpha, \lambda, K, \iota, X_1] \times$$

$$[Y_1|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}] \times$$

$$[Y_2|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, Y_1]$$

These components are specified in equations 9.7, 9.9, 9.11 and 9.12. This full conditional distribution is a multinomial distribution with probability vector π_{di} and index 1,

$$[d_{2i:}^{mis}|\cdot] = MN(1, \pi_{di})$$

where

$$\pi_{dij} = \frac{\kappa_{dij}}{\sum_{h=\mathfrak{L}_i}^{t} \kappa_{dih}},$$

$$\kappa_{dij} = \begin{cases} (1-S_{ij}){R'_{ij}}^{Y_{2ij}}(1-R'_{ij})^{1-Y_{2ij}}(1-r_{ij}), & j = \mathfrak{L}_i \\[2em] \displaystyle\prod_{h=\mathfrak{L}_i}^{j-1} S_{ih}(1-S_{ij})R_{i\mathfrak{L}_i}^{Y_{2i\mathfrak{L}_i}}(1-R_{i\mathfrak{L}_i})^{1-Y_{2i\mathfrak{L}_i}} \times \\[1em] \displaystyle\prod_{h=\mathfrak{L}_i+1}^{j-1}(1-R_{ih})\prod_{h=\mathfrak{L}_i+1}^{j}(1-p_{ih})^{z_{ih}}(1-r_{ij})(1-R'_{ij}) & \mathfrak{L}_i < j < k+1 \\[2em] \displaystyle\prod_{h=\mathfrak{L}_i}^{j-1} S_{ih}R_{i\mathfrak{L}_i}^{Y_{2i\mathfrak{L}_i}}(1-R_{i\mathfrak{L}_i})^{1-Y_{2i\mathfrak{L}_i}} \times \\[1em] \displaystyle\prod_{h=\mathfrak{L}_i+1}^{j-1}(1-R_{ih})\prod_{h=\mathfrak{L}_i+1}^{j-1}(1-p_{ih})^{z_{ih}}. & j = k+1 \end{cases}$$

The terms that includes $Y_{2i\mathcal{L}_i}$ are required because it is possible that individual i was last observed by re-sighting between period $\mathcal{L}_i$ and $\mathcal{L}_i + 1$. Note that if an individual is recovered dead then the values of d_2 for that individual are fully known and do not need to be updated. Therefore, we do not need to include a $(1 - Y_{1ij})$ term for the $(1 - R'_{ij})$ components as we know that none of the individuals being updated were recovered dead.

Full Conditional Distribution $[\theta_{Sh}|\cdot]$

The full conditional distribution for θ_{Sh}, $h = 0, \ldots, 4$ is

$[\theta_{Sh}|\cdot] \propto [d_2|\theta_S, \epsilon_S, \alpha, \lambda, K, \iota, X_1] \times$

$[\boldsymbol{Y_1}|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}] \times$

$[\boldsymbol{Y_2}|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, Y_1][\theta_{Sh}].$

The prior distribution $[\theta_{Sh}]$ is specified in table 9.2 and the other components are specified in equations 9.9, 9.11 and 9.12. This full conditional distribution is not of known form.

Full Conditional Distribution $[\epsilon_{Sh}|\cdot]$

The full conditional distribution for ϵ_{Sh}, $h = 1, \ldots, k'$ is

$[\epsilon_{Sh}|\cdot] \propto [d_2|\theta_S, \epsilon_S, \alpha, \lambda, K, \iota, X_1] \times$

$[\boldsymbol{Y_1}|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}] \times$

$[\boldsymbol{Y_2}|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, Y_1][\epsilon_{Sh}|\tau_S]$

The components are specified in equations 9.9, 9.11, 9.12 and 9.10. This full conditional distribution is not of known form.

Full Conditional Distribution $[\tau_S|\cdot]$

The full conditional distribution for τ_S is

$$[\tau_S|\cdot] \propto [\epsilon_S|\tau_S][\tau_S].$$

The prior distribution $[\tau_S]$ is specified in table 9.2 and the distribution $[\epsilon_S|\tau_S]$ is specified in equation 9.10. This gives a gamma full conditional distribution,

$$[\tau_S|\cdot] = Ga\left(\frac{k'}{2} + a, \frac{\sum_{h=1}^{k'}\epsilon_{Sh}^2}{2} + b\right).$$

Full Conditional Distribution $[\theta_{ph}|\cdot]$

The full conditional distribution for θ_{ph}, $h = 0, 1$ is

$$[\theta_{ph}|\cdot] \propto [X_2|\theta_p, \epsilon_p, d_2, z, sea, \alpha, \lambda, K, X_1][\theta_{ph}]$$

The prior distribution $[\theta_{ph}]$ is specified in table 9.2 and the component $[X_2|\theta_p, \epsilon_p, d_2, z, sea, \alpha, \lambda, K, X_1]$ is specified in equation 9.7. This full conditional distribution is not of known form.

Full Conditional Distribution $[\epsilon_{ph}|\cdot]$

The full conditional distribution for ϵ_{ph}, $h = 2, \ldots, k$ is

$$[\epsilon_{ph}|\cdot] \propto [X_2|\theta_p, \epsilon_p, d_2, z, sea, \alpha, \lambda, K, X_1][\epsilon_{ph}|\tau_p], \ h = 2, \ldots, k.$$

The components are specified in equations 9.7 and 9.8. This full conditional distribution is not of known form.

Full Conditional Distribution $[\tau_p|\cdot]$

The full conditional distribution for τ_p is

$$[\tau_p|\cdot] \propto [\epsilon_p|\tau_p][\tau_p].$$

The prior distribution $[\tau_p]$ is specified in table 9.2 and the distribution $[\epsilon_p|\tau_p]$ is specified in equation 9.8. This gives a gamma full conditional distribution,

$$[\tau_p|\cdot] = Ga\left(\frac{k-1}{2}+a, \frac{\sum_{h=2}^{k}\epsilon_{ph}^2}{2}+b\right).$$

9.7.3   Sex Model

A summary of the unknowns in the sex component, their prior distributions and the updater used to sample from the full conditional distribution is given in table 9.3.

Unknown	Prior Distribution	Updater
θ_g	$Be(a,b)$	Beta
sex_i^{mis}	–	Bernoulli

Table 9.3: Table summarizing the Gibbs sampler for sex component of the model for Rangitikei trout.

Full Conditional Distribution $[\theta_g|\cdot]$

The full conditional distribution for θ_g is

$$[\theta_g|\cdot] \propto [sex|\theta_g][\theta_g].$$

The prior distribution $[\theta_g]$ is specified in 9.3 and the distribution $[sex|\theta_g]$ is specified in equation 9.6. This gives a beta full conditional distribution,

$$[\theta_g|\cdot] = Be\left(\sum_i I(sex_i = 1) + a, \sum_i I(sex_i = 0) + b\right).$$

Full Conditional Distribution $[sex_i^{mis}|\cdot]$

The full conditional distribution for the missing sex observation for individual i is

$$[sex_i^{mis}|\cdot] \propto [\alpha|\theta_\alpha, \tau_\alpha, sex][\lambda|\theta_\lambda, \tau_\lambda, sex][sex|\theta_g]$$

These components are specified in equations 9.4, 9.3 and 9.6. This gives a Bernoulli full conditional distribution with $Pr(sex_i^{mis} = 1) = \pi_{gi}$, where

$$\pi_{gi} = \frac{\kappa_{g1i}}{\kappa_{g0i} + \kappa_{g1i}},$$

where

$$\kappa_{g1i} = \theta_g \exp\left(-\frac{1}{2}(\log(\lambda) - X_\lambda^\star\theta_\lambda)'\Sigma_\lambda^{-1}(\log(\lambda) - X_\lambda^\star\theta_\lambda)\right) \times$$
$$\exp\left(-\frac{1}{2}(\log(\alpha) - X_\alpha^\star\theta_\alpha)'\Sigma_\alpha^{-1}(\log(\alpha) - X_\alpha^\star\theta_\alpha)\right)$$
$$\kappa_{g0i} = (1 - \theta_g) \exp\left(-\frac{1}{2}(\log(\lambda) - X_\lambda^\diamond\theta_\lambda)'\Sigma_\lambda^{-1}(\log(\lambda) - X_\lambda^\diamond\theta_\lambda)\right) \times$$
$$\exp\left(-\frac{1}{2}(\log(\alpha) - X_\alpha^\diamond\theta_\alpha)'\Sigma_\alpha^{-1}(\log(\alpha) - X_\alpha^\diamond\theta_\alpha)\right)$$

where $X_\lambda^\star$ is the matrix X_λ with the missing value sex_i set to 1 and $X_\lambda^\diamond$ is the matrix X_λ with the missing value sex_i is set to 0. The same principle applies to $X_\alpha^\star$ and $X_\alpha^\diamond$.

9.7.4   Auxiliary Data

A summary of the unknowns in the recovery and re-sighting component, their prior distributions and the updater used to sample from the full conditional distribution is given in table 9.4.

Unknown	Prior Distribution	Updater
r_{win}	$Be(a, b)$	Beta
R_{win}	$Be(a, b)$	Beta
R'_{win}	$Be(a, b)$	Beta
θ_{Rh}	$N(\mu_0, \tau_0)$	MH
ϵ_{Rh}	$-$	MH
τ_R	$Ga(a, b)$	Gamma
θ_{fh}	$N(\mu_0, \tau_0)$	MH
ϵ_{fh}	$-$	MH
τ_f	$Ga(a, b)$	Gamma
$\theta_{\nu h}$	$N(\mu_0, \tau_0)$	MH
$\epsilon_{\nu h}$	$-$	MH
τ_ν	$Ga(a, b)$	Gamma

Table 9.4: Table summarizing the Gibbs sampler for recovery and re-sighting component of the model for Rangitikei trout.

Full Conditional Distribution $[r_{win}|\cdot]$

The full conditional distribution for r_{win} is

$$[r_{win}|\cdot] \propto [\mathbf{Y}_1|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}][r_{win}].$$

The prior distribution $[r_{win}]$ is specified in table 9.4 and the other distribution is specified in equation 9.11. This gives a beta full conditional distribution,

$$[r_{win}|\cdot] = Be(a', b')$$

where

$$a' = \sum_{i=1}^{u.} Y_{1i\mathfrak{D}_i} I(sea_{\mathfrak{D}_i} = 0)\iota_i + a,$$

$$b' = \sum_{i=1}^{u.} (1 - Y_{1i\mathfrak{D}_i}) I(sea_{\mathfrak{D}_i} = 0)\iota_i + b.$$

Full Conditional Distribution $[R_{win}|\cdot]$

The full conditional distribution for R_{win} is

$$[R_{win}|\cdot] \propto [Y_2|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, Y_1][R_{win}].$$

The prior distribution $[R_{win}]$ is specified in table 9.4 and the other distribution is specified in equation 9.12. This gives a beta full conditional distribution,

$$[R_{win}|\cdot] = Be(a', b')$$

where

$$a' = \sum_{i=1}^{u.} \sum_{j=\mathfrak{F}_i}^{\mathfrak{D}_i-1} Y_{2ij} I(sea_j = 0) + a,$$

$$b' = \sum_{i=1}^{u.} \sum_{j=\mathfrak{F}_i}^{\mathfrak{D}_i-1} (1 - Y_{2ij}) I(sea_j = 0) + b.$$

Full Conditional Distribution $[R'_{win}|\cdot]$

The full conditional distribution for R'_{win} is

$$[R_{win}|\cdot] \propto [Y_2|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, Y_1][R'_{win}].$$

The prior distribution $[R'_{win}]$ is specified in table 9.4 and the other distribution is specified in equation 9.12. This gives a beta full conditional distribution,

$$[R'_{win}|\cdot] = Be(a', b')$$

where

$$a' = \sum_{i=1}^{u.} Y_{2i\mathfrak{D}_i}(1 - Y_{1i\mathfrak{D}_i})I(sea_{\mathfrak{D}_i} = 0)\iota_i + a,$$

$$b' = \sum_{i=1}^{u.} (1 - Y_{2i\mathfrak{D}_i})(1 - Y_{1i\mathfrak{D}_i})I(sea_{\mathfrak{D}_i} = 0)\iota_i + b.$$

Full Conditional Distribution $[\theta_{Rh}|\cdot]$

The full conditional distribution for θ_{Rh}, $h = 0, 1$ is

$$[\theta_{Rh}|\cdot] \propto [\boldsymbol{Y}_1|\boldsymbol{\theta}_f, \boldsymbol{\theta}_R, \boldsymbol{\theta}_\nu, \boldsymbol{\theta}_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}] \times$$

$$[\boldsymbol{Y}_2|\boldsymbol{\theta}_f, \boldsymbol{\theta}_R, \boldsymbol{\theta}_\nu, \boldsymbol{\theta}_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, \boldsymbol{Y}_1][\theta_{Rh}].$$

The prior distribution $[\theta_{Rh}]$ is specified in table 9.4 and the other distributions are specified in equations 9.11 and 9.12. This full conditional distribution is not of known form.

Full Conditional Distribution $[\epsilon_{Rh}|\cdot]$

The full conditional distribution for ϵ_{Rh} for all values of h for which $sea_h = 1$,

$$[\epsilon_{Rh}|\cdot] \propto [\boldsymbol{Y}_1|\boldsymbol{\theta}_f, \boldsymbol{\theta}_R, \boldsymbol{\theta}_\nu, \boldsymbol{\theta}_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}][\epsilon_{Rh}|\tau_R].$$

The distributions are specified by equations 9.11, 9.12 and 9.13. This full conditional distribution is not of known form.

Full Conditional Distribution $[\tau_R|\cdot]$

The full conditional distribution for τ_R is

$$[\tau_R|\cdot] \propto [\epsilon_R|\tau_R][\tau_R].$$

The prior distribution $[\tau_\nu]$ is specified in table 9.4 and $[\epsilon_R|\tau_R]$ is specified in equation 9.13. This gives a gamma full conditional distribution,

$$[\tau_R|\cdot] = Ga\left(\frac{\sum_{h=1}^{k'} I(sea_h = 1)}{2} + a, \frac{\sum_{h=1}^{k'} I(sea_h = 1)\epsilon_{Rh}^2}{2} + b\right).$$

Full Conditional Distribution $[\theta_{fh}|\cdot]$

The full conditional distribution for θ_{fh}, $h = 0, 1$ is

$$[\theta_{fh}|\cdot] \propto [Y_1|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}] \times$$

$$[Y_2|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, Y_1][\theta_{fh}].$$

The prior distribution $[\theta_{fh}]$ is specified in table 9.4 and the other distributions are specified in equations 9.11 and 9.12. This full conditional distribution is not of known form.

Full Conditional Distribution $[\epsilon_{fh}|\cdot]$

The full conditional distribution for ϵ_{fh} for all h for which $sea_h = 1$,

$$[\epsilon_{fh}|\cdot] \propto [Y_1|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}] \times$$

$$[Y_2|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, Y_1][\epsilon_{fh}|\tau_f].$$

The distributions are specified in equations 9.11, 9.12 and 9.14. This full conditional distribution is not of known form.

Full Conditional Distribution $[\tau_f|\cdot]$

The full conditional distribution for τ_f is

$$[\tau_f|\cdot] \propto [\epsilon_f|\tau_f][\tau_f].$$

The prior distribution $[\tau_f]$ is specified in table 9.4 and $[\epsilon_f|\tau_f]$ is specified in equation 9.14. This gives a gamma full conditional distribution,

$$[\tau_f|\cdot] = Ga\left(\frac{\sum_{h=1}^{k'} I(sea_h = 1)}{2} + a, \frac{\sum_{h=1}^{k'} I(sea_h = 1)\epsilon_{fh}^2}{2} + b\right).$$

Full Conditional Distribution $[\theta_{\nu h}|\cdot]$

The full conditional distribution for $\theta_{\nu h}$, $h = 0, 1$ is

$$[\theta_{\nu h}|\cdot] \propto [Y_1|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}] \times$$

$$[Y_2|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, Y_1][\theta_{\nu h}].$$

The prior distribution $[\theta_{\nu h}]$ is specified in table 9.4 and the other components are specified in equations 9.11 and 9.12. This full conditional distribution is not of known form.

Full Conditional Distribution $[\epsilon_{\nu h}|\cdot]$

The full conditional distribution for $\epsilon_{\nu h}$, for all h for which $sea_h = 1$,

$$[\epsilon_{\nu h}|\cdot] \propto [Y_1|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, r_{win}] \times$$

$$[Y_2|\theta_f, \theta_R, \theta_\nu, \theta_S, d_2, X_1, sea, \iota, \alpha, \lambda, K, \epsilon_f, \epsilon_R, \epsilon_\nu, \epsilon_S, R_{win}, R'_{win}, Y_1][\epsilon_{\nu h}|\tau_\nu].$$

The components are specified in equations 9.11, 9.12 and 9.15. This full conditional distribution is not of known form.

Full Conditional Distribution $[\tau_\nu|\cdot]$

The full conditional distribution for τ_ν is

$$[\tau_\nu|\cdot] \propto [\epsilon_\nu|\tau_\nu][\tau_\nu].$$

The prior distribution $[\tau_\nu]$ is specified in table 9.4 and $[\epsilon_\nu | \tau_\nu]$ is specified in equation 9.15. This gives a gamma full conditional distribution,

$$[\tau_\nu | \cdot] = Ga\left(\frac{\sum_{h=1}^{k'} I(sea_h = 1)}{2} + a, \frac{\sum_{h=1}^{k'} I(sea_h = 1)\epsilon_{\nu h}^2}{2} + b\right).$$

9.7.5   Movement Model

A summary of the unknowns in the movement component, their prior distributions and the updater used to sample from the full conditional distribution is given in table 9.5.

Unknown	Prior Distribution	Updater
z_{ij}^{mis}	–	Bernoulli
θ_{z1}	$Be(a, b)$	Beta
θ_{z2}	$Be(a, b)$	Beta

Table 9.5: Table summarizing the Gibbs sampler for movement component of the model for Rangitikei trout.

Full Conditional Distribution $[z_{ij}^{mis} | \cdot]$

The full conditional distribution for the missing value z_{ij}^{mis} is

$$[z_{ij}^{mis}] \propto [X_2 | \theta_p, \epsilon_p, d_2, z, sea, \alpha, \lambda, K, X_1][z_2 | \theta_z, z_1]$$

These components are specified in equations 9.7 and 9.16. This full conditional distribution is a Bernoulli distribution with $Pr(z_{ij} = 1) = \pi_{zij}$ for all missing values of z_2, where

where

$$\pi_{zij} = \frac{\kappa_{z1ij}}{\kappa_{z0ij} + \kappa_{z1ij}},$$

$$\kappa_{z0ij} = (z_{ij-1}(1 - \theta_{z1}) + (1 - z_{ij-1})(1 - \theta_{z2}))(z_{ij+1}\theta_{z2} + (1 - z_{ij+1})(1 - \theta_{z2})),$$

$$\kappa_{z1ij} = (z_{ij-1}\theta_{z1} + (1 - z_{ij-1})\theta_{z2})(1 - p_{ij})(z_{ij+1}\theta_{z1} + (1 - z_{ij+1})(1 - \theta_{z1})).$$

$$(9.17)$$

When the value for z_{ik}^{mis} is updated, the terms involving z_{ij+1} are not included, hence equation 9.17 becomes,

$$\kappa_{z0ij} = (z_{ij-1}(1 - \theta_{z1}) + (1 - z_{ij-1})(1 - \theta_{z2})),$$

$$\kappa_{z1ij} = (z_{ij-1}\theta_{z1} + (1 - z_{ij-1})\theta_{z2})(1 - p_{ij}).$$

After individual i has died the value of κ_{z1ij} in equation 9.17 no longer has the $(1 - p_{ij})$ term and we update z_{ij}^{mis} from its prior distribution,

$$\kappa_{z1ij} = (z_{ij-1}\theta_{z1} + (1 - z_{ij-1})\theta_{z2})(z_{ij+1}\theta_{z1} + (1 - z_{ij+1})(1 - \theta_{z1})).$$

Full Conditional Distribution $[\theta_{z1}|\cdot]$

The full conditional distribution for θ_{z1} is

$$[\theta_{z1}|\cdot] \propto [z_2|\theta_z, z_1][\theta_{z1}].$$

The prior distribution $[\theta_{z1}]$ is specified in table 9.5 and $[z_2|\theta_z, z_1]$ is specified in equation 9.16. This gives a beta full conditional distribution,

$$[\theta_{z1}|\cdot] = Be(a', b')$$

where

$$a' = \sum_{i=1}^{u.} \sum_{j=\mathfrak{F}_i+1}^{\mathfrak{D}_i} I(z_{ij} = 1)I(z_{ij-1} = 1) + a,$$

$$b' = \sum_{i=1}^{u.} \sum_{j=\mathfrak{F}_i+1}^{\mathfrak{D}_i} I(z_{ij} = 0)I(z_{ij-1} = 1) + b.$$

Full Conditional Distribution $[\theta_{z2}|\cdot]$

The full conditional distribution for θ_{z2} is

$$[\theta_{z2}|\cdot] \propto [z_2|\theta_z, z_1][\theta_{z2}].$$

The prior distribution $[\theta_{z2}]$ is specified in table 9.5 and $[z_2|\theta_z, z_1]$ is specified in equation 9.16. This gives a beta full conditional distribution,

$$[\theta_{z2}|\cdot] = Be(a', b')$$

where

$$a' = \sum_{i=1}^{u.} \sum_{j=\mathfrak{F}_i+1}^{\mathfrak{D}_i} I(z_{ij} = 1)I(z_{ij-1} = 0) + a,$$

$$b' = \sum_{i=1}^{u.} \sum_{j=\mathfrak{F}_i+1}^{\mathfrak{D}_i} I(z_{ij} = 0)I(z_{ij-1} = 0) + b.$$

9.8   Results

The following prior distributions were used for the length model

$$[K] = U(0,5),$$

$$[\boldsymbol{\theta}_\lambda] = MVN\left((0,0)', 10000\boldsymbol{I}_2\right),$$

$$[\tau_\lambda] = Ga(0.001, 0.001),$$

$$[\boldsymbol{\theta}_\alpha] = MVN\left((0,0,0,0)', 10000\boldsymbol{I}_4\right),$$

$$[\tau_{\alpha h}] = Ga(0.001, 0.001), \ h = 1, 2,$$

$$[\tau_h] = Ga(0.001, 0.001), \ h = 1, 2,$$

$$[\mathcal{B}] = N(0, 0.0001)$$

where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. The following prior distributions were used for the capture-recapture model

$$[\theta_{Sh}] = N(0, 0.0001), \ h = 0, \ldots, 4,$$

$$[\tau_S] = Ga(0.001, 0.001),$$

$$[\theta_{ph}] = N(0, 0.0001), \ h = 0, 1,$$

$$[\tau_p] = Ga(0.001, 0.001).$$

The following prior distributions were used for the recoveries and re-sightings

$$[r_{win}] = Be(1,1),$$

$$[R_{win}] = Be(1,1),$$

$$[R'_{win}] = Be(1,1),$$

$$[\theta_{Rh}] = N(0, 0.0001), \ h = 0,1,$$

$$[\tau_R] = Ga(0.001, 0.001),$$

$$[\theta_{fh}] = N(0, 0.0001), \ h = 0,1,$$

$$[\tau_f] = Ga(0.001, 0.001),$$

$$[\theta_{\nu h}] = N(0, 0.0001), \ h = 0,1,$$

$$[\tau_\nu] = Ga(0.001, 0.001).$$

The following prior distributions were used for the movement model

$$[\theta_{zh}] = Be(1,1), \ h = 1,2.$$

The Gibbs sampler was run on 3 independent chains started from different values. Each chain had an adaptive phase of $20,000$ iterations to optimize the Metropolis Hastings and RJMCMC algorithms as specified in section 5.2.2. As there was high autocorrelation in the values, each chain was run for a $100,000$ iteration burn-in followed by a further $500,000$ iterations. The chains mixed and were combined to give a posterior sample of $1,500,000$ iterations.

The regulation appears to have improved the release rate of fish over 550mm (Figure 9.2), with a posterior probability of 0.9999 that the rate of release by an angler increased after the upper size limit for a fish larger than 550mm. However, even though there was an improvement in release rate, it

appears that not everybody upheld the new regulation (Figure 9.3).
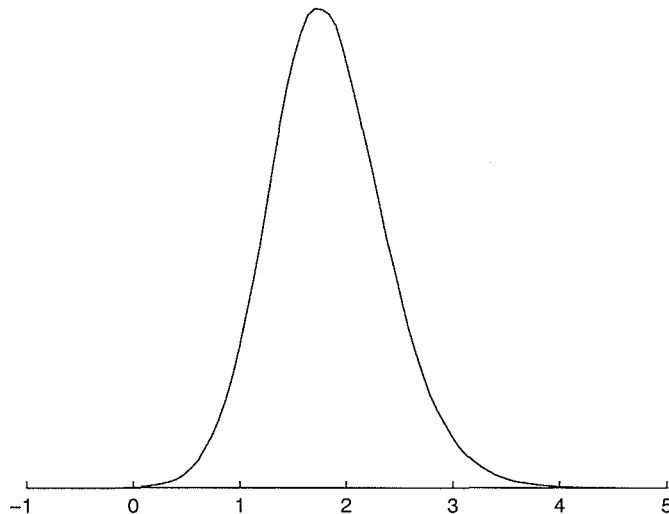
Figure 9.2: Estimate of the posterior distribution for $\theta_{\nu 1}$ the change in release probability on the logit scale for fish larger 550mm after the regulation.

Before the regulation it appears that neither age nor length affect the survival of the fish (Figure 9.4). However, it appears the survival of the larger fish improved after the upper size limit, with a posterior probability of 0.9535 that $\theta_{S3} > 0$ (Figure 9.5). These result differ from the analysis of Barker et al. (2001), who concluded that

1. Length has a significantly positive effect on survival before the regulation. The approximate 95% confidence interval for the length effect before the regulation is $(0.492, 1.790)$.

2. The effect of length on survival increased after the regulation. The approximate 95% confidence interval for the length effect after the regulation is $(1.676, 6.086)$.

These results not only yield the wrong conclusion for survival before the
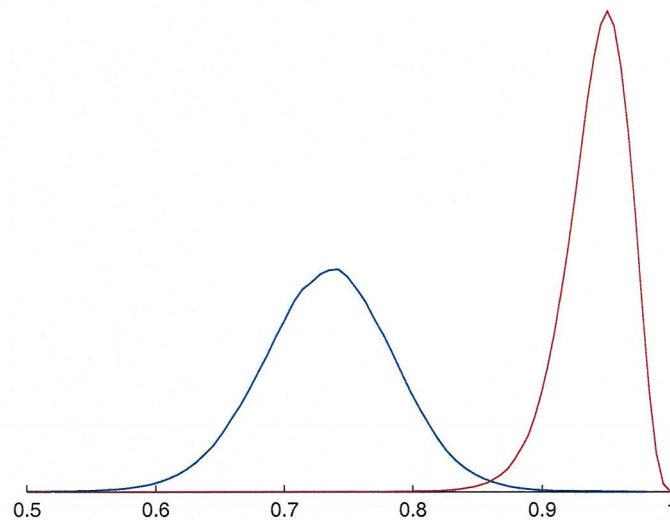
Figure 9.3: Estimate of the posterior distribution for ν before the regulation (blue line) and after the regulation (red line) for fish over 550mm.

regulation but also over-estimate the positive effect of the regulation for larger fish. The spurious results obtained in Barker et al. (2001) show the danger of not fully accounting for the missing data properly.

The changes in both survival and release rate suggest that the upper size limit imposed has worked effectively. Not only did the release rate for fish over 550mm appear to increase (Figures 9.2 and 9.3) but this increase appears to have led to an increased survival in the larger fish (Figure 9.5). As the larger fish in the headwaters are thought to be the breeders in the population, the regulation appears to have improved the survival of those individuals that sustain the Rangitikei rainbow trout fishery.

An interesting observation is that sex appears to have no effect on length. The marginal posterior distributions of parameters $\theta_{\alpha 1}$ and $\theta_{\lambda 1}$ both suggest that sex does not influence either the size or instantaneous growth rate at
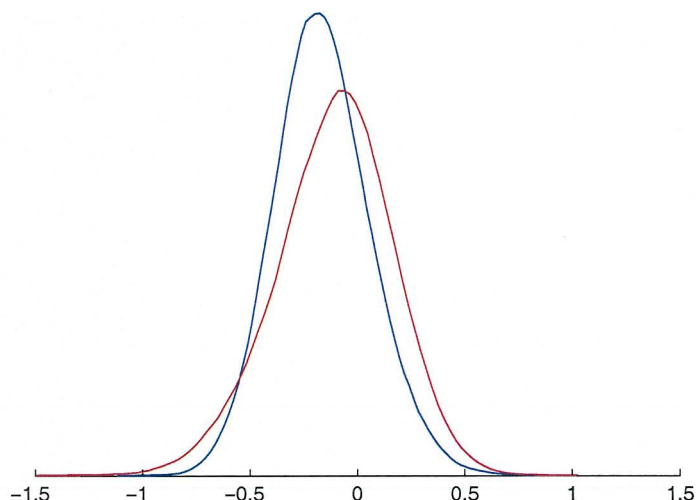
Figure 9.4: Estimate of the posterior distribution for θ_{S1}, the effect of length on the logit of survival before the regulation (blue line) and θ_{S4}, the effect of age on the logit of survival (red line).

first capture (Figure 9.6).

It appears that the angler observations are not only unbiased but also have lower error than the observations from fish and game officers (Figure 9.7). One possible explanation is that the screening process used by fish and game to ensure measurement were correct taken may have been too strict.

Another interesting aspect of the model is that the joint posterior distribution of parameters $(\theta_{z1}, \theta_{z2}, \theta_{p0}, \theta_{p1})$ appears to be bi-modal. As it is difficult to visualize in $4-5$ dimensions, the contours of the kernel density smooth[6] of the pairs $(\theta_{z1}, \mathrm{expit}(\theta_{p0}))$ and $(\theta_{z2}, \mathrm{expit}(\theta_{p0}))$ are given as an example in figure 9.8. These show that when the probability of capture in the autumn season is relatively low $\mathrm{expit}(\theta_{p0}) \approx (0.05, 0.15)$ then the

---

[6]Note that for computational reasons we thinned the posterior by 1 in 100 giving a posterior sample of 15000 for this calculation.
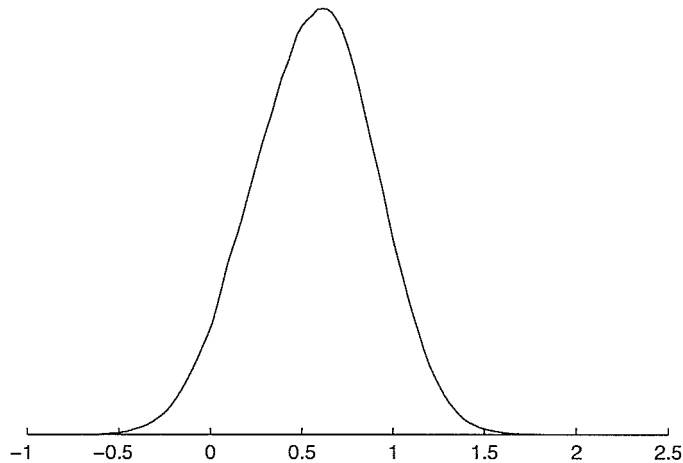
Figure 9.5: Estimate of the posterior distribution for θ_{S3}, the interaction between length and the regulation. There is an estimated posterior probability of 0.9589 that $\theta_{S3} > 0$.

movement rates are relatively high, $\theta_{z1} \approx (0.15, 0.45)$ and $\theta_{z2} \approx (0.5, 1)$. However, when probability of capture in the autumn season is relatively high $\text{expit}(\theta_{p0}) \approx (0.3, 0.8)$ then the movement rates are relatively low $\theta_{z1} \approx (0.025, 0.075)$ and $\theta_{z2} \approx (0.05, 0.15)$. In other words, the data are supporting two alternatives:

1. Relatively low capture probabilities with individuals moving freely between the alternative states of being available and unavailable for capture.

2. Relatively high capture probabilities with individuals strongly preferring to be unavailable for capture.

A summary of the marginal posterior distributions for all of the parameters is given in table 9.6.

Parameter	2.5%	50%	97.5%
θ_{S0}	0.68884	1.3438	1.9693
θ_{S1}	-0.58081	-0.16867	0.29796
θ_{S2}	-0.65822	0.075286	0.80839
θ_{S3}	-0.090616	0.58007	1.1946
θ_{S4}	-0.6886	-0.095583	0.39034
σ_S	0.027785	0.15848	0.93182
θ_{p0}	-3.3516	-2.1096	2.8787
θ_{p1}	-0.14422	0.97292	9.1337
σ_p	0.051169	0.61233	3.7678
θ_{R0}	-2.2446	-1.868	-1.5368
θ_{R1}	-0.38636	-0.029746	0.32181
σ_R	0.023098	0.075888	0.36572
θ_{f0}	-1.7549	-1.4934	-1.2415
θ_{f1}	-0.40243	-0.071213	0.23949
σ_f	0.022774	0.071534	0.28998
$\theta_{\nu 0}$	0.52841	1.0189	1.6159
$\theta_{\nu 1}$	0.83882	1.818	2.9772
σ_ν	0.026959	0.13795	0.77223
r_{win}	0.021391	0.045773	0.087734
R_{win}	0.0021409	0.0084038	0.019501
R'_{win}	0.0064037	0.042565	0.095496
θ_{z1}	0.033294	0.20669	0.82139
θ_{z2}	0.063517	0.44529	0.97784
$\theta_{\alpha 0}$	6.2923	6.3116	6.3308
$\theta_{\alpha 1}$	-0.01429	0.0049146	0.024217
$\theta_{\alpha 2}$	0.027018	0.047741	0.068451
$\theta_{\alpha 3}$	-0.196	-0.1331	-0.070327
$\sigma_{\alpha 1}$	0.091639	0.099568	0.10762
$\sigma_{\alpha 2}$	0.23036	0.26842	0.31777
$\theta_{\lambda 0}$	3.1258	3.596	4.0451
$\theta_{\lambda 1}$	-0.36292	-0.036537	0.28222
σ_λ	0.39667	0.56548	0.76092
K	0.39245	0.7722	1.3731
σ_1	0.026706	0.033453	0.042459
σ_2	0.011929	0.016918	0.024468
B	-0.0059469	0.0047047	0.015625
θ_g	0.37399	0.41523	0.45724

Table 9.6: Table of posterior 2.5%, 50% and 97.5% quantiles for all parameters from the Rangitikei trout model.
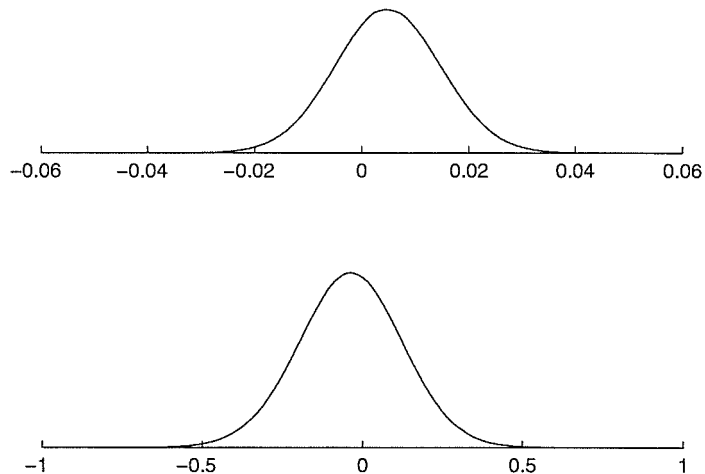
Figure 9.6: Estimates of the posterior distribution for $\theta_{\alpha 1}$, the effect of sex on the log of length at first capture (top) and $\theta_{\lambda 1}$, the effect of sex on the log of derivative at first capture (bottom).

9.9   Extensions

One appealing extension involves modeling the biomass of the fishery. The biomass is the total mass of all fish in the population. There are two steps required in order to include a model of this nature. The first is to include the continuous covariate weight into the model[7]. The second step is to include first captures into the model so that all fish in the population are included in the model. Once these components are specified, the biomass at time t is obtained by adding up the weight of each individual alive at t.

A joint model needs to be specified which can broken into conditional components. Denoting birth as B, length as L, weight as W and mortality

---

[7]It is not necessary to also have the covariate length, however, if it is available then we are able to model the joint distribution of weight and length, improving our model fit.
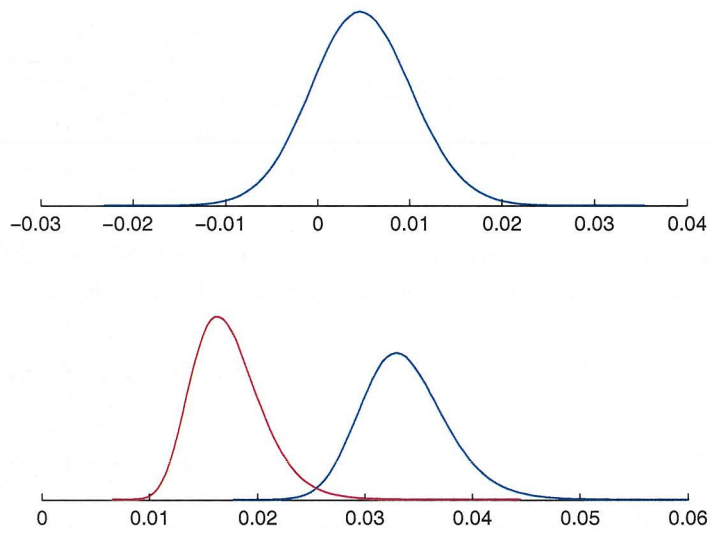
Figure 9.7: Estimate of the posterior distribution for $\mathcal{B}$, the bias in angler length observations (top) and the comparison between the estimated posterior distribution σ_1 (blue line), the fish and game officer standard deviation and σ_2 (red line), the angler standard deviation (bottom).
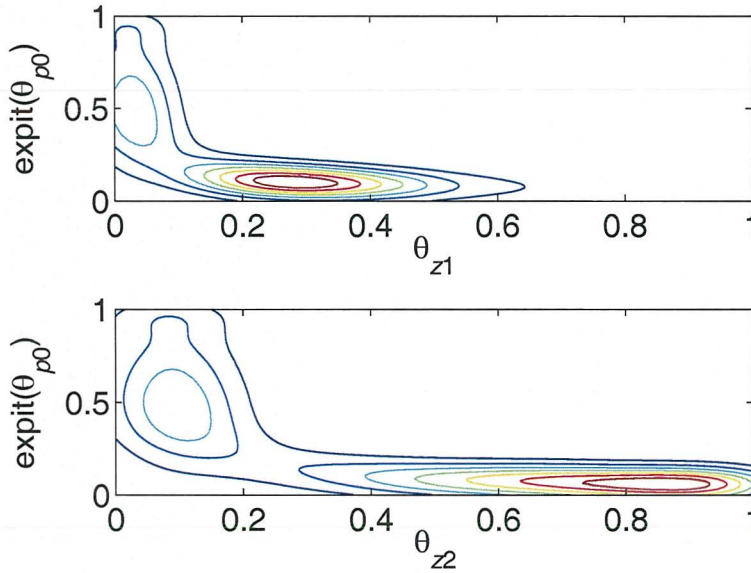
Figure 9.8: Contour plot of the estimate of the joint posterior distribution for the parameters θ_{z1} and $\text{expit}(\theta_{p0})$ (top) and θ_{z2} and $\text{expit}(\theta_{p0})$ (bottom).

as D, one possibility for a model would be,

$$[B, L, W, D] = [B][L|B][W|L, B][D|W, L, B],$$

where $[B]$ is a model for birth, $[L|B]$ models the length conditional on birth, $[W|L, B]$ is model for weight conditional on length and $[D|W, L, B]$ models the survival process.

A model of this kind allows the fisheries managers to derive biologically sensible models for each of the demographic components in turn. If the managers were worried about a perceived drop in the biomass they could specify a set of models that investigated which component(s) is responsible for the decrease and provide clarification on what area(s) resources need to be focused. This type of model empowers managers to target the key areas in improving the state of the fishery.

Part IV

Summary

Chapter 10

Discussion

We develop a common framework for capture-recapture models which includes all missing data directly into the model using data augmentation. The inclusion of the missing data means that we are able to model in terms of the complete data likelihood (CDL). The CDL has a natural biological factorization which separates the terms that specify the demographic processes from the terms that model the sampling process. This means that the focus can be on building interesting biological models instead of accounting for the complexities of the sampling process.

One of the major advantages of modeling in terms of the CDL is that all latent variables are in the likelihood available to be used in the model. We make use of the latent variables to model the birth process in terms of the per-capita birth rates, a natural demographic parameter. The approaches of Pradel (1996) and Link and Barker (2005) both approximate the per-capita birth rate with an index that finds the expected number of births per individual *expected* to be alive at the beginning of the period. Using the CDL means we no longer have to approximate the per-capita birth rate and are able to parameterize in terms of the expected number of births per

individual *actually* alive at the beginning of the sample. This is a natural parameter that many demographers would choose to use when modeling the fecundity of a population.

We also use the latent population size as a covariate in modeling. Even though population size can be derived using standard methods, it is impossible to then use this value without using an errors-in-variable approach. However, using the CDL means that it is possible to include population size as a covariate. In chapter 8 we presented an example where we had a density dependent model with both the per-capita birth rates and survival probabilities dependent on the population size at the beginning of the period.

The latent variables that have been included in the examples provided are far from an exhaustive list. Another group of latent variables that could be included into the model are future unobservable quantities that we wish to predict, usually to investigate population dynamics (Besbeas et al. 2005, Thomas et al. 2005, Newman et al. 2006). Barker, Schofield, Armstrong, and Armstrong (2008) do this when examining the population dynamics of a newly introduced population of North Island saddlebacks (*Philesturnus ru-fusater*) on Mokoia Island, New Zealand. A capture-resighting model is used with additional information on fecundity to fit a model where the survival probabilities and per-capita birth rates are density dependent. Of particular interest is quantifying the equilibrium population size on the island as well as discovering how many birds can be safely be removed every three years in order to establish populations elsewhere. To predict the population size, the number of births and deaths in each year for δ years after last capture are included into the model as latent variables.

Barker et al. (2008) predicted forward $\delta = 100$ years after the last sam-

pling period and found the median equilibrium population size to be around 130 females with a 95% credible interval of between 70 and 400 females. The population reached equilibrium around $10 - 15$ years after the last sampling period.

The predictions were repeated with a varying number of birds harvested every three years to establish new populations. The simulations suggested that up to 80 birds could safely be removed every three years as the removal of 80 birds (or less) resulted in a posterior probability of extinction of less than 5%.

A further advantage of the hierarchical framework developed is that many of the common capture recapture models can be obtained by selecting the appropriate conditional likelihood components and multiplying them together. Biologists are able to specify the appropriate conditional likelihood components that account for the process by which they collect the data and also specify the appropriate components that include the demographic processes of interest. The nice feature of this is that very complex models can be written in terms a series of simpler conditional likelihood components. A good example of this is the Rangitikei River trout model. This is a very complex CJS model that has length as an individual-specific time-varying covariate, as well as auxiliary information on re-sightings, recoveries and movement (though availability for capture). Using the CDL we are able to break this into conditional likelihood components that model:

- Length covariate.

- Mortality after first capture conditional on length.

- Movement after first capture.

- Sampling after first capture conditional on mortality, length and move-

ment.

- Dead recoveries conditional on length and mortality.

- Live re-sightings conditional on length and mortality.

The ability to break up complex problems in terms common conditional likelihood components has wide implications for the development of software. The currently used capture-recapture software, in particular MARK (White and Burnham 1999) and M-SURGE (Choquet et al. 2004) specify the likelihood for every distinct model separately. To include a new model, such as the one used for the Rangitikei trout, would require a new likelihood to be coded into the program. Coding models of this complexity using the observed data likelihood is very difficult and time consuming. In contrast, the only features that need to be specified when using the CDL is any new conditional likelihood components that are not previously specified. One obvious consequence is that once the initial work has been put into coding the standard conditional likelihood components, the maintainance of the program is far easier than for the standard software. The problem is that there is currently no computer program capable of implementing such a framework. WinBUGS can be used for a variety of models, but the lack of flexibility, particularly with no RJMCMC step, make it impractical for full implementation. One of the challenges in the future is developing software that is able to fit models of this kind, either through improving current software such as WinBUGS, or starting afresh. If an easy-to-use software package is not developed, then these methods, despite all their advantages, are unlikely to be widely adopted as users would need to write their own MCMC code.

A related issue for the future are the computational algorithms them-

selves. The development of flexible software requires flexible algorithms with which to fit the models. The problem is that the flexible algorithms, such as Metropolis-Hastings can often have slow convergence with high autocorrelation in the Markov chain. Even the complex models investigated in chapters 8 and 9 with specialized algorithms exhibited high autocorrelation in the Markov chain. Therefore, as biologists and ecologists continue to collect data of increasing complexity, the computational algorithms need to improve so that valid inference can be achieved quickly and easily.

Perhaps the most interesting and exciting change in the future will come from the data being collected. As was mentioned at the EURING technical meeting in Dunedin in January 2007, technological advances are making standard capture-recapture experiments obsolete. Researchers are preferring methods, such as radio telemetry, that claim to have perfect detection of individuals. In this case, we really do have the data we would like to observe. The only change required to incorporate this data in the CDL framework is that we would no longer need to include a conditional distribution for the capture histories X, as they are fully known. However, even though the technological advances claim to have perfect detection, it seems highly likely that there will still be missing data. It may be that we are not always able to pick up a signal of the radio tags. Perhaps the battery has died, or maybe the animal has strayed into an area where signal cannot be picked up. It seems reasonable to assume that irrespective of the technology that prevails, there will always be missing data that need to be accounted for.

References

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

Armstrong, D. P., Davidson, R. S., Perrott, J. K., Royguard, J., and Buchanan, L. (2005), "Density-dependent population growth in a reintroduced population of North Island saddlebacks," *Journal of Animal Ecology*, 74, 160 – 170.

Barker, R. J. (1997), "Joint Modeling of Live-Recapture, Tag-Resight, and Tag-Recovery Data," *Biometrics*, 53, 666 – 677.

Barker, R. J., Burnham, K. P., and White, G. C. (2004), "Encounter history modeling of joint mark-recapture, tag-resighting and tag-recovery data under temporary emigration." *Statistica Sinica*, 14, 1037 – 1055.

Barker, R. J., Fletcher, D. J., and Scofield, P. (2002), "Measuring density dependence in survival from mark-recapture data." *Journal of Applied Statistics*, 29, 305 – 313.

Barker, R. J. and Kavalieris, L. (2001), "Efficiency Gain from Auxiliary Data Requiring Additional Nuisance Parameters," *Biometrics*, 57, 563 566.

Barker, R. J. and Schofield, M. R. (2008), "Putting Markov Chain Back

Into Markov Chain Monte Carlo," *Journal of Applied Mathematics and Decision Sciences*, In Press.

Barker, R. J., Schofield, M. R., Armstrong, D. P., and Armstrong, R. S. (2008), "Bayesian Hierarchical Models for Inference About Population Growth," *Journal of Environmental and Ecological Statistics*, In Press.

Barker, R. J., Taylor, P. H., and Smith, S. C. (2001), "Effect of a change in fishing regulations on the survival and capture probabilities of rainbow trout in the upper Rangitikei river, New Zealand," *North American Journal of Fisheries Management*, 22, 465–473.

Barker, R. J. and White, G. C. (2004), "Towards the mother-of-all-models: customised construction of the mark-recapture likelihood function," *Animal Biodiversity and Conservation*, 27, 177 – 185.

Besbeas, P., Freeman, S. N., and Morgan, B. J. T. (2005), "The Potential of Integrated Population Modelling," *Australia and New Zealand Journal of Statistics*, 47, 35 – 48.

Bishop, J. A., Cook, L. M., and Muggleton, J. (1978), "The response of two species of moths to industrialization in northwest England. II Relative fitness of morphs and population size." *Philosophical Transaction of the Royal Society of London*, 281, 517 – 540.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis*, MIT Press.

Bonner, S. J. and Schwarz, C. J. (2006), "An Extension of the Cormack-Jolly-Seber Model for Continuous Covariates with Application to *Microtus pennsylvanicus*," *Biometrics*, 62, 142 – 149.

Briand, L. C., Emam, K. E., Freimut, B., and Laitenberger, O. (1997), "Quantitative Evaluation of Capture-Recapture Models to Control Software Inspections," *issre*, 00, 234.

Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2000a), "Bayesian Animal Survival Estimation," *Statistical Science*, 15, 357 – 376.

Brooks, S. P., Catchpole, E. A., Morgan, B. J. T., and Barry, S. C. (2000b), "On the Bayesian analysis of ring-recovery data." *Biometrics*, 56, 951 – 956.

Brooks, S. P. and Gelman, A. (1998), "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, 7, 434–455.

Brownie, C., Hines, J. E., Nichols, J. D., Pollock, K. H., and Hestbeck, J. B. (1993), "Capture-recapture studies for multiple strata including non-Markovian transition probabilities," *Biometrics*, 49, 1173 – 1187.

Burnham, K. P. (1991), "On a unified theory for release-resampling of animal populations." in *Proceedings 1990 Taipei Simposium in Statistics.*, eds. Chao, M. T. and Cheng, P. E., Institute of Statistical Science, Academia Sinica, pp. 11 – 35.

— (1993), "A theory for combined analysis of ring recovery and recapture data," in *Marked Individuals in Bird Population Studies.*, eds. Lebreton, J.-D. and North, P., Birkhauser Verlag, Basel., pp. 199 – 213.

Burnham, K. P. and Anderson, D. R. (1998), *Model Selection and Inference*, Springer.

Burnham, K. P. and Overton, W. S. (1978), "Estimation of the size of

a closed population when capture probabilities vary among animals," *Biometrika*, 65, 625 – 633.

Carothers, A. D. (1973), "The effects of unequal catchability on Jolly-Seber estimates." *Biometrics*, 29, 79 – 100.

Choquet, R., Reboulet, A. M., Pradel, R., Gimenez, O., and Lebreton, J. D. (2004), "M-SURGE: new software specifically designed for multistate capture-recapture models," *Animal Biodiversity and Conservation*, 27, 1 – 9.

Cormack, R. M. (1964), "Estimates of Survival from the Sighting of Marked Animals," *Biometrika*, 51, 429 – 438.

— (1989), "Log-linear Models for Capture-Recapture," *Biometrics*, 45, 395 – 413.

Cox, D. R. and Miller, H. D. (1965), *The Theory of Stochastic Processes*, New York: Chapman & Hall.

Crosbie, S. F. and Manly, B. F. J. (1985), "Parsimonious Modelling of Capture-Mark-Recapture Studies," *Biometrics*, 41, 385 – 398.

Damien, P., Wakefield, J., and Walker, S. (1999), "Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables," *Journal of the Royal Statistical Society*, 61, 331 – 334.

Dupuis, J. A. (1995), "Bayesian estimation of movement and survival probabilities from capture-recapture data," *Biometrika*, 82, 761 – 772.

Durban, J. W. and Elston, D. A. (2005), "Mark-Recapture With Occasion and Individual Effects: Abundance Estimation Through Bayesian Model

Selection in a Fixed Dimensional Parameter Space," *Journal of Agricultural, Biological, and Environmental Statistics*, 10, 291 – 305.

Eveson, J. P., Laslett, G. M., and Polacheck, T. (2004), "An integrated model for growth incorporating tag-recapture, length-frequency, and direct aging data," *Canadian Journal of Fisheries and Aquatic Science*, 61, 292 306.

Fabens, A. J. (1965), "Properties and fitting of the von Bertalanffy growth curve," *Growth*, 29, 265 – 289.

Fienberg, S. E. (1972), "The multiple recapture census for closed populations and incomplete 2^k contingency tables," *Biometrika*, 59, 591 – 603.

Francis, R. I. C. C. (1988a), "Are Growth Parameters Estimated from Tagging and Age-Length Data Comparable?" *Canadian Journal of Fisheries and Aquatic Science*, 45, 936 942.

— (1988b), "Maximum Likelihood estimation of growth and growth variability from tagging data," *New Zealand Journal of Marine and Freshwater Research*, 22, 42 – 51.

Freeman, M. F. and Tukey, J. W. (1950), "Transformations related to the angular and square root," *Annals of Mathematical Statistics*, 21, 607 – 611.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis*, Chapman & Hall/CRC, 2nd ed.

Geman, S. and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721 – 741.

Gilks, W. R. and Wild, P. (1992), "Adaptive rejection sampling for Gibbs sampling," *Applied Statistics*, 41, 337 – 348.

Green, P. (1995), "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711 – 732.

Han, C. and Carlin, B. (2001), "Markov chain Monte Carlo methods for computing Bayes factors: a comparative review." *Journal of the American Statistical Association*, 96, 1122–1133.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57, 97 – 109.

Higdon, D. M. (1998), "Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications," *Journal of the American Statistical Association*, 93, 585 – 595.

Hoadley, B. (1970), "A Bayesian Look at Inverse Linear Regression," *Journal of the American Statistical Association*, 65, 356 – 369.

Holzmann, H., Munk, A., and Zucchini, W. (2006), "On Identifiability in Capture-Recapture Models," *Biometrics*, 62, 934 – 939.

Hook, E. B. and Regal, R. R. (1995), "Capture-Recapture Methods in Epidemiology: Methods and Limitations," *Epidemiol Rev*, 17, 243 – 264.

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. (1989), "On the Statistical Analysis of Capture Experiments," *Biometrika*, 76, 133 – 140.

Hunter, W. G. and Lamboy, W. F. (1981), "A Bayesian Analysis of the Linear Calibration Problem," *Technometrics*, 23, 323 – 328.

James, I. (1991), "Estimation of von Bertalanffy Growth Curve Parameters from Recapture Data," *Biometrics*, 47, 1519 1530.

Johnson, D. S. and Hoeting, J. A. (2003), "Autoregressive Models for Capture-Recaputure Data: A Bayesian Approach," *Biometrics*, 59, 340 – 349.

Jolly, G. M. (1965), "Explicit Estimates from Capture-Recapture Data with Both Death and Immigration-Stochastic Model," *Biometrika*, 52, 225 – 247.

King, R. and Brooks, S. P. (2002), "Bayesian model discrimination for multiple strata capture-recapture data," *Biometrika*, 89, 785 – 806.

Laslett, G. M., Eveson, J., and Polacheck, T. (2002), "A flexible maximum likelihood approach for fitting growth curves to tag-recapture data," *Canadian Journal of Fisheries and Aquatic Science*, 59, 976 – 986.

Lebreton, J. D., Burnham, K., Clobert, J., and Anderson, D. (1992), "Modelling survival and testing biological hypotheses using marked animals: A unified approach with case studies," *Ecological Monographs*, 62, 67–118.

Link, W. A. (2003), "Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities," *Biometrics*, 59, 1123 – 1130.

Link, W. A. and Barker, R. J. (2005), "Modeling Association among Demographic Parameters in Analysis of Open Population Capture-Recapture Data," *Biometrics*, 61, 46 – 54.

— (2006), "Model Weights and the Foundations for Multimodel Inference," *Ecology*, 87, 2626 – 2635.

Lui, J. S. (2001), *Monte Carlo Strategies in Scientific Computing*, Springer-Verlag New York.

Maller, R. A. and deBoer, E. S. (1988), "An analysis of two methods of fitting the von Bertalanffy curve to capture-recapture data," *Australian Journal of Marine and Freshwater Research*, 39, 459 – 466.

Meng, X. L. and van Dyk, D. (1997), "The EM algorithm - an old folk song sung to a fast new tune (with discussion)," *Journal of the Royal Statistical Society B*, 59, 511 – 597.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equation of State Calculations by Fast Computing Machine," *Journal of Chemical Physics*, 21, 1087 – 1091.

Neal, P. and Roberts, G. (2005), "A case study in non-centering for data augmentation: Stochastic epidemics," *Statistics and Computing*, 15, 315 – 327.

Neal, R. (2003), "Slice sampling (with discussion)," *Annals of Statistics*, 31, 705 – 767.

Newman, K. B., Buckland, S. T., Lindley, S. T. Thomas, L., and Fernández, C. (2006), "Hidden Process Models for Animal Population Dynamics," *Ecological Applications*, 16, 74 – 86.

Osborne, C. (1991), "Statistical Calibration: A Review," *International Statistical Review*, 59, 309 – 336.

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.

Papaspiliopoulos, O., Roberts, G. O., and Skold, M. (2003), "Non-centered parameterisations for hierarchical models and data augmentation," in *Bayesian Statistics 7*, eds. Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., Oxford University Press.

Pledger, S. (2000), "Unified maximum likelihood estimates for closed capture-recapture models using mixtures," *Biometrics*, 56, 434 – 442.

Pledger, S., Pollock, K. H., and Norris, J. L. (2003), "Open Capture-Recapture Models with Heterogeneity: I. Cormack-Jolly-Seber Model," *Biometrics*, 59, 786–794.

Pollock, K. H. (1982), "A Capture-Recapture Design Robust to Unequal Probability of Capture," *Journal of Wildlife Management*, 46, 752 – 757.

Pollock, K. H., Hines, J. E., and Nichols, J. D. (1985), "Goodness-of-Fit Tests for Open Capture-Recapture Models," *Biometrics*, 41, 399 – 410.

Pollock, K. H., Nichols, J. D., Brownie, C., and Hines, J. E. (1990), "Statistical inference for capturerecapture experiments," *Wildlife Monographs*, 62, 1 – 135.

Pradel, R. (1996), "Utilization of Capture-Mark-Recapture for the Study of Recruitment and Population Growth Rate," *Biometrics*, 52, 703 – 709.

— (2005), "Multievent: An Extension of Multistate Capture-Recapture Models to Uncertain States," *Biometrics*, 61, 442 – 447.

Raftery, A. E., Newton, M. A., Satagopan, J. M., and Krivitsky, P. N. (2007), "Estimating the Integrated Likelihood via Posterior Simulation Using the Harmonic Mean Identity," in *Bayesian Statistics 8*, eds. Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., Oxford University Press, pp. 1 – 45.

Ratkowsky, D. A. (1986), "Statistical Properties of Alternative Parameterizations of the von Bertalanffy Growth Curve," *Canadian Journal of Fisheries and Aquatic Science*, 43, 742 – 747.

Robert, C. P. and Casella, G. (2004), *Monte Carlo Statistical Methods*, Springer.

Schofield, M. R. and Barker, R. J. (2007a), "Flexible hierarchical mark-recapture modeling for open populations using WinBUGS," Submitted to *Ecological and Environmental Statistics*.

— (2007b), "A unified capture-recapture model," Submitted to *Journal of Agricultural, Biological, and Environmental Statistics*.

— (2008), "A Further Step Toward the Mother-of-all-Models: Flexibility and Functionality in the Modeling of Capture-Recapture Data," *Journal of Environmental and Ecological Statistics*, In Press.

Schwarz, C. J. and Arnason, A. N. (1996), "A General Methodology for the Analysis of Capture-Recapture Experiments in Open Populations," *Biometrics*, 52, 860 – 873.

Schwarz, C. J., Bailey, R. E., Irvine, J. R., and Dalziel, F. C. (1993a), "Estimating Salmon Spawning Escapement Using Capture-Recapture Methods," *Canadian Journal of Fisheries and Aquatic Science*, 50, 1181 – 1197.

Schwarz, C. J., Schweigert, J. F., and Arnason, A. N. (1993b), "Estimating migration rates using tag-recovery data," *Biometrics*, 49, 177 – 193.

Seber, G. A. F. (1965), "A Note on the Multiple-Recapture Census," *Biometrika*, 52, 249 – 259.

Seber, G. A. F. and Schwarz, C. J. (2002), "Capture-recapture: before and after EURING 2000," *Journal of Applied Statistics*, 29, 5 – 18.

Spiegelhalter, D. J., Thomas, A., G., B. N., and Lunn, D. (2003), *WinBUGS Version 1.4 User Manual*, MRC Biostatistics Unit.

Tanner, M. A. and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation (with discussion)," *Journal of the American Statistical Society*, 82, 529 – 550.

Thomas, L., Buckland, S. T., Newman, K. B., and Harwood, J. (2005), "A Unified Framework for Modelling Wildlife Population Dynamics," *Australia and New Zealand Journal of Statistics*, 47, 19 – 34.

van Dyk, D. A. and Meng, X. L. (2001), "The Art of Data Augmentation," *Journal of Computational and Graphical Statistics*, 10, 1–50.

von Bertalanffy, L. (1938), "A quantitative theory of organic growth (Inquries on growth laws. II)," *Human Biology*, 10, 181 – 213.

Wang, Y. G. (2004), "Estimation of Growth Parameters from Multiple-Recapture Data," *Biometrics*, 60, 670–675.

White, G. C. and Burnham, K. P. (1999), "Program MARK: survival estimation from populations of marked animals." *Bird Study*, 46S, 120 – 139.

Williams, B., Nichols, J., and Conroy, M. (2002), *Analysis and management of animal populations.*, Academic Press.

Appendix A

Notation

A.1   Matrices and Vectors

Throughout the thesis, scalar values are distinguished from vectors and matrices. Any vector or matrix is specified with bold face, for example, $\boldsymbol{a}$. The ith value from the vector $\boldsymbol{a}$ is a scalar and is denoted by a_i.

The notation $\boldsymbol{b}_{:3}$ is the third column of the matrix $\boldsymbol{b}$. Likewise, the notation $\boldsymbol{b}_{5:}$ denotes the fifth row of the matrix $\boldsymbol{b}$. The notation $\boldsymbol{b}_{:(3:7)}$ denotes a matrix containing columns 3 to 7 of matrix $\boldsymbol{b}$.

A.2   Probability Distribution Functions Used

Normal Distribution

The normal distribution will be parameterized in terms of either the variance or the precision. The normal distribution written in terms of the precision is

$$[y|\mu, \tau] = N(\mu, \tau)$$
$$= \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau}{2}(y - \mu)^2\right).$$

The model in terms of the variance is

$$[y|\mu, \sigma^2] = N(\mu, \sigma^2)$$

$$= \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right),$$

where

$$\sigma^2 = \frac{1}{\tau}.$$

If a parameter other than σ or τ is used, then it will be made clear how the model is parameterized.

Multivariate Normal Distribution

The multivariate normal distribution for $y = (y_1, \ldots, y_n)'$ is

$$[y|\mu, \Sigma] = MVN(\mu, \Sigma)$$

$$= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right),$$

where μ is a $n \times 1$ mean vector and Σ is the $n \times n$ symmetric, positive definite variance-covariance matrix.

Log-Normal Distribution

The log-normal distribution for $y \geq 0$ is

$$[y|\mu, \sigma^2] = LN(\mu, \sigma^2)$$

$$= \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(\log(y)-\mu)^2\right) \frac{1}{y}.$$

As with the normal distribution, the log-normal distribution can also be written in terms of the precision τ.

Logit-Normal Distribution

The log-normal distribution for $0 \leq y \leq 1$ is

$$[y|\mu, \sigma^2] = \text{logit} N(\mu, \sigma^2)$$
$$= \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(\text{logit}(y) - \mu)^2\right) \frac{1}{y(1-y)}.$$

As with the normal distribution, the logit-normal distribution can also be written in terms of the precision τ.

Logistic Distribution

The logistic distribution for y is

$$[y|\mu, \sigma] = Logistic(\mu, \sigma)$$
$$= \frac{\exp\left(-\frac{y-\mu}{\sigma}\right)}{\sigma\left(1 + \exp\left(-\frac{y-\mu}{\sigma}\right)\right)^2}$$

where μ is the location and σ is the scale parameter.

Gamma Distribution

The gamma distribution for $y \geq 0$ is

$$[y|\alpha, \beta] = Ga(\alpha, \beta)$$
$$= \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp(-\beta y),$$

where $\alpha > 0$, $\beta > 0$.

Beta Distribution

The beta distribution for $0 \leq y \leq 1$ is

$$[y|\alpha, \beta] = Be(\alpha, \beta)$$
$$= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1},$$

where $\alpha > 0$, $\beta > 0$.

Dirichlet Distribution

The Dirichlet distribution is the multivariate extension of the beta distribution. The Dirichlet distribution for $\boldsymbol{y} = (y_1, \ldots, y_n)$ is

$$[\boldsymbol{y}|\boldsymbol{\alpha}] = Dir(\boldsymbol{\alpha})$$
$$= \frac{\Gamma\left(\sum_{j=1}^{n} \alpha_j\right)}{\prod_{j=1}^{n} \Gamma(\alpha_j)} \prod_{j=1}^{n} y_j^{\alpha_j - 1},$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ and $0 < y_j < 1$, $\alpha_j > 0$ for $j = 1, \ldots, n$.

Uniform Distribution

The uniform distribution for y is

$$[y|a, b] = U(a, b)$$
$$= \frac{1}{b-a},$$

where $y \in [a, b]$ and $b > a$.

Discrete Uniform Distribution

The discrete uniform distribution for y is

$$[y|a,b] = DU(a,b)$$
$$= \frac{1}{b-a},$$

where y, a, b are all integers, $y \in [a, b]$ and $b > a$.

Bernoulli Distribution

The Bernoulli distribution for y is

$$[y|p] = Bern(p)$$
$$= p^y (1-p)^{1-y},$$

where y either takes the value 0 or 1 and $p \in [0, 1]$.

Binomial Distribution

The binomial distribution for y is

$$[y|p, N] = Bin(N, p)$$
$$= \frac{N!}{y!(n-y)!} p^y (1-p)^{N-y},$$

where y and N are both integers, $y \in [0, N]$ and $p \in [0, 1]$.

Multinomial Distribution

The multinomial distribution is the multivariate extension of the binomial

distribution. The multinomial distribution for $y = (y_1, \ldots, y_n)$ is

$$[y|p, N] = MN(N, p)$$
$$= \frac{N!}{\prod_{j=1}^{n} y_j!} \prod_{j=1}^{n} p_j^{y_j},$$

where $p = (p_1, \ldots, p_n)$, $\sum_{j=1}^{n} p_j = 1$, N is an integer and the vector of observations y is a non-negative vector with $\sum_{j=1}^{n} y_j = N$.

Poisson Distribution

The Poisson distribution for y is

$$[y|\lambda] = Pois(\lambda)$$
$$= \frac{\lambda^y \exp(-\lambda)}{y!},$$

where $y \geq 0$ is an integer and $\lambda > 0$.

A.3   Summary of Data and Parameters

Capture-Recapture Data and Parameters

$\aleph$	Total number of individuals ever available for capture.
k	No. of capture occasions in the study.
k'	No. of re-sighting/recovery samples.
$u.$	The number of unique individuals caught during the study.
t_j	The time at sample j.
X^{obs}	A $u. \times k$ matrix of observed capture histories. The value $X_{ij}^{obs} = 1$ means that individual i was observed in sample j with $X_{ij}^{obs} = 0$ otherwise.
X	Matrix of the complete capture histories, including the $\aleph - u.$ observed individuals.

Capture-Recapture Data and Parameters

X_1 Information on capture up to and including first capture.

X_2 Information on capture after first capture.

b A matrix denoting birth. The value $b_{ij} = 1$ means that individual i is born between sample j and $j + 1$ and $b_{ij} = 0$ otherwise. The value $b_{i0} = 1$ means that individual i was alive at the beginning of the study.

d A matrix denoting death. The value $d_{ij} = 1$ means that individual i died between sample j and $j + 1$ and $d_{ij} = 0$ otherwise. The value $b_{ik} = 1$ means that individual i was alive at the end of the study.

d_1 Information on death up to first capture.

d_2 Information on death after first capture.

ι A vector denoting loss on capture. The value $\iota_i = 0$ means that an individual was lost on capture, with $\iota_i = 1$ otherwise.

z A vector/matrix of covariates. One such covariate is availability for capture.

Y_1 Matrix of dead recoveries. The value $Y_{1ij} = 1$ means that individual i was recovered dead between sample j and $j + 1$, with $Y_{1ij} = 0$ otherwise.

Y_2 Matrix of live re-sightings. The value $Y_{2ij} = 1$ means that individual i was re-sighted between sample j and $j + 1$, with $Y_{2ij} = 0$ otherwise.

U_j The number of unmarked individuals in the population immediately before sample j.

u_j The number of unmarked individuals caught in sample j.

U_j' The number of unmarked individuals in the population immediately after sample j that survived until sample $j + 1$.

Capture-Recapture Data and Parameters

B_j The number of individual born between sample j and $j+1$.

p_j Probability of capture in sample j.

S_j Probability of surviving from sample j to sample $j+1$.

χ_j Probability of an individual not being seen again after sample j.

η_j The per-capita birth rate between sample j and $j+1$. This is the expected number of births between sample j and $j+1$ for every individual alive in sample j.

β_j The multinomial probability of birth between sample j and $j+1$ given that $\aleph$ individuals were available for capture during the study.

N_j Population size in sample j.

$\mathfrak{B}_i$ Interval censored sample of birth for individual i. The value $\mathfrak{B}_i = j$ means that $b_{ij} = 1$.

$\mathfrak{F}_i$ Sample of first capture for individual i.

$\mathfrak{L}_i$ Sample of last capture for individual i.

$\mathfrak{D}_i$ Interval censored sample of death for individual i. The value $\mathfrak{D}_i = j$ means that $d_{ij} = 1$.

R_j Probability of an individual being re-sighted between sample j and $j+1$ given that it is alive in both j and $j+1$.

R'_j Probability of an individual being re-sighted between samples j and $j+1$ given that it died between j and $j+1$ and was not recovered dead.

r_j Probability that an individual who dies between sample j and $j+1$ will be recovered dead.

f_j Probability of an individual being re-sighted or recovered between j and $j+1$ given they were alive in sample j.

Capture-Recapture Data and Parameters

ν_j Probability of an individual being re-sighting alive given they were re-sighted or recovered between samples j and $j + 1$.

Table A.1: Notation used for the capture-recapture models.

VB Length Parameters

ω_{1i} The number of captures for individual i in which a length was recorded.

ω_{2i} The number of recoveries for individual i in which a length was recorded. Note that ω_{2i} must either be 0 or 1.

ω_{3i} The number of re-sightings for individual i in which a length was recorded.

δ_{1ih} The time at capture h for individual i.

δ_{2ih} The time at recovery h for individual i.

δ_{3ih} The time at re-sighting h for individual i.

L_1 Information of the length values obtained from the capture-recapture study. The value $L_{1i}(\delta_{1ih})$ is the recorded length for individual i in capture h.

L_2 Information of the length values obtained from the recovery observations. The value $L_{2i}(\delta_{2ih})$ is the recorded length for individual i in recovery h.

L_3 Information of the length values obtained from the re-sighting observations. The value $L_{3i}(\delta_{3ih})$ is the recorded length for individual i in recovery h.

$L_{\infty i}$ Asymptotic Length of individual i.

$A_i(t)$ Age of individual i at time t.

α_i The length of individual i at first capture.

VB Length Parameters

K_i Growth Rate Parameter for individual i.

λ_i The instantaneous growth rate of individual i at first capture

$\mathcal{B}$ Length bias when caught by an angler

$\Delta_i(t)$ Time between first capture of individual i and time t

Table A.2: Notation used for the VB model.

Appendix B

Appendix: CDL vs ODL

Consider a coin tossing experiment where we record the value of N coin tosses. The observation x_j takes the value 1 if toss j is a head and 0 otherwise. After the data were recorded but before they could be analyzed, some of the pages of data went missing. Instead of N trials, we now only observe N^{obs} of the coin tosses with $N^{mis} = N - N^{obs}$ missing values[1]. We are interested in estimating the unknown probability of tossing a head, denoted p. We can summarize the data through counting up the trials. The total number of heads is y,

$$y = \sum_{j=1}^{N} x_j,$$

of which y^{obs} heads were actually observed,

$$y^{obs} = \sum_{j=1}^{N^{obs}} x_j.$$

In order to determine whether we need to include the missingness mechanism we write $\mathcal{I} = (\mathcal{I}_1, \ldots, \mathcal{I}_N)$, where $\mathcal{I}_j = 1$ if toss j was observed and

---

[1]For convenience we index the observed data as $j = 1, \ldots, N^{obs}$ and the missing data as $j = N^{obs} + 1, \ldots, N$.

$\mathcal{I}_j$ otherwise. We assume that the model for $\mathcal{I}$ does not depend on the data,

$$[\mathcal{I}|\phi],$$

where ϕ is the parameter that describes how the data went missing. We assume that p and ϕ are distinct meaning that the data are MCAR and the missingness mechanism is ignorable.

The complete data likelihood treats the missing data as if we actually observed it. Specifying the coin tosses to be independent Bernoulli distributions,

$$\mathcal{L}_C(p; x, N) \propto [x|p]$$
$$= \prod_{j=1}^{N} [x_j|p]$$
$$= p_j^y (1 - p_j)^{N-y},$$

where $x = (x_1, \ldots, x_N)$. Note that we do not include the missingness mechanism in the CDL because it is ignorable.

The standard approach for analyzing this data is to use the observed data likelihood,

$$\mathcal{L}_O(p; x^{obs}, N) \propto \prod_{j=1}^{N^{obs}} [x_j|p] \prod_{j=N^{obs}+1}^{N} \sum_{x_j=0}^{1} [x_j|p]$$
$$= p_j^{y^{obs}} (1 - p_j)^{N^{obs}-y^{obs}} \prod_{j=N^{obs}+1}^{N} \sum_{x_j=0}^{1} p_j^{x_j} (1 - p_j)^{1-x_j}$$
$$= p_j^{y^{obs}} (1 - p_j)^{N^{obs}-y^{obs}}.$$

Unsurprisingly, this is the likelihood for p that would be used if we assumed we had only undertaken N^{obs} coin tosses in the first place. Placing a $Be(\alpha, \beta)$

prior distribution on p yields a

$$Be\left(y^{obs} + \alpha, N^{obs} - y^{obs} + \beta\right)$$

posterior distribution for p.

An alternative approach is to use data augmentation and fit the model from the CDL specified above. We obtain a posterior sample for p through Gibbs sampling (see section 5.2.2), alternating between

$$[p|\cdot] = Be\left(y + \alpha, N - y + \beta\right)$$

$$[x_j|\cdot] = Bern(p), \ j = N^{obs} + 1, \ldots, N,$$

where $y = \sum_{j=1}^{N} x_j$. One could wonder whether the Gibbs sampler will sample from the required posterior distribution for p. However, with data $N^{obs} = 100$, $N^{mis} = 20$, $y^{obs} = 43$ and a prior distribution

$$p \sim Be(1,1),$$

a kernel density smooth of a posterior sample of $50,000$ using the CDL approach is in excellent agreement to the direct posterior distribution (figure B.1), confirming that the two approaches are equivalent.
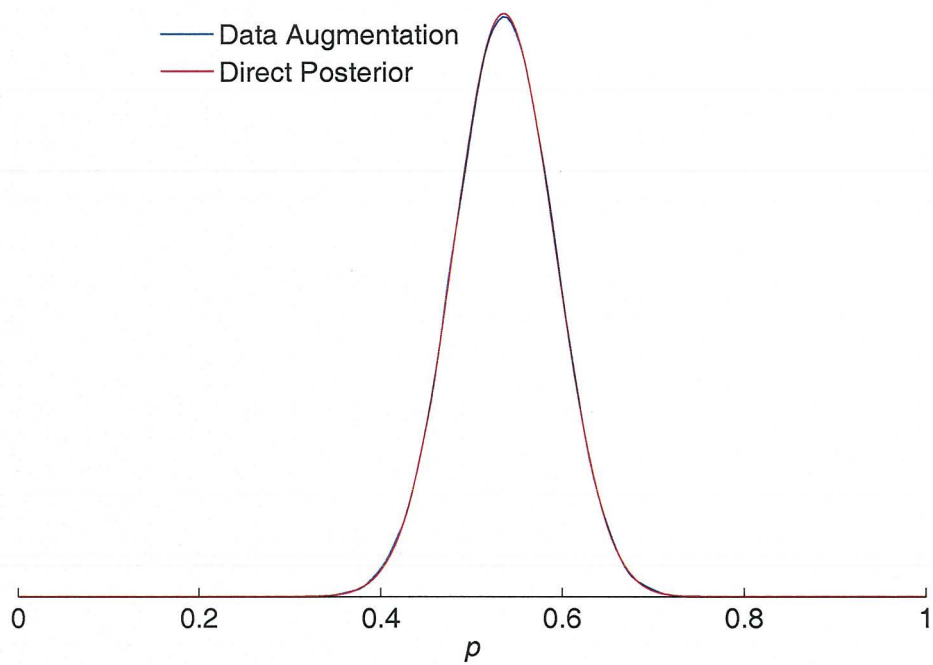
Figure B.1: Comparison between the true posterior distribution for p (red-line) and the posterior distribution estimated by a kernel density smooth of $50,000$ samples found using Gibbs sampling from the CDL (blue line).

Appendix C

Appendix: WinBUGS code

C.1 WinBUGS code for CJS Model

```
1   # The entire model is placed between model{}
2   model{
3     for(i in 1:udot){
4       for(j in first[i]+1:k){
5         # X[i,j] = 1 if individual i was caught in sample j
6         X[i,j] ~ dbern(peeX[i,j])
7         # individuals can only be caught while alive
8         peeX[i,j] <- pcap[i,j]*alive[i,j]
9       }
10
11      # Cumsv[i,last[i]] is the cumulative survival rate for
12      # individual i between first and last capture.
13      Cumsv[i,first[i]] <- 1
14      for(j in first[i]+1:last[i]){
15        Cumsv[i,j] <- Cumsv[i,j-1]*sv[i,j-1]
```

```
16    }

17

18    # Cumsvd[i,j] is the cumulative survival rate from

19    # last capture until sample j for individual i

20    Cumsvd[i,last[i]] <- 1

21    for(j in last[i]+1:k){

22       Cumsvd[i,j] <- Cumsvd[i,j-1]*sv[i,j-1]

23    }

24

25    # svdmat[i,j] is used in the multinomial for the missing

26    # d component. It is the conditional probability of

27    # dying in sample j given they were alive at last[i]

28    svdmat[i,last[i]] <- Cumsvd[i,last[i]]*(1-sv[i,last[i]])

29    for(j in last[i]+1:k){

30       svdmat[i,j] <- Cumsvd[i,j]*(1-sv[i,j])

31    }

32    svdmat[i,k+1] <- 0

33

34    # Generates the multinomial draw for the missing value of d

35    # conditional on being alive at last[i]

36    d[i,last[i]:k+1] ~ dmulti(svdmat[i,last[i]:k+1],1)

37

38    # alive[i,j] = 1 if individual i is alive in sample j.

39    # It is a deterministic function of d.

40    alive[i,first[i]] <- 1

41    for(j in first[i]+1:k){

42       alive[i,j] <- alive[i,j-1]*(1-d[i,j-1])
```

```
43    }

44

45    # includes the survival probabilities from first[i]

46    # to last[i]

47    wd[i] ~ dbern(Cumsv[i,last[i]]) # ones trick

48

49    sv[i,k] <- 0 # so that sum(d) = 1

50    # S[j] is the time-specific survival probability

51    # p[j] is the time-specific capture probabililty

52    for(j in first[i]:k-1){

53      sv[i,j] <- S[j]

54      pcap[i,j+1] <- p[j+1]

55    }

56    }

57

58    for(j in 1:k-1){

59      S[j] ~ dbeta(1,1) # equivalent to Unif(0,1)

60      p[j+1] ~ dbeta(1,1)

61    }

62  }
```

C.2   Data for CJS Model

```
list(

X=structure(.Data = c(

1,1,1,1

,1,0,0,0

,...
```

```
,0,1,1,1),.Dim=c(199,4)),

udot=199,

k=4,

first=c(1,1,...,2),

last=c(4,1,...,4),

wd=c(1,1,...,1),

d=structure(.Data = c(

0,0,0,NA,NA

,NA,NA,NA,NA,NA

,...

,0,0,0,NA,NA),.Dim=c(199,5))

)
```

C.3 WinBUGS Code for JS Model

The WinBUGS code that is required to fit the JS Model is

```
1   # The entire model is placed between model{}

2  model{

3    for(i in 1:udot){

4      for(j in first[i]+1:k){

5        # X[i,j] = 1 if individual i was caught in sample j

6        X[i,j] ~ dbern(peeX[i,j])

7        # individuals can only be caught while alive

8        peeX[i,j] <- pcap[i,j]*alive[i,j]

9      }

10

11     # Cumsv[i,last[i]] is the cumulative survival rate for
```

```
12      # individual i between first and last capture.

13      Cumsv[i,first[i]] <- 1

14      for(j in first[i]+1:last[i]){

15          Cumsv[i,j] <- Cumsv[i,j-1]*sv[i,j-1]

16      }

17

18      # Cumsvd[i,j] is the cumulative survival rate from

19      # last capture until sample j for individual i

20      Cumsvd[i,last[i]] <- 1

21      for(j in last[i]+1:k){

22          Cumsvd[i,j] <- Cumsvd[i,j-1]*sv[i,j-1]

23      }

24

25      # svdmat[i,j] is used in the multinomial for the missing

26      # d component. It is the conditional probability of

27      # dying in sample j given they were alive at last[i]

28      svdmat[i,last[i]] <- Cumsvd[i,last[i]]*(1-sv[i,last[i]])

29      for(j in last[i]+1:k){

30          svdmat[i,j] <- Cumsvd[i,j]*(1-sv[i,j])

31      }

32      svdmat[i,k+1] <- 0

33

34      # Generates the multinomial draw for the missing value of d

35      # conditional on being alive at last[i]

36      d[i,last[i]:k+1] ~ dmulti(svdmat[i,last[i]:k+1],1)

37

38      # alive[i,j] = 1 if individual i is alive in sample j.
```

```
39    # It is a deterministic function of d.

40    alive[i,first[i]] <- 1

41    for(j in first[i]+1:k){

42      alive[i,j] <- alive[i,j-1]*(1-d[i,j-1])

43    }

44

45    # includes the survival probabilities from first[i]

46    # to last[i]

47    wd[i] ~ dbern(Cumsv[i,last[i]]) # ones trick

48

49    sv[i,k] <- 0 # so that sum(d) = 1

50    # S[j] is the time-specific survival probability

51    # p[j] is the time-specific capture probabililty

52    for(j in first[i]:k-1){

53      sv[i,j] <- S[j]

54      pcap[i,j+1] <- p[j+1]

55      }

56    }

57

58    for(j in 1:k-1){

59      S[j] ~ dbeta(1,1) # equivalent to Unif(0,1)

60      p[j+1] ~ dbeta(1,1)

61    }

62  for(j in 1:k){

63    # first capture are binomially distributed

64    u[j] ~ dbin(p[j],capu[j])

65    # the use of round(.) ensures capu is integer
```

```
66    capu[j] <- round(capucont[j])

67    capucont[j] ~ dunif(0,5000)

68  }

69  p[1] ~ dbeta(1,1)

70  }
```

Note that for the program to work, the initial values for capucont[j] must be larger than the values of u[j].

C.4 WinBUGS Code for CJS Model with Continuous Covariates

The WinBUGS code required to fit the model for the continuous covariates is

```
1   # The entire model is placed between model{}

2  model{

3    for(i in 1:udot){

4      for(j in first[i]+1:k){

5        # X[i,j] = 1 if individual i was caught in sample j

6        X[i,j] ~ dbern(peeX[i,j])

7        # individuals can only be caught while alive

8        peeX[i,j] <- pcap[i,j]*alive[i,j]

9      }

10

11     # Cumsv[i,last[i]] is the cumulative survival rate for

12     # individual i between first and last capture.

13     Cumsv[i,first[i]] <- 1

14     for(j in first[i]+1:last[i]){
```

```
15        Cumsv[i,j] <- Cumsv[i,j-1]*sv[i,j-1]

16    }

17

18    # Cumsvd[i,j] is the cumulative survival rate from

19    # last capture until sample j for individual i

20    Cumsvd[i,last[i]] <- 1

21    for(j in last[i]+1:k){

22      Cumsvd[i,j] <- Cumsvd[i,j-1]*sv[i,j-1]

23    }

24

25    # svdmat[i,j] is used in the multinomial for the missing

26    # d component. It is the conditional probability of

27    # dying in sample j given they were alive at last[i]

28    svdmat[i,last[i]] <- Cumsvd[i,last[i]]*(1-sv[i,last[i]])

29    for(j in last[i]+1:k){

30       svdmat[i,j] <- Cumsvd[i,j]*(1-sv[i,j])

31    }

32    svdmat[i,k+1] <- 0

33

34    # Generates the multinomial draw for the missing value of d

35    # conditional on being alive at last[i]

36    d[i,last[i]:k+1] ~ dmulti(svdmat[i,last[i]:k+1],1)

37

38    # alive[i,j] = 1 if individual i is alive in sample j.

39    # It is a deterministic function of d.

40    alive[i,first[i]] <- 1

41    for(j in first[i]+1:k){
```

```
42        alive[i,j] <- alive[i,j-1]*(1-d[i,j-1])

43     }

44

45     # includes the survival probabilities from first[i]

46     # to last[i]

47     wd[i] ~ dbern(Cumsv[i,last[i]]) # ones trick

48

49     sv[i,k] <- 0 # so that sum(d) = 1

50     for(j in first[i]:k-1){

51        # survival and capture probability depend on the

52        # continuous covariate z.

53        logit(sv[i,j]) <- gamma[1] + gamma[2]*z[i,j]

54        logit(pcap[i,j+1]) <- gamma[3] + gamma[4]*z[i,j+1]

55     }

56     for(j in first[i]+1:k){

57        # the model for z must be included

58        z[i,j] ~ dnorm(muz[i,j],tau)

59        muz[i,j] <- z[i,j-1] + delta[j-1]

60     }

61   }

62   for(j in 1:4){

63    gamma[j] ~ dnorm(0,0.0001) # prior distributions

64   }

65   for(j in 1:k-1){

66     delta[j] ~ dnorm(0,0.0001) # prior distributions

67   }

68   tau ~ dgamma(0.001,0.001) # prior distribution
```

69 }

C.5   WinBUGS Code for Multistate Model

The additional code required to include the multistate model is

```
1    # The entire model is placed between model{}

2    model{

3      for(i in 1:udot){

4        for(j in first[i]+1:k){

5          # X[i,j] = 1 if individual i was caught in sample j

6          X[i,j] ~ dbern(peeX[i,j])

7          # individuals can only be caught while alive

8          peeX[i,j] <- pcap[i,j]*alive[i,j]

9        }

10

11       # Cumsv[i,last[i]] is the cumulative survival rate for

12       # individual i between first and last capture.

13       Cumsv[i,first[i]] <- 1

14       for(j in first[i]+1:last[i]){

15         Cumsv[i,j] <- Cumsv[i,j-1]*sv[i,j-1]

16       }

17

18       # Cumsvd[i,j] is the cumulative survival rate from

19       # last capture until sample j for individual i

20       Cumsvd[i,last[i]] <- 1

21       for(j in last[i]+1:k){

22         Cumsvd[i,j] <- Cumsvd[i,j-1]*sv[i,j-1]
```

```
23     }

24

25     # svdmat[i,j] is used in the multinomial for the missing

26     # d component. It is the conditional probability of

27     # dying in sample j given they were alive at last[i]

28     svdmat[i,last[i]] <- Cumsvd[i,last[i]]*(1-sv[i,last[i]])

29     for(j in last[i]+1:k){

30        svdmat[i,j] <- Cumsvd[i,j]*(1-sv[i,j])

31     }

32     svdmat[i,k+1] <- 0

33

34     # Generates the multinomial draw for the missing value of d

35     # conditional on being alive at last[i]

36     d[i,last[i]:k+1] ~ dmulti(svdmat[i,last[i]:k+1],1)

37

38     # alive[i,j] = 1 if individual i is alive in sample j.

39     # It is a deterministic function of d.

40     alive[i,first[i]] <- 1

41     for(j in first[i]+1:k){

42       alive[i,j] <- alive[i,j-1]*(1-d[i,j-1])

43     }

44

45     # includes the survival probabilities from first[i]

46     # to last[i]

47     wd[i] ~ dbern(Cumsv[i,last[i]]) # ones trick

48

49     sv[i,k] <- 0 # so that sum(d) = 1
```

```
50    for(j in first[i]:k-1){

51        # survival and capture probability depend on the

52        # continuous covariate z.

53        sv[i,j] <- s[z[i,j],j]

54        pcap[i,j+1] <- p[z[i,j+1],j+1]

55    }

56    for(j in first[i]+1:k){

57        # the model for z must be included.

58        z[i,j] ~ dcat(psi[z[i,j-1],1:nstate])

59    }

60    }

61    for(j in 1:nstate){

62        psi[j,1:nstate] ~ ddirch(alpha[1:nstate]) # prior distributions

63    }

64    for(j in 1:k-1){

65        for(h in 1:nstate){

66            s[h,j] ~ dbeta(1,1) # prior distributions

67            p[h,j+1] ~ dbeta(1,1) # prior distribution

68        }

69    }

70    for(j in 1:nstate){

71        alpha[j] <- 1 # parameters of the dirichlet in line 62 above

72    }

73 }
```

Note that the number of states, nstate, needs to be specified as data.

C.6   WinBUGS Code for JS Density Dependence Model

```
model{
  for(j in 2:k){
    # ones trick to include capture probability.
    # In sample j, m[j] are caught out of M[j].
    wr[j] ~ dbern(pier[j])
    pier[j] <- pow(p[j],m[j])*pow((1-p[j]),M[j]-m[j])
  }


  M[1] <- 0 # There are no marked animals before the first sample.
  # There are R[1] marked animals immediately after the first sample
  Mplus[1] <- R[1]
  for(j in 1:k-1){
    # The number of deaths is generated as a binomial
    D[j] ~ dbin(mort[j],candie[j])
    # candie[j] is the number of marked individuals that are able
    # to die between sample j and j+1.  We know that T[j] individuals
    # do not die because they were seen in later periods.
    candie[j] <- Mplus[j] - T[j]
    mort[j] <- 1-S[j]
    # The number of marked animals before the j+1th sample are the number
    # alive immediately after the jth sample less the number that die
    # between j and j+1
    M[j+1] <- Mplus[j] - D[j]
    # The number of marked animals immediately after sample j+1 is the
```

```
  # number of marked individuals immediately before sample j+1 (M[j+1])

  # plus  the number of unmarked individuals caught in sample j (R[j+1]-m[j+1])

  Mplus[j+1] <- M[j+1] - m[j+1] + R[j+1]

  # wd uses the ones trick to include [T|S], the individuals we know survive.

  wd[j] ~ dbern(pied[j])

  pied[j] <- pow(S[j],T[j])

}


# JS component
for(j in 1:k){

  u[j] ~ dbin(p[j],capu[j])

  # the number of individual alive in sample j is the number of unmarked

  # individuals alive at j (capu[j]) plus the number of marked

  # individuals alive at j (M[j]).

  capn[j] <- capu[j] + M[j]

}


# round ensure capu is an integer.
capu[1] <- round(capucont[1])

capucont[1] ~ dunif(0,5000) # prior distribution.

for(j in 2:k){

  # the number of unmarked individuals comprises unmarked individuals

  # from the previous sample that have survived (capuprime[j-1]) and new

  # births capb[j-1]

  capu[j] <- capuprime[j-1] + capb[j-1]

}
```

```
for(j in 1:k-1){

  # The number of survivors is binomial

  capuprime[j] ~ dbin(S[j],bigu[j])

  # bigu is the number of unmarked individuals immediately

  # after the jth sample.

  bigu[j] <- capn[j]-Mplus[j]

}


for(j in 1:k-1){

  # the number of new births is poisson

  # bmu[j] is the expected number of births between sample j and j+1

  capbcont[j] ~ dpois(bmu[j])

  # BUGS allows Poisson random variables to be continuous so we must round

  capb[j] <- round(capbcont[j])

  # eta[j] is the per-capita birth rate

  bmu[j] <- eta[j]*capn[j]

  log(eta[j]) <- leta[j]

  leta[j] ~ dnorm(mueta[j],tau[2])

  # the density dependent relationship on per-capita birth rates

  mueta[j] <- beta[3] + beta[4]*(log(capn[j])-5.5)

}


for(j in 1:k-1){

  # the per-capita birth rate on survival

  logit(S[j]) <- beta[1] + beta[2]*(log(capn[j])-5.5) + epss[j]

  epss[j] ~ dnorm(0,tau[1])

}
```

```
for(j in 1:k){

  logit(p[j]) <- lp[j]

  # random effect on capture probability

  lp[j] ~ dnorm(beta[5],tau[3])

}


for(j in 1:3){

  tau[j] ~ dgamma(0.001,0.001) # prior distribution

  sd[j] <- 1/sqrt(tau[j]) # prior distribution

}


for(j in 1:5){

  beta[j] ~ dnorm(0,0.00001) # prior distribution

}

}
```

C.7 Data for JS Density Dependence Model

```
list(

k = 17,

m = c(NA,8,17,...,8),

T = c(10,24,...,8),

wd = c(1,1,...,1),

wr = c(1,1,...,1),

littleu = c(15,44,...,0),

R = c(15,52,...,8))
```