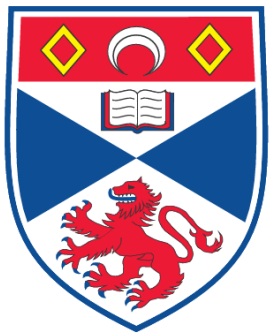
EMBEDDING POPULATION DYNAMICS IN MARK-RECAPTURE MODELS

Jonathan R. B. Bishop

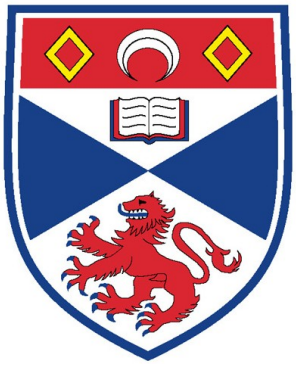
A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews



2009

EMBEDDING POPULATION DYNAMICS IN MARK-RECAPTURE MODELS

Jonathan R. B. Bishop



Thesis submitted for the degree of
DOCTOR OF PHILOSOPHY
in the School of Mathematics and Statistics
UNIVERSITY OF ST ANDREWS
ST ANDREWS
OCTOBER 2008

Abstract

Mark-recapture methods use repeated captures of individually identifiable animals to provide estimates of properties of populations. Different models allow estimates to be obtained for population size and rates of processes governing population dynamics. State-space models consist of two linked processes evolving simultaneously over time. The state process models the evolution of the true, but unknown, states of the population. The observation process relates observations on the population to these true states.

Mark-recapture models specified within a state-space framework allow population dynamics models to be embedded in inference ensuring that estimated changes in the population are consistent with assumptions regarding the biology of the modelled population. This overcomes a limitation of current mark-recapture methods.

Two alternative approaches are considered. The "conditional" approach conditions on known numbers of animals possessing capture history patterns including capture in the current time period. An animal's capture history determines its state; consequently, capture parameters appear in the state process rather than the observation process. There is no observation error in the model. Uncertainty occurs only through the numbers of animals not captured in the current time period.

An "unconditional" approach is considered in which the capture histories are regarded as observations. Consequently, capture histories do not influence an animal's state and capture probability parameters appear in the observation process. Capture histories are considered a random realization of the stochastic observation process. This is more consistent with traditional mark-recapture methods.

Development and implementation of particle filtering techniques for fitting these models under each approach are discussed. Simulation studies show reasonable performance for the unconditional approach and highlight problems with the conditional approach. Strengths and limitations of each approach are outlined, with reference to Soay sheep data analysis, and suggestions are presented for future analyses.

Declarations

I, Jonathan Richard Bryan Bishop, hereby certify that this thesis, which is approximately 80000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September, 2003 and as a candidate for the degree of Doctor of Philosophy in Statistics in September, 2003; the higher study for which this is a record was carried out in the University of St Andrews between 2003 and 2008.

date:_____ signature of candidate:_____

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in Statistics in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date:_____ signature of supervisor:_____

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. We have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to Printed copy and electronic publication of thesis through the University of St Andrews.

date:_____ signature of candidate:_____

signature of supervisor:_____

Acknowledgements

This thesis can be considered as a cumulative effort. Without the help, support and encouragement of key individuals this work would be of a significantly lower standard. I would like to take the opportunity to acknowledge their contributions. Therefore, thanks go to:

- Steve Buckland for his supervision during the course of my thesis. His support, encouragement and patience have been of great value during the progression of my research. An extremely accessible supervisor, the speed and detail of his replies to my queries were able to quickly correct my misconceptions. Equally, I would like to thank Steve for the opportunity to participate in workshops and to gain valuable experience of both tutoring and lecturing. His faith in both my ability and the quality of my work provided great reassurance during periods of doubt.

- Len Thomas for his co-supervision of my thesis. Never one to accept the "obvious" answer, Len constantly made me question my work and provided many insightful comments that significantly improved my research. His encouragement to consider alternative approaches, and his ability to explain them to me, proved invaluable. Thanks also to Len for the opportunity to participate in workshops and thereby gain excellent teaching experience, as well as the chance to meet a variety of interesting people and travel the world.

- Ken Newman for his co-supervision during the first years of my research. His thoroughness and attention to detail are qualities that have hopefully filtered down into my own work. His many useful suggestions led to significant improvements and refinements of the models developed in this thesis.

- Philip Le Feuvre and Rhona Rodger for their constant assistance over the duration of my research. Phil, for his help with regard to computational issues and Rhona

for her help with all administrative issues. Their friendliness and willingness to help greatly eased my progress during my time at St Andrews.

- My office-mates over the course of my studies. Thanks to Ciara Brewer, Joyce Yuan and Danielle Harris for their encouragement, support and provision of welcome, and much needed, occasional distractions from work. Special thanks are reserved for Tiago Marques. Six months ahead of me in his research he provided an excellent example to follow and was a great source of advice in the final stages of my thesis. He also proved to be a memorable traveling companion on several road-trips.

- All those at CREEM yet to be mentioned. The sense of community and level of encouragement given to research students by more senior researches provided a great atmosphere in which to work. Thanks to all for the entertaining coffee-time conversations, general friendliness and, more recently, offers of accommodation.

- A special thank you to Amanda Russell Beattie for providing so much support and helping to improve me as an academic and an individual. Her conduct during her own research provided an excellent example from which I was able to learn a great deal. Her understanding, support and love were invaluable to me during my thesis.

Naturally, despite the valuable contributions of those listed above, this thesis is not perfect and all mistakes contained herein are my sole responsibility.

Data acknowledgements

- Thanks to all those involved in the running and maintenance of the Soay sheep project. Specifically, thanks go to Tim Coulson for his kind permission to use the Soay sheep data analysed in section 7.3. Thanks also to Ruth King for her help in extracting the relevant information from the dataset.

Institutional funding acknowledgements

- Thanks to EPSRC for funding three years of my PhD research.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Growing international concern over the human impact on the environment has predicated an increasing need for the development of techniques to assess the current state of many populations, both plant and animal. Hence, there is a need to monitor biodiversity in order to determine which populations are endangered, and to develop management strategies that can impact positively on these populations. To be effective tools for managing populations, these models need to provide reliable predictions about the future impact of any management strategy. To be meaningful, these predictions need to be made with a quantifiable degree of precision in order to measure the uncertainty associated with any model-predicted outcome.

These management models will be formulated based on data collected on the animal population of interest. These models will then be used to answer important questions about the population of interest. Two of these will be: "How many animals are there?" and "Given how many there are now, how many animals will there be in the future?". These two questions have inspired a significant body of work: the issue of estimating animal abundance is addressed in detail in Seber (1982) and Borchers *et al.* (2002) with a comprehensive review of the different approaches given by Schwarz and Seber (1999).

The second question centres on investigating the relationship between the state of the population at two different points in time. Models are formulated in order to link a time series of population estimates via dynamic processes that are assumed to drive the evolution of the population over time. A unified general framework for embedding population dynamics models in inference is presented in Buckland *et al.* (2007).

To monitor the dynamics of a population over a period of time it is necessary to obtain data on that population over a series of time points. The method chosen to analyse these data is then determined by the requirements of the monitoring study, the nature of the animal being studied and its environment, the resources available to those performing the monitoring and the required precision of the estimates obtained. If individual animals can be identified uniquely, either by their distinctive physical, or chemical (e.g. DNA), markings or by the application of an identifying tag after initial observation, then mark-recapture approaches can be used. Mark-recapture methods are important and widely used methods for analysing time series data based on the repeated observation or capture of individual animals. These methods can be used to obtain estimates for the population abundance as well as the vital rates that specify the additions and removals that determine the changes in abundance over time.

An overview of mark-recapture methodology is given in chapter 2, with emphasis on obtaining inference on the dynamics of animal populations. Classical models, with focus on the estimation of survival and capture rates, are summarised and the main extensions and developments to these are also discussed. Alternative models that explicitly incorporate a recruitment process are discussed in detail. The remainder of the chapter then summarises the developments of mark-recapture models from a Bayesian perspective that allows more complex models to be fitted to mark-recapture data.

The existing mark-recapture analyses are limited with reference to inference on the population dynamics. The construction of the models is often determined by computational expediency rather than any requirement to embed a biologically realistic structure in inference. A review of statistical approaches to modelling the dynamics of wild animal populations is presented in chapter 3. Modelling population dynamics using matrix models (Caswell, 2001) is described and the benefits and limitations of this approach are outlined. Fitting matrix models to data using these methods can often fail to account for the various sources of uncertainty associated with the model specification and the population being studied. Although the statistical models reviewed in chapter 2 offer an integrated modelling approach to account for uncertainty, they typically fail to incorporate an explicit population dynamics model. The integrated approach presented by Buckland *et al.* (2007) is discussed in detail and will form the framework for the new approaches developed in this thesis. Suitable models will need to integrate the time-series of mark-recapture data with the assumed population dynamics model whilst accounting for the various sources of uncertainty. Within a Bayesian framework, state-space models satisfy these requirements and their general structure, fitting methods and inference are discussed in detail in chapter 3.

The chief aim of this thesis is to develop techniques that allow dynamic ecological processes to be embedded into mark-recapture analyses. A new approach is proposed whereby a population dynamics model is embedded into a state-space model structure for analysing mark-recapture data in order to ensure that the resulting model inference is consistent with what is assumed to be known about the biological processes of the system being studied. This approach incorporates the flexible state-space modelling framework that allows complex ecological processes to be specified as components of an embedded

population dynamics model. Using this approach it is possible to mitigate against some of the limitations of the existing mark-recapture methods described in chapter 2.

In chapter 4 one of two alternative approaches is presented. The "conditional" approach restricts the numbers of animals that were not captured at each time period to be consistent with the observed data. That is, under this "conditional" approach the model structure is conditional on the known numbers of animals that were captured at each time period. There is no observation error because the the model parameters relating to capture are included in the state process as opposed to the observation process. Stochasticity enters the model only through those animals that are not captured in the current time period. The general approach to formulating these models and their specific application to mark-recapture data is presented in chapter 4. Simulation methods are discussed and particle filtering techniques for fitting models under the "conditional" approach are described.

A more efficient fitting algorithm for the models constructed under the "conditional" approach is described in chapter 4. This "bottom-up" algorithm initialises model fitting by first focussing on the evolution of the population over the most recent time period. Inferences to earlier time periods then proceeds by simulating backwards in time. The justification for this is presented in more detail in chapter 5. It can be summarised as an attempt to increase the probability that the trial density, required under the particle-filtering fitting approach, respects the conditioning imposed by the observed animals at each time period. The states simulated under the model fitting algorithm need to be consistent with the observed mark-recapture data. The details of this alternative fitting algorithm when applied to a relatively simple state-space model are presented in chapter 5. This approach will not be possible for models for which the state process model depends on state at a previous time period (e.g. density dependence).

An alternative approach to embedding a population dynamics model within a mark-recapture analysis is introduced in chapter 6. The prime distinction from the "conditional" approach is that this "unconditional" approach does not condition on the recorded mark-recapture data. Stochasticity now occurs in the modelling framework via the observation process which models the relationship between simulated states and the observed measures on those states. In common with the "conditional" approach, a state-space modelling framework and particle-filter based fitting methods are both still used for this "unconditional" approach. This approach is compared in detail with the "conditional" approach in chapter 6 and the formulation of the model, fitting methods and inference are all presented in detail.

Applications of these two new approaches: the "conditional" and "unconditional" are discussed in chapter 7. Studies based on simulated data are presented with a detailed analysis provided on the efficacy of both approaches when fitting the same simulated data. Further simulation studies were used to investigate the general performance of each model fitting approach. Multiple populations were simulated and repeated analyses were performed on each of these populations using the model fitting methods specified under each approach. Analyses of variance were performed to investigate the relative contributions of different sources of error to the variability in the average state and parameter estimates produced by each model fitting approach. A real mark-recapture data set from a monitoring study on a population of Soay sheep was analysed under both approaches using existing studies to inform the choice of population dynamics models used in the analysis. The results are presented in chapter 7.

Chapter 8 provides a summary of the work presented and discussed in the thesis.

Limitations to the new approaches developed in chapters 4 to 6 are presented and areas requiring future research are discussed.

Chapter 2

Capture-Recapture Methods

2.1 Introduction

The study of an ecological system from a statistical perspective can often lead to the following questions of interest. What processes affect the system? How can these processes be described in a mathematical context? Therefore, one of the major roles of statistics in ecology is to construct and test models that can be used to simulate accurately the biological processes that are assumed to drive the dynamics of the system being studied.

The ecological systems discussed in this thesis will be limited to those that can be analysed using open population capture-recapture methods. Capture-recapture methods may also be referred to as Mark-Recapture methods during this thesis and these two terms can be assumed to be interchangeable. This chapter will provide an overview of the methods currently used to obtain inference on the dynamics of an animal population using capture-recapture methods.

In general, when investigating the dynamics of an animal population the choice of modelling approach will be determined by the aims of the study. A natural aim of such a study would be to obtain estimates of the total abundance of the population. Another

desirable aim would be to obtain estimates of the demographic parameters assumed to be fundamental to the changes in the size of the population. For example, birth rates and immigration rates determine the additions to the population whereas mortality rates and emigration rates determine the removals from the population over time. Capture-recapture models can be used to obtain estimates for the population size, addition rates and removal rates although not all models can yield all three types of estimate.

One of the main distinctions in the classification of capture-recapture models is between those suitable for closed or open population studies. Although this thesis will not consider closed population models it is useful to elucidate the distinction in a capture-recapture modelling context. Closed populations are ones that do not experience any temporary or permanent changes to the population size; that is, the total abundance remains constant for the entire duration of the sampling period. Open populations are defined as those that can experience additions in the form of births and/or immigration and permanent removals in the form of deaths and/or emigration during the course of the study. Consequently open population studies are defined as having sufficient time between sampling occasions to expect some additions or losses from the sampled population. Therefore, when modelling open populations extra parameters must be incorporated in addition to those that are used for closed populations, to model the changes to the population.

Closed population models, by themselves, are of limited usefulness when investigating the dynamics of animal populations as they require the assumption that births and deaths can effectively be ignored. This restrictive assumption naturally restricts their utility to short-term sampling periods during which constant abundance is a realistic possibility. However, if there is still significant animal movement from migration in and out of the

study area these closed population approaches can no longer be applied reliably. There has been a significant body of work produced on these closed population methods with recent summaries provided by Borchers *et al.* (2002) and Williams *et al.* (2002, pp. 289-314). Both open and closed models can be used with the same study to improve the derived estimates. Pollock's robust design (Pollock *et al.*, 1990; Kendall and Nichols, 1995; Williams *et al.*, 2002, pp.523-544) was conceived to overcome potential problems with bias in estimates of population size which can arise if the underlying assumptions of open population models are not met. Closed population estimates of abundance are robust to heterogeneity in the estimated capture probability and, under Pollock's design, are recommended for estimation of abundance with open population models being used for survival rate estimation. These models will not be discussed in detail in this thesis but are referred to as a future research direction in Chapter 8. From this point on the main focus of the thesis will be on open population models and the increased flexibility this paradigm offers.

We assume that the capture-recapture studies discussed in the following sections consist of $K \geq 1$ sampling occasions. During each sampling occasion some of the animals are "captured" and these can then be classified in two distinct categories: those that have been captured previously and those that have not. The newly captured animals either possess natural, individually distinct, markings or they are given unique distinguishing codes (or marks). These natural markings or issued codes are recorded and, if physically captured, the animals are then released back into the population. Previously caught (and, therefore marked) animals have their identifying codes recorded and are then also released back into the population. The linked, but separate, processes of capture and marking can involve deleterious physical contact with the animal. For example, the capture process may involve setting a grid of traps and waiting for the animals in the study area to enter them. There then exists the potential for some of the trapped animals to be affected

through stress or physical injury. Also, although somewhat less likely, the process of giving each animal a unique identifying code, or mark, can be physically damaging to the animal. If the mark consists of a tag fixed to an animal (e.g. by being punched through its ear) then the animal can be injured. Therefore, in the context of capture-recapture models, animals that are injured during the capture process and not subsequently released, or that are purposefully removed for other reasons, are referred to as "losses on capture".

The use of the phrase "captured" can imply that the animal is physically restrained in the form of a trap. Although this can be the case, the actual mechanism for capture and marking does not always involve physical restraint; there are many cases in which this would be impractical if not impossible. Some animals, or organisms more generally, can be "captured" from distance simply by the observation of some mark or tag. For example, neck collars or leg rings on birds are a common form of tag that can be observed from distance. In these cases, the sampling process consists of attempting to resight marked animals. The initial tagging of the animals may require them to be physically captured but after this initial capture, future captures consist only of the identifying mark being observed. Equally, there are scenarios in which the organism does not even need to be physically captured initially. For some species, individual animals can be identified by unique natural markings. For example, whales (Gowans and Whitehead, 2001), seals (Harrison *et al.*, 2006), and tigers (Karanth and Nichols, 2000) can be identified from photographs of their markings.

The $K \geq 1$ sampling occasions will then result in a record of captures and/or observations for each animal that is uniquely identified at some point during the study. The record for an individual animal is referred to as its *capture history* and consists of a row

vector of length K comprised of 1s and 0s indicating whether an animal was caught (denoted by a 1) or not caught (denoted by a 0) on each occasion.

Modelling based on this capture history form of data structure focuses on the development of probabilistic models that explain adequately the biological processes which can engender the observed capture history data. The development of appropriate probabilistic models requires a suitable choice of parameterisation. The simple closed population models for capture-recapture require only the specification of capture parameters due to the restrictive assumption that the population remains at some fixed, but unknown, total for the duration of the sampling study. The more flexible nature of open population models, in which the composition of the population is allowed to vary over the course of the study, necessitates a larger number of processes to be included in the probabilistic models, and consequently, a larger number of associated parameters. The most flexible and general open population capture-recapture models can incorporate both additions and losses of animals throughout the course of the sample study.

After the completion of a general long-term open population capture-recapture study, the main parameters and quantities of interest will consist of two types of estimates. Firstly, estimates of the abundance at various points throughout the study and secondly, estimates of the demographic parameters relating to additions (often referred to as recruitments in the literature) and removals between each sampling occasion. Depending on the assumptions made about the model structure, additions may incorporate both birth and immigration and removals may incorporate both mortality and permanent emigration. The ability to estimate these quantities of interest is then determined by the information collected in the capture-recapture study. In this general open population capture-recapture study there are two main sources of information; firstly, that obtained

from recapturing animals that have been previously marked and secondly, that obtained from comparing the numbers of previously marked animals with the numbers of unmarked animals captured during each sampling occasion. The first source of information is adequate if the aim of the study is solely to estimate survival rates of marked individuals, whereas both sources of information are required to estimate population sizes and numbers of additions at each sampling occasion.

The estimation of survival rate can be the prime aim of a study (Cormack, 1964; Buckland, 1982; Lebreton *et al.*, 1992). In these approaches, the fitted models are conditional on the initial capture of an animal with the emphasis then being on the processes that determine the subsequent entries in the animal's recorded capture history. As noted in Cormack (1964), there is no provision in these models for estimating the population abundance and, as described in Lebreton *et al.* (1992), the recapture of previously marked animals is simply modelled as a function of survival and capture probabilities and does not depend on population abundance. A particular form of this type of study is one in which inference is based solely on recoveries (recapture and subsequent removal from the population) of previously marked animals. Band recovery models are examples of this type and in this scenario the first recapture consists of the recovery of the identifying band or tag from the body of the animal. Clearly, there can be at most a single recapture, and inference is based on the losses of these individuals from the population over the period of the study. A general analysis using numerical solutions to maximum likelihood formulations for these band recovery models is given by White (1983), and a comprehensive description of the theory and application of band recovery models is given by Brownie *et al.* (1985). The mark-recapture survival analysis in Buckland (1982) incorporates both tag return and capture-recapture data to provide estimates of survival parameters that are less influenced by departures from the restrictive assumptions of homogeneous capture

probabilities across all animals.

This chapter presents a review of the development of capture-recapture methodology. The classical modelling approaches are introduced and explained with major extensions and developments to these approaches discussed in detail. Extensions that have particular relevance to obtaining inference on population dynamics within a mark-recapture analysis are examined in detail. More recent development have seen capture-recapture analyses embedded within a Bayesian framework. This section of this review chapter introduces some of the important recent developments in this area of research by focusses on each in turn.

The models discussed in Section 2.2 will be restricted to those that are conditional on the initial capture of an animal. The standard assumptions and resulting structure of simple capture-recapture models of this type will be introduced in the next section. The following sections broadly follow the structure of Chapters 17 and 18 of Williams *et al.* (2002) and provide an outline of the development of the use and analysis of open population capture-recapture studies.

2.2 Capture-Recapture Models Single Age

The simplest form of open population capture-recapture model consists of a single cohort of animals being monitored over the duration of the study. This is often referred to as a 'single-age' model and is defined such that the survival and capture parameters are assumed to be homogeneous for all sampled animals regardless of when they were captured or their age at the time of capture.

The capture history data can be expressed in the form of a capture history matrix,

X	=	The matrix of capture histories.
K	=	Number of sampling occasions.
M_{K+1}	=	Total number of animals that have been captured at least once during the study.
ω	=	A zero-one indicator vector denoting the capture history: $\omega = \left(\delta_1^{(\omega)}, \delta_2^{(\omega)}, \ldots, \delta_K^{(\omega)}\right)$ where $\delta_i^{(\omega)} = 1$ or 0 for$j = 1, \ldots, K$ if the animal was caught on the j^{th} occasion or not. There are 2^K possible patterns, one of which represents the null pattern: $\omega_1 = (0, 0, \ldots, 0)$ i.e. those animals that were present at some stage but never caught.
x_ω	=	Number of animals having the capture history pattern ω

Table 2.1: Notations and definitions for the x_ω representation of single-age capture-recapture models.

X. The total number of individuals caught during the study is denoted as M_{K+1}, thus the matrix X will have M_{K+1} rows. If there are a total of K sampling occasions during the study then X will have K columns. Therefore, if i represents an individual animal ($i = 1, 2, \ldots, M_{K+1}$) and j represents the sampling occasion ($j = 1, 2, \ldots, K$) then the element X_{ij} takes the value 1 if the i^{th} animal was caught during the j^{th} sampling occasion, and 0 if the i^{th} animal was not caught during the j^{th} sampling occasion. This notation is defined in Table 2.1.

Using the definitions in Table 2.1, inference from capture-recapture studies is based on the the number x_ω of animals having the capture history pattern ω. Although there are 2^K possible capture history patterns, the null pattern corresponding to an animal that exists in the population yet is never captured is not included in the analysis. This is because these models are constructed to be conditional on the first capture of the animals and thus require an animal to be captured on at least one occasion. The inference is then based only on the observable capture histories, that is, those histories that contain at least one capture. For example, in a three-period study there are eight possible capture histories, of which seven are observable and the other is the null pattern. The

m_{ij}	=	Number of animals released in period i that are next captured in period j $(j > i)$.
R_i	=	Number of releases in period i.

Table 2.2: Notations and definitions for the m_{ij}-array representation of single-age capture-recapture models.

numbers of animals attaining these observable capture histories can then be expressed as $x_{111}, x_{110}, x_{101}, x_{100}, x_{011}, x_{010}, x_{001}$ and the required likelihood can be constructed from these data.

An alternative, but equally common, representation for the data that can be fit to these single-age models is an m_{ij}-array. As defined in Table 2.2, the entries in this array correspond to the numbers of animals first captured and released during sampling period i that are then not captured again until sampling period j $(1 \leq i < j \leq K)$[1]. For example, m_{13} denotes the number of animals that were first captured during the first sampling occasion and subsequently marked and released but then were not captured during the second sampling occasion before being captured again during the third sampling occasion. The total, R_i, corresponds to the number of animals released back into the population after the i^{th} sampling occasion. If it is assumed that there are no losses on capture the number of animals released after the i^{th} sampling occasion is given by

$$R_i = \sum_{r=1}^{i-1} m_{ri} \quad 1 \leq i < K.$$

It should also be noted that the individual capture history patterns, ω, can contribute to multiple values in the m_{ij} array. For a three-period study an animal with the capture history pattern x_{111} will be released after the first, second and, assuming it is not lost on

[1]Note that the indexes i and j differ between the x_ω and m_{ij} summary representations.

p_i	=	The probability that a marked animal in the study population during sampling period i is captured during sampling period i
ϕ_i	=	The probability that a marked animal in the study population during sampling period i survives until $i+1$ and remains in the population
χ_i	=	The probability that an animal alive in the study population during sampling period i is not captured again during any future sampling period.

Table 2.3: Further notations and definitions for the single-age capture-recapture models.

capture, the third capture occasions and will therefore contribute to R_1, R_2 and potentially R_3. Equally, the animal will contribute to both m_{12} since it was released in period 1 before being captured and released in period 2, and m_{23} since it was then captured again in period 3.

2.2.1 The Cormack-Jolly-Seber Model

The above m_{ij} array representation is a common way to represent data for the single age Cormack-Jolly-Seber model (Cormack, 1964; Jolly, 1965; Seber, 1965). To express the conditional Cormack-Jolly-Seber[2] models for the capture-recapture data three further parameters are defined in Table 2.3.

It is not necessary to always use the parameter χ_i as it is constructed from previously specified parameters: the capture and survival probabilities p_i and ϕ_i respectively. However, the model expression is somewhat easier to interpret when the χ_i parameterisation is used. For a K sample study, no animals can be seen again after the K^{th} sampling occasion and therefore $\chi_K = 1$. For sampling occasions $i < K$, χ_i can be calculated recursively as:

$$\chi_i = (1 - \phi_i) + \phi_i(1 - p_{i+1})\chi_{i+1} \tag{2.2.1}$$

[2]hereafter referred to as CJS

The composition of Equation (2.2.1) manifests the two different ways in which an animal can avoid being recaptured after sampling occasion i: it can 'die' and therefore permanently leave the population (which has probability $1 - \phi_i$) or it can survive and remain in the population without ever being recaptured (which has probability $\phi_i(1-p_{i+1})\chi_{i+1}$). With regard to the 'survival' parameter ϕ_i it should be noted that the parameter refers to the probability that the i^{th} individual remains in the population available to be sampled. This happens as a combination of two separate processes: firstly the individual animal must survive and secondly the individual animal must not permanently leave the population by emigrating. Hence, the process referred to as 'survival' is a combination of both true survival and non-emigration.

Having defined the required survival and capture probabilities the modelling approach for the capture history data can now be specified. These models are conditional on first capture so the initial state of the model consists of animals that have been caught, marked and released. Then, a released animal can either remain alive until period 2 with probability ϕ_1 or it can die. If it survives, it can be caught with probability p_2 or it can remain uncaught. The development of this system over the first two sampling periods is shown in Figure 2.1.

This modelling structure is then repeated through all periods in the study and can be used to model an observed capture history in terms of the probabilities associated with the survival and capture processes. Consider a five period study and consider an individual animal that was captured during the 2^{nd} and 4^{th} sampling occasions but was not captured during the 3^{rd} occasion. The capture history pattern for this animal can then be written as 01010. This history is modeled by conditioning on the first observed capture which occurred in the 2^{nd} period and then expressing the remainder of the history

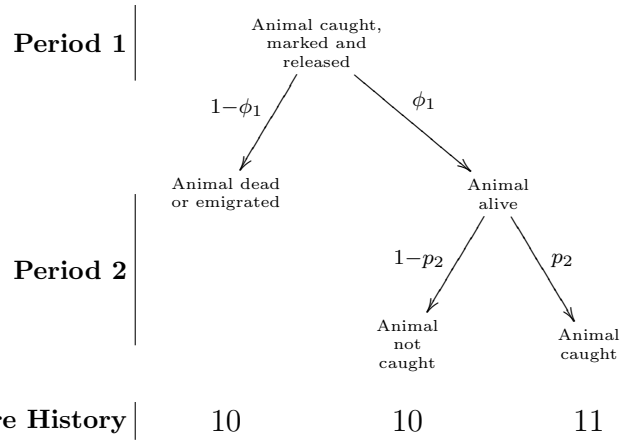


Figure 2.1: Diagram of events and associated probabilities for an animal released in period 1 of a two-period study under the CJS model for open populations

as:

$$Pr(01010|\text{First capture during period 2}) = \phi_2(1-p_3)\phi_3 p_4 \chi_4.$$

For the observed capture history 01010 the model must include survival from periods 2 to 3 (which has probability (ϕ_2), non-capture in the 3^{rd} period (which has probability $1-p_3$), survival from periods 3 to 4(which has probability ϕ_3), capture in the 4^{th} period (which has probability p_4) and the probability of not being captured again after period 4 which is given by $\chi_4 = (1-\phi_4) + \phi_4(1-p_5) = 1-\phi_4 p_5$.

The above expression for the capture history is actually conditional on both the first capture and the animal's fate after the last observed capture. For example, if the animal had been lost on capture or else removed from the population and not released after being captured in the 4^{th} period then the model for the 01010 history would be:

$$Pr(01010|\text{First capture during period 2 and removal at period 4}) = \phi_2(1-p_3)\phi_3 p_4.$$

The difference is that in the latter case the animal is removed from the population after

Capture History	Probability
100	$(1-\phi_1)+\phi_1(1-p_2)(1-\phi_2 p_3)$
101	$\phi_1(1-p_2)\phi_2 p_3$
110	$\phi_1 p_2(1-\phi_2 p_3)$
111	$\phi_1 p_2 \phi_2 p_3$

Table 2.4: Capture histories and associated probabilities, conditional on capture in the first sample, for R_1 animals under a three-period CJS model.

period 4 so there is no need to model any events following the last observed capture and the χ_4 term is no longer included.

Consider the R_1 animals; these animals have been captured and released into the population after the first capture occasion. If a three-period sampling study is assumed then the animal can experience one of the following capture histories: 100,101,110 or 111. Constructing the capture history pattern probabilities in the same way as for the five-period example above, Table 2.4 shows the probabilities associated with the four possible capture history patterns for the R_1 animals.

From the construction of the capture history probabilities in this manner it can be seen that $\chi_3 = 1$, $\chi_2 = (1-\phi_2)+\phi_2(1-p_3) = 1-\phi_2 p_3$ and $\chi_1 = (1-\phi_1)+\phi_1(1-p_2)(1-\phi_2 p_3) = (1-\phi_1)+\phi_1(1-p_2)\chi_2$. Since each of the R_1 animals must take one of these four mutually exclusive capture histories the probability distribution of these four histories can be constructed as a conditional multinomial distribution:

$$\begin{aligned} Pr(\{x_\omega\}\,|R_1) &= \frac{R_1!}{\prod_\omega x_\omega!}[(1-\phi_1)+\phi_1(1-p_2)(1-\phi_2 p_3)]^{x_{100}}[\phi_1(1-p_2)\phi_2 p_3]^{x_{101}} \\ &\quad \times[\phi_1 p_2(1-\phi_2 p_3)]^{x_{110}}[\phi_1 p_2 \phi_2 p_3]^{x_{111}} \end{aligned} \tag{2.2.2}$$

This can be expressed more concisely in terms of the χ_i $(i = 1, 2, 3)$ values:

$$\begin{aligned} Pr(\{x_\omega\}\,|R_1) &= \frac{R_1!}{\prod_\omega x_\omega!}[\chi_1]^{x_{100}}[\phi_1(1-p_2)\phi_2 p_3]^{x_{101}} \\ &\quad \times[\phi_1 p_2 \chi_2]^{x_{110}}[\phi_1 p_2 \phi_2 p_3]^{x_{111}} \end{aligned}$$

where $\chi_3 = 1$ and is thus omitted. The index ω in the denominator of Equation (2.2.2) denotes the range of the four possible capture history patterns: $100, 101, 110$ and 111.

The model for the entire data set of capture histories needs to include the probabilities for the animals that are released after the second sampling period having been uncaught during the first period. These animals are denoted as u_2 and will exhibit capture history patterns with 01 as their first two entries. For the three-sample study being considered there are only two such histories: 011 and 010, the animals are either caught again during the final capture period or they are not. The conditional multinomial[3] distribution for u_2 is then:

$$Pr(\{x_{010}, x_{011}\}\,|u_2)\frac{u_2!}{x_{010}!x_{011}!}[1-\phi_2 p_3]^{x_{010}}[\phi_2 p_3]^{x_{011}} \tag{2.2.3}$$

The probabilistic model for the three period study does not incorporate terms corresponding to animals that are first released after the third (final) sampling occasion. The aim of the model is to construct a probabilistic description of the fate of an animal (i.e. its capture history) conditional on the time of initial release. If the initial release is only after the final capture occasion then there is no remaining information in the capture history to model the animal's fate. Hence, the 001 history is not included in the probability distribution. Equally, the null history 000 is not included as this capture history pattern corresponds to animals that are never captured or observed and therefore can not be a component of a model that conditions on first capture.

[3]Actually a binomial distribution in this case

Eight possible capture history patterns can occur in the three period study under consideration, but only six are included in the conditional model. The conditional probability model for these six possible capture histories which are observed for animals released after initial capture in either of the first two periods is then given by the product of Eqs (2.2.2) and (2.2.3).

This form of model expression denoted in Eqs (2.2.2) and (2.2.3) specifies a fully parameterised CJS model. That is, the index i on both the capture parameters p_i and survival parameters ϕ_i indicates that both processes can have time-specific values. Reduced forms of CJS models exist and these can allow for fixed values of p_i, ϕ_i or both. These reduced model forms as well as other special cases and extensions to the CJS model will be discussed later in this chapter.

2.2.1.1 Model Assumptions

The probabilistic form of, and the inference obtained from, a model is determined by the inherent assumptions made about the system being studied. For the Cormack-Jolly-Seber model the assumptions (Pollock *et al.*, 1990; Williams *et al.*, 2002), with a more fully specified set in (Burnham *et al.*, 1987), are often given as:

1. Every marked animal present in the population at the time of sampling period i $(i = 1, 2, \ldots, K)$ has the same probability of being captured or resighted, p_i.

2. Every marked animal present in the population immediately after the i^{th} sample has the same probability, ϕ_i, of survival until the $(i+1)^{th}$ sampling period (i=1,2,...,K-1).

3. Marks are recorded accurately and are neither lost, overlooked or misread.

4. All sampling periods are instantaneous and recaptured animals are released immediately after the sample.

5. All emigration from the sampled area is permanent.

6. The fate of each animal with respect to capture and survival probability is independent of the fate of any other animal.

Assumption (4) is that all sampling periods are instantaneous. In practice this is unlikely to be the case and instead the assumption is that the duration of the sampling period is very small relative to the time between sampling periods, that is the interval over which the rate of survival will be estimated. If the sampling period is long then this could increase the heterogeneity of survival amongst released animals. For example, if mortality can occur during the sampling period then an animal released at the beginning of a long sampling period may have a lower probability of surviving to some future point in time than an animal released at the end of the sampling period.

Assumptions (1) and (2) relate to the implicit assumption of homogeneity for the capture and survival probabilities used to construct the probabilistic model for the observed capture history patterns. This assumption is not always a reasonable one and the true survival and capture probabilities can often be better described as functions of demographic or environmental factors attributed to the individual captured animal rather than assuming common rates across all captured animals in a given time period. Williams *et al.* (2002) categorise these characteristics as belonging to one of four possible classes determined by both the measurement of the characteristic (discrete or continuous) and the range of values taken by the attribute for an individual animal (static or dynamic). A brief summary of each is given in the following paragraphs.

A static attribute is defined as one that does not change for an individual animal over the course of the study. For example, gender is typically static for vertebrates and the capture and survival rates may differ between males and females of a particular species. Gender is also a discrete attribute and therefore the capture history data can be modelled in two ways. Firstly, the data could be separated by gender and two independent analyses could be carried out, one for the males and one for the females. However a more parsimonious approach would be to analyse all of the data in one model in which some parameters are specific to gender but others are assumed to be common across genders.

Another static attribute that may affect survival probability throughout an animal's existence could be the animal's weight at birth. Birth weight can be treated as either discrete or continuous. If the animals are grouped into a different categories of weight then separate studies could be carried out for each weight class. This can lead to some loss of information from the data and also may result in models that have an excessive number of parameters. An alternative to this approach would be to view the birth weight as a continuous variable and to model survival as a function of it. This scenario is therefore an example of a static, continuous variable.

Dynamic attributes are ones that change for a specific animal over time. These dynamic characteristics can affect the survival and capture parameters for an individual animal and can be classified as either discrete or continuous. For discrete, dynamic variables the probability model that determines how the variable changes over time can determine the manner in which they are incorporated into the capture history probability model. Some discrete dynamic variables evolve over time in an entirely predictable way. For example, the age of animal in a dynamic variable that is typically measured discretely and changes in a deterministic fashion; the age of animal in period $i+1$ will be predicted

with certainty if its age in period i is known. Capture-recapture models for multiple age classes that are comprised of age-specific parameters are discussed from a band-recovery perspective in Brownie *et al.* (1985) and from a more general CJS perspective in Pollock *et al.* (1990) and Lebreton *et al.* (1992). Alternatively discrete, dynamic variables can evolve in a stochastic manner in which case a model will need to be developed to predict the probabilities of the transitions between the different values taken by the dynamic variable. For example, if it can be assumed that capture and survival rates vary by location (as a proxy for habitat type) then the probability of capture during period i and the probability of survival from period i to $i+1$ are dependent on the location of the animal at time i. If the location of an animal can be categorized into discrete location categories (rather than recording the exact co-ordinates), then an animal in location A at time i may remain in the same location at time $i+1$ or may have moved to location B. These transitions are modelled probabilistically such that an animal in location A at time i is in location A at time $i+1$ with probability ψ_i^{AA} and is in location B at time $i+1$ with probability ψ_i^{AB}. Models in which animals can move between discrete states during the course of the study are referred to in the literature as multistate models (Brownie *et al.*, 1993; Nichols *et al.*, 1994; Nichols and Kendall, 1995; Schwarz and Arnason, 1996).

The final class of characteristics are those which are both continuous and dynamic, these are discussed in detail in Section 2.2.5.

The fifth assumption, that all emigration from the sampled area is permanent, is necessary to interpret the model-based inference associated with the estimate of capture $\hat{p}$. If a capture history contains the sub-pattern 101 then the 0 in the middle is assumed to represent the event of non-capture with probability $1-p$. It is assumed that the animal was present in the population sample area but remained uncaught during that period.

If the assumption of permanent emigration was relaxed and temporary emigration was allowed then this could result in an alternative interpretation of the middle 0: that the animal was unavailable for capture at that time. In this case the interpretation of $\hat{p}$ must be modified and estimates of capture and survival rates may well be biased (Kendall, 1999), although the bias for survival may be quite low.

The final assumption is that the fates of any two individual animals are independent. This can be violated, for example, animals that habitually aggregate in family groups or herds may not exhibit independent fates. Violation of this assumption does not typically impact on the point estimates of the capture and survival rates but it can induce bias with regard to the variance assumptions made implicit by the use of a multinomial model. To accommodate the extra variation induced by the potential non-independence of animal fates, quasilikelihood methods can be considered (Burnham *et al.*, 1987; Lebreton *et al.*, 1992).

2.2.1.2 Model Estimation

The probability model defined as the product of Equations (2.2.2) and (2.2.3) and the associated data consisting of the observed capture history patterns ω and the number of animals possessing that history, x_ω, is used to estimate the model parameters. The basis of this inference is the construction of the appropriate maximum likelihood expression for the model (Burnham *et al.*, 1987). Using the same notation as Williams *et al.* (2002) the general expression for conditional models of a single-age CJS study can be expressed as follows. Define $\pi_\omega = f(\{\phi_i\}, \{p_i\}) \quad (i = 1, 2, \ldots, K)$ as the probabilities associated with each capture history; the use of $f(\cdot)$ simply denotes that these capture history probabilities are functions of the survival and capture probabilities across all K periods. Then, the

conditional probability distribution of the observed capture histories $\{x_\omega\}$ is:

$$Pr(\{x_\omega\} \mid \{u_i\}; \{\phi_i\}, \{p_i\}) = \frac{\prod_{i=1}^{K-1} u_i!}{\prod_\omega x_\omega!} \prod_\omega \pi_\omega^{x_\omega} \tag{2.2.4}$$

As before, the index ω ranges over all possible observed capture histories, but with the null pattern corresponding to an animal never being captured $(00\ldots00)$ and the pattern corresponding to a first release after the final capture occasion $(00\ldots01)$ not included in the general expression. From a maximum likelihood perspective the terms on the right-hand side of Equation (2.2.4) give the likelihood function for the survival and capture parameters, $L(\{\phi_i\}, \{p_i\})$. Hence this expression can be maximised to obtain maximum likelihood estimates of the survival and capture probabilities.

The parameterisation of the model expressed in Equation (2.2.4) allows for both capture and survival probabilities to be fully time-specific. This full model was examined initially by Cormack (1964) who focussed on a survival model, and then by Jolly (1965) and Seber (1965) who included the estimation of abundance and births. Together these formalised methods give rise to the Cormack-Jolly-Seber name for models of this type. Their examination of these models led to the derivation of closed-form estimators for the identifiable survival and capture parameters. They show that a K-period sampling study with time specific capture and survival probabilities will contain $2K-3$ identifiable and estimable parameters. Closed-form estimators can only be obtained separately for the survival parameters $\phi_1, \phi_2, \ldots, \phi_{K-2}$ and the capture parameters $(p_2, p_3, \ldots, p_{K-1})$. The parameters p_K and ϕ_{K-1} can only be estimated as the product $\phi_{K-1}p_K$ and cannot be estimated separately.

The lack of identifiability of these parameters can be understood using the intuitive explanation in Williams *et al.* (2002, pp. 424). To estimate the capture probability during

sampling period i, p_i, two quantities are required: firstly, the number of animals that are known to be alive during the i^{th} sampling period and secondly, the proportion of these animals that are then captured during sampling period i. The first quantity is obtained from the numbers of animals caught during both of the sampling periods preceding and following the i^{th} one. Conditional on this total, the second quantity is then just the number of these animals that were caught in the i^{th} sampling period. Thus, the number of animals known to be alive during the i^{th} sampling period is given, in the ω notation, as

$$\sum_{\omega_{1..0..1}} x_{..1..0..1..} + \sum_{\omega_{..1..1..1..}} x_{..1..1..1..}$$

where the central element in each pattern correspond to the capture status during sampling period i. The ..1.. notation simply indicates a sub-pattern consisting of any combination of 0s and 1s for a set of sampling occasions that includes at least one 1. The index $\omega_{1..0..1}$ ranges over all possible capture history patterns ω that include capture both before and after the i^{th} sample; and similarly for the ω_{111} index. Then an intuitive estimate for the capture parameter in the i^{th} sample would be:

$$p_i = \frac{\sum_{\omega_{1..1..1}} x_{..1..1..1..}}{\sum_{\omega_{1..0..1}} x_{..1..0..1..} + \sum_{\omega_{1..1..1}} x_{..1..1..1..}}$$

However, from the data there exists no known total of animals of animals that existed in the population prior to the initial sample and therefore there is no subset of animals to condition on. Thus, the first quantity is unobtainable and, consequently, no estimate can be obtained for the initial capture probability p_1. Similarly, for the final sampling occasion, K, there are no future sampling occasions so there are no means to establish the total number of animals known to be alive at K. Therefore, the first quantity is again unobtainable and the final capture probability p_K cannot be estimated separately.

2.2.1.3 Reduced-Parameter Models

The full CJS model in which both survival and capture probabilities are time-specific provides an extremely flexible modelling formulation. However, due to the requirement to estimate each parameter separately this generality incurs a penalty by yielding estimates that may lack precision. The estimates of time-specific model parameters will typically have larger associated variances than the equivalent estimates of parameters whose values are fixed over time. The gain in precision available when capture and/or survival parameters are constrained to be constant over time has led to the development of reduced-parameter CJS models (Jolly, 1982). The number of parameters that can be identified and estimated separately will change depending on which parameters are constrained to be constant over time.

As a notational convention, the full time-specific CJS model can be denoted ϕ_t, p_t where the t subscript indicates that parameter can take a different value for each separate period. Therefore, ϕ_t, p represents a model in which survival probabilities are still time specific but capture probabilities are assumed to be constant over time. That is $p_i = p$ for $(i = 2, 3, \ldots, K)$ which yields a model consisting of $K - 1$ survival probabilities and a single capture probability. By imposing the constraint that $p_K = p$ the elements of the product $\phi_{K-1}p_K$ can now be separated to obtain an estimate for ϕ_{K-1}. Thus, all $K - 1$ survival probabilities can be estimated along with the single capture probability p. Similarly, for the model ϕ, p_t the capture probabilities are time dependent and take separate values for each of the time periods 2 to K but the survival probabilities are now constant yielding a single survival probability which is the same for each of the time periods 1 to $\ldots, K - 1$. By imposing the restriction that all survival rates are equal such that $\phi_{K-1} = \phi$ an estimate for p_K can be obtained. All $K - 1$ capture probabilities can be estimated along with the single survival probability ϕ.

The element of time is important in determining the interpretation of constant probabilities. Capture is assumed to occur instantaneously and involve a constant amount of effort but for models in which this is not the case it can be reasonable to ask what the assumption $p_i = p$ actually means. If the duration of the capture occasion varied greatly over different sampling occasions then the assumption of a constant probability of capture may not be realistic. If in one month the capture occasion consisted of a single trapping day but the following month it consisted of an entire week of trapping then the assumption that the capture probabilities are the same on both occasions is probably quite unreasonable. A reasonable question to then ask is whether the per-unit time rate of capture is similar across the two sampling periods. One approach to answering this question is to model capture as $p_i = p^{t_i}$ where t_i denotes the duration, in units of time, of the capture occasion and p is the per-unit-time capture rate measured in the same units as t_i. A similar, and more common occurrence, is when the length of time between sampling occasions varies significantly. In this case the assumption of a constant probability of survival between consecutive sampling occasions is not realistic and the solution is, as before, to model survival in terms of unit time such that $\phi_i = \phi^{t_i}$ where ϕ is the per-unit-time survival rate (Pollock *et al.*, 1990).

Finally, the simplest CJS capture-recapture model is denoted ϕ, p and represents a study in which both survival and capture parameters are assumed to be constant over time and the sampling periods. There are only two parameters to be estimated for this version of the CJS model. In general, the reduced-parameter models can be expressed in exactly the same form as in Equation (2.2.4) with the constant parameters being used in place of the time-specific ones in the full expression for each π_ω term. There exist no general closed-form expressions for these reduced-parameter models and computational

numerical approximation methods are required to obtain maximum likelihood estimates of the parameters in these models. The assumptions of these models can be tested using Likelihood-Ratio-Tests (LRT) (Lebreton *et al.*, 1992; Williams *et al.*, 2002, pp. 53-54) which typically compare nested models against a more general alternative to determine whether the fit of the model can be improved by removing the constraints on parameters. For example a LRT can be constructed to investigate if survival is time-specific by comparing the constrained model ϕ, p_t against the general model ϕ_t, p_t.

The modelling approaches discussed in Section 2.2.1.1 allowed capture and survival probabilities to be expressed as functions of variables that could be either discrete or continuous and either static or dynamic. These approaches can also be incorporated into the reduced-parameter models just discussed. The developments of Lebreton *et al.* (1992) provide a detailed framework in which these models can be specified. For example, if a set of time-specific covariates $\mathbf{X} = \{x_{1i}, x_{2i}, \dots\}$ for each period i is specified then the GLM-like approach of Lebreton *et al.* (1992) involves expressing the parameter of interest as a linear function of the explanatory covariates. For example if the belief was that capture was a function of multiple covariates, $\mathbf{X}$, and associated parameters $\boldsymbol{\beta} = \{\beta_0, \beta_1, \dots\}$, the relationship can be expressed as

$$p_i = f\left(\mathbf{X}\boldsymbol{\beta}\right) = f\left(\beta_0 + \sum_j \beta_j x_{ij}\right)$$

where the β_0 term corresponds to an intercept term in the model and a column of 1s in the design matrix $\mathbf{X}$. Then the functional form of the link function f needs to be assumed to obtain the inverse function f^{-1} which will allow the model for capture during i to be expressed in a form that is linear in the parameters to be estimated (the $\boldsymbol{\beta}$). This linearisation is crucial in order for the estimates of $\boldsymbol{\beta}$ to be obtained using GLM theory. A commonly chosen function is the logit or logistic function as it is invertible and provides a

flexible bounded form that constrains the estimates of the response variable (the capture probability p_i in this case) to lie between 0 and 1. For example, using the logit link the capture probability for sampling occasion i can be written as

$$\hat{p}_i = \frac{\exp\left(\hat{\beta}_0 + \sum_j \hat{\beta}_j x_{ij}\right)}{1 + \exp\left(\hat{\beta}_0 + \sum_j \hat{\beta}_j x_{ij}\right)}. \tag{2.2.5}$$

This approach is equally viable for the full time-specific CJS model ϕ_t, p_t and the reduced-parameter model in which survival is constant across all pairs of consecutive periods ϕ, p_t. The approach of Lebreton *et al.* (1992) goes on to formulate suitable CJS model parameterisations for multiple groups if it is believed that capture and survival parameters may be common within distinct cohorts (e.g. gender or spatial location). Their "parallelism" approach provides a useful constraint on the parameters by specifying an assumed relationship between the way in which the parameters vary over time across cohorts. For example, male and female capture may not take the same values at the same sampling occasions but they may exhibit the same pattern of temporal variation. The general CJS model formulation in Eq. (2.2.4) can also be parameterised to accommodate a trap response in both capture and survival probabilities (Lebreton *et al.*, 1992). The use of physical tagging or baited traps may, respectively, decrease or increase the probability that a previously trapped animal is trapped again, given that it is available to be captured. Equally, trapping or tagging an animal may be a traumatic event that decreases the animal's subsequent survival rate. For both of these scenarios the trap-response is assumed to be temporary with the dependence only occurring for the period immediately following a capture occasion. Thus, if there is trap-response for survival then captured animals will, post-capture, survive from period i to period $i + 1$ with one probability, uncaptured animals will survive the same period with a different probability but they will share the same survival probabilities after period $i + 1$.

Individual covariate models are an even more flexible approach to trying to account for the heterogeneity in the data. The equality of survival and capture probabilities across all individuals in a population is unlikely to be true. Grouping animals into cohorts that are assumed to possess common probabilities is still unlikely to remove all the heterogeneity in the data. As discussed in section 2.2.1.1, it may be possible to explain individual level differences in capture and survival rates using a series of measurable GLM-like covariates. These individual covariate models are again constructed using a GLM-framework with the emphasis usually being on static, continuous variables. Dynamic variables can be more difficult to model in a non-Bayesian context although, as discussed in Section 2.2.3, a multistate modelling can be used to model capture and survival rates as functions of discrete and dynamic covariates.

2.2.1.4 Model Selection

The selection of models in classical (non-Bayesian) capture-recapture analyses requires that an initial model be found that provides an acceptable fit to the data. Conditional on this initial acceptable model, alternative models can be selected on the basis of LRTs. Under this paradigm, model selection is framed as a hypothesis test in which a more general model is compared to a nested[4] model. These model comparisons proceed sequentially with the more general models being compared to reduced-parameter models using LRTs (Brownie *et al.*, 1985; Burnham *et al.*, 1987; Lebreton *et al.*, 1992). Equally, the simplest model can be chosen as the starting point and can then be compared with increasingly complex models with LRTs used to determine if the increase in complexity is warranted. Score tests (Catchpole and Morgan, 1996) can also be used in this manner, starting from

[4]the more general model with constraints on its parameters

the simplest model, with the aim being to reduce the chance of fitting overly complicated models to data.

An optimisation framework can also be used to determine the choice of model. The use of information-theoretic approaches to model selection requires the specification of an optimisation criterion and Akaike's Information Criterion (Akaike, 1974) is the most commonly used one. The information-theoretic criterion combines a goodness-of-fit test by favouring the model that produces the largest log-likelihood value (evaluated at the maximum likelihood estimates) with a penalty term that favours a parsimonious model parameterisation. The model producing the smallest AIC value will be selected, thus, the more simple of two models with equal likelihood evaluations will be preferred. The AIC provides a relative goodness-of-fit measure rather than an absolute measure as provided by the standard contingency-table based tests. This means that a model with the smallest AIC may be the best fitting of the group of models it was compared with but this does not mean the model necessarily matches the data well.

Rather than using a sequential procedure to simply select a single best model on which to base all inference it is more justifiable to incorporate the uncertainty in choosing between a range of models into the precision associated with an estimator. The use of relative AIC values can allow the uncertainty of selecting models to be incorporated into the parameter estimation process. If there are a set of models then AIC_{min} denotes the minimum AIC value across all models in the set and AIC_s denotes the AIC value for model s. Firstly, the differences between these values for each model are defined, $\delta_s = \mathrm{AIC}_s - \mathrm{AIC}_{min}$. Then, the weights are calculated as $\exp(-\delta_s/2)$ and can be considered as the relative support for model s given the data and the other models. Normalising these

values produces the weights:

$$w_s = \frac{\exp(-\delta_s/2)}{\sum_{s=1}^{S} \exp(-\delta_s/2)}$$

where S denotes the number of models being compared. Then if there exists a parameter θ that can be assumed to be common to all compared models, the weighted estimate of θ is given by

$$\hat{\theta}_s = \sum_{s=1}^{S} w_s \hat{\theta}_s$$

where $\hat{\theta}_s$ is the estimate of the parameter θ under model s and w_s is the weight corresponding to model s. Buckland *et al.* (1997) provide a detailed discussion of how to incorporate model uncertainty into inference and how potential bias in inference due to model misspecification can be considered to be a component of the variance associated with model based estimates. Model selection issues are covered in detail in Burnham and Anderson (2002).

2.2.2 Multiple-Age Models

The previous sections reviewed the classical CJS models for populations in which the restrictive assumption was made that survival and capture probabilities are not functions of an animal's age. In many biological populations this assumption is not always valid and there has been substantial work done to develop models in which survival and capture probabilities can vary with age.

The discussion in Section 2.2.1.1 already noted that age can be considered to be a discrete and dynamic covariate which allows the animals to be classified into age-based cohorts. The evolution of this dynamic covariate is deterministic so that, assuming the duration between sampling occasions is known, the animals in age class j in period i will,

assuming they survive, be in age class $j+1$ in period $i+1$. Of course, age classes can span a range of years and can span different ranges, depending on the assumed biology of the animal being studied, to define relatively homogeneous cohorts.

The structure of the data is broadly similar to that for the single-age models (Section 2.2). For the age-specific models the age of each animal at initial capture must be known, as well as the animal's capture history. By knowing an animal's age at first capture the deterministic nature of the evolution of an animal's age means that its age will be known on all future occasions. If the animal's age is not known on first capture then the types of model which can be used to analyse the data are reduced in number. If the age-structure of a model is relatively crude consisting of only a small number of age classes (e.g. new-borns and adults) which are clearly distinguishable on capture then it may be possible to use certain types of multiple age models (Pollock, 1981). However, in more complicated models it may be the case that a more detailed form of age-structure is used (e.g. age classes for each year of birth). If it is only possible to distinguish between a small number of distinct age classes on capture but not to distinguish animals by year of birth then only those animals caught and marked in the first age class identifiable on capture will have known ages in subsequent years. These cohort models (Buckland, 1982; Pollock *et al.*, 1990) then require animals to be marked in the first identifiable age class to produce estimates of age-specific capture and survival parameters. To assess any variation in capture and survival probabilities across different age classes it is important to consider the requirement for the duration between sampling periods to closely match the time taken for animals to move between consecutive age-classes. The general CJS model structured by cohorts has capture and survival probabilities that are specific to both time and age cohort. A result of this formulation allows the model to be viewed as a series of separate CJS models for each cohort. Each cohort is based on those animals caught and

marked in the first age class (typically "young" or age-0 animals) and a separate multinomial distribution conditional on the releases at each time period can be constructed for each cohort of first age class releases. Since the parameters are specific to time and age each combination of time and age will be unique to specific cohorts, for example animals that are in the second age class in period 2 will have different parameters for capture and survival compared to animals in all other age classes in period 2.

2.2.3 Multistate Models

The age-cohort models of the previous section can be thought of as a special case of the more general multistate model extension to the single-age CJS models. Multistate models can be defined as models that allow transitions between states. Generally these transitions are assumed to be stochastic rather than deterministic as in the age-cohort models. As noted in Section 2.2.1.1, the CJS model can be extended to accommodate dynamic, discrete variables such as location or phenotype. If it is assumed that capture and survival probabilities vary according to some measurable variable then it may be appropriate to stratify the population into distinct groups, which are assumed to be relatively homogenous, according to the values taken by the variable with separate analyses conducted for each group. This approach can work for static, discrete variables (e.g., gender) but for dynamic variables that change stochastically over time it can be difficult to define appropriate groups. For example, if breeding status was assumed to impact on the health and activity of animals it may be reasonable to assume there are different probabilities of capture and survival for breeders compared to non-breeders. However, in this scenario simply stratifying the population into two groups determined by breeding status on first capture may not be appropriate as animals that breed in period i are unlikely to breed in every single subsequent period, equally, animals that are non-breeders in period i will

not necessarily stay as non-breeders for the remainder of the study. The breeding status of animals not captured in period i will not be known either, providing further modelling complications. A multistate modelling approach allows the capture and survival probabilities to vary by state and also allows animals to move between those states from one time period to another and is more appropriate for these scenarios.

Schwarz *et al.* (1993) use matrix model extensions of the tag-recovery models developed by Brownie *et al.* (1985) to develop a general likelihood based approach to formulating multistate models. Brownie *et al.* (1993) also adopt a likelihood based approach and expand on the approach of Schwarz *et al.* (1993) by considering "Markovian" models in which state transition probabilities between periods i and $i+1$ are functions of the current state at i as well as "memory" models in which the transition probabilities depend not only on the state at i but also on the state at $i-1$. The multistate model formulation can also be used to address ecological questions concerning the transition rates between states (e.g., estimating site fidelity) or the influence of life history on vital rates (e.g., testing the association between breeding history and survival rates) (Nichols *et al.*, 1994; Nichols and Kendall, 1995). A detailed discussion of the structure, assumptions and estimation techniques for multistate models is provided by Brownie *et al.* (1993) and Williams *et al.* (2002, pp. 454-468).

2.2.4 Reverse Time Models

Another important development in the analysis of capture-recapture data is the reverse-time approach (Pradel, 1996). Under this approach the model structure is conditional on the last observed capture histories and the data are considered in reverse time order to obtain inference on the "recruitment" process. The data and model structure are both fairly similar to those discussed in Section 2.2 with the data, as before, consisting only of

observable capture histories. However, in contrast to the previously examined models the capture history pattern $00\ldots01$ is now included whereas $10\ldots0$ will not be. The model structure differs only through the temporal direction and this necessitates the definition of two new parameters for the model (Table 2.5). The parameters γ_i are referred to by Pradel (1996) as "seniority" parameters as they model the probability that an animal present in the population immediately prior to period i was present immediately after $i-1$ and must therefore be "old" in some sense.

γ_i	=	The probability that an animal present in the population immediately prior to period i is was present immediately after the $(i-1)^{th}$ sample. - the seniority probabilities
p'_i	=	The probability that an animal present in the population immediately after the i^{th} sample was captured during the i^{th} sample.
ξ_i	=	The probability of not being seen prior to the i^{th} sampling occasion for an animal present in the population immediately prior to the i^{th} sample.

Table 2.5: Notations and definitions for the Pradel's reverse-time models

It should be noted that losses on capture are handled slightly differently in these reverse-time models compared to the usual conditional forward-time CJS approach. Under the forward-time approach animals that were not released after capture in sampling period i were simply not included in the R_i releases and their capture history pattern would not include any events following the observed capture in i. For the reverse-time model, given that the model is conditioned on the last-observed capture an animal captured in sampling period i cannot have been lost on capture on any previous occasion. Thus, the fate of an animal following its last observed capture does affect the forward-time CJS models, but it does not affect the reverse-time models which ignore all events after the last observed capture.

The parameter ξ_i is analogous to the forward-time CJS model parameter χ_i (see Eq. (2.2.1)) and can be defined recursively as

$$\xi_i = (1 - \gamma_i) + \gamma_i(1 - p'_{i+1})\xi_{i+1} \quad \text{for } i = 2, 3, \ldots, K \tag{2.2.6}$$

with $\xi_K = 1$. The construction of this recursive definition is similar to that for the χ_i but in this case comprises the two different ways in which an animal can be seen prior to the i^{th} sample. Firstly, the animal may be a new entry to the populations, that is it is not a survivor from the $(i-1)^{th}$ period (which has the probability $1 - \gamma_i$). Alternatively, the animal is a survivor (which has probability γ_i), was not caught in the previous sample (with probability $1 - p'_i$) and was not seen prior to the $(i-1)^{th}$ sample (which has probability ξ_{i-1}).

Using the parameters defined in Table 2.5 conditional multinomial models can be constructed in a similar manner to those for the forward-time CJS models. The models condition on the animals caught for the last time on each sampling occasion and then use the numbers of these animals, x_ω, exhibiting each of the capture history patterns ω along with the probabilities of attaining these histories to construct the appropriate conditional multinomial model. As with the CJS models only the observed capture histories are included in the model, thus the null pattern corresponding to an animal that is present in the population at some point but never observed is not included in the model. Also, as the history pattern consisting of $K - 1$ 0's followed by a 1 was not included in the CJS formulation as no events occurred after the final possible capture, the analogous case holds here where the pattern consisting of a 1 followed by $K - 1$ 0's is not included as no events occur prior to the last observed capture.

Many of the assumptions and estimation procedures for these reverse-time models

are broadly analogous to the forward-time versions. Also the reverse-time models can be extended to accommodate a multistate structure (Nichols and Hines, 2002) which allows investigation into a wide range of estimation questions. The relative influence of animals across different states on the evolution of the population in a specific state can be investigated using these models and thus provide an appropriate framework for metapopulation analysis.

2.2.5 Incorporating Auxiliary Data

Capture-recapture studies can benefit significantly from incorporating auxiliary data into the model structure. As noted in Section 2.2.1.1, the use of covariates enables a more parsimonious approach to be taken in preference to conducting entirely separate analyses for cohorts of animals that are assumed to exhibit homogeneity with regard to the parameters of interest. The parsimonious approach may also increases the precision of all parameter estimates as all of the data is used in the estimation process. Information on animal level phenotypic covariates or environmental variables may, when incorporated into the capture-recapture analysis, provide a much improved prediction of the capture and survival rates compared to the standard CJS model analysis. A second reason is that the investigator may wish to explore a variety of biological hypotheses about the relationships between the parameters being estimated and the covariates.

Most studies that incorporate auxiliary information (Pollock, 2002) use covariates that can commonly be classified as either individual animal level covariates or group or environmental covariates. The group or environmental covariates will often be assumed to be dynamic variables which vary between sampling occasions but remain constant over the animals in the group (e.g., environmental indexes such as the North Atlantic Oscillation). Individual animal level covariates are, by contrast, usually assumed to be static variables

which are constant over time (e.g., birth weight). In much of the existing literature on capture-recapture studies that incorporate auxiliary information the analyses are conditional on the observed values of the covariates; typically no distribution of the covariate is specified. If the auxiliary information consists of a static variable then the modelling is relatively simple. Since CJS models condition on marked animals no unmarked animals are included in the model. Therefore, for static variables that can be recorded at any point in time there will be no missing covariate values and the analysis can be relatively straightforward.

For continuous and stochastic covariates the analysis becomes more problematic. For variables of this type some essentially arbitrary discretisation can be imposed to create a series of states and the transitions between them can be modelled using the multistate approach. For example, if it was believed that an animal's weight may influence its probabilities of survival and capture then this effect could be modelled by defining a sequence of discrete weight classes and specifying a model for the transition probabilities between them. These multistate models have the advantage that they can include time-specific covariates which vary on an individual level. Also, there is no distribution specified that constrains the covariate. The parameters (capture, survival and state transition rates) for animals in each individual state can vary independently from the parameters in any of the other states. However, this multistate approach of discretising continuous variables (e.g. Nichols *et al.* (1992)) can lead to a loss of information that may conceal any relationship, present in the data, between the variable and capture or survival rates. The large number of parameters required for the multiple states may mean it is not the most parsimonious model. Equally problems may arise with the fit of the chosen model. The choice of discretisation can be effectively arbitrary and the assumption of homogeneity within states may be violated if the continuous auxiliary variable is classified into too few

discrete states resulting in decreased precision for the parameter estimates. Conversely, if a large amount of grouping is imposed then the data may not contain enough information for all the parameters in the model, thus rendering some of them unidentifiable. As noted in Bonner and Schwarz (2006), another downside to this approach is that multistate models implicitly assume that all animals in a single state behave identically but will behave differently if they belong to different states. Therefore, animals with observed covariate values either side of a state classification boundary will be assumed to behave differently when their covariate values may in fact be very similar.

An alternative approach (Lebreton *et al.*, 1992) is to use a GLM-like modelling structure which allows the capture and survival rates to depend on linear combinations of variables (or covariates) through a "link" function of specified form. Lebreton *et al.* (1992) advocate using a scaled logistic link function as it constrains the estimated probabilities to lie in the plausible range of 0 to 1. The advantage to their approach over the multistate models is that it utilises the flexibility of GLMs allowing both discrete and continuous covariates to be incorporated in the linear predictor. An assumption implicit in this approach is that the relationships between the parameters of interest and the auxiliary information can be described using these relatively simple link functions. Typically, the form of the relationship determined by the link function is assumed to be constant over time which may not always be realistic. Also, as noted in Bonner and Schwarz (2006) the common use of the logistic function imposes the constraint that the relationship between the parameter being estimated and an explanatory covariate is monotonic. They suggest exploring the use of piecewise functions such as polynomials or splines to obtain more flexible relationships. One of the main issues with the GLM-like modelling approach is that whilst it can be extremely effective for static covariates problems arise for dynamic time-dependent ones when an animal is not captured. For static covariates that can be

observed at any time the value remains the same, and therefore known, for the duration of the study regardless of whether the animal is captured again after the initial release or not. For dynamic covariates that evolve deterministically, such as age, the missing values of the covariate can be interpolated.

For dynamic covariates that evolve stochastically on an individual basis, any sampling occasion during which the individual is not captured will result in a missed value of the covariate. Equally, for static variables that can only be recorded at a specific time (e.g. birth weight) the issue of missing covariates makes analysis more problematic. A full model analysis requires some method of modelling the distribution of the missing values taken by the covariate over the duration of the study (Pollock, 2002; Bonner and Schwarz, 2006). Pollock (2002) suggests that, in addition to the discretising continuous variables approach of Nichols *et al.* (1992) in which the weights of meadow voles were classified into discrete weight states, an alternative would be to develop a full likelihood approach by integrating out unobserved covariates. As with the GLM-like approach (Lebreton *et al.*, 1992) this method requires the distribution of the covariate to be specified with regard to how it evolves over time. Alternatively Pollock (2002) suggests conducting an analysis conditional on the covariates in the context of a missing data problem. Under this scenario the EM algorithm (Dempster *et al.*, 1977) could be used to model the change in covariates over time.

Pollock (2002) also advocates Bayesian methods for the full likelihood approach and as a way of dealing with missing covariate values. The model likelihood can be evaluated using numerical integration to integrate out the missing values or Markov Chain Monte Carlo (MCMC) methods can be used to impute the missing values which can be treated as variables to be estimated in a Bayesian framework. A summary of Bayesian approaches

to the analysis of Capture-Recapture studies is given in Section 2.5. The approach of Bonner and Schwarz (2006) utilises a Bayesian methodology to provide a solution to the problem of modelling capture and survival rates using continuous stochastic covariates that vary with both time and individual.

The use of auxiliary information in capture-recapture models presents some issues regarding goodness-of-fit tests (White, 2002). If models are not fitting well then it is recommended (Pollock, 2002) to adjust the AIC and the variance of the parameter estimates to correct for possible over-dispersion in the data. A correction parameter, c, is required but White (2002) notes that

> *no general, robust, procedures are currently available for estimating c.*

White (2002) also warns against placing too much faith in the model-based relationships between auxiliary variables and estimated parameters unless the auxiliary variable is included as a component of a "manipulative investigation". Without this the relationship can only be described as correlative.

2.2.6 Summary

The previous sections have introduced Cormack-Jolly-Seber models which provide open-population estimates based on observed capture history data. The open-population form of these models requires an extension of the closed population formulation by means of incorporating non-stationary survival rates. The models are conditioned on the first capture of an animal and therefore the data consists of all capture histories containing at least one capture. The numbers of animals recaptured during the study are conditioned on the number of releases during each sampling period of the number of unmarked animals, only once it has been marked initially is an animal available to be released for resampling. As

a result the model only uses part of the information that is available in the sample and a more efficient use of the data would require the uncaptured animals to be incorporated into the model. This will be discussed in more detail in Section 2.3. The data structure used for the CJS model results in a conditional statistical form consisting of the product of a series of multinomial distributions for each cohort of recaptured animals. This model form can be constructed as a likelihood from which maximum likelihood estimates and their variances can be derived.

Numerous extensions to the CJS models were discussed with incorporating a cohort-structure one of the most prominent extensions covered in the literature. Using flexible GLM-type models or framing the CJS models in a multistate context allowed age, phenotypic and geographical covariates to be incorporated into the general modelling approach and enabled capture and survival parameters to be modelled as functions of auxiliary data. By imposing constraints on the time-specific, cohort-based CJS models, a wide variety of biological hypotheses could be examined through hypothesis testing, via likelihood ratio tests, to compare nested models.

Although the models discussed so far can provide an extremely flexible framework in which to study the capture and survival parameters of a population, the models are still limited to inference on these rates. By incorporating additional stochastic elements into the CJS models the events before initial capture become part of the model framework which expands the range of biologically interesting quantities that can be estimated from the model. Inference on recruitment, mortality and population abundance can all be made using this expanded modelling approach and these ideas are discussed in Section 2.3.

2.3 Extending Mark-Recapture Models

The previous chapter focussed on the development of the classical Cormack-Jolly-Seber model for estimating survival and capture rates in an open population. This chapter considers the estimation of both recruitment into the population and of population size. Much of the methodology used to formulate and analyse these models is an extension of that in Section 2.2. The data itself are constructed in the same way with animals being marked, released and recaptured at a series of discrete sampling periods throughout the study. The models in Section 2.2 were conditional on first capture and the capture histories were then expressed in terms of the probabilities of subsequent capture and survival events. This conditional component appears in some of the models in this section.

This section will compare the classical CJS model with the broader techniques discussed here. Different approaches to formulating models that can produce estimates of population abundance and recruitment will be discussed and compared. Extensions to the models and the use of auxiliary information will also be covered briefly. The structure of this section will broadly follow the outline of Williams *et al.* (2002, 495-522).

2.3.1 The Jolly-Seber Model

The estimation of abundance and recruitment using the models discussed in this section requires information on the numbers of animals caught on each occasion that have remained uncaught previously. From this information the ratio of marked animals to previously unmarked animals in a sample is used to obtain an estimate for the population size during that sampling period.

The models in this section will be restricted to focus on the single-age class structure. One reason for this is that abundance cannot be estimated for the first age-class in an

age-cohort structured model using only count data. To estimate the abundance for the first age-class an estimate of the associated capture probability is required. For a time-specific age-structured study the capture probability for the first age-cohort in the i^{th} sample is estimated conditionally on the number of animals in the first age-cohort that are known to be alive (through capture) during the i^{th} sampling period. However, there can be no total of first age-cohort animals known to exist in the population owing to previous and subsequent capture since there were no animals in a younger age-class in a previous sampling occasion. Consequently, there is no subset of animals to condition on and no estimate can be obtained for the capture probability of the first age class. Consequently, single-age models have been the basis for much of the formal development of abundance estimation techniques (Jolly, 1965; Seber, 1965; Pollock *et al.*, 1990; Schwarz and Arnason, 1996). Equally, it can be difficult to assign "older" animals to the correct age-cohort on first capture.

The Cormack-Jolly-Seber model was discussed in section 2.2 and comprised a single-age model with time-specific capture and survival probabilities. Of the three studies the model was based on Jolly (1965) and Seber (1965) include components in their models that allow for the estimation of population size and abundance whereas Cormack (1964) focuses on the conditional modelling of survival and capture parameters. Hence, if population abundance and recruitment are incorporated into the modelling framework the resulting classical model is referred to as the Jolly-Seber (henceforth JS) model.

2.3.1.1 Model Formulation

The data structure for the single-age JS model is broadly similar to that of Section 2.2. Capture-history data once again consist of the set $\mathbf{x} = \left[x_{\omega_1}, x_{\omega_2}, \ldots, x_{\omega_{2^K}}\right]$ where x_ω

denotes the number of animals exhibiting capture history pattern ω and there are 2^K possible capture history patterns for a K sample study. The notation of Williams *et al.* (2002, 495-522) includes terms to explicitly model losses on capture whereby individuals are removed from the population. These terms are omitted here and the notational conventions of Pollock *et al.* (1990) are adopted instead to provide a more direct comparison with the expressions in Section 2.2. The definition of several of the parameters used in the formulation for the CJS models (see Section 2.2 and Table 2.3) differs slightly for the JS model. For the JS model, the u_i represent the number of unmarked animals captured in the i^{th} sample whereas for the CJS model they were defined as the number of animals that are released after the i^{th} sampling period having been previously uncaught. Thus, for the CJS model they were restricted to only those animals that had been captured and then subsequently released, whereas under the JS model the u_i can be either released or removed following capture. Table 2.6 contains the notation for the JS models discussed in this section and consists of both model parameters that are unknown random variables to be estimated and observed statistics.

To compare the model formulation for the CJS and JS models it is constructive to compare the probabilistic representations for analogous components of the respective models. To compare with Equation (2.2.2), the JS formulation of the model component for unmarked animals captured in the first period, u_1 is presented. Considering the capture histories available for a three-sample study yields the model formulation:

$$\begin{aligned} Pr(\{x_\omega\}\,|u_1) \;=\; & \left[\frac{U_1!}{u_1!\,(U_1-u_1)!}p_1^{u_1}(1-p_1)^{U_1-u_1}\right] \\ & \times\left\{\frac{u_1!}{\prod_\omega x_\omega!}[\chi_1]^{x_{100}}[\phi_1(1-p_2)\phi_2 p_3]^{x_{101}}\right. \\ & \left.\times\,[\phi_1 p_2\chi_2]^{x_{110}}[\phi_1 p_2\phi_2 p_3]^{x_{111}}\right\} \end{aligned} \tag{2.3.1}$$

Parameters		
M_i	=	The number of marked animals in the population at the time of the i^{th} sampling period ($i = 1, \ldots, K$).
N_i	=	The total number of animals in the population available to be captured in the i^{th} sampling period ($i = 1, \ldots, K$).
U_i	=	The number of unmarked animals in the population at the time of the i^{th} sampling period ($i = 1, \ldots, K$).
B_i	=	The number of new animals entering the population between sampling periods i and $i+1$ and present at the time of the $(i+1))^{th}$ sampling period ($i = 1, \ldots, K-1$).
p_i	=	The probability that an animal in the study population during sampling period i is captured during sampling period i ($i = 1, \ldots, K$).
ϕ_i	=	The probability that an animal in the study population during sampling period i survives until $i+1$ and remains in the population ($i = 1, \ldots, K-1$).
χ_i	=	The probability that an animal alive in the study population during sampling period i is neither captured nor observed again during any future sampling period ($i = 1, \ldots, K$).
Statistics		
m_i	=	The number of marked animals captured in the i^{th} sample ($i = 1, \ldots, K$).
u_i	=	The number of unmarked animals captured in the i^{th} sample ($i = 1, \ldots, K$).
n_i	=	m_i+u_i, the total number of animals captured in the i^{th} sample ($i = 1, \ldots, K$).
R_i	=	The number of the n_i that are released after the i^{th} sample ($i = 1, \ldots, K-1$).
r_i	=	The number of the R_i releases at i that are recaptured ($i = 1, \ldots, K-1$).
z_i	=	The number of animals caught before sample period i, not caught during i, and recaptured on some sampling period after i ($i = 2, \ldots, K-1$).

Table 2.6: Parameters and Statistics for the single-age Jolly-Seber model.

where the index ω ranges over the capture histories including capture on the first sampling occasion: 100, 101, 110 and 111.

From comparison with Equation (2.2.2) it can be seen that, under the JS approach, the representation of the model component for animals first caught in sampling period 1 does not condition on the new releases from the first sample. Instead the first term is a binomial term that models the number of unmarked animals that are captured during the first sample u_1 from the available population of initially unmarked animals U_1. The initial capture probability p_1 did not appear in Eq. 2.2.2 but it does in Eq. 2.3.1. For a three-sample study the full model for the capture history data is obtained by considering each group of previously unmarked animals captured on each sampling occasion, u_1, u_2 and u_3 formulating their probabilistic expressions, as in Eq. 2.3.1, and taking their product.

2.3.1.2 Model Assumptions

The modelling assumptions for the Jolly-Seber model are analogous to those listed for the Cormack-Jolly-Seber model in Section 2.2.1.1. However, an important difference occurs for the first assumption. Under the CJS model it was assumed that every marked animal in the population at the time of the i^{th} sample has the same probability of being recaptured. This is modified for the JS model since the capture parameter p_i applies to both marked and unmarked animals as can be seen in Eq. 2.3.1. Therefore, under the JS model, the first assumption is modified to state that every animal present in the population at the time of the i^{th} sample $i = 1, 2, \ldots, K$ has the same probability, p_i of being captured. By including both marked and unmarked animals in the assumption of homogeneous capture probability there are now additional ways in which this assumption can be violated. If certain animals within the population are more likely to be caught than others and this distinction is upheld for the duration of the study then the sample of marked animals obtained will tend to consist of those individuals with the higher capture

probabilities. This leads to a difference in the average capture probabilities for the marked and unmarked animals, even though the assumption is that they are equal, and will lead to a biased estimate of population abundance. One example would be if capture probability is determined by permanent trap-response. In this scenario the marked and unmarked animals will have different capture probabilities. For the conditional CJS models this is not a problem as they condition on first capture and so every observed animal included in the data will have the same recapture probability. For the JS models the marked animals are used to estimate the capture probability under the assumption that the same capture probability applies to the unmarked animals; the presence of trap-response violates this assumption.

As discussed for the Cormack-Jolly-Seber model (see Sections 2.2.1.1 and 2.2.5) there are a variety of approaches that have been developed to allow the models to meet the required assumptions. Many of these suggestions can be implemented for the Jolly-Seber models also. Grouping the population into (assumed) homogeneous strata is a common way of accounting for variation in the capture and survival probabilities associated with static and discrete variables. Discrete dynamic variables can be analysed using further extensions of the models. If the value of the variable changes deterministically (e.g. age), special reduced forms of multistate models can be formed in which transition probabilties between states are known *a priori*. The more general case of discrete dynamic variables that evolve stochastically can be approached using a multistate model formulation (see Section 2.3.2.2). Incorporating auxiliary information into the JS model formulation is discussed in more detail in Section 2.3.3.

2.3.1.3 Estimation

As for the Cormack-Jolly-Seber models (see Section 2.2.1.2) estimation techniques are based on numerically maximimising the appropriate likelihood functions. In particular the component of Eq. (2.3.1) that does not contain the unknown random variables U_i can be used to obtain maximum likelihood estimates of the survival, ϕ_i, and capture, p_i, parameters respectively. Although the historical emphasis has been on obtaining analytical solutions, modern methods emphasise the use of numerical solutions.

The extensions to the CJS model allow the JS model to estimate both recruitment and population abundance. The remaining sections will primarily focus on the estimators for these random variables. The definitions provided in Table 2.6, can be used to express the closed-form estimators for the relevant parameters. For example, a moment estimator for abundance (Pollock, 2002) can be shown to be:

$$\hat{N}_i = \frac{n_i \hat{M}_i}{m_i} \tag{2.3.2}$$

where m_i, n_i and $\hat{M}_i$ are the marked and total numbers of animals captured and the estimated total number of marked animals in the population in the i^{th} sample respectively. The approximately unbiased analogue of Eq. (2.3.2) is often used for JS models with time-specific capture and survival rates and is given (Pollock *et al.*, 1990) as

$$\tilde{N}_i = \frac{(n_i + 1)\tilde{M}_i}{m_i + 1} \tag{2.3.3}$$

where

$$\tilde{M}_i = m_i + \frac{(R_i + 1)z_i}{r_i + 1} \tag{2.3.4}$$

and $\sim$ denotes approximately unbiased estimators. For the estimator of recruitment

between i and $i+1$, the relationship

$$E\left[N_{i+1}|N_i, B_i\right] = B_i + \phi_i\left(N_i - n_i + R_i\right) \tag{2.3.5}$$

is constructed and is comprised of two components: the new recruits in the i^{th} sample (B_i) and the survivors from the previous sample ($\phi_i\left(N_i - n_i + R_i\right)$). The term $-n_i + R_i$ represents losses on capture. Using Eq. 2.3.5, an estimator of recruitment between i and $i+1$ is given (Pollock *et al.*, 1990) as:

$$\hat{B}_i = \hat{N}_{i+1} - \hat{\phi}_i\left(\hat{N}_i - n_i + R_i\right), \tag{2.3.6}$$

where

$$\hat{\phi}_i = \frac{\hat{M}_{i+1}}{\hat{M}_i - m_i + R_i}. \tag{2.3.7}$$

The approximately unbiased version (Pollock *et al.*, 1990) is given as:

$$\tilde{B}_i = \tilde{N}_{i+1} - \tilde{\phi}_i\left(\tilde{N}_i - n_i + R_i\right), \tag{2.3.8}$$

where

$$\tilde{\phi}_i = \frac{\tilde{M}_{i+1}}{\tilde{M}_i - m_i + R_i}. \tag{2.3.9}$$

It should be noted that all of the above abundance and recruitment estimators require the estimation of the capture probability p_i. As capture probability can only be estimated for periods $2, 3, \ldots, K-1$ (see Section 2.2.1.2), the abundance and recruitment estimates are only available for periods $2, 3, \ldots, K-1$ and $2, 3, \ldots, K-2$ respectively. Approximate variances and covariances for the estimated parameters are described in Pollock *et al.* (1990). It should also be noted (Buckland, 1980) that the survival estimates can exceed unity and estimated recruitment may be negative under the Jolly-Seber model. Equally, analytic confidence intervals may contain values outwith the biologically feasible range, even if the point estimates of the parameters themselves lie within the permissible range.

To address these issues Buckland (1980) uses a parametric bootstrap to estimate both confidence intervals and variances. However, Buckland (1980) notes that, in general, the simulation-based approach will not always work on the standard Jolly-Seber estimates as some may take impossible values. Instead the method will only work for modified estimates such as those presented in Buckland (1980). Also, the use of the parametric bootstrap in Buckland (1980) fails to provide variance estimates for parameters corresponding to the first or last sample periods in the study owing to the non-identifiability of these parameters under the Jolly-Seber model. A solution to this involves the use of bootstrapping encounter histories (Buckland and Garthwaite, 1991) and conditioning on the number of observed capture histories. This approach yields variance estimates and robust confidence intervals for all estimated parameters in the classical Jolly-Seber model.

2.3.1.4 Special Cases

So far the focus has been on single-age Jolly-Seber models with a time-specific parameterisation for both survival and capture rates. Extensions for this model have been investigated and are summarised in Pollock *et al.* (1990). Firstly, the special cases of partially open models are considered.

The first type of partially open model is the death-only model in which births and immigration are assumed to be negligible. This form of this model is now similar to that of the CJS model, however in this case the initial number of animals is not known. In this scenario deaths and (permanent) emigration are both allowable processes. By reducing the number of parameters required to be estimated (no estimates of the births are needed) the precision of the remaining parameters increases. These models are described in Jolly (1965) as a special case of his general open model and it should be noted that the population abundance estimates can now be made for $i = 1, 2, \ldots, K-1$ unlike for

the JS model estimates in Eq. (2.3.2) which were not defined for the first time period. Expression for the parameter estimates and their associated variances are presented in Jolly (1965) and Pollock *et al.* (1990).

The second form of model is the births-only model in which there are no deaths or emigration in the model and change is the result of recruitment (births and immigration) alone. In this scenario the number of marked animals in the population at sampling period i, M_i, becomes a known statistic rather than an unknown random variable. The known nature of the M_i results in Lincoln-Petersen type estimators (Williams *et al.*, 2002, pp. 291-293) of population size and is defined for $i = 2, 3, \ldots, K$. If no losses on capture are allowed in this model then the estimate of new recruits during period i simply becomes the difference in population estimates between periods i and $i + 1$. That is, $\hat{B}_i = \hat{N}_{i+1} - \hat{N}_i$. Expressions for the parameter estimates and their associated variances are again given in Jolly (1965) and Pollock *et al.* (1990).

Reduced-parameter models (Jolly, 1982) have also been considered as restricted versions of the general Jolly-Seber model. These models are suitable for scenarios in which both births and deaths may occur and it may be reasonable to assume that survival and/or capture are constant throughout the study. The approaches in Crosbie and Manly (1985) and Schwarz and Arnason (1996) also describe reduced-parameter model formulations and will be discussed in Section 2.4. The three alternative models can be classified as:

1. The Jolly-Seber model with time-specific survival and capture parameters. (ϕ_t, p_t)
2. The Constant Survival model with constant survival parameters $\phi_1 = \phi_2 = \ldots \phi_{K-1}$. (ϕ, p_t)

3. The Constant Capture model with constant capture parameters $p_1 = p_2 = \ldots p_K$. (ϕ_t, p)

4. The Constant Survival and Capture model with both survival and capture parameters constant over the study. (ϕ, p)

If the assumptions made in the reduced-parameter models are reasonable then Jolly (1982) demonstrates that the parameter estimates will experience a gain in precision. Using reduced-parameter models can also lead to a larger number of sampling occasions for which the parameters can be estimated. Under the model (ϕ, p), abundance can be estimated for each time period. Under these reduced parameter models there are no closed form expressions for the maximum likelihood estimates and these need to be obtained using iterative numerical optimisation techniques. Jolly (1982) provides estimators for the variances and covariances under the reduced-parameter models. It should also be noted that the requirement to use numerical methods provides all possible intermediate possibilities.

2.3.2 Extensions

The Jolly-Seber model can also be extended and generalised in similar ways to the Cormack-Jolly-Seber model through the specification of an age-cohort structure or a state-based model classification. Many of the ideas behind these models were introduced in Sections 2.2.2-2.2.3 and only a brief review will be presented here.

2.3.2.1 Multiple Age Models

The focus of the preceding sections has been on the single-age Jolly-Seber model (Jolly, 1965; Seber, 1965) which defines animals as belonging to a single state. Fully age-specified models were considered in Section 2.2.2 where the sampling design allowed the age of each

captured animal to be accurately determined. Consequently, it is possible to obtain age-specific abundance estimates for each of the central $K-2$ sampling periods and for all but the first age class. Pollock's robust design (Pollock *et al.*, 1990; Kendall and Nichols, 1995; Williams *et al.*, 2002, pp.523-544) can be used to obtain estimates of abundance for the age class 0.

The other form of age-cohort models discussed in Section 2.2.2 in which age is known only for marked animals typically do not allow abundance estimation. These cohort models (Buckland, 1982; Pollock *et al.*, 1990) are typically used for animals marked as young or age 0. These models are appropriate when it is not possible to easily age animals but there is reason to believe that the survival rate may be age-dependent. Due to the fact that the unmarked animals caught in period i, u_i, cannot be assigned an age-class no age-specific capture rates can be determined and, consequently, no abundance estimation is possible for these forms of cohort-models. It is possible to obtain cohort and time-specific survival estimates (Buckland, 1982; Pollock *et al.*, 1990). Pollock *et al.* (1990) obtains survival estimates for a study in which the newly marked animals in each year are defined as a cohort and the time since marking is recorded. For this form of study Buckland (1982) notes that survival rates are functions of both time since marking and temporal effects and describes analyses which emphasise one or other of these effects.

2.3.2.2 Multistate Models

Once again, the discussion on extending the Cormack-Jolly-Seber model to a multistate modelling framework (see Section 2.2.3) covered many of the issues that are relevant to performing a similar expansion for the Jolly-Seber model. When populations were classified by states for both the Markovian and memory models discussed in Section 2.2.3,

the capture probability was modelled as being both state and time-specific. Hence, if the focus of inference was on the animals belonging to state g during sampling period i, a Horvitz-Thompson like estimate (Borchers *et al.*, 2002) of the abundance for that state would be given as

$$\hat{N}_i^g = \frac{n_i^g}{\hat{p}_i^g}$$

where $\hat{n}_i^g$ is the number of animals in state g at time i that are caught during period i, and $\hat{p}_i^g$ is the estimated capture probability for that specific group of animals. If the population was classified into a total of G states then the total abundance for animals across all states is obtained by summing the state-specific abundance estimates:

$$\hat{N}_i^{\Sigma} = \sum_{g=1}^{G} \hat{N}_i^g.$$

Once again, a state specific estimate of capture, $\hat{p}_i^g$, is required to obtain an abundance estimate and this implicitly assumes that the marked and unmarked animals in state g during sample period i possess the same capture probabilities.

This is still an active area of research and no efficient parameterisation of the multistate Jolly-Seber model has yet been published. A framework respecting proper conditioning has yet to be developed.

2.3.3 Auxiliary Data

Many of the ideas for incorporating auxiliary information into Jolly-Seber models for analysing capture-recapture data have been covered in the discussion on Cormack-Jolly-Seber models in Section 2.2.5. Many of the techniques outlined in Lebreton *et al.* (1992) and reviewed by Pollock (2002) are applicable to the Jolly-Seber models. By extending the analysis of capture-recapture data to estimate population abundance Jolly-Seber

models incorporate unmarked animals which can result in auxiliary information being unrecorded for animals that are not observed. The problem of modelling missing covariate values is well suited to Bayesian modelling techniques and recent developments of extensions to capture-recapture models (Link and Barker, 2004, 2005; Bonner and Schwarz, 2006; Schofield and Barker, 2008; Dupuis and Schwarz, 2007) use a Bayesian framework to impute missing covariate values when modelling demographic parameters as functions of covariates (see Section 2.5).

2.3.4 Model Selection

An alternative approach to Eq. (2.3.1) for expressing the distribution function for the observed data in a Jolly-Seber model is now presented. This approach is described in detail in Williams *et al.* (2002, 495-522) and is based on the approach of Seber (1982). This approach models the data using the m_{ij}-array statistics introduced in Section 2.2. The significant property of this approach is that it decomposes the distribution function into three separate likelihood components. Using the original notation in Table 2.6 and the extra notation in Table 2.7 the model can be expressed as

$$\begin{aligned} Pr\left(\{u_i\},\{d_i,d_i'\},\{m_{ij}\}\right) &= P_1(\{u_i\}\,|\,\{U_i\},\{p_i\}) \\ &\quad \times P_2(d_i,d_i'|m_i,u_i),\eta_i,\eta_i') \\ &\quad \times P_3(m_{ij}|R_i,\phi_i,p_i) \end{aligned} \tag{2.3.10}$$

The first component of this decomposition models the capture of unmarked animals and takes the same binomial form as the first term in Eq 2.3.1:

$$P_1(\{u_i\}\,|\,\{U_i\},\{p_i\}) = \prod_{i=1}^{K}\left[\frac{U_i!}{u_i!\,(U_i-u_i)!}p_i^{u_i}(1-p_i)^{U_i-u_i}\right] \tag{2.3.11}$$

The second component of Eq. 2.3.10 models the marked and unmarked animals that are not released back into the population after capture and can be expressed as:

$$\begin{aligned} P_2(\{d_i, d_i'\} \mid \{m_i, u_i\}), \{\eta_i, \eta_i'\}) &= \prod_{k=1}^{K} \left[\binom{m_i}{d_i} (\eta_i)^{d_i} (1-\eta_i)^{m_i - d_i} \right] \\ &\times \prod_{k=1}^{K} \left[\binom{u_i}{d_i'} (\eta_i')^{d_i'} (1-\eta_i')^{u_i - d_i'} \right] \end{aligned} \tag{2.3.12}$$

The third component of Eq. 2.3.10 models the conditional probability distribution of the animals recaptured at each period and is given by:

$$\begin{aligned} P_3(\{m_{ij}\} \mid \{R_i\}, \{\phi_i, p_i\}) &= \prod_{i=1}^{K-1} \frac{R_i!}{(m_{i,i+1})!(m_{i,i+2})! \ldots m_{i,K}!(R_i - r_i)!} (\phi_i p_{i+1})^{m_{i,i+1}} \\ &\times \{[\phi_i(1-p_{i+1})\phi_{i+1}p_{i+2}]^{m_{i,i+2}} \ldots \qquad (2.3.13) \\ &\quad \ldots [\phi_i(1-p_{i+1})\phi_{K-1}p_K]^{m_{i,K}\chi_i^{R_{i-r_i}}} \Big\} \qquad (2.3.14) \end{aligned}$$

This component is simply the Cormack-Jolly-Seber conditional model for the marked animals that are released back into the population R_i. It could also be written using the expression in Eq. (2.2.4) if the conditioning was on the number of unmarked animals u_i caught in each sampling period.

Assessing the goodness-of-fit for the JS models and comparing nested models is based on component P_3 of the likelihood representation in Eq.(2.3.10). This component conditions on the number of releases in each time period and incorporates the subsequent capture-history data on these marked animals. Therefore, since the classical goodness-of-fit and model comparison tests are calculated using the capture histories of the animals, they are typically based on the P_3 component of the likelihood. The second component P_2 is included for completeness but typically does not form a part of classical inference on Jolly-Seber models, although tests could be performed if issues such as the temporal

Parameters		
η_i	=	The probability that a marked animal captured during sampling period i is released back into the population.
η_i'	=	The probability that an unmarked animal captured during sampling period i is released back into the population.
Statistics		
m_{ij}	=	The number of marked animals captured and released on the i^{th} sample that are next caught on sampling period j ($i = 1, \ldots, K-1$ and $j = i+1, \ldots, K$).
d_i	=	The numbers of m_i that are not released back into the population at i.
d_i'	=	The numbers of u_i that are not released back into the population at i.

Table 2.7: Extra Parameters and Statistics for an alternative representation of the single-age Jolly-Seber model.

variation of release rates or equality between the release rates for marked and unmarked animals were of interest. The first component, P_1, is useful in the estimation of population abundance but is not used to assess model goodness of fit. However in alternative parameterisations (Crosbie and Manly, 1985; Schwarz and Arnason, 1996) the entry of previously unmarked animals, u_i, is explicitly modelled using entry probabilities. These parameterisations then allow the u_i to be incorporated into the model selection and fit assessment inference.

The mechanics of model selection and goodness-of-fit testing for the Jolly-Seber model follow much the same procedures as discussed in Section 2.2.1.4 for the Cormack-Jolly-Seber models. The use of an information-theoretic approach to model selection such as AIC and its small-sample analogues are preferred to the more traditional approaches involving likelihood ratio tests. Incorporating model uncertainty into inference by averaging across multiple models and weighting each parameter estimate by the model AIC

can provide more robust parameter estimates and is also recommended.

2.4 Alternative Models

Expressing the Jolly-Seber model in the decomposed likelihood form of Eq. (2.3.10) enables a direct comparison to be made with some of the major alternative developments of the Jolly-Seber model. Many of these alternatives are likelihood-based and adopt a similar form of likelihood decomposition to that given for the m_{ij}-array sufficient statistics based representation of Eq. (2.3.10). The three components P_1, P_2 and P_3 can be thought of as modelling initial capture, removals and recaptures respectively. The alternative models discussed in the following sections typically retain the form of the P_2 and P_3 components but offer alternative models for the way in which animals first enter the population, these models are encapsulated in the P_1 component of the likelihood.

The open population models covered so far allow for a population to experience both recruitment and removals. The removals are typically defined as deaths, emigration or losses on capture and, although there is often confounding of estimates of mortality as distinguished from emigration, these processes are explicitly included in the CJS and JS models through the survival parameters ϕ and the losses on capture parameters η, η'. The estimation of recruitment is a more complicated issue and the existing model structure (see Eqs (2.2.3),(2.3.10) does not include any stochastic component with which to explicitly model recruitment. Instead, as in Eq. (2.3.6), the estimate of recruitment is obtained deterministically from the difference between the estimated abundance at one time period minus the estimated number of survivors from the previous time period. Therefore, the B_i's are not incorporated into the model and their estimates are obtained from conditioning of the number of unmarked animals, U_i, estimated to exist in the population at each time period.

The alternative models for the initial capture component, P_1, have focused on the recruitment process (Crosbie and Manly, 1985; Cormack, 1989; Schwarz and Arnason, 1996). The approach of Schwarz and Arnason (1996) synthesises the earlier developments of Crosbie and Manly (1985) and expands on their approach to provide a general modelling framework that explicitly incorporates a recruitment process and can incorporate all the simple Jolly-Seber models into a single modelling paradigm. This section reviews the alternative models that have been proposed to incorporate a specific recruitment model and begins with a description of some of the problems with the previous models before describing the Schwarz and Arnason (1996) approach.

2.4.1 Model Structure

The Jolly-Seber (Jolly, 1965; Seber, 1965) approach assumes that the unmarked animals in the population, the U_i are fixed parameters and that the recruits to the population can be expressed in terms of the U_i's as $\hat{B}_i = \hat{U}_{i-1}\hat{\phi}_i(\hat{U}_i - u_i)$. This then allowed the first component, P_1, to be expressed in the products of binomials form given in Eq. (2.3.11). From P_1, an estimate of U_i can be obtained: $\hat{U}_i = \dfrac{u_i}{\hat{p}_i}$. The relationship between B_i and U_i can then be used to estimate the net recruits $\hat{B}_i$. However, Schwarz and Arnason (1996) noted several problems with this approach. With the births not appearing in the likelihood it is difficult to impose any restrictions on the estimates of B_i. For example, there is no simple way to ensure that known periods experiencing no recruitment can be implemented as a model constraint. The numerical evaluation of the likelihood can result in negative estimates for the B_i; there is no clear way to maximize the likelihood ensuring that all estimated B_i are non-negative. The reduced-parameter models of the general Jolly-Seber model, in which the model is restricted to either no-births, no-deaths

or both, cannot be expressed using the reduced form of P_1.

One approach to ensuring the numerical optimisation of the likelihood results in non-negative birth estimates centered on log-linear models (Cormack, 1989). These models included a parameter Ψ_i defined such that the number of unmarked individuals in the population at the time of the $(i+1)^{\text{st}}$ sample is Ψ_i times the number of unmarked animals surviving from the i^{th} sample. From this formulation, the relationship between uncaught animals and births is given by

$$U_{i+1} = U_i(1 - p_i)\phi_i\Psi_i = U_i(1 - p_i)\phi_i + B_i.$$

The parameterisation of the models formulated by Cormack (1989) has a direct correspondence with the rates for capture, survival and recruitment in the usual JS models. The advantage of this approach is that negative estimates of recruits can be avoided by constraining the Ψ_i to be non-negative. However, there are some disadvantages to this approach. The Ψ_i are only an indirect estimate of the B_i and it can be difficult to obtain the equivalent restriction on the Ψ_i to impose a constraint on the B_i. Deriving standard errors of the back-transformed parameters of biological interest is a non-trivial process and no method is presented to do so. This issue is later resolved (Cormack, 1993).

The approach taken by Schwarz and Arnason (1996) builds on the work of Crosbie and Manly (1985) who reparameterised the Jolly-Seber model by defining a new "super-population" parameter. This parameter, N_i, is defined as the total number of animals that exist in the population at i and survive until the next sample time, thus it gives the number of animals available to be sampled in the population at any point during the study. They define the parameter, B_i, as the number of animals that enter the population between sampling occasions i and $i + 1$ and are available to be sampled at $i + 1$. The

relationship between these two parameters is then

$$N = \sum_{i=0}^{K-1} B_i \tag{2.4.1}$$

where $B_0 = N_1$ since all the animals in the population initially, N_1, are new entries relative to the first sampling occasion. Schwarz and Arnason (1996) then model these random variables B_i as realisations from a multinomial distribution. The B_i's are referred to as births and in this context the term birth is used to define any mechanism by which new animals can enter the population and can encompass birth, immigration, etc. The multinomial model for the B_i includes entry parameters β_i which determine the point at which the new births enter the sampling population. Hence the births model can be expressed as

$$\{B_1, \ldots, B_{K-1}\} \sim \text{Multinomial}(N; \beta_1, \ldots, \beta_{K-1}) \quad \text{and} \quad B_0 = N - \sum_{i=1}^{K-1} B_i. \tag{2.4.2}$$

They then define another set of parameters which are given, recursively, as $\Psi_1 = \beta_0$ and $\Psi_{i+1} = \Psi_i(1 - p_i)\psi_i + \beta_i$. These parameters are used to model the number of unmarked animals that, as a proportion of the superpopulation N, are in the population and captured on each sampling occasion. The sum for Ψ_{i+1} consists of the two components, firstly the probability that an animal was in the previous population, avoided capture in i and survived until $i + 1$ and secondly the probability that the animal was a new recruit in period i. From this definition it can be seen that $\Psi_i p_i$ corresponds to the capture of unmarked animals at sampling period i. This results in a multinomial model for the capture of previously unmarked animals at each sampling occasion

$$\{u_1, \ldots, u_K\} \sim \text{Multinomial}(N; \Psi_1 p_1, \Psi_2 p_2, \ldots, \Psi_K p_K).$$

This corresponds to the first component P_1 of the standard Jolly-Seber distribution function (Eq. 2.3.11) and can be written as:

$$P_1'(\{u_i\}|N,\{\beta_i\},\{p_i\},\{\phi_i\}) = \frac{N!}{u_1!u_2!\ldots u_K!(N-u_\cdot)!}\left[1-\sum_{i=1}^{K}\Psi_1 p_i\right]^{N-u_\cdot}\prod_{i=1}^{K}(\Psi_i p_i)^{u_i} \tag{2.4.3}$$

where P_1' corresponds to the Schwarz-Arnason form of P_1, and $u_\cdot$ denotes the total number of unmarked animals captured during the study,

$$u_\cdot = \sum_{i=1}^{K} u_i.$$

The $N-u_\cdot$ term represents all the animals that exist in the superpopulation but are never caught. The expression P_1' is a function of the parameters $N,\{\beta_i\},\{p_i\}$ and $\{\phi_i\}$, which are subject to the constraint:

$$\sum_{i=0}^{K-1}\beta_i = 1.$$

Then, by conditioning on the total number of unmarked animals observed during the study ($u_\cdot$) a further factorisation of P_1' can be performed:

$$\begin{aligned} P_1'(\{u_i\}|N,\{\beta_i\},\{p_i\},\{\phi_i\}) & \\ &= P_{1a}'(\{u_\cdot\}|N)P_{1b}'(\{u_i\}|u_\cdot,\{\beta_i\},\{p_i\},\{\phi_i\}) \\ &= \left\{\frac{N!}{u_\cdot!(N-u_\cdot)!}\left[\prod_{i=1}^{K}\Psi_i p_i\right]^{u_\cdot}\left[1-\prod_{i=1}^{K}\Psi_i p_i\right]^{N-u_\cdot}\right\} \\ &\quad\times\left\{\frac{u_\cdot}{u_1!u_2!\ldots u_K!}\prod_{i=1}^{K}\left(\frac{\Psi_i p_i}{\sum_{i=1}^{K}\Psi_i p_i}\right)^{u_i}\right\} \end{aligned} \tag{2.4.4}$$

Of these two components the first, P_{1a}', models the split between the animals in the superpopulation that are caught at some point during the study and those that are not. The second component, P_{1b}', is restricted to the animals that are caught at least once and models the temporal distribution of their initial captures.

Combining the standard forms for P_2 and P_3 (see Eq. 2.3.10) with the new Schwarz-Arnason (Schwarz and Arnason, 1996) representation of P_1 produces the entire distribution function:

$$\begin{aligned} Pr\left(\{u_i\},\{d_i,d_i'\},\{m_{ij}\}\right) &= P'_{1a}(u.|N) \\ &\quad \times P'_{1b}(\{u_i\}|u.,\{\beta_i\},\{p_i\},\phi_i) \\ &\quad \times P_2(\{d_i,d_i'\}\,|\,\{m_i,u_i\}),\{\eta_i,\eta_i'\}) \\ &\quad \times P_3(\{m_{ij}\}\,|\,\{R_i\},\{\phi_i,p_i\}) \end{aligned} \tag{2.4.5}$$

The likelihood model formulations in Eq. (2.3.11) and Eq. (2.4.4)) are different in the parameterisations but they are both based on exactly the same observed capture-recapture data. They also have common components P_2 and P_3. Therefore, the alternative expressions in Eq. (2.3.11) and Eq. (2.4.4) are statistically equivalent.

2.4.2 Model Assumptions

The alternative expression for the model component P_1 required the specification of some new parameters, the entry probabilities β_i. These parameters model the probabilities that a member of the superpopulation is not present on the study area (i.e. available to be sampled) until sampling period i, but then enters the population and can be caught on sampling occasion $i+1$. The key assumption relating to these entry probabilities is that of homogeneity. It is assumed that all animals in the superpopulation that have yet to enter the population have an equal chance of belonging to the next cohort of recruits. For example, animals of different gender are not assumed to have different probabilities of making the transition to the population during the same period.

The use of the multinomial distribution implicitly requires the assumption that the fates of the animals are independent. By including the entry probabilities, the "fate" of an animal incorporates not only the usual capture and survival parameters but also the entry of an animal into the sampling population. The independence assumption should still hold unless there are strong reasons why the entry of one animal to the population may influence the entry of another, for example, multiple births per family.

2.4.3 Estimation

The estimation of the model parameters is performed chiefly through numerical maximisation of the likelihood equations (Schwarz and Arnason, 1996). The full likelihood can be expressed as the product of a sequence of multinomial and binomial terms (Schwarz and Arnason, 1996) and it can be demonstrated that the conditional MLEs derived in this way are asymptotically equivalent to the unconditional MLEs[5].

Schwarz and Arnason (1996) chose to estimate the model parameters using a logit link function as a means of ensuring that all estimates are kept within the parameter space. The estimates are then derived from the following components of the data: the estimates of survival and capture rates are based on the recaptures of previously marked animals, the estimates of new recruits to the sampling population are based on the proportions of unmarked animals that are captured on each occasion and the estimate of the superpopulation size is based on the total number of unmarked animals captured during the sampling study. Hence, the capture $\{\hat{p}_i\}$, survival $\left\{\hat{\phi}_i\right\}$ and entry $\left\{\hat{\beta}_i\right\}$ probability estimates can be obtained from numerical maximisation of the product $P'_{1b}P_2P_3$. These

[5]The conditional likelihood estimates will differ from the MLEs if constraints are applied to the model parameters.

estimates $\hat{p}_i$, $\hat{\phi}_i$ and $\hat{\beta}_i$ can then be used with P'_{1a} to estimate the superpopulation size N.

$$\hat{N} = \frac{u_.}{\sum_{i=1}^{K} \Psi_i \hat{p}_i} \tag{2.4.6}$$

As with both the Cormack-Jolly-Seber (Section 2.2.1.2) and the Jolly-Seber (Section 2.3.1.3) models not all of the parameters are identifiable. The capture probability on the first sample, p_1, cannot be estimated and $\phi_{K-1}p_K$ can only be estimated as a product. Along with the constraint that $\sum_{i=0}^{K-1} \beta_i = 1$ there are $3K - 3$ parameters that can be estimated under the full time-specific model (Schwarz and Arnason, 1996). The lack of identifiable estimates for p_1 and p_K makes estimation of the superpopulation total $\hat{N}$ difficult for the fully time-specific model. Imposing the constraints $p_1 = p_K = 1$ is recommended by Schwarz and Arnason (1996). Estimates of time-specific abundance and recruitment are also given and it is demonstrated that the usual estimates (Crosbie and Manly, 1985; Pollock *et al.*, 1990) are obtained.

The variances are derived asymptotically and those obtained for ϕ_i and p_i are the same as in Pollock *et al.* (1990). The recruits B_i are assumed to be random variables which adds a component of variance to the asymptotic variances of $\hat{B}_i$ and $\hat{N}_i$. If there are no constraints on the β_i, or only simple constraints such as $\beta_i =$ constant then this extra source of variance is removed. However, for more general constraints Schwarz and Arnason (1996) note that there is no general method for deriving the variance.

2.4.4 Other Models

As with the Jolly-Seber models the alternative model proposed by Schwarz and Arnason (1996) can be simplified into a reduced parameter form by the use of constraints on the parameters. Setting all the survival parameters equal to 1 will obtain the birth-only model.

Setting $\beta_1 = 1$ and all other entry probabilities to zero will give the death-only model. Both of these simplified models yield the same estimators as given in Pollock *et al.* (1990).

Constraining the parameters can also be used to investigate temporal effects on capture, survival and recruitment rates. However, Schwarz and Arnason (1996) note that parameters in models with temporal variation in capture probabilities can experience identifiability problems and advocate using constant capture probabilities where possible.

Extending the model to incorporate covariates is also possible. Discrete, static covariates can be modelled by expanding the model to multiple groups and discretised dynamic covariates can be modelled using a multistate modelling approach. Alternatively a GLM-based approach (Lebreton *et al.*, 1992) can be used to model the parameters as a function of multiple covariates.

2.5 Bayesian Methods for Capture-Recapture

2.5.1 Introduction

Capture-recapture methods have undergone considerable development since what are now seen as the archetypes of capture-recapture modelling were first formally represented (Cormack, 1964; Jolly, 1965; Seber, 1965). With recent extensions of these models often requiring ever-more complicated likelihood formulations (e.g. extensions to hierarchical models (Link and Barker, 2004; Clark *et al.*, 2005)) the numerical evaluation of the models using classical likelihood methods can become prohibitively complex (Link and Barker, 2005). One alternative approach to investigating these models is to embed the inference within a Bayesian framework.

Pollock (1991) reviewed the use of capture-recapture in the estimation of demographic parameters and noted that the structure of Bayesian analysis of time series data, in which some statement of prior knowledge is updated sequentially as new information enters the system, is well suited to the analysis of capture-recapture data. The studies typically take place over a series of sampling occasions thus making it appropriate to update the model upon receipt of the new observations.

Another potential advantage of analysing capture-recapture data using Bayesian models is the issue of parsimony. Link and Barker (2005) note that for even relatively simple models, such as the fully time-specific Jolly-Seber model, for any sampling study conducted over a significant number of occasions, the number of parameters to be estimated can soon become prohibitive. Even the most parsimonious models, as determined by careful model selection, may still consist of a large number of parameters precluding succinct descriptions of the data. Depending on the purpose of the study many of the parameters may even be considered to be nuisance parameters (Poole, 2002). These can be complicated to handle in the classical modelling context and the use of Monte Carlo Bayesian methods can provide a more simple, yet still formal, approach to analysing the models.

This section will review the use of Bayesian techniques in capture-recapture analyses focusing on important recent developments.

2.5.2 Closed Populations

Although closed population experiments are outwith the scope of this thesis, there has been some notable work performed on investigating behavioural response of animals using Bayesian techniques to model capture-recapture data. Yang and Chao (2005) define models to analyse both permanent and temporary effects on capture probabilities and fit

these models using a bivariate Markov chain approach. In the case of permanent effects an animal's capture probability changes permanently after first capture and thus capture probability only differs across marking status. For short term effects it is assumed that only the most recent capture occasion affects the current capture probability. Pollock (1991) and Borchers *et al.* (2002) cover the traditional closed-population behavioural response model, M_b, in which an animal's capture probability is determined by its marked status. There exist two probabilities: one for animals that have been marked and one for animals that have yet to be captured. Yang and Chao (2005) create a simple bivariate state-space model (see Chapter 3) in which states are determined by both marking and current capture status. They define three states:

$$\begin{aligned} \text{a} &= \{0,0\} \text{ The animal is uncaught and unmarked.} \\ \text{b} &= \{0,1\} \text{ The animal is uncaught and is marked.} \\ \text{c} &= \{1,1\} \text{ The animal is caught and marked.} \end{aligned}$$

and the associated probabilities for plausible transitions:

$$\begin{aligned} P_{ac} &= \text{The probability an animal unmarked in period } i \text{ is caught in } i+1. \\ P_{bc} &= \text{The probability a marked animal uncaught in period } i \text{ is caught in } i+1 \\ P_{cc} &= \text{The probability a marked animal caught in period } i \text{ is caught again in } i+1. \end{aligned}$$

Then the difference $P_{bc} - P_{ac}$ measures the permanent trap response effect. The difference $P_{cc} - P_{bc}$ measures the temporary trap response as any difference is determined by the most recent capture history. If $P_{cc} = P_{bc}$ then the model reduces to the classical behavioural-response model M_b. A conditional likelihood based on only the observable capture histories is chosen over the full likelihood for four principle reasons. Firstly, the full likelihood considers uncaught animals but because these animals are never captured no covariate values can be observed for them. This makes extending the usual likelihood method to include covariates difficult, however missing values can be dealt with within an MCMC process (Link and Barker, 2005; King *et al.*, 2006; Dupuis and Schwarz, 2007).

By making the likelihood conditional on observable animals this problem is avoided. Secondly, for large population sizes the computationally more simple conditional MLEs will be asymptotically similar to the unconditional MLEs, although Yang and Chao (2005) don't specify how large the population needs to be for adequate closeness. Thirdly the conditional MLE is equivalent to a Horvitz-Thompson type estimator as discussed in Borchers *et al.* (2002). Fourthly, the conditional MLE is invariant to rescaling the data, whereas the unconditional MLE is not (Chao *et al.*, 2000). For example if data $\{u_1, u_2, u_3\}$ are analysed and yield an estimate $\hat{N}$ then if the observations are scaled by a factor (say 10, to give $\{10 * u_1, 10 * u_2, 10 * u_3\}$) then the corresponding estimate $\hat{N}$ will not be scaled by that same factor.

Yang and Chao (2005) discuss extensions to their behavioural model and note that the classic M_{tb} model is unidentifiable as is the temporal effects extension to their bivariate permanent behavioural effects model. However, an extended model with temporal effects and temporary behavioural effect can be identified under the definition of appropriate constraints. They also discuss the incorporation of covariates to establish a bivariate model with temporal, behavioural and heterogeneity effects analogous to the classical M_{tbh} model. The restrictive assumption that the covariates are static is made to avoid missing-covariate problems on the occasions the animal is not captured. The transition probabilities can then be considered in the GLM-framework described in Lebreton *et al.* (1992) and logistic models are used to relate the probabilities to their explanatory covariates. The results of their investigations suggest that the classical models perform poorly when analysing data generated from a population exhibiting a temporary behavioural effect. They suggest their Markov models allow new interpretations of the behaviour of animals during the study and provide more satisfactory population estimates when there are only temporary behavioural effects.

2.5.3 Survival Analysis

Poole (2002) provides a description of a Bayesian approach to modelling survival in a capture-recapture study. The models are fitted using MCMC methods to obtain samples from the joint posterior distribution of the model parameters. The techniques described are then extended to account for missing data which can arise when sampling occasions were missed during the experiment. The proposed extension is deemed to be simpler to apply for CJS models than other standard approaches to dealing with missing data such as the EM Algorithm (Dempster *et al.*, 1977) or MCMC data augmentation (Gilks and Roberts, 1996).

Poole (2002) focuses on the CJS model, restricting inference to the survival and capture parameters so the data consists only of animals that have been captured at least once. Therefore, the CJS model likelihood as described in Cormack (1964) is used as the basis for Monte Carlo fitting via a Markov chain. The Markov chain is constructed so that it produces an approximation to the joint posterior of the capture/sighting probabilities p_i and the survival probabilities ϕ_i. For the full data set, in which the mark-recapture data is available for all years, the target posterior distribution is simply the product of the CJS likelihood and the prior on the survival and capture parameters. An algorithm for component-wise Metropolis-Hastings updates is then described. For the missing-data case Poole (2002) notes that the usual solution is simply to regard the missing data as extra variables that need to be estimated. Although effective in many situations, Poole (2002) argues that it cannot be applied here due to high interdependence in the data. For example, under the CJS model every animal included in the model must have been captured at least once and this imposes the constraint that $\sum b_j = \sum c_j$ since both sum

to the number of captured animals where c_j is the number of marked animals seen for the last time at time j and b_j is the number of animals marked for the first time at time j. Simply proposing values for any missing capture histories will change the values of the CJS likelihood for the other time periods. Poole (2002) suggests dealing with a missing capture occasion by modelling the survival from the preceding sampling occasion as a rate that extends over the missing occasion. The prior can also be altered in accordance with this re-paramaterisation.

From the example analysis detailed in Poole (2002), the posterior Bayesian estimates of the survival parameters are in strong agreement with analytic CJS estimates. For simplicity the chosen priors on all survival and capture parameters were uniform but Poole (2002) acknowledges that more complex prior specifications are possible. Specifically, if correlations among survival probabilities are known to exist the results are likely to be improved by explicitly modelling these correlations as part of the Bayesian analysis, although Poole (2002) states that the feasibility of this approach is unknown. However, Link and Barker (2005) model demographic parameters by specifying a bivariate normal prior on the joint distribution of survival and fecundity rates and successfully fit their resulting model to capture-recapture data.

2.5.4 Extending the Model

An extension to the Jolly-Seber model developed by Crosbie and Manly (1985) and Schwarz and Arnason (1996) was presented in Section 2.4 in which the birth process was explicitly incorporated into the likelihood formulation. A similar extension is investigated by Link and Barker (2005) who develop a hierarchical extension to the CJS model to model relationships amongst demographic parameters. They develop an analysis of

capture-recapture data that allows inference to be made about the relationships between survival and birth processes rather than focussing on survival or abundance alone. The Schwarz-Arnason model (Schwarz and Arnason, 1996) is analysed and an alternative parameterisation is presented. They introduce the parameter f_i which is defined as a birth rate parameter and is given by $f_i = \frac{\beta_i}{d_i}$ where

$$d_{i+1} = d_i\phi_i + \beta_i \tag{2.5.1}$$

for $i = 1, 2, \ldots, K-1$ and $d_1 = \beta_0$. The f_i are then used to replace the entry probabilities β_i in the Schwarz-Arnason model. The d_i can be thought of as approximating the proportion of the superpopulation N that constitute the population in the i^{th} period N_i under the assumption of no losses on capture. Link and Barker (2005) note that $f_i \approx \frac{B_i}{N_i}$ represents an index to per capita birth rates as opposed to the index of total births represented by the Schwarz-Arnason parameter $\beta_i \approx \frac{B_i}{N}$ where both these approximations are based on assuming no losses on capture. They also demonstrate that all of the information on the parameters β_0 and N is contained in the Schwarz-Arnason likelihood component $P'_{1a}(\{u.\}\,|N)$ and note that, without making some untestable assumptions, the capture histories contain no information about β_0 and N. Consequently, they determine the statistic $u.$ to be approximately ancillary to the parameters in the model that require estimation. Hence, their extensions to the CJS model are based on the P'_{1b} (see Eq. (2.4.4)) likelihood component of the Schwarz-Arnason model (Schwarz and Arnason, 1996).

The formulation of their likelihood is analogous to that of the Schwarz-Arnason extension (see Eq. 2.4.5) to the three component product for the Jolly-Seber model (see Eq. 2.3.10). Their reparameterisation is chosen so that the model is expressed entirely in terms of identifiable parameters.

They define

$$\lambda_i = \phi_i p_{i+1} + \phi_i(1 - p_{i+1})\lambda_{i+1}, \quad i = 1, 2, \ldots, K-1 \tag{2.5.2}$$

where $\lambda_{K-1} = \phi_{K-1}p_K$. Thus, λ_i represents the probability of an animal remaining in the population and being recaptured after sample i conditional on it being released after capture in sample i. Another new parameterisation is established with the definition

$$\tau_i = \frac{\phi_{i-1}p_i}{\lambda_{i-1}}, \quad i = 2, 3, \ldots, K-1 \tag{2.5.3}$$

which represents the probability of being recaptured in sample i relative to the probability of being recaptured after sample $i-1$. These two definitions are used in an alternative parameterisation of the CJS component of the likelihood formulation. Link and Barker (2005) decompose this component (which is equivalent to P_3 - see Eq. (2.3.14)) into conditionally independent binomial distributions. The first component

$$L_{2a} = \prod_{i=1}^{K-1} \binom{R_i}{r_i} (\lambda_i)^{r_i} (1 - \lambda_i)^{R_i - r_i} \tag{2.5.4}$$

then models the future recaptures of each cohort of animals that are marked and released following the i^{th} sampling occasion and takes the product of these terms. The second component

$$L_{2b} = \prod_{i=2}^{K-1} \binom{T_i}{m_i} (\tau_i)^{m_i} (1 - \tau_i)^{T_i - m_i} \tag{2.5.5}$$

includes the new term T_i which is formally defined as the number of animals marked and released prior to sampling period i that are then recaptured at some sampling occasion j, $i \leq j \leq K$. It is defined recursively with $T_2 = r_1$, and $T_{i+1} = T_i - m_i + r_i$, for $i = 2, 3, \ldots, K-2$. This allows for the less formal definition that T_i is the number of animals

released before sampling period i that are known to be recaptured on some future sampling occasion but have not yet been recaptured (hence the subtraction of the m_i terms). Therefore, the component P'_{3b} (Eq. (2.5.5)) models the split between animals that are recaptured in the i^{th} sample and those that are recaptured after the i^{th} sample and again takes the product of these terms.

Then the likelihood, as a function of all model parameters $\boldsymbol{\theta}$, specified by (Link and Barker, 2005) is expressed as the product of three components:

$$L(\theta) \propto L_1(\{\nu_i\})L_2(\{p_i\},\{\phi_i\})L_3(\{p_i\},\{\phi_i\},\{f_i\}), \tag{2.5.6}$$

where the first component L_1 contains the ν_i parameters which are the release probabilities for animals following capture at the i^{th} sample. It is analogous to the losses on capture component P_2 (Eq. (2.3.12)) of the Jolly-Seber model, although the losses on capture are not modelled separately for marked and unmarked animals in the Link and Barker (2005) formulation. As noted previously, the second component $L_2 = L_{2a} \times L_{2b}$ represents the CJS model and is equivalent to P_3 (Eq. (2.3.14)) in the Jolly-Seber model. The third component L_3 is equivalent to P'_{1b} (see Eq. (2.4.4)) and models the temporal distribution of the animals' initial captures. The likelihoods are then expressed solely in terms of identifiable parameters and MLEs can be obtained in closed form. Link and Barker (2005) note that the MLEs for the parameters ϕ_i, p_i and ν_i are the classical Jolly-Seber solutions (Jolly, 1965; Seber, 1965; Pollock *et al.*, 1990). Details of the MLE form of the other parameters are provided by Link and Barker (2005).

Link and Barker (2005) also specify a bivariate distribution for transformations of the fecundity and survival parameters f_i and ϕ_i respectively. They choose a bivariate

normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$ and specify a log transform for the f_i and a logit transform for the ϕ_i. Therefore, their full likelihood consists of the statistics $\mathbf{D} = [\{u_i\}, \{m_i\}, \{R_i\}, \{r_i\}]$ and the parameters $\boldsymbol{\theta} = [\{\phi_i\}, \{p_i\}, \{\nu_i\}, \{f_i\}, \{\mu_i\}, \{\Sigma_i\}]$. The full model expression also requires the specification of a prior due to the use of Bayesian model fitting methods. Link and Barker (2005) briefly detail the assumptions they make and the choice of prior this results in.

As discussed in his review paper (Schwarz, 2001) notes the lack, at the time, of investigations into the utility of the JS model in estimating population abundance. As noted by Link and Barker (2005), one possible reason for this, as suggested by Schwarz (2001), is the lack of appropriate parameterisation of the model to obtain direct inference about birth rates. Specifically, Schwarz (2001) notes the lack of comparable procedures for the birth parameters when compared to those developed by Lebreton *et al.* (1992) for survival and capture parameters. By focussing on modelling the interactions of demographic parameters, Link and Barker (2005) chose to express the likelihood in terms of Eq. (2.5.6) which enables them to express the model in terms of parameters relating directly to fecundity f_i instead of the less biologically relevant β_i used by Schwarz and Arnason (1996). Link and Barker (2005) also claim that the factorisation of the likelihood simplifies the incorporation of hierarchical modelling and is necessary for the generation of suitable candidates as part of the MCMC fitting-algorithm. If losses on capture do occur then they can be modelled using the L_1 component (see Eq. (2.5.6)) and, although f_i can still be considered a birth parameter, its interpretation is modified to being the number of births per animals that would have existed in the population during period i if there had been no losses on capture.

2.5.5 Bayesian Multistate Models

A further extension to the general multistate Jolly-Seber capture-recapture model (Schwarz and Arnason, 1996) is developed by Dupuis and Schwarz (2007). To address potential problems with estimator bias incurred by heterogeneity in the capture and survival rates methodology has been developed in which animals are classified into states. This classification can be based on static variables (e.g. gender) (Schwarz and Arnason, 1996) or on stochastic variables (e.g. movement between geographical regions) (Schwarz *et al.*, 1993). When there is movement between states these models are unable to produce estimates of abundance. A Bayesian treatment for the movement case was presented by King and Brooks (2002b) although they did not address the issue of abundance. The superpopulation approach described in Schwarz and Arnason (1996) is extended to model the appearance of new entrants into the population between sampling occasions. The missing value problem is given an alternative approach owing to the model structure requiring that both unrecorded states prior to an animal's first capture as well as the full set of unobserved states for animals that are never marked are incorporated into the modelling framework.

The observed data are decomposed into two interlinked vectors for the i^{th} animal: $\boldsymbol{x}_i$ simply records the standard capture history pattern of 1's and 0's and $\boldsymbol{z}_i$ records the covariate value observed on each occasion corresponding to a 1 (the animal is observed) in $\boldsymbol{x}_i$. No covariate is recorded on occasions corresponding to 0's in $\boldsymbol{x}_i$. Independence is assumed with respect to the processes of capture, survival and movement and movement is assumed to be a first-order Markov process. The model probabilities are now defined to be a function of state as well as time which would increase the complexity of the likelihood-based models defined in Schwarz and Arnason (1996). Dupuis and Schwarz (2007) define $z_{i,t}$ as the state of animal i at time t such that if $z_{i,t} = r$ then animal i

is present in the sampling population and is alive in state r at time t. Also, if $z_{i,t} = *$, animal i has not yet entered the sampling population at time t. Using these variables, the number of animals from the superpopulation that enter into the sampling population in state r between times t and $t+1$ is given by: $E_t(r) = \sum_{i=1}^{N} I_{z_{i,t}=*, z_{i,t+1}=r}$, where $I_{(c)}$ is an indicator function taking the value 1 if condition c is true and 0 otherwise. Similarly, the number of animals in the population in state r at time t is given by $N_t(r) = \sum_{i=1}^{N} I_{z_{i,t+1}=r}$. Hence, the number of animals per state and the number of new entrants are not parameters in their own right but simply functions of other variables. The priors are assumed to be independent of each other with specific emphasis placed on the independence of the priors on the survival parameters and the superpopulation.

Dupuis and Schwarz (2007) introduce a blocking design to simulate missing covariate data. The blocks are determined by capture history with blocks of Type I corresponding to the occasions before first capture, Type II corresponding to occasions in which the animal was not captured that occur between occasions of the animal's first and last observed capture, Type III corresponding to occasions after the last observed capture until the end of the study and Type IV corresponding to the scenario where an animal is not seen on any of the sampling occasions. Thus, each of these blocks corresponds to a set of missing values and the distribution of these blocks, conditional on the observed capture histories can be derived (Dupuis and Schwarz, 2007). The joint distribution of the missing covariate blocks for an animal that has been captured at least once is shown to be the product of the distribution of the missing covariates in the individual blocks. The missing data are simulated conditionally on the superpopulation N and a set of sufficient statistics summarising all the marks and captures attributed to animals across all states are simulated using multinomial distributions. They note that their blocking approach to data-augmentation produces an ergodic Markov chain even when some transitions between

states are impossible. In comparison, the component-by-component data-augmentation approach (King and Brooks, 2002b) can be non-ergodic.

The posterior distribution for the superpopulation is simulated and, for the two alterntative priors used, can be shown to be negative binomial with parameters determined by the number of animals marked during the study and derived probability λ of any animal being marked during the population. The value of λ is simply the probability of an animal moving into a particular state (from another state or entering from the superpopulation) and surviving summed across all states and all time periods. For posteriors of the other parameters the complete data likelihood is derived and decomposed into three separate components: one for the survival and state transition parameters, one for the entry parameters and one for the capture parameters.

The imposition of constraints on parameters and the identifiability of the parameters is also investigated. A reparameterisation is required on the entry probabilities to allow suitable constraints to be imposed. Dupuis and Schwarz (2007) also note that there exists a correspondence between the unidentifiable parameters in the classical Jolly-Seber model (see Sections 2.2.1.2 and 2.3.1.3) and in their multistate extension. The full time-specific model has confounded initial immigration, state distribution and capture parameters which cannot be estimated separately without imposing identifiability constraints on the parameters.

The approach developed by Dupuis and Schwarz (2007) allows the combination of a super-population approach with a Bayesian analysis. The approach extends the existing methods of analysing multistate capture-recapture studies and proposes a more computationally efficient algorithm for dealing with missing covariate information within a MCMC

framework. There is no attempt at model selection but Dupuis and Schwarz (2007) note that one approach to MCMC model selection is covered in detail in King and Brooks (2002b) in which RJMCMC is used to move between a potentially vast array of alternative model specifications. Also, the model could incorporate a prior covariance structure to investigate interactions between the model parameters.

2.5.6 Unified Approaches

2.5.6.1 A Flexible General Model

A general approach to constructing a factorised likelihood for a fully time-specific multistate capture-recapture model is developed by Barker and White (2004). The model formulation allows for integrated analyses involving recoveries of dead animals and resightings of marked animals in addition to the usual recapture data. The model contains recruitment, survival, movement, capture, recovery and resighting probabilties, all of which can be modelled as time-specific or can be restricted for special reduced-parameter models. For the classical approaches it is often the case that the parameterisations that are convenient for computing estimates from a likelihood are not necessarily the most appropriate for exploring biological relationships. The factorisation approach decomposes the likelihood so that different components correspond to different information in the data allowing the likelihood to be formulated dependent on the inferential aims of the investigator and the information available in the data. Barker and White (2004) refer to their general approach as the "mother-of-all-models" (hereafter MOAM) approach. They detail the range of mark-recapture modelling techniques available in the powerful software package MARK (White and Burnham, 1999; White *et al.*, 2001) and note that the correct choice of model is determined by the available data and the parameters of

interest. Choosing the correct model is a non-trivial exercise and the choice of parameterisation can influence the manner in which constraints can be imposed on the model. MARK incorporates constraints for capture-recapture models in a GLM-framework (Lebreton *et al.*, 1992) and requires constraints to be able to be expressed linearly so that they can be incorporated into the design matrix. Alternative model parameterisation can lead to non-linear constraints which then cannot be applied within this framework. They recommend developing automated modelling routines that satisfy four main requirements: firstly, they should incorporate integrated analyses of data from multiples sources, secondly, they should incorporate non-linear constraints and allow estimates of functions of parameters, thirdly, they should be able to overlook nuisance parameters and finally, they should enable hierarchical extensions to be incorporated. Hierarchical models (Clark, 2003; Link and Barker, 2004; Clark *et al.*, 2005; Link and Barker, 2005) allow the relationship between parameters and their explanatory covariates to be modelled to investigate questions of biological interest.

Barker and White (2004) consider the Schwarz-Arnason (Schwarz and Arnason, 1996) refinement of the work by Crosbie and Manly (1985) and their extension to the JS likelihood. By expressing the likelihood as the product of a standard CJS component for modelling the observable capture histories and a component for modelling the temporal distribution of first captures, Barker and White (2004) note that incorporating the second component provides a way of including birth and population growth models in many of the open population models in MARK at the time. They describe a very general model consisting of the core CJS component with multistate factors incorporated into the model. This flexible form can be applied to most of the models in MARK. The multistate models can present some difficulties (Lebreton and Pradel, 2002) in terms of numerical optimisation as the complexity of their surface can result in multiple local maxima, especially

when constraints are imposed. The complexity and multi-dimensionality of the surface also makes graphical exploration of the likelihood prohibitively difficult. The approach of sequentially optimising the likelihood function (Schwarz and Arnason, 1996) over capture occasions could help to highlight problem areas of the likelihood surface allowing for more targeted exploration. Nuisance parameters are often required in the full likelihood of capture-recapture models even when inference may be focused on a small subset of all model parameters. Barker and White (2004) suggest model selection to reduce the number of nuisance parameters could be an option although they note that choosing one from a potentially vast array of plausible models, all with different restrictions on the parameters, can obscure biological inference. Model averaging (Buckland *et al.*, 1997) would seem more appropriate depending on the focus of the inference. Specifying a general model with few constraints and comparing it to reduced models with constraints on the parameters corresponding to the biological hypothesis of interest may be a more efficient exploration of *a priori* desirable regions of model space. However, this approach may result in biologically interesting models being missed if they lay in a region of parameter space not deemed to be worth exploring. A Bayesian approach to model selection that explores vast regions of model space relatively efficiently is referred to as RJMCMC (King and Brooks, 2002b). Other alternatives suggested by Barker and White (2004) include specifying a hierarchical model in which nuisance parameters can be modelled as random effects and using a conditional likelihood approach. The latter approach is only applicable when the sufficient statistics can be modelled with conditional distributions that are only functions of the parameters of interest.

Barker and White (2004) suggest that hierarchical models are becoming increasingly important but comment that the formulation of a hierarchical likelihood for recent extensions to Jolly-Seber models can result in extremely complex models whose solution

requires multi-dimensional numerical integration techniques. An alternative is to use Bayesian methods in which Monte Carlo simulation can be used to approximate the integration process and obtain a posterior distribution. Summaries of this posterior can then provide the experimenter with the distributions of the quantities of interest. With the specification of a uniform prior the resulting posterior distribution is the scaled likelihood and the mode of this is equivalent to the MLE. The use of hierarchical models sits naturally within the Bayesian paradigm (Link and Barker, 2004) and the specification of prior distributions on the model parameters can be considered analogous to specifying a hierarchical model in the classical JS model context.

2.5.6.2 A Unified Model

As something of a companion piece to the work of Barker and White (2004) in developing a flexible general likelihood formulation, Schofield and Barker (2008) develop an approach to specifying a unified capture-recapture model using a hierarchical framework. As with Link and Barker (2004, 2005) a flexible hierarchical parameterisation is advocated as a means to specify relationships between parameters of interest and external covariates to investigate biological hypotheses. For example, it may be assumed that the vital rates in the population are a function of the abundance or density of the animals being studied. Capture-recapture models such as the Jolly-Seber model and its extensions can be used to obtain abundance estimates and then these estimates can be incorporated into a model for density dependence. However, as discussed in the review paper of Seber and Schwarz (2002), classical methods for the analysis of capture-recapture data have possessed limited tools with which to estimate density dependence. Although the flexible hierarchical likelihood developed by Link and Barker (2005) allows for some demographic relationships to be specified their models do not allow the relationships to be a function of abundance, thus limiting their utility for investigating density dependence. Schofield and Barker (2008)

note that it may be the case that some constraints cannot be expressed as "deterministic functions of parameters that are explicitly expressed in the likelihood", especially if the biological parameters of interest have been transformed to make the likelihood evaluation more tractable.

The model developed by Schofield and Barker (2008) can incorporate missing information and is fitted using Bayesian multiple imputation which allows demographic parameters to be explicitly specified in the model. The model formulation considers a superpopulation N and explicitly defines matrices for the capture histories, times of birth and times of death for all animals in the superpopulation. Simple biological constraints are imposed on the births and deaths to ensure that each animal can only be born and die once and that the birth event must occur prior to the death event. By specifying the birth and death matrices the demographic processes of biological interest can be modelled directly and can be summarised as a function of the elements of the matrices. The capture process is modelled using a binomial distribution analogous to P'_{1a} in the Schwarz-Arnason expanded form of the Jolly-Seber likelihood decomposition (see Eq. (2.4.4)). Deaths are modelled individually as a series of Bernoulli trials conditional on the animal being alive at the beginning of the sample. Births are modelled using a per capita birth rate analogous to the fecundity parameter in Pradel (1996) and the birth rate used by Link and Barker (2005). This gives a more biologically natural birth rate than the Schwarz and Arnason (1996) parameterisation (see (2.4.2)).

From the form specified for their model Schofield and Barker (2008) use a Bayesian approach to obtain the joint posterior distribution on the unknown parameters (capture, survival, per capita birth rates and superpopulation total) and missing data (unobserved times of birth and death). This posterior is proportional to the full likelihood which

can be expressed as the product of the conditional distributions for the capture, survival and birth processes. If the assumption is made that the demographic and capture rates are fully time-specific and modelled as fixed-effects a careful choice of prior distribution (Beta's in this case) and appropriate reparameterisation[6] yields full conditional distributions of known functional form thus allowing a more efficient MCMC fitting algorithm.

Partially unobserved covariates are given the usual Bayesian treatment and are updated with the parameters during each fitting iteration. The models developed by Schofield and Barker (2008) incorporate individual-specific time-varying categorical covariates (e.g. breeding status) and can accommodate both the typical Markovian models as well as Memory models (see Section 2.2.3 and Brownie *et al.* (1993)). Movement models can also be included and are defined as modelling an animal's availability for capture or movement into and out of the sampling population. This can be modelled explicitly using Markovian transitions between states of availability or can be subsumed into a "no movement" model under two scenarios. Firstly, permanent migration occurs and times of birth and immigration are combined to give recruits and times of death and emigration combine to give removals. Hence survival probabilities are now removal rates and birth rates become recruitment rates. Secondly, random migration occurs and the probability of capture now combines capture and availability for capture. Continuous stochastic individual-level covariates require a continuous model to simulate the unobserved covariates (see Section 2.2.5 and Bonner and Schwarz (2006)). The explicit data structure in which times of birth and death are recorded allows the model parameters to be functions of the random variables defined using the birth and death times. Population vital rates can then be related to functions of the numbers of births and deaths in the previous time period, which is effectively density-dependence.

[6] of the birth rates/entry probabilities

The model framework developed by Schofield and Barker (2008) is extremely flexible and covers standard capture-recapture models as well as missing data models. The composition of the model as the product of conditional distributions can be extended to accommodate more complex models. The standard core of the model remains the same with additional components representing the extension (e.g. density dependence) added to the model. This structure is the component-based MOAM framework (Barker and White, 2004).

Chapter 3

Modelling the dynamics of wild animal populations

The previous chapter gave a detailed review of the development of research into the construction and analysis of models for capture-recapture data. However, the formulation of many of these modelling approaches focused on the form taken by the likelihood and were designed with a view to tractability and computational expediency rather than a full attempt to create biologically realistic models that capture adequately the underlying physical dynamics of the population of study.

This chapter reviews existing developments in the formulation of models that attempt to embed population dynamics into their structure. A particular modelling framework which constitutes a flexible component-based approach to formulating models that can be fit to discrete time-series data is then explained in detail.

We begin this section by discussing the traditional approaches to fitting models to the dynamics of animal populations.

3.1 Overview

Under these traditional approaches existing knowledge about the animal population of interest would be used to construct models in an attempt to capture adequately the behaviour of those physical processes believed to dominate the dynamics of the population. The specification of the model that is assumed to approximate the system of interest can be summarised as consisting of three features (Krzanowski, 1998).

Firstly, a statistical modeller wishes to investigate the behaviour of a particular system and is able to observe and measure output produced by this system. Having collected a sample of data from the target population of interest the statistical modeller is then usually interested in testing some personal idea, or hypothesis, about the dominant behaviour of the system. The modeller is required to specify the mathematical form of the process which they believe best describes the dominant behaviour and which is consistent with the collected data. This mathematical form is often referred to as the systematic or deterministic component of the model.

Secondly, the modeller also needs to consider the random fluctuations that can occur in the system. Conducting the same experiment repeatedly and measuring the outputs from the system will very rarely produce the same output on each occasion. The modeller can observe the level of variation in the collected outputs and can then incorporate their beliefs about the nature of this stochastic component into the model. Thirdly, the modeller defines the relationship between the deterministic and stochastic components.

When constructing these types of models it is necessary for the modeller to establish the relationships and dependencies between the variables of interest. It is also vital that

the modeller has a good understanding of the inputs to the model, that they understand the nature of the transformation implemented on these inputs by the model and that they can identify accurately the outputs from these transformations.

These traditional approaches based on mathematical models typically offer no formal framework for quantifying the level of uncertainty associated with the model parameters or the conclusions obtained from the analyses. In the majority of the traditional approaches there are two main sources of uncertainty that are often inadequately accounted for (Thomas *et al.*, 2005). Firstly, there is uncertainty surrounding the specification of input parameters. The input parameters are either obtained from previous studies of survey data deemed to have suitably close parallels with the current situation of interest or from expert opinion when no such suitable studies exist. Secondly, the choice of form for the underlying process believed to adequately capture the dominant characteristics of the true population dynamics can be quite subjective. The choice of model is typically motivated by the desire to find the simplest model that is able to explain the greatest proportion of the recorded observations. The modeller will also usually wish to derive a model which produces outputs that are relatively insensitive to small changes in the model inputs. Another desirable characteristic of any candidate model is that it should be easy to use; the form of the model should, ideally, be easily interpretable and the implementation of the calculations required by the model should be reasonably straightforward. Thomas *et al.* (2005) sum up this approach succinctly by saying that the process of formulating a model "represents an attempt to construct a parsimonious, robust and tractable characterization of the system under study".

In traditional approaches based on mathematical modelling, the modeller will typically specify a population dynamics model. This model can be either deterministic or

stochastic. The performance of these models is then assessed either through analytical or simulation-based methods. On the basis of this form of assessment the parameters in the model can then be adjusted, or "tuned" until they provide a suitable match to the observed data. The quality of this matching is often determined in a relatively informal way. The matrix models of (Caswell, 2001) can be classified under this type of approach. Uncertainty in parameters has been investigated by simulating ranges of plausible values for each of the model parameters, projecting forward according to the specified model and measuring the uncertainty in the resulting output. Assessing the extent to which the outputs are affected by the alternative choice of parameter values is often referred to as a "Sensitivity Analysis". The "ad hoc" nature of this process arises from the often subjective choice of alternative parameter values used in the sensitivity analysis as well as the criteria used to interpret the significance of the results. Also, there is no formal relationship that specifies the relationship between the model and the observed data under this approach. The significant advance in this area was the development of a unified framework that would account for the uncertainty in the structure and parameterisation of the model as part of the model fitting (Thomas *et al.*, 2005).

3.2 Recent Developments

These traditional approaches to incorporating stochastic variation into the model formulating process will often fail to adequately account for significant sources of uncertainty. Modelling approaches which incorporate both process error and measurement error have been developed for animal population dynamics. The fisheries industry is an area that experienced significant development in the formulation of new modelling techniques (Schnute, 1994; Newman, 1998; Meyer and Millar, 1999; Millar and Meyer, 2000a). Schnute (1994) introduces a general approach to developing sequential models for the

dynamics of fisheries. The traditional approaches to constructing fisheries models can be separated into two main classes; stock recruitment models (Millar and Meyer, 2000b) or catch-at-age models (Millar and Meyer, 2000a). The stock recruitment approach models the current level of fish stock recruitment as a function of the sizes of fish stocks at previous time periods. The catch-at-age models necessitate the derivation of a function that describes the linkage between the number of fish of a particular age category (say, A) caught at the current time period with the number of fish in the previous age category (A-1), caught in the previous time period.

Schnute (1994) describes three main approaches to formulating sequential dynamics models. The first approach incorporates error only from the observations; the parameters of the process model completely determine the dynamics of the fish stocks. For these types of models Schnute (1994) describes the linkage between theoretical and observed age structures and how the parameters of the deterministic process model can be adjusted to obtain an optimal fit to the data. The second approach allows the sequential population process equations to include error but regards the observations as being known exactly. Schnute (1994) raises the issue that, for models created under this paradigm, the output can be highly sensitive to the manner in which the process error is introduced to the modelling framework. The third approach describes techniques for simultaneously allowing both measurement and process error to be incorporated into the models; an example of this approach is described by Millar and Meyer (2000b).

The models Schnute (1994) develop are sequential in form whereby the characteristics of interest for the population being modelled, and their linkage with the recorded observations, are defined at a sequence of distinct time points. These models allow for the accommodation of non-stationary behaviour providing that the requisite process equations

describing the evolution of the dynamics can be specified for each distinct time period. To make the calculations described by Schnute (1994) tractable a number of simplifying assumptions are required. The distributions of the stochastic elements of process and observation are assumed to be normal (Gaussian) and the form taken by the process and observation equations is assumed to be linear. Models of this set are typically referred to as normal dynamic linear models and the necessary calculations for estimating the parameters can be performed using the Kalman filter (Besbeas *et al.*, 2002, 2003, 2005), with specific application to a fisheries model presented in Newman (1998).

Although the normal dynamic linear models can incorporate both process and observation error simultaneously the required assumptions of normality and linearity are fairly restrictive. By sacrificing reality for computational expediency these models can still often fail to capture adequately the true stochastic nature inherent in the underlying population dynamics. A relaxation of the assumption of linearity allows the use of the extended Kalman filter which uses linear approximations to fit non-linear process and observation equations to the observed data. When the stochastic elements of the model are Gaussian, the extended Kalman filter is optimal; however when the dynamic system is non-Gaussian the extended Kalman filter is only the best linear predictor (Carlin *et al.*, 1992). There exists the potential for the difference between the best linear predictor and the optimal forecast to be substantial.

When attempting to transform realistic population dynamics models into a framework of normal dynamic linear models the resulting transformed models often convey behaviour that is significantly altered and no longer realistic. Realistic population dynamics models that are non-linear and non-Gaussian would often be far less tractable than their normal and linear counterparts. Recent developments in computer intensive simulation methods

have allowed complicated statistical estimation procedures to be applied to these problems. Sequential Monte Carlo Methods (Doucet *et al.*, 2001; Liu, 2001) and Markov Chain Monte Carlo (Gilks *et al.*, 1996), hereafter MCMC, represent extremely powerful tools for fitting more flexible, and potentially more realistic, animal population dynamics models to observed data. These statistical estimation procedures have been applied to a variety of animal population dynamics models which incorporate non-Gaussian process and measurement errors and have non-linear form. Recent applications include models for red deer (*Cervus elaphus*) (Trenkel *et al.*, 2000), South Atlantic albacore (*Thunnus alalunga*) (Millar and Meyer, 2000b), and grey seals (*Halichoerus grypus*) (Thomas *et al.*, 2005).

In many of these recent applications the modelling approaches have shared certain common features. Although the complete specification of the models differed between the studies, there were several components of the models that were common to each of the projects. This motivated the development of a unified framework that allows the joint definition of both population dynamics models and the measurements taken on a population (Buckland *et al.*, 2004; Thomas *et al.*, 2005; Newman *et al.*, 2006).

In the following section a generalised approach to formulating a framework that allows stochastic population dynamics models to be embedded into statistical inference is described. Although the modelling approach can be applied to a wide variety of situations, the following section will describe the structure of these models only in the context of modelling animal population dynamics.

3.3 State-space models for animal population dynamics

3.3.1 Introduction

The general framework outlines a procedure for formulating models which possess useful features that have frequently not been included in published models. One of the main advantages of this approach is the flexibility it offers in specifying the underlying population dynamics model. The model structure allows for a number of processes to be combined into a single model and each process can be either deterministic or it can incorporate stochastic variation. The flexibility derives from the manner in which the model is specified. By identifying distinct sub-processes that are believed to drive the population dynamics, the models can be constructed in a modular way. Each sub-process is modelled separately but then linked together to define the complete population dynamics model. Conceptually this can be thought of as connecting a series of different building blocks together to construct the required overall structure.

A second feature of this generalised approach is the ability to allow for model uncertainty to be part of the inferential framework. Using this approach it is reasonably straightforward to examine the outputs from a series of plausible models and quantifying the uncertainty associated with selecting a suitable model is now couched in a formal statistical framework. From a practical viewpoint another advantage of the formulation of these types of models is that the risk associated with potential population management strategies can be quantified in a formal manner. The modelling framework incorporates three of the main sources of uncertainty (stochastic variation of the process model, measurement error and model uncertainty) and alternative management strategies can be investigated with the outputs reflecting this uncertainty.

The modelling approach can be considered as an extension of the matrix modelling framework described by Caswell (2001). The departures from that framework can be classified as three main aspects. Firstly, an observation model that can be either deterministic or stochastic is explicitly added to the model framework. The observed measurements are directly linked to the underlying states of the population via this observation model and therefore become fully integrated into the modelling framework. The manner in which the model is defined is the second main departure. Each process that is believed to influence the underlying dynamics of the population (e.g. birth, survival, movement, maturation, etc.) is modelled separately and can be either stochastic or deterministic. Each model of an individual process is often referred to as a sub-process model. Separating the processes enables each individual sub-process to be defined easily before they are linked together to form the overall population dynamics model of the required structure. This approach to constructing the models is far simpler to implement than attempting to specify fully complex dynamics models in a single stage. The third main departure is that each of the sub-process models can be represented in the form of a matrix. The matrix will represent the expected values of the sub-process and an associated random error term can be incorporated depending on whether the sub-process is stochastic or deterministic. However, it should be noted that the matrix representation will sometimes only be an approximation to the dynamics modelled by the subprocesses. When the expected values of properties of the population (e.g. abundances) at the current time are a non-linear function of the same properties at a previous time period, the matrix representation is only approximate. This approach is taken further in Buckland *et al.* (2007) where non-linear operators are considered for the subprocesses; these become matrices in the linear case.

An example of this non-linearity would be if a sub-process for survival was modelled

as being dependent on the density of the population. Although this matrix representation can be conceptually convenient as an approach to constructing complex models, it typically does not form part of the process of fitting the models to the observed data. The sub-process models can be represented as either probability density functions (pdfs) or probability mass functions (pmfs) for continuous or discrete sub-process models respectively and it is these probability functions that are used in the fitting process.

The product of the matrices representing the individual subprocesses is a generalized Leslie matrix (Caswell, 2001). In its simplest form, when all subprocesses are modelled without a stochastic element, the product is a deterministic population projection matrix (Buckland *et al.*, 2004; Thomas *et al.*, 2005) and the matrix based approach to model definition is illustrated with an example of this kind:

Consider a population that can be classified into three age classes. We then define a vector

$$\mathbf{n}_t = \begin{bmatrix} n_{0,t} \\ n_{1,t} \\ n_{2,t} \end{bmatrix}$$

where $n_{i,t}$ is the number of animals of age i in year t for $i = 0, 1$ and $n_{2,t}$ is the number of animals of age 2 or greater in year t.

Assume that there are three processes believed to govern the dynamics of this animal population. Firstly, there is a maturation process whereby animals that begin the year in one age class have matured to the next age class over the course of the year. The matrix

representation for this process is:

$$M = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

This models the ageing process of the population. The animals that are age 0 at the beginning of the year become age 1 at the end of the year. Animals that are age one at the beginning become age two or greater by the end of the year and animals already aged two or greater remain in that age class.

The composition of this population immediately after the maturation process can be represented in matrix notation. Post-multiplying the matrix representing the maturation process by the vector of states at time $t-1$ yields the following post-maturation vector:

$$\begin{bmatrix} 0 \\ n_{0,t-1} \\ n_{1,t-1} + n_{2,t-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} n_{0,t-1} \\ n_{1,t-1} \\ n_{2,t-1} \end{bmatrix}.$$

Thus, after maturation, all animals that were one year old at time $t-1$ ($n_{1,t-1}$) now move to element denoting those animals aged two or greater joining those already there ($n_{2,t-1}$). Also, after maturation there will be no animals of age zero in the population until the birth process has occurred hence the 0 in the first element of the post-maturation vector above. All the newborn animals in time $t-1$ ($n_{0,t-1}$) have now matured to become one-year-olds and thus occupy the second element of the post-maturation vector.

Secondly, there is a birth process where the parameter π is the mean number of young produced by each mature animal each year. Only animals one year old or more are capable

of breeding. This birth process is represented in matrix form as follows:

$$B = \begin{bmatrix} 0 & \pi & \pi \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The 1's on the diagonal of the matrix indicate that the numbers of animals aged one year old or more do not change. The first row of the matrix generates the number of new young produced by the animals capable of breeding.

Thirdly, there is a survival process where ϕ_i is the annual probability of survival for an animal belonging to age class i. This process is represented in matrix form as:

$$S = \begin{bmatrix} \phi_0 & 0 & 0 \\ 0 & \phi_1 & 0 \\ 0 & 0 & \phi_2 \end{bmatrix}$$

This matrix generates the numbers of animals in each age class that survive to the next year.

The dynamics model for this population can then be written as:

$$\begin{bmatrix} n_{0,t} \\ n_{1,t} \\ n_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_0 & 0 & 0 \\ 0 & \phi_1 & 0 \\ 0 & 0 & \phi_2 \end{bmatrix} \begin{bmatrix} 0 & \pi & \pi \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} n_{0,t-1} \\ n_{1,t-1} \\ n_{2,t-1} \end{bmatrix} = SBA\mathbf{n}_{t-1}$$

The population dynamics model can then be expressed as a product of the three matrices representing the individual sub-processes:

$$\begin{bmatrix} n_{0,t} \\ n_{1,t} \\ n_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_0\pi & \phi_0\pi & \phi_0\pi \\ \phi_1 & 0 & 0 \\ 0 & \phi_2 & \phi_2 \end{bmatrix} \begin{bmatrix} n_{0,t-1} \\ n_{1,t-1} \\ n_{2,t-1} \end{bmatrix} = L\mathbf{n}_{t-1}$$

The matrix L, which is the product of the matrices S,B and A, is an example of a generalised Leslie matrix (Buckland *et al.*, 2004). It should be noted that this particular expression differs from the standard Leslie matrix in that the standard representation the processes of adult survival and reproduction are convolved into a single fecundity process. Fecundity simply models the breeding adults that survive whereas the above formulation separates the survival and reproduction processes. Explicitly modelling each subprocess in this way increases the flexibility of the modelling approach.

This particular formulation of the model is deterministic as stochastic variation has not been incorporated in any of the subprocesses. The extension of this approach to encompass fully stochastic models is described in the next section.

3.3.2 State-space model structure

State-space models (Harvey, 1989) describe dynamic systems as consisting of two linked processes which are regarded as a pair of time series running in parallel. One of these processes is the state process which describes the true underlying state of the population at a sequence of successive time steps. A realisation of the state process at time t, here defined as $\mathbf{n}_t$, $t = 0, 1, 2, \ldots, T$, is a vector of states, some of which may be unobservable. The elements of these vectors correspond to the numbers of animals belonging to particular, mutually exclusive, categories. For example, elements of the state vector may represent the number of five year old males in region A or the number of immature females in region B. The other time series is the observation process, a realisation of which at time t is here defined as $\mathbf{y}_t$, $t = 1, 2, \ldots, T$. The observation process provides a correspondence between the unobserved true states and the recorded measurement on the population and is completely observable; each element of the observation vector corresponds to a measurement or estimate of some category of the population. For both vectors, $\mathbf{n}_t$ and $\mathbf{y}_t$, the index

t represents an instant in time. The duration between $t-1$ and t is the period of time during which all distinct subprocesses occur once only. For example, consider modelling the dynamics of a population by yearly intervals and let this evolution include a survival process that includes two temporally defined survival rates within a year (e.g. summer and winter survival rates). For this situation the survival process would need to split into two distinct subprocesses; one corresponding to each temporal survival rate.

Although the notation for the index t implies equally spaced intervals it is perfectly possible for the intervals to represent different lengths of time. For models of animal population dynamics that demonstrate regular behaviour and that are observed at a series of regularly spaced time points (e.g. an annual trapping survey) the assumption that the state and observation processes follow regular time series is justifiable. However, if the intervals are not equally spaced care must be taken when interpreting estimates of parameters. For example, in a capture-recapture analysis it may be assumed that the probability of survival is constant across time (i.e. between consecutive sampling occasions) and only a single constant parameter (say, ϕ) is used to model survival. If the sampling intervals are not evenly spaced then the use of constant ϕ is not justified and a more appropriate approach would be to interpret ϕ as a constant per-unit-time survival rate and to model inter-period survival as a function of time such as $\phi_d = \phi^d$ (Pollock *et al.*, 1990) where d is the time between consecutive sampling occasions and is measured on the same scale as per-unit-time survival (e.g. days, weeks, months etc.).

Although the state and observation vectors are linked, they can be of different dimension. The components of $\mathbf{y}_t$ do not always correspond to each individual component of the state vector, $\mathbf{n}_t$. They usually represent the combined total of several elements of the state vector. For example, consider a population of adult red deer being surveyed using

the analysis of dung counts; the observations would not contain any information about the age of the deer producing the dung. The state vector, however, may well categorise the animals by age; in this case a component of the observation vector would represent the aggregation of a subset of components of the state vector.

Full realisations of the state and observation processes over the entire time series are denoted $\mathbf{n}_{0:T}$ and $\mathbf{y}_{1:T}$ respectively, where $\mathbf{n}_{0:T} = \mathbf{n}_0, \mathbf{n}_1, \ldots, \mathbf{n}_T$ and $\mathbf{y}_{1:T} = \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T$. To be able to model the evolution of the states over time the values taken by the states at the very beginning of the period of time covered by the model need to specified. The evolution of these initial states, $\mathbf{n}_0$, during the first time period is then determined by the subprocesses. The values of states at the end of this first time period are denoted by the state vector $\mathbf{n}_1$. The first observation vector, $\mathbf{y}_1$, will then contain components that correspond to either some elements of the state vector, $\mathbf{n}_1$, or to combined totals of multiple elements.

The state and observation processes will be represented as probability functions. These functions will be referred to as pdfs without making a distinction between probability mass functions representing discrete distributions and probability density functions representing continuous distributions. In terms of mathematical representation, when the pdfs are expressed they will be written as integrals on continuous parameter space rather than summations on discrete parameter space. The complete formulation of a state-space model consists of three probability density functions: one that generates the initial state vector, one that models the evolution of the state vector from one time period to the next and one that describes the correspondence between the observation vector and the state

vector. They are written as:

$$\text{Initial state distribution} = g_0(\mathbf{n}_0|\Theta) \tag{3.3.1a}$$

$$\begin{aligned}\text{State Process distribution} &= g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \mathbf{n}_{t-2}, \ldots, \mathbf{n}_0, \Theta) \\ &= g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \Theta)\end{aligned} \tag{3.3.1b}$$

$$\text{Observation Process distribution} = f_t(\mathbf{y}_t|\mathbf{n}_t, \Theta) \tag{3.3.1c}$$

where $t = 1, 2 \ldots, T$ is the time index and Θ represents a vector of all the parameters involved in the model from both the observation and state processes. For state-space models of animal population dynamics these parameters may often include survival probabilities, population carrying capacity, removal probabilities and parameters for observation noise which reflects potential uncertainty in the recorded observations. The state process distribution (3.3.1b) is assumed to be first-order Markovian which means that the current state vector, $\mathbf{n}_t$, is assumed to depend only on the value of the previous state vector, $\mathbf{n}_{t-1}$, and the parameters Θ and not on the value of the states in any previous time periods. A more general form of models, Hidden Process models (Newman *et al.*, 2006), do not require this first-order Markovian condition but otherwise have many parallels with the state-space model structure.

3.3.3 State-space model inference

The nature of the statistical inference obtained from these types of models often involves summarising the state vectors $\mathbf{n}_t$ and the model parameters Θ, both of which are conditional on the observed data $\mathbf{y}_t$. When making inferences about the states in the population, there are three principal distributions that are often of interest.

Firstly, if the object of inference is to examine the distribution of the current states

conditional on all observed data up to the current time period t ($1 \leq t \leq T$) then the distribution $p(\mathbf{n}_t|\mathbf{y}_{1:t}, \Theta)$ will be required. This distribution is referred to as the *filtered* distribution and is used if the object of inference is to estimate the current state of the population given all of the data up until the current time period. This approach to inference would be appropriate for population monitoring schemes that wish to obtain updated estimates of the current population as soon as new observations are made.

Secondly, the distribution of some particular past state of the population conditional on all the observed data up to the current time period t is useful when retrospective investigation of a population is the object of inference. This distribution is denoted as $p(\mathbf{n}_r|\mathbf{y}_{1:t}, \Theta)$ where $r < t$ and is referred to as the *smoothed* distribution. Inference based on smoothed estimates of the state involves the use of the full time series of observations made across the entire duration of the study from time 1 to time t to estimate the state vector at time r. Therefore smoothed estimates can be used to modify the estimates of the population at some past time r ($r < t$) given the data that was observed both post and prior to r.

Thirdly, the distribution of future states of the population given all the observed data, up to time T, is useful when the object of inference is to obtain an estimate of some future state of the population based only on the data observed up to the current period. This distribution is denoted as $p(\mathbf{n}_r|\mathbf{y}_{1:T}, \Theta)$ where $r > T$ and is referred to as the *predicted* distribution. These distributions may be used for population management strategy decisions such as examining the future effect on the population of current levels of harvesting.

The evaluation of these distributions can be viewed as the result of integrating or

summing over the appropriate portion of the states. For example, if the object of inference was to attain the maximum likelihood estimates of the parameters Θ then the likelihood needs to be evaluated. To do this, the joint distribution of the states, $\mathbf{n}_{1:T}$, and observations, $\mathbf{y}_{1:T}$, is obtained:

$$p(\mathbf{n}_{1:T}, \mathbf{y}_{1:T}|\Theta) = g_0(\mathbf{n}_0, \Theta) \prod_{t=1}^{T} f_t(\mathbf{y}_t|\mathbf{n}_t, \Theta) g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \Theta) \quad \text{for } t = 0, 1, \ldots, T \tag{3.3.2}$$

The marginal distribution of the observations $\mathbf{y}_{1:T}$, conditional on the parameters Θ, is then obtained by either summation or integration over the states, $\mathbf{n}_{1:T}$, depending on whether the distribution is discrete or continuous. This distribution is the likelihood and is given as:

$$p(\mathbf{y}_{1:T}|\Theta) = \int_{\mathbf{n}_0} \cdots \int_{\mathbf{n}_T} \left\{ \prod_{t=1}^{T} f_t(\mathbf{y}_t|\mathbf{n}_t, \Theta) g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \Theta) d\mathbf{n}_t \right\} g_0(\mathbf{n}_0, \Theta) d\mathbf{n}_0 \quad \text{for } t = 0, 1, \ldots, T \tag{3.3.3}$$

As a second example, the smoothed distribution $p(\mathbf{n}_{1:T}|\mathbf{y}_{1:T}, \Theta)$ can be obtained using a similar approach based on integration and an application of Bayes Theorem. The smoothed distribution can be expressed in the following way using Bayes Theorem:

$$p(\mathbf{n}_{1:T}|\mathbf{y}_{1:T}, \Theta) = \frac{p(\mathbf{n}_{1:T}, \mathbf{y}_{1:T}|\Theta)}{p(\mathbf{y}_{1:T}|\Theta)} \tag{3.3.4}$$

which, using (3.3.2), can be expressed as:

$$p(\mathbf{n}_{1:T}|\mathbf{y}_{1:T}, \Theta) = \frac{g_0(\mathbf{n}_0, \Theta) \prod_{t=1}^{T} f_t(\mathbf{y}_t|\mathbf{n}_t, \Theta) g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \Theta)}{p(\mathbf{y}_{1:T}|\Theta)} \tag{3.3.5}$$

Consider a random variable X which has a continuous distribution for which the pdf is f, the expectation $E[X]$ is denoted as:

$$E[X] = \int_s x f(x) dx \tag{3.3.6}$$

where S is the set of possible values X can take.

To make inferences based on the smoothed distribution it is useful to calculate the expectation of the states at some time t, $(1 < t < T)$, conditional on all the observations $\mathbf{y}_{1:T}$ and the parameters Θ. Using (3.3.6) the expectation can be expressed as follows:

$$E[\mathbf{n}_t|\mathbf{y}_{1:T}, \Theta] = \int_{\mathbf{n}_0} \cdots \int_{\mathbf{n}_T} \mathbf{n}_t p(\mathbf{n}_t|\mathbf{y}_{1:T}, \Theta) d\mathbf{n}_0 \ldots \mathbf{n}_T \tag{3.3.7}$$

Then, using (3.3.5), the expectation can be written as:

$$E[\mathbf{n}_t|\mathbf{y}_{1:T}, \Theta] = \frac{\int_{\mathbf{n}_0} \cdots \int_{\mathbf{n}_t} \mathbf{n}_t \left\{ \prod_{t=1}^{T} f_t(\mathbf{y}_t|\mathbf{n}_t, \Theta) g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \Theta) d\mathbf{n}_t \right\} g_0(\mathbf{n}_0, \Theta) d\mathbf{n}_0}{p(\mathbf{y}_{1:T}|\Theta)} \tag{3.3.8}$$

This expression allows inferences to be made about some state $\mathbf{n}_t$ conditional on the realisation of the entire observation process.

Another application of Bayes' Theorem is used to obtain the posterior distribution of the model parameters Θ conditional on all the observed data $\mathbf{y}_{1:T}$. This distribution is central to Bayesian inference and is calculated by:

$$p(\Theta|\mathbf{y}_{1:T}) = \frac{p(\mathbf{y}_{1:T}|\Theta)p(\Theta)}{\int_\Theta p(\mathbf{y}_{1:T}|\Theta)p(\Theta)} = \frac{p(\mathbf{y}_{1:T}|\Theta)p(\Theta)}{f(\mathbf{y}_{1:T})} \tag{3.3.9}$$

where the numerator is the product of the prior $p(\Theta)$ and the likelihood $p(\mathbf{y}_{1:T}|\Theta)$ (see Eq. 3.3.3). The denominator $f(\mathbf{y}_{1:T})$ is then obtained by integrating the numerator over the range of the parameters Θ. This evaluates to a constant and therefore the posterior distribution is proportional to the product of the prior for the model parameters Θ and the likelihood of the parameters given the data. The evaluation of this constant can be intractable and methods to evaluate the posterior distribution that can overcome this problem are discussed in Section 3.3.6.

3.3.4 The State Process

The pdf $g_t(\cdot)$ represents the state process model which models the processes driving the evolution of the system under study. Under the state-space model structure this pdf is typically stochastic and represents the variation engendered by the processes. The state-process can be considered to consist of a number of separate sub-processes each modelled with their own pdf (Buckland *et al.*, 2004; Thomas *et al.*, 2005; Buckland *et al.*, 2007). One advantage of this approach is that each individual sub-process can be modelled separately from the others which allows a greater degree of flexibility in the approach to constructing the models. This approach provides a useful framework for testing hypotheses about the particular structure of the model. For example, a population dynamics model for red deer is assumed to be governed by three processes: birth, movement and survival. There may be broad agreement amongst biologists with regard to the birth and movement processes, but there could be some debate with regard to whether the same survival model is appropriate for both males and females. Modularising the state process thus allows these alternative theories to be more easily accommodated and allows hypotheses to be tested.

The sub-processes are assumed to be discrete and to occur sequentially rather than simultaneously. Under these assumptions the pdf g_t can be expressed as a series of linked probability functions, each of which represents a separate sub-process, with $g_{r,t}$ denoting the probability function for the r^{th} sub-process. The link between consecutive probability functions is defined such that the input to one sub-process is taken to be the output from the previous sub-process with $\mathbf{u}_{r,t}$ denoting the resulting state of the population following the r^{th} sub-process occurring between time periods $t-1$ and t. Then, if there are k such sub-processes that operate between these time periods the evolution of the state vector

over this time can be denoted as:

$$\mathbf{u}_{1,t} \sim H_{1,t}(\mathbf{n}_{0:t-1}, \Theta), \mathbf{u}_{2,t} \sim H_{2,t}(\mathbf{n}_{0:t-1}, \mathbf{u}_{1,t}, \Theta), \ldots, \mathbf{n}_t \equiv H_{k,t}(\mathbf{n}_{0:t-1}, \mathbf{u}_{1,t}, \ldots, \mathbf{u}_{k,t}, \Theta)$$

where each distribution denoted by $H_{r,t}$ corresponds to the appropriate sub-process pdf. The dependency on $\mathbf{n}_{0:t-1}$ in all but the first of the distributions given above does not arise if the state process pdf is assumed to be first-order Markov. In this case, if a time period is redefined as the duration between sub-processes, the structure is conceptually that of a state-space model that does not have observations at each time period. This resulting model could then be fitted using standard state-space model methods.

Although this modular approach to model construction does allow greater flexibility it also results in an increased level of complexity for the state process distribution function and the model likelihood. This is because the evaluation of the state process pdf involves integrating over the states immediately following each of the separate sub-processes. This integration is constrained by the values of the current states, $\mathbf{n}_t$, and the state from the previous time period, $\mathbf{n}_{t-1}$. By adopting this modular approach to constructing the state process model the state pdf can be written as:

$$g_t(\mathbf{n}_t|\mathbf{n}_{0:t-1}, \theta) = \int_{\mathbf{u}_{k-1,t}} \prod_{r=1}^{k} g_{r,t}(\mathbf{u}_{r,t}|\mathbf{n}_{0:t-1}, \mathbf{u}_{1,r-1:t}, \Theta) d\mathbf{u}_{1,k-1:t}$$

where $\mathbf{u}_{1:r,t} = \mathbf{u}_{1,t}, \mathbf{u}_{2,t}, \ldots, \mathbf{u}_{r,t}$ denotes the first r sub-processes of the sequence that generates the current state vector, $\mathbf{n}_t$, from the previous state vector, $\mathbf{n}_{t-1}$.

If each sub-process is assumed to be first-order Markov the state process pdf can be

written as:

$$g_t(\mathbf{n}_t|\mathbf{n}_{t-1},\Theta) = \int_{\mathbf{u}_{k-1,t}} \cdots \int_{\mathbf{u}_{1,t}} g_{1,t}(\mathbf{u}_{1,t}|\mathbf{n}_{t-1},\Theta) \times \left[\prod_{r=2}^{k-1} g_{r,t}(\mathbf{u}_{r,t}|\mathbf{u}_{r-1,t},\Theta)\right] \times g_{k,t}(\mathbf{n}_t|\mathbf{u}_{k-1,t},\Theta) d\mathbf{u}_{1,t},\ldots,d\mathbf{u}_{k-1,t} \quad (3.3.10)$$

The pdf for the r^{th} individual sub-process generates a realisation, $\mathbf{u}_{r,t}$, that is now conditional on the model parameters and the realisation, $\mathbf{u}_{r-1,t}$, of the immediately preceding sub-process only. Evaluating the state process pdf is necessary to calculate the likelihood (3.3.3). The complexity of performing the required calculations for any state process defined by multiple sub-processes can often prove to be prohibitive. Alternative approaches, such as sequential Monte Carlo procedures, that do not require the explicit evaluation of the likelihood are therefore extremely attractive.

To fully define the structure of the state process model there are a couple of conventions required. Firstly, it is assumed that a single survey of the animal population of interest occurs in the time interval $(t-1,t]$ and this observation occurs just as the system advances from time period $t-1$ to time period t. It is possible to extend the modelling approach to include a variable number of survey occasions (Buckland *et al.*, 2004) but for ease of representation the examples in this section will be assumed to have one survey corresponding to each time period. Also, the time periods between surveys can be irregularly spaced since the time intervals in which all intermediate sub-processes are realised do not have to be constant. The time period from the end of the $(i-1)^{\text{th}}$ interval to the beginning of the i^{th} interval does not have to be the same for each value of i. Furthermore, the duration of the i^{th} interval can itself vary across the different intervals. Secondly, the starting point of the model is assumed to be the initial state vector $\mathbf{n}_0$ as opposed to any of the intermediate states $\mathbf{u}_{\cdot,1}$ between $\mathbf{n}_0$ and $\mathbf{n}_1$. The values of the initial states $\mathbf{n}_0$ are

required to implement the evaluation of the model likelihood but they are unknown. The first recorded observations, $\mathbf{y}_1$, only correspond to the states at $t = 1$ and therefore no elements of $\mathbf{n}_0$ can be known. Instead they are drawn from the initial state distribution Eq. (3.3.1a).

Referring back to the Leslie matrix representation discussed in section 3.3.1 it can be seen that the modular approach to modelling the intermediate sub-processes can, under certain conditions, result in a very clear representation of the appropriate sequence of the sub-processes as well as their cumulative effect. If the sub-processes are first-order Markov and if the realisation of a sub-process can be expressed as a linear function of the preceding sub-processes then the conditional expectation of the states at time t, given the states at time $t-1$, can be expressed as the product of matrices. Assume that there are three sub-processes in the model and that they form a sequence of three discrete processes with the input to one process being taken as the output from the preceding process. The evolution of the system from the states at time $t-1$ to the states at time t is illustrated schematically below:

$$\mathbf{n}_{t-1} \xrightarrow{g_{1,t}} \mathbf{u}_{1,t} \xrightarrow{g_{2,t}} \mathbf{u}_{2,t} \xrightarrow{g_{3,t}} \mathbf{u}_{3,t} = \mathbf{n}_t$$

The conditional expectation of the states at time t, conditional on the states at time $t-1$ is denoted by $E_{\mathbf{n}_t|\mathbf{n}_{t-1}}$. Using similar notation to denote the conditional dependencies of the expectations the expression can be written as:

$$E_{\mathbf{n}_t|\mathbf{n}_{t-1}} = E_{\mathbf{u}_{1,t}|\mathbf{n}_{t-1}} \left\{ E_{\mathbf{u}_{2,t}|\mathbf{u}_{1,t}} \left[E_{\mathbf{n}_t|\mathbf{u}_{2,t}}[\mathbf{n}_t] \right] \right\}$$

If all the sub-processes are first-order Markov and if the expectation of the i^{th} sub-process $\mathbf{u}_{i,t}$ conditional on $\mathbf{u}_{i-1,t}$ is a linear function of $\mathbf{u}_{i-1,t}$ for any $i > 0$, then the conditional

expectation of the states at t can be written as follows:

$$E_{\mathbf{n}_t|\mathbf{n}_{t-1}} = D_k D_{k-1} D_{k-2} \dots D_1 \mathbf{n}_{t-1}$$

where each matrix D represents a single sub-process and the expectation is given directly as the product of the matrices representing these individual sub-processes.

3.3.5 The Observation Process

The schematic diagram in Fig. 3.1 describes the manner in which the state and observation processes are linked via the modular sub-process approach. The states at the previous time period, $\mathbf{n}_{t-1}$, evolve to the current states, $\mathbf{n}_t$, by means of the intermediate sub-processes, the $\mathbf{u}_{\cdot,t}$s. The observation process model pdfs, f_{t-1} and f_t, connect the observations $\mathbf{y}_{t-1}$ and $\mathbf{y}_t$ to the corresponding states $\mathbf{n}_{t-1}$ and $\mathbf{n}_t$ respectively.

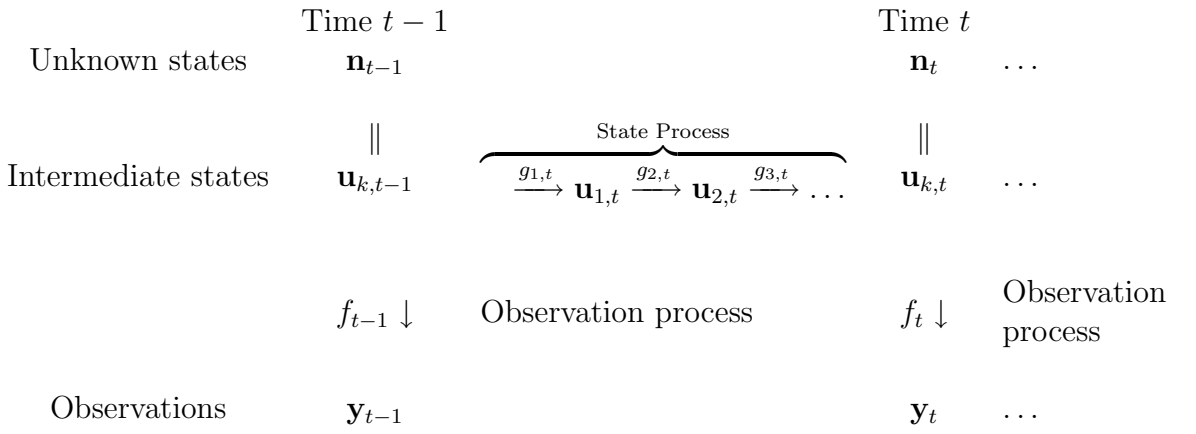


Figure 3.1: Schematic diagram of state and observation processes with linked sub-processes mapping the evolution of the state.

Observational data are collected on the population of interest and are used to make inferences about the states of the population and the parameters of the sub-processes which are assumed to drive the underlying population dynamics. The observational data

are not necessarily assumed to come from a single source, for example, studies of red deer *Cervus elaphus* have incorporated both census and cull information (Trenkel *et al.*, 2000) and some avian studies (northern lapwing (*Vanellus vanellus*) and grey heron (*Ardea cinerea*)) have incorporated census data, survival rate data and information on weather conditions (Besbeas *et al.*, 2002). The observational data is also not assumed to be collected at regular intervals and, as discussed previously, the approach to formulating these state-space models allows for varying durations of the intervals. In a similar fashion to the state process model, the observation process model can be either stochastic or deterministic. The function of the observation process is to model the relationship between the collected observational data and the unknown underlying states in the population. The observational data is stored in the observation vector and this will occasionally consist of components that are exactly equal to components of the state vector. Equally, the components in the observation vector could represent the summation of multiple elements of the state vector with no associated stochastic error. An example of the latter situation would be if the observational data consisted of a complete census in which no distinction was made across different components of the population. In this case the observational data would be the total population and the observation process would just sum all of the components of the state vector. In the cases when the observation process model is deterministic the form of the likelihood (Eq. (3.3.3)) becomes simplified. The component of the likelihood corresponding to the observation process, f_t, will simply equal 1 for the states, $\mathbf{n}_t$, which correspond exactly to the relevant components of the observation vector $\mathbf{y}_t$. Otherwise, $f(\mathbf{y}_t|\mathbf{n}_t, \Theta)$ will equal 0. In these cases the process of evaluating the integral to obtain the likelihood consists of integrating just the state process pdf over the appropriate subspace of all state vectors over the time intervals, $\mathbf{n}_{0:T}$. This appropriate subspace is defined as those states that map correctly to the corresponding observation vector. Then, (using the notation of Buckland *et al.* (2004)), the mapping from the

state vector $\mathbf{n}_t$ to the observation vector $\mathbf{y}_t$ is represented as $m(\mathbf{n}_t)$ and the appropriate subspace can then be expressed as

$$\nu_t = \{\mathbf{n}_t : m(\mathbf{n}_t) = \mathbf{y}_t\}. \tag{3.3.11}$$

Then, the posterior distribution (see Eq. 3.3.9) can be expressed as

$$p(\Theta|\mathbf{y}_{1:T}) = \int_{\mathbf{n}_0}\int_{\nu_1}\cdots\int_{\nu_T}\left\{\prod_{t=1}^{T} g_t(\mathbf{n}_t|\mathbf{n}_{t-1},\Theta)d\mathbf{n}_t\right\} \times g_0(\mathbf{n}_0,\Theta)d\mathbf{n}_0 \tag{3.3.12}$$

This would be the form of the posterior obtained for a single-state population on which measures were observed without error.

3.3.6 Fitting the models.

In section 3.3.3 the expression for the posterior distribution of parameters (Eq. (3.3.9)) was defined and required the calculation of the denominator term $f(\mathbf{y}_{1:T})$. The evaluation of this term is typically a high-dimensional integration problem due to the number of parameters in the models. Even for cases in which the integration need only be performed over a subspace of the states (see Section 3.3.5) the resulting integrals are still high dimensional. Analytic solutions can be obtained in some special cases as mentioned in Section 3.2. For the normal dynamic linear models (Harvey, 1989; Schnute, 1994) the required integration can be performed analytically using the Kalman filter. For example, these models have been applied to fisheries data (Newman, 1998) and avian population time series data (Besbeas *et al.*, 2002, 2003, 2005).

In wildlife populations it is often unrealistic to assume the dynamics of the model of interest can be adequately approximated using linear expressions or that the associated error distribution can be assumed to be normal. Without making the restrictive assumptions required by the Kalman filter, the direct evaluation of the integrals is still

problematic. A solution to this problem is the use of Monte Carlo integration techniques. These methods are used to approximate the posterior distribution using simulation methods which avoid the need for explicit evaluation of $f(\mathbf{y}_{1:T})$.

Computer intensive Monte Carlo methods are useful for fitting complex likelihood models such as those outlined in the previous sections and typically couch model fitting and inference in a Bayesian framework. The Monte Carlo methods can be split into two general, albeit overlapping, classes of inferential procedures: MCMC (Gilks *et al.*, 1996) and sequential importance sampling (hereafter SIS)(Doucet *et al.*, 2001) with both approaches described in Liu (2001).

Using the approach in Newman *et al.* (2006), a general description is given on the way in which both approaches can be applied to fit models that can be classified within the state-space framework and this is summarised in the following paragraphs.

The expected value of a random variable X is calculated using (3.3.6). More generally, if the pdf for X is p, the expectation of some function of X, $E[\vartheta(X)]$ is defined as:

$$E[\vartheta(X)] = \int \vartheta(X)p(x)dx \tag{3.3.13}$$

These integrals can be evaluated using Monte Carlo simulation techniques and samples can be generated for the target pdf $p(x)$. For Bayesian inference $p(x)$ typically represents the posterior distribution (Eq. 3.3.9). If it is possible to simulate directly from the target pdf $p(x)$, simple Monte Carlo integration can be used. The procedure is conceptually simple; a large number (say, N) of samples $x_1^*, x_2^*, \ldots, x_N^*$ are drawn from the target pdf $p(x)$ and the expectation in Eq. 3.3.13 is obtained by evaluating $\vartheta(x^*)$ for each x^ϑ in the

sample and then averaging over all samples. Thus,

$$\hat{E}[\vartheta(X)] = \frac{1}{N}\sum_{i=1}^{N}\vartheta(x_i^*). \tag{3.3.14}$$

However, it is not always possible to directly sample from the the target pdf $p(x)$. In this case one alternative method is to sample from some alternative, feasible pdf $q(x)$. This alternative distribution is typically referred to as the trial pdf. Since $q(x)$ is unlikely to be equivalent to $p(x)$ the evaluation of the integral in Eq. (3.3.14) is no longer valid. Instead, the sample values drawn from the trial pdf $x_1^*, x_2^*, \ldots, x_N^*$ must be adjusted to take account of the fact that they came from $q(x)$ not $p(x)$. The adjustment is determined by the ratio between the target and trial pdfs evaluated for each drawn value. Formally, the weight $w_1(x^*)$ for a particular x^* is given by $\frac{p(x^*)}{q(x^*)}$. The estimates of the integral can now be expressed as:

$$\hat{E}[\vartheta(X)] = \frac{1}{N}\sum_{i=1}^{N}\vartheta(x_i^*)\frac{p(x_i^*)}{q(x_i^*)} = \frac{1}{N}\sum_{i=1}^{N}\vartheta(x_i^*)w_1(x_i^*). \tag{3.3.15}$$

This technique is usually referred to as importance sampling in the literature where the weight or "importance" of a drawn value is determined by how closely the trial pdf mirrors the target pdf when evaluated at the drawn value.

A problem with importance sampling as described above can arise in certain instances. If $p(x)$ cannot be sampled from directly it can also be difficult to evaluate $p(x)$ which is necessary for the evaluation of the weight $w_1(x)$. The difficulty will arise when the target pdf $p(x)$ can be expressed as the product of components of which some are intractable. For example, if $p(x) = h(x)c$ it may be that $h(x)$ is tractable but the constant c is not. This is precisely the case that can arise in Bayesian inference where $p(x)$ represents the posterior distribution (Eq. 3.3.9) and the intractable constant c is the denominator

$f(\mathbf{y}_{1:T})$. The necessary adjustment to the calculation of the weights then requires the intractable constant c to be cancelled out. Thus, if the weights are defined as

$$w_2(x^{*[i]}) = \frac{c\frac{h(x^{*[i]})}{q(x^{*[i]})}}{\sum_{j=1}^{N} c\frac{h(x^{*[j]})}{q(x^{*[j]})}} = \frac{\frac{h(x^{*[i]})}{q(x^{*[i]})}}{\sum_{j=1}^{N} \frac{h(x^{*[j]})}{q(x^{*[j]})}}$$

the evaluation of the integral can be obtained using

$$\sum_{i=1}^{N} \vartheta(x^{*[i]}) w_2(x^{*[i]})$$

Having obtained a sample $x^{*[1]}, x^{*[2]}, \ldots, x^{*[N]}$ from a trial pdf $q(x)$ and calculated the weights $w_2(x^{*[1]}), w_2(x^{*[2]}), \ldots, w_2(x^{*[N]})$ it is possible to convert the sample into a sample from the target pdf $p(x)$. This can be performed using a procedure referred to as bootstrap resampling or sampling importance resampling (SIR) (Smith and Gelfand, 1992). The procedure is, again, conceptually simple and involves the x^*s being resampled with replacement according to the weights $w_2(x^*)$s. The draws from the sample whose evaluation under the trial pdf most closely mirrors the evaluation under the tractable component of the target pdf are given the highest weights. These high weighted particles will tend to be the ones chosen multiple times in the bootstrap resample and therefore the draws deemed to be most closely representative of the target pdf are the ones that dominate the resampled set.

With specific focus on the state-space model formulation, inference on unknown elements in the model can be obtained using a combination of importance sampling and bootstrap resampling. The approach is referred to as sequential importance sampling with resampling (SISR) and can be applied for the situation in which both the parameters, Θ, and the states, $\mathbf{n}_t$, are unknown. For the simple case, assume that the parameters Θ are known and only the states, $\mathbf{n}_t$, need to be estimated. Hence, the target distribution is the

posterior of the states $\mathbf{n}_t$ conditional on all the data observed up to time t, $\mathbf{y}_{1:t}$. In the simplest application of the SISR algorithm the trial pdf for the unknown states at time t, $\mathbf{n}_t$, is simply the state process pdf (Eq. (3.3.1b)) at time t. That is, $q(\mathbf{n}_t) = g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \Theta)$. The weights will then be

$$w_2(\mathbf{n}_t) \propto \frac{g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \Theta) \times f_t(\mathbf{y}_t|\mathbf{n}_t, \Theta)}{g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \Theta)} = f_t(\mathbf{y}_t|\mathbf{n}_t, \Theta) \tag{3.3.16}$$

Thus, the weights are proportional to the observation pdf $f_t(\mathbf{y}_t|\mathbf{n}_t, \Theta)$ at time t and the resampled states, $\mathbf{n}_t$, will be ascribed higher weights the better the fit they provide to the observations. This approach yields filtered estimates of the states, $\mathbf{n}_t$, given $\mathbf{y}_{1:t}$ for each value of t.

The extension of the approach to the situation in which both the states, $\mathbf{n}_t$, and the parameters, Θ, are unknown is conceptually similar to that described for obtaining filtered estimates of the states alone. The basic sequential importance sampling (SIS) algorithm when applied to state-space models requires the joint prior distribution on the parameters and initial states, $g_0(\mathbf{n}_0, \Theta) = g(\Theta) \times g_0(\mathbf{n}_0|\Theta)$, to be defined. A large number, R, of "particles" are then simulated from this prior distribution. The r^{th} $(r = 1, 2, \ldots, R)$ particle is defined as a pair of state and parameter vectors $(\mathbf{n}_0^{[r]}, \Theta^{[r]})$ where $\mathbf{n}_0^{[r]}$ is the r^{th} realisation of the initial population $\mathbf{n}_0$ and $\Theta^{[r]}$ is the r^{th} realisation of the parameters of the model. The algorithm proceeds with each particle being projected forward stochastically according to the state process model for time 1 so that $\mathbf{n}_1^{[r]}$ is simulated from $g_1(\mathbf{n}_1|\mathbf{n}_0^{[r]}, \Theta^{[r]})$. The target pdf to be estimated is the filtered state distribution $g_1(\mathbf{n}_1|\mathbf{y}_1, \Theta)$, and the trial distribution is the state process $g_1(\mathbf{n}_1|\mathbf{n}_0^{[r]}, \Theta^{[r]})$ for each particle. Then, from Eq. (3.3.16)

$$\frac{h(\mathbf{n}_1^{[r]})}{q(\mathbf{n}_1^{[r]})} \propto \frac{f_1(\mathbf{y}_1|\mathbf{n}_1^{[r]}, \Theta^{[r]}) g_1(\mathbf{n}_1|\mathbf{n}_0^{[r]}, \Theta^{[r]})}{g_1(\mathbf{n}_1|\mathbf{n}_0^{[r]}, \Theta^{[r]})} = f_1(\mathbf{y}_1|\mathbf{n}_1^{[r]}, \Theta^{[r]}).$$

Normalising these weights yields

$$w_3^{[r]} = \frac{f_1(\mathbf{y}_1|\mathbf{n}_1^{[r]}, \Theta^{[r]})}{\sum_{i=1}^{R} f_1(\mathbf{y}_1|\mathbf{n}_1^{[i]}, \Theta^{[i]})}. \tag{3.3.17}$$

Thus, from Eq. 3.3.17, the particles are resampled with their weights proportional to the likelihood evaluated for any observations at $t = 1$ and the surviving particles $(\mathbf{n}_0^{[r]}, \mathbf{n}_1^{[r]}, \Theta^{[r]})$ form an approximate sample from the posterior filtered distribution of $g_1(\mathbf{n}_1|\mathbf{y}_1, \Theta)$. The process is then repeated with surviving resampled particles projected forward to the next time period using the state process distribution. The resulting predicted state distributions are then adjusted using the weighted resample where weights are calculated using the observation process distribution. At each time point the resampled particles provide an estimate of the distribution of the filtered states and parameters. Repeating this forward projection followed by weighted resampling until the resampling at the last time period T has been carried out produces an estimate of the posterior parameter density given all the data (see Eq. (3.3.9)).

Although the SIS approach is applicable to obtaining inference on unknown parameters as well as states, there are some well-known problems with the basic approach (Doucet *et al.*, 2000, 2001). The principal problem that can arise is "particle depletion". The resampling step in the SIS algorithm involves N particles being drawn with the probability of selection given by the weight associated with the particle. Those particles with the highest weights will occur most frequently in the sample and particles with only very small weights will rarely occur. Hence, a small number of high weighted particles can come to dominate the weighted bootstrap resample resulting in an approximate posterior distribution that may be a very poor representation of the true posterior state and parameter densities. This problem propagates through time with the dominant particles being

resampled at each time step and constituting a greater proportion of the resampled particles as the number of time steps increases. Thus, over a long time series with a modest number of initial particles the posterior may be dominated by only a handful of unique particles leading to poor approximations to the posterior. Within each sampling step the problem becomes worse if the distribution of particle weights is very skewed with most of the weight accorded to a small number of particles. Running the simulation again could result in a very different group of particles becoming dominant and thus result in high variation of the posterior distribution across simulations. This variation is often referred to as Monte Carlo variation and there are various techniques that have been developed to reduce it.

Various techniques to mitigate particle depletion are discussed in Liu (2001) and Doucet *et al.* (2001). One of the most commonly adopted approaches is kernel smoothing (Trenkel *et al.*, 2000; Liu and West, 2001; Newman *et al.*, 2006). This aims to reduce the impact of particle depletion by introducing some new parameter values in proximate parameter space of 'good' particles, that is those that are supported by the observed data and thus have high likelihood weight. The parameters of each particle are perturbed whereby values for the parameters are drawn from a specified density kernel (typically Multivariate Normal and centered on the mean of the original parameter values) and are added to the original parameters. A smoothing parameter λ (or d in Liu and West (2001)) is used to control the degree of perturbation and effectively makes the perturbed parameter set a weighted average of the original parameters and the values simulated from the density kernel. Typically, if the smoothing parameter is 0 there is no smoothing and the perturbed parameters are just the original parameters and the algorithm is just the non-parametric weighted bootstrap (Smith and Gelfand, 1992). If $\lambda = 1$ there is

maximal smoothing and the perturbed parameters are determined by the choice of density kernel with the original parameters only influencing the perturbed parameters in the specification of the mean and variance-covariance structure of the density kernel. The algorithm is then a parametric weighted bootstrap. Simply perturbing the parameters induces increased variance in the distributions of the parameters. The approach of Liu and West (2001) incorporates shrinkage into a kernel smoothing approach to maintain the mean and variance of the parameters, at least on some scale. It should also be noted that kernel smoothing introduces bias into inference due to the perturbation of parameters; the link between the simulated states and associated parameter values in a given particle is broken. This results in approximate Bayesian inference for the posterior parameter distributions. Finally, kernel smoothing is only really necessary when the focus is on static unknown parameters. It is not typically required for the elements of the state vectors as they regenerate themselves via the state process.

The auxiliary particle filter (Pitt and Shephard, 1999) is another refinement to the basic SIS algorithm and an application of this technique with reference to a grey seal (*Halichoerus grypus*) population is presented in (Thomas *et al.*, 2005). At time step t an "auxiliary" resample of the particles is taken with the sampling weights determined by the expected likelihood of the states at time step $t+1$, conditional on the observed data at $t+1$. The set of resampled particles is projected forward from t to $t+1$ and is then updated (or corrected) using likelihood weights. This is very similar to the weighted bootstrap resample but in this case the likelihood weights are calculated differently and must incorporate the contribution from the auxiliary resampling step. Again, this refinement hopes to reduce Monte Carlo variation by increasing the number of resampled particles that are expected to be projected into the region of state and parameter space best supported by the data. Increasing the number of parameters in the "good" regions

of state and parameter space reduces the variance of the likelihood weights attributed to the resampled particles and therefore reduces the effect of particle depletion.

Residual sampling (Liu and Chen, 1998; Liu, 2001) can also be used to mitigate the effects of particle depletion. By using residual sampling instead of simple random sampling at the weighted resampling stage of the fitting algorithm the obtained set of resampled particles will have the same expected distribution as under simple random sampling but will have smaller Monte Carlo variance. The SIS approaches invoke a resampling step to generate a sample from the distribution of the current state of the population given all of the data up until the current time period. (Carpenter *et al.*, 1999) also propose an improvement to the basic SIR particle filter via a stratified resampling approach. This filtering approach exhibits greater accuracy and precision when compared to the SIR filter. As a result of the resampling step, SIS provides filtered estimates and a smoothing process needs to be implemented to obtain the same output as would be obtained under an MCMC fitting approach. Filtered estimates are useful when dealing with online dynamical systems with states that require updating as soon as new observations are recorded. However, for offline problems such as wildlife population dynamics, interest is typically greatest in inference based on smoothed estimates of states and parameters that takes into account the observed data over the entire study.

Markov chain Monte Carlo methods are an alternative approach that can be used to approximate the posterior distribution (Eq. (3.3.9)). A Markov chain is simulated such that the stationary distribution, the values the iterative scheme converges toward, is the desired posterior distribution of the states $\mathbf{n}_{1:T}$ and parameters Θ. The strength of these approaches are based on the construction of the chain and the manner in which it

can be decomposed into conditional blocks of states and parameters. As described earlier, sequential importance sampling approximates the posterior by iteratively calculating weights at each time period conditional on previous weights. This reduces the problem of estimating high dimensional posterior distributions by breaking up the approximation process into a series of more tractable evaluation steps. Analogously, Markov chain Monte Carlo methods also reduce the estimation problem to a series of evaluation steps. For the MCMC approach blocks of states and parameters are simulated conditional on the values of the other blocks of states and variables in the chain. The choice of how to classify parameters and states in different blocks then becomes a crucial component of implementing MCMC algorithms. Many issues surrounding the choice of blocking schemes and the specification of suitable proposal distributions are covered in Gilks and Roberts (1996).

Both SIS and MCMC based fitting approaches are considered in Buckland *et al.* (2007). A detailed comparison of the relative strengths of SIS and MCMC approaches to inference on state-space models for the dynamics of wild animals populations is given in Newman *et al.* (2009). The performance of both approaches is compared for both real and simulated data sets. A set of criteria are defined to assess the relative performance of the methods: the ease of implementing the fitting routine, the computational efficiency of running the algorithm and the accuracy of estimates produced. A bespoke MCMC approach to fitting a state-space model for which the data are informative relative to the priors is observed to be the "best" general choice. However, they also note that the choice of approach should be determined by the formulation of the state-space model being fitted and the observed data to which the model is fitted.

Chapter 4

The Conditional Approach to Embedding Population Dynamics into Mark-Recapture Models

4.1 Overview of existing approaches

The previous chapters have reviewed some of the major developments in the analysis of capture-recapture data and have provided a detailed introduction to the theory underpinning state-space models: an extremely flexible framework that can be used to fit complex ecological models to data. This chapter will focus on an alternative approach to modelling mark-recapture data in open populations compared to those previously described.

As discussed in Chapter 2 the approaches to analysing capture-recapture data have typically focussed on the estimation of survival. The existing approaches based on the Cormack-Jolly-Seber (Cormack, 1964; Jolly, 1965; Seber, 1965) formulation model capture-recapture data by conditioning on the first capture of an animal and modelling the probability of the animal being recaptured. From this type of analysis it is possible to obtain estimates of survival rates and capture rates but not abundance. The more general Jolly-Seber models (Jolly, 1965; Seber, 1965) include a likelihood component that

models the unmarked (or uncaptured animals) in the population and this allows estimates of abundance, as well as survival and capture, to be obtained.

The typical process used to obtain estimates of survival and abundance is to obtain the values of the model parameters that maximise the model likelihood; that is, the values that can be considered the most likely given the data in the likelihood. These estimates can have closed-form expressions for certain model parameterisations (i.e. fully time-specific parameters in the JS model) see Eqs. (2.3.9) and Pollock *et al.* (1990). From these closed forms it can be seen that a natural estimate for survival (Eq. (2.3.9)) is the ratio between the number of marked animals estimated to be alive in the population just before the end of the current period and the number of marked animals in the population at the start of the current period. Similarly, estimates of abundance (Eq. (2.3.2)) are obtained using the number of marked animals in the population at the start of the current period scaled by the proportion of animals caught in the current sample that were marked. The former quantity is unknown in open population studies and needs to be estimated using the number of animals that survive to be recaptured in future samples. Recruits to the population, often referred to as births, are then estimated deterministically conditional on the estimates of abundance and survival. Thus, the number of births is typically estimated as the difference between the population size at the current sample and the number of survivors from the previous sample.

With these conventional approaches there are some obvious limitations. From Eq. (2.3.6) it can be seen that, assuming there are no losses on capture (that is $n_i - R_i = 0$), it is quite possible for the estimate of births in the current time period to be negative. More generally, these conventional approaches do not prohibit biologically infeasible estimates from being obtained. If it is assumed that breeding females in a population can give

birth to singletons only, then a biologically realistic model should constrain the estimated number of births to not exceed the number of breeding females in the population. These problems arise from the deterministic calculation of the estimated births; the manner in which births are calculated is a by-product of the form assumed for the model likelihood. The actual birth process itself is not explicitly incorporated into the model.

It is possible to impose constraints when fitting models using existing techniques to respect certain assumed biological relationships. For example, adding a constraint to enforce the number of animals at time $t + 1$ to be no greater than the number of animals at time t plus the number of breeders at time t would allow the biological assumption that you can have no more than one birth per adult on average to be respected. This constrained maximisation approach may well lead to convergence problems due to the complexity of the engendered likelihood surface. Equally, the constrained maximum likelihood no longer respects the asymptotic properties such as normality and unbiasedness. Similarly, it can be possible to obtain estimates of survival that exceed unity under the Jolly-Seber model. This chapter describes a new approach which ensures the estimated parameters in the model and the estimated changes in the population over time are consistent with what is assumed to be known about the biology of the population. By formulating capture-recapture models in a state-space modelling framework it is possible to embed a population dynamics model into the inferential procedure and thus restrict the model estimates to the biologically feasible regions of state-space.

In Chapter 2 the general form of the Jolly-Seber model with time-specific survival ϕ_t and capture p_t was introduced and an expanded version of the likelihood formulation was given in Eq. (2.3.10). This split the likelihood into three components that modelled the

capture of unmarked animals, the losses on capture of marked animals and the conditional probability of recapturing marked animals. The basic form and its extensions were shown to form the basis for much of the work that has been done on open population capture-recapture studies. The extension shown in Eq. (2.3.10) allowed loss (or death) of animals on capture to be incorporated into the model. Known deaths on capture can arise when animals are marked with identifying tags that are returned when found on a dead animal. Thus, the JS likelihood can be extended to incorporate tag-return information to model known deaths (Buckland, 1980). The likelihood can also be simplified somewhat for reduced parameter models in which either survival ϕ_t and/or capture p_t are constant over time (Jolly, 1982) (see Section 2.3.1.4). To address potential heterogeneity problems extensions to the likelihood by classifying the population into approximately homogeneous groups have been developed. Age-structured cohort models have been developed (Pollock, 1981; Buckland, 1982; Brownie *et al.*, 1985; Pollock *et al.*, 1990; Lebreton *et al.*, 1992) (see Section 2.2.1.1). More general multistate models in which animals can move between states stochastically have been developed (Schwarz *et al.*, 1993; Brownie *et al.*, 1993; Nichols *et al.*, 1994; Nichols and Kendall, 1995) (see Section 2.2.3). Alternatively, unexplained variation in the data can be incorporated into the model using auxiliary information to make ϕ_t and/or p_t functions of covariates (Lebreton *et al.*, 1992; Pollock, 2002). If the covariates are measured on the individual level then time-varying individual covariates can also be incorporated into the Cormack-Jolly-Seber model framework (Bonner and Schwarz, 2006) (see Section 2.2.5) with Dupuis and Schwarz (2007) developing a multistate Jolly-Seber capture-recapture model that uses data augmentation to accommodate missing covariates in the model.

All of these refinements and extensions of the basic Jolly-Seber likelihood have provided the modeller with a flexible suite of techniques that can allow more biologically

realistic models to be fit to data. However, there are still limitations owing to the lack of any embedded population dynamics model in these extensions. The extensions to the Jolly-Seber likelihood developed by Crosbie and Manly (1985) and refined by Schwarz and Arnason (1996) explicitly incorporate the birth process. They consider a superpopulation of all animals that are available to be sampled at some point during the study and model the proportion of this superpopulation that enter the sampling population at each time (see Section 2.4). The sampling population is simply defined as the animals that are present in the population and available to be captured during a particular time step. Therefore the new entrants from the superpopulation at each time step can be regarded as births or immigrants. This approach was further extended by Link and Barker (2005) who reparameterised the models under the Schwarz and Arnason (1996) approach by replacing the parameter controlling the proportion of superpopulation entries at each time step with a fecundity parameter that indexed the per-capita recruitment rates. This reparameterisation is felt to yield a more biologically realistic interpretation with regard to per-capita recruitment rates.

General approaches to providing a unified capture-recapture model (Barker and White, 2004; Schofield and Barker, 2008) emphasise the generality of the extensions to the Jolly-Seber modelling framework and propose the construction of component-based models in which the likelihood is formulated dependent on the inferential aims of the study and the limitations in the data. Both approaches allow for a birth process to be included in the likelihood and the unified approach of Schofield and Barker (2008) explicitly constructs matrices of times of birth and death for the animals in the population thus allowing certain demographic processes of interest to be modelled as functions of the summary statistics on births and deaths. For more details of these extensions that incorporate birth processes see Sections 2.5.4-2.5.6.2.

However, under these modelling approaches it is not always possible to incorporate certain forms of constraints on the model parameters. This can be especially problematic if the biological parameter of interest has been transformed for computational expediency. Without imposing extra constraints on these models, these approaches can still yield inference that fails to respect what is assumed to be biological reality. For example, if it is assumed that mature females can produce only singletons during a breeding season, there are no internal constraints in these models that will ensure the estimated number of births does not exceed the estimated number of mature females. Also, the modeller may wish to construct a system in which birth rate was constant over time but the number of births was constrained by the number of mature females in a time period. Barring constrained maximisation, there is no mechanism to include these joint constraints in any of the these extended models, that is, there is no population dynamics model embedded in inference for these approaches. State-space models (Chapter 3) provide an extremely flexible framework that can be used to fit complex ecological models and can accommodate embedded population dynamics models. Thus, by adopting a state-space approach to modelling capture-recapture data the limitations of the existing Jolly-Seber based approaches can be mitigated.

4.2 Formulating the Models

The state-space modelling approach to capture-recapture models can be split into two alternative approaches: the conditional or the unconditional approach. The focus of this chapter will be on the former approach but it is useful to elucidate the philosophical differences between the two alternative approaches. The data used to fit the models will be the usual matrix of individual capture histories as described in Section 2.2. The manner

in which the mark-recapture data are conditioned on in the model is the key distinction between the two approaches and leads to two different ways of incorporating the data into the modelling framework.

For the conditional approach, the conditioning is on those capture history patterns that include capture in the current time period. Hence, the capture history associated with an animal determines the state the animal is assigned to. For each time period the model is conditioned on the number of animals that were observed during that time period. That is, the number of animals to have a capture history pattern which includes capture in the current time period is known exactly for each time period. Consequently, there is no error associated with this observation process and there is a direct correspondence between elements in the state vector and the observed number of capture histories containing capture in the current period. The unknown elements of the state vector correspond to the capture histories that do not contain capture in the current period. Hence, stochasticity enters the model framework through the modelling of the number of animals that are present in the population but remain unobserved in the current time period. Thus, propagating the state vectors through time conditional on the capture history data is what gives rise to the name for the conditional approach. It is possible to categorise the state classification by other covariates (e.g. gender, location, age) in addition to capture history, but for the examples presented in the following chapters the principle variable used to determine state allocation is the observed capture history pattern associated with each animal.

For the unconditional approach there is no conditioning on the matrix of capture history patterns. Consequently, the capture histories are now regarded as observations on the evolving state process and whether an animal is captured or not does not influence its

state. Under this modelling approach stochasticity enters the model through the capture process and the observed capture histories are then regarded as a single random realisation of this stochastic capture process. This is more consistent with the classical Jolly-Seber capture-recapture methods summarised earlier and in recent work on developing state-space models for capture-recapture data (Giminez *et al.*, 2007) this is also the approach taken.

In the standard definition of the three model processes that constitute a state-space model (Eqs (3.3.1a)-(3.3.1c)) the vector of parameters Θ contained all the parameters used in the model for both the state and the observation processes. If $\boldsymbol{\theta}$ now denotes the vector of parameters that are used solely in the state process and $\boldsymbol{\psi}$ denotes the vector of parameters that are used solely in the observation process then Θ can be defined as $\Theta = (\boldsymbol{\theta}, \boldsymbol{\psi})$. In the standard state-space formulation the observation process (see Chapter 3.3.5) models the assumed relationship between the observed data and the states in the population. For a state-space representation of a capture-recapture model (Giminez *et al.*, 2007) the observation process would be given by the capture process and would be a function of the probability of capture p_t. Thus the parameters relating to capture would appear in $\boldsymbol{\psi}$. Under the conditional approach the capture process determines the observed capture histories and, consequently, the state of an animal. Hence, the parameters relating to the probability of capture will appear in the vector of state parameters $\boldsymbol{\theta}$. However, under the unconditional approach, the parameters relating to capture do appear in the vector of observation model parameters $\boldsymbol{\psi}$.

A further important distinction needs to be made between the conditional and unconditional approaches with reference to the interpretation of the model structure. By restricting the conditional model to operate on the observed capture histories at each time

point the model inference is restricted to the population that generated this specific set of observed capture histories. The conditional approach can therefore be said to assume a "population model" in which the population is of some fixed but unknown size. The unconditional model allows for stochasticity in the capture process and therefore allows for two conceptually different interpretations of the model inference. A population model may still be assumed in which, as for the conditional case, the population is of some unknown but fixed size and, contrasting with the conditional case, each realisation of the stochastically modelled capture histories corresponds to this fixed population. Alternatively a superpopulation model may be assumed in which each realisation of the stochastically modelled capture histories corresponds to a different population. For the superpopulation approach a state process model also needs to be specified as a component of the likelihood. This then stochastically models the entry probabilities of births into the population for each time period thus determining the possible range of observed capture histories for each realisation of the population. A more detailed discussion of the distinction between population and superpopulation approaches is given in section 6.2.1.

4.3 The Conditional Approach

As described in Section 4.2, for the conditional approach $\Theta = \boldsymbol{\theta}$, since due to the parameters relating to the capture process now belonging to the vector of state parameters $\boldsymbol{\theta}$, there are no observation parameters $\boldsymbol{\psi}$. There is still an observation process model but it is deterministic and degenerate because for year t the number of animals with capture histories that include capture in time t are observed without error. Thus there is an equivalence between these observed totals and the corresponding elements of the state vector $\mathbf{n}_t$. The state vector for time t also contains elements corresponding to animals that are present in the sampling population but are not caught in time t and it is these elements

that are modelled through the state process pdf at time t, $g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \theta)$ (see Eq. (3.3.1b)).

The particular structure of the probability functions for the observation and state processes can be best understood using a simple example. Consider a conditional model in which the state an animal belongs to is determined solely by its capture history. Then each element of the state vector $\mathbf{n}_t$ will correspond to the numbers of animals possessing a possible capture history pattern at time t. The state vector can be decomposed as shown in Table 4.1.

$$
\begin{array}{rcl}
\mathbf{s}_t & = & \text{the vector of numbers of animals for all capture history} \\
 & & \text{patterns at time } t \text{ that include capture in time } t. \\
\breve{\mathbf{s}}_t & = & \text{the vector of numbers of animals for all capture history} \\
 & & \text{patterns at time } t \text{ that exclude capture in time } t. \\
\mathbf{n}_t & = & \begin{bmatrix} \mathbf{s}_t \\ \breve{\mathbf{s}}_t \end{bmatrix} \quad = \quad \text{the vector of numbers of animals for each possible cap-} \\
 & & \text{ture history pattern in time } t.
\end{array}
$$

Figure 4.1: Notation for decomposition of state vector $\mathbf{n}_t$.

The notation $\breve{\mathbf{s}}_t$ is used rather than $\mathbf{u}_t$ for the unseen elements in the state vector at time t to avoid confusion with the intermediate vector after process k in time t, $\mathbf{u}_{k,t}$.

This decomposition splits the state vector into an observed vector, $\mathbf{s}_t$, and an unobserved vector, $\breve{\mathbf{s}}_t$. This state vector representation can then be used for the observation process model. From Eq. (3.3.1c), the observation process distribution is typically expressed as $f_t(\mathbf{y}_t|\mathbf{n}_t, \psi)$. Under the conditional approach the observation pdf at time t is deterministic and degenerate and can be expressed as:

$$
f_t(\mathbf{y}_t|\mathbf{n}_t) = f_t(\mathbf{y}_t|\mathbf{s}_t) = \begin{cases} 1 & \text{if } \mathbf{s}_t = \mathbf{y}_t \\ 0 & \text{otherwise} \end{cases} .
$$

It is this conditioning on the observed components $\mathbf{s}_t$ of the states at time t that defines

the conditional approach.

4.3.1 Simulation Methods

This section develops the model fitting approach using the sequential importance sampling (SIS) methods that were introduced in Section 3.3.6.

In the context of SIS, model fitting methods require the simulation of state vectors. Therefore, for the conditional approach, model fitting methods must ensure that for any simulated state vector at time t $\mathbf{n}_t^*$, the elements corresponding to $\mathbf{s}_t$ need to match exactly to the corresponding totals in the observed data at time t $\mathbf{y}_t$ otherwise the weight in the importance sampling step will be zero. For example, consider a two sample study for a single group of animals in which state elements are determined solely by capture history. At time $t = 2$ there are four possible capture histories, $00, 01, 10$ and 11. Hence,

$$\mathbf{n}_2 = \begin{bmatrix} n_{2,00} \\ n_{2,01} \\ n_{2,10} \\ n_{2,11} \end{bmatrix} = \begin{bmatrix} \mathbf{s}_2 \\ \breve{\mathbf{s}}_2 \end{bmatrix} = \begin{bmatrix} n_{2,01} \\ n_{2,11} \\ n_{2,00} \\ n_{2,10} \end{bmatrix}$$

Then, the observed elements are $\mathbf{s}_2 = \begin{bmatrix} n_{2,01} \\ n_{2,11} \end{bmatrix}$ and the unobserved elements are $\breve{\mathbf{s}}_2 = \begin{bmatrix} n_{2,00} \\ n_{2,10} \end{bmatrix}$. The observation vector at time $t = 2$ is $\mathbf{y}_2 = \begin{bmatrix} y_{2,01} \\ y_{2,11} \end{bmatrix}$. Thus, any simulated vector for $\mathbf{n}_2^*$ must generate $\mathbf{s}_2$ such that $\mathbf{s}_2 = \mathbf{y}_2$ for each element. For a population classified as a single group (e.g. no classification by different genders, ages or locations etc.) at time t the observation vector $\mathbf{y}_t$, and therefore $\mathbf{s}_t$, will consist of t elements corresponding to capture history patterns including capture at the time t. The observation pdf is then

degenerate and will evaluate to 0 unless every simulated state element corresponding to a capture history pattern which includes capture on the most recent time period is exactly equal to the corresponding element of the observation vector. That is, if $s_{j,t} \neq y_{j,t}$ for any $j = 1, \ldots, t$ then $f_t(\mathbf{y}_t|\mathbf{s}_t) = 0$.

The vector $\check{\mathbf{s}}_t$ consists of the unknown elements in the state vector at time t and it is these elements that are modelled stochastically through the state process pdf $g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \theta)$. Assuming the capture-recapture data are collected on T sampling occasions, any simulated element of $\check{\mathbf{s}}_t$ must be consistent with the previously determined state elements $\mathbf{s}_{1:t-1}$, the current known state elements $\mathbf{s}_t$ and the known values of future state elements $\mathbf{s}_{t+1:T}$. The fitting algorithm therefore must be able to simulate the stochastic elements of the state process pdf under the constraints imposed by conditioning on the entire series of known state elements, $\mathbf{s}_{1:T}$. This requirement makes fitting the models a complex and involved process.

To simplify the notation it is assumed that the parameters Θ and the initial states $\mathbf{n}_0$ are known and dependencies on these will be implicit in the following expressions. The aim of using these sequential importance sampling methods is to obtain a sample from some specified target distribution. For the conditional approach to fitting state-space models to data the target pdf will typically be the conditional distribution for the filtered estimates of the states at time t, $g_t(\mathbf{n}_t|\mathbf{y}_{1:t})$.

As described in Section 3.3.6, the simplest implementation of the SIS fitting algorithm occurs when, at each sampling step, the trial density $q()$ is simply the state process pdf at time t, that is $q(\mathbf{n}_t) = g_t(\mathbf{n}_t|\mathbf{n}_{t-1})$ (omitting the dependency on Θ). This simple approach is highly inefficient for fitting state-space models under the conditional approach. The

state process pdf simply models the stochastic progression of the states of the population through time irrespective of the observations or measures that have been made on the system. Thus, any generated state vectors $\mathbf{n}_t$ are unlikely to have an $\mathbf{s}_t$ sub-vector that corresponds exactly with the observation vector $\mathbf{y}_t$, hence, for the majority of the simulated $\mathbf{n}_t = [\mathbf{s}_t, \breve{\mathbf{s}}_t]'$ (where $'$ denotes the transpose of the vector) it will be the case that $s_{j,t} \neq y_{j,t}$ for some $j \in (1, \ldots, t)$. Consequently, $f_t(\mathbf{y}_t|\mathbf{n}_t) = 0$ for the majority of $\mathbf{n}_t$ simulated from the state process pdf. Therefore, the majority of the weights under the most simple SIS implementation will be zero making the procedure highly inefficient.

The necessary solution is to develop a trial pdf that respects the constraints imposed on the simulated state vectors by the observed data. Let $\mathbf{n}_t^*$ denote a simulated state vector from a trial pdf. The aim is to develop a suitable trial density so that any generated state vector $\mathbf{n}_t^*$ satisfies $f(\mathbf{n}_t^*|\mathbf{y}_t) = 1$. Conceptually, this can be seen as the logical endpoint of the auxiliary particle filter with all resampled auxiliary particles being restricted to the region of state and parameter space defined by the data.

If such a trial density is denoted as $h(\mathbf{n}_t|\mathbf{y}_t, \mathbf{n}_{t-1})$ then, from Eq. (3.3.16), the weights accorded to the simulated state vector $\mathbf{n}_t^*$ are

$$w \propto \frac{f(\mathbf{y}_t|\mathbf{n}_t^*)g_t(\mathbf{n}_t^*|\mathbf{n}_{t-1})}{h(\mathbf{n}_t^*|\mathbf{y}_t, \mathbf{n}_{t-1})} \tag{4.3.1a}$$

$$= \frac{g_t(\mathbf{n}_t^*|\mathbf{n}_{t-1})}{h(\mathbf{n}_t^*|\mathbf{y}_t, \mathbf{n}_{t-1})} \tag{4.3.1b}$$

where $f(\mathbf{y}_t|\mathbf{n}_t^*) = 1$ since all the constraints imposed by the data have been met by the trial density $h()$; that is $\mathbf{s}_t^* \equiv \mathbf{y}_t$. Note the inclusion in the trial pdf $h()$ of the state at the previous time period $\mathbf{n}_{t-1}$. This is included because it can contain information on $\mathbf{n}_t$ that is not contained in the observed data $\mathbf{y}_t$, for example uncaptured animals at t that were caught at $t-1$ are not included in $\mathbf{y}_t$ but can act as a constraint on $\mathbf{n}_t$.

Under the basic SIS approach, where the state pdf was the trial density and the simulation of state vectors was not constrained by the observed data, it could be shown that to calculate the weights associated with a simulated state vector $\mathbf{n}_t^*$ only the likelihood needed to be evaluated (Eq. (3.3.17)). The trial pdf $g_t(\mathbf{n}_t^*|\mathbf{n}_{t-1})$ was used solely to generate candidate particles, $\mathbf{n}_t^*$, and did not need to be evaluated. However, under the trial density proposed to meet the constraints the weight evaluation becomes more complicated. From the expression for Eq. (4.3.1b) it can be seen that the evaluation of the weights involves evaluating both $g_t(\mathbf{n}_t^*|\mathbf{n}_{t-1})$ and $h(\mathbf{n}_t^*|\mathbf{y}_t, \mathbf{n}_{t-1})$ or at least their ratio up to some constant of proportionality. Directly evaluating $g(\mathbf{n}_t^*|\mathbf{n}_{t-1})$ can be complicated, especially when inference on the model processes may be convolved (Buckland *et al.*, 2004). Convolutions can arise when a state element in $\mathbf{n}_t$ represents the summation of elements generated via intermediate sub-processes. The sub-process elements that are summed may have belonged to different states in $\mathbf{n}_{t-1}$ and therefore, for a multistate model with state-specific transition rates, their probabilities of evolving from $\mathbf{n}_{t-1}$ to $\mathbf{n}_t$ are not necessarily the same. These differing ancestral paths need to be accounted for when evaluating the trial pdf $h()$. The need to evaluate probabilities over multiple convolved scenarios can make evaluation of $g(\mathbf{n}_t^*|\mathbf{n}_{t-1})$ difficult.

To increase the tractability of the calculations required for the importance sampling weights it may be necessary to calculate the weight based on elements of the intermediate vectors $\mathbf{u}_{1:k,t}$ for which the state process pdf can be more easily evaluated. For example, consider a z-element state vector $\mathbf{n}_t$ from which one element is obtained by summing components of the intermediate vector $\mathbf{u}_{1,t}$ that preceded it, e.g. $n_{1,t} = u_{1,1,t} + u_{5,1,t} + u_{9,1,t} + u_{13,1,t}$ where $u_{1,1,t}$ is the first element in the vector $\mathbf{u}_{1,t}$ which denotes the intermediate states after the first sub-process, $u_{5,1,t}$ is the fifth element in $\mathbf{u}_1$ and so on.

In this case, to evaluate the weights the values $(u_{1,1,t}, u_{5,1,t}, u_{9,1,t}, u_{13,1,t}, n_{2,t}, n_{3,t}, \ldots, n_{3,t})$ need to be simulated using the trial pdf $h()$ and evaluated under both the state pdf $g(\mathbf{n}_t^*|\mathbf{n}_{t-1})$ and the trial pdf $h(\mathbf{n}_t^*|\mathbf{y}_t, \mathbf{n}_{t-1})$ where $\mathbf{n}_t^*$ now represents the expanded vector, including the intermediate state elements that are simulated subject to the constraints imposed by the data.

The implementation of this model fitting importance sampling procedure is then defined by the choice of trial pdf $h()$ and the appropriate expansion of the simulated state vector. These choices can be somewhat ad-hoc leading to bespoke fitting procedures for particular modelling problems, however there are some general commonalities between possible approaches. The following discussion investigates these issues from a heuristic perspective.

To investigate potential expansions of the state vector it can be constructive to produce directed graphs that represent the evolution of a state vector over a single time period. An initial starting point $\mathbf{n}_{t-1}$ is considered and the evolution of this state vector via each of the sub-processes assumed to be in the model is then graphed as a process tree diagram. The tree therefore begins with each element of the state vector representing a 'parent node' at the top of a tree. Then branches are added to each node to denote the various fates that can be attributed to the animals in the parent node. For each of these intermediate nodes further branches are added to correspond to the possible outcomes of a particular sub-process. After all the branches attributable to the various sub-processes have been added the resulting final set of nodes will represent the final fates of the animals classified in the initial parent nodes. For example, consider a relatively simple population

consisting of two states at time $t-1$,

$$\mathbf{n}_{t-1} = \begin{bmatrix} n_{1,t-1} \\ n_{2,t-1} \end{bmatrix}$$

corresponding to animals that were unobserved and observed in time period $t-1$ respectively. There are two sub-processes assumed to model the dynamics of this system: survival and capture. It is assumed that the animals in the population survive from $t-1$ to t with probability ϕ_t. It is assumed that the animals in the population are captured at sampling period t with probability p_t. This example can be represented schematically by the simple trees displayed in Figure 4.2. For this example, the state vector at time t, $\mathbf{n}_t$ is then given by collapsing the elements of $\mathbf{u}_{2,t}$ such that

$$\mathbf{n}_t = \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix} = \begin{bmatrix} u_{1,2,t} + u_{3,2,t} \\ u_{2,2,t} + u_{4,2,t} \end{bmatrix}$$

This collapsing process is also illustrated in Figure 4.2. As a result of this collapsing process the intermediate states generated by the capture sub-process are combined into two groups: one corresponding to animals captured in t, the other corresponding to animals not captured in t. It can then be seen that attempting to evaluate the weights using direct evaluation of the state pdf $g(\mathbf{n}_t^*|\mathbf{n}_{t-1})$ involves convolutions as both the captured and uncaptured elements in $\mathbf{n}_t$ are obtained by summing across elements of the intermediate states. In this case "expanding" the state vector would be equivalent to replacing $\mathbf{n}_t^*$ with $\mathbf{u}_{2,t}^*$.

For many population dynamics models constructed in a state-space modelling framework the expanded state vector can be obtained using a semi-automated procedure. In Section 3.3.1 the generalised Leslie matrix approach to representing the evolution from $\mathbf{n}_{t-1}$ to $\mathbf{n}_t$ (Buckland *et al.*, 2004; Thomas *et al.*, 2005) was discussed. For the simple

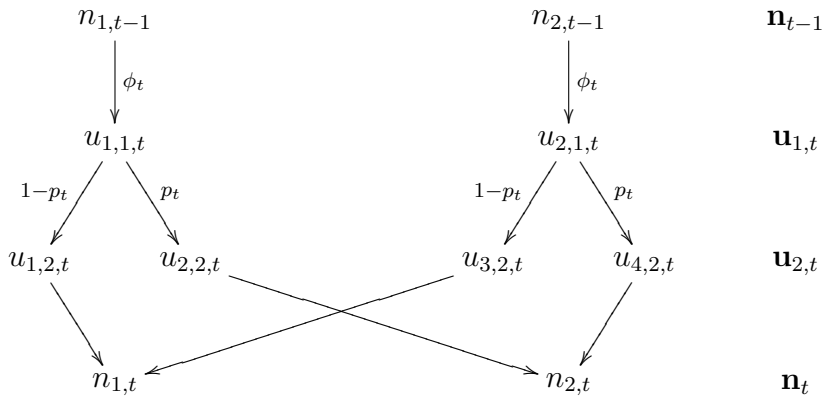


Figure 4.2: Process Tree diagram for the evolution from $\mathbf{n}_{t-1}$ to $\mathbf{n}_t$ via the sub-processes for the simple example.

example discussed above, if S denotes the survival matrix and C denotes the capture matrix the expectation of the states $E[\mathbf{n}_t]$ could then be written

$$E[\mathbf{n}_t] = \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix} = \begin{bmatrix} 1-p_t & 1-p_t \\ p_t & p_t \end{bmatrix} \begin{bmatrix} \phi_t & 0 \\ 0 & \phi_t \end{bmatrix} \begin{bmatrix} n_{1,t-1} \\ n_{2,t-1} \end{bmatrix} = CS\mathbf{n}_{t-1}$$

The semi-automated approach then involves expanding the matrix representation of each sub-process to ensure that each row of the matrix contains at most a single positive entry. If a single component of an intermediate or state vector is obtained by performing operations on multiple components of the preceding vector then there is a convolution. Expanding the matrices can remove this problem. For the simple example above, the number of animals that were captured at t is a convolution of both observed and unobserved animals at $t-1$ that then survived to be captured at t. Hence, the expanded capture matrix will now look like

$$C_{ex} = \begin{bmatrix} 1-p_t & 0 \\ 0 & p_t \\ 1-p_t & 0 \\ 0 & p_t \end{bmatrix}$$

It can then be seen that $C_{ex}S\mathbf{n}_{t-1} = E[\mathbf{u}_{2,t}]$, which was the earlier suggestion for the 'expanded' state vector.

Choosing the trial pdf $h()$ is again non-trivial. For animal populations in general and often those analysed in a capture-recapture framework, the functional form of the sub-processes is often multinomial. Consequently, the elements in the state vector $\mathbf{n}_t$ are functions of multinomial random variables as are the elements in the expanded state vector. The correspondence between elements of the state vector $\mathbf{n}_t$ and the observation vector $\mathbf{y}_t$ can lead to another source of stochasticity in the modelling framework. As noted in Section 3.3.5 the observation vector can have components that correspond exactly to those in the state vector but it could also be the case that some elements of $\mathbf{y}_t$ represent the summation of multiple elements of $\mathbf{n}_t$. Therefore, there needs to be a splitting process which allocates $\mathbf{y}_t$ across suitable elements of $\mathbf{n}_t$. For example, in the simple case described above the observation vector $\mathbf{y}_t$ consists of a single element corresponding to the number of animals captured during t. This observed total acts as a constraint on the simulation of the intermediate vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ and therefore $\mathbf{y}_t$ needs to be allocated across appropriate elements of the extended state vector, which in this case is equivalent to $\mathbf{u}_{2,t}$. An obvious choice to split a discrete total into a set of discrete elements would be a multinomial distribution. Hence, multinomial densities are often used for both the sub-process and splitting functions thus making multinomial trial densities an appropriate choice for $h()$. Having determined the functional form of the densities it is then necessary to determine how to specify the appropriate parameters for the multinomial distributions.

The generalised Leslie matrix representation of the models can also be used in one approach to determining the rate parameters of the multinomial splitting functions. Conceptually, the multinomial rates represent the relative probabilities of an individual animal

being assigned to each of the possible elements. These relative probabilities can be approximated by the ratios of the expected values, under the state process model, for each of the appropriate elements of the expanded state vector $\mathbf{n}_t$. The generalised Leslie matrix representation immediately provides a straightforward method of obtaining approximate expected values by multiplying together the matrix components representing each subprocess.

Using the simple example again, the expected values of the extended state vector are:

$$E[\mathbf{u}_{2,t}] = \begin{bmatrix} u_{1,2,t} \\ u_{2,2,t} \\ u_{3,2,t} \\ u_{4,2,t} \end{bmatrix} = \begin{bmatrix} 1-p_t & 0 \\ 0 & p_t \\ 1-p_t & 0 \\ 0 & p_t \end{bmatrix} \begin{bmatrix} \phi_t & 0 \\ 0 & \phi_t \end{bmatrix} \begin{bmatrix} n_{1,t-1} \\ n_{2,t-1} \end{bmatrix} = \begin{bmatrix} (1-p_t)\phi_t n_{1,t-1} \\ p_t\phi_t n_{1,t-1} \\ (1-p_t)\phi_t n_{2,t-1} \\ p_t\phi_t n_{2,t-1} \end{bmatrix}$$

The total observed number of animals caught at t is given as $\mathbf{y}_t = y_{1,t}$. From the above Leslie matrix approximation to the expected values, the appropriate elements are

$$\begin{aligned} E[u_{2,2,t}] &= p_t\phi_t n_{1,t-1} \\ E[u_{4,2,t}] &= p_t\phi_t n_{2,t-1} \end{aligned}$$

Then, the expected proportion of $y_{1,t}$ that is assigned to $u_{2,2,t}$ is approximated by

$$r_{2,2,t} = \frac{E[u_{2,2,t}]}{E[u_{2,2,t}] + E[u_{4,2,t}]} = \frac{n_{1,t-1}\phi_t p_t}{n_{1,t-1}\phi_t p_t + n_{2,t-1}\phi_t p_t} = \frac{n_{1,t-1}}{n_{1,t-1} + n_{2,t-1}}$$

Since the total $y_{1,t}$ is only split over two elements the distribution here is binomial. Thus, a possible trial density for simulating a value $u^*_{2,2,t}$ could be:

$$u^*_{2,2,t} \sim Binomial(y_{1,t}, r_{2,2,t})$$

Then, $u^*_{4,2,t}$ is simply $y_{1,t} - u^*_{2,2,t}$. This still leaves the intermediate vector corresponding

to survival $\mathbf{u}_t$ and the elements $\{u_{1,2,t}, u_{3,2,t}\}$ corresponding to non-capture to be estimated from the trial pdf $h()$. Given that the vector $\mathbf{n}_{t-1}$ is assumed known and the values $\left\{u^*_{2,2,t}, u^*_{4,2,t}\right\}$ have been assigned it is possible to simulate values for the remaining unknown states conditional on the values that are known or have been simulated so far. Binomial distributions are assumed for the sub-processes relating to survival and capture. From Figure 4.2, it can be noted that

$$p(u_{1,1,t}|n_{1,t-1}) = \binom{n_{1,t-1}}{u_{1,1,t}}(\phi_t)^{u_{1,1,t}}(1-\phi_t)^{n_{1,t-1}-u_{1,1,t}} \tag{4.3.2a}$$

$$p(u_{2,1,t}|n_{2,t-1}) = \binom{n_{2,t-1}}{u_{2,1,t}}(\phi_t)^{u_{2,1,t}}(1-\phi_t)^{n_{2,t-1}-u_{2,1,t}} \tag{4.3.2b}$$

$$p(u^*_{2,2,t}|u_{1,1,t}) = \binom{u_{1,1,t}}{u^*_{2,2,t}}(p_t)^{u^*_{2,2,t}}(1-p_t)^{u_{1,1,t}-u^*_{2,2,t}} \tag{4.3.2c}$$

$$p(u^*_{4,2,t}|u_{2,1,t}) = \binom{u_{2,1,t}}{u^*_{4,2,t}}(p_t)^{u^*_{4,2,t}}(1-p_t)^{u_{2,1,t}-u^*_{4,2,t}} \tag{4.3.2d}$$

Due to the survival and capture processes both being specified as binomial functions it can also be noted that

$$p(u^*_{2,1,t}|n_{1,t-1}) = \left[\binom{n_{1,t-1}}{u^*_{2,2,t}}(\phi_t p_t)^{u^*_{2,2,t}}(1-\phi_t p_t)^{n_{1,t-1}-u^*_{2,2,t}}\right]$$

$$p(u^*_{4,1,t}|n_{2,t-1}) = \left[\binom{n_{2,t-1}}{u^*_{4,2,t}}(\phi_t p_t)^{u^*_{4,2,t}}(1-\phi_t p_t)^{n_{2,t-1}-u^*_{4,2,t}}\right]$$

Then, the distribution of $u_{1,1,t}$ conditional on the known value $n_{1,t-1}$ and the just simulated $u^*_{2,2,t}$ is given by:

$$p(u_{1,1,t}|u^*_{2,2,t}, n_{1,t-1}) = \frac{p(u^*_{2,2,t}|u_{1,1,t})p(u_{1,1,t}|n_{1,t})}{\sum_{u^*_{2,2,t}\leq u'\leq n_{1,t-1}} p(u^*_{2,2,t}|u')p(u'|n_{1,t-1})} \tag{4.3.3}$$

Once the value $u^*_{1,1,t}$ has been simulated, the value $u^*_{1,2,t}$ is then obtained deterministically:

$$u^*_{1,2,t} = u^*_{1,1,t} - u^*_{2,2,t}.$$

A similar process yields the conditional distribution of $u^*_{2,1,t}$ conditional on the known

value $n_{2,t-1}$ and the just simulated $u^*_{4,2,t}$.

Then, the component of the weights contributed by the state process pdf $g(\mathbf{n}^*_t|\mathbf{n}_{t-1})$ is given by the product of the binomials in Eq's. (4.3.2a)-(4.3.2d):

$$\begin{aligned} g(\mathbf{n}^*_t|\mathbf{n}_{t-1}) &= \binom{n_{1,t-1}}{u^*_{1,1,t}}(\phi_t)^{u^*_{1,1,t}}(1-\phi_t)^{n_{1,t-1}-u^*_{1,1,t}} && (4.3.4)\\ &\times\binom{n_{2,t-1}}{u^*_{2,1,t}}(\phi_t)^{u^*_{2,1,t}}(1-\phi_t)^{n_{2,t-1}-u^*_{2,1,t}} \\ &\times\binom{u^*_{1,1,t}}{u^*_{2,2,t}}(p_t)^{u^*_{2,2,t}}(1-p_t)^{u^*_{1,1,t}-u^*_{2,2,t}} && (4.3.5)\\ &\times\binom{u^*_{2,1,t}}{u^*_{4,2,t}}(p_t)^{u^*_{4,2,t}}(1-p_t)^{u^*_{2,1,t}-u^*_{4,2,t}} && (4.3.6) \end{aligned}$$

and the component of the weights from the trial density pdf $h(\mathbf{n}^*_t|\mathbf{y}_t,\mathbf{n}_{t-1})$ are:

$$\begin{aligned} h(\mathbf{n}^*_t|\mathbf{y}_t,\mathbf{n}_{t-1}) &= \binom{y_{1,t}}{u^*_{2,2,t}}(r_{2,2,t})^{u^*_{2,2,t}}(1-r_{2,2,t})^{y_{1,t}-u^*_{2,2,t}} \\ &\times\frac{p(u^*_{2,2,t}|u_{1,1,t})p(u_{1,1,t}|n_{1,t})}{\sum_{u^*_{2,2,t}\le u'\le n_{1,t}}p(u^*_{2,2,t}|u')p(u'|n_{1,t})} \\ &\times\frac{p(u^*_{4,2,t}|u_{2,1,t})p(u_{2,1,t}|n_{2,t})}{\sum_{u^*_{4,2,t}\le u'\le n_{2,t}}p(u^*_{4,2,t}|u')p(u'|n_{2,t})} && (4.3.7) \end{aligned}$$

Hence, using the expressions in Eq's. (4.3.6) and (4.3.7), the weights for the entire vector $\mathbf{n}_t$ are evaluated as in Eq. (4.3.1b):

$$w \propto \frac{g_t(\mathbf{n}^*_t|\mathbf{n}_{t-1})}{h(\mathbf{n}^*_t|\mathbf{y}_t,\mathbf{n}_{t-1})}.$$

In general, evaluating the weights at each time step will consist of two main steps. Firstly, in the event of convolutions, observed totals need to be split across the appropriate elements of the expanded state vector using rates determined by ratios of expected values. Secondly, the remaining unknown elements of the expanded state vector are simulated from trial densities that are functions of the multinomial (often binomial) distributions

assumed for the sub-processes. These trial densities are conditional on known values from data and the values simulated from the de-convolving splitting process. At each time step t, every term that is evaluated under the proposed trial density $h_t()$ should also be evaluated under the state process pdf $g_t()$. The following section describes a model consisting of multiple sub-processes such that the direct evaluation of $g_t(\mathbf{n}_t|\mathbf{n}_{t-1})$ is difficult and contains convolutions. A fitting algorithm is developed and approaches to simulating the expanded state vector using appropriate conditional distributions are suggested.

4.4 An Application to Models for Capture-Recapture Data

The notation and structure of this model is similar to that developed in Section 3.3 of Buckland *et al.* (2004). A model for a capture-recapture survey is constructed with the population classified into two age-cohorts: juveniles and adults. Animals are classified as adults when they reach age 1 and become sexually mature. To simplify the model structure, only the female population will be modelled but the methods are readily extensible to include males.

It is assumed that the general open-population capture-recapture processes are present in the model: survival, capture and recruitment. To keep the model relatively simple it is assumed that there is no migration between the population area and any external area. Therefore, survival relates solely to animals that die rather than including emigrants and, more importantly in view of a constrained model, the term births relates solely to animals that are born to surviving adults in the population and does not include immigrants.

The survival process is assumed to be binomial for both juveniles and adults. The

parameters ϕ_j and ϕ_a denote, for juveniles and adults respectively, the probabilities of an individual animal surviving from the beginning of one time period t to immediately before the beginning of the next time period $t+1$. It is assumed that this juvenile survival probability is constant for all juveniles across all time periods and the adult survival probability is also assumed constant across all adults and all time periods, although extending the model to time-specific survival $(\phi_{j,t}, \phi_{a,t})$ is straightforward.

The birth process is also modelled as a binomial density with the parameter π denoting the probability of a surviving animal giving birth to a single offspring. The assumption that a breeding animal is only able to produce, at most, a single young may well be unrealistic for some animal populations. However, this assumption does simplify the exposition of the algorithm implementation and should serve to make the mechanics of this example more transparent.

Following the birth process a single sample is taken as part of the capture-recapture study. Juvenile animals are assumed to be distinguishable from adults when captured. Once an animal matures it is assumed that its age cannot be identified on capture if it has not been previously captured as a juvenile with known period of birth. In year t juveniles are caught with probability p_j and adults are caught with probability p_a. There is no t subscript to denote time dependence in the capture rates as it is assumed that the juvenile capture probability is constant across all juveniles and all time periods and, similarly, that the adult capture probability is assumed constant across all adults and all time periods. It is also assumed that there is no mortality between the birth process and the capture process (although a sub-process could be incorporated to model this if it was thought to be viable biologically) and that there are no losses on capture. An animal that is alive during the birth process will still be alive and available for capture during the sampling

period. As with survival and births, capture is modelled using binomial densities for both juvenile and adult capture. From the parameters defined for these processes the vector of model parameters to be estimated is then $\boldsymbol{\theta} = \{\phi_j, \phi_a, \pi, p_j, p_a\}$.

4.4.1 State and Observation Vector

In the conditional approach to modelling capture-recapture data, as defined in Section 4.2, the elements of the state vector $\mathbf{n}_t$ correspond to the capture-history patterns that are possible at time t. Therefore, $\mathbf{n}_t$ consists of the numbers of animals (abundances) that exhibit each possible capture-history pattern. In classical capture-recapture analyses the capture-history patterns for animals that were not caught in the current time period make no distinction between animals that were alive but remained undetected and those that were dead and were unavailable to be sampled. However since survival is modelled as a component of the processes that generate the state vector $\mathbf{n}_t$, only animals that are alive at t will be included in the elements of $\mathbf{n}_t$. For example, the element $n_{100,3}$ would correspond to the animals that were caught on the first capture occasion and survived to the current year but were uncaught in the last two years.

For animals grouped into a single state there are 2^t possible capture histories. However, this is conceptually a multistate model in which the animals are classified as either juveniles or adults. Therefore, if the distinction between animals first caught as juveniles and those first caught as adults is maintained, animals in each of these two groups can exhibit any of the observed capture history patterns and the state vector $\mathbf{n}_t$ will then contain $2 * 2^t$ elements. For this model a reduced version of this full state vector $\mathbf{n}_t$ is used. The distinction is only maintained for one year after initial capture between those animals that are first captured as juveniles and those that are first captured as adults to make the

parameter controlling juvenile survival, ϕ_j, identifiable. As discussed in Section 2.2.1.2 the identifiability of parameters in a classical capture-recapture analysis depends on the information contained in the data. To estimate survival rates for a particular cohort both initial captures and future recaptures of that cohort need to be recorded. Hence, the number of juveniles captured during the sampling period for time t and the number of these that are recaptured at $t+1$ both need to be known to estimate ϕ_j. The marked juveniles in time t that survive through until the capture process in time $t+1$ are labelled as yearlings. The unmarked juveniles in time t that survive until $t+1$ are not observed and therefore do not contribute to the estimate of ϕ_j; they are labelled as adults in $t+1$. As a notational convention the subscripts j, y and a correspond to juveniles, yearlings and adults respectively.

With these definitions in place the initial states $\mathbf{n}_0$ and the state vectors for the first three time periods are represented in Figure 4.3.

$$\mathbf{n}_0 = \begin{bmatrix} n_{j,0} \\ n_{a,0} \end{bmatrix} \rightarrow \mathbf{n}_1 = \begin{bmatrix} n_{j0,1} \\ n_{j1,1} \\ n_{a0,1} \\ n_{a1,1} \end{bmatrix} \rightarrow \mathbf{n}_2 = \begin{bmatrix} n_{j0,2} \\ n_{j1,2} \\ n_{y10,2} \\ n_{y11,2} \\ n_{a00,2} \\ n_{a01,2} \\ n_{a10,2} \\ n_{a11,2} \end{bmatrix} \rightarrow \mathbf{n}_3 = \begin{bmatrix} n_{j0,3} \\ n_{j1,3} \\ n_{y10,3} \\ n_{y11,3} \\ n_{a000,3} \\ n_{a001,3} \\ n_{a010,3} \\ n_{a100,3} \\ n_{a011,3} \\ n_{a101,3} \\ n_{a110,3} \\ n_{a111,3} \end{bmatrix}$$

Figure 4.3: The elements of the state vectors $\mathbf{n}_t$ for $t = 0, 1, 2, 3$.

The subscripts indicate the age class of the animal, j, y or a, the capture history pattern and the time t. The length of the capture histories is determined by the known length of time an animal has been in the population. For example, animals born in time t will not have been available for capture prior to the sampling occasion on time t and thus only exhibit a capture history pattern with entries corresponding to the years in which they were alive. This distinction no longer holds once an animal has matured and is classified as an adult which occurs at age 1 for unmarked juveniles and at age 2 for yearlings. Biologically there is no distinction between adults and yearlings; yearlings are capable of breeding and are considered to be mature adults. The distinction is maintained solely for modelling purposes to allow the estimation of juvenile survival. For example, at time $t = 3$ the vector $\mathbf{n}_t$ contains the element $n_{y10,3}$ which corresponds to animals that were born and captured during the second time period, survived through to the third time period but were not captured. In the fourth time period surviving animals from $n_{y10,3}$ will be assigned to either $n_{a0100,4}$ or $n_{a0101,4}$ if they are not caught or if they are caught respectively. Equally, surviving animals from $n_{a010,3}$ will be assigned to one of those two elements dependent on capture status during the fourth time period. Therefore both $n_{a0100,4}$ and $n_{a0101,4}$ represent fates that can be attained by animals following two different progression paths through the model. In general, elements of $\mathbf{n}_t$ correspond to sums over the surviving numbers of animals from multiple elements of $\mathbf{n}_{t-1}$. The initial state vector consists of juveniles and adults that have yet to be marked.

The observation vector $\mathbf{y}_t$ consists of the numbers of animals that are observed to display a capture history pattern that includes capture in sampling period t. As such, it represents a subset of the elements in $\mathbf{n}_t$. The observation vectors for the first three time periods are given in Figure 4.4.

$$\mathbf{y}_1 = \begin{bmatrix} n_{j1,1} \\ n_{a1,1} \end{bmatrix} \rightarrow \mathbf{y}_2 = \begin{bmatrix} n_{j1,2} \\ n_{y11,2} \\ n_{a01,2} \\ n_{a11,2} \end{bmatrix} \rightarrow \mathbf{y}_3 = \begin{bmatrix} n_{j1,3} \\ n_{y11,3} \\ n_{a001,3} \\ n_{a011,3} \\ n_{a101,3} \\ n_{a111,3} \end{bmatrix}$$

Figure 4.4: The observation vectors $\mathbf{y}_t$ for $t = 1, 2, 3$

From Figure 4.4 it can be seen that no animals with capture-history patterns ending in 0 are included. These elements do exist in the state vectors and, under the conditional approach, it is these unobserved elements that induce stochasticity in the modelling framework.

The sub-processes assumed to drive the underlying dynamics of the model are survival, birth and capture. The birth process also coincides with the deterministic maturation process which advances all surviving juveniles from the previous time period to be adults or yearlings depending on their capture status in the previous time period. Using the basic structure discussed in Section 3.3.4 the sub-processes are assumed to be discrete and generate $\mathbf{n}_t$ from $\mathbf{n}_{t-1}$ by acting sequentially with each sub-process taking as its input the output from the preceding sub-process. It is assumed that a period begins immediately after the capture-recapture sample from the previous time period has been taken. Then the order in which the sub-processes are assumed to be applied is survival first followed by birth and maturation followed by capture. The definitions of the intermediate vectors are as follows:

$\mathbf{u}_{1,t}$ = the number of animals in $\mathbf{n}_{t-1}$ that survive through until time t.

$\mathbf{u}_{2,t}$ = the number of juveniles produced by the surviving animals $\mathbf{u}_{1,t}$ and the number, after age incrementation, of adults in $\mathbf{u}_{1,t}$.

$\mathbf{u}_{3,t}$ = the numbers of animals in $\mathbf{u}_{2,t}$ that are marked or recaptured in time t.

It is assumed that the state process at time t, g_t, can be expressed as a series of linked probability functions, each of which corresponds to a specified sub-process, with $g_{r,t}$ denoting the probability function for the r^{th} sub-process. For this model, $g_{1,t}$ corresponds to survival, $g_{2,t}$ corresponds to birth and $g_{3,t}$ corresponds to capture. If the distribution $H_{r,t}$ then corresponds to the sub-process pdf $g_{r,t}$ the evolution of the states can be expressed as follows:

$$\mathbf{u}_{1,t} \sim H_{1,t}(\mathbf{n}_{t-1}, (\phi_j, \phi_a)), \mathbf{u}_{2,t} \sim H_{2,t}(\mathbf{u}_{1,t}, \pi), \mathbf{u}_{3,t} \sim H_{3,t}(\mathbf{u}_{2,t}, (p_j, p_a))$$

The state and intermediate vectors for the evolution from the initial states $\mathbf{n}_0$ to the states in the system immediately after the capture sub-process are given in Figure 4.5.

$$\mathbf{n}_0 = \begin{bmatrix} n_{j,0} \\ n_{a,0} \end{bmatrix} \xrightarrow{g_{1,1}} \mathbf{u}_{1,1} = \begin{bmatrix} u_{1(1),1} \\ u_{2(1),1} \end{bmatrix} \xrightarrow{g_{2,1}} \mathbf{u}_{2,1} = \begin{bmatrix} u_{1(2),1} \\ u_{1(1),1} \\ u_{2(2),1} \\ u_{2(1),1} \end{bmatrix} \xrightarrow{g_{3,1}} \mathbf{u}_{3,1} = \begin{bmatrix} u_{1(3j0),1} \\ u_{1(3j1),1} \\ u_{1(3a0),1} \\ u_{1(3a1),1} \\ u_{2(3j0),1} \\ u_{2(3j1),1} \\ u_{2(3a0),1} \\ u_{2(3a1),1} \end{bmatrix}$$

Figure 4.5: Evolution of the initial states $\mathbf{n}_0$ via the intermediate sub-processes.

The subscripts in the intermediate state elements contain three pieces of information.

The first value indicates the "parent node" (see Section 4.3.1) associated with that element. The first number in the parentheses denotes which of the sub-process vectors the element is a component of and, if it is present, the two character alpha-numeric label indicates age and capture status. The final value following the comma indicates the time period. Using the directed graph approach introduced in Section 4.3.1 a process tree illustrating the evolution of this system from the initial states $\mathbf{n}_0$ is produced in Figure 4.6.

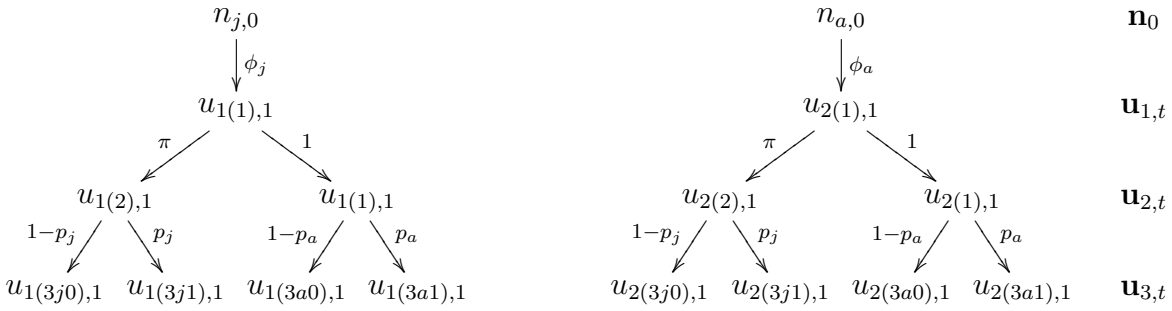


Figure 4.6: Process Tree diagram for the evolution from $\mathbf{n}_0$ via the sub-processes for the capture-recapture study.

In Figure 4.6 the parameters next to the branches indicate the rate parameters of the binomial distributions that have been specified for the survival, birth and capture processes. The process tree diagram gives a clear indication of the dynamics of the system. It should be noted that in the representation of the intermediate vector $\mathbf{u}_{2,1}$, which denotes the state of the system after both the survival process and the birth and maturation process have occurred, the second and fourth elements correspond to the number of survivors from $t = 0$ that are classified as adults at $t = 1$. For the intermediate states in general time period t, there is no mortality during the birth process and age incrementation is a deterministic process so all parent nodes that survive from time t, denoted by $\mathbf{u}_{1,t+1}$, will be classified as adults at time $t + 1$ and will constitute half of

the elements of $\mathbf{u}_{2,t+1}$ with the other half denoting new juveniles born to the survivors. An alternative representation of $\mathbf{u}_{2,1}$ with subscripts that perhaps make this relationship more obvious is:

$$\mathbf{u}_{2,1} = \begin{bmatrix} u_{1(2j),1} = u_{1(2),1} \\ u_{1(2a),1} \equiv u_{1(1),1} \\ u_{2(2j),1} = u_{2(2),1} \\ u_{2(2a),1} \equiv u_{2(1),1} \end{bmatrix}$$

An explanation of the notation in Figures 4.5 and 4.6 along with the interpretation of each component of the intermediate vectors is given in Table 4.1. It should be noted

Element		Description
$n_{j,0}$	=	The initial number of juvenile animals in the population at the start of the study. These animals are all unmarked.
$u_{1(1),1}$	=	The number of initial juveniles that survive from $t = 0$.
$u_{1(2),1}$	=	The number of new juveniles produced by the initial juveniles that have survived from $t = 0$.
$u_{1(3j0),1}$	=	The number of new juveniles that are not caught during the sampling period at $t = 1$.
$u_{1(3j1),1}$	=	The number of new juveniles that are caught during the sampling period at $t = 1$.
$u_{1(3a0),1}$	=	The number of animals that were initial juveniles at $t = 0$ but have survived and have not been caught during the sampling period at $t = 1$.
$u_{1(3a1),1}$	=	The number of animals that were initial juveniles at $t = 0$ but have survived and have been caught during the sampling period at $t = 1$.

Table 4.1: Definitions of the notation for the evolution of initial juveniles from $\mathbf{n}_0$ via the three sub-processes of survival, birth and capture. An intermediate node element is denoted as $u_{\alpha(\beta),\gamma}$ where α indexes the parent node, β indicates the sub-process (1 = survival, 2 = birth, 3 = capture) and γ denotes the time period.

that $\mathbf{n}_t \neq \mathbf{u_{3,t}}$. Elements of $\mathbf{n}_t$ represent sums across the elements of $\mathbf{u}_{3,t}$ and therefore an appropriate mapping of $\mathbf{u}_{3,t} \rightarrow \mathbf{n}_t$ needs to be determined. For example, each parent node produces a child node $u_{.(3j1),t}$ that represents the number of new juveniles that are marked in t having been born to the surviving members of the parent node at $t-1$. The

observed number of marked juveniles at time t is given by $y_{1,t} = n_{j1,t}$. When recorded in the capture-history data these marked juveniles are captured independently from their parents so the ancestry of these new juveniles is unknown. Therefore, the value of $n_{j1,t}$ is obtained by summing across all child nodes that correspond to marked juveniles. An equivalent sum across all unmarked juvenile nodes is required to obtain $n_{j0,t}$. For this conditional system, and for $t \geq 2$, there will be eight elements of $\mathbf{n}_t$ that involve summing over multiple elements of $\mathbf{u}_{3,t}$. As can be seen from Figure 4.3, for $t \geq 2$ the state vector $\mathbf{n}_t$ will consist of elements corresponding to unmarked and marked juveniles, to uncaught and caught yearlings and to 2^t distinct capture-history patterns. Thus, $\mathbf{n}_t$ will consist of $2^t + 4$ elements for $t \geq 2$. The summations arise because animals that are designated as juveniles or yearlings at $t-1$ will, if they survive, be designated as adults at t and will share common capture history patterns with surviving animals that were designated as adults at $t-1$. In addition to the sums for the unmarked and marked juveniles, for $t > 2$, the other three sets of sums occur for the following capture-history patterns:

1. Animals from both $n_{j0,t-1}$ and $n_{a\{q_1\},t-1}$ will contribute to both $n_{a\{q_1\}0,t}$ and $n_{a\{q_1\}1,t}$, where the pattern $\{q_1\}$ is a string of $t-1$ zeros.

2. Animals from both $n_{y10,t-1}$ and $n_{a\{q_2\}10,t-1}$ will contribute to both $n_{a\{q_2\}100,t}$ and $n_{a\{q_2\}101,t}$ where the pattern $\{q_2\}$ is a string of $t-3$ zeros.

3. Animals from both $n_{y11,t-1}$ and $n_{a\{q_2\}11,t-1}$ will contribute to both $n_{a\{q_2\}110,t}$ and $n_{a\{q_2\}111,t}$ where the pattern $\{q_2\}$ is a string of $t-3$ zeros.

Hence, for the progression from $t = 0$ to $t = 1$ the state vector $\mathbf{n}_1$ is obtained from the intermediate vector $\mathbf{u}_{3,1}$ representing the states in the system immediately after the

capture process using the following collapsing structure:

$$\mathbf{n}_1 = \begin{bmatrix} n_{j0,1} \\ n_{j1,1} \\ n_{a0,1} \\ n_{a1,1} \end{bmatrix} = \begin{bmatrix} u_{1(3j0),1} + u_{2(3j0),1} \\ u_{1(3j1),1} + u_{2(3j1),1} \\ u_{1(3a0),1} + u_{2(3a0),1} \\ u_{1(3a1),1} + u_{2(3a1),1} \end{bmatrix}$$

Using the notational definitions in Table 4.1 and the state vector and intermediate state vector structure represented in Figures 4.3 and 4.5 the deterministic collapsing structure to obtain $\mathbf{n}_2$ from $\mathbf{u}_{3,2}$is represented in Figure 4.7.

$$\mathbf{n}_2 = \begin{bmatrix} n_{j0,2} \\ n_{j1,2} \\ n_{y10,2} \\ n_{y11,2} \\ n_{a00,2} \\ n_{a10,2} \\ n_{a01,2} \\ n_{a11,2} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{4} u_{j(3j0),2} \\ \sum_{j=1}^{4} u_{j(3j1),2} \\ u_{2(3j0),2} \\ u_{2(3j1),2} \\ u_{1(3a0),2} + u_{3(3a0),2} \\ u_{1(3a1),2} + u_{3(3a1),2} \\ u_{4(3a0),2} \\ u_{4(3a1),2} \end{bmatrix}$$

Figure 4.7: Relationship between the state vector $\mathbf{n}_2$ and the intermediate states after capture at time $t = 2$: $\mathbf{u}_{3,2}$.

Therefore, Figure 4.7 contains the expression for the eight summations that constitute the elements of $\mathbf{n}_2$. Those elements in $\mathbf{n}_2$ that do not comprise the sum of multiple components correspond to the uncaught and caught yearlings at $t = 2$. These animals are only generated by marked juveniles at time $t = 1$; no adult parent node can contribute to the yearlings so there is no convolution when determining the ancestry of yearlings at $t = 2$. This holds for all time periods $t = 2, 3, \ldots, T$. In general, for $t > 2$, any capture history pattern (consisting of t elements) that includes a 1 anywhere in the first $t - 3$ entries can only have been generated by a single element of the state vector at $t - 1$.

These state elements will be referred to as *mono-parent* elements whereas the eight state elements at time t that are obtained from summing over multiple elements of the intermediate state vector $\mathbf{u}_{3,t}$ are *multi-parent* elements. The evaluation of the trial density for simulated abundances for these capture-history patterns corresponding to *mono-parent* elements does not involve convolutions.

Let $\mathbf{n}_t^{[j]}$ represent the set of elements in the state vector $\mathbf{n}_t$ that correspond to the abundances of juvenile animals at time t. Similarly, let $\mathbf{n}_t^{[a]}$ represent the set of elements in the state vector $\mathbf{n}_t$ that correspond to the abundances of adult animals, including yearlings, at time t. Using this superscript notation the distributions of the sub-processes can then be written as:

$$u_{1,t,[j]}|n_{t-1,[j]} \sim \text{Binomial}(n_{t-1,[j]}, \phi_j) \tag{4.4.1a}$$

$$u_{1,t,[a]}|n_{t-1,[a]} \sim \text{Binomial}(n_{t-1,[a]}, \phi_a) \tag{4.4.1b}$$

$$u_{2,t,[j]}|u_{1,t} \sim \text{Binomial}(u_{1,t}, \pi) \tag{4.4.1c}$$

$$u_{2,t,[a]}|u_{1,t} = u_{1,t} \tag{4.4.1d}$$

$$u_{3j1,t}|u_{2,t,[j]} \sim \text{Binomial}(u_{2,t,[j]}, p_j) \tag{4.4.1e}$$

$$u_{3a1,t}|u_{2,t,[a]} \sim \text{Binomial}(u_{2,t,[a]}, p_a) \tag{4.4.1f}$$

Then, from Eq. (4.4.1e), the deterministic relationship $u_{3j0,t} = u_{2,t,[j]} - u_{3j1,t}$ yields the values for the components of $\mathbf{u}_{3,t}$ corresponding to unmarked juveniles. Similarly, from Eq. (4.4.1f), the deterministic relationship $u_{3a0,t} = u_{2,t,[a]} - u_{3a1,t}$ yields the values for the components of $\mathbf{u}_{3,t}$ corresponding to adults that are not captured during the sampling period at t.

As discussed in Section 4.3.1 the generalised Leslie matrix expansion provides two

useful tools for fitting these conditional models to data. Firstly, they have utility for choosing a trial density $h()$ and an associated expanded state vector to make the evaluation more tractable. Secondly, the de-convolution of components of the evaluation process can require splitting observed totals across multiple elements of the expanded state vector. Taking the product of the individual matrix representations of the sub-processes produces the (approximate) conditional expectations of the elements of the expanded state vector. These approximate expectations can be used to determine appropriate proposed splitting rates for the observed totals.

In terms of matrix representations let S_t denote the survival process matrix, B_t the birth and maturation process matrix and C_t the capture process matrix where t denotes the process acting from time $t-1$ to t. Then, the approximate conditional expectations can be expressed as $E_{\mathbf{n}_t|\mathbf{n}_{t-1}}[\mathbf{n}_t] = C_t B_t S_t \mathbf{n}_{t-1}$. For $t = 1$ the matrix representations are:

$$S_1 = \begin{bmatrix} \phi_j & 0 \\ 0 & \phi_a \end{bmatrix} \quad B_1 = \begin{bmatrix} \pi & 0 \\ 0 & 1 \\ \pi & 0 \\ 0 & 1 \end{bmatrix} \quad C_1 = \begin{bmatrix} 1-p_j & 0 & 1-p_j & 0 \\ p_j & 0 & p_j & 0 \\ 0 & 1-p_a & 0 & 1-p_a \\ 0 & p_a & 0 & p_a \end{bmatrix}$$

The matrix representation of C_1 incorporates both the capture process and the collapsing of the intermediate vector $\mathbf{u}_{3,1}$ to obtain $\mathbf{n}_{t-1}$. As discussed in Section 4.3.1, expanding the matrices to have at most one non-zero entry per row helps to avoid convolutions in the evaluation of the state pdf and trial pdf densities. The specification of C_1 will lead to

convolutions and an obvious expansion yields:

$$C_{ex,1} = \begin{bmatrix} 1-p_j & 0 & 0 & 0 \\ p_j & 0 & 0 & 0 \\ 0 & 1-p_a & 0 & 0 \\ 0 & p_a & 0 & 0 \\ 0 & 0 & 1-p_j & 0 \\ 0 & 0 & p_j & 0 \\ 0 & 0 & 0 & 1-p_a \\ 0 & 0 & 0 & p_a \end{bmatrix}$$

Using this expanded matrix the approximate conditional expected values of the expanded state vector are

$$E_{\mathbf{n}_1^{ex}|\mathbf{n}_0}[\mathbf{n}_1^{ex}] = C_{ex,1}B_1S_1\mathbf{n}_0 = \begin{bmatrix} (1-p_j)\pi\phi_j n_{j,0} \\ p_j\pi\phi_j n_{j,0} \\ (1-p_a)\phi_j n_{j,0} \\ p_a\phi_j n_{j,0} \\ (1-p_j)\pi\phi_a n_{a,0} \\ p_j\pi\phi_a n_{a,0} \\ (1-p_a)\phi_a n_{a,0} \\ p_a\phi_a n_{a,0} \end{bmatrix}$$

where $\mathbf{n}_1^{ex}$ denotes the expanded state vector. It should be noted that for this particular time step $\mathbf{n}_1^{ex} \equiv \mathbf{u}_{3,1}$. The same form of expansion can be applied to C_t for all time steps $t \in (1, T)$ to yield an expanded state vector. As mentioned previously, the convolutions only arise for *multi-parent* elements of $\mathbf{n}_t$. The state vector $\mathbf{n}_{t-1}$ will contain $2^{t-1} + 4$ elements which will generate an intermediate vector $\mathbf{u}_{3,t}$ containing $2^{t+1} + 16$ elements. Of these elements in $\mathbf{u}_{3,t}$, $2^{t-1} + 4$ of them will be summed to give $n_{j0,t}$ and another $2^{t-1} + 4$ of them will be summed to give $n_{j1,t}$. Also, two elements each will be summed for each of the other six sums, meaning another 12 elements of $\mathbf{u}_{3,t}$ are involved in the

summations. Therefore, a total of $2^t + 20$ elements of $\mathbf{u}_{3,t}$ are used for the sums required by the mapping from $\mathbf{u}_{3,t}$ to the *multi-parent* elements of $\mathbf{n}_t$. The remaining $2^t - 4$ elements of $\mathbf{u}_{3,t}$ represent *mono-parent* elements of $\mathbf{n}_t$. This leads to a heuristic approach to determining the expanded state vector for this particular capture-recapture population dynamics model. The expanded state vector should consist of the $2^t + 20$ elements of $\mathbf{u}_{3,t}$ that are components of sums and the $2^t - 4$ elements of $\mathbf{n}_t$ that are *mono-parent* nodes. This is equivalent to the intermediate vector $\mathbf{u}_{3,t}$ and is exactly the extended state vector that is obtained using the generalised Leslie Matrix representation of the model with the expanded capture matrix $C_{ex,t}$.

4.4.2 Fitting the Model

Having developed a procedure for determining the appropriate form for the extended state vector it is now necessary to define a trial density $h()$ from which simulated state vectors $\mathbf{n}_t^*$ are drawn that obey all constraints imposed by conditioning on the entire series of known state elements, $\mathbf{s}_{1:T}$. As discussed in Section 4.3, any simulated state vectors that fail to obey the constraints will have weight zero in the importance sampling step of model inference. Expressing this formally, from Eq. (4.3.1a) the weights for the simulated state vectors are given as

$$w \propto \frac{f(\mathbf{y}_t|\mathbf{n}_t^*)g_t(\mathbf{n}_t^*|\mathbf{n}_{t-1})}{h(\mathbf{n}_t^*|\mathbf{y}_t, \mathbf{n}_{t-1})} \tag{4.4.2}$$

where

$$f_t(\mathbf{y}_t|\mathbf{n}_t^*) = f_t(\mathbf{y}_t|\mathbf{s}_t^*) = \begin{cases} 1 & \text{if } \mathbf{s}_t^* = \mathbf{y}_t \\ 0 & \text{otherwise} \end{cases} .$$

Therefore, the aim of choosing an appropriate trial density $h_t()$ is to ensure that $f_t(\mathbf{y}_t|\mathbf{n}_t^*) = 1$. As will be discussed in more detail later on in this section, the specification of a model that generates expanded state vectors which satisfy every single constraint imposed by the

entire set of observed capture-histories and the model structure is prohibitively complex and laborious for a non-trivial model. The aim of choosing $h_t()$ is now to increase the probability that $f_t(\mathbf{y}_t|\mathbf{n}_t^*) = 1$.

Using the heuristic approach introduced in 4.3.1 the trial density $h_t()$ will consist of functions of binomial and multinomial distributions to model the constrained evolution of the system from $t-1$ to t. For example, a multinomial will be used to split the observed total of marked juveniles at time t across the appropriate elements of the expanded state vector $\mathbf{n}_t^*$ with the splitting rates determined by the ratios of the expected values $E[\mathbf{u}_{i(3j1),t}]$ for $i = 1, 2, \ldots, 2^{t-1} + 4$. Binomial splitting functions will be used to split the *multi-parent* elements of $\mathbf{n}_t$ across the appropriate elements of $\mathbf{n}_t^*$. Elements in the intermediate vectors can then be obtained using deterministic processes conditional on the simulated splits and observed data or they can be drawn stochastically using constrained conditional distributions that are functions of the binomials specified in Eqs. (4.4.1a)-(4.4.1f).

Under the sequential approach to model fitting discussed in Section 3.3.6 the initial step typically involves simulating the initial states from their prior distribution (Eq. (3.3.1a)) $g_0(\mathbf{n}_0|\Theta)$. Alternatively they can be assumed to be known. The directed process tree

diagram in Figure 4.6 illustrates the progression of the model from $\mathbf{n}_0$ to $\mathbf{n}_1$. From Section 4.4.1 the expanded state vector is equivalent to $\mathbf{u}_{3,1}$ and

$$E_{\mathbf{u}_{3,1}|\mathbf{n}_0}[\mathbf{u}_{3,1}] = \begin{bmatrix} (1-p_j)\pi\phi_j n_{j,0} \\ p_j\pi\phi_j n_{j,0} \\ (1-p_a)\phi_j n_{j,0} \\ p_a\phi_j n_{j,0} \\ (1-p_j)\pi\phi_a n_{a,0} \\ p_j\pi\phi_a n_{a,0} \\ (1-p_a)\phi_a n_{a,0} \\ p_a\phi_a n_{a,0} \end{bmatrix}$$

The observation $y_{1,1} = n_{j1,1}$ needs to be split across $u_{13j1,1}$ and $u_{23j1,1}$. A binomial splitting function is used to simulate the values. As for the example in Section 4.3.1, the rate parameter is determined using the ratio of the approximate expected values of the appropriate expanded state vector elements:

$$u^*_{1(3j1),1}|n_{j1,1} \sim \text{Binomial}(n_{j1,1}, \frac{\phi_j n_{j,0}}{\phi_j n_{j,0} + \phi_a n_{a,0}})$$

with $u_{2(3j1),1} = n_{j1,1} - u_{1(3j1),1}$. Similarly, the observation $y_{2,1} = n_{a1,1}$ needs to be split across $u_{13a1,1}$ and $u_{23a1,1}$:

$$u^*_{1(3a1),1}|n_{a1,1} \sim \text{Binomial}(n_{a1,1}, \frac{\phi_j n_{j,0}}{\phi_j n_{j,0} + \phi_a n_{a,0}})$$

with $u_{2(3a1),1} = n_{a1,1} - u_{1(3a1),1}$. The values in Figure 4.8 in green font denote elements that are assumed known or have been simulated from the trial density pdf $h_1()$. Unknown values that are still to be simulated are denoted in plain black font. The current structure of the process tree is one that reoccurs frequently throughout the model and a general algorithm can be developed to generate the unobserved elements in trees demonstrating this structure. The key features of this structure for a process tree generated from the i^{th}

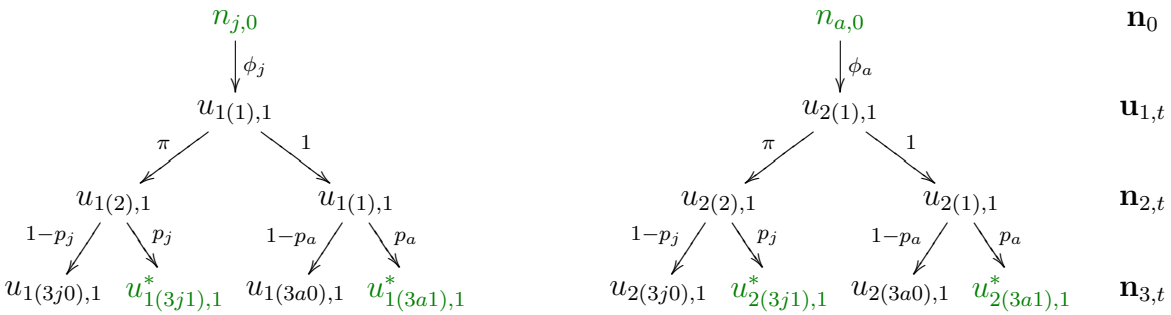


Figure 4.8: Process Tree diagram for the evolution from $\mathbf{n}_0$ after splitting over convolutions. "Known" values denoted in green font.

$(i = 1, 2, \ldots, 2^t + 4)$ element of the state vector at time $t - 1$ are:

1. The abundance of the parent node (an element of $n_{i,t-1}$) is assumed known.

2. The abundance of marked juveniles in the i^{th} tree $u_{i(3j1),t}$ has been simulated via splitting.

3. The abundance of captured adults in the i^{th} tree $u_{i(3a1),t}$ is assumed known.

4.4.3 The Conditional Generation Approach Algorithm

A general procedure, which shall be referred to as the conditional generation approach (CGA) algorithm, can now be specified. Again, for the process tree generated from the i^{th} element of the state vector at time $t - 1$ there are two objectives under the CGA algorithm: firstly, to stochastically simulate a value for $u_{i(1),t}$ subject to the constraints imposed by $n_{i,t-1}$, $u^*_{i(3j1),t}$ and $u^*_{i(3a1),1}$ and secondly, to stochastically simulate a value for $u_{i(2j),t}$ subject to the constraints imposed by $u^*_{i(1),t}$ and $u^*_{i(3j1),t}$. The dependence on the parameters $\boldsymbol{\Theta}$ is implicit in the following distributions to reduce the notation.

Simulating survivors

The first objective is to simulate $u_{i(1),t}$, the number of animals from $n_{i,t-1}$ that survive until the end of time period $t = 1$. To reduce notation define $v = n_{i,t-1}$, $w = u_{i(1),t}$,

$x = u_{i(2j),t}$, $y = u_{i(3j1),t}$, and $z = u_{i(3a1),t}$. Also, assume that $n_{i,t-1}$ represents juvenile animals born in time $t-1$. Then, the distributions are:

$$w|v \sim \text{Binomial}(v, \phi_j) \tag{4.4.3}$$

$$z|w \sim \text{Binomial}(w, p_a) \tag{4.4.4}$$

$$x|w \sim \text{Binomial}(w, \pi) \tag{4.4.5}$$

$$y|x \sim \text{Binomial}(x, p_j) \tag{4.4.6}$$

$$y|w \sim \text{Binomial}(w, \pi p_j) \tag{4.4.7}$$

For the entire tree generated from v, the unknown elements of interest are w and x, whilst the elements with values assumed to be known (from simulation) are v, y and z. Then, the conditional joint distribution of w and x given y, z and v is

$$\begin{aligned} p(w,x|y,z,v) &= \frac{p(w,x,y,z,v)}{p(y,z)} = \frac{p(y,z|w,x,v)p(w,x,v)}{p(y,z)} \\ &= \frac{p(y|x)p(z|w)p(x|w)p(w|v)}{\sum_{w'=\max(y,z)}^{v}\sum_{x'=y}^{w'} p(y|x')p(z|w')p(x'|w')p(w'|v)} \end{aligned} \tag{4.4.8}$$

The objective is to simulate $w = u_{i(1),t}$ which requires the marginal distribution of w given y, z and v. This is given by

$$p(w|y,z,v) = \frac{p(y|w)p(z|w)p(w|v)}{\sum_{w'=\max(y,z)}^{v} p(y|w')p(z|w')p(w'|v)} \quad \text{for } max(x,y) \le w \le v \tag{4.4.9}$$

Thus, under the CGA the trial density h_t consists of a function of the binomial distributions specified in Eqs. (4.4.1a) - (4.4.1f). The range of plausible values that w can take is conditional on the simulated values for x, y and v. Therefore, to simulate a value for w all possible values of w need to be enumerated and the marginal density $p(w|y,z,v)$ evaluated for each. This process produces a set of all plausible values of w, each with an evaluated probability, thus forming a non-uniform finite discrete distribution on the range of w. A procedure is then needed to draw random deviates from this distribution

and a method based on the alias method with table look-up (Devroye, 1973) is used. See Appendix A for details.

Simulating births

The second objective is to simulate $u_{i(2j),t}$, the number of new juveniles that are born to the survivors from $n_{i,t-1}$. Using the notation and distributions specified above, the unknown element of interest is now just x, with w assumed known having been simulated in the previous procedure. Therefore to generate x, the distribution of x conditional on the simulated w,y and z is required:

$$\begin{aligned} p(x|w,y,z) &= \frac{p(w,x,y,z)}{p(w,y,z)} \\ &= \frac{p(y|x)p(z|w)p(x|w)}{p(y,z|w)} \\ &= \frac{p(y|x)p(z|w)p(x|w)}{p(y|w)p(z|w)} \\ &= \frac{p(y|x)p(x|w)}{p(y|w)} \quad \text{for } y \leq x \leq w \end{aligned} \tag{4.4.10}$$

As before, x can be considered a random variable with its support constrained by y (obtained from the simulated split of the observed value of $n_{j1,t}$) and w (the number of survivors from the parent node simulated previously). The distribution for x will again be non-uniform finite discrete and a value for x^* can be drawn using the alias method.

Figure 4.9 displays the elements of a single arbitrary (juvenile) process tree. The green font denotes elements whose values are now assumed known, either directly from the data or from simulation. The only remaining unknown values are $u_{i(3j0),t}$ and $u_{i(3a0),t}$ the unmarked juveniles and uncaught adults generated by the i^{th} tree respectively. These

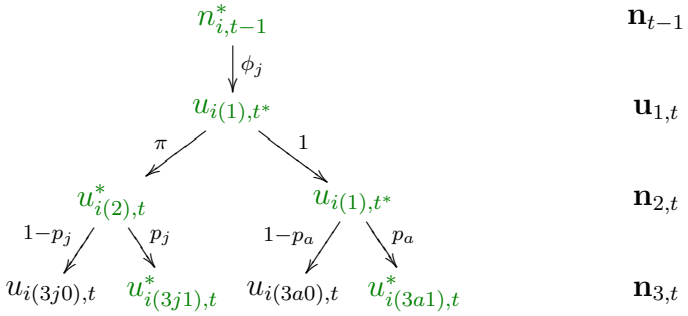


Figure 4.9: Process Tree diagram for the evolution from $\mathbf{n}_{t-1}$ after simulation of elements using the CGA algorithm. "Known" values denoted in green font.

are obtained deterministically as:

$$\begin{aligned} u_{i(3j0),t} &= u_{i(2j),t} - u_{i(3j1),t} \\ u_{i(3a0),t} &= u_{i(1),t} - u_{i(3a1),t} \end{aligned}$$

This conditional generation approach can be applied to the unknown elements of any process tree exhibiting the three key features specified earlier. In the above explanation of the CGA it was assumed that $n_{i,t-1}$ represents juvenile animals born in time $t-1$, equally it could represent adult animals. In the latter case the juvenile survival probability ϕ_j in the above expressions and diagrams should be replaced with an adult survival probability ϕ_a.

Referring back to the earlier example, the CGA can be applied to the sets of elements generated by each parent node in $\mathbf{n}_0$. Thus, an application of the CGA algorithm can be used to obtain a simulated value for $u_{1(1),1}$ conditional on the values $\left\{n^*_{j,0}, u^*_{1(3j1),1}, u^*_{1(3a1),1}\right\}$ and a simulated value for $u_{1(2j),1}$ conditional on the simulated

values $\left\{u^*_{1(1),1}, u^*_{1(3j1),1}\right\}$. The same process is applied to generate values of $u_{2(1),1}$ and $u_{2(2j),1}$. The construction of the weights for this model fitting algorithm is an involved process and, to avoid repetition with material in the next Chapter, details are omitted in this section. One approach to calculating weights for a particular model fitting algorithm, constructed according to the assumptions of the conditional approach, is presented in detail in the next Chapter.

It should be noted that this conditional generation approach is not fully conditional in that it does not automatically condition on the full set of known state elements, $\mathbf{s}_{1:T}$, across the duration of the study. For the first time period only the simulated initial states $\mathbf{n}_0$ and the observations

$$\mathbf{s}_1 = \begin{bmatrix} n_{j1,1} \\ n_{a1,1} \end{bmatrix}$$

act as constraints on the values that are simulated for the expanded state vector. For example, there is no conditioning on the simulation of $n_{a0,1} = u_{1(3a0),1} + u_{2(3a0),1}$ to ensure that the observed future values stemming from this node are biologically feasible. The evolution of the elements $n_{j0,1}$ and $n_{a0,1}$ of the state vector at time $t = 1$ through until time $t = 2$ will produce the elements in the intermediate vector $\mathbf{u}_{3,2}$, $u_{1(3a1),2}$ and $u_{4(3a1),2}$. These elements correspond to surviving animals from $n_{j0,1}$ and $n_{a0,1}$ that were captured as adults during the sampling occasion in $t = 2$. Summing these elements of $\mathbf{u}_{3,2}$ yields $n_{a01,2} = u_{1(3a1),2} + u_{4(3a1),2}$. This capture-history pattern 01 contains a 1 in the rightmost position which means this element is observed and $n_{a01,2}$ represents the known abundance of adult animals that were captured during $t = 2$ but were not captured during $t = 1$. There is no constraint on the simulation of $n_{j0,1}$ and $n_{a0,1}$ to ensure that $n_{j0,1} + n_{a0,1} \geq n_{a01,2}$. In general, a sequential approach to fitting these population dynamics models to capture-recapture data will simulate state elements at some time $t = t_\alpha$ that are implausible given

the observation at some future time $t = t_\beta$. A particle that contains these implausible state elements will have a zero importance sampling weight at time $t = t_\beta$. The occurrence of zero-weighted particles then increases the rate of particle depletion and can result in a skewed distribution of the final weights as discussed in Section 3.3.6. One approach that aims to reduce the particle depletion is discussed in the next Chapter.

Chapter 5

The Bottom Up Implementation of the Conditional Approach

5.1 Introduction

The following section details the implementation of the bottom-up approach to fitting the conditional capture-recapture model framework to the observed capture-histories. The example will consider a three period study, $T = 3$, in which the duration of each period is assumed to be equal. Within each of these three periods the population is assumed to evolve subject to survival and recruitment with measures taken on this population modelled by the capture processes. As for the general approach detailed in the previous chapter the same assumptions are made and reviewed briefly here. The first such assumption is that the population experiences no migration and therefore mortality refers solely to animals that die rather than animals that emigrate. Secondly, recruitment is assumed to be determined solely by the number of new juveniles born into the population during a single time period, no animals are assumed to immigrate from outside of the population study area. The order of the processes is, as in the general method, assumed to be survival, birth and capture.

Time period t ($t = 1, 2, 3$) begins immediately after the conclusion of the sampling

occasion in time period $t-1$. The survival process then models the survival of the animals from the end of the sampling occasion until the beginning of the breeding season. The birth and maturation process then models the recruitment to the population of new juveniles born to the surviving mature adults. Following the birth process the capture process then models the observation of the animals in the population. It is assumed that there is no mortality during either the birth or the capture processes. The population is classified into two age-cohorts: juveniles and adults. Juveniles born in time $t-1$ mature to become adults immediately prior to the birth process in time t. The group of adults known to be one-year-old in time t is defined as those animals that were marked as juveniles in time period $t-1$. This distinction allows the estimation of juvenile survival rates.

The general theory for the bottom-up approach to model fitting will be described for fitting capture-history data obtained from a general T period study. Examples will be given for a three period study to illustrate the appropriate application for the various steps involved in fitting the model.

5.2 Simulation from T to T-1

As outlined in Sections 3.3.2 fitting state-space models under traditional approaches typically involves simulating the initial states $\mathbf{n}_0$ from some prior distribution (e.g. see Eq. (3.3.1a)) before using a specified trial distribution (often the state process distribution, Eq. (3.3.1b)) to simulate the states at time 1, $\mathbf{n}_1^*$, conditional on these initial states. As previously defined, the $\mathbf{n}_t^*$ notation denotes a state vector simulated from some trial density and distinguishes the simulated states from the true, but unknown, state vector $\mathbf{n}_t$. Then, assuming an SIS fitting approach is being implemented (see Section 3.3.6), the filtered states are obtained by resampling the particles using a weighted sampling scheme.

The weights are determined by the choice of trial density and, assuming the state process pdf is the chosen trial density, the weights are proportional to the model likelihood. This likelihood, $f_1(\mathbf{y}_1|\mathbf{n}_1^*)$, is evaluated for the elements of the state vector $\mathbf{n}_1^*$ corresponding to the observed measures on the states $\mathbf{y}_1$. Once this resampling process has been completed the trial distribution is then used to simulate the states at time 2, conditional on the filtered states at time 1 and the same pattern of simulation, weighting and resampling occurs for each time period until the final sample at time T.

Fitting the model using the bottom-up approach requires an alternative strategy that begins with the most recent time period $t = T$. Therefore, the initial focus of the fitting algorithm is on the state vectors $\mathbf{n}_{T-1}$ and $\mathbf{n}_T$ and the intermediate states describing the evolution between them. These state and intermediate elements will be simulated from a trial density $h_T()$ that aims to increase the probability that all constraints imposed by the observed abundances of capture-history patterns are met. From Section 4.3 it was seen that the state vector at time t can be decomposed into two mutually exclusive vectors, $\mathbf{n}_t = [\mathbf{s}_t, \breve{\mathbf{s}}_t]'$. $\mathbf{s}_t$ contains the abundances of animals for all capture-history patterns that include capture in year t. $\breve{\mathbf{s}}_t$ contains the abundances for all capture-history patterns that do not include capture in year t. The first step in the bottom-up approach is to set the observed elements of the state vector $\mathbf{n}_T$ to the values of the observations $\mathbf{y}_T$. Thus, for this example $T = 3$ $\mathbf{s}_T^* \equiv \mathbf{y}_T$, and from Figures 4.3 and 4.4 the initial allocation of elements in $\mathbf{n}_T$ can be represented as shown in Figure 5.1. A similar process is performed to set the observed elements $\mathbf{s}_t^*$ equal to the observations $\mathbf{y}_t$ at all time periods $t = 1, \ldots, T-1$.

The next step is then to consider simulating elements of the expanded vector $\mathbf{n}_T^*$ to make the evaluation of the trial density $h_T()$ for the simulated states tractable. From Section 4.4.1 it was demonstrated that an appropriate expansion of the state vector $\mathbf{n}_T$ was

$$\mathbf{y}_3 = \begin{bmatrix} y_{j1,3} \\ y_{y11,3} \\ y_{a001,3} \\ y_{a011,3} \\ y_{a111,3} \end{bmatrix} \equiv \mathbf{s}_3^* = \begin{bmatrix} n^*_{j1,3} \\ n^*_{y11,3} \\ n^*_{a001,3} \\ n^*_{a011,3} \\ n^*_{a111,3} \end{bmatrix} \Longrightarrow \mathbf{n}_3^* = \begin{bmatrix} n_{j0,3} \\ y_{j1,3} \\ n_{y10,3} \\ y_{y11,3} \\ n_{a000,3} \\ y_{a001,3} \\ n_{a010,3} \\ n_{a100,3} \\ y_{a011,3} \\ n_{a101,3} \\ n_{a110,3} \\ y_{a111,3} \end{bmatrix}$$

Figure 5.1: Assigning the observed elements of $\mathbf{n}_T$ for $T = 3$.

equivalent to the intermediate vector $\mathbf{u}_{3,T}$ that describes the state of population immediately after the sampling occasion in time T. Thus, the next task is to simulate values for the elements of the intermediate vector $\mathbf{u}_{3,T}$ conditional on the observed and unobserved states $\mathbf{s}^*_{T-1}$ and $\breve{\mathbf{s}}^*_{T-1}$ respectively.

5.2.1 Splitting the Juveniles

The total number of new juveniles marked at time T is assumed to be known exactly and corresponds to the sum across all elements of the intermediate vector $\mathbf{u}_{3,T}$ that correspond to marked juveniles. This can be expressed as

$$n_{j1,T} = \sum_{k=1}^{2^{T-1}+4} u_{k(3j1),T}$$

and the requirement is then to split $n_{j1,T}$ across the relevant elements of $\mathbf{u}_{3,T}$. The splitting is performed using the same approach as that described in 4.4.2. The generalised Leslie matrix representation of the population dynamics of this model allows the approximate expected values of each intermediate and state vector to be obtained. The process is made

more complicated under the bottom-up approach because the elements of $\breve{\mathbf{s}}_{T-1}$ will not have been approximated at this stage of model fitting. The calculation of approximate expected values is then performed in two stages:

1. Define $\mathbf{u}^{[\mathbf{s}]}_{\cdot(3j1),T}$ as the set of elements of $\mathbf{u}_{3,T}$ corresponding to marked juveniles that are generated by parent nodes belonging to $\mathbf{s}_{T-1}$. Then the expectation of these elements is attained in the usual way:

$$E[\mathbf{u}^{[\mathbf{s}]}_{\cdot(3j1),T}|\mathbf{s}_{T-1},\theta] = \mathbf{s}_{T-1}\phi_{\cdot}\pi p_j$$

where $\phi_{\cdot} = \begin{cases} \phi_j & \text{if the element of } \mathbf{s}_t \text{ corresponds to an abundance of juvenile animals} \\ \phi_a & \text{if the element of } \mathbf{s}_t \text{ corresponds to an abundance of adult animals} \end{cases}$

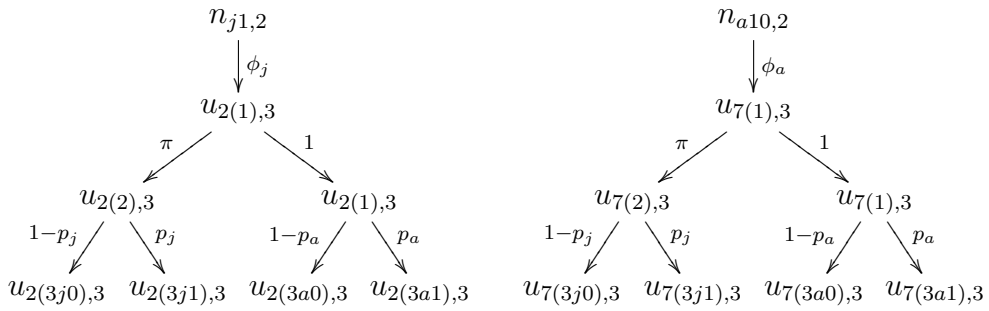
2. Define $\mathbf{u}^{[\breve{\mathbf{s}}]}_{\cdot(3j1),T}$ as the set of elements of $\mathbf{u}_{3,T}$ corresponding to marked juveniles that are generated by parent nodes belonging to $\breve{\mathbf{s}}_{T-1}$. Also, define $\mathbf{u}^{[\breve{\mathbf{s}}]}_{\cdot(3a1),T}$ as the set of elements of $\mathbf{u}_{3,T}$ corresponding to captured adults that are generated by parent nodes belonging to $\breve{\mathbf{s}}_{T-1}$. The expectation of $\mathbf{u}^{[\breve{\mathbf{s}}]}_{\cdot 3j1,T}$ is then

$$E[\mathbf{u}^{[\breve{\mathbf{s}}]}_{\cdot 3j1,T}|\mathbf{s}_{T-1},\theta] = \pi p_j \frac{\mathbf{u}^{[\breve{\mathbf{s}}]}_{\cdot 3a1,T}}{p_a}$$

The second stage requires some more explanation. The two trees in Figure 5.5 denote the evolution of the population between the second and third time periods for the animals represented in the parent nodes. The first tree (Figure 5.5a) begins with $n_{j1,2}$, the total abundance of marked juveniles after capture in the second time period. This state element is observed and therefore belongs to $\mathbf{s}_2$. Consequently the expected value can be obtained by following the path along the branches in Figure 5.5a to obtain

$$E[u_{2(3j1),2}|n_{j1,2},\theta] = n_{j1,2}\phi_j\pi p_j.$$

The second tree (Figure 5.5b) begins with $n_{a10,2}$, the abundance of adult animals that were captured during the first time period and survived, without being captured, through the second time period. The state element is unobserved and belongs to $\breve{\mathbf{s}}_2$. At this point in the fitting process it will not yet have been simulated and therefore its value is unknown. Consequently the expected value is obtained by conditioning on the value of $u_{7(3a1),3}$ which represents the number of the $n_{a10,2}$ animals that survive the third time period and are captured during the third sampling occasion. Dividing this total by the adult capture rate p_a approximates the number of $n_{a10,2}$ that survived from the second time period. Multiplying this by the birth rate π then approximates the number of juveniles that are produced by these surviving adults. Multiplying this by the juvenile capture rate p_j then yields the approximate number of marked juveniles that were produced by the survivors from the initial $n_{a01,2}$ adults.



(a) Tree with observed parent node (b) Tree with unobserved parent node

Figure 5.5: Examples of trees generated from observed and unobserved parent nodes describing the evolution of the population from time $t = 2$ until time $t = 3$.

There is one further complication with this approach; convolutions. As discussed on page 157 for the capture-recapture models under the conditional approach there are *multi-parent* state elements at time t that are obtained from summing over multiple elements

of the intermediate state vector $\mathbf{u}_{3,t}$. Therefore, the value of $u_{\cdot(3a1),t}$ will not always correspond to a single element of $\mathbf{y}_t$ and, consequently, will not be known exactly. This is the case for all the elements of the intermediate vector $\mathbf{u}_{3,3}$ that correspond to animals that are not captured during the second sampling occasion but survive the third time period and are captured during the third sampling occasion. These animals are generated from the elements of $\breve{\mathbf{s}}_\mathbf{2}$ which are the parent nodes $n_{j0,2}, n_{y10,2}, n_{a00,2}$ and $n_{a10,2}$. It can be seen that surviving adults from both the $n_{j0,2}$ and $n_{a00,2}$ state elements will contribute to the observed state element $n_{a001,3}$. Equally the surviving adults from both $n_{y10,2}$ and $n_{a10,2}$ will contribute to $n_{a101,3}$. Therefore, in general, for unobserved parent nodes in $\mathbf{n}_{T-1}$ that produce captured adults which contribute towards a *multi-parent* element of $\mathbf{n}_T$, the expected value of marked juveniles will only be obtained for the sum across multiple elements of $\mathbf{u}^{[\breve{\mathbf{s}}_{T-1}]}_{(3a1),T-1}$. For the three time period example, the expected values of neither $u_{13j1,3}$ nor $u_{53j1,3}$ can be estimated directly as only the sum $n_{a001,3} = u_{13j1,3} + u_{53j1,3}$ is known. Therefore the expectation of $u_{1(3j1),3} + u_{5(3j1),3}$ is obtained using the technique in stage 2 detailed previously:

$$E[u_{1(3j1),3} + u_{5(3j1),3} | n_{a001,3}, \theta] = \frac{\pi p_j n_{a001,3}}{p_a}.$$

A similar situation occurs when estimating the expected values of $u_{3(3j1),3}$ and $u_{7(3j1),3}$. Thus, the approximate expected values for juveniles marked in the third sampling occasion

are obtained as follows:

$$\begin{aligned}
E[u_{1,3j1,3} + u_{5,3j1,3}|n_{a001,3}] &= e_1 = \frac{\pi p_j n_{a001,3}}{p_a} \\
E[u_{2,3j1,3}|n_{j1,2}] &= e_2 = n_{j1,2}\phi_j \pi p_j \\
E[u_{3,3j1,3} + u_{7,3j1,3}|n_{a101,3}] &= e_3 = \frac{\pi p_j n_{a101,3}}{p_a} \\
E[u_{4,3j1,3}|n_{y11,2}] &= e_4 = n_{y11,2}\phi_a \pi p_j \\
E[u_{6,3j1,3}|n_{a01,2}] &= e_5 = n_{a01,2}\phi_a \pi p_j \\
E[u_{8,3j1,3}|n_{a11,2}] &= e_6 = n_{a11,2}\phi_a \pi p_j
\end{aligned}$$

These approximate expected values are then used to obtain the rates for a multinomial split of the observed total of marked juveniles $n_{j1,3}$. Let

$$r_i = \frac{e_i}{\sum_{i=1}^{6} e_i} \quad \text{for } i = 1, 2, \ldots, 6$$

then

$$\mathbf{u}^*_{\bullet(3j1),3} \sim \text{Multinomial}\left(n_{j1,3}; \{r_1, r_2, \ldots, r_6\}\right) \tag{5.2.1}$$

where $\mathbf{u}^*_{\bullet(3j1),3}$ denotes the six elements

$$\left\{u^*_{1,3j1,3} + u^*_{5,3j1,3}, u^*_{2,3j1,3}, u^*_{3,3j1,3} + u^*_{7,3j1,3}, u^*_{4,3j1,3}, u^*_{6,3j1,3}, u^*_{8,3j1,3}\right\}.$$

Let $\mathbf{u}^{[\mathbf{s}_2]}_{(3j1),3}$ denote the elements of $\mathbf{u}_{3,3}$ which record the numbers of marked juveniles generated from the elements in $\mathbf{s}_2$. The elements in the vector $\mathbf{s}_2$ correspond to the abundances of animals with a capture history pattern that included capture in time $t = 2$. Then, each element of $\mathbf{u}^{[\mathbf{s}_2]}_{(3j1),3}$ can be assigned after the multinomial split in Equation (5.2.1). For the general T period study, after splitting the marked juveniles in this way it is then necessary to consider assigning the adults captured at time T to the appropriate elements of the intermediate vector $\mathbf{u}_{3,T}$.

5.2.2 Splitting Observed Adults

In general, convolutions also occur when simulating values for elements of the intermediate vector $\mathbf{u}_{3,T}$ that correspond to animals captured at time T. As discussed on page 157, for conditional models in which the state classification is determined by capture history and age-cohort, there are several instances in which elements of the state vector $\mathbf{n}_T$ are obtained by summing multiple elements of $\mathbf{u}_{3,T}$. The general form of the capture history patterns for these *multi-parent* nodes are enumerated on page 155. For the three time period example it can be seen that $n_{a001,3}$, $n_{a101,3}$ and $n_{a111,3}$ are the *multi-parent* elements of $\mathbf{s}_3$ and are given by the following sums over elements of $\mathbf{u}_{3,3}$:

$$n_{a001,3} = u_{1,(3a1),3} + u_{5,(3a1),3} \tag{5.2.2a}$$

$$n_{a101,3} = u_{3,(3a1),3} + u_{7,(3a1),3} \tag{5.2.2b}$$

$$n_{a111,3} = u_{4,(3a1),3} + u_{8,(3a1),3}. \tag{5.2.2c}$$

It can also be seen that the *multi-parent* nodes are the same as those that led to convolutions when splitting the marked juveniles. The only observed *multi-parent* node at time T which has observed parent nodes at time $T-1$ is $n_{a\{q_2\}111,T}$ where the pattern $\{q_2\}$ is a string of $T-3$ zeros. Thus, for the three period study, $T=3$, and it can be seen that animals from both $n_{y11,2}$ and $n_{a11,2}$ will contribute to $n_{a111,3}$. Define $\mathbf{u}^{[\mathbf{s}_2]}_{(3a1),3}$ as the set of elements of $\mathbf{u}_{3,3}$ which correspond to adult animals captured in sample period 3 that were also captured in period 2. Then it is necessary to split the observed total $n_{a111,3}$ over two elements of $\mathbf{u}^{[\mathbf{s}_2]}_{(3a1),3}$: $u_{4,(3a1),3}$ and $u_{8,(3a1),3}$. The directed graph in Figure 5.6 illustrates the relationship between the elements that need to be estimated and known state element values that act as constraints in the estimation process. The intermediate processes are implicit in the graph and the parameters on the branches represent the product of the parameters $\phi_a, 1, p_a$ over the three consecutive sub-processes: survival, maturation and

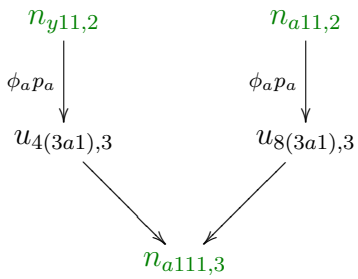


Figure 5.6: A directed graph to illustrate the convolution of $n_{a111,3}$ from the parent nodes $n_{y11,2}$ and $n_{a11,2}$. Intermediate nodes are omitted for clarity. Green text denotes known values.

capture. It is assumed that

$$u_{4(3a1),3}|n_{y11,2} \sim \text{Binomial}(n_{y11,2}, \phi_a p_a)$$
$$u_{8(3a1),3}|n_{a11,2} \sim \text{Binomial}(n_{a11,2}, \phi_a p_a).$$

Then, the first task is to simulate a value for $u_{4(3a1),3}$. The range of possible simulated integer values is given by:

$$\max(0, n_{a111,3} - n_{a11,2}) \leq u^*_{4(3a1),3} \leq \min(n_{a111,3}, n_{y11,2})$$

Thus, define

$$\mathbf{a} = \{\max(0, n_{a111,3} - n_{a11,2}), \ldots, \min(n_{a111,3}, n_{y11,2})\}.$$

Then, a splitting rate using the ratio of the approximate expected values for $u_{4(3a1),3}$ and $u_{8(3a1),3}$ is calculated

$$r_{sp} = \frac{\phi_a p_a n_{y11,2}}{\phi_a p_a n_{y11,2} + \phi_a p_a n_{a11,2}}.$$

The density associated with each possible split of $n_{a111,3}$ into the pairs $\left\{u_{4(3a1),3}, u_{8(3a1),3}\right\}$ is then evaluated

$$p(a_i|n_{a111,3}, n_{y11,2}, n_{a11,2}) = \binom{n_{a111,2}}{a_i}(r_{sp})^{a_i}(1 - r_{sp})^{n_{a111,2}-a_i} \quad \text{for all } a_i \in \mathbf{a}$$

Then, these densities are normalised:

$$p_{i,norm} = \frac{p(a_i|n_{a111,3}, n_{y11,2}, n_{a11,2})}{\sum_i p(a_i|n_{a111,3}, n_{y11,2}, n_{a11,2})} \quad \text{for all } a_i \in \mathbf{a} \tag{5.2.3}$$

Therefore, the values $\mathbf{a}_i$ and associated probabilities from Equation (5.2.3), $\mathbf{p}_{norm}$, form a non-uniform finite discrete distribution on the range of $\mathbf{a}$. The Alias method with table look-up (Appendix A) can then be used to draw a value for $u_{4(3a1),3}$. Once this is done, the value for $u_{8(3a1),3}$ is obtained deterministically as $u_{8(3a1),3} = n_{a111,3} - u_{4(3a1),3}$. For the general T period study, the observed total $n_{a\{q_2\}111,T}$ needs to be split over the elements $u_{4(3a1),3}$ and $u_{T+5(3a1),3}$.

5.2.3 Filling in the Trees

Having completed both the splitting of the observed marked juveniles at time T, $n_{j1,T}$, and the splitting of convoluted observed numbers of marked animals $n_{a\{q_2\}111,T}$, it is now possible to calculate elements of the intermediate vectors $\mathbf{u}_1$ and $\mathbf{u}_2$. For all process trees generated by observed parent nodes at time $T-1$, $\mathbf{s}_{T-1}$, the elements $\mathbf{u}_{3j1}^{[\mathbf{s}_{T-1}]*}$ and $\mathbf{u}_{3a1}^{[\mathbf{s}_{T-1}]*}$ will be known or will have been simulated by this stage. Hence, there will be $K = 2^{T-2}+2$ elements in $\mathbf{s}_{T-1}$ and for $k = 1, 2, \ldots, K$ the k^{th} process tree in Figure 5.7 can be constructed. The notation in Figure 5.7 has been simplified and the superscript $[\mathbf{s}_{T-1}]$ has been omitted. The index k refers only to the elements in $[\mathbf{s}_{T-1}]$ rather than all elements in $[\mathbf{n}_{T-1}]$. For example $u_{2(1),T}$ denotes the number of $n_{y11,T-1}$ animals that survive from the previous year, rather than the survivors from $n_{j1,T-1}$.

By applying the CGA algorithm (see page 163), values for $u_{k(1),T}$ and $u_{k(2),T}$ can then

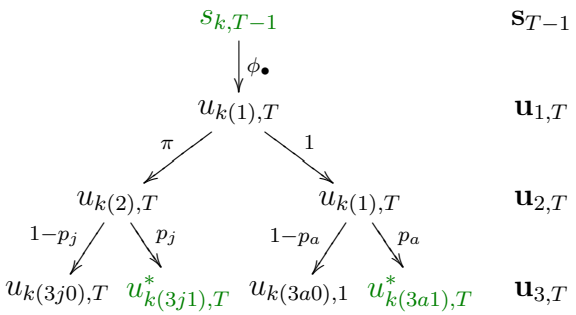


Figure 5.7: Process Tree diagram for the evolution from $\mathbf{s}_{k,T-1}$ after splitting over convolutions. "Known" values denoted in green font.

be drawn and the remaining unknown elements of $\mathbf{u}_{3,T}^{[\mathbf{s}_{T-1}]*}$ can be obtained deterministically:

$$u^*_{k(3j0),T} = u^*_{k(2),T} - u^*_{k(3j1),T}$$
$$u^*_{k(3a0),T} = u^*_{k(1),T} - u^*_{k(3a1),T}.$$

The next step is to complete the split of all remaining observed elements of $\mathbf{s}_T$ that are yet to be assigned to individual elements of $\mathbf{u}_{3,T}$.

5.2.4 Splitting across Convolutions

Following on from section 5.2.1, $u^*_{1,3j1,3} + u^*_{5,3j1,3}$ and $u^*_{3,3j1,3} + u^*_{7,3j1,3}$ both represent convolutions, each of which requires an appropriate split to assign simulated values to the remaining elements of $\mathbf{u}_{\bullet 3j1,3}$. Also, it can be seen from Eqn.'s (5.2.2a) and (5.2.2b) that the *multi-parent* elements of $\mathbf{n}_3$ with unobserved parents nodes in $\mathbf{n}_2$ will also require splitting.

For Observed Ancestral Nodes

In general, for the T period study it can be seen that $u^*_{3(3a1),T} + u^*_{7(3a1),T} = n_{a\{q2\}101,T}$ where q_2 denotes a sequence of $T-3$ zeroes. The current task is to then determine appropriate splitting rates for both the marked juveniles and captured adults that are generated by the third and seventh nodes of $\mathbf{n}_{T-1}$. It can be seen in the second enumerated convolution on page 155 that both $n_{y10,T-1}$ and $n_{a\{q_2\}10,T-1}$ will contribute to the *multi-parent* element $n_{a\{q_2\}101,T}$ where $\{q_2\}$ is a string of $T-3$ zeros. Also, for $T \geq 3$, it can be seen that the element $n_{y10,T-1}$ is generated from the process tree with $n_{j1,T-2}$ as a parent node such that $u_{2(3a0),T-1} = n_{y10,T-1}$. Equally, for $T \geq 3$, the element $n_{a\{q_2\}10,T-1}$ is generated from the process tree with $n_{a\{q_2\}1,T-2}$ as a parent node such that $u_{6(3a0),T-1} = n_{a\{q_2\}10,T-1}$.

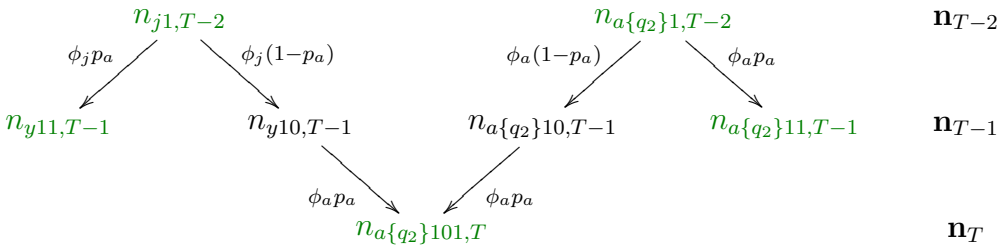


Figure 5.8: Process Tree diagram for splitting the convoluted element $n_{a\{q_2\}101,T}$ conditional on observed ancestral elements. "Known" values denoted in green font.

The process tree diagram in Figure 5.8 illustrates the relationship between the observed elements $\mathbf{n}_{j1,T-2}, \mathbf{n}_{a\{q_2\}1,T-2}$, the unobserved nodes $\mathbf{n}_{y10,T-1}, \mathbf{n}_{a\{q_2\}10,T-1}$ and the total that requires splitting $\mathbf{n}_{a\{q_2\}101,T}$. The observed elements at $T-2$ and $T-1$ that contribute to the convolved total $\mathbf{n}_{a\{q_2\}101,T}$ can be thought of as the *ancestral* elements. To ensure that simulated values for the state elements $\mathbf{n}_{y10,T-1}$ and $\mathbf{n}_{a\{q_2\}10,T-1}$ satisfy the constraints imposed by the observed capture-history data $\mathbf{s}_{1:T}$ it is necessary to include the observed elements of $\mathbf{n}_{T-1}$ generated by the *ancestral* nodes in $\mathbf{n}_{T-2}$. For example; a

simulated value of $n_{y10,2}$ obtained from splitting $n_{a101,3}$ needs to be constrained so that it cannot exceed either the value of its parent node, $n_{j1,1}$, or the value $n_{j1,1} - n_{y11,2}$ which denotes the number of surviving marked juveniles from time $t = 1$ that have not been recaptured in the second sampling period.

The simulated values of $\mathbf{n}_{y10,T-1}$ and $\mathbf{n}_{a\{q_2\}10,T-1}$ are then determined using the following approach. Firstly, to simplify the notation, define the following terms:

$$\begin{aligned} N_1 &= n_{j1,T-2} \\ N_2 &= n_{a\{q_2\}1,T-2} \\ v &= n_{y11,T-1} \\ w &= n_{y10,T-1} \\ x &= n_{a\{q_2\}10,T-1} \\ y &= n_{a\{q_2\}11,T-1} \\ z &= n_{a\{q_2\}101,T}. \end{aligned}$$

Then, define the required distributions:

$$\begin{aligned} w|N_1 &\sim \text{Binomial}(N_1, \phi_j(1-p_a)) \\ x|N_2 &\sim \text{Binomial}(N_2, \phi_a(1-p_a)) \\ z|\{w,x\} &\sim \text{Binomial}((w+x), \phi_a p_a). \end{aligned}$$

Using these definitions the range of values that w can take to respect the constraints imposed by the observed capture-history data is:

$$\max(z - N_2 + y, 0) \leq w \leq (N_1 - v)$$

and for x is

$$\max(z - N_1 + v, 0) \leq x \leq (N_2 - y).$$

Given these ranges, define $\mathbf{w}$ as the set of integer values $\{\max(z - N_2 + y, 0), \ldots, (N_1 - v)\}$ consisting of L elements and $\mathbf{x}$ as the set of integer values $\{\max(z - N_1 + v, 0), \ldots, (N_2 - y)\}$ consisting of M elements. Then the joint conditional density for each possible pair $\{w_l, x_m\}$ for $l = 1, \ldots, L$ and $m = 1, \ldots, M$ is evaluated:

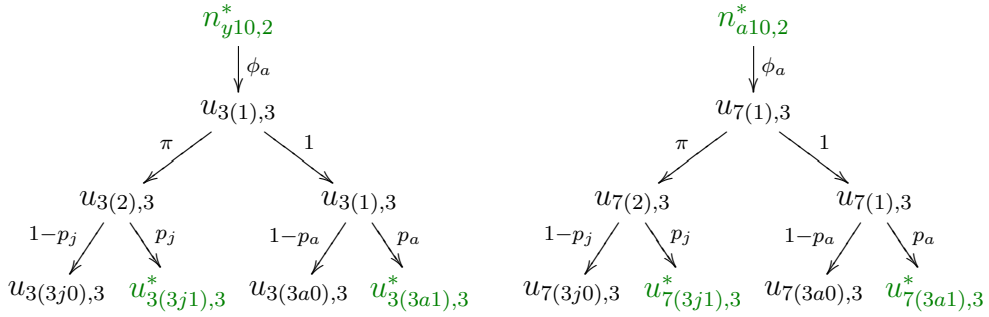
$$p(w_l, x_m | N_1, N_2, v, y, z) = \binom{N_1}{w_l} (\phi_j p_a)^{w_l} (1 - \phi_j p_a)^{N_1 - w_l} \times \binom{N_1}{x_m} (\phi_a p_a)^{x_m} (1 - \phi_a p_a)^{N_1 - x_m} \times \binom{w_l + x_m}{z} (\phi_a p_a)^z (1 - \phi_a p_a)^{w_l + x_m - z}$$

These densities are then normalised:

$$p_{lm,norm} = \frac{p(w_l, x_m | N_1, N_2, v, y, z)}{\sum_{\substack{\max(z-N_2+y,0) \leq w' \leq (N_1 - v) \\ \max(z-N_1+v,0) \leq x' \leq (N_2 - y)}} p(w', x' | N_1, N_2, v, y, z)}. \tag{5.2.4}$$

Define $\mathbf{a}_{wx}$ as the set of length $L \times M$ of pairs $\{w_l, x_m\}$ (for $l = 1, \ldots, L$ and $m = 1, \ldots, M$). Then these pairs $\mathbf{a}_{wx}$ each have an associated probability given by Equation (5.2.4) and thus form a non-uniform finite discrete distribution on the range of $\mathbf{a}_{wx}$. A pair of values $\{w_l, x_m\}$ can then be drawn using the Alias algorithm with table look-up.

Given simulated values $n^*_{y10,T-1}$ and $n^*_{a\{q_2\}10,T-1}$ it is now possible to apportion the observed number of marked adults $n_{a\{q_2\}101,T}$ to the intermediate vector elements $u_{3(3a1),T}$ and $u_{7(3a1),T}$. This is done using the splitting approach, described in section 5.2.2, that was used to apportion $n_{a111,3}$ to $u_{4(3a1),3}$ and $u_{8(3a1),3}$. Similarly the convolved number of marked juveniles, $u^*_{3(3j1),T} + u^*_{7(3j1),T}$, obtained from the juvenile splitting in section 5.2.1 can be allocated to $u^*_{3(3j1),T}$ and $u^*_{7(3j1),T}$ separately using the splitting approach from section 5.2.2. For the example when $T = 3$, consider the process tree diagram in Figure 5.12.



(a) Tree with parent node $n_{y10,2}$ (b) Tree with parent node $n_{a10,2}$

Figure 5.12: The process trees with parent nodes $n_{y10,2}$ and $n_{a10,2}$ from $t = 2$ to $t = 3$ after the convolved totals of marked juveniles and observed adults have been split across the appropriate elements of the intermediate vector $\mathbf{u}_{3,3}$. Known or simulated values are in green font.

The parent nodes and the observed child nodes (the marked juveniles and the captured adults) are all now either known from the data or have been simulated. The remaining unknown elements: the numbers of surviving adults, the numbers of new juveniles and the numbers of these that are not captured during the sampling at time $t = 3$ for each tree, can be simulated using the CGA algorithm as in section 5.2.3.

For Unobserved Ancestral Nodes

In the previous section the splitting of the marked juveniles and captured adults attributable to the third and seventh elements of $\mathbf{n}_{T-1}$ involved the use of observed ancestral parent nodes (the $n_{j1,T-2}$ and $n_{a\{q_2\}1,T-2}$ elements) and observed ancestral brethren (the $n_{y11,T-1}$ and $n_{a\{q_2\}11,T-1}$ elements). These observed ancestral nodes were then used to restrict the plausible range of values for the simulated elements $n_{y10,T-1}$ and $n_{a\{q_2\}10,T-1}$ and ensure the values were consistent with the constraints imposed by the capture-history data.

In general, for the T period study it can be seen that $u^*_{1(3a1),T} + u^*_{5(3a1),T} = n_{a\{q_2\}001,T}$

where q_2 denotes a sequence of $T-3$ zeroes. The first enumerated convolution on page 155 shows that both $n_{j0,T-1}$ and $n_{a\{q_2\}00,T-1}$ will contribute to this *multi-parent* element $n_{a\{q2\}001,T}$. It can then be seen that both of these parent nodes are unobserved and the splitting method of section 5.2.2 cannot be used. The ancestral parent node for $n_{a\{q_2\}00,T-1}$ is $n_{a\{q_2\}0,T-2}$ which is unobserved and currently unknown. Similarly $n_{j0,T-1}$ is obtained by summing all $u_{(3j0),T-2}$ elements which are generated from both known and unknown elements at time $T-2$. Hence, the previous approach in which observed ancestral nodes were used to determine appropriate simulated values cannot be used either.

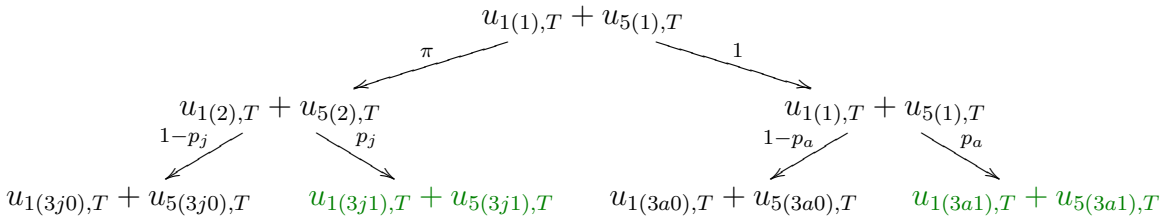


Figure 5.13: Combined intermediate nodes generated by parent nodes $n_{j0,T-1}$ and $n_{a\{q_2\}0,T-1}$. "Known" values are in green font.

To make the necessary evaluation tractable it is expedient to combine the intermediate nodes that are generated by the parent nodes $n_{j0,T-1}$ and $n_{a\{q_2\}0,T-1}$. The process tree diagram for these intermediate elements is shown in Figure 5.13 and illustrates the feasible combinations. The parent elements are not included in this combined tree as the survival probabilities for the juveniles, ϕ_j, and the adults, ϕ_a, are allowed to differ. It is assumed that the birth parameter π and the capture probabilities p_j, p_a do not vary by the age of the animals in the parent nodes and are therefore the same for the processes that cause the population to evolve from each parent node.

Using this combined form for the intermediate elements the total $u_{1(1),T} + u_{5(1),T}$ can be simulated using the Negative Binomial to approximate the true distribution. Given

the observed number of captured adults at time T that have the capture history pattern $0\,\{q_2\}\,001$ (where q_2 denotes a string of $T-3$ zeroes), the negative binomial distribution can be used to approximate the distribution of the total number of these adults that were available for capture at time T. Hence, it is assumed:

$$u_{1(1),T} + u_{5(1),T} \approx \text{NegBinomial}\left(u_{1(3a1),T} + u_{5(3a1),T}, p_a\right).$$

The approximation is necessary because $u_{1(3a1),T}+u_{5(3a1),T}$ can feasibly be zero which leads to computational errors when simulating from this negative binomial density using the standard function in the statistical software package R. The implemented approximation is then

$$u_{1(1),T} + u_{3(1),T} + 1 \approx \text{NegBinomial}\left(u_{1(3a1),T} + u_{5(3a1),T} + 1, p_a\right)$$

and once a simulated value of $(u_{1(1),T} + u_{5(1),T} + 1)^*$ is obtained, one is subtracted from this total to give the simulated number of animals that were uncaught in the first $T-1$ sampling periods and are currently alive.

Having simulated the value $(u_{1(1),T} + u_{5(1),T})^*$ the next task is to determine an appropriate split. This can be obtained using a technique similar to that described in stage 2 in section 5.2.1. Although $n_{j0,T-1}$ is not known, its observed counterpart $n_{j1,T-1}$ is an element of $\mathbf{s}_{T-1}$ and is therefore known exactly. Conditioning on this known value it is then possible to obtain an approximate expected value for $u_{1(1),T}$:

$$E[u_{1(1),T}|n_{j1,T-1}] = \phi_j(1-p_j)\frac{n_{j1,T-1}}{(p_j)}.$$

Similarly, although $n_{j0,T-1} \in \breve{\mathbf{s}}_{T-1}$, the counterpart $n_{j1,T-1} \in \mathbf{s}_{T-1}$ and is therefore known, hence an approximate expected value for $u_{1(1),T}$ is given by

$$E[u_{5(1),T}|n_{a\{q_2\}1,T-1}] = \phi_a(1-p_a)\frac{n_{a\{q_2\}1,T-1}}{(p_a)}.$$

Then, a binomial distribution can be used to split $(u_{1(1),T} + u_{3(1),T})^*$ where the rate parameter is determined by the ratio of these two approximate conditional expectations:

$$u^*_{1(1),T} \left| \left\{ (u_{1(1),T} + u_{3(1),T})^*, n_{j1,T-1}, n_{a\{q_2\}1,T-1} \right\} \sim \right.$$
$$\text{Binomial}\left((u_{1(1),T} + u_{3(1),T})^*, \frac{n_{j1,T-1}\phi_j(1-p_j)p_a}{n_{j1,T-1}\phi_j(1-p_j)p_a + n_{a\{q_2\}1,T-1}\phi_a(1-p_a)p_j} \right) \tag{5.2.5}$$

Once a value is simulated for $u^*_{1(1),T}$, the value of $u^*_{5(1),T}$ is obtained deterministically since

$$u^*_{5(1),T} = (u_{1(1),T} + u_{5(1),T})^* - u^*_{1(1),T}.$$

Having split the sum across the appropriate elements of the intermediate vector $\mathbf{u}_{1,T}$,

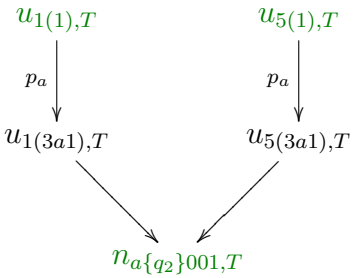


Figure 5.14: A directed graph to illustrate the convolution of $n_{a001,T}$ across the trees containing the simulated elements $u^*_{1(1),T}$ and $u^*_{5(1),T}$. Green text denotes known values.

the next step is to split the surviving number of adults $(u_{1(3a1),T} + u_{5(3a1),T})^* = n_{a\{q2\}001,T}$ across the first and fifth trees. The relationship between these elements is represented in Figure 5.14. The process used in section 5.2.2 to split summations of elements of $\mathbf{u}_{3,T}$ that corresponded to elements of $\mathbf{s}_T$ can be used here. The sub-process distributions are

again assumed to be Binomial such that

$$u_{1(3a1),T}|u^*_{1(1),T} \sim \text{Binomial}(u^*_{1(1),T}, p_a)$$
$$u_{5(3a1),T}|u^*_{5(1),T} \sim \text{Binomial}(u^*_{5(1),T}, p_a)$$

The range of plausible values for $u_{1(3a1),T}$ is then given by

$$\max(0, n_{a\{q_2\}001,T} - u_{5(1),T}) \leq u_{1(3a1),T} \leq \min(n_{a\{q_2\}001,T}, u_{1(1),T})$$

and the set $\mathcal{A}$ is then defined as

$$\mathcal{A} = \left\{\max(0, n_{a\{q_2\}001,T} - u_{5(1),T}), \ldots, \min(n_{a\{q_2\}001,T}, u_{1(1),T})\right\}.$$

The density associated with each possible split of $n_{a\{q2\}001,T}$ across $u_{1(3a1),T}$ and $u_{5(3a1),T}$ is then evaluated as

$$p(\mathcal{A}_i|n_{a\{q2\}001,T}, u^*_{1(1),T}, u^*_{5(1),T}) = \binom{u^*_{1(1),T}}{\mathcal{A}_i}(p_a)^{\mathcal{A}_i}(1-p_a)^{u^*_{1(1),T}-\mathcal{A}_i}$$
$$\times \binom{u^*_{5(1),T}}{n_{a\{q_2\}001,T} - \mathcal{A}_i}(p_a)^{n_{a\{q_2\}001,T}-\mathcal{A}_i}(1-p_a)^{u^*_{5(1),T}-n_{a\{q_2\}001,T}-\mathcal{A}_i}. \quad (5.2.6)$$

These densities are then normalised

$$p_{\mathcal{A}_i,norm} = \frac{p(\mathcal{A}_i|n_{a\{q2\}001,T}, u^*_{1(1),T}, u^*_{5(1),T})}{\sum_i p(\mathcal{A}_i|n_{a\{q2\}001,T}, u^*_{1(1),T}, u^*_{5(1),T})} \quad \text{for all } \mathcal{A}_i \in \mathcal{A} \quad (5.2.7)$$

such that the values $\mathcal{A}$ and their associated probabilities as obtained in Equation (5.2.7) form a non-uniform finite discrete distribution on the range of $\mathcal{A}$. A pair of values $\left\{u^*_{1(3a1),T}, u^*_{5(3a1),T}\right\}$ can then be drawn using the Alias method with table look-up. Having

drawn these values, their unknown complements can be obtained deterministically as

$$u^*_{1(3a0),T} = u^*_{1(1),T} - u^*_{1(3a1),T}$$
$$u^*_{5(3a0),T} = u^*_{5(1),T} - u^*_{5(3a1),T}$$

A similar approach is used for splitting $u_{1(3j1),T} + u_{5(3j1),T}$ across the first and fifth trees. There is a slight added complexity for this splitting process in that the densities need to be evaluated over the birth and juvenile capture processes rather than just the adult capture process. However the general process is much the same as for the split of the captured adults. If the model for the birth process allows an individual animal to produce more than a single young in a breeding season then the simulation and evaluation of state and intermediate elements under the trial density $h()$ becomes more complicated. This is discussed in further detail in section 5.7

Having completed the split of both the captured adult state element $n_{a\{q_2\}001,T}$ and the marked juveniles $(u_{1(3j1),T} + u_{5(3j1),T})^*$ across the first and fifth trees it is now possible to simulate values of new juveniles born to the surviving animals $u_{1(1),T}$ and $u_{5(1),T}$. This is done for both trees using the 'Simulating births' section of the CGA algorithm described on page 165. Having simulated values for $u^*_{1(2),T}$ and $u^*_{5(2),T}$ it is then simple to assign values to $u^*_{1(3j0),T}$ and $u^*_{5(3j0),T}$ using the deterministic relationship:

$$u^*_{1(3j0),T} = u^*_{1(2),T} - u^*_{1(3j1),T}$$
$$u^*_{5(3j0),T} = u^*_{5(2),T} - u^*_{5(3j1),T}$$

5.2.5 Simulating Unobserved Parent Nodes

For observed ancestral nodes

After dealing with the required splitting over the convolutions of observed adults detailed in section 5.2.2 and the associated convolutions of marked juveniles, there still remain elements of $\check{\mathbf{s}}_T$ that need to be simulated. If the study is based on more than three sampling occasions, $T > 3$, then certain elements of $\check{\mathbf{s}}_{T-1}$ can be estimated using their ancestral parent nodes (if known) and related elements of $\mathbf{s}^T$ and $\mathbf{s}^{T-1}$. Simulating these values can then allow elements of $\check{\mathbf{s}}_T$ to be estimated.

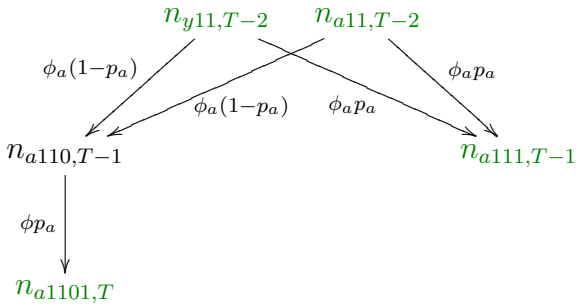


Figure 5.15: A directed graph to illustrate the relationship between the unobserved element $n_{a110,T-1}$ and its observed ancestral parent, child and brethren nodes. Green text denotes known values.

Consider an example for a study consisting of four sampling occasions $T = 4$. The state element $n_{a110,T-1}$ belongs to $\check{\mathbf{s}}_{T-1}$ and is therefore unobserved. It is also a *multi-parent* element as the surviving animals from both $n_{y11,T-2}$ and $n_{a11,T-2}$ will contribute to the total $n_{a110,T-1}$. It is also the sole parent node for the state element $n_{a1101,T}$ which is an element of $\mathbf{s}_T$ and is known exactly. The relationships between these state elements are illustrated in the directed graph in Figure 5.15.

Then the following distributions are assumed:

$$n_{a110,T-1}|(n_{y11,T-2}+n_{a11,T-2}) \sim \text{Binomial}(n_{y11,T-2}+n_{a11,T-2}, \phi_a(1-p_a))$$
$$n_{a1101,T}|n_{a110,T-1} \sim \text{Binomial}(n_{a110,T-1}, \phi_a p_a).$$

Equally, it is assumed that

$$n_{a1100,T}|(n_{y11,T-2}+n_{a11,T-2}) \sim \text{Binomial}\left((n_{y11,T-2}+n_{a11,T-2}), (\phi_a(1-p_a))^2\right).$$

As can be seen in Figure 5.15, the ancestral parent nodes $n_{y11,T-2}+n_{a11,T-2}$ contribute to both the observed state element, $n_{a111,T-1}$, and the unknown state element of interest, $n_{a110,T-1}$. Thus, conditioning on the observed value, $n_{a111,T-1}$, the viable range for $n_{a110,T-1}$ is determined as

$$n_{a1101,T} \leq n_{a110,T-1} \leq (n_{y11,T-2}+n_{a11,T-2}-n_{a111,T-1}).$$

Defining the set of values $\mathbf{a}_{ns} = \{n_{a1101,T}, \ldots, (n_{y11,T-2}+n_{a11,T-2}-n_{a111,T-1})\}$ the densities associated with each element of $\mathbf{a}_{ns}$ are obtained using the earlier distributions: Let

$$A_i = \binom{a_{ns,i}}{n_{a1101,T}} (\phi_a p_a)^{n_{a1101,T}} (1-\phi_a p_a)^{a_{ns,i}-n_{a1101,T}}$$
$$B_i = \binom{n_{y11,T-2}+n_{a11,T-2}}{a_{ns,i}} (\phi_a(1-p_a))^{a_{ns,i}} (1-\phi_a(1-p_a))^{n_{y11,T-2}+n_{a11,T-2}-a_{ns,i}}$$
$$C = \binom{n_{y11,T-2}+n_{a11,T-2}}{n_{a1101,T}} (\phi_a(1-p_a)\phi_a p_a)^{n_{a1101,T}} (1-\phi_a(1-p_a)\phi_a p_a)^{n_{y11,T-2}+n_{a11,T-2}-n_{a1101,T}}$$

Then,

$$p(a_{ns,i}|n_{a1101,T}, n_{a111,T-1}, n_{a11,T-2}, n_{y11,T-2}) = \frac{A_i \times B_i}{C}. \tag{5.2.8}$$

These densities are then normalised

$$p(a_{ns.norm,i}) = \frac{p(a_{ns,i}|n_{a1101,T}, n_{a111,T-1}, n_{a11,T-2}, n_{y11,T-2})}{\sum_i p(a_{ns,i}|n_{a1101,T}, n_{a111,T-1}, n_{a11,T-2}, n_{y11,T-2})}$$
$$= \frac{A_i B_i}{\sum_i (A_i B_i)} \qquad \forall a_{ns,i} \in \mathbf{a}_{ns,i}. \tag{5.2.9}$$

As in the approaches covered previously, the values $\mathbf{a_{ns}}$ and the associated probabilities obtained from Equation (5.2.9) form a non-uniform finite discrete distribution on the range of $\mathbf{a_{ns}}$. A value for $n^*_{a110,T-1}$ is then drawn using the Alias method with table look-up.

From Figure 4.3 it can be seen that $n^*_{a110,T-1}$ is the 11^{th} element of $\mathbf{n}_{T-1}$, when $T = 4$. Hence, the surviving number of $n^*_{a110,T-1}$ adults produce $u_{11(3j1),T}$ new juveniles that are marked during the capture occasion in time T. The simulated value $u^*_{11(3j1),T}$ will have been obtained using the technique described in step 2 in Section 5.2.1 since $n_{a1101,T}$ is known exactly. Thus, $u^*_{11(3j1),T}, u_{11(3a1),T} = n_{a1101,T}$ and $n^*_{a110,T-1}$ are all either simulated or known at this point. Hence the CGA algorithm (see page 163) can be applied to simulate values of $u_{11(1),T}$ and $u_{11(2),T}$. Having done this the values of $u_{11(3j0),T}$ and $u_{11(3a0),T}$ are obtained deterministically as described in Section 5.2.3.

This process is then repeated for all elements of $\check{\mathbf{s}}_{T-1}$ that are constrained by elements of $\mathbf{s}_{T-2}$ and, in turn, constrain elements of $\mathbf{s}_T$. Each of the *mono-parent* elements belonging to $\check{\mathbf{s}}_{T-1}$ that have yet to be simulated will, by definition, have only one ancestral parent node in $\check{\mathbf{s}}_{T-2}$. The simulated process is ultimately the same for these cases as for the example with convolutions given above and represented in Figure 5.15.

For unobserved ancestral nodes

For the elements of $\breve{\mathbf{s}}_{T-1}$ that do not have observed ancestral parent nodes, that is they have parent nodes that are elements of $\breve{\mathbf{s}}_{T-2}$, the above approach cannot be implemented. Denote the set of these elements by $\breve{\mathbf{s}}_{T-1}^{[\breve{\mathbf{s}}_{T-2}]}$ and assume it consists of V elements. At this stage all *multi-parent* elements of $\mathbf{n}_T$ will have been split over the appropriate elements of $\mathbf{u}_{3,T}$. Therefore, each element $\breve{s}_{v,T-1}^{[\breve{\mathbf{s}}_{T-2}]}$ will produce a number of surviving adults that are captured at T, $u_{v(3a1,T)}$, that is known exactly. Then, the approximate negative binomial approach, as described on page 185, can be used to simulate a value for $u_{v(1),T}$. Therefore, this proposed component of the trial density is

$$u_{v(1),T} + 1 \approx \text{NegBinomial}\left(u_{v(3a1),T} + 1, p_a\right)$$

and once the value of $(u_{v(1),T} + 1)^*$ is obtained one is subtracted from it to give the simulated number of $\breve{s}_{v,T-1}^{[\breve{\mathbf{s}}_{T-2}]}$ animals that survived after the capture process in $T-1$ until time T.

Having simulated $u^*_{v(1),T}$ and with $u^*_{v(3j1)}$ obtained from the split of the marked juveniles, $n_{j1,T}$, using the technique described in step 2 in Section 5.2.1, a simulated value for $u^*_{v(2),T}$ can be obtained using the CGA algorithm. The number of new juveniles, $u^*_{v(2),T}$, born to the surviving adults, $u^*_{v(1),T}$, is simulated using the technique described in the "Simulating Births" section on page 4.4.3. In terms of the simplified notation used in the description, x represents the value that needs to be simulated, $u^*_{v(2),T}$, conditional on both w, which represents $u^*_{v(1),T}$, and y, which represents $u^*_{v(3j1),T}$. Therefore, $u^*_{v(2),T}$ will have a non-uniform finite discrete distribution and a value can be drawn using the Alias algorithm.

At this stage all of the elements of the intermediate vectors denoting new juveniles, $\mathbf{u}_{2,T}$, and surviving animals, $\mathbf{u}_{1,T}$, will have been simulated. All elements of $\mathbf{u}_{3,T}$ will

also now have been simulated, assigned known values or set deterministically, conditional on the simulated elements of $\mathbf{u}_{1,T}$ and $\mathbf{u}_{2,T}$. To complete the simulation of all state and intermediate elements that describe the population's evolution between times $T-1$ and T, the state elements belonging to $\breve{\mathbf{s}}_{T-1}^{[\breve{\mathbf{s}}_{T-2}]}$ need to be generated. The element of $\breve{s}_{v,T-1}^{[\breve{\mathbf{s}}_{T-2}],*}$ can be simulated using the approximate negative binomial approach described in the above section. It is assumed that

$$\breve{s}_{v,T-1}^{[\breve{\mathbf{s}}_{T-2}],*} + 1 \approx \text{NegBinomial}\left(u_{v(1),T}^{*} + 1, \phi_{\cdot}\right)$$

where $\phi_{\cdot}$ is ϕ_j if $\breve{s}_{v,T-1}^{[\breve{\mathbf{s}}_{T-2}],*}$ corresponds to the unmarked juveniles at time $T-1$ and is ϕ_a for all other elements of $\breve{\mathbf{s}}_{T-1}^{[\breve{\mathbf{s}}_{T-2}],*}$. As before, one is subtracted from the simulated value of $(\breve{s}_{v,T-1}^{[\breve{\mathbf{s}}_{T-2}]})^*$ to obtain a simulated value for the unobserved state element $\breve{s}_{v,T-1}^{[\breve{\mathbf{s}}_{T-2}],*}$ at time $T-1$.

5.2.6 Completion of state vector at T

After the completion of the preceding steps all elements of the intermediate vectors, $\mathbf{u}_{1,T}$,$\mathbf{u}_{2,T}$ and $\mathbf{u}_{3,T}$ will have been simulated as will all previously unknown elements of the state vector at time $T-1$, $\mathbf{n}_{T-1}$. It still remains to specify the elements of $\breve{\mathbf{s}}_T$ that have yet to be assigned values. Given that all elements of the intermediate vector, $\mathbf{u}_{3,T}$, that corresponds to the state of population immediately after the sampling occasion in time T have been specified, it is a simple process to specify the remaining elements of $\breve{\mathbf{s}}_T$. For example, the number of unmarked juveniles that are alive in the population after the sampling occasion at time T is given by

$$n_{j0,T} = \sum_{i=1}^{2^{T-1}+4} u_{i(3j0),T}$$

The other elements of $\breve{\mathbf{s}}_T$ are either *mono-parent* elements, thus having direct correspondence with a single element of $\mathbf{u}_{3,T}$, or are *multi-parent* elements obtained by summing

across two elements of $\mathbf{u}_{3,T}$.

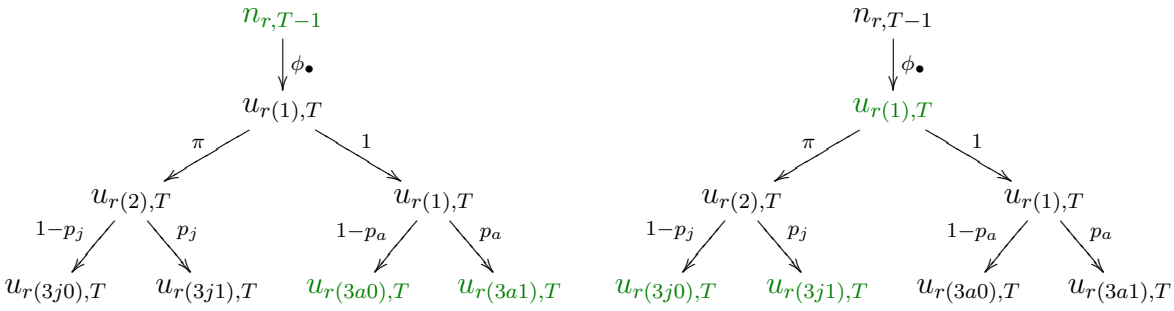
This final step completes the process of simulating all elements of the state vectors at times T and $T-1$ as well as all elements of the intermediate vectors $\mathbf{u}_{1,T}$,$\mathbf{u}_{2,T}$ and $\mathbf{u}_{3,T}$ that map the evolution of the population from the end of the sampling period in time $T-1$ to the end of the sampling period in time T. The final state vector $\mathbf{n}_T^*$ then contains elements that correspond to the most recent observations on the population as well as elements that correspond to estimates of the abundance of subsets of the population that are classified as alive but unobserved during the sampling process at time T.

5.2.7 Evaluating the densities

A general approach to fitting state-space models to mark-recapture data using the conditional approach was discussed in section 4.4.2. As specified in Eq. (4.4.2), the weight for a simulated state vector calculated for a single time period is proportional to the ratio of the state process distribution, $g_T()$, to the trial density $h_T()$. Given that all the required state and intermediate elements have been simulated by this stage to produce the expanded vector, the evaluation of the densities under both the state process distribution (Eq. (3.3.1b)) and the trial density $h()$ now becomes tractable.

For this model, in which it was assumed an adult animal could have at most one young per breeding season, it is assumed, under the state process $g_t()$, that the survival, birth and capture processes are all modelled with binomial distributions. Therefore, the cumulative effect of these sub-processes on the evolution of the animals in a tree corresponding to a single parent node can be modelled with two multinomial distributions. One multinomial is used to represent the fates of the initial animals in the tree; that is

those animals represented in the parent node (the element of $\mathbf{n}_{T-1}$). The other multinomial then represents the newborn animals, the corresponding elements of $\mathbf{u}_{2,T}$, that are introduced into the population as the progeny of the surviving animals from the parent node. These two multinomials are evaluated for each tree (i.e. for each element of $\mathbf{n}_{T-1}$) and the product of all these multinomials is taken to give the density under the state process distribution $g_T(\mathbf{n}_T|\mathbf{n}_{T-1})$.



(a) Nodes for multinomial 1 (b) Nodes for multinomial 2

Figure 5.19: A single process tree with the elements used in the evaluation of the first multinomial are in green font in graph (a), the elements used in the evaluation of the second multinomial are in green font in graph (b). $\phi_\bullet$ corresponds to ϕ_j/ϕ_a if $n_{r,T-1}$ denotes juveniles/adults.

Consider the tree generated by the r^{th} element of $\mathbf{n}_{T-1}$ $(1 \leq r \leq 2^{T-1}+4)$. The elements used in the evaluation of each of the two multinomials for this r^{th} tree are illustrated in Figure 5.19. The first multinomial models the fates of the animals represented by the parent node $n^*_{r,T-1}$. During the evolution of the population from the states at time $T-1$ to those at time T there are three fates that can befall the animals in the parent nodes. Firstly, the animals can survive from time $T-1$ and are captured during the sampling process at time T, these animals are denoted by $u^*_{r(3a1),T}$. Secondly, the animals can survive from time $T-1$ but are not captured during the sampling process at time T, these animals are denoted by $u^*_{r(3a0),T}$. Finally, the animals can fail to survive the period

between the end of sampling at time $T-1$ and the beginning of the breeding season at time T, these animals are then obtained as the difference between the parent node and the number of animals that are known or simulated to have survived. Hence, if the number of animals from the r^{th} parent node $n_{r,T-1}$ that die are denoted as $u^*_{r(dead),T-1}$ this value can be obtained as

$$u^*_{r(dead),T} = n^*_{r,T-1} - (u^*_{r(3a0),T} + u^*_{r(3a1),T})$$

Hence, the multinomial (more precisely, trinomial) distribution for the fates of the animals in the parent node $n^*_{t,T-1}$ is given as

$$u^*_{r(3a0),T}, u^*_{r(3a1),T}, u^*_{r(dead),T} \sim \text{Multinomial}(n^*_{r,T-1}; \phi.(1-p_a), \phi.(p_a), 1-\phi.).$$

The second multinomial models the fates of the newborn animals represented by the intermediate state element $u^*_{r(2),T}$ that are produced by the surviving animals $u^*_{r(1),T}$. There are, as before, three mutually exclusive and exhaustive ways in which the new animals can be classified. Firstly, the newborn animals can be captured during the sampling process at time T, these animals are denoted by $u^*_{r(3j1),T}$. Secondly, the newborn animals can remain uncaught during the sampling process at time T, these animals are denoted by $u^*_{r(3j0),T}$. Finally, the surviving adults, $u^*_{r(1),T}$, can be split between those animals that breed and those that do not produce any young. This distinction needs to be incorporated into the model and necessitates the definition of the number of non-breeding adults. These animals are obtained as the difference between the number of survivors from the parent node and the number of new juveniles that are known or simulated to have been produced. Hence, if the number of non-breeding adults is denoted as $u^*_{r(nb),T}$ this value is obtained as

$$u^*_{r(nb),T} = u^*_{r(1),T-1} - (u^*_{r(3j0),T} + u^*_{r(3j1),T})$$

Hence, the multinomial (more precisely, trinomial) distribution for the fates of the new

born juveniles in time T is given as

$$u^*_{r(3j0),T}, u^*_{r(3j1),T}, u^*_{r(nb),T} \sim \text{Multinomial}(u^*_{r(1),T-1}; \pi(1-p_j), \pi p_j, 1-\pi).$$

Then, let $ad_{r,T} = (u^*_{r(3a0),T}, u^*_{r(3a1),T}, u^*_{r(dead),T})$ denote the set of intermediate elements corresponding to the three classifications of possible fates for the animals represented by the parent node $n_{r,T-1}$. Similarly, define the set of intermediate elements corresponding to the three classifications of new juveniles for the r^{th} tree as $nj_{r,T} = (u^*_{r(3j0),T}, u^*_{r(3j1),T}, u^*_{r(nb),T})$. Then the evaluation of the state process pdf is given by

$$g_t(\mathbf{n}_T|\mathbf{n}_{T-1}) = \prod_{r=1}^{2^{T-1}+4} p(ad_{r,T}|n_{r,T-1}) p(nj_{r,T}|u_{r(1),T}) \tag{5.2.10}$$

where

$$p(ad_{r,T}|n_{r,T-1}) = \frac{n^*_{r,T-1}}{u^*_{r(3a0),T}!\, u^*_{r(3a1),T}!\, u^*_{r(dead),T}!} \times [\phi_\bullet(1-p_a)]^{u^*_{r(3a0),T}} [\phi_\bullet(p_a)]^{u^*_{r(3a1),T}} [1-\phi_\bullet]^{u^*_{r(dead),T}} \tag{5.2.11a}$$

$$p(nj_{r,T}|u_{r(1),T}) = \frac{u^*_{r(1),T}}{u^*_{r(3j0),T}!\, u^*_{r(3j1),T}!\, u^*_{r(nb),T}!} \times [\pi(1-p_j)]^{u^*_{r(3j0),T}} [\pi(p_j)]^{u^*_{r(3j1),T}} [1-\pi]^{u^*_{r(nb),T}} \tag{5.2.11b}$$

This expression can be shown to be equivalent, but more concise notationally, to taking the products of the pdfs representing each individual subprocess.

The evaluation of the trial density $h(\mathbf{u}^*_{3,T}|\mathbf{y}_T, \mathbf{n}^*_{T-1})$ where $\mathbf{u}^*_{3,T}$ is the expanded state vector (see page 160) is then achieved by evaluating the series of probability functions that are specified in sections 5.2.1:5.2.6.

However, due to the "bottom-up" nature of the model fitting process the trial density

can more accurately be written as

$$h_T(\mathbf{u}^*_{3,T}, \breve{\mathbf{s}}^*_{T-1} | \mathbf{y}_T, \mathbf{y}_{T-1}, \mathbf{y}_{T-2}).$$

The elements that need to be specified to evaluate the trial density are those contained in the intermediate vector $\mathbf{u}^*_{3,T}$ and those in $\breve{\mathbf{s}}^*_{T-1}$ which are the unobserved components of the state vector at time $T-1$. Therefore, all of these elements combine to form the expanded state vector necessary to make the density evaluation tractable. It can be seen that the simulation of the intermediate vector $\mathbf{u}^*_{3,T}$ is conditional only on the constraints imposed by the observed capture histories recorded during the samples at times T and $T-1$. As seen in section 5.2.5 the generation of the elements in $\breve{\mathbf{s}}^*_{T-1}$ is typically conditional on the already simulated elements of $\mathbf{u}^*_{3,T}$ as well as the observed capture histories during the sampling at time $T-2$. Therefore the only elements of $\mathbf{n}^*_{T-1}$ that act as constraints on the simulated values of $\mathbf{u}^*_{3,T}$ are themselves determined by the observed capture histories $\mathbf{y}_{T-1}$. Hence the simulation of the expanded vector $\left[\mathbf{u}^*_{3,T}, \breve{\mathbf{s}}^*_{T-1}\right]$ is conditional on these three sets of observations and by specifying the trial density in this form the dependency on the observed capture histories is made explicit.

The list of processes described in sections 5.2.1:5.2.6 can be summarised as follows:

a) Section 5.2.1 - splitting the observed juveniles captured during sampling at T over individual and convolved trees.

b) Section 5.2.2 - splitting the observed adults captured during sampling at T over convolved nodes where possible.

c) Section 5.2.3 - the CGA algorithm is used to simulate the intermediate nodes for trees with a known parent node element and with elements corresponding to captured juveniles and adults also assumed known.

d) Section 5.2.4 - splitting convolved elements of intermediate vectors over trees emanating from either observed or unobserved ancestral nodes.

e) Section 5.2.5 - simulating parent nodes corresponding to animals that were alive but not captured during sampling at $T-1$ for trees emanating from either observed or unobserved ancestral nodes.

f) Section 5.2.6 - completing the specification of the state vector $\mathbf{n}_T$ through assigning individual or summed elements of $\mathbf{u}_{3,T}$.

Each element in the state vector $\mathbf{n}_T$ and the preceding intermediate vectors, $\mathbf{u}_{1,T}$, $\mathbf{u}_{2,T}$ and $\mathbf{u}_{3,T}$, should appear in the density evaluation under both the state process pdf $g_T()$ and the trial density pdf $h_T()$. Due to constraints imposed by the observed data $\mathbf{y}_T$, many of the state and intermediate elements are set deterministically, which means that any component of the trial density associated with their simulation can only evaluate to unity. As a consequence, these components are not explicitly included in the evaluation of the trial density pdf.

Under a standard sequential importance sampling scheme, see Section 3.3.6, the ratio between the evaluated state process distribution, $g_T()$, and the trial density $h_T()$ determines the weights by which the particles are resampled to obtain the filtered states. These filtered states would then be projected forward across the next time period and resampled according to the updated importance sampling weights. However, under the bottom-up approach this (re)sampling step cannot be applied due to the impact any new observations $\mathbf{y}_{T+1}$ will have on the previously simulated states $\mathbf{n}^*_{1:T}$.

This problem can be illustrated by considering the impact a new observation would have on the mechanism used to calculate the sampling weights under both the basic

sequential importance sampling approach and the bottom-up fitting approach. Consider a set of n particles (as defined in Section 3.3.6) and consider again the definition of the weight assigned to a simulated state vector (see Eq (4.3.1)). The weight accorded to the i^{th} particle $(i = 1, \ldots, n)$ for the simulated states $\mathbf{n}^*_{T+1}$ conditional on the new observations $\mathbf{y}_{T+1}$ under the basic SIS approach is obtained as follows where the $*$ superscript is omitted for clarity:

$$
\begin{aligned}
w^{(i)}_{T+1} &\propto \frac{p(\mathbf{n}^{(i)}_{T+1}|\mathbf{y}_{1:T+1})}{h_{T+1}()} \\
&\propto \frac{\frac{f_{T+1}(\mathbf{y}_{T+1}|\mathbf{n}^{(i)}_{T+1})p(\mathbf{n}^{(i)}_{T+1}|\mathbf{y}_{1:T})}{p(\mathbf{y}_{T+1}|\mathbf{y}_{1:T})}}{h_{T+1}(\mathbf{n}^{(i)}_{T+1}|\mathbf{n}^{(i)}_{T}, \mathbf{y}_{1:T+1})} \\
&\propto \frac{\frac{f_{T+1}(\mathbf{y}_{T+1}|\mathbf{n}^{(i)}_{T+1})p(\mathbf{n}^{(i)}_{T}|\mathbf{y}_{1:T})g_{T+1}(\mathbf{n}_{T+1}|\mathbf{n}^{(i)}_{T})}{p(\mathbf{y}_{T+1}|\mathbf{y}_{1:T})}}{h_{T+1}(\mathbf{n}^{(i)}_{T+1}|\mathbf{n}^{(i)}_{T}, \mathbf{y}_{1:T+1})} \\
&\propto \frac{f_{T+1}(\mathbf{y}_{T+1}|\mathbf{n}^{(i)}_{T+1})p(\mathbf{n}^{(i)}_{T}|\mathbf{y}_{1:T})g_{T+1}(\mathbf{n}_{T+1}|\mathbf{n}^{(i)}_{T})}{h_{T+1}(\mathbf{n}^{(i)}_{T+1}|\mathbf{n}^{(i)}_{T}, \mathbf{y}_{1:T+1})} \\
&\propto \frac{p(\mathbf{n}^{(i)}_{T}|\mathbf{y}_{1:t})g_{T+1}(\mathbf{n}_{T+1}|\mathbf{n}^{(i)}_{t})}{h_{T+1}(\mathbf{n}^{(i)}_{T+1}|\mathbf{n}^{(i)}_{T}, \mathbf{y}_{1:T+1})} \\
&\propto w^{(i)}_{T} \frac{g_{T+1}(\mathbf{n}_{T+1}|\mathbf{n}^{(i)}_{T})}{h_{T+1}(\mathbf{n}^{(i)}_{T+1}|\mathbf{n}^{(i)}_{T}, \mathbf{y}_{1:T+1})}.
\end{aligned}
\tag{5.2.12}
$$

In the above formulation $f_{T+1}()$ denotes the observation process distribution (see Eq. (3.3.1c)), $g_{T+1}()$ denotes the state process distribution (see Eq. (3.3.1b)), $h_{T+1}()$ denotes the trial density or proposal distribution, $p(\mathbf{n}^{(i)}_T|\mathbf{y}_{1:T})$ denotes the pdf of the filtered states $\mathbf{n}_T$ conditional on the observed data up to time T and the other distributions $p()$ do not need to be made explicit. The original definition of $h_{T+1}()$ is used to maintain consistency with the notation introduced in section 4.3.1 as well as to preserve a degree of clarity in the above equations. The alternative expression for h_{T+1} can be substituted in and the following underlying argument remains valid.

The sequential nature of the weight calculation can be seen from Eq. (5.2.12). The filtered states at time T, $\mathbf{n}_T^*$, are projected forward to time $T+1$ under the trial density $h_{T+1}()$ and are therefore constrained by the new observations $\mathbf{y}_{T+1}$. The updated weight $w_{T+1}^{(i)}$ is then simply obtained by scaling the previous weight $w_T^{(i)}$ by the ratio between the evaluated state process and trial distributions, $g_{T+1}()/h_{T+1}()$. Therefore, for any t $(t = 1, \ldots, T-1)$, given the current weight $w_t^{(i)}$, the updated weight $w_{t+1}^{(i)}$ is given by

$$w_{t+1}^{(i)} \propto w_t^{(i)} \frac{g_{t+1}(\mathbf{n}_{t+1}|\mathbf{n}_t^{(i)})}{h_{t+1}(\mathbf{n}_{t+1}^{(i)}|\mathbf{n}_t^{(i)}, \mathbf{y}_{1:t+1})}. \tag{5.2.13}$$

Thus, once the weights at t for the i^{th} particle are obtained under the basic SIS approach, the weighting for any future resampling due to the addition of new observations does not require any recalculation of these existing weights. That is, under the sequential approach, the simulated states $\mathbf{n}_{0:t}^*$ are conditional only on the observations up to time t, $\mathbf{y}_{1:t}$, and therefore do not change in the presence of new observations $\mathbf{y}_{t+1}$. This is not the case for the bottom-up fitting approach.

Consider again the three time period example. Section 5.2.4 described the simulation of the convolved intermediate vector components $u_{3(3a0),3}^*$ and $u_{7(3a0),3}^*$ that sum to give the state element $n_{a100,3}^*$. This state element is effectively conditioned on the simulated values of $u_{3(3j1),3}^* + u_{7(3j1),3}^*$ and the observed value $n_{a101,3}$. However there is no conditioning that ensures the simulated value of the convolved state element $n_{a100,3}^*$ will be biologically feasible given any future observation. For example, if the population was monitored for a further time period and a mark-recapture sampling process occurred at $t = 4$ then this new vector of observed capture histories $\mathbf{y}_4$ would contain the element $n_{a1001,4}$ corresponding to adult animals that were marked initially in the first sampling occasion but were not captured again until the fourth sample. To be biologically feasible the number of animals that were marked initially in the first sample but not captured

in the second or third samples ($n_{a100,3}$) must be at least as large as the number of these animals that survive from the first time period and are then recaptured in the fourth time sample: ($n_{a1001,4}$).

The aim of simulating states under the trial density $h_t()$ can be thought of as an attempt to increase the probability that $f_t(\mathbf{y}_t|\mathbf{n}_t^*) = 1$. For this example it is quite possible that the states simulated under $h_3()$ will lead to $f_4(\mathbf{y}_4|\mathbf{n}_4^*) = 0$ since there is no feature in the trial density $h_3()$ that ensures $n_{a100,3}^* \geq n_{a1001,4}$; a condition that is required due to the constraints imposed by $\mathbf{y}_4$. Thus, for any particle containing the simulated state element $\mathbf{n}_{a000,3}^*$ such that $n_{a100,3}^* < n_{a1001,4}$, under the conditional generation approach, the observation process distribution $f_4(\mathbf{y}_4|\mathbf{n}_4^*)$ will evaluate to zero and the updated weight for that particle will be zero. Hence, proceeding in a sequential manner and updating the weights whenever new observations occur can result in significant particle depletion.

To try and reduce the potential particle depletion it is necessary to adopt an approach that increases the probability that all simulated states $\mathbf{n}_{0:t+1}$ respect the constraints imposed by all observed data $\mathbf{y}_{0:t+1}$. Under the bottom-up approach the inclusion of new observations $\mathbf{y}_{t+1}$ requires the entire set of state vectors over all time periods $\mathbf{n}_{0:t+1}$ to be simulated to replace the previous set $\mathbf{n}_{0:t}^*$. That is, the simulated states are updated whenever new observations enter into the model and can therefore be considered smoothed estimates of the true states. By adding new observations to the model the previously simulated states must change to respect the constraints imposed by the new observation. The existing particle weights will therefore change in accordance with the new observation at time $t+1$ and, consequently, cannot be updated sequentially. The particle filtering approach used to fit the models to the observed data under the bottom-up conditional generation approach is then akin to an importance sampling approach (see Section 3.3.6).

Under an importance sampling approach the posterior distribution of the states and associated parameters is then obtained by a single weighted sampling step that occurs only after the state process densities $g_t\left(\mathbf{n}_t^*|\mathbf{n}_{t-1}^*\right)$ and trial densities $h_t(\mathbf{n}_t^*|\mathbf{n}_{t-1}^*, \mathbf{y}_t)$ are evaluated at each time period t for $1 \leq t \leq T$.

Therefore, before a sampling step can be implemented it is necessary to simulate the remaining unknown state elements $\breve{\mathbf{s}}_{0:T-2}$ across all time periods conditional on the observations $\mathbf{y}_{1:T}$ and the currently simulated states $\mathbf{n}_{T-1:T}^*$. The fitting procedure is detailed in the remainder of this chapter.

5.3 Simulation from $t = T - 1$ to $t = 2$

Following the completion of the processes described in Section 5.2 the next step is to use those processes to simulate the elements of the expanded vector such that the evaluation of the trial density $h_t()$ is tractable for $3 \leq t \leq T - 1$. As was the case for the previous time period an appropriate expanded vector is the combined vector $\left[\mathbf{u}_{3,t}^*, \breve{\mathbf{s}}_{t-1}^*\right]$ for $3 \leq t \leq T - 1$. The elements of $\mathbf{u}_{3,t}$ and $\breve{\mathbf{s}}_{t-1}^*$ are then simulated using a sequence of processes similar to those detailed on page 200.

The simulation from $T - 1$ to 2 can then be thought of as a sequential process in which the state vector $\mathbf{n}_{t-1}$ and intermediate state vectors $\mathbf{u}_{1,t}$,$\mathbf{u}_{2,t}$ and $\mathbf{u}_{3,t}$ are simulated conditional on the previously simulated state vector $\mathbf{n}_t^*$ and the observations $\mathbf{y}_{t-1:T}$. The processes used to simulate the expanded state vector $\left[\mathbf{u}_{3,t}^*, \breve{\mathbf{s}}_{t-1}^*\right]$ under the trial density $h_t()$ are identical for each time period t between $T - 1$ and 2. Due to the definition of yearlings and the effect this has on the structure of the state vector $\mathbf{n}_1$, the processes required to simulate the state and intermediate vector elements mapping the transition

between the second and first time periods differ to those described here. This is discussed in more detail in Section 5.4.

The methods detailed in the following sections are only implemented if the bottom-up conditional generation approach is used to fit models to data gathered from at least four separate mark-recapture sampling periods. For three samples or less, $T \leq 3$ and therefore $T-1 \leq 2$ meaning that model fitting is completed using the methods of Sections 5.4 and 5.5 without needing to use the approach covered in section 5.3.

5.3.1 Assigning Known Values

The elements of $\mathbf{n}_{T-1}$ have all been assigned or simulated using the methods in the previous section. These values can then be assigned to the appropriate elements of the intermediate vector $\mathbf{u}_{3,T-1}$ which represents the state of the population immediately after the sampling process during time period $T-1$. To do this an index needs to be produced of the *mono-parent* elements of $\mathbf{n}_{T-1}$. By identifying these nodes a direct correspondence between the parent node in $\mathbf{n}_{T-1}$ and the intermediate node in $\mathbf{u}_{3,T-1}$ can be established and the values then assigned.

Section 5.3.2 discusses the automated approach to identifying the *multi-parent* elements. However, before performing this step it is useful to first focus on splitting the elements of $\mathbf{n}_{T-1}$ that correspond to adults into two lists; one for those elements also in the vector of observed elements $\mathbf{s}_{T-1}$ and the other for those elements also in the vector of unobserved elements $\breve{\mathbf{s}}_{T-1}$.

This automated method can be applied to the state vectors in each time period t for $2 \leq t \leq T-1$. Consider the elements of the state vector at time $t-1$, $\mathbf{n}_{t-1}$, that

	NoCE	Elements	Index
Unobserved	1	$n_{a000,3}$	5
	2	$n_{a010,3}, n_{a100,3}$	7,8
	1	$n_{a110,3}$	11
Observed	1	$n_{a001,3}$	6
	2	$n_{a011,3}, n_{a101,3}$	9,10
	1	$n_{a111,3}$	12

Table 5.1: Runs of consecutive observed/unobserved elements in $\mathbf{n}_3$. NoCE represents the Number of Consecutive Elements.

correspond to adult animals. From Figure 4.3 it can be seen that for $t \geq 2$ the first four elements of $\mathbf{n}_t$ correspond to juveniles ($n_{j0,t}$ and $n_{j1,t}$) and adults that were marked as juveniles in the previous sampling period ($n_{y10,t}$ and $n_{y11,t}$). The remaining elements represent the surviving adults in the population classified by capture-history pattern. Due to this form of classification it is relatively straightforward to automate the procedure for determining which of the elements of $\mathbf{n}_t$ belong to either list. It can be seen that, in terms of the sequences of consecutive elements corresponding to either observed or unobserved animals, the adult elements observe a pattern based on binomial coefficients given by appropriate rows of Pascal's triangle. For example, at time $t = 3$ it can be seen (see Figure 4.3) that the numbers of consecutive elements corresponding to animals that were captured during the sampling process at t follows a $(1 : 2 : 1)$ pattern; the third row of Pascal's triangle. This is replicated for the runs of consecutive elements corresponding to animals that were observed at time t. The sequences of elements are presented in Table 5.1. For time $t = 4$ the same pattern can be seen with both sequences following a $(1 : 3 : 3 : 1)$ pattern (see Table 5.2).

By observing the "Index" column in Tables 5.1 and 5.2 it can be seen that the runs of

	NoCE	Elements	Index
Unobserved	1	$n_{a0000,4}$	5
	3	$n_{a0010,4}, n_{a0100,4}, n_{a1000,4}$	7,8,9
	3	$n_{a0110,4}, n_{a1010,4}, n_{a1100,4}$	13,14,15
	1	$n_{a1110,4}$	19
Observed	1	$n_{a0001,4}$	6
	3	$n_{a0011,4}, n_{a0101,4}, n_{a1001,4}$	10,11,12
	3	$n_{a0111,4}, n_{a1011,4}, n_{a1101,4}$	16,17,18
	1	$n_{a1111,4}$	20

Table 5.2: Runs of consecutive observed/unobserved elements in $\mathbf{n}_4$. NoCE represents the Number of Consecutive Elements.

consecutive elements of $\mathbf{n}_t$ alternate between the observed and unobserved sequences. For example, in the eight adults elements of $\mathbf{n}_3$ (indexed from 5 to 12), there is 1 unobserved element followed by 1 observed element which is then followed by 2 unobserved elements which are followed by 2 observed elements and so on. This regular pattern allows the generation of the two lists to be automated using a simple recursive procedure. For the state vector at time t, define the following sequences:

$$U_1 = \binom{t-1}{0} = 1$$

$$U_2 = 1 + \binom{t-1}{1} : 2 \times \binom{t-1}{0} + \binom{t-1}{1} = 3 : t+1$$

$$U_\eta = \max(U_{\eta-1}) + \binom{t-1}{\eta-2} + 1 : \max(U_{\eta-1}) + \binom{t-1}{\eta-2} + \binom{t-1}{\eta-1} \quad \text{for } 3 \leq \eta \leq t-1$$

for $\eta \in 3, 4, \ldots, t-1$. Then define the sequence $U_{unobs} = \{U_1, U_2, \ldots, U_t\}$. Add 4 to each of these values and this gives the indexes of the elements of $\mathbf{n}_t$ that correspond to adult animals that are not captured during the sample process at time t. To obtain the corresponding indexes for the adult animals observed at time t the following sequences

are defined:

$$S_1 = 1 + \binom{t-1}{0} = 2$$
$$S_\eta = \max(S_{\eta-1}) + \binom{t-1}{\eta-1} + 1 : \max(S_{\eta-1}) + \binom{t-1}{\eta-1} + \binom{t-1}{\eta}$$

for $\eta \in 3, 4, \ldots, t-1$. Then, similarly to above, the required composite sequence is given by defining $S_{obs} = \{S_1, \ldots, S_t\}$ and adding 4 to each of the values. The sequence S_{obs} will then denote which elements of $\mathbf{n}_t$ corresponding to adult animals are contained in the vector $\mathbf{s}_t$ and will correspond to the vector of observations at time t, $\mathbf{y}_t$, excluding the first two elements $y_{j1,t}$ and $y_{y11,t}$.

Hence, the two required sequences of indexes are obtained. The next step is to automate the fitting algorithm to identify which of these elements correspond to *multi-parent* nodes.

5.3.2 Assigning Known Multi-Parent Values

For the general model introduced in Section 5.1 the population was defined to consist of two age cohorts meaning that the state vectors were classified primarily by capture-history pattern but also by age-class structure. As was described on page 155, for this type of population classification, there are three different scenarios in which *multi-parent* elements of $\mathbf{n}_{t-1}$ that do not correspond to juvenile animals are obtained by summing across two elements of $\mathbf{u}_{3,t-1}$ for $2 \leq t \leq T$. This section describes the automated creation of the indexes defining the association between the *multi-parent* elements of $\mathbf{n}_t$ and their connection to the elements of $\mathbf{u}_{3,t}$ and $\mathbf{n}_t$.

As noted on page 155, the state vector at time t, $\mathbf{n}_t$, consists of $2^t + 4$ elements. Of these, 2^t elements correspond to the non-yearling adult animals. The three scenarios leading to convolved capture history patterns as detailed on page 155 lead to the multi-parent elements of $\mathbf{n}_t$. Each of these three scenarios describes a pair of parent nodes at $t-1$ that will contribute to a pair of elements contained in $\mathbf{n}_t$ of which one element will correspond to adult animals that are unobserved at time t with the other element being the observed analogue. Consequently, six of the adult nodes in each state vector will be *multi-parent* elements. For the age and capture history pattern classification used to construct the state vectors in the preceding examples these six elements can be identified along with their parent nodes in $\mathbf{n}_{t-1}$. Consequently, the correspondence between the *multi-parent* elements in $\mathbf{n}_t$ and the elements in the intermediate vector $\mathbf{u}_{3,t}$ can now be determined. This allows the known or simulated values of the six *multi-parent* elements to be split across the appropriate elements of $\mathbf{u}_{3,t}$. The relationships between the *multi-parent* elements of $\mathbf{n}_t$, their corresponding elements of $\mathbf{u}_{3,t}$ and their parent nodes in $\mathbf{n}_t$ are given in Table 5.3.

For example, the elements $n_{1,t-1}$ and $n_{5,t-1}$ denote uncaught juveniles at $t-1$ and adult animals that remain unmarked after $t-1$ capture occasions respectively. Each of these parent nodes produces child nodes that represent the number of initial animals that survive through to the capture occasion at time t. The surviving animals from the parent node $n_{1,t-1}$ $(= n_{j0,t-1})$ are then split into those animals that are not captured at time t $(u_{1(3a0),t})$ and those that are $(u_{1(3a1),t})$; the third and fourth elements of $\mathbf{u}_{3,t}$ respectively. Similarly, the surviving animals from $n_{5,t-1}$ are split across $u_{5(3a0),t}$ and $u_{5(3a1),t}$: the 19^{th} and 20^{th} elements of $\mathbf{u}_{3,t}$ respectively. It can be seen that both $u_{1(3a0),t}$ and $u_{5(3a0),t}$ correspond to adult animals that have not been marked during any of the first t capture occasions. Hence, the element $n_{5,t}$ is obtained by the sum $u_{1(3a0),t} + u_{5(3a0),t}$: that is the sum of the 3^{rd} and 19^{th} elements of $\mathbf{u}_{3,t}$. Similarly, the element $n_{6,t}$ is obtained by the sum

$\mathbf{n}_{t-1}$	$\mathbf{u}_{3a0,t}$	$\mathbf{u}_{3a1,t}$	Index Sum	$\mathbf{n}_t$
1	3	4	$3+19$	5
5	19	20	$4+20$	6
3	11	12	$11+27$	8
7	27	28	$12+28$	$t+7$
4	15	16	$15+(4t+19)$	$2\left(1+\binom{t-1}{1}\right)+5$
$t+5$	$4t+19$	$4t+20$	$16+(4t+20)$	$2\left(1+\binom{t-1}{1}\right)+\binom{t-1}{2}+5$

Table 5.3: Index for splitting *multi-parent* elements of $\mathbf{n}_t$ across the intermediate vector $\mathbf{u}_{3,t}$. The fifth column indexes the *multi-parent* nodes elements of $\mathbf{n}_t$. The fourth column gives the indexes of elements of $\mathbf{u}_{3,t}$ that sum to give the *multi-parent* nodes. The second and third columns give the indexes of elements of $\mathbf{u}_{3,t}$ corresponding to unobserved and observed adults respectively. The first column indexes the elements of $\mathbf{n}_{t-1}$ that contribute to the *multi-parent* nodes elements of $\mathbf{n}_t$.

$u_{1(3a1),t} + u_{5(3a1),t}$: that is the sum of the 4^{th} and 20^{th} elements of $\mathbf{u}_{3,t}$. From Table 5.3 it can be seen that the appropriate pair of indices for summing the elements of $\mathbf{u}_{3,t}$ is fixed for the first three rows but becomes a function of t for the remaining rows. This allows the process of assigning the *multi-parent* elements of $\mathbf{n}_t$ to $\mathbf{u}_{3,t}$ to be implemented automatically in the fitting algorithm.

By using the indexes obtained in Sections 5.3.1 and 5.3.2 it is then possible to identify the *multi-parent* nodes and determine which of these correspond to observed animals and which denote unobserved animals. Section 5.3.4 describes the methods used to simulate an appropriate allocation of *multi-parent* elements of $\mathbf{n}_t$ to the related nodes of the intermediate vector $\mathbf{u}_{3,t}$.

5.3.3 Splitting the Juveniles

At this point all state elements in $\mathbf{n}_{T-1}$ have been simulated and thus all elements of $\mathbf{u}_{3,T-1}$ corresponding to a single element of $\mathbf{n}_T$ will also be set to that value. The next step is to allocate the number of juveniles at time $T-1$ to the appropriate elements of $\mathbf{u}_{3,T-1}$.

By applying the techniques in Section 5.2.5 the unknown elements of the state vector $\mathbf{n}_{T-1}$ have been simulated. Thus, the element $\mathbf{n}^*_{j0,T-1}$ denotes the simulated number of juvenile animals that were uncaught during the capture occasion in period $T-1$. At this stage both $\mathbf{n}^*_{j0,T-1}$ and $\mathbf{n}_{j1,T-1}$ are known; through simulation and directly from the observed data $\mathbf{y}_{T-1}$ respectively. Both of these totals can then be divided and allocated to the appropriate elements of $\mathbf{u}_{3,T-1}$ using the approach detailed in Section 5.2.1. The required rates for the multinomial split of the unmarked juveniles are obtained using the same techniques as for the marked animals. For the simple form of the model described in section 5.1 in which the juvenile and adult capture rates, p_j and p_a respectively, are constant within a particle the calculation of the splitting rates is simplified. In this case the required rates for the multinomial splits of both the marked and unmarked juveniles will be equivalent. The only difference in calculating the splitting rates for each of these totals is the use of the parameters p_j or $1-p_j$ for the observed and unobserved juveniles respectively. Since these parameters are constant across all nodes within a particle they cancel out in the calculation of the ratios of expected values. More generally, for the simple form of the model the required rates for the multinomial splits of both the marked and unmarked juveniles do not depend on the values of either the birth rate parameter π or the juvenile capture rate parameter p_j. In more complex models either one or both of the birth and juvenile capture probabilities may vary according to the element of the state vector $\mathbf{n}_{T-2}$. These rates may be set independently or they may be modelled as

functions of external covariates (see section 2.2.5). For example, it may be believed that the probability of producing a juvenile in year t is influenced by the parent's capture history in year $t-1$. For these more complex cases the simplification no longer applies and the splitting rates for marked and unmarked juveniles need to be calculated separately.

For example, consider the split of the marked and unmarked juveniles at time $T-1$. These juveniles need to be split over the relevant elements of the intermediate vector $\mathbf{u}_{3,T-1}$. For the simple model, as noted in section 5.3.2, the state vector at $T-2$ $\mathbf{n}_{T-2}$, consists of $2^{T-2}+4$ elements. Thus, both the marked and unmarked juveniles need to be split over $2^{T-2}+4$ elements. However, as noted in the earlier description of splitting juveniles (section 5.2.1), the convolutions of capture histories that result in *multi-parent* elements mean the splitting rates cannot be obtained for all elements, hence the full split cannot yet be calculated. From the generation of the indexes used to classify the observed (section 5.3.1) and multi-parent (section 5.3.2) elements it can be seen that the splitting rates, for both marked and unmarked juveniles, cannot be obtained separately for the first, fifth, third or seventh elements. Hence, due to these convolutions, for this initial multinomial based split the marked and unmarked juveniles can only be apportioned into $2^{T-1}+2$ distinct parts. Thus, with reference to Equation (5.2.1), if r_i denotes the multinomial rate parameter for the i^{th} scaled expected value then the distributions used to simulate the split of unmarked and marked juveniles are:

$$\mathbf{u}^*_{\bullet(3j0),T-1} \sim \text{Multinomial}\left(n^*_{j0,T-1}; \{r_1, r_2, \ldots, r_{2^{T-1}+2}\}\right)$$
$$\mathbf{u}^*_{\bullet(3j1),T-1} \sim \text{Multinomial}\left(n_{j1,T-1}; \{r_1, r_2, \ldots, r_{2^{T-1}+2}\}\right)$$

where $\mathbf{u}^*_{\bullet(3j0),3}$ and $\mathbf{u}^*_{\bullet(3j1),3}$ denote the $2^{T-1}+2$ elements that the unmarked and marked juveniles respectively are split across. As discussed earlier, for the simple model form the splitting rates $\mathbf{r}$ are the same for both distributions.

5.3.4 Completing the loop

Having completed the initial split of the marked and unmarked juveniles the model fitting continues by broadly following the same series of processes that are outlined on page 200.

Splitting across Convolutions

Using the indexes from sections 5.3.1 and 5.3.2 it can be seen that the only *multi-parent* element observed at $T-1$ that is obtained from parent elements that have both been observed at $T-2$ is $n_{\varsigma,T-1}$ where $\varsigma = 2\left(1+\binom{T-2}{1}\right)+\binom{T-2}{2}+5$. Again, in this context, "observed at $T-1$" means that the capture history pattern associated with that element will contain a 1 in the $(T-1)^{\text{th}}$ position. The process of splitting the observed adults $n_{\varsigma,T-1}$ across the known parent nodes is conducted using the technique described in section 5.2.2.

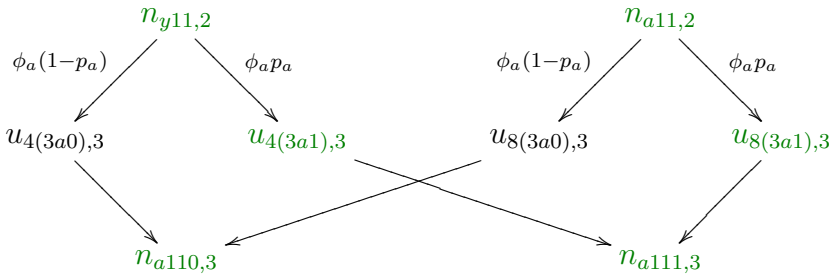


Figure 5.20: A directed graph to illustrate the convolution of $n_{a110,3}$ from the parent nodes $n_{y11,2}$ and $n_{a11,2}$ after splitting the value $n_{a111,3}$. Intermediate nodes are omitted for clarity. Green text denotes known/simulated values.

Having split the observed adults it is now necessary to incorporate the splitting of the associated *multi-parent* element that is unobserved at $T-1$ and is obtained from observed

parent nodes. This split is conditional on the values obtained from splitting the observed adults $n_{\varsigma,T-1}$. This conditional splitting can be illustrated with an example. Assume that the time-series of observed capture histories is extended to four capture occasions (i.e. $T = 4$) and the initial model fitting from time T to $T-1$ had been completed. Then the observed *multi-parent* element $n_{a111,3}$ will have just been split over the trees generated by the observed parent nodes $n_{y11,2}$ and $n_{a11,2}$. It now remains to split the unobserved *multi-parent* elements $n_{a110,3}$ over those same trees. The directed graph in Figure 5.20 is similar to Figure 5.6 but now incorporates the *multi-parent* element $n_{a110,3}$ that is unobserved at time 3 (i.e. $T-1$). At this stage both $u_{4(3a1),3}$ and $u_{8(3a1),3}$ will have been simulated and are therefore assumed known. Analogously to the example in Section 5.2.2 it is assumed that

$$u_{4(3a0),3}|n_{y11,2} \sim \text{Binomial}(n_{y11,2}, \phi_a(1-p_a))$$
$$u_{8(3a0),3}|n_{a11,2} \sim \text{Binomial}(n_{a11,2}, \phi_a(1-p_a)).$$

Then, conditional on the simulated split of the observed *multi-parent* element $n_{a111,3}$, the range of possible simulated integer values for $u_{4(3a0),3}$ is:

$$\max(0, n_{a110,3} - n_{a11,2} + u_{8(3a1),3}) \leq u^*_{4(3a0),3} \leq \min(n_{y11,2} - u_{4(3a1),3}, n_{a110,3})$$

The same techniques as in section 5.2.2 are then used with the splitting rate being obtained from the ratio of the approximated expected values for $u_{4(3a0),3}$ and $u_{8(3a0),3}$. The densities for each possible pair of $\left(u_{4(3a0),3}, u_{8(3a0),3}\right)$ values are then obtained and normalised to form a non-uniform discrete distribution on the possible range of $u_{4(3a0),3}$. The Alias method with table look-up is then used to simulate the pair of elements $\left(u^*_{4(3a0),3}, u^*_{8(3a0),3}\right)$.

For this example, having now simulated the splitting of all the relevant child nodes for the trees generated by the observed parent nodes $n_{y11,3}$ and $n_{a111,3}$ the intermediate

elements $u_{4(1)}$, $u_{8(1)}$, $u_{4(2)}$ and $u_{8(2)}$ can all be set deterministically. This property can be described more generally. For the trees generated by elements of $\mathbf{s}_{T-2}$, it is not necessary to use the CGA algorithm to simulate the values of the associated intermediate nodes $\mathbf{u}_{1,T-1}$ or $\mathbf{u}_{2,T-1}$. This contrasts with the techniques described in section 5.2.3 and is due to the splitting of both the unmarked and marked juveniles as described in section 5.3.3. Consider the trees generated by elements of $\mathbf{s}_{T-2}$; the child nodes corresponding to unmarked and marked juveniles have both been simulated by this stage whereas in the initial simulation from time T to $T-1$ the unmarked juveniles were unknown. Therefore, using the same notational syntax as in Figure 5.7, the elements $u_{k(3j0),T-1}$ and $u_{k(3j1),T-1}$ have both been simulated at this stage. This means the element representing the number of new-born juveniles $u_{k(2),T-1}$ attributed to the k^{th} element of $\mathbf{n}_{T-2}$ is set deterministically as

$$u_{k(2),T-1} = u_{k(3j0),T-1} + u_{k(3j1),T-1}$$

thus negating the use of the CGA algorithm for this component of the fitting algorithm. Similarly, after the necessary splitting of the *multi-parent* elements of $\mathbf{n}_{T-1}$ corresponding to adults, the elements in $\mathbf{u}_{k(1),T-1}$ can be set deterministically.

Studying the indexes in sections 5.3.1 and 5.3.2 it can also be seen that the *multi-parent* elements $n_{8,T-1}$ and $n_{T+6,T-1}$ are obtained from the unobserved parent elements $n_{3,T-2}$ and $n_{7,T-2}$. These parent elements are themselves generated by the observed ancestral elements $n_{2,T-3}$ and $n_{6,T-3}$. It is then necessary to simulate these *multi-parent* elements which can be done using the techniques described in section 5.2.4. The difference in this case is that both $n_{8,T-1}$ and $n_{T+6,T-1}$ will have been simulated during the initial model fitting from T to $T-1$. Hence, the simulated values for the unobserved *multi-parent* elements need to be constrained by both the observed and previously simulated values at $T-1$. Continuing the earlier example it can be seen that $n_{a100,3}$ and $n_{a101,3}$ are

both convolved elements and their sum $n_{a100,3} + n_{a101,3}$ denotes the number of surviving animals, including both those animals that were juveniles during the first capture occasion as well as those that were adults, that were caught during the first but not the second capture occasions. Then, with reference to the notation on page 182, the following general definitions are made:

$$\begin{aligned} N_1 &= n_{j1,T-3} \\ N_2 &= n_{a\{q_2\}1,T-3} \\ v &= n_{y11,T-2} \\ w &= n_{y10,T-2} \\ x &= n_{a\{q_2\}10,T-2} \\ y &= n_{a\{q_2\}11,T-2} \\ z &= n_{a\{q_2\}100,T-1} + n_{a\{q_2\}101,T-1}. \end{aligned}$$

Where the pattern $\{q_2\}$ is a string of $T-4$ zeros. For this example $T = 4$ and therefore q_2 is not needed. Hence, for this example the required distributions are:

$$\begin{aligned} w|N_1 &\sim \text{Binomial}(N_1, \phi_j(1-p_a)) \\ x|N_2 &\sim \text{Binomial}(N_2, \phi_a(1-p_a)) \\ z|\{w,x\} &\sim \text{Binomial}((w+x), \phi_a). \end{aligned}$$

Note the change of the rate parameter in the distribution $z|\{w,x\}$. The generation of a plausible pair of values for the parent nodes $(n_{y10,3}, n_{a010,3})$ is then performed using the methods discussed in section 5.2.4. For the general case, given these simulated values the *multi-parent* nodes $n_{8,T-1}$ and $n_{T+6,T-1}$ can then be apportioned to the elements of the intermediate vector $\mathbf{u}_{3,T-1}$ specified by the indexes in the third and fourth rows of Table 5.3. This splitting is performed as described in section 5.2.2 and these techniques are also used

to allocate the convolved sums of juveniles, both unmarked ($u^*_{3(3j0),T-1} + u^*_{7(3j1),T-1}$) and marked ($u^*_{3(3j1),T-1} + u^*_{7(3j1),T-1}$) to the relevant individual elements of $\mathbf{u}_{3,T-1}$. The remaining intermediate elements for the trees generated by the parent nodes $n_{3,T-2}$ and $n_{7,T-2}$ are then obtained deterministically as previously described.

The indexes in sections 5.3.1 and 5.3.2 show that the final pair of *multi-parent* elements $n_{5,T-1}$ and $n_{6,T-1}$ are also obtained from unobserved parent elements: $n_{1,T-2}$ and $n_{5,T-2}$. However, there is no direct correspondence between these parent elements and observed ancestral elements at time $T-3$; therefore the previous method of conditional splitting cannot be applied in this case. During the initial stage of model fitting these *multi-parent* elements were split using the methods introduced in section 5.2.4. Given that the state element $n^*_{5,T-1}$ will have been simulated during the initial fitting stage it is not necessary to simulate the sum $u^*_{1(1),T-1} + u^*_{5(1),T-1}$ using the adjusted negative-binomial approach introduced in section 5.2.4. Instead, the splitting of the *multi-parent* nodes can be obtained using multinomial distributions with rate parameters determined by the expected values of the relevant elements of the intermediate vector $\mathbf{u}_{3,T-1}$. These expectations are obtained using the technique described earlier in stage 2 of section 5.2.1. For example:

$$E[u_{1(3a0),T-1}|n_{2,T-2}] = \frac{n_{2,T-2}}{(p_j)} \times (1-p_j) \times \phi_j \times \pi \times (1-p_a).$$

and

$$E[u_{5(3a0),T-1}|n_{6,T-2}] = \frac{n_{6,T-2}}{(p_a)} \times (1-p_a) \times \phi_a \times \pi \times (1-p_a).$$

The ratio of these two conditional expectations is then used to determine the rate parameter for the binomial distribution used to split the simulated *multi-parent* element $n^*_{5,T-1}$ over the intermediate nodes $u_{1(3a0),T-1}$ and $u_{5(3a0),T-1}$. A similar procedure is invoked to

simulate the splitting of the observed adult element $n_{6,T-1}$ as well as that of the simulated components of both the observed and unobserved juveniles: $u^*_{1(3j1)} + u^*_{5(3j1)}$ and $u^*_{1(3j0)} + u^*_{5(3j0)}$ respectively. If the state process parameters do not vary by state element then it can be seen that the splitting rate will be the same for each of the four simulated splits. Once simulated, these split values can then be summed appropriately to give the values assigned to the numbers of surviving adults and new juveniles, contained in $\mathbf{u}_{1,T-1}$ and $\mathbf{u}_{2,T-1}$ respectively.

Simulating Unobserved Parent Nodes

Having both completed the splitting of the *multi-parent* elements of $\mathbf{n}_{T-1}$ and simulated the associated intermediate nodes the next step is to consider any remaining elements of $\check{\mathbf{s}}_{T-2}$ that are still to be simulated. Using similar arguments to those in section 5.2.5 it can be shown that, if $T \geq 5$, there will be elements of $\check{\mathbf{s}}_{T-2}$ that can be estimated by conditioning on observed or simulated elements of $\mathbf{s}^{T-1}$, $\mathbf{s}^{T-2}$ and $\mathbf{u}^{1,T-1}$.

In section 5.2.5, where $t = T-1$, the constraint on the minimum simulated value of $\check{\mathbf{s}}_t$ was imposed by the appropriate elements of s_{t+1}. For the main loop, where $t \in (T-2, 2)$, this minimum constraint is imposed by the simulated value of $\mathbf{u}_{1,t+1}$. This is because both $\mathbf{s}_t$ and $\check{\mathbf{s}}_t$ will have been simulated by this stage and summing the elements, after the appropriate splitting if required, yields the values of $\mathbf{u}_{1,t+1}$ which denote the numbers of surviving animals from $\mathbf{n}_t$. Accordingly, the rate parameter for the binomial distribution relating the unobserved parent node in $\check{\mathbf{s}}_t$ to its minimum allowable value is adjusted. Now it will simply consist of the survival rate (ϕ_j or ϕ_a) rather than the product of survival and capture as was the case in Equation A_i on page 192. For example, consider a model being fitted to capture history data obtained from 5 capture occasions. In this scenario $T = 5$

and the interest lies in simulating a value for $n_{a110,T-2}$. The relationship between this element and the constraints imposed upon it by observed or simulated state elements is illustrated in the directed graph in Figure 5.21. By comparing Figure 5.21 to Figure 5.15 it can be seen how the methods of section 5.2.5 can be applied in these scenarios.

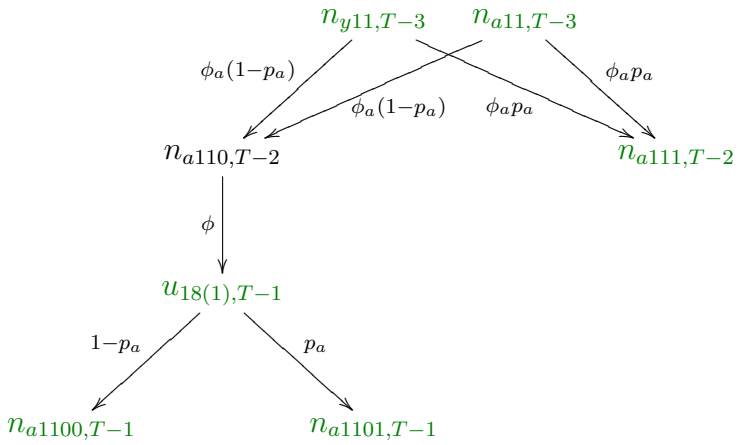


Figure 5.21: A directed graph to illustrate the relationship between the unobserved element $n_{a110,T-2}$ and its observed ancestral parent, child and brethren nodes. Green text denotes known values.

To implement this approach there needs to be an observed ancestral element of $\mathbf{s}_{t-1}$ to act as a constraint on the maximum allowable value for the corresponding simulated unobserved element of $\breve{\mathbf{s}}_t^*$. As was noted in section 5.2.5 this is not always the case. For this scenario, given the minimum constraints of $\mathbf{u}_{1,t+1}$, the values of the unobserved state elements in $\breve{\mathbf{s}}_t$ that have unobserved parent nodes in $\mathbf{s}_{t-1}$ can be simulated using the approximate negative binomial approach described in section 5.2.5.

At this stage all elements of the state vector $\mathbf{n}_{T-2}$ will have been simulated along with all the elements of the intermediate vectors $\mathbf{u}_{1,T-2}$, $\mathbf{u}_{2,T-2}$ and $\mathbf{u}_{3,T-2}$ thus mapping the assumed evolution of the population between the sampling period in $T-2$ and that in

$T-1$.

5.3.5 Evaluating the Densities

Having simulated all required elements from $T-2$ to $T-1$ it is now necessary to evaluate the densities under both the state process distribution and the trial density. The evaluation of the density under the state process once again consists of taking the product of two multinomial densities for each 'tree' generated by an element of $\mathbf{n}_{T-2}$: one modelling the fate of the initial adults and the other modelling the fate of the new-born juveniles. The product of all these multinomial density pairs is then evaluated to obtain $g_{T-1}\left(\mathbf{n}_{T-2}|\mathbf{n}_{T-1}\right)$.

Evaluating the trial density,

$$h_{T-1}(\mathbf{u}^*_{3,T-1}, \breve{\mathbf{s}}^*_{\mathbf{T-2}}|\mathbf{n}^*_{T-1}, \mathbf{y}_{T-2}, \mathbf{y}_{T-3}),$$

requires an approach similar to that for the fitting of the model evolution from time T to $T-1$. As in section 5.2.7, the expanded vector, required to make the evaluation of the densities tractable, is the combination of intermediate vector denoting the capture status at time $T-1$ and the unobserved elements of the state vector at time $T-2$: $\left[\mathbf{u}^*_{3,T-1}, \breve{\mathbf{s}}^*_{T-2}\right]$.

The trial density is comprised of the probability functions specified in sections 5.3.3:5.3.4. These are broadly similar to those that were listed previously (page 200) and can be summarised as follows:

a) Section 5.3.3 - splitting both the observed and unobserved juveniles alive during sampling at $T-1$ over individual and convolved trees.

b) Section 5.3.4 - this section described the following processes:

- splitting the observed adults captured during sampling at $T-1$ over convolved nodes.
- splitting the unobserved adults that were alive but not captured during sampling at $T-1$ over convolved nodes.
- splitting convolved elements of intermediate vectors over trees emanating from either observed or unobserved ancestral nodes.
- simulating non-convolved parent nodes corresponding to animals that were alive but not captured during sampling at $T-1$ for trees emanating from either observed or unobserved ancestral nodes.

As can be seen from comparing this list to that on page 200, the main difference between these sets of processes is that the requirement to simulate intermediate nodes is much reduced for fitting the model to the observations recorded between sampling occasions $t = 2$ and $t = T - 1$. When modelling the evolution between times $t - 1$ and t the elements of the state vector $\mathbf{n}_t$ will have already been simulated during the fitting of the previous time period. Hence, due to the correspondence between $\mathbf{n}_t$ and $\mathbf{u}_{3,t}$, the only simulation required is that to split convolutions across multiple elements of $\mathbf{u}_{3,t}$. The remaining intermediate nodes are then obtained deterministically and the technique described in Section 5.2.3 is not required in this part of the fitting process.

Having simulated the state vector $\mathbf{n}^*_{T-2}$ using the techniques in sections 5.3.3:5.3.4 the time index is reduced from $T-2$ to $T-3$ and the same fitting procedure is then repeated to simulate $\mathbf{n}^*_{T-3}$ and all associated intermediate nodes. This incremental fitting approach is then continued until $\mathbf{n}^*_2$, the states denoting the population immediately following the capture occasion at time $t = 2$, has been simulated. This looping approach can be implemented between time periods $t = T - 1$ and $t = 2$ for any general T. Hence this chapter

describes a general algorithm that can be applied to fit models, of the structure specified previously, to capture-recapture data consisting of T capture occasions for general T, such that $T > 2$.

The looping procedure can no longer be applied as the transition of the population between times $t = 2$ and $t = 1$ does not involve the simulation of yearlings at time $t = 1$. This feature of the model and its impact on the fitting algorithm are discussed in the next section.

5.4 Simulation from 2 to 1

This section of the model fitting algorithm describes the procedures necessary to simulate the state vector at time $t = 1$, $\mathbf{n}_1^*$, and the intermediate nodes $\mathbf{u}_{1,2}$, $\mathbf{u}_{2,2}$ and $\mathbf{u}_{3,2}$. Many of the fitting techniques used for this section of the model will be identical in structure to those previously discussed. Hence, only a relatively brief outline will be given. As with the previous sections, the first required process is to establish the relationship between the elements of the state vector at t, in this case $\mathbf{n}_2^*$, and the elements of the final intermediate vector, in this case $\mathbf{u}_{3,2}$. This relationship is described in the following section.

5.4.1 Assigning Known Adult Values

The elements of $\mathbf{n}_2^*$ that summarise the population after the second sampling occasion were first represented in Figure 4.3. Also, the relationship between these state elements and the intermediate elements in $\mathbf{u}_{3,2}$ was earlier illustrated in Figure 4.7. These figures illustrate that of the eight elements in $\mathbf{n}_2^*$, four correspond exactly with individual elements of $\mathbf{u}_{3,2}$.

$\mathbf{n}_1$	$\mathbf{u}_{3a0,2}$	$\mathbf{u}_{3a1,2}$	Index Sum	$\mathbf{n}_2$
1	3	4	$3+11$	5
3	11	12	$4+12$	6

Table 5.4: Index for splitting *multi-parent* elements of $\mathbf{n}_2$ across the intermediate vector $\mathbf{u}_{3,2}$. The fifth column indexes the *multi-parent* nodes elements of $\mathbf{n}_2$. The fourth column gives the indexes of elements of $\mathbf{u}_{3,2}$ that sum to give the *multi-parent* nodes. The second and third columns give the indexes of elements of $\mathbf{u}_{3,2}$ corresponding to unobserved and observed adults respectively. The first column indexes the elements of $\mathbf{n}_1$ that contribute to the *multi-parent* nodes elements of $\mathbf{n}_2$.

For the previous time periods $3 \leq t \leq T$ it could be seen, in Table 5.3 and the accompanying description in section 5.3.2, that there were always six convolved capture histories denoted by *multi-parent* elements that needed to be assigned to the appropriate elements of the intermediate vector after capture $\mathbf{u}_{3,t}$. From Figures 4.3 and 4.7 it can be seen that the fitting process will involve splitting only two convolved capture histories for those non-juvenile elements of $\mathbf{n}_2$. Those two *multi-parent* elements are $n^*_{a00,2}$ and $n^*_{a01,2}$ and the indexes for splitting these elements are given in Table 5.4.

Having identified the *multi-parent* elements of $\mathbf{n}^*_2$ the next two sections describe how this affects the simulated split of both juvenile elements of $\mathbf{n}^*_2$ as well as the method for splitting the *multi-parent* elements across the appropriate nodes of the intermediate vector $\mathbf{u}_{3,2}$.

5.4.2 Splitting the Juveniles

Each state vector $\mathbf{n_t}$ (for $1 \leq t \leq T$) contains elements corresponding to unmarked and marked juveniles, $n_{j0,t}$ and $n_{j1,t}$ respectively. At this stage of the model fitting both of these elements will be assumed known, either from direct observation in the case of $n_{j1,2}$ or from simulation in the case of $n^*_{j0,2}$. The splitting process is then performed using the methods described in section 5.3.3. For this time period each of the elements corresponding to juvenile animals at time $t = 2$ can only be apportioned to three distinct categories due to the convolution discussed in section 5.4.1. The expectation for the number of juveniles apportioned to these convolved unobserved parent nodes is then obtained using the technique in stage 2 of section 5.2.1. As discussed in section 5.3.3 the rates for splitting both the observed and unobserved juveniles will be equivalent for the simple model.

5.4.3 Splitting across Convolutions

From section 5.4.1 it was seen that $n_{a00,2}$ and $n_{a01,2}$ are *multi-parent* elements that are both generated from the unobserved parent nodes $n_{j0,1}$ and $n_{a0,1}$. The necessary splitting is then implemented using the techniques described in the final paragraphs of section 5.3.4. Multinomial distributions are constructed with rate parameters again determined by the expected values of the relevant elements of $\mathbf{u}_{3,2}$. For this section of the model the appropriate conditional expectations are then:

$$E[u_{1(3a0),2}|n_{2,1}] = \frac{n_{2,1}}{(p_j)} \times (1 - p_j) \times \phi_j \times \pi \times (1 - p_a)$$

and

$$E[u_{3(3a0),T-1}|n_{4,1}] = \frac{n_{4,1}}{(p_a)} \times (1 - p_a) \times \phi_a \times \pi \times (1 - p_a).$$

The calculation of the rates and the simulation of the resulting model splitting then

follow the techniques in section 5.3.4. As before, both the observed *multi-parent* element $n_{a01,2}$ and the simulated convolved components of the observed and unobserved juveniles ($u^*_{1(3j1),2} + u^*_{3(3j1),2}$ and $u^*_{1(3j0),2} + u^*_{3(3j0),2}$) are split across the tree emanating from $n_{j0,1}$ and $n_{a0,1}$. Using the same argument as in section 5.3.4, the rates for all four of these simulated splits will be equivalent.

Once these remaining elements of the intermediate vector $\mathbf{u}^*_{3,2}$ have been simulated the usual deterministic procedure can occur with the appropriate summations of $\mathbf{u}^*_{3,2}$ enabling the other intermediate vectors $\mathbf{u}_{2,2}$ and $\mathbf{u}_{1,2}$ to be specified. All that remains is to simulate the unobserved elements of the state vector at time $t = 1$, $\mathbf{n}_1$, and this is discussed in the following section.

5.4.4 Simulating Unobserved Parent Nodes

From section 5.4.1 it can be seen that the vector $\breve{\mathbf{s}}_1$ of unobserved elements of the state vector $\mathbf{n}_1$ consists of just two elements; the juvenile animals that were not caught during the first sampling occasion and the adult animals that were alive but were not caught during the first sampling occasion. Hence,

$$\breve{\mathbf{s}}_1 = [n_{j0,1}, n_{a0,1}] \,.$$

Once again, as described in section 5.2.5, there are no elements in $\mathbf{s}_0$ and thus no constraints on the simulated values of $\breve{\mathbf{s}}^*_1$. Instead the approximate negative binomial technique introduced in section 5.2.5 is used here with $\mathbf{u}_{1,1}$ and $\mathbf{u}_{3,1}$ acting as the minimum constraints on $n_{j0,1}$ and $n_{a0,1}$ respectively.

Having completed this step all elements of the state vector $\mathbf{n}^*_1$ and the associated intermediate vectors describing the population after survival $\mathbf{u}^*_{1,2}$, birth and maturation

$\mathbf{u}^*_{2,2}$ and capture $\mathbf{u}^*_{3,2}$ have now been simulated.

5.4.5 Evaluating the Densities

As in the previous sections the densities associated with the fitting process from $t = 2$ to $t = 1$ detailed in the preceding section now need to be evaluated. The density for the state process distribution is calculated in precisely the same manner as in section 5.2.7. The usual pair of multinomial distributions are evaluated for each tree generated by the elements of $\mathbf{n}_1$ and their product is taken across all trees to give the density $g_2\left(\mathbf{n}_2|\mathbf{n}_1\right)$.

To evaluate the trial density, $h_2(\mathbf{u}^*_{3,2}, \breve{\mathbf{s}}^*_1|\mathbf{n}^*_2, \mathbf{y}_1)$, the expanded vector is the usual combination of the intermediate vector denoting the capture status at time 2: $\mathbf{u}^*_{3,2}$ and $\breve{\mathbf{s}}^*_1$ the unobserved elements of the state vector at time $t = 1$. Then, the techniques of section 5.3.5 are replicated and the probability functions for the stochastic processes described in sections 5.4.2:5.4.4 are evaluated. Those processes can be summarised as follows:

a) Section 5.4.2 - splitting both the observed and unobserved juveniles alive during sampling at $t = 2$ over individual and convolved trees.

b) Section 5.4.3 - splitting the unobserved adults that were alive but not captured during sampling at $t = 2$ over convolved nodes.

c) Section 5.4.4 - simulating non-convolved parent nodes corresponding to animals that were alive but not captured during sampling at $T - 1$ for trees emanating from unobserved ancestral nodes.

A comparison with the processes listed in sections 5.2.7 and 5.3.5 reveals the number of separate processes required decreases over (reverse) time in line with the dimension of the expanded state vector. With fewer elements in the expanded state vector to simulate

fewer processes are needed to specify the trial density at time t: $h_t()$.

The final section of the model fitting algorithm maps the transition between the states at time $t = 1$ and the initial states at time $t = 0$. The details of this procedure are given in the next section.

5.5 Simulation from 1 to 0

As discussed in section 3.3.3 the Bayesian approach to formulating and fitting this form of state-space model requires a prior distribution to be specified for the parameters of the model and the initial states $\mathbf{n}_0$. Therefore, the state elements $n^*_{j0,0}$ and $n^*_{a0,0}$ can be assumed to be known and thus provide a further constraint to the simulated values of the intermediate nodes $\mathbf{u}_{1,1}$, $\mathbf{u}_{2,1}$ and $\mathbf{u}_{3,1}$.

The simulation, from the prior, of the state vector $\mathbf{n}_0 = \left[n^*_{j0,0}, n^*_{a0,0}\right]'$ can be implemented in several ways. The simplest is to simply specify independent priors on each of the two elements and simulate from those distributions. An alternative approach would be to specify a prior on the total number of animals, both juveniles and adults, that existed at time $t = 0$ and then to split this total into two categories: juveniles and adults. This latter approach requires the specification of a splitting parameter, ν, that will determine how the total animals are distributed across the two elements. Hence, if N_0 denotes the total number of animals at time $t = 0$, then the split can be simulated using the following distribution:

$$n^*_{j0} \sim \text{Binomial}\left(N_0, \nu\right). \tag{5.5.1}$$

Once n^*_{j0} is simulated it is clear to see that n^*_{a0} is obtained deterministically as $n^*_{a0} = N_0 - n^*_{j0}$.

Given that the state vectors $\mathbf{n}_1^*$ and $\mathbf{n}_0^*$ have now both been simulated the next step is to simulate the intermediate nodes.

5.5.1 Splitting across convolutions

The evolution of the population between $\mathbf{n}_0$ and $\mathbf{n}_1$ via the intermediate sub-processes was represented in the tree diagram in Figure 4.6. From this diagram and the known structure of the state vector $\mathbf{n}_1$ the relationship between $\mathbf{u}_{3,1}$ and $\mathbf{n}_1$ can be seen to be the following:

$$
\begin{aligned}
n_{j0,1} &= u_{1(3j0),1} + u_{2(3j0),1} \\
n_{j1,1} &= u_{1(3j1),1} + u_{2(3j1),1} \\
n_{a0,1} &= u_{1(3a0),1} + u_{2(3a0),1} \\
n_{a1,1} &= u_{1(3a1),1} + u_{2(3a1),1}.
\end{aligned}
$$

From these equations it can be seen that each element of $\mathbf{n}_1^*$ is a *multi-parent* element as it is convolved over two elements of the intermediate vector $\mathbf{u}_{3,1}$. Since both parent elements $n_{j0,0}^*$ and $n_{a0,0}^*$ are assumed known after simulation these *multi-parent* elements can be split using the techniques described in sections 5.2.2 and 5.3.4. The splits of the adults and juveniles can be considered separately as in the previous sections. Firstly the observed adults $n_{a1,1}^*$ are split across $u_{1(3a1),1}^*$ and $u_{2(3a1),1}^*$ by drawing from the appropriate non-uniform discrete distribution using the Alias method as in section 5.2.2. The split of the unobserved, but previously simulated, adults $n_{a0,1}^*$ across $u_{1(3a0),1}^*$ and $u_{2(3a0),1}^*$ is then conditional on the simulated split of $n_{a1,1}^*$ and proceeds using the approach described in section 5.3.4. The same approach holds for the juveniles with the split of the unobserved juveniles $n_{j0,1}^*$ being simulated conditional on the split of the observed juveniles $n_{j1,1}^*$.

Having simulated the splits of all four *multi-parent* elements of $\mathbf{n}_1$ the intermediate vector $\mathbf{u}_{3,1}$ is fully specified and the elements of the remaining intermediate vectors can be specified deterministically using the relationships illustrated in Figure 4.6.

5.5.2 Evaluating the densities

The evaluation of both the state and trial densities is relatively straightforward for the final time period. The usual approach for evaluating the density under the state process model as described in section 5.2.7 is applied again here. Each of the two elements in $\mathbf{n}_0$ generated trees that represent the stochastic processes that model the evolution of the population from $t = 0$ to $t = 1$. The processes for a single tree are modelled using the pair of multinomial distributions defined in Eq. (5.2.11a) and Eq. (5.2.11b). The density $g_1(\mathbf{n}_1|\mathbf{n}_0)$ is then obtained by taking the product over the four multinomials used to model the two trees.

The trial density can be expressed in the following form:

$$h_1(\mathbf{u}^*_{3,1}, \mathbf{n}^*_0|\mathbf{n}^*_1, N^*_0).$$

The vector $\mathbf{n}^*_0$ can be used instead of $\breve{\mathbf{s}}^*_0$ in the above density expression as they are equivalent. No animals are captured in year 0 and thus $\breve{\mathbf{s}}^*_0$, which denotes the abundances for all unobserved capture-history patterns that do not include capture in year 0, is equivalent to the state vector at time 0: $\mathbf{n}^*_0$. The simulation of $\mathbf{n}_0$ is necessarily conditional on the value of N^*_0 drawn from the prior and this is specified explicitly in the trial density. The expanded vector required for the evaluation of the trial density is then the combination of the final intermediate vector at time 1: $\mathbf{u}^*_{3,1}$ and the initial state vector $\mathbf{n}^*_0$.

For fitting the observations in the initial time period of this model there are only two

sets of stochastic processes that are used to simulate the unknown state elements under the trial density. Firstly, the splitting process used to apportion N_0^* to $n_{j0,0}^*$ and $n_{a0,0}^*$ and secondly those processes involved in splitting the *multi-parent* elements of $\mathbf{n}_1$ across the intermediate vector $\mathbf{u}_{3,1}$. The techniques of section 5.3.5 are then used to evaluate the densities for each process and the product of these yields the component of the trial density for this initial time period: h_1.

5.6 Calculating the Weights

The final processes described in section 5.5 completed the simulation of the elements of all state and intermediate vectors across the entire time series from $t = 0$ to $t = T$. The densities under the state process and the proposed trial process have also both been evaluated for the evolution of the population during each of the T time periods. These individual densities then need to be combined to give an overall density associated with the set of simulated state and intermediate vectors. The concept of a "particle" was introduced in section 3.3.6 and the term is used here to define the set of state, intermediate and parameter vectors $\{\mathbf{n}_0, \mathbf{u}_{1,1}, \mathbf{u}_{2,1}, \mathbf{u}_{3,1}, \ldots, \mathbf{n}_{T-1}, \mathbf{u}_{1,T}, \mathbf{u}_{2,T}, \mathbf{u}_{3,T}, \Theta\}$. Then, for this particular particle, the density under the state process, $g(\mathbf{n}_{1:T}|\mathbf{n}_{0:T-1})$, is obtained by taking the product of the state densities over all time periods. Using the notation defined in Eq. (5.2.11) this can be expressed as follows:

$$g(\mathbf{n}_{1:T}^*|N_0) = \prod_{t=1}^{T} g_t(\mathbf{n}_t^*|\mathbf{n}_{t-1}^*) = \prod_{t=1}^{T} \prod_{r=1}^{2^{t-1}+4} p(ad_{r,t}^*|n_{r,t-1}^*) p(nj_{r,t}^*|u_{r(1),t}^*). \tag{5.6.1}$$

Similarly, the density under the proposed trial process is obtained by taking the product over the trial densities evaluated for each time period:

$$\begin{aligned} h(\mathbf{u}^*_{3,1:T}, \breve{\mathbf{s}}^*_{\mathbf{0:T-1}} | \mathbf{y}_{1:T}, N_0) = & \, h_T(\mathbf{u}^*_{3,T}, \breve{\mathbf{s}}^*_{T-1} | \mathbf{y}_T, \mathbf{y}_{T-1}, \mathbf{y}_{T-2}) \\ & \times \prod_{t=3}^{T-1} h_t(\mathbf{u}^*_{3,t}, \breve{\mathbf{s}}^*_{\mathbf{t-1}} | \mathbf{n}^*_t, \mathbf{y}_{t-1}, \mathbf{y}_{t-2}) \\ & \times h_2(\mathbf{u}^*_{3,2}, \breve{\mathbf{s}}^*_1 | \mathbf{n}^*_2, \mathbf{y}_1) \\ & \times h_1(\mathbf{u}^*_{3,1}, \mathbf{n}^*_0 | \mathbf{n}^*_1, N^*_0). \end{aligned} \tag{5.6.2}$$

The general concept of deriving a weight associated with an individual particle was introduced in section 3.3.6 and was refined in the context of a model fitting approach that conditioned on the full series of observations over all time periods, $\mathbf{y}_{1:T}$, in section 4.4.2. From the general expression given in Eq (4.4.2), if the condition that $f(\mathbf{y}_{1:T} | \mathbf{n}^*_{1:T}) = 1$ is satisfied, the weight for a single particle is obtained using the following ratio of densities:

$$w \propto \frac{g(\mathbf{n}^*_{1:T} | N_0)}{h(\mathbf{u}^*_{3,1:T}, \breve{\mathbf{s}}^*_{\mathbf{0:T-1}} | \mathbf{y}_{1:T}, N_0)}.$$

Assume that a large number, N_p, of these particles are simulated using the approach described in this chapter. Omitting the superscript $*$ to make the notation clearer; using Eq's (5.6.1) and (5.6.2) the weight for the i^{th} ($i = 1, 2, \ldots, N_p$) particle can be expressed

as follows:

$$
\begin{aligned}
w^{[i]} \propto & \frac{g_T\left(\mathbf{n}_T^{[i]}|\mathbf{n}_{T-1}^{[i]}\right)}{h_T(\mathbf{u}_{3,T}^{[i]}, \breve{\mathbf{s}}_{T-1}^{[i]}|\mathbf{y}_T, \mathbf{y}_{T-1}, \mathbf{y}_{T-2})} \\
& \times \prod_{t=3}^{T-1} \frac{g_t\left(\mathbf{n}_t^{[i]}|\mathbf{n}_{t-1}^{[i]}\right)}{h_t(\mathbf{u}_{3,t}^{[i]}, \breve{\mathbf{s}}_{\mathbf{t-1}}^{[i]}|\mathbf{n}_t^{[i]}, \mathbf{y}_{t-1}, \mathbf{y}_{t-2})} \\
& \times \frac{g_2\left(\mathbf{n}_2^{[i]}|\mathbf{n}_1^{[i]}\right)}{h_2(\mathbf{u}_{3,2}^{[i]}, \breve{\mathbf{s}}_1^{[i]}|\mathbf{n}_2^{[i]}, \mathbf{y}_1)} \\
& \times \frac{g_1\left(\mathbf{n}_1^{[i]}|\mathbf{n}_0^{[i]}\right)}{h_1(\mathbf{u}_{3,1}^{[i]}, \mathbf{n}_0^{[i]}|\mathbf{n}_1^{[i]}, N_0^{[i]})}.
\end{aligned} \tag{5.6.3}
$$

These weights are then normalised such that

$$
w^{[i]} = \frac{w^{[i]}}{\sum_{i=1}^{N_p} w^{[i]}}
$$

although the actual calculation of these normalised weights is often performed on the log scale for computational reasons when dealing with very small weights for some particles. This is often the case for the bottom-up conditional approach as the weights are constructed from the product over the multiple densities specified under the trial density $h()$.

If these particles are then resampled using a single importance sampling step with their weights calculated as in Eq (5.6.3), the surviving particles will form an approximate sample from the posterior distribution of $g(\mathbf{n}_{1:T}|\mathbf{y}_{1:T}, \Theta)$. Equally, due to only a single importance sampling step being viable, an equivalent approach to obtaining estimates of individual parameters is to take the weighted average across all particles of the values of that parameter in the particles. The weights are, again, the importance sampling weights. This equivalent approach then avoids the need to resample and discard information present in the data. As noted in section 5.2.7, the bottom-up conditional approach

to simulating elements of state and intermediate vectors produces a smoothed approximation to the posterior distribution.

5.7 Extension to Multiple Births

The population dynamics model outlined in section 5.1 included a model for the birth process that allowed mature animals to produce, at most, a single young during the breeding period. Although appropriate for certain animal populations, this restriction is not always realistic and the model can be easily extended to incorporate multiple births. As mentioned in section 5.2.4, by extending the birth process to allow an individual animal to produce more than a single young in a breeding season, the construction of the trial density $h()$, and subsequent implementation of the fitting algorithm, becomes more complicated.

The underlying techniques used to constrain the simulated states are the same as those described in the chapter for the single birth model, however the details of the implementation do vary. Consider a birth process process that allows a single mature female to produce either 0, 1 or 2 young per breeding season with probabilities π_0, π_1 and π_1 respectively. Consider a single particle where $u_{1,t}$ represent an arbitrary element of the intermediate vector $\mathbf{u}_{1,t}$ that denotes the surviving animals from the state vector at time $t-1$, $\mathbf{n}_{t-1}$. Then, the distribution of the birth model can be specified as follows:

$$u_{2,t,[j]}|u_{1,t} \sim \sum_{z=1}^{u_{1,t}} \varphi \times \text{Multinomial}(1, c(\pi_0, \pi_1, \pi_2)) \tag{5.7.1}$$

where φ denotes the set of the possible number of juveniles produced by an individual mature animal: 0, 1 or 2 for this model. Simulating a draw from this multinomial model with yield a string of three values, two 0's and a 1 with the position of the 1 corresponding

to the element in φ. For example, a multinomial draw of 001 corresponds to an adult producing twins. Taking the product of φ with the draw from the multinomial then yields the number of young produced by that individual animal. This process is repeated for each individual animal in the element $u_{1,t}$ and the resulting simulated births are summed to give $u_{2,t,[j]}$, the total number of juveniles produced by the surviving animals in $u_{1,t}$. Equation (5.7.1) then replaces Equation (4.4.1c) in the earlier list of distributions for each component of the process model.

The implementation of the fitting algorithm using this new birth process model begins with the construction of expected values. The expected number of juveniles produced by a single mature female is $0 * \pi_0 + 1 * \pi_1 + 2 * \pi_2 = \pi_1 + 2 * \pi_2$. Thus, in the calculation of the expected values in the previous sections of this chapter the parameter π should be replaced by $\pi_1 + 2 * \pi_2$.

Further changes are necessary with regard to implementing the conditional generation algorithm (section 4.4.3). Consider again equations (4.4.9) and (4.4.10), used to simulate the number of survivors and births respectively. The range of valid values for survivors was defined as $\max(y, z) \leq w \leq v$. For the multiple birth situation the y captured juveniles could have been produced by fewer than y surviving adults. Hence, for the example in which adults can produce two young at most, the range of valid values for survivors is redefined as $\max(y/2, z) \leq w \leq v$. Similarly, for simulating births, the range of valid values is redefined as $y/2 \leq x \leq w$. The evaluation of $p(x|w)$ and $p(y|w)$ will also change in accordance with the new birth process model defined in Equation (5.7.1). For a particular value of w there will be multiple ways in which x juveniles could have been produced and each needs to be evaluated with summation over all such possible combinations giving the required probability. For example, if 10 adults produced 4 juveniles then, if a breeding

adult could produce no births, singletons or twins, this could be achieved by either four adults each having singletons, one adult producing twins and two producing singletons or two adults both producing twins. This increases the complexity of the evaluation procedure.

The construction of the trial density needs to be amended for every component that requires the evaluation of the probability of a certain number of juveniles being produced conditional on observed counts of capture history patterns. This approach may result in the confounding of the birth parameters π_1 and π_2 for populations that exhibit average birth rates of less than one young per breeding adult per time period. For this situation there may not be enough information to obtain good inference on the relative contributions of singletons and twins to the observed counts of capture history patterns. The study on the real dataset in chapter 7 allows this issue to be explored in more detail.

5.8 Discussion

One of the limitations to the conditional generation approach, as described in section 4.4.2, is that it does not generate particles according to the full conditional distribution. It is fully conditional for the evolution of the system within a single time period but not for the entire duration of the study. This meant that, for a particular particle, state elements could be simulated at some time period t_α that were not consistent with observations at some later time period t_β. That particular particle would therefore be implausible and would be assigned a weight of zero; it could not be included in the weighted resample of particles used to approximate the smoothed posterior distribution of states. The motivation behind the "bottom-up" approach to simulating the state and intermediate vectors was to develop a trial density that increases the probability of all constraints, as imposed

by the observed abundances of capture-history patterns, being met when compared to the approach described in section 4.4.2. However, the implementation of this "bottom-up" approach again uses proposal distributions that do not fully respect the conditioning on all observed abundances. The algorithm does not ensure that the state elements simulated at each time-step are consistent with the entire series of observed capture history patterns $\mathbf{y}_{1:T}$. Consequently, as with the approach in section 4.4.2, the bottom-up conditional approach to simulating the states can produce implausible particles. These implausible particles will be assigned zero weight and are therefore not included in the approximate posterior distribution. So, the method itself is fully conditional in that all particles with non-zero weight will respect the constraints. The issue arises with regard to the efficiency of the algorithm and the proportion of the particles simulated via the proposal density that do not fully respect the conditioning. Applications of this approach based on both real and simulated data are presented in Chapter 7.

Chapter 6

The Unconditional Generation Approach to Embedding Population Dynamics into Mark-Recapture Models

6.1 Introduction

The conceptual distinction between the conditional and unconditional approaches to modelling capture-recapture data within a state-space modelling framework was discussed in section 4.2. However, both approaches possess certain key commonalities. As with the conditional approach, the unconditional approach also allows for a population dynamics model to be embedded in the modelling framework and thus attempts to overcome some of the limitations with the existing open population models for mark-recapture data as discussed in section 4.1.

As was described in Section 3.3.2, formulating a state-space model requires the specification of models for the initial states as well as the state and observation processes at time t ($0 \leq t \leq T$). These probability distributions then map the evolution of the system over a period of time that was demarcated into discrete intervals with observations

or measures taken on the system at these discrete points. One approach to developing algorithms for fitting state-space models to mark-recapture data was developed under a set of modelling assumptions defined as the conditional approach; this is covered in detail in chapters 4 and 5. A salient feature of this approach is that the probability distribution used to specify the observation process is degenerate; the state elements corresponding to animals captured at time t are observed without error. Thus, the uncertainty in the model for period t relates only to the elements of the state vector at t which correspond to animals that are not caught during that period. As was illustrated in Section 5.1 the requirement to generate simulated particles that satisfy the constraints imposed by the observed measures at each time period complicates the model fitting process in comparison with the more traditional state-space modelling approaches and generally requires an 'ad hoc' approach to specifying the details of the fitting algorithm. The unconditional approach relaxes this restriction and, consequently, the model fitting process is somewhat simpler and can be generalised more easily.

Under the conditional generation approach to fitting state-space models to mark-recapture data the state elements were determined by an animal's capture history together with any other factors such as age or gender that could be used to specify the composition of the state vectors during each period t. The stochasticity in this approach then related, at each time period t, only to those abundances associated with capture history patterns that did not include capture during sampling at t. Consequently, the model parameters relating to the capture process appeared in the model for the state process rather than the observation process. In the unconditional approach this is no longer the case as there is no conditioning on the capture history data and the states are not determined by capture history pattern. Under this approach, stochasticity arises in the modelling framework via the relationship between the simulated states and the observed measures on those states.

This relationship is modelled by the observation process and the observed capture history patterns can be regarded as one of many possible realisations of the capture process. Hence, for the unconditional approach, the state and observation process models are now specified in the more traditional manner with the parameters relating to capture probability occurring in the observation process model only. Therefore this approach is more compatible with the traditional state-space modelling approaches reviewed in Section 3.

The unconditional approach is described, formalised and discussed in more detail in the following sections. In the following sections the terms "observed" and "captured" are used interchangeably.

6.2 Comparison to Existing Approaches

By incorporating a state-space modelling framework into the analysis of capture-recapture data the form of the likelihood representing the capture history patterns is a simplified version of that in more conventional mark-recapture analyses. To illustrate this point, consider the conventional mark-recapture likelihood for the Jolly-Seber models (Section 2.3.1). The required notation was defined in Tables 2.6 and 2.3. Consider again the three-sample study example used in section 2.3.1.1. The Jolly-Seber model formulation for the component relating to the unmarked animals that were first captured in the first sampling period was given in Eq. (2.3.1). The full Jolly-Seber likelihood for this three-sample study is obtained by taking the product over the likelihood components corresponding to unmarked animals captured for the first time on each sampling occasion. Hence, considering each component in turn it is necessary to separate the capture history patterns into sets corresponding to the time of first capture and thus condition on the time of first capture when constructing the likelihood component. Therefore, the

index ω_1 denotes the set of capture histories including capture on the first sampling occasion: $\{100, 101, 110, 111\}$. Similarly, $\omega_2 = \{010, 011\}$ and $\omega_3 = \{001\}$. Hence x_{ω_t} denotes the abundances associated with the capture history patterns in the set ω_t. Given these definitions, the three components can be expressed as follows:

$$\begin{aligned} Pr(\{x_{\omega_1}\}|u_1) = & \left[\frac{U_1!}{u_1!\,(U_1-u_1)!}p_1^{u_1}(1-p_1)^{U_1-u_1}\right] \\ & \times\left\{\frac{u_1!}{\prod_{\omega_1}x_{\omega_1}!}[\chi_1]^{x_{100}}[\phi_1(1-p_2)\phi_2p_3]^{x_{101}}\right. \\ & \left.\times\,[\phi_1p_2\chi_2]^{x_{110}}[\phi_1p_2\phi_2p_3]^{x_{111}}\right\} \qquad (6.2.1a) \\ Pr(\{x_{\omega_2}\}|u_2) = & \left[\frac{U_2!}{u_2!\,(U_2-u_2)!}p_2^{u_2}(1-p_2)^{U_2-u_2}\right] \\ & \times\left\{\frac{u_2!}{\prod_{\omega_2}x_{\omega_2}!}[\chi_2]^{x_{010}}[\phi_2p_3]^{x_{011}}\right\} \qquad (6.2.1b) \\ Pr(\{x_{\omega_3}\}|u_3) = & \left[\frac{U_3!}{u_3!\,(U_3-u_3)!}p_3^{u_3}(1-p_3)^{U_3-u_3}\right] \\ & \times\left\{\frac{u_3!}{\prod_{\omega_3}x_{\omega_3}!}[\chi_3]^{x_{001}}\right\} \qquad (6.2.1c) \end{aligned}$$

The value of u_t in the above equations is given by the sum of the abundances in the set x_{ω_t}. It can also be seen that χ_3, the probability that an animal released in the final sample period is not recaptured, is, by definition, unity; the animals released following capture in the final period cannot be re-captured as there are no future capture occasions in this study.

Then, the full Jolly-Seber likelihood is obtained as the product of equations (6.2.1a),(6.2.1b) and (6.2.1c):

$$L(\{p_1,p_2,p_3\},\{\phi_1,\phi_2\}|\omega_1,\omega_2,\omega_3) = \prod_{t=1}^{3} Pr(\{x_{\omega_t}\}|u_t).$$

This expression is for a single-age population in which no distinction is made between different age cohorts in terms of the parameterisation of the model. The t suffix allows

the parameter to vary by time period but every animal, regardless of age, will have a common set of parameters for a given time period.

An alternative formulation of the conventional mark-recapture likelihood can be obtained in which the distinction between the components relating to the capture process and those relating to survival is emphasised. The required notation is given in Table 6.1 and borrows heavily from the earlier definitions used in the Jolly-Seber likelihood above. The only significant difference is the use of s_t rather than n_t to denote the number of animals caught at the start of period t. This is done simply to avoid confusion with the vector notation $\mathbf{n}_t$ which denotes the state elements at time t.

N_t	=	number of animals in the population immediately before births in period t;
M_t	=	number of marked animals, before capture at t, in the population immediately before births in period t;
U_t	=	$N_t - M_t$ = number of unmarked animals in the population immediately before births in period t;
s_t	=	number of animals caught at the start of period t;
m_t	=	number of marked animals caught at the start of period t;
u_t	=	$s_t - m_t$ = number of unmarked animals caught at the start of period t;
r_t	=	number of animals captured and then released in period t and caught at some future occasion;
z_t	=	number of animals that are caught before period t, not caught during t but are recaptured at some future occasion;
χ_t	=	probability that an animal released in period t is not recaptured.

Table 6.1: Notation for the conventional mark-recapture model.

For this alternative formulation it is assumed that the population evolves according to the combination of survival and birth processes. The capture process then determines what proportion of the population is observed during each time period. As with the

conventional Jolly-Seber formulation it is assumed that each animal in the population has the same capture probability, p_t, in period t, and the same survival probability, ϕ_t, from year t to year $t+1$. Under these assumptions the mark-recapture model based on T capture occasions can be expressed as in Seber (1965) as a multinomial model which takes the following likelihood form (omitting constants):

$$\prod_{t=1}^{T} \frac{U_t!}{(U_t - u_t)!} p_t^{s_t} (1-p_t)^{z_t + U_t - u_t} \prod_{t=1}^{T-1} \phi_t^{r_t + z_t} \chi_t^{s_t - r_t}. \tag{6.2.2}$$

From comparison with Equation (6.2.1) it can be seen that the omitted constant terms are the products across all time periods of the factorials of the abundances for the observed capture history patterns: $\prod_{t=1}^{T} \prod_{\omega_t} x_{\omega_t}!$. These observed patterns are known exactly and therefore do not need to be estimated and can be excluded from the likelihood formulation. It can also be seen that, constants aside, the expressions in Equations (6.2.1) and (6.2.2) are equivalent. The utility of expressing the likelihood in the form of Equation (6.2.2) is that it allows a concise but flexible expression that can be used to fit state-space models to mark-recapture data. This will be discussed in more detail in section 6.5.

6.2.1 Population and Superpopulation modelling

As noted in section 4.2 the conditional approach is said to assume a "population model" in which the population is of fixed but unknown size whereas the unconditional approach may assume either a population model or a superpopulation model. That section provided a relatively brief description of the distinction between the "population" and the "superpopulation" approaches to modelling mark-recapture data. This can be expanded upon by first charting the existence of these terms in the mark-recapture literature.

In the context of analysing mark-recapture data the term "superpopulation" first appears in reference to the model formulation and parameterisation developed by Crosbie and Manly (1985). The term doesn't appear in that specific paper but rather in the work of Schwarz and Arnason (1996) that was based on the parameterisation first developed by Crosbie and Manly (1985). A definition of the distinction between the superpopulation and population approaches based on the work by Schwarz and Arnason (1996) is provided by Williams *et al.* (2002, 508-510). Kendall (1999) offers a similar definition of the terms. Adopting the definitions given in Williams *et al.* (2002, 508-510) a "superpopulation" is defined as the total number of animals present over the entire study, some fraction of which are available to be caught during each time period. This available fraction is then defined as the "population". In the Schwarz-Arnason approach the superpopulation N for a K sample study is given by the sum $\sum_{i=0}^{K-1} B_i$, where B_i denotes the number of new animals that are available to be sampled in time period $i+1$ that were not available to be sampled at time period i. These new animals can be thought of as recruitments or additions to the animals available to be sampled and will often be assumed to represent a realisation of some birth or immigration process. Conceptually the superpopulation N is some fixed, but typically unknown, constant. Also, the B_i can be thought to denote the members of this superpopulation that are assumed to become part of the population available to be sampled during the $(i+1)^{\text{th}}$ time period, with their entry probability depending on the entry parameters β_i. These entry probabilities are simply defined as $\beta_i = \frac{B_i}{N}$ which is just the relative frequency of total net births that enter the population between sample occasions i and $i+1$.

Therefore, under the unconditional approach, if a population model is assumed the resulting analysis will focus inference from the model fitting on parameter estimates that are based solely on the animals captured at each sampling occasion. The likelihood function

for the observations then simply equates to the observation process model; the observed data in the likelihood comprises only the animals captured at each sampling occasion, consequently only the parameters relating to the capture process need to be included. Alternatively, the approach to analysis based on a superpopulation model focuses the inference on parameter estimates that are now based on the model for unmarked animals entering the population and being available for capture as well as the observed numbers of captures on a given sampling occasion. Equally, the model needs to include a component for the marked and unmarked animals leaving the population. In this case, the likelihood function for the observations then needs to incorporate the process by which these unmarked animals enter the population and become available to be sampled. Therefore, under the unconditional approach, the distinction between the population and superpopulation models can be conceptualised as the difference in the components of their respective likelihoods. For the population model approach only the observation process model appears in the likelihood because the probability of observing is determined solely by the observation process. Each simulated realisation of the population assumes the same fixed population and therefore any differences in the samples of captured animals obtained during each realisation are due to the observation process. For the superpopulation model approach, the probability functions for both the observation and state processes are used to construct the likelihood because each realisation is assumed to correspond to a different population. Thus, for the superpopulation approach, the number of animals available to be sampled, as well as the sample itself, differs across realisations. Hence, any differences in the samples of captured animals obtained during each realisation are due to both the state and observation processes.

As noted in section 4.2, for the conditional approach the model formulation can only be conceptualised in terms of a "population model"; that is, the population being modelled

is of some fixed but unknown size. In this context a fixed population means that each simulated realisation of the population is assumed to correspond to the same true, but unknown, value at each particular time period. The population is not assumed to be fixed temporally, that is the population can vary in abundance over time. The observed abundances of captured animals at each time period then correspond to observations on a single population. For the conditional approach the observation process model is degenerate and due to the state elements being determined by capture history pattern the parameters associated with the capture process appear in the model for the state process rather than the observation process. The conditional approach then only allows a population model to be assumed due to the requirement to condition on the exact totals of animals observed at each capture occasion. If the capture process is allowed to be stochastic then the observed abundances of capture histories can be viewed as just a single realisation of all possible capture histories that could have obtained under the assumed model. Then, by assuming each realisation corresponds to a different population a superpopulation model can be used to analyse the observed data from a mark-recapture study.

The methods in the next section will focus on the construction and implementation of a fitting algorithm for analysing mark-recapture data using the unconditional approach assuming a population model.

6.3 Formulating the Model

The unconditional approach is relatively simple when compared to the conditional approach in both underlying concept and model formulation. As with the conditional approach the models to be fitted are constructed using the standard state-space methodology. Under the unconditional approach the models are then fitted to the observed capture

histories by essentially evaluating a single multinomial distribution. This is in marked contrast to the requirement to evaluate multiple densities for each of the component processes that constitute the trial density, as necessitated by the conditional approach.

The usual standard state-space model definitions of an initial state distribution, a state process distribution and an observation process distribution are required to define the assumed form of the underlying process. For consistency with earlier models the time index is defined as $t = 1, 2, \ldots, T$ and Θ represents the vector of all the parameters involved in the model from both the observation and state processes. More precisely, $\Theta = (\theta, \psi)$ where θ denotes the state process parameters and ψ denotes the observation process parameters. Framing the analysis within a Bayesian approach calls for a prior distribution $\varpi(\theta, \psi)$ to be specified on the model parameters. Using these definitions the distributions under the state space model formulation are:

$$\text{Initial state distribution} = g_0(\mathbf{n}_t|\mathbf{n}_0\theta) \quad (6.3.1a)$$

$$\text{State Process distribution} = g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \mathbf{n}_{t-2}, \ldots, \mathbf{n}_0, \theta) = g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \theta) \quad (6.3.1b)$$

$$\text{Observation Process distribution} = f_t(\mathbf{y}_t|\mathbf{n}_t, \psi). \quad (6.3.1c)$$

These equations are, as expected, virtually identical to those of Equation (3.3.1) with the only difference being the explicit parameterisation of the state and observation process distributions. The specification of Equation (6.3.1b) assumes a first-order Markov process: the states at time t depend only the states in the preceding time period $t - 1$. By relaxing this assumption and retaining a dependence on earlier years the models are no longer state-space but rather "hidden process" models.

Under the conditional approach to fitting these state-space models the capture history of an individual animal determined the state it was assigned to. One consequence of this was that the parameters corresponding to probability of capture $\{p_1, p_2, \ldots, p_T\}$ appeared in θ the set of state process parameters. For the unconditional approach capture is treated as a stochastic process that only enters the model through the specification of the probability distribution for the observation process. However, the form of the data used to fit the models under the unconditional approach means that the evaluation of the likelihood is dependent on both the observation process parameters ψ and the state process parameters θ.

Under the unconditional approach it is assumed that each animal that is captured on at least one occasion can be identified uniquely. Under this assumption each individual animal will be assigned an associated capture history pattern that spans the duration of the entire study $[0, T]$. As discussed in section 2.2, for each capture occasion the corresponding element of the capture history pattern for each animal will consist of a 1 if it was captured during that sampling occasion and a 0 if it was not. It is assumed that there are T sampling occasions during the study and that N animals have existed in the population during the study period, of which a total of M distinct individuals are marked on at least one occasion. The data consisting of the possible capture history patterns can then be represented using the format demonstrated in Table 6.2.

The "ID" column in Table 6.2 denotes the index value assigned to each individual animal that occurs in the population of animals available for capture at some point during the study. The first M rows represent the capture history patterns associated with each of the distinct animals observed on at least one occasion during the study. For example, consider the first three sampling occasions: the first row indicates that an animal was

ID	Sampling Occasion				
	1	2	3	$\cdots$	T
1	1	0	1	$\cdots$	0
2	0	0	1	$\cdots$	0
3	0	1	0	$\cdots$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$M-1$	1	0	1	$\cdots$	1
M	0	0	0	$\cdots$	1
$M+1$	0	0	0	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$N-1$	0	0	0	$\cdots$	0
N	0	0	0	$\cdots$	0

Table 6.2: Example of the structure of capture histories.

observed, and consequently first marked, during the first sampling occasion, was not captured during the second sampling occasion but was recaptured during the third sampling occasion. The first M rows will all contain at least a single 1 as these are the capture histories for the M individuals that were marked at least once during the study. The following $N-M$ rows will consist of 0s in every column as they represent the animals that were part of the population and were available for capture during the study but were not caught at any point. Therefore, N represents the fixed but unknown total population over the duration of the study and is one of the parameters that can be estimated during inference.

The total M capture history patterns for animals that were observed at least once are then summarised to give the abundances of animals corresponding to each of the possible distinct capture history patterns that include at least one capture. The number of animals that are assigned to the capture history corresponding to the null pattern, $\omega_1 = (0, 0, \ldots, 0)$ (Table 2.1), representing no captures during the study period, cannot

be known unless some necessary constraints are imposed. For example, the total number of animals alive in the study area immediately before the study began would need to be known and some strong assumptions about the sub-processes that drive the dynamics of the population over time would need to be made. If it can be assumed reliably that the population is completely closed; that is no animals can enter or leave the population during the study period then these assumptions, together with the knowledge of the initial number of animals will allow the number of animals that were present in the population at some point during the study period but never captured to be calculated. For many studied ecological systems, these assumptions can prove to be unrealistically restrictive and possessing that degree of knowledge about the population would render the mark-recapture experiment redundant.

6.4 The State Process

The state process under the unconditional approach operates in a manner similar to that outlined in Section 3.3 for the general formulation of state-space models. The difference between the state process for the unconditional approach compared to the conditional approach is that the sub-process modelling capture is no longer included in the state process. Instead, as with the traditional state-space model formulation, the capture process is incorporated into the model structure through the observation process.

The specification of the state process is consistent with that presented in Section 3.3. It is constructed from individual components for each assumed biological process which allows the state process to be specified in terms of a series of linked sub-processes that are assumed to operate sequentially with the output of one sub-process being the input to the following sub-process. Again, it shall be assumed that each specified sub-process operates on discrete time and that the order of the sub-processes remains the same across

each distinct time period.

In the following illustrative examples the dynamics of the population over time are assumed to be driven by two sub-processes: a survival process and a birth and maturation process. Both the survival and birth processes are modelled stochastically whereas maturation is specified as a deterministic process; if a juvenile animal survives from the previous time period then it will become mature and will be classified as either a yearling or an adult. This age-structure for the population under the unconditional approach is the same as that for the conditional approach section 4.4.1. The population is classified by age and is divided into three distinct, mutually exclusive categories: juveniles, yearlings and adults. Juveniles animals are those that are born in time period t immediately before sampling occasion t. Those uncaught juveniles that survive the post-capture period until the start of the next time period $t+1$ will mature and be classed as adults by the time of the sampling occasion $t+1$. Animals classified as yearlings immediately before the sampling occasion at $t+1$ will be defined as adult animals that were caught as juveniles in the sampling occasion during the time period t. Therefore, as was seen for the conditional approach, the distinction between "yearling" and "adult" exists as an artefact of the birth and capture processes and, as explained in section 4.4.1, allows the estimation of juvenile capture rates. More generally, the parameterisation and distributions for these processes are the same as those given for the conditional approach in section 4.4.

Therefore, for the unconditional approach there are only two sub-processes that drive the underlying dynamics of the model. Consequently, there will be only two intermediate vectors required to denote the state of the population after evolving according to the sub-process. This compares to the three intermediate vectors required for the conditional approach. The intermediate states of the system will be stored in the vectors $\mathbf{v}_{1,t}$ and $\mathbf{v}_{2,t}$

and they are defined as follows:

$$\begin{aligned} \mathbf{v}_{1,t} &= \text{the number of animals in } \mathbf{n}_t \text{ that survive through until time } t+1. \\ \mathbf{v}_{2,t} &= \text{the number of juveniles produced by the surviving animals } \mathbf{v}_{1,t} \text{ and age incrementation of the those animals in } \mathbf{v}_{1,t}. \end{aligned}$$

The state process distribution, as defined in section (6.3.1b), can then be expressed as $g_t = g_{1,t} * g_{2,t}$ where $g_{1,t}$ and $g_{2,t}$ are the probability distributions used to model the survival and birth processes respectively. Then, using the definitions of the intermediate state vectors $(\mathbf{v}_{1,t}, \mathbf{v}_{2,t})$ and the probability distributions of the sub-processes that generate them, the evolution of the initial population to time $t = 1$ is illustrated schematically in Figure 6.1.

$$\mathbf{n}_0 = \begin{bmatrix} n_{j,0} \\ n_{a,0} \end{bmatrix} \xrightarrow{g_{1,1}} \mathbf{u}_{1,1} = \begin{bmatrix} u_{1(1),1} \\ u_{2(1),1} \end{bmatrix} \xrightarrow{g_{2,1}} \mathbf{u}_{2,1} = \begin{bmatrix} u_{1(2),1} \\ u_{1(1),1} \\ u_{2(2),1} \\ u_{2(1),1} \end{bmatrix}$$

Figure 6.1: Evolution of the initial states $\mathbf{n}_0$ via the intermediate sub-processes.

The definitions for the subscripts in the notation for the state elements are the same as those specified in section 4.4.1. Comparing Figures 6.1 and 4.5 the equivalence between the intermediate state vectors for the evolution of the population from the initial states $\mathbf{n}_0$ via the survival and birth sub-processes under both the conditional and unconditional approaches can be seen. The equivalence can be expressed as $\mathbf{u}_{1,1} \equiv \mathbf{v}_{1,1}$ and $\mathbf{u}_{2,1} \equiv \mathbf{v}_{2,1}$. The difference between the approaches in terms of the structure of the intermediate vectors used to summarise the population after each sub-process is the lack of a capture process component in the composition of the state process under the unconditional approach.

Again, adopting the approach of section 4.4.1, the tree diagrams in Figure 6.2 provide a schematic representation of the evolution of the population over the first time period. The rate parameters associated with the survival and birth processes are next to the relevant branches of the trees.

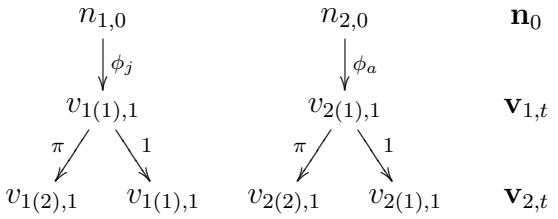


Figure 6.2: Evolution from $\mathbf{n}_0$ to $\mathbf{n}_1$ for the unconditional approach

The schematic representation in Figure 6.2 can be expressed in terms of the distributions of the sub-process. The notation of section 4.4.1 is once again replicated to highlight the distinction between the survival rates ϕ_j and ϕ_a. Let $\mathbf{n}_t^{[j]}$ and $\mathbf{n}_t^{[a]}$ represent the set of elements in the state vector $\mathbf{n}_t$ that correspond to the abundances of juvenile and adult animals respectively at time t. Then the distributions of the sub-processes can then be written as:

$$u_{1,t}^{[j]} | n_{t-1}^{[j]} \sim \text{Binomial}(n_{t-1}^{[j]}, \phi_j) \tag{6.4.1a}$$

$$u_{1,t}^{[a]} | n_{t-1}^{[a]} \sim \text{Binomial}(n_{t-1}^{[a]}, \phi_a) \tag{6.4.1b}$$

$$u_{2,t}^{[j]} | u_{1,t} \sim \text{Binomial}(u_{1,t}, \pi) \tag{6.4.1c}$$

$$u_{2,t}^{[a]} | u_{1,t} \equiv u_{1,t}. \tag{6.4.1d}$$

Using Equations (6.4.1c) and (6.4.1d), Figure 6.2 and the equivalence between $\mathbf{v}_{2,1}$ and

$\mathbf{u}_{2,1}$ it can be seen that the even numbered elements of $\mathbf{v}_{2,1}$ correspond to the survivors from the animals represented in $\mathbf{n}_0$ and the odd numbered elements correspond to the new-born juveniles produced by these survivors. This equivalence also means that the definitions of the elements of the intermediate vectors are the same as those given previously in Table 4.1.

Having examined the evolution of the population over the first time period it is constructive to examine the structure of the state vector at future time periods. Figure 6.3 represents the state vectors for the initial animals $t = 0$ and for the population at the end of the first and second time periods.

$$\mathbf{n}_0 = \begin{bmatrix} n_{0,j} \\ n_{0,a} \end{bmatrix} \rightarrow \mathbf{n}_1 = \begin{bmatrix} n_{1,jj} \\ n_{1,ja} \\ n_{1,aj} \\ n_{1,aa} \end{bmatrix} \rightarrow \mathbf{n}_2 = \begin{bmatrix} n_{2,jjj} \\ n_{2,jja} \\ n_{2,jaj} \\ n_{2,jaa} \\ n_{2,ajj} \\ n_{2,aja} \\ n_{2,aaj} \\ n_{2,aaa} \end{bmatrix}$$

Figure 6.3: The elements of the state vectors $\mathbf{n}_t$ for $t = 0, 1, 2$ for the unconditional approach. The rightmost character of the subscripts on each element denotes current age class: either juvenile j or adult a. The preceding characters denote the ancestry of the animal by giving the time of entry of its ancestors. A jja suffix denotes animals that are adults at time 2 that were born in time 1 to animals that were initially juveniles (at time 0).

From a comparison between Figures 6.1 and 6.3 it can be seen that there is an equivalence between the state vector at time 1, $\mathbf{n}_1$, and the intermediate states that summarise

the population after the first survival and birth processes, $\mathbf{u_{2,1}}$. Hence,

$$\begin{bmatrix} u_{1(2),1} \\ u_{1(1),1} \\ u_{2(2),1} \\ u_{2(1),1} \end{bmatrix} \equiv \begin{bmatrix} n_{1,jj} \\ n_{1,ja} \\ n_{1,aj} \\ n_{1,aa} \end{bmatrix}$$

and this relationship, that $\mathbf{u_{2,t}} \equiv \mathbf{n}_t$, holds for any time period t. Therefore, there is no need to perform any summing of elements of $u_{2,t}$ to obtain the state vector $\mathbf{n}_t$. In short, using the parlance of the conditional approach, there are no convolutions in the correspondence between the "parent" and "child" nodes. Each element of the state vector $\mathbf{n}_{t-1}$ will produce two "child" node elements in $\mathbf{u}_{2,t}$ and equivalently $\mathbf{n}_t$. Thus, with the initial vector $\mathbf{n}_0$ consisting of two elements it is easily seen that the state vector at time t will consist of 2^{t+1} elements.

With regard to the notation used in Figure 6.3, the elements of the state vectors represent the abundances of animals classified by their ancestry. The rightmost character of the subscripts on each element in the state vectors denotes the current age class of the animal, indicating whether it is a juvenile j or an adult a. The preceding characters then denote the ancestry of the animal by denoting the times periods during which its ancestors entered the population. For example, the subscript jja would represent an animal that is an adult during $t = 2$ but was born in the previous time period $t = 1$ to an animal that was a juvenile at the start of the study $t = 0$. As will be discussed in section 4.4.2, recording the time of entry into the population is necessary to correctly evaluate the probabilities of occurrence for each of the capture history patterns. This requirement to distinguish between groups of animals by their ancestral paths prevents combining together animals that are subject to the same model (i.e. share common parameters) from their time of entry into the population. This therefore adds a necessary

level of complexity to the classification of the elements in the state vectors.

6.5 The Observation Process

One of the key distinctions between the unconditional and conditional approaches is the difference in the structure of the data that is used in the model fitting process. As described in Section 6.3, the full series of observations at time t, $\mathbf{y}_t$, consists of the abundances for the capture history patterns that include capture on at least one occasion between $0:t$. This contrasts with the conditional approach in which the vector of observations at time t, $\mathbf{y}_t$, consisted of the known abundances of animals with capture history patterns that included capture in year t.

The observations in $\mathbf{y}_t$ are obtained by considering the capture histories exhibited by each individual animal that was observed at least once during the first t capture occasions and then summing across common capture history patterns. Thus, the first M rows of the matrix of capture histories, as illustrated in Table 6.2, provide the necessary information to construct the vector of abundances of capture history patterns containing at least once capture. Care should be taken to distinguish between juveniles, yearlings and adults for the population structure described in section 6.4. This issue is discussed in detail in section 6.6.3. The observation vectors specified under the conditional and unconditional approaches are represented in Figures 6.4 and 6.5 respectively.

Comparing Figure 6.5 with Figure 6.4 it can be seen that by relaxing the requirement to condition on the animals that were observed at each time period the dimension of the observation vector for time t is increased. It should also be noted that the capture history patterns that do not contain a single capture are, by definition, unobserved throughout

$$\mathbf{y}_1 = \begin{bmatrix} n_{j1,1} \\ n_{a1,1} \end{bmatrix} \rightarrow \mathbf{y}_2 = \begin{bmatrix} n_{j1,2} \\ n_{y11,2} \\ n_{a01,2} \\ n_{a11,2} \end{bmatrix} \rightarrow \mathbf{y}_3 = \begin{bmatrix} n_{j1,3} \\ n_{y11,3} \\ n_{a001,3} \\ n_{a011,3} \\ n_{a101,3} \\ n_{a111,3} \end{bmatrix}$$

Figure 6.4: The observation vectors $\mathbf{y}_t$ for $t = 1, 2, 3$ for the conditional approach

$$\mathbf{y}_1 = \begin{bmatrix} n_{j1,1} \\ n_{a1,1} \end{bmatrix} \rightarrow \mathbf{y}_2 = \begin{bmatrix} n_{j1,2} \\ n_{y10,2} \\ n_{y11,2} \\ n_{a01,2} \\ n_{a10,2} \\ n_{a11,2} \end{bmatrix} \rightarrow \mathbf{y}_3 = \begin{bmatrix} n_{j1,3} \\ n_{y10,3} \\ n_{y11,3} \\ n_{a001,3} \\ n_{a010,3} \\ n_{a100,3} \\ n_{a011,3} \\ n_{a101,3} \\ n_{a110,3} \\ n_{a111,3} \end{bmatrix}$$

Figure 6.5: The observation vectors $\mathbf{y}_t$ for $t = 1, 2, 3$ for the unconditional approach

the study and are therefore not included in the observation vectors. Thus, the juveniles that are not detected during the sampling process at time t, $n_{j0,t}$, and the adult animals with the null capture history pattern, $n_{a0\ldots0,t}$, are the only capture history patterns not to be incorporated into the observation vector. The reduced dimension of the vectors of observations under the conditional approach results from the requirement to condition on observed capture history pattern abundances that are known with certainty. This can be explained by examining the relationship between the state and observation vectors under both the conditional and unconditional approaches.

For the unconditional approach the uncertainty in period t relates to the probability of capture which is modelled through the observation process at t as defined in (6.3.1c).

The capture process, as defined in section 4.4, is then used to model the relationship between the elements of the state and observation vectors at time t. The stochastic nature of this binomial capture process thus introduces uncertainty at each time period t into the correspondence between the states $\mathbf{n}_t$ and the observed capture history patterns $\mathbf{y}_t$. This uncertainty means that the form of the correspondence between the elements of $\mathbf{n}_t$ and $\mathbf{y}_t$ is less obvious under the unconditional approach than it was under the conditional approach. For the conditional approach there was a direct correspondence between the individual elements of the vectors; the elements of the observed vector corresponded, either exactly or through summation, to the state elements that represented abundances of observed animals. This relationship thus determinined the constraints for the conditional generation algorithm.

For the unconditional approach the state vector, $\mathbf{n}_t$, simply gives the numbers of animals, as categorised by their ancestry, that are available to be captured immediately after the new juveniles enter the population during time period t. As indicated in Table 6.2, the observed capture histories can be expressed in matrix form and consist of M rows with each row corresponding to the capture history for a unique individual caught at least once. The key point about this representation of the population is that a 0 in the i^{th} element of the row simply indicates that the animal was not captured during the i^{th} sampling occasion. It does not distinguish between those animals that were present in the population but evaded capture and those animals that were not available to be captured (i.e. they had died before the i^{th} capture occasion or were born after the i^{th} capture occasion). The state vector only includes the numbers of animals that are alive during each period hence the disparity between the observation and state vectors. It is this disparity that is modelled by the observation process. The fitting algorithm will then need to account for animals that have died at some stage during the study yet still make

a contribution to the observed abundances of capture history patterns.

The likelihood formulation for conventional mark-recapture models was discussed in section 6.2. From the expression of Equation (6.2.2) it can be seen that the likelihood includes the parameters relating to both the survival and capture processes. By applying the unconditional approach within a state-space modelling framework the likelihood for the observations becomes simpler when compared to the conventional likelihood. The survival and birth processes do not need to be specified as part of the observation process model because they are already accounted for in the state process model. Hence, the state model determines the abundances of animals assigned to each state in the population with the observation model then determining the relationship between the underlying states and the measures taken on these states. For example, if it is assumed that all animals in the population can be grouped into a single state at each period t, the likelihood given in Equation (6.2.2) can be re-written as:

$$L\left(\psi|s_1,\ldots,s_T\right)=\prod_{t=1}^{T}\binom{N_t}{s_t}(p_t)^{s_t}(1-p_t)^{N_t-s_t} \tag{6.5.1}$$

where $p_t \equiv p_t(\psi)$ is defined to allow the probability of capture to be a function of the observation model parameters ψ.

The likelihood of the observations in Equation (6.5.1) is constructed from the product of components, with each component corresponding to a single time period t. Each of these components represents the probability of observing the number of animals caught at the start of a period, s_t, given the number of animals in the population available to be captured in that period, N_t. This is simply the observation process model, as defined in

Equation (6.3.1c), therefore:

$$f_t(\mathbf{y}_t|\mathbf{n}_t,\psi) = f_t(s_t|N_t,\psi) = \binom{N_t}{s_t}(p_t)^{s_t}(1-p_t)^{N_t-s_t}.$$

This specification of the likelihood of the observation allows the model to be fitted using any of the model algorithms described in Section 3; specifically, the approaches developed in Buckland *et al.* (2004), Thomas *et al.* (2005), Newman *et al.* (2006) and Buckland *et al.* (2007).

The simple likelihood of Equation (6.5.1) requires the assumption that all animals can be grouped into a single state. This means the observation vector $\mathbf{y}_t$ simply becomes a scalar s_t which is defined, as in Table 6.1, as the number of animals caught at the start of period t. Also, the state vector $\mathbf{n}_t$ consists only of a single element and thus was written as the scalar N_t. The value of N_t denotes the total number of animals in the population at the beginning of period t and this is determined by the state model. This simple model can be readily expanded to the vector case. Consider state and observation vectors consisting of K states such that

$$\mathbf{n}_t = \begin{bmatrix} n_{1,t} \\ n_{2,t} \\ \vdots \\ n_{K,t} \end{bmatrix} \qquad \text{and} \qquad \mathbf{y}_t = \begin{bmatrix} s_{1,t} \\ s_{2,t} \\ \vdots \\ s_{K,t} \end{bmatrix}.$$

The resulting likelihood for the observations can now be expressed as:

$$L\left(\psi|\mathbf{y}_1,\ldots,\mathbf{y}_T\right) = \prod_{t=1}^{T} f_t(\mathbf{y}_t|\mathbf{n}_t,\psi) \tag{6.5.2}$$

where the observation process for period t is defined as:

$$f_t(\mathbf{y}_t|\mathbf{n}_t,\psi) = \prod_{k=1}^{K} \binom{n_{k,t}}{s_{k,t}} (p_{k,t})^{s_{k,t}} (1-p_{k,t})^{n_{k,t}-s_{k,t}}. \tag{6.5.3}$$

For this k-state model $p_{k,t}$ is defined as the probability of an animal in state k being captured in period t. The definition of $p_{k,t}$ is an extremely flexible one. As discussed in section 2.2.1.3 the probability of capture can be defined to be either constant over time and state, $(p_{k,t} = p)$, constant over state $p_{k,t} = p_t$, constant over time $p_{k,t} = p_k$ or, as currently defined, dependent on both. Equally, as discussed in section 2.2.5, the probability of capture can be modelled as a function of covariates, such as previous capture history for example, which can include auxiliary data if available. Alternatively, a random effects model could be used to model the capture rates.

The formulations of the likelihoods in Equations (6.5.1) and (6.5.2) serve to illustrate the distinction between the observation processes under the conditional and unconditional approaches. For the unconditional approach the observed captures ($\mathbf{s}_t$ and $\mathbf{y}_t$ in Equations (6.5.1) and (6.5.2) respectively) are included in the likelihood rather than the state process and therefore do not influence an animal's state.

6.6 Fitting the Models

This section describes an alternative approach to formulating the likelihood for the observed data using the unconditional approach and presents a detailed explanation of the implementation of the model fitting algorithm. This unconditional fitting algorithm builds on the existing techniques for state-space and mark-recapture modelling approaches discussed earlier. Comparing the conditional and unconditional approaches it can be seen that the model formulation under the conditional approach is closer in structure to the

N	=	Total number of animals that have formed part of the population at any time during the interval $[0, T]$
ω	=	A zero-one indicator vector denoting the capture history: $\omega = \left(\delta_1^{(\omega)}, \delta_2^{(\omega)}, \ldots, \delta_T^{(\omega)}\right)$ where $\delta_i^{(\omega)} = 1$ or 0 for $i = 1, \ldots, T$ if the animal was caught on the i^{th} occasion or not. There are 2^T possible patterns, one of which represents the null pattern: $\omega_1 = (0, 0, \ldots, 0)$ i.e. those animals that were present at some stage but never caught.
x_ω	=	Number of animals having the capture history pattern ω

Table 6.3: Definitions and notation for the unconditional approach.

state-space model as opposed to conventional mark-recapture models. The reverse is true for the unconditional approach. The fitting process centers on the construction of a multinomial likelihood that models the partitioning of the total population over the study into the distinct capture history patterns. It is the manner in which the rates for this multinomial likelihood are calculated that marks a departure from existing methods of fitting models to mark-recapture data.

To begin describing the fitting algorithm it is first necessary to formally define the terms used. The notation in Table 6.3 shares much in common with the definitions given in Table 2.1 and Equation (6.2.1). This reflects the strong parallels between the basic data structure for the unconditional approach and conventional mark-recapture models.

From the definitions in Table 2.1 it can be seen that the total number of animals that have appeared in the population throughout the duration of the study is given by the abundances for each capture history pattern summed across all possible patterns. That is

$$N = \sum_{j=1}^{2^T} x_{\omega_j} \tag{6.6.1}$$

where x_{ω_j} denotes the abundance for the j^{th} capture history pattern ω_j. For a T sample study there are 2^T distinct and mutually exclusive capture history patterns. The

implementation of the unconditional fitting algorithm then requires that the probabilities of each capture history pattern occurring need to be evaluated. The technique used to do this is discussed in the next section.

6.6.1 Evaluating Probabilities

To illustrate this component of the fitting algorithm the example model as introduced in section 6.4 will again be used. The probabilities for each capture history pattern are determined by the combined effect of the state and observation processes. Thus, the first step is to obtain a population of N animals by generating a single realisation of the state vectors for the entire study, $\mathbf{n}_{0:T}$. As with the conditional generation approach, the initial states can be obtained via simulation from a prior using the methods described in section 5.5. Given this starting point sequential applications of the survival and birth processes, using the distributions in Equations (6.4.1a):(6.4.1d), for each time period are then performed to simulate the remaining states and intermediate vectors as illustrated in Figures 6.1 and 6.3.

Given the full set of state vectors $\mathbf{n}_{0:T}$ the total number of animals that were available to be captured at some point during the study is given by summing the components of the initial states, $\mathbf{n}_0$, and all components of the states at time t that correspond to juvenile animals born in that time period. As can be seen from Fig 6.1, for this example, the state vector at time $t, (t > 0)$ contains 2^t elements of which the odd-numbered ones denote the new-born juveniles. Hence,

$$N = n_{j,0} + n_{a,0} + \sum_{t=1}^{T} \sum_{i=1}^{2^t} n_{2i-1,t}. \tag{6.6.2}$$

The next required step is to establish the correspondence between the components of

$n_{jaa,2} - n_{jaaa,3}$	=	Animals that were juveniles at $t = 0$ and died before the third capture occasion.
$n_{aaa,2} - n_{aaaa,3}$	=	Animals that were adults at $t = 0$ and died before the third capture occasion.
$n_{jaaa,3}$	=	Animals that were juveniles at $t = 0$ and were available for capture during the third capture occasion.
$n_{aaaa,3}$	=	Animals that were adults at $t = 0$ and were available for capture during the third capture occasion.
$n_{jja,2} - n_{jjaa,3}$	=	Animals that were born during $t = 1$ to an initial juvenile and died before the third capture occasion.
$n_{aja,2} - n_{ajaa,3}$	=	Animals that were born during $t = 1$ to an initial adult and died before the third capture occasion.
$n_{jjaa,3}$	=	Animals that were born during $t = 1$ to an initial juvenile and were available for capture during the third capture occasion.
$n_{ajaa,3}$	=	Animals that were born during $t = 1$ to an initial adult and were available for capture during the third capture occasion.

Table 6.4: The number of animals, by state element, that can exhibit a 110 capture history pattern.

Equations (6.6.1) and (6.6.2). Effectively this is equivalent to establishing the correspondence between the state elements $\mathbf{n}_t$ and the observations $\mathbf{y}_t$ at each time period. The elements of $\mathbf{n}_t$ denote animals that are alive at t, whereas the abundances for the capture history patterns denoted in y_t include animals that died before the capture process during time t. Hence the equivalence between the components involves summing over those animals that die but still exhibit the particular capture history pattern of interest. For example, consider the pattern 110. For a population structured with two age classes: juveniles and adults, there are eight combinations of state elements that could possibly exhibit this pattern. These combinations then sum to give the total number of animals that could exhibit a 110 capture history pattern. The elements of $\mathbf{n}_2$ and $\mathbf{n}_3$ that are combined for this pattern are given in Table 6.4. Summing these elements gives the following relationship:

$$x_{110} = n_{jaa,2} + n_{aaa,2} + n_{jja,2} + n_{aja,2}.$$

Thus, by summing the eight components the expression becomes simplified as there is no cumulative contribution from the elements in $\mathbf{n}_3$. Including all eight combinations may seem somewhat redundant as the same information is contained in just four elements of $\mathbf{n}_2$. In terms of determining the correspondence between the state elements and the observed abundances for the capture history patterns this specification is indeed redundant. However, including all possible combinations is necessary to calculate the rates of the multinomial likelihood for the observed data.

The rates for the multinomial likelihood are the set of probabilities $\left\{r_{\omega_1}, \ldots, r_{\omega_{2^T}}\right\}$ where the probability r_{ω_j} corresponds to the capture history pattern ω_j. There are several equivalent approaches to calculating these rates with the equivalence determined by the specification of those parameters ψ used by the capture process. The most general approach is perhaps most easily explained by considering each individual animal that occurred in the population simulated under the state process. During this implementation of the state process via stochastic simulation, the point at which the animal entered the population, t_{ent} and the point at which it departed the population, t_{ex}, should both be stored. This can be achieved by assigning an index value to each animal in a new cohort of juveniles that enter the population in each time period. The index values corresponding to the animals from these cohorts that then fail to survive to the next period are stored with this process being repeated for each time period. Thus, once the state process has simulated the population of N animals over the study period, these N individuals are grouped into cohorts defined by their particular entry and exit times. In general, the cohort indexed by (t_{ent}, t_{ex}) denotes the number of animals that were born during the time period t_{ent} and died during the time period t_{ex}. For example, the cohort indexed by $(t_{ent} = 1, t_{ex} = 3)$ denotes the number of animals that were born during the first time period (i.e. were juveniles during the capture occasion in time $t = 1$) but died during

Capture History	Probability
010	$(1-p_{i,1})p_{i,2}(1-p_{i,3})$
100	$p_{i,1}(1-p_{i,2})(1-p_{i,3})$
110	$p_{i,1}p_{i,2}(1-p_{i,3})$
000	$(1-p_{i,1})(1-p_{i,2})(1-p_{i,3})$

Table 6.5: Capture histories and associated probabilities for the animal A_i given that it belongs to the $(t_{ent}=1, t_{ex}=3)$ cohort.

the third time period (i.e. were not available to be caught during the during the capture occasion in time $t=3$).

After assigning each animal in the population to one of these cohorts the next stage is to consider which possible capture history patterns these individual animals could have displayed. For example, those animals in the $(t_{ent}=1, t_{ex}=3)$ cohort had died before the third capture occasion and therefore could not exhibit any capture history pattern that indicated a capture in the third element. So, in the case of a three sample study, the only viable capture history patterns for the animals in the $(t_{ent}=1, t_{ex}=3)$ cohort are $010, 100, 110$ and 000. None of the patterns $001, 011, 101$ or 111 could be exhibited by this cohort. Having determined the viable capture history patterns, the next step is to consider each individual animal in this cohort separately and then evaluate the probability of that animal exhibiting each of the viable capture history patterns. Let $p_{i,t}$ denote the probability of animal A_i $(i=1,2,\ldots,N)$ being captured during the capture occasion at time t. Then, if animal A_i belongs to the $(t_{ent}=1, t_{ex}=3)$ cohort the probabilities of this animal exhibiting each of the four viable histories are presented in Table 6.5.

This process is then repeated for the animals in each of the possible $\frac{(T+1)(T+2)}{2}$ cohorts

until the probability of exhibiting each possible capture history pattern has been calculated for all N individuals. Let $Pr\left(\omega_j|A_i\right)$ denote the probability of animal A_i exhibiting the capture history pattern ω_j. Then, the rate r_{ω_j} for the capture history pattern ω_j is obtained by taking a weighted sum of the probabilities $Pr\left(\omega_j|A_i\right)$ over all N animals. The weight attributed to animal A_i is denoted as $Pr(A_i)$. For this population model approach it is assumed *a priori* that each animal has the same weight in this sum; that is $Pr(A_i) = \frac{1}{N}$. Under this weighting approach the rate r_{ω_j} can be thought of as the "average" probability of any animal in the population exhibiting the capture history pattern ω_j. The rates for the multinomial likelihood are now defined formally in Equation (6.6.3) with ψ_i denoting the set of parameters for the observation process associated with animal A_i.

$$r_{\omega_j} = Pr\left(\omega_j\right) = \sum_{i=1}^{N} Pr\left(\omega_j|A_i, \psi_i\right) Pr\left(A_i\right) \tag{6.6.3}$$

By allowing the capture rate to vary both by individual and by time period the resulting model is saturated with $N \times T$ capture parameters. By making more restrictive assumptions about the parameterisation of the capture process this cohort-based approach can be used to simplify the process of calculating the probabilities for each capture history pattern. An example is presented in the next section.

6.6.2 The General Approach

As discussed in the previous section, the general approach to fitting models under the unconditional approach requires each animal to be considered in turn. This process is simplified by grouping animals that share common capture probabilities for each time period. All the animals belonging to an individual group will then have the same probability of exhibiting a specific capture history pattern. If the population can be partitioned into K such groups the rate associated with capture history pattern ω_j can now be calculated

using the expression in Equation (6.6.4). N_k is defined as the number of distinct animals that are assigned to group K such that $N = \sum_{k=1}^{K} N_k$. In this case the weights are determined by group size and are simply the relative frequencies of the K groups.

$$r_{\omega_j} = Pr\left(\omega_j\right) = \frac{1}{N}\sum_{k=1}^{K} N_k Pr\left(\omega_j | N_k, \psi_k\right) \tag{6.6.4}$$

The rates calculated using Equations (6.6.3) and (6.6.4) will be equivalent but the approach indicated in Equation (6.6.4) will typically be more efficient.

This approach to calculating the rates can be applied to the usual example population. The parameterisation and structure of the state and observation models are the same as for the model outlined in section 4.4. The population is subject to evolution according to binomial survival and birth processes as specified in Equations (6.4.1a):(6.4.1d). Measures taken on this system are modelled using a binomial capture process and consist of the observed abundances for the possible capture history patterns exhibited by the animals. It is assumed that the population consists of two distinct age classes: juveniles and adults and that the model parameters relating to the survival and capture processes differ by age class.

For this population the groups that share common capture probabilities for each time period will be the cohorts defined by the entry and exit times. For a study consisting of three sampling occasions, $T = 3$, there are $\frac{(T+1)(T+2)}{2} = 10$ distinct cohorts. Using the (t_{ent}, t_{ex}) cohort indexing from section 6.6.1 the 10 cohorts are

$$\{(0,1),(0,2),(0,3),(0,\bullet),(1,2),(1,3),(1,\bullet),(2,3),(2,\bullet),(3,\bullet)\}$$

where $t_{ex} = \bullet$ corresponds to animals that did not exit the population during the duration of the study. Then, each capture history pattern ω is consider in turn and the probabilities,

as in Table 6.5, are evaluated for each (t_{ent}, t_{ex}) cohort. These probabilities are presented in Table 6.6. It should be noted that the 0 entries in Table 6.6 relate to capture history patterns that are not attainable by the animals in the corresponding cohort.

Then, to calculate the rates using Equation (6.6.4) it is necessary to establish the abundances, N_k, that correspond to each of the 10 cohorts. Table 6.4 effectively represents the total number of animals that could exhibit a 110 capture history pattern partitioned by t_{ent}, t_{ex} cohort. The same process used to generate that break-down is then applied to the entire population of N animals and the N_k's can then be expressed as functions of the elements of the state vectors $\mathbf{n}_{0:T}$. For this three occasion example, the contributions to the abundances for each of the 10 t_{ent}, t_{ex} cohorts are given in Table 6.7. Summing the components then returns the abundance for that cohort. For example, the animals that entered the population during the first time period and left during the third time period would belong to the $1, 3$ cohort and the abundance of this cohort is then

$$N_{(1,3)} = n_{jja,2} - n_{jjaa,3} + n_{aja,2} - n_{ajaa,3}.$$

Repeating this process for each cohort then yields a vector of abundances.

The probabilities from Table 6.6 and the abundances obtained from Table 6.7 are then combined to calculate the probability associated with each capture history pattern. The entries in Table 6.6 form a $2^T \times \frac{(T+1)(T+2)}{2}$ matrix. Switching the rows and columns and multiplying this resulting matrix by the vector obtained from summing the rows in Table 6.7 nearly completes the calculation of the multinomial rates $\left\{r_{\omega_1}, \ldots, r_{\omega_{2^T}}\right\}$. For example, using Equation (6.6.4) and Tables 6.7 and 6.6 the probability associated with

Cohort	**Pattern**			
(t_{ent}, t_{ex})	000	001	010	100
$(0,1)$	1	0	0	0
$(0,2)$	$1-p_a$	0	0	p_a
$(0,3)$	$(1-p_a)^2$	0	$(1-p_a)p_a$	$p_a(1-p_a)$
$(0,\bullet)$	$(1-p_a)^3$	$(1-p_a)^2p_a$	$(1-p_a)p_a(1-p_a)$	$p_a(1-p_a)^2$
$(1,2)$	$1-p_j$	0	0	p_j
$(1,3)$	$(1-p_j)(1-p_a)$	0	$(1-p_j)p_a$	$p_j(1-p_a)$
$(1,\bullet)$	$(1-p_j)(1-p_a)^2$	$(1-p_j)(1-p_a)p_a$	$(1-p_j)p_a(1-p_a)$	$p_j(1-p_a)^2$
$(2,3)$	$(1-p_j)$	0	p_j	0
$(2,\bullet)$	$(1-p_j)(1-p_a)$	$(1-p_j)p_a$	$p_j(1-p_a)$	0
$(3,\bullet)$	$1-p_j$	p_j	0	0

(a) First four histories.

Cohort	**Pattern**			
(t_{ent}, t_{ex})	011	101	110	111
$(0,1)$	0	0	0	0
$(0,2)$	0	0	0	0
$(0,3)$	0	0	p_a^2	0
$(0,\bullet)$	$(1-p_a)p_a^2$	$p_a(1-p_a)p_a$	$p_a^2(1-p_a)$	p_a^3
$(1,2)$	0	0	0	0
$(1,3)$	0	0	p_jp_a	0
$(1,\bullet)$	$(1-p_j)p_a^2$	$p_j(1-p_a)p_a$	$p_jp_a(1-p_a)$	$p_jp_a^2$
$(2,3)$	0	0	0	0
$(2,\bullet)$	p_jp_a	0	0	0
$(3,\bullet)$	0	0	0	0

(b) Remaining histories.

Table 6.6: Probabilities for capture history patterns by entry and exit times

Cohort (t_{ent}, t_{ex})	**Components of Abundance**			
$(0,1)$	$n_{j,0} - n_{ja,1}$	$n_{a_0} - n_{aa,1}$		
$(0,2)$	$n_{ja,1} - n_{jaa,2}$	$n_{aa_1} - n_{aaa,2}$		
$(0,3)$	$n_{jaa,2} - n_{jaaa,3}$	$n_{aaa_2} - n_{aaaa,3}$		
$(0,\bullet)$	$n_{jaaa,3}$	n_{aaaa_3}		
$(1,2)$	$n_{jj,1} - n_{jja,2}$	$n_{aj_1} - n_{aja,2}$		
$(1,3)$	$n_{jja,2} - n_{jjaa,3}$	$n_{aja_2} - n_{ajaa,3}$		
$(1,\bullet)$	$n_{jjaa,3}$	n_{ajaa_3}		
$(2,3)$	$n_{jjj,2} - n_{jjja,3}$	$n_{jaj_2} - n_{jaja,3}$	$n_{ajj_2} - n_{ajja,3}$	$n_{aaj_2} - n_{aaja,3}$
$(2,\bullet)$	$n_{jjja,3}$	$n_{jaja,3}$	$n_{ajja,3}$	$n_{aaja,3}$
$(3,\bullet)$	$n_{jjjj,3}$	$n_{jjaj,3}$	$n_{jajj,3}$	$n_{jaaj,3}$
	$n_{ajjj,3}$	$n_{ajaj,3}$	$n_{aajj,3}$	$n_{aaaj,3}$

Table 6.7: Abundances for each (t_{ent}, t_{ex}) cohort.

capture history pattern 110 is:

$$
\begin{aligned}
r_{110} = Pr\,(110) &= \frac{1}{N}\sum_{k=1}^{10} N_k Pr\,(110|N_k,(p_j,p_a)) \\
&= \frac{1}{N}\Big\{\; p_a^2\left[(n_{jaa,2}-n_{jaaa,3})+(n_{aaa,2}-n_{aaaa,3})\right] \\
&\quad + p_a^2(1-p_a)\left[n_{jaaa,3}+n_{aaaa,3}\right] \\
&\quad + p_j p_a\left[(n_{jja,2}-n_{jjaa,3})+(n_{aja,2}-n_{ajaa,3})\right] \\
&\quad + p_j p_a(1-p_a)\left[n_{jjaa,3}+n_{ajaa,3}\right] \;\Big\}.
\end{aligned}
\tag{6.6.5}
$$

This procedure is then repeated for each of the capture history patterns and, using the definitions from Table 6.3, these rates can then be used in the multinomial likelihood given in Equation (6.6.6).

$$
L(\Theta|\omega_1,\ldots,\omega_{2^T}) = \frac{N!}{\prod_{j=1}^{2^T} x_{\omega_j}!}\prod_{j=1}^{2^T} r_{\omega_j}^{x_{\omega_j}} \tag{6.6.6}
$$

The set of capture history patterns used in the evaluation of the multinomial likelihood consists of 2^T individual patterns, one of which is the null pattern corresponding to those animals that were present in the population but never captured during the study. These animals are represented by the $N - M$ individuals from Table 6.2. The abundance of animals with the null capture history pattern is then obtained by taking the difference between the total number of animals that were present during the study, as determined by Equation (6.6.2), and the number of animals that were captured at least once during the study. Thus, for the three sampling occasion example,

$$
x_{000} = n_{j,0} + n_{a,0} + \sum_{t=1}^{3}\sum_{i=1}^{8} n_{2i-1,3} - \sum_{j=2}^{8} x_{\omega_j} \tag{6.6.7}
$$

and the likelihood in Equation (6.6.6) can be evaluated.

6.6.3 Extending The General Approach

The general approach in section 6.6.2 set out the required approach to evaluating the model likelihood for the observed abundances of the capture history patterns. This approach was valid when the observations made no distinction between the age of the animals exhibiting those capture history patterns. However for the state-space model defined by the observation and state processes given in sections 6.4 and 6.5 it can be seen from Figure 6.5 that the vector of observations $\mathbf{y}_t$ maintains a distinction between juveniles, yearlings and adults. Under this structure there are $4 + 2^t$ possible capture history patterns that can be exhibited by animals that have existed in the population between times 0 and t.

This then requires the multinomial likelihood to be expressed as in Equation (6.6.8) where the sums and products range over $4 + 2^t$ separate cohorts.

$$L(\Theta|\omega_1, \ldots, \omega_{4+2^T}) = \frac{N!}{\prod_{j=1}^{4+2^T} x_{\omega_j}!} \prod_{j=1}^{4+2^T} r_{\omega_j}^{x_{\omega_j}} \tag{6.6.8}$$

Using the approach of section 6.6.2 and partitioning the population into cohorts defined by their entry and exit times allows this extended likelihood to be calculated relatively easily. The four extra capture history patterns that now appear in Equation (6.6.8) correspond to capture history patterns denoting unobserved juveniles $j0$, observed juveniles $j1$, unobserved yearlings y_{10} and observed yearlings y_{11} at time T. This reduced notation is used only in reference to the capture history patterns that include all T capture occasions. In other words, for a three sample study the $j0$ notation could be equivalently expressed as $j000$ but the first two zeros denoting non-capture due to absence from the

population are omitted for simplicity.

Three of these four extra patterns are observed and thus their totals are fixed whereas the abundance of unobserved juveniles needs to be simulated. It is determined in a manner similar to that discussed in section 6.6.2 for the abundance of adults exhibiting the null capture history pattern. The total number of juveniles produced in the final time period is

$$n_{j0,T} + n_{j1,T} = \sum_{i=1}^{2^T} n_{2i-1,T}$$

and the number of juveniles captured during time T is simply the first elements of the observation vector $\mathbf{y}_T$. Hence,

$$n_{j0,T} = \sum_{i=1}^{2^T} n_{2i-1,T} - n_{j1,T}.$$

The rates r_ω correspond to the proportions of animals in the population exhibiting the capture history patterns ω. Hence, the rates for each of these four extra patterns need to be determined. Consider again the matrix of probabilities in Table 6.6. Multiplying each column of this matrix by the vector of abundances obtained from summing the rows in Table 6.7 and dividing the resulting vector by the population size N gives the relative probabilities by (t_{ent}, t_{ex}) cohort for each of the 2^T capture history patterns. Then, summing together the rows according to time of entry results in a $(T+1) \times 2^T$ matrix, D, that gives the relative probabilities by time of entry into the population for each of the 2^T capture history patterns. These probabilities can then be used to obtain the rates r_ω for these four extra capture history patterns. For example; for a three sample study the rate r_{y10} corresponds to yearlings that were not observed in the third sample and, by definition, a yearling was captured as a juvenile in the previous time period. Hence, the subscript notation for this cohort, $y10$, assumes implicitly that any yearling at time

t would not have been present in the population at $t-2$ and therefore an equivalent alternative subscript notation for this example is $y010$. The x_{y10} animals exhibiting this capture history pattern will then be those animals that were born during the second period and captured as juveniles during the second capture occasion. Hence, the relevant probability for r_{y10} is the element of D that corresponds to those animals that exhibit a 010 capture history pattern and entered the population during the second period. The value of r_{y10} is given in Equation (6.6.9).

$$r_{y_{10}} = \frac{1}{N}\Big\{ p_j\left[(n_{jjj,2} - n_{jjja,3}) + (n_{jaj_2} - n_{jaja,3}) + (n_{ajj_2} - n_{ajja,3}) + (n_{aaj_2} - n_{aaja,3})\right] + p_j(1-p_a)\left[n_{jjja,3} + n_{jaja,3} + n_{ajja,3} + n_{aaja,3}\right] \Big\} \tag{6.6.9}$$

A similar procedure is then used to obtain the other three multinomial rates. Thus, the relevant probability for r_{j0} is the element of D that corresponds to those animals that exhibit a 000 capture history pattern and entered the population during the third period. For r_{j1} the relevant element corresponds to the 001 capture history pattern and entry during $t=3$. Finally, for r_{y11} the relevant element corresponds to the 011 capture history pattern and entry during $t=2$. It should be noted that for this example the observed abundances in $\mathbf{y}_T$ for the patterns $a000$, $a001$, $a010$ and $a011$ do not include the juvenile or yearling animals that exhibit those patterns. Therefore the rates for the eight capture history patterns $r_{a000}, r_{a001}, \ldots, r_{a111}$ will not all be equivalent to those used in the likelihood that was given in Equation (6.6.6). Due to the required distinction between juveniles, yearlings and adults in this extended approach the relationship between the

rates in Equations (6.6.6) and (6.6.8) is

$$r_{000} = r_{j0} + r_{a000}$$
$$r_{001} = r_{j1} + r_{a001}$$
$$r_{010} = r_{y10} + r_{a010}$$
$$r_{011} = r_{y11} + r_{a011}.$$

Alternatively, and as for any model, the rates for the multinomial likelihood can be obtained using the fully general approach in Equation (6.6.3). Ultimately, both approaches will allow the likelihood in Equation (6.6.8) to be evaluated for a realisation of the population.

The construction of the likelihood in Equation (6.6.8) means that sequential importance sampling is not an appropriate inferential procedure for fitting the models specified under the unconditional approach. In Equation (6.5.2) it was shown how the likelihood determined by that model structure can be expressed as the product of year-specific components. This is not the case for the likelihood determined by Equation (6.6.8). To evaluate the set of multinomial rates $\mathbf{r}_\omega$ for the likelihood based on the observations at time t it is necessary to include the parameters that are used to construct the likelihood at time $t-1$. This is because the observations at time t will include the data that was used to determine the observed capture histories, and consequently the likelihood, at time $t-1$. The likelihood at time t is based on observed data from the first t capture occasions rather than just data from the capture occasion at t alone. Therefore, the likelihoods at each time t are not independent and the likelihood based on the abundances of observed capture history patterns for the entire study cannot be expressed as their product.

As a result of this, model fitting for the unconditional approach then uses the importance sampling techniques discussed in section 3.3.6. For the unconditional approach a particle refers to the set of state and parameter vectors $\{\mathbf{n}_{0,T}, \Theta\}$. The likelihood for the observations $\mathbf{y}_T$ in Equation (6.6.8) is evaluated for each particle and the weights required for the sampling procedure are then proportional to the likelihood. That is:

$$w \propto \frac{N!}{\prod_{j=1}^{4+2^T} x_{\omega_j}!} \prod_{j=1}^{4+2^T} r_{\omega_j}^{x_{\omega_j}}$$

and these weights are then normalised as described in section 5.6.

Using an importance sampling approach to the fitting procedure will, as discussed in section 5.6, produce an approximate sample from the posterior distribution of $g(\mathbf{n}_{1:T}|\mathbf{y}_{1:T}, \Theta)$. It should be noted that the unconditional approach can, as with the conditional approach, suffer from the simulation of implausible particles. From Equation (6.6.7) it can be seen that the simulated number of animals that were present in the population during the study without ever being observed is obtained by taking the difference between the model-based number of animals that have entered the population and the total number of animals that have been observed. It is possible that, for an individual particle, the application of the state process may simulate a total population size N that is smaller than the actual number of animals observed throughout the study. This scenario would result in a negative number of unseen animals and thus produce an implausible particle which would then be assigned zero weight and excluded from the posterior sample. A similar problem can occur when calculating the value of x_{j0} in the extended likelihood of Equation (6.6.8).

One of the defining features of the unconditional approach is that, in contrast to the conditional approach, the simulation of state elements is not constrained by the observed

abundances of capture history patterns. This may result in the population at a particular time period being implausible when compared with the observed capture history patterns. For example, the number of animals that enter the population at time $T-1$ and survive to the end of the study may not be large enough to have generated the number of yearlings that exhibited a y_{11} capture history pattern in the observed data. From the simulation results presented in Chapter 7, this problem is not that pronounced for the relatively simple models considered so far. However this potential limitation should not be ignored and incorporating some form of conditioning on the state process, but not to the extent and complexity of that required in the conditional approach, is a potential area for further research.

This chapter has focussed on developing the methodology for fitting model developed under the unconditional population approach. This allows a comparison to be made between the the fitting algorithms developed under the population models assumed under both the conditional and unconditional approaches. Developing the methodology for fitting models under the unconditional superpopulation approach is an area of future research. Applications of the unconditional approach to fitting models based on both real and simulated data are presented in Chapter 7.

Chapter 7

Simulation Studies and Analysis of Real Data

7.1 Outline

The previous chapters have presented the conditional and unconditional approaches to fitting state-space models to mark-recapture data by way of embedding a population dynamics model into the inference. The motivation for each approach has been presented and the structure of each fitting algorithm has been specified with illustrative examples based on a simple model structure. These examples can be developed still further by way of simulation studies.

Simulation studies allow the performance of the model and the efficacy of the philosophy of the approach to be investigated. Simulating data allows the investigator to analyse model output by comparing it to a known truth. Knowing truth, in the context of model parameters and state elements, provides a simple metric by which to determine how well the model is able to distinguish the signal from the noise. The aim of developing these approaches is to devise flexible integrated modelling techniques that draw from both existing mark-recapture and state-space modelling theory. By embedding a

population dynamics model into the mark-recapture analysis the resulting parametric inference should be consistent with what is assumed to be known about the population being studied. Simulation studies allow this issue to be investigated and can serve to highlight practical issues that arise when applying the approaches. Simulation studies are performed for both the conditional and unconditional approaches in sections 7.2.3 and 7.2.4.

Whilst simulation studies are an excellent tool for investigating the performance of a model, the artificially constructed data used in the simulation may not accurately represent real data. Therefore, attempting to fit models to real data under both approaches is constructive and can reveal potential limitations in the modelling framework that were not apparent from the simulation studies. Section 7.3 describes an application to Soay Sheep data using the same model form as was investigated in the simulation studies.

In section 7.4 the performance of the model fitting algorithms over multiple simulations is investigated. The degree of uncertainty relating to posterior summaries is often of interest and performing multiple simulations allows the investigator to distinguish between the component of this uncertainty that can be attributed to the state process and that which is the result of Monte Carlo variation. It should also be noted that, for this chapter, all analyses based on the unconditional approach assume a population model.

7.2 Analysis of Simulated Data

The simulation study begins with a realisation of a single population. The structure of this population will be the same as that defined in section 4.4. The evolution of the population over a series of T time periods can then be simulated using the sub-processes of survival and birth. The simulation begins by specifying an initial cohort of N_0 animals

that is then split into two age cohorts: juveniles (with probability τ) and adults. These two cohorts are then projected forward stochastically through the survival process. Juveniles survive with probability ϕ_j, adults with probability ϕ_a. It is assumed that these rates are constant over time. The surviving animals then produce new juveniles via the stochastic birth process with π denoting the probability that an animal produces a single young for that time period. The surviving animals are subject to a deterministic ageing process; surviving juveniles from time $t-1$ are classified as adults at time t, surviving adults from time $t-1$ are, naturally, still classified as adults at time t. Each surviving animal is then subjected to a stochastic capture process and is observed with probability p_j if a juvenile and p_a if an adult. Following the capture process the cycle of sub-processes is repeated beginning again with survival. This process continues until the final capture occasion at time T at which point all elements of the state and intermediate vectors will have been simulated. The set of Equations (4.4.1a):(4.4.1f) define the functional forms of the stochastic sub-processes used to simulate the population, all of which are binomial for this model.

From Figures 6.4 and 6.5 it can be seen that the structure of the observations made on this simulated population differ under the two approaches. For the conditional approach the observed data at time t consists of the counts of capture history patterns including capture in the most recent time period. Under the unconditional approach the observed data at time t consists of the counts of capture history patterns that include capture at least once during the study. So, although the structure of the observation vectors $\mathbf{y}_{0:T}$ may differ between approaches, the same data is used for both the conditional and unconditional approaches. The parameter values used to simulate the observed data will be fixed *a priori* and can be chosen to create populations displaying certain characteristics. For example, setting the birth rate π and the splitting rates τ both to 0 will result in the

Parameter	Symbol	Value
Juvenile Survival	ϕ_j	0.7
Adult Survival	ϕ_a	0.8
Juvenile Capture	p_j	0.5
Adult Capture	p_a	0.6
Birth	π	0.4
Population Split	τ	0.3

Table 7.1: Fixed parameters used to simulate the population.

simulation of a population consisting solely of adults that declines in abundance over time.

For the simulation studies covered in this section a population, and the resulting observed data, will be simulated using the parameters specified in Table 7.1. An initial total N_0 will also need to be specified and will be taken to be 100 for the simulations in this section.

Model fitting under each approach is initiated by simulating a large number of particles. This requires drawing values for the parameters, $\Theta = (\phi_j, \phi_a, \pi, p_j, p_a, \tau)$, from some prior distribution. One possible approach to this is discussed in the next section.

7.2.1 Simulation of priors

The model-fitting process is embedded within a Bayesian framework and this approach necessitates the specification of a prior distribution for the parameters and states of interest (see (3.3.1a)). This prior is specified as a joint prior distribution on the priors and states and is typically comprised as the product of prior distributions for each individual parameter if the parameters are assumed to be independent *a priori*.

For these simulation studies the derivation of these component prior distributions is kept relatively simple. All of the parameters $\Theta = (\theta, \psi)$ that were specified in the state and observation processes in the previous chapters correspond to rates of binomial distributions with the exception of the multinomial parameters that are required for a birth process that allows a single adult to produce more than one progeny during each breeding occasion. Consequently, a necessary restriction is that all of these parameters must be bounded between 0 and 1 and the mechanism used to simulate from the prior distribution must respect this constraint.

One suitable approach is to specify the mean value of the prior for each parameter of interest, along with the associated variance. The specification of the variance then allows the prior distribution for the parameter to be "spiked", where the majority of the probability mass is spread over a small range of values, or "flat", where the probability mass is distributed evenly between the boundary values 0 and 1.

To simulate a general parameter of interest, α, the mean μ_α and variance σ^2_α are both specified. The prior on α is then obtained using the following process:

a Define μ_l as the inverse logistic transform of μ_α such that $\mu_l = \log(\mu_\alpha) - \log(1 - \mu_\alpha)$.

b Using this transformed variable μ_l define a random variable X such that: $X \sim \text{Normal}\,(\mu_l, \sigma^2_\alpha)$.

c Draw a sample of size N_α from the normal pdf X.

d Transform the resulting sample $x_1, x_2, \ldots, x_{N_\alpha}$ using the logistic transformation to obtain a sample for α. The i^{th} element of the sample is thus defined as

$$\tilde{\alpha}_i = \frac{\exp(x_i)}{1 + \exp(x_i)}$$

This sample of N_α draws $\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_n$ is then centered about μ_α and bounded between 0 and 1 as required.

This is just one possible approach to simulating the values of the priors. Examples of more detailed prior elicitation for state-space or hidden process models for biological populations include studies on red deer (*Cervus elaphus*) (Trenkel *et al.*, 2000), winter-run chinook salmon (*Oncorhunchus tshawytsha*) (Newman *et al.*, 2006) and grey seals (*Halichoerus grypus*) (Thomas *et al.*, 2005).

7.2.2 Fitting the models

Having drawn a set of N_{part} values for each parameter in $\Theta = (\phi_j, \phi_a, \pi, p_j, p_a, \tau)$ the model fitting algorithms are then applied to simulate the state elements and produce posterior distributions on the parameters of interest.

A simplified pseudo-algorithm then outlines the importance-sampling based model fitting approach with the specific details varying for the conditional and unconditional approaches:

a Specify a set of values for the parameters $(\phi_j, \phi_a, \pi, p_j, p_a, \tau)$ that will be regarded as the "true" values.

b Choose an initial population size N_0 and simulate splitting this total into juveniles and adults.

c Simulate a single realisation of a population from the state process pdf $g_t(\mathbf{n}_t|\mathbf{n}_{t-1}, \Theta)$ which is given by the product of the survival and birth processes.

d Simulate the observed animals in this population using the stochastic capture process. For the conditional and unconditional approaches store those abundances

corresponding to the capture history patterns that comprise the observed vector at each time period.

e Generate N_{part} sets of $\mathbf{n}_0$ which is defined as

$$\mathbf{n}_0 = \begin{bmatrix} n_{j,0} \\ n_{a,0} \end{bmatrix}$$

and N_{part} sets of Θ using the method in section 7.2.1. Do this for both approaches.

f For the conditional approach generate a sample of size N_{part} of the unobserved elements $\check{\mathbf{s}}_T$ of $\mathbf{n}_T$ from the trial pdf $h_T()$. These simulated observed elements will be constrained to be consistent with the observed data obtained in step d. Increment the index in decreasing time to generate N_{part} samples of $\check{\mathbf{s}}_{T-1}$ from the trial pdf $h_{T-1}()$ conditional on $\check{\mathbf{n}}_T$ and the observed data. Repeat until $\check{\mathbf{s}}_1$ has been simulated and thus all elements of $\mathbf{n}_{0:T}$ have been specified. Call this sample $\mathbf{n}^{*,i}_{0:T,cga}, i = 1, \ldots, N_{part}$.

g For the unconditional approach simulate a sample of size N_{part} of $\mathbf{n}_{1:T}$ from the trial pdf which is defined to be the state process pdf and does not include a capture process. Call this sample $\mathbf{n}^{*,i}_{0:T,uga}, i = 1, \ldots, N_{part}$.

h Calculate N_{part} weights for the conditional approach that are proportional to the ratio of the product of the trial densities $h_t()$ at time t to the product of the state process pdf at time t. Normalise the weights and assign the weight $w^{*,i}_{cga}$ to the sample $\mathbf{n}^{*,i}_{0:T,cga}, \theta^{*,i}$.

i Calculate N_{part} weights for the unconditional approach that are the proportional to the multinomial likelihood for the observed capture history patterns as expressed by Equation (6.6.8). Normalise the weights and assign the weight $w^{*,i}_{uga}$ to the sample $\mathbf{n}^{*,i}_{0:T,uga}, \theta^{*,i}$.

j Use the particles $\mathbf{n}^{*,i}_{0:T,cga}, \theta^{*,i}$ and associated weights $w^{*,i}_{cga}$ to obtain weighted summaries on the posterior distributions on $\mathbf{n}_{0:T,cga}, \theta$ from the conditional approach.

k Use the particles $\mathbf{n}^{*,i}_{0:T,uga}, \theta^{*,i}$ and associated weights $w^{*,i}_{uga}$ to obtain weighted summaries on the posterior distributions on $\mathbf{n}_{0:T,uga}, \theta$ from the unconditional approach.

Implementing fitting routines using the steps of this pseudo-algorithm will yield approximations to the the smoothed posterior distributions for the parameters and states under both the conditional and unconditional approaches. As explained in section 5.2.7, the bottom-up conditional generation approach does not allow a sequential importance sampling scheme to be applied. The unconditional approach is also unable to accommodate a sequential resampling scheme. A single importance sampling step, via a weighted bootstrap resample, could be utilised for both approaches to obtain posterior summaries. However, this resampling step increases Monte Carlo error by removing particles that provide information about the posterior distribution. All required inferences can be made using weighted averages that retain all the information in the data and this is the approach used for the model fitting processes in this chapter. It should be noted that, even with the restrictions imposed on the simulation of states for the conditional approach, this fitting algorithm can be quite inefficient as many of the particles may be assigned low weights. This means computer resources have been effectively wasted on projecting these "bad" particles through time that will have little contribution to the posterior. Various approaches to boosting the number of "good" particles in a posterior sample have been developed, details of which can be found in Liu and Chen (1998); Carpenter *et al.* (1999); Pitt and Shephard (1999); Liu and West (2001); Liu (2001).

One tool that can be used to assess the extent of this problem is the effective sample size. This provides a metric by which to judge the performance of the model fitting

algorithm in terms of particle depletion and the distribution of the weights associated with each particle. Assume that N_{part} particles are simulated and, after an application of the model fitting algorithm, they are assigned the set of weights $\{w_1, w_2, \ldots, w_{N_{part}}\}$. These weights will be assigned the value zero if they correspond to invalid particles. The set of weights is then used to obtain the effective sample size (ESS) using the expression in Equation (7.2.1).

$$\begin{aligned} cv^2 &= \frac{\sum_{i=1}^{N_{part}} (w_i - \bar{w})^2}{(N_{part} - 1) \times \bar{w}^2} \\ ESS &= \frac{N_{part}}{1 + cv^2} \end{aligned} \tag{7.2.1}$$

Therefore, the less variation present in the weights, the smaller the value of cv^2 and, consequently, the larger the value of the effective sample size ESS. This determines the effective number, out of the initial N_{part} particles, of independent particles that constitute the sample from the posterior. The smaller the ESS is, the more variable the weights are and the larger the Monte Carlo variation which simply means different simulation runs give appreciably different results.

Each model fitting algorithm is now applied to the same simulated population with the same set of priors used for each approach. The priors are generated using the technique in section 7.2.1 with the means (μ_α) and variances (σ^2_α) as specified in Table 7.2. The values for each σ^2_α are on the untransformed scale rather than the probability scale. The value σ^2_α is simply a parameter that is used to control the precision of the priors after transformation rather than give the variance of the transformed sample. Plots of the distributions of the six independent priors are presented in Figure 7.1

Both fitting approaches are initialised with 50000 particles consisting of simulated

	ϕ_j	ϕ_a	p_j	p_a	π_0, π_1	τ
Prior Means: μ_α	0.5	0.85	0.65	0.5	0.7,0.3	0.4
Prior Variances: σ^2_α	1	1	1	1	1,1	1

Table 7.2: Specification of parameters μ_α and σ^2_α used to simulate priors for the unconditional approach. Values for σ^2_α are on the logit scale.

states $\mathbf{n}_0$ and draws from the priors for the parameters Θ. The results are presented in sections 7.2.3 and 7.2.4.

7.2.3 Conditional Approach

This conditional fitting approach had 21973 plausible particles remaining after the model had been fitted to all the observed data. Fitting the model using 50000 particles took 16 hours and 7 minutes. The effective sample size based on these surviving particles is 6.153. This is a rather small effective sample size and inference should be undertaken carefully in this situation. The small effective sample size obtained for this analysis is a feature that was observed under many alternative analyses based on the fitting algorithm for the conditional approach. This may imply that a small effective sample size is an artefact of the proposal densities constructed in the fitting algorithm rather than the model not fitting the data well. For the analysis of this simulated data, the proposed model being fit to data is of the same form as the model used to generate the data; hence, a low effective sample size is unlikely to be a result of poor model fit.

Sample output from the model fitting process under the conditional approach is now presented to illustrate the performance of the process and to allow direct comparison with the results from the unconditional approach as presented in Section 7.2.4. Table 7.3 summarises the posterior distributions of the state elements that denote the surviving

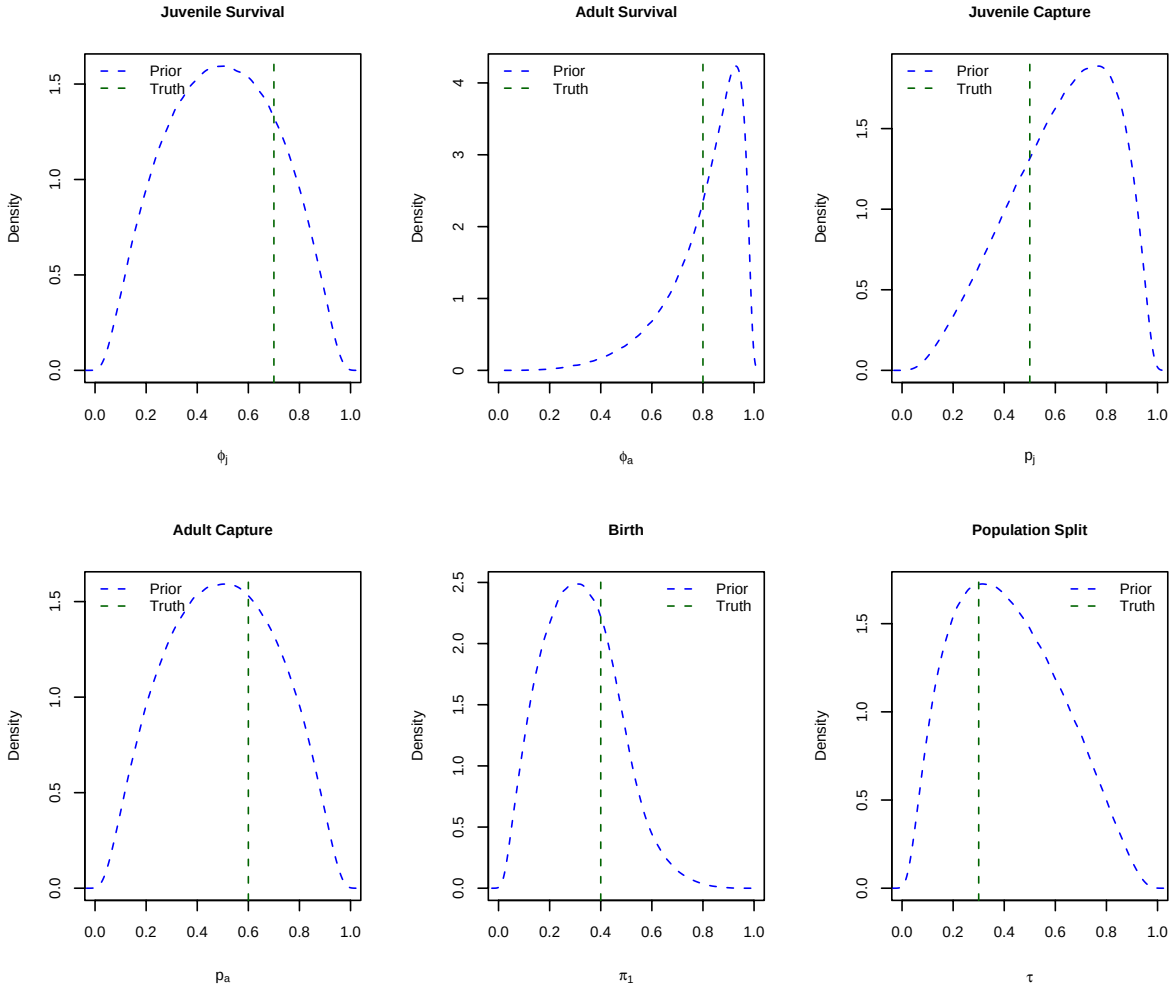


Figure 7.1: Density plots of priors for survival, capture, birth and population split rates simulated using the parameters specified in Table 7.2. Truth denotes the value of the parameter used to simulate the population.

population immediately following the second capture occasion. The column headed "Expected" contains the expected value of the state elements. This is obtained by taking the initial 100 animals and multiplying this total by the true population parameters. This process is iterated over each time period and can be thought of as providing the abundance for each state and intermediate vector that would be obtained if the population was projected forward deterministically with the the parameters fixed at their true values.

State	Median	Mean	Truth	LCL	UCL	Expected	CV
n_{j0}	15	14.60	17	6	25	16.60	0.392
n_{j1}	18	18.00	18	18	18	16.60	0
n_{y10}	7	6.38	10	1	12	4.31	0.446
n_{y11}	3	3.00	3	3	3	6.47	0
n_{a00}	21	20.60	12	10	26	14.20	0.279
n_{a01}	22	22.00	22	22	22	21.30	0
n_{a10}	23	23.50	16	17	28	14.80	0.121
n_{a11}	20	20.00	20	20	20	22.20	0

Table 7.3: Summary for state vector at $t = 2$ for the conditional approach. Expected values denote projected state totals obtained by multiplying the initial 100 animals by the true population parameters and iterating this process over each time period.

For example, $E_\Theta[n_{a1,1}|N_0, \Theta] = N_0(\tau(\phi_j p_a) + (1-\tau)(\phi_a p_a))$. For a single realisation of a population the column headed "Truth" is the more important to compare the posterior means against. The expected value of the states can be used to assess the performance of the model fitting algorithm by averaging the posterior means over multiple simulations. This is investigated in Section 7.4.

It is immediately clear that all state elements corresponding to animals that were captured at time $t = 2$ are fixed at the true, observed, values and have a variance of zero. The posterior summaries for the states in $\mathbf{n}_2$ corresponding to animals that were alive but remained uncaught at $t = 2$ seem to match well with the true population values. The only exception to this is the posterior for n_{a10} which denotes adult animals that were captured at time $t = 1$ and were alive, but uncaught, during the capture process at time $t = 2$. The posterior mean of 23 is significantly larger than the true value of 16 and the interval for n_{a10} excludes this true value.

Having examined some of the states, it is worth investigating the model's performance with regard to the posterior distributions of the model parameters Θ. These distributions are summarised in Table 7.4 and can be compared with the priors as well as the true values given in Table 7.1. It can be seen that the posterior means seem to give a reasonable summary of the true underlying parameters that were used to simulate the data. The intervals for all but one parameter include the true value; the only exception being the interval for ϕ_a adult survival. The lower limit of this interval just excludes the true parameter value of 0.8 and it can also be noted that the posterior mean for ϕ_a moves away from the true value relative to the location of its prior. For all other parameters the means of their posteriors move towards the true value relative to the location of their priors. For this particular realisation of a population it seems that the conditional generation approach is working; albeit the small effective sample size should induce a note of caution into this analysis.

It can also be noted that the posterior distributions for the parameters corresponding to processes acting on adult animals have lower variances than the posteriors for rates associated with juvenile animals. Adult survival (ϕ_a) has a lower variance than juvenile survival (ϕ_j) and the same relationship is observed between adult and juvenile capture rates. This reflects the relative amounts of information in the data in the observed capture history patterns for juveniles and adults. The largest CV is that for the posterior of the population splitting parameter τ which should come as no real surprise as there is very little information in the data to allow this parameter to be estimated precisely. It should also be noted that Table 7.4 displays only the six independent parameters. The value of π_1 (the average number of singletons produced by a mature female in one year) will determine the value of π_0 since $\pi_0 + \pi_1 = 1$. Hence the posterior variance will be equal for both parameters.

	Prior			Posterior					
Parameter		**Mean**	**Var**	**Median**	**Mean**	**Var**	**CV**	**LCL**	**UCL**
ϕ_j	(0.7)	0.500	0.04343	0.768	0.708	0.01820	0.191	0.4550	0.854
ϕ_a	(0.8)	0.811	0.02003	0.875	0.889	0.00206	0.051	0.8040	0.966
p_j	(0.5)	0.628	0.03910	0.501	0.538	0.01170	0.201	0.3750	0.736
p_a	(0.6)	0.501	0.04330	0.522	0.541	0.00328	0.106	0.4630	0.667
π_1	(0.4)	0.322	0.02132	0.312	0.340	0.00437	0.195	0.2540	0.480
τ	(0.3)	0.417	0.04153	0.276	0.294	0.01820	0.459	0.0841	0.545

Table 7.4: Summary of model parameters for the conditional approach. True parameter values denoted in ( ).

To enable a direct comparison to be made between the output from the conditional and unconditional model fitting processes it is expedient to summarise the data into two age-based cohorts at each time period: juveniles and adults (including yearlings). No distinction is now made between those animals that were captured at time t and those that weren't. Table 7.5 now summarises the posterior distributions for the total juveniles and total adults at each time step. The large CVs at time $t = 0$ for both juveniles and adults are a result of the split of the initial 100 animals being made irrespective of any conditioning on observed data. The conditional fitting approach again seems to be performing quite well over each time period with all intervals including the true totals for the two cohorts. As noted before, there is more information in the data regarding the adult animals and this is illustrated by the smaller CVs for the adult cohorts when compared with the juvenile cohort at time time step.

The bottom-up conditional model fitting approach seems to have performed reasonably well, albeit having incurred significant particle depletion from the initial 50000 particles. The same true population and set of priors was then used to investigate the efficacy of

Time t	Median	Mean	Truth	LCL	UCL	Expected	CV
$t = 0$	29	29	40	9	56	30	0.475
$t = 1$	28	28	37	22	38	30	0.136
$t = 2$	33	32.6	35	24	43	33	0.176
$t = 3$	34	34.4	34	23	46	35.9	0.175

(a) Simulated juveniles.

Time t	Median	Mean	Truth	LCL	UCL	Expected	CV
$t = 0$	71	70.5	60	44	91	70	0.199
$t = 1$	87	86.4	79	74	92	77	0.0566
$t = 2$	99	95.5	83	78	103	83.2	0.0798
$t = 3$	112	108	92	87	117	89.8	0.0855

(b) Simulated adults.

Table 7.5: Posterior summary of age classes over time obtained under the conditional approach. Expected values denote projected state totals obtained by multiplying the initial 100 animals by the true population parameters and iterating this process over each time period.

the unconditional model fitting approach. By simulating a matrix of capture history patterns and indexing the entry and exit times of each animal that occurred in the simulated population over the duration of the study, it was possible to appropriately format the data to enable analysis under both modelling approaches. Specifically, it was necessary to summarise the true states at each time period in the formats required by each modelling approach. Section 7.2.4 presents summary output from using the unconditional approach to modelling the data obtained from this true population.

7.2.4 Unconditional Approach

This fitting approach was implemented with an initial 50000 particles of which 26610 survived until the end of the fitting process without simulating an invalid state element. These survivors produced an effective sample size of 115.98. This is considerably larger than the effective sample size obtained under the conditional approach in the previous section and is indicative of a pattern observed from fitting multiple simulated data sets as noted in section 7.4. Fitting models using the restrictive conditional approach seems to result in a more variable distribution of posterior weights for the particles than when the unconditional approach is used. A greater range of values is obtained for the weights under the conditional approach when compared with the unconditional approach.

Inference based on these surviving particles begins with summary data on the posterior samples of the state elements $\mathbf{n}_{0:T}$. As an example of the model output, Table 7.6 contains the summary data for the state element representing the population after the birth process during the second time period. The posterior means and medians for $\mathbf{n}_2$ seem to closely match the true states that were simulated in the original population with the intervals for each state element containing the true value. The posterior obtained for

State	Median	Mean	Truth	LCL	UCL	Expected	CV
n_{jjj}	1	1.83	3	0	7	2.35	1.170
n_{jja}	5	6.08	12	0	21	5.88	0.859
n_{jaj}	5	6.45	10	0	16	6.72	0.715
n_{jaa}	19	22.00	19	3	65	16.80	0.657
n_{ajj}	3	3.66	5	0	9	6.27	0.651
n_{aja}	13	12.90	16	3	21	15.70	0.350
n_{aaj}	16	16.40	17	4	29	17.90	0.353
n_{aaa}	58	55.40	36	9	82	44.80	0.308

Table 7.6: Summary for state vector at $t = 2$ for the unconditional approach. Expected values denote projected state totals obtained by multiplying the initial 100 animals by the true population parameters and iterating this process over each time period.

the analysis under the conditional approach is dominated by significantly fewer particles than the posterior under the unconditional approach. This is reflected in the relatively wide intervals and moderately large CVs observed in Table 7.6 compared to those obtained under the conditional approach.

The posterior distributions of the model parameters Θ are given in Table 7.7. The posterior means can then be compared with the true values from Table 7.1 as well as the means of the priors. Once again the posterior means are consistent with the true values although only three of the six independent posterior means are closer to the true value when compared to the location of the prior. The posteriors for adult survival ϕ_a, juvenile capture p_j and birth rate π_1 all have posterior probability masses centered further away from truth than their corresponding prior masses. The intervals for each parameter do contain the true values and it can be observed that the adult intervals are narrower than their juvenile analogues; the posteriors for both adult survival and capture display lower variance than the posteriors for juvenile survival and capture. It can also be seen that the

	Prior			Posterior					
Parameter		**Mean**	**Var**	**Median**	**Mean**	**Var**	**CV**	**LCL**	**UCL**
ϕ_j	(0.7)	0.500	0.04343	0.733	0.715	0.02150	0.2050	0.422	0.937
ϕ_a	(0.8)	0.811	0.02003	0.911	0.902	0.00311	0.0618	0.763	0.983
p_j	(0.5)	0.628	0.03910	0.664	0.675	0.01930	0.2060	0.424	0.911
p_a	(0.6)	0.501	0.04330	0.513	0.514	0.00267	0.1000	0.410	0.623
π_1	(0.4)	0.322	0.02132	0.285	0.298	0.00526	0.2440	0.182	0.463
τ	(0.3)	0.417	0.04153	0.298	0.329	0.03310	0.5520	0.084	0.871

Table 7.7: Summary of model parameters for the unconditional approach. True parameter values denoted in ( ).

largest variance is associated with the posterior mean for the splitting parameter τ. This is not particularly surprising as there is little information in the data to set the population splitting rate.

For this particular population, when comparing Tables 7.4 and 7.7, the performance of the conditional approach seems to be superior in that the posteriors are usually closer to the true values under the conditional approach when compared to the unconditional approach. It can also be noted that the posterior variances for each parameter are generally greater under the unconditional approach than under the conditional approach. The conditional approach may be expected to exhibit a higher level of precision as a result of the different, more restrictive, assumptions that are made when specifying the fitting algorithm for the chosen state-space model. However, both approaches produce intervals on the posterior distributions that include truth for most parameters.

Posterior summaries by age-cohort and time period are presented in Table 7.8. Once again the unconditional approach performs well and produces summaries of states that closely match truth for the time periods containing observed animals. The initial state

Time t	Median	Mean	Truth	LCL	UCL	Expected	CV
$t = 0$	30	32.6	40	7	87	30	0.57
$t = 1$	27	27.7	37	18	39	30.8	0.196
$t = 2$	28	28.3	35	19	42	33.3	0.209
$t = 3$	28	29.1	34	19	48	35.9	0.261

(a) Simulated juveniles.

Time t	Median	Mean	Truth	LCL	UCL	Expected	CV
$t = 0$	70	67.4	60	13	93	70	0.276
$t = 1$	87	86.1	79	74	96	77	0.0649
$t = 2$	97	96.4	83	79	109	83.2	0.0833
$t = 3$	109	108	92	85	127	89.8	0.105

(b) Simulated adults.

Table 7.8: Posterior summary of age classes over time obtained under the unconditional approach. Expected values denote projected state totals obtained by multiplying the initial 100 animals by the true population parameters and iterating this process over each time period.

vector $\mathbf{n}_0$ is simulated by splitting the 100 initial animals with no influence from the recorded observations. This resulting freedom is reflected in the large CVs at $t = 0$ for both juveniles and adults. The greater amount of information contained in the observed capture history patterns that relates to adult animals is reflected by the smaller CVs at each time period for the adults compared to the juveniles.

From comparing Tables 7.5 and 7.8 there is little difference between the performance of the modelling approaches when applied to this population. Both the conditional and unconditional approaches simulate totals of juveniles and adults at each time step that match closely with the true data. It could be argued that the conditional approach does a better job of simulating the juveniles at times $t = 2$ and $t = 3$ but the differences are

not especially pronounced. It should be noted that there is greater uncertainty associated with the posterior distributions produced by the unconditional approach when compared to those produced by the conditional approach and this can be seen by comparing the CVs.

7.3 Application to Soay Sheep

This section investigates the utility of the model fitting algorithms developed under the assumptions of the conditional and unconditional approaches with regard to fitting real data. This is not an attempt to reveal new insight into the data but instead should be viewed as an illustrative example of the efficacy of the modelling approaches when fitting a relatively simple model structure to the observed capture history data.

The data for this application consist of the population of Soay sheep (*Ovis aries*) on Hirta (St Kilda). This population has been well documented and examples of existing studies can be found in Milner *et al.* (1999), Catchpole *et al.* (2000), Coulson *et al.* (2001) and King *et al.* (2006). These studies will provide the summary information used for the elicitation of the prior distributions for the model parameters. The large number of analyses performed on these Soay sheep data provide considerable information about the parameters and population size over time and give a basis against which the performance of the two modelling approaches can be judged.

7.3.1 The Data and Model Formulation

The analysis is based on mark-recapture-recovery data for the Soay sheep from the Village Bay area of Hirta, in the St. Kilda archipelago off the west coast of Scotland. Further details are provided in the previously cited studies. The animals are classified by gender

and their year of birth is recorded. The extensive nature of the monitoring scheme has resulted in more than 95% of the sheep in the Village Bay area having been marked by 2000 (Catchpole *et al.*, 2000). The data on the individual sheep consists of capture history patterns as described in Chapter 2. However, in addition to the usual sequence of 1's and 0's denoting capture or non-capture respectively, some capture history patterns may also contain a 2 which denotes the recovery of a dead animal. The simple model structure used in the following analysis does not incorporate a recovery component, although such an extension is possible in theory, and for the purpose of the model fitting process it is assumed that recoveries are regarded as non-captures and labelled as 0's.

The data also contain various individual level covariates that include time-invariant measures, genetic information and time-varying covariates. A discussion of time-invariant and time-varying covariates with regard to modelling assumptions is presented in section 2.2.1.1. The analysis in King *et al.* (2006) focuses on the time-invariant covariates: sex, coat type, horn type and birth weight. Environmental covariates that apply to all animals in the population, such as population density and weather variables, are also considered. The weather variables studied in King *et al.* (2006) are the winter North Atlantic Oscillation, the total rainfall in both March and the autumn, and the average March temperature. The justification for the choice of these covariates is given in King *et al.* (2006).

The analyses conducted under the conditional and unconditional approaches will be based on mark-recapture data for the three years from 1996 to 1999. A total of 640 female sheep are included in this subset of the full Soay data. Analyses on longer time series of the data were attempted but required an excessive number of initial particle streams to avoid extreme particle depletion. Due to the restrictions imposed by the relatively simple

population dynamics model used in the model fitting process the data will be restricted to the females in the population.

The population dynamics model assumed for the following Soay sheep analyses takes the same form and parameterisation as that described in the examples in Chapters 4 and 5 and defined in section 4.4. The initial population in 1996 is assumed to consist of juvenile and adult females. An initial total population is specified and then split into the juvenile and adult cohorts using a binomial process with rate τ. For this process τ represents the probability that an animal in the initial population is assigned to be a juvenile. After splitting, both cohorts are subject to an annual survival process with ϕ_j and ϕ_a denoting juvenile and adult survival respectively. Juvenile females alive at time t that survive to time $t+1$ are assumed to mature and become capable of breeding during time period t. These mature females will then be classified as adults at $t+1$. The surviving females are then subject to a birth process that stochastically simulates new juveniles. For the Soay sheep a multiple birth process model is specified as the sheep are capable of producing twins as well as singletons. This requires the specification of the birth parameters (π_0, π_1, π_2) where π_j for $j = 0, 1, 2$ denotes the probability of an individual mature female producing j juveniles during the breeding season. Then, for each mature female, a draw is made from the multinomial distribution where the value $0, 1$ and 2 are simulated with probabilities π_0, π_1 and π_2 respectively. These simulated births are then summed to give the total number of juveniles produced by mature females for that time period.

The survival, maturation and birth processes are the biological processes included in the state process models under both the conditional and unconditional approaches. The capture process also needs to be specified. Each animal present in the population at time t is subjected to a stochastic capture process and is captured with probability p_j if a

juvenile and p_a if an adult. This same sequence of sub-processes then occurs until the final capture occasion at time T (1999). For the Soay analyses conducted in the next sections $T = 3$. The functional forms of the stochastic sub-processes that constitute the embedded population dynamics model are specified in the sets of Equations (4.4.1a):(4.4.1b) and (4.4.1d):(4.4.1f). Due to the multinomial birth process the functional form specified in Equation (4.4.1c) is replaced by that in Equation (7.3.1):

$$u_{2,t,[j]}|u_{1,t} \sim \sum_{r=1}^{u_{1,t}} \text{Multinomial}(1; \pi_0, \pi_1, \pi_2). \tag{7.3.1}$$

The capture history patterns for the individually identifiable Soay sheep that were observed in the population between 1996 and 1999 are used to construct the observed data under the different formats required by the two model fitting approaches. The usual distinctions apply and are briefly reviewed again. The observed data at each time point t for the conditional approach consists of the abundances corresponding to the distinct capture history patterns that include capture during time period t. For the unconditional approach the observed data at time t consists of the abundances of capture history patterns that include capture at some point between 1996 and 1999.

7.3.2 Prior Elicitation

Due to the multiple birth process in this model structure there are now eight parameters in the model: $\tau, \phi_j, \phi_a, \pi_0, \pi_1, \pi_2, p_j, p_a$. Only seven of these are independent as by definition $\pi_0 + \pi_1 + \pi_2 = 1$. For an initial analysis on the Soay data a relatively parsimonious model is assumed in which all parameters are time-invariant, remaining constant across all three time periods. However, age-specific survival and capture rates are assumed, with distinct rates assumed for juveniles and adults (including yearlings). The analyses conducted in Catchpole *et al.* (2000) are used to inform the choice of prior means for each of these

seven parameters.

The survival rates can be obtained from Table 6 in Catchpole *et al.* (2000). The rates are specified for two different classifications of horn-type, either scurred or non-scurred, and for four distinct age categories: lambs, yearlings, adults and seniors. It is assumed that horn-type affects the survival of lambs only and therefore the rates for the other three age categories are the same across horn-type. Hence, a point estimate of adult survival (ϕ_a) is 0.919. The survival rates for lambs are estimated as 0.752 and 0.518 for scurred and non-scurred horns respectively. It is reported in Catchpole *et al.* (2000) that there were 188 animals with scurred horns and 538 animals with non-scurred horns. Using these totals to take a weighted average of survival for all lambs without distinguishing by horn-type yields a juvenile survival rate (ϕ_j)of 0.578.

In their discussion section Catchpole *et al.* (2000) obtain an average capture probability for all females of 0.93. The elicitation of estimates for a juvenile capture rate is a little more convoluted. Due to the model structure described in section 7.3.1, juvenile capture can be regarded as a confounding of juvenile post-natal survival and subsequent capture. The embedded population dynamics model specified in section 7.3.1 it is assumed that there is no juvenile mortality between the moment of their birth and the following capture occasion. However, the existing studies identify a degree of post-natal mortality in the Soay population and this needs to be incorporated into the estimate of juvenile capture for this model structure. Then, given a capture probability of 0.93 for females and a survival probabilty of 0.578 for juveniles, the probability of a female juvenile surviving after birth and being captured is then $0.578 * 0.93 = 0.538$. Hence, a prior mean for juvenile capture rate is 0.538.

The existing studies on the Soay sheep data typically focus their inference on estimates of survival and capture rates. There are no explicit estimates provided for the distribution of birth rates in these studies. The supplied Soay data includes records of the numbers of juveniles born to observed animals during the monitoring study. Using the first 11 years of data the estimates of the distribution of birth rates can be obtained. By examining the relative frequencies of births the estimates $0.496, 0.439, 0.065$ are obtained for π_0, π_1, π_2. However the recorded births include both males and females whereas the observed capture history patterns that constitute the data used in the following analyses correspond only to females. Hence, for the specified population dynamics model in section 7.3.1, the birth parameters are a convolution of birth rates and gender split.

Table 8 in the supplementary material associated with Coulson *et al.* (2001) classifies the composition of the population of Soay sheep by age and gender at each time period. When averaged across all time periods, approximately 18% of the population consisted of female lambs. For males lambs this figure was 16%. Thus, the relative split of juveniles was assumed to be roughly $1 : 1$. Using this estimate the birth rates could be adjusted to represent the probability of a single mature female giving birth to no female lambs, a female singleton or female twins. The resulting estimates for the birth rates of female juveniles are 0.748, 0.219 and 0.033.

The same source of information (Table 8 in Coulson *et al.* (2001)) can also be used to obtain an estimate for the population splitting parameter τ. The average proportion of mature females in the population over the course of the study is obtained by summing the averages specified for female yearlings, female adults and older females. From this it can be estimated that, on average, 47% of the population of Soay sheep are classified as mature females in a given time period. The corresponding figure for female juveniles was earlier seen to be 18%. The ratio between these two proportions yields the relative

	ϕ_j	ϕ_a	p_j	p_a	π_0, π_1	τ
Prior Means: μ_α	0.580	0.919	0.538	0.930	0.748,0.219	0.277
Prior Variances: σ^2_α	1	1	1	1	1,1	1

Table 7.9: Specification of parameters μ_α and σ^2_α used to simulate priors for the Soay sheep analysis. Values for σ^2_α are on the logit scale.

proportion of juvenile females to adult females in the Soay population. Thus, a prior mean for the population splitting rate τ can be set at 0.277.

The final prior value to be specified is that of the initial population of female Soay sheep in 1996. From the observed Soay data a total of 270 individual animals were captured in 1996. Due to the high adult capture rates obtained from previous studies it was assumed that the 270 individual represented the majority of the sheep that were available to be caught in 1996. Hence, a prior was then simulated for the initial population using a left-truncated Poisson distribution with parameter 290. The left-truncation was set at 270 as this provided a lower limit on the population known to have existed at the start of the study.

The priors are then simulated using the approach detailed in section 7.2.1. The means (μ_α) and variances (σ^2_α) used to implement this prior simulation algorithm are given in Table 7.9. These values of 1 for each σ^2_α is on the untransformed logit scale rather than the probability scale. With regard to the birth rates; once prior values have been simulated for π_0 and π_1 given π_0, the prior value for π_2 is set deterministically to respect the constraint that $\pi_0 + \pi_1 + \pi_2 = 1$. The same set of priors will be used for fitting models under both the conditional and unconditional approaches.

An initial analysis was performed on the full Soay data set consisting of all individuals captured on at least one occasion between 1996 and 1999. A model was fitted to this observed data using a fitting algorithm based on the assumptions of the conditional approach. Of the 500000 particles that were simulated from priors centered on the values specified in Table 7.9, 461861 plausible particles remained after model fitting and these particles yielded an effective sample size of only 1.346. A second analysis was then conducted in which the model fitting approach developed under the unconditional algorithm was used. In this instance the same 500000 initial particles from the analysis under the conditional modelling approach were used and 270448 of them remained after model fitting. The effective sample size based on these particles was 7.916. Although an improvement on the result obtained under the conditional approach, this effective sample size is still too small for reliable inference to be made on the output from the model fitting. The results from these analyses do not adequately represent the utility of the fitting algorithms. To improve the results of analyses on this dataset two options were considered: either repeating these analyses increasing the number of particles until the effective sample size attained a suitable level or sub-sampling the data to allow more flexibility in the fitting process.

All analyses were performed using algorithms written for the statistical software *R* (see http://cran.r-project.org/). With regard to the first option, running both fitting algorithms simultaneously for the initial analyses took 132 hours using an *i*686 processor with CPU speed of approximately 3.06GHz. From monitoring the weights associated with the particles it was observed that even an exponential increase in the number of particles resulted in little growth in the effective sample size under the conditional generation approach. The computational burden and time required to obtain a suitably large effective sample size were deemed to be excessive and a more efficient solution was sought.

The second option was the chosen solution. The high adult survival and capture probabilities, means of 0.930 and 0.919 respectively, resulted in a proposal density that severely restricts the viable state and parameter space the posterior could occupy. Under this parameterisation the observed data effectively represented the majority of adult animals that were present in the population during each sampling occasion. The fitting algorithm under the conditional approach therefore had very little flexibility with regard to the number of unobserved animals that could be simulated subject to the constraints imposed by the observed counts. Similarly, 46% of the initial particles were implausible for the fitting algorithm specified under the unconditional approach. The complexities in the data were not adequately accounted for in the structure of the model. The simple form of the state-space model specified for the Soay population dynamics and the high adult capture and survival rates used to set the priors were not consistent with the mark-recapture data. One approach to addressing this problem is to sub-sample from the original dataset to reduce the proportion of the animals that were detected during each capture occasion. With the current parameterisation, conditional on survival, the probability of not seeing an adult animal during the three period study is $(1 - 0.919)^3 \approx 05$. Removing a proportion of observed captures would result in the detection probability falling to a level that corresponds to some of the sheep not being observed. For example, retaining only 30% of the captures would result in the probability of missing a sheep over the duration of the study being approximately $(1 - (0.3 \times 0.919))^3 \approx 0.38$.

To further increase the flexibility of the model time-varying parameters were allowed. From smaller scale simulations, not presented here, it was observed that temporal variability in survival provided the best results in terms of increasing the effective sample

	$\phi_{j,1}$	$\phi_{a,1}$	$\phi_{j,2}$	$\phi_{a,2}$	$\phi_{j,3}$	$\phi_{a,3}$
Prior Means: μ_α	0.580	0.919	0.580	0.919	0.580	0.919
Prior Variances: σ^2_α	0.5	0.5	0.5	0.5	0.5	0.5

Table 7.10: Specification of parameters μ_α and σ^2_α used to simulate priors for survival rates in the analysis of the sub-sampled Soay data. Values for σ^2_α are on the logit scale.

size relative to the earlier analysis based on constant parameters. Alternative parameterisations may well provide a superior fit to the data but they would require a new fitting algorithm to be written to fit models under the conditional approach. Due to the complexity of constructing these fitting algorithms this is left as an area for future research.

Allowing survival rates to vary by time requires the definition of some new notation. Let $\phi_{j,t}$ and $\phi_{a,t}$ denote the probabilities of survival from time $t-1$ to time t for juveniles and adults respectively. Table 7.10 now contains the means and variances used to generate draws for the priors for the six parameters required to model time-varying survival. To decrease the capture rate of the Soay sheep 70% of the captures were discarded. The matrix of capture-history patterns for the sheep data was examined and each entry of '1' (corresponding to a capture) was re-assigned a value of either 0 or 1 with probabilities 0.7 and 0.3 respectively. Accordingly, the means used to generate the priors for the non-survival parameters are given in Table 7.11. From the initial analysis it was noted that the relatively wide priors may contribute to the efficiency of the algorithm. The variance of the priors was reduced accordingly; each prior variance was calculated using value of 0.5 for the parameter σ^2_α in contrast to the value of 1 used in the initial analysis. These values of 0.5 and 1 are on the untransformed scale rather than the probability scale.

	p_j	p_a	π_0, π_1	τ
Prior Means: μ_α	0.1614	0.279	0.748,0.219	0.277
Prior Variances: σ^2_α	0.5	0.5	0.5,0.5	0.5

Table 7.11: Specification of parameters μ_α and σ^2_α used to simulate priors for capture, birth and population split for the Soay sheep analysis. Values for σ^2_α are on the logit scale.

7.3.3 Conditional Approach

The model with time-varying survival was fit to the sub-sample of the Soay data using the algorithm based on the assumptions of the conditional approach. An initial 250000 particles were simulated from the priors and their distributions are summarised in Table 7.12. Following the fitting process a total of 220796 particles were still plausible with respect to the constraints imposed by the data. The effective sample size for these remaining particles was only 1.349. This is virtually identical to the effective sample size obtained under the initial conditional analysis. This extremely small value renders any attempted subtle interpretation of the summaries on the posterior virtually meaningless.

The usual summary output of the posterior distributions is produced to investigate the behaviour of the fitting approach. Table 7.13 presents a summary of the posterior distribution of the state elements defining the population of female Soay sheep immediately after the second capture occasion. The states corresponding to animals that are captured during the second capture occasion are, as required, fixed at the values specified in the observed data. This is reflected in the CVs of zero for these four elements. The small ESS is reflected in the narrow intervals for the four elements that are simulated during the importance sampling fitting approach. Three of the elements, n_{j0}, n_{y10} and n_{a10}, have medians that match either the upper or lower limits of their corresponding intervals. Therefore, due to such a small ESS these intervals are rendered virtually meaningless.

Parameter	Prior Mean	Prior SD
$\phi_{j,1}$	0.5760	0.1158
$\phi_{a,1}$	0.9114	0.04109
$\phi_{j,2}$	0.5756	0.1157
$\phi_{a,2}$	0.9112	0.04111
$\phi_{j,3}$	0.5758	0.1157
$\phi_{a,3}$	0.9113	0.04105
p_j	0.1723	0.07026
p_a	0.2897	0.09877
π_0	0.7368	0.06823
π_1	0.2263	0.06846
π_2	0.0370	0.01831
τ	0.2872	0.09855

Table 7.12: Summary of priors for each parameter for the Soay sheep application. Survival rates are allowed to vary by time period.

State	Median	Mean	LCL	UCL	CV
n_{j0}	43	45.8	43	81	0.3730
n_{j1}	31	31.0	31	31	00
n_{y10}	16	15.0	6	16	0.3630
n_{y11}	4	4.0	4	4	00
n_{a00}	114	115.0	105	130	0.0804
n_{a01}	50	50.0	50	50	00
n_{a10}	49	50.0	49	57	0.1060
n_{a11}	17	17.0	17	17	00

Table 7.13: Summary for state vector at $t = 2$ for the conditional approach to the Soay sheep analysis.

The distribution of the animals across these eight state elements reflects the posterior means for the adult capture and survival rates denoted in Table 7.14. In the initial analysis on the full data set the posterior means for the captured and uncaptured juveniles were not consistent with the priors on the parameters or the simple structure of the model. For this analysis, albeit from a very small effective sample size, the posterior summaries for the elements of the state vector at $t = 2$, presented in Table 7.13, are more consistent with the data and the model structure when compared to the initial analysis. Having examined the posterior distributions on one state vector, attention can now be turned to the posterior distributions on the parameters Θ.

The posteriors for Θ are summarised in Table 7.14 and should be compared with the prior distributions summarised in Table 7.12. The posterior means are, in general, reasonably close to the prior means for many of the parameters but there is some change of location. The posterior means for juvenile survival at each time period are 0.5300, 0.5420 and 0.5390 respectively and these are compared to their prior means of 0.5760, 0.5756

Parameter	Median	Mean	Var	CV	LCL	UCL
$\phi_{j,1}$	0.5310	0.5300	0.005560	0.1410	0.4400	0.6460
$\phi_{a,1}$	0.9320	0.9290	0.001020	0.0344	0.8610	0.9420
$\phi_{j,2}$	0.5460	0.5420	0.004610	0.1250	0.4360	0.5790
$\phi_{a,2}$	0.8700	0.8750	0.001260	0.0405	0.8700	0.9470
$\phi_{j,3}$	0.5440	0.5390	0.004470	0.1240	0.4250	0.5830
$\phi_{a,3}$	0.9580	0.9510	0.001870	0.0455	0.8710	0.9580
p_j	0.3950	0.3760	0.012000	0.2910	0.2040	0.3950
p_a	0.3050	0.3050	0.001010	0.1040	0.2460	0.3290
π_0	0.7310	0.7260	0.001430	0.0520	0.6660	0.7310
π_1	0.2390	0.2410	0.001280	0.1480	0.2150	0.2930
π_2	0.0299	0.0325	0465	0.6640	0.0274	0.0859
τ	0.1660	0.1870	0.016600	0.6900	0.1660	0.3920

Table 7.14: Summary of model parameters for the Soay sheep example under the conditional approach. Survival rates were allowed to vary by time period.

and 0.5758. This suggests that juvenile survival may be slightly lower at each time period than initially assumed. The posterior means for adult survival increase, relative to the priors, in the first and third time periods but decreases in the second time period. This could suggest that the mean adult survival was lower during the second time period than in the first and third time periods. However, due to the small sample size it is unwise to draw too many conclusions of this nature. The low effective sample size may well reflect the susceptibility of this model fitting framework to high Monte Carlo error. A repeated analysis could yield a posterior that is, once again, dominated by a single particle and this particle could contain quite different parameters and states to that which dominates the posterior in this current analysis. Hence, it is dubious to assume that this posterior is a good approximation to the true distribution of states and parameters.

The posterior means for capture rates, for both juveniles and adults, do move away from their priors with the relative magnitudes of the capture rates switching between the age classes. That is, the posterior mean for juvenile capture is greater than that for adult capture which is in contrast to the prior. The posterior means for the birth parameters do not change very much when compared to the priors. The precision of the posterior distributions of the parameters noticeably increases, as evidenced by the decrease in the posterior variances, when compared to the priors. Again, this is an artefact of the significant particle depletion observed when fitting this model. The fact that the posterior means are either at, or in close proximity to, the limits of the intervals reiterates the effect of particle depletion and the resulting skewed distribution of weights in this analysis.

In the original analysis both adult capture and survival were assumed to be very high. A consequence of this was that the abundances of states corresponding to unobserved mature animals would be very low. At each time period t the element with the largest abundance was that which corresponded to the adult female sheep that are captured during t and all preceding time periods. This element was fixed at its observed total and, because it constitutes the majority of adult animals alive in the population at time t, this resulted in the elements corresponding to non-capture at time t exhibiting very low abundances for that first analysis. This presented some complications for the model fitting algorithm implemented under the assumptions of the conditional approach. As was discussed in Section 5.6, even for the bottom-up implementation of the fitting algorithm, it is possible to produce implausible particles due to the use of proposal distributions that do not fully respect the conditioning on the three years of observed data. When simulating states with very low abundances it is quite possible that a state element simulated at some time t_β would not be consistent with observations at some earlier time period t_α. For example, a simulated value of 1 in an unobserved element n_{i,t_β} at time t_β may

Time t	**Median**	**Mean**	**LCL**	**UCL**	**CV**
$t = 0$	45	51.6	45	111	0.669
$t = 1$	65	67.3	65.8	94	0.221
$t = 2$	74	76.8	74	112	0.223
$t = 3$	88	91.2	88	123	0.202

(a) Simulated juveniles.

Time t	**Median**	**Mean**	**LCL**	**UCL**	**CV**
$t = 0$	242	237	195	242	0.118
$t = 1$	242	241	226	251	0.0339
$t = 2$	250	250	242	264	0.0419
$t = 3$	273	274	269	301	0.0509

(b) Simulated adults.

Table 7.15: Posterior summary of age classes over time for the conditional approach to the Soay sheep data.

be inconsistent with a value of 0 in the observed element n_{k,t_α} at time t_α when n_{i,t_β} is a direct descendant of n_{k,t_α} via *mono-parent* elements only. For a three period study the duration is too brief for many of these potentially problematic relationships between ancestral elements to develop but these problems have been observed in studies of simulated data.

Summaries of the posterior distributions on the states aggregated into two age-cohorts are presented in Table 7.15. These summaries reflect aspects of the model fitting that have been identified previously in this section. Once again the probability masses of the posterior distributions are centred on narrow intervals with the posterior means or medians attaining, or lying close to, the limits of the intervals for both juveniles and adults at all time periods. The most variable aggregated states are the initial juveniles and initial

adults. This mirrors the behaviour observed in the earlier simulation studies and reflects the lack of information available in the data to inform the estimate of the population splitting parameter τ. The aggregated states in the first time period are obtained by simulating a random split of the initial population with no conditioning on the observed data. Thus, the animals in the initial time period are subject to the fewest constraints under the model fitting process and consequently exhibit the largest CVs. Comparing Tables 7.15a and 7.15b, it can be seen that the CVs are smaller for the adults than the juveniles during the same time period. This is to be expected as an artefact of the data and bottom-up fitting algorithm. The observed counts of capture history patterns contain more information about adult animals than juveniles. The low rates drawn for the priors on capture allow the model to exhibit some flexibility in the posterior for the aggregated states; a capture rate of 1 would result in the aggregated states being fixed at the total counts of observed animals. Thus, the model structure and the amount of information in the data place a greater restriction on the posterior for total adult females than for the total juvenile females. This is reflected in the CVs.

7.3.4 Unconditional Approach

The same sub-sampled Soay data are now analysed using the fitting algorithm based on the assumptions of the unconditional approach. The same 250000 particles simulated from the priors (Table 7.12) were used to initialise the approach. After the model fitting had been completed there were 214011 plausible particles remaining. The effective sample size based in these particles is 378.74. This ESS is considerably larger than that obtained under the conditional approach and allows more meaningful inference to be made. It can also be seen that by sub-sampling the Soay data to reduce the proportion of animals that are observed, the ESS has dramatically increased from the value of 7.916 obtained for the

State	Median	Mean	LCL	UCL	CV
n_{jjj}	2	2.11	0	6	0.923
n_{jja}	6	6.93	1	17	0.675
n_{jaj}	10	11.80	2	26	0.524
n_{jaa}	34	36.90	11	75	0.441
n_{ajj}	11	11.60	4	22	0.397
n_{aja}	34	34.30	19	51	0.229
n_{aaj}	59	59.40	39	85	0.201
n_{aaa}	188	186.00	133	228	0.129

Table 7.16: Summary for state vector at $t = 2$ for the unconditional approach to the Soay sheep analysis.

initial analysis under the unconditional approach.

As for the conditional approach, summary output of the posterior distributions of the states at time $t = 2$ is obtained. These states define the population of female Soay sheep immediately after the birth process during the second time period. Table 7.16 contains the summary data for these eight state elements. The extremely small number of particles that contributed to the posterior under the conditional approach meant that the intervals constructed from the weighted particles were extremely narrow. The increased ESS reflects lower Monte Carlo variation in the estimates of the posterior means when compared to the output from the model fitting under the conditional approach. However, the larger ESS results in wider intervals due to the greater number of particles that contribute to the posterior. This is an artefact of the exceptionally small ESS obtained under the conditional approach. With a slightly larger ESS the variance estimates would be expected to exhibit very high Monte Carlo variation and, simply by chance, could be very large or very small. The small ESS for the analysis under the conditional approach

Time t	Median	Mean	LCL	UCL	CV
$t = 0$	293	293	271	320	0.046
$t = 1$	320	320	284	355	0.0578
$t = 2$	350	349	293	400	0.0737
$t = 3$	384	384	302	461	0.0969

Table 7.17: Posterior summary of total population over time for the unconditional approach to the Soay sheep data.

is likely to result in posterior summaries that exhibit greater bias than those obtained for the analysis under the unconditional approach. Due to the single particle dominating the posterior for the conditional analysis it is not meaningful to compare the posterior variances for the state elements obtained under each approach.

Obviously, the states in Table 7.16 do not reflect the animals' capture status so there are no states fixed at observed values under this approach. The element corresponding to those mature females that are still present in the population at time $t = 2$ having entered the initial population as adults exhibits a large abundance. This is consistent with the high initial population and the high survival rates for mature females. The posterior summaries for the total population abundance over time are given in Table 7.17. From this is can be seen that the posterior means of 320, 350 and 384, at times $t = 1$, $t = 2$ and $t = 3$ respectively, are reasonably consistent with the known minimum abundances of 331, 332 and 360. The posterior mean for the population abundance seems to underestimate the true value at time $t = 1$ but otherwise there is a good correspondence between the posterior means and the known minimum abundances over time.

Parameter	Median	Mean	Var	CV	LCL	UCL
$\phi_{j,1}$	0.5460	0.544	0.011600	0.1970	0.3430	0.7500
$\phi_{a,1}$	0.9240	0.918	0.001200	0.0378	0.8330	0.9670
$\phi_{j,2}$	0.5770	0.577	0.014200	0.2070	0.3380	0.8230
$\phi_{a,2}$	0.9170	0.906	0.002330	0.0533	0.7820	0.9640
$\phi_{j,3}$	0.5900	0.578	0.013500	0.2010	0.3270	0.7840
$\phi_{a,3}$	0.9210	0.916	0.001450	0.0415	0.8310	0.9710
p_j	0.3300	0.310	0.012000	0.3530	0.0882	0.4870
p_a	0.2850	0.285	0.003460	0.2070	0.1380	0.4200
π_0	0.6770	0.680	0.005020	0.1040	0.5510	0.8360
π_1	0.2850	0.284	0.005300	0.2560	0.1230	0.4120
π_2	0.0319	0.036	0295	0.4770	0.0131	0.0789
τ	0.2600	0.265	0.007510	0.3270	0.1200	0.4670

Table 7.18: Summary of model parameters for the unconditional approach to the Soay sheep analysis.

The summaries of the posteriors for the model parameters Θ are presented in Table 7.18. Once again there is no significant difference between the posterior and prior means for many of the model parameters. The posterior means for juvenile survival at each time period are 0.544, 0.577 and 0.578 respectively and these are compared to their prior means of 0.5760, 0.5756 and 0.5758. The intervals for juvenile survival at each time period overlap considerably suggesting that the model fitting process has not yielded any large differences between the time-specific rates. Adult survival also does not vary across the time periods although all the posterior means for three parameters $\phi_{a,1}, \phi_{a,2}$ and $\phi_{a,3}$ are slightly greater than their prior means. The posterior summaries for the capture rates move from the priors in a similar fashion to that exhibited in the conditional analysis. As for that analysis the posterior mean for the juvenile capture rate is greater than that for adult Soay sheep although both parameters have overlapping intervals. The point estimates are reasonably similar across analyses although the posterior means for both the

juvenile and adult capture rates are lower under the unconditional analysis than under the conditional analysis.

The posterior means for birth rates mark a departure from the prior means and suggest that the relative frequencies of singletons and twins increase and decrease respectively when compared to the prior means. The average number of female yearlings produced by mature females increases slightly relative to the prior. The analyses, both conditional and unconditional, seem to produce broadly similar posterior distributions for Θ. The posterior survival rates for both analyses across all time periods do not move far away from their prior locations. The posterior capture rates change, relative to the priors, in the same directions and with similar magnitudes. The posterior birth rates demonstrate some contrast with very little change exhibited under the conditional approach but a more noticeable one occurring under the unconditional approach. The precision of the posterior distributions of the parameters again increases relative to the priors. The ESS of 1.3 obtained under the conditional approach explained the low estimated posterior variances observed for that analysis. In general, particle depletion will typically lead to higher Monte Carlo variation in the posteriors for the estimates, until approximately unitary ESSs are obtained. For this analysis under the unconditional approach the posterior variances are significantly smaller than those for the priors. Due to the reasonably large ESS of 370, it is likely that this reflects the data providing information about the value of the parameters, relative to what was specified in the priors.

Table 7.19 presents summaries of the posterior distributions on the aggregated juveniles and adults for each time period. The pattern exhibited in Table 7.19 is similar to that seen for the conditional approach in Table 7.15. For the model fitting under the conditional approach a single particle dominated the posterior for the aggregated mature

Time t	Median	Mean	LCL	UCL	CV
$t = 0$	72	74.6	32	130	0.328
$t = 1$	77	78.1	55	105	0.163
$t = 2$	85	84.9	58	119	0.173
$t = 3$	94	95.4	67	132	0.183

(a) Simulated juveniles.

Time t	Median	Mean	LCL	UCL	CV
$t = 0$	218	218	164	259	0.11
$t = 1$	241	242	207	270	0.0629
$t = 2$	264	264	227	301	0.0738
$t = 3$	289	289	231	347	0.0953

(b) Simulated adults.

Table 7.19: Posterior summary of age classes over time for the unconditional approach to the Soay sheep data.

females and the narrow intervals in Table 7.15b supported this. The corresponding intervals in Table 7.19b are wider and reflect the combined effect of the lack of required conditioning, the larger ESS and the posterior variation under the unconditional approach. Due to the small ESS obtained under the conditional approach it is difficult to quantify whether the wider intervals in Table 7.19b are due to the lower Monte Carlo variation or if it is a function of the model and data under the unconditional approach.

In both Table 7.19a and Table 7.19b the CV decreases between the initial time period and the remaining times which then display broadly similar CVs. This may be an artefact of the fitting algorithm developed under the assumptions of the unconditional approach. The observed capture history data consists of the abundances for the capture history patterns that include capture on at least one occasion during the study. This data provides

information about the composition of the population at each time period during which a capture occasion occurred but does not contain any information that can be used to determine the relative frequencies of juveniles and adults in the initial population. However, this assumed bias due to the fitting algorithm may be a particular feature of this state-space model with the given priors. Further investigations should be conducted to establish the relative veracity of these ideas.

7.3.5 Discussion

The structure of the models fitted to the observed capture history patterns in sections 7.3.3 and 7.3.4 is much simplified when compared to models selected in the analyses performed by Catchpole *et al.* (2000) and King *et al.* (2006). The classical models fitted by Catchpole *et al.* (2000) use maximum likelihood techniques to estimate survival, capture and recovery probabilities. The final model chosen for female Soay sheep models the survival parameters as age-dependent functions of the size of the population, the environmental covariate March rainfall and the individual covariate horn-type. Capture and recovery probabilities are modelled as being fully time-dependent.

The age structure for female Soay sheep is analysed in Section 4.1 of Catchpole *et al.* (2000) with models exhibiting a variety of age structures being fitted to the data and then compared using AIC. The age structure specified for the models in section 7.3.1 consisted of separate survival rates for juveniles and adults and is denoted as the ϕ_1, ϕ_a model in Catchpole *et al.* (2000). This model has the second-highest AIC of all age structured models considered by Catchpole *et al.* (2000). Thus, the evidence from existing studies is that, even discounting the effect of other covariates, the survival rates for female Soay sheep require a more complicated age structured model than that specified in section 7.3.1.

The Bayesian analysis of King *et al.* (2006) suggests that there is strong posterior support that age structured survival parameters are functions of coat type, birth weight, North Atlantic Oscillation (NAO) index and population size. The contribution of horn-type, March rainfall, March temperature and autumn rainfall are also incorporated into the model-averaged parameter estimates. Again, there is clear evidence that the assumption of an age structure consisting of juveniles and adults alone is far too simplistic to capture adequately the true underlying model for survival rates.

Attempting to fit these more complex model structures using the algorithms developed under the assumptions of the conditional approach can be problematic. Specifically, incorporating age dependency into the models for the survival, capture or birth processes cannot be performed in the traditional way when a bottom-up fitting algorithm is used. Denote the population at time t by N_t. If the probability of an animal surviving from time $t-1$ to time t is dependent on the population at $t-1$, N_{t-1}, then this density dependence can be incorporated into more traditional "top-down" sequential approaches with relative ease. However, for the bottom-up approach, once the state vector $\mathbf{n}_t$ has been simulated, the estimate of N_t and the survival rates are used to estimate the unknown population N_{t-1}. Therefore, due to the reverse-time progression of the model fitting algorithm, the population at $t-1$, N_{t-1}, cannot be used to determine survival rates between time $t-1$ and t because it is currently unknown. One possible approach to overcome this issue would be to use the total number of animals captured at time t as a proxy for the population size. This total would be known exactly at each time period for the model structure specified under the conditional fitting approach. However, if capture rates were assumed to be highly variable over time this observed total would not provide a good indication of the total population at each time period.

Attempts to fit more complex models have yet to be successful. Computational problems and extremely small effective sample sizes have prevented any meaningful analyses from being obtained. Developing more complicated models that allow a greater degree of flexibility in the age structure of a population is an area that requires further research. Specifying extra age classes would require the state and intermediate vectors to be expanded and would result in a more complex fitting algorithm. If the age on first capture of every recorded animal was able to be determined accurately then this expanded model would remove some of the convolutions of observed adults due to *multi-parent* nodes that exist in the existing model structure under the conditional approach. For example, if the elements of the state vector now classified animals by both capture history pattern and age then only yearlings would mature to become two-year-old adults with either a 100 or a 101 capture history pattern. Adults with a 100 or 101 capture history pattern would be at least three years old and would therefore contribute to a different state element. If the exact age of each observed animal was not recorded then attempting to incorporate this form of age structure into the model would result in many more convolutions. For example, the observed abundance of all adult animals with a 0001 capture history pattern would need to be split across all possible age classes that could produce adults with that capture history pattern.

More work needs to be done to develop these model fitting algorithms under both approaches to allow more complicated model structures to be incorporated into the analysis.

7.4 General Simulation Analysis of Approaches

This section of the investigation provides a more detailed examination of the performance of the model fitting algorithms under each of the conditional and unconditional

approaches. The same techniques described in section 7.2 can be applied again for this study. A single population is simulated via the processes specified for survival, birth and capture using specified values for these model parameters. Vectors of observed data are then obtained from this population realisation in the formats required for each approach. An application of each of the model fitting algorithms ensues based on N_{part} initial particles and the resulting posterior summaries for the states and parameters are stored. This process is then repeated a series of n_{rep} times; that is, the same population is analysed under both approaches n_{rep} times resulting in n_{rep} sets of samples from the posterior parameters and states. This is equivalent to a single analysis based on $N_{part} \times n_{rep}$ initial particles. The initial particles here will simply consist of the initial total N_0 and the set of parameters used in the survival, birth and capture processes. By conducting multiple analyses on the same observations it is possible to quantify the degree of Monte Carlo variation engendered by the fitting algorithm.

After the set of n_{rep} analyses are completed a new population is simulated thus generating new observed data and this new population is analysed n_{rep} times with N_{part} initial particles drawn for both approaches for each replicated analysis. A total of n_{pop} populations are simulated which means the entire simulation study will consist of $n_{pop} \times n_{rep}$ samples from the posterior distributions for each of the model fitting approaches. This two-level replication, of multiple analyses on multiple populations generated from the same initial parameters, allows the uncertainty in the posterior to be apportioned into two components: that which can be attributed to Monte Carlo variation and that which is due to process error. This can be formally analysed using the standard application of ANOVA tests.

For consistency with the single simulated populations generated in section 7.2 the parameters used to simulate each of the n_{pop} populations are the same as those in Table 7.1. The specification and simulation from the prior distributions for each of these six model parameters proceeds in the manner outlined in section 7.2.1 where the mean and variance controlling parameters for each prior need to be specified. This specification will vary depending on both the population structure being investigated and focus of the inference. Using the notation from section 7.2.1 it can be seen that setting the variance controlling parameter σ_α^2 to zero, results in a corresponding prior for α that is fixed at μ_α; that is, all the probability mass is placed on μ_α. By fixing the priors in this way the inference of the model fitting algorithm is restricted to the state elements. This approach then provides a clearer illustration of the ability of the model algorithms to re-create the population from which the observed data was obtained.

The next sections present the results from simulation studies used to investigate model performance in a variety of situations. For the first scenario 25 separate populations will be simulated with each analysis being repeated 4 times on each of these populations to produce 100 samples from the posterior distribution on states and parameters. Each model fitting will be based on 5000 initial particles. In summary, $n_{pop} = 25$, $n_{rep} = 4$ and $N_{part} = 5000$ for the following study.

7.4.1 Two Age Classes. Fixed Priors

The first situation focuses on the simulation of populations composed as outlined in section 7.2. The priors are fixed at the values used to simulate each of the populations, hence there is no need to produce posterior summaries for these parameters.

	ϕ_j	ϕ_a	p_j	p_a	π_0, π_1	τ
Prior Means: μ_α	0.7	0.8	0.5	0.6	0.6,0.4	0
Prior Variances: σ^2_α	0	0	0	0	0,0	0

Table 7.20: Specification of parameters μ_α and σ^2_α used to simulate priors for each parameter. Values for σ^2_α are on the logit scale.

For this simulation the means and variances specified in the prior generation routine are given in Table 7.20. The initial population size for this and all simulation studies discussed in this chapter is 100 animals.

The 100 model fittings under the conditional approach produced an average ESS of 59.250; with the corresponding figure for the unconditional approach being 1372.757. Thus, as was noted in section 7.2, the average effective sample size is greatly increased by relaxing the requirement to condition on the observed capture history patterns at each time step. This also helps to illustrate the effect of particle depletion on the estimates; on average a significantly larger proportion of the initial particles simulated for the unconditional approach had non-zero weight compared to those for the conditional approach. The weights produced for the conditional approach have most of their probability mass distributed around a narrow range of positive values but also exhibit a spike at zero. Whereas, on average, the weights under the unconditional approach have a less pronounced spike at zero and exhibit a more uni-modal distribution. Table 7.21 defines the column headings used in Table 7.22 which summarises the effects of the different distribution of particle weights under each approach.

Due to the parameters being fixed there is nothing of interest in the posterior summary on the parameters and this part of the model output is omitted for this set of simulations.

Label		Definition
Exp. Val	=	The expected value of the quantity of interest.
$\hat{\mu}$	=	The posterior means of the quantity of interest.
$\hat{\mu}_{\bar{x}}$	=	The mean of the posterior means obtained from the 100 simulations.
$\hat{\mu}_{SD}$	=	The standard deviation of the posterior means obtained from the 100 simulations.
$\hat{\sigma}$	=	The posterior standard deviation of the quantity of interest.
$\hat{\sigma}_{\bar{x}}$	=	The mean of the posterior standard deviations obtained from the 100 simulations.
$\hat{\sigma}_{SD}$	=	The standard deviation of the posterior standard deviations obtained from the 100 simulations.

Table 7.21: Definitions of the symbols and labels used in the summary tables.

7.4.1.1 State Summary

Each of the 100 model fittings results in a posterior sample of state elements under each approach. For each of these samples the weights for the 5000 particles and the simulated values contained in the particles are used to produce point estimates and associated standard errors for the model parameters and the unknown state elements. The point estimates will just be the weighted mean across all surviving particles and the distribution of these means across simulations is then summarised.

The population generated by the fitting algorithms in each of the 100 simulations can be summarised by aggregating the states at each time period into two elements classified by age-cohort: all juveniles and all adults that are alive in the population following the sampling occasion at t. This data is then presented in Table 7.22. The column headings used in Table 7.22 are defined in Table 7.21. For this state summary the "quantity of interest" will be the aggregated age cohort abundances at each time period. Thus, the "Exp. Value" column denotes the expected abundances for those state aggregations that

			CGA				UGA			
Time	Total	Exp. Val	$\hat{\mu}_{\bar{x}}$	$\hat{\mu}_{SD}$	$\hat{\sigma}_{\bar{x}}$	$\hat{\sigma}_{SD}$	$\hat{\mu}_{\bar{x}}$	$\hat{\mu}_{SD}$	$\hat{\sigma}_{\bar{x}}$	$\hat{\sigma}_{SD}$
t=0	J	30	30.01	0.9603	4.414	0.5029	30.04	0.2416	4.577	0.08086
	A	70	69.99	0.9603	4.414	0.5029	69.96	0.2416	4.577	0.08086
t=1	J	30.8	29.78	3.124	3.042	0.4194	30.33	1.833	3.651	0.07538
	A	77	76.57	2.073	2.979	0.3682	76.61	1.713	3.343	0.08231
t=2	J	33.26	33.44	4.183	3.143	0.4008	33.32	3.143	3.589	0.08073
	A	83.16	82.42	3.638	3.498	0.3491	82.84	3.239	4.495	0.06683
t=3	J	35.93	35.79	3.611	3.773	0.4712	35.84	3.578	3.91	0.1224
	A	89.81	89.06	5.654	4.639	0.7601	89.39	5.044	5.631	0.115

Table 7.22: Summaries on the Posteriors of aggregated states for the multiple simulations analysis with priors fixed at truth. J denotes aggregated juveniles, A denotes aggregated adults. CGA denotes the summaries obtained under the conditional generation approach. UGA denotes summaries obtained under the unconditional approach.

are based on a population with an initial 100 animals. This population is simulated using the approach in section 7.2 based on the parameter values in Table 7.20. The expected values are then obtained by multiplying the elements of the state vectors by the parameters associated with the sub-processes that drive the population. A similar idea was used to obtain the expected values in the splitting rates derived in section 5.2.1.

It can be seen from Table 7.22 that both the conditional generation approach and the unconditional approach are performing extremely well, on average, in terms of simulating the expected aggregated totals for the two age classes at each time. There is less variability in the posterior means under the unconditional approach compared to the conditional approach for all age cohorts and time periods which supports the disparity in average

effective sample sizes for both approaches. However the difference decreases as the time period increases; that is the variability about the posterior means was roughly the same under both approaches at time $t = 3$ but was markedly more for the conditional approach relative to the unconditional approach at time $t = 0$. The standard deviation associated with the posterior means is roughly equal for both approaches although the variability of the standard deviation associated with the posterior means is, again, greater for the conditional approach.

7.4.1.2 Analysis of Variance for States

As discussed in section 7.4 the model fitting algorithms are applied to each of the 25 populations 4 times. Then, an analysis of variance can be conducted to investigate the composition of the uncertainty associated with variability of the posterior mean for the quantity of interest which, as before, is the aggregated age cohort abundances at each time period for this set of simulations. Define inter-population variation as the variance of the posterior mean for the quantity of interest across the 25 populations, and intra-population variation as the variance of the posterior mean across the four replicates of a given population. Then, this ANOVA approach can determine whether the inter-population or intra-population variation makes the larger contribution to the overall variance.

Table 7.23 contains the relevant ANOVA output for both approaches with the third, fourth and fifth columns representing the relative degree of variation due to the repeated analyses on the same population, analysing different populations and the interaction between them respectively. It can be seen that, for the conditional generation approach in Table 7.23a, the F-Ratio for including population as a factor is highly significant but the F-Ratio for including replicate as a factor is not. This is supported by examining the

fourth column and noting that over 90% of the variance in the posterior means for the aggregated states is explained by using different populations. Thus, the inter-population variation comprises a much greater component of variation than the intra-population variation. Process error seems to significantly outweigh Monte Carlo error for this set of simulations. Table 7.23b demonstrates this even more strongly.

This approach is extended in the next section by allowing the model parameters Θ to be drawn from prior distributions rather than set at fixed values.

7.4.2 Two Age Classes. Variable Priors

This modelling scenario once again focuses on populations exhibiting the same structure as in section 7.2. Initial draws of size N_{part}, in this case 10000, are made for each parameter from the prior distributions specified using the values in Table 7.24. For this scenario 10 distinct populations will be simulated with each model fitting analysis being repeated 10 times per population. This will yield 100 samples from the posterior distribution on the states and parameters. In summary, for this scenario, $n_{pop} = 10$, $n_{rep} = 10$ and $N_{part} = 10000$ for the following study. As before it is assumed that there are 100 animals present in the population at the start of the mark-recapture study.

The simulated draws from the prior distributions are summarised in Table 7.25. The fourth column in that table headed "True Value" simply denotes the value taken by that parameter when used to simulate the true population. This true population yields the data that is then used by the model fitting algorithms. Again it is noted that the birth parameters π_0, π_1 have the same prior standard deviation due to the constraint that $\pi_0 + \pi_1 = 1$.

		Scaled Sums of Squares						
		Rep.	Pop.	Int.	Total	Est sd	$Pr(\geq F)$ Pop	$Pr(\geq F)$ Rep.
t=0	J	0.03727	0.4016	0.5611	1	0.9555	0.006883	0.1983
	A	0.03727	0.4016	0.5611	1	0.9555	0.006883	0.1983
t=1	J	0.003511	0.9483	0.04815	1	3.108	4.36e-38	0.1645
	A	0.00168	0.9234	0.07495	1	2.063	2.488e-31	0.6577
t=2	J	08051	0.9845	0.01471	1	4.162	1.636e-56	0.2765
	A	0.002552	0.9708	0.02668	1	3.62	3.153e-47	0.08496
t=3	J	0.001562	0.9667	0.03174	1	3.593	1.495e-44	0.323
	A	06753	0.987	0.01234	1	5.625	3.016e-59	0.2767

(a) ANOVA output for the conditional approach.

		Scaled Sums of Squares						
		Rep.	Pop.	Int.	Total	Est sd	$Pr(\geq F)$ Pop	$Pr(\geq F)$ Rep.
t=0	J	0.01979	0.7915	0.1887	1	0.2404	3.093e-17	0.06488
	A	0.01979	0.7915	0.1887	1	0.2404	3.093e-17	0.06488
t=1	J	2.428e-05	0.998	0.002004	1	1.824	1.262e-87	0.8319
	A	01277	0.9977	0.002171	1	1.704	2.237e-86	0.2462
t=2	J	1.32e-06	0.9994	05939	1	3.127	1.236e-106	0.9836
	A	3.161e-05	0.9987	0.001245	1	3.222	4.608e-95	0.6111
t=3	J	2.476e-06	0.9995	04693	1	3.56	2.572e-110	0.944
	A	2.604e-05	0.9993	07023	1	5.019	5.167e-104	0.4506

(b) ANOVA output for the unconditional approach.

Table 7.23: ANOVA output for conditional and unconditional approaches for the multiple simulations analysis with priors fixed at truth. J denotes aggregated juveniles, A denotes aggregated adults.

	ϕ_j	ϕ_a	p_j	p_a	π_0, π_1	τ
Prior Means: μ_α	0.5	0.85	0.65	0.5	0.7,0.3	0.4
Prior Variances: σ^2_α	1	1	1	1	1,1	1

Table 7.24: Specification of parameters μ_α and σ^2_α used to simulate priors for each parameter where priors are neither centered on truth nor fixed.

Parameter	**Prior Mean**	**Prior SD**	**True Value**
ϕ_j	0.5012	0.2072	0.7
ϕ_a	0.8101	0.1422	0.8
p_j	0.6231	0.1969	0.5
p_a	0.4989	0.2090	0.6
π_0	0.6791	0.1453	0.6
π_1	0.3209	0.1453	0.4
τ	0.4158	0.2045	0.3

Table 7.25: Summary of prior distributions for each parameter where priors are neither centered on truth nor fixed.

The 100 model fitting simulations were then conducted using a procedure similar to that outlined in the pseudo-algorithm of section 7.2.2. This resulted in an average ESS obtained for the conditional approach of 4.737; with the unconditional approach yielding an average of 32.500. Both of these values are notably smaller than their corresponding values in section 7.4.1 and such a small ESS for the conditional approach may result in the model fitting process exhibiting large Monte Carlo error. Equally, it would be unwise to place too much faith in inference based on such a small effective sample size. As seen in Table 7.23a, a small ESS can result in a significant proportion of the uncertainty in the model being explained by random variation across each simulation rather than by a population or replicate effect.

The stochastic simulation of parameters from a joint prior distribution means that for this scenario the posterior sample of the model parameters will be of interest. The main issue will be to determine the relative performance of the model fitting algorithms with regard to their ability to correctly identify the true population parameters.

7.4.2.1 Posterior Summary

The weighted bootstrap samples obtained from each of the 100 simulations are used to approximate the posterior distributions for each quantity of interest under each modelling approach. As for the state summary in section 7.4.1 point estimates and associated standard errors are obtained for each model parameter. The distributions of the point estimates are then summarised in Table 7.26 in the same format as was defined in Table 7.21. Then for each parameter under both modelling approaches the distributions of the posterior means and standard deviations are summarised.

	CGA				UGA			
Parameter	$\hat{\mu}_{\bar{x}}$	$\hat{\mu}_{SD}$	$\hat{\sigma}_{\bar{x}}$	$\hat{\sigma}_{SD}$	$\hat{\mu}_{\bar{x}}$	$\hat{\mu}_{SD}$	$\hat{\sigma}_{\bar{x}}$	$\hat{\sigma}_{SD}$
ϕ_j	0.6935	0.07530	0.09936	0.02999	0.6905	0.05486	0.11860	0.01581
ϕ_a	0.8428	0.05587	0.06424	0.02574	0.8379	0.05684	0.07011	0.01364
p_j	0.5885	0.09739	0.10320	0.03225	0.5881	0.07624	0.11860	0.01426
p_a	0.5750	0.04500	0.06186	0.01684	0.5800	0.05204	0.06313	0.01064
π_0	0.6320	0.06390	0.07924	0.02334	0.6263	0.06023	0.08969	0.01920
π_1	0.3680	0.06390	0.07924	0.02334	0.3737	0.06023	0.08969	0.01920
τ	0.4105	0.12170	0.20040	0.06393	0.3723	0.05868	0.20400	0.03173

Table 7.26: Summaries on the posterior distribution of the model parameters Θ for the multiple simulations analysis with priors neither centered on truth nor fixed. CGA denotes summaries obtained under the conditional generation approach. UGA denotes summaries obtained under the unconditional generation approach.

It can be seen that, on average, the conditional generation approach performs reasonably but sometimes struggles to accurately estimate the true parameters that generated the 10 simulated populations. Estimated adult capture rate p_a is smaller than the juvenile capture rate p_j for both the conditional and unconditional approaches although the capture rates are quite close to one another in both cases. This could also be an effect of the small effective sample size; with only 4 effective particles in the posterior sample it is not surprising that some atypical behaviour is exhibited. It should also be noted that the posterior mean of p_j does not move very far from the prior mean and the average standard deviation for p_j is the second largest under the conditional approach and the largest under the unconditional approach. This may suggest that there was relatively weak information in the data for estimating the juvenile capture rate. It should also be noted that the relatively short time series of observations may result in the parameters being confounded; for example the birth rate π_1 and juvenile capture p_j may well be confounded.

7.4.2.2 Analysis of Variance for Parameters

As for the aggregated states in the previous section the simulation output can be analysed to investigate what influence different sources of variance that contribute to the uncertainty in the posterior estimates of the parameters. Specifically the division between inter-population and intra-population variance is of interest.

Consider the conditional approach; it can be seen from the F-Ratios in Table 7.27a that population should be included as a factor but replicate should not be for all parameters. There is strong evidence that including population as a factor explains a significant proportion of the variation in the posterior estimates for that parameter. However, the proportion of the variation that is attributed to the interaction term is quite sizeable for the conditional approach supporting the idea that the small effective sample size will lead to Monte Carlo variation. Examining the output for the unconditional approach it can be seen that the larger effective sample size corresponds to a much greater significance attached to the F-Ratio for including population as a factor. However there is some weak evidence to suggest that replicate should be included as a factor for juvenile capture p_j. Apart from τ all other parameters have over 90% of the variance in their posterior means explained by the difference between simulated populations. Process error then outweighs Monte Carlo error under the unconditional approach but the small average ESS of the conditional approach sees a less pronounced difference between the two sources of variation.

7.4.2.3 State Summary

Similar behaviour is then also displayed in Table 7.28. Looking at the distribution of the means on the aggregated states it seems that both the unconditional and conditional

	Sum of Squares						
	Rep.	Pop.	Int.	Total	Est sd	$Pr(\geq F)$ Pop.	$Pr(\geq F)$ Rep.
ϕ_j	0.04897	0.6354	0.3156	1	0.07492	4.443e-16	0.2036
ϕ_a	0.03427	0.6186	0.3471	1	0.05559	9.712e-15	0.5392
p_j	0.02131	0.7202	0.2585	1	0.09690	5.933e-20	0.6695
p_a	0.04439	0.5612	0.3944	1	0.04477	1.97e-12	0.4369
π_0	0.01380	0.6342	0.3520	1	0.06358	7.462e-15	0.9537
π_1	0.01380	0.6342	0.3520	1	0.06358	7.462e-15	0.9537
τ	0.08645	0.2634	0.6501	1	0.12110	07278	0.3085

(a) ANOVA output for the conditional approach.

	Sum of Squares						
	Rep.	Pop.	Int.	Total	Est sd	$Pr(\geq F)$ Pop.	$Pr(\geq F)$ Rep.
ϕ_j	0.012430	0.9000	0.08755	1	0.05459	7.7e-39	0.2616
ϕ_a	0.004037	0.9487	0.04727	1	0.05655	9.151e-50	0.6454
p_j	0.009847	0.9478	0.04232	1	0.07586	1.343e-51	0.0394
p_a	0.001966	0.9669	0.03116	1	0.05178	4.167e-57	0.8195
π_0	0.010630	0.9193	0.07009	1	0.05993	9.345e-43	0.2179
π_1	0.010630	0.9193	0.07009	1	0.05993	9.345e-43	0.2179
τ	0.008957	0.6236	0.36740	1	0.05838	3.212e-14	0.9909

(b) ANOVA output for the unconditional approach.

Table 7.27: ANOVA output for model parameters for the conditional and unconditional approaches for the multiple simulations analysis with priors neither centered on truth nor fixed.

approaches are doing quite well at matching the expected average totals. Each approach produces average means that are extremely close to one another for both age cohorts and at all time periods. From comparing the standard deviations of the posterior means, $\hat{\mu}_{SD}$, under each approach it seems that the variability of the posterior means for the aggregated states between simulations is greater under the conditional generation approach. However this difference is more pronounced for the states at $t = 0$ and $t = 1$ with the differences being quite small for times $t = 2$ and $t = 3$. However the variability within simulations, as given by the mean of the posterior standard errors $\hat{\sigma}_{\bar{x}}$ is typically greater for the unconditional approach. This is also observable in Table 7.22. These results make intuitive sense. The conditional approach rejects any simulated particle that contains a state element that is inconsistent with the observed capture histories. This restrictive condition means that there is a smaller region of state-space that can be occupied under the conditional approach compared to the unconditional approach. Consequently, the aggregated juveniles and adults are restricted under the conditional approach and this reduced variability is reflected in the smaller values of $\hat{\sigma}_{\bar{x}}$ under the conditional approach.

7.4.2.4 Analysis of Variance for States

The ANOVA output in Table 7.29a corresponds to the two-way analysis on the posterior samples obtained under the conditional approach. Consistent with previous analyses, it can be seen that, under the conditional approach, there is little information in the data, for any simulated population, that allows the accurate determination of the distribution of the initial animals. There is a population effect but no replicate effect on the initial animals; the variation in the posterior distribution for the adults and juveniles at time $t = 0$ arises from the stochastic splitting process applied to each individual initial population. This population based variation is greater than that which can be attributed to repeating

			CGA				UGA			
Time	Total	Exp. Val	$\hat{\mu}_{\bar{x}}$	$\hat{\mu}_{SD}$	$\hat{\sigma}_{\bar{x}}$	$\hat{\sigma}_{SD}$	$\hat{\mu}_{\bar{x}}$	$\hat{\mu}_{SD}$	$\hat{\sigma}_{\bar{x}}$	$\hat{\sigma}_{SD}$
t=0	J	30	40.71	12.63	20.67	6.724	37.06	6.043	20.8	3.262
	A	70	59.29	12.63	20.67	6.724	62.94	6.043	20.8	3.262
t=1	J	30.8	30.22	6.204	5.028	1.417	30.29	5.645	5.974	0.9608
	A	77	80.27	5.275	6.067	2.129	79.82	4.066	6.432	0.9816
t=2	J	33.26	33.14	7.447	5.538	1.928	33.09	6.814	6.44	1.402
	A	83.16	87.72	8.359	8.045	3.083	87.46	8.335	8.36	1.278
t=3	J	35.93	34.82	9.022	6.914	2.398	34.96	8.789	8.008	2.114
	A	89.81	97.92	12.49	11.08	4.077	96.92	13.05	11.48	1.936

Table 7.28: Summaries on the Posteriors of aggregated states for the multiple simulations analysis with priors neither centered on truth nor fixed. J denotes aggregated juveniles, A denotes aggregated adults.

analyses on the same population. The F-Ratios for including population as a factor are significant for both juveniles and adults for all time periods $t = 0 : 3$. As with the earlier simulation study it appears that process error comprises a larger proportion of the total variance exhibited by the aggregated states than Monte Carlo error. Equally, Table 7.29b demonstrates that process error is the main source of variation in the posterior samples of the aggregated states under the unconditional approach. Once again it can be observed that, of the total error, process error owing to fitting different populations forms a greater proportion of the total error under the unconditional approach than under the conditional approach.

7.4.3 Two Age Classes. Variable Priors centred on truth.

A third analysis of multiple simulations was conducted to investigate the issue of prior sensitivity. As with the scenario presented in section 7.4.2, 10 distinct populations will be simulated and each model fitting analysis is repeated 10 times per population for both the conditional and unconditional approaches. Each of these 100 model fittings will be based on initial samples of 10000 particles, with the prior distributions specified using the values in Table 7.30. Thus, for this third analysis $n_{pop} = 10$, $n_{rep} = 10$ and $N_{part} = 10000$. Once again the initial population is assumed to consist of 100 animals at the start of the simulated mark-recapture study.

The simulated draws from these prior distributions are summarised in Table 7.31.

After performing the 100 model fittings average ESS's of 5.958 and 61.488 were obtained under the conditional and unconditional approaches respectively. These values are still smaller than those obtained in section 7.4.1 for model fitting based of priors fixed at the true parameter values. However, there is an appreciable increase in the average

		Scaled Sums of Squares						
		Rep.	Pop.	Int.	Total	Est sd	$Pr(\geq F)$ Pop	$Pr(\geq F)$ Rep.
t=0	J	0.07561	0.27	0.6544	1	12.57	06128	0.4162
	A	0.07561	0.27	0.6544	1	12.57	06128	0.4162
t=1	J	0.01562	0.8118	0.1726	1	6.173	5.392e-27	0.6043
	A	0.0118	0.6605	0.3277	1	5.249	4.283e-16	0.9647
t=2	J	0.01062	0.849	0.1404	1	7.41	1.166e-30	0.7241
	A	0.0136	0.7712	0.2152	1	8.317	3.14e-23	0.8189
t=3	J	0.007072	0.8262	0.1667	1	8.977	9.592e-28	0.9407
	A	0.0111	0.806	0.1829	1	12.42	4.544e-26	0.8364

(a) ANOVA output for the conditional approach.

		Scaled Sums of Squares						
		Rep.	Pop.	Int.	Total	Est sd	$Pr(\geq F)$ Pop	$Pr(\geq F)$ Rep.
t=0	J	0.012530	0.6288	0.35870	1	6.012	1.456e-14	0.9681
	A	0.012530	0.6288	0.35870	1	6.012	1.456e-14	0.9681
t=1	J	0.005181	0.9490	0.04582	1	5.616	2.725e-50	0.4331
	A	0.003611	0.8999	0.09652	1	4.045	2.7e-37	0.9601
t=2	J	0.002684	0.9687	0.02858	1	6.780	1.302e-58	0.5770
	A	0.002789	0.9702	0.02698	1	8.293	1.283e-59	0.5037
t=3	J	0.002267	0.9736	0.02414	1	8.745	1.398e-61	0.5770
	A	0.002750	0.9792	0.01805	1	12.980	1.131e-66	0.2149

(b) ANOVA output for the unconditional approach.

Table 7.29: ANOVA output for conditional and unconditional approaches for the multiple simulations analysis with priors neither centered on truth nor fixed. J denotes aggregated juveniles, A denotes aggregated adults.

	ϕ_j	ϕ_a	p_j	p_a	π_0, π_1	τ
Prior Means: μ_α	0.7	0.8	0.5	0.6	0.6,0.4	0.3
Prior Variances: σ^2_α	1	1	1	1	1,1	1

Table 7.30: Specification of parameters μ_α and σ^2_α used to simulate priors for each parameter where priors are centered on truth but not fixed.

Parameter	**Prior Mean**	**Prior SD**	**True Value**
ϕ_j	0.6702	0.1891	0.7
ϕ_a	0.7597	0.1633	0.8
p_j	0.4970	0.2063	0.5
p_a	0.5819	0.2048	0.6
π_0	0.5893	0.1570	0.6
π_1	0.4107	0.1570	0.4
τ	0.3301	0.1903	0.3

Table 7.31: Summary of prior distributions for each parameter where priors are centered on truth but not fixed.

	CGA				UGA			
Parameter	$\hat{\mu}_{\bar{x}}$	$\hat{\mu}_{SD}$	$\hat{\sigma}_{\bar{x}}$	$\hat{\sigma}_{SD}$	$\hat{\mu}_{\bar{x}}$	$\hat{\mu}_{SD}$	$\hat{\sigma}_{\bar{x}}$	$\hat{\sigma}_{SD}$
ϕ_j	0.7045	0.08796	0.09701	0.02517	0.6988	0.07078	0.1224	0.019930
ϕ_a	0.8069	0.05618	0.06191	0.01675	0.8138	0.05123	0.0696	0.009661
p_j	0.5059	0.06449	0.09093	0.03006	0.5078	0.05201	0.1025	0.015420
p_a	0.6154	0.05867	0.06155	0.01659	0.6053	0.04447	0.0612	0.006867
π_0	0.5903	0.06535	0.08438	0.02539	0.5897	0.05008	0.0926	0.008894
π_1	0.4097	0.06535	0.08438	0.02539	0.4103	0.05008	0.0926	0.008894
τ	0.3274	0.11940	0.18040	0.06379	0.3222	0.03505	0.1871	0.021210

Table 7.32: Summaries on the posterior distribution of the model parameters Θ for the multiple simulations analysis with priors centered on truth but not fixed.

ESS under both approaches when compared to the values in section 7.4.2 based on priors centered away from the true values. Unsurprisingly the average model performance, in the context of the ESS metric, is best when using fixed priors centered on truth and decreases as variability is introduced into the priors and decreases still further when priors are centered away from truth. These simulations also illustrate the increased level of particle depletion experienced under the conditional approach when compared to the unconditional approach. The average ESS of 5.958 under the conditional approach is, once again, quite small and results in Monte Carlo error forming a larger component of the total variation exhibited in the posteriors across the simulations than is observed under the unconditional approach.

7.4.3.1 Posterior Summary

The posterior summaries are obtained under each approach and averaged across all 100 simulations. The resulting distributions are summarised in Table 7.32 in the usual format.

Both model fitting approaches seems to behave well and don't produce posteriors that move significantly away from the true parameter values. Although the movement is minimal in some cases it can be observed that all parameters do move closer to the true parameters that were used to simulate each of the 10 population realisations. By comparing Tables 7.26 and 7.26 it seems that, for this set of simulations, there is less variation in the point estimates as indicated by the $\hat{\mu}_{SD}$ column. That is, when the priors are centered on truth the posterior means seem to vary less across simulations than when the priors are not centered on truth. It can also be seen that, when the priors are centered on truth, there is generally less variation in the precision of the posterior point estimates as indicated by the $\hat{\sigma}_{SD}$. The exceptions to this are the parameters associated with juvenile survival (ϕ_j) and capture (p_j). That is, when the priors are centered on truth the posterior standard deviations for each parameter seem to vary less across simulations than when the priors are not centered on truth. Hence, for the unconditional approach, this suggests that, across simulations, the posterior distributions for the parameters exhibit less variability in both location and precision when the priors are centered on truth than when they are not.

This distinction is not observed for the conditional approach. This may be a result of the low ESS's observed for the both sets of simulations: those based on priors centered on truth and those based on priors not centered on truth. It can be observed that there is greater variability in both the posterior means and posterior standard deviations for all parameters across the 100 simulations under the conditional approach when compared to the unconditional approach.

7.4.3.2 Analysis of Variance for Parameters

The general interpretation of Table 7.33 is similar to that for Table 7.27. For the conditional output an examination of the F-Ratios in Table 7.33a again provides strong evidence that population, but not replicate, should be included as a factor when explaining the variation across simulations in the posterior means for each parameter. A comparison with Table 7.27a suggests that centering the priors on truth reduces the proportion of the variation that is attributed to the interaction term for all parameters apart from the splitting parameter τ. Once again, when compared to the output for the unconditional approach in Table 7.33b, the small ESS results in a greater level of Monte Carlo variation as evidenced by the greater proportion of variation attributed to the interaction term.

Comparing Table 7.33b with Table 7.27b again yields little marked difference between the average performance of the unconditional approach for the simulations based on priors centered on truth and those based on priors not centered on truth. When compared to Table 7.27b it can be seen that the proportion of variation explained by the differences between simulated populations generally increases for the simulations based on priors centered on truth. The exception to this is, again, the population splitting parameter τ. Inspecting the F-Ratios it is seen, for most parameters, that population, but not replicate, explains the variation in posterior means induced by the different simulations. For this set of simulations there is some evidence to suggest that replicate could also be included as a factor to explain the variation observed for the birth rate π_0.

As with the earlier analysis of the simulations based on priors that were not centered on truth it can be seen from this analysis that process error significantly outweighs Monte Carlo error under the unconditional approach. A similar conclusion can be reached for the conditional approach although the differences between the error sources is less pronounced.

	Sum of Squares						
	Rep.	Pop.	Int.	Total	Est sd	$Pr(\geq F)$ Pop.	$Pr(\geq F)$ Rep.
ϕ_j	0.01923	0.7564	0.2243	1	0.08752	2.051e-22	0.6431
ϕ_a	0.01638	0.7567	0.2269	1	0.05590	2.866e-22	0.7513
p_j	0.01501	0.7333	0.2517	1	0.06416	1.622e-20	0.8438
p_a	0.02808	0.7994	0.1725	1	0.05838	8.744e-27	0.1750
π_0	0.02524	0.7127	0.2621	1	0.06502	1.189e-19	0.5581
π_1	0.02524	0.7127	0.2621	1	0.06502	1.189e-19	0.5581
τ	0.02972	0.2173	0.7529	1	0.11880	0.01099	0.9527

(a) ANOVA output for the conditional approach.

	Sum of Squares						
	Rep.	Pop.	Int.	Total	Est sd	$Pr(\geq F)$ Pop.	$Pr(\geq F)$ Rep.
ϕ_j	0.003724	0.9611	0.03519	1	0.07042	6.09e-55	0.48530
ϕ_a	0843	0.9842	0.01492	1	0.05098	4.691e-70	0.86440
p_j	0.012270	0.9426	0.04517	1	0.05175	2.057e-50	0.01628
p_a	0.001565	0.9862	0.01219	1	0.04424	1.371e-73	0.33500
π_0	0.008739	0.9657	0.02551	1	0.04983	1.7e-60	0.00314
π_1	0.008739	0.9657	0.02551	1	0.04983	1.7e-60	0.00314
τ	0.018490	0.5983	0.38320	1	0.03488	2.348e-13	0.91280

(b) ANOVA output for the unconditional approach.

Table 7.33: ANOVA output for model parameters for the conditional and unconditional approaches for the multiple simulations analysis with priors centered on truth but not fixed.

			CGA				UGA			
Time	Total	Exp. Val	$\hat{\mu}_{\bar{x}}$	$\hat{\mu}_{SD}$	$\hat{\sigma}_{\bar{x}}$	$\hat{\sigma}_{SD}$	$\hat{\mu}_{\bar{x}}$	$\hat{\mu}_{SD}$	$\hat{\sigma}_{\bar{x}}$	$\hat{\sigma}_{SD}$
t=0	J	30	33.23	11.72	18.22	6.171	32.34	3.656	19.05	2.103
	A	70	66.77	11.72	18.22	6.171	67.66	3.656	19.05	2.103
t=1	J	30.8	29.59	3.818	5.29	1.586	30.87	3.783	6.071	0.6538
	A	77	76.93	3.709	6.186	1.484	77.5	2.827	6.585	0.6456
t=2	J	33.26	34.55	6.724	5.672	1.764	34.42	4.881	6.645	0.7832
	A	83.16	83.18	5.833	7.311	2.001	84.64	5.258	8.248	0.7262
t=3	J	35.93	38.45	7.156	6.991	2.265	38.37	6.127	8.038	1.051
	A	89.81	92.88	10.03	9.77	2.93	94.07	8.929	11.23	1.293

Table 7.34: Summaries on the Posteriors of aggregated states for the multiple simulations analysis with priors centered on truth but not fixed. J denotes aggregated juveniles, A denotes aggregated adults. CGA denotes summaries obtained under the conditional generation approach. UGA denotes summaries obtained under the unconditional approach.

7.4.3.3 State Summary

The summary presented in Table 7.34 leads to similar conclusions to those based on the previous simulation study as summarised in Table 7.28. Both modelling approaches seem to be doing a very good job, on average, of accurately matching the expected aggregated state totals representing juveniles and adults at each time step. Inspecting the standard deviations of the posterior means, $\hat{\mu}_{SD}$, again indicates that the variability of the posterior means for the aggregated states is greater under the conditional approach. This difference is most pronounced for the posteriors on the initial states; the differences are relatively small for the other time periods. A similar pattern is observed for the standard deviations of the posterior standard deviations, $\hat{\sigma}_{SD}$, in that the variability of the posterior standard

deviations is greater under the conditional approach. This mirrors what was observed for posterior summaries for the model parameters.

Comparing Table 7.28 and Table 7.34 it can be seen that the standard deviations of the posterior means, $\hat{\mu}_{SD}$, are smaller for the simulations based on priors centered on truth than for those based on priors that are not centered on truth. This holds for both the unconditional and conditional approaches. Similarly, the standard deviations of the posterior standard deviations, $\hat{\sigma}_{SD}$, exhibit a similar pattern of behaviour. When the priors are centered on truth there is less variation in the precision of the posterior point estimated for the aggregated states at each time step. This holds for the unconditional approach and holds for the conditional approach over all aggregated states apart from the total number of juveniles at time $t = 1$. Hence, for both the modelling approaches, when compared to the simulations based on priors not centered on truth, it seems that, across simulations, the posterior distributions for the aggregated states exhibit less variability in both location and precision when the priors are centered on truth.

7.4.3.4 Analysis of Variance for States

The summary information presented in Table 7.35a is consistent with that presented in Table 7.29a for the previous simulation study. As before, there is evidence for a population effect but no replicate effect on the distribution of posterior means for the aggregated states at each time step. The variation that occurs in these posterior means is primarily explained by the differences due to the 10 simulated populations rather than the repeated analyses of those populations. However, the variation in the posterior means for the initial animals is mostly explained by the differences between the 100 individual analyses which again supports the idea that there is little information in the data to estimate the

splitting parameter τ. Comparing Tables 7.29a and 7.35a does not produce a clear effect on the composition of the total variation exhibited by the posterior means. For example, centering the priors on truth reduces the proportion of the variation that is attributed to the different populations for the initial animals. However, the effect differs for animals at the second and third time periods. For both of these time periods, centering the priors on truth results in more of the variation in aggregated juveniles being explained by the population term. However, for the posterior means of aggregated adults, centering the priors on truth results in less of the total variability being explained by the population term.

A comparison with Table 7.27a suggests that centering the priors on truth reduces the proportion of the variation that is attributed to the interaction term for all parameters apart from the splitting parameter τ. Once again, when compared to the output for the unconditional approach in Table 7.33b, the small ESS results in a greater level of Monte Carlo variation as evidenced by the greater proportion of variation attributed to the interaction term.

A comparison between Table 7.35b and Table 7.29b results in similar conclusions to those made when comparing the analysis of variance on the posterior means of the parameters. As with that analysis it can be seen that, for simulations based on priors that are centered on truth, the proportion of the total variation that is attributed to the difference between populations increases for all but the initial aggregated states. It can be seen in Table 7.35b that there is some weak evidence to suggest that replicate could be included as a factor to explain the variation observed in the posterior means for the aggregated juveniles at times $t = 1$,$t = 2$ and $t = 3$. The F-ratios for all aggregated states suggest that population should be included as a factor.

		Scaled Sums of Squares						
		Rep.	Pop.	Int.	Total	Est sd	$Pr(\geq F)$ Pop	$Pr(\geq F)$ Rep.
t=0	J	0.02591	0.2259	0.7482	1	11.66	0.008065	0.9691
	A	0.02591	0.2259	0.7482	1	11.66	0.008065	0.9691
t=1	J	0.01765	0.6727	0.3096	1	3.799	5.969e-17	0.8611
	A	0.05943	0.5246	0.4160	1	3.690	2.722e-11	0.2575
t=2	J	0.008116	0.8754	0.1165	1	6.690	6.11e-34	0.7708
	A	0.04871	0.6754	0.2759	1	5.804	2.313e-18	0.1325
t=3	J	0.02030	0.8382	0.1415	1	7.121	2.416e-30	0.2547
	A	0.01818	0.7983	0.1835	1	9.980	6.827e-26	0.5362

(a) ANOVA output for the conditional approach.

		Scaled Sums of Squares						
		Rep.	Pop.	Int.	Total	Est sd	$Pr(\geq F)$ Pop	$Pr(\geq F)$ Rep.
t=0	J	0.01329	0.6119	0.37480	1	3.638	8.223e-14	0.9665
	A	0.01329	0.6119	0.37480	1	3.638	8.223e-14	0.9665
t=1	J	0.006849	0.9670	0.02610	1	3.764	3.977e-60	0.0201
	A	0.006855	0.9569	0.03629	1	2.813	2.398e-54	0.1024
t=2	J	0.005448	0.9803	0.01422	1	4.856	8.042e-71	0.0012
	A	0.003623	0.9775	0.01885	1	5.231	6.795e-66	0.0955
t=3	J	0.003704	0.9827	0.01360	1	6.096	1.243e-71	0.0160
	A	06222	0.9855	0.01389	1	8.884	2.583e-71	0.9300

(b) ANOVA output for the unconditional approach.

Table 7.35: ANOVA output for conditional and unconditional approaches for the multiple simulations analysis with priors centered on truth but not fixed. J denotes aggregated juveniles, A denotes aggregated adults.

When compared with the analysis based on simulations that used priors not centered on the true parameter values the same general conclusions can be reached. Process error still significantly outweighs Monte Carlo error under both the conditional and unconditional approaches for all aggregated states bar those for the initial animals. The effect is less pronounced under the conditional approach but it still apparent. Using priors centered on truth does increase the disparity between the relative contributions of these two error sources to the total variability. However this increase is not particularly large and does not significantly alter the interpretation of the efficacy of the modelling approach.

7.4.4 General Summary

In general, there does appear to be some degree of prior sensitivity in the model with regard to the proximity of the posterior means to the true values for both the model parameters and the aggregated states of the expected population. The extremely small average ESS exhibited by the model fitting under the conditional approach requires these results to be interpreted with a degree of caution. Further larger-scale comparisons could be conducted using more initial particles to address the issue of particle depletion.

Both modelling approaches seem to perform well on average. There is an appreciable improvement in the accuracy of the posterior means for both parameters and states when the priors are centered on truth. These simulations yielded larger effective sample sizes than the simulations based on priors that were not centered on truth and they exhibited less variability in both the location and precision of the posterior distributions. However these improvements were mostly fairly minor and the model fitting algorithms performed well when using priors that were not centered on truth.

7.5 Discussion

The simulation study has allowed the performance of these fitting algorithms to be investigated in detail. It can be seen that, for this particular population structure, simulations based on fixed priors set at the true values return posterior densities that have their probability mass centered close to the true values of the states. Thus, the fitting algorithms replicate truth quite successfully. Even when variability is introduced by simulating the model parameters and initial states from prior distributions it was seen that the models still performed reasonably well. However, it was noted that both approaches can still be affected by particle depletion caused by the simulation of implausible states. As expected, the effect is stronger for the conditional approach than for the unconditional approach. It also seems that the model fitting algorithms are reasonably consistent and differences between the posterior samples based on different analyses are more pronounced when different populations are analysed rather than when the same population is analysed repeatedly.

These exploratory analyses, including that on the Soay sheep data, provide numerous avenues for further research. The nature of the model fitting implemented under the conditional generation approach automatically raises questions about suitable goodness-of-fit diagnostics. The complexity of the trial pdf that was constructed for the relatively simple population used in these simulation studies reduces the viability of specifying a general and flexible approach to applying the conditional generation techniques. The current approach is extremely model specific: both the traditional state and observation processes are required to be known or at least well approximated to develop an effective conditional generation algorithm. The unconditional approach is the more flexible of the two and does typically return a larger effective sample size when both models are applied to the same

population. However, the current implementation of the approach does not require the states to be consistent with the observed capture history patterns and can consequently suffer from particle depletion. One approach to investigating whether or not the fitting algorithm code works as intended has been developed by Cook *et al.* (2006). This involves a simulation-based method using the uniformity of posterior quantiles to establish the efficacy of the code used to run the model fitting algorithms. The approach is embedded within a Bayesian framework and illustrated with an application to an MCMC fitting routine. The methods should be adaptable for any iterative Bayesian fitting process and would be worth examining in the context of the algorithms developed under the conditional and unconditional approaches.

Chapter 8

General Discussion and Future Directions

This thesis described the formulation of approaches to modelling that allowed population dynamics to be embedded into inference on mark-recapture data. Fitting algorithms were developed under two alternative approaches: (1) The "conditional" approach in which inference was conditional on the numbers of animals known to be captured at each time point thus leading to the parameters relating to the capture process to be incorporated into the state process equation and (2) The "unconditional" approach in which the capture process is included in the observation process equation and there is no conditioning on the known abundances of captured animals. From the simulation analyses presented in chapter 7 it could be seen that the fitting algorithms under both the conditional and unconditional approaches produced posterior distributions on the states and parameters that were, on average, consistent with the true values.

The analysis on the Soay sheep dataset, also presented in chapter 7, yielded disappointing results. The effective sample sizes generated by both fitting algorithms were extremely small; with the effect being more pronounced under the conditional approach when compared with the unconditional approach. Due to the posterior distribution on

the states and parameters being dominated by a single particle under the conditional approach there is a danger that the results can be over-interpreted. It is difficult to draw meaningful conclusions regarding the performance of the model-fitting algorithm under the conditional approach. The situation is not greatly improved for the analysis based on fitting the mark-recapture data on the Soay sheep population using the model-fitting algorithm proposed under the unconditional approach. This immediately suggests further avenues for research.

Further investigation is needed into the efficiency of the fitting algorithms developed under both approaches with greater emphasis on those developed under the conditional approach. The nature of the requisite conditioning that defines the conditional approach results in a fitting algorithm that typically produces posterior distributions of very high dimension. All parameters that are fixed at a particular time interval and all elements of the state and intermediate vectors for each time period are included in the posterior distribution. The constraints imposed by conditioning on those animals that were known to be caught at each time period results in posterior distributions that are restricted to regions of viable state and parameter space. In this context viable state space is defined as the set of state elements that are consistent with the observed abundances of captured animals over the duration of the monitoring study. Accounting for these restrictions in the fitting algorithm seems to result in smaller effective sample sizes being obtained when compared with the unconditional approach.

Examining the output from the analyses in chapter 7 it is clear that, when compared to the unconditional approach, fitting models under the conditional approach produces weights that are more variable about their mean value. Under the importance sampling

fitting procedures it can be seen that the conditional approach produces a smaller proportion of implausible particles. However the distribution of the weights associated with these plausible particles is more skewed under the conditional approach than for the equivalent distribution under the unconditional approach. Thus, it seems that although the conditional generation of states results in fewer implausible particles being obtained, the unconditional generation results in a more even distribution of the weights over the surviving particles. This is, most likely, a reflection of the different assumptions under each approach to constructing suitable models, specifically the manner in which the observed data are incorporated into the analysis. Under the unconditional approach the observed data enters the model framework through the capture history patterns that have been recorded in the final time period. This allows a greater level of flexibility in the possible paths taken by the animals when exhibiting a particular capture history pattern. That is, the unconditional approach requires the evaluation of the probabilities of an individual animal exhibiting each of the possible capture histories recorded in the final time period.

For a study consisting of T capture occasions, consider a single capture history pattern. If there are multiple paths that would result in an animal displaying this capture history pattern and each of these paths has approximately the same probability of occurrence then each path would provide an approximately equal contribution to the weight for that particular particle. Under the conditional approach the possible paths that an animal could have travelled to exhibit a particular capture history pattern are typically restricted to a smaller range. Under the conditional approach the data enters the model framework through the known captures at each time period. This results in a far more prescribed path for the animals simulated under the fitting methods for the conditional approach. This restriction on the plausible life histories displayed by the simulated animals results in weights that are more skewed than those obtained under a less restrictive

model fitting approach. In effect, the particles simulated under the conditional approach need to satisfy a much larger set of criteria than do those simulated under the unconditional approach. Therefore, simulating particles subject to these criteria increases the number of particles that are consistent with the observed data. However, the restrictions imposed on the viable parameter and state space by these criteria results in fewer particles that meet all criteria well and, consequently, fewer particles that are assigned high weights.

This is reflected in the analysis on the multiple simulation study presented in section 7.4. It can be seen that the smaller effective sample size for the model fitting conducted under the conditional approach, as opposed to the unconditional approach, results in smaller average weighted standard deviations for the posteriors of the model parameters and aggregated states.

A further issue arises with regard to the efficiency of the model fitting algorithms developed under the conditional approach. The specification of the trial density presented in chapter 5 means that the duration of the fitting algorithm is a function of the population size. Increasing the number of animals that are either known or simulated to exist in the population increases the time it takes for the model fitting algorithm to run. This can be understood by considering section 5.2.2. In this section the observed number of adults that correspond to convolved totals are required to be split across the appropriate trees. The method for performing this split involved obtaining a non-uniform finite discrete distribution on the range of possible splits. The probability associated with each of these possible values then needed to be calculated to use the alias method with table look-up as a means of drawing a particular split. Hence, assuming the model parameters remain the same, increasing the known or simulated abundance of the population at each time period will result in a larger range of plausible splits for a convolved total. A larger range

of splits engenders a larger range of associated probabilities and this increased computation burden increases the time taken for the fitting algorithm to be performed. A similar situation occurs when applying the conditional generation approach (see section 4.4.3) to simulate survivors and births for certain trees.

One potential future avenue for research is to develop a more efficient trial density that doesn't require all possible probabilities to be calculated when the fitting algorithm needs to draw a value from a non-uniform finite discrete distribution on the range of possible values. Potentially a flexible probability function could be specified that would provide an adequate approximation to the discrete-valued probability mass function for this distribution. This in turn raises questions regarding the desired level of "adequacy" of any approximation and the value of the trade-off between the increased efficiency of the algorithm and the loss of accuracy. It should also be noted that the fitting algorithm under the conditional approach is of greater utility when dealing with smaller populations. It is more important to ensure that all state and intermediate nodes are consistent with the known number of animals alive at each time period if those known abundances are small. As a rather trivial example, if the abundance of a simulated state element is two less than the true value this is less of an issue if truth is 1000 than if truth is 2. The probability mass functions that arise due to the form of the trial density specified in the fitting algorithm constructed under the conditional approach need to be identified. Then, if these functions can be approximated adequately this may increase the efficiency of the algorithm with no crucial loss of precision. As previously mentioned, this trade-off needs to be investigated in more detail.

The fitting algorithm specified under the unconditional approach experiences a similar functional dependency on simulated population size but for different reasons. For

this algorithm the dependency arises from the need to evaluate the probabilities of each individual animal exhibiting each of the possible capture history patterns. This effect is somewhat mitigated by grouping the animals into cohorts defined by their times of entry and exit into the population (see section 6.6.1). Performing this classification means that only the numbers of animals belonging to those cohorts need to be calculated with each cohort member exhibiting the same probabilities. This approach can also accommodate a model with time-specific survival and capture rates. Models that incorporate animal level covariates (e.g. horn type in Soay sheep) would require an alternative classification of cohorts to be specified to group animals by common life-history parameters. The original, less efficient, evaluation of particle weights, by calculating the probabilities of each individual animal exhibiting each capture history pattern, easily accommodates models with animal level covariates.

The fitting algorithms developed under each of the approaches can be inefficient with regard to both the level of particle depletion they exhibit and the resulting skewed distribution of the weights. The Soay sheep analyses in chapter 7, was conducted by running each fitting algorithm simultaneously, with an initial 250000 particles for each. This took approximately 16 hours using parallel processing across three processors with CPU speeds ranging between 1 and 2.2GHz. As mentioned previously, these analyses yielded effective sample sizes that were extremely low under the conditional approach but adequate for inference under the unconditional approach . Attempts to increase the number of particles resulted in memory allocation errors when using the software *R* (see http://cran.r-project.org/). Further use of parallel processing can increase the speed of the fitting algorithms. However, to avoid memory allocation errors, this would necessitate a more advanced routine to be written for harvesting the particles with non-zero weights. Equally, during the development of these fitting algorithms it was often the case that excessive

particle depletion was experienced when fitting to more than three years of capture history data.

Further research is needed to investigate alternative fitting algorithms. The methods discussed in chapter 4 presents the general theory behind the conditional approach and chapter 5 describes one possible methodology for developing a fitting algorithm under the assumptions of the conditional approach. The construction of the trial density was key to the fitting algorithm, and it was seen (section 5.6) that this particular form of the trial density does not generate particles according to the full conditional distributions as specified under the conditional approach. Alternative trial densities could be investigated in which the conditioning was extended to a larger window of observed capture history patterns. Under the current "bottom-up" approach, the densities associated with the simulated states and intermediate vectors at t are conditional on the observed abundances at time $t-1$. If the parent node at time $t-1$ is unknown then, depending on the status of the appropriate ancestral node, the fitting algorithm conditions on observed abundances at time $t-2$. When simulating the number of surviving animals and the number of new-born juveniles these surviving animals produced, it seems reasonable to restrict conditioning to those observed values that provide natural limits, where available. That is, the observed number of animals at time $t-1$, the number of those animals that were encountered at time t and the number of juveniles produced by the survivors that were encountered at time t. This component of the trial density is described in detail in section 4.4.3. However, if the ancestral nodes are not observed, an alternative model fitting algorithm could be considered in which the most closely related element corresponding to a known abundance is used as a constraint. A state element that corresponds to a known abundance but is only loosely connected to the element being simulated will not provide a great deal of information with which to constrain the range of plausible values for that element. As

before, the computational burden of incorporating this constraint on the fitting process should be weighed against the gain in accuracy resulting from restricting the viable state space.

Developing a fully conditional fitting algorithm over multiple time periods of data is extremely complicated for the high-dimensional state-space models that result from the model structure specified under the conditional approach. The fully conditional approach would require a large number of extra constraints to be specified. For example, any simulated value for a state element corresponding to animals that were not observed at time t would need to meet constraints imposed by each and every related known abundance. The simulated element would need to be large enough to produce adequate numbers of descendants that were observed for all future years $t+1, t+2, \ldots, T$ and small enough not to be implausible given the known numbers of ancestors in previous time periods $t-1, t-2, \ldots, 1$. Specifying all such constraints is a non-trivial process and incorporating them all into a fitting algorithm will significantly increase its complexity. As the interrelation between state and intermediate elements decreases, the amount of information provided by one element about the other is also likely to decrease. That is, the greater the number of sub-process realisations that need to occur to model the life history path between two elements, the less information they provide about the plausible range of values each can take. In other words, the more closely related two elements are, the more information they are likely to provide about one another. With this in mind it is worth considering what a suitable level of conditioning would be. Increasing the complexity of a model fitting algorithm to make it fully conditional over four time periods may not add a significant amount of information to that which is obtainable from a model fitting algorithm that is fully conditional over only three time periods. Further investigation is

required into developing alternative trial densities under the assumptions of the conditional approach that respect an increased level of conditioning. The trade-off between decreased particle depletion and increased computational burden is worth examining.

An alternative avenue of research is to consider incorporating Pollock's robust design (Pollock *et al.*, 1990; Kendall and Nichols, 1995; Williams *et al.*, 2002, pp.523-544) into the algorithms under each approach. Nichol's and Kendall's refinements and developments of the robust design aim to produce estimates of capture rates that are more robust to heterogeneity than the equivalent estimates obtained under an open population model. This results in estimates of abundance that are also less biased by heterogeneity in capture rates. The robust design uses both the long-term open population models that are fitted to data from 'primary' sampling occasions as well as the short-term closed population models fitted to short-term 'secondary' sampling occasions. It is assumed that these repeated secondary sampling occasions occur over a time period in which the population can be assumed to be closed. By considering both primary and secondary sampling occasions, it is possible to identify more model parameters and to test certain modelling assumptions; for example, the initial capture rate can be distinguished from the recapture rate under this approach. The capture histories obtained from the secondary sampling occasions also provide extra information about the model parameters and results in improved precision of parameter estimates. By improving the inference on the population estimates at each primary sampling occasion the robust design may help to improve the performance of the conditional and unconditional fitting algorithms. Heterogeneity in capture probabilities may be negatively impacting the model fit and a robust design approach would help to investigate this issue. By obtaining more precise estimates of a greater number if identifiable parameters incorporating a robust approach can help to identify limitations in the conditional and unconditional algorithms.

The performance of the model fitting algorithms can be investigated by examining the likelihood surface. Relatively simple procedures such as obtaining the MLEs of model parameters can potentially be used to tune the performance of the algorithms. Comparing the MLEs to posteriors modes also allows algorithm performance to be monitored; if the modes are far from the MLEs then this would suggest the algorithm is not functioning as expected. Optimisation routines such as simulated annealing can be used to approximate the likelihood surface and obtain inference on marginal distributions of parameters of interest. These approaches can provide information on the efficacy of the algorithms: are they producing the 'right' answers? Allied to these issues is the consideration of the covariance structure of the model. The prior structure for the Bayesian implementation of the conditional and unconditional approaches is predicated on the assumption of independent parameters. Although expedient in terms of computational complexity this may not be a realistic assumption. Consideration should be given to specifying a more complex prior structure on the parameters that incorporates covariance terms. For example, a multivariate normal distribution could be considered and the variance-covariance matrix given by the Hessian matrix obtained from maximum likelihood estimation of the parameters. By obtaining an improved approximation to the underlying covariance structure in the models, the fitting process may be greatly improved. A more careful selection of proposal distributions to reflect these issues would be worth investigation.

One of the overarching aims of this thesis was to develop a general methodology for constructing state-space models that allows population dynamics to be embedded into a mark-recapture analysis. This outline of this approach borrows heavily from that which is described by Buckland *et al.* (2007). They advocate a flexible 'building-block' approach to formulating models that can incorporate population dynamics and are fitted

to a discrete time series of data. The details of this and other, similar, approaches are given in chapter 3. Within the context of model fitting algorithms developed under the conditional approach, the form taken by the algorithm is too dependent on the choice of biological sub-processes, the order in which they occur and the parameterisation of those processes to allow for highly automated fitting of the models. Steps 5 to 11 of the simplified pseudo-algorithm in section 7.2.2 outline a general model fitting approach but the details in step 6 will be determined by the specified structure of the model. The structure of the state vector and the information contained in the mark-recapture data will determine the conditioning that is contained in the model fitting algorithm. As discussed in Buckland *et al.* (2007), environmental stochasticity can be incorporated into the model by allowing parameters to vary across time periods. Equally, the probabilities for assumed biological processes could be modelled as functions of covariates; Buckland *et al.* (2004) use logistic functions for this purpose. Buckland *et al.* (2007) also note that individual level covariates can be incorporated into the analysis by modelling individual states in the population. The choice of parameterisation can then determine the structure of the state and intermediate vectors for the state-space models constructed under the conditional approach. The relationship between the observed data and the state vector then determines the level of convolution present in the model. In the relatively simple state-space models formulated in chapter 4, the elements of the state vector were classified by age-cohort (juveniles and adults including yearlings) and capture history pattern. An alternative state-space model could be specified in which the elements of the state vector are classified by individual level covariates as well as age-cohort and capture history pattern. Although both of these models could be fit to the same observed data, the fitting algorithm for the alternative model will incorporate different constraints from those under the more simple model. The sensitivity of the model fitting algorithm to the structure of the state vectors and the manner in which parameters are assumed to vary (temporally

or as a function of covariates) reduces the extent to which the fitting algorithm can be automated.

Further research needs to be conducted to examine the extent to which the model fitting algorithms can be automated under the conditional approach. The absence of conditioning in the fitting algorithms constructed under the unconditional approach makes automating the algorithm more viable. The "full" unconditional method involves calculating the probabilities for each of the possible capture history patterns for each animal that has been part of the population over the duration of the study. This method will allow any model to be incorporated into the model fitting algorithm as evaluation is based on individual animals. Grouping a nimals with common life-history parameters allows for a more efficient implementation of this model-fitting algorithm. The question then becomes, is it possible to automate this grouping over common parameters? The grouping will naturally depend on the choice of parameterisation specified for the state-space model. If individual-level covariates are included in the model then the state structure can be expanded to include a single animal in each state element; this would then be equivalent to the implementation of the full method. Environmental stochasticity modelled by temporal variation in parameters can be incorporated by grouping animals by their times of entry and exit. Similarly, grouping by discrete covariate value will allow the model fitting algorithm to incorporate different forms of environmental stochasticity. Automating model fitting algorithms developed under the assumptions of the conditional approach is complex due to the issue of conditioning on known data. The absence of this requirement under the unconditional approach allows model fitting algorithms to be developed that can be automated once the choice of parameterisation is specified.

The conditional and unconditional approaches are both embedded with a Bayesian

inferential framework. This allows more complex models to be formulated and fitted and allows various sources of uncertainty to be accounted for. Typically, existing fitting methods for population dynamics (e.g. matrix models (Caswell, 2001)) do not intrinsically account for uncertainty in either the model parameters or the model structure. Statistical models can account for various sources of uncertainty but are often formulated from an empirical standpoint with no explicit embedded population dynamics component to the model structure. State-space modelling within a Bayesian framework provides an integrated approach to the formulation and fitting of models. An importance sampling approach used to implement the model fitting algorithms under both approaches was chosen due to the success of SIS methods in fitting state-space, or hidden-process, models with embedded animal population dynamics (Buckland *et al.*, 2004; Thomas *et al.*, 2005; Newman *et al.*, 2006; Buckland *et al.*, 2007). The results from the Soay analysis indicate there are problems with the implementation of the importance sampling approach for the fitting algorithms constructed under the conditional and unconditional approaches. Clearly, further work needs to done on investigating both the choice of an appropriate trial density, or importance distribution in the nomenclature of Newman *et al.* (2009), and the choice of Monte Carlo fitting procedure. The relative performance of SIS and MCMC approaches to making inferences, within a Bayesian context, on the states and parameters of state-space models for animal population dynamics is discussed in Buckland *et al.* (2007) with a detailed treatment given in Newman *et al.* (2009).

The approach used by Newman *et al.* (2009) to determine which of the Monte Carlo approaches is better involved assessment of relative performance for three criteria: ease of implementation, computational efficiency and accuracy. Denoting a state-space model with the acronym SSM, they concluded that "for both SIS and MCMC, ease, efficiency, and accuracy are a function of both the SSM formulation and the available data.". The

recommendations given in Newman *et al.* (2009) are worth investigating to see if an appreciable improvement in model performance can be obtained for the models developed in this thesis. From their investigations they conclude that implementing an MCMC approach that has been designed specifically for the particular state-space model and where the recorded data are informative relative to the priors is the "gold-standard" in terms of satisfying the three criteria. It provides the best trade-off between the complexity of the fitting algorithm and the resulting computational efficiency.

Careful selection of an appropriate expanded state vector (see section 4.3.1) to include latent states increases the tractability of the calculations required for the weights used in an importance sampling based fitting approach. Highly complex state-space models that include multiple intermediate subprocesses can be more complicated to fit under MCMC based approaches than under SIS approaches. However, Newman *et al.* (2009) note that, as for SIS approaches, a careful choice of expanded vector increases the tractability of the state process pdf such that the acceptance probabilities in a Metropolis-Hastings algorithm can be more easily obtained, thus allowing MCMC approaches to be implemented. Newman *et al.* (2009) suggest that a carefully specified MCMC sampler often exhibits a greater degree of efficiency when compared to SIS. However the performance can be similar if the data provide little information relative to the priors. One possible future direction of research could be to investigate the performance of MCMC based fitting algorithms under both the conditional and unconditional approaches. The requirement to condition on the observed data may make the specification of an efficient proposal distribution extremely difficult. Existing software such as OpenBUGS (MRC Biostatistics Unit, Cambridge, UK) can be used to fit state-space models to data using MCMC (Rivot *et al.*, 2004). However the complexity of the model form obtained under the conditional approach is likely to require model-specific software to be written. The independence of

the model parameters is also worthy of investigation as Newman *et al.* (2009) note that the MCMC based fitting algorithms can experience poor rates of convergence if there are high correlations between parameters and states. Alternative proposal distributions can be investigated to address this issue.

Further investigation is also required into the use of model diagnostics to assess the performance of the algorithm and the quality of fit produced by the model fitting algorithm with reference to its utility for model selection. Classical goodness-of-fit tests for capture-recapture experiments are described in Pollock *et al.* (1990). As discussed in section 2.2.1.4, the classical methods use information-theoretic approaches to model selection such as AIC. A weighted average over multiple models, with weights proportional to model AICs, can accommodate model uncertainty into the parameter and state estimates produced by the models. Further research into Bayesian analogues of the classical goodness-of-fit tests and model selection approaches is required. Bayesian p-values (Gelman *et al.*, 1996) are regarded as a useful method for assessing goodness-of-fit. Bayesian model averaging is illustrated with application to survival models on mark recovery and recapture data in Brooks *et al.* (2000). It can also be incorporated into SIS fitting algorithms (Buckland *et al.*, 2004) in which priors are specified to reflect the relative plausibility of each member of a set of defined models. For MCMC based fitting algorithms RJMCMC techniques can be used to move between potentially viable alternative models (King and Brooks, 2002a,b). Applying these model selection techniques to the types of models developed in this thesis allows further analysis of model uncertainty under both approaches.

Section 6.2 compares the superpopulation and population approaches to formulating models under the conditional and unconditional approach. The conditional approach is

restricted to assume a population model whereas the unconditional approach allows for both population and superpopulation models to be specified. The models developed in chapter 6 assumed a population model. Further research is needed into the development of superpopulation models under the unconditional approach.

In the course of this thesis theoretical models, under two alternative approaches, have been successfully developed that incorporate an embedded population dynamics component. The general behaviour of the models when fitted to simulated data was encouraging. Limitations of the modelling approaches were exposed by their application to a real-life dataset. Further work needs to be done to increase the utility of these modelling approaches when applied to real, often complex, animal populations that have been monitored as a mark-recapture study.

Bibliography

Akaike, H., "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19, (1974), 716–723.

Barker, R.J. and White, G.W., "Towards the mother-of-all-models: customised construction of the markrecapture likelihood function," *Animal Biodiversity and Conservation*, 27, 1, (2004), 177–185.

Besbeas, P., Freeman, S.N., and Morgan, B.J.T., "The Potential of Integrated Population Modelling," *Australian and New Zealand Journal of Statistics*, 47, (2005), 35–48.

Besbeas, P., Freeman, S.N., Morgan, B.J.T., and Catchpole, E.A., "Integrating Mark-Recapture-Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters," *Biometrics*, 58, (2002), 540–547.

Besbeas, P., Lebreton, J.D., and Morgan, B.J.T., "The efficient integration of abundance and demographic data," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52, (2003), 95–102.

Bonner, S.J. and Schwarz, C.J., "An Extension of the Cormack-Jolly-Seber Model for Continous Covariates with Application to *Microtus pennsylvanicus*." *Ecology*, 62, 1, (2006), 142–149.

Borchers, D.L., Buckland, S.T., and Zucchini, W., *Estimating Animal Abundance - Closed Populations*, Springer-Verlag, London, UK, 2002, chapter 6, pp. 104–130.

Brooks, S.P., Catchpole, E.A., and Morgan, B.J.T., "Bayesian animal survival estimation," *Statistical Science*, 15, 4, (2000), 357–376.

Brownie, C., Anderson, D.R., Burnham, K.P., and Robson, D.S., *Statistical Inference from Band Recovery Data - A Handbook*, 156, U.S. Fish and Wildlife Service Resource Publication, 1985, 2nd edition.

Brownie, C., Hines, J.E., Nichols, J.D., Pollock, K.H., and Hestbeck, J.B., "Capture-Recapture Studies for Multiple Strata including Non-Markovian Transitions." *Biometrics*, 49, 4, (1993), 1173–1187.

Buckland, S.T., "A Modified Analysis of the Jolly-Seber Capture-Recapture Model," *Biometrics*, 36, 3, (1980), 419–435.

Buckland, S.T., "A Mark-Recapture Survival Analysis," *The Journal of Animal Ecology*, 51, 3, (1982), 833–847.

Buckland, S.T., Burnham, K.P., and Augustin, N.H., "Model Selection: An Integral Part of Inference," *Biometrics*, 53, 2, (1997), 603–618.

Buckland, S.T. and Garthwaite, P.H., "Quantifying the Precision of Mark-Recapture Estimates Using the Bootstrap and Related Methods," *Biometrics*, 47, 1, (1991), 255–268.

Buckland, S.T., Newman, K.B., Fernández, C., Thomas, L., and Harwood, J., "Embedding Population Dynamics Models in Inference," *Statistical Science*, 22, 1, (2007), 44–58.

Buckland, S.T., Newman, K.B., Thomas, L., and Koesters, N.B., "State-space models for the dynamics of wild animal populations," *Ecological Modelling*, 171, (2004), 157–175.

Burnham, K.P. and Anderson, D.R., *Model selection and multimodel inference: a practical information-theoretic approach*, Springer, New York, 2002, second edition.

Burnham, K.P., Anderson, D.R., White, G.C., Brownie, C., and Pollock, K.H., "Design and Analysis Methods for Fish Survival Experiments Based on Release-Recapture," *American Fisheries Society Monographs*, 5, (1987), 1–437.

Carlin, B.P., Polson, N.G., and Stoffer, D.S., "A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling," *Journal of the American Statistical Association*, 87, 418, (1992), 493–500.

Carpenter, J., Clifford, P., and Fearnhead, P., "Improved particle filter for nonlinear problems," *Radar, Sonar and Navigation, IEE Proceedings*, 146, 1, (1999), 2–7.

Caswell, H., *Matrix Population Models: Construction, Analysis, and Interpretation*, Sinauer Associates, Inc., Sunderland, Massachusetts, 2001, 2nd edition.

Catchpole, E.A. and Morgan, B.J.T., "Model Selection in Ring-Recovery Models using Score-Tests," *Biometrics*, 52, 2, (1996), 664–672.

Catchpole, E.A., Morgan, B.J.T., Coulson, T.N., Freeman, S.N., and Albon, S.D., "Factors Influencing Soay Sheep Survival," *Applied Statistics*, 49, 4, (2000), 453–472.

Chao, A., Chu, W., and Hsu, C.H., "Capture-Recapture When Time and Behavioral Response Affect Capture Probabilities," *Biometrics*, 56, 2, (2000), 427–433.

Clark, J.S., "Uncertainty and variability in demography and population growth rates: a hierarchical approach," *Ecology*, 84, (2003), 1370–1381.

Clark, J.S., Ferraz, G., Oguge, N., Hays, H., and DiCostanzo, J., "Hierarchical Bayes for Structured Variable Populations: From Recapture Data to Life-History Prediction," *Ecology*, 86, 8, (2005), 2232–2244.

Cook, S.R., Gelman, A., and Rubin, D.B., "Validation of Software for Bayesian Models Using Posterior Quantiles," *Journal of Computational & Graphical Statistics*, 15, 3, (2006), 675–692.

Cormack, R.M., "Estimates of Survival from the Sighting of Marked Animals," *Biometrika*, 51, 3/4, (1964), 429–438.

Cormack, R.M., "Log-Linear Models for Capture-Recapture," *Biometrics*, 45, 2, (1989), 395–413.

Cormack, R.M., "Variance of Mark-Recapture Estimates," *Biometrics*, 49, 4, (1993), 1188–1193.

Coulson, T., Catchpole, E.A., Albon, S.D., Morgan, B.J.T., Pemberton, J.M., Clutton-Brock, T.H., Crawley, M.J., and Grenfell, B.T., "Age, Sex, Density, Winter Weather, and Population Crashes in Soay Sheep," *Science*, 292, (2001), 1528–1531.

Crosbie, S.F. and Manly, B.J.F., "Parsimonious Modelling of Capture-Mark-Recapture Studies," *Biometrics*, 41, (1985), 385–398.

Dempster, A.P., Laird, N.M., and Rubin, D.B., "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1, (1977), 1–39.

Devroye, L., *Non-Uniform Random Variate Generation*, Springer-Verlag, 175 Fifth Avenue, New York, New York 10010, U.S.A, 1973, section 3.2-3.3, pp. 102–110, first edition.

Doucet, A., de Freitas, N., and Gordon, N. (editors), *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, Springer, New York, 2001, first edition.

Doucet, A., Godsill, S., and Andrieu, C., "On sequential Monte Carlo sampling methods for Bayesian filtering," *Journal Statistics and Computing*, 10, 3, (2000), 197–208.

Dupuis, J.A. and Schwarz, C.J., "A Bayesian Approach to the Multistate Jolly-Seber Capture-Recapture Model," *Biometrics*, 63, 4, (2007), 1015–1022.

Gelman, A., Meng, X.l., and Stern, H., "Posterior predictive assessment of model fitness via realized discrepancies (with discussion)," *Statistica Sinica*, 6, (1996), 733–807.

Gilks, W.R. and Roberts, G.O., *Strategies for improving MCMC in Markov Chain Monte Carlo in Practice*, Chapman and Hall, New York, 1996, chapter 6, pp. 90–114.

Gilks, W., Richardson, S., and Spiegelhalter, D. (editors), *Strategies for improving MCMC in Markov Chain Monte Carlo in Practice*, Chapman and Hall, New York, 1996.

Giminez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H., Vila, J.P., and Pradel, R., "State-space modelling of data on marked individuals," *Ecological Modelling*.

Gowans, S. and Whitehead, H., "Photographic Identification Of Northern Bottlenose Whales (*Hyperoodon Ampullatus*): Sources Of Heterogeneity From Natural Marks," *Marine Mammal Science*, 17, 1, (2001), 76–93.

Harrison, P.J., Buckland, S.T., Thomas, L., Harris, R., Pomeroy, P.P., and Harwood, J., "Incorporating movement into models of grey seal population dynamics," *Journal of Animal Ecology*, 75, 3, (2006), 634–645.

Harvey, A.C., *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, 1989.

Jolly, G.M., "Explicit Estimates from Capture-Recapture Data with Both Death and Immigration-Stochastic Model," *Biometrika*, 52, 1/2, (1965), 225–247.

Jolly, G.M., "Mark-Recapture Models with Parameters Constant in Time," *Biometrics*, 38, 2, (1982), 301–321.

Karanth, K.U. and Nichols, J.D. (2000), "Ecological Status and Conservation of Tigers in India," Final technical report, Division of International Conservation, US Fish and Wildlife Service and Wildlife Conservation Society and Centre for Wildlife Studies, Washington, DC and New York and Bangalore, India.

Kendall, W.L., "Robustness of Closed Capture-Recapture Methods to Violations of the Closure Assumption," *Ecology*, 80, 8, (1999), 2517–2525.

Kendall, W.L. and Nichols, J.D., "On the use of secondary capture-recapture samples to estimate temporary emigration and breeding proportions," *Journal of Applied Statistics*, 22, 5-6, (1995), 751–762.

King, R. and Brooks, S.P., "Model selection for integrated recovery/recapture data," *Biometrics*, 58, (2002a), 841–851.

King, R. and Brooks, S.P., "Bayesian model discrimination for multiple strata capture-recapture data," *Biometrika*, 89, 4, (2002b), 785–806.

King, R., Brooks, S.P., Morgan, B.J.T., and Coulson, T., "Factors Influencing Soay Sheep Survival: A Bayesian Analysis," *Biometrics*, 62, 1, (2006), 211–220.

Krzanowski, W.J., *An Introduction to Statistical Modelling*, Arnold, London, UK, 1998.

Lebreton, J.D., Burnham, K.P., Clobert, J., and Anderson, D.R., "Modeling Survival and Testing Biological Hypotheses Using Marked Animals: A Unified Approach with Case Studies," *Ecological Monographs*, 62, 1, (1992), 67–118.

Lebreton, J.D. and Pradel, R., "Multistate recapture models: modelling incomplete individual histories," *Journal of Applied Statistics*, 29, 1-4, (2002), 353–369.

Link, W.A. and Barker, R.J., "Hierarchical mark-recapture models: a framework for inference about demographic processes," *Animal Biodiversity and Conservation*, 27, 1, (2004), 441–449.

Link, W.A. and Barker, R.J., "Modeling Association among Demographic Parameters in Analysis of Open Population CaptureRecapture Data," *Biometrics*, 61, 1, (2005), 46–54.

Liu, J. and West, M., *Combined Parameter and State Estimation in Simulation Based Filtering*, Springer, New York, 2001, chapter 10, pp. 197–223, Statistics for Engineering and Information Science, first edition.

Liu, J.S. (editor), *Monte Carlo Strategies in Scientific Computing*, Springer Series in Statistics, Springer, New York, 2001, first edition.

Liu, J. and Chen, R., "Sequential Monte Carlo Methods for Dynamic Systems," *Journal of the American Statistical Association*, 93, 443, (1998), 1032–1044.

Meyer, R. and Millar, R.B., "Bayesian stock assessment using a state-space implementation of the delay difference model," *Canadian Journal of Fisheries and Aquatic Sciences*, 56, 1, (1999), 37–52.

Millar, R.B. and Meyer, R., "Bayesian state-space modeling of age-structured data: fitting a model is just the beginning," *Canadian Journal of Fisheries and Aquatic Sciences*, 57, 1, (2000a), 43–50.

Millar, R.B. and Meyer, R., "Non-Linear State Space Modelling of Fisheries Biomass Dynamics by Using Metropolis-Hastings within-Gibbs Sampling," *Applied Statistics*, 49, 3, (2000b), 327–342.

Milner, J.M., Elston, D.A., and Albon, S.D., "Estimating the Contributions of Population Density and Climatic Fluctuations to Interannual Variation in Survival of Soay Sheep," *The Journal of Animal Ecology*, 68, 6, (1999), 1235–1247.

Newman, K.B., "State-Space Modeling of Animal Movement and Mortality with Application to Salmon," *Biometrics*, 54, 4, (1998), 1290–1314.

Newman, K.B., Buckland, S.T., Lindley, S.T., Thomas, L., and Fernández, C., "Hidden Process Models for Animal Population Dynamics," *Ecological Applications*, 16, 1, (2006), 74–86.

Newman, K.B., Fernández, C., Thomas, L., and Buckland, S.T., "Monte Carlo Inference for State-Space Models of Wild Animal Populations," *In press.*

Nichols, J.D. and Hines, J.E., "Approaches for the direct estimation of λ, and demographic contributions to λ, using capture-recapture data," *Journal of Applied Statistics*, 29, 1-4, (2002), 539–568.

Nichols, J.D., Hines, J.E., Pollock, K.H., Hinz, R.L., and Link, W.A., "Estimating Breeding Proportions and Testing Hypotheses about Costs of Reproduction with Capture-Recapture Data," *Ecology*, 75, 7, (1994), 2052–2065.

Nichols, J.D. and Kendall, W.L., "The use of multi-state capture-recapture models to address questions in evolutionary ecology," *The Journal of Applied Statistics*, 22, 5 & 6, (1995), 835–846.

Nichols, J.D., Sauer, J.R., Pollock, K.H., and Hestbeck, J.B., "Estimating Transition Probabilities for Stage-Based Population Projection Matrices Using Capture-Recapture Data." *Ecology*, 73, 1, (1992), 306–312.

Pitt, M. and Shephard, N., "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94, 446, (1999), 590–599.

Pollock, K.H., "Capture-Recapture Models Allowing for Age-Dependent Survival and Capture Rates," *Biometrics*, 37, 3, (1981), 521–529.

Pollock, K.H., "Modeling Capture, Recapture, and Removal Statistics for Estimation of Demographic Parameters for Fish and Wildlife Populations: Past, Present, and Future," *Journal of the American Statistical Association*, 86, 413, (1991), 225–238.

Pollock, K.H., "The use of auxiliary variables in capture-recapture modelling: an overview." *Journal of Applied Statistics*, 29, 1-4, (2002), 85–102.

Pollock, K.H., Nichols, J.D., Brownie, C., and Hines, J.E., "Statistical Inference for Capture-Recapture Experiments," *Wildlife Monographs*, 107.

Poole, D., "Bayesian Estimation of Survival from Mark-Recapture Data," *Journal of Agricultural, Biological, and Environmental Statistics*, 7, 2, (2002), 264–276.

Pradel, R., "Utilization of Capture-Mark-Recapture for the Study of Recruitment and Population Growth Rate," *Biometrics*, 52, 2, (1996), 703–709.

Rivot, E., Prévost, E., Parent, E., and Baglinière, J.L., "A Bayesian state-space modelling framework for fitting a salmon stage-structured population dynamic model to multiple time series of field data," *Ecological Modelling*, 179, 4, (2004), 463–485.

Schnute, J.T., "A General Framework for Developing Sequential Fisheries Models," *Canadian Journal of Fisheries and Aquatic Sciences*, 51, (1994), 1676–1688.

Schofield, M.R. and Barker, R.J. (2008), "A Unified Capture-Recapture Model," .

Schwarz, C.J., "The Jolly-Seber Model: More than Just Abundance," *Journal of Agricultural, Biological and Environmental Statistics*, 6, 2, (2001), 195–205.

Schwarz, C.J. and Arnason, A.N., "A General Methodology for the Analysis of Capture-Recapture Experiments in Open Populations," *Biometrics*, 52, 3, (1996), 860–873.

Schwarz, C.J., Schweigert, J.F., and Arnason, A.N., "Estimating Migration Rates Using Tag-Recovery Data," *Biometrics*, 49, 1, (1993), 177–193.

Schwarz, C.J. and Seber, G.A.F., "Estimating animal abundance: review III," *Statistical Science*, 14, (1999), 427–456.

Seber, G.A.F., "A Note on the Multiple-Recapture Census," *Biometrika*, 52, 1/2, (1965), 249–259.

Seber, G.A.F., *The Estimation of Animal Abundance and Related Parameters*, MacMillan, New York, N. Y., 1982, 2nd edition.

Seber, G.A.F. and Schwarz, C.J., "Capture-Recapture: Before and after EURING 2000," *Journal of Applied Statistics*, 29, 1, (2002), 5–18.

Smith, A.F.M. and Gelfand, A.E., "Bayesian Statistics without Tears: A Sampling-Resampling Perspective," *The American Statistician*, 46, 2, (1992), 84–88.

Thomas, L., Buckland, S.T., Newman, K.B., and Harwood, J., "A unified framework for modelling wildlife population dynamics," *Australian and New Zealand Journal of Statistics*, 47, (2005), 19–34.

Trenkel, V.M., Elston, D.A., and Buckland, S.T., "Fitting population dynamics models to count and cull data using sequential importance sampling." *Journal of the American Statistical Association*, 95, (2000), 363–374.

White, G.C., "Numerical Estimation of Survival Rates from Band-Recovery and Biotelemetry Data," *The Journal of Wildlife Management*, 47, 3, (1983), 716–728.

White, G.C., "Discussion comments on: the use of auxiliary variables in capture-recapture modelling. An overview." *Journal of Applied Statistics*, 29, 1-4, (2002), 103–106.

White, G.C. and Burnham, K.P., "Program MARK: survival rate estimation from both live and dead encounters," *Bird Study*, 46, (1999), 120–139.

White, G.C., Burnham, K.P., and Anderson, D.R., *Advanced features of program MARK*, Wildlife Society, Bethesda, Maryland, USA, 2001, pp. 368–377, Wildlife, land, people: priorities for the 21st century.

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

Yang, H.C. and Chao, A., "Modeling Animals' Behavioral Response by Markov Chain Models for Capture-Recapture Experiments," *Biometrics*, 61, 4, (2005), 1010–1017.

Appendix A

The Alias Method

The Alias Method is a means by which we can draw random deviates from a non-uniform finite discrete distribution. This section describes the method and provides examples to illustrate its application.

If we wished to draw a random deviate x from a Uniform distribution on the range $(1, n)$. Since each of the integers $1, .., n$ has the same chance, $p = 1/n$, of being selected one method of generating x would be by drawing a value y where Y Uniform$(0, 1)$ and setting $x = \lceil ny \rceil$. In a geometric framework this approach is equivalent to partitioning the unit square by dividing the x-axis into n strips of equal width to describe slabs in the plane, as illustrated in Fig A.1.

We then draw random uniform (x, y) coordinates in this rectangle and return the value of the strip the coordinate lies in. For a uniformly distributed range of values the y coordinates are not important, it is only the value of the x-coordinate that determines which value we draw from the range $(1, n)$.

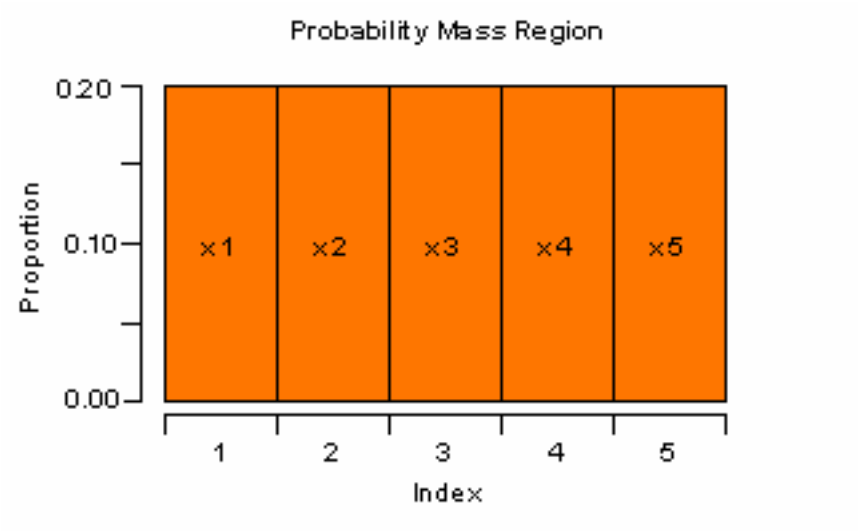
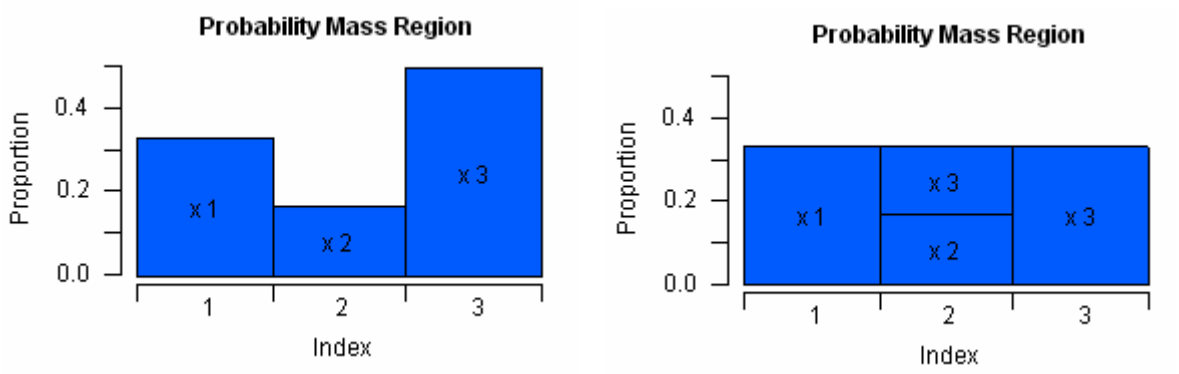


Figure A.1: Partitioning of the unit square when $n = 5$

If we expand this approach to drawing random variates from ranges of discrete values with non-uniform probabilities the usefulness of the geometric interpretation becomes more apparent. If we have a discrete range of values $x_1, \ldots, x_n$ and assign the probabilities $p_1, \ldots, p_n$ to these values the idea now is to rearrange the probability mass so that we have a rectangle subdivided into strips of equal width, as before, but this time these strips are partitioned into vertical subregions.

For example, Figure A.2 illustrates the re-allocation of the probability mass on x_1, x_2, x_3 (Figure A.2a) to create a rectangle with the appropriate vertical partitioning (Figure A.2b).

Any single strip will never need to be divided into more than two vertical subregions, one part associated with the original value x_i and the other associated with an alternative value x_j for some $j \neq i$. This alternative value will be referred to as the alias value. Once we have rearranged the probability mass into strips subdivided into vertical subregions we then need to devise a method to select one of the values. As before, conceptually, we draw random uniform (x, y) coordinates in the probability mass rectangle and return the



(a) We want to go from this...

(b) ...to this.

Figure A.2: Example figure of re-arranging the probability mass to form a rectangular region demarcated into sub-regions.

value of the region the coordinate lies in. An efficient way of doing this is to create an alias table containing the values in the discrete range ($\mathbf{x} = x_1, \ldots, x_n$), their associated probabilities ($\mathbf{p} = p_1, \ldots, p_n$), the alias values ($\mathbf{a} = a_1, \ldots, a_n$) and threshold values indicating the proportion of the strip's total probability mass taken by the original value ($\mathbf{q} = q_1, \ldots, q_n$). Thus, the elements of the vector $\mathbf{q}$ are then defined as

$$q_i = \frac{\text{height of lower part of rectangle } i}{\text{height of rectangle } i}.$$

The alias table for the above example is:

i	x_i	p_i	a_i	q_i
1	x_1	1/3	x_1	1
2	x_2	1/6	x_3	0.5
3	x_3	1/2	x_3	1

To obtain a draw from a range of n non-uniformly distributed discrete values we generate

two variates from a uniform distribution on $(0, 1)$. So we have:

$$\begin{aligned} U_1, U_2 &\sim U(0,1) \\ \text{Let } M &= \lceil nU_1 \rceil \\ \text{Then if } U_2 &= q_M \text{ return } x_i \text{ else return } a_M. \end{aligned}$$

In the random uniform (x, y) coordinates framework the value of U_1 represents the x-coordinate and determines which strip we look in and the value of U_2 represents the y-coordinate and determines which vertical sub-region we choose in the strip. An efficient method of creating the alias table is described in the following section.

A.1 Basic Alias Algorithm:

We wish to draw random deviates from a range containing n discrete values $x_1, x_2, \ldots, x_n$. The i^{th} value in the range, x_i, has a probability mass of p_i. We begin by creating an alias table which is achieved using the following procedure:

(1) Create vectors $\mathbf{x}, \mathbf{p}, \mathbf{a}$ and $\mathbf{q}$ all of length n so that:

- $\mathbf{x}$ contains the n discrete values: $\mathbf{x}[i] = x_i$
- $\mathbf{p}$ contains the probability masses of the discrete values: $\mathbf{p}[i] = p_i$
- $\mathbf{a}$ contains the n discrete values: $\mathbf{a}[i] = \mathbf{x}[i]$
- $\mathbf{q}$ contains the scaled probability masses: $\mathbf{q}[i] = Np_i$

(2) Set up two lists: L and H to contain indices for lower and higher values of q_i such that:

- If $q_i \begin{cases} < 1 & i \in L \quad \text{(assign the value } i \text{ to the list } L\text{)} \\ \geq 1 & i \in H \quad \text{(assign the value } i \text{ to the list } H\text{)} \end{cases}$

- If L is empty then we have a Uniform distribution and we can stop.

(3) While the list L is non-empty we continue with the following steps:

(a) Select an index j from L and an index k from H.

(b) Set = $\mathbf{a}[j] = \mathbf{x}[k]q[k] = q[k] - (1 - q[j])$

(c) If $\mathbf{q}[k] < 1$ $\begin{cases} \text{then remove the index } k \text{ from the list } H \\ \text{and add the index } k \text{ to the list } L \end{cases}$

(d) Remove the index j from L.

If the list L is empty then stop, otherwise go to (3)(a).

Once the alias table is complete we then proceed as described before and draw two variates from a Uniform distribution on $(0, 1)$.

A.2 An Example

This section provides an example of how to implement the Alias algorithm.

We begin with a set of 5 values $x_1, \ldots, x_5$ and their associated probability masses 0.15, 0.05, 0.3, 0.2 and 0.3. We set up the vectors as directed and display them in a table. Our initial table is then

i	x_i	p_i	a_i	q_i
1	x_1	0.15	x_1	0.75
2	x_2	0.05	x_2	0.25
3	x_3	0.3	x_3	1.5
4	x_4	0.2	x_4	1
5	x_5	0.3	x_5	1.5

We have $L = \{1, 2\}$
$H = \{3, 4, 5\}$

We select the index 2 from L and the index 3 from H. So, $\mathbf{a}[2] = 3$ and $\mathbf{q}[3] = 1.5 - (1 - 0.25) = 0.75$. $q[3]$ is less than one so we remove index 3 from H and add it to L. We also remove index 2 from L. The table now becomes:

i	x_i	p_i	a_i	q_i
1	x_1	0.15	x_1	0.75
2	x_2	0.05	x_3	0.25
3	x_3	0.3	x_3	0.75
4	x_4	0.2	x_4	1
5	x_5	0.3	x_5	1.5

We have $L = \{1, 3\}$, $H = \{4, 5\}$

L is non-empty so we go back to step (3)(a). and choose two more indices. Choosing index 1 from L and index 5 from H we find: $\mathbf{a}[1] = 5$ and $\mathbf{q}[5] = 1.5 - (1 - 0.75) = 1.25$. Now $\mathbf{q}[5]$ is not less than one so all we can do is to remove the index 1 from L. The table then becomes:

i	x_i	p_i	a_i	q_i
1	x_1	0.15	x_5	0.75
2	x_2	0.05	x_3	0.25
3	x_3	0.3	x_3	0.75
4	x_4	0.2	x_4	1
5	x_5	0.3	x_5	1.25

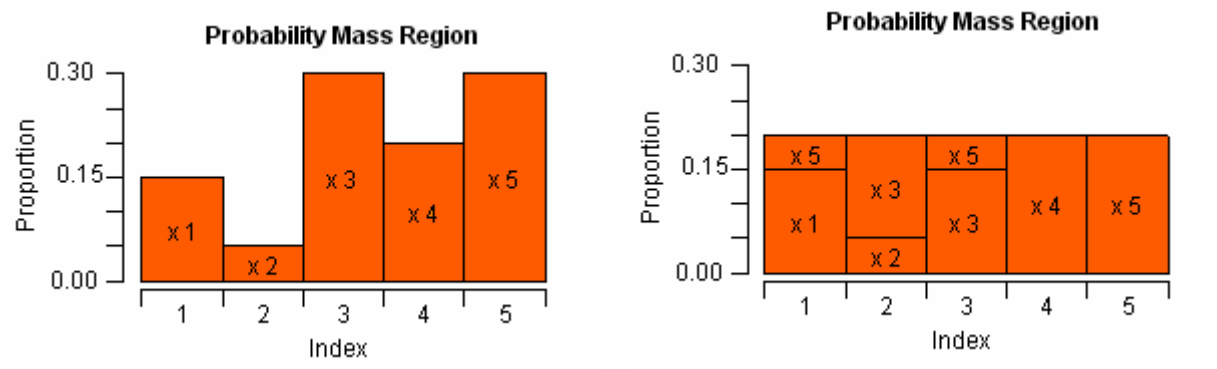
We have $L = \{3\}$, $H = \{4, 5\}$

L is still non-empty so we go back to step (3)(a). and choose two more indices. Choosing index 3 from L and index 5 from H we have: $\mathbf{a}[3] = 5$ and $\mathbf{q}[5] = 1.25 - (1 - 0.75) = 1$. Since $\mathbf{q}[5]$ is still not less than one all we can do is the remove the index 3 from L. The table then becomes:

i	x_i	p_i	a_i	q_i
1	x_1	0.15	x_5	0.75
2	x_2	0.05	x_3	0.25
3	x_3	0.3	x_5	0.75
4	x_4	0.2	x_4	1
5	x_5	0.3	x_5	1

We have $L = \{\}$
$H = \{4, 5\}$

L is empty so we now have our Uniform distribution. We have rearranged the probability mass region (Figure A.3a) into the uniform rectangle shown in Figure A.3b.



(a) Initial distribution

(b) Re-arranged uniform distribution

Figure A.3: Example figure of re-arranging the probability mass for a more complicated distribution to form a rectangular region demarcated into sub-regions.

We now generate two random deviates from a uniform distribution on $(0, 1)$ and obtain $U_1 = 0.461$ and $U_2 = 0.842$. We set $M = \lceil nU_1 \rceil = \lceil 5 \times 0.461 \rceil = \lceil 2.305 \rceil = 3$. Since $U_2 = 0.842 > 0.75 = q_3$ we return $a_3 = x_5$.