SAPIENZA
UNIVERSITÀ DI ROMA

Bayesian finite mixture models to account for latent heterogeneity in Capture-Recapture analysis

School of Statistical Sciences
Ph.D. program in "Methodological Statistics" (XXXV cycle)

Gianmarco Caruso
ID number 1654516

Advisors
Prof. Giovanna Jona Lasinio
Prof. Luca Tardella

Academic Year 2022/2023

Thesis defended on 30 May 2023
in front of a Board of Examiners composed by:

Marco Alfò (chairman)
Enea Giuseppe Bongiorno
Roberto Di Mari

Bayesian finite mixture models to account for latent heterogeneity in Capture-Recapture analysis
PhD thesis. Sapienza University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Version: May 2023

Author's email: gianmarco.caruso@uniroma1.it

To all those who dare to doubt
and keep pushing forward.

Acknowledgments

I would like to thank my supervisors, Prof. Giovanna Jona Lasinio and Prof. Luca Tardella, since their guidance and mentoring during my PhD journey helped shape my development as a researcher. I thank Dr. Pierfrancesco Alaimo Di Loro and Dr. Marco Mingione for sharing with me their expertise in simulation studies and more advanced R programming, and Prof. Daniela Silvia Pace for providing me with the data about the bottlenose dolphins and giving me biological insights about the behaviour of this species. I want to express my gratitude to the two referees who reviewed my PhD thesis, Prof. Alessio Farcomeni and Prof. Andrea Tancredi, who gave me very good feedback about my work and whose precious comments and suggestions helped me to improve the quality of the manuscript. I am also grateful for their time and effort in evaluating my work.

Additionally, I want to thank Prof. Marco Perone Pacifico and Dr. Francesco Iafrate, with whom I collaborated as a teaching assistant at LUISS during these three years. The supplementary activity of teaching assistance gave me a lot of satisfaction and helped me stay refreshed with the basics of Statistics.

Starting a PhD project during a pandemic which has disrupted all aspects of life may be hard and stressful, so I want to share my gratefulness also to those people who helped me maintain – consciously or unconsciously – a healthy work-life balance throughout this challenging journey. For this reason, I want to thank my modern family, who have always had faith in me, even though they may not fully understand my work. Their love has been a source of strength and comfort for me. My dear sister Valentina has always been my first emotional support when difficulties have come, and I am grateful with her for this.

I also want to thank all the friends and dear people who have been beside me during these intense three years, full of events and changes. In particular, I thank my good old friends, Ilaria S., Roberta, Antonio, Federica and Luca, with whom I have grown up and shared many lovely moments: despite our diverse paths in life, we have all stayed connected and supported each other over the years; Matteo, Giulia and Ilaria V. for the countless (mis)adventures we have experienced together during these years, which have always been source of laughter: to many more years of fun together; Andrés, Carmen and their friends in UK, who have become my Spanish family in UK and from whom I have learnt that the struggles and fears I have faced are common to PhD students around the world: I am particularly grateful to Andrés who showed me how fortune favours the bolds and helped me in regaining confidence and trust in my own abilities; my climbing friends, Luigi, Elisa, Raissa, Marco, Daniele and Paolo for all the outdoor adventures, the laughter and the beers after climbing on rock; and, finally, Vittoria L., Vittoria M., Riccardo, Giorgia and all my PhD colleagues with whom I have exchanged mutual emotional support during these years which helped us cope with the challenges and stresses of the PhD journey.

Contents

Abstract

Gaining knowledge about the size of an animal population in a given area is of particular interest for wildlife management and conservation. Indeed, over the last decades, thousands of species worldwide have been experiencing either an outsize expansion or, more often, a dramatic shrinkage in their abundances: in the worst cases, the latter trend has even led to their extinction. Since carrying out a complete count of animal populations is generally a challenging task, Capture-Recapture models have arisen as valuable tool to estimate the population abundance in the chosen study area, along with some other demographically meaningful parameters. A further ecological key issue for wildlife managers involves the identification of distinct groups of individuals that share similar biological patterns. In this spirit, we bring to light how finite mixtures can be easily embed into Capture-Recapture models in order to carry out jointly the estimation and the classification task. We adopt a Bayesian modelling perspective and this requires ad-hoc solutions in this specific context. Indeed, the literature about Bayesian finite mixture Capture-Recapture models is scarce in addressing some issues that arise in the implementation of the model, such as the common label-switching problem that affects finite mixtures and the specification of suitable prior distributions on component-specific parameters. Notably, we deal with these two issues by proposing two novel flexible classes of joint priors for parameters bounded in the $[0, 1]$ set. The idea is to specify joint priors that both retain the flexibility to induce the desired marginal behaviour on the component-specific parameters and help the correct identification of their posterior distributions. The proposals are enhanced by the derivation of some theoretical results. Moreover, we propose a class of parsimonious cross-classified mixture models which can be successfully used to identify different residency patterns in wildlife populations. Notably, when the existence of such patterns is known in advance, finite mixtures can be leveraged to model the structure of the population under study. For each proposed methodology, a simulation study is carried out to investigate its inferential benefits and pitfalls. The application of the outlined models and methods is illustrated on wildlife datasets, revealing their merits and validity in real-world examples and giving insights that may be useful to practitioners.

Chapter 1

Motivation and overview

Climate change is recognised as a major threat to global biodiversity, since it may lead to a very dramatic redistribution - and even to the extinction - of thousands of animal species over the next decades (Thomas et al., 2004; Pecl et al., 2017). Besides, the effects of global warming have already been visible over the past decades, producing numerous shifts in the distributions and abundances of species. For example, a recent study (Soroye et al., 2020) pointed out that an increasing frequency of unusually hot days is leading to the decline of lots of bumble bee species across North America and Europe. A very recent paper (Huvier et al., 2023) warns about the possible future extinction of the boreal linx (*Lynx lynx*) in the French Jura Mountains: severe inbreeding depression has led to drastic reduction of their size and any environmental change can have serious negative effects, since this population is not strong enough to adapt to these new changes.

In contrast, some wild animal species are seen as growing too fast: for example, while previously confined to natural areas with low human presence, in the last decades the wild boar has colonised urban environments, establishing a permanent presence in several European cities. This is the case of the municipality of Rome, which is currently facing a serious issue in the boar-human cohabitation, often due to collisions of the wild animal with vehicles, damages to gardens and ransacking of rubbish containers (Amendolia et al., 2019). There is an increasing international awareness about the need both to conserve as much as possible of the current biological diversity and to control the too rapid expansion of some wild species. Indeed, government institutions may adopt strategies to preserve some endangered species. An example is the *Western Shield*, a widespread fox control campaign organised by the Government of Western Australia to protect smaller native species of mammals from the European red fox (*Vulpes vulpes*): this campaign has been successful, having led to an increase in abundance and range of several species (Burbidge and Manly, 2002).

A central focus of conservation biology is, thus, the maintenance of biodiversity by preventing the uncontrolled expansion of some species and the total extinction of some others. To achieve this goal, wildlife managers need simple quantitative approaches and tools that could help them to make better management decisions. In this regard, the monitoring of marked animals and the analyses of the resulting data by Capture-Recapture statistical models have become a common and powerful

wildlife conservation paradigm. Indeed, in absence of a census of the individuals belonging to a wildlife population in a limited area, multiple Capture-Recapture samplings of these individuals - in combination with suitable statistical models - allows to estimate the population size and to obtain an associated measure of uncertainty (Amstrup et al., 2005; McCrea and Morgan, 2014).

A further concern consists in understanding the diversity between the animals that constitute a population: in ecology, this diversity is called *individual heterogeneity* (Gimenez et al., 2018). This heterogeneity may be due to observable and measurable traits of the individual (e.g. age, sex, location, weight, breeding status) or to latent traits, and plays a key role in population dynamics. Therefore, along with the estimation of the population abundance and of its evolution over time, it is often interesting to identify groups of individuals that share similar biological patterns. When covariates are available, these can be used to try to explain differences between individuals. However, unexplained heterogeneity is often present among individuals and several approaches have been proposed in literature to account for it. Notably, finite mixtures have arisen in the Capture-Recapture context as a means to capture heterogeneity across group of individuals. In these models, a latent variable is used to assign individuals to one of the mixture components characterised by specific parameters.

In Capture-Recapture literature, the grouping induced by finite mixtures is often deemed to be an artifact to allow for heterogeneity and correct for the bias in the estimate of the population size (e.g. Pledger (2000); Pledger et al. (2010)). In this thesis we restore finite mixture models in the context of Capture-Recapture analysis as a valuable model-based clustering tool to jointly estimate model's parameters and identifying groups of individuals that share similar profiles. In general, finite mixtures can be useful as a mean for finding groups of individuals that are similar: what *similar* means strictly depends on the underlying parametric distribution that is assumed for each component of the finite mixture model. Remarkably, when suitable considerations lead to a grouping being plausible biologically, the groupings from a finite mixture model can also give insights into the nature of the population.

We adopt a Bayesian approach throughout the whole thesis, following the original idea of Royle et al. (2007) for successfully fitting Bayesian Capture-Recapture models and embedding finite mixtures within this framework. We believe that literature about such Bayesian implementation of this class of models is rather scarce and it does not usually address some related challenges that may arise, such as the elicitation of prior distributions or the correct identification of the component-specific parameters: we feel that these topics would deserve more attention since, after all, these kinds of models are mostly used by ecologists and biologists, who may generally have limited statistical knowledge and skills. In this spirit, we try to address some issues related to the implementation of Capture-Recapture finite mixture models in a Bayesian context, for example by focusing on the prior specifications of their component-specific parameters. Moreover, we propose and investigate parsimonious mixture models whose components are characterised by a partially common set of parameters. Parsimony represents a trade-off between too few parameters and too little model structure - which may induce to model bias and underestimates of sampling variation - versus too many parameters and too much model structure - which may lead to overfitting and a lack of precision in estimates. For this reason, parsimonious models

should be pursued whenever possible. In addition, finite mixtures are not only useful to let emerge clusters of individuals sharing similar patterns, but they can also be used to model the already known structure of a population of individuals. Specifically, we adopt an ad-hoc finite mixture model to estimate and classifying individuals belonging to a bottlenose dolphin (*Tursiops truncatus*) population. Several ecological studies (Dinis et al., 2016; Estrade and Dulau, 2020; Haughey et al., 2020; Pace et al., 2021; La Manna et al., 2022) on this kind of population have identified three different types of individuals showing distinct residency patterns and we try to provide a model that takes into account this particular structure.

1.1 Structure of the thesis

After this general introduction to the motivations that inspire this work, we outline the structure of the thesis.

Chapter 2 is an introduction to Capture-Recapture analysis: basic models are presented, along with the necessary definitions, hypotheses and notation. A first main distinction is made between closed and open populations, depending whether the assumption of no births, no deaths and no migration during the whole survey may hold or not. The main closed-population and open-population models (Jolly-Seber) are illustrated, along with the sources of variability that may affect the models' parameters and lead to relax some basic Capture-Recapture hypotheses. Finally, the parameter-expanded data augmentation approach is carefully detailed: it consists in a valuable alternative formulation of Capture-Recapture models that allows an easy way to fit Bayesian models in this context.

Bayesian finite mixture models in the context of Capture-Recapture analysis are introduced in Chapter 3 as a tool to handle unexplained heterogeneity across groups of individuals. Emphasis is given to closed-population models with individuals presenting heterogeneous capture probabilities. Challenges related to these kinds of models are discussed by focusing on specifications of the prior distributions of the model's parameters. Notably, when dealing with finite mixtures, Bayesian machinery is often affected by the so-called label-switching problem. A common solution for practitioners to overcome this issue consists in assigning ordering constraints between class-specific parameters, so as to identify their relative roles. That is usually achieved by specifying conditionally uniform densities that respect such constraints, preventing the possibility to shape the prior according to available prior knowledge; however, the implication of this choice on the marginal prior distributions of the class-specific parameters is not usually discussed in literature. We address this issue and we generalise this approach by proposing two novel flexible classes of joint priors based on manipulating Beta distributions. The idea is to specify a joint prior that retains the flexibility to induce the desired marginal behaviour while still guaranteeing the desired ordering. Our proposals are enhanced by the proof of some original theoretical results concerning Beta truncated and Beta restricted probability distributions. Then, a simulation study that compares several alternative prior specifications is carried out to investigate the inferential benefits of the proposed methodology. Finally, the practical use of this approach is illustrated on a real dataset concerning a closed population.

Chapter 4 starts with the review of the finite mixtures within the Jolly-Seber model framework. These models are used to handle heterogeneity between recruitment, survival and detection across group of individuals. Along with the classical single mixture grouping in G classes, a double mixture grouping in $G \times H$ classes can be specified (these are the so-called *cross-classified models*). The latter allows to consider components that share some common parameters. In this spirit, we propose a class of more parsimonious cross-classified models that may turn to be useful to identify different residency patterns in animal open populations. We illustrate their use on a real data example based on a robust design analysis, a particular sampling scheme in which a series of closely spaced samples are separated by longer intervals. Motivated by the real data example, we further investigate the performance of the considered class of models in a simulation experiment.

On the wake of what discussed in the previous chapters, in Chapter 5 finite mixtures are adopted to characterise the residency patterns in marine wildlife populations (e.g. dolphin populations). In this case, several ecological studies have identified a recurring structure in some marine species, which tend to dwell in the study area over time with different propensities. They have recognised three distinct behavioural patterns that are followed by these species: specifically, they can be deemed to be either residents, long-term visitors (or part-times) or short-term visitors (transients). In the light of this structure, we propose an original parsimonious cross-classified model to jointly estimate the population abundance - along with the main ecological parameters of interest - and classify the encountered individuals according to their residency pattern. Model performances are assessed through a simulation experiment and, finally, an application to a common bottlenose dolphin (*Tursiops truncatus*) population is illustrated at the end of the chapter.

The final remarks at the end of the chapters briefly resume their main content, highlight some benefits and pitfalls of the presented methodologies and give some insights about future interesting directions to follow and further developments that might be welcomed to expand our proposals.

As customary, we include in appendices those written parts that would otherwise have overloaded the reading of the main part of the thesis. Appendix A presents BUGS code listings for implementing all the presented models with through the Bayesian software JAGS. Appendix B shows the mathematical derivation of the multinomial likelihood for the basic Capture-Recapture closed-population model. Appendix C contains original theoretical proofs on prior specifications of class-specific probabilities. Finally, Appendix D reports the details of the time-lags between the sampling occasions considered in the simulation study of Chapter 5.

Chapter 2

Introduction to Capture-Recapture analysis

The abundance of wildlife populations, along with other demographic information (e.g. birth, death, migration, etc.), is a key factor in ecology to help conservation managers in understanding the underlying dynamics of an ecosystem, monitoring the evolution of the population over time and developing suitable strategies for biodiversity preservation (Lebreton et al., 1992; Morrison et al., 2012; Fryxell et al., 2014). Researchers are typically concerned with the analysis of a specific wildlife species in a limited natural area. However, a census of all the individuals belonging to a population is usually infeasible, so that statistical methods are needed to estimate the total abundance of the population and an associated measure of uncertainty (Williams et al., 2002; Schmidt, 2005; Nichols et al., 2009). In this context, Capture-Recapture (CR, henceforth) methods have been extensively used to estimate the size of wildlife populations which are subject to multiple capture occasions: recapture information can be then exploited to estimate the number of uncaptured individuals and, consequently, the total size of the population.

If samples are repeated over a short time period, the population can be considered *closed*, i.e. there are neither births, deaths nor migrations beyond the study area throughout the whole study period. If this cannot hold, population should be considered *open* and the process of entry and exit must be taken into consideration. A key assumption is that all individuals in the population behave independently from one another, so that the capture of an individual does not affect the catchability of the others. Researchers may design a trapping experiment in a way that guarantees the reliability of the main assumptions of the CR. For example, captures may be carried out in a very short time period and in a well-bounded area, so that the population can be reasonably considered as closed.

The term *Capture-Recapture* is due to the most common way to identify wild animals across multiple detection occasions, namely through capture, marking and release: the animal is typically caught using a bait, marked with a unique identifier (e.g. a numbered tag) and then released unharmed into the environment. Once animals have been marked, then recapturing some of them in subsequent samples provides information about the total number of animals in the area (Pradel, 1996). Observations do not only correspond to physical captures, but they may also consist

in less invasive techniques, such as photographic identification of natural markings (Auger-Méthé et al., 2010; Gardner et al., 2009; Pace et al., 2021) or DNA matching (Lukacs and Burnham, 2005; Morin et al., 2016), where genetic material (e.g. hairs, faeces) is collected in the field and used to identify individuals. Less invasive or non-invasive techniques usually tend to implicitly exclude the possibility of individuals' behavioral response, since the researcher do not physically interact with the animal.

Interestingly, Capture-Recapture analysis can be more widely applied to multiple data sources through the linkage of individuals across multiple lists. This approach - often referred to as *multiple systems estimation* - is preferred when estimating hard-to-reach human populations, including those who are sclerosis patients (Farcomeni, 2020), individuals affected by Covid-19 infections (Böhning et al., 2020), victims of modern slavery (Sharifi Far et al., 2021), refugees and migrants (Farcomeni, 2022) or drug dealers (Altieri et al., 2022). Although the paramount relevance of such scope - which also presents interesting challenges due to the higher complexity of human behaviours -, in the following we only focus on Capture-Recapture models in the ecological field.

2.1 Closed populations

We start considering a closed population whose size is assumed to be fixed, where all units are supposed to act independently each other and no misidentification of the units occurs.

In the simplest case in which only two visits are made in the study area, the population size of a closed population can be estimated through the Lincoln-Petersen estimate (Seber et al., 1982),

$$\hat{N}_{LP} = \frac{n_1\, n_2}{m_2}\,, \tag{2.1}$$

where n_1 is the number of animals captured and marked in the first visit, n_2 is the number of animals captured in the second visit and m_2 is the number of marked animals that were recaptured in the second visit. By assuming that all individuals have the same probability to be recaptured in the second visit, the idea behind this simple estimate is that the proportion of marked individuals that are caught in the second sample should be approximately equal to the proportion of individuals that were caught in the first visit, namely that $\frac{m_2}{n_2} \simeq \frac{n_1}{N}$. Of course, basic assumptions are that the researcher correctly records all marks and that animals do not lose their marks.

Although the intuitive derivation, the previous estimator is biased at small sample sizes and can be replaced by the Chapman estimator, i.e.

$$\hat{N}_C = \frac{(n_1+1)\,(n_2+1)}{(m_2+1)} - 1\,,$$

which is less biased than (2.1) (Brittain and Böhning, 2009).

This simple scheme can be generalized to the case of $T > 2$ sampling occasions. Let D be the number of distinct units observed at least once during T sampling occasions. Observed data can be thus represented by a $D \times T$ binary matrix,

$\mathbf{Y} = [y_{it}]$, where

$$y_{it} = \begin{cases} 1 & \text{unit } i \text{ has been captured at occasion } t \\ 0 & \text{otherwise} \end{cases}$$

The matrix $\mathbf{Y}$ contains all the observed capture histories for the D observed individuals on its rows. An example of matrix $\mathbf{Y}$ is provided in Table 2.1.

1	1	1
1	0	1
1	0	1
0	1	0
0	1	0
0	1	1
0	0	1

Table 2.1. Capture histories of $D = 7$ individuals encountered during $T = 3$ sample occasions.

In this case, estimator (2.1) can be generalised by the Schnabel estimator (Schnabel, 1938; Overton, 1965), i.e.

$$\hat{N}_S = \frac{\sum_{t=1}^{T} n_t \, A_t}{\sum_{t=1}^{T} m_t} \,,$$

where n_t is the number of individuals captured on the t-th occasion, m_i is the number of marked individuals captured on the t-th occasion and A_t is the number of marked individuals available to the capture on the t-th occasion.

Under independence both among capture occasions and among individuals, it is reasonable to assume that

$$y_{it} \overset{iid}{\sim} Bernoulli(p_{it}) \,,$$

where $p_{it} = Pr(y_{it} = 1)$ is the probability that individual i is observed at occasion t ($i = 1, \ldots, N$ and $t = 1, \ldots, T$). These probabilities are often referred to as *capture probabilities* or *detection probabilities*. The first basic CR model assumes that capture probabilities do not vary with time and are constant for all the individuals, namely $p_{it} = p$, $\forall i, t$. This is often referred to as model M_0.

2.1.1 Sources of variability in capture probabilities

It most of the cases, it is unreasonable to use the same capture probability for all individuals and all the sampling periods. Otis et al. (1978) distinguish between three factors that may affect the detection probability, namely the time effect (model M_t), the individual heterogeneity effect (M_h) and the behavioural effect (M_b), or a combinations of two (M_{th}, M_{tb} M_{bh}) or all of them (M_{tbh}).

Time effect makes detection probability varying among different encounter occasions. This consists in assuming that every animal in the population has the

same probability of capture at each sampling occasion, namely $p_{it} = p_t$, $\forall i, t$. These temporal changes in capture probabilities may be due, for instance, to different weather conditions that may affect the behaviour of the individuals or different efforts among sampling occasions.

Individual effect highlights differences in detection among individuals; therefore, they are supposed to have different detection probabilities, namely $p_{it} = p_i$, $\forall i, t$: some may appear more elusive than others, according to observable or latent factor. Heterogeneity of captures among individuals often causes negative bias in estimates of population abundance (N), when one does not account for it: intuitively, if some individuals are more likely to be captured on each occasion than the rest of the population, the estimated overall capture probability will tend to be higher, by causing an underestimation of the uncaptured individuals. This problem has been treated by several authors (Burnham and Overton, 1979; Alho, 1990; Tardella, 2002; Gimenez et al., 2018).

Individual effect may be fixed or random, categorical or continuous, latent or explained by means of measurable covariates (e.g. sex, age, weight, breeding status, location). For example, when individual heterogeneity depends on K fixed covariates, we can model p_i as

$$logit(p_i) = \alpha + \beta_1 x_{i1} + \cdots + \beta_k x_{iK} \,, \tag{2.2}$$

where $\alpha, \beta_1, \ldots, \beta_K$ are regression parameters and $\mathbf{x}_i = (x_{i1}, \ldots, x_{iK})$ is the vector of K covariates giving the characteristics of individual i relevant to the capture (Huggins, 1989; Alho, 1990).

To model unobserved (or latent) heterogeneity, popular choices have been finite (Pledger, 2000) and continuous mixtures (Burnham, 1972; Coull and Agresti, 1999; Dorazio and Andrew Royle, 2003). In particular, finite mixture (or latent-class) models consider the population as divided in two or more classes, each characterised by a different capture probability. Class membership is latent and, thus, the proportion of individuals in each class has to be estimated. These kinds of models seem to be appropriate when the target population indeed contains groups of individuals with different catchability (e.g. juveniles and adults). The fact that a latent-class model fits the data well does not imply that those classes really exists: in general, the grouping may serve as artefact to allow for heterogeneity and correct for the bias in the estimate of the population size (Pledger, 2000). Of course, when there is some reason to postulate the existence of such a population, posterior probabilities of class membership can be useful to get insights about the population's structure.

Since additional variation may exist among individuals within each class, a possible alternative is to consider a different capture probability for each one by using, for instance, the following logistic-normal model (Coull and Agresti, 1999),

$$logit(p_i) = \eta_i \,, \tag{2.3}$$

where $\eta_i \sim N(\mu, \sigma^2)$. Of course, the greater σ, the more heterogeneous are the capture probabilities.

Behavioral response denotes the situation where the detection probability of an individual depends on whether this has been previously detected. This kind of effect is reasonable in live-trapping experiments, where individuals are captured, marked and then released: in fact, animals can react in a positive (*trap-happiness*) or negative (*trap-shyness*) way to the capture and, consequently, be less or more elusive in next capture occasions (Yang and Chao, 2005; Farcomeni, 2011; Alunni Fegatelli and Tardella, 2016).

2.1.2 Multinomial model formulations

The multinomial distribution is fundamental in Capture-Recapture analysis, since it is a natural choice for modelling frequencies of discrete outcomes. Consider the classical problem of estimating the size N of a closed population, whose individuals are sampled with replacement on T occasions. Under the assumption of model M_0, each binary observation y_{it} is the result of a Bernoulli trial with probability of success (namely, capture probability) p. Let

$$y_i = \sum_{t=0}^{T} y_{it}$$

be the capture frequency of the i-th individual in T capture attempts.

By using this data representation, the matrix in Table 2.1 is rewritten as a D-dimensional observed vector of detections, namely $(y_1, \ldots, y_D) = (3, 2, 2, 1, 1, 2, 1)$. By considering the whole population of N individuals, we have that

$$y_i \sim Binom(T,\, p)\,, \quad i = 1, \ldots, N\,.$$

The corresponding likelihood function is thus

$$L(p|y_1, \ldots, y_N, D) = \prod_{i=1}^{N} Binom(y_i|T, p)\,,$$

where, of course, $y_{D+1}, \ldots, y_N = 0$, corresponding to the detection frequency of the $N - D$ uncaptured individuals.

Alternatively, it is possible to group the information contained in $y_1, \ldots, y_N$ by considering the frequency of individuals detected in k out of T sampling occasions,

$$f_k = \sum_{i=1}^{N} \mathbb{1}(y_i = k)\,,$$

for $k = 0, 1, \ldots, T$. The matrix of all the observed capture histories in Table 2.1 can be rewritten as a T-dimensional vector, namely $(f_1, \ldots, f_T) = (3, 3, 1)$, while $f_0 = N - D$ is unknown. Notice that $D = \sum_{k=1}^{T} f_k$.

The joint distribution of $(f_0, f_1, \ldots, f_T)$ is multinomial and the corresponding likelihood function is

$$L(N, p|f_1, \ldots, f_T) = \frac{N!}{(N-D)!\, f_1! \cdots f_T!}\, \pi_0^{N-D}\, \pi_1^{f_1} \cdots \pi_T^{f_T}\,,$$

where $\pi_k = \binom{T}{k} p^k (1-p)^{T-k}$ corresponds to a binomial probability mass function (Sanathanan, 1972). By exploiting the relation between π_k and p, the likelihood for M_0 can be expressed in a more straightforward way, namely

$$L(N, p|f_1, \ldots, f_T) \propto \frac{N!}{(N-D)!}\, p^{\sum_{k=1}^{T} k f_k} (1-p)^{TN - \sum_{k=1}^{T} k f_k}\,.$$

Observing that $\sum_{i=1}^{D} y_i = \sum_{k=1}^{T} k f_k$, the previous likelihood can be easily expressed in terms of the individual frequencies $y_1, \ldots, y_D$, namely

$$L(N, p|y_1, \ldots, y_D) \propto \frac{N!}{(N-D)!}\, p^{\sum_{i=1}^{D} y_i} (1-p)^{TN - \sum_{i=1}^{D} y_i}\,. \tag{2.4}$$

A sketch of the proof is provided in Appendix B.

When the assumption of time-constant capture probabilities cannot hold, a more general way to summarise capture histories is required. A convenient way consists in building encounter history frequencies, by listing all the possible 2^T unique capture histories which can occur in T capture attempts. Let n_k be the number of individuals having a certain capture history k, with $k = 0, 1, \ldots, 2^T - 1$. For convenience, n_0 is the frequency of the unobserved individuals. Table 2.2 summarises the information contained in Table 2.1 in a encounter history frequencies fashion. It is important

history	n_k	cell probability
$(0,0,0)$	?	$(1-p_1)(1-p_2)(1-p_3)$
$(1,0,0)$	0	$p_1(1-p_2)(1-p_3)$
$(0,1,0)$	2	$(1-p_1)p_2(1-p_3)$
$(0,0,1)$	1	$(1-p_1)(1-p_2)p_3$
$(1,1,0)$	0	$p_1p_2(1-p_3)$
$(1,0,1)$	2	$p_1(1-p_2)p_3$
$(0,1,1)$	1	$(1-p_1)p_2p_3$
$(1,1,1)$	1	$p_1p_2p_3$

Table 2.2. Encounter history frequencies derived by the data matrix in Table 2.1. Here, a model with time-varying capture probabilities (i.e. model M_t) is assumed. The parameter p_t is the capture probability at occasion t ($t = 1, 2, 3$).

to notice that this kind of formulation can be used only in absence of individual effects on p; otherwise, pooling all the individuals in groups of unique encounter histories would make information about observed heterogeneity unusable. The vector of encounter history frequencies $(n_0, n_1, \ldots, n_{2^T-1})$ is multinomial with vector of cell probabilities $(\pi_0, \pi_1, \ldots, \pi_{2^T-1})$, where π_k is the probability of observing the k-th encounter history. Therefore, the corresponding likelihood has the following form:

$$L(N, \pi_1, \ldots, \pi_{2^T-1}|n_1, \ldots, n_{2^T-1}) \propto \frac{N!}{(N-D)!} \left(\prod_{k=1}^{2^T-1} \pi_k^{n_k} \right) \pi_0^{N-D}\,,$$

where $D = \sum_{k=1}^{2^T-1} n_k$ and

$$\pi_0 = \left(1 - \sum_{k=1}^{2^T-1} \pi_k\right)$$

is the probability of not encountering an individual during the T attempts.

Notice that the multinomial cell probabilities are function of the parameters that describe the detection process. For example, in the case of model M_t, the multinomial cell probabilities correspond to the probabilities of T independent Bernoulli trials, as shown in the last column of Table 2.2.

2.1.3 Bayesian estimation via PX-DA

In a Bayesian framework, we require prior distributions for the parameters that control the detection process and for the population size. Notably, the population abundance N itself - which is often the main object of inference - determines the dimension of the model parameter space and it can vary within D and infinity, making the Capture-Recapture estimation a challenging task. In fact, if we attempt to fit this model using Bayesian methods, such as Markov chain Monte Carlo (MCMC), each update of N changes the number of parameters in the model because some parameters (e.g. p_i's, in the case of models like (2.3)) may be created or deleted depending on whether N increases or decreases from its previous value. This challenging MCMC problem has been resolved using specialized algorithms such as reversible-jump MCMC (Brooks et al., 2000), regarding the problem as a model selection task (Durban and Elston, 2005). However, these methods appear to be inaccessible to many ecologists since their implementation may require a high level of statistical and computational knowledge.

A quite appealing alternative has been provided by Royle et al. (2007) who fix the dimension of the parameter space by embedding the original dataset into a larger zero-inflated version. Notably, the original dataset $\mathbf{Y}$ containing the capture histories of the D encountered individuals is augmented with $M - D$ rows of zeros (i.e. all-zero histories), where $M \gg N$ is an arbitrarily large upper bound for N. Royle et al. (2007) refer to this approach as parameter-expanded data augmentation (PX-DA, henceforth), since the Capture-Recapture model must be expanded to account for the number of all-zero histories in the augmented dataset $\mathbf{Y}_{\text{aug}}$.

Differently by the reversible-jump MCMC, the PX-DA is not an MCMC algorithm but it is rather a reformulation of the CR model which aims at simplifying the analysis and providing a conventional implementation of MCMC. For this reason, complex CR PX-DA models can be easily fitted with MCMC softwares like JAGS (Plummer et al., 2003), a popular program which allows for analysis of complex hierarchical models using Markov Chain Monte Carlo (MCMC) simulations. JAGS is an implementation of the BUGS language, introduced by Gilks et al. (1994).

Notice that the $N - D$ among the $M - D$ rows of zeros are *sampling zeros*, namely corresponding to individuals who belong to the population but have never been encountered during the whole study. The remaining $M - N$ rows of zeros are *structural zeros*, which correspond to pseudo-individuals who do not belong to the

population. The augmented dataset may be as referred to a *pseudo-population* of M individuals that potentially could belong to the real population of size N.

Let ψ be the probability that an individual in the pseudo-population of size M is a member of the population of N individuals exposed to sampling. Therefore, the parameter ψ can be thought as an *inflation parameter* or *inclusion probability* and it is straightforward to assume that

$$N|\psi \sim Bin(M,\, \psi)\,. \tag{2.5}$$

Royle et al. (2007) shows that the likelihood function for the PX-DA reparameterised version of model M_0 can be written as

$$L(N, p, \psi|y_1, \ldots, y_D) \propto \left[\frac{N!}{(N-D)!}\, p^{\sum_{i=1}^{D} y_i}\, (1-p)^{TN-\sum_{i=1}^{D} y_i}\right]\left[\frac{M!}{N!(M-N)!}\, \psi^N (1-\psi)^{M-N}\right],$$

where the first term is the joint likelihood for p and N in (2.4) and the second term is the probability mass function corresponding to (2.5).

In a Bayesian context, a standard $Unif[0, 1]$ can be assumed for p. Moreover, if $\psi \sim Unif[0, 1]$, then the resulting marginal prior distribution for N is a discrete uniform on $\{1, 2, \ldots, M\}$ (Royle et al., 2007). In effect, under data augmentation, the parameter N is replaced by the parameter ψ.

In order to obtain a convenient parameterisation which allows to easily implement a CR PX-DA model in BUGS, one can introduce a further level in the hierarchy of the model by introducing a set of independent and identically distributed (iid) latent binary variables $z_1, \ldots, z_M$, which indicates whether the i-th individual of the augmented dataset belongs to the real population ($z_i = 1$) or not ($z_i = 0$). We assume that $z_i \overset{iid}{\sim} Bern(\psi)$, $\forall i$. Notice that this set of variables is actually partially observed, since an amount of D individuals has been encountered at the end of the survey: indeed, $z_1 = \cdots = z_D = 1$. Under this formulation, it is straightforward to observe that the true population abundance is a derived parameter, i.e.

$$N = \sum_{i=1}^{M} z_i\,.$$

The observation process of the PX-DA version of model M_0 is, thus, given by

$$y_i \sim Bin(T,\, z_i \cdot p)\,, \quad i = 1, \ldots, M\,,$$

which formally represents a zero-inflated binomial model, since y_i is almost surely equal to 0 when the individual i does not belong to the real population (i.e. when $z_i = 0$). The BUGS code of the PX-DA version of model M_0 is provided in Listing A.1 in Appendix A.

2.2 Open populations

In many long-term Capture-Recapture studies it is not reasonable to assume that the population remains the same for the whole survey, since individuals may be subject to additions and to removals from the population throughout the considered time-frame.

2.2.1 The Jolly-Seber model

The Jolly-Seber (JS, henceforth) model (Jolly, 1965; Seber, 1965) extends the closed-population models by assuming the presence of recruitment, mortality and migration of the individuals. Individuals are thus allowed to enter (i.e. via birth or immigration) and exit (i.e. via death or emigration) the population during the study. Migration cannot be separated from the birth and death processes without additional information; thus, one does not distinguish between sources of new animals or between the ways animals leave the population. For identifiability purposes, the emigration must be considered permanent (once an individual has left the population, it cannot get back into it). All the encounters are assumed to be independent across individuals and along time.

When one considers an open population, the object of the study is a dynamic population which changes over time. Along with the estimate of the population size at some intervals (e.g. a month, a year, etc.), ecologists are often interested in the size of the *super-population*, N_{super}, that is the number of individuals ever alive during all sampling periods: if an individual has been part of the population for at least one sampling period, then it does belong to the super-population (Schwarz and Arnason, 1996). It is reasonable to assume that a fraction of the N_{super} individuals is already into the population at the first capture occasion and the remaining individuals will join the population by the end of the survey.

To model ecological process, JS model requires the definition of two parameters that control the way individuals are added to the population (i.e. recruitment process) and the way individuals remain into the population (i.e. survival process): they are, respectively, the *entry probability* and the *apparent survival probability*.

The entry probability b_t ($t = 1, \ldots, T$) is the probability that an individual is new in the population at time t, that is, that it has entered the population between occasion $t-1$ and t. This parameter can be also interpreted as the probability that a member of the super-population of size N_{super} enters the population at occasion t. Notably, since at the end of the T occasions all individuals in the super-population have been alive for at least once, the entry probabilities $b_1, \ldots, b_T$ must sum to 1. Notice that the first entry probability b_1 has not a clear ecological meaning because it includes all the individuals that were already part of the population before the start of the survey and, thus, that the model *virtually* adds to the population at the first occasion.

The apparent survival probability (survival probability, henceforth) ϕ_t ($t = 2, \ldots, T$) is the probability that a recruited individual who belongs to the population at occasion $t-1$ remains in the population at occasion t. The adjective *apparent*

is included as permanent emigration and mortality are indistinguishable and, thus, treated as the same phenomenon in JS-type models. Marshall et al. (2004) show that, generally, the larger a study area the closer the match between apparent and true survival probability since dispersing individuals tend to have a higher probability to remain in the study area.

Of course, the capture probability p_t is the probability of encountering (i.e. catching or sighting) an individual, provided that the individual is part of the population. The basic JS model assumes that recruitment, survival and capture probabilities remain constant over time and are identical for all individuals; however, we will see in the next chapters that, in practice, one may want to relax some of these assumptions.

2.2.2 Robust design

Pollock (1982)'s *robust design* is a sampling scheme that combines the features of both closed and open population surveys. One assumes a sampling scheme divided in T periods (*primary sampling sessions*) and, for each period $t = 1, \ldots, T$, a number J_t of encounter occasions (*secondary sampling sessions*). The idea is that secondary sampling sessions within the same primary sampling session are close in time, so close that the population can be assumed as closed between those encounter occasions. Each primary sampling session can be thus regarded as a closed detection survey. Between primary sessions the population is considered open, as more time goes by between each period.

Let D be the number of distinct individuals that have been observed at least once during the T sampling periods. In this case, data are collected in a $D \times T$ matrix, $\mathbf{Y}$, where its generic element $y_{it} \in [0, J_t]$ is the detection frequency of individual i at period t. Each row of Y contains the detection history of an individual. It is reasonable to assume that

$$y_{it} \sim Binom(J_t, p_{it}) , \quad \forall i, t .$$

When $J_t = 1$, $\forall t$, the previous model corresponds to the original JS model described in Subsection 2.2.1, which does not consider a hierarchical structure in the sampling design.

By pooling the data of a series of short-term studies, the robust design improves the estimation of the demographic characteristics of the population and allow the estimation of abundances for each primary period. Notably, with these two levels of sampling one aims at minimising the influence of unequal catchability on the population abundance estimates (Pollock et al., 1990). An example of data collected through a robust design is provided in Section 4.3.

2.2.3 PX-DA implementation of the Bayesian JS model

When we deal with open populations, individual encounter histories are the result of the combination between two distinct processes: the detection process and the population dynamics. The former describes how individuals appear in the sample, while the latter decides when individuals can appear in the sample.

Modelling the population dynamics

The population dynamics is controlled through two time-varying latent binary variables: the first one indicates the *recruitability* of the individual into the population, that is

$$r_{it} = \begin{cases} 1 & \text{individual } i \text{ is recruitable at time } t \\ 0 & \text{otherwise} \end{cases} , \forall i, t \,,$$

while the second one indicates the *presence* of the individual in the population, that is

$$z_{it} = \begin{cases} 1 & \text{individual } i \text{ belongs to the population at time } t \\ 0 & \text{otherwise} \end{cases} , \forall i, t \,.$$

All the M individuals are recruitable at the first time (i.e. $r_{i1} = 1,\ i = 1, \ldots, M$), while they become permanently non-recruitable once they have entered the population.

Let ρ_t be the *recruitment probability*, i.e. the probability that an available (not yet entered) individual in the augmented dataset is recruited in the population at time t. This parameter is the product of the probability of entry in the population, given not previously entered, and the inclusion probability. In formulas,

$$\rho_1 = \psi \, b_1 \qquad \text{and} \qquad \rho_t = \psi \, \frac{b_t}{1 - \sum_{i=1}^{t-1} b_i}, \quad t > 1 \,,$$

where b_t is the probability of entry the population at period t and ψ is the inclusion probability. Thus, the PX-DA formalization in the open-population models merely changes the interpretation of the recruitment parameter, which now controls the potential entrance in the real population of the individuals in the pseudo-population.

At the first occasion we have that

$$z_{i1} \sim Bern(\rho_1) \,,$$

while, for $t = 2, \ldots, T$,

$$\begin{aligned} r_{it} &= \min\{r_{i,t-1}, 1 - z_{i,t-1}\} \\ z_{it} \,|\, z_{i,t-1}, r_{it} &\sim Bern(\phi_t \cdot z_{i,t-1} + \rho_t \cdot r_{it}) \end{aligned}$$

for $i = 1, \ldots, M$.

Marginally, this hierarchical specification implies that $N_{super} \sim Binom(M, \psi)$. Royle and Young (2008) show that ψ is linked to the recruitment probabilities through the following equation:

$$\psi = 1 - \prod_{t=1}^{T} (1 - \rho_t),$$

which implies that the conditional expectation of the super-population abundance is

$$\mathbb{E}[N_{super}|M, \rho_1, \ldots, \rho_T] = M\left[1 - \prod_{t=1}^{T}(1-\rho_t)\right] = M\,\psi\,.$$

Hence, the choice of the prior distribution for $\rho_1, \ldots, \rho_T$ is crucial as it determines the marginal prior on N_{super}. Dorazio (2020) demonstrates that the prior specification

$$\rho_t \sim Beta\left(\frac{1}{T}, 2 - \frac{t}{T}\right), \quad t = 1, \ldots, T, \tag{2.6}$$

induces an objective marginal prior on N_{super}.

In the case of constant recruitment, namely $\rho_1 = \cdots = \rho_T = \rho$, (2.6) can be replaced by

$$\rho \sim Beta(1, T)\,,$$

which yields

$$\psi = 1 - (1-\rho)^T \sim Unif[0, 1]\,,$$

and, consequently, a discrete uniform prior distribution for N_{super} (Dorazio, 2020).

In terms of practical inference, the estimated population size at each time t and the overall super-population size can be derived through the latent variables z_{it}'s, namely

$$N_t = \sum_{i=1}^{M} z_{it}$$
$$N_{super} = \sum_{i=1}^{M} \mathbb{1}_{\{\sum_{t=1}^{T} z_{it} > 0\}}\,.$$

An alternative multi-state formulation

The population dynamics described so far can be also seen as a multi-state process, where at each occasion t each individual (or pseudo-individual) is into one and only one state. We first notice that when an individual becomes part of the population, it cannot be recruited any more: this implies that for $t > 1$, r_{it} and z_{it} cannot be simultaneously equal to 1. At period $t > 1$, the i-th individual is into one of the following three states:

1. it has never been part of the population ($r_{it} = 1 \wedge z_{i,t} = 0$);
2. it is part of the population ($r_{it} = 0 \wedge z_{i,t} = 1$);
3. it was part of the population, but now it is not ($r_{it} = 0 \wedge z_{it} = 0$).

In the JS modelling framework, individuals that leave the population cannot get back and hence state 3 is an absorbing state. If one allows temporal heterogeneity, then the transition probability matrix associated to the three states at times $t = 2, \ldots, T$ is

$$\begin{array}{c|ccc} & 1 & 2 & 3 \\ 1 & 1-\rho_t & \rho_t & 0 \\ 2 & 0 & \phi_t & 1-\phi_t \\ 3 & 0 & 0 & 1 \end{array}$$

where rows and columns represent, respectively, the states at time t and $t+1$. At time $t = 1$, all individuals can be recruited in the population. As t increases, more and more individuals enter the population or, equivalently, are *removed* from state 1. All the observed individuals will eventually be recruited in the population before time T, but not all of them will leave it (they may have survived to future, unobserved, periods). Figure 2.1 illustrates the multi-state formulation of the population dynamics.

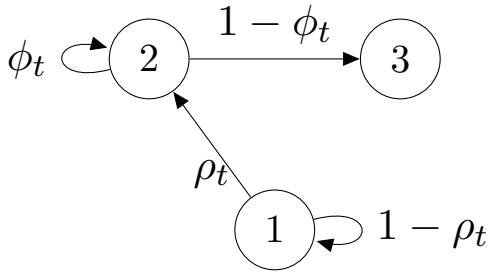


Figure 2.1. Multi-state formulation of the PX-DA Jolly-Seber population dynamics.

Remember that individuals are exposed to capture, with probability p_t, only during their transitory stay in state 2. Hence, also a portion of never observed individuals may have been recruited in the population at some point without never being captured. They represent the unknown part one wants to estimate.

The detection process

As for the case of closed populations (cfr. Subsection 2.1.3), we can express the distribution of the generic element of the augmented data matrix conditionally on z_{it} as

$$y_{it}|z_{it} \sim Bern(p_t \cdot z_{it}). \tag{2.7}$$

Since $y_{it} = 0$ almost surely when $z_{it} = 0$, (2.7) is a zero-inflated binomial model.

A BUGS code of the basic JS-model with a PX-DA reparameterisation is provided in Listing A.2 in Appendix A, where we suppose constant recruitment, survival and detection parameters.

Chapter 3

Modelling latent heterogeneity via finite mixtures

Finite mixtures models are often used in CR analyses to account for latent heterogeneity between individuals (e.g. Norris III and Pollock (1996); Pledger (2000)). However, literature about Bayesian implementation of finite mixture models within the CR framework seems to be scarce: more importantly, common issues like identifiability or prior specification of the component-specific parameters are often disregarded.

This chapter illustrates the use of finite mixture models within the PX-DA Capture-Recapture framework presented in Chapter 2. Notably, we propose and investigate a strategy to properly identify mixture components which allows a practitioner to directly specify suitable prior distributions on the component-specific parameters and avoid label-switching, without adopting post-hoc relabelling schemes. The proposed approach involves order constraint on the component-specific parameters which are employed through a meaningful specification of conditional prior distributions on the parameters. In particular, two alternative classes of prior specifications are introduced and some theoretical results are provided (see Appendix C for the formal proofs). The inferential benefits of the proposed methodology is highlighted in the simulation study, by comparing different alternative prior specifications on component-specific parameters. Finally, an illustration based on real data is presented to show the practical use of the method.

3.1 A general introduction to Finite mixture models

Let $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$ be n realizations of a l-dimensional random vector with distribution $\boldsymbol{Y} \sim f(\cdot \,|\, \boldsymbol{p}, \boldsymbol{w})$, where $\boldsymbol{p} = (p_1, \dots, p_G)$, $\boldsymbol{w} = (w_1, \dots, w_G)$ and

$$f(\boldsymbol{y} \,|\, \boldsymbol{p}, \boldsymbol{w}) = \sum_{g=1}^{G} w_g \cdot h(\boldsymbol{y} \,|\, p_g). \tag{3.1}$$

This is a finite mixture distribution with G components and $w_1, \dots, w_G$ are the *component weights* such that $w_g > 0$, $\forall g$ and $\sum_g w_g = 1$. The probability (density or mass) function $h(\boldsymbol{y} \,|\, p_g)$ is often called g-th *mixture component*. There are mainly to ways in which a mixture model can be used: either for clustering or as a tool to

approximate nonstandard distributions in a parametric model framework.

In the following we will employ finite mixture models as a tool to perform model-based clustering. In this regard, the model in (3.1) can be more efficiently represented through a missing data formulation, by augmenting the space with a latent indicator variable $\zeta \in \{1, \dots, G\}$ that denotes from which of the G densities each observation come from. The prior weight w_g can be interpreted as the prior probability that a unit belongs to the g-th class (or group), namely

$$P(\zeta_i = g) = w_g, \quad i = 1, \dots, n,\ g = 1, \dots, G\,.$$

Mixture weights, thus, represent the relative proportion of observations coming from the G groups.

The idea behind this formulation is that each component is responsible for generating a particular observation (or vector of subsequent observations, in the case of longitudinal data as the ones provided by CR experiments) and the true component label of each unit is unknown: therefore, the goal is to determine which component each individual belongs to.

An appealing advantage of the model-based clustering is that it performs soft classification, that is, once the model has been fitted, each unit is provided with a probability to belong to a given component: this avoids potential biases introduced by a crisp classification, yielding a well-interpretable measure of uncertainty in the classification (McLachlan et al., 2019).

In this context, Bayesian inference is usually pursued through the following hierarchical specification:

$$\begin{aligned}
&\boldsymbol{Y} \,|\, \boldsymbol{p}, \zeta \ \sim\ f(\cdot \,|\, p_\zeta) \\
&\zeta \,|\, \boldsymbol{w} \sim Mult(\boldsymbol{w}) \\
&\boldsymbol{w} \sim Dir(\boldsymbol{\alpha}), \quad \boldsymbol{p} \sim \pi_{\boldsymbol{p}}(\cdot)
\end{aligned}$$

where $Mult(\cdot)$ is the multinomial distribution, $Dir(\boldsymbol{\alpha})$ the Dirichlet distribution with hyperparameter vector $\boldsymbol{\alpha}$, and $\pi_{\boldsymbol{p}}(\cdot)$ is the joint prior on the class-specific probabilities $p_1, \dots, p_G$.

3.2 Finite mixtures in CR experiments

Detection heterogeneity among animals of the same population can exist due to many different reasons, like heterogeneous sampling effort, unbalanced observer skills or variation in the individual behaviour. Ideally, covariates could be measured and incorporated in ecological models to account for detection heterogeneity. However, latent heterogeneity may still remain or measuring the relevant covariates may simply be impossible. When unexplained heterogeneity exists, it can be accommodated using finite mixtures in which discrete latent variables are used to assign individuals to mixture components (i.e. homogeneous groups), each characterised by group-specific parameters (Pledger, 2000; Royle, 2006; Louvrier et al., 2018). In simulation studies, finite mixtures have appeared to be successful in decreasing bias in abundance estimates that was introduced by heterogeneity in the detection

process (e.g. Pledger (2005)).

Finite mixture models have been successfully employed both in closed-population (Norris III and Pollock, 1996; Pledger, 2000; Dorazio and Andrew Royle, 2003; Pledger, 2005) and open-population (Pledger et al., 2003, 2010; Guéry et al., 2017) studies to model unobserved heterogeneity in detection and survival between different classes of individuals. The main contributions have been made from a frequentist perspective. In particular, Pledger (2000) proposes a unified theory for closed populations using finite mixture models.

A Bayesian CR analysis which allows for both heterogeneity of detection probabilities between animals and trap effects through discrete finite mixtures have been conducted by Ghosh and Norris (2005), but restricted to closed populations. More importantly, practitioners who deals with Bayesian finite mixture models in the context of CR data may have to cope with several challenges, ranging from non-identifiability issues to a suitable definition of the priors distribution on the model's parameters. In this regard, we feel that literature on how to implement these types of models in a Bayesian framework is scarce.

Pledger et al. (2003) use finite mixture models as an artifact to capture heterogeneity in the detection and survival processes, by discouraging the use of mixture models to bring evidence of the true existence of classes. On the other hand, model-based clustering may be worth to be considered as a useful tool to identify classes of individual histories sharing similar profiles when one has reasonable belief that a particular structure in the target population may exist (Cubaynes et al., 2012; Guéry et al., 2017).

3.2.1 A Bayesian model for closed populations

We consider the component-specific parameters' prior specification in the context of Capture-Recapture methods. In particular, we embed finite mixture models within the PX-DA formalization of the CR closed-population model illustrated in Subsection 2.1.3 by adding one layer of hierarchy in the original hierarchical specification.

Let y_i be the capture frequency of the i-th unit across the T occasions, for $i = 1, \dots, N$, and let D be the total number of distinct *observed* individuals (i.e. captured at least once). The unknown population size is N and we bound its parameter space with a large (at will) upper bound M and augment the data sample with null capture frequencies $(y_{D+1}, \dots, y_M) = (0, \dots, 0)$.

With the purpose of modelling the capture heterogeneity between individuals, we consider finite mixture of G binomial distributions (Pledger, 2000) for the counts y_i's, that is

$$y_i | z_i \sim \sum_{g=1}^{G} w_g \cdot Bin(T, p_g \cdot z_i), \quad i = 1, \dots, M,$$

where w_g is the probability that individual i belongs to the g-th mixture component and p_g is the capture probability of the g-th component. Notice that $y_i = 0$ almost surely when $z_i = 0$, so that the previous model corresponds to a finite mixture of zero-inflated binomial distributions.

The Bayesian estimation of this kind of model can be performed via JAGS. The BUGS code to fit the model in JAGS is presented in Listing A.3 in Appendix A, where we consider the case of $G = 2$ mixture components and standard Uniform priors for probability parameters.

3.3 Challenges in fitting Bayesian finite mixtures

The implementation of Capture-Recapture methods in a Bayesian context involves the specification of distributions to model the prior beliefs on the capture (or survival) probabilities of the individuals. The Beta distribution is a very common candidate as a prior for such parameters: it is compactly supported on $[0, 1]$. It can be shaped to represent non-informative or informative settings by manipulating its two shape parameters. When dealing with finite mixture models, a suitable Beta prior distribution should be placed on each component-specific parameter, p_g, $g = 1, \ldots, G$, where G is the number of components of the mixture model.

The naive application of finite mixture models within the Bayesian machinery is affected by the so-called *label-switching* problem, due to the fact that the likelihood function is invariant under permutation of the indices of the components: at each Markov Chain Monte Carlo (MCMC) step, the groups may interchange their relative role (Jasra et al., 2005). An example is provided in Figure 3.1, where a 2-component finite mixture model is considered. A review of the label switching issue and a wide

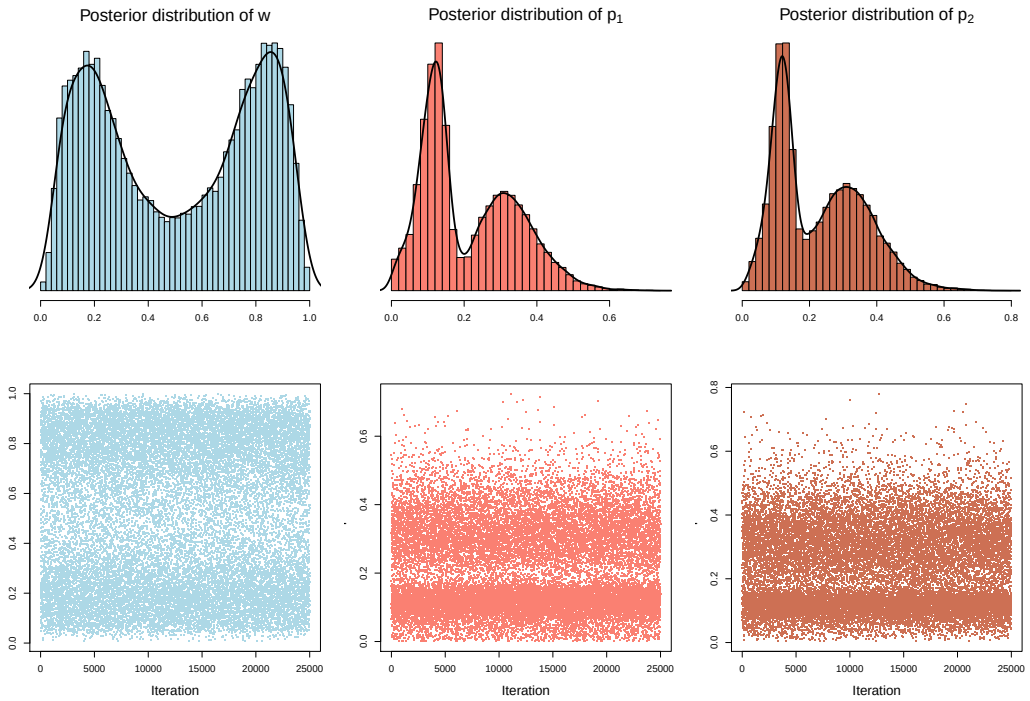


Figure 3.1. Example of posterior distributions of the component-specific parameters of a 2-component mixture model. Posterior distributions have been estimated via MCMC.

range of potential solutions can be found in Stephens (2000) and Jasra et al. (2005).

A first effective solution to this issue is to uniquely identify the components by including prior information about their marginal and relative behaviour, as by imposing ordering constraints that define the group-specific parameters' positions in a hierarchy (Diebolt and Robert, 1994). An increasingly popular approach is to employ a relabelling scheme, such as that proposed by Celeux (1998), where the posterior samples of the parameter of interest are clustered according to a k-means algorithm. This method converges to local minima, so the results based on multiple starting points are compared to identify the optimal solution. A possible extension have been proposed by Marin et al. (2005). Another strategy is to use label invariant loss functions (e.g. Hurn et al. (2003)), although this approach can incur a high computational cost for mixtures with many components and rapidly become impractical.

In the following, we investigate the first type of solution (i.e. the one which exploits order constraints) and we focus on a finite mixture model with $G = 2$ components, whose density is given by

$$f(\boldsymbol{y}|p_1, p_2, w) = w \cdot h(\boldsymbol{y}|p_1) + (1 - w) \cdot h(\boldsymbol{y}|p_2)\,.$$

In order to avoid the label-switching problem, the prior on $\boldsymbol{p}$ shall envision an ordering constraint to solve the symmetry in the posterior distribution and make the two components identifiable. For instance, we could enforce $p_1 < p_2$ by specifying the joint prior in a conditional fashion:

$$\pi_{\boldsymbol{p}}(p_1, p_2) = \pi_{p_1}(p_1) \cdot \pi_{p_2\,|\,p_1}(p_2), \tag{3.2}$$

where $\pi_{p_2\,|\,p_1}(\cdot)$ guarantees $p_2 > p_1$, e.g. has a support depending on p_1.

When p_1 and p_2 are probabilities (i.e. $\in (0, 1)$), the most straightforward solution is to have a Beta prior on p_1 and then specify $\pi_{p_2\,|\,p_1}(\cdot) = Unif\,(\cdot\,|\,p_1, 1)$. In practice, any other specification of $\pi_{p_2\,|\,p_1}(\cdot)$ complying with the ordering constraint could be valid. Nevertheless, it is rather uncommon to find literature that explores the marginal distribution induced on $p_2 \sim \pi_{p_2}(\cdot)$ obtained by integrating p_1 out from (3.2). This inevitably jeopardizes the possibility of eliciting informative priors whenever the information is available.

For example, in the context of Capture-Recapture experiments, Turek et al. (2021) propose to impose order constraints on the component-specific parameters, simply by using the specification $p_1 \sim Unif[0, 1]$ and $p_g \sim Unif[p_{g-1}, 1]$, for $g = 2, \ldots, G$, but no mention is made about the consequence of this choice on the resulting marginal prior distributions of the component-specific parameters. An example is provided by Figure 3.2, which considers this prior specification of component parameters when $G = 2$: such kind of truncation may indeed radically modify the prior modelling, by inducing a prior on p_2 which places most of its density close to the boundary 1. This may occasionally turn out to be undesired and beyond the real intention of the practitioner.

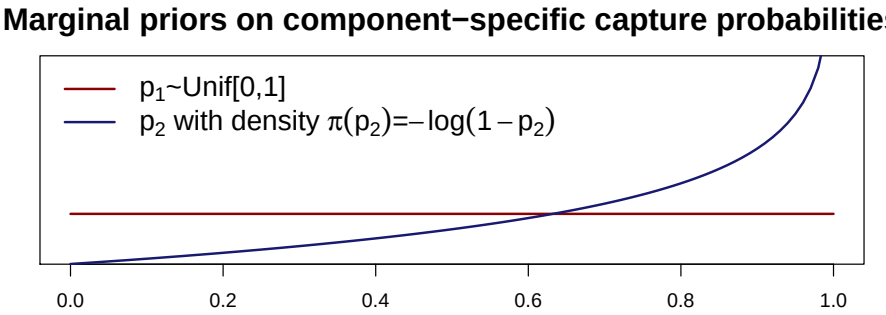


Figure 3.2. Marginal prior induced on the parameter p_2 of a 2-component mixture model when $p_1 \sim Unif[0,1]$ and $p_2|p_1 \sim Unif[p_1,1]$.

3.4 Flexible prior specifications on class-specific probability parameters

We investigate some alternative specifications that allow for controlling the marginal expected value and variance, and possibly also the form of the two marginal prior distributions of the component-specific parameters. These tools can be exploited to embed prior information in the estimation process, which can, in turn, favour the proper identification of the two components. This extends the work sketched in Alaimo Di Loro et al. (2022).

3.4.1 The Beta and truncated Beta

It is possible to derive the marginal prior distribution of p_2, when

$$\begin{aligned} p_1 &\sim Beta(\alpha_1, \beta_1) \\ p_2|p_1 &\sim tBeta(\alpha_2, \beta_2, p_1, 1) \end{aligned}$$

where $tBeta(\cdot,\cdot,l,u)$ denotes the *truncated Beta* in (l,u), whose probability density function is given by the following expression:

$$tBeta(p_2 \mid \alpha_2, \beta_2, p_1, 1) = \frac{1}{B(\alpha_2,\beta_2)} \frac{p_2^{\alpha_2-1}(1-p_2)^{\beta_2-1}}{1-F_{Beta(\alpha_2,\beta_2)}(p_1)} \mathbb{1}_{\{p_1,1\}}(p_2)\,.$$

However, the expression of the resulting marginal $\pi_{p_2}(\cdot)$ is not trivial unless we set $\alpha_2 = 1$, for which we obtain:

$$\pi_{p_2}(p_2) = \frac{B(\alpha_1, \beta_1-\beta_2)}{B(\alpha_1,\beta_1)} \beta_2 (1-p_2)^{\beta_2-1} F_{Beta(\alpha_1,\beta_1-\beta_2)}(p_2)\,,$$

with $\beta_1 > \beta_2$.

In particular, when $\alpha_1 = \beta_2 = k$ and $\beta_1 = k+1$, i.e.

$$\begin{aligned} p_1 &\sim Beta(k, k+1) \\ p_2|p_1 &\sim tBeta(1, k, p_1, 1)\,, \end{aligned} \tag{3.3}$$

then we obtain the following convenient result for the marginal prior distribution of p_2:

$$p_2 \sim Beta(k+1, k), \quad k > 0.$$

Notice that, in this specific case, the marginal distribution induced on p_2 is symmetrical with respect to the distribution of p_1 around the vertical line $p^* = 0.5$; equivalently, we can say that $p_2 \overset{d}{=} 1 - p_1$. Formal proof of the previous results is provided in Appendix C (Theorem 1). Examples of prior distributions resulting from this kind of specification are provided in Figure 3.3, where the four different couple of marginal density functions of p_1 and p_2 were derived by choosing different values for k.

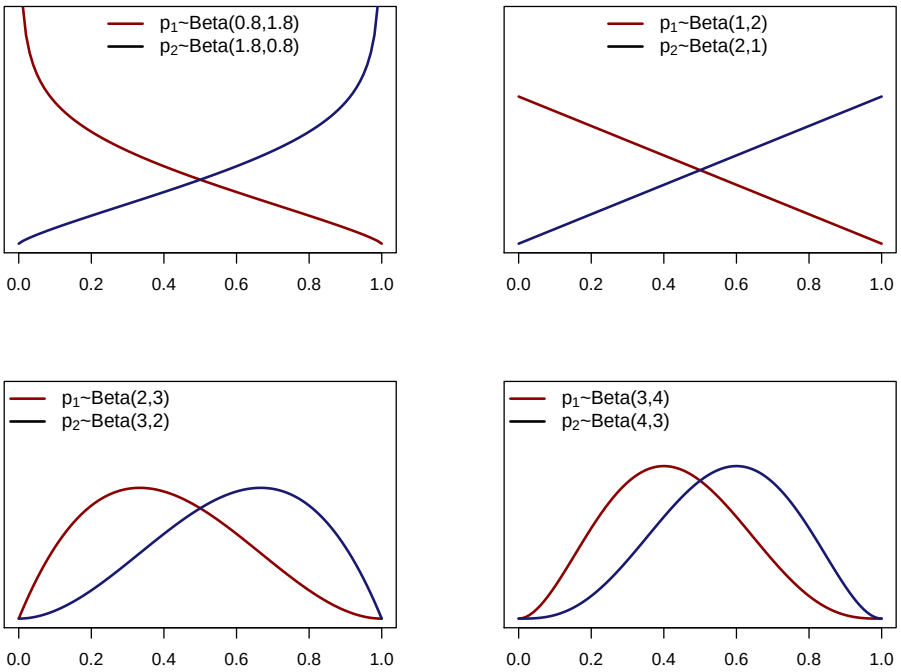


Figure 3.3. Marginal prior densities for component-specific parameters p_1 and p_2 resulting from $p_1 \sim Beta(k, k+1)$ and $p_2|p_1 \sim tBeta(1, k, p_1, 1)$.

This kind of prior specification may turn out to be useful as regularization tool which helps in separating the distributions of the couple of component-specific parameters. In this sense, the hyperparameter k may be seen as a measure of the degree with which the two prior distributions are separated: the smaller is k the more *repulsed* are the two priors.

3.4.2 The Beta and restricted Beta

The *restricted Beta* (also called *4-parameters Beta*) is a Beta random variable which has been shifted and scaled to lie on a different domain (l, u). In particular, if $X \sim Beta(\alpha, \beta)$ and $Z = l + X \cdot (u - l)$, then $Z \sim rBeta(\alpha, \beta, l, u)$, whose probability density function is as follows:

$$rBeta\left(z \,|\, \alpha, \beta, l, u\right) = \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{(z-l)^{\alpha-1} \cdot (u-z)^{\beta-1}}{(u-l)^{\alpha+\beta-1}} \mathbb{1}_{\{l,u\}}(z).$$

The expected value and variance of Z are

$$\begin{aligned} \mathbb{E}\left[Z\right] &= \frac{\alpha}{\alpha+\beta} \cdot (u-l) + l = \frac{u\alpha + l\beta}{\alpha+\beta}, \\ \mathbb{V}\left[Z\right] &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \cdot (u-l)^2 \end{aligned}$$

In this context, the restricted Beta can be exploited to specify a compelling joint prior for p_1 and p_2, such that $p_1 < p_2$.

If we consider

$$\begin{aligned} p_1 &\sim Beta(\alpha_1, \beta_1) \\ p_2 \,|\, p_1 &\sim rBeta\left(\alpha_2, \beta_2, p_1, 1\right) \end{aligned}$$

then the corresponding marginal expected value and variance of p_2 can be derived as a function of mean and variance of p_1 using the *Law of iterated expectations*:

$$\begin{aligned} \mathbb{E}\left[p_2\right] &= \frac{\alpha_2}{\alpha_2+\beta_2} + \mu_1 \frac{\beta_2}{\alpha_2+\beta_2} \\ \mathbb{V}\left[p_2\right] &= \sigma_1^2 \cdot \frac{\beta_2^2}{(\alpha_2+\beta_2)^2} \left(1 + \frac{\alpha_2}{\beta_2(\alpha_2+\beta_2+1)}\right) \\ &\quad + \left(1-\mu_1\right)^2 \cdot \frac{\alpha_2\beta_2}{(\alpha_2+\beta_2)^2(\alpha_2+\beta_2+1)} \end{aligned} \tag{3.4}$$

where $\mu_1 = \mathbb{E}[p_1]$ and $\sigma_1^2 = \mathbb{V}[p_1]$. Thus, once chosen α_1 and β_1 to comply with some prior information on μ_1 and σ_1^2, we can elicit the values of α_2 and β_2 to respect prior knowledge on μ_2 and σ_2^2 simply by solving a linear system based on Eq. (3.4).

Examples of marginal prior distributions resulting from this kind of specification are provided in Figure 3.4. Formal proof is provided in Appendix C (Theorem 2).

A convenient parameter setting

A very convenient parameter setting turns up if we fix $\alpha_2 = 1$ and let $\beta_1 = \beta + 1$ and $\beta_2 = \beta$. Indeed, if

$$\begin{aligned} p_1 &\sim Beta(\alpha_1, \beta+1) \\ p_2 \,|\, p_1 &\sim rBeta\left(1, \beta, p_1, 1\right) \,, \end{aligned}$$

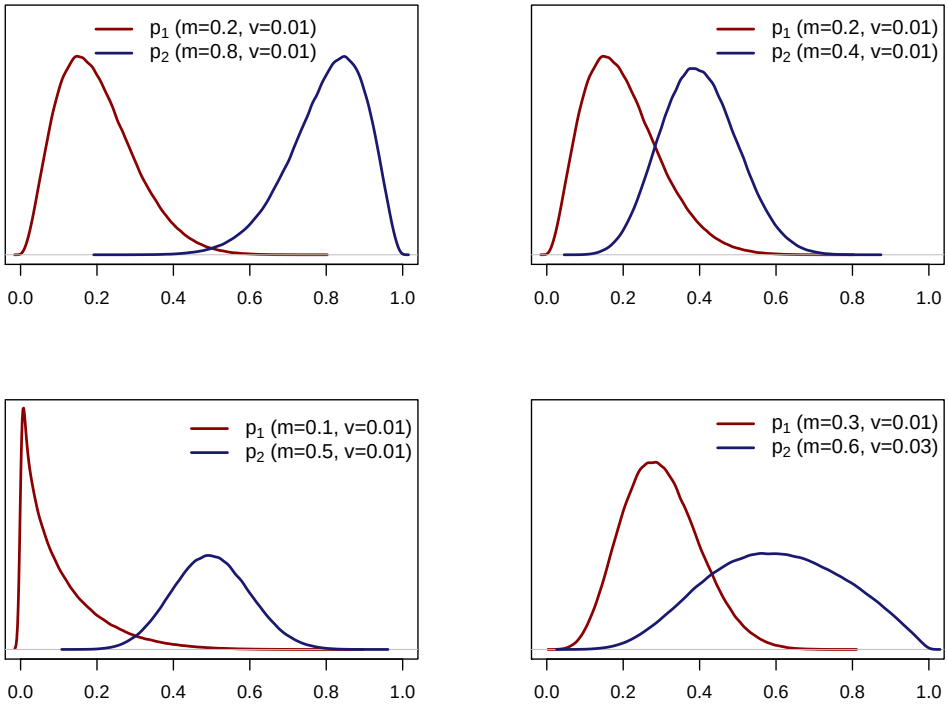


Figure 3.4. Marginal prior densities for component-specific parameters p_1 and p_2 resulting from $p_1 \sim Beta(\alpha_1, \beta_1)$ and $p_2|p_1 \sim rBeta(\alpha_2, \beta_2, p_1, 1)$.

then the joint distribution for (p_1, p_2) is

$$p_{p_1,p_2}(p_1, p_2) = \frac{(\alpha_1 + \beta)\Gamma(\alpha_1 + \beta)}{\Gamma(\alpha_1)\Gamma(\beta)} \cdot p_1^{\alpha_1 - 1}(1 - p_2)^{\beta},$$

and the marginal $\pi_{p_2}(\cdot)$ is a Beta density; more specifically,

$$\pi_{p_2}(p_2) = Beta(p_2 \,|\, \alpha_1 + 1, \beta)\,.$$

The relevance of this result lies on the fact that, taking a Beta prior on p_1 and a particular Beta on p_2 restricted to the set $[p_1, 1]$, the resulting marginal prior for p_2 is still a Beta distribution.

3.5 Simulation experiment

This section shows comparative merits of alternative prior specifications in dealing with label switching and exploiting prior information on group-specific parameters.

We simulate $K = 50$ alternative Capture-Recapture datasets from closed populations with two groups (cfr. the model illustrated in Subsection 3.2.1). The simulation scheme mimics results which can be obtained from real data applications, such as the one that will be illustrated in Section 3.6.

We repeat the same experiments for an increasing number of occasions, i.e. $T = \{5, 10, 15, 20\}$, with the aim of assessing the results as the amount of information increases. Captures are generated by sampling individuals from a *super-population* of size $M = 500$, with a probability for a pseudo-individual to be included in the actual population of $\psi = 0.3$, for all datasets. This yields an expected value of $\mathbb{E}[N_k] = 150$ across all datasets.

Individuals are allocated to the two groups according to $w = 0.85$, where w is the weight of the first component. The group-specific capture probabilities are set to $p_1 = 0.05$ and $p_2 = 0.20$. Under the closure hypothesis, the capture histories of each dataset can be collapsed into the vector of individual overall capture frequencies $\boldsymbol{y}_k = (y_1, \ldots, y_{D_k})$, $k = 1, \ldots, K$, where D_k is the number of distinct observed individuals of the simulated set k.

The following seven prior specifications on the group-specific capture probabilities are considered, which correspond to those represented in Figure 3.5:

I) $p_1 \sim Unif[0, 1]$ and $p_2 \sim Unif[0, 1]$;

II) $p_1 \sim Beta(1.091,\ 1.230)$ and $p_2 \sim Beta(1.230,\ 1.091)$, with means $\mathbb{E}[p_1] = 0.47$ and $\mathbb{E}[p_2] = 0.53$ and variance $\mathbb{V}(p_1) = \mathbb{V}(p_2) = 0.075$;

III) $p_1 \sim Beta(1.70,\ 15.30)$ and $p_2 \sim Beta(2.39,\ 6.46)$, with modes 0.05 and 0.20, respectively, and variances $\mathbb{V}(p_1) = 0.005$ and $\mathbb{V}(p_2) = 0.02$;

IV) $p_1 \sim Unif[0, 1]$ and $p_2|p_1 \sim Unif[p_1, 1]$;

V) $p_1 \sim Beta(1.091,\ 1.230)$ and $p_2|p_1 \sim rBeta(0.139, 1.091, p_1, 1)$, inducing a marginal prior distribution on p_2 with the same mean and variance of the one considered in setting II;

VI) $p_1 \sim Beta(1.222,\ 4.889)$ and $p_2|p_1 \sim rBeta(1.295, 11.653, p_1, 1)$, inducing a marginal prior distribution on p_2 with the same mode and variance of the one considered in setting III;

VII) $p_1 \sim Beta(1, 2)$ and $p_2|p_1 \sim tBeta(1, 1, p_1, 1)$ which induces a marginal prior $p_2 \sim Beta(2, 1)$ favoring a slight repulsion between the two components.

Notice that the identifiability constraint $p1 < p2$ - commonly adopted for finite mixture modelling - is not considered for specifications I, II and III.

In particular, setting I consists in two priors which are both uniform on the interval $[0, 1]$ and can be seen as a default choice; one could be tempted to compare this naive prior setting with the one resulting from setting IV: however, in the latter case, p_1 is still $Unif[0, 1]$ while the marginal prior induced on p_2 by the constraint is a distribution which gives support to higher values: 85% of its density is placed on values greater than 0.5, while 20% on values greater than 0.95.

Setting II and setting V both consider marginal priors on p_1 and p_2 which tend to be quite vague, reflecting a lack of information *a priori* on the true parameter values. Though the two settings results similar, marginal prior distributions resulting from setting V are obtained by respecting the constraint $p_1 < p_2$; this should encourage the identifiability of the two components.

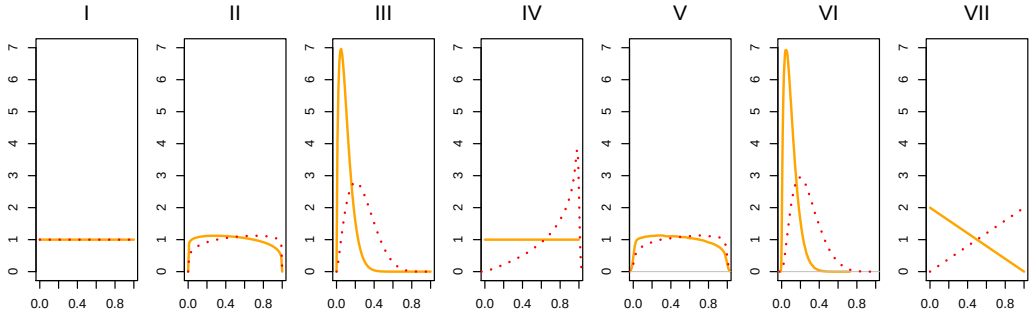


Figure 3.5. Marginal prior distributions on the two component-specific capture probabilities according to the seven considered settings: I) $p_1 \sim Unif[0,1]$ and $p_2 \sim Unif[0,1]$; II) $p_1 \sim Beta(1.091, 1.230)$ and $p_2 \sim Beta(1.230, 1.091)$; III) $p_1 \sim Beta(1.70, 15.30)$ and $p_2 \sim Beta(2.39, 6.46)$; IV) $p_1 \sim Unif[0,1]$ and $p_2|p_1 \sim Unif[p_1,1]$; V) $p_1 \sim Beta(1.091, 1.230)$ and $p_2|p_1 \sim rBeta(0.139, 1.091, p_1, 1)$; VI) $p_1 \sim Beta(1.222, 4.889)$ and $p_2|p_1 \sim rBeta(1.295, 11.653, p_1, 1)$; VII) $p_1 \sim Beta(1,2)$ and $p_2|p_1 \sim tBeta(1,1,p_1,1)$.

Similarly, setting III and setting VI both lead to marginal prior distributions which are centered around the true values of the parameters (of course, this information is available only when the data are indeed simulated), but setting VI involves the prior constraint.

Finally, setting VII may be seen as a choice who merely favours a repulsion between the two components, which can turn out to be useful in a situation where a regularization tool may be required.

The model illustrated in Sec. 3.2.1 is then estimated in JAGS separately on each dataset and, *mutatis mutandis*, for each prior specification. We run four parallel chains from different random starting points; notably, capture probabilities are initialised such that the value provided for p_1 is always smaller than the one provided for p_2. For each chain, the MCMC algorithm is ran for $20,000$ iterations, discarding the first $15,000$ as burn-in and thinning by 4 the remainder to save storage space (Brooks et al., 2004).

Estimated posterior distributions for each parameter have been obtained by merging all the four parallel chains. However, it should be pointed out that Gelman-Rubin's $\hat{R}$ convergence statistics is occasionally larger than 3 for prior settings where order constraints have not been used and always smaller than 1.01 in the other considered prior settings: high values of $\hat{R}$ in this context are ascribable to the presence of label-switching, an issue which appears to be fixed when ordering constraints are placed.

Posterior samples of the capture probabilities p_1 and p_2 corresponding to different datasets have been merged (by prior setting) to analyze their overall behavior: in particular, for each prior specification and for all the considered number of capture occasions, this merge allows to obtain an average of the marginal posterior distributions obtained with the $K = 50$ different samples. Figure 3.6 shows the

resulting average estimated posterior distributions, highlighting how they seem to fail to be correctly identified when an ordering constraint between p_1 and p_2 is not considered (cfr. settings I-III). As T increases, information provided by the

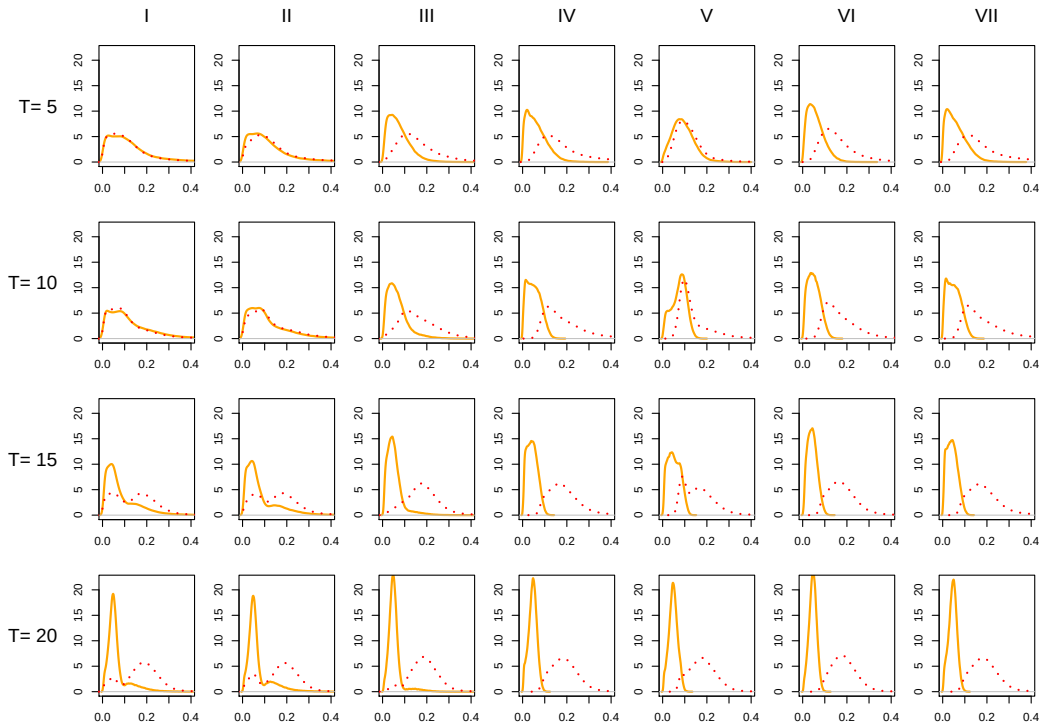


Figure 3.6. Marginal estimated posterior densities for p_1 (*solid orange line*) and p_2 (*dotted red line*), for several number of capture occasions ($T = 5, 10, 15, 20$) and for several prior settings (I-VII).

likelihood tends to prevail over the prior specifications and the estimated posterior distributions tend to set around the true values of the parameters; however, slight multimodality persists in settings I-III. It is interesting to see that prior settings which introduce the same kind of prior information on the capture probabilities (i.e. II vs V and III vs VI) lead to substantially different component-specific posterior distributions, above all when T is small and the prior contribution is more relevant to the inference: in those cases, posterior densities corresponding to priors with constraints show a more pronounced separation.

Additionally, the *overlapping index* (OV, Pastore (2018); Pastore and Calcagnì (2019)) between estimated posterior distributions on p_1 and p_2 is shown in Figure 3.7: this measure gives an estimate of the area shared by two densities and, in this context, it is used to measure the degree of separability between the two posteriors. If $OV = 0$, the two distributions are completely separated, while if $OV = 1$, the two distributions are identical. Notably, while for constrained prior settings (IV-VII) the overlapping index tends to vanish in most of the datasets as T increases, it keeps persisting for higher T in settings where constraints have not been used (i.e. I-III). Notice that marginal priors on p_1 and p_2 corresponding to setting II and V have the same high overlapping index between priors, but overlapping between posteriors

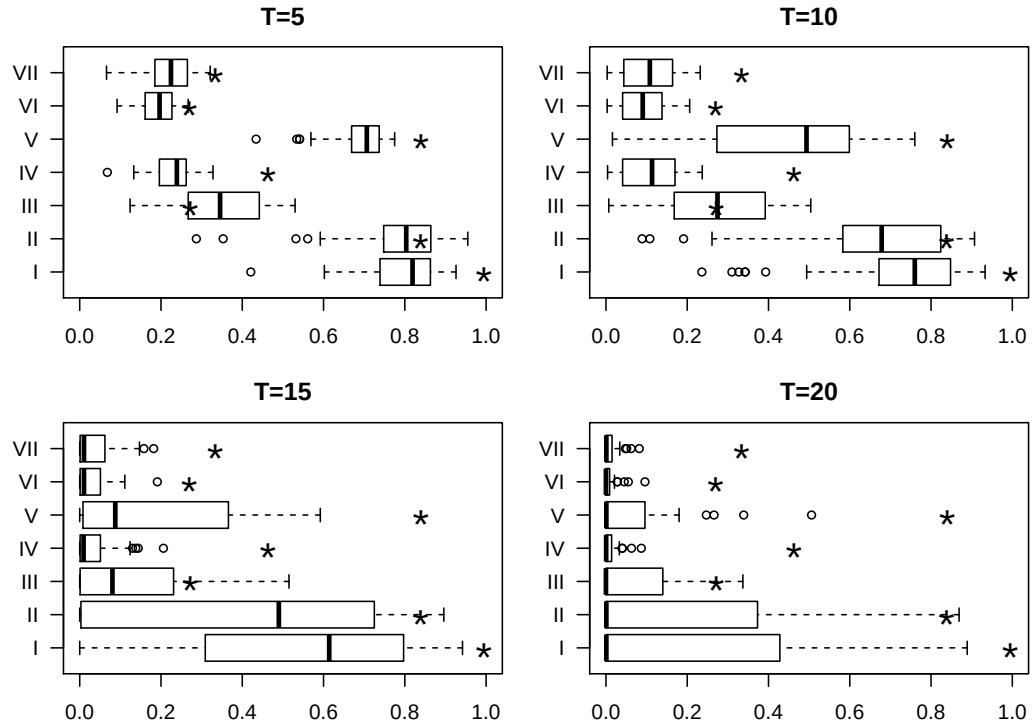


Figure 3.7. Boxplots of the overlapping index between the estimated posteriors of p_1 and p_2 for the $K = 50$ replicas and for each prior setting. The four panels represent scenarios with different number of sampling occasions (i.e. $T = 5, 10, 15, 20$). Overlapping index between component-specific marginal prior distributions is indicated with an asterisk.

tends to decrease more rapidly with T for setting V. The same happens for settings III and VI, where the latter leads to a larger separation between the posteriors of the two component-specific parameters.

For different number of capture occasions T, the overall accuracy on estimating the true component-specific capture probabilities of the different prior settings has been compared in terms of *Mean Absolute Error* (MAE). The results are provided in Table 3.1, where setting VI seems to perform slightly better on average than its direct competitor (i.e. setting III) and among the other constrained settings. As expected, settings I and II both correspond to the worst performances in estimating p_1 and p_2.

Posterior inference on the population abundance N shows a general robustness against different prior specifications on the component-specific capture probabilities, as shown by the distributions of estimated errors in Figure 3.8. This was somehow expected since - unlike the component-specific parameters - the population abundance parameter does not suffer identifiability issues under these mixture model assumptions. In addition, setting V seems to lead to a little underestimate of the true abundance for smaller values of T, but which tends to vanish as T increases. Additionally, distributions of estimated errors are also provided for p_1 (Figure 3.9) and for p_2 (Figure 3.10), by showing - *ceteris paribus* - the better performances of the proposed class of Beta prior specifications over the classical unconstrained Beta

Parameter	T	Prior setting						
		I	**II**	**III**	**IV**	**V**	**VI**	**VII**
p_1	5	0.058	0.054	0.017	0.018	0.045	0.012	0.018
	10	0.059	0.048	0.013	0.010	0.029	0.008	0.009
	15	0.027	0.030	0.012	0.012	0.017	0.010	0.013
	20	0.017	0.020	0.008	0.010	0.009	0.008	0.009
p_2	5	0.097	0.087	0.063	0.062	0.091	0.060	0.059
	10	0.104	0.093	0.058	0.058	0.079	0.054	0.060
	15	0.072	0.071	0.042	0.046	0.057	0.041	0.046
	20	0.055	0.058	0.037	0.039	0.043	0.036	0.039

Table 3.1. MAEs associated to the estimates of p_1 and p_2 for each prior setting and number of capture occasions. Posterior medians have been used as point estimates.

priors. Reasonably, such gain is more evident for smaller values of T.

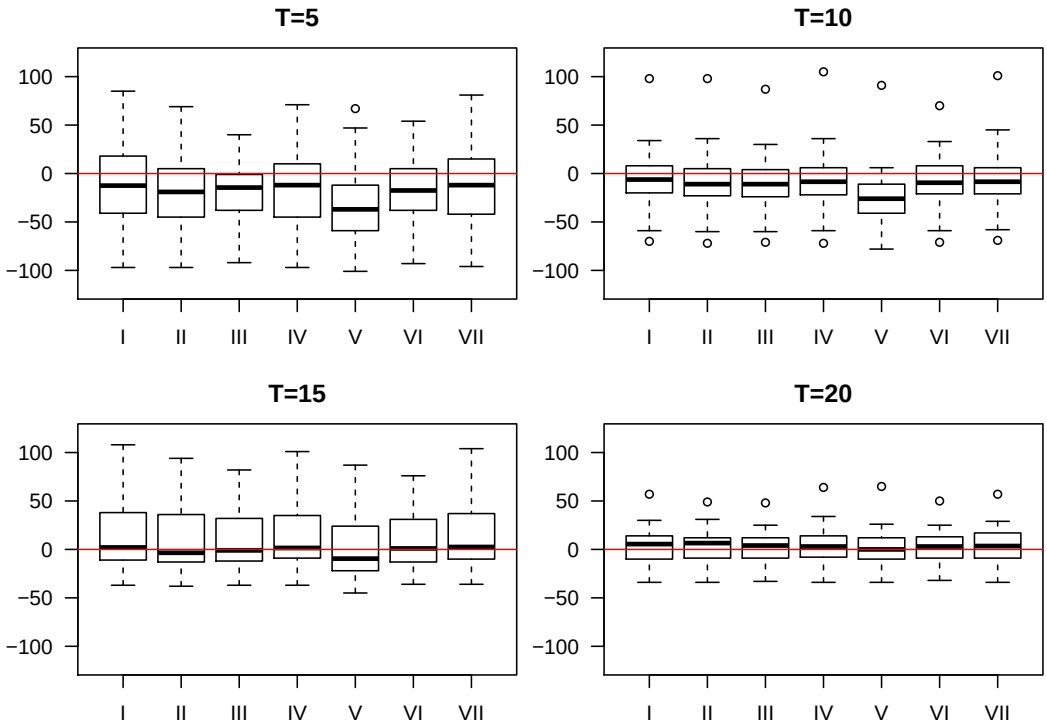


Figure 3.8. Boxplots of the errors (posterior median - true value) in estimating N for the $K = 50$ replicas and for each prior setting. The four panels represent scenarios with different number of sampling occasions (i.e. $T = 5, 10, 15, 20$).

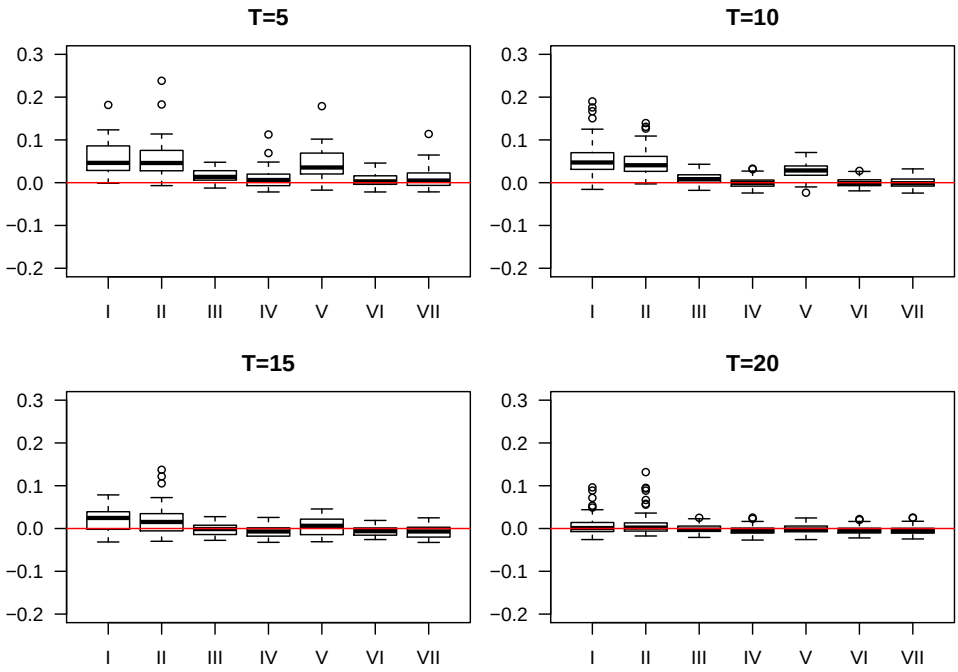


Figure 3.9. Boxplots of the errors (posterior median - true value) in estimating p_1 for the $K = 50$ replicas and for each prior setting. The four panels represent scenarios with different number of sampling occasions (i.e. $T = 5, 10, 15, 20$).

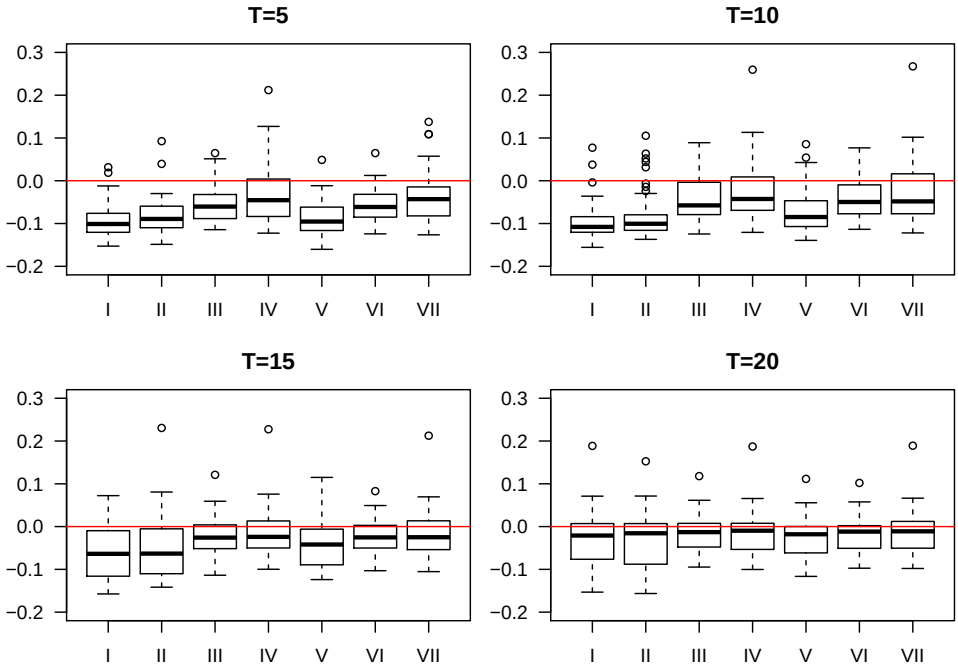


Figure 3.10. Boxplots of the errors (posterior median - true value) in estimating p_2 for the $K = 50$ replicas and for each prior setting. The four panels represent scenarios with different number of sampling occasions (i.e. $T = 5, 10, 15, 20$).

3.6 Real data illustration: *Giant Day Geckos*

We consider giant day gecko's capture histories recorded in Masoala rain-forest exhibit at the Zurich Zoo (Wanger et al., 2009), where $D = 68$ individuals were encountered during $T = 30$ capture occasions.

Data have been collected through a photographic capture method: since it is harder to spot and photograph juvenile compared to adult geckos, at least two groups of individuals with different capture probabilities are expected. Notably, Wanger et al. (2009) found that age indeed has an almost significant effect on capture probability.

Here, we are mainly interested in illustrating our methodology in absence of additional information that may be used as covariate. Thus, we only make use of the individuals' capture frequencies to make inference on the component-specific capture probabilities and on the population abundance.

We fit the CR model for closed population illustrated in Subsection 3.2.1 by using JAGS. We fix $M = 500$: this value allows to widely explore the space of N, yet avoiding a cumbersome computational burden. We consider the same prior settings presented in Section 3.5 and shown in Figure 3.5, with the exception of settings III and VI which were only suitable in the simulation framework where the ground truth was indeed available. We run five parallel chains for each setting. For each chain, the MCMC algorithm is ran for $20,000$ iterations, discarding the first $15,000$ as burn-in. Thinning is not necessary in this context.

Table 3.2 shows point and interval estimates of the main parameters of interest. Population abundance estimates are in line with those found by Alunni Fegatelli and Tardella (2016)), though a quite different model is considered. Notably, point

Prior Setting	$\hat{N}$	$\hat{p}_1$	$\hat{p}_2$
I	111 (77, 165)	0.06 (0.01, 0.20)	0.15 (0.01, 0.23)
II	108 (78, 156)	0.06 (0.01, 0.20)	0.15 (0.02, 0.23)
IV	110 (76, 161)	0.03 (0.01, 0.05)	0.18 (0.13, 0.24)
V	108 (77, 152)	0.03 (0.01, 0.05)	0.18 (0.12, 0.23)
VII	112 (75, 166)	0.03 (0.01, 0.05)	0.18 (0.12, 0.24)

Table 3.2. Posterior means and 95% credible intervals for the main model's parameters of interest for the five considered prior settings.

estimates on N range between 108 and 112 - according to different prior settings -, with the smaller lower bound of the 95% credible interval being 75 and the largest upper bound of the interval being 166 (both correspond to the prior setting VII, which is thus the one associated with the largest credible interval width on N). Point and interval estimates differ for prior settings without (I and II) and with order constraint (IV, V and VII): the former present much larger credible intervals, presumably because of the lack of identifiability due to label-switching.

All the prior settings which involves constraints (i.e. IV, V and VII) agree in assigning 78% of the observed individuals to the group of more elusive geckos and

the rest to the group of less elusive ones, according to the Maximum a Posteriori criterion (MAP). The median probability with which an individual is assigned to a given group via MAP is 0.93 for setting VII and 0.94 for settings IV and V.

3.7 Discussion

In this chapter we have proposed two flexible classes prior specifications on component-specific parameters based on manipulating Beta distributions. This is a general setting which can conveniently be applied in the case of component-specific parameters which lie in $[0, 1]$, which is the case of the probability parameters of a mixture of binomial distributions. Finite mixture CR models are thus an appealing and straightforward example of application of the proposed methodology.

The outlined strategy allows to simultaneously avoid label-switching and elicit prior distributions in a quite flexible way. In particular, the explicit form of the mean and the variance of the marginal prior distribution of p_2 have been derived for those cases where a restricted Beta prior is considered on $p_2|p_1$: when prior information is available and conveniently used to elicit such a marginal prior on p_2, a more accurate posterior inference on component-specific parameters is yielded. Indeed, results on simulated data have shown how the proposed methodology can perform better than the unconstrained Beta prior specifications, above all when the number of occasions T is small and when some prior knowledge on the parameters is available. When scarce prior information is available, a conveniently specified conditional truncated Beta prior on $p_2|p_1$ may lead to a repulsion between the marginal priors on p_1 and p_2, working as a regularisation tool. We thus suggest the use of the specification (3.3) as a default choice when no (or very few) prior information is available: if needed, the hyperparameter k can be tuned to select the desired degree of repulsion between the two marginal priors (cfr. Figure 3.3). A rigorous sensitive analysis to investigate the way the posterior inference on p_1 and p_2 change with k can be of interest for future research. Interestingly, the prior specification $p_1 \sim Unif[0, 1]$ and $p_2|p_1 \sim Unif[p_1, 1]$ - which is often exploited in literature without any discussion about the consequences of this choice on the resulting marginal priors (e.g. Turek et al. (2021)) - is a special case of the more general class of prior specifications illustrated in Subsection 3.4.1.

Theoretical results have been derived only for the case of couples of component-specific parameters and this can sound as restrictive at a first glance. However, Capture-Recapture datasets are typically not large enough to be fitted by models that consider too many component-specific parameters (Pledger, 2000; Cubaynes et al., 2012), therefore we believe that this situation was worth to be investigated (also in light of its application to the models illustrated in the following chapters). Anyway, an extension to the case of $G > 2$ could be necessary to provide a generalisation of this methodology to some broader contexts.

Chapter 4

Modelling different residency patterns in open populations

In this chapter we review Jolly-Seber finite mixture models introduced by (Pledger et al., 2010) and we define a way to express them in the PX-DA context described in Subsection 2.2.3. In particular, we focus on models based on a single mixture grouping in G classes and models based on a double mixture grouping in $G \times H$ classes (cross-classified models).

Moreover, we propose and investigate more parsimonious models based on cross-classification. These types of models can turn out to be useful to characterise different residency patterns in animal open populations in a more economic way with respect to the models proposed by Pledger et al. (2003, 2010).

We illustrate the use of the considered models on a CR dataset from a live-trapping study, by highlighting the advantage in using our proposal over existing models. Motivated by the real data example, we further investigate the performance of the considered class of models in a simulation experiment. Data based on a robust design are considered both on the real application and the simulation experiment.

4.1 Modelling class heterogeneity using finite mixtures within the Jolly-Seber framework

In open-population CR studies individuals belonging to different groups may be allowed to have different capture and survival probabilities, eventually time-varying. In the most general specification, the relative order among the parameters of groups can change at each time t (i.e. group 1 could have the highest detection rate at the first period, but the lowest at the second one). This model is known as Interactive Heterogeneous Models (IHM) and its very rich specification may be an issue, since it has too many parameters for successful model fitting (for further discussion, see Pledger et al. (2010)).

Pledger et al. (2003, 2010) explore simpler specifications that could adequately represent the population structure and introduce a convenient notation to navigate through all possible sub-models. Let t and h be the time and group heterogeneity effects, respectively. Different expressions would correspond to different modelling structures: constant in time and homogeneous across groups $(\cdot)$; time-varying, but

homogeneous across groups (t); constant in time, but heterogeneous across groups (h); separable interaction of time-varying and heterogeneous across groups effects ($t+h$); non-separable interaction of time-varying and heterogeneous across groups effects ($t \times h$). For example, the IHM corresponds to $\{[\rho_{t\times h},\, \phi_{t\times h},\, p_{t\times h}]_G\}$, where G is the number of mixture components. If we want to specify a model whose heterogeneous group effect lies in the capture probabilities only, we write $\{\rho_t,\, \phi_t,\, p_{t\times h_G}\}$. The subscript is moved to highlight that the mixture of G components is related only to detection. Conversely, if the heterogeneous effect also applies to the survival probabilities, the we write $\{\rho_t,\, [\phi_{t\times h},\, p_{t\times h}]_G\}$.

The separable interaction between time and heterogeneity effect can be modelled through a logit link on the probability parameter to avoid problems on estimates going outside the set $[0, 1]$: for example, if we consider a model with time and heterogeneous across group additive effects like $\{\rho_t,\, \phi_t,\, p_{t+h_G}\}$, then the capture probability at time t of an individual belonging to the g-th class may be modelled as

$$\text{logit}(p_{gt}) = \mu + \eta_g + \tau_t\,, \tag{4.1}$$

where two constraints like $\sum_g \eta_g = 0$ and $\sum_t \tau_t = 0$ are often needed (Pledger et al., 2003).

This choice leads to additive effects, where an average capture probability fluctuating through time is modified either up or down by adding a different constant for each class. Of course, it is possible to exploit the logit link just to model either the time or the heterogeneity across group effect too. Similarly, such a kind of modelling could be applied also to recruitment and survival probabilities.

Anyway, it is relevant to notice that in terms of survival probabilities, constant survival implies that the same value should hold in identical time scales. For example, when capture occasions are not equally spaced but present uneven time differences between subsequent pairs, the survival probabilities should be appropriately compounded. Let $l_t = \tau_t - \tau_{t-1}$ $(t = 2, \ldots, T)$ be the *time-lag* between occasion $t-1$ and occasion t. Once the time scale is set (e.g. days, weeks, months, years, etc.), we have that $\phi_t = \phi^{l_t}$, where ϕ represents the survival probability across a single time-unit on the chosen scale.

PX-DA formalization

We can embed finite mixture models within the PX-DA formalization presented in Subsection 2.2.3 by adding one layer of hierarchy in the Bayesian model specification.

Let $\zeta_i \in \{1, \ldots, G\}$ be the latent membership label of each individual $i = 1, \ldots, M$ in $\mathbf{Y_{aug}}$. The hierarchical specification follows:

$$\begin{aligned}
&y_{it} \,|\, z_{it}, \zeta_i = g \sim Binom(J_t,\, p_{gt} \cdot z_{it}),\\
&z_{it} \,|\, z_{i,t-1}, r_{it}, \zeta_i = g \sim Bern(\phi_{gt} \cdot z_{i,t-1} + \rho_t \cdot r_{it}), \quad r_{it} = \min\{r_{i,t-1}, 1 - z_{i,t-1}\},\\
&p_{gt} \sim \pi_{p_g}(\cdot), \quad \phi_{gt} \sim \pi_{\phi_g}(\cdot), \quad \rho_t \sim \pi_\rho(\cdot), \quad \zeta_i \sim Multinom\,(1, (w_1, \ldots, w_G))\,,\\
&(w_1, \ldots, w_G) \sim \pi_{\mathbf{w}}(\cdot)\,.
\end{aligned} \tag{4.2}$$

The hierarchical formulation of (4.2) includes a multitude of possible specifications, according to what varies with time and across groups. As example, a graphical

visualization of the main components of model $\{[\rho, \phi, p_{h_G}]\}$ is provided by Figure 4.1, where a directed acyclic graph (DAG) highlights the relations between deterministic, observable and latent variables of the model.

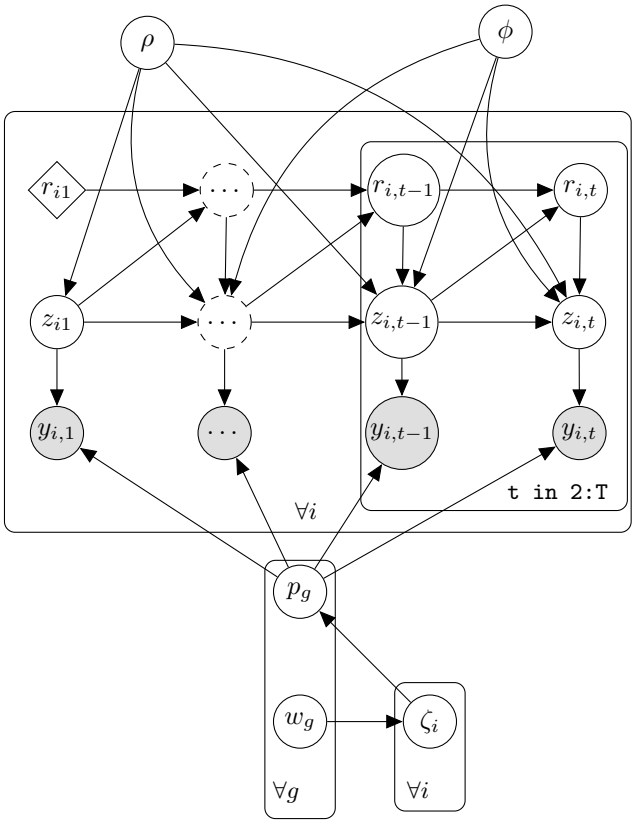


Figure 4.1. Bayesian DAG with the main components of the model $\{\rho, \phi, p_{h_G}\}$. White rhombi represent deterministic variables. White circles represent latent variables and parameters. Grey circles represents observable variables.

4.2 Cross-classified parsimonious CR mixture models

In this chapter we will consider homogeneous recruitment parameter and survival and capture probability that are constant over time, that is the subclass of models like $\{\rho_t, [p_h, \phi_h]_G\}$ or with less parameters. These models use a single mixture grouping into G classes, even when both survival and capture probability are heterogeneous across groups. This means that, for example, the 2-class model $\{\rho_t, [p_h, \phi_h]_2\}$ only considers two mixture components characterised by (p_1, ϕ_1) and (p_2, ϕ_2). This may be unsatisfactory in the case one wants to allow for all the possible combinations between capture and survival probabilities (i.e. (p_1, ϕ_1), (p_2, ϕ_1), (p_1, ϕ_2) and (p_2, ϕ_2)).

As alternative, one can consider a $(G \times H)$-class model which shows cross-classification of individuals into G capture probability classes and H survival probability classes. To characterise each parameter without any ambiguity (i.e. class 1 is the one characterised by the lowest survival and capture probabilities), we will state that $p_1 < \cdots < p_G$ and $\phi_1 < \cdots < \phi_H$. In the following, we will restrict to the

case of $G = H = 2$, which is a more reasonable situation in a real world application where the number of encountered individuals (and, thus, the size of the resulting data matrix) is not large. This choice yields the 4-class model illustrated in Table 4.1, which can be written as $\{\rho_t,\, p_{h_2}, \phi_{h_2}\}$. These kinds of models allows to express a 4-class model that is much more parsimonious than $\{\rho_t,\, [p_h, \phi_h]_4\}$ (Pledger et al., 2003).

	ϕ_1	ϕ_2
p_1	Class 1	Class 2
p_2	Class 3	Class 4

Table 4.1. Description of 4-class cross-classified model, i.e. $\{\rho_t,\, p_{h_2}, \phi_{h_2}\}$.

In the same spirit, a parsimonious 3-class alternative to model $\{\rho_t,\, [p_h, \phi_h]_3\}$ may adopt the model described in Table 4.1, but collapsing one of the four classes. Such a 3-class model may be specified in four different way (see Table 4.2), depending on the combination $(p_\cdot,\, \phi_\cdot)$ which is excluded. For example, $\{\rho_t,\, p_{h_2}, \phi_{h_2}\}_{-(p_1,\, \phi_1)}$ indicates

	ϕ_1	ϕ_2
p_1		Class 1
p_2	Class 2	Class 3

(a)

	ϕ_1	ϕ_2
p_1	Class 1	
p_2	Class 2	Class 3

(b)

	ϕ_1	ϕ_2
p_1	Class 1	Class 2
p_2		Class 3

(c)

	ϕ_1	ϕ_2
p_1	Class 1	Class 2
p_2	Class 3	

(d)

Table 4.2. Description of model $\{\rho_t,\, p_{h_2}, \phi_{h_2}\}_{(p_\cdot,\, \phi_\cdot)}$, where (a) excludes combination $(p_1,\, \phi_1)$, (b) excludes combination $(p_1,\, \phi_2)$, (c) excludes combination $(p_2,\, \phi_1)$ and (d) excludes combination $(p_2,\, \phi_2)$.

a 3-class cross-classified model where the couple $(p_1,\, \phi_1)$ has been excluded. This alternative specification configures a more parsimonious and less complex model, with some mixture components sharing common parameters.

4.3 Real data illustration: *Microtus Pennsylvanicus*

We illustrate the proposed class of models using Capture-Recapture data from a livetrapping study of meadow voles (*Microtus Pennsylvanicus*) at Patuxent Wildlife Research Center in Laurel, Maryland (Nichols et al., 1994). As illustration, we only consider the reduced version of the data set, which is available in the R package `Rcapture` (Baillargeon and Rivest, 2007). This dataset concerns a robust design with 30 capture occasions pertaining to $T = 6$ primary periods having $J = 5$ capture

occasions each. These capture occasions represent five consecutive days of trapping every month from June to December 1981. A total of $D = 171$ meadow voles have been encountered in the considered time frame.

We consider different alternative models, all having time-varying recruitment parameters and time-constant capture and survival probabilities. In the simplest case, we consider a model with homogeneous capture and survival probability ($\{\rho_t, \phi, p\}$). Regarding single mixture grouping into G classes, we consider models with components that varies by capture probability ($\{\rho_t, \phi, p_{h_G}\}$, for $G = 2, 3$), by survival probability ($\{\rho_t, \phi_{h_G}, p\}$, for $G = 2, 3$) or by both ($\{\rho_t, [\phi_h, p_h]_G\}$, for $G = 2, 3, 4$). Regarding cross-classified models, we consider the 4-class cross-classified models (i.e. $\{\rho_t, \phi_{h_2}, p_{h_2}\}$) and the four 3-class cross-classified more parsimonious alternative depicted in Table 4.2.

Prior distribution in (2.6) is chosen for ρ_t for all the considered models. A flat Dirichlet distribution is used for the vector of mixture weights, i.e. $\mathbf{w} \sim Dirichlet(\alpha_1, \ldots, \alpha_G)$ with $\alpha_1 = \cdots = \alpha_G = 1$. When p (or ϕ) is the same for all the individuals, a standard Uniform prior is placed on the parameter. When two capture (or two survival) probabilities are considered, prior specification of setting VII presented in Section 3.5 is chosen to favour a slight repulsion between the two components. When one considers more than three capture or survival parameters, constrained Uniform priors are instead chosen. These different prior choices for the models does not substantially affect the model selection criterion associated with each model; conversely, such a prior choice for models with two capture (or two survival) probabilities turns out to be useful to induce a better separation between couples of capture (or couple of survival) parameters.

All models are fitted in JAGS (BUGS code is provided in Appendix A), by using two parallel chains of $20,000$ iterations each and discarding the first $10,000$ of them as burn-in. The overall goodness-of-fit is measured via Watanabe-Akaike Information Criterion (WAIC, Watanabe and Opper (2010)), following the good practice of Gelman et al. (2014) whenever complex hierarchical models are fitted. WAIC is measured on the scale of deviance, and therefore lower values of WAIC indicate a more parsimonious fit to the data. This measure has been recently used in the context of Capture-Recapture mixture models by Turek et al. (2021).

Table 4.3 shows posterior estimates on N_{super} for the 12 fitted models that are listed in an increasing order according to the corresponding WAIC. The lowest WAIC value is attained by model $\{\rho_t, p_{h_2}, \phi_{h_2}\}_{-(p_2, \phi_1)}$, which yields an estimated super-population abundance of 186 individuals. It is interesting to notice that all the models that assume two capture probabilities yield similar estimates of N_{super}. The same happens for all the models that consider a higher or lower number of capture probabilities: notably, all models with homogeneous capture probability tend to agree with an estimate of 173 individuals and a very short 95% credible interval, while the two models that assume three capture probabilities yield the highest abundance estimates. The model with the second lowest WAIC is the 4-class cross-classified model, which considers one more class than the *best* model, i.e. the class characterised by the combination (p_2, ϕ_1). Nevertheless, this class is associated with a mixture weight whose estimated posterior density tends to concentrate close to

Model	$\hat{N}_{super}$	95% CI	WAIC
$\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_2,\, \phi_1)}$	186	(176, 197)	1493.72
$\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}$	185	(176, 196)	1500.34
$\{\rho_t,\, [\phi_h,\, p_h]_2\}$	185	(176, 196)	1503.15
$\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_1,\, \phi_2)}$	185	(176, 196)	1503.31
$\{\rho_t,\, [\phi_h,\, p_h]_3\}$	196	(179, 218)	1511.52
$\{\rho_t,\, \phi,\, p_{h_2}\}$	185	(176, 196)	1522.00
$\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_2,\, \phi_2)}$	185	(175, 196)	1524.98
$\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_1,\, \phi_1)}$	185	(175, 196)	1525.43
$\{\rho_t,\, [\phi_h,\, p_h]_4\}$	227	(181, 302)	1534.46
$\{\rho_t,\, \phi,\, p_{h_3}\}$	202	(178, 248)	1535.90
$\{\rho_t,\, \phi_{h_3},\, p\}$	173	(171, 176)	1601.17
$\{\rho_t,\, \phi_{h_2},\, p\}$	173	(171, 176)	1601.27
$\{\rho_t,\, \phi,\, p\}$	173	(171, 176)	1606.34

Table 4.3. Posterior estimates (posterior median and 95% high posterior density credible interval) for N_{super} and WAIC. Models are sorted by increasing values of WAIC.

0 (cfr. bottom left panel of Figure 4.2b): this may suggest that a more parsimonious model is needed. We thus select $\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_2,\, \phi_1)}$ as most suitable model among the considered ones.

For the selected model, the estimated capture probabilities are $\hat{p}_1 = 0.26$ (0.19, 0.33) and $\hat{p}_2 = 0.70$ (0.65, 0.75), while the estimated survival probabilities are $\hat{\phi}_1 = 0.40$ (0.09, 0.67) and $\hat{\phi}_2 = 0.76$ (0.69, 0.83). The estimated posterior density of ϕ_1 is bell-shaped but quite vague, as suggested by the large 95% credible interval: indeed, the latent nature of the survival process and the fact that ϕ_1 controls the survival of those individuals that remain in the population for a shorter period may make the estimate of this parameter a challenging task. The estimate is expected to improve with either a much larger number of primary periods or with the elicitation of strong prior information on the parameter. Anyway, the choice of mild repulsive priors on ϕ_1 and ϕ_2 does help to separate the posterior densities of the two survival parameters, whose credible intervals do not overlap.

Figure 4.3a represents the evolution of the estimated population size over the six considered months. The trend is in line with findings of Baillargeon and Rivest (2007), who apply different types of CR models on these data. Figure 4.3b shows the estimated recruitment parameters. Notice that $\hat{\rho}_1$ and $\hat{\rho}_6$ are higher than the other estimates: in the first case, this may attributed to the fact that all the individuals belonging to the population before June 1981 are *virtually* added to the population at time $t = 1$; in the second case, such a large recruitment parameter may be due to the exceptional number of individuals which have been observed for the first time in the last primary period (42% of the individuals encountered in only one

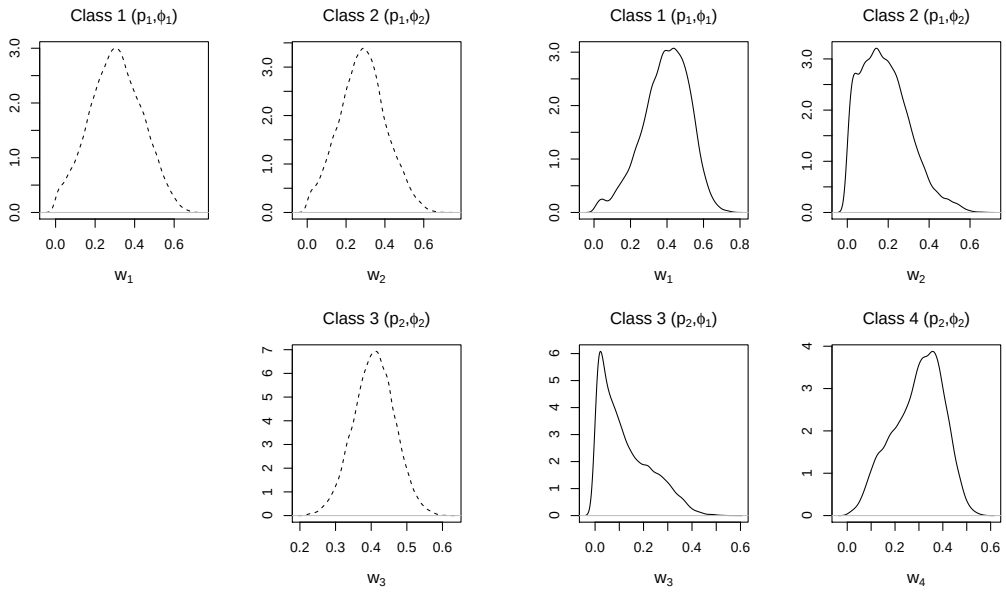
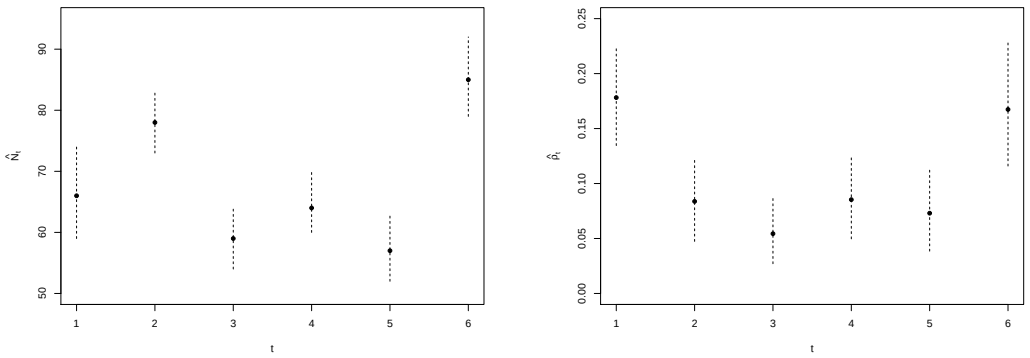


Figure 4.2. Estimated posterior densities of the mixture weights for the 3-class (a) and the 4-class (b) cross-classified models. Bottom left panel in (a) is blank, since combination (p_2, ϕ_1) does not exist in the considered 3-class model.

primary period have been encountered at $t = 6$). Moreover, the wider 95% credible interval for ρ_6 reflects the combination between the exceptional number of first encounters in the last primary period and the lack of further information about the new encountered individuals due to censoring.

Post-hoc inspection of the probabilities of belonging to each class using the selected model shows that long dense capture histories are clearly allocated to the third class, characterised by the combination of high capture and high survival, i.e. (p_2, ϕ_2). For example, an individual with capture history 1 3 1 4 5 2 is assigned to class 3 with probability 0.91. Notably, 52% of the observed individuals allocated to the third class via Maximum a Posteriori (MAP) are assigned to that class with probability greater than 0.9. On the other hand, long sparse capture histories are assigned to class 2, characterised by the couple (p_1, ϕ_2): for example, an individual with capture history 0 1 1 1 0 1 is allocated to this class with probability 0.85. On average, observed individuals have been assigned to this second class with a posterior probability of 0.65. Individuals that have been encountered a few number of times during one or few primary periods are all assigned to the third class, characterised by small capture and small survival, i.e. (p_1, ϕ_1). As expected, the situation associated with the highest uncertainty of allocation is the one where the individual is encountered for the first time in the last primary period and for very few secondary periods. In this case, the individuals is likely to be either a short-term survivor or a long-term survivor that have just entered the population. Short capture histories are of course less informative since they can be attributed to either elusive long-term survivors which enter the population at some occasion $t > 1$ or to elusive short-term



(a) **(b)**

Figure 4.3. Estimated population abundance (a) and recruitment parameter (b) at each occasion t, corresponding to months from June to December 1981. Dashed lines represent 95% credible intervals.

survivors.

4.4 Simulation experiment

We conduct a simulation experiment to assess the performances of the model selected in Section 4.3 whenever it is well-specified. Therefore, we generate multiple sets of artificial data from model $\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_2,\phi_1)}$ and then estimate a pool of models on them. These candidate models are the thirteen models presented in the previous real data illustration and reported in Table 4.3. The main objective is to evaluate the ability to recover the true value of the super-population abundance (N_{super}).

Design details and true values of the parameters have been inspired by the results obtained in Section 4.3. We consider a number of primary sampling occasions $T \in \{4, 6, 8, 10\}$ and constant secondary sampling occasions $J_t = J = 5$. The choice of different values for T aims at assessing models' performance with increasing sample size. True survival probabilities are $\phi_1 = 0.35$ and $\phi_2 = 0.75$, while true capture probabilities are $p_1 = 0.25$ and $p_2 = 0.70$. The recruitment parameters are fixed to $\rho_1 = 0.25$ and $\rho_t = 0.05$ for $t = 2, \ldots, T$. Such a higher value for ρ_1 aims at taking into account of individuals which are already part of the population before the beginning of the study, but that are virtually added to the population at that occasion. The three mixture components are chosen to be balanced on the pseudo-population, i.e. $w_1 = w_2 = w_3 = \frac{1}{3}$. We simulate independent encounter histories for $K = 50$ pseudo-populations, assuming an expected super-population size of $\mathbb{E}[N_{super}] = 200$. That corresponds to $\psi = 0.36,\, 0.42,\, 0.48,\, 0.53$ for $T = 4,\, 6,\, 8,\, 10$, and then requires creating pseudo-populations with $M = 561,\, 477,\, 420,\, 380$ pseudo-individuals, respectively. Any of the M pseudo-individuals not recruited into the population or not captured are, of course, excluded to produce an observable set of capture histories.

We fit all the considered models on the K simulated datasets, augmenting each dataset by 200 all-zero capture histories for the implementation of the PX-DA approach. The prior setting is the one considered in Section 4.3. Estimation is

carried out using JAGS (Plummer et al., 2003). We run 2 parallel chains with $20,000$ iterations each, discarding $10,000$ as burn-in.

We report median posterior estimates as a robust measure to outlier point estimate, mitigating the effect of anomalies that can result in occasionally low-informative datasets. In the same spirit, we rely on the *mean absolute error* (MAE) as an accuracy measure for posterior estimates. We also report the overlapping index OV (see Section 3.5) between the posterior distributions of the couple of capture (or survival) probabilities, averaged over $K = 50$ replicas. This metric is particularly appealing to understand whether the posterior distributions of these parameters are well-separated or not. Finally, we investigate the classification ability of the true model. Recall that classification performances can be evaluated in a simulation setting as the true group labels are known, while they are not known in real data applications. We compute the multiclass AUC (mAUC) to assess the overall performance of the fuzzy classification (Hand and Till, 2001) when the true model is considered. Notice that classification performances are only evaluated for individuals that have been actually encountered, since all-zero histories do not provide useful information about the class an individual belongs to (we are only provided with the posterior probability for an uncaptured individual to belong to a particular class).

Table 4.4 reports summaries of the estimated posterior on N_{super} for the thirteen considered models, highlighting how the true data generating process tends to return the lowest MAEs as T increases. Models that do not take into account the existing heterogeneity in capture probabilities do fail in recovering the true super-population abundance in all the cases: indeed, their MAEs associated with the estimated of N_{super} are large, while their 95% posterior credible intervals are short and never cover the true value of the parameter (cfr. xi-xiii).

		$T=4$			$T=6$			$T=8$			$T=10$		
	Model	**MAE**	**Cov.**	**CIW**	**MAE**	**Cov.**	**CIW**	**MAE**	**Cov.**	**CIW**	**MAE**	**Cov.**	**CIW**
(i)	$\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_1,\phi_1)}$	4.70	0.88	18.10	6.30	0.78	16.78	5.10	0.80	16.92	4.32	0.80	15.80
(ii)	$\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_1,\phi_2)}$	4.34	0.90	19.36	5.42	0.82	18.10	4.60	0.88	18.14	3.48	0.94	17.00
(iii)	$\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_2,\phi_1)}$	4.40	0.92	20.32	4.62	0.90	19.82	4.58	0.94	20.04	3.30	1.00	19.02
(iv)	$\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_2,\phi_2)}$	4.58	0.88	18.22	6.20	0.78	16.90	5.00	0.80	17.16	4.10	0.90	16.08
(v)	$\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}$	4.44	0.88	20.06	4.92	0.88	19.36	4.60	0.94	19.56	3.32	0.98	18.50
(vi)	$\{\rho_t,\, [\phi_h,\, p_h]_2\}$	4.30	0.90	19.52	5.32	0.86	18.28	4.56	0.88	18.42	3.54	0.94	17.14
(vii)	$\{\rho_t,\, [\phi_h,\, p_h]_3\}$	5.50	1.00	43.68	6.12	0.96	45.58	6.82	0.98	47.46	5.06	1.00	46.16
(viii)	$\{\rho_t,\, [\phi_h,\, p_h]_4\}$	6.17	1.00	52.74	5.92	1.00	55.66	7.44	1.00	52.60	5.54	1.00	49.64
(ix)	$\{\rho_t,\, \phi,\, p_{h_2}\}$	4.60	0.88	18.36	6.14	0.78	16.96	4.92	0.80	17.16	4.06	0.92	16.06
(x)	$\{\rho_t,\, \phi,\, p_{h_3}\}$	7.82	0.98	52.56	7.36	0.94	53.06	9.76	0.88	52.04	9.50	0.96	66.14
(xi)	$\{\rho_t,\, \phi_{h_2},\, p\}$	14.36	0.00	6.84	15.82	0.04	6.76	14.06	0.00	7.04	13.34	0.00	6.60
(xii)	$\{\rho_t,\, \phi_{h_3},\, p\}$	14.36	0.00	6.86	15.74	0.04	6.78	13.96	0.00	7.06	13.26	0.00	6.66
(xiii)	$\{\rho_t,\, \phi,\, p\}$	14.58	0.00	6.70	16.14	0.02	6.44	14.40	0.00	6.62	13.72	0.00	6.16

Table 4.4. Estimates of MAE, coverage (Cov.) and width of the 95% credible intervals (CIW) for N_{super}, obtained when data are simulated from model $\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_2,\phi_1)}$ (the corresponding row is highlighted in light grey). For convenience, labels (i)-(xiii) are assigned to the thirteen models.

Similar considerations can be supported by Figure 4.4, which shows the differences

between the estimated ($\hat{N}_{super}$) and true (N_{super}) super-population size for each of the $K = 50$ replicas by all considered models. We notice that, for some replicas, the true value of $\hat{N}_{super}$ tends to be dramatically overestimated by models that consider a larger number of capture probabilities than necessary, i.e. $\{[\phi_g, p_g]_{G=3}\}$ and $\{[\phi_g, p_g]_{G=3}\}$. On the other hand, as already pointed out, models that do not consider heterogeneity between capture probabilities persistently underestimate the super-population abundance.

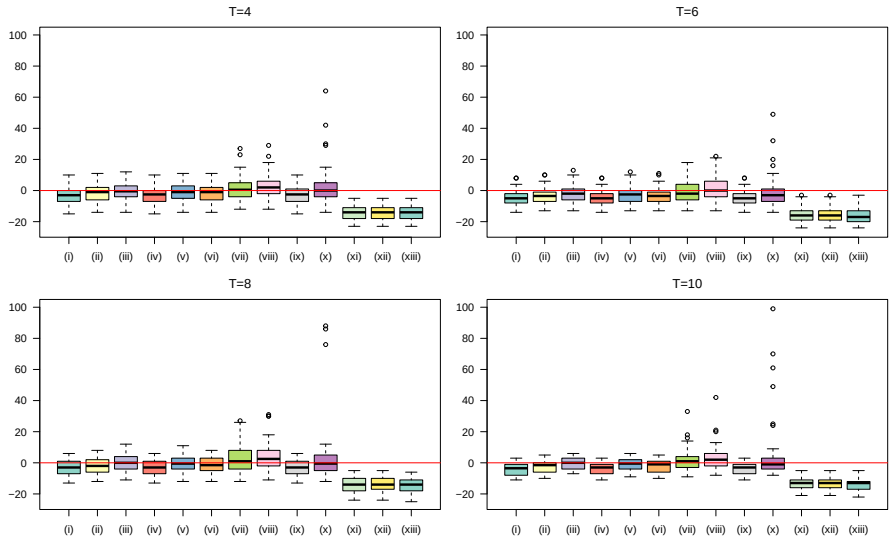


Figure 4.4. Difference between the estimated ($\hat{N}_{super}$) and true (N_{super}) super-population abundance for each of the $K = 50$ independent replicas and for the candidate models listed in Table 4.4. Each panel reports the values for a different number of primary occasions T. All the replicas are simulated from the same data generating process - i.e. $\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_2,\,\phi_1)}$ - which corresponds to (iii).

The lowest values of the WAIC are systematically those related to the 3-class model $\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_2,\,\phi_1)}$ and to the 4-class model $\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}$, attaining values for the two models whose difference is negligible in most of the replicas. Anyway, if we look at the mixture weight of the latter model corresponding to the combination $(p_2,\, \phi_1)$, we notice that its estimated posterior density tends to concentrate around 0 in most of the simulated datasets (see Figure 4.5). This highlights a potential overfitting of model $\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}$ and suggests a reasonable collapse of the component characterised by the combination $(p_2,\, \phi_1)$ in favour of the more parsimonious model $\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_2,\,\phi_1)}$.

Regarding the estimation of group-specific detection and survival probabilities, the high coverage of the credible intervals in Table 4.5 shows that model $\{\rho_t,\, \phi_{h_2},\, p_{h_2}\}_{-(p_2,\,\phi_1)}$ correctly recovers the true parameter values in most of the cases. Although ϕ_1 has quite wide 95% credible intervals in all the scenarios (cfr. Section 4.3), the estimated posterior densities of the two survival probabilities turn to be well-separated ($\text{OV}_{\phi_1,\,\phi_2} \leq 0.03$ for each T).

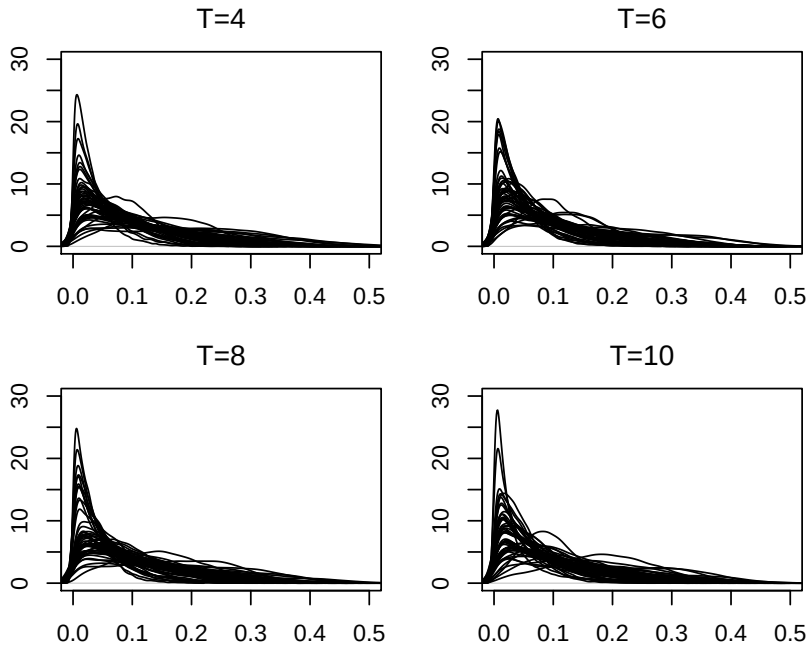


Figure 4.5. Estimated posterior distributions of the mixture weights of the component of model $\{\rho_t, \phi_{h_2}, p_{h_2}\}$ which is characterised by the combination (p_2, ϕ_1).

	p_1		p_2			ϕ_1		ϕ_2		
T	**Cov.**	**CIW**	**Cov.**	**CIW**	$\mathbf{OV}_{p_1,p_2}$	**Cov.**	**CIW**	**Cov.**	**CIW**	$\mathbf{OV}_{\phi_1,\phi_2}$
4	0.96	0.08	0.88	0.08	0.00	0.94	0.53	0.92	0.16	0.03
6	0.98	0.07	0.96	0.07	0.00	0.96	0.52	1.00	0.13	0.01
8	0.86	0.06	0.98	0.07	0.00	0.96	0.54	0.94	0.12	0.02
10	0.96	0.06	0.94	0.06	0.00	0.96	0.52	0.94	0.11	0.01

Table 4.5. Estimates of coverage (Cov.) and width of the 95% credible intervals (CIW) for capture and survival probabilities and average overlapping index between posterior densities of p_1 and p_2 and between the ones of ϕ_1 and ϕ_2. The fitted model coincides with the data generating process, i.e. $\{\rho_t, \phi_{h_2}, p_{h_2}\}_{-(p_2,\phi_1)}$.

Finally, classification performances of the true model are quantified by the mAUC score, which is pretty high also in the case of few primary capture occasions (i.e. average mAUC is 0.87 when $T = 4$) and tends to slowly increase with T. It is important to remark that there are aspects of CR data that might unavoidably affect the performance of the classification. For example, encounter histories are right-censored and this makes it hard to classify an individual that is captured for the first time close to the end of the study (Cubaynes et al., 2012).

4.5 Discussion

In this chapter we have reviewed the Pledger et al. (2010)'s class of JS finite mixture models and we have extended it to the Bayesian PX-DA framework of Subsection 2.2.3. These types of models can be used to jointly estimate the population abundance and classify individuals according to their capture and survival tendencies.

Along with the classical single mixture grouping in G classes, we considered model based on a double mixture grouping too. In the latter case, for example, different mixture components may share some common parameters by avoiding the use of too many parameters. Notably, we explored a parsimonious class of cross-classified JS-type models, which can be used to highlight groups of individuals with different patterns of residency in the study area. Notably, this model comes from a parsimonious reduction of the 4-class cross-classified model $\{\rho_t, \phi_{h_2}, p_{h_2}\}$ obtained by collapsing one of the mixture component: therefore, the proposed 3-class cross-classified model was named $\{\rho_t, \phi_{h_2}, p_{h_2}\}_{-(p_\cdot, \phi_\cdot)}$, that corresponds to model $\{\rho_t, \phi_{h_2}, p_{h_2}\}$ without the combination $(p_\cdot, \phi_\cdot)$.

When true data are generated by the 3-class cross-classified model, the 4-class cross-classified model still performs well in estimating the true super-population abundance and attains values of WAIC which are comparable with those of the true model; however, from a further inspection of the mixture weights, it is possible to find out that the 4-class model introduces an additional but unnecessary mixture component which seems to capture some random noise in the data.

Mild repulsive priors turned to be useful in separating the posterior densities of the two survival parameters. However, the latent nature of the survival process makes the estimation of the lower survival probability quite challenging, above all in presence of few primary sampling periods: whenever ecological information is available, suitable prior should be elicited to reduce the uncertainty on the final parameter estimate.

The model selection has been supported by the WAIC, a measure of the goodness-of-fit of a hierarchical model structure to a given dataset. This measure has been recently used by Turek et al. (2021) in the context of CR mixture models and seems a valid alternative to the most common DIC, whose use is not recommended for models involving mixtures of distributions (Spiegelhalter et al., 2003). An interesting alternative which may be worth exploring in the future is the Gelfand and Ghosh (1998)'s model selection criterion, which has been adopted in the context of Bayesian Capture-Recapture models by Ghosh and Norris (2005).

For simplicity, throughout all the applications in this chapter, we have considered the recruitability parameter ρ_t to vary across primary sampling periods but to be homogeneous across groups. However, for example, it can happen that two different groups of individuals may share the same capture and survival probabilities but being characterised by different recruitment parameters; or even more complex situations may be taken into account, as shown in Chapter 5, where constant recruitments are assumed for a group of individuals and time-varying recruitments are assumed for the rest of the population. Summing up, there are plenty of options with which one can built a cross-classified model and reasonable model assumptions should be based on several factors, such as previous knowledge about the population under study and the way Capture-Recapture samplings have been designed and carried

out. Here, we have focused on a situation which will serve as the basis for what will be presented afterwards: specifically, in the next chapter we will move a step forward by extending the 3-class cross-classified model to depict the structure of a population whose residency patterns are already known in literature.

Chapter 5

Modelling residency patterns in marine wildlife populations

In Chapter 4 we introduced parsimonious cross-classified JS mixture models which allow to jointly estimate the population abundance and classify individuals according to their capture and survival tendencies: individuals which show similar capture histories tend to be assigned to the same group.

In some cases, ecological information about the population - or, more generally, about the species - of interest is available to the researcher and it can be suitably used to specify a statistical model which attempts to describe the real phenomenon in mathematical terms. Common practice in this very general setting is to include covariates to explain differences in population's members and use them to induce changes in the detection or survival mechanisms through link functions. However, informative covariates are often unavailable and, as seen in Chapter 3 and 4, a natural solution is represented by finite mixture models. In addition, we believe that finite mixtures represent a valuable tool to outline - in statistical terms - classes of individuals sharing similar profiles whenever strong scientific evidence of population groups' existence is available. The identification of common profiles is crucial to describe different individual residency patterns and quantify the number and proportion of individuals that correspond to each profile, a task that is highly relevant in animal conservation management.

In this spirit, in this chapter we focus on the study of an open population of bottlenose dolphins composed of individuals that are known to cluster in groups which vary according to their level of site-fidelity. This motivates the subsequent definition of a tailored Bayesian mixture model to describe and characterising this known structure. Notably, we show how prior scientific knowledge on the structure of the population of interest can be leveraged to specify a more parsimonious model and to improve on the final estimates of the population groups and sizes. We assess the performance of our proposal through a simulation study, comparing it with two alternative models: the first assuming a similar but less parsimonious structure, while the second not assuming a group-wise structure of the population. We finally provide an illustration on a dataset from a recent survey on a population of bottlenose dolphins.

Part of this chapter is currently published (pre-print) in Caruso et al. (2023),

while undergoing the review process.

5.1 Residency patterns of common bottlenose dolphins

Most of the recent literature about common bottlenose dolphins converges towards the identification of three groups characterized by different levels of site-fidelity with respect to a specific study area: from the most to the least frequently present (Dinis et al., 2016; Estrade and Dulau, 2020; Haughey et al., 2020; La Manna et al., 2022).

The first group is composed by individuals that almost never leave the study area; these are usually referred to as *resident* individuals, that are observable throughout many occasions and for long periods of time (e.g. across multiple years). The second group includes individuals that are not continuously present in the study area but regularly visit it; these are called *part-time* resident individuals, that are observable throughout a wide time window but are usually encountered at occasions far apart in time. The third and last group is composed of individuals that enter the study area only once in their lifetime and stay there for a short time window; these are *transient* individuals, that are observable only at occasions occurring on close dates.

This population structure is of utter interest for marine biologists, who are mostly interested in disentangling the permanent or semi-permanent population from the transient one. They usually apply clustering methods on some site-fidelity metrics derived from the capture histories to deterministically assign each individual to its group. After that, the group labels are used in the Jolly-Seber framework to model heterogeneity in the entrance, detection, and survival probabilities of the individuals (see, for example, Estrade and Dulau (2020); Pace et al. (2021)). While practical for understanding the underlying structure of the population of interest, this two-step approach has some relevant issues. First of all, there is no quantification nor propagation of the uncertainty of the classification step onto the modeling step; this can bias the final estimates and yield over-confident conclusions. Second, the same information set is used twice in two different statistical procedures, where the latter is performed after conditioning on the first; this can lead to a confirmation bias in favor of the original hypothesis.

Here, we propose to unify the two steps into a joint statistical procedure which allows population size estimation and classification at the same time. This lets the same estimation procedure find the best fuzzy classification to explain the observed capture histories and properly propagate the uncertainty at all levels (Clark and Gelfand, 2006). In particular, what we propose is a tailored version of the parsimonious cross-classified models illustrated in Chapter 4 which incorporates the prior knowledge about the population's structure.

5.2 RPT model for characterising residency patterns

The population structure illustrated in Section 5.1 and characterised by individuals with either *resident* (R), *part-time* (P) or *transient* (T) pattern can be translated in mathematical terms to reflect variations in the parameters of the open-population PX-DA framework.

We can see individuals as separated in two large distinct groups with two different ecological behaviours: specifically, they can be either short-term survivors (transient) or long-term survivors (non-transient, that is either resident or part-time) according to their tendency to stay in the study area for a shorter or wider time window, respectively. Therefore, we suppose that transient individuals has a survival probability (ϕ_{TR}) which is smaller than the one of the other two groups (ϕ_{NTR}), that is $\phi_{TR} < \phi_{NTR}$. Notice that these two survival parameters represent the survival probabilities across a single time-unit on the chosen scale. In fact, when subsequent capture occasions are not equally spaced, the survival probabilities should be appropriately compounded (cfr. Section 4.1), that is $\phi_{TR,\,t} = \phi_{TR}^{l_t}$ and $\phi_{NTR,\,t} = \phi_{NTR}^{l_t}$, where $l_2, \ldots, l_T$ are the time differences between the T subsequent occasions.

Moreover, we assume that transient individuals are characterised by a constant recruitability parameter $\rho_{TR,\,t} = \rho_{TR}$, $\forall t$, while resident and part-time individuals are characterised by a time-varying recruitability parameter, i.e. $\rho_{NTR,\,t}$. In fact, while transient individuals may randomly pop up in the population at any occasion, most of the resident and part-time individuals are expected to be already part of the population before the start of the survey and, consequently, to be virtually added into the population in the very first capture occasions.

Notice that these two sets of recruitment parameters induce a clusterisation in the relationship between the recruitment parameters and the super-population size N_{super}. In practice, it yields group-specific inflation parameters that modify the expected super-population size as

$$\mathbb{E}\left[N_{super} | \psi_{TR}, \psi_{NTR}, w_3\right] = M \cdot \left[w_3 \psi_{TR} + (1 - w_3)\psi_{NTR}\right],$$

where w_3 is the mixture weight of transient individuals, $\psi_{TR} = 1 - (1 - \rho_{TR})^T$ is the inflation parameter for transient individuals and $\psi_{NTR} = 1 - \prod_{t=1}^{T} (1 - \rho_{NTR,\,t})$ is the inflation parameter for non-transient individuals.

The only difference between resident and part-time individuals lies in the partial undetectability of the latter group, whose individuals (if alive) may not be present in the study area at some random occasions with probability $\delta \in (0, 1)$. The biological interpretation of a time-constant *undetectability parameter* is that the probability of being in the study area during the current sighting is the same for those part-time individuals in and those out of the study area during the previous sighting. Interestingly, this parameter plays a similar role to the completely random emigration parameter of Kendall et al. (1997). Anyway, although the undetectability parameter is part of the ecological process, it directly affects the capture probability of part-time individuals. In fact, if the capture probability at time t for resident and transient individuals is

$$p_{NPT,\,t} = \text{logit}^{-1}\left(\mu + \tau_t\right), \quad t = 1, \ldots, T,$$

the capture probability at time t for part-time individuals is given by

$$p_{PT,\,t} = (1 - \delta) \cdot p_{NPT,\,t} = (1 - \delta) \cdot \text{logit}^{-1}\left(\mu + \tau_t\right), \quad t = 1, \ldots, T.$$

Of course, $p_{PT,t} < p_{NPT,t}$, $\forall t$. Notice that the larger δ the more separated the part-time group from the resident group. General-purpose and weakly informative priors can be ascribed to μ. In order to induce a little time variation (in the logit-scale) on the capture probabilities, we suggest considering

$$\tau_t \sim N(0,\, \sigma^2), \forall t\,, \tag{5.1}$$

with σ small.

We name this model RPT as it encompasses the three illustrated residency patterns ordered on a progressively lower degree of site fidelity. However, we would like to stress how the model does not enforce this interpretation. For instance, both survivals could be estimated to be high, or the undetectability parameter estimated to be close to 0. Table 5.1 provides a summary of the parameters which characterise each group of the RPT model, highlighting the fact that some mixture components share common parameters.

	mix. weight			**recruitment**		**zero-inflation**		**survival**		**capture**		**undetect.**
Group	w_1	w_2	w_3	ρ_{TR}	$\rho_{NTR,t}$	ψ_{TR}	ψ_{NTR}	ϕ_{TR}	ϕ_{NTR}	$p_{PT,t}$	$p_{NPT,t}$	δ
Resident	✓				✓		✓		✓		✓	
Part-time		✓			✓		✓		✓	✓		✓
Transient			✓	✓		✓		✓			✓	

Table 5.1. Description of the main parameters of the RPT model.

5.3 Simulation experiment

We conduct a simulation experiment to assess the performances of the RPT model whenever it is well-specified, i.e. the data are generated according to the structure described in Section 5.2. We generate multiple sets of artificial data under alternative scenarios from the RPT model with fixed parameters and then estimate a pool of models on them. Our main objective is twofold: i) to evaluate the ability of recovering the true values of the parameters, with a particular focus on N_{super}; ii) to verify whether the RPT model is chosen as the *best* one among other alternatives according to some model selection criterion.

5.3.1 Simulation setup

Experimental design

We consider four scenarios characterised by different number of sampling occasions, i.e. $T \in \{10, 20, 30, 40\}$, to verify the model performances with an increasing time horizon. We suppose that in the first scenario (i.e. $T = 10$) all the captures are recorded within a relatively short period (e.g. within a year). Longer time horizons are considered in the subsequent scenarios, where a larger time-gap (year gap) is assumed to occur every 10 occasions. Further details about the time lags are available in Appendix D.

The month (and portion of months) is taken as the basic time unit to avoid the possible numerical instability related to the large values of the lags in terms of days. Note that this has an effect on the interpretation of the survival probability parameter that will represent the probability to survive one month. We adopt the following parameters' values in all the scenarios. The (monthly) survival probabilities are set to $\phi_{TR} = 0.01$ and $\phi_{NTR} = 0.997$. These two values may seem quite extreme at a first glance, but they actually assure that the short-time survivors (transient individuals) almost surely stay in the population for less than a year and that the long-term survivors (non-transient individuals) stay in the population for more than three months with very high probability (> 0.99). Furthermore, notice that a monthly survival probability equal to 0.01 corresponds to a survival probability equal to a probability of 0.86 on a daily scale and of 0.34 on a weekly scale. On the other hand, a monthly survival probability equal to 0.997 corresponds to a probability of 0.87 on a quadrennial scale. Therefore, the monthly scale appears as a good compromise to avoid both a too low value for ϕ_{TR} and a too high value for ϕ_{NTR}.

The capture probabilities are obtained by setting $\mu = 0$ and $\delta = 0.7$, and generating τ_t's from a $N(0, 0.25)$ in each scenario.

The recruitment parameter for transient individuals is fixed to $\rho_{TR} = 0.02$ for all capture occasions, while for resident and part-time individuals we consider time-varying recruitment parameters. Specifically,

$$\rho_{NTR,t} = \begin{cases} 0.4\,, & t = 1 \\ 0.04\,, & t = 10k + 1\,,\ k = 1, 2, 3 \\ 0.005\,, & \text{otherwise} \end{cases} .$$

The mixture weights for the three groups in the pseudo-population are set to $w_1 = 0.2$ (resident), $w_2 = 0.45$ (part-time) and $w_3 = 0.35$ (transient).

We envision a pseudo-population of $M = 500$, that yields an expected super-population size $\mathbb{E}[N_{super}] \in \{171, 213, 248, 278\}$, as T increases. Notice that N_{super} increases with T as more individuals can visit the study area during a longer time horizon.

We simulate independent encounter histories for $K = 50$ pseudo-populations from the RPT model by using the above parameters' specification.

Model's competitors

We compare the proposed model with two competitors: an under-parameterised alternative, $\{\rho_t, \phi, p_t\}$, and an over-parameterised alternative, $\{[\rho_{t\times h}, \phi_h, p_{t+h}]_{G=3}\}$, with capture probabilities modelled as in equation (4.1) and temporal random effects τ_t supposed to be shared by the three components.

Notably, model $\{[\rho_{t\times h}, \phi_h, p_{t+h}]_{G=3}\}$ resembles the 3-class structure of the RPT model, but without the use of cross-classification. Thus, in our opinion, this model may represent a Pledger et al. (2010)'s benchmark alternative to the proposed model.

On the other hand, model $\{\rho_t, \phi, p_t\}$ excludes the presence of different classes of individuals in the population. As already noted (cfr. Section 5.2), such a kind of model is probably unsuitable to characterise the target population, since it does not account for heterogeneity; nevertheless, it can still be interesting to see what happens if the underlying structure of the population is ignored in this context.

Prior specifications

We briefly illustrates the prior specification for the RPT model's parameters.

Following the methodology illustrated in Chapter 3, we use the following conditional specification for the survival probabilities:

$$\begin{aligned}\phi_{TR} &\sim Beta(1,\, 2)\\ \phi_{NTR}|\phi_{TR} &\sim tBeta(1,\, 1,\, \phi_{TR},\, 1)\,,\end{aligned}$$

with the latter marginally yielding $\phi_{NTR} \sim Beta(2,\, 1)$. This enforces $\phi_{TR} < \phi_{NTR}$ and also induces a slight repulsion between the two parameters. The two sets of recruitment parameters are ascribed the priors $\rho_{TR} \sim Beta(1,\, T)$ and $\rho_{NTR,\,t} \sim Beta(1/T, 2 - t/T)$, $t = 2, \ldots, T$. A standard Dirichlet distribution is placed on the group weights vector, namely $w \sim Dirichlet(3,\, (1, 1, 1))$. In absence of relevant prior information, we consider $\delta \sim Unif[0,\, 1]$. Regarding the parameters of the detection process, we consider $\mu \sim N(0,\, 10)$ and $\tau_t \sim N(0,\, 0.25)$, $\forall t$.

The priors on the parameters of the other two candidate models basically resemble their counterparts used for the RPT model. Notably, the survival parameter of model $\{\rho_t, \phi, p_t\}$ is provided with a standard $Unif[0, 1]$, while constrained Uniform priors are chosen for the survival parameters of model $\{[\rho_{t\times h}, \phi_h, p_{t+h}]_{G=3}\}$ (cfr. Chapter 3).

Model fitting

Each simulated dataset is augmented by 500 all-zero capture histories for the implementation of the PX-DA approach and we fit the three considered models on each of them, separately. In particular, estimation is carried out using JAGS and the BUGS codes for the three considered model is reported in Listing A.14 and A.15 in Appendix A. We run 2 parallel chains with $20,000$ iterations each, discarding $15,000$ as burn-in.

5.3.2 Results

Figure 5.1 shows the differences between the estimated ($\hat{N}_{super}$) and true (N_{super}) super-population size for each of the $K = 50$ replicas by all considered models. To allow for a meaningful comparison between scenarios having expected super-population sizes of different magnitudes, the error is divided by the expected value of N_{super} of the corresponding scenario. We notice that the RPT uniformly provides the best results overall. On the contrary, the super-population size is consistently overestimated by the model $\{[\rho_{t\times h}, \phi_h, p_{t+h}]_{G=3}\}$, while it is always underestimated by the simpler model $\{\rho_t, \phi, p_t\}$.

Table 5.2 reports some summaries useful to assess models' performance in estimating N_{super}, along with the WAIC associated to each model. When the true model is fitted, we also report the median multi-AUC (mAUC) for each scenario. Notice that MAE and CIW are scale-dependent measures and do not allow for comparisons between scenarios. Again, we consider relative versions of these measures by dividing both by the corresponding expected value of N_{super}. RPT model returns

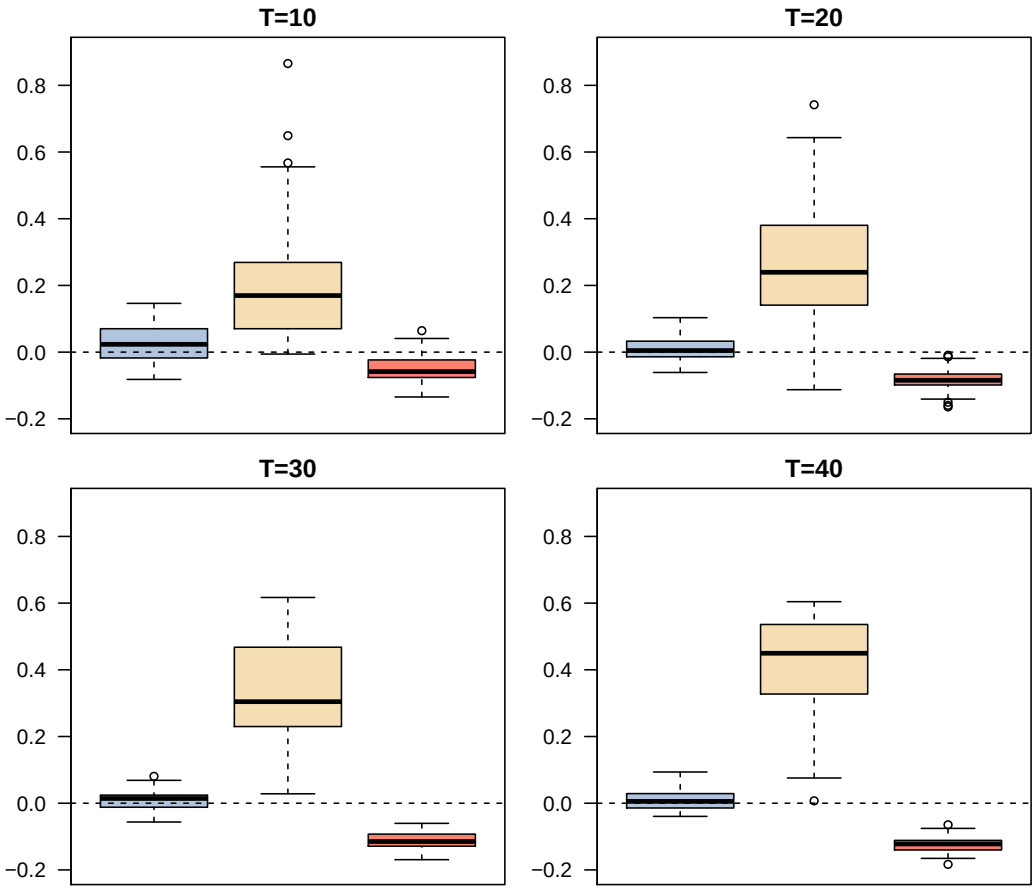


Figure 5.1. Relative estimation error of the super-population (N_{super}) abundance for increasing number of sampling occasions (T), calculated for each of the $K = 50$ independent replicas, by the three competing models: RPT (*lightblue*), $\{[\rho_{t\times h}, \phi_h, p_{t+h}]_{G=3}\}$ (*wheat*) and $\{\rho_t, \phi, p_t\}$ (*salmon*).

the lowest relative MAEs in each scenario, with the error decreasing as the number of occasions increases. The WAIC tends to favour model $\{\rho_t, \phi, p_t\}$ over the other two alternatives only when the number of sampling occasions is rather small, i.e. $T = 10$. This is not alarming as it is very hard to estimate such a complex set of interplaying parameters when the amount of capture occasions is scarce. In particular, it is extremely complicated to distinguish individuals with low and high survivals when the occasions span only one year. The WAIC favours the true model in all other scenarios, yielding the lowest median score and also selecting it (i.e. returning the lowest WAIC) in most of the replicas when $T > 10$ (96% when the number of sampling occasions is the largest).

The classification performances of the RPT model are quite good in all the scenarios, with the mAUC improving as T increases. Notably, the resulting median mAUC is always greater or equal than 0.84 and it is above 0.95 when at least one change of year occurs. Furthermore, by assigning the labels to each individual according to the Maximum a Posteriori (MAP) rule, the median accuracy (across replicas) is between 72% and 92% in all scenarios (once again improving as T

	RPT						$\{\rho_t, \phi, p_t\}$					$\{[\rho_{t\times h}, \phi_h, p_{t+h}]_{G=3}\}$				
T	$\mathbf{MAE_{rel}}$	**Cov.**	$\mathbf{CIW_{rel}}$	**WAIC**	**%waic**	**mAUC**	$\mathbf{MAE_{rel}}$	**Cov.**	$\mathbf{CIW_{rel}}$	**WAIC**	**%waic**	$\mathbf{MAE_{rel}}$	**Cov.**	$\mathbf{CIW_{rel}}$	**WAIC**	**%waic**
10	0.05	1	0.23	1613.7	16	0.84	0.06	0.70	0.14	1589.0	68	0.21	0.84	0.69	1603.5	16
20	0.03	1	0.15	3173.7	80	0.95	0.08	0.12	0.09	3293.2	2	0.27	0.64	0.67	3229.0	18
30	0.03	1	0.15	4640.7	92	0.97	0.11	0	0.07	4984.1	0	0.34	0.38	0.64	4782.7	8
40	0.02	0.98	0.14	6365.6	96	0.98	0.12	0	0.06	6869.8	0	0.42	0.12	0.57	6542.2	4

Table 5.2. Estimates of relative MAE ($\mathrm{MAE_{rel}}$), coverage (Cov.) and relative average width of the 95% credible intervals ($\mathrm{CIW_{rel}}$) for N_{super}, median WAIC and percentage of times each competing model has achieved best WAIC (%waic). When RPT model is fitted, median multi-AUC (mAUC) is reported. All these summaries have been obtained when data are simulated from the RPT model. Summaries concern data that are simulated from the RPT model.

increases).

Table 5.3 shows the RPT model performance in estimating some time-constant parameters. The estimates of ϕ_{NTR} and δ have a very small MAE but fail in attaining the nominal credible interval coverage of 95% in most scenarios. This indicates a good accuracy of the point-estimates but a slight over-confidence (visible in the low average CIWs), which leads to a reduction of the nominal coverage. Nonetheless, they both settle to a fair and acceptable level. All things considered, this is not particularly surprising as such component-specific parameters are governing a latent process and are potentially mutually confounded. The average credible interval width of ϕ_{TR} is high for the first scenario (i.e. $T = 10$), but it rapidly decreases in the following scenarios, indicating that the transient component is indeed well separated from the non-transient one whenever the study is conducted for a large period (i.e. at least two year of sightings).

Finally, we verify the overlapping between the posterior distribution of the two survival probabilities, i.e. ϕ_{TR} and ϕ_{NTR}. They are well separated in all scenarios, with an overlapping index (OV) estimated to be equal to 0.015 when $T = 10$ and equal to 0 for larger T.

	ϕ_{TR}			ϕ_{NTR}			δ			μ		
T	**MAE**	**Cov.**	**CIW**	**MAE**	**Cov.**	**CIW**	**MAE**	**Cov.**	**CIW**	**MAE**	**Cov.**	**CIW**
10	0.06	1	0.59	0.04	0.72	0.08	0.08	0.82	0.29	0.15	1	1.04
20	0.02	1	0.11	0	0.82	0.01	0.05	0.50	0.10	0.10	0.98	0.56
30	0.02	0.96	0.11	0	0.96	0.01	0.03	0.60	0.07	0.05	1	0.44
40	0.01	1	0.07	0	0.88	0	0.02	0.72	0.06	0.04	1	0.37

Table 5.3. Estimates of MAE, coverage (Cov.) and average width of the 95% credible intervals (CIW) for N_{super} for some time-constant parameters of model RPT. Summaries concern data that are simulated from the RPT model.

5.4 Real data illustration: *Tursiops truncatus*

We consider a population of common bottlenose dolphins (*Tursiops truncatus*) inhabiting the area of the Tiber river estuary (Mediterranean Sea, Rome, Italy)

(Pace et al., 2021). The coastal area and nearby regions surrounding the Tiber River estuary represent a suitable habitat for bottlenose dolphins, despite their proximity to the city of Rome - one of the major urban centers in Europe.

Boat-based daily surveys to collect photographic data of the dolphins were conducted between June 2018 and November 2020 in favorable weather conditions to reduce the detection probability bias. The photo-identification technique was used to identify unique individuals over multiple sampling occasions and to build single encounter histories. The individual distinctiveness was scored as *well-marked* (individuals with highly distinctive dorsal fins and scars on the body), *fairly marked* (individuals with moderately distinctive dorsal fins), and *unmarked* (individuals with no distinctive features on dorsal fins). We will only consider well-marked individuals in this analysis, since they are the ones for whom identification is deemed to be reliable. The resulting data matrix consists in the detection histories of $D = 195$ well-marked dolphins that have been sighted in the area between June 2018 and November 2020, for a total of $T = 87$ occasions.

We apply the RPT model to estimate the total population size of the common bottlenose dolphins inhabiting the Tiber river estuary in the considered time-window. Since time-lag between capture occasions is really variable, we adapt the survival mechanism to describe not equally spaced capture occasions, as already discussed in Section 5.2.

After some preliminary runs with numbers of all-zero rows, we finally choose to add 500 rows of pseudo-individuals (i.e. null capture histories), thus yielding an augmented data matrix of 695 rows. Further additions of all-zero rows led to similar results but strain the computational burden, both in terms of runtime and storage.

As for the simulation study, we compare the performances of the RPT model with a lower and higher-parameterised alternative: $\{\rho_t, \phi, p_t\}$ and $\{[\rho_{g\times t}, \phi_g, p_{g+t}]_{G=3}\}$, respectively. The comparison among these three models allows testing whether there is unobserved heterogeneity or not, and if the more parsimonious RPT is a better fit than the more general $\{[\rho_{g\times t}, \phi_g, p_{g+t}]_{G=3}\}$ model.

We consider the same prior settings used in Section 5.3 and we run 2 parallel chains, each with $20,000$ iterations with a burn-in of $10,000$ and no thinning.

Table 5.4 reports the results on abundance estimation along with the WAIC associated to each competing model. Notice that the same labels of the RPT model are used for the classes of model $\{[\rho_{g\times t}, \phi_g, p_{g+t}]_{G=3}\}$, since its estimated mixture components identify three classes of individuals that resembles the ones which characterised the RPT model: the first with low survival and capture probabilities, the second with similar capture probability but higher survival and the third with both higher survival and capture probabilities. It is important to stress that one could have placed order constraints on the group-dependent intercepts η_g's of model $\{[\rho_{g\times t}, \phi_g, p_{g+t}]_{G=3}\}$ - i.e. $\eta_1 < \eta_2 < \eta_3$ - to better identify the three components; however, in this case it was not necessary since the estimated $\hat{\eta}_1$, $\hat{\eta}_2$ and $\hat{\eta}_3$ do respect this sorting, with a difference between η_1 and η_2 that is actually negligible. On the other hand, the estimated posterior distributions of the three survival parameters do not overlap each other.

Notably, the RPT model yields an estimated abundance of well-marked in-

Model	WAIC	$\hat{N}_{super}$	$\hat{N}_{\mathrm{R}}$	$\hat{N}_{\mathrm{P}}$	$\hat{N}_{\mathrm{T}}$
RPT	5045.5	299 (256, 343)	62 (51, 73)	87 (64, 105)	150 (99, 204)
$\{[\rho_{t\times g}, \phi_g, p_{t+g}]_{G=3}\}$	5071.4	517 (357, 638)	53 (45, 60)	90 (56, 169)	379 (154, 506)
$\{\rho_t, \phi, p_t\}$	5101.8	227 (213, 242)	-	-	-

Table 5.4. WAIC and estimated super-population abundance for the whole population and by group. 95% high posterior credible interval is reported along with the point estimate.

dividuals of $\hat{N}_{super} = 299$ ($CI_{0.95} = [256, 343]$), with yearly variations $\hat{N}_y$ ($y = 2018, 2019, 2020$) that show a peak in 2019 and a decrease in the last year of observation. The same trend is observed for the other two candidate models, as shown by Figure 5.2. However, the estimated abundance of well-marked individuals in 2019 is much higher and associated with much more uncertainty when model $\{[\rho_{g\times t}, \phi_g, p_{g+t}]_{G=3}\}$ is adopted: this yearly estimate seems to be the main component that affects the final abundance estimate, by dragging it to a very high value (i.e. $\hat{N}_{super} = 517$).

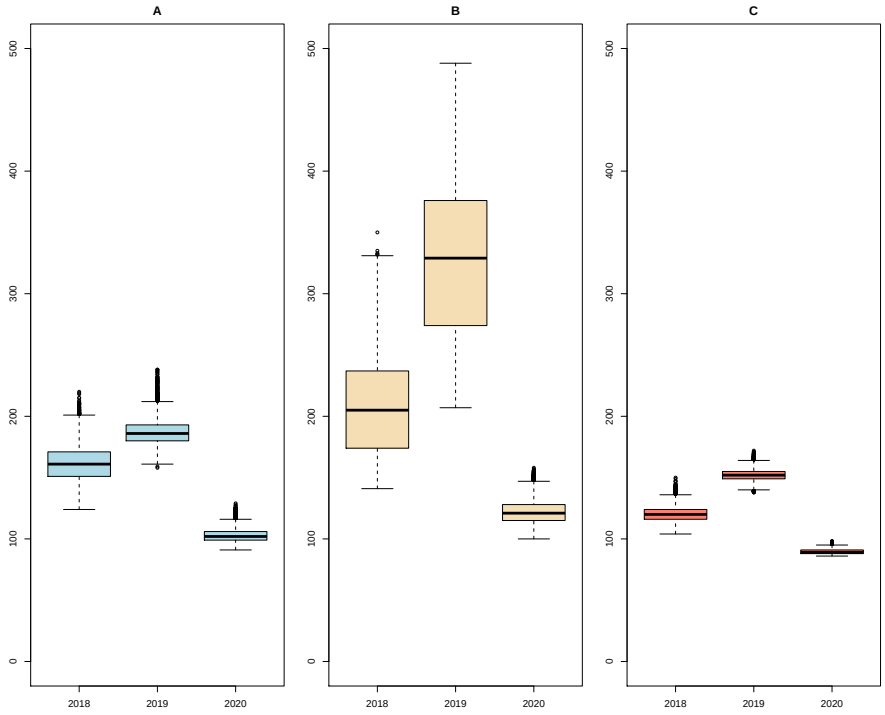


Figure 5.2. Boxplots of the estimated super-population size by year for model RPT (A), $\{[\rho_{t\times g}, \phi_g, p_{t+g}]_{G=3}\}$ (B) and $\{\rho_t, \phi, p_t\}$ (C).

As previously shown by Table 5.4, the RPT model does yield the lowest WAIC score. This suggests the presence of unobserved heterogeneity in the population, which is better described by the more parsimonious structure of the RPT model than by the more generic 3-class specification. The RPT model is thus selected and the associated estimates are analysed more in depth.

The two annual survival probabilities are well separated, with an overlapping index equal to 0. Posterior estimates are $\hat{\phi}_{TR} = 6.6 \times 10^{-7}$ ($CI_{0.95} = [0, 5.5 \times 10^{-6}]$) and $\hat{\phi}_{NTR} = 0.69$ ($CI_{0.95} = [0.60, 0.78]$). Notice that the estimated smaller survival parameter $\hat{\phi}_{TR} = 6.6 \times 10^{-7}$ is of little interpretability in annual scale, but it corresponds to a probability of 0.76 on weekly scale and of 0.31 in monthly scale.

The parameter δ regulating the undetectability of the part-time group is estimated to be $\hat{\delta} \approx 0.75$. This means that individuals in that group - when alive - are present in the area approximately 25% of the times. The resulting average capture probability of part-time individuals is 0.0425, while for the rest of the individuals is about four times larger. The corresponding temporal variations, captured by τ_t, results in the time-dependent posterior distributions of $p_{PT,t}$ and $p_{NPT,t}$ reported in Figure 5.3.

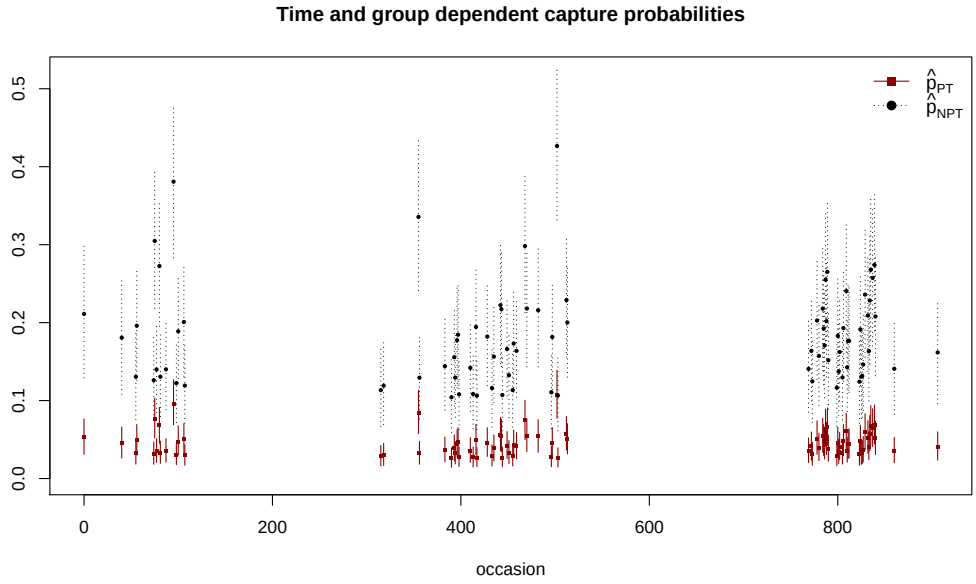


Figure 5.3. Posterior median and 95% credible intervals of the two capture probabilities.

It is interesting to see in Figure 5.4 how the recruitment of resident and part-time individuals is high at the first occasion and then rapidly decreases to low values. This indeed accounts for all individuals that were already present in the population before the starting of the survey and that are virtually recruited at the first occasion. On the other hand, the estimated recruitment parameter of transient individuals is 0.032, with 95% credible interval (0.019, 0.046).

By assigning the 195 well-marked individuals that were encountered between June 2018 and November 2020 to a single group according to the *Maximum a Posteriori* (MAP) allocation, we have that: 48 are assigned to the group of resident individuals, 56 to the group of part-time individuals, and 91 to the group of transient individuals. The fuzzy nature of this clustering allows to quantify how decisively each individual is assigned to one group or the other. For instance, 90% of the individuals classified into the first group have been assigned to it with a probability greater than 0.9. This probability is greater than 0.54 and than 0.53 for the individuals assigned to the

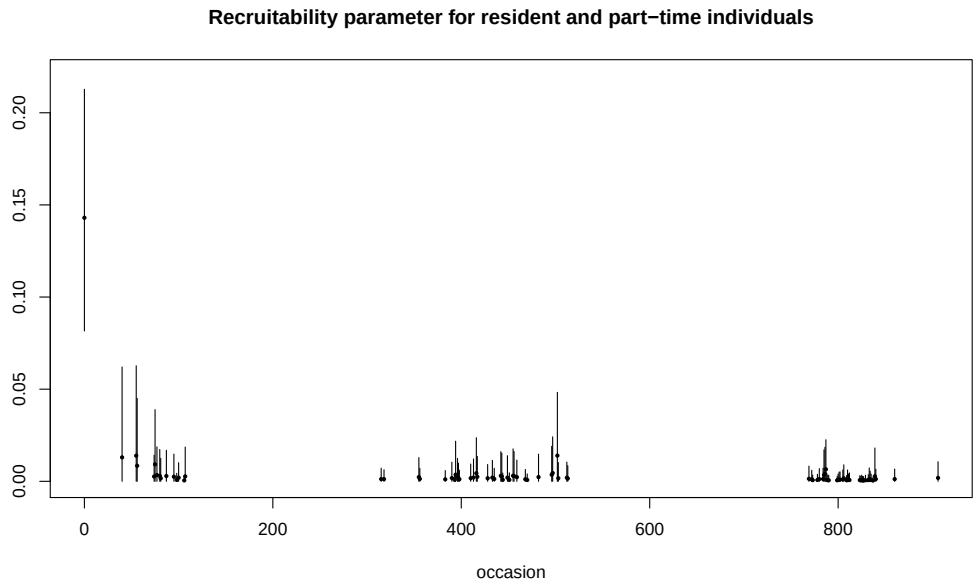


Figure 5.4. Posterior median and 95% credible intervals of the recruitment parameter for resident and part-time individuals (i.e. $\rho_{NTR,t}$).

second and to the third group, respectively. Given the little information provided by their capture histories, the classification of the individuals deemed to be transient is slightly more vague compared to the other groups, especially for those that have only been observed in the last sampling occasions. Remarkably, the classification of the individuals who have been assigned to the part-time group is crisp if they have been encountered for the first time toward the beginning of the study period, while it is dicey when they have been only encountered for the first time at the end of the study: they indeed might be either new part-time individuals that have joined the population from a year to another or transient individuals who are bounded not to show up in future occasions. A longer time-window should be able to clarify the role of these individuals in the population.

It is relevant to notice that the proportions of the three groups do not align with the estimated weights $\hat{\mathbf{w}} = (0.32, 0.45, 0.23)$. This is because the prior weights are not a good indicator of the true impact of each component in the super-population, but only of their impact on the pseudo-population: individuals with all-zero histories cannot be reliably assigned to any group since they are merely the result of the data augmentation (indeed, they have never been encountered - or even been part of the population).

Figure 5.5 show that the classification of the observed individuals is able to highlight three well-distinguishable patterns of capture histories, that indeed seem to comply with the typical RPT behavior. Resident individuals tend to be available in the area until the end of the study period and are spotted very frequently in subsequent sampling occasions. Part-time individuals are available in the area for the most part of the study period and are frequently spotted, but not as often as resident ones. Finally, transient individuals show short capture histories with few

captures that never cross two years; this reflects the fact that they are spotted only few times and tend to stay in the population only for a short time period.

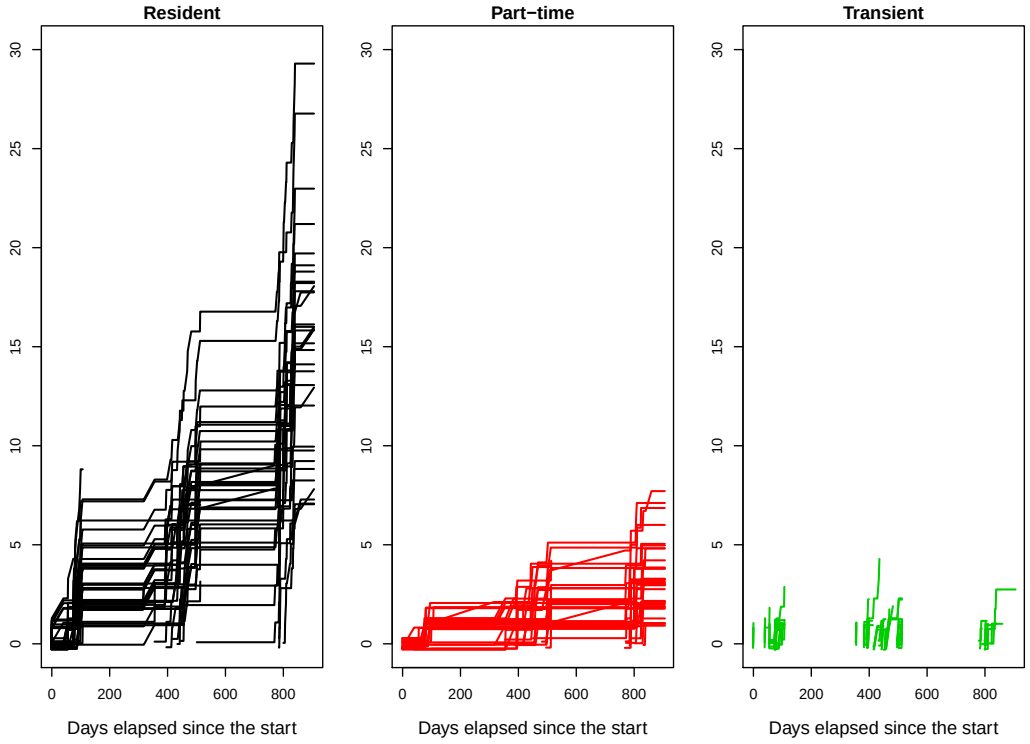


Figure 5.5. Cumulative frequencies of detection of the encountered individuals divided by group. The x-axis report the number of days elapsed since the start of the survey: the begin and the end of each line correspond to the days when an individual is estimated to have (respectively) joined and left the population.

5.5 Discussion

Estimating the abundance of a marine wildlife species is a challenging but crucial activity that can tell much about the undergoing ecological processes. Thus, combining high-quality data with solid analytical approaches is essential to improve our knowledge of these dynamics and increase the potential for management actions (Vella et al., 2021).

In this chapter, we have pursued the estimation of the size of a common bottlenose dolphin population which is known to be organised in three groups with different residency and site-fidelity patterns. Accounting for such unobserved heterogeneity is a very common problem in the environmental literature. In this context, the main concern has not been whether such a kind of population structure exists or not, since the issue had already been investigated in several works. Here, we have decided to move a step forward, by proposing a suitable and parsimonious model to characterise such a type of population. Indeed, most of the current literature on the topic tends to adopt two-step approaches, where parameter estimation and classification are

carried out separately. We believe that this does not yield a quantification nor a propagation of the uncertainty of the classification step onto the modeling step. Thus, we have considered a one-step approach based on finite mixture models, where estimation and classification are performed jointly. Moreover, we have carried out all the analysis in a Bayesian framework; this is pretty unusual in most of the marine ecology literature about Capture-Recapture, where only few approach the problem from the Bayesian perspective and develop ad-hoc solutions based on prior scientific knowledge of the population of interest. Indeed, we believe that this has been an opportunity to give some insights about the Bayesian implementation of models that are commonly used by practitioners and, more specifically, ecologists.

Notably, we have proposed a parsimonious specification of a finite mixture model within the PX-DA setting for Capture-Recapture analysis, that we have named RPT. This specification reflects the typical bottlenose dolphin residency pattern, with individuals showing high, partial or low site-fidelity (Dinis et al., 2016; Hunt et al., 2017; Haughey et al., 2020; La Manna et al., 2022). Its adoption, characterised by less free-parameters, simplifies the identification of all the models' components with respect to more generic and flexible alternatives (Pledger et al., 2010). In addition, it also allows for the convenient use of the classes of prior specifications presented in Chapter 3. We have verified the performances of our proposal through a simulation study in a setting which attempts to retrace a real situation, where encounters do not necessarily occur at evenly spaced sampling occasions. We have considered scenarios with increasing lengths of the study period, and hence of the capture histories (i.e. mimicking a continuous monitoring of a population of interest over time). We have observed the influence of the number of sampling occasions on the estimates' quality: as expected, the larger the sample size the better is the accuracy and coverage of the estimates. Comparison with alternative model versions over many replicas has shown that the model estimation exhibits satisfying and robust performances when the observation window is long enough (more than one year of monitoring). Finally, we have estimated the RPT model and two other alternatives on a real dataset concerning a common bottlenose dolphin population. The analysis has shown that the proposed model outplays both under and over-parameterised versions in terms of the WAIC score. The results correspond to biologically meaningful findings and are in line with those of previous research works. The estimated abundance is of $\hat{N}^*_{super} = 299$, close to the one obtained by Pace et al. (2021) for the well-marked population in the triennium 2018-2020. The new estimates better quantify the overall uncertainty of the model and this is particularly important in this context as the major interest lies in estimating a reasonable range that account for both the worst and best case scenario.

Recall that the reported estimates only refer to well-marked individuals, who represent only a proportion of the whole population (approximately 60% of all sighted individuals across the three years). This choice has been made so as to consider those individuals that were correctly identifiable through natural markings. Although this was not the focus here, inferring the overall size from the well-marked abundance estimate is possible; further details can be found in Wilson et al. (1999). However, it is important to point out that the process of photo-identification of the individuals may be affected by the temporal evolution of natural markings. In this regard, several solutions have been proposed in literature to account for uncertainty

on individual photo-identification (Yoshizaki et al., 2009; Tancredi et al., 2013).

One basic assumption of CR experiments is that captures of different individuals are independent one another. However, this assumption may be sometimes unrealistic for species that are gregarious: for example, bottlenose dolphin populations can form structured societies with complex social networks (Pace et al., 2022). In future developments, we would like to build a model which relax the assumption of independence among captures by incorporating - when available - information on the social structure of the population. It would also be interesting to include the effect of external or individual covariates. For instance, it is possible for some individuals to recognise their gender and age-class through the photo-identification. This partial information may be incorporated in the Bayesian framework and could help for a better assessment, for example, of the membership of different individuals in different groups or of their marking probability (Wu et al., 2021). Last but not least, it would be extremely interesting to conduct the survey on a larger spatial scale and include external information to account for the spatial heterogeneity (Wu and Holan, 2017).

Appendix A

BUGS code listings

In this appendix, we report the BUGS code listings of the models which have been fitted via JAGS throughout the thesis. All the models are formulated in terms of the PX-DA formalisation of Royle et al. (2007).

A.1 Notation

In the following, we define some notation that will be used in most of the considered BUGS code listings. Specifically, `T` is the number of encounter occasions or - in the case of a robust design - number of primary sampling periods, while `M` is the number of pseudo-individuals in the augmented data matrix. Moreover, we will consider the following nodes:

- `psi`: inclusion probability;
- `p`: constant capture probability;
- `phi`: constant survival probability;
- `rho`: constant recruitment parameter.
- `z[i]`: binary variable that indicates whether the i-th individual is part of the population (for closed populations);
- `r[i,t]`: binary variable that indicates whether the i-th individual is recruitable into the population at time t (for open populations);
- `z[i,t]`: binary variable that indicates whether the i-th individual is part of the population at time t (for open populations);
- `Nind[i]`: number of occasions the i-th individual was in the population;
- `Nalive[i]`: binary variable which counts whether the i-th individual belongs to the super-population;
- `N[t]`: number of individuals in the population at time t;
- `Nsuper`: size of the super-population.

Notice that in BUGS the function `dnorm(mu, tau)` is associated with a normal distribution with mean `mu` and precision `tau`. Further details about the BUGS language and JAGS are available in Plummer et al. (2003).

A.2 Models from Chapter 2:

A.2.1 Basic closed-population model

Since data consists in a zero-inflated vector of detection frequencies, `y[i]` indicates the frequency of detection of the i-th individual. The population size is the node `N`.

```
# Priors

p ~ dunif(0,1)
psi ~ dunif(0,1)

# Likelihood

for (i in 1:M){

    z[i] ~ dbern(psi)
    y[i] ~ dbin(z[i]*p,T)

}

# Derived parameters

N = sum(z[1:M])
```

Listing A.1. Closed-population CR model with constant capture probability (model M_0).

A.2.2 Basic open-population model

Notation:

- `y[i,t]`: binary variable that indicates whether the i-th individual is detected at occasion t;
- `muy[i,t]`: capture probability at time t for the i-th individual of the pseudo-population. It is null if the individual i does not belong to the population at time t;
- `muz[i,t]`: probability for the i-th individual of being in the population at time t. It is null if, at time t, the individual i was already recruited in the past but was not into the population at time $t-1$.

```
#Priors

rho ~ dbeta(1,T)  #prior Dorazio (2020)
phi ~ dunif(0,1)
p ~ dunif(0,1)

# Likelihood

for (i in 1:M){

    y[i,1] ~ dbin(muy[i,1],1)
```

```
    muy[i,1] = z[i,1]*p
    z[i,1] ~ dbin(rho,1)
    r[i,1] = 1

    for (t in 2:T){

      y[i,t] ~ dbin(muy[i,t],1)
      muy[i,t] = z[i,t]*p
      z[i,t] ~ dbin(muz[i,t],1)
      muz[i,t] = phi*z[i,t-1] + rho*r[i,t]
      r[i,t] = r[i,(t-1)]*(1-z[i,t-1])

    }

  }

  # Derived parameters

  for(t in 1:T){
    N[t] = sum(z[1:M,t])
  }

  for (i in 1:M){

    Nind[i] = sum(z[i,1:T])
    Nalive[i] = 1-equals(Nind[i],0)

  }

Nsuper = sum(Nalive[1:M])
```

Listing A.2. Open-population Jolly-Seber model with constant recruitment, survival and detection parameters.

A.3 Models from Chapter 3:

Here, we introduce the node `clust[i]` which indicates the cluster the i-th individual belongs to. The constant `G` is the number of components of the mixture models. Node `w[g]` contains the weight of the g-th component of the mixture, while node `p[g]` contains the capture probability of individuals in the g-th class.

```
# Priors

p[1] ~ dunif(0,1)
p[2] ~ dunif(0,1)

psi ~ dunif(0,1)

w ~ ddirch(rep(1,2))

# Likelihood

for (i in 1:M){

    clust[i] ~ dcat(w[1:2])
    z[i] ~ dbern(psi)
    y[i] ~ dbin(z[i]*p[clust[i]],T)

}

# Derived parameters

```

```
N = sum(z[1:M])
```

Listing A.3. 2-class finite mixture CR model for closed populations.

Notice that the priors on p_1 and p_2 correspond to setting I of Section 3.5, but other prior specifications can be used simply by modifying the code above. For example, one could use a constrained specification based on Uniform priors, such as

```
p[1] ~ dunif(0,1)
p[2] ~ dbeta(1,1) T(p[1],1)
```

which corresponds to setting IV of Section 3.5. Notice that the truncation `T( , )` is used to bound the support of the uniform distribution (i.e. $Beta(1,1)$) into the set $[p_1, 1]$.

A.4 Models from Chapter 4:

All the following open-population models consider a time-varying recruitment parameter, ρ_t, indicated through the node `rho[t]`. The constant node `J[t]` indicates the number of secondary sampling occasions in the t-th primary sampling period, for $t = 1, \ldots, T$.

Notice that, here, the stochastic node `y[i,t]` corresponds to the number of secondary sampling occasions the i-th individual is encountered during the t-th primary sampling period.

A.4.1 Homogeneous capture and survival probabilities

```
#Priors

phi ~ dunif(0,1)
p ~ dunif(0,1)

for(t in 1:T){
    rho[t] ~ dbeta(1/T,2-t/T)  #prior Dorazio (2020)
}

# Likelihood

for (i in 1:M){

    y[i,1] ~ dbin(muy[i,1],J[1])
    muy[i,1] = z[i,1]*p
    z[i,1] ~ dbin(rho[1],1)
    r[i,1] = 1

    for (t in 2:T){

      y[i,t] ~ dbin(muy[i,t],J[t])
      muy[i,t] = z[i,t]*p
      z[i,t] ~ dbin(muz[i,t],1)
      muz[i,t] = phi*z[i,t-1] + rho[t]*r[i,t]
      r[i,t] = r[i,(t-1)]*(1-z[i,t-1])

    }

  }

```

```
  # Derived parameters

  for(t in 1:T){
    N[t] = sum(z[1:M,t])
  }

  for (i in 1:M){

    Nind[i] = sum(z[i,1:T])
    Nalive[i] = 1-equals(Nind[i],0)

  }

Nsuper = sum(Nalive[1:M])
```

Listing A.4. CR model for open populations with homogeneous capture and survival probabilities, i.e. $\{\rho_t, \phi, p\}$.

A.4.2 Finite mixture models with single mixture grouping

Nodes `p[g]` and `phi[g]` will be used to indicate the g-th component-specific capture and survival probability, respectively.

```
# Priors

phi ~ dunif(0,1)

p[1] ~ dunif(0,1)

for (g in 2:G){
    p[g] ~ dbeta(1,1) T(p[g-1],1)
}

for(t in 1:T){
    rho[t] ~ dbeta(1/T,2-t/T)
}

w ~ ddirch(rep(1,G))

# Likelihood

for (i in 1:M){

    clust[i] ~ dcat(w[1:G])

    y[i,1] ~ dbin(muy[i,1],J[1])
    muy[i,1] = z[i,1]*p[clust[i]]
    z[i,1] ~ dbin(rho[1],1)
    r[i,1] = 1

    for (t in 2:T){

      y[i,t] ~ dbin(muy[i,t],J[t])
      muy[i,t] = z[i,t]*p[clust[i]]
      z[i,t] ~ dbin(muz[i,t],1)
      muz[i,t] = phi*z[i,t-1] + rho[t]*r[i,t]
      r[i,t] = r[i,(t-1)]*(1-z[i,t-1])

    }

  }

  # Derived parameters

```

```
  for(t in 1:T){
    N[t] = sum(z[1:M,t])
  }

  for (i in 1:M){

    Nind[i] = sum(z[i,1:T])
    Nalive[i] = 1-equals(Nind[i],0)

  }

Nsuper = sum(Nalive[1:M])
```

Listing A.5. Finite mixture CR model for open populations with heterogeneous capture probabilities, i.e. $\{\rho_t, \phi, p_{h_G}\}$. When $G = 2$, alternative constrained prior specification can be also used (see Section 3.4).

```
# Priors

p ~ dunif(0,1)

phi[1] ~ dunif(0,1)

for (g in 2:G){
    phi[g] ~ dbeta(1,1) T(phi[g-1],1)
}

for(t in 1:T){
    rho[t] ~ dbeta(1/T,2-t/T)
}

w ~ ddirch(rep(1,G))

# Likelihood

for (i in 1:M){

    clust[i] ~ dcat(w[1:G])

    y[i,1] ~ dbin(muy[i,1],J[1])
    muy[i,1] = z[i,1]*p
    z[i,1] ~ dbin(rho[1],1)
    r[i,1] = 1

    for (t in 2:T){

      y[i,t] ~ dbin(muy[i,t],J[t])
      muy[i,t] = z[i,t]*p
      z[i,t] ~ dbin(muz[i,t],1)
      muz[i,t] = phi[clust[i]]*z[i,t-1] + rho[t]*r[i,t]
      r[i,t] = r[i,(t-1)]*(1-z[i,t-1])

    }

  }

  # Derived parameters

  for(t in 1:T){
    N[t] = sum(z[1:M,t])
  }

  for (i in 1:M){

    Nind[i] = sum(z[i,1:T])
    Nalive[i] = 1-equals(Nind[i],0)
```

```
50
51    }
52
53  Nsuper = sum(Nalive[1:M])
```

Listing A.6. Finite mixture CR model for open populations with heterogeneous survival probabilities, i.e. $\{\rho_t, \phi_{h_G}, p\}$. When $G = 2$, alternative constrained prior specification can be also used (see Section 3.4).

```
1   # Priors
2
3   p[1] ~ dunif(0,1)
4   phi[1] ~ dunif(0,1)
5
6   for (g in 2:G){
7       p[g] ~ dbeta(1,1) T(p[g-1],1)
8       phi[g] ~ dbeta(1,1)
9   }
10
11  for(t in 1:T){
12      rho[t] ~ dbeta(1/T,2-t/T)
13  }
14
15  w ~ ddirch(rep(1,G))
16
17  # Likelihood
18
19  for (i in 1:M){
20
21      clust[i] ~ dcat(w[1:G])
22
23      y[i,1] ~ dbin(muy[i,1],J[1])
24      muy[i,1] = z[i,1]*p[clust[i]]
25      z[i,1] ~ dbin(rho[1],1)
26      r[i,1] = 1
27
28      for (t in 2:T){
29
30        y[i,t] ~ dbin(muy[i,t],J[t])
31        muy[i,t] = z[i,t]*p[clust[i]]
32        z[i,t] ~ dbin(muz[i,t],1)
33        muz[i,t] = phi[clust[i]]*z[i,t-1] + rho[t]*r[i,t]
34        r[i,t] = r[i,(t-1)]*(1-z[i,t-1])
35
36      }
37
38    }
39
40    # Derived parameters
41
42    for(t in 1:T){
43      N[t] = sum(z[1:M,t])
44    }
45
46    for (i in 1:M){
47
48      Nind[i] = sum(z[i,1:T])
49      Nalive[i] = 1-equals(Nind[i],0)
50
51    }
52
53  Nsuper = sum(Nalive[1:M])
```

Listing A.7. Finite mixture CR model for open populations with heterogeneous capture and survival probabilities, i.e. $\{\rho_t, [\phi_h, p_h]_G\}$. When $G = 2$, alternative constrained prior specification can be also used (see Section 3.4).

A.4.3 Cross-classified models

When some components share common parameters, a harmless trick is needed in BUGS to select the correct survival or capture probability for each individual (see lines 21-22 of Listing A.8). This trick is necessary as an *if-else* construct is not provided in BUGS. Notably, for this purpose, we introduce two further classes of nodes:

- `p_ind[i]`: capture probability for the i-th individual;
- `phi_ind[i]`: survival probability for the i-th individual.

```
1  # Priors
2
3  phi[1] ~ dbeta(1,2)
4  phi[2] ~ dbeta(1,1) T(phi[1],1)
5
6  p[1] ~ dbeta(1,2)
7  p[2] ~ dbeta(1,1) T(p[1],1)
8
9  for(t in 1:T){
10     rho[t] ~ dbeta(1/T,2-t/T)
11  }
12
13 w ~ ddirch(rep(1,4))
14
15 # Likelihood
16
17 for (i in 1:M){
18
19     clust[i] ~ dcat(w[1:4])
20
21     p_ind[i] = (3-clust[i])*(4-clust[i])*3^(clust[i])/18*p[1]+
22                (clust[i]-1)*(clust[i]-2)*3^(5-clust[i])/18*p[2]
23     phi_ind[i] = (clust[i]-2)^2*(4-clust[i])*clust[i]/3*phi[1]+
24                  (clust[i]-3)^2*(1-clust[i])*(clust[i]-5)/3*phi[2]
25
26     y[i,1] ~ dbin(muy[i,1],J[1])
27     muy[i,1] = z[i,1]*p_ind[i]
28     z[i,1] ~ dbin(rho[1],1)
29     r[i,1] = 1
30
31     for (t in 2:T){
32
33       y[i,t] ~ dbin(muy[i,t],J[t])
34       muy[i,t] = z[i,t]*p_ind[i]
35       z[i,t] ~ dbin(muz[i,t],1)
36       muz[i,t] = phi_ind[i]*z[i,t-1] + rho[t]*r[i,t]
37       r[i,t] = r[i,(t-1)]*(1-z[i,t-1])
38
39     }
40
41   }
42
43   # Derived parameters
```

```
44
45   for(t in 1:T){
46     N[t] = sum(z[1:M,t])
47   }
48
49   for (i in 1:M){
50
51     Nind[i] = sum(z[i,1:T])
52     Nalive[i] = 1-equals(Nind[i],0)
53
54   }
55
56 Nsuper = sum(Nalive[1:M])
```

Listing A.8. CR 4-class cross-classified model for open populations, i.e. $\{\rho_t, \phi_{h_2}, p_{h_2}\}$.

```
1  # Priors
2
3  phi[1] ~ dbeta(1,2)
4  phi[2] ~ dbeta(1,1) T(phi[1],1)
5
6  p[1] ~ dbeta(1,2)
7  p[2] ~ dbeta(1,1) T(p[1],1)
8
9  for(t in 1:T){
10     rho[t] ~ dbeta(1/T,2-t/T)
11  }
12
13 w ~ ddirch(rep(1,3))
14
15 # Likelihood
16
17 for (i in 1:M){
18
19     clust[i] ~ dcat(w[1:3])
20
21     p_ind[i] = (clust[i]-2)*(clust[i]-1)/2*p[1]+
22                (3-clust[i])*(clust[i]/2)*p[2]
23     phi_ind[i] = (3-clust[i])*(clust[i]-1)*phi[1]+
24                  (clust[i]-2)^2*phi[2]
25
26     y[i,1] ~ dbin(muy[i,1],J[1])
27     muy[i,1] = z[i,1]*p_ind[i]
28     z[i,1] ~ dbin(rho[1],1)
29     r[i,1] = 1
30
31     for (t in 2:T){
32
33       y[i,t] ~ dbin(muy[i,t],J[t])
34       muy[i,t] = z[i,t]*p_ind[i]
35       z[i,t] ~ dbin(muz[i,t],1)
36       muz[i,t] = phi_ind[i]*z[i,t-1] + rho[t]*r[i,t]
37       r[i,t] = r[i,(t-1)]*(1-z[i,t-1])
38
39     }
40
41   }
42
43   # Derived parameters
44
45   for(t in 1:T){
46     N[t] = sum(z[1:M,t])
47   }
48
49   for (i in 1:M){
50
```

```
    Nind[i] = sum(z[i,1:T])
    Nalive[i] = 1-equals(Nind[i],0)

  }

Nsuper = sum(Nalive[1:M])
```

Listing A.9. CR 3-class cross-classified model for open populations that excludes the combination (p_1, ϕ_1), i.e. $\{\rho_t, \phi_{h_2}, p_{h_2}\}_{-(p_1,\phi_1)}$.

```
# Priors

phi[1] ~ dbeta(1,2)
phi[2] ~ dbeta(1,1) T(phi[1],1)

p[1] ~ dbeta(1,2)
p[2] ~ dbeta(1,1) T(p[1],1)

for(t in 1:T){
    rho[t] ~ dbeta(1/T,2-t/T)
 }

w ~ ddirch(rep(1,3))

# Likelihood

for (i in 1:M){

    clust[i] ~ dcat(w[1:3])

    p_ind[i] = (clust[i]-2)*(clust[i]-1)/2*p[1]+
                (3-clust[i])*(clust[i]/2)*p[2]
    phi_ind[i] = (3-clust[i])/2*(2-clust[i])*phi[2]+
                  (clust[i]-1)/2*(4-clust[i])*phi[1]

    y[i,1] ~ dbin(muy[i,1],J[1])
    muy[i,1] = z[i,1]*p_ind[i]
    z[i,1] ~ dbin(rho[1],1)
    r[i,1] = 1

    for (t in 2:T){

      y[i,t] ~ dbin(muy[i,t],J[t])
      muy[i,t] = z[i,t]*p_ind[i]
      z[i,t] ~ dbin(muz[i,t],1)
      muz[i,t] = phi_ind[i]*z[i,t-1] + rho[t]*r[i,t]
      r[i,t] = r[i,(t-1)]*(1-z[i,t-1])

    }

  }

  # Derived parameters

  for(t in 1:T){
    N[t] = sum(z[1:M,t])
  }

  for (i in 1:M){

    Nind[i] = sum(z[i,1:T])
    Nalive[i] = 1-equals(Nind[i],0)

  }

Nsuper = sum(Nalive[1:M])
```

Listing A.10. CR 3-class cross-classified model for open populations that excludes the combination (p_1, ϕ_2), i.e. $\{\rho_t, \phi_{h_2}, p_{h_2}\}_{-(p_1, \phi_2)}$.

```
# Priors

phi[1] ~ dbeta(1,2)
phi[2] ~ dbeta(1,1) T(phi[1],1)

p[1] ~ dbeta(1,2)
p[2] ~ dbeta(1,1) T(p[1],1)

for(t in 1:T){
    rho[t] ~ dbeta(1/T,2-t/T)
 }

w ~ ddirch(rep(1,3))

# Likelihood

for (i in 1:M){

    clust[i] ~ dcat(w[1:3])

    p_ind[i] = (3-clust[i])/2*(2-clust[i])*p[2]+
               (clust[i]-1)/2*(4-clust[i])*p[1]
    phi_ind[i] = (clust[i]-2)*(clust[i]-1)/2*phi[1]+
                 (3-clust[i])*(clust[i]/2)*phi[2]

    y[i,1] ~ dbin(muy[i,1],J[1])
    muy[i,1] = z[i,1]*p_ind[i]
    z[i,1] ~ dbin(rho[1],1)
    r[i,1] = 1

    for (t in 2:T){

      y[i,t] ~ dbin(muy[i,t],J[t])
      muy[i,t] = z[i,t]*p_ind[i]
      z[i,t] ~ dbin(muz[i,t],1)
      muz[i,t] = phi_ind[i]*z[i,t-1] + rho[t]*r[i,t]
      r[i,t] = r[i,(t-1)]*(1-z[i,t-1])

    }

  }

  # Derived parameters

  for(t in 1:T){
    N[t] = sum(z[1:M,t])
  }

  for (i in 1:M){

    Nind[i] = sum(z[i,1:T])
    Nalive[i] = 1-equals(Nind[i],0)

  }

Nsuper = sum(Nalive[1:M])
```

Listing A.11. CR 3-class cross-classified model for open populations that excludes the combination (p_2, ϕ_1), i.e. $\{\rho_t, \phi_{h_2}, p_{h_2}\}_{-(p_2, \phi_1)}$.

```
# Priors

phi[1] ~ dbeta(1,2)
phi[2] ~ dbeta(1,1) T(phi[1],1)

p[1] ~ dbeta(1,2)
p[2] ~ dbeta(1,1) T(p[1],1)

for(t in 1:T){
    rho[t] ~ dbeta(1/T,2-t/T)
 }

w ~ ddirch(rep(1,3))

# Likelihood

for (i in 1:M){

    clust[i] ~ dcat(w[1:3])

    p_ind[i] = (3-clust[i])/2*(2-clust[i])*p[2]+
               (clust[i]-1)/2*(4-clust[i])*p[1]
    phi_ind[i] = (3-clust[i])*(clust[i]-1)*phi[2]+
                 (clust[i]-2)^2*phi[1]

    y[i,1] ~ dbin(muy[i,1],J[1])
    muy[i,1] = z[i,1]*p_ind[i]
    z[i,1] ~ dbin(rho[1],1)
    r[i,1] = 1

    for (t in 2:T){

      y[i,t] ~ dbin(muy[i,t],J[t])
      muy[i,t] = z[i,t]*p_ind[i]
      z[i,t] ~ dbin(muz[i,t],1)
      muz[i,t] = phi_ind[i]*z[i,t-1] + rho[t]*r[i,t]
      r[i,t] = r[i,(t-1)]*(1-z[i,t-1])

    }

  }

  # Derived parameters

  for(t in 1:T){
    N[t] = sum(z[1:M,t])
  }

  for (i in 1:M){

    Nind[i] = sum(z[i,1:T])
    Nalive[i] = 1-equals(Nind[i],0)

  }

Nsuper = sum(Nalive[1:M])
```

Listing A.12. CR 3-class cross-classified model for open populations that excludes the combination (p_2, ϕ_2), i.e. $\{\rho_t, \phi_{h_2}, p_{h_2}\}_{-(p_2,\phi_2)}$.

A.5 Models from Chapter 5:

In Chapter 5 we considered the more general case of not evenly spaced sampling occasions. To account for this and to compute a yearly super-population abundance, we have to feed the model with some additional information about the structure of the structure of the encounter occasions. More specifically, we should provide two information: first, the information about the time-lag between the $t-1$-th and the t-th occasion (`t_lag[t-1]`), $t = 2, \ldots, T$; second, the information about which is the first occasion of each year (`year_start`).

In addition, three new nodes are included in the BUGS model in order to compute yearly estimates of the population size:

- `Nind.y[i,year]`: number of occasions the i-th individual is in the population in a given year;
- `Nalive.y[i,year]`: binary variable which counts whether the i-th individual belongs to the population in a given year;
- `N.y[year]`: population size in a given year.

Finally, to model the logit on the capture probability, we add a node `mu` for the intercept and a vector of nodes `tau` for the temporal random effect.

A.5.1 RPT model

We introduce the node `delta` which represents the undetectability parameter. Temporal random effects τ_t's may be set to sum to zero for identifiability purpose (Pledger et al., 2003).

```
# Priors

phi_TR ~ dbeta(1,2)
phi_NTR ~ dbeta(1,1) T(phi_TR,1)

delta ~ dunif(0,1)
mu ~ dnorm(0,0.1)
tau[1] = -sum(tau[2:T]) #to help identifiability tau's sum to 0
p_NPT[1] = ilogit(mu + tau[1])
p_PT[1] = (1-delta)*p_NPT[1]

rho_TR ~ dbeta(1,T)
rho_NTR[1] ~ dbeta(1/T,2-1/T)

for(t in 2:T){

    tau[t] ~ dnorm(0,4)
    p_NPT[t] = ilogit(mu + tau[t])
    p_PT[t] = (1-delta)*p_NPT[t]

    rho_NTR[t] ~ dbeta(1/T,2-t/T)

  }

w ~ ddirch(rep(1,3))

# Likelihood

```

```
for (i in 1:M){

    clust[i] ~ dcat(w[1:3])

    p_ind[i,1] = (3-clust[i])*(clust[i]-1)*p_PT[1]+
                 (clust[i]-2)^2*p_NPT[1]
    phi_ind[i] = (clust[i]-2)*(clust[i]-1)/2*phi_TR+
                 (3-clust[i])*(clust[i]/2)*phi_NTR
    rho_ind[i,1] = (clust[i]-2)*(clust[i]-1)/2*rho_TR+
                   (3-clust[i])*(clust[i]/2)*rho_NTR[1]

    y[i,1] ~ dbin(muy[i,1],1)
    muy[i,1] = z[i,1]*p_ind[i,1]
    z[i,1] ~ dbin(rho_ind[i,1],1)
    r[i,1] = 1

    for (t in 2:T){

      p_ind[i,t] = (3-clust[i])*(clust[i]-1)*p_PT[t]+
                   (clust[i]-2)^2*p_NPT[t]
      rho_ind[i,t] = (clust[i]-2)*(clust[i]-1)/2*rho_TR+
                     (3-clust[i])*(clust[i]/2)*rho_NTR[t]

      y[i,t] ~ dbin(muy[i,t],1)
      muy[i,t] = z[i,t]*p_ind[i,t]
      z[i,t] ~ dbin(muz[i,t],1)
      muz[i,t] = phi_ind[i]^t_lag[t-1]*z[i,t-1] + rho_ind[i,t]*r[i,t]
      r[i,t] = r[i,(t-1)]*(1-z[i,t-1])

    }

  }

  # Derived parameters

  for(year in 1:(length(year_start)-1)){

      Nind.y[i,year] = sum(z[i,year_start[year]:(year_start[year+1]-1)])
      Nalive.y[i,year] = 1-equals(Nind.y[i,year],0)

      N.y[year] = sum(Nalive.y[1:M,year])
    }

  for (i in 1:M){

    Nind[i] = sum(z[i,1:T])
    Nalive[i] = 1-equals(Nind[i],0)

  }

Nsuper = sum(Nalive[1:M])
```

Listing A.13. RPT for bottlenose dolphin populations.

A.5.2 Alternatives to RPT model

For the next model, we consider `p[t]` as the detection probability at time t.

```
# Priors

phi ~ dunif(0,1)

mu ~ dnorm(0,0.1)

tau[1] = -sum(tau[2:T])
```

```
p[1] = ilogit(mu + tau[1])

rho[1] ~ dbeta(1/T,2-1/T)

for(t in 2:T){

    tau[t] ~ dnorm(0,4)
    p[t] = ilogit(mu + tau[t])

    rho[t] ~ dbeta(1/T,2-t/T)

  }


# Likelihood

for (i in 1:M){

    y[i,1] ~ dbin(muy[i,1],1)
    muy[i,1] = z[i,1]*p[1]
    z[i,1] ~ dbin(rho[1],1)
    r[i,1] = 1

    for (t in 2:T){

      y[i,t] ~ dbin(muy[i,t],1)
      muy[i,t] = z[i,t]*p[t]
      z[i,t] ~ dbin(muz[i,t],1)
      muz[i,t] = phi^t_lag[t-1]*z[i,t-1] + rho[t]*r[i,t]
      r[i,t] = r[i,(t-1)]*(1-z[i,t-1])

    }

  }

  # Derived parameters

  for(year in 1:(length(year_start)-1)){

      Nind.y[i,year] = sum(z[i,year_start[year]:(year_start[year+1]-1)])
      Nalive.y[i,year] = 1-equals(Nind.y[i,year],0)

      N.y[year] = sum(Nalive.y[1:M,year])
    }

  for (i in 1:M){

    Nind[i] = sum(z[i,1:T])
    Nalive[i] = 1-equals(Nind[i],0)

  }

Nsuper = sum(Nalive[1:M])
```

Listing A.14. JS model with homogeneous parameters, i.e. $\{\rho_t, \phi, p_t\}$. Temporal effect affects the capture probability through a logit link.

Finally, for the next model we introduce the group-dependent intercept `eta[g]` and the time-dependent intercept `tau[g]` to model the logit on the capture probabilities. The capture probability at time t for an individual in group g is `p[g,t]`.

```
# Priors

phi[1] ~ dunif(0,1)
phi[2] ~ dunif(phi[1],1)
phi[3] ~ dunif(phi[2],1)
```

```
6
7  eta[1] = -sum(eta[2:3])
8  eta[2] ~ dnorm(0,0.1)
9  eta[3] ~ dnorm(0,0.1)
10
11 tau[1] = -sum(tau[2:T])
12
13 p[1,1] = ilogit(eta[1] + tau[1])
14 p[2,1] = ilogit(eta[2] + tau[1])
15 p[3,1] = ilogit(eta[3] + tau[1])
16
17 rho[1,1] ~ dbeta(1/T,2-1/T)
18 rho[2,1] ~ dbeta(1/T,2-1/T)
19 rho[3,1] ~ dbeta(1/T,2-1/T)
20
21 for(t in 2:T){
22
23     tau[t] ~ dnorm(0,4)
24     p[1,t] = ilogit(eta[1] + tau[t])
25
26     rho[1,t] ~ dbeta(1/T,2-t/T)
27
28     for(g in 2:3){
29
30         p[g,t] = ilogit(eta[g] + tau[t])
31
32         rho[g,t] ~ dbeta(1/T,2-t/T)
33
34     }
35
36 }
37
38 w ~ ddirch(rep(1,3))
39
40 # Likelihood
41
42 for (i in 1:M){
43
44     clust[i] ~ dcat(w[1:3])
45
46     y[i,1] ~ dbin(muy[i,1],1)
47     muy[i,1] = z[i,1]*p[clust[i],1]
48     z[i,1] ~ dbin(rho[clust[i],1],1)
49     r[i,1] = 1
50
51     for (t in 2:T){
52
53       y[i,t] ~ dbin(muy[i,t],1)
54       muy[i,t] = z[i,t]*p[clust[i],t]
55       z[i,t] ~ dbin(muz[i,t],1)
56       muz[i,t] = phi[clust[i]]^t_lag[t-1]*z[i,t-1] + rho[clust[i],t]*r[i,t]
57       r[i,t] = r[i,(t-1)]*(1-z[i,t-1])
58
59     }
60
61   }
62
63   # Derived parameters
64
65   for(year in 1:(length(year_start)-1)){
66
67       Nind.y[i,year] = sum(z[i,year_start[year]:(year_start[year+1]-1)])
68       Nalive.y[i,year] = 1-equals(Nind.y[i,year],0)
69
70       N.y[year] = sum(Nalive.y[1:M,year])
71     }
```

```
72
73   for (i in 1:M){
74
75     Nind[i] = sum(z[i,1:T])
76     Nalive[i] = 1-equals(Nind[i],0)
77
78   }
79
80 Nsuper = sum(Nalive[1:M])
```

Listing A.15. JS 3-class mixture model with component-specific recruitment, survival and capture parameters, i.e. $\{[\rho_{t\times g}, \phi_g, p_{t+g}]_{G=3}\}$. Temporal and heterogeneous across group effect affect the logit of capture probabilities in an additive way.

Appendix B

Proof of likelihood for multinomial grouped data

Let y_i be the capture frequency of the i-th individual in T capture attempts and let $f_k = \sum_{i=1}^{N} \mathbb{1}(y_i = k)$ be the frequency of individuals detected in k out of T sampling occasions.

Consider the following likelihood function for the abundance N and the capture probability p

$$L(N, p|f_1, \ldots, f_T) = \frac{N!}{(N-D)!\, f_1! \cdots f_T!}\, \pi_0^{N-D}\, \pi_1^{f_1} \cdots \pi_T^{f_T}\,, \tag{B.1}$$

which is written in terms of $f_1, \ldots, f_T$. Of course,

$$\pi_k = \binom{T}{k} p^k (1-p)^{T-k}\,.$$

The likelihood in (B.1) can be easily expressed as function of p, that is

$$\begin{aligned} L(N, p|f_1, \ldots, f_T) &= \frac{N!}{(N-D)! \prod_{k=1}^{T} f_k!}\, (1-p)^{T(N-D)} \prod_{k=1}^{T} \left[\binom{T}{k} p^k (1-p)^{T-k} \right]^{f_k} \\ &= \frac{N!}{(N-D)! \prod_{k=1}^{T} f_k!}\, (1-p)^{T(N-D) + \sum_{k=1}^{T} (T-k) f_k}\, p^{\sum_{k=1}^{T} k f_k} \prod_{k=1}^{T} \binom{T}{k}^{f_k}\,. \end{aligned}$$

Exploiting the fact that

$$\sum_{i=1}^{D} y_i = \sum_{k=1}^{T} k f_k\,,$$

the previous likelihood can be alternatively formulated in terms of the individual frequencies $y_1, \ldots, y_D$, namely

$$\begin{aligned} L(N, p|y_1, \ldots, y_D) &= \frac{N!}{(N-D)! \prod_{k=1}^{T} f_k!}\, p^{\sum_{i=1}^{D} y_i}\, (1-p)^{TN - TD + T\sum_{k=1}^{T} f_k - \sum_{k=1}^{T} k f_k} \prod_{k=1}^{T} \binom{T}{k}^{f_k} \\ &\propto \frac{N!}{(N-D)!}\, p^{\sum_{i=1}^{D} y_i}\, (1-p)^{TN - \sum_{i=1}^{D} y_i}\,. \end{aligned}$$

Appendix C

Proof of results on prior specifications of class-specific probabilities

C.1 Beta and truncated Beta

Theorem 1. *Suppose that* $X \sim Beta(\alpha_1, \beta_1)$ *and* $Y|X = x \sim tBeta(\alpha_2, \beta_2, x, 1)$, *namely a truncated Beta whose density is*

$$p_{Y|X}(y|x) = \frac{1}{B(\alpha_2, \beta_2)} \frac{y^{\alpha_2 - 1}\,(1-y)^{\beta_2 - 1}}{1 - F_{Beta(\alpha_2,\beta_2)}(x)}\, \mathbb{1}_{\{x,1\}}(y)\,,$$

with $F_{Beta(1,\beta_2)}(\cdot)$ *being the cdf of a* $Beta(1, \beta_2)$. *Then, if* $\alpha_2 = 1$, *the marginal density function of* Y *is given by*

$$p_Y(y) = \frac{B(\alpha_1, \beta_1 - \beta_2)}{B(\alpha_1, \beta_1)}\, \beta_2 (1-y)^{\beta_2 - 1}\, F_{Beta(\alpha_1,\beta_1-\beta_2)}(y)\,, \quad \textit{with } \beta_1 > \beta_2\ ,$$

where $F_{Beta(\alpha_1,\beta_1-\beta_2)}(\cdot)$ *is the cdf of a* $Beta(\alpha_1, \beta_1 - \beta_2)$.

In particular, when $\alpha_1 = \beta_2 = k$ *and* $\beta_1 = k + 1$, *that is*

$$X \sim Beta(k, k+1) \quad \textit{and} \quad Y|X = x \sim tBeta(1, k, x, 1)\,,$$

the marginal prior induced on Y *is* $Beta(k + 1, k)$.

Proof. Consider $X \sim Beta(\alpha_1, \beta_1)$ and $Y|X = x \sim tBeta(1, \beta_2, x, 1)$, whose probability density functions are, respectively,

$$p_X(x) = \frac{1}{B(\alpha_1, \beta_1)}\, x^{\alpha_1 - 1}(1-x)^{\beta_1 - 1}$$

and

$$p_{Y|X}(y|x) = \frac{\beta_2\,(1-y)^{\beta_2 - 1}}{1 - F_{Beta(1,\beta_2)}(x)}\, \mathbb{1}_{\{x,1\}}(y)\,,$$

with $F_{Beta(1,\beta_2)}(\cdot)$ being the cdf of a $Beta(1,\beta_2)$.

Then, the marginal prior distribution induced on Y is given by

$$\begin{aligned}
p_Y(y) &= \int_{-\infty}^{\infty} p_{Y|X}(y|x)p_X(x)\,dx \\
&= \int_0^1 \frac{\beta_2\,(1-y)^{\beta_2-1}}{1-F_{Beta(1,\beta_2)}(x)}\,\mathbb{1}_{\{x,1\}}(y)\frac{1}{B(\alpha_1,\beta_1)} \times x^{\alpha_1-1}(1-x)^{\beta_1-1}\,dx \\
&= \frac{\beta_2\,(1-y)^{\beta_2-1}}{B(\alpha_1,\beta_1)} \int_0^y \frac{x^{\alpha_1-1}(1-x)^{\beta_1-1}}{1-F_{Beta(1,\beta_2)}(x)}\,dx \\
&= \frac{\beta_2\,(1-y)^{\beta_2-1}}{B(\alpha_1,\beta_1)} \int_0^y x^{\alpha_1-1}(1-x)^{(\beta_1-\beta_2)-1}\,dx \\
&= \frac{B(\alpha_1,\beta_1-\beta_2)}{B(\alpha_1,\beta_1)}\,\beta_2\,(1-y)^{\beta_2-1} \int_0^y \frac{x^{\alpha_1-1}(1-x)^{(\beta_1-\beta_2)-1}}{B(\alpha_1,\beta_1-\beta_2)}\,dx \\
&= \frac{B(\alpha_1,\beta_1-\beta_2)}{B(\alpha_1,\beta_1)}\,\beta_2(1-y)^{\beta_2-1}\,F_{Beta(\alpha_1,\beta_1-\beta_2)}(y)\,,
\end{aligned}$$

where we exploited the fact that $F_{Beta(1,b)}(z) = \int_0^z (1-t)^{b-1}\,dt = 1-(1-z)^b$. Observe that the constraint $\beta_1 > \beta_2$ is essential to avoid the divergence of the beta function.

Now, if we consider $\alpha_1 = \beta_2 = k$ and $\beta_1 = k+1$, the previous probability density function becomes

$$\begin{aligned}
p_Y(y) &= \frac{1}{B(k,k+1)}\,\frac{1}{k}\,k\,(1-y)^{k-1}\,F_{Beta(k,1)}(y) \\
&= \frac{1}{B(k+1,k)}\,(1-y)^{k-1}\,y^k\,,
\end{aligned}$$

which implies that $Y \sim Beta(k+1,k)$.

□

C.1.1 Beta and restricted Beta

Theorem 2. *If $X \sim Beta(\alpha_x,\beta_x)$ and $Y \mid X = x \sim rBeta\,(\alpha_y,\beta_y,x,1)$, then the joint prior is*

$$p_{XY}\,(x,y) = \frac{\Gamma(\alpha_x+\beta_x)\Gamma(\alpha_y+\beta_y)}{\Gamma(\alpha_x)\Gamma(\beta_x)\Gamma(\alpha_y)\Gamma(\beta_y)}\,x^{\alpha_x-1}(1-x)^{\beta_x-\beta_y-\alpha_y}(x-y)^{\alpha_y-1}(1-y)^{\beta_y-1}\,.$$

There is no easy way to express the marginal distribution of Y, but using the Law of iterated expectation we can derive the marginal expected value and variance for p_2:

$$\begin{aligned}
\mathbb{E}\,[Y] &= \frac{\alpha_y}{\alpha_y+\beta_y} + \mu_x\frac{\beta_y}{\alpha_y+\beta_y} \\
\mathbb{V}\,[Y] &= \sigma_x^2\frac{\beta_y^2}{(\alpha_y+\beta_y)^2}\left(1+\frac{\alpha_y}{\beta_y(\alpha_y+\beta_y+1)}\right) + (1-\mu_x)^2\frac{\alpha_y\beta_y}{(\alpha_y+\beta_y)^2(\alpha_y+\beta_y+1)}
\end{aligned}$$

where $\mu_x = \mathbb{E}[X]$ *and* $\sigma_x^2 = \mathbb{V}[X]$.

Proof. The joint distribution is obtained by direct application of the conditional equivalence of Equation (3.2).
Marginal expected value and variance are obtained by the application of the Law of iterated expectations. First, the expected value and variance of $Y|X = x \sim rBeta(\alpha_y, \beta_y, x, 1)$ are:

$$\mathbb{E}[Y \mid X] = \frac{\alpha_y}{\alpha_y + \beta_y} * (1 - X) + X = \frac{\alpha_y}{\alpha_y + \beta_y} + X\frac{\beta_y}{\alpha_y + \beta_y}$$
$$\mathbb{V}[Y \mid X] = \frac{\alpha_y \beta_y}{(\alpha_y + \beta_y)^2(\alpha_y + \beta_y + 1)}(1 - X)^2$$

Then:

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}[\mathbb{E}[Y \mid X]] \\ &= \frac{\alpha_y}{\alpha_y + \beta_y} + \mathbb{E}[X]\frac{\beta_y}{\alpha_y + \beta_y} \\ &= \frac{\alpha_y}{\alpha_y + \beta_y} + \frac{\alpha_x \beta_y}{(\alpha_x + \beta_x)(\alpha_y + \beta_y)} \\ &= \frac{\alpha_y}{\alpha_y + \beta_y} + \mu_x\frac{\beta_y}{\alpha_y + \beta_y}\end{aligned}$$

$$\begin{aligned}\mathbb{V}[Y] &= \mathbb{E}[\mathbb{V}[Y \mid X]] + \mathbb{V}[\mathbb{E}[Y \mid X]] \\ &= \frac{\alpha_y \beta_y}{(\alpha_y + \beta_y)^2(\alpha_y + \beta_y + 1)}\mathbb{E}\left[(1 - X)^2\right] + \mathbb{V}[X]\frac{\beta_y^2}{(\alpha_y + \beta_y)^2} \\ &= \frac{\alpha_y \beta_y}{(\alpha_y + \beta_y)^2(\alpha_y + \beta_y + 1)}\left(1 - 2\mu_x + \mathbb{E}\left[X^2\right]\right) + \sigma_x^2\frac{\beta_y^2}{(\alpha_y + \beta_y)^2} \\ &= \frac{\alpha_y \beta_y}{(\alpha_y + \beta_y)^2(\alpha_y + \beta_y + 1)}\left(1 - 2\mu_x + \mu_x^2 + \sigma_x^2\right) + \sigma_x^2\frac{\beta_y^2}{(\alpha_y + \beta_y)^2} \\ &= \sigma_x^2\frac{\beta_y^2}{(\alpha_y + \beta_y)^2}\left(1 + \frac{\alpha_y}{\beta_y(\alpha_y + \beta_y + 1)}\right) + (1 - \mu_x)^2\frac{\alpha_y \beta_y}{(\alpha_y + \beta_y)^2(\alpha_y + \beta_y + 1)}\end{aligned}$$

□

C.1.2 A convenient parameter setting

Theorem 3. *Let* $X \sim Beta(\alpha_x, \beta + 1)$ *and* $Y \mid X = x \sim rBeta(1, \beta, x, 1)$*, then the joint distribution for* (X, Y) *is*

$$p_{XY}(x, y) = \frac{(\alpha_x + \beta)\Gamma(\alpha_x + \beta)}{\Gamma(\alpha_x)\Gamma(\beta)}x^{\alpha_x - 1}(1 - y)^{\beta}.$$

Hence, the marginal on Y *is a Beta density, namely*

$$Y \sim Beta(\alpha_x + 1, \beta)$$

with marginal expected value and variance given by

$$\mathbb{E}[Y] = \frac{\alpha_x + 1}{\alpha_x + \beta + 1}$$
$$\mathbb{V}[Y] = \frac{(\alpha_x + 1)\beta_y}{(\alpha_x + \beta + 1)^2(\alpha_x + \beta + 2)} .$$

Proof. The derivation of the joint density is trivial.

The marginal distribution of Y is:

$$\begin{aligned} p_Y(y) &= \int_{\mathcal{X}} p_{XY}(x, y)\mathbb{1}_{x,1}(y)dx \\ &= \int_0^1 p_{XY}(x, y)\mathbb{1}_{0,y}(x)dx \\ &= \int_0^y p_{XY}(x, y)dx \\ &= \frac{(\alpha_x + \beta)\Gamma(\alpha_x + \beta)}{\Gamma(\alpha_x)\Gamma(\beta)} \cdot (1 - y)^{\beta-1} \cdot \int_0^y x^{\alpha_x - 1}dx \\ &= \frac{\Gamma(\alpha_x + \beta + 1)}{\Gamma(\alpha_x)\Gamma(\beta)} \cdot (1 - y)^{\beta-1} \cdot \frac{y^{\alpha_x}}{\alpha_x} \\ &= \frac{\Gamma(\alpha_x + \beta + 1)}{\Gamma(\alpha_x + 1)\Gamma(\beta)} \cdot y^{\alpha_x + 1 - 1}(1 - y)^{\beta-1} , \end{aligned}$$

which corresponds to a $Beta(\alpha_x + 1, \beta)$.

Expected value and variance follow from basic properties of the Beta distribution. □

Appendix D

Details about time-lags used in Subsection 5.3.1

The number of days between two consecutive capture occasions (*daily time lags*, henceforth) within a single year have been simulated from a *shifted* geometrical distribution with probability 0.05, which has an expected value equal to 20 and a standard deviation equal to 19.5. The resulting random sequence of time lags is

$$(20, 1, 12, 15, 56, 9, 9, 12, 10)$$

and, for scenarios that contemplate more than one year of study, the same sequence is repeated during each new year. The shift of year occurring each 10 occasions is achieved by using a higher constant time lag (i.e. 240 days) between the $(10k)$th occasion and the $(10k + 1)$th occasion, with $k = 1, 2, 3$. This results in a scenario k composed by k years of study, for $k = 1, 2, 3, 4$. For example, scenario 2 ($T = 20$) is composed by the following sequence of time lags, resulting in 2 years of capture occasions:

$$(20, 1, 12, 15, 56, 9, 9, 12, 10, 240, 20, 1, 12, 15, 56, 9, 9, 12, 10)\,.$$

Bibliography

Alaimo Di Loro, P., Caruso, G., Mingione, M., Jona Lasinio, G., and Tardella, L. (2022). Specification of informative priors for capture-recapture finite mixture models. In *Book of short paper – SIS 2022*. Pearson.

Alho, J. M. (1990). Logistic regression in capture-recapture models. *Biometrics*, pages 623–635.

Altieri, L., Farcomeni, A., and Fegatelli, D. A. (2022). Continuous time-interaction processes for population size estimation, with an application to drug dealing in italy. *Biometrics*.

Alunni Fegatelli, D. and Tardella, L. (2016). Flexible behavioral capture–recapture modeling. *Biometrics*, 72(1):125–135.

Amendolia, S., Lombardini, M., Pierucci, P., and Meriggi, A. (2019). Seasonal spatial ecology of the wild boar in a peri-urban area. *Mammal Research*, 64:387–396.

Amstrup, S. C., McDonald, T. L., and Manly, B. F. (2005). *Handbook of capture-recapture analysis*. Princeton University Press.

Auger-Méthé, M., Marcoux, M., and Whitehead, H. (2010). Nicks and notches of the dorsal ridge: promising mark types for the photo-identification of narwhals. *Marine Mammal Science*, 26(3):663–678.

Baillargeon, S. and Rivest, L.-P. (2007). Rcapture: loglinear models for capture-recapture in r. *Journal of Statistical Software*, 19:1–31.

Böhning, D., Rocchetti, I., Maruotti, A., and Holling, H. (2020). Estimating the undetected infections in the covid-19 outbreak by harnessing capture–recapture methods. *International Journal of Infectious Diseases*, 97:197–201.

Brittain, S. and Böhning, D. (2009). Estimators in capture–recapture studies with two sources. *AStA Advances in Statistical Analysis*, 93(1):23–47.

Brooks, S., King, R., and Morgan, B. (2004). A bayesian approach to combining animal abundance and demographic data. *Animal biodiversity and conservation*, 27(1):515–529.

Brooks, S. P., Catchpole, E. A., Morgan, B. J., and Barry, S. (2000). On the bayesian analysis of ring-recovery data. *Biometrics*, 56(3):951–956.

Burbidge, A. A. and Manly, B. F. (2002). Mammal extinctions on australian islands: causes and conservation implications. *Journal of biogeography*, 29(4):465–473.

Burnham, K. P. (1972). Estimation of population size in multiple capture-recapture studies when capture probabilities vary among animals.

Burnham, K. P. and Overton, W. S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology*, 60(5):927–936.

Caruso, G., Di Loro, P. A., Mingione, M., Tardella, L., Pace, D. S., and Lasinio, G. J. (2023). Finite mixtures in capture-recapture surveys for modelling residency patterns in marine wildlife populations. *arXiv preprint arXiv:2304.06999*.

Celeux, G. (1998). Bayesian inference for mixture: The label switching problem. In *Compstat*, pages 227–232. Springer.

Clark, J. S. and Gelfand, A. E. (2006). *Hierarchical modelling for the environmental sciences: statistical methods and applications*. OUP Oxford.

Coull, B. A. and Agresti, A. (1999). The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics*, 55(1):294–301.

Cubaynes, S., Lavergne, C., Marboutin, E., and Gimenez, O. (2012). Assessing individual heterogeneity using model selection criteria: how many mixture components in capture–recapture models? *Methods in Ecology and Evolution*, 3(3):564–573.

Diebolt, J. and Robert, C. P. (1994). Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(2):363–375.

Dinis, A., Alves, F., Nicolau, C., Ribeiro, C., Kaufmann, M., Cañadas, A., and Freitas, L. (2016). Bottlenose dolphin tursiops truncatus group dynamics, site fidelity, residency and movement patterns in the madeira archipelago (north-east atlantic). *African Journal of Marine Science*, 38(2):151–160.

Dorazio, R. M. (2020). Objective prior distributions for jolly-seber models of zero-augmented data. *Biometrics*, 76(4):1285–1296.

Dorazio, R. M. and Andrew Royle, J. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics*, 59(2):351–364.

Durban, J. W. and Elston, D. A. (2005). Mark-recapture with occasion and individual effects: abundance estimation through bayesian model selection in a fixed dimensional parameter space. *Journal of agricultural, biological, and environmental statistics*, 10(3):291–305.

Estrade, V. and Dulau, V. (2020). Abundance and site fidelity of bottlenose dolphins off a remote oceanic island (reunion island, southwest indian ocean). *Marine Mammal Science*, 36(3):871–896.

Farcomeni, A. (2011). Recapture models under equality constraints for the conditional capture probabilities. *Biometrika*, 98(1):237–242.

Farcomeni, A. (2020). Population size estimation with interval censored counts and external information: Prevalence of multiple sclerosis in rome. *Biometrical Journal*, 62(4):945–956.

Farcomeni, A. (2022). How many refugees and migrants died trying to reach europe? joint population size and total estimation. *The Annals of Applied Statistics*, 16(4):2339–2351.

Fryxell, J. M., Sinclair, A. R., and Caughley, G. (2014). *Wildlife ecology, conservation, and management*. John Wiley & Sons.

Gardner, B., Royle, J. A., and Wegan, M. T. (2009). Hierarchical models for estimating density from dna mark–recapture studies. *Ecology*, 90(4):1106–1115.

Gelfand, A. E. and Ghosh, S. K. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika*, 85(1):1–11.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and computing*, 24(6):997–1016.

Ghosh, S. K. and Norris, J. L. (2005). Bayesian capture-recapture analysis and model selection allowing for heterogeneity and behavioral effects. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(1):35–49.

Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994). A language and program for complex bayesian modelling. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 43(1):169–177.

Gimenez, O., Cam, E., and Gaillard, J.-M. (2018). Individual heterogeneity and capture–recapture models: what, why and how? *Oikos*, 127(5):664–686.

Guéry, L., Descamps, S., Pradel, R., Hanssen, S. A., Erikstad, K. E., Gabrielsen, G. W., Gilchrist, H. G., and Bêty, J. (2017). Hidden survival heterogeneity of three common eider populations in response to climate fluctuations. *Journal of Animal Ecology*, 86(3):683–693.

Hand, D. J. and Till, R. J. (2001). A simple generalisation of the area under the roc curve for multiple class classification problems. *Machine learning*, 45(2):171–186.

Haughey, R., Hunt, T., Hanf, D., Rankin, R. W., and Parra, G. J. (2020). Photographic capture-recapture analysis reveals a large population of indo-pacific bottlenose dolphins (tursiops aduncus) with low site fidelity off the north west cape, western australia. *Frontiers in Marine Science*, 6:781.

Huggins, R. (1989). On the statistical analysis of capture experiments. *Biometrika*, 76(1):133–140.

Hunt, T. N., Bejder, L., Allen, S. J., Rankin, R. W., Hanf, D., and Parra, G. J. (2017). Demographic characteristics of australian humpback dolphins reveal important habitat toward the southwestern limit of their range. *Endangered Species Research*, 32:71–88.

Hurn, M., Justel, A., and Robert, C. P. (2003). Estimating mixtures of regressions. *Journal of computational and graphical statistics*, 12(1):55–79.

Huvier, N., Moyne, G., Kaerlé, C., and Mouzon-Moyne, L. (2023). Time is running out: Microsatellite data predict the imminent extinction of the boreal lynx (lynx lynx) in france. *Frontiers in Conservation Science*, 4.

Jasra, A., Holmes, C. C., and Stephens, D. A. (2005). Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. *Statistical Science*, 20(1):50–67.

Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, 52(1/2):225–247.

Kendall, W. L., Nichols, J. D., and Hines, J. E. (1997). Estimating temporary emigration using capture–recapture data with pollock's robust design. *Ecology*, 78(2):563–578.

La Manna, G., Rako-Gospić, N., Pace, D. S., Bonizzoni, S., Di Iorio, L., Polimeno, L., Perretti, F., Ronchetti, F., Giacomini, G., Pavan, G., et al. (2022). Determinants of variability in signature whistles of the mediterranean common bottlenose dolphin. *Scientific Reports*, 12(1):1–16.

Lebreton, J.-D., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological monographs*, 62(1):67–118.

Louvrier, J., Chambert, T., Marboutin, E., and Gimenez, O. (2018). Accounting for misidentification and heterogeneity in occupancy studies using hidden markov models. *Ecological modelling*, 387:61–69.

Lukacs, P. M. and Burnham, K. P. (2005). Review of capture–recapture methods applicable to noninvasive genetic sampling. *Molecular ecology*, 14(13):3909–3919.

Marin, J.-M., Mengersen, K., and Robert, C. P. (2005). Bayesian modelling and inference on mixtures of distributions. *Handbook of statistics*, 25:459–507.

Marshall, M. R., Diefenbach, D. R., Wood, L. A., and Cooper, R. J. (2004). Annual survival estimation of migratory songbirds confounded by incomplete breeding site–fidelity: study designs that may help. *Animal Biodiversity and Conservation*, 27(1):59–72.

McCrea, R. S. and Morgan, B. J. (2014). *Analysis of capture-recapture data*. CRC Press.

McLachlan, G. J., Lee, S. X., and Rathnayake, S. I. (2019). Finite mixture models. *Annual review of statistics and its application*, 6:355–378.

Morin, D. J., Kelly, M. J., and Waits, L. P. (2016). Monitoring coyote population dynamics with fecal dna and spatial capture–recapture. *The Journal of Wildlife Management*, 80(5):824–836.

Morrison, M. L., Marcot, B., and Mannan, W. (2012). Wildlife-habitat relationships: concepts and applications.

Nichols, J. D., Hines, J. E., Pollock, K. H., Hinz, R. L., and Link, W. A. (1994). Estimating breeding proportions and testing hypotheses about costs of reproduction with capture-recapture data. *Ecology*, 75(7):2052–2065.

Nichols, J. D., Thomas, L., and Conn, P. B. (2009). Inferences about landbird abundance from count data: recent advances and future directions. *Modeling demographic processes in marked populations*, pages 201–235.

Norris III, J. L. and Pollock, K. H. (1996). Nonparametric mle under two closed capture-recapture models with heterogeneity. *Biometrics*, pages 639–649.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife monographs*, (62):3–135.

Overton, W. S. (1965). A modification of the schnabel estimator to account for removal of animals from the population. *The Journal of Wildlife Management*, pages 392–395.

Pace, D. S., Di Marco, C., Giacomini, G., Ferri, S., Silvestri, M., Papale, E., Casoli, E., Ventura, D., Mingione, M., Alaimo Di Loro, P., et al. (2021). Capitoline dolphins: residency patterns and abundance estimate of tursiops truncatus at the tiber river estuary (mediterranean sea). *Biology*, 10(4):275.

Pace, D. S., Ferri, S., Giacomini, G., Di Marco, C., Papale, E., Silvestri, M., Pedrazzi, G., Ventura, D., Casoli, E., and Ardizzone, G. (2022). Resources and population traits modulate the association patterns in the common bottlenose dolphin living nearby the tiber river estuary (mediterranean sea). *Frontiers in Marine Science*, page 1506.

Pastore, M. (2018). Overlapping: a r package for estimating overlapping in empirical distributions. *Journal of Open Source Software*, 3(32):1023.

Pastore, M. and Calcagnì, A. (2019). Measuring distribution similarities between samples: A distribution-free overlapping index. *Frontiers in psychology*, 10:1089.

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I.-C., Clark, T. D., Colwell, R. K., Danielsen, F., Evengård, B., et al. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, 355(6332):eaai9214.

Pledger, S. (2000). Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics*, 56(2):434–442.

Pledger, S. (2005). The performance of mixture models in heterogeneous closed population capture–recapture. *Biometrics*, 61(3):868–873.

Pledger, S., Pollock, K. H., and Norris, J. L. (2003). Open capture-recapture models with heterogeneity: I. cormack-jolly-seber model. *Biometrics*, 59(4):786–794.

Pledger, S., Pollock, K. H., and Norris, J. L. (2010). Open capture–recapture models with heterogeneity: Ii. jolly–seber model. *Biometrics*, 66(3):883–890.

Plummer, M. et al. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, pages 1–10. Vienna, Austria.

Pollock, K. H. (1982). A capture-recapture design robust to unequal probability of capture. *The Journal of Wildlife Management*, 46(3):752–757.

Pollock, K. H., Nichols, J. D., Brownie, C., and Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife monographs*, pages 3–97.

Pradel, R. (1996). Utilization of capture-mark-recapture for the study of recruitment and population growth rate. *Biometrics*, pages 703–709.

Royle, J. A. (2006). Site occupancy models with heterogeneous detection probabilities. *Biometrics*, 62(1):97–102.

Royle, J. A., Dorazio, R. M., and Link, W. A. (2007). Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics*, 16(1):67–85.

Royle, J. A. and Young, K. V. (2008). A hierarchical model for spatial capture–recapture data. *Ecology*, 89(8):2281–2289.

Sanathanan, L. (1972). Estimating the size of a multinomial population. *The Annals of Mathematical Statistics*, 43(1):142–152.

Schmidt, B. R. (2005). Monitoring the distribution of pond-breeding amphibians when species are detected imperfectly. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 15(6):681–692.

Schnabel, Z. E. (1938). The estimation of the total fish population of a lake. *The American Mathematical Monthly*, 45(6):348–352.

Schwarz, C. J. and Arnason, A. N. (1996). A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics*, pages 860–873.

Seber, G. A. (1965). A note on the multiple-recapture census. *Biometrika*, 52(1/2):249–259.

Seber, G. A. F. et al. (1982). The estimation of animal abundance and related parameters.

Sharifi Far, S., King, R., Bird, S., Overstall, A., Worthington, H., and Jewell, N. (2021). Multiple systems estimation for modern slavery: Robustness of list omission and combination. *Crime & Delinquency*, 67(13-14):2213–2236.

Soroye, P., Newbold, T., and Kerr, J. (2020). Climate change contributes to widespread declines among bumble bees across continents. *Science*, 367(6478):685–688.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). Winbugs user manual.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):795–809.

Tancredi, A., Auger-Méthé, M., Marcoux, M., and Liseo, B. (2013). Accounting for matching uncertainty in two stage capture–recapture experiments using photographic measurements of natural marks. *Environmental and ecological statistics*, 20(4):647–665.

Tardella, L. (2002). A new bayesian method for nonparametric capture-recapture models in presence of heterogeneity. *Biometrika*, 89(4):807–817.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F., De Siqueira, M. F., Grainger, A., Hannah, L., et al. (2004). Extinction risk from climate change. *Nature*, 427(6970):145–148.

Turek, D., Wehrhahn, C., and Gimenez, O. (2021). Bayesian non-parametric detection heterogeneity in ecological models. *Environmental and Ecological Statistics*, 28(2):355–381.

Vella, A., Murphy, S., Giménez, J., de Stephanis, R., Mussi, B., Vella, J. G., Larbi Doukara, K., and Pace, D. S. (2021). The conservation of the endangered mediterranean common dolphin (delphinus delphis): Current knowledge and research priorities. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 31:110–136.

Wanger, T. C., Motzke, I., Furrer, S. C., Brook, B. W., and Gruber, B. (2009). Movement patterns and habitat selection of the giant day gecko (phelsuma madagascariensis grandis) in the masoala rainforest exhibit, zurich zoo. *Salamandra*, 45(3):147–153.

Watanabe, S. and Opper, M. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, 11(12).

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and management of animal populations*. Academic press.

Wilson, B., Hammond, P. S., and Thompson, P. M. (1999). Estimating size and assessing trends in a coastal bottlenose dolphin population. *Ecological applications*, 9(1):288–300.

Wu, G. and Holan, S. H. (2017). Bayesian hierarchical multi-population multistate jolly–seber models with covariates: application to the pallid sturgeon population assessment program. *Journal of the American Statistical Association*, 112(518):471–483.

Wu, G., Holan, S. H., Avril, A., and Waldenström, J. (2021). A bayesian semi-parametric jolly–seber model with individual heterogeneity: An application to migratory mallards at stopover. *The Annals of Applied Statistics*, 15(2):813–830.

Yang, H.-C. and Chao, A. (2005). Modeling animals' behavioral response by markov chain models for capture–recapture experiments. *Biometrics*, 61(4):1010–1017.

Yoshizaki, J., Pollock, K. H., Brownie, C., and Webster, R. A. (2009). Modeling misidentification errors in capture–recapture studies using photographic identification of evolving marks. *Ecology*, 90(1):3–9.