Capture-recapture methods and models: Estimating population size

Capture-recapture methods and models: Estimating population size

Ruth King and Rachel McCrea

Abstract

This book chapter describes ecological capture-recapture studies and associated models often fitted to capture-recapture data to obtain estimates of total population size. Such estimates can be important for numerous reasons, including for example, conservation and management purposes. We focus on different forms of heterogeneity that may affect the propensity of individuals to be observed within the study period. Failing to account for such heterogeneity can lead to significant bias in the population estimates. We focus on different types of heterogeneity corresponding to recorded (discrete-valued) covariates/characteristics of individuals that are observed within the study period; in addition to unobserved heterogeneity in the form of mixture distributions. The different models are motivated and discussed, including the specification of the likelihood functions, before being applied to a real dataset. Finally we conclude with a discussion including the modern challenges which are arising due to technological advances.

Key words: Abundance; Capture histories; Discrete covariates; Heterogeneity; Likelihood; Mixtures; Models

1 Introduction

Population sizes are of interest across a wide spectrum of disciplines and geographical scales. For example, at the macro-scale, such as the number of individuals on the planet; to the micro-scale, such as the number of specific animals that inhabit a localised area; and everything in between. Many different data collection techniques are used to look at population sizes and their trends over time. These include for example national censuses sent to all households (carried out periodically in many countries, such as the UK and US every 10 years) or local sub-sampling of individual animals within a small survey area. In practice, observing all members of a given population is often infeasible for a combination of resource and pragmatic issues. For example, marine animals primarily spend their time underwater and traverse a very wide location so that survey boats may not observe the animals despite being in the same area; alternatively problem drug users live in communities and their drug use may be well hidden. Thus, reliable estimation of total population sizes, for hidden or difficult to observe populations, is an important aspect in many different areas. For example, within wildlife conservation the IUCN Red List of Threatened

Species (http://www.iucnredlist.org) classifies species and plants in relation to their conservation status and uses, amongst other criteria, population size (and trends over time of these) in categorising the different species. Alternatively, assessing the impact of different governmental policies on populations such as the homeless or problem drug users will require robust estimation of population sizes in order to detect associated changes over time; similarly financial socio-economic planning may be dependent on these sub-populations, with areas with a higher level of problem drug use requiring additional healthcare resources. See for example, Böhning *et al* (2018), for further discussion and associated examples of this type.

In this book chapter we focus on ecological populations, and the estimation of wildlife population sizes. There are many different ways that data can be collected on such populations. The most accurate, in terms of a complete census count of all individual animals within the population, is typically not possible due to the nature of the population combined with time and financial resource limitations. In such circumstances partial observation and enumeration of the population under study may be possible. These may take multiple different forms including, for example, distance sampling (where for line transect sampling observers traverse some transect and record the number of animals observed and their distance to the transect; Buckland *et al* (2004)); aerial survey data over a given region (i.e. a snapshot of individuals within the given location and time; Fleming and Tracey (2008)); and nets/traps to capture animals that pass through a given area (Ralph and Dunn, 2004). Typically, multiple types of surveys may be conducted and a combination of both spatial and temporal replications used.

We focus on repeated surveys over time, such that there are a series of "capture" occasions where researchers go into the field at distinct times, and subsequently identify and record all individuals observed from the population of interest. Such studies are generally referred to as capture-recapture (or mark-recapture/resight) studies (see for example King (2014); McCrea and Morgan (2014)). We note that for the analogous case within epidemiological studies, the capture occasions correspond to different lists of individual names, such as hospital records, police records etc.; the data are consequently often referred to as multi-list data, and the approach of using such data to estimate the total population size "multiple systems estimation" (see for example, Bird and King (2018) for a review).

For standard capture-recapture studies it is assumed that individuals are uniquely identifiable at each capture occasion. This assumption permits the construction of individual capture histories for each individual observed within the study, detailing which capture occasions they are (or are not) observed. Traditionally, to uniquely identify individuals, marks are applied to an individual the first time it is observed, such as a tag attached (e.g. sheep, seals) or a ring applied to its leg (e.g. common for many bird species). However, new techniques for uniquely identifying individuals include the use of DNA (for hair/scat samples; Lukacs and Burnham (2005a)) or natural unique markings and photographic identification (e.g. great crested newts, seals, tigers; Worthington *et al* (2018b); Smout *et al* (2011); Royle *et al* (2009)) - the latter becoming increasingly common with the use of motion sensor traps. There are several attractions for the use of these latter new technologies as they do not disturb the animals by the addition of an artificial mark and the associated marks cannot

generally be lost; however there is a trade-off as uniquely identifying individuals can be more difficult and can lead to additional uncertainty in the capture histories (see Section 5 for further discussion of modern techniques).

The primary parameter of interest within such capture-recapture studies is the total population size. However, in estimating this we also need to consider the capture probabilities of the animals, i.e. the probability that a given individual is observed at a given capture occasion - though these parameters are essentially regarded as nuisance parameters. The capture probabilities may be dependent on a number of different factors such as the effort expended at the given capture occasion (the longer amount of time, or increased number of observers, the more likely an individual may be seen); the time of day that animals are attempted to be observed (animals may be more active at different times of day, such as dawn or dusk, dependent on the species); geographical location (seeing an individual in areas with long grass is typically much harder than seeing an individual in open short grassland); and many other additional factors including individual characteristics. Despite the capture probabilities being nuisance parameters, and typically not of interest themselves, the dependence structure of the capture probabilities may have an appreciable impact on the overall estimate of total population size - see for example Cubaynes *et al* (2010). We discuss different forms of heterogeneity that may affect the capture probabilities, and hence estimate of the total population size. We focus on models associated with discrete forms of heterogeneity (both at the global and individual levels), which may be either observed or unobserved. In this book chapter we focus on the models themselves and for illustration apply a classical approach to the data. For a discussion of the use of Bayesian methods, see for example, King (2012); King *et al* (2009).

In Section 2 we give a brief history of the use of capture-recapture data in terms of the number of capture occasions used, from the simplest, minimum, case of just two occasions to multiple occasions and associated modelling issues, focusing on closed populations where it is assumed that there no births/deaths or migrations within the study period. In Section 3 we discuss in more detail the different forms of heterogeneity (observed and unobserved; time-invariant and time-varying) that may be present in the system, particularly in reference to the capture probabilities, and the associated impact on the statistical analysis. We extend the ideas to open populations in Section 4. In Section 5 we provide a general discussion and briefly describe further extensions to the models described within this book chapter. Throughout these sections we primarily focus on a simple capture-recapture example relating to deer mice for comparison of the different models.

2 Closed population capture-recapture studies

Within this section we begin by describing the simplest form of capture-recapture studies, where there are just two sampling occasions, before discussing issues relating to this two-sample approach. Some of these shortcomings, and necessary associated assumptions, can be addressed by increasing the number of capture occasions within the study. We define the multiple occasion approach and discuss in further detail the associated heterogeneities that may arise, that, if unaccounted for, can lead

to significant bias within the estimates of the total population size. Models that account for different types of heterogeneity given the available information within the data are discussed. For the different models described, we fit the models to data from a capture-recapture study of mice and discuss the corresponding results.

2.1 Simple beginnings - Lincoln-Petersen estimator

We begin by describing the simplest capture-recapture study where there are just two capture occasions. We assume that the population is closed so that the population size is fixed over the study period with no entries to, or exits from, the population. The capture history for a given individual observed within the study corresponds to the pair (x_1, x_2) where $x_t = 0, 1$ corresponding to whether the individual was not observed ($x_t = 0$) or observed ($x_t = 1$) at time $t = 1, 2$. There are three possible observable capture history combinations:

$$(1,1) \qquad (1,0) \qquad (0,1).$$

The first pair corresponds to an individual being observed at both capture occasions; the second to only being observed at the first occasion; and the final pair to only being observed at the second occasion. Clearly there is a further *unobservable* capture history: $(0, 0)$.

The data can be summarised as the number of individuals with each given observed capture history (or pair). Let n_{ij} denote the number of individuals with capture history (i, j) for $i, j = 0, 1$. The observed data correspond to n_{11}, n_{10} and n_{01} - the number of individuals with each distinct combination of observable capture histories. The number of individuals not observed at either capture occasion, denoted by n_{00} is unknown. The total number of observed individuals is given by $n = n_{11} + n_{10} + n_{01}$. The true total population size is given by $N = \sum_{i=0}^{1} \sum_{j=0}^{1} n_{ij} = n_{00} + n$ and is the primary parameter of interest to be estimated.

The observed data are often usefully represented in the form of an incomplete (contingency) table, where the cell entries correspond to the number of individuals with the given capture history - see Table 1.

		Capture occasion 2	
		0	1
Capture	0	$n_{00} = ?$	n_{01}
occasion 1	1	n_{10}	n_{11}

Table 1: An example of the observed data of a capture-recapture study with two capture occasions. The cell n_{00} corresponding to the number of individuals not observed by either occasion is unobserved and hence unknown.

The pattern of the observed cell entries permits the estimation of the total population size, N, (or equivalently the number of individuals in the population not observed at either occasion, n_{00}). For example, the following argument can be applied, given the assumption that the probability an individual is observed at occasion

2 is independent of whether or not they are observed at occasion 1. In this case we would expect the ratio of observed individuals to unobserved individuals at occasion 2 to be approximately equal for the two subsets of animals corresponding to those (i) observed at occasion 1, and (ii) not observed at occasion 1. Mathematically, this means that we would expect:

$$\frac{n_{11}}{n_{10}} \approx \frac{n_{01}}{n_{00}}.$$

Rearranging the equation we obtain the corresponding estimates:

$$\hat{n}_{00} = \frac{n_{01}n_{10}}{n_{11}}; \quad \text{and} \quad \hat{N} = \frac{(n_{11}+n_{10})(n_{11}+n_{01})}{n_{11}}.$$

This estimate can be obtained using analogous arguments by equating different ratios or proportions of the cell entries which should be (approximately) equal if the probability an individual is observed on the second occasion is independent of whether or not they were observed on the first capture occasion.

This estimate for the total population size, $\hat{N}$, is typically referred to as the Lincoln-Petersen estimate (Lincoln, 1930; Petersen, 1896). For a discussion of the history of the estimate see for example, Goudie and Goudie (2007). The estimate pre-dates both Lincoln and Petersen by several hundred years, dating back to at least the 1600s, when Graunt used a similar idea to estimate the size of the population of England and the effect of the plague. However, the most famous early application of the estimate is typically attributed to Laplace (Laplace, 1786) - and the application of these ideas to estimate the total population of France in 1802 using "occasions" corresponding to birth records across France and local complete municipality census records (Bird and King, 2018).

The Lincoln-Petersen estimate is also the maximum likelihood estimate (MLE) for the total population size so that it is a consistent estimate, however it is not unbiased. This led to the proposal of the Chapman estimator (Chapman, 1951) that is less biased and given by:

$$\hat{N} = \frac{(n_{11}+n_{10}+1)(n_{11}+n_{01}+1)}{n_{11}+1} - 1.$$

2.1.1 Example: mice data

We consider an example of capture-recapture data relating to deer mice (*peromyscus maniculatus*) collected using traps over a period of 6 consecutive days at East Stuart Gulch, Colorado by V. Reid. See Huggins (1991) for further discussion of the data. For the purposes of demonstrating the application of the two sample estimators, we simplify the data by defining capture occasion 1 to be days 1-3 and capture occasion 2 to be days 4-6. Thus an animal caught in a trap on any of the days 1, 2 or 3 was recorded as being observed at capture occasion 1; and similarly an animal caught in a trap on any of the days 4, 5 or 6 was recorded as being observed at capture occasion 2. Note that we will consider the data more fully in Sections 2.2 and 3, and in particular we will remove the combining of days implemented here. The corresponding, simplified, two sample capture-recapture data are provided in Table 2.

		Capture occasion 2	
		0	1
Capture	0	?	9
occasion 1	1	2	27

Table 2: Observed mice data using a two sample capture-recapture study.

Applying the Lincoln-Petersen estimator we obtain an estimate of the total population size of 38.41. The corresponding 95% confidence interval for the total population using a non-parametric bootstrap approach is given by (38.19, 38.81). We note that applying the Chapman estimator provides a very similar estimate of 38.64 and associated 95% non-parametric bootstrap confidence interval of (38.00, 40.09). Both estimates suggest that the majority of individuals have been observed within the study with little uncertainty.

Throughout this chapter we apply a non-parametric bootstrap approach for calculating the confidence intervals of the parameters. To calculate the associated confidence interval of the parameters, within each bootstrap replication, we simulate "new" data by sampling with replacement over the set of observed capture histories, and calculate the associated MLEs of the model parameters. Thus, we take the capture histories as the sampling level. We use a percentile bootstrap confidence interval by taking the lower and upper 2.5% quantiles of the associated MLEs obtained over the bootstrap replicates. The non-parametric bootstrap approach is used for several reasons: (i) the bootstrap approach does not rely on the asymptotic normality assumption which is perhaps questionable in this case given the amount of available data; (ii) the estimates of population size are close to the boundary leading to unreliable confidence intervals (using the delta method to obtain standard errors of transformed estimates even when this may appear reasonable may lead to naive confidence intervals with bounds outside the permissible range); (iii) for ease of use for calculating associated confidence intervals of functions of the estimated parameters (see for example Section 3.2 where we individually estimate the population sizes of sub-groups corresponding to males and females, and wish to also obtain an estimate of the total population size).

For the above observed data, with only two capture occasions, the estimates rely on a number of assumptions. For example, this includes that the population is homogeneous with regard to capture probabilities, so that all individuals have the same capture probability at a given occasion. In particular this implies that the capture occasions are independent of each other so that an individual observed at the first capture occasion does not affect the associated probability that the individual is observed on the second capture occasion - recall the rationale underlying the Lincoln-Petersen estimator. In other words the probability that an individual is observed at the second capture occasion is independent of whether or not they are observed at the first capture occasion. Deviations in the system from such assumptions will lead to unreliable and biased estimates of the total population

size. However, unfortunately, for two sample capture-recapture studies it is not possible to test such assumptions - we have 3 observed data values permitting the estimation of 3 parameters: the total population size, the probability an individual is observed at capture occasion 1, and the probability an individual is observed at capture occasion 2. There are no additional degrees of freedom to model or test for independence or additional heterogeneity. We have however simplified the data for this example for comparability with later analyses. A simple visual inspection of the more detailed data suggests that these assumptions are likely to be violated. For example, some mice are captured on all 6 occasions whilst others are captured only once, leading us to question the assumption of equal detectability of all individuals. An obvious way to investigate such issues and address them within the modelling process is to extend the study design and consider multiple capture occasions (as is the case with the more detailed mice data).

2.2 Multiple capture occasions

Expanding the number of capture occasions is a natural extension to the two-sample case and associated Lincoln-Petersen estimator. Suppose that there are a total of T capture occasions. The data collection process extends immediately so that on each capture occasion $t = 1, \ldots, T$ individuals are uniquely identified and recorded before being released. A number of assumptions are typically made with regard to the population, including:

- The population is closed (though we discuss how this may be relaxed in Section 4);
- Animals cannot lose or change their unique identifying marks so that there is no misidentification within the study (i.e. there are no false positives or false negatives with regard to matching individuals across sightings);
- All individuals behave independently of each other.

For further discussion of these (and additional) assumptions, see for example, McCrea and Morgan (2014).

2.2.1 Schnabel census

Early studies to look at multiple (> 2) capture occasions simply recorded the number of times that each individual was observed within the study period, and such data are typically referred to as a Schnabel census (Schnabel, 1938). This can be thought of as similar to a binomial coin-tossing example - where we simply record the number of "successes" (i.e. heads or tails) but not where the successes occur in the T coin tosses. The data were recorded as the number of individuals observed a total of t times within the study period, denoted by f_t, for $t = 1, \ldots, T$. For notational simplicity we let $\boldsymbol{f} = \{f_1, \ldots, f_T\}$. The unobserved value f_0 corresponds to the number of individuals not observed within the study. The summary statistics f_t for $t = 1, \ldots, T$ are the minimal sufficient statistics for individual heterogeneity models which we will discuss in Section 3.1. We note that, as for a binomial experiment, it is assumed

that for each individual the associated capture (or non-capture) at each capture occasion are assumed to be independent Bernoulli trials with the same probability of "success" (i.e. capture probability). Recording only the number of times an individual is observed means that we cannot investigate any time dependence on the capture probabilities (as we do not know at what times individuals are observed). To more fully explore dependence structures we need the full encounter (or capture) histories of individuals over all the capture occasions.

2.2.2 Capture histories

For T capture occasions there are a total of $2^T - 1$ distinct observable capture histories (though in general we may not observe all of these possible histories within a given dataset) corresponding to an individual being observed or not at each possible capture occasion. Thus, here we consider the exact capture occasions that an individual is observed, rather than simply the summary of how many times an individual is observed, as in the Schnabel census. Notationally, we let:

$$x_{it} = \begin{cases} 0 & \text{if individual } i \text{ is not observed at time } t; \\ 1 & \text{if individual } i \text{ is observed at time } t. \end{cases}$$

The capture history for individual i is denoted by $\boldsymbol{x}_i = \{x_{i1}, \ldots, x_{iT}\}$. For notational convenience, we let $\boldsymbol{x}_0$ denote the unobserved capture history, i.e. $\boldsymbol{x}_0 = \{0, \ldots, 0\}$. The set of all observed capture histories is denoted by $\boldsymbol{x} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_D\}$, where D denotes the number of individuals observed within the study. The raw data (x_{it} values) for the individuals observed in the study can be displayed in the form:

	Capture occasion (t)				
unique ID	1	2	3	$\cdots$	T
1	1	1	0	$\ldots$	1
2	0	1	0	$\ldots$	1
3	0	1	0	$\ldots$	0
$\vdots$					
D	0	0	0	$\ldots$	1

The "unique ID" corresponds to the unique identifier (in this case number) assigned to each individual; and a total of D individuals are observed within the study period. Thus D provides a lower bound on the total population size, but of course does not provide any indication regarding the number of individuals not observed within the study. Alternatively these data can be summarised in the form of an incomplete 2^T contingency table, analogous to Table 1 extended to multiple capture occasions, where the cell corresponding to not being observed at any capture occasion is again unknown.

2.2.3 Model formulations

There are two common formulations for the general model that arise for capture-recapture data corresponding to the number of individuals observed by each combination of capture occasions (i.e. contingency table cell): a Poisson model; and a

multinomial model. Both of these formulations are examples of a generalised linear model for the data.

Poisson model

Mathematically, we index the set of possible contingency table cells by $j = 0, 1, \ldots, 2^T - 1$ and for simplicity assume that cell $j = 0$ corresponds to the unobserved capture history. The associated cell entries are denoted by $n_0, \ldots, n_{2^T-1}$. For notational simplicity we let $\boldsymbol{n} = \{n_0, \ldots, n_{2^T-1}\}$, so that the number of observed individuals is given by $D = \sum_{j=1}^{2^T-1} n_j$. We assume that for $j = 0, \ldots, 2^T - 1$,

$$n_j | \mu_j \sim Poisson(\mu_j),$$

independently, where μ_j denotes the mean cell counts. The cell counts μ_j are assumed to be a function of a number of model parameters, such as an underlying prevalence level and propensity of being observed at each capture occasion. In general, we let the mean cell count model parameters be denoted by $\boldsymbol{\theta}$.

We note that only the cell entries for $j = 1, \ldots, 2^T - 1$ are observed. We let the set of observed cell entries be denoted by $\boldsymbol{n}^{obs} = \{n_1, \ldots, n_{2^T-1}\}$. The number of individuals in the population but not observed in the study, denoted n_0, is unknown. However, since the cell entries are *independent* given the model parameters, we can apply a two-step estimation process. First we fit the model to the observed cell entries $j = 1, \ldots, 2^T - 1$ and estimate the model parameters. The corresponding likelihood for the *observed* data is given by,

$$L(\boldsymbol{\theta}; \boldsymbol{n}^{obs}) = \prod_{j=1}^{2^T-1} \frac{\exp(-\mu_j)(\mu_j)^{n_j}}{n_j!}.$$

For example we may obtain the maximum likelihood estimates (MLEs) $\hat{\boldsymbol{\theta}}$ and associated MLE for the mean number of individuals not observed, denoted, $\hat{\mu}_0$. Secondly, we estimate the unobserved cell entry using $\hat{n}_0 = \hat{\mu}_0$, via the property of the MLE for a Poisson random variable (recall that the cell entries are assumed to be independent Poisson random variables). The overall estimate of the total population size is given by,

$$\hat{N} = \hat{n}_0 + \sum_{j=1}^{2^T-1} n_j = \hat{n}_0 + D.$$

Using the same form of two-step process a Bayesian approach can be implemented and a posterior distribution for $\boldsymbol{\theta}$ obtained up to proportionality. A sample from the posterior distribution can be obtained using standard techniques, such as as Markov chain Monte Carlo; and hence a sample from the posterior distribution of n_0 can be obtained by subsequently simulating a Poisson random variable with given mean (calculated from the sampled $\boldsymbol{\theta}$ values).

Multinomial model

Alternatively we can consider a multinomial model formulation. Suppose that we randomly select an individual from the population of interest - that individual must

have one of the possible 2^T capture histories. The probability associated with capture history $j = 0, \ldots, 2^T - 1$ is denoted by q_j; and we set $\boldsymbol{q} = \{q_0, \ldots, q_{2^T-1}\}$. Summing over the number of individuals in the population leads to,

$$\boldsymbol{n}|N, \boldsymbol{q} \sim Multinomial(N, \boldsymbol{q}).$$

The corresponding likelihood of the capture histories, conditional on the total population size, is given by,

$$L(N, q; \boldsymbol{n}^{obs}) = \frac{N!}{\prod_{j=0}^{2^T-1} n_j} \prod_{j=0}^{2^T-1} q_j^{n_j},$$

where $N = n_0 + D$ (or conversely $n_0 = N - D$) and N corresponds to the total population size. This likelihood can be directly maximised in order to obtain unconditional MLEs of the parameters, and in particular the total population size.

We note that we can factorise the multinomial likelihood as follows:

$$\begin{aligned} L(N, \boldsymbol{q}; \boldsymbol{n}^{obs}) &= \frac{N!}{D!(N-D)!} q_0^{N-D}(1-q_0)^D \times \frac{D!}{\prod_{j=1}^{2^T-1} n_j} \prod_{j=0}^{2^T-1} \left(\frac{q_j}{1-q_0}\right)^{n_j} \\ &= \frac{N!}{D!(N-D)!} q_0^{N-D}(1-q_0)^D L(q; \boldsymbol{n}^{obs}), \end{aligned} \tag{1}$$

where q_0 denotes the probability that an individual is not observed within the study; so that $1 - q_0$ corresponds to the probability an individual is observed at least once within the study. The first term is a binomial component corresponding to the number of individuals not observed (or conversely observed at least once), given the total population size N; the second term, $L(q; \boldsymbol{n}^{obs})$ is a conditional multinomial component corresponding to the number of individuals observed by each combination of sources (i.e. cells $j = 1, \ldots, 2^T$), conditional on an individual being observed at least once within the study.

Relationship between Poisson and multinomial models

It is straightforward to show that for observations from independent Poisson distributions, if we condition on the sum of the observations, then the distribution of the observations has a multinomial distribution. The associated probabilities of the multinomial distribution are equal to the Poisson mean for that cell divided by the sum of all Poisson mean values. Thus, suppose that for $j = 0, \ldots, 2^T - 1$ we let:

$$\mu_j = N p_j.$$

In other words the Poisson mean cell counts are equal to the total population size multiplied by the corresponding cell probabilities. Conditional on the total population size the cell counts have a multinomial distribution with probabilities given by,

$$\frac{\mu_j}{\sum_{j=0}^{2^T-1} \mu_j} = \frac{N p_j}{N \sum_{j=0}^{2^T-1} p_j} = p_j.$$

The Poisson model can then be factorised (after algebra) as follows:

$$\begin{aligned} L(N, \boldsymbol{q}; \boldsymbol{n}) &= \frac{\exp(-N(1-q_0))(N(1-q_0))^D}{D!} \times \left\{ \frac{D!}{\prod_{j=1}^{2^T-1} n_j!} \prod_{j=1}^{2^T-1} \left(\frac{q_j}{1-q_0} \right)^{n_j} \right\} \\ &= \frac{\exp(-N(1-q_0))(N(1-q_0))^D}{D!} \times L(q; \boldsymbol{n}^{obs}). \end{aligned}$$

The first term corresponds to a Poisson component for the total number of observed individuals, n, (recall that the sum of independent Poisson random variables is again Poisson with mean equal to the sum of the means); and the second component corresponds to the distribution of the cell counts given the number of observed individuals (which is thus independent of the total population size N). Comparing to equation (1) we can see that this Poisson likelihood shares the same conditional multinomial component, $L(q; \boldsymbol{n}^{obs})$, (corresponding to the likelihood of the observed cell counts conditional on only the observed individuals) - the difference lies with the first term - either a Poisson term for the total number of individuals observed; or a binomial component for the number of observed individuals (or equivalently unobserved individuals) given the total population size. Sandland and Cormack (1984) showed that the MLEs of the these multinomial and Poisson formulations are the same. For further discussion see for example, McCrea and Morgan (2014).

Note that within this chapter we will focus on the multinomial formulation as we will consider heterogeneity at the individual level, and the most intuitive way of specifying the multinomial cell probabilities corresponds to individuals (or groups of individuals) within the population. In particular, we can rewrite the multinomial formulation as follows. Let p_{it} denote the probability that individual i is observed at time t; and let the set of capture probabilities be given by $\boldsymbol{p} = \{p_{it} : i = 1, \ldots, D; t = 1, \ldots, T\}$. In addition we let p^* denote the probability that an individual is observed at least once within the study period; so that $1 - p^*$ is the probability that an individual is not observed within the study period. Thus $p^* = 1 - q_0$ in the previous notation. The general multinomial form of the unconditional likelihood, specified at the individual level, is given by

$$L(N, \boldsymbol{p}; \boldsymbol{x}) \propto \frac{N!}{(N-D)!} \prod_{i=1}^{D} \Pr(\boldsymbol{x}_i|\boldsymbol{p}) \times \Pr(\boldsymbol{x}_0|\boldsymbol{p})^{N-D} \tag{2}$$

where $\Pr(\boldsymbol{x}_i|\boldsymbol{p})$ denotes the probability of capture history $\boldsymbol{x}_i$ given the set of capture probabilities $\boldsymbol{p}$; and $\Pr(\boldsymbol{x}_0|\boldsymbol{p})$ the probability of an encounter history with no captures. For notational simplicity, here and throughout the remainder of this chapter, we drop the dependence on the capture probabilities and simply write $\Pr(\boldsymbol{x}_i)$ and $\Pr(\boldsymbol{x}_0)$ instead of $\Pr(\boldsymbol{x}_i|\boldsymbol{p})$ and $\Pr(\boldsymbol{x}_0|\boldsymbol{p})$, respectively. We note that we can write,

$$\Pr(\boldsymbol{x}_i) = \prod_{t=1}^{T} p_{it}^{x_{it}} (1 - p_{it})^{(1-x_{it})};$$

and

$$\Pr(\boldsymbol{x}_0) = 1 - p^*.$$

In general we specify restrictions on the capture probabilities, to represent different possible dependencies. For example, we may assume that $p_{it} = p$ for all individuals i and times t - so that the capture probability is constant over time and individuals; in this case $(1 - p^*) = (1 - p)^T$. Alternatively the capture probabilities may vary across capture occasions, but be consistent across individuals, so that $p_{it} = p_t$ for all individuals i; and $(1 - p^*) = \prod_{t=1}^{T}(1 - p_t)$. We discuss some of these models in the context of the mice data considered previously, before discussing these dependencies in more detail in Section 3.

2.2.4 Example: mice data

We return to the mice data discussed in Section 2.1.1. For this study there were 6 separate trapping occasions on consecutive days - one for each day of the study. Previously we reduced the data to 2 capture occasions by combining the trapping occasions on days 1-3 and days 4-6. Now we consider each day of trapping as a separate capture occasion. The data are provided by Huggins (1991) - we consider a subset of the full data (removing further covariate information of the individual animals) and present these in Table 3.

A total of 38 individuals are observed over the study period such that of these 17 are female and 21 are male. We ignore this additional individual characteristic data for the moment - but will return to this in Section 3.2. Of the 38 individuals observed only 1 individual is observed at every capture occasion; with 9 individuals observed only once. However, the number of individuals in the study population but not observed is unknown.

The previous estimate of the total population size using the Lincoln-Petersen estimator on the reduced two-sample data provided an estimate of 38.41 with 95% confidence interval (38.19, 38.81). We now extend the statistical analysis and use the individual days to be 6 separate capture occasions. In general, there is no closed form expression for the MLE of the total population size or capture probability, and a numerical optimisation algorithm is applied to maximise the (log-)likelihood to obtain the associated MLEs and associated confidence intervals. We consider 2 different models here (i) the model with a constant capture probability over time (denoted model M_0); and (ii) the model with a different capture probability for each capture occasion (denoted model M_t) - we discuss temporal heterogeneity in further detail in Section 3.2.

Model M_0

The simplest model for such multiple capture occasion data assumes a constant capture probability, p, for all individuals. Thus we have that $p_{it} = p$ for all individuals i and time t. The likelihood can be simplified to be of the form:

$$L(N, p; R, D) \propto \frac{N!}{(N - D)!} p^R (1 - p)^{NT - R}, \tag{3}$$

where,

$$R = \sum_{i=1}^{D} \sum_{t=1}^{T} x_{it} = \sum_{t=1}^{t} t f_t;$$

Individual number	Capture occasion (t) 1	2	3	4	5	6	Individual number	Capture occasion (t) 1	2	3	4	5	6
1	1	0	0	1	1	1	18	1	1	1	1	1	1
2	1	1	1	0	1	1	19	1	1	0	0	1	1
3	1	0	0	1	0	0	20	1	1	0	1	1	1
4	0	1	0	0	1	0	21	1	1	1	1	1	1
5	0	1	1	0	0	1	22	1	1	0	1	1	1
6	0	1	0	1	0	1	23	1	1	1	1	1	0
7	0	1	0	0	0	1	24	1	1	1	0	0	1
8	0	0	1	0	0	0	25	1	1	1	1	1	1
9	0	0	1	1	1	1	26	1	1	0	1	1	1
10	0	0	1	0	1	1	27	1	1	0	1	1	1
11	0	0	1	1	1	1	28	1	1	1	1	1	1
12	0	0	1	0	1	0	29	1	0	1	1	1	0
13	0	0	1	0	0	0	30	0	1	0	0	0	1
14	0	0	0	1	0	0	31	0	1	0	1	0	1
15	0	0	0	1	1	0	32	0	1	1	0	1	0
16	0	0	0	0	1	0	33	0	1	0	0	1	1
17	0	0	0	0	0	1	34	0	0	0	1	1	1
							35	0	0	0	0	1	0
							36	0	0	0	0	1	0
							37	0	0	0	0	0	1
							38	0	0	0	0	0	1

Table 3: The capture histories of the 38 deer mice observed (arbitrarily numbered) over the 6 capture occasions. A "0" indicates the individual is unobserved at the given capture occasion; and a "1" that the individual is observed at the given capture occasion. Individuals 1-17 are female (left-hand side); individuals 18-38 are male (right-hand side).

such that $D = 38$ and $T = 6$. Recall, from Section 2.2.1, f_t denotes the number of animals observed at t different capture occasions within the study. We note that $R = 120$ simply corresponds to the number of captures observed within the study period.

For model M_0 the summary statistics R and D are minimally sufficient. The likelihood can be easily interpreted as a form of binomial expression where, for each individual in the population, we have a series of simple independent and identical Bernoulli trials corresponding to whether or not the individual is observed at each of the capture occasions. Under this binomial model (i.e. model M_0) we obtain an MLE for N of 38 and associated 95% non-parametric bootstrap confidence interval of (38.00, 38.22). We note that this estimate is slightly smaller than the previous Lincoln-Petersen estimator with a reduced 95% confidence interval width. It is perhaps not surprising that the confidence interval is narrower, since we are estimating the total population size using all 6 capture occasions (and not aggregated to simply two occasions), so that we are using greater amount of information and have

only 2 parameters to estimate (i.e. N and p). We also note that the estimate of the total population is on the boundary of the parameter space. Further the MLE of the (common) capture probability is 0.526 with 95% non-parametric bootstrap confidence interval (0.434, 0.614).

We note that the Lincoln-Petersen estimator does not assume the same capture probability at each of the two capture occasions, simply that each individual has the same probability of capture within a single capture occasion. The capture probabilities themselves are assumed to be different. In particular, the MLEs (95% non-parametric bootstrap confidence intervals) of the capture probabilities are 0.729 (0.595, 0.889) for capture occasion 1 and 0.989 (0.856, 0.994) for capture occasion 2. Thus this suggests that assuming a constant capture probability may not be appropriate, for example, as sampling conditions can vary over time. Of course the Lincoln-Petersen estimator has simplified the data by aggregating the capture occasions, but, for consistency, the same number of capture occasions are aggregated within each of the two sampling periods.

We now consider the six capture occasions in more detail. Under the assumption of a constant capture probability we would expect approximately the same number of individuals to be observed at each capture occasion. We let R_t denote the number of individuals observed at capture occasion $t = 1, \ldots, T$, and set $\boldsymbol{R} = \{R_1, \ldots, R_T\}$. For the mice data we have the following number of observed captures at each occasion:

	Capture occasion (t)					
	1	2	3	4	5	6
Number observed (R_t)	15	20	16	19	25	25

These data would perhaps suggest that the capture probability is not constant over time - so that we consider an alternative model allowing for different capture probabilities at each capture occasion.

Model M_t

For this model we assume that each of the capture occasions have a different capture probability (but that the capture probabilities are the same for each individual at a given capture occasion). We let p_t denote the capture probability at time $t = 1, \ldots, T$, and for notational simplicity we let the set of all capture probabilities be denoted by $\boldsymbol{p} = \{p_1, \ldots, p_T\}$. The multinomial likelihood formulation for this model can be reduced to:

$$L(N, \boldsymbol{p}; D, \boldsymbol{R}) \propto \frac{N!}{(N-D)!} \prod_{t=1}^{T} p_t^{R_t} (1-p_t)^{N-R_t}, \tag{4}$$

where for the mice data $D = 38$; $T = 6$; and the R_t values are given above. Using a numerical optimisation algorithm we obtain the MLE (95% non-parametric bootstrap confidence interval) of N to be 38.00 (38.00, 38.05). This estimate is similar to the previous estimate for model M_0, but with a slightly shorter confidence interval (i.e. less uncertainty), most likely due to the improved fit of the model to the

data due to the greater flexibility of the additional parameters in the model. The associated MLEs and associated 95% non-parametric bootstrap confidence intervals of the capture probabilities are:

Capture occasion	MLE $\hat{p}_t$	95% confidence interval
1	0.395	(0.237, 0.553)
2	0.526	(0.364, 0.684)
3	0.421	(0.263, 0.579)
4	0.500	(0.342, 0.658)
5	0.658	(0.500, 0.816)
6	0.658	(0.500, 0.816)

Thus we note that there is some variability with regard to the estimated capture probabilities, although the confidence intervals are overlapping for all capture occasions.

To formally compare the models, M_0 and M_t, (and additional models throughout this chapter), we will consider Akaike's Information Criterion (AIC; Akaike (1974)). The AIC statistic is defined to be equal to the deviance (-2 $\times$ log-likelihood evaluated at the MLE of the parameters) plus 2 $\times$ the number of parameters. For further discussion of model selection techniques, and in particular with reference to the AIC statistic, see for example, Burnham and Anderson (2002). The associated AIC statistics of model M_0 and M_t are 113.51 and 113.67, respectively. The AIC statistic is a relative goodness of fit measure, and so in the calculation of the deviance within this expression, we only use the multinomial combinatorial terms containing the unknown population size (the other combinatorial terms are constants). There is very little difference between the models, M_0 and M_t with regard to the AIC statistic, and the additional time dependence of the capture probabilities have not significantly increased the fit of the model to the data, allowing for the additional model complexity (i.e. number of parameters) - this is perhaps not surprising given the overlapping confidence intervals of the capture probabilities. Further model refinements could be considered such as considering "low" catchable times (occasions 1-4) and "high" catchable times (occasions 5-6) - however this is fairly arbitrary and without some ecological justification regarding knowledge of the data collection process is difficult to sensibly interpret. Clearly such results can lead to useful discussions with the data collectors and ecologists to understand (and possibly improve) future study protocols or additional data being recorded (such as effort expended at each capture occasion; or weather conditions etc.). Thus for simplicity we will generally assume no temporal heterogeneity for these data.

As previously mentioned, the AIC statistic is a *relative* goodness-of-fit measure - given the set of models considered which model fits the data best. In order to assess *absolute* goodness-of-fit (does the given model fit the observed data well?) it is possible to compare the observed and expected values for the number of individuals with each possible capture history (i.e. cell entries) and apply a Pearson chi-squared test, for example. Considering such tests will typically lead to issues of sparsity with small observed and expected values. However for some models there are sensible

ways of pooling observations to overcome this issue. For example, consider the Schnabel census for the mice data - here the data can be summarised via the use of the sufficient statistics $\boldsymbol{f} = \{9, 6, 7, 6, 6, 4\}$. The expected values of $\boldsymbol{f}$ under model M_0 are $\{2.9, 8.0, 11.9, 9.9, 4.4, 0.8\}$, thus by simple inspection there would appear to be a large discrepancy between the observed and expected values, under the given model. Formally, applying Pearson's chi-squared goodness-of-fit test we obtain an observed test statistic of 30.27; and an associated p-value of 1.2×10^{-6}, providing strong evidence of a lack of fit under model M_0. Thus for these data there appear to be other factors that have not been accounted for - we discuss next different forms of heterogeneity that may impact the capture probabilities (and hence total population size).

3 Heterogeneity

Heterogeneity can arise in many different forms and can have a noticeable impact on the data observed and hence a significant effect on the estimate for the total population size. Let us temporarily return to the two sample Lincoln-Petersen experiment. Suppose capture heterogeneity exists within the population, so that a proportion of the population is much more likely to be captured on each occasion than the rest of the population. This means that on the second sampling occasion the proportion of already marked individuals is likely to be higher (since marked individuals have a higher probability overall of being caught than unmarked individuals), and the resulting population size estimate will be negatively biased (i.e. an underestimate of the population size). Conversely suppose that the marking of an individual is an unpleasant experience, which changes the behaviour of the individual so that they avoid future captures. This means that on the second capture occasion a marked individual has a lower probability of being observed than an unmarked individual - often referred to as a "trap shy" response. This subsequently leads to a population size estimate that is positively biased (i.e. an overestimate of the population size).

The manifestation of heterogeneity in capture is not always simply due to study design, and therefore it is not always possible to predict prior to model fitting that you need to consider models incorporating capture heterogeneity. A relatively simple diagnostic test, proposed in Jeyam *et al* (2018), can be applied to capture-recapture data with multiple occasions to detect whether the capture histories exhibit capture heterogeneity. The test outcome can guide which models should be considered as potential candidate models for the given data set. The test compares the previous number of captures to the number of future captures and calculates a non-parametric test to test for positive association between these values. No model structure is assumed for the computation of this test. The cause of capture heterogeneity is not always clear. For example, some individuals may be more likely to be captured because of their role within a population. For other populations it is not obvious what the differences are and they may be as subtle as some individuals showing no fear to being captured and therefore would be attracted to baited traps; or some animals being more curious than others when humans approach.

Otis *et al* (1978) identified and described three primary sources of heterogeneity corresponding to temporal (t); behavioural (b); and individual heterogeneity (h).

We briefly discuss each of these different forms of heterogeneity:

Temporal heterogeneity: Temporal heterogeneity naturally reflects changes in capture probabilities across the different capture occasions - as previously discussed in Section 2.2.4. Such changes may reflect for example, different weather conditions that affects the behaviour of individuals or search strategy used by the observers; or different efforts that may be expended within the capture occasions - so that the capture probabilities at the different capture occasions are not constant but vary.

Behavioural heterogeneity: Behavioural effects remove the independence assumption between captures and allows for *trap happy* or *trap shy* responses. In other words the behaviour of an individual may change following their initial capture - they may become more likely to be observed at future capture occasions if the experience is profitable such as providing available food (a *trap happy* response); or less likely to be observed at future capture occasions if the initial capture experience is unpleasant such as being physically handled by an observer or unable to escape a trap (a *trap shy* response). In the simplest case there are two distinct capture probabilities corresponding to the initial capture probability of an individual and the subsequent recapture probability of the individual.

Individual heterogeneity: Individual heterogeneity assumes that the population is not homogeneous and so the capture probabilities of individual animals differ. This may be a result of many different factors including, for example, natural variability of individual behaviour (some individuals may simply be more adventurous than others); relationship between the study region and an individual's home range; or characteristics that may affect behaviour, for example males and females often display different behaviours.

Models may incorporate any combination of these heterogeneities. To represent the different possible models the notation M with the set of subscripts t, b and h (used in this order presented) is typically used to denote the set of heterogeneities present. For example, M_b denotes the model with only behavioural heterogeneity present; the model M_{th} the model with both temporal and individual heterogeneity present. We note that the model notation does not correspond to a single unique model - but simply denotes the heterogeneities present in the model - there are many possible model formulations for the different heterogeneities. For example, for M_t we may specify arbitrary capture probabilities for each capture occasion (to be estimated) or a catch-effort model where the capture probabilities are written as a function of the recorded effort used at each capture occasion. Alternatively for model M_b, the trap response may be permanent, or the behavioural effect may decrease over time since last capture, as an individual "forgets" its previous capture.

In this chapter we focus on individual heterogeneity models where different individuals in the population may have different capture probabilities (though note that in Section 3.3 that a behavioural response can be regarded as a special case of individual time-varying heterogeneity). In particular, within this section we explore three particular types of heterogeneity: unobserved heterogeneity; observed (time-

invariant) heterogeneity; and time-varying observed heterogeneity; and in each case focus on the case where the heterogeneity can be discretely categorised.

3.1 Unobserved heterogeneity

We begin by considering unobserved heterogeneity where we have no additional information on the observed capture histories but recognise that some individual animals may be more "catchable" than others. For example, the catchability of an individual may be related to the closeness of the individual's home range to the study area; or the position of the individual in the hierarchical social structure of the population which may influence their behaviour and/or movement patterns. We focus on the case where we assume that the overall population is heterogeneous, but is composed of an (unknown) number of homogeneous sub-populations. We consider the model $M_{h(k)}$ denoting the heterogeneity model with k homogeneous (unobserved) subpopulations.

3.1.1 Model $M_{h(k)}$

The model can be mathematically described as a finite mixture model on the individuals, where it is unknown which mixture, or sub-population, each individual belongs to - such models were initially proposed by Pledger (2000). We assume that there are a total of k mixture components, such that an individual belongs to component (or sub-population) j with probability $\pi(j)$, where $0 \leq \pi(j) \leq 1$ and $\sum_{j=1}^{k} \pi(j) = 1$. We set $\boldsymbol{\pi} = \{\pi(1), \ldots, \pi(k)\}$. In other words the different sub-populations correspond to a partition such that each individual in the population belongs to one and only one sub-population. Each individual in sub-population $j = 1, \ldots, k$ has a capture probability $p(j)$, assumed to be constant over time. The set of capture probabilities is given by $\boldsymbol{p} = \{p(1), \ldots, p(k)\}$. For this model, the sufficient statistics correspond to the Schnabel census, $\boldsymbol{f} = \{f_1, \ldots, f_T\}$, where f_t corresponds to the number of animals observed a total of t times within the study period.

In this case the likelihood of the multinomial formulation given in equation (2) simplifies to be of binomial form:

$$L(N, \boldsymbol{p}, \boldsymbol{\pi}; \boldsymbol{f}) \propto \frac{N!}{(N-D)!} \prod_{t=1}^{T} \Pr(\text{individual observed } t \text{ times})^{f_t} \times \Pr(\boldsymbol{x}_0)^{N-D}.$$

To calculate the probability that an individual is seen t times (for $t = 0, \ldots, T$) we use the basic laws of probability. In particular we sum over all possible sub-populations that an individual may belong to of the joint probability of belonging to the given sub-population and having the given capture probability. Mathematically,

we have that, for $t = 0, \ldots, T$,

$$\begin{aligned}
\Pr(\text{individual observed } t \text{ times}) &= \sum_{j=1}^{k} \Pr(\text{individual observed } t \text{ times and belongs to group } j) \\
&= \sum_{j=1}^{k} \Pr(\text{in group } j) \times \Pr(\text{observed } t \text{ times} \mid \text{in group } j) \\
&= \sum_{j=1}^{k} \pi(j) \times p(j)^t (1 - p(j))^{T-t}.
\end{aligned}$$

We note that this likelihood is simply a mixture of binomial components.

Although finite mixture models are structurally straightforward to construct there are some computational challenges with their use in practice. Commonly referred to as a label-switching problem, the finite mixture model can have a local identifiability issue if the parameter space of the proportion parameter is not restricted; see Kim and Lindsay (2015) for a full discussion of this issue. Further, there exist some challenges in the context of model selection. The issues (which are discussed in full within Pledger (2000)) arise because if the models being compared differ in the number of groups of animals, the parameters at the boundary under the null model mean that conditions of the standard likelihood ratio test are not met. However, a non-standard likelihood ratio test can be used - see Self and Liang (1987) for details. In practice Cubaynes *et al* (2012) showed that the AIC statistic (or other information criteria approaches), work reasonably well.

Alternative approaches for the modelling of unobserved heterogeneity exist. In particular, Dorazio and Royle (2003) proposed the use of an infinite mixture model, by considering a beta-binomial distribution, and Morgan and Ridout (2008) suggested combining the use of a finite and infinite mixture. Alternative infinite mixture models have been proposed, specifying the individual capture probabilities in the form of a logistic regression with an individual (normal) random effect component and using a numerical integration approach to evaluate the likelihood (Coull and Agresti, 1999) or a Bayesian data augmentation (sometimes referred to as a complete-data likelihood) approach (King and Brooks, 2008; King *et al*, 2016). For further discussion of the strengths and weaknesses of some of these approaches, see for example Pledger (2005) and the associated response.

3.1.2 Example: mice data

For the mice data the corresponding Schnabel census sufficient statistics for model $M_{h(k)}$ are given by:

t	1	2	3	4	5	6
f_t	9	6	7	6	6	4

We fit models of the form $M_{h(k)}$ (i.e. mixtures of k binomial components). In particular, we consider the cases $k = 2$ and $k = 3$. The estimates of the parameters, associated 95% non-parametric bootstrap confidence intervals and AIC statistics are provided in Table 4. The analyses indicate that the model with $k = 2$ is preferred

over the model where $k = 3$ using the AIC statistic. We note further that these are nested models (although the null model has parameters on the boundary) and considering only the fit of the model to the data in terms of the deviance (or equivalently the log-likelihood evaluated at the MLE of the parameters), there is very little difference between these two models. The difference in AIC statistics is thus dominated by the number of parameters associated with each model (4 parameters when $k = 2$ and 6 parameters when $k = 3$). Thus there is clear support for $k = 2$ and we focus our discussions on this model.

	Model $M_{h(2)}$		Model $M_{h(3)}$	
AIC	100.767		104.479	
Parameter	MLE	95% confidence interval	MLE	95% confidence interval
N	40.52	(38.00, 83.59)	41.91	(38.11, 149.62*)
$\pi(1)$	0.577	(0.351, 0.885)	0.467	(0.219, 0.840)
$\pi(2)$	-	-	0.387	(0.016, 0.675)
$p(1)$	0.290	(0.038, 0.475)	0.223	(0.015, 0.452)
$p(2)$	0.771	(0.609, 0.976)	0.629	(0.414, 0.875)
$p(3)$	-	-	0.890	(0.563, 1.000)

Table 4: Model estimates from fitting binomial mixture models, $M_{h(2)}$ (2 components) and $M_{h(3)}$ (3 components) to the mice data. *Note that the upper 2.5% quantile may not have completely converged due to the very long right-hand tail of this distribution.

For the case $k = 2$, there clearly appears to be a large difference in the detectability of the two different sub-populations assumed to be present in the population. The associated AIC statistic (100.767) is significantly smaller than for model M_0 fitted in Section 2.2.4, indicating some level of unobserved heterogeneity present within the population. The estimate for the total population size is 40.52, suggesting that not quite all members of the population were observed within the study (recall that a total of 38 individuals were observed). We note that there appears to be a very long right-hand tail for the estimate of the total population size. This is often observed in such models, particularly where the capture probability of one of the components (with reasonable mixture weight) is small - capture probabilities and population size (for the given sub-population) are inversely proportional to each other, so that variations in the small capture probabilities lead to generally large variability of the estimate of the total population size. (We note that this is particularly marked for the case $k = 3$ where we appear to be over-estimating the number of sub-populations). The expected values of $\boldsymbol{f}$ under the $k = 2$ model are $\{7.5, 8.0, 6.0, 6.1, 6.7, 3.7\}$. Applying Pearson's chi-squared goodness-of-fit test we obtain a test statistic of 1.07; and an associated p-value of 0.3, suggesting that the observed and expected values are not significantly different from one another. We note that sparsity of the available data means that it is often not straightforward to assess absolute goodness-of-fit due to the requirement for all expected values to be greater than 5, and even for this fairly straightforward heterogeneity model we have

an expected value of 3.7. Therefore for the more complex models we consider in the remainder of this chapter we do not consider absolute goodness-of-fit further. This issue of sparsity of data for goodness-of-fit assessment is well-documented in many areas of ecological modelling - see for example Amstrup *et al* (2005); Stanley and Burnham (1999).

For the heterogeneity studied here we assume that we do not observe the sub-populations. These sub-populations may relate to physical characteristics of the individuals such as gender, breeding status, age, etc. In many studies, such characteristics may also be recorded when individuals are observed. These may be subsequently used to create observable sub-populations to incorporate such heterogeneity.

3.2 Observed (time-invariant) heterogeneity

We now consider the case where additional information may be recorded on individuals observed within the study. For example, this may correspond to gender, age group or breeding status. We let $\mathcal{G} = \{1, \ldots, G\}$ denote the set of G mutually exclusive and exhaustive individual physical characteristics (or groups) within a population. Here we assume that the group membership of an individual is constant within the study period - we relax this assumption in Section 3.3. When individuals are captured they are identified as belonging to group $g \in G$ without error, and we assume that a total of $D(g)$ individuals in group g are observed within the study period. For notational convenience we let $\boldsymbol{D} = \{D(1), \ldots, D(g)\}$ denote the set of observed animal numbers corresponding to each group. As before, a total of $D = \sum_{g \in \mathcal{G}} D(g)$ unique individuals are observed within the study period. We wish to estimate the total sub-population sizes for each group denoted by $\boldsymbol{N} = \{N(1), \ldots, N(G)\}$ and associated overall total population size $N = \sum_{g \in \mathcal{G}} N(g)$. For the observed heterogeneity models, we use the notation $h(\mathcal{G})$ to denote individual heterogeneity attributed to the set of characteristics $\mathcal{G}$.

3.2.1 Model $M_{0(\mathcal{G})}$

For model $M_{0(\mathcal{G})}$ we include the additional covariate information, and condition on this when forming the likelihood, but assume that the population is homogeneous across the different sub-populations or groups, i.e. the capture probabilities are independent of group membership. The associated likelihood expression is given by:

$$L(\boldsymbol{N}, p; R, \boldsymbol{D}) \propto \left[\prod_{g \in \mathcal{G}} \frac{N(g)!}{(N(g) - D(g))!} \right] \times p^R (1-p)^{NT-R},$$

where we recall that R corresponds to the number of observed sightings within the whole study period (and so in this case across all groups); and $N = \sum_{g \in \mathcal{G}} N(g)$.

The expression is very similar to that given in equation (3) corresponding to model M_0 when there is no additional observed covariate information. However the initial combinatorial terms are different. For model M_0 there is no information relating to the groups and so all individuals can be simply reordered as the order is

unimportant in the likelihood calculation, leading to a single multinomial coefficient in the likelihood expression. However, for model $M_{0(\mathcal{G})}$ we also condition on the observed group that each individual belongs to so that individual animals are distinguishable between groups, though the order that they are numbered within the groups are exchangeable, thus leading to a multinomial coefficient term for each individual group within the likelihood expression. The two models, M_0 and $M_{0(\mathcal{G})}$ are using different levels of information - with M_0 taking an "unconditional" approach with regard to gender and model $M_{0(\mathcal{G})}$ conditioning on the additional covariate information within the likelihood. Consequently we cannot directly compare such models with regard to AIC statistics or likelihood ratio tests.

3.2.2 Model $M_{h(\mathcal{G})}$

We consider the model $M_{h(\mathcal{G})}$, where the only heterogeneity acting on the system corresponds to the observed group membership. For any given capture occasion, we let $p(g)$ denote the probability that an individual from group $g \in \mathcal{G}$ is observed, i.e. $p(g)$ is the capture probability of an individual in group $g \in \mathcal{G}$. For notational simplicity we let $\boldsymbol{p} = \{p(1), \ldots, p(G)\}$ denote the set of capture probabilities across all the different groups. Conditioning on the observed group membership of each individual we can extend equation (3) and express the likelihood of the data as a product of the likelihood contributions of the capture histories of the individuals associated with each group $g \in G$ and write:

$$L(\boldsymbol{N}, \boldsymbol{p}; \boldsymbol{R}, \boldsymbol{D}) \propto \prod_{g \in \mathcal{G}} \frac{N(g)!}{(N(g) - D(g))!} p(g)^{R(g)} (1 - p(g))^{N(g)T - R(g)}, \tag{5}$$

where $\boldsymbol{R} = \{R(1), \ldots, R(G)\}$, such that $R(g)$ corresponds to the number of captures of individuals in sub-population $g \in \mathcal{G}$ and can be expressed as

$$R(g) = \sum_{t=1}^{T} t f_t(g);$$

such that $f_t(g)$ denotes the number of individuals from group g captured t times within the study; and the parameters of primary interest, $N(g)$, correspond to the (unknown) population size of group $g \in \mathcal{G}$. For each individual group $f_0(g)$ denotes the number of individuals from group g that were not observed within the study, so that $f_0(g) = N(g) - D(g)$ and $\sum_{g \in \mathcal{G}} f_0(g) = N - D$. We note that in this approach we essentially stratify the individuals into the separate groups and analyse each group independently of each other. The estimates of $N(g)$ for each group can then be combined to obtain an overall estimate for N. Constructing the confidence intervals using a bootstrap approach means that it is trivial to obtain an associated bootstrap confidence interval for N, using the relationship that $N = \sum_{g \in \mathcal{G}} N(g)$ applied to the MLEs of $N(g)$ for each bootstrap replicate.

3.2.3 Model $M_{th(\mathcal{G})}$

Here we consider further both temporal and observed individual heterogeneity, denoted by the model notation $M_{th(\mathcal{G})}$. We extend the above group stratification, so

that the capture probabilities for each group are also time-dependent. The data can be most easily summarised in terms of the set of the total number of individuals observed in each group, $\boldsymbol{D} = \{D(1), \ldots, D(G)\}$, and the number of individuals from group g that are observed at time t, denoted $\boldsymbol{R} = \{R_t(g) : t = 1, \ldots, T; g \in \mathcal{G}\}$. The corresponding likelihood can be expressed in the form:

$$L(\boldsymbol{N}, \boldsymbol{p}; \boldsymbol{R}, \boldsymbol{D}) \propto \prod_{g \in \mathcal{G}} \frac{N(g)!}{(N(g) - D(g))!} \prod_{t=1}^{T} p_t(g)^{R_t(g)} (1 - p_t(g))^{N(g) - R_t(g)}, \qquad (6)$$

where $p_t(g)$ denotes the capture probability at time $t = 1, \ldots, T$ for group $g \in \mathcal{G}$ and $\boldsymbol{p} = \{p_t(g) : t = 1, \ldots, T; g \in \mathcal{G}\}$. We once again note the similarity with equation (4) allowing for the additional group membership.

Arbitrarily estimating $p_t(g)$ for each time and group once again essentially stratifies the data into the different groups and leads to independent analyses for each of the groups. However, this may lead to small sample sizes for each of the different groups with a relatively large numbers of parameters to be estimated - in this case there are $(T+1)G$ parameters to estimate (TG recapture probabilities and G population sizes) and consequently leading to potentially poor precision of the estimates. However, alternative capture probabilities can be specified that reduces the number of parameters to be estimated, and permits the borrowing of information across the different groups. For example, we may specify an additive logistic model for the capture probabilities, such that,

$$\text{logit } p_t(g) = \alpha(g) + \beta_t,$$

where $\alpha(g)$ denotes the effect associated with group $g \in \mathcal{G}$ and β_t the temporal effect for $t = 1, \ldots, T$. For identifiability we typically set one of the parameters to be equal to 0, for example, without loss of generality, we may set $\alpha(1) = 0$. The remaining $\alpha(g)$ terms, for $g = 2, \ldots, G$ are then interpreted as the effect of group g *relative* to group 1. This model assumes that the capture probabilities associated with each group display a similar temporal pattern over the capture occasions - but allowing for different levels of catchability for the different groups. The model also assumes that these differences in catchability across groups are additive on the logistic scale - taking the logistic link function is often primarily for mathematical convenience to ensure that the capture probabilities are constrained to the interval [0,1]. Alternative relationships can also be specified, such as $p_t(g) = \beta_t^{\alpha(g)}$. However, additional care needs to be taken in this case to ensure that $p_t(g) \in [0, 1]$ for all $g \in \mathcal{G}$ and $t = 1, \ldots, T$. Thus, for simplicity we retain the additive logistic models described above. For this model there are a total of $2G + T - 1$ parameters (G population sizes; T temporal effects and $G - 1$ group effects to estimate, since we specify the constraint $\alpha(1) = 0$ for identifiability); this is compared to $(T + 1)G$ for the case above with arbitrary group and time effects.

This logistic model for the capture probabilities no longer stratifies the data into the separate sub-populations and considers each stratified dataset independently. Instead the model uses information from all of the groups to estimate the temporal model parameters, borrowing information across each of the different groups, but with the given assumed structure across times and groups.

3.2.4 Model $M_{t\times h(\mathcal{G})}$

The final model that we consider is of a similar form to model $M_{th(\mathcal{G})}$ with the associated likelihood expression given in equation (6). However, here we consider an interaction term in the logistic model specified for the capture probabilities with regard to the group and time effects, expressed as model $M_{t\times h(\mathcal{G})}$. Thus for $g \in \mathcal{G}$ and $t = 1, \ldots, T$, we specify:

$$\text{logit } p_t(g) = \alpha(g) + \beta_t + \delta_t(g),$$

with the standard constraints for identifiability (for example $\alpha(1) = 0$ and $\delta_1(1) = 0$). We note that this model is equivalent to the previous model discussed above with arbitrary capture times $p_t(g)$ to be estimated - and hence again removes this borrowing of information across the different groups, and associated issues discussed above. The sufficient statistics are again the number of individuals from group $g \in \mathcal{G}$ that are observed at time $t = 1, \ldots, T$, and denoted by, $\boldsymbol{R} = \{R_t(g) : t = 1, \ldots, T; g \in \mathcal{G}\}$.

3.2.5 Example: mice

We return to the mice example, and now consider the additional information of gender within our analyses. There are naturally two groups corresponding to females and males. Thus we set $\mathcal{G} = \{f, m\}$ with obvious notation (so that in the previous group labelling $1 \equiv f$; and $2 \equiv m$). We consider four different models:

- Model $M_{0(\mathcal{G})}$ - capture probabilities are constant over all times and groups.
- Model $M_{h(\mathcal{G})}$ - capture probabilities are different for each group but constant over time.
- Model $M_{th(\mathcal{G})}$ - capture probabilities logistically regressed with additive group and time effects.
- Model $M_{t\times h(\mathcal{G})}$ - capture probabilities logistically regressed with additive group and time effects with interactions - equivalent to arbitrary capture probabilities for each group and time.

For model $M_{h(\mathcal{G})}$ the sufficient statistics are given by $R(f) = 41$; $R(g) = 79$; $R = 120$; $D(f) = 17$; and $D(m) = 21$. For models $M_{th(\mathcal{G})}$ and $M_{t\times h(\mathcal{G})}$ the sufficient statistics are given by,

t	1	2	3	4	5	6
$R_t(f)$	3	5	8	7	9	9
$R_t(m)$	12	15	8	12	16	16

We note that there does appear to be some variability over capture occasions with regard to the number of females and males observed - with a general mild positive correlation (correlation coefficient = 0.19) between the number of observed captures for females and males over capture occasions. However the correlation is difficult to assess with the relatively few data points being considered.

The associated estimates of the population sizes and AIC statistics are provided in Table 5. In the calculation of the non-parametric bootstrap confidence intervals we apply a stratified bootstrap to preserve the proportion of male and female histories within each bootstrap replication (i.e. for these data we sample with replacement 17 female capture histories from the set of all observed female histories; and 21 male capture histories from the set of all observed male histories). The model identified as optimal with regard to the AIC statistic is model $M_{th(\mathcal{G})}$ with additive group and time effects - although we note that $\Delta AIC = 0.31$ for the nested time-independent model $M_{h(\mathcal{G})}$. The estimates of the capture probabilities are given in Table 6 for $M_{th(\mathcal{G})}$. We note that the gender difference ("male-female") on the logit scale is estimated to be 0.984, with 95% non-parametric bootstrap confidence interval (0.299, 1.814), signifying a difference between the capture probabilities for males and females, such that males are more likely to be observed than females. Similarly, for the second ranked model, $M_{h(\mathcal{G})}$, with no temporal heterogeneity, the corresponding MLE (and 95% confidence interval) for the capture probability of females is 0.394 (0.261, 0.500) and for males is 0.627 (0.500, 0.746). This again suggests a substantial gender difference, assuming that there is no temporal heterogeneity.

	$N(f)$		$N(m)$		$N^{\dagger}$		
Model	MLE	95% CI	MLE	95% CI	MLE	95% CI	AIC
$M_{0(\mathcal{G})}$	17.00	*	21.00	*	38.00	*	163.67
$M_{h(\mathcal{G})}$	17.34	(17.00, 19.79)	21.00	*	38.34	(38.00, 40.79)	154.03
$M_{th(\mathcal{G})}$	17.23	(17.00, 19.04)	21.00	(21.00, 21.01)	38.34	(38.00, 40.04)	153.72
$M_{t\times h(\mathcal{G})}$	17.00	(17.00, 17.21)	21.00	*	38.00	(38.00, 38.21)	156.74

Table 5: The MLEs and associated 95% non-parametric bootstrap confidence intervals (CIs) for population sizes and associated AIC statistic for the four fitted models.
* denotes that the confidence interval was essentially ill-defined with the parameter lying on the boundary of the parameter space with the width of the 95% confidence interval less than 0.01.
†Note that N is a derived parameter from $N(f)$ and $N(m)$ calculated by the simple relationship $N = N(f) + N(m)$.

If group information is missing for some individuals an EM algorithm (Little and Rubin, 2002) or Bayesian data augmentation approach (Tanner and Wong, 1987) can be used to impute the group membership within the computational algorithms. Alternatively, a mixture modelling approach can be implemented, combining the theory from Sections 3.1 and 3.2 - see McCrea *et al* (2013) for an application of this procedure in the related field of ring-recovery modelling for individuals with missing age-class information.

3.3 Time-varying observed individual heterogeneity

So far we have only focused on "fixed" forms of individual heterogeneities - in other words observed or unobserved characteristics of the individuals that are static over

Parameter	MLE	95% CI	Parameter	MLE	95% CI
$p_1(f)$	0.268	(0.151, 0.407)	$p_1(m)$	0.494	(0.305, 0.683)
$p_2(f)$	0.391	(0.240, 0.549)	$p_2(m)$	0.632	(0.447, 0.798)
$p_3(f)$	0.291	(0.147, 0.457)	$p_3(m)$	0.523	(0.341, 0.711)
$p_4(f)$	0.363	(0.210, 0.527)	$p_4(m)$	0.605	(0.405, 0.787)
$p_5(f)$	0.533	(0.330, 0.729)	$p_5(m)$	0.753	(0.605, 0.881)
$p_6(f)$	0.533	(0.330, 0.729)	$p_6(m)$	0.753	(0.605, 0.881)

Table 6: The derived MLEs and 95% non-parametric bootstrap confidence intervals (CIs) for the capture probabilities at each capture occasion for females and males for model $M_{th(\mathcal{G})}$.

the period of the study. For example, characteristics such as gender or species which are invariant over all time; or other factors that do not vary over the study period itself, although are not invariant over longer periods of time, such as age or breeding status. However, other individual level factors leading to heterogeneity in the capture probabilities may vary over the time of the study period itself, such as location or group size. For simplicity, we refer to these as covariates. If an individual is observed, the associated time-varying covariate is generally observed (though this may not always be the case); however, if an individual is not observed the corresponding time-varying covariate cannot be recorded and is thus "missing". An exception to this is where the covariate can be inferred without error from the observed data - this includes, for example, where the covariate corresponds to age or for describing behavioural responses (i.e. trap happy and trap shy) as an observed individual time-varying covariate (we consider this as a special case in Section 3.3.2). We initially consider this latter, simpler, case before the more general case where we account for the missing covariate values explicitly within the likelihood expression.

3.3.1 Deterministic covariate values

Here we consider the case where the covariates vary deterministically over time. For closed populations (typically over a short study period) this situation is not very common, except for the special case described in Section 3.3.2 corresponding to behavioural effects. However, we describe the situation in this general case for completeness and for motivation for Section 3.3.4 where we extend the idea to stochastic time-varying covariate values. For deterministic covariate values, if the covariate is observed at the initial capture of an individual, (and possibly conditional on the given observed capture history), the corresponding covariate values are known for all capture occasions for that individual. We let the set of possible covariate values be denoted by $\mathcal{Z} = \{1, \ldots, Z\}$. Further, we let, z_{it} denote the known covariate value for individual $i = 1, \ldots, D$ at capture occasion $t = 1, \ldots, T$; and set $\boldsymbol{z} = \{z_{it} : i = 1, \ldots, D; t = 1, \ldots, T\}$.

We again specify the capture probabilities to be a function of the covariate values. In particular, assuming no additional heterogeneities, the capture probability for

observed individual $i = 1, \ldots, D$ at time $t = 1, \ldots, T$ is given by

$$\text{logit } p_{it} = \alpha(z_{it}).$$

The covariate values for the unobserved individuals are however, unknown. Thus, in this general case we specify a distribution for the initial covariate value for all individuals at time $t = 1$. We note that since the covariate values are deterministic, given the observed capture history, the choice of time is arbitrary - we could have chosen the final state distribution for example, at time $t = T$, but this, in general, is less intuitive. For observed individuals, $i = 1, \ldots, D$, we specify,

$$z_{i1} \sim Multinomial(1, \boldsymbol{\pi}),$$

where $\boldsymbol{\pi} = \{\pi(1), \ldots, \pi(Z)\}$ such that $\pi(z)$ denotes the probability that an individual is initially in state $z \in \mathcal{Z}$ at time $t = 1$.

The likelihood expression given in equation (2) can be extended to incorporate these deterministic covariate values to give

$$L(N, \boldsymbol{p}, \boldsymbol{\pi}; \boldsymbol{x}, \boldsymbol{z}) \propto \frac{N!}{(N-D)!} \prod_{i=1}^{D} \pi(z_{i1}) \prod_{t=1}^{T} p_{it}^{x_{it}} (1 - p_{it})^{1-x_{it}} \times (Pr(\boldsymbol{x}_0))^{N-D}.$$

The likelihood function is a product of terms corresponding to the multinomial coefficient; initial state distribution and capture history probabilities associated with observed individuals; and finally the contribution associated with unobserved individuals. In particular, assuming that the unobserved individuals have the same initial state distribution as the observed individuals, we have that,

$$\begin{aligned} \Pr(\boldsymbol{x}_0) &= \sum_{z \in \mathcal{Z}} \Pr(\text{initial state} = z) \Pr(\text{unobserved in study} \mid \text{initial state} = z) \\ &= \sum_{z \in \mathcal{Z}} \pi(z) \prod_{t=1}^{T} (1 - p_t(z)), \end{aligned}$$

where $p_t(z)$ corresponds to the capture probability at time t, given that the initial covariate value of an individual at time $t = 1$ is equal to z. Suppose that we let $h(z, t)$ denote the covariate value at time t, given that the initial covariate value at time $t = 1$ is equal to z. Then, we have that,

$$\text{logit } p_t(z) = \alpha(h(z, t)).$$

Recall that the covariate values are deterministic, conditional on the capture histories, so for any given initial covariate value at time $t = 1$, all other covariate values as times $t = 2, \ldots, T$ are a deterministic function. We note that the above likelihood can be extended to the case where we do not necessarily observe the covariate values for all observed individuals. In this case, for those individuals, we can apply a similar argument to that of the unobserved individuals and sum over all possible initial covariate values within the associated likelihood expression.

As discussed previously, recording deterministic time-varying covariate values that may influence the capture probabilities is not particularly common for capture-recapture studies over a short period of time where it is reasonable to assume that the

population is closed (i.e. there are no births/deaths/migrations within the study period). The most common examples of such covariates are often age- or stage-related, that simply increases deterministically over time. One particular example where the covariate is time-varying and known for all individuals and capture occasions (given the observed capture histories) is with regard to behavioural effects.

3.3.2 Behavioural heterogeneity

Behavioural heterogeneity permits an individual's capture probability to change following its initial capture - either increasing (a trap happy response) or decreasing (a trap shy response) the capture probability at subsequent capture occasions. The response is typically assumed to be a permanent response effect, but can be easily extended to allow for the memory of the capture to decrease over time. In the simplest model, M_b, there are just two capture probabilities: an initial capture probability; and a recapture probability (for individuals who have been observed at least once in the study). In this model only the initial capture has any effect on the associated capture probability of the individual - multiple recaptures do not lead to any further modification.

For this model, we set the individual time-varying covariate values to be equal to $z_{it} = 0$ up to and including the initial capture of the individual; and $z_{it} = 1$ for all times after initial capture for individual $i = 1, \ldots, N$. Thus these values are known for *all* individuals both those observed within the study and those unobserved - for those unobserved the covariates values are equal to 0 for all capture times. The capture probability of individual i at time t is specified to be of the form,

$$\text{logit } p_{it} = \begin{cases} \text{logit } p = \beta & \text{for } z_{it} = 0; \\ \text{logit } c = \beta + \gamma & \text{for } z_{it} = 1. \end{cases}$$

The term p correspond to the initial capture probability; and c to the recapture probability. The model is parameterised in the above form for interpretability of the difference between initial and subsequent capture probabilities (on the logit scale); and for consistency with the extension to include additional forms of heterogeneity, such as temporal heterogeneity which we discuss below. The parameter γ directly corresponds to the behavioural response effect: if $\gamma > 0$ there is a trap happy response; if $\gamma < 0$ there is a trap shy response. The corresponding likelihood simplifies to,

$$L(N, p, c|R, U, D) \propto \frac{N!}{(N-D)!} p^D (1-p)^{U+(N-D)T} c^{R-D} (1-c)^{DT-R-U},$$

where,

$$R = \sum_{i=1}^{D} \sum_{t=1}^{T} x_{it}; \qquad \text{and} \qquad U = \sum_{i=1}^{D} \sum_{t=1}^{T} (1 - z_{it}) - D.$$

We note that R corresponds to the total number of captures within the study period; thus $R - D$ corresponds to the total number of recaptures (excluding the initial capture of each observed individual). Similarly U corresponds to the total number of non-captures of individuals that occur prior to the individuals being initially

observed (unobserved individuals are not included in this calculation since they are not initially observed within the study); $(N-D)T$ the number of non-captures relating to unobserved individuals; and $DT-R-U$ the total number of non-captures following the initial capture of the individuals (equal to the total number of possible captures for individuals observed within the study minus the number of captures minus the number of non-captures prior to initial capture). We note that for this model, the likelihood can be partitioned into two distinct components relating to (i) up to and including initial capture (if any); and (ii) after initial capture. It is only the initial captures (and associated initial capture probability, p) that affects the estimate of the total population size ($N-D$ and c are independent; similarly p and c are independent of each other). For this simplest model, M_b, the statistics D, R and U are the minimal sufficient statistics.

As discussed previously, there can be multiple heterogeneities present within the same system. For example, consider the model M_{tb}, with both temporal and behavioural effects present. We note that there is no unique model specification for M_{tb}, with several different parameterisations possible. We consider an additive logistic model here, such that the (logit of the) capture probability of an individual at time t is specified to be of the form,

$$\text{logit } p_{it} = \begin{cases} \text{logit } p_t = \beta_t & \text{for } z_{it} = 0; \\ \text{logit } c_t = \beta_t + \gamma & \text{for } z_{it} = 1. \end{cases}$$

The term p_t corresponds to the initial capture probability at time t; and c_t the recapture probability of an individual at time t, given that they have already been observed within the study period. The corresponding sufficient statistics for the model are given by:

$$R_t = \sum_{i=1}^{D} x_{it}; \quad A_t = \sum_{i=1}^{D} x_{it}(1-z_{it}); \qquad \text{and} \qquad U_t = \sum_{i=1}^{D}(1-z_{it}) - A_t.$$

for $t = 1, \ldots, T$. The statistics can be easily interpreted as follows: R_t corresponds to the total number of captures at time t; A_t the number of initial captures at time t (by definition $A_1 = R_1$ and $\sum_{t=1}^{T} A_t = D$); and U_t the number of individuals unobserved at time t that have not been observed at any previous occasion. For notational convenience we set $\boldsymbol{R} = \{R_1, \ldots, R_T\}$, $\boldsymbol{A} = \{A_1, \ldots, A_T\}$ and $\boldsymbol{U} = \{U_1, \ldots, U_T\}$. The corresponding likelihood follows similarly to the above for M_b, but allowing for the additional temporal dependence. In particular we obtain:

$$L(N, \boldsymbol{p}|\boldsymbol{R}, \boldsymbol{U}, \boldsymbol{A}, D) \propto \frac{N!}{(N-D)!} \prod_{t=1}^{T} p_t^{A_t}(1-p_t)^{U_t+(N-D)} c_t^{R_t-A_t}(1-c_t)^{D-R_t-U_t}. \quad (7)$$

Similarly additional individual heterogeneity can be incorporated into such models - such as unobserved or observed individual heterogeneity. We see an example of this for the mice data, where we consider the observed covariate of gender.

3.3.3 Example: mice

We return to the mice data and consider models with additional behavioural effects. Given the previous importance of including the observed (static) covariate of gender,

we again condition on the gender within all of the models that we consider. We consider a number of competing models corresponding to the additional inclusion of behavioural and/or temporal effects present in the model.

Model $M_{bh(0)}$

In this model we assume that there is a behavioural effect, but no gender effects on the capture probabilities, denoted $M_{bh(0)}$. The sufficient statistics for this model are given by $D(f) = 17$ (number of females observed); $D(m) = 21$ (number of males observed); $U = 56$ and $R = 120$. The corresponding likelihood expression is given by,

$$L(\boldsymbol{N}, p, c|R, U, \boldsymbol{D}) \propto \prod_{g \in \mathcal{G}} \frac{N(g)!}{(N(g) - D(g))!} p^D (1-p)^{U+(N-D)T} c^{R-D} (1-c)^{DT-R-U},$$

where $D = D(f) + D(m)$ (and corresponds to the total number of individuals observed). We note that the only difference between this expression and that given in Section 3.3.2 (with no gender effects) corresponds to the multinomial coefficient terms for the different groups.

Model $M_{bh(\mathcal{G})}$

Here there is both a behavioural effect, specified to be additive on the logistic scale, and a gender effect for the capture probabilities. Thus we assume that the behavioural effect is independent of gender. In particular we specify the capture probability for an individual of gender $g \in \mathcal{G}$ at time t to be of the form:

$$\text{logit } p_{it}(g) = \begin{cases} \text{logit } p(g) = \alpha(g) & \text{for } z_{it} = 0; \\ \text{logit } c(g) = \alpha(g) + \gamma & \text{for } z_{it} = 1. \end{cases}$$

As usual for notational convenience we set $\boldsymbol{p} = \{p_t(g) : g \in \mathcal{G}\}$ and $\boldsymbol{c} = \{c_t(g) : g \in \mathcal{G}\}$. The corresponding sufficient statistics are given by $\boldsymbol{D} = \{D(g) : g \in \mathcal{G}\}$; $\boldsymbol{U} = \{U(g) : g \in \mathcal{G}\}$; and $\boldsymbol{R} = \{R(g) : g \in \mathcal{G}\}$, with obvious notation (conditioning on the gender of the individual animals). For the mice data, $U(f) = 31$; $U(m) = 25$; $D(f) = 17$; $D(m) = 21$; $R(f) = 41$; and $R(m) = 79$. The corresponding likelihood is given by,

$$\begin{aligned} L(\boldsymbol{N}, \boldsymbol{p}, \boldsymbol{c}|\boldsymbol{R}, \boldsymbol{U}, \boldsymbol{D}) \quad \propto \quad & \prod_{g \in \mathcal{G}} \frac{N(g)!}{(N(g) - D(g))!} p(g)^{D(g)} (1-p(g))^{U(g)+(N(g)-D(g))T} \\ & \times c(g)^{R(g)-D(g)} (1-c(g))^{D(g)T-R(g)-U(g)}. \end{aligned}$$

This model assumes that the behavioural response, denoted by the additive γ term is the same for all observed covariate values (in this case gender). This may not be the case, for example, if the different observed sub-populations respond differently to the capture process. In this case we can consider an interaction between the behavioural effect and observed covariate.

Model $M_{b\times h(\mathcal{G})}$

Within this model we assume that the behavioural response is gender specific. The capture probabilities are specified similarly to model $M_{bh(\mathcal{G})}$, but where the behavioural response is now a function of the observed covariate:

$$\text{logit } p_{it}(g) = \begin{cases} \text{logit } p(g) = \alpha(g) & \text{for } z_{it} = 0; \\ \text{logit } c(g) = \alpha(g) + \gamma(g) & \text{for } z_{it} = 1. \end{cases}$$

The sufficient statistics and associated likelihood expression are of the same form as for $M_{bh(\mathcal{G})}$ - it is simply the specification of the capture probabilities that differ (and the estimation of an additional parameter). We note that in this case (as for the model $M_{t\times h(\mathcal{G})}$) the likelihood once again decomposes into separate, and independent, components for the distinct groups $g \in \mathcal{G}$.

Model $M_{tbh(\mathcal{G})}$

Here we consider temporal, behavioural and observed heterogeneity via an additive logistic model. The capture probabilities are specified in the form,

$$\text{logit } p_{it}(g) = \begin{cases} \text{logit } p(g) = \alpha(g) + \beta_t & \text{for } z_{it} = 0; \\ \text{logit } c(g) = \alpha(g) + \beta_t + \gamma(g) & \text{for } z_{it} = 1, \end{cases}$$

and we set $\alpha(f) = 0$ for identifiability. The corresponding likelihood function is an extension of equation (7) following the previous idea of extending to include group membership by producting over groups and specifying the parameters (capture probabilities and total population sizes) and sufficient statistics to be conditional on the observed groups. The sufficient statistics are given by $\boldsymbol{R} = \{R_t(g) : t = 1, \ldots, T; g \in \mathcal{G}\}$; $\boldsymbol{A} = \{A_t(g) : t = 1, \ldots, T; g \in \mathcal{G}\}$; $\boldsymbol{U} = \{U_t(g) : t = 1, \ldots, T; g \in \mathcal{G}\}$; and $\boldsymbol{D} = \{D(g) : g \in \mathcal{G}\}$.

Results

The four models described above are fitted to the data, and the associated results (in terms of MLEs, 95% non-parametric confidence intervals and AIC statistics) are presented in Table 7. The model deemed optimal has both behavioural and individual heterogeneity effects in terms of observed group, i.e. model $M_{bh(\mathcal{G})}$, with an associated AIC statistic of 145.38. Recall that the model previously identified as optimal via the AIC statistic without behavioural effects was model $M_{th(\mathcal{G})}$ with an associated AIC statistic of 153.72 (see Table 5). Thus, this suggests that the behavioural effects are more important than arbitrary temporal effects. Further, we note that including the behavioural effects consistently leads to a significantly improved fit of the model to the data, irrespective of the other effects that are present in the model. The behavioural response for this model (and for all models fitted with a behavioural effect present) is positive, corresponding to a "trap happy" response. In other words an individual who is observed at a given capture time is more likely to be observed at future capture occasions. Finally, we note that the estimates of the total population size are also generally dependent on the model being fitted - with greater variability with regard to the female sub-population.

	$N(f)$		$N(m)$		$N^{\dagger}$		
Model	MLE	95% CI	MLE	95% CI	MLE	95% CI	AIC
$M_{b(\mathcal{G})}$	17.61	(17.00, 22.03)	21.88	(21.00, 27.33)	39.49	(38.00, 49.36)	155.53
$M_{bh(\mathcal{G})}$	21.25	(17.51, 34.62)	21.00	*	42.25	(38.51, 55.62)	145.38
$M_{b\times h(\mathcal{G})}$	19.59	(17.00, 48.60)	21.00	(21.00, 21.01)	38.00	(38.00, 69.60)	147.17
$M_{tbh(\mathcal{G})}$	20.95	(17.04, 33.45)	21.00	(21.00, 21.13)	41.95	(38.04, 54.54)	151.00

Table 7: The MLE and associated 95% non-parametric bootstrap confidence intervals (CIs) for population sizes and associated AIC statistic for the four fitted models.
* denotes that the confidence interval was essentially ill-defined with the parameter lying on the boundary of the parameter space with the width of the 95% confidence interval less than 0.01.
†Note that N is a derived parameter from $N(f)$ and $N(m)$ calculated by the simple relationship $N = N(f) + N(m)$.

3.3.4 Stochastic covariate values

In practice many individual covariates may change over time in a stochastic manner, such as their breeding status, hunger levels, health, location, etc. The temporal scale acting on the changes is dependent on the actual covariate - for example breeding status (breeding/not breeding) may change at an annual level; whereas hunger levels may vary over hours or days, dependent on the species. The capture probability of an individual may be highly dependent on the covariate - animals that are hungry are more likely to forage for food and hence may be more likely to be seen during a given capture occasion, compared to an individual who is satiated and hence stays within some home location. Alternatively, the location of an animal may also influence the probability that they are seen - with individuals more likely to be observed in open grassland, compared to dense forest. Here we consider such covariates where changes in their values may occur in individuals at a finer scale than the study period.

Notationally, we let $\mathcal{Z} = \{1, \ldots, Z\}$ denote the set of possible discrete covariate values (or states). Recall that we let z_{it} denote the covariate value for individual $i = 1, \ldots, D$ at time $t = 1, \ldots, T$. For simplicity, we initially assume that if an individual is observed, the corresponding covariate value is observed without error. However, if an individual is not observed, the corresponding covariate value is also unknown. We note that the model can be extended to allow for misclassification of states, and such a model is referred to as the multievent model (Pradel, 2005), and a related model exists for partially-observed data (King and McCrea, 2014).

We define the set of observed covariate values, denoted, $\boldsymbol{z}^{obs} = \{z_{it} : z_{it} \text{ is known}\}$. Further we set:

$$\mathcal{Z}^*_{it} = \begin{cases} z_{it} & \text{for } z_{it} \text{ known;} \\ \mathcal{Z} & \text{for } z_{it} \text{ unknown.} \end{cases}$$

Thus $\mathcal{Z}^*_{it}$ corresponds to the set of values that the value z_{it} may be, given the observed data. If the covariate value is observed then as we assume that these values are correctly recorded so that the state of the individual is known without error; else if the covariate is unknown then it may take any possible covariate value,

i.e. any value in $\mathcal{Z}$. The set of possible covariate values over the capture occasions for individual i, given the observed data is given by $\boldsymbol{\mathcal{Z}}_i^* = \{\mathcal{Z}_{it}^* : t = 1, \ldots, T\}$.

We specify the capture probabilities to be a function of the covariate values, and other possible effects. For example we may specify the capture probabilities to be of the following forms:

- Model M_h: capture probabilities of the form logit $p_t(z_{it}) = \alpha(z_{it})$ - dependence only on state;
- Model M_{th}: capture probabilities of the form logit $p_t(z_{it}) = \alpha(z_{it}) + \beta_t$ - dependence on state and time assumed to be additive on logistic scale (for identifiability we set $\alpha(1) = 0$);
- Model $M_{t\times h}$: capture probabilities of the form logit $p_t(z_{it}) = \alpha(z_{it}) + \beta_t + \delta_t(z_{it})$ - dependence on state and time with an interaction (i.e. arbitrary time and state dependence) (for identifiability we set $\alpha(1) = 0$ and $\delta_1(1) = 0$).

The above models, and associated capture probabilities, can be extended, for example, to allow for group effects (observed deterministic effects) or behavioural effects in a similar way.

We also need to specify some form of model for the change in state over time. We assume a first-order Markov model, so that the state at time t depends only on the state at time $t-1$ (and no previous states). We define the $Z \times Z$ state transition probabilities at time t to be,

$$\begin{aligned} \psi_t(z, z^\dagger) &= \Pr(\text{individual in state } z^\dagger \text{ at time } t+1 \mid \text{in state } z \text{ at time } t) \\ &= \Pr(z_{it+1} = z^\dagger | z_{it} = z). \end{aligned}$$

Notationally, we let the set of all transition probabilities be denoted by $\boldsymbol{\psi} = \{\psi_t(z, z^\dagger) : t = 1, \ldots, T-1; z, z^\dagger \in \mathcal{Z}\}$. Finally, we define the parameters corresponding to the initial state distribution, and let $\pi(z)$ denote the probability that an individual is in state z at time $t = 1$. We note that $\sum_{z\in\mathcal{Z}} \pi(z) = 1$ (an individual must belong to one of the possible states at time $t = 1$). The set of initial state probabilities is given by $\boldsymbol{\pi} = \{\pi(1), \ldots, \pi(Z)\}$.

The corresponding likelihood of the models can be constructed, summing over all possible unknown states, and taking into account the transition probabilities between the possible states:

$$\begin{aligned} L(N, \boldsymbol{p}, \boldsymbol{\pi}, \boldsymbol{\psi}; \boldsymbol{x}, \boldsymbol{z}^{obs}) \propto \frac{N!}{(N-D)!} \prod_{i=1}^{D} \sum_{\boldsymbol{z}_i \in \boldsymbol{\mathcal{Z}}_i^*} \Big(& \pi(z_{i1}) p_1(z_{i1})^{x_{i1}} (1 - p_1(z_{i1}))^{1-x_{i1}} \\ & \times \prod_{t=1}^{T-1} \psi_t(z_{it}, z_{it+1}) p_{t+1}(z_{it+1})^{x_{it+1}} (1 - p_{t+1}(z_{it+1}))^{1-x_{it+1}} \Big) \\ & \times \Pr(\boldsymbol{x}_0)^{N-D}, \end{aligned}$$

where,

$$\Pr(\boldsymbol{x}_0) = \sum_{z_1 \in \mathcal{Z}} \cdots \sum_{z_T \in \mathcal{Z}} \pi(z_1)(1 - p_1(z_1)) \prod_{t=1}^{T-1} \psi_t(z_t, z_{t+1})(1 - p_{t+1}(z_{t+1})).$$

Summing over all the possible missing covariate values will often be computationally expensive, or even infeasible, due to the number of possible combinations of unobserved values. For example, for the probability of not being observed, $\Pr(\boldsymbol{x}_0)$, the summation is over Z^T possible values - this is the most "expensive" capture history in terms of the calculation, but the number of unique capture histories will also typically be "large" for which the associated probability will need to be calculated.

The (first-order) Markovian assumption for the state transitions, however, allows us to significantly simplify the likelihood expression, so that it can be written in the form of an efficient hidden Markov model (see Zucchini *et al* (2016) for an introduction to general hidden Markov models). Due to the first order Markov structure we can consider "pairs" of state values at times t and $t+1$. Intuitively, this is due to the memoryless property of the Markov process - what happens after time t only depends on the state of an individual at time t (and not anything that has happened before). Thus we do not need to consider how an individual arrives at its state at time t (i.e. its previous covariate values or capture history) but simply what its state is at time t.

The subsequent likelihood is most easily expressed in matrix notation. We describe the three different components corresponding to (i) the initial covariate (or state) distribution; (ii) observation process; and (iii) transition process between the different possible discrete covariate values (or states). We consider each in turn.

(i) Initial state distribution: We consider two different cases, corresponding to whether or not the initial state of the individual is known or not i.e. for individual i whether the covariate value is known, so that $\mathcal{Z}_{i1}^* = z_{i1} \in \mathcal{Z}$; or unknown so that $\mathcal{Z}_{i1}^* = \mathcal{Z}$. We define the initial state distribution row vector:

$$\boldsymbol{\pi}(\mathcal{Z}_{i1}^*) = \begin{cases} (\pi(1)I(z_{i1}=1), \ldots, \pi(Z)I(z_{i1}=Z)) & \text{for } \mathcal{Z}_{i1}^* = z_{i1} \text{ (i.e. } z_{it} \text{ is known);} \\ (\pi(1), \ldots, \pi(Z)) & \text{for } \mathcal{Z}_{i1}^* = \mathcal{Z} \text{ (i.e. } z_{it} \text{ is unknown),} \end{cases}$$

where $I(\cdot)$ denotes the indicator function.

(ii) Observation process: We define the observation matrix for individual $i = 1, \ldots, D$, at time $t = 1, \ldots, T$ to be of the form,

$$\boldsymbol{P}_t(\mathcal{Z}_{it}^*) = \begin{cases} \text{diag}(p_t(1)I(z_{it}=1), \ldots, p_t(Z)I(z_{it}=Z)) & \text{for } \mathcal{Z}_{it}^* = z_{it} \in \mathcal{Z}; \\ \text{diag}(1-p_t(1), \ldots, 1-p_t(Z)) & \text{for } \mathcal{Z}_{it}^* = \mathcal{Z}. \end{cases}$$

As for the initial state distribution, the observation matrix depends on whether the individual is observed or not at time t. We assume that if an individual is observed then the associated covariate value is known without error, so that if an individual is observed the contribution to the likelihood is simply $p_t(z_{it})$; if an individual is not observed, we need to consider all possible states that an individual may be in (i.e. we will need to sum over all possible states), with associated contribution of the probability of not being observed (conditional on the covariate value).

(iii) Transition process: Finally the state transition matrix is specified as follows, dependent on whether or not the covariate values are observed or not at time t and

$t+1$ for individual $i = 1, \ldots, D$. We initially consider the case where the covariate value is unknown at times t and $t+1$, so that $\mathcal{Z}^*_{it} = \mathcal{Z}^*_{it+1} = \mathcal{Z}$, then:

$$\boldsymbol{\Psi}_t(\mathcal{Z}^*_{it}, \mathcal{Z}^*_{it+1}) = \begin{pmatrix} \psi_t(1,1) & \ldots & \psi_t(1,Z) \\ \vdots & & \vdots \\ \psi_t(Z,1) & \ldots & \psi_t(Z,Z) \end{pmatrix}.$$

Now if the covariate value is known at time t, so that z_{it} is known ($\mathcal{Z}^*_{it} = z_{it}$), but the covariate value is unknown at time $t+1$ ($\mathcal{Z}^*_{it+1} = \mathcal{Z}$), the corresponding transition matrix is the same as $\boldsymbol{\Psi}_t(\mathcal{Z}^*_{it}, \mathcal{Z}^*_{it+1})$ but with 0 replacing all the values in the rows except for the z_{it}th row. Similarly if the covariate value at time $t+1$ is known, ($\mathcal{Z}^*_{it+1} = z_{it+1}$) but the covariate value at time t is unknown ($\mathcal{Z}^*_{it} = \mathcal{Z}$), the transition matrix is the same $\boldsymbol{\Psi}_t(\mathcal{Z}^*_{it}, \mathcal{Z}^*_{it+1})$ but with the 0 replacing all values except for the z_{it+1}th column. Finally if both z_{it} and z_{it+1} are both known, the matrix reduces such that all values are set equal to 0 except $\psi_t(z_{it}, z_{it+1})$.

The corresponding likelihood can be expressed in the form:

$$L(N, \boldsymbol{p}, \boldsymbol{\pi}, \boldsymbol{\psi}; \boldsymbol{x}, \boldsymbol{z}^{obs}) \propto \frac{N!}{(N-D)!} \prod_{i=1}^{D} \left(\boldsymbol{\pi}(\mathcal{Z}^*_{i1}) \mathbf{P}_1(\mathcal{Z}^*_{i1}) \prod_{t=1}^{T-1} \left[\boldsymbol{\Psi}_t(\mathcal{Z}^*_{it}, \mathcal{Z}^*_{it+1}) \boldsymbol{P}_{t+1}(\mathcal{Z}^*_{it+1}) \right] \mathbf{1}_Z^T \right) \times \Pr(\boldsymbol{x}_0)^{N-D},$$

where $\mathbf{1}_Z^T$ corresponds to the column vector with all elements equal to unity; and

$$\Pr(\boldsymbol{x}_0) = \pi(\mathcal{Z}) \mathbf{P}_1(\mathcal{Z}) \prod_{t=1}^{T-1} \boldsymbol{\Psi}_t(\mathcal{Z}, \mathcal{Z}) \boldsymbol{P}_{t+1}(\mathcal{Z}) \mathbf{1}_Z^T.$$

The likelihood expression is of the same form as a hidden Markov model, but allows for known observed covariate values (for a hidden Markov model typically no true states are observed). A similar approach is taken by Langrock and King (2013) for capture-recapture data for an open population with continuous time-varying individual covariates using a discrete approximation (and allowing for covariate values being unknown when an individual is observed); and King and Langrock (2016), permitting semi-Markov transition processes for multi-state capture-recapture data (and where covariate values may be observed with error), removing the first-order Markov assumption. Worthington *et al* (2018b) describe the associated summary statistics for these data, and an efficient likelihood specification, using an alternative recursive formulation to the matrix approach.

4 Open populations

We have, so far, considered closed populations where it is assumed that the population does not change over the duration of the study period, i.e. there are no births/deaths or emigration/immigration acting on the system. In other words all individuals in the study are available for capture at every capture occasion within the study. In order to satisfy this assumption, the study typically covers only a

short duration of time to minimise the chances of this assumption being violated - for example for the mice data studied throughout, the capture-recapture study was undertaken over 6 consecutive days. It is possible to perform a test for closure: Otis *et al* (1978) present a test for closure which is unaffected by the presence of heterogeneous capture probabilities, whilst Stanley and Burnham (1999) propose a test for closure for cases where capture probabilities vary over time. We have previously seen how failing to account for heterogeneity in capture may lead to biased estimates of population size. Similarly, if closed population models are fitted to data from populations which exhibit birth/immigration and death/emigration, capture probabilities will generally be underestimated and hence population size overestimated. The literature of open population capture-recapture models is extensive (for a review see for example King (2014) and McCrea and Morgan (2014)). However, for such open population models, the focus has generally been on the estimation of demographic parameters such as survival, and the factors that may affect the survival probabilities, for example, to gain an understanding of the system and the primary factors driving the population which may be important for conservation and management purposes.

The original open capture-recapture model which allowed inference to be made on population size was the Jolly-Seber model (Jolly, 1965; Seber, 1965). However for this original model the births in the populations were not directly incorporated into the model but were inferred from the relationship between survival and the numbers of unmarked individuals captured on consecutive capture occasions. The population size estimate was obtained through the use of a Horvitz-Thompson-like estimate (Borchers *et al*, 2002; Horvitz and Thompson, 1952) as a ratio of the number of captured individuals and estimated probability that an individual is observed within the study. The disadvantage of such an approach is that it is not possible to constrain births to be positive, and hence negative birth rates can be obtained. In addition it is not possible to formally model the birth rates, for example, as a function of population size and/or additional covariate information.

Schwarz and Arnason (1996), proposed an alternative approach for open population models introducing the idea of a "super-population" from which individuals may enter and leave the study area. The extended Jolly-Seber-type model builds on the likelihood given in equation (2), permitting arrivals to the study area (after which individuals can be subsequently observed); and departures from the study site (from which time individuals are no longer able to be observed). It is assumed that once each individual arrives at the study site they remain there until they depart and cannot return thereafter (i.e. there is no temporary migration).

To account for the additional processes (arrivals and departures), we define the additional parameters:

β_t : the probability that an individual in the super-populaton and available for capture at some point within the study period is first available for capture at occasion $t+1$;

ϕ_t : the probability an individual present in the study area at occasion t remains in the study area until occasion $t+1$.

We note that by definition, $\sum_{j=0}^{T-1} \beta_j = 1$, since we condition on the individual

being available for capture at some point within the study period. For notational convenience we set $\boldsymbol{\beta} = \{\beta_0, \ldots, \beta_{T-1}\}$ and $\boldsymbol{\phi} = \{\phi_1, \ldots, \phi_{T-1}\}$.

The likelihood expression can be written in the form:

$$L(N, \boldsymbol{p}, \boldsymbol{\beta}, \boldsymbol{\phi}; \boldsymbol{x}) \propto \frac{N!}{(N-D)!} \prod_{i=1}^{D} \Pr(\boldsymbol{x}_i) \times \Pr(\boldsymbol{x}_0)^{N-D},$$

where $\Pr(\boldsymbol{x}_i)$ denotes the probability of the capture history of individual i; and $\Pr(\boldsymbol{x}_0)$ the associated probability of not being observed within the study (given the model parameters). Both of these probability terms need to account for the arrival and departure of the individuals - an individual cannot be recaptured before they arrive at the study site; or after they leave the study site (or alternatively the associated capture probabilities at these times are equal to 0). For notational convenience, we let g_i, and l_i, denote the occasions where individual i is observed for the first time, and last time, respectively. Recall that we assume that there is no temporary migration, so that under this assumption, an individual must be present (and available for capture) for times $g_i, \ldots, l_i$. However, we do not necessarily know when the individual arrives - it could be that the individual is first available for capture at time g_i, when it was observed; or the individual may have arrived earlier but have been unobserved until time g_i. Similarly, we know that the individual is in the population and available for capture between times g_i and l_i, inclusive; but after l_i it is again unknown whether the individual leaves the study area before capture occasion $l_i + 1$ (and so cannot be observed), or remains in the study area available for capture but is simply unobserved until it leaves at a future time (or remains in the study area until the end of the study). Thus we need to account for the unknown arrival and departures times (if any) from the study - this can be done by summing over all possible arrival and departure times. For example, for model M_t, with simply a temporal dependence on the capture probabilities, the probability of the capture history of individual $i = 1, \ldots, D$, denoted $\boldsymbol{x}_i$, is given by,

$$\Pr(\boldsymbol{x}_i) = \sum_{b=1}^{g_i} \beta_{b-1} \sum_{d=l_i}^{T} (1-\phi_d) \prod_{t=b}^{d-1} \phi_t \prod_{t=b}^{d} p_t^{x_{it}} (1-p_t)^{1-x_{it}},$$

where we define $\phi_T = 0$; and apply the mathematical convention that the product over the empty set is equal to unity, i.e. $\prod_{t=b}^{b-1} \equiv 1$. The first term relates to an individual joining the study area; the second term to the individual leaving the study area; the third term for remaining in the study from arrival to departure; and the final term to the capture of the individual given they are in the study area. Similarly, we let p^* denote the probability that an individual is observed at least once within the study. Then, $\Pr(\boldsymbol{x}_0) = 1 - p^*$, such that

$$1 - p^* = \sum_{b=1}^{T} \beta_{b-1} \sum_{d=b}^{T} (1-\phi_d) \prod_{t=b}^{d-1} \phi_t \prod_{t=b}^{d} (1-p_t).$$

This model has been successfully extended to account for the probability that an animal leaves the study area is a function of its arrival time (or equivalently

the duration of time at the study location; Pledger *et al* (2009)). This model is particularly useful for animals stopping over at breeding sites - see Matechou *et al* (2013, 2014) for some applications - and is thus often referred to as a *stopover* model. Additional behavioural and/or individual heterogeneity effects can be incorporated into the model, by extending the definition of the capture probabilities. For further discussion of these models, including the extension to discrete time-varying individual heterogeneity, see Worthington *et al* (2018b); and for a Bayesian approach of fitting a model with discrete time-varying heterogeneity see, for example, Dupuis and Schwarz (2007).

The theory of robust design (Kendall *et al*, 1995; Pollock, 1982) combined both open and closed population models by defining primary periods of the study which are open and secondary sampling periods which occur within a primary period within which the population is assumed to be closed. This study design facilitates the estimation of abundance, whilst allowing for the estimation of important demographic parameters; however the assumption of closure is again required within the secondary sampling periods, which may be unrealistic. The open robust design model Kendall and Bjorkland (2001) removed this closure assumption for the secondary periods, by conditioning on the number of individuals capture in each primary occasion, which in turn removed the ability to estimate abundance directly. The concept of robust design has been extended to the related area of removal modelling (Zhou *et al*, 2018) which is a special case of the behavioural model discussed in this chapter. Worthington *et al* (2018b) extends the robust design models in two ways: by removing the conditional requirement of robust design models by directly modelling all arrival and departure times to the study site within both the primary and secondary sampling periods; and introducing additional time-varying discrete individual covariates, using the analogous approach to Section 3.3.4.

5 Discussion

The accurate estimation of population sizes is an important aspect of conservation and wildlife management. There are several common methods used to obtain estimates of population size, including capture-recapture studies. The associated capture-recapture models rely on a number of assumptions - we briefly discussed the implications of the assumption of closure in Section 4 and described extended models to allow for such open populations. The failure of the study to satisfy the model assumptions can have a significant impact on the estimates of total population size. This has led to the development of numerous capture-recapture models, such as those described in Section 3 to allow for different forms of heterogeneity that may arise. These models have been constructed based on an understanding of the different factors that may influence the systems: the possible effects on the system are traditionally described in terms of temporal, behavioural and individual heterogeneity. These models are generally fairly rich in nature and can be adapted to incorporate specific knowledge of the system under study. For example, catch-effort models can be applied when the known resources expended at each capture occasion is recorded - in such cases it may be appropriate to reparameterise the

model in terms of the recapture probability per unit effort (see for example, King and Brooks (2004) for an example of such a parameterisation).

Traditionally, the capture-recapture models that have been developed rely on the data collection techniques involving a number of observers going into the field and observing (and subsequently marking) individual animals. However, recent technological advances have had a significant impact on the way that data may be collected - for example using photographic identification rather than physically adding a mark to the animal to minimise disruption/impact on the animals; using alternative non-visual queues such as scat/hair samples; or even motion-sensor cameras instead of physical traps. These advances in data collection techniques provide new forms of data and associated challenges that the traditional models in their current form do not address. We discuss some of these challenges and the associated statistical developments to combat these.

In the construction of the capture histories of the individuals observed within the study, it is assumed that each individual carries a unique identification, and that these are recorded without error when an individual is observed. However, within mark-resight studies it may be the case that an individual can be observed to have a mark, but that the unique identifier cannot be read. Thus, the individual capture histories cannot be constructed completely and a given number of observations of marked individuals can be recorded but not assigned to individuals. To address this issue McClintock *et al* (2014) consider a Bayesian data augmentation technique (and discuss previous approaches). Alternatively, modern techniques for identifying individuals, such as DNA matching from scat/hair samples or photo-identification can lead to non-ignorable mismatches within the study - including false negatives when a single individual is observed at two separate capture occasions but is not matched and hence recorded as two different individuals; or false positives where two different individuals seen at different capture occasions are incorrectly matched to be the same individual. The nature of the scheme for identifying individuals, and associated protocol, provides information on whether false positives and/or false negatives may occur. For example, when using DNA to match individuals, due to genotyping errors that may occur in the process of the given sample, false negatives may occur, but not false positives. For further discussion of issues using DNA for matching individuals, see for example Lukacs and Burnham (2005a,b); Wright *et al* (2009).

The problem of not being able to form individual capture histories arises when individuals are not able to be uniquely identified. A traditional marking method where this arises relates to *batch* marking. This is simply where an individual is marked such that its initial capture time is recorded, for example a colour-coded system may be used such that the first time an individual is observed a (permanent) marker/paint may be used - with a different colour applied at each capture occasion. The colour is only applied the first time an individual is observed; if an individual already has a colour no additional colour is added at a future capture occasion. At each capture occasion, the number of individuals are recorded that were first observed at each previous capture occasion - but we do not have the individual capture histories of each individual observed within the study (just the total number of individuals observed that were first observed at each previous capture occasion).

Estimates of the total population size can be obtained using for example estimating equations and a pseudo-likelihood (Huggins *et al*, 2010) or hidden Markov model approach (Cowen *et al*, 2017).

Location data may also be recorded in addition to an individual being observed/not observed. For example, capture-recapture studies may involve an array of traps over some spatial region - the capture histories of individuals will then typically detail not only whether an individual was observed at a capture occasion - but also in which trap the individual was observed. The traditional, standard, capture-recapture models collapse these data to simply observed/not observed, discarding the associated spatial information. However, incorporating the spatial information permits the estimation of spatial densities of animals, in addition to the estimation of total population size (Borchers and Efford, 2008; Efford, 2004; Royle and Young, 2008). The development of models to explicitly incorporate this spatial information have exploded in the last decade - and are now referred to as spatially explicit capture-recapture (SECR); or spatial capture-recapture (SCR). Typically, for these models the capture probability of an individual at a given time and trap location is expressed as a function of the distance of the given trap to the (unobserved) home range centre of the individual. The unobserved home range centre can be regarded as a two-dimensional individual (continuous) random effect - the likelihood is calculated by integrating out these unobserved locations. A common aspect of modern SECR data relates to the captures being observed in continuous time - with the spatial array of traps corresponding to motion sensors, so that the data recorded relates to both the individual and the exact time that the individual is observed by the sensor. Two issues immediately arise with regard to correctly identifying individuals from the photographs/videos and the continuous time recordings as opposed to the traditional discrete observation times. For discussion of many of these issues, and associated models, see for example, Borchers *et al* (2019); Royle *et al* (2013).

The technological advances associated with the collection of capture-recapture-type data will continue. However, there is an inevitable delay between the advanced technological tools for collecting the data and the necessary statistical tools to fully analyse the data to maximise the information that can be obtained. In addition, not only may the types of data change over time, but also the amount of available data, for example, multiple types of data may be collected on the same species/ecosystem; or the number of individuals on which data are collected may increase from 10s or 100s of individuals to 1000s of individuals. This leads to additional statistical challenges such as developing integrated models that incorporate different types of data within a single robust analysis which may be particularly challenging when the datasets are not independent of each other; or where the use of traditional methods are computationally infeasible due to the number of individuals within the study. Although the types of data, and associated size and/or and complexity are changing, the need for accurate estimation of population size across many different areas of application will continue to drive forward the statistical innovation in developing appropriate models and corresponding efficient computational model-fitting techniques.

References

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

Amstrup, S. C., McDonald, T. L. and Manly, B. F. J. (2005) *Handbook of Capture-Recapture Analysis*. Princeton University Press.

Bird, S. M. and King, R. (2018) Multiple systems estimation (or capture-recapture estimation) to inform public policy. *Annual Review of Statistics and its Application* **5**, 95-118.

Borchers, D. L., Buckland, S. T. and Zucchini W. (2002) *Estimating Animal Abundance, Closed Populations*. Springer. 314 pages.

Borchers, D. L. and Efford, M. G. (2008) Spatially explicit maximum likelihood methods for capture-recapture studies. *Biometrics*, **64**, 377–85.

Borchers, D. L., Stevenson, B. C., Howe, E. J., Distiller, G. and Glennie, R. (2019) *Spatial Capture-recapture by Maximum Likelihood*. Springer.

Böhning, D., van Der Heijden, P. G. M. and Bunge, J. (2018) *Capture-recapture Methods for the Social Sciences*. Chapman and Hall. 429 pages.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. and Thomas, L. (2004) *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press. 448 pages.

Burnham, K. P. and Anderson, D. R. (2002) *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*, Second Edition. Springer. 488 pages.

Chapman, D. G. (1951) Some properties of the hypergeometric distribution with applications to zoological sample censuses. *University of California Press*.

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

Cowen, L. L. E., Besbeas, P., Morgan, B. J. T. and Schwarz, C. J. (2017) Hidden Markov models for extended batch data. *Biometrics*, **73**, 1321-1331.

Cubaynes, S., Pradel, R., Choquet, R., Duchamp, C., Gaillard, J.-M., Lebreton, J.-D., Marboutin, E., Miquel, C., Reboulet, A., Poillot, C., Taberlet, P. and Gimenez, O. (2010) Importance of accounting for detection heterogeneity when estimating abundance: the case of the French wolves. *Conservation Biology* **24**, 621-626.

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**, 564-573.

Dorazio, R. M. and Royle, J. A. (2003) Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* **55**, 294-301.

Dupuis, J. A. and Schwarz, C. J. (2007) A Bayesian approach to the multistate Jolly-Seber capture-recapture model. *Biometrics* **63**, 1015-1022.

Efford, M. G. (2004) Density estimation in live-trapping studies. *Oikos* **106**, 598–610.

Fleming, P. and Tracey, J. (2008) Some human, aircraft and animal factors affecting aerial surveys - how to enumerate animals from the air. *Wildlife Research*, **35**, 258–267.

Goudie, I. B. J and Goudie M. (2007) Who captures the marks for the Petersen estimator? *Journal of the Royal Statistical Society: Series A*, **170**, 825-839.

Horvitz, D. G. and Thompson, D. J. (1952) A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, **47**, 663-85.

Huggins, R. (1991) Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics*, **47**, 725-732.

Huggins, R., Wang, Y. and Kearns, J. (2010) Analysis of an extended batch marking experiment using estimating equations. *Journal of Agricultural, Biological and Environmental Statistics*, **15**, 279-289.

Jeyam, A., McCrea, R. S. and Pradel, R. (2018) A test of positive association for detecting heterogeneity in capture for capture-recapture data. *Journal of Agricultural, Biological and Environmental Statistics*, **23**, 1-19.

Jolly, G. M. (1965) Explicit estimates from capture-recapture data with both death and immigration stochastic model. *Biometrika*, **52**, 225-47.

Kendall, W. L., Pollock, K. H. and Brownie, C. (1995) A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics*, **51**, 293-308.

Kendall, W. L. and Bjorkland, R. (2001) Using open robust design models to estimate temporary emigration from capture-recapture data. *Biometrics*, **57**, 1113–1122.

Kim, D. and Lindsay, B. G. (2015) Empirical identifiability in finite mixture models. *Annals of the Institute of Statistical Mathematics*, **67**, 745-772.

King, R. (2012) A review of Bayesian state-space modelling of capture-recapture-recovery data. *Interface Focus* **2**, 190-204.

King, R. (2014) Statistical ecology. *Annual Review of Statistics and its Application*, **1**, 410-426.

King, R. and Brooks, S. P. (2008) On the Bayesian estimation of a closed population size in the presence of heterogeneity and model uncertainty. *Biometrics* **64**, 816–24.

King, R. and Brooks, S. P. (2004) A classical study of catch-effort models for Hector's dolphins. *Journal of the American Statistical Association*, **90**, 325-333.

King, R. and Langrock, R. (2016) Semi-Markov Arnason-Schwarz models. *Biometrics*, **72**, 619-628.

King R., McClintock, B., Kidney, D. and Borchers, D. L. (2016) Abundance estimation using a semi-complete data likelihood approach. *Annals of Applied Statistics*, **10**, 264-285.

King, R. and McCrea, R. S. (2014) A generalised likelihood framework for partially observed capture-recapture-recovery data. *Statistical Methodology*, **17**, 30-45.

King, R., Morgan, B. J. T., Gimenez, O. and Brooks, S. P. (2009) *Bayesian Analysis for Population Ecology*. CRC Press. 456 pages.

Langrock, R. and King, R. (2013) Hidden Markov-type maximum likelihood estimation of mark-recapture-recovery models in the presence of continuous covariates. *Annals of Applied Statistics* **7**, 1709-1732.

Laplace, P. S. (1786), Sur les Naissances, les Mariages et les Morts. *Histoire de LAcademie Royale des Sciences*, 1783, Paris.

Lincoln, F. C. (1930) Calculating waterfowl abundance on the basis of banding returns. *US Department of Agriculture Circular* **118**, 1-4.

Little, R. J. A. and Rubin, D. B. (2002) *Statistical Analysis with Missing Data*. John Wiley & Sons. 389 pages.

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

Lukacs, P. M. and Burnham, K. P. (2005b) Estimating population Size from DNA-based closed capture-recapture data incorporating genotyping error. *The Journal of Wildlife Management*, **69**, 396-403.

Matechou, E., Morgan, B. J. T., Pledger, S., Callazo, J. A. and Lyons, J. E. (2013) Integrated analysis of capture-recapture-resighting data and counts of unmarked birds at stop-over sites. *Journal of Agricultural, Biological, and Environmental Statistics*, **18**, 120-135.

Matechou, E., Dennis, E., Freeman, S. N. and Brereton, T. (2014) Monitoring abundance and phenology in (multivoltine) butterfly species: a novel mixture model. *Journal of Applied Ecology*, **51**, 766-775.

McClintock, B. T., Hill, J. M., Fritz, L., Chumbley, K., Luxa, K. and Diefenbach, D. R. (2014) Mark-resight abundance estimation under incomplete identification of marked individuals. *Methods in Ecology and Evolution*, **5**, 1294-1304.

McCrea, R. S., Morgan, B. J. T. and Cole, D. J. (2013) Age-dependent models for recovery data on animals marked at unknown age. *Journal of the Royal Statistical Society Series C: Applied Statistics*, **62**, 101-113.

McCrea, R. S. and Morgan, B. J. T. (2014) *Analysis of Capture-Recapture Data*. Chapman and Hall. 314 pages.

Morgan, B. J. T. and Ridout, M. S. (2008) A new mixture model for capture heterogeneity. *Journal of the Royal Statistical Society Series C: Applied Statistics*, **57**, 433-446.

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.

Petersen, C. G. J. (1896) The yearly immigration of young plaice into the Limfjord from the German sea. *Report of the Danish Biological Station (1895)*, **6**, 5-84.

Pledger, S. (2000) Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics*, **56**, 434–42.

Pledger, S. (2005) The performance of mixture models in heterogeneous closed population capture–recapture. *Biometrics*, **61**, 868-876.

Pledger, S., Efford, M., Pollocj, J., Callazo, J. and Lyons, J. (2009) Stopover duration analysis with departure probability on unknown time since arrival. *Environmental and Ecological Statistics*, **3**, 1071-1082.

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

Pradel, R. (2005) Multievent: an extension of multistate capture-recapture models to uncertain states. *Biometrics*, **61**, 442-447.

Ralph, C. J. and Dunn, E.J. (Editors) (2004) *Monitoring bird populations using mist nets*. Studies in Avian Biology No. 29, Cooper Ornithological Society.

Royle, J. A., Chandler, R. B., Sollmann, R. and Gardner, B. (2013) *Spatial Capture-Recapture*. Academic Press. 612 pages.

Royle, J.A., Karanth, K. U., Gopalaswamy, A. M., Kumar, N. S. (2009) Bayesian inference in camera trapping studies for a class of spatial capture-recapture models. *Ecology*, **90**, 3233-3244.

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

Sandland, R. L. and Cormack, R. M. (1984) Statistical inference for Poisson and multinomial models for capture-recapture experiments. *Biometrika*, **71**, 27-33.

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

Schwarz, C. J. and Arnason, A. N. (1996) A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics*, **52**, 860-73.

Seber, G. A. F. (1965) A note on the multiple-recapture census. *Biometrika*, **52**, 249-59.

Self, S. G. and Liang, K.-Y. (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**, 605-610.

Smout, S. C., King, R. and Pomeroy, P. P. (2011) Estimating Demographic Parameters for Capture-Recapture Data in the Presence of Multiple Mark Types. *Environmental and Ecological Statistics*, **18**, 331-347.

Stanley, T. R. and Burnham, K. P. (1999) A closure test for time-specific capture-recapture data. *Environmental and Ecological Statistics*, **6**, 197-209.

Tanner, M. and Wong, W. (1987) The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, **81**, 528-550.

Worthington, H., McCrea, R. S., King, R. and Griffiths, R. (2018a) Estimating abundance from multiple sampling capture-recapture data using hidden Markov models in the presence of discrete-state information Technical report, University of St Andrews.

Worthington, H., McCrea, R. S., King, R. and Griffiths, R. (2018b) Estimating population size for multi-state closed capture-recapture data. Technical report, University of St Andrews. arXiv: https://arxiv.org/abs/1708.00348.

Wright, J. A., Barker, R. J., Schofield, M. R., Frantz, A. C., Byrom, A. E. and Gleeson, D. M. (2009) Incorporating genotyping uncertainty into mark-recapture-type models for estimating abundance using DNA samples. *Biometrics* **65**, 833–40.

Zhou, M., McCrea, R. S., Matehcou, E., Cole, D. J. and Griffiths, R. A. (2018) Removal models accounting for temporary emigration. *Biometrics*. In press. 10.1111/biom.12961.

Zucchini, W., MacDonald, I. L., Langrock, R. (2016) *Hidden Markov Models for Time Series - An Introduction Using R* (2nd Edition). Chapman and Hall/CRC. 370 pages.