8-2-2019 10:30 AM

Exploring the Estimability of Mark-Recapture Models with Individual, Time-Varying Covariates using the Scaled Logit Link Function

Jiaqi Mu, *The University of Western Ontario*

Supervisor: Bonner, Simon J., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Statistics and Actuarial Sciences

Recommended Citation

Mu, Jiaqi, "Exploring the Estimability of Mark-Recapture Models with Individual, Time-Varying Covariates using the Scaled Logit Link Function" (2019). *Electronic Thesis and Dissertation Repository*. 6385.
https://ir.lib.uwo.ca/etd/6385

Abstract

Mark-recapture studies are often used to estimate the survival of individuals in a population and identify factors that affect survival in order to understand how the population might be impacted by changing conditions. Factors that vary between individuals and over time, such as body mass, present a challenge because they can only be observed when an individual is captured. Several models have been proposed to deal with this missing-covariate problem and commonly impose a logit link function which implies that the survival probability varies between 0 and 1. In this thesis I explore the estimability of four possible models when survival is linked to the covariate through a scaled logit link function which imposes some upper limit, $c < 1$. Through a combination of theoretical analysis and simulation I show that the binomial model is not estimable under the scaled link while the other three models remain estimable.

Keywords: model estimability, scaled logit link function, missing data, maximum likelihood, Bayesian inference

Summary

Wildlife conservation has become a global task in the past few decades. A large number of studies and field experiments have been conducted to assist in wildlife management and conservation. Ecologists are often interested in monitoring the abundance or understanding factors that affect survival of an animal population, and have proposed many statistical models to study these properties.

The Cormack-Jolly-Seber (CJS) model with covariates is often used to estimate the survival of individuals in a population and identify factors that affect survival in order to understand how individuals might behave differently and how the population might be affected by changing conditions. Factors that vary between individuals and over time, like body mass, present a challenge because they can only be observed when an individual is captured. Several extensions of the CJS models have been proposed to deal with the missing-covariate problem and to understand the effects that a factor may have on survival even when it cannot be observed at all times for all individuals. Moreover, these models all impose the assumption that the survival probability varies between 0 and 1.

In this thesis, I examine the behaviour of the different models when the survival probability is modeled via a scaled logit link function that restricts the survival probability to be less than some constant, $c < 1$, that must be estimated from the data. In particular, I explore the estimability of four possible models: the binomial model, trinomial model, full-likelihood model, and alternative trinomial model. Through a combination of theoretical analysis and simulation I show that effects of an individual time-varying covariate on survival cannot be estimated from the binomial model when survival is scaled to be less than some $c < 1$, but the other three models remain estimable. For me, the work in the thesis is important because it helps us to understand what types of models we can use, and what type of data biologists need to collect, in order to examine the effects of different types of factors that may have a significant impact on individual survival and, hence, the sustainability of wild animal populations.

Acknowlegements

In the process of writing this thesis, I received a great deal of support and assistance. I would first like to thank my supervisor, Dr. Simon Bonner, for presenting the ideas behind the thesis and helping me solve the problems I encountered. He always encouraged me and pushed me forward. Then I would like to thank Dr. Wei Zhang, for his helpful advice and comments on my thesis.

I really enjoyed my time at the Western University. I want to thank my friends and classmates who have brought me a lot of fun in my study life. In addition, I want to thank my parents who gave me unconditional support and love.

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview

Wildlife conservation has become a worldwide task in the past decades. A lot of research and fieldwork have been done to help with wildlife management and protection. In wildlife research, one is often interested in the abundance or survival of an animal population. Since it is unrealistic to count or observe all individuals in the population, one first conducts experiments to study the population of interest. Statistical models are then developed to analyze the data from the experiment. Over the past century, capture-recapture (CR) methods have been widely used to estimate the abundance (i.e., the total number of animals in a population) or other demographic parameters (e.g., survival probability). This CR method is also the basis of other more complex models, which can obtain some auxiliary information (e.g., individual covariates) to estimate relevant parameters more accurately. Many statistical and computational techniques can be applied to these models to deal with more complicated cases and improve the accuracy of the parameter estimation, which is an active research area.

In this thesis, I focus on extensions of the Cormack-Jolly-Seber (CJS) model (Cormack, 1964; Jolly, 1965; Seber, 1965) where survival probability is linked to an individual, continuous, and time-dependent covariate. For example, I am interested in how an individual's body mass influences the survival probability which usually uses a logit link function to explore the relationship. However, covariates are missing if individuals are not captured on some occasions. In this case, the survival probabilities cannot be expressed in terms of covariates, thus we fail to write down the likelihood function for the CJS model.

So far, two methods have been developed to deal with the missing-covariate problem. One is deleting unknown terms in the likelihood function such as the trinomial model based on the three-state process for mark-recapture-recovery data (Catchpole et al., 2008). The other is the full-likelihood approach which models the covariates to construct a likelihood function that can be optimized by Monte Carlo EM (MCEM) algorithm or incorporated into a Bayesian analysis (Bonner and Schwarz, 2006). Note that the MCEM algorithm is a modification of the EM algorithm where the expectation in the E-step is computed numerically through Monte Carlo simulation.

All of these previous methods assume that the survival is linked to the covariate through a simple link function, by default the logit link, which assumes that survival can vary between 0 and 1. However, the logit link function can be questionable because we cannot promise the individual will survive due to, for example, a large body mass. Instead, the scaled logit link function has been chosen to explore the relationship between survival and covariate.

In similar work, Knape and Korner-Nievergelt (2015) showed that estimates of abundance using the single-visit model with covariates of Sólymos et al. (2012) rely on accurately specifying the link function for detection probability. Their findings indicate that the choice of the link function has a great impact on the estimation of parameters, and particularly that the scaled link function makes it impossible to estimate the parameters of the basic single-visit N-mixture models. For our problem, if we employ the scaled logit link function, the models we investigate may become non-estimable – meaning that the data does not provide significant information about the true parameter values. Futher details on the definition of estimability are provided in Section 2.5.

The purpose of this thesis is to investigate the effects of using a scaled logit link function on the estimability of the models developed to address the missing data problem that arises when covariates vary both by time and individual and can only be observed when individuals are captured. I conduct simulations to explore the estimability of the binomial model, trinomial model, alternative trinomial model, and full likelihood model.

1.2 Closed-Population Models

Closed-population models, whose main goal is to estimate the population size of the study area, assume that no individuals enter the population by births or immigration or leave the population through deaths or emigration during the experiment. Lincoln-Petersen estimator (Lincoln et al., 1930), Chapman estimator (Chapman, 1951), and other more complex closed-population models (Pollock, 1974) were proposed over the last century, which are based on the capture-recapture experiment.

Capture-recapture (CR) is a popular method for estimating the abundance or other demographic parameters of an animal population (Pollock et al., 1990). The CR method is useful when counting all individuals in a population is impractical. Pearson (1896) first used the method to estimate the size of the population of plaice (*Pleuronectes platessa*) from the German sea. Since then CR experiments have been used to study populations of many different species. The same methods are also used in medical studies, and in the study of (Chartier et al., 2015) to estimate the prevalence of chronic kidney disease based on cases defined in administrative and laboratory data. Furthermore, CR data are typically a set of capture histories, each of which is a string of binary digits of length K with 1 standing for capture and 0 for non-capture. For example, an individual with history 01001 indicates that the individual was captured on occasions 2 and 5, while not captured on other occasions in a 5-occasion study.

The Lincoln-Petersen (LP) estimator (Lincoln et al., 1930; Petersen, 1896) and the Chapman estimator (Chapman, 1951) can be constructed from the simplest CR study, which only consists of two capture occasions. On the first occasion, a sample of animals of size n_1 is captured from the study area. Researchers put a unique tag on each of these animals and then release them back to the population. After the marked animals sufficiently mix back with the rest of the population, another sample of size n_2, which is independent of the first sample, is captured. If the number of previously marked individuals in the second sample is m, then the LP estimator is

$$\hat{N} = \frac{n_1 n_2}{m}, \tag{1.1}$$

where N denotes the population size, and we assume that the population is closed, being marked does not affect the marked animals and the marks will not be lost. The LP estimator is very simple but there are many limitations to this estimator. For example, some species are not easy to be observed such that $m = 0$ or some small value. Chapman (1951) indicated that the LP estimator is asymptotically unbiased as the sample size approaches infinity, but is biased at small sample sizes and suggested an alternative estimator of population size, the Chapman estimator:

$$\hat{N} = \frac{(n_1 + 1)(n_2 + 1)}{m + 1} - 1, \tag{1.2}$$

This is less biased than the Lincoln-Petersen estimator for small samples (Chapman, 1951).

In practice, to make more accurate inferences of the population, the closed-population models employ the CR data containing more than two capture occasions. Eight models were presented by Pollock (1974) and were fully developed by Otis et al. (1978). For example, the simplest model, commonly referred to as model M_0, assumes all individuals have the same probability p of being captured on all occasions. Let M denotes the number of individuals captured at least one time during in the study, let K denote the number of occasions and let n_t denote the number of individuals captured on occasion t. The MLE (maximum likelihood estimation) of the population size N can be found through solving equation (1.3) (Darroch, 1958):

$$\left(1 - \frac{M}{N}\right) = \prod_{t=1}^{K} \left(1 - \frac{n_t}{N}\right). \tag{1.3}$$

Given the value of N by the equation (1.3), the MLE of all parameters can be obtained by maximizing the $\ln L$ (1.4) (Otis et al., 1978):

$$\ln L = \ln\left(\frac{N!}{(N-M)!}\right) + \left(\sum_{t=1}^{K} n_t\right)\ln p + \left(KN - \sum_{t=1}^{K} n_t\right)\ln(1-p). \tag{1.4}$$

and then the MLE of p is

$$\hat{p} = \frac{\sum_{t=1}^{K} n_t}{KN}. \tag{1.5}$$

1.3 Open-Population Models

Individual survival probability is a vital rate and often measured in the open-population CR models. The open-population models allow individuals to die, be born, immigrate, and emigrate during the experiment. The Jolly-Seber model (Jolly, 1965; Seber, 1965) serves as the starting point for open-population CR models which allows estimation of survival and capture probabilities, population size, and the number of new individuals entering the population. In contrast, the Cormack-Jolly-Seber model is limited to estimation of survival and capture probabilities, serves as the basis for the thesis and is described in the following paragraphs.

The Cormack-Jolly-Seber (CJS) model, formulated by Cormack (1964), Jolly (1965), and Seber (1965), models the open-population CR data and is widely used in ecological studies. By July 2019, Seber (1965) has been cited almost 2000 times according to Google Scholar. Furthermore, many extensions to the CJS model have been developed. For example, in the multi-state model, the capture and survival probabilities can change independently between animals in a set of finite states defined by some factors (Arnason, 1973). Here I am interested in models which allow the probability of survival to depend on continuous covariates that vary between individuals and over time.

Let $\boldsymbol{\omega_i} = (\omega_{i1}, \omega_{i2}, \ldots, \omega_{iK})$ denote the capture history of individual i, where $\omega_{it} = 1$ if the individual was captured on occasion t ($1 \leq t \leq K$) and 0 otherwise, and let n denote the number of different individuals captured in the study. It is also useful to introduce a set of latent variables indicating when the individual was alive and marked. Specifically, the survival history is represented as $\boldsymbol{d_i} = (d_{i1}, d_{i2}, \ldots, d_{iK})$ where $d_{it} = 1$ if individual i is alive and $\sum_{t=1}^{K} \omega_{it} > 0$, and $d_{it} = 0$ otherwise.

The key assumptions of the CJS model are that all individuals alive on given capture occasion are equally likely to be captured and equally likely to survive to the next occasion. More specifically, the model assumes that (Williams et al., 2003):

1. Capture probability $p_t = P(\omega_{i,t} = 1 | d_{i,t} = 1)$ is the same for all individuals on occasion t.
2. Survival probability $\phi_t = P(d_{i,t+1} = 1 | d_{i,t} = 1)$ is the same for all individuals on occasion t.
3. Marks are neither lost nor overlooked and are recorded correctly throughout the study.
4. Sampling periods are instananeous and recaptured individuals are released immediately.
5. Emigration from the study area is permanent.
6. Individuals are independent from each other.
7. Captures of the same individual on different occasions are independent.

Figure 1.1 illustrates the structure of the model for a single individual. Note that p_1 cannot be estimated because the likelihood function for the CJS model (see equation (1.6)) conditions on the first capture of each individual. Also, the final capture and survival probabilities, p_K and ϕ_{K-1}, are confounded and cannot be estimated separately.

Figure 1.1: The structure of the CJS model.

$$\begin{tikzcd}
d_{i1} \arrow[r, "\phi_1"] \arrow[d, "p_1"] & d_{i2} \arrow[r, "\phi_2"] \arrow[d, "p_2"] & d_{i3} \arrow[r, "\phi_3"] \arrow[d, "p_3"] & \cdots \arrow[r, "\phi_{K-1}"] & d_{iK} \arrow[d, "p_K"] \\
\omega_{i1} & \omega_{i2} & \omega_{i3} & & \omega_{iK}
\end{tikzcd}$$

As an example, we consider an individual with capture history 01010. The likelihood contribution for this individual is

$$\phi_2(1 - p_3)\phi_3 p_4((1 - \phi_4) + \phi_4(1 - p_5)), \tag{1.6}$$

and is formed as follows. First, we condition on the second occasion because that was when the individual was first captured. We know that the individual survived from occasion two to three and three to four, and the probabilities of these events are ϕ_2 and ϕ_3. Also, the individual was alive but not captured on occasion three (probability $1 - p_3$) and was alive and captured on occasion four (probability p_4). On the last occasion, we don't know whether the individual died or survived but simply was not captured. The combined probability of these two events is $(1 - \phi_4) + \phi_4(1 - p_5)$.

Since independence is assumed between individuals, the full likelihood for the whole set of capture histories is the product of all these single conditional probabilities. Jolly (1965) initially proposed method of moments type estimators for the capture and survival probabilities. In fact, Jolly's model also provides estimates of recruitment (birth and immigration). Seber (1965) showed that the maximum likelihood estimators of capture and survival probabilities have closed forms. Later, Bayesian inference was developed by Poole (2002). There are now several stand-alone software packages available, for example, Program MARK (White and Burnham, 1999), or packages for R, like `marked`, (Jeff Laake, 2018) that allow researchers to fit the CJS model easily, obtain confidence intervals for the parameters, conduct hypothesis testing (e.g., $p_2 = p_3 = \cdots = p_{K-1}$ (Lebreton et al., 1992)), and perform model selection.

1.4 CJS Model with Covariates

Ecologists often wish to incorporate some measures as covariates to account for the situation where assumptions of CJS model cannot be satisfied or more complicated situations arise. This idea was first proposed by Pollock et al. (1984) to estimate capture probability for the closed-population models. Pollock et al. (1984) linked the capture probability p_t to a scalar variable x_t through the logit link function:

$$\log\left(\frac{p_t}{1 - p_t}\right) = \beta_0 + \beta_1 x_t, \tag{1.7}$$

where β_0 is the intercept and β_1 is the slope. As we are interested in modeling the survival probability, model of Pollock et al. (1984) cannot be used because it assumes the population is closed. Later Lebreton et al. (1992) described the CJS model with external variables through the logit link function. Recall that the basic CJS model assumes that all individuals in the population share the same survival probability on the same sampling occasion. However, this assumption can be questionable in many situations because survival probability may depend not only on extrinsic conditions but also intrinsic characteristics such as body mass, which may vary between individuals and depend on sampling time. Williams et al. (2003) summarized all possible covariates and classified them according to two descriptors: discrete vs continuous and static vs dynamic. In my project, I am interested in modeling the effects of time-dependent covariates (i.e., continuous and dynamic), which are stochastic for each individual.

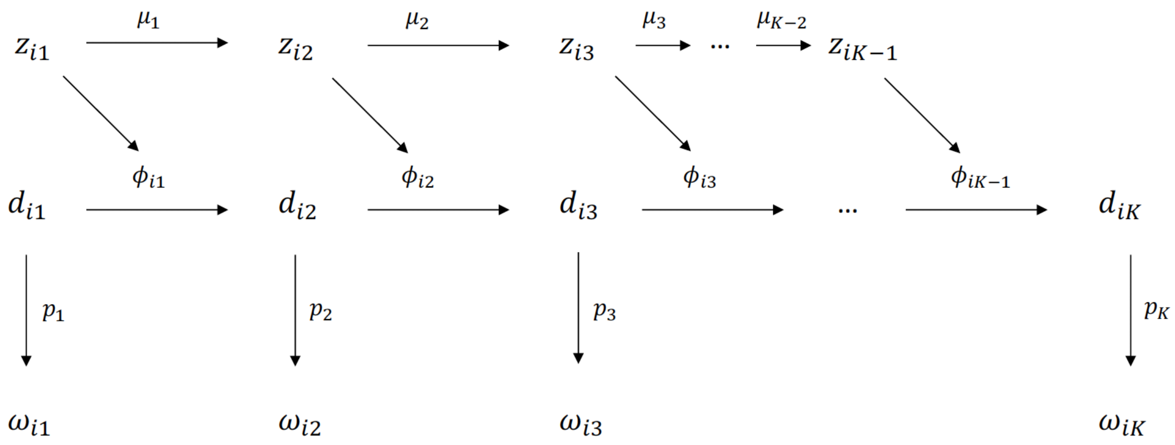
For the CJS model with covariates, data consists of the normal CR data and the corresponding covariates which are collected by measuring the characteristic of each captured individual. Suppose we consider a continuous, individual, and time-dependent covariate such as body mass. Let z_{it} denote the value of the covariate for individual i on occasion t, and let ϕ_{it} denote the survival probability from occasion t to $t+1$ for individual i. Here we only model the effect of the covariate on survival and assume that the capture probability p_t is continuous time-dependent but common to all individuals. Other variables are defined in the same manner as for the CJS model.

A key difference between the basic CJS model and the CJS model with covariates is that the survival probability for the latter is no longer assumed to be equal for all individuals on each occasion. Since the survival probability lies between 0 and 1, it is natural to use a logit link function to relate the survival probability to the covariate as we usually do for the logistic regression model. Indeed, there are many other choices for the link function such as the probit link functions, which are also widely used for generalized linear models. Under the logit link function, we have

$$\phi(z_{it}|\boldsymbol{\beta}) = \frac{\exp(\beta_0 + \beta_1 z_{it})}{1 + \exp(\beta_0 + \beta_1 z_{it})} \tag{1.8}$$

for $t = 1, \ldots, K - 1$. The model structure shown in Figure 1.2 is similar to that of the CJS model except the upper part, where the covariates are added.

Figure 1.2: The structure of the CJS model with covariates.



One considerable challenge of fitting the new model is that the covariate value, z_{it}, can only be observed if the individual is captured on occasion t, $\omega_{it} = 1$. Missing covariates prevent the construction of the complete likelihood function, so the inference methods presented in the last section cannot be applied. No matter which link function is used, we need the covariates of all individuals on all occasions except for the last one. This is because the covariate on the last occasion, z_{iK}, only influences the survival probability $\phi_{i,K+1}$, which is not involved in the likelihood of the model (Figure 1.2). Unfortunately, we have no way to collect covariate information of an individual if it was not encountered on some occasions. For example, if individual i has capture history 010010, then covariates z_{i3} and z_{i4} are missing. We can ignore the fact that z_{i6} was not observed because it does not influence any survival probability. Due to the missing covariates (e.g., z_{i3} and z_{i4} in the above example), we cannot write down the full likelihood function. To deal with this problem, various methods were proposed. For instance, Abraham and Russell (2004) developed a complete-case analysis which omitted all individuals involving missing covariates. Alternatively, Catchpole et al. (2004a) proposed a method in which all missing covariates were replaced with the last available values. However, both these two methods result in severe bias (Catchpole et al., 2008). In chapter 2, I introduce four models to handle missing covariates. These include three existing models: the binomial model (Catchpole et al., 2004b), the trinomial model (Catchpole et al., 2008) and the full likelihood model (Bonner, 2003). I also introduce an alternative model, which is a special case of the truncated model described by (Burchett, 2017).

Chapter 2

Methodology

2.1 Binomial Model

The binomial model (Catchpole et al., 2004b) is based on the CJS model with covariates but constructs the likelihood function by deleting all the unknown transitions which include missing covariates from the full likelihood function. As discussed in Section 1.4, missing covariates cause the problem that the corresponding survival probabilities cannot be calculated and thus the full likelihood and parameter estimates cannot be obtained. Suppose individual *i* has capture history 101001. The likelihood contribution of this individual is

$$\phi_{i1}(1-p_2)\phi_{i2}p_3\phi_{i3}(1-p_4)\phi_{i4}(1-p_5)\phi_{i5}p_6. \tag{2.1}$$

Catchpole et al. (2008) considered a partial-case analysis by omitting all unknown terms from the above likelihood contribution. Since this individual was not captured on occasions 2, 4, and 5, covariates z_{i2}, z_{i4}, and z_{i5} were not recorded. Consequently, ϕ_{i2}, ϕ_{i4}, and ϕ_{i5} cannot be computed and are removed from the full likelihood. Then the resulting reduced likelihood contribution for this individual is

$$\phi_{i1}(1-p_2)p_3\phi_{i3}(1-p_4)(1-p_5)p_6. \tag{2.2}$$

The complete-case analysis (Abraham and Russell, 2004) simply omits all animals which have any missing covariate. Compared to the complete-case analysis, the binomial model is less biased because it only omits unknown transitions instead of all animals with missing covariates.

We introduce a two-state process to help derive the likelihood for the binomial model. Recall that $\omega_{it=1}$ if the individual i was captured on occasion t and 0 otherwise. Define the transition $\pi_{i,t}(a, b|\boldsymbol{p}, \boldsymbol{\beta}, \boldsymbol{z_i})$ as:

$$\pi_{i,t}(a, b|\boldsymbol{p}, \boldsymbol{\beta}, \boldsymbol{z_i}) = \Pr(\omega_{i,t+1} = b \mid \omega_{i,t} = a, \omega_{i,t-1}, \ldots, \omega_{i,1}, \boldsymbol{p}, \boldsymbol{\beta}, \boldsymbol{z_i}), \quad a, b \in \{0, 1\}. \tag{2.3}$$

In addition, we define $\chi_{i,r,s}$ to be the probability that individual i was not captured from occasion $r + 1$ to s conditional on that it was alive at occasion r. Then we have the following recurrence relation equation:

$$\chi_{i,r,s} = (1 - \phi_{i,r}) + \phi_{i,r}(1 - p_{r+1})\chi_{i,r+1,s}, \quad a_i \leq r < s \leq K, \tag{2.4}$$

where we let $\chi_{i,r,r} = 1$. Now we have, for $a_i \leq r \leq K - 1$,

$$\pi_{i,r}(a, b|\boldsymbol{p}, \boldsymbol{\beta}, \boldsymbol{z_i}) = \begin{cases} \phi_{i,r}(1 - p_{r+1}) + (1 - \phi_{i,r}) & a = 1, b = 0 \\ \phi_{i,r}p_{r+1} & a = 1, b = 1 \\ \chi_{i,l_{i,r},r+1}/\chi_{i,l_{i,r},r} & a = 0, b = 0 \\ \prod_{s=l_{i,r}}^{r-1} \phi_{i,s}(1 - p_{s+1})\phi_{i,r}p_{r+1}/\chi_{i,l_{i,r},r} & a = 0, b = 1 \end{cases} \tag{2.5}$$

where $l_{i,r}$ is the last occasion on which an individual was captured before occasion r.

Using the transition probabilities defined above, the full likelihood function can be written as:

$$L(\boldsymbol{p}, \boldsymbol{\beta}; \boldsymbol{z}, \boldsymbol{\omega})) = \prod_{i=1}^{n} \prod_{r=a_i}^{K-1} \pi(\omega_{i,r}, \omega_{i,r+1}|\boldsymbol{p}, \boldsymbol{\beta}, \boldsymbol{z_i}), \tag{2.6}$$

where $a_i = \min\{t : \omega_{it} = 1\}$ and $b_i = \max\{t : \omega_{it} = 1\}$ denote the first and last occasion on which individual i was captured. Recall the previous example for the capture history 101001.

The contribution of this history to the two-state likelihood is

$$\pi_{i,1}(1,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})\pi_{i,2}(0,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})\pi_{i,3}(1,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})\pi_{i,4}(0,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})\pi_{i,5}(0,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i}). \tag{2.7}$$

It is obvious that if individual i was not recaptured on occasion t, the covariate z_{it} is unknown so that ϕ_{it} cannot be obtained. For this reason, the transition probabilities $\pi_{i,t}(0,b|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})$ $(b = 0, 1)$ are unknown. Therefore, $\pi_{i,2}(0,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})$, $\pi_{i,4}(0,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})$, and $\pi_{i,5}(0,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})$, which include $\phi_{i,2}$, $\phi_{i,4}$, and $\phi_{i,5}$, cannot be computed through the logit link function. After that omitting these three transitions from the likelihood, we have $\pi_{i,1}(1,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})$ and $\pi_{i,3}(1,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})$ left in the partial likelihood. Furthermore, the full likelihood presented by the equation (2.6) reduces to a partial likelihood by removing all transitions starting from 0 for all captured individuals:

$$L(\boldsymbol{p},\boldsymbol{\beta};\boldsymbol{z},\boldsymbol{\omega}) = \prod_{i=1}^{n} \prod_{r\in\{a_i,\ldots,K-1:\ \omega_{i,r}=1\}} \pi(1,\omega_{i,r+1}|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i}). \tag{2.8}$$

By maximizing this partial likelihood, we can find the maximum likelihood estimates (MLEs) of the parameters.

2.2 Trinomial Model

The trinomial model (Catchpole et al., 2008) is used to analyze mark-recapture-recovery (MRR) data with missing covariates for which the likelihood construction is based on a three-state process. The trinomial model resembles the binomial model as they both delete unknown transitions in the likelihood, is easy to implement, and performs well according to the simulation results of (Catchpole et al., 2008). The key difference is that the trinomial model analyzes the MRR data.

MRR data comprise a set of capture histories that consist of a new state, the dead recovery state 2 representing that an individual is found dead, in addition to the previously-defined states 0 and 1 on each capture occasion. Thus a capture history in MRR data is a string of ternary values of length K with 0 standing for non-capture, 1 for capture, and 2 for dead recovery. For example, capture history 01020 denotes that the individual was first captured on occasion 2, not encountered on occasion 3, and found dead on occasion 4.

The trinomial model involves all the assumptions of the CJS model with covariates and two more assumptions for the state of dead recovery. One is that if an individual dies between occasions t and $t + 1$, it is either found dead in this time interval or never found (i.e., the individual cannot be found dead after occasion $t + 1$). The other is that the dead recovery probability $\lambda_t = Pr(\omega_{i,t} = 2, d_{i,t} = 0 \mid d_{i,t-1} = 1)$ is the same for all individuals that die between occasions t and $t+1$. Under the trinomial model, we see that the individual with capture history 01020 was still alive on occasion 3, died between occasion 3 and 4, and was found dead with probability λ_3.

The full likelihood of the trinomial model conditional on the initial captures is (Catchpole et al., 2008),

$$L(\boldsymbol{p}, \boldsymbol{\beta}, \boldsymbol{\lambda}; \boldsymbol{z}, \boldsymbol{\omega}) = \prod_{i=1}^{n} \{(1 - \phi_{i,b_i})\lambda_{b_i}\}^{d_{i,bi}} {\chi_{i,b_i}}^{1-d_{i,b_i}} \prod_{t=a_i}^{b_i-1} \{\phi_{i,t} p_{t+1}^{w_{i,t}} (1 - p_{t+1})^{1-w_{i,t}}\}. \tag{2.9}$$

Similar to the basic CJS model with covariates, we cannot evaluate the likelihood (2.9) because of missing covariates. To deal with this issue, Catchpole et al. (2008) introduced a three-state process. Now the transitions $\pi_{i,t}(a, b)$ and $\chi_{i,r,s}$ defined for the binomial model have new expressions with a new state introduced. We add a “$*''$ to the top right corner of each transition to represent the new expressions. Let

$$\begin{aligned} \chi^*_{i,r,s} &= \Pr(\omega_{i,r+1} =, \ldots, \omega_{i,s} = 0 \mid d_{i,r} = 1) \\ &= (1 - \phi_{i,r})(1 - \lambda_r) + \phi_{i,r}(1 - p_{r+1})\chi^*_{i,r+1,s}, \end{aligned} \tag{2.10}$$

for $a_i \leq r < s \leq K$ with $\chi^*_{i,r,r} = 1$. Then we have,

$$\pi_{i,r}(a, b|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^* = \begin{cases} \phi_{i,r}(1-p_{r+1}) + (1-\phi_{i,r})(1-\lambda_r) & a=1, b=0 \\ \phi_{i,r}p_{r+1} & a=1, b=1 \\ (1-\phi_{i,r})\lambda_r & a=1, b=2 \\ \chi^*_{i,l_{i,r},r+1}/\chi^*_{i,l_{i,r},r} & a=0, b=0 \\ \prod_{s=l_{i,r}}^{r-1} \phi_{i,s}(1-p_{s+1})\phi_{i,r}p_{r+1}/\chi^*_{i,l_{i,r},r} & a=1, b=1 \\ \prod_{s=l_{i,r}}^{r-1} \phi_{i,s}(1-p_{s+1})(1-\phi_{i,r})\lambda_r/\chi^*_{i,l_{i,r},r} & a=0, b=2 \\ 1 & a=2, b=0 \end{cases} \tag{2.11}$$

for $a_i \leq r \leq K-1$, where $l_{i,r}$ is the last time an individual is captured alive before occasion r. Using these transitions, the full likelihood function for the trinomial model, equation (2.9), can be written as

$$L(\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda}; z,\boldsymbol{\omega}) = \prod_{i=1}^{n}\prod_{r=a_i}^{K-1} \pi_{i,r}(\omega_{i,r}, \omega_{i,r+1}|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^*. \tag{2.12}$$

Finally, all the transitions in (2.10) containing unknown terms are omitted to construct the new partial likelihood. From equation (2.11), transitions $\pi_{i,r}(0, b|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^*$, $b \in \{0, 1, 2\}$ contain unknown terms $\phi_{i,r}$ while other transitions do not. Note that $\pi_{i,r}(2, 0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^* = 1$, so we ignore it directly. It is immediate that only transitions starting from state 1 are kept in the likelihood and thus the partial likelihood becomes

$$L(\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda}; z,\boldsymbol{\omega}) = \prod_{i=1}^{n}\prod_{r\in\{a_i,\ldots,K-1:\ \omega_{i,r}=1\}} \pi(1, \omega_{i,r+1}|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^*. \tag{2.13}$$

Then the parameters are estimated by maximizing this partial likelihood.

Consider the capture history 0100112 as an example to illustrate the procedure above. Associated with this history is the following contribution to the full likelihood:

$$\phi_{i2}(1-p_3)\phi_{i3}(1-p_4)\phi_{i4}p_5\phi_{i5}p_6(1-\phi_{i6})\lambda_6, \tag{2.14}$$

which can also be written as

$$\pi_{i,2}(1,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^{*}\pi_{i,3}(0,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^{*}\pi_{i,4}(0,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^{*}\pi_{i,5}(1,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^{*}\pi_{i,6}(1,2|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^{*}. \tag{2.15}$$

Since the individual was not seen on occasions 2 and 3, covariates on these two occasions are missing and thus the transitions $\pi_{i,3}(0,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^{*}$ and $\pi_{i,4}(0,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^{*}$ cannot be obtained. To address this, the trinomial model deletes $\pi_{i,3}(0,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^{*}$ and $\pi_{i,4}(0,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^{*}$ directly from (2.15) so that the likelihood contribution of this individual is

$$\pi_{i,2}(1,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^{*}\pi_{i,5}(1,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^{*}\pi_{i,6}(1,2|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{z_i})^{*}. \tag{2.16}$$

2.3 Full Likelihood Model

The full likelihood approach of Bonner (2003) aims to model the missing covariates rather than deleting the components of the likelihood that depend on the missing covariates. In this approach, the CJS model is augmented by modeling the distribution of the covariate values, both missing and observed. As I assume the covariate to be continuous, individual, and time-dependent (Section 1.4), the covariate distribution should enable us to model the change of the covariate over time for the individual. To define the distribution, Bonner (2003) developed a drift process similar to the Brownian motion, which is a continuous extension of the Arnason-Schwarz model (Arnason, 1973). The complete data likelihood function is then formed by the joint distribution of the capture histories and the covariate values and, in theory, the likelihood function is constructed by integrating across the unobserved covariate values. In practice, this is too difficult and Bonner and Schwarz (2006) apply Bayesian inference via MCMC instead.

The specific process that Bonner and Schwarz (2006) suggested for modeling the covariate is based on the Wiener process following Cox and Miller (1965), which is an extension of the continuous-time random walk. The key result of the drift process is that the covariates are assumed to satisfy:

$$z_{i,t+1} - z_{i,t} \sim N(\mu_t, \sigma^2), \quad t = 1, \ldots, K-1 \tag{2.17}$$

for any individual i, i.e., $z_{i,t+1} - z_{i,t}$ is normally distributed with mean μ_t and variance σ^2. Furthermore, for any $t, r \in \{a_i, \ldots, K\}$, $z_{i,t+1} - z_{i,t}$ is independent from $z_{i,r+1} - z_{i,r}$, and $z_{i,t}$ can be written as $z_{i,t-1} + (z_{i,t} - z_{i,t-1})$, so that $z_{i,t}$ only depends on $z_{i,t-1}$. Hence, $z_{i,1}, z_{i,2}, ..., z_{i,K}$ form a Markov chain with transition kernel:

$$Z_{i,t}|Z_{i,t-1} = z_{i,t-1} \sim N(z_{i,t-1} + \mu_{t-1}, \sigma^2), \tag{2.18}$$

where μ_t is the drift parameter determining the trend of the z_{it} while σ^2 is the variance parameter determining the differences between these generated chains. It follows that $Z_{i,t+\delta}$ given $Z_{i,t}$ is actually a sum of independent normal random variables, for $\delta = 0, 1, ..., K-t$:

$$Z_{i,t+\delta} = Z_{i,t} + \sum_{r=t}^{\delta-1} (Z_{i,r+1} - Z_{i,r}) \,. \tag{2.19}$$

This implies

$$Z_{i,t+\delta} \sim N\left(z_{i,t} + \sum_{r=t}^{\delta-1} \mu_r, \delta\sigma^2\right). \tag{2.20}$$

To construct the likelihood of the full likelihood model, we first find the joint density of the vector of covariates using the drift process. For instance, the vector of covariates for capture history $\omega_i = 010011000$ is $z_i = (NA, z_{i,2}, \cdot, \cdot, z_{i,5}, z_{i,6}, \cdot, \cdot, \cdot)$. This vector can be divided into two parts, $(z_{i2}, \cdot, \cdot, z_{i5})$ and $(z_{i,6}, \cdot, \cdot, \cdot)$. The first part can be classified into the form $(z_{i,t}, \cdot, \ldots, \cdot, z_{i,t+s+1})$ and the second can be classified into the form $(z_{i,t}, \cdot, \ldots)$. For the covariate vector of the form $(z_{i,t}, \cdot, \cdot, \cdot)$, the joint density of the covariates $(\cdot, \cdot, \cdot)$ depends on $z_{i,t}$ according to the drift model. Thus it is necessary to derive the conditional distribution of covariates given the observed data $z_{i,t}$ from occasion $t+1$ to ocassion K. Following the drift process, the random vector

$\mathbf{Z_i}^* = (Z_{i,t+1}, \ldots, Z_{i,K})$ given $z_{i,t}$ follows a multivariate normal distribution with mean vector:

$$E(\mathbf{Z_i}^*|Z_{i,t} = z_{i,t}) = (z_{i,t} + \mu_t, z_{i,t} + \mu_t + \mu_{t+1}, \cdots, z_{i,t} + \sum_{r=t}^{k-1} \mu_r) \tag{2.21}$$

and covariance matrix:

$$Var(\mathbf{Z_i}^*|Z_{i,t} = z_{i,t}) = \sigma^2 \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2 & \cdots & 2 \\ 1 & 2 & 3 & \cdots & 3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 2 & 3 & \cdots & k-t \end{bmatrix}. \tag{2.22}$$

Another possibility is that missing covariates appear in the middle of a capture history, for example, $(z_{i,t}, \cdot, ..., \cdot, z_{i.t+s+1})$. In this case, the joint distribution of the missing covariates depends on both the first and the last covariates since $z_{i,t+1} - z_{i,t} \sim N(\mu_t, \sigma^2)$ and $z_{i,t+s+1} - z_{i,t+s} \sim N(\mu_{t+s}, \sigma^2)$. It is a general form of the Brownian Bridge (Revuz and Yor, 1999) so that the vector of the unknown covariates $\mathbf{Z_i}^\bullet = (Z_{i,t+1}, ..., Z_{i,t+s})$ conditional on $z_{i,t}$ and $z_{i,t+s+1}$ is also multivariate normal with mean:

$$E(\mathbf{Z_i}^\bullet|Z_{i,t} = z_{i,t}, Z_{i,t+s+1} = z_{i,t+s+1}) = \frac{1}{s+1}\begin{pmatrix} (s)(z_{i,t} + \mu_t) + (z_{i,t+s+1} - \sum_{r=t+1}^{t+s} \mu_r) \\ (s-1)(z_{i,t} + \mu_t + \mu_{t+1} + 2(z_{i,t+s+1} - \sum_{r=t+2}^{t+s} \mu_r)) \\ \vdots \\ (1)(z_{i,t} + \sum_{r=t}^{t+s+1} \mu_r) + s(z_{i,t+s+1} - \mu_{t+s}) \end{pmatrix} \tag{2.23}$$

and covariance matrix:

$$Var(\mathbf{Z_i}^\bullet | Z_{i,t} = z_{i,t}, Z_{i,t+s+1} = z_{i,t+s+1}) = \frac{\sigma^2}{s+1} \begin{bmatrix} s & s-1 & s-2 & \cdots & 1 \\ s-1 & 2(s-2) & 2(s-3) & \cdots & 2 \\ s-2 & 2(s-2) & 3(s-3) & \cdots & 3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 2 & 3 & \cdots & s \end{bmatrix}. \qquad (2.24)$$

Then these two multivariate normal distributions are used to model the distribution of missing covariates.

In comparison with the previous models, the likelihood construction of the full likelihood model has extra terms modeling the density of covariates. I will use $f(z_{i,t}|z_{i,t-1})$ to represent the density of the covariate at time t conditional on the value at time $t-1$. We still assume that the capture probability p_t is the same for all individuals on occasion t while survival probability ϕ_{it} is linked to z_{it} through the logit function in the full likelihood model. It is a challenge to fit the model via maximum likelihood because we need to integrate over the whole space of the missing covariates to compute the full likelihood function. For example, if individual i has capture history 01001 then the likelihood contribution for this individual conditional on the initial capture is

$$L_i = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_{i,2}(1-p_3)\phi(z_3)(1-p_4)\phi(z_4)p_5 f(z_3|z_{i,2}) f(z_4|z_3) f(z_{i,5}|z_4) dz_3 dz_4. \qquad (2.25)$$

This becomes even more complicated if an individual is not captured after some occasion $t < K$ in which case the individual may have died before the experiment ended.

Bonner (2003) used the complete data likelihood to avoid considering all possible unobserved transitions. In this case, the complete data set contains the capture histories, observed covariates, the missing covariates, and the known death of all individuals on each occasion. Bonner (2003) fit the full likelihood model using two methods, in the maximum likelihood framework using MCEM and in the Bayesian framework using MCMC to sample from the posterior distribution. However, MCEM does not provide direct estimates of the standard errors, meaning that further computational techniques are needed for inference. Hence, in my simulation, I fit the full likelihood model in the Bayesian framework using MCMC implemented in JAGS to obtain samples from the posterior distribution of the parameters.

2.4 Alternative Trinomial Model

As a final possibility, Burchett (2017) developed an alternative model to address the missing-covariate problem for open-population CR data by modifying the CJS likelihood to allow the truncation of capture histories. A tuning parameter $T \in \{1, 2, ..., K - 1\}$ was introduced to determine the number of occasions to truncate the likelihood contribution after each capture or recapture. The maximum value of T is $K - 1$ and all values of T build a spectrum. Moreover, Burchett (2017) showed that the alternative model includes the binomial ($T = 1$) and full likelihood models ($T = K - 1$) as special case since they fall at the two ends of the spectrum. Assumptions of the alternative model are all the same as those of the full likelihood model so that the mean rate change μ_t between occasions and the rate of variance σ^2 are considered and the likelihood function involves the contribution of the covariate values. I introduce the alternative trinomial model (the case when $T = 2$) in more details in the following paragraph.

Heuristically, the alternative trinomial model is constructed by considering three possible events when an animal is captured and released on occasion t:

(1) It is recaptured on occasion $t + 1$,

(2) It is not recaptured on occasion $t + 1$ but is recaptured on occasion $t + 2$, or

(3) It is not recaptured on occasions $t + 1$ or $t + 2$.

Note that the first event includes two capture occasions, which can be represented by the previously defined transition $\pi_{i,t}(\omega_{i,t}, \omega_{i,t+1}|\boldsymbol{p}, \boldsymbol{\beta}, \boldsymbol{z_i})$. However, the other two events include three capture occasions. Thus we need to define the new transition, $\pi_{i,t}(a,b,c|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})$, which contains three occasions, as:

$$\pi_{i,t}(a,b,c|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i}) = Pr(\omega_{i,t+2}=c,\omega_{i,t+1}=b|\omega_{i,t}=a,\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i}), \quad a,b,c \in \{0,1\}. \tag{2.26}$$

Moreover, this model needs us to start with every capture so that all the transitions from $\omega_{i,r}=0$ will be omitted. In addition, if $\omega_{i,r+1}=1$ we will no longer take $\omega_{i,r+2}$ into consideration. That is, for the alternative model, we only consider three forms of transitions, $\pi_{i,r}(1,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})$, $\pi_{i,r}(1,0,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})$, and $\pi_{i,r}(1,0,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})$, whose expressions are:

$$\begin{aligned}
&\pi_{i,r}(1,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i}) = \phi_{i,r}p_{r+1}\\
&\pi_{i,r}(1,0,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i}) = \phi_{i,r}(1-p_{r+1})\phi_{i,r+1}p_{r+2}\\
&\pi_{i,r}(1,0,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i}) = 1-\phi_{i,r}p_{r+1}-\phi_{i,r}(1-p_{r+1})\phi_{i,r+1}p_{r+2}.
\end{aligned} \tag{2.27}$$

For example, if individual i has capture history 0110010, the full likelihood is:

$$\begin{aligned}
&\pi_{i,2}(1,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})\cdot\pi_{i,3}(1,0,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})\cdot\pi_{i,5}(0,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})\cdot\pi_{i,6}(1,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})\cdot f(z_{i,3}|z_{i,2})\\
&f(z_{i,4}|z_{i,3})f(z_{i,5}|z_{i,4})\cdot f(z_{i,6}|z_{i,5})\cdot f(z_{i,7}|z_{i,6}).
\end{aligned} \tag{2.28}$$

After deleting transitions from 0, the truncated alternative trinomial likelihood is

$$\begin{aligned}
&\pi_{i,2}(1,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})\cdot\pi_{i,3}(1,0,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})\cdot\pi_{i,6}(1,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})\cdot f(z_{i,3}|z_{i,2})\cdot f(z_{i,4}|z_{i,3})\\
&f(z_{i,5}|z_{i,4})\cdot f(z_{i,6}|z_{i,5})\cdot f(z_{i,7}|z_{i,6}).
\end{aligned} \tag{2.29}$$

It is immediate that $\phi_{i,r+1}$ in $\pi_{i,r}(1,0,1|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})$ and $\pi_{i,r}(1,0,0|\boldsymbol{p},\boldsymbol{\beta},\boldsymbol{z_i})$ cannot be computed due to the missing covariate $z_{i,r+1}$. Consequently, it is necessary to model the missing covariates using the drift model for the truncated likelihood of the alternative trinomial model. Once again, I fit the alternative trinomial model in the Bayesian framework using JAGS to perform MCMC sampling and characterize the posterior distribution.

Similar to the alternative trinomial model, we can build more different models by changing the value of T. For the model with $T = m$, when an individual was captured on occasion t but not recaptured from occasion $t + 1$ to $t + m - 1$, m more occasions are considered. If the individual was recaptured on occasion r $(t + 1 \leq r \leq t + m - s)$, we truncate the transition on this occasion. In conclusion, every capture initializes a transition, which contains occasions from this capture to the next but the maximum number of occasions contained in this transition is $m + 1$. Then the product of all the transitions and the densities of covariates give the truncated likelihood. For example, when $T = 3$ for the alternative model, the truncated likelihood for the capture history 011000110 is

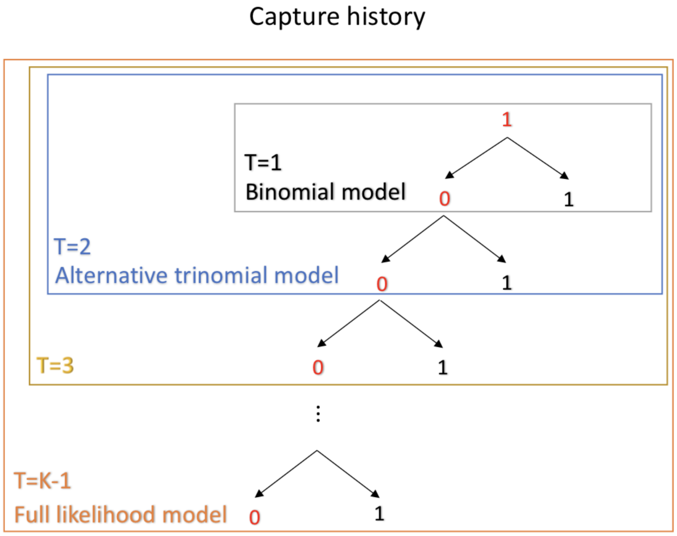
$$\begin{aligned} &\pi_{i,2}(1, 1|\boldsymbol{p},\boldsymbol{\beta}, \boldsymbol{z_i}) \cdot \pi_{i,3}(1, 0, 0, 0|\boldsymbol{p},\boldsymbol{\beta}, \boldsymbol{z_i}) \cdot \pi_{i,7}(1, 1|\boldsymbol{p},\boldsymbol{\beta}, \boldsymbol{z_i}) \cdot \pi_{i,8}(1, 0|\boldsymbol{p},\boldsymbol{\beta}, \boldsymbol{z_i}) \cdot f(z_{i,3}|z_{i,2}) \\ &f(z_{i,4}|z_{i,3}) \cdot f(z_{i,5}|z_{i,4}) \cdot f(z_{i,6}|z_{i,5}) \cdot f(z_{i,7}|z_{i,6}) \cdot f(z_{i,8}|z_{i,7}) \cdot f(z_{i,9}|z_{i,8}), \end{aligned} \tag{2.30}$$

where

$$\pi_{i,3}(1, 0, 0, 0|\boldsymbol{p},\boldsymbol{\beta}, \boldsymbol{z_i}) = Pr(\omega_{i,6} = 0, \omega_{i,5} = 0, \omega_{i,4} = 0|\omega_{i,3} = 1, \boldsymbol{p},\boldsymbol{\beta}, \boldsymbol{z_i}). \tag{2.31}$$

Figure 2.1 illustrates the association between the binomial model, the full likelihood model, and the alternative trinomial model. The likelihood contribution related to each release of a marked individual is constructed by following the tree to a terminal node - either a recapture (1) or the edge of the bounding box defined by the number of occasions considered after release. When T is equal to $K - 1$, the model is exactly the full likelihood model because once you cannot recapture an animal the entire capture history will be involved in the likelihood. If we do not consider any occasion forward when encountering $\pi_{i,r}(1, 0|\boldsymbol{p},\boldsymbol{\beta}, \boldsymbol{z_i})$, the model reduces to the binomial model. In addition, the number of the missing covariates needed to be computed for the likelihood decreases as T decreases.

Figure 2.1: The structure of the alternative model.



2.5 Scaled Logit Link Function

The goal of my thesis is to see how these different models behave when the logit link function relating survival to the covariate is replaced with a scaled logit function such that survival probability is bounded in $(0, c)$ for some $0 < c < 1$. Under the logit link function the survival probability tends to be one as z_{it} increases to infinity if covariate coefficient β_1 is positive or as z_{it} decreases to infinity if β_1 is negative. However, it may not be realistic to assume that survival probabilities approach one in some situations. For example, if an individual's weight is the covariate of interest, then it may be reasonable to assume that an individual with a larger weight has a higher survival probability, but it is unreasonable to assume that the individual will undoubtedly survive from one occasion to the next when its weight is sufficiently large. For this reason, we scale the logit link function using a specific value c $(0 < c < 1)$ such that

$$\phi^*(z_{i,t}|\boldsymbol{\beta}, c) = c\frac{\exp(\beta_0 + \beta_1 z_{i,t})}{\exp(\beta_0 + \beta_1 z_{i,t}) + 1}. \tag{2.32}$$

It follows that the scaled survival probability (2.32) varies between 0 and c.

As mentioned before, we consider the parameters in the model to be estimable if the likelihood provides strong information about the value of the parameters. In the case of MLE, we consider the parameters to be estimable if there is a single, unique setting of the parameters that maximize the likelihood. Contrary to this, the likelihood is non-estimable if we can find at least two sets of parameters which generate the same maximum likelihood value. In our specific situation, we actually find that there are infinitely many sets of parameters defined by an interval I for the scalar parameter c such that for any c in this range we can find values of the other parameters that maximize the likelihood. This is easily visualized as a flat spot in the profile likelihood computed by fixing the value of c and then maximizing over the remaining parameters.

For the Bayesian method, non-estimability implies two things: first, that the posterior distribution provides little information about the value of the parameter and second, that the shape of the posterior distribution depends heavily on the choice of prior (a situation called prior sensitivity). Consider a simple example in which there is only one parameter, θ, in the model. The posterior distribution are beliefs about the value of the parameter given the observed data and it is proportional to $\pi(\theta)L(x|\theta)$, where x represents the observed data. Hence, if the prior distribution of θ is a constant, then the posterior distribution is proportional to the likelihood $L(x|\theta)$ and has the same shape as the likelihood. Therefore, if the likelihood has a flat spot, as occurs in our case, then a flat spot will also appear on the posterior distribution, leading to the non-estimability of the parameter. Moreover, if the likelihood does not provide strong information about the parameter then the posterior distribution will change considerably when the prior is replaced by another distribution (e.g., if the uniform prior is replaced by a beta prior with parameters α, β, defined on the interval [0,1]). In problems with multiple parameters we can examine the estimability of one parameter by considering the marginal posterior distribution of this parameter. In our case, the marginal posterior distribution of c is $\pi(c|x) \propto \pi(c) \int_{\theta_{-c}} L(x|\theta)\pi(\theta_{-c})d\theta_{-c}$, where θ represents the full set of parameters and θ_{-c} the set of all parameters excluding c. Though the structures of the posterior distribution of multi and single parameter models are different, the properties of the non-estimable model like prior sensitivity and non-estimability should not be changed. If the model is estimable, then the posterior distribution should mainly depend on the likelihood rather than the prior distribution, and should provide a reasonable estimate and credible interval for any sensible specification of the prior.

Chapter 3

Results

3.1 Binomial Model

3.1.1 Theoretical results

The key result for the binomial model is that parameters are not estimable when survival is modeled via the scaled logit link function. This can be proved directly from the theoretical analysis by showing that there is an alternative set of parameters that produce exactly the same likelihood.

Theorem 3.1.1 *For the bionomial model, where we assume the survival probability* $\phi^*(z_{i,t}|\boldsymbol{\beta}, c)$ *is linked with the covariate* z_{it} *through the scaled logit link function,*

$$\phi^*(z_{i,t}|\boldsymbol{\beta}, c) = c\frac{\exp(\beta_0 + \beta_1 z_{i,t})}{\exp(\beta_0 + \beta_1 z_{i,t}) + 1}, \tag{3.1}$$

we have

$$L(\boldsymbol{\beta}, \mathbf{p}, 1; \mathbf{z}) = L(\boldsymbol{\beta}, \mathbf{p}^*, c; \mathbf{z}), \quad \textit{for any} \max_{2 \le r \le K} p_r < c \le 1, \tag{3.2}$$

where $p^*_t = p_t/c$, $t = 1, \ldots, K$. *More generally, for any values of the parameters* $\boldsymbol{\beta}$, $\boldsymbol{p}_{(1)}$ *and* *any pair* $c_{(1)}, c_{(2)} \in \left(\max_{2\le r\le K} p_r, 1\right)$*, we can find another set of parameters* $\boldsymbol{\beta}$*,* $\boldsymbol{p_{(2)}}$ *such that*

$$L(\boldsymbol{\beta}, \mathbf{p_{(1)}}, c_{(1)}; \mathbf{z}) = L(\boldsymbol{\beta}, \mathbf{p_{(2)}}, c_{(2)}; \mathbf{z}). \tag{3.3}$$

Hence, the binomial model is not estimable under the scaled link function.

Proof Recall from Section 2.1 that the likelihood for the binomial model using the logit link ignores all transitions starting from 0. Hence, there are only two kinds of transitions left, which are

$$\begin{aligned}
&\pi_{i,r}(1,0|\boldsymbol{p},\boldsymbol{\beta},1) = \phi^*(z_{i,r}|\boldsymbol{\beta},1)(1-p_{r+1}) + (1-\phi^*(z_{i,r}|\boldsymbol{\beta},1)) = 1-\phi^*(z_{i,r}|\boldsymbol{\beta},1)p_{r+1}\\
&\pi_{i,r}(1,1|\boldsymbol{p},\boldsymbol{\beta},1) = \phi^*(z_{i,r}|\boldsymbol{\beta},1)p_{r+1}.
\end{aligned} \tag{3.4}$$

Under the scaled logit link function (3.1), these two transitions can be written as:

$$\begin{aligned}
\pi_{i,r}(1,0|\boldsymbol{p},\boldsymbol{\beta},1) &= 1-\phi^*(z_{i,r}|\boldsymbol{\beta},1)p_{r+1}\\
&= 1-\frac{\exp(\beta_0+\beta_1 z_{i,t})}{\exp(\beta_0+\beta_1 z_{i,t})+1}p_{r+1}\\
&= 1-c\frac{\exp(\beta_0+\beta_1 z_{i,t})}{\exp(\beta_0+\beta_1 z_{i,t})+1}\frac{p_{r+1}}{c}\\
&= 1-\phi^*(z_{i,r}|\boldsymbol{\beta},c)p^*_{r+1}\\
&= \pi_{i,r}(1,0|\boldsymbol{p^*},\boldsymbol{\beta},c)
\end{aligned} \tag{3.5}$$

$$\begin{aligned}
\pi_{i,r}(1,1|\boldsymbol{p},\boldsymbol{\beta},1) &= \phi^*(z_{i,r}|\boldsymbol{\beta},1)p_{r+1}\\
&= \frac{\exp(\beta_0+\beta_1 z_{i,t})}{\exp(\beta_0+\beta_1 z_{i,t})+1}p_{r+1}\\
&= c\frac{\exp(\beta_0+\beta_1 z_{i,t})}{\exp(\beta_0+\beta_1 z_{i,t})+1}\frac{p_{r+1}}{c}\\
&= \phi^*(z_{i,r}|\boldsymbol{\beta},c)p^*_{r+1}\\
&= \pi_{i,r}(1,1|\boldsymbol{p^*},\boldsymbol{\beta},c)
\end{aligned}$$

From equation (3.5), we can see that $\pi_{i,r}(1,0|\boldsymbol{p},\boldsymbol{\beta},1) = \pi_{i,r}(1,0|\boldsymbol{p^*},\boldsymbol{\beta},c)$ and $\pi_{i,r}(1,1|\boldsymbol{p},\boldsymbol{\beta},1) = \pi_{i,r}(1,1|\boldsymbol{p^*},\boldsymbol{\beta},c)$. Hence, the partial likelihoods $L(\boldsymbol{\beta},\mathbf{p},1;\mathbf{z})$ and $L(\boldsymbol{\beta},\mathbf{p}^*,c;\mathbf{z})$ which are defined by equation (2.8) and constructed by $\pi_{i,r}(1,b|\boldsymbol{p},\boldsymbol{\beta},1)$ and $\pi_{i,r}(1,b|\boldsymbol{p^*},\boldsymbol{\beta},c)$ are also the same. Moreover, the capture probability should lie between 0 and 1, so $\boldsymbol{p}/c$ is required to be less than 1 (i.e., $\max_{2\le r\le K} p_r < c < 1$). Note that the likelihood is conditional on the first capture thus p_1 is not involved in the likelihood. Therefore, all parameters $p_r^* = \boldsymbol{p}/c$ and $\boldsymbol{\beta}, c$ where $c \in \left(\max\limits_{2\le r\le K} p_r, 1\right)$ result in the same likelihood. More generally, for any set of parameters $(\boldsymbol{\beta}, \boldsymbol{p}_{\mathbf{(1)}}, c_{(1)})$ or $(\boldsymbol{\beta}, \boldsymbol{p}_{\mathbf{(2)}}, c_{(2)})$ both satisfying the relatinships in equation (3.5), we have

$$L(\boldsymbol{\beta},\mathbf{p}_{(\mathbf{1})},c_{(1)};\mathbf{z}) = L(\boldsymbol{\beta},\mathbf{p},1;\mathbf{z}) = L(\boldsymbol{\beta},\mathbf{p}_{(\mathbf{2})},c_{(2)};\mathbf{z}).$$

Hence, the binomial model is not estimable under the scaled logit link function.

3.1.2 Analysis of a simulated data set

To numerically confirm the Theorem 3.1.1, I conducted the analysis of a single simulated data set. I simulated a CR data set for a total of $n = 500$ individuals over $K = 6$ capture occasions using the following setting of parameters:

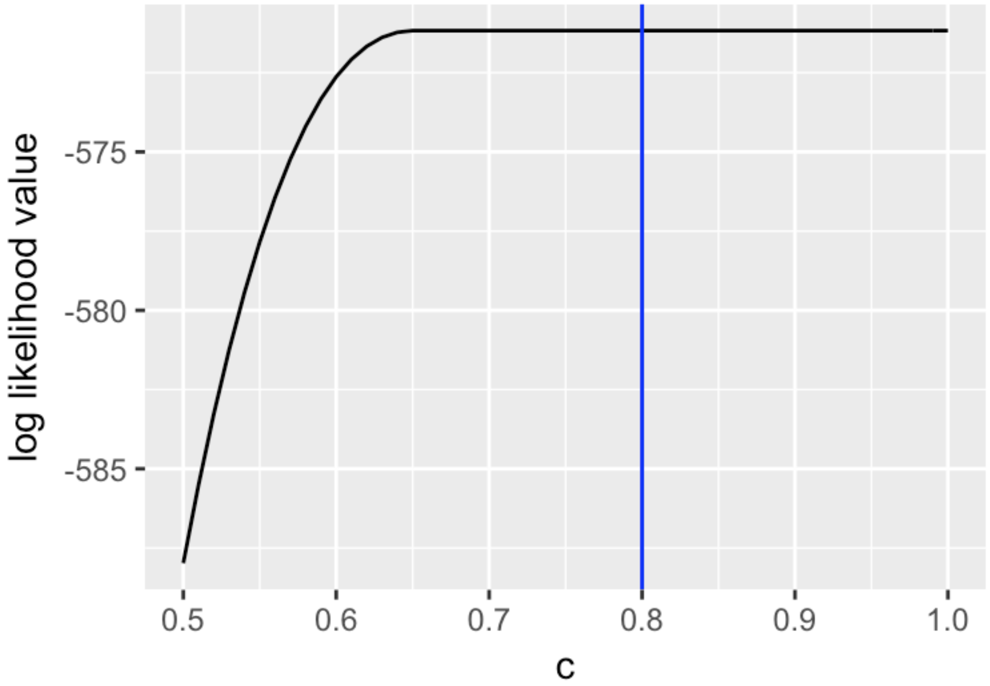
$$\begin{aligned}
&\beta_0 = 2, \quad \beta_1 = 2 \\
&c = 0.8 \\
&p = 0.8 \\
&(\mu_1,\mu_2,\mu_3,\mu_4,\mu_5) = (0,0,0,0,0) \\
&\sigma = 0.1 \\
&\mu_0 = 0, \quad \sigma_0 = 1
\end{aligned} \tag{3.6}$$

where I assume the capture probability p is the same on all occasions for simplicity of implementation. I first generated the initial covariates for the 500 individuals with the noraml distribution of mean μ_0 and variance σ_0^2, and then used the mean rate vector $\boldsymbol{\mu}$ and variance σ^2 to generate 500 individual covariate vectors with length K. For each individual, the survival probabilities were computed for the first $K - 1$ covariate values, given β_0 and β_1. Then I sampled D_i which was the last occasion before the individual died. Finally, the capture history for each individual was generated based on the capture probability p but if the individual died before occasion K then the capture history from the death occasion to occasion K was set as 0. Then the binomial model was fit to the data repeatedly with values of fixed c. That is, I fixed the value of c first, and then maximized the likelihood function to get MLEs of the other parameters including p, β_0, and β_1.

The analysis of this simulated data set exactly reproduced what we expected based on the theoretical results. In Figure 3.1, when c increases from 0.5 to 0.65, the maximum log-likelihood under fixed c increases and then remains unchanged at the maximum value (-571.173). The appearence of the flat spot confirms that the binomial model is not estimable under the scaled link function because all c's and other corresponding parameters in the flat spot can generate the same maximum likelihood value.

The appearance of the growth of the log likelihood value at the beginning is due to the truncation of parameters and Figure 3.2 explains the phenomenon in a more visual way. The blue curve illustrates how the estimate of p varies with the change of c while the red curve describes the product of $\hat{p}$ and c. Let $\hat{p}(1)$ denote the estimate of p under the logit link function. From the theoretical analysis, we know that $\hat{p}$ is related to c (i.e., $p(c)$), and is equal to the $\hat{p}(1)/c$ for $\hat{p}(1) < c < 1$. The red curve confirms this result since $\hat{p}(1) = 0.65$ and $\hat{p}(c)$ remains a constant from $c = 0.65$ to 1. Therefore, when $c \leq \hat{p}(1)$, $\hat{p}(c) = \hat{p}(1)/c > 1$, so $\hat{p}(c)$ is truncated at 1. This means that the profile likelihood is below the maximum when $c < 0.65$, but as c increases from 0.5 to 0.65, the reduced value for $\hat{p}(c)$ (i.e., $\hat{p}(1)/c - 1$) decreases, thus the log likelihood increases in this interval.

Figure 3.1: The log-likelihood value of the Binomial model under the scaled link function with the scalar c. The black line is the profile log-likelihood value as a function of c for the Binomial model. The blue line is the true value of c ($c = 0.8$).

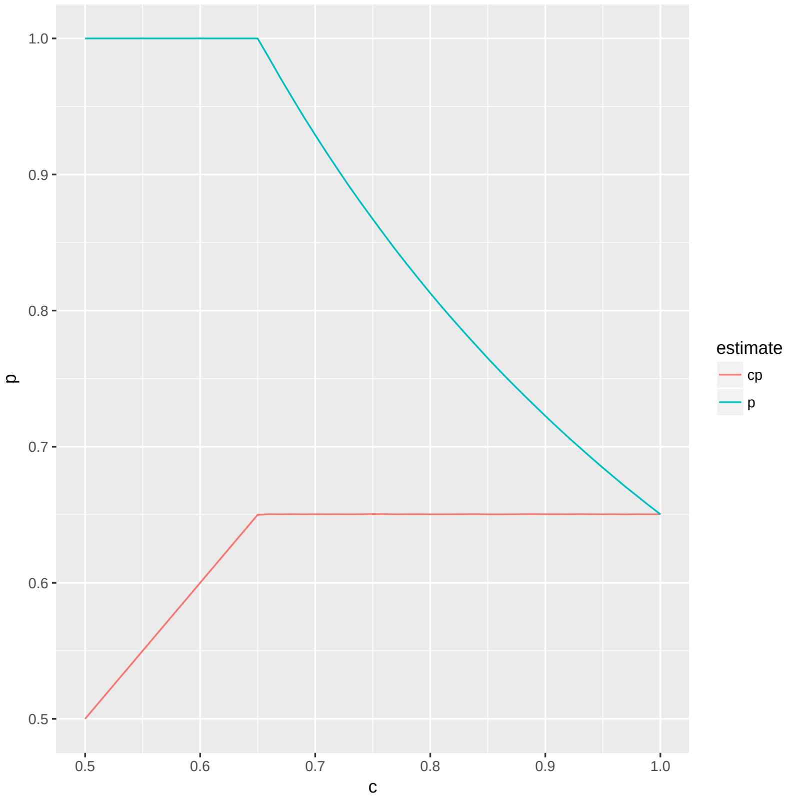


3.2 Trinomial Model

The complexity of the trinomial model makes it difficult to explore the estimability mathematically, and so we only examine the model by simulation. I simulated a new data set (MRR data) for a total of $n = 500$ individuals over $K = 6$ occasions using the same parameter values as in the above simulation but adding a constant recovery probability $\lambda = 0.4$. Also, we assume that the capture probability p and the recovery probability λ are the same on all occasions for ease of implementation. Then the trinomial model was fit to the data repeatedly with values of fixed c like we did in the binomial model. Then this model was fit to the data again but c served as a parameter to be estimated.

The analysis of the simulated data in this case showed that the trinomial model is estimable, which can be seen in Figure 3.3. The black line in Figure 3.3 shows how the maximum log-likelihood value changes as the value of c changes. The solid vertical red line ($\hat{c}$=0.88) is plotted at the maximum likelihood estimate of c when c is regarded as an unknown parameter. The

Figure 3.2: The estimates of capture probability p of the Binomial model under the scaled link function with the scalar c. The blue line is the estimate of p as a function of c for the Binomial model. The red line is the product of the estimate of p and c.

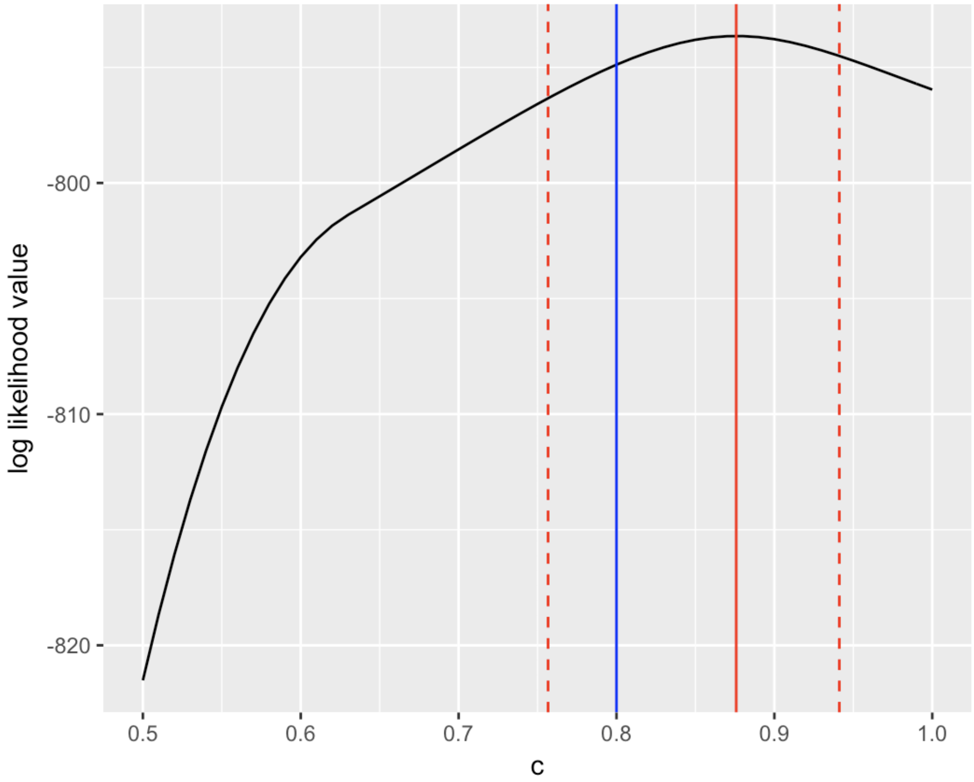


two dashed lines indicate the 95% confidence interval (0.76 0.94) of $\hat{c}$. When c rises from 0.5 to 0.88, the log-likelihood value increases from -821.51 to -793.65. Then the log-likelihood value decreases as c increases from 0.88 to 1. The biggest value among all log-likelihood values is -793.65 and the corresponding fixed c is 0.88, exactly equal to the maximum likelihood estimate $\hat{c}$ when c is treated as an unknown parameter to be estimated. That is to say, $\hat{c}$ is the only value of c that generates the maximum likelihood value. Hence, the trinomial model is estimable under the scaled link function.

3.3 Alternative Trinomial Model and Full Likelihood Model

One challenge with the trinomial model is that it requires the recovery of dead individuals, which is not always possible. Here we show that the alternative trinomial model is also es-

Figure 3.3: The log-likelihood value of the Trinomial model under the scaled link function with the scalar c. The black line is the profile log-likelihood value as a function of c for the Trinomial model. The blue line is the true value of c ($c = 0.8$). The red solid line is the mle of c. The two dashed red lines indicates the 95% confidence interval of $\hat{c}$.



timable. In this section, the estimability of the binomial model, the alternative trinomial model, and the full likelihood model are analyzed through Bayesian inference. As with the trinomial model, we can only assess the estimability of the alternative model by the simulation study due to the complexity of the likelihood. Three models: the binomial model, the alternative trinomial model, and the full likelihood model were fit to the data where c is an unknown parameter to be estimated through the Bayesian inference implemented by MCMC sampling. Following the same setting of parameters used for the simulation under the binomial model, we simulated a data set for a total of $n = 500$ individuals over $K = 6$ capture occasions. For each model, in order to ensure the accuracy of the inference, the burn-in length and MCMC sample size are

10000 and 100000, respectively. The prior distributions of all parameters are set as folllows:

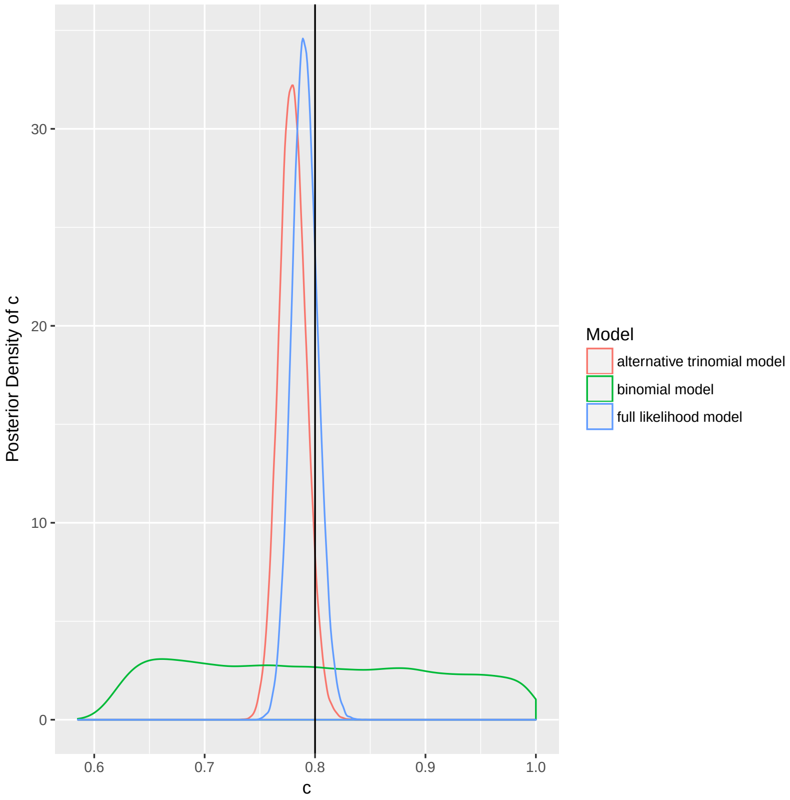
$$
\begin{aligned}
&c \sim U[0,1] \\
&p \sim U[0,1] \\
&\mu_t \sim N(0,100^2) \\
&\sigma = 1/\tau, \quad \tau \sim \text{Gamma}(0.001,0.001), \quad t = 1,...,K-1 \\
&\beta_0 \sim N(0,10^2) \\
&\beta_1 \sim N(0,10^2).
\end{aligned}
\tag{3.7}
$$

Then the posterior distributions of c under the three models can be obtained, from which I can assess the estimability.

As expected, the posterior distributions of c show the estimability of three models. The appearance of the flat spot of the binomial model under the scaled link illustrates the posterior distribution provides limited information about the parameter when the prior is uniform which is consistent with the result of the non-estimable model discussed in Section 2.5. In contrast, based on Figure 3.4 we can see from the posterior distributions of the alternative trinomial model and the full likelihood model that c is highly likely to be in the interval $(0.75, 0.82)$ because 99% mass of the distributions lies below 0.82 and above 0.75. Hence, these explain the binomial model is not estimable while the alternative trinomial and the full likelihood model are estimable under the scaled logit link.

Compared with the full likelihood model, the alternative trinomial model has less runtime, and can estimate c accurately since the point estimate of c is 0.78 and the 95% interval estimate is (0.76,0.80). The point estimate of c of the full likelihood model is 0.79 and the 95% interval estimate is (0.77,0.81), which is almost the same with the alternative trinomial.

Figure 3.4: Posterior density of c for three models based on Bayesian inference. The black line is the true value of c ($c = 0.8$).

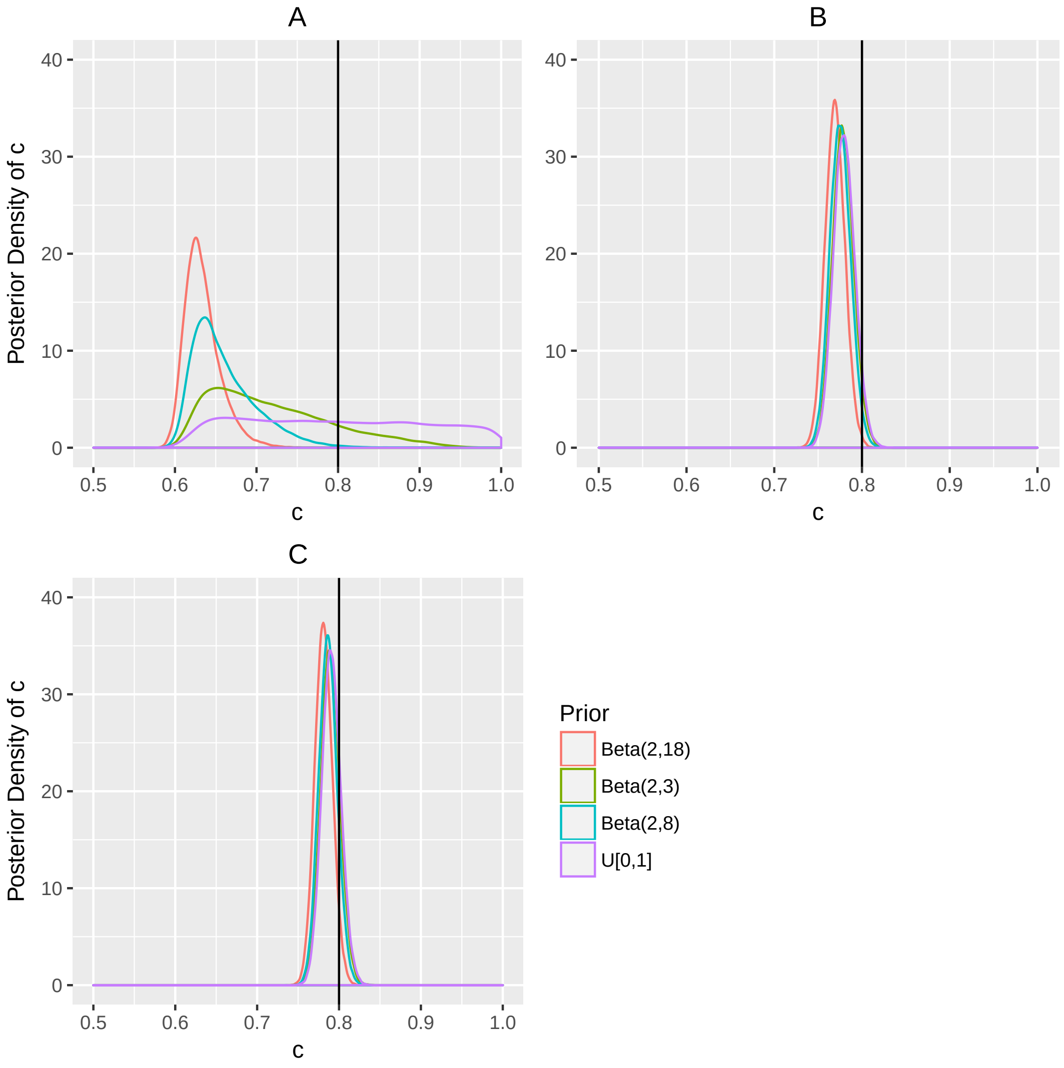


3.4 Prior Sensitivity

Another way to illustrate the non-estimability of the binomial model is to consider prior sensitivity in a Bayesian framework. The non-estimable model is sensitive to the change of the prior distribution according to our analysis in Section 2.5. To show this, I continued to use the data set in the last section and parameter settings but changed the prior distribution of c to the Beta distribution. Then I ran the binomial model, the alternative trinomial model, and the full likelihood model using three different priors, Beta$(2, 3)$, Beta$(2, 8)$ and Beta$(2, 18)$. Figure 3.5 combines the results from Beta priors and the result from the uniform prior, which is actually a Beta distribution with parameters $\alpha = \beta = 1$, indicating the prior sensitivity of the binomial model.

As the prior distribution changes, the posterior distribution of a non-estimable model changes dramatically while that of an estimable model is almost unchanged. In Figure 3.5, when the second parameter β is larger, the posterior density curve of c for the Binomial model is thinner and higher, consistent with the characteristics of the beta distribution. However, for the alternative trinomial and full likelihood models, the posterior density curves change very little although the prior distribution of c are changed a lot. So we still can determine the estimate of c. This further indicates that the binomial model with the scaled logit link cannot be used reliably in practice.

Figure 3.5: Posterior density of c of different prior distributions for three models: the binomial model (A), the alternative trinomial model (B), and the full likelihood model (C). The black line is the true value of c (c = 0.8).



Chapter 4

Conclusion

In this thesis, I have investigated the estimability of the binomial model, trinomial model, alternative trinomial model, and the full likelihood model when the scaled logit link function is used for estimating population parameters with individual, continuous, and time-dependent covariate. The estimability of models can be checked under the frequentist (MLE) and the Bayesian paradigm. The likelihood of the binomial model is simple to construct and so I have mathematically investigated the behaviour of the MLE for this model. Since the standard MLE technique does not work for the models with unknown covariates, i.e., the alternative trinomial and full likelihood models, I consider a Bayesian approach for these models. In addition, one may be curious about the difference between the estimability and the identifiability. The distinction between identifiability and estimability is essentially determined by whether or not a parameter can never be estimated or can be estimated for some data sets but not others. Non-identifiability says that there are different combinations of the parameters that produce the same distribution of the data (i.e., $f(x|\theta)$). Note that there is nothing about the observed data in this statement. For the estimability, on the other hand, it depends on the configuration of the data. In MLE, it occurs when there are different values of the parameters that produce the same maximum likelihood for the specific data that is observed. In Bayesian framework, non-identifiability occurs when the posterior distribution cannot provide strong information about the value of parameters and is sensitive to the change of the prior distribution. The binomial model and the trinomial model use the reduced likelihood not the full likelihood, and thus have nothing to do with the identifiability. The alternative trinomial model and the full likelihood model are implemented in the Bayesian framework where a sample is obtained from the posterior distribution $\pi(\theta|x)$ of the parameters and a density plot of the sample is made to see how parameters of interest are distributed along with the domain of definition. That is, the posterior distribution $\pi(\theta|x)$ only determine whether or not the estimate can be estimated and we cannot conclude anything about the likelihood. Therefore, I explore the estimability of the four models, not the identifiability.

The non-estimability of the binomial model under the scaled logit link function is proved by theoretical analysis and confirmed through the simulation analysis. Theorem 3.1.1 shows that the reduced binomial likelihood values generated by two different sets of parameters are equal, indicating the non-estimability of the binomial model. As for the simulation analysis, the non-estimability behaves as the curve of the profile likelihood value at different fixed scaler c has a flat spot, which is consistent with the result of the theoretical analysis. Moreover, the binomial model is sensitive to the prior distribution, which is an important characteristic of non-estimable models under the Bayesian framework. When the prior distribution of c changes, significant changes have taken place in the shape of the posterior distribution. It is very noteworthy that when the prior distribution of c is the uniform distribution on [0,1], the posterior distribution of c also has a flat spot resulting in the parameters cannot be estimated and we cannot ascertain the location of the true parameter in the wide region of the parameters.

The trinomial model is still estimable when the scaled logit link function is used, which can be seen from the curve of the likelihood value at different values of c. The likelihood value on the curve increases until c reaches the estimate of c and then decreases. This indicates that, no other set of parameters can generate the same likelihood as the set of true parameters, thus the model remains estimable under the scaled logit link function.

The alternative trinomial model and the full likelihood model are both estimable under the scaled link function. Most mass of these two posterior distributions lies in the interval (0.76, 0.81), which offers a precise estimate of c. Furthermore, changing the prior distribution hardly affects the distribution of the posterior distribution of c. Therefore, the parameters of the alternative trinomial model are estimable and there is no prior sensitivity for the posterior distribution. Both show the alternative trinomial is estimable under the scaled logit link function.

The trinomial, alternative trinomial, and full likelihood models are all alternatives when the binomial model is not estimable under the scaled logit link function. The trinomial model is based on the trinomial distribution whose likelihood is constructed by omitting transitions that we are unable to compute. The construction of the likelihood of the alternative trinomial model includes part of the unknown covariates because transitions starting from 0 are deleted and transitions starting from 1 also contain the missing covariates. To solve this problem, the drift model is proposed to model the missing covariates. Then we perform the alternative trinomial model through MCMC sampling. Compared with the trinomial model, the alternative model is more time-consuming because unknown covariates are involved. Moreover, the burn-in length and length of the MCMC chain are required to be large enough to ensure the accuracy of inference, which can be improved by the parallel computing but still can be time-consuming. The biggest issue of the trinomial model is that we need to check whether each animal is still alive between occasions but some species of animals are not easy to observe. So collecting the MRR data is difficult and time-consuming. As for the estimation accuracy, from the simulation results, the estimate $\hat{c} = 0.88$ (CI: (0.76, 0.94)) provided by the trinomial model is not as accurate as those provided by the full likelihood model ($\hat{c} = 0.79$, CI: (0.77, 0.81)) and the alternative trinomial model ($\hat{c} = 0.78$, CI: (0.76, 0.80)). The full likelihood model is almost the same as the alternative trinomial model except for that the full likelihood model incorporates all the unknown covariates. For this reason, the full likelihood model is complicated and more time-consuming but provides similar results, compared with the alternative trinomial model. Therefore, for the scaled logit link function, I recommend use of the alternative trinomial model.

In future work, I hope to explore other link functions to allow more flexibility in the relationships between survival and covariate. This is because, the scaled logit link function only allows the relationship to be monotonic but this can be unrealistic in some situations. For example, the survival probability may first increase and then decrease, as the covariate increases. In addition, in Section 2.4, I mentioned that MCEM failed when I wanted to obtain the mles of parameters for the full likelihood model. Therefore, I hope to find another technique to obtain the estimates of parameters through the MLE.

Bibliography

Abraham, W. T. and Russell, D. W. (2004). Missing data: a review of current methods and applications in epidemiological research. *Current Opinion in Psychiatry*, 17(4):315–321.

Arnason, A. N. (1973). The estimation of population size, migration rates and survival in a stratified population. *Researches on population ecology*, 15(2):1–8.

Bonner, S. and Schwarz, C. (2006). An extension of the cormack–jolly–seber model for continuous covariates with application to Microtus pennsylvanicus. *Biometrics*, 62(1):142–149.

Bonner, S. J. (2003). Continuous, individual, time-dependent covariates in the cormack-jolly-seber model. Master's thesis, Simon Fraser University. `www.stat.sfu.ca/alumni`.

Burchett, W. (2017). *Improving the Computational Efficiency in Bayesian Fitting of Cormack-Jolly-Seber Models with Individual, Continuous, Time-Varying Covariates*. PhD thesis, University of Kentucky. `https://uknowledge.uky.edu/statistics_etds/27/`.

Catchpole, E. A., Fan, Y., Morgan, B. J., Clutton-Brock, T., and Coulson, T. (2004a). Sexual dimorphism, survival and dispersal in red deer. *Journal of Agricultural, Biological, and Environmental Statistics*, 9(1):1–26.

Catchpole, E. A., Morgan, B. J., and Coulson, T. (2004b). Conditional methodology for individual case history data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(1):123–131.

Catchpole, E. A., Morgan, B. J. T., and Tavecchia, G. (2008). A new method for analysing discrete life history data with missing covariate values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2):445–460.

Chapman, D. G. (1951). Some properties of hyper-geometric distribution with application to zoological census. *University of California Publications Statistics*, 1:131–160.

Chartier, M. J., Dart, A., Tangri, N., Komenda, P., Walld, R., Bogdanovic, B., Burchill, C. A., Koseva, I., McGowan, K.-L., and Rajotte, L. (2015). *Care of Manitobans Living With Chronic Kidney Disease*. Manitoba Centre for Health Policy, College of Medicine, Faculty of Health Science. `http://deslibris.ca/ID/248841`.

Cormack, R. (1964). Estimates of survival from the sighting of marked animals. *Biometrika*, 51(3/4):429–438.

Cox, D. and Miller, H. (1965). The theory of stochastic processes, methuen & co. *Ltd, London, UK*.

Darroch, J. N. (1958). The multiple-recapture census: I. estimation of a closed population. *Biometrika*, 45(3/4):343–359.

Jeff Laake, Devin Johnson, P. C. e. f. s. f. J. R. (2018). *Mark-Recapture Analysis for Survival and Abundance Estimation*. R Foundation for Statistical Computing. `https://cran.r-project.org/web/packages/marked/index.html`.

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

Knape, J. and Korner-Nievergelt, F. (2015). Estimates from non-replicated population surveys rely on critical assumptions. *Methods in Ecology and Evolution*, 6(3):298–306.

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

Lincoln, F. C. et al. (1930). Calculating waterfowl abundance on the basis of banding returns.

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.

Pearson, C. (1896). The yearly immigration of young plaice into the Limfjord from the german sea, etc. *Rept. Danish Biol. Sta. for 1895*, 6:1–48.

Petersen, C. G. J. (1896). The yearly immigration of young plaice in the limfjord from the german sea. *Rept. Danish Biol. Sta.*, 6:1–48.

Pollock, K. H. (1974). *The assumption of equal catchability of animals in tag-recapture experiments*. PhD thesis, Cornell University, June. `https://newcatalog.library.cornell.edu/catalog/307906`.

Pollock, K. H., Hines, J. E., and Nichols, J. D. (1984). The use of auxiliary variables in capture-recapture and removal experiments. *Biometrics*, 40(2):329–340.

Pollock, K. H., Nichols, J. D., Brownie, C., and Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife monographs*, (107):3–97.

Poole, D. (2002). Bayesian estimation of survival from mark-recapture data. *Journal of Agricultural, Biological, and Environmental Statistics*, 7(2):264.

Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion Springer*, volume 293 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg, 3 edition.

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

Sólymos, P., Lele, S., and Bayne, E. (2012). Conditional likelihood approach for analyzing single visit abundance survey data in the presence of zero inflation and detection error. *Environmetrics*, 23(2):197–205.

White, G. C. and Burnham, K. P. (1999). Program mark: survival estimation from populations of marked animals. *Bird study*, 46(sup1):S120–S139.

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2003). Analysis and management of animal populations. 67(3):654–655.

Curriculum Vitae

Name: Jiaqi Mu

Post-Secondary Education and Degrees:
University of Western Ontario
London, ON
2017 - Present M.Sc. candidate

South China University of Technology
Guangzhou, China
2013 - 2017 B.Sc.

Related Work Experience:
Teaching Assistant
The University of Western Ontario
2017 - 2018