Methods for estimating occupancy

by Natalie Karavarsamis

Submitted in total fulfilment of the requirements
of the degree of Doctor of Philosophy

Supervisors: Dr Andrew Robinson, Professor Richard Huggins,
Associate Professor Ray Watson and Associate Professor
Graham Hepworth

Department of Mathematics and Statistics
Faculty of Science
University of Melbourne
Victoria 3010
Australia

August 2014

Abstract

The estimation of the probability of occupancy of a site by a species is used to monitor the distribution of that species. Occupancy models have been widely applied and several limitations have been identified. In this thesis we resolve some of these. In particular we focus on limitations of maximum likelihood estimators and the associated interval estimators, and the difficulties associated with the extension from linear to generalised additive models for the relationship between occupancy and covariates. Initially we consider in detail the basic occupancy model which includes two parameters: ψ and p. Our primary concern is the probability that the species occupies a particular site, ψ. The other parameter, the detection probability p, is a nuisance parameter. We first derive the joint probability mass function for the sufficient statistics of occupancy which allows the exact evaluation of its mean and variance, and hence its bias. We show that estimation near the boundaries of the parameter space is difficult. For small values of detection, we show that estimation of occupancy is not possible and specify the region of the parameter space where maximum likelihood estimators exist, and give the equations for the MLEs in this region. We next demonstrate that the asymptotic variance of the estimated occupancy is underestimated, yielding interval estimators that are too narrow. Methods for constructing interval estimators are then explored. We evaluate several bootstrap-based interval estimators for occupancy. Finally, instead of the full likelihood we consider a partial likelihood approach. This gives simple closed form estimators in a basic model with only a small loss of efficiency. It greatly simplifies the inclusion of linear and nonlinear covariates by allowing the use of standard statistical software for GLM and GAM frameworks and in our simulation study there is little loss of efficiency compared to the full likelihood.

Declaration

This is to certify that:

1. *the thesis comprises only my original work towards the PhD except where indicated in the Preface,*
2. *due acknowledgment has been made in the text to all other material used,*
3. *the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices*

Natalie Karavarsamis

Preface

This thesis is submitted to the University of Melbourne in support of my application for admission to the degree of Doctor of Philosophy. No part of it has been submitted in support of an application for another degree or qualification of this or any other institution of learning. Chapter 2 is based on joint work with Associate Professor Ray Watson that is being prepared for publication. Chapter 3 is a modification of Karavarsamis et al. (2013) which has been published in the Australian and New Zealand Journal of Statistics. Chapter 4 is based on joint work with Professor Richard Huggins that is being prepared for publication.

Acknowledgments

First, I acknowledge my supervisors, Ass. Prof. Ray Watson, Prof. Richard Huggins, Dr Andrew Robinson and Ass. Prof. Graham Hepworth for their ongoing support. It was a privilege to work with them. I am grateful to Prof. Huggins and Ass. Prof. Watson, who as eminent researchers and educators gave me the opportunity to learn from their extraordinary knowledge and vast experience.

Next, I acknowledge the support of the Department of Mathematics and Statistics at the University of Melbourne, the broader University community, as well as funding provided by Parks Victoria.

In closing, I would like to thank the ongoing support of my family and friends, for their loving care, which included nourishing, nurturing and comforting home-cooked meals.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Statistical inference on abundance and dispersion, the distribution of an animal or plant population, has been identified as critical for the protection of animal populations and for ecosystems, in general (Hoffman et al., 2010). Historically, statistical methods have focussed on estimating abundance of species, for example, the body of work including Burnham and Overton (1978), Burnham and Overton (1979), Burnham et al. (1987), Pollock (2002), etc. A comprehensive review is given in Chao (2001). Methods for estimating abundance generally involve intensive sampling plans, such as capture-recapture sampling, which are usually possible only on small areas owing to the high costs involved. Furthermore, these studies are time consuming and labour intensive. As a result, there has been a shift in focus from estimating the abundance of a species to monitoring the dispersion of populations, possibly on larger scales (Pollock, 2002).

Often it is the dispersion of a species rather than its abundance that is of interest and this shift has led to the development of alternative statistical methods that meet changed objectives, such as monitoring various population characteristics.

Monitoring the dispersion of populations provides insights into the behaviour of species and how they interact with their environment. To this end, large-scale studies, which may assist in conservation planning, are often conducted. These studies can yield habitat suitability maps, which assist in the understanding of biogeographical issues, such as invasive species biology, and trends

in population dispersion; they can give the ecological community insights into the preferences that a species displays for a particular habitat and add to the broader understanding of ecosystems (Hoffman et al., 2010; MacKenzie et al., 2011). Sound management of wildlife populations is an important consideration in monitoring species. Inaccurate assessments and inferences about the species population and their habitat could lead to detrimental decision making practices. The misclassification of occupancy status is an example which will result in such errors (Binns et al., 2000; Lindenmayer and Burgman, 2005; Miller et al., 2011).

In contrast to abundance estimation, which was developed from capture, mark then recapture of individual animals, studies for dispersion require only the observance of the presence and absence of the species. Thus, studies of dispersion are less intensive, which means they are less time consuming, and therefore less costly (MacKenzie et al., 2002; Royle and Nichols, 2003; Wintle et al., 2005).

The remainder of this chapter is laid out as follows. Section 1.2 contains a description of data that lend themselves to occupancy models and introduces the data that will be used in applications throughout this thesis. This is followed by section 1.3, where methods for occupancy models are reviewed. A description of relevant methods is provided together with their historical accounts. Finally, in Section 1.4, the chapter ends with a summary of known problems and how these will be addressed in this thesis.

1.2 Data

Presence-absence data are collected by recording detections and non-detections on the sampling units (or sites). When a site is visited the species is either seen or it is not and the event is recorded with, for example, 1 (seen) or 0 (not seen) for each site on every survey occasion, i.e., for each site-occasion.

In the homogeneous case it is assumed that there is no migration, deaths or births, at the site level. That is, it is assumed that the population is closed at the site level, but closure is not assumed at the species level. So if the species is detected on any visit (=1), then the site is considered occupied at all visits i.e. the species is said to inhabit that site, irrespective of whether or not a detection is recorded on other visits. However, the occupancy status of the site will be inconclusive if the species has not been detected over any of the survey occasions. Two possibilities exist in this case, either the species does

not inhabit that site, or the species does inhabit that site but was not detected on the survey occasion. In many instances the data that are collected contain a high number of zeros.

In addition to recording detections of a species, it is common to record spatial and temporal characteristics. These may be used as covariates in the fitted models, although doing so is not always straightforward, particularly when fitting covariates nonparametrically.

We will analyse presence-absence data collected on a frog population from Victoria, and examine publicly available data about two fish populations. The complexities of the structures of these data motivate extensions to existing methods we present for the basic occupancy model; this is the homogenous case of the full likelihood function.

1.2.1 Frogs

We now describe the data about the endangered growling grass frog (*Litoria raniformis*). The study area encompassed the Merri Creek Corridor and adjacent catchments of Yuroke and Darebin Creeks on the northern outskirts of the Melbourne metropolitan area, around 350km^2. A detailed description is given in Heard et al. (2006).

Here we use diurnal presence-absence data that was collected on the adult frog population during the 2002-2003 season. Specifically there were 27 sites, each visited on 4 occasions with complete data i.e. no missing data for presences or absences.

Covariate information was also collected during the study. We consider water and land vegetation of sites in our models.

1.2.2 Crossbills

The methods presented in this thesis are demonstrated on data from the European crossbill (Loxia curvirostra) collected in 267 1 km^2 sample quadrats in Switzerland, 1999 (Schmidt, 2004). We analyse the data from the example presented in Fiske and Chandler (2011) that involves the site covariates elevation and forest. The models are fitted to detections from the first year of the study, in 1999, which comprise three survey occasions (or visits to each site).

1.2.3 Fish

We use publicly available data on two fish populations, although these data have not been published. James Peterson presents data on Coosa bass collected via electrofishing from four 50-80 m sections in streams at 54 sites in the Upper Coosa River basin, USA[1]. Covariates were the stream link magnitude and the coefficient of variation of streamflow during the spring and summer proceeding the sampling. For our purposes we considered presence of under one year old fish as presence of the species to reduce the data to a single state.

Another data set concerns detections of Brook trout collected via electrofishing in three 50 m sections of streams at 57 sites in the Upper Chattachochee River basin, USA. Information was recorded for covariates on elevation and stream mean cross-sectional area at each sample section. These smaller data sets are convenient for the evaluation of the methods developed here.

[Note: These data were collected in preparation of a senior thesis at the University of Georgia (Athens, USA) they have not been published but are publicly available and we have obtained these with written consent from Jim Peterson[2].]

1.3 Methods

The estimation of dispersion of a species is based on aggregated population measures. One such measure involves estimating the probability that a species occupies a discrete sampling unit, such as a patch of habitat. Simply put, we want to estimate occupancy of a species. However, estimation of occupancy is complicated by unreliable detection. The problem then becomes one of jointly estimating the probability of occupation (or occupancy), ψ, also referred to as the probability of presence, and the probability of detection or detectability, p. In this context, the detection probability can be treated as a nuisance parameter: it is not itself of primary interest, but it must be estimated in order to estimate the probability of occupancy.

Related studies have developed methods which may be readily adapted for simultaneously estimating the probabilities of detectability and occupancy (see, e.g., Geissler and Fuller, 1987; Azuma et al., 1990; Pereira and Itami, 1991; Hall, 2000; Hanski, 1994; Bayley and Peterson, 2001; MacKenzie et al., 2002;

[1] http://people.oregonstate.edu/~peterjam/occupancy_workshop/hands_on.html
[2] http://people.oregonstate.edu/~peterjam/

Nichols and Karanth, 2002; Tyre et al., 2003). A class of these methods has come to be known commonly as occupancy models (Hall, 2000; MacKenzie et al., 2002; Tyre et al., 2003). These models may be described by a likelihood function. The likelihood is constructed from the number of detections and nondetections made at the sites of the study. The model differs from that used by conventional capture-recapture methods in that the capture of an individual is defined as the sighting (or detection) of the species at a site (for example Cormack, 1964; Jolly, 1965; Seber, 1965; Pollock, 2002; Huggins, 1989; Huggins and Hwang, 2007). Specifically, the *site* constitutes the *individual*, and the *capture* corresponds with the *sighting* of the species at that site. The analogy with capture-recapture models is not perfect because the individuals that are never captured do not contribute to the likelihood, whereas information is included for sites at which the species was never detected.

As mentioned earlier, the data that are collected may present a high number of zeros. The zero-inflated-binomial (ZIB) distribution which underpins occupancy likelihood models is suitable for binary data with increased zeros. The zero-inflated Poisson (ZIP) likelihood (Lambert, 1992; Welsh et al., 1996; Ridout et al., 1998) is the basis of the ZIB likelihood; it is a direct adaptation of the ZIP likelihood framework for two parameters. Some examples for the ZIB likelihood constructed in a way appropriate for occupancy models were first seen in Hanski (1994); Hall (2000); MacKenzie et al. (2002); and Tyre et al. (2003), etc. The basis for the ZIB formulation found in the MacKenzie et al. (2002) likelihood, is that there are two classes of sites, namely those that are occupied and those that are not. Royle and Nichols (2003) adopt this same basis to formulate a mixture of binomial random variables with different values of p, and mixing proportion ψ, similar to Lambert (1992) and Welsh et al. (1996). They show that for a closed population with constant detection and survey occasions, this formulation is equivalent to the MacKenzie et al. (2002) construction as a mixture of a binomial random variable conditional on occupancy, and a point mass at 0. This construction is related to the way that Huggins (1989) partitioned the full likelihood of a capture-recapture experiment into two parts: the first describes nondetections, for occupied and unoccupied sites; and the second describes detections made at sites that directly form a binomial function.

Since the basic occupancy model, in which the 'world' was viewed as static, and estimation was focussed on occupancy and detectability, the model has been extended to allow for the estimation of parameters such as colonisation

and local extinction, as well as to allow for changes in the relationship between the species population and its environment via covariates (MacKenzie et al., 2003; Royle, 2006; Nichols et al., 2007; MacKenzie et al., 2009, 2011; Guillera-Arroita et al., 2011; Welsh et al., 2013). In particular, Welsh et al. (1996) and Welsh et al. (2013) develop models for the inclusion of covariates and derive approximate standard errors, and the latter study identifies major limitations with the full likelihood function which is based on the ZIB formulation. See also Guillera-Arroita et al. (2014) who refute some of these limitations.

Here, we focus on models that estimate occupancy ψ and detection p. For example, we consider models in which the parameters are constrained so that $0 \leqslant \psi \leqslant 1$ and $0 \leqslant p \leqslant 1$.

Bayesian methods maybe be developed directly, or easily adapted, for estimating occupancy and detectability, and their standard errors, is commonplace, for example Milne et al. (1989); Lunn et al. (2000); Wintle et al. (2003); MacKenzie et al. (2006); Gimenez et al. (2007); Royle and Dorazio (2008); Gimenez et al. (2009); MacKenzie et al. (2009); Fiske and Chandler (2011); Martin et al. (2011); Aing et al. (2011); Hui et al. (2011). This approach helps to overcome known restrictions with the full likelihood approach, for example to constrain probabilities to between 0 and 1, particularly when covariates are included in models. However there are other limitations with these approaches. For example, empirical Bayes methods can underestimate the variance of the posterior distribution (Royle and Dorazio, 2008; Fiske and Chandler, 2011).

A full likelihood approach, simultaneously estimating detectability and occupancy, is of course possible, as is also when including covariates (e.g. MacKenzie et al., 2002; Fiske and Chandler, 2011; Welsh et al., 2013). However, even in the simple homogenous case we reveal problems with the standard likelihood and propose solutions. Others have identified restrictions with the full likelihood, and have evaluated its bias and measures of uncertainty (Tyre et al., 2003; Wintle et al., 2004; Guillera-Arroita et al., 2010; Welsh et al., 2013).

Sample-based uncertainty about the parameter estimates must be considered in order to provide a complete picture of a study. The full likelihood function does not have a readily available closed form solution for the estimates of detectability or occupancy. Further, the asymptotic variance for occupancy given in MacKenzie et al. (2006) may not always yield estimates. It is noted in MacKenzie et al. (2002) that the likelihood-based, large-sample standard errors may not be appropriate for estimating the uncertainty, especially when

sample sizes are small (MacKenzie et al., 2002). The observed information (or hessian) matrix is not always invertible. Consequently, MacKenzie et al. (2002) used an ordinary nonparametric bootstrap estimator, also called the nonparametric basic bootstrap, though it may not be appropriate for rare (small population size) and clustered species distributions (Efron, 1982). We verify this and explore other interval estimators in Chapter 3, the work of which has been published in Karavarsamis et al. (2013).

In Chapter 2 we examine the equations for the maximum likelihood estimators (MLEs) and identify when the score equations of the basic occupancy likelihood do not yield solutions for the ML estimates. For example the score equations will give estimates for occupancy and detection that are greater than 1. We find that the score equations do not apply on the boundary (or edges) of the sample space and we define the relationships between the sufficient statistics, the number of sites and survey occasions that yield the boundaries. We then relate the edges to the parameter space and determine when the score equations do not apply.

Furthermore, we define the region of the sample space and parameter space for which estimation is not possible, and identify a region within which the MLEs always exist and where estimates are less biased. We give a rough approximation to the convex hull for all possible solutions for the MLE.

To evaluate the bias associated with occupancy we derive the probability mass function of the sufficient statistics, which leads to an expression for the exact expectation of occupancy and for the exact variance. We evaluate the asymptotic variance against the exact variance, and evaluate the bias of the basic occupancy estimator and explore ways of correcting for it. Overall, the score equations and asymptotic variance for occupancy will perform well for large N and large T.

Often investigators will wish to include auxiliary (or covariate) information into the estimation process to better understand relationships between the species and its environment. This may include site characteristics (e.g. geographic), or species characteristics. Species characteristics may relate the probability of occupancy to the species via covariates on habitat type, patch size, gender, age etc; or relate the probability of detecting the species to weather conditions, site accessibility, detection methods etc.

When incorporating auxiliary information for the estimation of occupancy and detection into the full likelihood function the problems discussed above are even

more serious. Welsh et al. (2013) investigate this scenario, they found many problems when including covariates into the full likelihood model. Specifically, they remark on problems when estimating in some regions of the parameter space, as we investigate in Chapter 2 for the non-covariate case. In addition, limitations imposed by the full likelihood do not guarantee that the solutions to the score equations are between 0 and 1.

In summary, the full likelihood gives estimates which are highly variable, and it is difficult to include covariate information. Welsh et al. (2013) show that issues with the basic occupancy model extend to the covariate model, although some of these limitations may be due to the use of the VGAM package. Further, the full likelihood is not a generalised linear model (GLM) so the GLM machinery is not easily applicable; specialist programs are needed. Many of the limitations mentioned may be overcome by including covariate information, in that covariates assist in stabilising estimates. As well, the nature of presence-absence data that is characterised by repeated visits to a site is a source of heterogeneity which may also be overcome with covariates.

We exploit repeated visits to a site in order to consider occupancy separately to detection. In particular, detections and nondetections are recorded at each site so that there is more information on the detection probabilities. We apply partial likelihoods that are often used to simplify complex likelihoods and to deal with nuisance parameters, here the detection probabilities. This way we have full accessibility to the GLM framework, as well as its extensions.

We show, in Chapter 4, a number of approaches for incorporating covariate information into the likelihood under site inhomogeneity, i.e., when presence and detection are not the same among sites, but do remain constant across the duration of the survey, i.e., constant within sites. We show that it is possible to estimate ψ and p in two separate stages, rather than maximising the full likelihood, and how this is done by using standard statistical methods and software. The proposed two-stage method resolves the problems identified with the full likelihood, both in the site homogenous case and the site inhomogeneous case. We explore presence-absence type data and how we may use standard statistical techniques (e.g. GLM, generalised additive models (GAM) and vector generalised linear and additive models (VGLMs/VGAMs)) to analyse these (Hastie and Tibshirani, 1986; McCullagh and Nelder, 1989; Yee and Wild, 1996). We derive standard errors for the estimates of the coefficients for models with, and without, covariates. The use of standard GLM techniques (in the two-stage procedure) give computational advantages over the

full likelihood approach, and, GAMs may easily incorporate semi parametric modelling of occupancy and detection probabilities. We show that the gain in computational efficiency outweighs some small loss of efficiency in estimating the detection probabilities.

1.4 Statement of the Problem

The main problems involved with occupancy models we address, then resolve, in this thesis, are:

1. evaluate boundary estimates then give correct expressions and a plausible region for estimation
2. examine nonconvergence of the full likelihood and propose an alternative two-stage approach for partial likelihoods
3. assess the asymptotic variance and bias of occupancy
4. resolve computational limitations within modern software.

Chapter 2

The bias of the occupancy estimator $\widehat{\psi}$

2.1 Overview

Site occupancy ψ is used for monitoring species populations, as mentioned previously. Further, to accurately estimate occupancy the uncertainty associated with nondetections must also be considered. Inherently, occupancy is biased by the uncertainty of imperfect detection.

In this chapter we explore, in detail, the basic occupancy model as proposed by MacKenzie et al. (2002) with emphasis on, and motivated by, evaluating the bias of the occupancy estimator. This model is specified by two parameters: ψ and p. Our primary concern is the occupancy (probability of occupation or presence), ψ, which denotes the probability that the species occupies a particular site. The other parameter, p, is effectively a nuisance parameter: the detectability p denotes the probability that the species is observed, given that it is present at a site.

Overall, it is difficult to estimate near the boundaries of the parameter space of occupancy and detectability. Guillera-Arroita et al. (2010) give the sufficient statistics and we verify this here with a complete derivation. Guillera-Arroita et al. (2010) give the conditions for which the score equations do not apply and give solutions that yield the MLEs. We identify an additional condition ('boundary' case) and give the complete set of boundary solutions, referred to as the 'edge' solutions. The special case when the sufficient statistics are zero is addressed throughout the chapter. We derive the joint distribution

of the sufficient statistics that is necessary to derive the exact expectation, the exact variance and bias for occupancy. These allow for the evaluation of the asymptotic variance estimator given in MacKenzie et al. (2006, p. 96) for which there is no closed form and which is known to under-perform. We give the derivations for the asymptotic variance estimator for occupancy.

Let K = the number of sites at which the species was observed (at least once) and X = the total number of (site-occasion) times that the species was observed. We show that (X, K) is sufficient for (ψ, p), and derive the joint probability mass function (pmf) for (X, K). This is used to obtain an expression for the expectation of the occupancy estimator, $\mathrm{E}(\widehat{\psi})$. We evaluate the expectation of $\widehat{\psi}$ and hence its bias. We consider also a conditional expectation, conditional on $(X, K) \neq (0, 0)$, and compare the conditional and unconditional expectations.

For small values of the detection probability p, we show that estimation of ψ is unfeasible, in that the ML estimates do not exist everywhere in the parameter space. We produce a rule specifying a region where estimation of (ψ, p) is 'plausible', and within this region we give a correction for the occupancy estimator to reduce its bias.

The pmf of (X, K) is used to compute the exact variance and mean squared error (MSE) of $\widehat{\psi}$. The asymptotic variance estimate of the occupancy estimator is evaluated and found to seriously underestimate the variance, unless N, T, and p are large. This will lead to confidence intervals for $\widehat{\psi}$ that are too narrow. Furthermore, the MSE will be underestimated. Instead, we recommend that the exact standard deviation is used i.e. the exact variance.

To evaluate these results in a practical setting we use several test cases and two real data sets. These cover a range of real study sizes and test scenarios.

2.2 Definition and derivation of likelihood function

Consider a set of N sites that are each surveyed over T occasions. Let x_{ij} denote a detection indicator at site i on occasion j, i.e. if $x_{ij} = 1$ then the species is detected at site i on occasion j; and $x_{ij} = 0$ otherwise. Thus, the row vector $\mathbf{x}_i = (x_{i1}\, x_{i2}\, \ldots\, x_{iT})$ describes the site history, i.e. the detection history vector for site i, $(i = 1, 2, \ldots, N)$.

The data set $\mathbf{X} = [x_{ij}]$ is the $N \times T$ detection (or history) matrix of responses for N sites surveyed on T occasions. The matrix $\mathbf{X}$ describes the detections and nondetections of the study:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1T} \\ x_{21} & x_{22} & \dots & x_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{NT} \end{bmatrix}$$

The basic occupancy-detection model depends on the parameters ψ and p. The distribution of the random matrix $\mathbf{X}$ can be defined in terms of independent Bernoulli random variables. We let $\mathrm{B}(\theta)$ denote a Bernoulli random variable, i.e. 1 with probability θ, and 0 with probability $1 - \theta$.

Then X_{ij} can be regarded as being obtained from the following unobserved independent random variables:

- P_i = presence (occupation) at site i; $P_1, P_2, \dots, P_N$ are assumed to be independent $\mathrm{B}(\psi)$. If $P_i = 1$ then the species is present at (i.e. occupies) site i for all occasions in the study.

- D_{ij} = detection at site i on occasion j, where the D_{ij} are independent $\mathrm{B}(p)$: if $D_{ij} = 1$, then species presence would be detected at site i on occasion j. It follows that $D_{i\bullet} = \sum_{j=1}^{T} D_{ij} \stackrel{\mathrm{d}}{=} \mathrm{Bi}(T, p)$, where Bi denotes the binomial distribution consisting of T trials with probability p of a detection on any trial.

Then $X_{ij} = P_i D_{ij}$, and

$\Pr(\mathbf{X}_i = \mathbf{0}) = \Pr(P_i = 0) + \Pr(P_i = 1, D_{i\bullet} = 0) = 1 - \psi + \psi(1-p)^T$;

$\Pr(\mathbf{X}_i = \mathbf{x}_i) = \Pr(P_i = 1, D_{ij} = x_{ij}) = \psi p^{x_{i\bullet}}(1-p)^{T - x_{i\bullet}}$, for $\mathbf{x}_i \neq \mathbf{0}$.

Let $x_{i\bullet} = \sum_{j=1}^{T} x_{ij}$ be the row sums of the history matrix, i.e. the number of occasions for which the species was detected at site i; and let $x_{\bullet j} = \sum_{i=1}^{N} x_{ij}$ denote the number of sites for which the species was detected at occasion j.

We define an indicator $u_i = \mathrm{I}(x_{i\bullet} \geqslant 1)$ to determine whether the species was detected at site i on at least one occasion. Then the likelihood for site i, based on observing $\mathbf{x}_i$, is given by

$$L_i(\psi, p \,|\, \mathbf{x}_i) = \left(\psi p^{x_{i\bullet}}(1-p)^{T - x_{i\bullet}}\right)^{u_i} \left(1 - \psi + \psi(1-p)^T\right)^{1-u_i} \tag{2.1}$$

Then, assuming independence between sites, the joint likelihood is

$$\begin{aligned} & L(\psi, p \,|\, \mathbf{X}) \\ &= \prod_{i=1}^{N} \left(\psi p^{x_i \cdot} (1-p)^{T-x_i \cdot}\right)^{u_i} \left(1 - \psi + \psi(1-p)^T\right)^{1-u_i} \\ &= \left(\psi(1-p)^T + (1-\psi)\right)^{N-k} \psi^k p^x (1-p)^{kT-x} \end{aligned} \tag{2.2}$$

where $k = \sum_{i=1}^{N} u_i$ = number of sites at which some detection is made during the study; and $x = x_{\cdot\cdot} = \sum_{i=1}^{N} x_{i\cdot}$ = total number of site-occasions for which a detection is made (MacKenzie et al., 2002).

Writing the likelihood in this way is akin to Huggins (1989) who suggested partitioning the likelihood into detections (equivalent to the captures) and nondetections.

It follows from the factorisation theorem that (X, K) is sufficient for (ψ, p) for a study of N sites and T survey occasions. In other words, the conditional distribution of the data, given the statistic (X, K), does not depend on the parameters (ψ, p), so (X, K) are sufficient statistics for (ψ, p). All the information concerning (ψ, p) that is contained in the data is captured by (X, K) (Casella and Berger, 2002, p.272 & p.274), and hence $(\widehat{\psi}, \widehat{p}) = (\widehat{\psi}(x, k), \widehat{p}(x, k))$. This observation is important in what comes later as it enables us to specify the exact distribution of the MLE, and hence to evaluate the expectation and the variance of $\widehat{\psi}$. Knowing the exact distribution for X and K, enables us to explore the bias and propose a correction for it; and to assess the asymptotic approximations for the variance.

Taking logs we obtain

$$\begin{aligned} \log L = k \log(\psi) + x \log(p) + (kT - x) \log(1-p) \\ + (N-k) \log\left\{\psi(1-p)^T + (1-\psi)\right\}. \end{aligned} \tag{2.3}$$

Define $\theta = g(p) = 1 - (1-p)^T$, the probability of detecting the species in at least one of the T survey occasions at an occupied site. Then

$$\log L = k \log(\psi) + x \log(p) + (kT - x) \log(1-p) + (N-k) \log(1 - \psi\theta). \tag{2.4}$$

2.3 Maximum likelihood estimators

2.3.1 The score equations and edge solutions

The score equations have been previously given in MacKenzie et al. (2002).

Maximising the log-likelihood function gives the maximum likelihood estimates (MLEs) $\widehat{\psi}$ and $\widehat{p}$. The score equations for the log-likelihood are

$$\frac{\partial \log L}{\partial \psi} = \frac{k}{\psi} - \frac{(N-k)\theta}{1-\psi\theta}, \tag{2.5}$$

$$\frac{\partial \log L}{\partial p} = \frac{x}{p} - \frac{(kT-x)}{(1-p)} - \frac{(N-k)\,\psi\,T(1-p)^{T-1}}{(1-\psi\,\theta)}, \tag{2.6}$$

recalling that $\theta' = g'(p) = T(1-p)^{T-1}$.

Equating (2.5) to zero and solving for ψ we obtain

$$\widehat{\psi} = \frac{k}{N\,\widehat{\theta}}\,. \tag{2.7}$$

Since $(1-p)^{T-1} = \dfrac{1-\theta}{1-p}$, the second score Equation (2.6) can be re-written as follows:

$$\begin{aligned}\frac{\partial \log L}{\partial p} &= \left(\frac{x}{p} - \frac{kT-x}{1-p}\right) - \left(\frac{(N-k)\theta}{1-\psi\theta}\right)\frac{\psi T(1-\theta)}{\theta(1-p)} \\ &= \frac{x-kTp}{p(1-p)} - \frac{k}{\psi}\frac{\psi T(1-\theta)}{\theta(1-p)}, \quad \text{using (2.5);} \\ &= \frac{x\theta - kTp}{\theta p(1-p)}.\end{aligned} \tag{2.8}$$

So, equating (2.8) to zero leads to

$$x\widehat{\theta} = kT\widehat{p},$$

and hence

$$\frac{\widehat{p}}{\widehat{\theta}} = \frac{x}{kT}. \tag{2.9}$$

Thus the score equations give the two equations:

$$\widehat{\psi} = \frac{k}{N\widehat{\theta}} \quad \text{and} \quad \frac{\widehat{p}}{\widehat{\theta}} = \frac{x}{kT}, \tag{2.10}$$

where $\widehat{\theta} = g(\widehat{p}) = 1 - (1 - \widehat{p})^T$. Thus the second equation of (2.10) yields a (numerical) solution for $\widehat{p}$, and hence for $\widehat{\theta}$. Then, putting $\widehat{\theta}$ in the first equation gives a solution for $\widehat{\psi}$.

Note that the values of (x, k) occupy a finite grid, so if (x, k) are specified, then we can evaluate $\widehat{\psi}\,(x, k)$, since N and T are given: we solve $\dfrac{1 - (1 - \widehat{\theta}\,)^{1/T}}{\widehat{\theta}} = \dfrac{x}{kT}$ to give $\widehat{\theta}\,(x, k)$; and then $\widehat{\psi}(x, k) = \dfrac{k}{N\widehat{\theta}\,(x, k)}$.

However, the solution to the score equations is not always the MLE. We let $(\widehat{\psi}_s, \widehat{p}_s)$ denote the solution of the score equations. This gives the MLE if the maximum occurs at a turning point in the interior of the parameter space, the unit square: $[0, 1]^2$. If the maximum of the likelihood function occurs on the edge of the parameter space, then the maximum may not occur at a turning point and one (or both) of the score equations may not be satisfied, and $(\widehat{\psi}, \widehat{p}\,) \neq (\widehat{\psi}_s, \widehat{p}_s)$.

This will tend to happen when (x, k) is on the edge of the sample space ($k = N$ or $x = k$ or $x = kT$). This results in three sets of 'edge' solutions: edge solution (1) for $k = N$, edge solution (2) for $x = k$, and edge solution (3) for $x = kT$; see Figure 2.1. Let Ω denote the sample space, i.e. the set of all possible values of (x, k):

$$\Omega = \{(x, k) : x = k, k{+}1, \ldots, kT : k = 0, 1, 2, \ldots, N\}.$$

Let $\Omega' = \Omega \backslash (0, 0)$, i.e. the sample space, excluding ($x{=}0, k{=}0$):

$$\Omega' = \{(x, k) : x = k, k{+}1, \ldots, kT : k = 1, 2, \ldots, N\}.$$

On the boundaries OA ($x{=}k$) or AB ($k{=}N$) the maximum is obtained by assuming that occupancy is perfect (or complete, $\psi = 1$), and all sites are occupied ($\widehat{\psi} = 1$) and our failure to observe at all site-occasions is explained by imperfect (or incomplete) detectability ($p < 1$): $\widehat{p}$ is the proportion of all NT site-occasions that animals are detected, i.e.

$$\text{if } x = k \text{ or if } k = N, \text{ then } (\widehat{\psi}, \widehat{p}\,) = (1, \frac{x}{NT}).$$

On the boundary OB ($x{=}kT$), the maximum is obtained by assuming that detection is perfect ($\psi = 1$), and that the failure to observe at all site-occasions is due to imperfect ($\psi < 1$) occupancy: $\widehat{\psi}$ is the number of sites that are

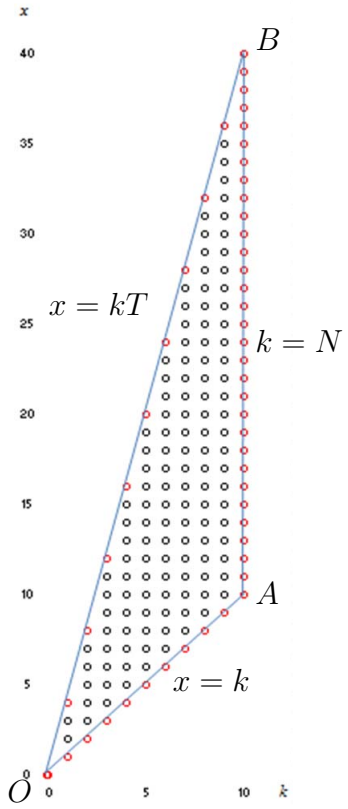


Figure 2.1: Diagram of the sample space, Ω, and its edges.

(perfectly) detected, i.e. $\widehat{\psi} = \frac{k}{N}$. Thus,

$$\text{if } x = kT, \text{ then } (\widehat{\psi}, \widehat{p}) = (\frac{k}{N}, 1).$$

We derive these 'edge' results next.

2.3.2 Edge solutions

Let Π denote the parameter space, i.e. the set of possible values of (ψ, p):

$$\Pi = \{(\psi, p) : 0 \leqslant p \leqslant 1; 0 \leqslant \psi \leqslant 1\}.$$

Let Π_s denote the "extended" parameter space, encompassing the set of possible values of $(\widehat{\psi}_s, \widehat{p}_s)$:

$$\Pi_s = \{(\psi, p) : 0 \leqslant p \leqslant 1,\ 0 \leqslant \psi \leqslant 1/\theta\}, \quad \text{where } \theta = 1 - (1-p)^T,$$

see Figure 2.2. The function L, as specified by (2.2), is non-negative for $(\psi, p) \in \Pi_s$. We seek to maximise L on Π.

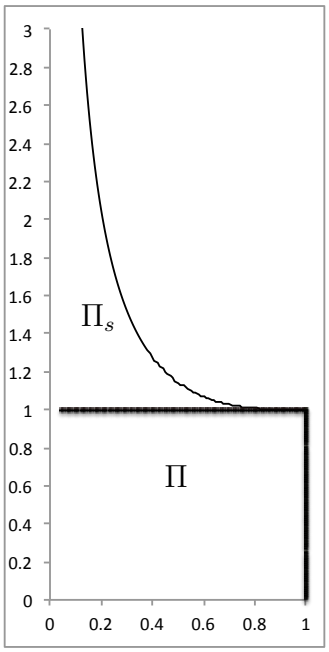


Figure 2.2: The sets: Π, the parameter space of possible (ψ, p), and Π_s, the "extended" parameter space of possible score estimates $(\widehat{\psi}_s, \widehat{p}_s)$: the region of the graph to the left of the curve and inside the square.

The score equations (see Equations (2.10)),

$$\widehat{\psi} = \frac{k}{N\widehat{\theta}} \quad \text{and} \quad \frac{\widehat{p}}{\widehat{\theta}} = \frac{x}{kT},$$

give a unique maximum on Π_s, denoted by $(\widehat{\psi}_s, \widehat{p}_s)$, allowing $(\widehat{\psi}_s, \widehat{p}_s) = (\infty, 0)$ in the case that $x = k$.

Let $u(p) = \dfrac{p}{\theta} = \dfrac{p}{g(p)}$, then u is an increasing function of p, such that $u(0) = \lim_{p\to 0} u(p) = \dfrac{1}{T}$ and $u(1) = 1$. Therefore $\dfrac{\widehat{p}}{\widehat{\theta}} = \dfrac{x}{kT}$ always has a solution $\widehat{p}_s \in [0, 1]$, since $k \leqslant x \leqslant kT$; $\widehat{\theta}_s = 1 - (1 - \widehat{p})^T$. The value of $\widehat{\psi}_s$ is then given by $\widehat{\psi}_s = \dfrac{k}{N\widehat{\theta}_s}$.

The MLEs for the edges (or boundaries) of the sample space Ω comprised of the (X, K) pairs (see Figure 2.1) are found next. The edge (OA+) considers the (x, k) points between the vertices excluding the special case $(0, 0)$, which is addressed separately in (O) below. The remainder of the edges below are defined according to the vertices in Figure 2.1.

(OA+) For $x = k$, and possibly some other values of x $(k{+}1, k{+}2, \ldots)$, close

to the boundary of the sample space, the value of $\widehat{\psi}_s$ is greater than 1 (see Figure 2.2). In that case, as L has no other local maxima in Π^*, it follows that $\widehat{\psi} = 1$. To find $\widehat{p}$, we therefore seek the point for which L is a maximum along the $\widehat{\psi} = 1$ boundary.

Now, $\log L(1, p) = x \log(p) + (NT - x) \log(1 - p)$, which has a maximum at $p = \dfrac{x}{NT}$. Therefore $(\widehat{\psi}, \widehat{p}) = (1, \dfrac{x}{NT})$.

(AB) If $k = N$, then $\dfrac{\partial \log L}{\partial \psi} = \dfrac{k}{\psi} > 0$; and therefore $\widehat{\psi} = 1$, from which it follows that $\widehat{p} = \dfrac{x}{NT}$, i.e. $(\widehat{\psi}, \widehat{p}) = (1, \dfrac{x}{NT})$ as above.

(OB) In the case that $x = kT$, the score equations give $\dfrac{\widehat{p}}{\widehat{\theta}} = 1$, so that $\widehat{p}_s = \widehat{\theta}_s = 1$ and $\widehat{\psi} = \dfrac{k}{N\widehat{\theta}_s} = \dfrac{k}{N}$. Hence, $(\widehat{\psi}, \widehat{p}) = (\dfrac{k}{N}, 1)$.

(O) The odd case $(x, k) = (0, 0)$. In this case,

$$L = [\psi(1-p)^T + (1-\psi)]^N = (1 - \psi\theta)^N.$$

This is maximised for any $(\widehat{\psi}, \widehat{p})$ such that $\widehat{\psi} \times \widehat{p} = 0$, i.e. when either $\widehat{\psi} = 0$ or $\widehat{p} = 0$, or both. Thus there is no unique MLE in this case, without further restriction or assumption.

We arbitrarily choose to define $(\widehat{\psi}, \widehat{p}) = (0, 0)$ in this case.

This case is odd in many ways. It means that N sites have been observed on T occasions ... all for no result, i.e. no species sightings are made on any of the NT site-occasions. This outcome conveys no information about occupation, as it is not known whether the failure to sight the species is due to its absence, or due to a failure of detection. This is reflected in the non-uniqueness of the MLE in this case.

Several approaches could be used:

1. If we assume that $p > 0$, i.e. that the detection probability is non-zero, then $\widehat{\psi} = 0$. This seems like a plausible assumption for any sensible observational procedure. For example, conducting a trial run in which a sighting is made would be enough to ensure $p > 0$. In this case $\widehat{p}$ remains undefined. [We, somewhat absurdly, may choose to retain the definition $\widehat{p} = 0$, for convenience.]

2. The $(x, k) = (0, 0)$ case can be avoided entirely by considering the behaviour of $\widehat{\psi}$ conditional on $(x, k) \neq (0, 0)$, i.e. $\widehat{\psi} > 0$.

Any sensible experiment would endeavour to avoid the $(0,0)$ outcome. It is unlikely that such a result would be reported. If N, T are moderately large, and ψ, p are not too small, then the probability of $(0,0)$ is very small.

In any event, the $(0,0)$ outcome is to be avoided — and we reflect this in our analysis. Any study for which $(0,0)$ (i.e. no detections are ever observed during the study) is a likely outcome would be regarded as a poor study, and so some indication of the 'quality' of the study is indicated by the magnitude of $\pi = \Pr(X = 0, K = 0)$: better 'quality' corresponding to smaller π.

We report two sets of results (1) allowing the possibility of the $(0,0)$ outcome, and (2) conditioning on $(x,k) \neq (0,0)$. We would want π to be small, and the difference between the conditional and unconditional results to be small.

Some examples of $(\widehat{\psi}, \widehat{p})$:

$N = 2$, $T = 2$

4			$(1,1)$
3			$(1,\frac{3}{4})$
2		$(\frac{1}{2},1)$	$(1,\frac{1}{2})$
1		$(1,\frac{1}{4})$	
$x=0$	(0,0)		
	$k=0$	1	2

$N = 3$, $T = 2$

6				$(1,1)$
5				$(1,\frac{5}{6})$
4			$(\frac{2}{3},1)$	$(1,\frac{2}{3})$
3			$(\frac{3}{4},\frac{2}{3})$	$(1,\frac{1}{2})$
2		$(\frac{1}{3},1)$	$(1,\frac{1}{3})$	
1		$(1,\frac{1}{6})$		
$x=0$	(0,0)			
	$k=0$	1	2	3

$N = 5$, $T = 3$

15						$(1, 1)$
14						$(1, \frac{14}{15})$
13						$(1, \frac{13}{15})$
12					$(0.8, 1)$	$(1, \frac{12}{15})$
11					$(0.801, 0.916)$	$(1, \frac{11}{15})$
10					$(0.804, 0.829)$	$(1, \frac{10}{15})$
9				$(0.6, 1)$	$(0.815, 0.736)$	$(1, \frac{9}{15})$
8				$(0.601, 0.888)$	$(0.841, 0.634)$	$(1, \frac{8}{15})$
7				$(0.608, 0.768)$	$(0.901, 0.518)$	$(1, \frac{7}{15})$
6			$(0.4, 1)$	$(0.631, 0.634)$	$(1.000, 0.400)^*$	$(1, \frac{6}{15})$
5			$(0.402, 0.829)$	$(0.701, 0.475)$	$(1.000, 0.333)^*$	$(1, \frac{5}{15})$
4			$(0.421, 0.634)$	$(0.969, 0.275)$	$(1, \frac{4}{5})$	
3		$(0.2, 1)$	$(0.524, 0.382)$	$(1, \frac{3}{15})$		
2		$(0.21, 0.63)$	$(1, \frac{2}{15})$			
1		$(1, \frac{1}{15})$				
$x = 0$	(0,0)					
	$k = 0$	1	2	3	4	5

* The asterisked entries produced $\widehat{\psi}_s > 1$ and so were truncated to $\widehat{\psi} = 1$. In turn, the $\widehat{p}$ were adjusted according to the edge solutions described above. This example demonstrates well how the edge solutions are applied to find estimates for both ψ and p and what happens when score equation solutions are greater than 1. We outline the details of the algorithm in Appendix 2.8.

2.3.3 Plausible region

In proceeding with our analysis, we must restrict ourselves to the region of the parameter space in which the non-zero ML estimates exist i.e. a region of the parameter space $\Pi = \{(\psi, p) : 0 \leqslant \psi \leqslant 1, 0 \leqslant p \leqslant 1\}$ in which unbiased estimation is plausible. We avoid the region of the parameter space in which there are no ML estimates.

We specify that the set of non-zero ML estimates for ψ and p exist in

$$\widehat{Q} = \{(\widehat{\psi}, \widehat{p})(x, k),\ (x, k) \in \Omega'\}, \quad \Omega' = \Omega \backslash (0, 0)$$

excluding $(x, k) = (0, 0)$. Note: we have defined $(\widehat{\psi}, \widehat{p})(0, 0) = (0, 0)$.

Let $\widehat{Q}_s$ denote the set of score equation solutions, excluding $(x{=}0, k{=}0)$:

$$\widehat{Q}_s = \{(\widehat{\psi}_s, \widehat{p}_s)(x,k), (x,k) \in \Omega'\}.$$

Then $\widehat{Q} \subset \Pi$ and $\widehat{Q}_s \subset \Pi_s$.

The set $\widehat{Q}$ (i.e. the values of ψ and p in $\widehat{Q}$) is bounded below by $\psi\theta = 1/N$, where $\theta = 1-(1-p)^T$, the $(k=1)$-line in Figure 2.3. This is found from the score equation $\widehat{\psi}_s = k/N\widehat{\theta}_s$ corresponding to $k=1$. This lower bounding curve is populated by only $T-2$ (i.e. $kT-2$, with $k=1$) 'internal' points (i.e. points not on the boundary of $\widehat{Q}$), so if T is small this lower bound curve may be well under most of the actual estimate points (Figure 2.3). The 'next' curve is $\psi\theta = 2/N$ populated by (at most) $2T-2$ internal points, which may also only be relatively few points for small T.

The actual lower limit for parameter values which may be unbiasedly estimated, excluding $(0,0)$ in some way, is the convex hull of the set of possible estimates. This is indicated by the solid line in Figure 2.3 which connects the $(\widehat{\psi}, \widehat{p}\,)$ for the observations at (x,k) for $k=1$ and $x = k, k+1, \ldots, kT = 1,2,3$. We apply the score equations i.e. $\widehat{\psi}_s = k/N\widehat{\theta}_s$ and $\widehat{p}_s = x\widehat{\theta}_s/kT$, to find the *internal* points corresponding to the (x,k) pairs. The $\widehat{\theta}_s$ was found previously via numerical approximation (see Section 2.3.1).

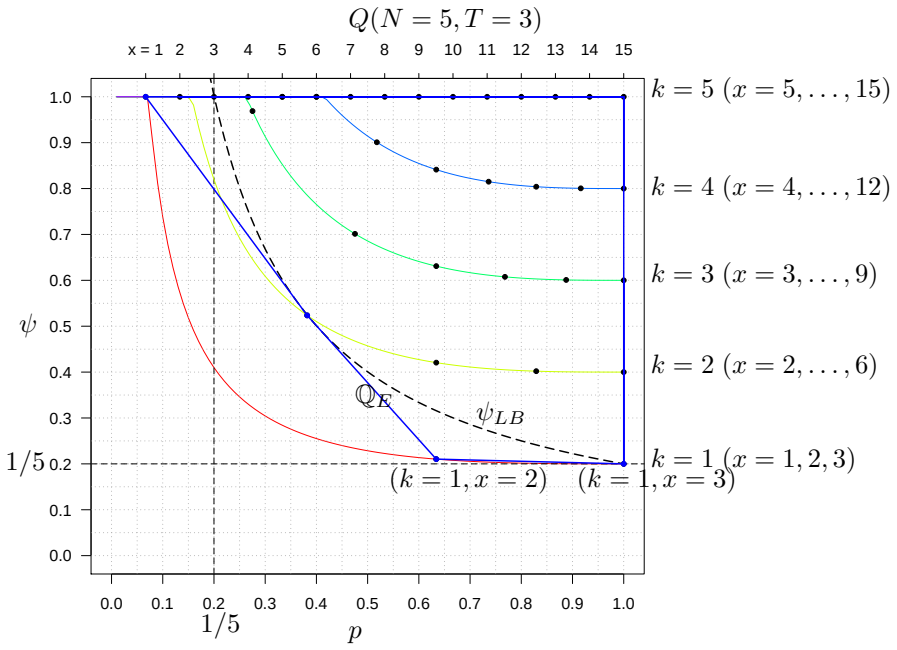


Figure 2.3: Convex hull $\mathbb{Q}_E$ for $N=5, T=3$. The solid lines mark the L_k-lines for $k=1,\ldots,N$ and the bullets along the L_k mark the MLE corresponding to (x,k) where for each $k: x = k, k+1, \ldots, kT$.

The convex hull is defined by all points in $\mathbb{Q}_E(N,T) = \text{Conv}(\widehat{Q}(N,T)) = \text{Conv}\{(\widehat{\psi},\widehat{p})(x,k),\ (x,k) \in \Omega'\}$. We would like to find an approximation for the bottom edge of the convex hull of the set of MLEs $\mathbb{Q}_E(N,T)$, so that for points above this, unbiased estimation is plausible.

The top–right region of the parameter space (large p, large ψ) is mostly populated by the estimated points with the bottom–left region (small p, small ψ) being the least populated. This is illustrated, for example, in Figures 2.3, 2.4 and 2.5, where for larger N and T the plausible region becomes more densely populated with points. For example, in Figure 2.3 when $N = 5, T = 3$ and $p = 0.4$ there would need to be at least $k = 2$ detected sites with $x = 3$ total detections and a level of occupancy of at least $\psi = 0.45$ to achieve an estimate in the convex hull (a $\widehat{\psi}$ in $\mathbb{Q}_E$) i.e. an estimate for occupancy that may be unbiased. Other studies have identified similar conditions necessary for estimates to exist, see, for example Wintle et al. (2004) and Guillera-Arroita et al. (2010).

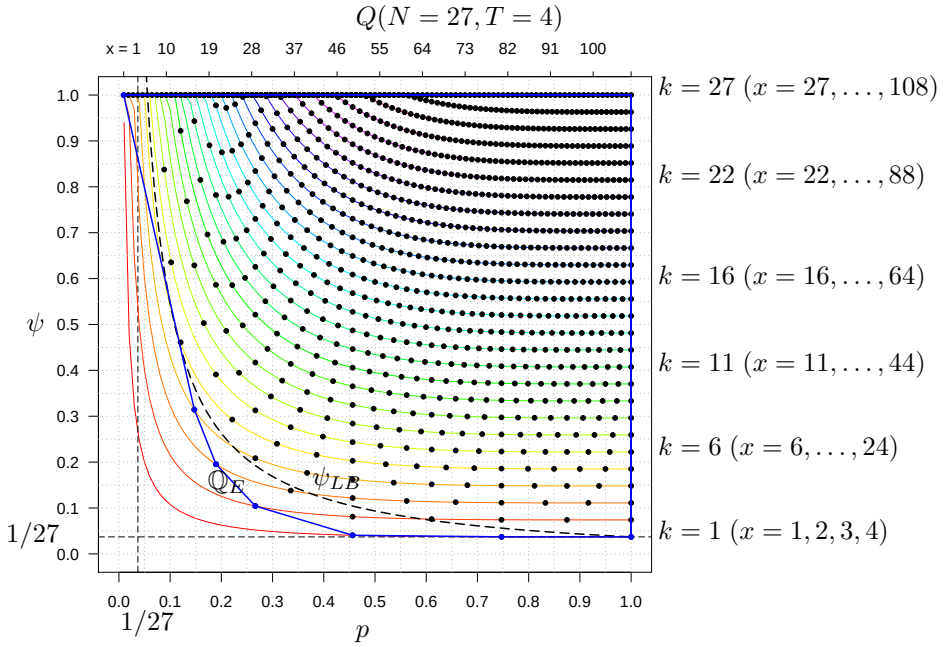


Figure 2.4: Convex hull $\mathbb{Q}_E$ for $N = 27, T = 4$. The solid lines mark the L_k-lines for $k = 1, \ldots, N$ and the bullets along the L_k mark the MLE corresponding to (x,k) where for each $k : x = k, k+1, \ldots, kT$.

In general, the set of MLEs, $\widehat{Q}(N,T)$ consists of points on the lines $L_k : \psi\theta = k/N$ for $k = 1, 2, \ldots, T$. These lines increase with k : $L_1 < L_2 < \ldots < L_T$. So we know that $\mathbb{Q}_E$ has a lower bound given by the line $\psi = 1/(N\theta)$, corresponding to $k = 1$. But, as L_1, the $(k = 1)$-line, has few points on it (at

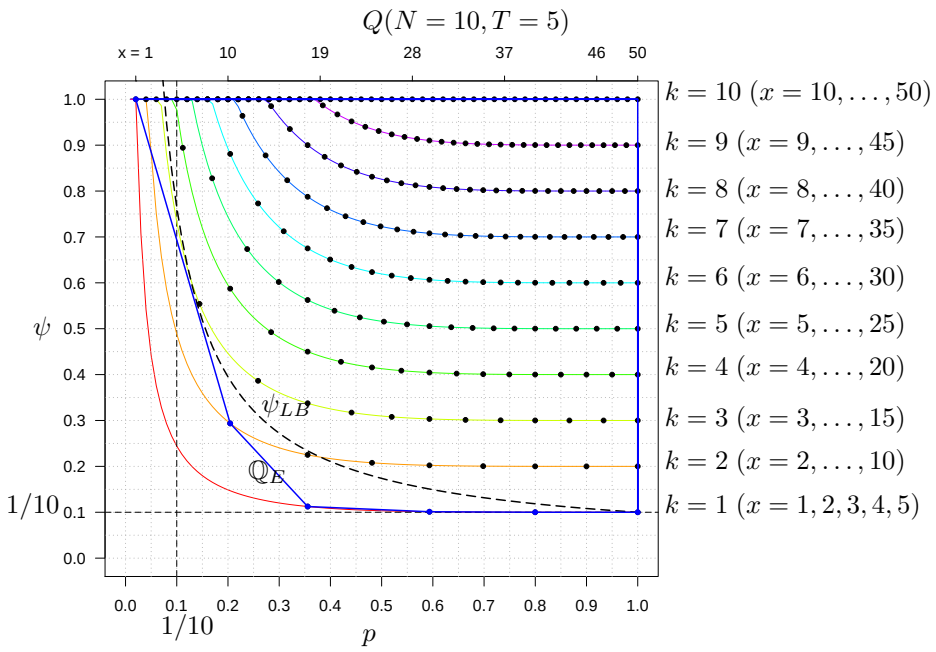


Figure 2.5: Convex hull $\mathbb{Q}_E$ for $N = 10, T = 5$. The solid lines mark the L_k-lines for $k = 1, \ldots, N$ and the bullets along the L_k mark the MLE corresponding to (x, k) where for each $k : x = k, k + 1, \ldots, kT$.

most $T - 2$), and the L_2 not many more, we would like a curve that is closer to $\mathbb{Q}_E$, and includes most of the points.

So, the approximation for the bottom edge of the convex hull of the set of ML estimates $\mathbb{Q}_E(N, T)$ consists of the region that is not densely populated so that for points above this, unbiased estimation is plausible. If this equation is expressed in the form $\psi = f(p)$, then given p, we can estimate ψ for $\psi \geqslant f(p)$, i.e. $f(p)$ represents a lower bound for unbiased estimation of ψ.

We find a lower bound for ψ, ψ_{LB}. Set ε to be a small number greater than zero (i.e. $\varepsilon > 0$), for example $\varepsilon = 0.01$.

Suppose that $p(0, 0) \leqslant \varepsilon$, and solve for ψ:

$$(1 - \psi\theta)^{1/N} \leqslant \varepsilon \quad \Rightarrow \quad \psi \geqslant \frac{1 - \varepsilon^{1/N}}{\theta}, \quad \text{where } \theta = 1 - (1 - p)^T.$$

Now, we have

$$\psi \;\geqslant\; \frac{1 - \varepsilon^{1/N}}{1 - (1 - p)^T} \;\approx\; \frac{c_\varepsilon}{p}, \quad \text{for small } p,$$

where $c_\varepsilon = (1 - \varepsilon^{1/N})/T$. We adjust this so that it goes through $(1, \frac{1}{N})$, and define the approximate lower bound

$$\psi_{LB} = \frac{c_\varepsilon}{p} + (\frac{1}{N} - c_\varepsilon), \tag{2.11}$$

which gives a reasonable approximation for small p, and tends to 1 as $p \to 1$.

This approximate lower bound is intended to provide a reasonable practical lower bound for 'plausible' estimation of ψ for a given value of p. It is found that when N and T are small $\varepsilon = 0.01$ provides a useful practical approximation. For example see Figure (2.5) for $N = 10, T = 5$ and Figure (2.3) when $N = 5, T = 3$.

If we consider p to be a nuisance parameter, and our primary interest is the estimation of ψ, then it will suffice to use the approximate lower bound for ψ, $\psi_{LB}(p)$ given by equation (2.11).

For example, in Figure 2.3 we would say that for a value for $p = 0.4$, the ML estimates for occupancy exist within the convex hull marked by $\mathbb{Q}_E$ and that it is plausible to estimate ψ provided $\psi \gtrsim 0.45$. [The region below ψ_{LB} is too sparsely populated by observations and must lead to biased estimates.] As another example, in Figure 2.5 for a study of size $N = 10$ sites and $T = 5$ survey occasions there are seven (x, k) pairs (marked with bullets in the figure) which lie below the boundary ψ_{LB}. There are three pairs at $\widehat{\psi} = 1$, if we specify $\psi^*_{LB} = \min(1, \psi_{LB})$, then only four pairs are below ψ^*_{LB}. So ψ_{LB} gives a reasonable practical lower bound for estimation of ψ. In the following sections we demonstrate with examples the practical use of the ψ lower bound when evaluating the expectation, bias and variance of occupancy.

2.4 Asymptotic variance of $\widehat{\psi}$

Following MacKenzie et al. (2006, p. 96), in this section we derive expressions for the asymptotic variance of $\widehat{\psi}$ and $\widehat{p}$ of Equations (2.10), respectively. The result of this asymptotic estimate is given in the literature, however the derivations are not published, thus we show our derivations here in order to understand the limitations with the estimator. We use the variances to evaluate the precision of the estimators.

To obtain the expressions for the asymptotic variances we need the expected information matrix. This matrix is obtained from the partial second derivatives of the log-likelihood function.

We calculate each entry of the expected Fisher information matrix, $\mathbf{I}(\psi, p)$, which corresponds to the basic occupancy likelihood function, shown below in Equation (2.13),

$$\mathbf{I} = \begin{bmatrix} \mathrm{I}_{\psi\psi} & \mathrm{I}_{\psi p} \\ \mathrm{I}_{p\psi} & \mathrm{I}_{pp} \end{bmatrix} \tag{2.12}$$

Each element is the expected value of the observed information, i.e. the negative of the partial second derivative of the log-likelihood. For example, $\mathrm{I}_{pp}(\psi, p) = -E\left(\dfrac{\partial^2 \log L(\psi, p \mid X, K)}{\partial\, p^2}\right)$.

Recall the log-likelihood equation from above (Equation (2.4))

$$\log L = k\log(\psi) + x\log(p) + (kT - x)\log(1-p) + (N-k)\log(1-\psi\theta) \tag{2.13}$$

where $\theta = \theta(p) = 1-(1-p)^T$, $x =$ the total number of site-occasions on which the species was observed, and $k =$ the number of sites on which the species was observed (at least once) — see Section 2.3.1 for details.

We give an alternative derivation of the likelihood function, specifying the distributions of X and K: the expectations of X and K are used to derive expressions for the expected information.

The basic occupancy model assumes that the detection, p, remains constant over all N sites and T survey occasions. Under this assumption, we may construct the likelihood by modelling the number of detections at each site $X_{i\bullet}$ as a ZIB$(T, p; \psi)$ random variable (that is, a zero-inflated binomial (ZIB) random variable) (Hall, 2000; Royle and Nichols, 2003; Royle, 2006; MacKenzie

et al., 2006, p. 94). We assume that the number of detections $X_{i\cdot}$ at site i is distributed as

$$X_{i\cdot} \overset{d}{=} \begin{cases} \text{Bi}(T,p), & \text{with probability } \psi, \\ 0, & \text{with probability } (1-\psi). \end{cases} \tag{2.14}$$

Thus

$$\Pr(X_{i\cdot} = x_{i\cdot}) = \begin{cases} \psi\binom{T}{x_{i\cdot}} p^{x_{i\cdot}} \cdot (1-p)^{T-x_{i\cdot}}, & x_{i\cdot} = 1, 2, \ldots, T; \\ \psi(1-p)^T + (1-\psi), & x_{i\cdot} = 0. \end{cases} \tag{2.15}$$

To construct the likelihood function we adopt the approach of Hall (2000, Equation 3). Then we sum over the site specific likelihoods similar to Royle and Nichols (2003). Let the indicator function $u_{i\cdot} = \text{I}(x_{i\cdot} \geqslant 1)$ denote whether a detection was made at the i^{th} site,

$$u_{i\cdot} = \text{I}(x_{i\cdot} \geqslant 1) = \begin{cases} 1, & \text{if } x_{i\cdot} > 0, \\ 0, & \text{if } x_{i\cdot} = 0. \end{cases} \tag{2.16}$$

The likelihood for data at the ith site is given by:

$$\begin{aligned} & L_i(\psi, p \,|\, \mathbf{x}_i) \\ &= \Pr(X_{i\cdot} = x_{i\cdot} \,;\, \psi, p) \\ &= \left[\psi\binom{T}{x_{i\cdot}} p^{x_{i\cdot}} \cdot (1-p)^{T-x_{i\cdot}}\right]^{u_{i\cdot}} \left[\psi(1-p)^T + (1-\psi)\right]^{(1-u_{i\cdot})}. \end{aligned}$$

Then, the full likelihood is

$$\begin{aligned} & L(\psi, p \,|\, \mathbf{X}) \\ &= \textstyle\prod_{i=1}^{N} L_i(\psi, p \,|\, x_{i\cdot}) \\ &= \textstyle\prod_{i=1}^{N} \Pr(X_{i\cdot} = x_{i\cdot}) \\ &= \textstyle\prod_{i=1}^{N} \left[\psi\binom{T}{x_{i\cdot}} p^{x_{i\cdot}} \cdot (1-p)^{T-x_{i\cdot}}\right]^{u_{i\cdot}} \left[\psi(1-p)^T + (1-\psi)\right]^{(1-u_{i\cdot})} \\ &= \textstyle\prod_{i=1}^{N} \binom{T}{x_{i\cdot}}^{u_{i\cdot}} \psi^{\sum_{i=1}^{N} u_{i\cdot}} p^{\sum_{i=1}^{N} x_{i\cdot} u_{i\cdot}} (1-p)^{\sum_{i=1}^{N} u_{i\cdot}(T-x_{i\cdot})} (1-\psi\theta)^{\sum_{i=1}^{N}(1-u_{i\cdot})}, \end{aligned}$$

where

$$p^{\sum_{i=1}^{N} x_{i\cdot} u_{i\cdot}} = p^{\sum_{i=1}^{N} x_{i\cdot} \text{I}(x_{i\cdot} > 0)} = p^{\sum_{i=1}^{N} x_{i\cdot}} = p^x \quad \text{and} \quad k = \sum_{i=1}^{N} u_{i\cdot}.$$

It follows that,

$$L\left(\psi, p \,|\, \mathbf{X}\right) = \kappa\ \psi^k p^x (1-p)^{kT-x} (1-\psi\theta)^{N-k}, \tag{2.17}$$

where $\kappa = \prod_{i=1}^{N} \binom{T}{x_{i\bullet}}$, which is independent of the parameters. Hence Equation (2.2) is equivalent to Equation (2.17).

To evaluate the expectation required for the information matrix, we need an expression for the expectations of X and K.

Now $X = \sum_{i=1}^{N} X_{i\bullet}$, where $X_{i\bullet} \stackrel{\mathrm{d}}{=} \mathrm{ZIB}(T, p; \psi)$, and so $E(X_{i\bullet}) = Tp\psi$. Consequently, $E(X) = NTp\psi$, since the X_i are identically distributed.

The random variable $K = \sum_{i=1}^{N} U_i = \sum_{i=1}^{N} \mathrm{I}(X_{i\bullet} > 0)$; and $E(\mathrm{I}(X_{i\bullet} > 0)) = \Pr(X_{i\bullet} > 0) = 1 - \psi(1-p)^T - (1-\psi) = \psi(1-(1-p)^T) = \psi\theta$.

It follows that $E(K) = N\psi\theta$.

Evaluation of $\mathbf{I}_{\psi\psi}$

The $\mathrm{I}_{\psi\psi}$ entry of the information matrix is the expectation of the negative of the second partial derivative of $\widehat{\psi}$,

$$\mathbf{I}_{\psi\psi}(\psi, p) = E\left(-\frac{\partial^2 \log L(\psi, p \,|\, x, k)}{\partial \psi^2}\right), \tag{2.18}$$

where $K = \sum_{i=1}^{N} \mathrm{I}(X_i > 0)$. The first partial derivative of the log-likelihood with respect to ψ is

$$\frac{\partial \log L(\psi, p \,|\, x, k)}{\partial\, \psi} = \frac{K}{\psi} - \frac{(N-K)\theta}{1-\psi\theta}. \tag{2.19}$$

Note that $E(\frac{\partial \log L}{\partial \psi}) = 0 \ \Rightarrow\ E(K) = N\psi\theta$.

The second partial derivative, again with respect to ψ gives

$$\frac{\partial^2 \log L(\psi, p \,|\, x, k)}{\partial \psi^2} = -\frac{K}{\psi^2} - \frac{(N-K)\theta^2}{(1-\psi\theta)^2}. \tag{2.20}$$

We use the expression for $E(K)$ to obtain $\mathrm{I}_{\psi\psi}$:

$$\begin{aligned}\mathrm{I}_{\psi\psi}(\psi,p) &= \frac{E(K)}{\psi^2} + \frac{(N-E(K))\theta^2}{(1-\psi\theta)^2} \\ &= \frac{N\psi\theta}{\psi^2} + \frac{(N-N\psi\theta)\theta^2}{(1-\psi\theta)^2} \\ &= \frac{N\theta}{\psi(1-\psi\theta)}.\end{aligned} \tag{2.21}$$

Evaluation of I_{pp}

Recall that $\theta = g(p) = 1-(1-p)^T$, so that $\theta' = g'(p) = T(1-p)^{T-1}$ and $\theta'' = g''(p) = -T(T-1)(1-p)^{T-2}$.

The first partial derivative with respect to p gives

$$\frac{\partial \log L}{\partial p} = \frac{x}{p} - \frac{kT-x}{1-p} - \frac{(N-k)\,\psi\,T(1-p)^{T-1}}{1-\psi\theta}. \tag{2.22}$$

Note that $E(\frac{\partial \log L}{\partial p}) = 0 \;\Rightarrow\; E(X) = NTp\psi$, using the previously derived expression for $E(X)$.

And, the second partial derivative again with respect to p is

$$\frac{\partial^2 \log L}{\partial p^2} = -\frac{x}{p^2} - \frac{kT-x}{(1-p)^2} - (N-k)\psi T\frac{(1-p)^{T-2}}{1-\psi\theta}\left(1-T+\frac{\psi T(1-p)^T}{1-\psi\theta}\right). \tag{2.23}$$

The final step in obtaining I_{pp} requires the expectation of Equation (2.23). The only random variables in this expression are X and K, so using the results $E(X) = NTp\psi$ and $E(K) = N\psi\theta$, we obtain

$$\begin{aligned}&\mathrm{I}_{pp} \\ &= -E\left(\frac{\partial^2 \log L(\psi,p\,|\,X,K)}{\partial p^2}\right) \\ &= \frac{E(X)}{p^2} + \frac{TE(K)-E(X)}{(1-p)^2} + (N-E(K))\psi T\frac{(1-p)^{T-2}}{1-\psi\theta}\left[1-T+\frac{\psi T(1-p)^T}{1-\psi\theta}\right] \\ &= \frac{NTp\psi}{p^2} + \frac{TN\psi\theta - NTp\psi}{(1-p)^2} + (N-N\psi\theta)\psi T\frac{(1-p)^{T-2}}{1-\psi\theta}\left[1-T+\frac{\psi T(1-p)^T}{1-\psi\theta}\right] \\ &= \frac{NT\psi}{p(1-p)} - \frac{N\psi T^2(1-\theta)(1-\psi)}{(1-p)^2(1-\psi\theta)}.\end{aligned} \tag{2.24}$$

Evaluation of $\mathrm{I}_{p\psi}$

The entries required to complete the information matrix are $\mathrm{I}_{p\psi}$ and $\mathrm{I}_{\psi p}$. Since these two quantities are equal we need only find one. The final entry, $\mathrm{I}_{p\psi} = -E\left(\frac{\partial^2 \log L(\psi,p\,|\,x,k)}{\partial\psi\partial p}\right)$; and we have, from (2.19):

$$\frac{\partial \log L(\psi, p\,|\,x,k)}{\partial\psi} = \frac{k}{\psi} - \frac{(N-k)\theta}{1-\psi\theta}. \tag{2.25}$$

So

$$\begin{aligned}
\frac{\partial^2 \log L(\psi, p\,|\,x,k)}{\partial\psi\,\partial p} &= \frac{\partial}{\partial p}\left[\frac{k}{\psi} - \frac{(N-k)\theta}{1-\psi\theta}\right], \\
&= -\frac{(N-k)T(1-p)^{T-1}}{1-\psi\theta} - \frac{(N-k)\psi\theta T(1-p)^{T-1}}{(1-\psi\theta)^2} \\
&= -\frac{(N-k)T(1-p)^{T-1}}{1-\psi\theta}\left[1 + \frac{\psi\theta}{1-\psi\theta}\right] \\
&= -\frac{(N-k)T(1-p)^{T-1}}{(1-\psi\theta)^2}.
\end{aligned} \tag{2.26}$$

Then the entry for the expected information matrix is

$$\begin{aligned}
\mathrm{I}_{\psi p} &= -E\left(\frac{\partial^2 \log L(\psi, p\,|\,X,K)}{\partial\psi\,\partial p}\right) \\
&= \frac{(N - E(K))T(1-p)^{T-1}}{1-\psi\theta} \\
&= \frac{(N - N\psi\theta)T(1-p)^{T-1}}{1-\psi\theta}, \\
&= \frac{NT(1-p)^{T-1}}{1-\psi\theta}.
\end{aligned} \tag{2.27}$$

Finally

$$\mathbf{I} = \begin{bmatrix} \mathrm{I}_{pp} & \mathrm{I}_{p\psi} \\ \mathrm{I}_{p\psi} & \mathrm{I}_{\psi\psi} \end{bmatrix}^{-1} = \begin{bmatrix} v_{pp} & v_{p\psi} \\ v_{p\psi} & v_{\psi\psi} \end{bmatrix}.$$

If we invert the information matrix $\mathbf{I}$, we find the asymptotic variance,

$$\sigma_a^2(\psi, p) = \mathrm{asvar}(\widehat{\psi}) = \frac{\psi}{N}\left((1-\psi) + \frac{1-\theta}{\theta - Tp(1-p)^{T-1}}\right). \tag{2.28}$$

Note that $\text{asvar}(\widehat{\psi}) = v_{\psi\psi}$. The asymptotic standard error is denoted by $ase(\widehat{\psi}) = \sqrt{\text{asvar}(\widehat{\psi})} = \sqrt{\sigma_a^2(\psi, p)}$. As mentioned previously, this asymptotic variance is given in MacKenzie et al. (2006, p.96).

2.5 Distributions of the sufficient statistics

Here we derive the distribution of the sufficient statistics. We show that $(\widehat{\psi}, \widehat{p})$ is a function of (x, k). Thus, to determine the distribution of $(\widehat{\psi}, \widehat{p})$, and in particular the expectation of $\widehat{\psi}$, we need to specify the distribution of (X, K). Then we may go on to evaluate the bias and variance of occupancy.

2.5.1 Introduction

We reorder the sites of the history matrix so that the *detected* sites are listed from $1, \ldots, k$ and the *undetected* sites from $k+1, \ldots, N$. By *detected* we mean a site at which the species was detected at least once during the study ($x_{ij} \geqslant 1$ for some j). It follows, that on *undetected* sites, the species was not detected on any occasion ($x_{ij} = 0$, for all j), resulting in a row of zeros for that site,

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1T} \\ x_{21} & x_{22} & \cdots & x_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kT} \\ x_{(k+1)1} & x_{(k+1)2} & \cdots & x_{(k+1)T} \\ x_{(k+2)1} & x_{(k+2)2} & \cdots & x_{(k+2)T} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \ldots & x_{NT} \\ \\ x_{\cdot 1} & x_{\cdot 2} & \cdots & x_{\cdot T} \end{bmatrix} \begin{matrix} x_{1\cdot} \\ x_{2\cdot} \\ \vdots \\ x_{k\cdot} \\ x_{(k+1)\cdot} \\ x_{(k+2)\cdot} \\ \vdots \\ x_{N\cdot} \\ \\ x_{\cdot\cdot} \end{matrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} \neq 0 \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} = 0 \\ \\ \\ \end{matrix} \tag{2.29}$$

where $k = 0, \ldots, N$. As before $k = \sum_{i=1}^{N} \mathrm{I}(x_{i\cdot} \geqslant 1)$, the number of detected sites, where $x_{i\cdot} = \sum_{j=1}^{T} x_{ij}$ are the number of detections at a site, i.e. the i^{th} row sum of the history matrix, and the row vectors $\mathbf{x}_i = (x_{i1}\, x_{i2}\, \ldots\, x_{iT})$ describe the N detection history vectors for each site i of the study (2.5.1). $x_{\cdot j} = \sum_{i=1}^{N} x_{ij}$ is the total number of detections on the j^{th} occasion, i.e. the sum of the j^{th} column of the history matrix. Contributions to the sums $x_{i\cdot}$ and $x_{\cdot j}$ are made exclusively by the k detected sites, whereas the (re ordered) sites $(k+1), \ldots, N$ do not contribute since all entries for these sites are zero. Finally, the total number of detections in the history matrix are $x = x_{\cdot\cdot} = \sum_{j=1}^{T} x_{\cdot j}$; the sum of the column sums. Equivalently, x may be described

as the sum of the row sums to which only the k detected sites contribute: $x = x_{\cdot\cdot} = \sum_{i=1}^{N} x_{i\cdot} = \sum_{i=1}^{k} x_{i\cdot}$.

As previously shown in Section 2.2, (x, k) is sufficient for (ψ, p), with the log-likelihood given by

$$\begin{aligned}\log L(\psi, p \,|\, x, k) = k\, \log(\psi) + x\, \log(p) + (k\,T - x) \log(1-p) \\ + (N-k) \log(1 - \psi\,\theta),\end{aligned} \tag{2.30}$$

where $\theta = \theta(p) = 1 - (1-p)^T$.

It follows that $(\widehat{\psi}, \widehat{p})$ is a function of (x, k). In the following section we find the distribution of (X, K) so that we may determine the distribution of $(\widehat{\psi}, \widehat{p})$ that will give the exact variance and bias for $\widehat{\psi}$.

2.5.2 The joint pmf of X and K

The number of detections at site i, $X_{i\cdot} = \sum_{j=1}^{T} X_{ij}$ is such that

if $i = 1, 2, \ldots, k$, then $X_{i\cdot} \stackrel{\mathrm{d}}{=} \mathrm{Bi}^{+}(T, p)$, and,

if $i = k{+}1, k{+}2, \ldots$, then $X_{i\cdot} = 0$;

where Bi^{+} denotes a Binomial distribution without zero. That is, conditional on $X_{i\cdot}$ being positive:

if $Z \stackrel{\mathrm{d}}{=} \mathrm{Bi}^{+}(n, p)$, then $\Pr(Z = z) = \binom{n}{z} \dfrac{p^z (1-p)^{n-z}}{1 - (1-p)^n}, \quad z = 1, 2, \ldots, n.$

It follows that the probability generating function (pgf) of Z is given by

$$\mathrm{P}_Z(s) = \frac{(1 - p + ps)^n - (1-p)^n}{1 - (1-p)^n}.$$

The joint probability mass function (pmf) of (X, K) is given by

$$\begin{aligned} p(x,k) &= \mathrm{P}(K = k) \times \mathrm{P}(X = x \,|\, K = k) \\ &= c(x,k)\, \psi^k (1 - \psi\theta)^{N-k} p^x (1-p)^{kT-x} \\ &= \underbrace{c_1(k) (\psi\theta)^k (1 - \psi\theta)^{N-k}}_{\substack{\text{pmf of } K \\ (k = 0, 1, \ldots, N)}} \times \underbrace{c_2(x,k) \frac{p^x (1-p)^{kT-x}}{\theta^k}}_{\substack{\text{pmf of } X \,|\, K=k \\ (x = k, k+1, \ldots, kT)}}, \end{aligned} \tag{2.31}$$

where $c(x, k)$, $c_1(k)$ and $c_2(x, k)$ are constants, i.e. they do not depend on (ψ, p).

The pmf of K, the distribution of the number of detected sites in the history matrix (Equation (2.2)), and the first term in Equation (2.31), follows a Binomial distribution, $K \stackrel{\mathrm{d}}{=} \mathrm{Bi}(N, \psi\theta)$. It follows that $c_1(k) = \binom{N}{k}$. The second term in Equation (2.31) corresponds to a sum of independent random variables each having the positive binomial distribution, Bi^+; the binomial distribution for which zero is not in the support space. Furthermore, the domain of X is restricted to between k and kT, $k \leqslant x \leqslant kT$.

Hence, the pgf of $(X|K=k) \stackrel{\mathrm{d}}{=} X_1 + X_2 + \cdots + X_k$ is the sum of k $\mathrm{Bi}^+(T,p)$ distributions. So conditional on $K = k$,

$$\begin{aligned} \mathrm{P}_X(s \,|\, K=k) &= E(s^X \,|\, K=k) \\ &= \mathrm{P}_{X_1}(s)^k, \quad \text{since the } X_i \text{ are iid} \\ &= \left(\frac{(ps+1-p)^T - (1-p)^T}{\theta}\right)^k \\ &= \left[\frac{(q+ps)^T - q^T}{1-q^T}\right]^k, \quad \theta = 1-q^T = 1-(1-p)^T \\ &= \frac{q^{kT}}{\theta^k}\left[\left(1+\frac{ps}{q}\right)^T - 1\right]^k \end{aligned} \tag{2.32}$$

Let $\left[\left(1+\frac{ps}{q}\right)^T - 1\right]^k = [(1+z)^T - 1]^k$, to which we apply the binomial expansion:

$$[(1+z)^T - 1]^k = \sum_{j=0}^{k} \binom{k}{j} (1+z)^{Tj} (-1)^{k-j}. \tag{2.33}$$

Next, expand $(1+z)^{Tj}$, so that

$$[(1+z)^T - 1]^k = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \sum_{x=0}^{Tj} \binom{Tj}{x} z^x, \tag{2.34}$$

where $x = 0$ when, and only when, $k = 0$ (which would mean that no positive observations were made in the study). Thus, we exclude $k = 0$ as a special case and, for $k = 1, 2, \ldots$

$$[(1+z)^T - 1]^k = \sum_{x=k}^{Tk} \left\{ \sum_{j=1}^{k} (-1)^{k-j} \binom{k}{j} \binom{Tj}{x} \right\} z^x. \tag{2.35}$$

Furthermore, $x = k, \ldots, Tk$, since the smallest power permissible is k,

$$\begin{aligned}[(1+z)^T - 1]^k &= [\,1 + \binom{T}{1} z + \binom{T}{2} z^2 + \cdots - 1\,]^k \\ &= [\,Tz + \frac{T(T-1)}{2} z^2 + \cdots]^k \\ &= T^k\, z^k + \cdots .\end{aligned} \tag{2.36}$$

In conclusion, the expression for the conditional probability generating function is

$$\begin{aligned}\mathrm{P}_X(s|K=k) &= \frac{q^{Tk}}{\theta^k} \sum_{x=k}^{Tk} \left\{ \sum_{j=1}^{k} (-1)^{k-j} \binom{k}{j} \binom{Tj}{x} \right\} z^x \\ &= \frac{q^{Tk}}{\theta^k} \sum_{x=k}^{Tk} \left\{ \sum_{j=1}^{k} (-1)^{k-j} \binom{k}{j} \binom{Tj}{x} \right\} \left(\frac{ps}{q}\right)^x, \\ &\qquad \text{where } q = 1-p,\ z = \frac{ps}{q} \text{ and } \theta = 1-(1-p)^T, \\ &= \sum_{x=k}^{Tk} \left\{ \sum_{j=1}^{k} \binom{k}{j} \binom{Tj}{x} (-1)^{k-j} \right\} \frac{p^x (1-p)^{Tk-x}}{\theta^k} s^x.\end{aligned} \tag{2.37}$$

Hence, the coefficient required to complete the expression for the joint pmf of (X, K) (2.31) is

$$c_2(x,k) = \sum_{j=\lceil \frac{x}{T} \rceil}^{k} \binom{k}{j} \binom{Tj}{x} (-1)^{k-j}, \tag{2.38}$$

where $k = 1, \ldots, N$ and $x = k, k+1, \ldots, Tk$. The summation begins at $j = \lceil \frac{x}{T} \rceil$ as $\binom{Tj}{x} = 0$ if $x > Tj$, since Tj is an integer.

The joint probability mass function of X and K is thus

$$p(x,k) = c_1(k)(\psi\theta)^k (1-\psi\theta)^{N-k} \ \times\ c_2(x,k) \frac{p^x (1-p)^{Tk-x}}{\theta^k}$$

$$= \binom{N}{k} (\psi\theta)^k (1-\psi\theta)^{N-k} \ \times \sum_{j=\lceil \frac{x}{T} \rceil}^{k} \binom{k}{j} \binom{Tj}{x} (-1)^{k-j}\ \frac{p^x (1-p)^{Tk-x}}{\theta^k}. \tag{2.39}$$

2.5.3 Expectation of the occupancy estimator, $\widehat{\psi}$

Having obtained the pmf of (X, K), since $\widehat{\psi} = \widehat{\psi}(X, K)$, it follows that the expectation for $\widehat{\psi}$ is given by

$$\begin{aligned} E\left(\widehat{\psi} \,|\, \psi, p\right) &= \sum_{k=0}^{N} \sum_{x=k}^{Tk} \widehat{\psi}(x,k)\; p(x,k) \\ &= \sum_{k=0}^{N} \sum_{x=k}^{Tk} \widehat{\psi}(x,k) \binom{N}{k} (\psi\theta)^k (1-\psi\theta)^{N-k} \quad \times \\ &\quad \left\{ \sum_{j=\lceil \frac{x}{T} \rceil}^{k} \binom{k}{j} \binom{Tj}{x} (-1)^{k-j} \right\} \frac{p^x (1-p)^{Tk-x}}{\theta^k}. \end{aligned} \tag{2.40}$$

The expectation can be re-written as

$$\mu(\psi, p) = E(\widehat{\psi}) \;=\; \sum_{k=1}^{N} \sum_{x=k}^{Tk} \widehat{\psi}(x,k)\; p(x,k), \tag{2.41}$$

since we have assumed $\widehat{\psi}(0,0) = 0$.

An alternative approach is to exclude the $(0,0)$-case to give a conditional expectation for the occupancy estimator, i.e.

$$\mu^*(\psi, p) = E(\widehat{\psi} \,|\, \widehat{\psi} > 0) = \sum_{k=1}^{N} \sum_{x=k}^{Tk} \frac{\widehat{\psi}(x,k)\; p(x,k)}{1 - p(0,0)}, \tag{2.42}$$

where $p(0,0) = (1-\psi\theta)^N$. So $\mu^* = \dfrac{\mu}{1-p(0,0)}$.

Graphs of μ and μ^* for various cases of N and T are shown in Figures 2.6a, 2.6b for $(N,T) = (5,3)$, 2.6c, 2.6d for $(N,T) = (10,5)$, 2.8a, 2.8b for $(N,T) = (27,4)$, and 2.8c, 2.8d for $(N,T) = (27,12)$.

In all these figures the estimation of ψ is not possible to the left of the vertical line or below the solid horizontal line, marked $1/N$ (where $k = 1$ and $\theta = 1$ in $\psi = k/N\theta$). We propose that estimation is restricted to within the plausible region, the visible portions of curves corresponding to a value for p shown in figures. The visible portion of each curve is determined by the lower bound for ψ, ψ_{LB}, defined in Section 2.3.3 and is based on the plausible region for estimating unbiased values for ψ when estimates exist, given by $\psi \geqslant \psi_{LB}$. Recall that the lower bound for ψ is defined by a weighted average function $\psi_{LB} = c_\varepsilon/p + (\frac{1}{N} - c_\varepsilon)$, where $c_\varepsilon = (1-\varepsilon^{1/N})/\theta$ and ε is set to 0.01.

For example, for $N = 5$, $T = 3$ and $p = 0.1$, the lower boundary for ψ is greater than 1, it is outside the parameter space outside the plot boundaries and is not visible in Figures 2.6a and 2.6b. This means that estimation for ψ when $p = 0.1$ is not possible: either the ML estimates do not exist, or the expectation function is non-monotone increasing or the parameter space is sparsely populated (if at all populated) causing strong bias. Recall the convex hull $\mathbb{Q}_E$ of the sample space for $N = 5, T = 3$ is sparse, Figure 2.3. The ψ lower bound for $p = 0.2$ is also greater than 1, so that no portion of the expectation curve appears in the figure, indicating that unbiased estimation for ψ is again not possible, see for example Figure 2.6b.

For some studies the mean function of μ, and its conditional version μ^*, are not monotone increasing. This will depend on the combination of the overall number of sites surveyed N and survey occasions T.

For some p, μ is non-monotonic in ψ. For example, for $N = 27, T = 12$ and $p = 0.1$ in Figure 2.7. For both the unconditional and conditional case (Figures 2.7c and 2.7d) the expectation function for occupancy decreases between $\psi = 0.1$ and 0.2. The function then increases again after $\psi = 0.2$. For values of $p \geq 0.2$ the mean function μ is monotonic increasing. The bias is not much of a problem in this case except if p is small, since all curves rapidly approach the diagonal line as p increases and the bias rapidly diminishes.The smallest ψ that give functions that are monotone increasing contribute to determining an appropriate lower bound of ψ, ψ_{LB}. The resulting curves for values of the mean functions greater than ψ_{LB} for $N = 27, T = 12$ are shown in Figures 2.8c and 2.8d.

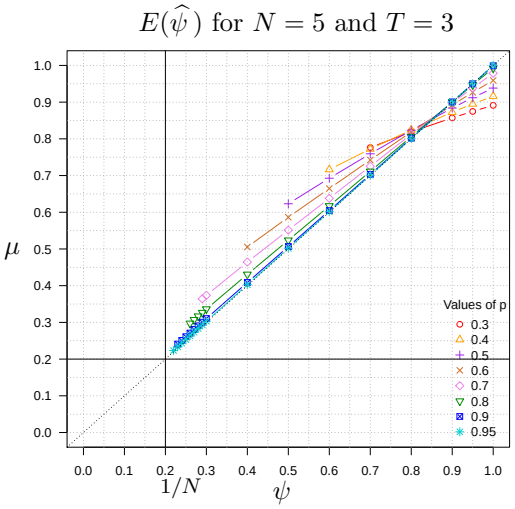
When there is a largish difference (or ratio) between N and T the function of expectation is non-monotonic for a greater range of values of p. We see that the curves are non-monotonic, for example, when $N = 27$, $T = 4$ and $p = 0.1, 0.2, \ldots, 0.5$ in Figure 2.7a and, for the conditional expectation function, Figure 2.7b. So in such cases, there needs to be a moderate level of detectability ($p > 0.5$) before the expectation function is monotonic increasing, $\psi > 0.4$. The functions that are above ψ_{LB} for given p are shown in Figures 2.8a and 2.8b.

Further, the bias is generally greater when the difference between N and T is larger. For example, when $N = 27$ and for the same values of ψ and p in Figures 2.8a and 2.8c the bias is considerably greater when $T = 4$ compared to when $T = 12$. Also, the bias persists for a larger range for ψ when T is

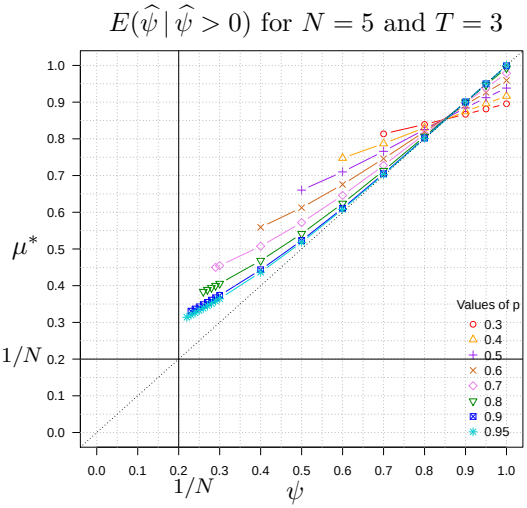
considerably smaller than N. For example when $N = 27, T = 4$ and $p = 0.2$ the estimator is more biased over the range of ψ compared to the same case when $T = 12$. Other studies have explored the dependence of the expectation on N and T, for example see Wintle et al. (2004); Guillera-Arroita et al. (2010); Karavarsamis et al. (2013).

On the other hand, when $N = 5$ and T is not much smaller e.g. when $T = 3$ the function of the expectation for occupancy is monotonic for all values of p and the entire range of ψ in [0,1] (figure not shown here), whereas the same is not true when $(N = 27, T = 4)$ or $(N = 27, T = 12)$ as seen in Figure 2.7. When $(N = 10, T = 5)$ the function is monotonic for the unconditional case but is non-monotonic for the conditional expectation function for ψ, $\psi > 0$ (figure not shown here). Non-monotonicity would be a problem, but this is avoided by the cut-offs at the ψ lower limit i.e. remaining within the plausible region. Compare Figure 2.7 which ignores the plausible region to Figure 2.8, once the plausible region is applied. In general, the magnitude of N and T individually, as well as their relative magnitude, will affect monotonicity of the occupancy expectation function. For example, if N is much greater than T ($N \gg T$), then monotonicity is almost surely achieved. However, if N and T are close in value, and neither is very large, then it is likely that there is a problem with monotonicity. So that the expectation function is non-monotonic for $N = 27$ and $T = 4$ or 12, i.e. where $N \gg T$ in both cases (see Figure 2.7).

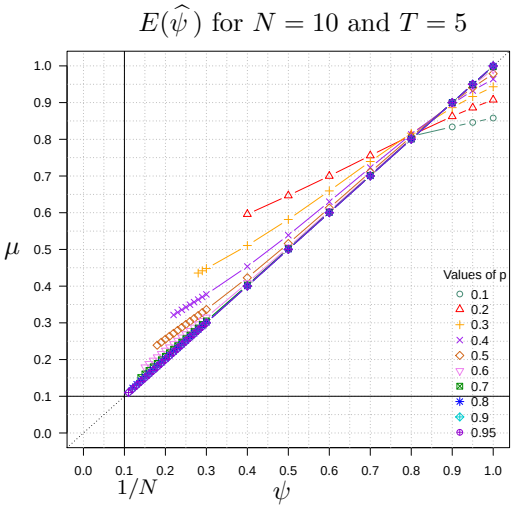
When the expectation is adjusted by $p(0, 0)$, effectively excluding the $(x, k) = (0, 0)$ case, the conditional mean function μ^* varies little in terms of the bias compared to μ. So for $N = 27, T = 4$ and for the same values of p and ψ in Figures 2.8a and 2.8b there is generally little difference in the bias between the curves for μ and μ^*, when we consider the plausible estimates for ψ i.e. for the visible section of the curves that is above the lower bounds for ψ for each value of p, $\psi \geqslant \psi_{LB}$. Similar comments may be made for the case $N = 5$ and $T = 3$ that μ and μ^* are not very different for values within the plausible region i.e. the visible section of the curves determined by $\psi \geqslant \psi_{LB}$.



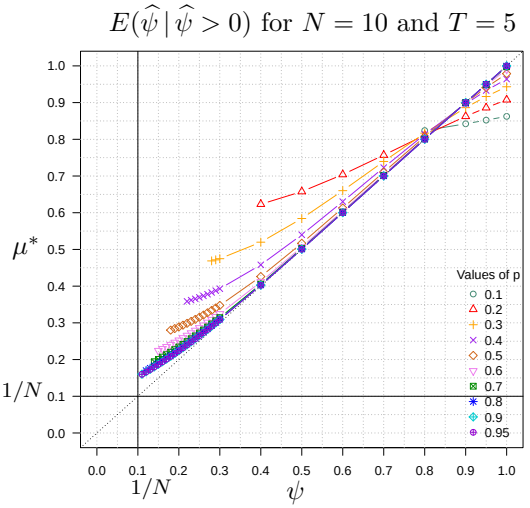
(a) Exact unconditional expectation for $\widehat{\psi}$.



(b) Exact conditional expectation for $\widehat{\psi}$.

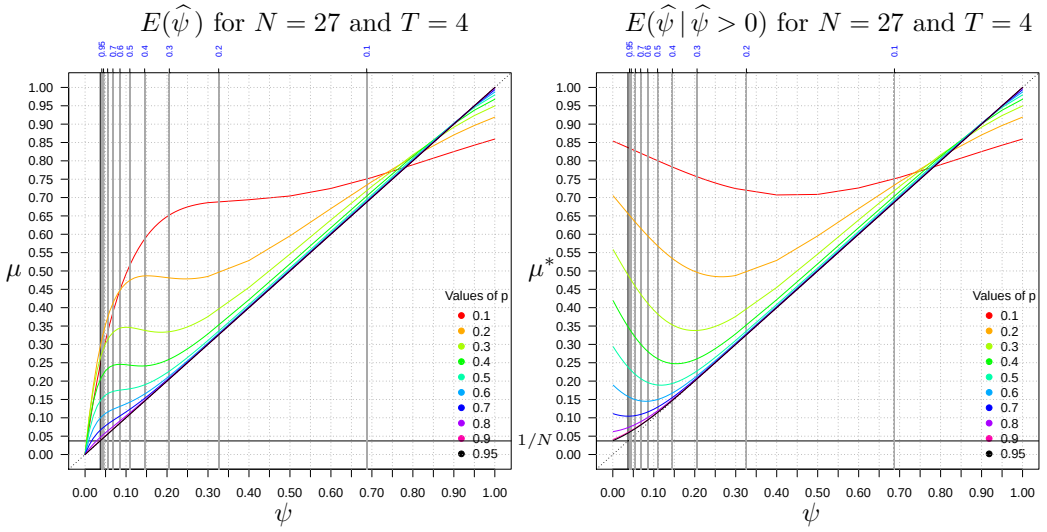


(c) Exact unconditional expectation for $\widehat{\psi}$.

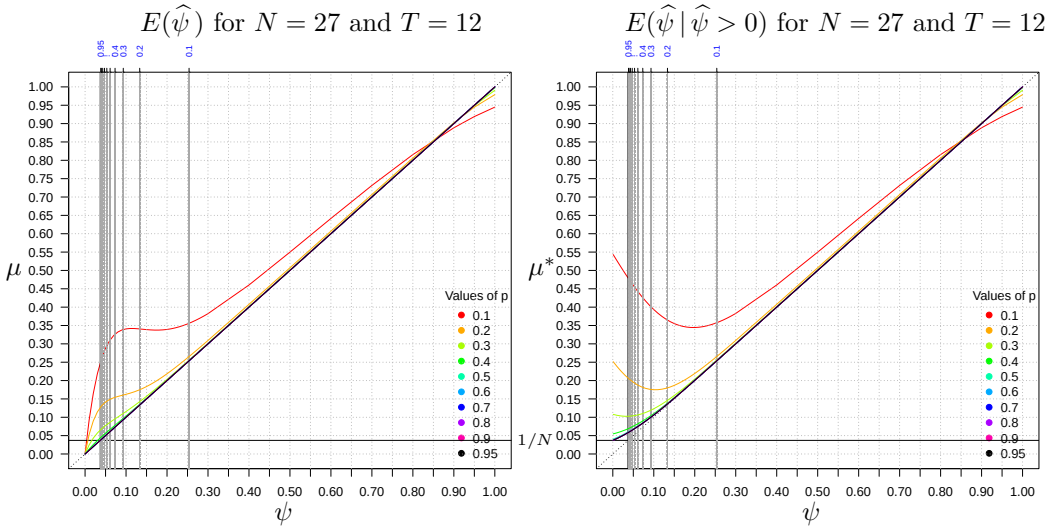


(d) Exact conditional expectation for $\widehat{\psi}$.

Figure 2.6: The exact unconditional and conditional expectations for occupancy when $(N = 5, T = 3)$ and $(N = 10, T = 5)$, respectively. We show the portion of the curves determined by $\psi \geqslant \psi_{LB}$ for the p in the legend.

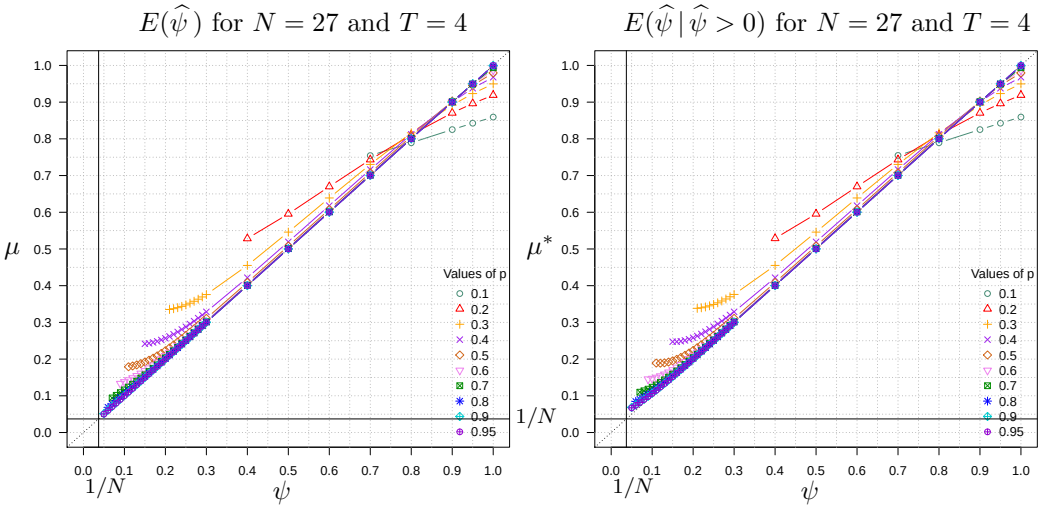


(a) Exact unconditional expectation for $\widehat{\psi}$. (b) Exact conditional expectation for $\widehat{\psi}$.

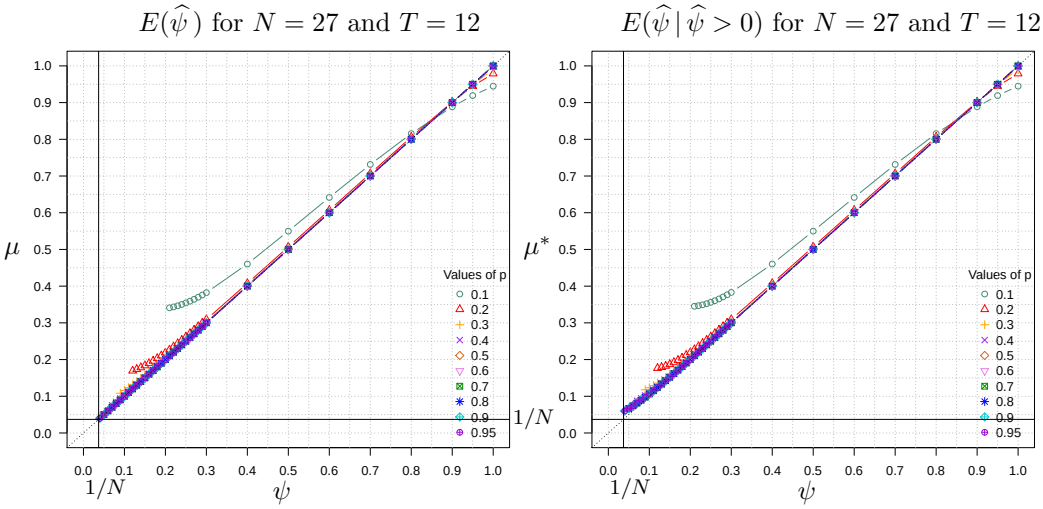


(c) Exact unconditional expectation for $\widehat{\psi}$. (d) Exact conditional expectation for $\widehat{\psi}$.

Figure 2.7: The full curves for the exact unconditional and conditional expectations for occupancy when $N = 27, T = 12$. The dashed vertical lines mark the ψ lower bounds ψ_{LB} for the values of p.



(a) Exact unconditional expectation for $\widehat{\psi}$. (b) Exact conditional expectation for $\widehat{\psi}$.



(c) Exact unconditional expectation for $\widehat{\psi}$. (d) Exact conditional expectation for $\widehat{\psi}$.

Figure 2.8: The exact unconditional and conditional expectations for occupancy when $(N = 27, T = 4)$ and $(N = 27, T = 12)$, respectively. We show the portion of the curves determined by $\psi \geqslant \psi_{LB}$ for the p in the legend.

2.5.4 The exact variance of $\widehat{\psi}$

Using the pmf of (X, K), since $\widehat{\psi} = \widehat{\psi}(X, K)$, it follows that the expectation for $\widehat{\psi}^2$ is given by

$$\lambda(\psi, p) = E(\widehat{\psi}^2 \,|\, \psi, p) = \sum_{k=0}^{N} \sum_{x=k}^{Tk} \widehat{\psi}^2(x, k)\; p(x, k). \tag{2.43}$$

And hence that

$$\sigma^2(\psi, p) = \text{var}(\widehat{\psi} \,|\, \psi, p) = \lambda(\psi, p) - \mu(\psi, p)^2, \tag{2.44}$$

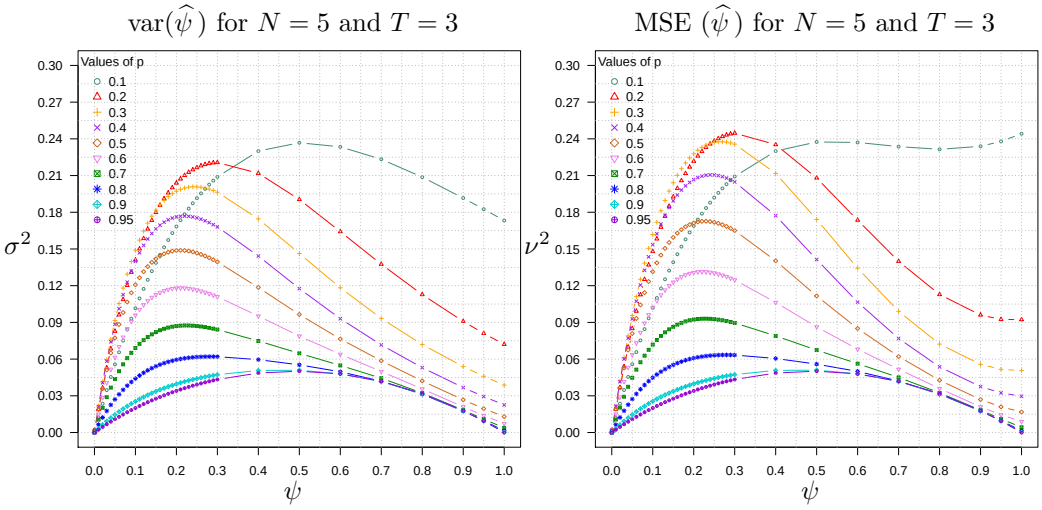
$$\nu^2(\psi, p) = \text{MSE}(\widehat{\psi} \,|\, \psi, p) = \lambda(\psi, p) - 2\psi\mu(\psi, p) + \psi^2. \tag{2.45}$$

Of course, $\text{MSE}(\widehat{\psi} \,|\, \psi, p) = \text{var}(\widehat{\psi} \,|\, \psi, p) + \text{Bias}(\widehat{\psi} \,|\, \psi, p)^2$, so that the MSE and variance of $\widehat{\psi}$ will not be equal i.e. $\text{Bias}(\widehat{\psi} \,|\, \psi, p) = \mu(\psi, p) - \psi$.

These results enable computation of the exact values of the variance and mean squared error — and hence comparison with the expressions for the asymptotic variance.

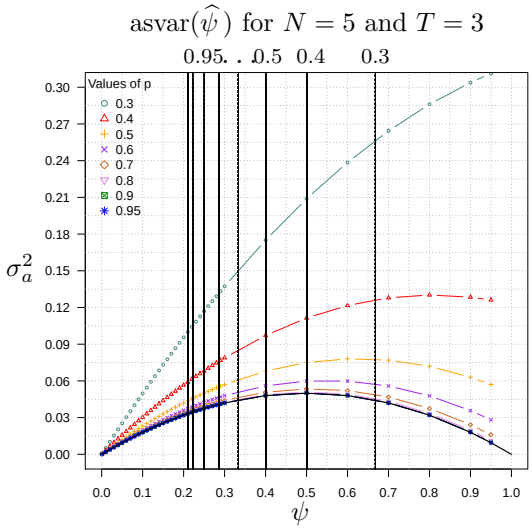
As an example, compare the plots for the asymptotic variance and exact variance for $N = 5, T = 3$ (Figures 2.9c and 2.9a) and $N = 27, T = 4$ (Figures 2.10c and 2.10a). In all plots of the asymptotic variances, we omit the curves for $p = 0.1$ as these gave results which were unreasonable. The variance functions for the same N and T are plotted on the same scale to make these comparable. So for some values of p the function may not appear on the diagram. In each case we address why these may be ignored.

The asymptotic variance for $N = 5, T = 3$ is shown in Figure 2.9c, where the exact variance (at $p = 1$) is marked by the dashed curved line and overlaps the asymptotic variance for $p = 0.95$. Its exact variance, for the same values of p, is plotted in Figure 2.9a and a plot of the MSE of $\widehat{\psi}$ is in Figure 2.9b. What is striking in this case is that the asymptotic variance performs poorly for small N, T.



(a) Exact variance for $\widehat{\psi}$, $\sigma^2(\psi, p)$.

(b) Exact MSE for $\widehat{\psi}$, $\nu^2(\psi, p)$.



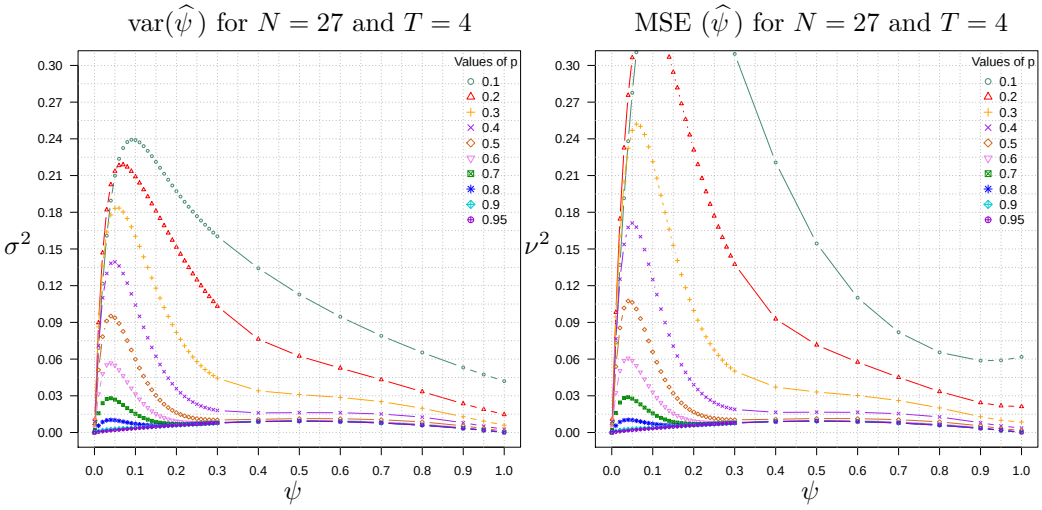
(c) Asymptotic variance for $\widehat{\psi}$, $\sigma_a^2(\psi, p)$.

Figure 2.9: The exact, and asymptotic, variance and exact MSE for occupancy, for $N = 5$ and $T = 3$. We show the curves for values of p in each legend. The dashed vertical lines mark the ψ lower bounds ψ_{LB} for the values of p. In Figure c) the exact variance for perfect detection $p = 1$ is shown by the solid black curve.

For $N = 27$ and $T = 4$ the asymptotic and exact variances are given in, respectively, Figures 2.10c and 2.10a, and 2.10b shows the MSE of $\widehat{\psi}$. For this case, when N is considerably larger than T, the asymptotic variance seems to misjudge the size and shape of the error when compared to the exact variance. For when p is between 0.2 and 0.8 the exact variance grows rapidly

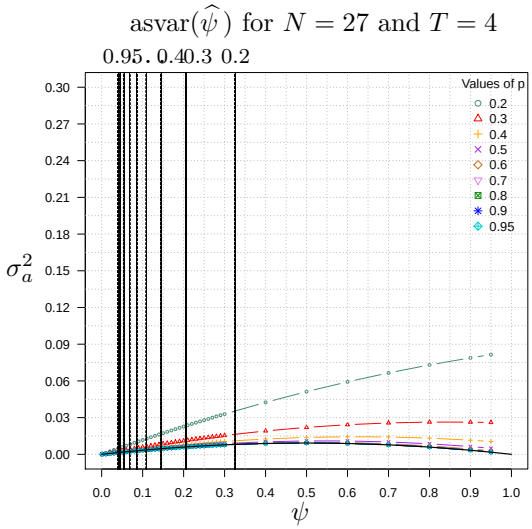
for ψ between 0 and 0.4. Then there is a steep descent beyond. The same characteristic is not present in the asymptotic variance. For this reason we use the ψ lower bounds to indicate where the sensible estimates are found. In general the shape of the two plots between the asymptotic and exact variances are distinctly different for small p. However for larger p and ψ they appear to be close. Overall, the asymptotic variance appears to perform poorly except for large values of p. Instead, we recommend using the standard error, i.e. $se(\widehat{\psi}) = \sqrt{\sigma^2(\psi, p)}$. The standard error is the estimate of the standard deviation (sd) obtained by substituting the estimates $\widehat{\psi}, \widehat{p}$ into the functional formula for the exact standard deviation, i.e. $\sigma(\widehat{\psi}, \widehat{p})$, where $sd(\widehat{\psi}) = \sigma(\psi, p)$. In other words, σ denotes a function of two variables.

For example, Figure 2.10 shows that when $N = 27, T = 4$, $p = 0.3$ and $\psi = 0.4$ the asymptotic variance for $\widehat{\psi}$ is much smaller that the actual (exact) variance, $\text{asvar}(\widehat{\psi})$ is approximately 0.018, whereas the exact variance is $\text{var}(0.4, 0.3) = 0.034$ and the (exact) MSE is approximately 0.035. Similarly, this may be observed from Figure 2.9 when $N = 5$ and $T = 3$. The case studies in Section 2.7 calculate the asymptotic (ase) and exact (se) standard error. In cases where p is large, the ase and se are close, as expected. In conclusion, the asymptotic variance for $\widehat{\psi}$ seriously underestimates the variance of $\widehat{\psi}$, and even more, underestimates the MSE. This means that the confidence interval for $\widehat{\psi}$ will be too narrow.



(a) Exact variance for $\widehat{\psi}$, $\sigma^2(\psi, p)$.

(b) Exact MSE for $\widehat{\psi}$, $\nu^2(\psi, p)$.



(c) Asymptotic variance for $\widehat{\psi}$, $\sigma_a^2(\psi, p)$.

Figure 2.10: The exact, and asymptotic, variance and exact MSE for occupancy, for $N = 27$ and $T = 4$. We show the curves for values of p in each legend. The dashed vertical lines mark the ψ lower bounds ψ_{LB} for the values of p. In Figure c) the exact variance for perfect detection $p = 1$ is shown by the solid black curve.

2.6 The Bias of the occupancy estimator $\widehat{\psi}$

2.6.1 Exact bias

In this section we examine the exact bias of $\widehat{\psi}$, based on $E(\widehat{\psi}) = \mu$ and $E(\widehat{\psi} \,|\, \widehat{\psi} > 0) = \mu^*$, as derived in the previous section; and we use these expressions to obtain a corrected estimator with reduced bias.

We evaluate μ and μ^* exactly on a grid of values $(\psi_i, p_j) \in G$, where $G = \{\psi_1, \psi_2, \ldots, \psi_r\} \times \{p_1, p_2, \ldots, p_s\} = \{0, 0.001, 0.01, 0.02, (0.01) \ldots, 0.3, (0.1) \ldots, 0.9, 0.95, 1\} \times \{0, 0.05, 0.1, \ldots, 0.9, 0.95, 1\}$; and, if values of μ or μ^* for other values of (ψ, p) are required as part of a computational procedure, we use (bi-)linear interpolation between the grid points, as outlined in Appendix 2.9.

We focus on the bias-corrected estimates for $\widehat{\psi}$ for values of ψ and p which fall within the 'plausible region' — the visible portion of the curves in the figures to come. Effectively, this means the section of the bias curves that are determined by $\psi \geqslant \psi_{LB}$ for a particular value for p i.e. the visible sections of the curves, for example, in Figures 2.11a and 2.11c).

If we [assume that $\widehat{\psi}(0,0) = 0$ or] exclude the case $(x, k) = (0, 0)$, the conditional expectation $\mu^*(\psi, p)$ produces more biased estimates than $\mu(\psi, p)$ for some cases, for example, when $N = 5$ and $T = 3$, μ^* (Figure 2.11b) is more biased above the lower bound ψ_{LB} (i.e. the visible portion of the curves) compared to μ (Figure 2.11a). For example, in Figure 2.11a and $p = 0.1$ there is no section of the curve which falls within the plausible region, as the region is outside the parameter space for ψ i.e. $\psi_{LB}(0.1) > 1$ and thus is not visible. This means that the bias curve for $p = 0.1$ should be ignored. And, effectively the bias curve should also be ignored for $p = 0.2$ as its ψ_{LB} falls almost at $\psi = 1$. Whereas for $N = 27, T = 4$ and $p = 0.1$ of Figure 2.11d the plausible region for occupancy is defined from $\psi \geqslant 0.7$. However, the bias is large and negative, especially for the bias function above $\psi \approx 0.8$ and is pretty much the same for unconditional ψ, $\widehat{\psi}$ (Figure 2.11c).

Thus, the exact bias when considering the unconditional expectation of $\widehat{\psi}$ is $\text{Bias}(\widehat{\psi})$, defined as $\text{B}(\psi, p) = \text{Bias}(\widehat{\psi} \,|\, \psi, p) = E(\widehat{\psi}) - \psi = \mu - \psi$. For the conditional expectation, the bias is $\text{Bias}(\widehat{\psi} \,|\, \widehat{\psi} > 0, \psi, p)$ defined as $\text{B}^*(\psi, p) = E(\widehat{\psi} \,|\, \widehat{\psi} > 0) - \psi = \mu^* - \psi$.

We apply a bias correction for each of the unconditional and conditional ex-

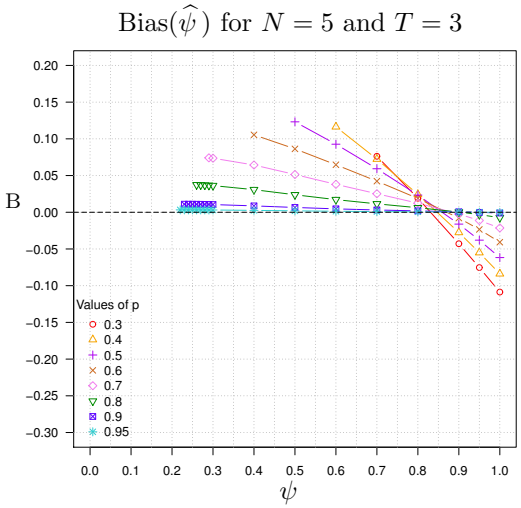
pectations to give $\widehat{\psi}_c$ and $\widehat{\psi}_c^*$. Our aim is to choose $\widehat{\psi}_c$ so that $\text{Bias}(\widehat{\psi}_c) \approx 0$ (which will give $E(\widehat{\psi}_c) \approx \psi$), and $\widehat{\psi}_c^*$ so that $\text{Bias}(\widehat{\psi}_c^* \,|\, \text{NZ}) \approx 0$. [$\text{NZ} = \{(x,k) : (x,k) \neq (0,0)\}$]. Then we find the expectation followed by the bias for each of the corrected estimators.

Although there are two standard ways of correcting for the bias (vertical and horizontal, described in Section 2.6.2) we use $\widehat{\psi}_c$ to denote the vertical bias correction method. We denote the bias for the corrected estimator as $\text{Bias}(\widehat{\psi}_c \,|\, \psi, p) = \text{B}_c(\psi, p)$ and for the conditional corrected estimator as $\text{Bias}(\widehat{\psi}_c^* \,|\, \widehat{\psi} > 0, \psi, p) = \text{B}_c^*(\psi, p)$.

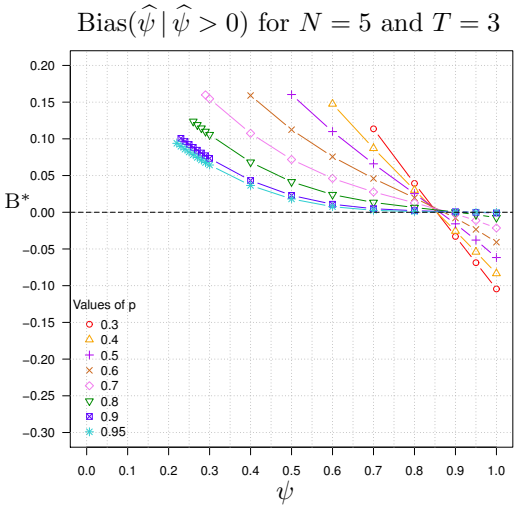
We examine and comment on the performance of the bias-corrected occupancy estimators, namely $\widehat{\psi}_c$ and $\widehat{\psi}_c^*$, in Section 2.7.

An underestimate for occupancy, or a negative bias, $\widehat{\psi}(x,k) - E(\widehat{\psi} \,|\, \psi, p) < 0$, may occur when occupancy is perfect, $\psi = 1$, but detectability is not i.e. $p < 1$. Practically, this means that we do not detect every animal ($\theta < 1$) that is actually present ($\psi = 1$). The quantity $\psi\theta$ in $p(x,k)$ affects $E(\widehat{\psi}) = \sum\sum \widehat{\psi}(x,k)p(x,k)$. So that, $\widehat{\psi} < E(\widehat{\psi})$ as $\psi \to 1$ since $\theta < 1$ in $\psi\theta$.

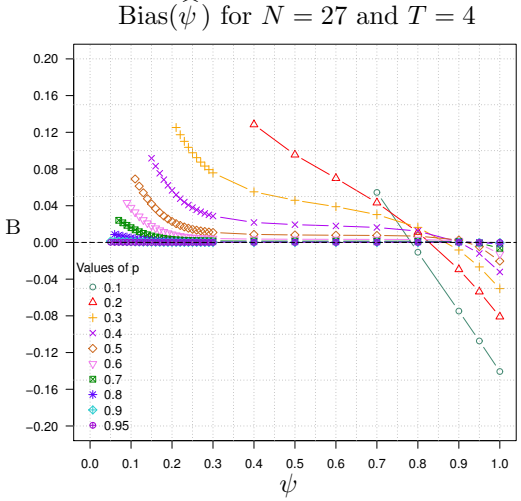
We adapted and applied an additional bias correction method based on a general method given in Bartlett (1955). The simplified version of the Bartlett (1955) bias correction method was proposed by Levin and Kong (1990) and evaluated by Hu and Lachin (2003, equation 2.2). However, we found that Bartlett's method was intractable in our case. The other bias correction method that is explored by Hu and Lachin (2003) is based on Lindsay's conditional likelihood method (Lindsay, 1982). This method does not apply in this case, since there is no sufficient statistic for p, the nuisance parameter.



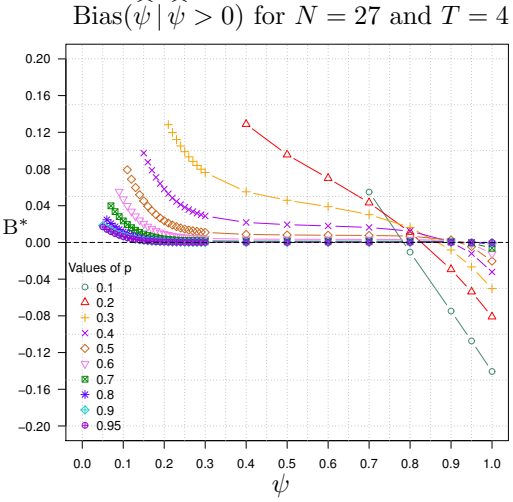
(a) Exact bias for $\widehat{\psi}$, $\mathrm{B} = \mu - \psi$.



(b) Exact bias for conditional $\widehat{\psi}$, $\mathrm{B}^* = \mu^* - \psi$.



(c) Exact bias for $\widehat{\psi}$, $\mathrm{B} = \mu - \psi$.



(d) Exact bias for conditional $\widehat{\psi}$, $\mathrm{B}^* = \mu^* - \psi$.

Figure 2.11: The exact bias from the unconditional and conditional expectation for occupancy when $(N = 5, T = 3)$ and $(N = 27, T = 4)$. We show the portion of the curves determined by $\psi \geqslant \psi_{LB}$ for the p in the legend.

2.6.2 Numerical bias correction based on exact expectation

Consider the one parameter case. Suppose T is an estimator of α. Suppose that its mean is $E(T) = \mu(\alpha)$, and if $\mu(\alpha) \neq \alpha$, then T is said to be biased. Its bias is given by $\mathrm{B}(\alpha) = \mu(\alpha) - \alpha$.

We could work this out using something like $E(T) = \sum p(\underset{\sim}{x} \,|\, \alpha)\, T(\underset{\sim}{x})$. This specifies the value of the mean of the estimator, for a specified value of α, and hence the bias $B(\alpha)$. This defines the bias function B.

Now, $T - \mathrm{B}(\alpha)$ is "unbiased", since $E(T - \mathrm{B}(\alpha)) = \mu(\alpha) - [\mu(\alpha) - \alpha] = \alpha$. But it's not an estimator, since it involves α. So, we replace $\mathrm{B}(\alpha)$ by $\mathrm{B}(T)$, assuming that T will be "close enough" to α to make the approximation reasonable.

Thus we propose a "bias-corrected" estimator $T_c = T - \mathrm{B}(T)$.

An alternative is to make the correction 'at the other end': i.e. to use T_c' such that $T_c' + \mathrm{B}(T_c') = T$.

Hepworth and Watson (2009) refer to these estimators as the vertical and horizontal corrections, based on a graphical representation of the correction.

In our situation, the idea is the same, but now we have two parameters: α and a nuisance parameter β.

The bias function is evaluated as a function of the parameters: $\mathrm{B}(\alpha, \beta)$, and we use this function to "correct" the estimator:

$$T_c = T - \mathrm{B}(T, U),$$

where U denotes an estimator of the nuisance parameter β; or in more familiar terms, the vertical method for bias adjustment yields the corrected estimates

$$\widehat{\psi}_c = \widehat{\psi} - \mathrm{B}(\widehat{\psi}, \widehat{p}).$$

We find $\mathrm{B}(\widehat{\psi}, \widehat{p})$ by evaluating $\mathrm{B}(\psi, p)$ at $(\psi, p) = (\widehat{\psi}, \widehat{p})$. For those values not on the grid G, we use (bi-)linear interpolation to obtain $\mathrm{B}(\widehat{\psi}, \widehat{p})$.

An alternative estimator, corresponding to the horizontal correction above, is given by:

$$T_c', \text{ such that } T_c' + \mathrm{B}(T_c', U) = T,$$

or in the present case:

$$\widehat{\psi}_c^*, \text{ such that } \widehat{\psi}_c^* + \mathrm{B}(\widehat{\psi}_c^*, \widehat{p}) = \widehat{\psi}.$$

The horizontal method requires solving the equation

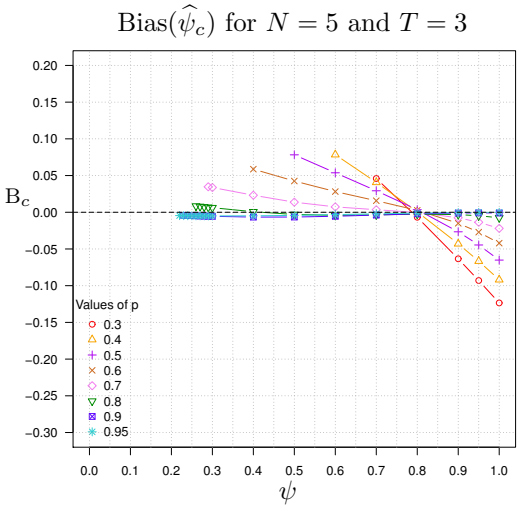
$$\widehat{\psi}_c^* + \mathrm{B}(\widehat{\psi}_c^*, \widehat{p}) = \widehat{\psi}, \quad \text{i.e.} \quad \mu(\widehat{\psi}_c^*, \widehat{p}) = \widehat{\psi}. \tag{2.46}$$

We use (bi-)linear interpolation for μ, where necessary. Specifically, values for ψ and p not found on the 'grid' will need to be found by interpolation.

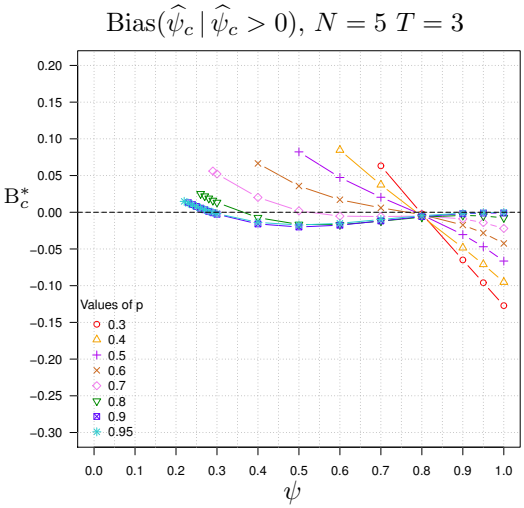
In practice, we use the table of $\mu(\psi, p)$ values previously calculated, and (bi-)linear interpolation as required, to find $m_{\widehat{p}}^{-1}(\widehat{\psi})$, where $m_{\widehat{p}}(\psi) = \mu(\psi, \widehat{p})$ (see Appendix 2.9).

We decided not to use the horizontal method as it was difficult to implement and for simple cases it turned out that it yielded no advantage over the vertical method.

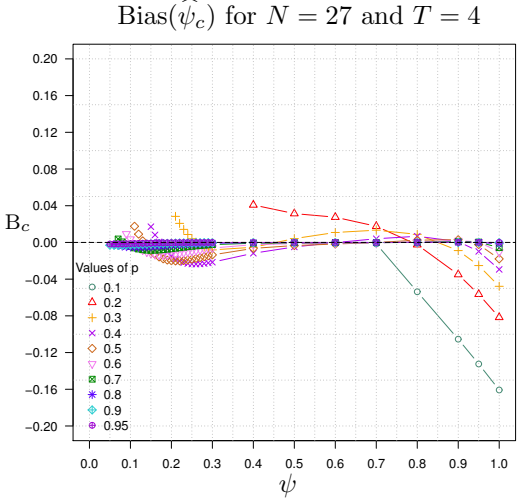
Some examples of the vertical bias correction method are shown in Figures: 2.12a, 2.12c, 2.12b, 2.12d.



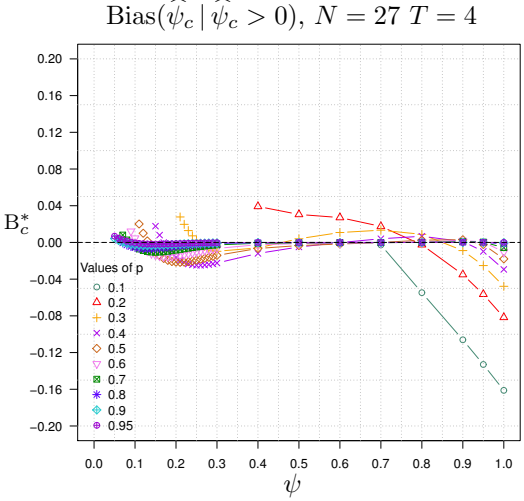
(a) Exact bias for $\widehat{\psi}_c$, $B_c = \mu_c - \psi$.



(b) Exact bias for $\widehat{\psi}_c$, $B_c^* = \mu_c^* - \psi$.



(c) Exact bias for $\widehat{\psi}_c$, $B_c = \mu_c - \psi$.



(d) Exact bias for $\widehat{\psi}_c^*$, $B_c^* = \mu_c^* - \psi$.

Figure 2.12: The exact bias for the bias-corrected occupancy of the unconditional and conditional expectation for occupancy when $(N = 5, T = 3)$ and $(N = 27, T = 4)$. We show the portion of the curves determined by $\psi \geqslant \psi_{LB}$ for the p in the legend.

2.7 Examples

We apply bias correction for occupancy estimation to studies that vary in size, occupancy, and detectability levels. We give a range of values for N and T, which would cover the range where data are likely to be observed and where the asymptotic behaviour may be a problem. The small study presented is a useful illustrative example of the procedure and is included for that reason. We restrict ourselves to moderate sample sizes, as larger sample sizes would entail a major project, and would be beyond the scope of this thesis. For each study, unconditional and conditional occupancy expectation is considered for bias correction. Corrected estimates are compared to uncorrected estimates with assistance from plots of MSE functions. Also, exact (numerical) variance is compared to asymptotic variance.

2.7.1 Example 1: Small N and T, moderate ψ, low p

A study is simulated in which $N = 5$ sites each are visited on $T = 3$ occasions, with a total of $x = 3$ detections made on $k = 2$ sites for $\psi = 0.5$ and $p = 0.3$. This simulation was done for illustrative purposes.

Estimates for uncorrected ($\widehat{\psi}$), corrected ($\widehat{\psi}_c$) and conditional corrected ($\widehat{\psi}_c^*$) occupancy, *se* and RMS, are presented in Table 2.1. Detectability was estimated to $\widehat{p} = 0.38$. The bias correction is too strong, the uncorrected estimate $\widehat{\psi} = 0.52$ is closest to $\psi = 0.5$. The standard error, $se = 0.382$, is too large for the occupancy estimate to be of much use. The bias correction is too strong, the *se* as well as the RMS are larger once the bias correction has been applied. The bias functions in Figures 2.11 and 2.12 show a larger bias with a correction, which even greater for the conditional expectation of ψ. Results for *se* and bias are summarised in the MSE plots in Figures: 2.13a, 2.13b, 2.13c and 2.13d.

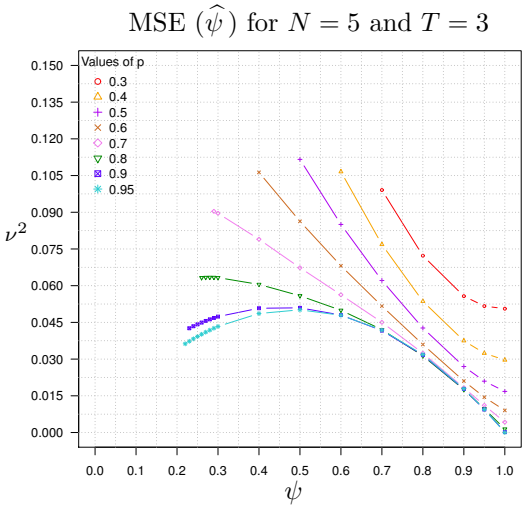
The (x, k) pair of this example lies on the boundary of the convex hull (Figure 2.3); L_k-line for $k = 2$ at $x = 3$. This means that the score equations did not apply for $\widehat{\psi}$ and $\widehat{p}$. Instead, estimates are found using equations given in Section 2.3.2 (occupancy estimates for the entire sample space are given in an example — page 21 of Section 2.3.2). Furthermore, estimates for ψ are not within the plausible region i.e. ψ estimates are not greater than the lower bound for ψ, ψ_{LB}. Thus, the bias function and MSE function do include estimates found here (Figures 2.11, 2.12 and 2.13).

	Estimate	se	ν
$\widehat{\psi}$	0.52	0.382	0.417
$\widehat{\psi}_c$	0.38	0.408	0.430
$\widehat{\psi}_c^*$	0.32	0.383	0.436

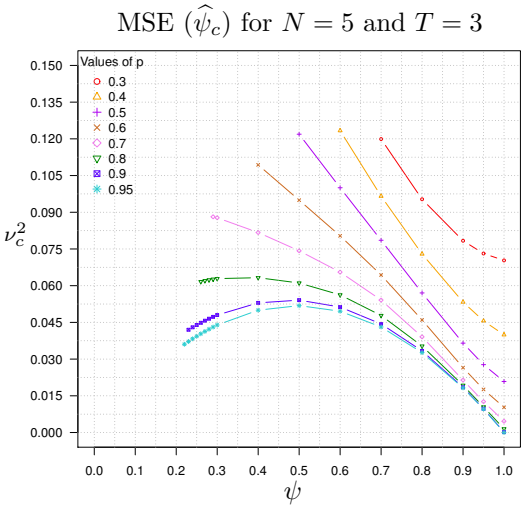
Table 2.1: Estimates and standard errors (estimate of the sd, se) and RMS (ν) for occupancy for a simulated history matrix from a study with $N = 5, T = 3$, and $x = 3, k = 2$ for $\psi = 0.5$ and $p = 0.3$. The detection estimate is $\widehat{p} = 0.38$.

The next set of MSE plots represent $(N = 5, T = 3)$ in Figures: 2.13a, 2.13b, 2.13c and 2.13d. These are helpful since they encapsulate the overall shapes of the MSE functions within the plausible region, for this study: $N = 5$ and $T = 3$.

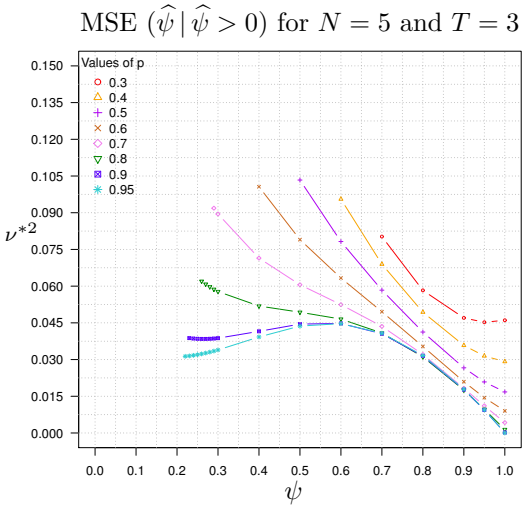
The exact expectation and estimates of variance, bias and MSE for occupancy are found for these scenarios, for parameter values on a grid defined by $(\psi_i, p_j) \in G$ as defined in Section 2.6. We use all valid combinations for (x, k), according to the criteria outlined previously (Section 2.3.1). Subsequently, unconditional and conditional estimates for occupancy are adjusted by the amount of bias obtained from the exact expectation of $\widehat{\psi}$. Where necessary, (bi-)linear interpolation is applied to calculate bias of occupancy from estimates of ψ and p $(\widehat{\psi}, \widehat{p})$ not on the grid G of parameter values (ψ_i, p_j).



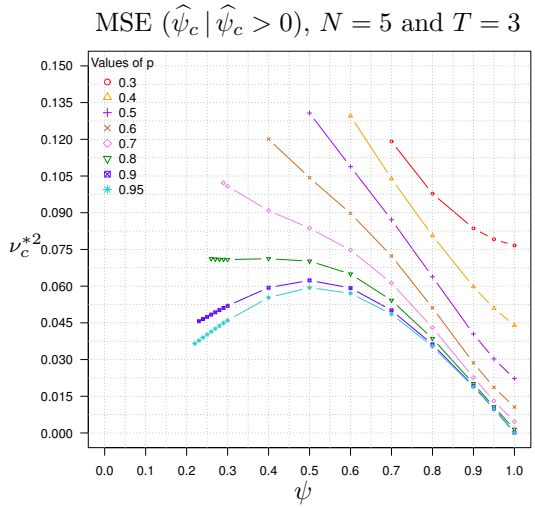
(a) Exact MSE for $\widehat{\psi}$, $\nu^2 = \sigma^2 + \mathrm{B}^2$.



(b) Exact MSE for $\widehat{\psi}_c$, $\nu_c^2 = \sigma_c^2 + \mathrm{B}_c^2$.



(c) Exact MSE conditional $\widehat{\psi}$, $\nu^{*2} = \sigma^{*2} + \mathrm{B}^{*2}$.



(d) Exact MSE for $\widehat{\psi}_c^*$, $\nu_c^{*2} = \sigma_c^{*2} + \mathrm{B}_c^{*2}$.

Figure 2.13: The exact MSE for the uncorrected and bias-corrected occupancy of the unconditional and conditional expectation for occupancy when $N = 5$ and $T = 3$. We show the portion of the curves determined by $\psi \geqslant \psi_{LB}$ for $p = 0.3, \ldots, 0.9, 0.95$.

2.7.2 Example 2: Frogs

We apply bias corrections to estimators for occupancy of a frog data set. This data is described in Chapter 3, Section 3.3.2 (and in Karavarsamis et al. (2013)). A complete detection history matrix for $N = 27$ sites each surveyed on $T = 4$ occasions, shown in Table 2.2, arise from a larger presence-absence study conducted on the the endangered growling grass frog (GGF) (*Litoria raniformis*) in southern Victoria, Australia (Heard et al., 2006). Detection histories were collected on catchment sites along the Merri Yuroke and Darebin Creeks on the northern outskirts of the Melbourne metropolitan area, each of which was visited multiple times.

	Survey 1	Survey 2	Survey 3	Survey 4
Site 1	0	0	0	0
Site 2	0	0	0	0
Site 3	0	0	0	0
Site 4	0	0	0	0
Site 5	0	0	0	0
Site 6	0	0	0	0
Site 7	0	0	0	0
Site 8	1	1	1	1
Site 9	0	0	1	1
Site 10	1	1	1	1
Site 11	0	1	0	0
Site 12	0	1	0	0
Site 13	1	1	1	1
Site 14	1	1	1	1
Site 15	1	1	1	1
Site 16	0	0	0	0
Site 17	0	0	0	0
Site 18	1	1	1	1
Site 19	0	0	0	0
Site 20	1	1	1	1
Site 21	1	1	1	1
Site 22	0	0	0	0
Site 23	0	1	1	1
Site 24	0	1	1	1
Site 25	0	0	0	1
Site 26	1	1	1	1
Site 27	0	0	0	0

Table 2.2: Detection histories for the growling grass frog. The 27 independent sites each were surveyed on 4 occasions within the 2002-2003 season.

Bias is corrected for unconditional and conditional occupancy estimates, results are shown in Table 2.3. Correcting estimates for ψ is ineffective, probably due to a high detectability estimate, $\widehat{p} = 0.782$. Occupancy is moderate for the frog history matrix, approximately a 56% chance that any site is occupied. The asymptotic standard error for unconditional occupancy, given by Equation (2.28), $ase(\widehat{\psi}) = \sqrt{\sigma_a^2(\widehat{\psi}, \widehat{p})} = 0.096$ is practically equal to the exact

standard error (an estimate from exact variance), $se = 0.095$. Considering high p and moderate ψ it is expected that the asymptotic variance is close to the exact variance, as is seen in Figure 2.10. For corrected unconditional and conditional occupancy estimators, Figure 2.14 allows for a comparison of MSE for the full range of ψ and p within $(0, 1)$. It shows that, for $p \geq 0.5$, MSE stabilises to ≈ 0.012 given $\psi \geq 0.3$, regardless of estimator. And that, for $p < 0.4$ the estimators are unreliable.

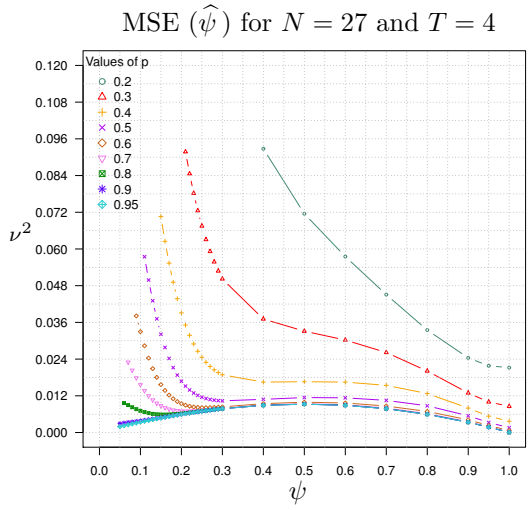
In comparison, estimates of occupancy obtained here from score equations are within the 95% credible interval of occupancy given in Heard et al. (2006); $\widehat{\psi} = 0.558$ $(0.408, 0.757)$ (Table 3). Although, the same cannot be said for detectability; $\widehat{p} = 0.696$ $(0.585, 0.757)$, compared to our estimate $\widehat{p} = 0.782$. Heard et al. (2006) use a Markov Chain Monte Carlo (MCMC) method in WinBUGS to fit a model to presence-absence that includes covariate information.

A direct maximisation of the likelihood results in estimates which are closer to those presented here than MCMC. In their study, Karavarsamis et al. (2013) use the `optim` function in R to produce $\widehat{\psi} = 0.56$ $(0.347, 0.787)$ and $\widehat{p} = 0.78$ $(0.577, 0.904)$. The interval limits result from the method of bootstrap-based studentised interval. Comparisons of several interval estimators are evaluated in the next chapter, Chapter 3 (the work of Karavarsamis et al. (2013)).

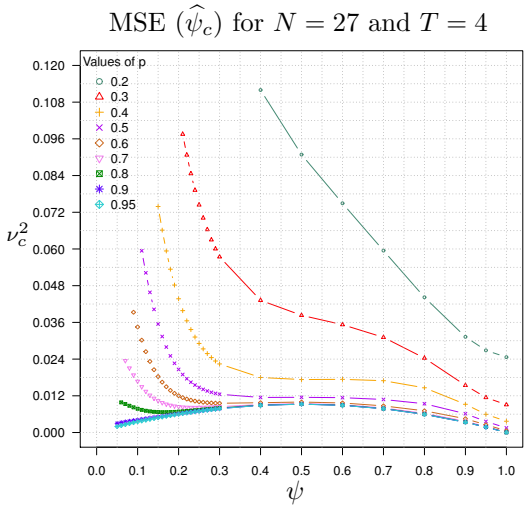
For this history matrix the number of sites with any detections is $k = 15$. And, the number of total (site-occasion) detections during the study is $x = 47$. The pair (x, k) does not fall onto an edge of the sample space: $k \neq N$, $x \neq k$ or $x \neq kT$. The score equations give estimates for occupancy and detectability: $\widehat{\psi} = k/(N\,\widehat{\theta}) = 15/(27\,\widehat{\theta})$ and $\widehat{p} = x\,\widehat{\theta}/(k\,T) = (47\,\widehat{\theta})/(15 \times 4)$. We solve $\widehat{\theta}$ numerically according to the method described in Section 2.3.2. A numerical solution to the function $\dfrac{(1-(1-\theta)^{1/T})}{\theta} = \dfrac{x}{kT}$ i.e. $\dfrac{(1-(1-\theta)^{1/4})}{\theta} = \dfrac{47}{15 \times 4}$ returns $\widehat{\theta} = 0.997$. Thus occupancy is estimated at $\widehat{\psi} = 0.557$ and detectability at $\widehat{p} = 0.782$.

	Estimate	se	ν
$\widehat{\psi}$	0.557	0.095	0.095
$\widehat{\psi}_c$	0.556	0.096	0.096
$\widehat{\psi}_c^*$	0.556	0.096	0.096

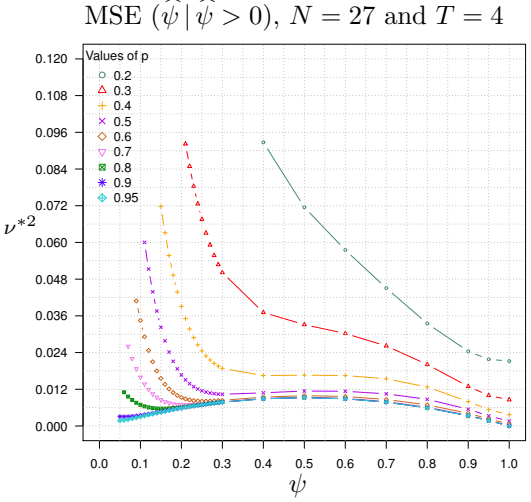
Table 2.3: Estimates, standard errors (estimate of the sd, se) and RMS (ν) for GGF ($N = 27, T = 4, x = 47$, $k = 15$), $\widehat{p} = 0.782$ and $ase(\widehat{\psi}) = 0.096$, for: $\widehat{\psi}$, unconditional corrected ($\widehat{\psi}_c$) and conditional corrected ($\widehat{\psi}_c^*$).



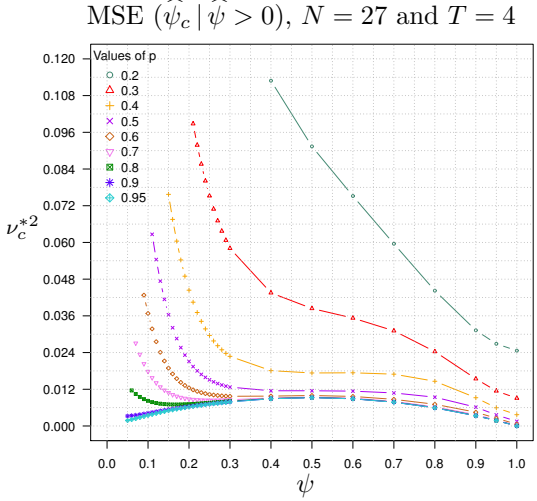
(a) Exact MSE for $\widehat{\psi}$, $\nu^2 = \sigma^2 + B^2$.



(b) Exact MSE for $\widehat{\psi}_c$, $\nu_c^2 = \sigma_c^2 + \mathrm{B}_c^2$.



(c) Exact MSE conditional $\widehat{\psi}$, $\nu^{*2} = \sigma^{*2} + \mathrm{B}^{*2}$.



(d) Exact MSE for $\widehat{\psi}_c^*$, $\nu_c^{*2} = \sigma_c^{*2} + \mathrm{B}_c^{*2}$.

Figure 2.14: The exact MSE for the uncorrected and bias-corrected occupancy of the unconditional and conditional expectation for occupancy when $N = 27$ and $T = 4$. We show the portion of the curves determined by $\psi \geqslant \psi_{LB}$ for $p = 0.2, \ldots, 0.9, 0.95$.

2.7.3 Example 5: Large N, moderate T, low ψ and p

This example investigates bias of occupancy for a study for a realistic scenario in which ($N =$) 55 sites are each surveyed ($T =$) 8 times, and low occupancy and detectability.

The estimation space $Q(N, T)$ for all possible (x, k) values for $N = 55$ and $T = 8$ are shown in Figure 2.15. The (x, k) pair which best matches moderate $\psi(= 0.35)$ and $p(= 0.36)$ occurs at $x = 57$ and $k = 19$. The L_k–line corresponding to this example is marked in the figure. The convex hull $\mathbb{Q}_E$ as well as the plausible region, with lower bound shown by ψ_{LB}, encompass the majority of the sample space, which is to be expected given the size of the study.

The (x, k) at $(57, 19)$ gives an occupancy estimate from the score equation of $\widehat{\psi} = k/(N\,\widehat{\theta}\,) = 19/(55 \times \widehat{\theta}\,)$ where $\widehat{\theta} = 0.974$ is obtained via numerical approximation. Thus, $\widehat{\psi} = 0.355$ and the score equation for detectability gives $\widehat{p} = 1 - (1 - \widehat{\theta}\,)^{1/T} = 0.365$. The bias-corrected occupancy estimates are shown in Table 2.4. We find that the estimates and standard errors for occupancy are very similar for all occupancy estimators: bias-corrected and uncorrected, conditional and unconditional. It is no surprise that bias correction had little effect, shown by investigations of this chapter (for example, see Section 2.6 which evaluates exact bias and numerical bias correction) and concluded that bias corrections, especially for large studies, are unnecessary. The MSE functions, in Figure 2.16, verify that bias is negligible even for moderate ψ and p. The asymptotic standard error $ase(\widehat{\psi}\,) = 0.066$ is not dissimilar from the exact standard error (i.e. the estimate of standard deviation), $se(\widehat{\psi}\,) = 0.063$. We showed earlier in this chapter that asymptotic results are quite reliable for large studies.

	Estimate	se	ν
$\widehat{\psi}$	0.355	0.063	0.063
$\widehat{\psi}_c$	0.353	0.063	0.063
$\widehat{\psi}_c^*$	0.353	0.063	0.063

Table 2.4: Estimates and standard errors (estimate of the sd, se) and RMS (ν) for occupancy for a study where $N = 55, T = 8, k = 19, x = 57$. The estimate for detection is $\widehat{p} = 0.366$. Shown are $\widehat{\psi}$, unconditional corrected ($\widehat{\psi}_c$) and conditional corrected ($\widehat{\psi}_c^*$) occupancy estimates. The asymptotic standard error is $ase(\widehat{\psi}\,) = 0.066$.

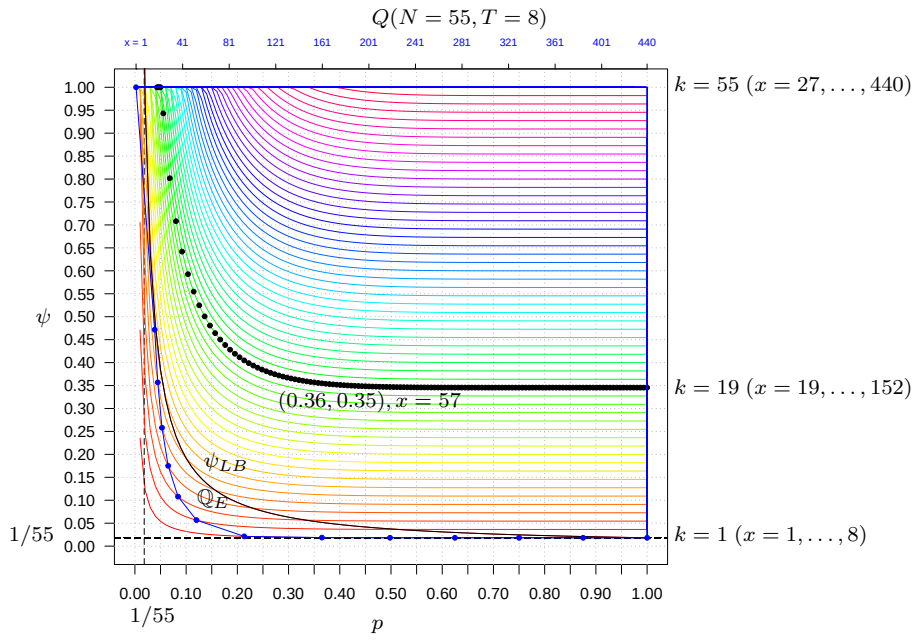


Figure 2.15: Convex hull $\mathbb{Q}_E$ for $(N = 55, T = 8)$. The solid lines mark the L_k-lines for $k = 1, \ldots, N$ and the bullets along the L_k mark the MLE corresponding to (x, k) where for each $k : x = k, k + 1, \ldots, kT$. The lower bound ψ_{LB} for the plausible region is shown.

2.7.4 Example 6: Large N, moderate T, high ψ and p

This example investigates whether high ψ and p will affect estimation for the same scenario presented above i.e. where $N = 55, T = 8$. We use simulation to generate a history matrix to illustrate the procedure of estimation and bias correction. Estimates are given in Table 2.5. Exact results give a standard error (an estimate of the standard deviation) $se = 0.054$.

	Estimate	se	ν
$\widehat{\psi}$	0.891	0.054	0.054
$\widehat{\psi}_c$	0.891	0.054	0.054
$\widehat{\psi}_c^*$	0.891	0.054	0.054

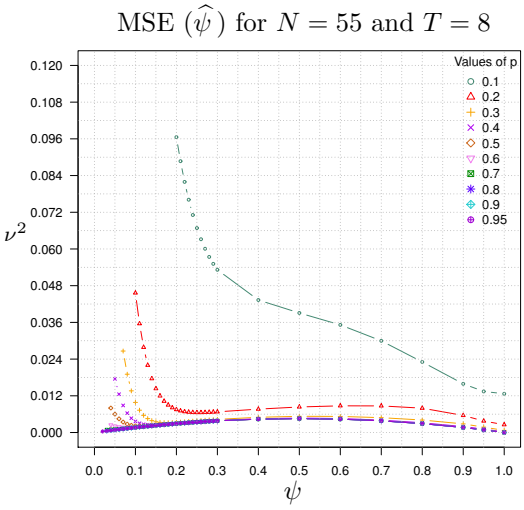
Table 2.5: Estimates and standard errors (estimate of the sd, se) and RMS (ν) for occupancy for a simulated history matrix, $N = 55, T = 8, k = 49, x = 291$, assuming $\psi = 0.8, p = 0.7$. The estimate for detection is $\widehat{p} = 0.742$. Shown are $\widehat{\psi}$, the unconditional corrected ($\widehat{\psi}_c$) and conditional corrected ($\widehat{\psi}_c^*$) occupancy estimates. The asymptotic standard error is $ase(\widehat{\psi}) = 0.042$).

Set $N = 55, T = 8, \psi = 0.8$ and $p = 0.7$. A vector of independent site-

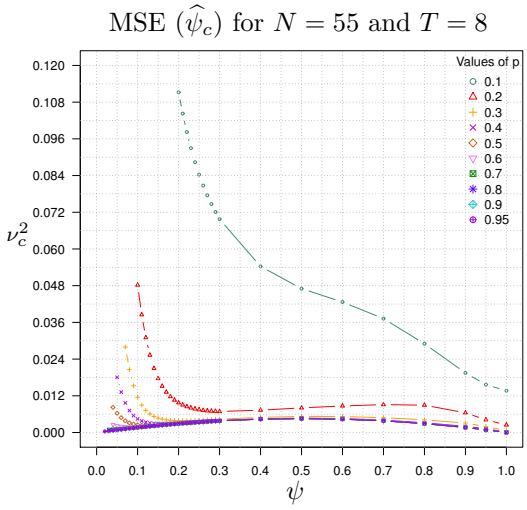
detections are generated from $D_{i\bullet} \stackrel{\mathrm{d}}{=} \mathrm{Bi}(T = 8, p = 0.7), i = 1, \ldots, N(= 55)$. A vector of independent site occupancy probabilities are generated from $P_i \stackrel{\mathrm{d}}{=} \mathrm{Bi}(\psi = 0.8)$. From this, the number of sites with at least one detection is $k = \sum_i^N p_i = 49$, and the total number of detections is given by $x = \sum_i^N x_{i\bullet} = 291$. Numerical approximation gives $\widehat{\theta} = 0.999$. Finally, the score equations give $\widehat{\psi} = 0.891$ and $\widehat{p} = 0.742$.

These two examples (Sections 2.7.3 and 2.7.4) verify that estimating occupancy for large N and moderate T is effectively unbiased. For example, the expectation function in Figure 2.8 (Section 2.5.3) and bias function in Figure 2.11 (Section 2.6). It is no surprise given that the sample space and convex hull for this study size are densely populated, bias correction is redundant in this case (Figure 2.15). The sample points are shown for the previous example at L_k-line $k = 19$. These figures of the expectation, and bias, function and the sample space, are testament to bias diminishing as N and T become larger. So much so, that, for this example, where detectability is high, estimates for occupancy effectively are equal to the parameter.

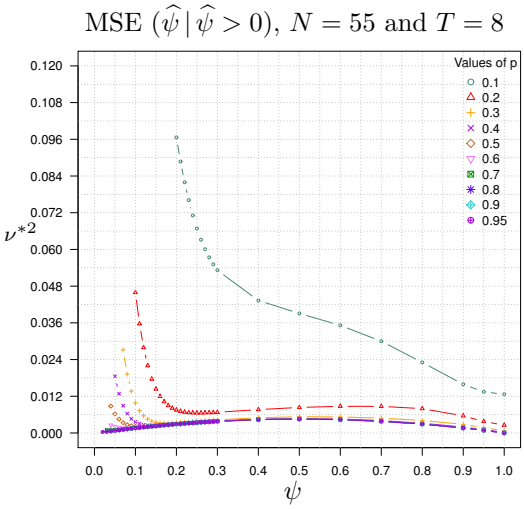
MSE functions in Figure 2.16 show that the MSE are small (< 0.012) for $\psi \geq 0.2$ and $p > 0.1$. They show a very small bias.



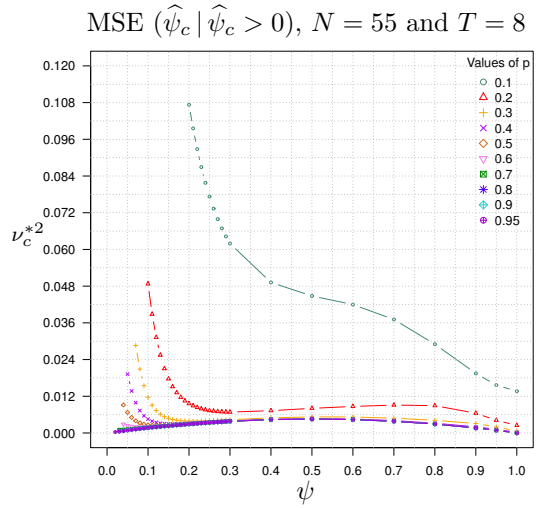
(a) Exact MSE for $\widehat{\psi}$, $\nu^2 = \sigma^2 + B^2$.



(b) Exact MSE for $\widehat{\psi}_c$, $\nu_c^2 = \sigma_c^2 + \mathrm{B}_c^2$.



(c) Exact MSE conditional $\widehat{\psi}$, $\nu^{*2} = \sigma^{*2} + \mathrm{B}_c^{*2}$.



(d) Exact MSE for $\widehat{\psi}_c^*$, $\nu_c^{*2} = \sigma_c^{*2} + \mathrm{B}_c^{*2}$.

Figure 2.16: The exact MSE for the uncorrected and bias-corrected occupancy of the unconditional and conditional expectation for occupancy when $N = 55$ and $T = 8$. We show the portion of the curves determined by $\psi \geqslant \psi_{LB}$ for $p = 0.1, \ldots, 0.9, 0.95$.

2.8 Appendix: Algorithms

This appendix gives a summary of methods, and algorithms, developed in this chapter for obtaining the expressions for appropriate estimators occupancy and detectability, the joint pmf of the sufficient statistics X and K, and the expectation for the occupancy estimator. We show how to obtain the variance and bias for occupancy based on the vertical method of bias-correction for occupancy described in Section 2.6.2. Bias-correction is shown for the conditional and unconditional expected occupancy. Appendix 2.9 outlines in detail the method of interpolating bias when the $(\hat{\psi}, \hat{p})$ are not on the grid G of values.

Our methods are described in this appendix with flowcharts. Figure 2.17 outlines the estimation method for occupancy and detectability, where x and k are determined by $x = 0, \ldots, N \times T$ and $k = 0, \ldots, N$. The sample space Ω, shown in Figure 2.1, determines which MLE equation will give the correct estimator. If the (x, k) co–ordinate is on an edge then an edge solution that we derived in Section 2.3.2 will give the correct estimator, otherwise the score equations are used if the co–ordinate sits within the boundaries of Ω.

To find a numerical approximation to $\hat{\theta}$ we use the `uniroot` function in R based on Brent's method, an algorithm without derivatives combining the bisection method, the secant method and inverse quadratic interpolation (Brent, 1973).

The joint pmf for X and K, $p(x, k)$ is calculated in parallel to the method for the estimators in Figure 2.17. The flowchart for finding $p(x, k)$ for a specific (ψ, p) pair is shown in Figure 2.18.

The estimate of occupancy generated from the method in Figure 2.17 is combined with the joint pmf of (X, K) from the flowchart in Figure 2.18 which gives the exact expectation, variance, bias and MSE for $\widehat{\psi}$. This results in the flowchart shown in Figure 2.19 and applies to a single (ψ, p) combination. To find the expectation, variance and bias for the conditional case, their expressions are adjusted by $1 - p(0, 0)$ as shown in Figures 2.20. These produce λ^*, μ^*, λ_c^* and μ_c^*. Then the MSE and bias are obtained for the conditional estimates. For both Figures 2.19 and 2.20, the bias is interpolated for occupancy estimates that are not on the grid G of values (ψ_i, p_j), defined in Section 2.6. The interpolation method in our case is described in next, in Appendix 2.9.

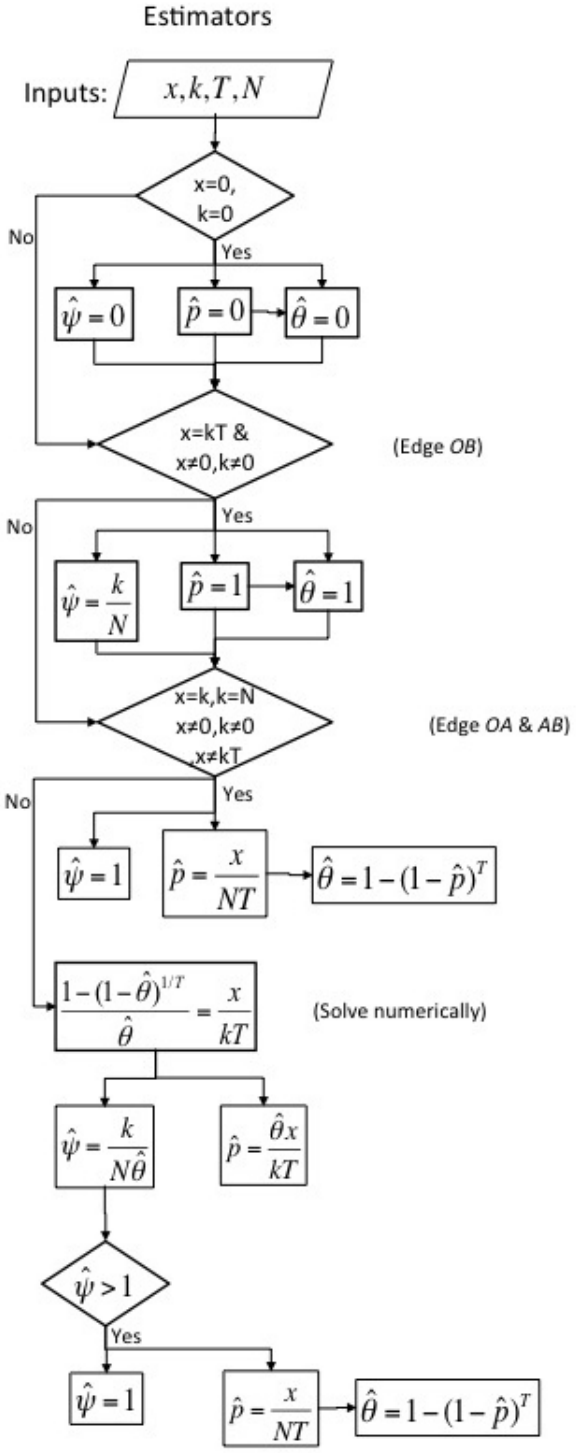


Figure 2.17: Flowchart of the algorithm for finding $\widehat{\psi}$, $\widehat{p}$.

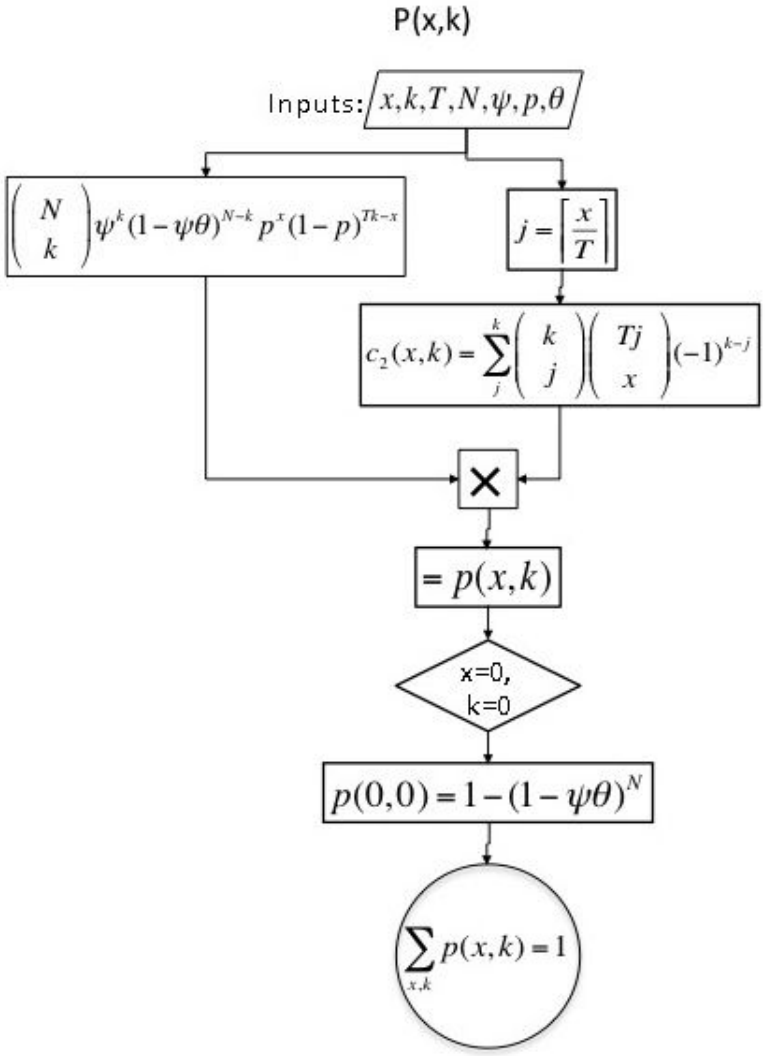


Figure 2.18: Flowchart of the algorithm for the pmf of X and K.

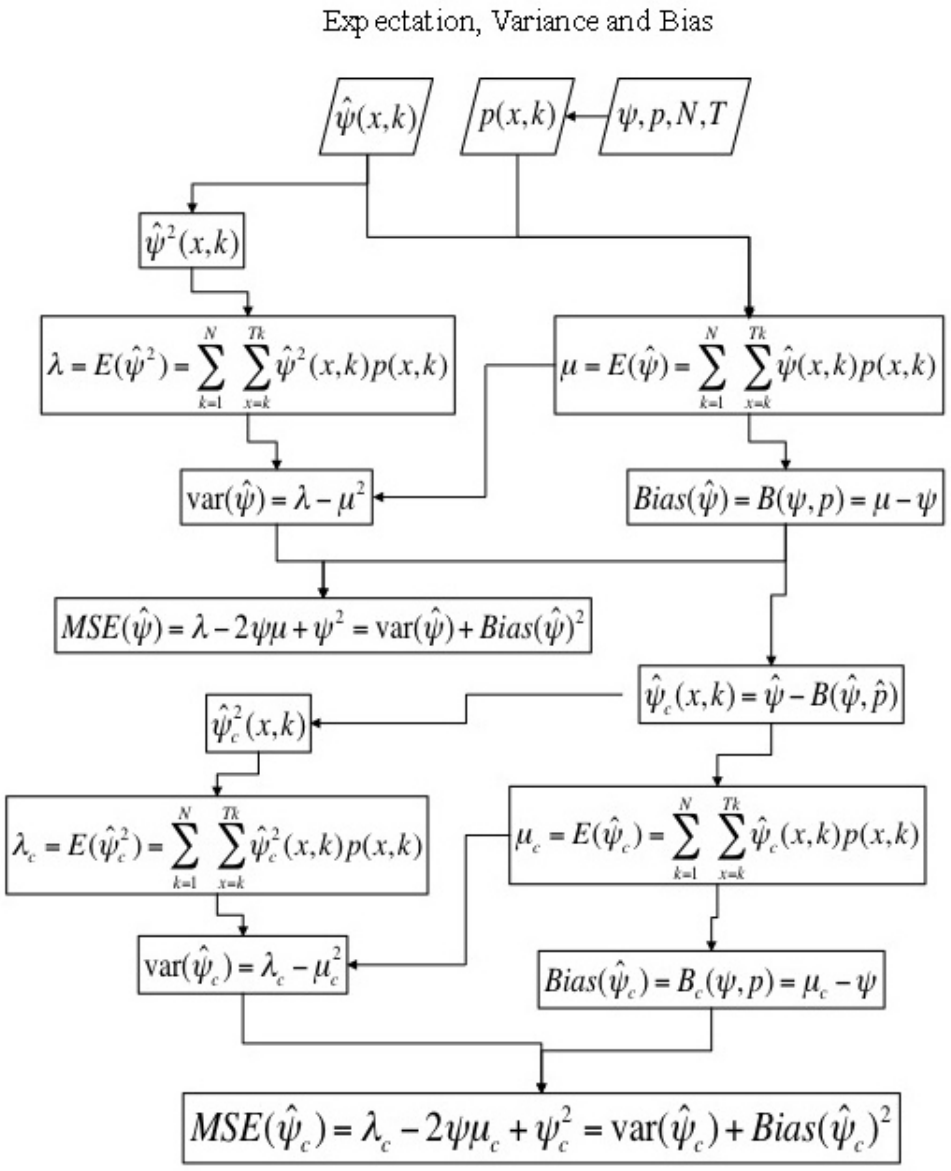


Figure 2.19: Flowchart for the expectation, variance, bias and MSE for unconditional $\widehat{\psi}$.

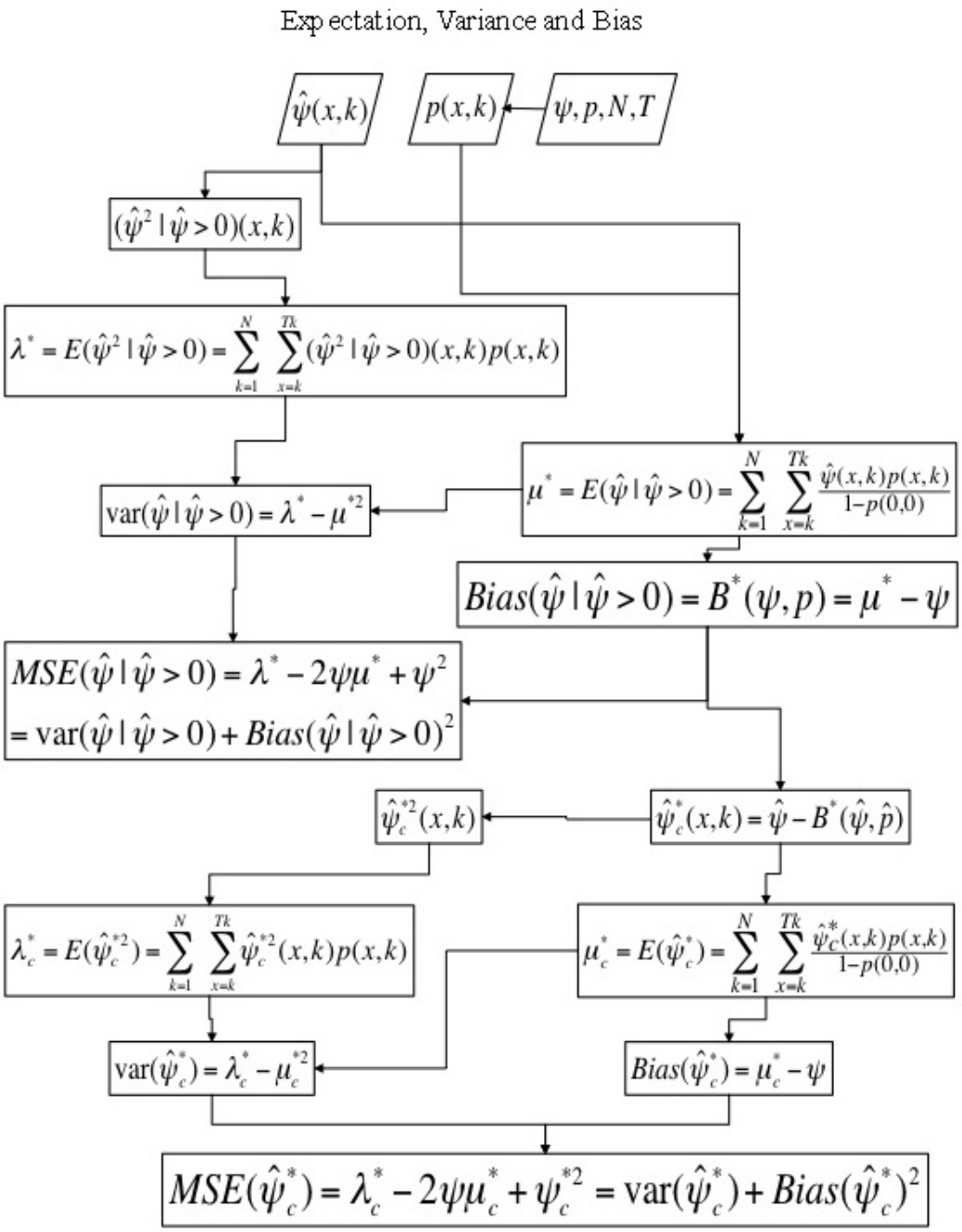


Figure 2.20: Flowchart for the conditional expectation, variance, bias and MSE for $\widehat{\psi}$.

2.9 Appendix: Bias Interpolation

Here we outline how we use the (bi-)linear interpolation method in our case.

In Section 2.6.2 we described the vertical and horizontal methods of adjusting bias to obtain the bias-corrected estimates for ψ. Linear interpolation is used if the ψ estimates fall on an edge of the 'grid', G, of predefined (ψ, p) co-ordinates. For the internal 'grid' estimates of ψ i.e. not on the (ψ, p) co-ordinates, we interpolate in two directions, both the ψ and p direction. In other words we use bi-linear interpolation for $\mathrm{B}(\psi, p)$. We adjusted the estimate of occupancy, $\widehat{\psi}$, by the numerical value of the bias using (bi-)linear interpolation to determine the amount by which a ψ estimate should be adjusted. The method of (bi-)linear interpolation for finding the amount of bias is shown in Figure 2.21.

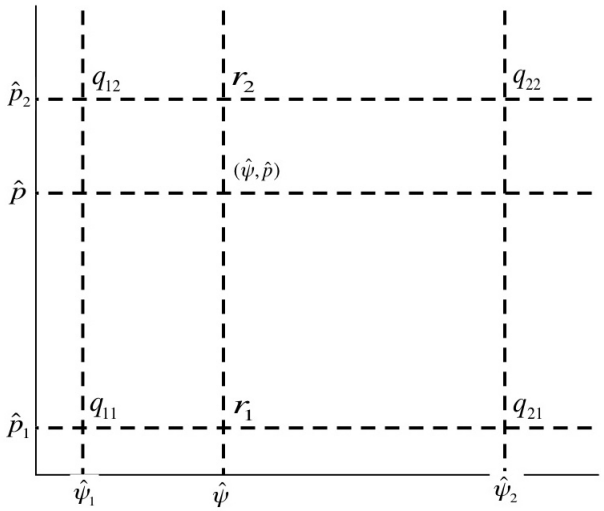


Figure 2.21: diagram of the bi-linear interpolation method.

For example, when $(\widehat{\psi}, \widehat{p})$ is not equal to a pre-defined (ψ, p) co-ordinate of G and is not on an edge of G, then an approximation for the bias is:

$$\begin{aligned}\mathrm{B}(\widehat{\psi}, \widehat{p}) \approx \quad & \frac{\mathrm{B}(q_{11})}{(\widehat{\psi}_2 - \widehat{\psi}_1)(\widehat{p}_2 - \widehat{p}_1)}(\widehat{\psi}_2 - \widehat{\psi})(\widehat{p}_2 - \widehat{p}) \\ & + \frac{\mathrm{B}(q_{21})}{(\widehat{\psi}_2 - \widehat{\psi}_1)(\widehat{p}_2 - \widehat{p}_1)}(\widehat{\psi} - \widehat{\psi}_1)(\widehat{p}_2 - \widehat{p}) \\ & + \frac{\mathrm{B}(q_{12})}{(\widehat{\psi}_2 - \widehat{\psi}_1)(\widehat{p}_2 - \widehat{p}_1)}(\widehat{\psi}_2 - \widehat{\psi})(\widehat{p} - \widehat{p}_1) \\ & + \frac{\mathrm{B}(q_{22})}{(\widehat{\psi}_2 - \widehat{\psi}_1)(\widehat{p}_2 - \widehat{p}_1)}(\widehat{\psi} - \widehat{\psi}_1)(\widehat{p} - \widehat{p}_1). \end{aligned} \tag{2.47}$$

Then the estimate for occupancy corrected for bias using the vertical method of adjustment is: $\widehat{\psi}_c = \widehat{\psi} - \mathrm{B}(\widehat{\psi}, \widehat{p})$.

When $(\widehat{\psi}, \widehat{p})$ falls on a vertex of $G : V_G = \{(0,0), (0,1), (1,0), (1,1)\}$ the bias is assumed zero: i.e. $\mathrm{B}(\psi, p) = 0$.

And, when $(\widehat{\psi}, \widehat{p})$ is on and edge of the grid then linear interpolation finds the amount of bias. For example, we find the amount of bias which corresponds to $(\widehat{\psi}, 0)$ (Figure 2.21) by amending Equation 2.47:

$$\mathrm{B}(\widehat{\psi}, \widehat{p}) = \mathrm{B}(q_{11})\frac{\widehat{\psi}_2 - \widehat{\psi}}{\widehat{\psi}_2 - \widehat{\psi}_1} + \mathrm{B}(q_{21})\frac{\widehat{\psi} - \widehat{\psi}_1}{\widehat{\psi}_2 - \widehat{\psi}_1} \tag{2.48}$$

A similar approach is used for the three remaining edges of Figure 2.21. As an example: $\mathrm{B}(0.57) = 0.3 * \mathrm{B}(0.5) + 0.7 * \mathrm{B}(0.6)$.

The same methods of linear bi-linear interpolation outlined here are used for finding the bias for $\widehat{\psi}_c$, $\widehat{\psi}^*$ and $\widehat{\psi}^*_c$.

When any of the occupancy estimators are greater than 1 or less than zero, say $\widehat{\psi} > 1$ or $\widehat{\psi} < 0$, we truncate to 1 and 0, respectively. This occurs usually as a result of floating point error, an accumulation of computational arithmetic errors.

2.10 Appendix: Simulation for Frogs

We give a detailed numerical example to illustrate methods outlined in Appendix 2.8 that are used to find estimates for occupancy in the case of the conditional and unconditional expectation for occupancy. And for the method of interpolation given in Appendix 2.9. Then we find their asymptotic variances, exact variances and finally their bias-corrected versions.

Let $\psi = 0.557$ and $p = 0.782$ for simulation of a detection history matrix. Simulate a vector of $(N=)$ 27 independent site-detections from $D_{i\bullet} \stackrel{\text{d}}{=} \text{Bi}(T = 4, p = 0.782)$, $i = 1, \ldots, N(= 27)$ (given in Section 2.2),

$$\{d_{i\bullet}\} : \{4, 4, 4, 1, 2, 3, 3, 3, 4, 2, 4, 3, 3, 3, 2, 2, 3, 3, 4, 3, 3, 4, 3, 3, 3, 4, 2\}.$$

A simulated vector for occupancy for site i, where $P_i \stackrel{\text{d}}{=} \text{Bi}(\psi = 0.557)$ (Section 2.2), is

$$\{p_i\} : \{0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1\}.$$

There are $k = \sum_i^N p_i = 13$ detected sites where the number of detections at each site i, given by $X_{i\bullet} = P_i D_{i\bullet}$, occurs with probability of occupancy $\psi = 0.557$,

$$\{x_{i\bullet}\} : \{0, 0, 0, 0, 0, 3, 0, 3, 4, 2, 4, 0, 3, 3, 2, 0, 3, 0, 4, 3, 0, 4, 0, 0, 0, 0, 2\}.$$

Thus in total, $x = \sum_i^N x_{i\bullet} = 40$ detections were recorded for this study. Occupancy and detectability estimates are obtained from the score equations (Equations (2.10)), since $(x, k) = (40, 13)$ is *internal* to the sample space Ω i.e. it is not on an edge of Ω (Figure 2.1).

The numerical solution $\widehat{\theta} = 0.997$ gives,

$$\widehat{\psi} = 0.483 \quad \text{and} \quad \widehat{p} = 0.767.$$

Table 2.6 summarises the estimates for occupancy and their standard errors.

For exact variance and bias, and by extension the MSE, it is necessary first to obtain μ, the exact expectation for ψ (Equation (2.40) or (2.41)). For this to occur, the joint probability for *every* (x, k) pair, $p(x, k)$, is to be calculated (Equation (2.39)). The estimate pair $(\widehat{\psi}, \widehat{p}) = (0.482, 0.767)$ is not a co-ordinate on the grid G of the parameter space (ψ, p), thus, exact variance and bias are approximated with bi-linear interpolation as outlined in Appendix 2.9.

The co-ordinates of the vertices of the smallest square that encloses $(\widehat{\psi}, \widehat{p}) = (0.482, 0.767)$ (see Figure 2.21) are:

$$q_{11} = (\widehat{\psi}_1, \widehat{p}_1) = (0.4, 0.7),$$
$$q_{12} = (\widehat{\psi}_1, \widehat{p}_2) = (0.4, 0.8),$$
$$q_{21} = (\widehat{\psi}_2, \widehat{p}_1) = (0.5, 0.7),$$
$$q_{22} = (\widehat{\psi}_2, \widehat{p}_2) = (0.5, 0.8).$$

The exact variance is found by replacing B for σ^2 in Equation (2.47) for interpolation. Then, the interpolated standard error (estimate of the *sd*) is $se(\widehat{\psi}) = 0.096$, which is equal to the asymptotic standard error, given by Equation (2.28), $ase(\widehat{\psi}) = \sqrt{\sigma_a^2(\widehat{\psi}, \widehat{p})} = 0.096$. As expected, the asymptotic standard error is a good approximation to the standard error *se* for large p, in this case $p > 0.7$. However, both are large in magnitude, which indicates that occupancy has not been estimated well.

An approximation of the exact bias for the occupancy estimate is found by interpolating the exact bias for the four co-ordinates: q_{11}, q_{12}, q_{21} and q_{22}:

$$\mathrm{B}(q_{11}) = \mathrm{B}(0.4, 0.7) = 0.00142,$$
$$\mathrm{B}(q_{12}) = \mathrm{B}(0.4, 0.8) = 0.0004307,$$
$$\mathrm{B}(q_{21}) = \mathrm{B}(0.5, 0.7) = 0.001374,$$
$$\mathrm{B}(q_{22}) = \mathrm{B}(0.5, 0.8) = 0.0004163.$$

Then, the method outlined in Appendix 2.9 gives the approximate bias, $\mathrm{B}(\widehat{\psi}, \widehat{p}) \approx 0.000737$, and the interpolated MSE is $\nu^2(\widehat{\psi}, \widehat{p}) \approx 0.009$ and its exact RMS is $\sqrt{\nu^2(\widehat{\psi}, \widehat{p})} \approx 0.096$ (see Table 2.6 for results of all estimators).

Then the vertical method (Section 2.6.2) corrects the occupancy estimate to become

$$\widehat{\psi}_c = \widehat{\psi} - \mathrm{B}(\widehat{\psi}, \widehat{p}) = 0.483 - 0.000737 = 0.482.$$

There is practically no difference, only a slight reduction is made by the correction but it is no closer to the actual occupancy parameter, $\psi = 0.557$. It has overcorrected and has further underestimated the actual occupancy. That is, the bias correction may cause the estimate to move from a positive bias, to a negative bias once corrected. For example, Figures 2.12c and 2.12d compared to the uncorrected bias estimate in Figures 2.11c and 2.11d. With a standard error of $se \approx 0.1$, we 'expect' that with $\psi = 0.58$, a rough probability interval is $0.38 \leqslant \widehat{\psi} \leqslant 0.78$ i.e. $\widehat{\psi} \pm 2 \times 0.1$.

In a similar way the bias-corrected estimate for the conditional case is, $\widehat{\psi}_c^* = \widehat{\psi} - \mathrm{B}^*(\widehat{\psi}, \widehat{p}) = 0.482$, where $\mathrm{B}^*(\widehat{\psi}, \widehat{p}) = \mathrm{Bias}(\widehat{\psi} \,|\, \widehat{\psi} > 0)$ and is approximated by its interpolated value. Its standard error is: $se(\widehat{\psi}_c^*) = 0.096$. Clearly there

is no difference in the corrected occupancy estimate between the conditional and unconditional ψ.

	Estimate	se	ν
$\widehat{\psi}$	0.483	0.096	0.096
$\widehat{\psi}_c$	0.482	0.096	0.096
$\widehat{\psi}_c^*$	0.482	0.096	0.096

Table 2.6: Estimates and standard errors (estimate of the sd, se) and RMS (ν) for occupancy for a simulated history matrix from the frog data with $N = 27, T = 4, x = 40$, $k = 13$, $\psi = 0.557$, and $p = 0.782$. The estimate for detection is $\widehat{p} = 0.767$. Shown are $\widehat{\psi}$, unconditional corrected ($\widehat{\psi}_c$) and conditional corrected ($\widehat{\psi}_c^*$) occupancy estimates. The asymptotic standard error is $ase(\widehat{\psi}) = 0.096$.

Chapter 3

Comparison of four interval estimators

The material in this chapter has been published in Karavarsamis et al. (2013).

3.1 Overview

Sample-based uncertainty about the parameter estimates must be reported to provide a complete picture of a study. The MacKenzie et al. (2002) likelihood function does not have a closed form solution for the variance associated with the probability of detection. Likelihood-based large-sample standard error estimates may not be appropriate for estimating the uncertainty, especially when sample sizes are small (MacKenzie et al., 2002). Consequently, MacKenzie et al. (2002) used ordinary nonparametric bootstrap estimators, also called the nonparametric basic bootstrap. However the basic bootstrap may not be appropriate for rare and clustered species distributions (Efron, 1982).

The goal of this chapter is to assess a suite of interval estimators. Here we compare the performance of three nonparametric, finite-population bootstrap-based interval estimators: 1) the basic bootstrap 2) the studentised bootstrap (or bootstrap-t) and 3) the percentile interval, against the normal approximation (or asymptotic method) (Davison and Hinkley, 1997, Section 5). The performances of the estimators are tested using two simulation studies. The first study is carried out under the assumption that there is no spatial clustering of the observations. The second study assesses the performance of the interval estimators when the observations show over-dispersion, using bootstrap-style

replicates from an ecological study.

These estimators are demonstrated using a presence-absence data set of the growling grass frog (*Litoria raniformis*), an endangered species found also in Southern Victoria, Australia.

Similarly to MacKenzie et al. (2002), we apply the nonparametric bootstrap resampling scheme proposed by Efron (1982) to the sites (rows) of the detection matrix to generate an estimate of the empirical distribution of the log-likelihood function (2.4). The resulting sample of bootstrap estimates for ψ and p inform the subsequent interval estimation. We used the `boot.ci` (Canty and Ripley, 2009) procedure in `R` (R Development Core Team, 2009) to provide the approximate confidence intervals.

The rest of the chapter is organised as follows, Sections 3.2 and 3.3 outline the methods used in the study i.e. the interval estimators we compared, and details of the simulation and cases studies. In Sections 3.4 we present results of the studies with corresponding tables and figures. The chapter ends with our conclusions (Section 3.5)

3.2 Interval estimators

The bootstrap sample variance estimate may overstate the precision for parameter values which are close to 1, leading to positively biased estimates. Similarly, the asymptotic variance from the normal approximation of large sample theory may not be appropriate for studies of small sample size. Furthermore, asymptotic normal approximation is inappropriate for populations that are rare or clustered, as is the case for our study population (MacKenzie et al., 2002).

The three bootstrap-based interval estimators we consider are derived from basic methods for confidence limits for a parameter η and compared to the large sample normal approximation. Suppose that H is an estimator for η and we seek an equi-tailed interval with tails of probability $\alpha/2$. The lower and upper boundaries of the $(1-\alpha/2)$ equi-tailed confidence interval are, respectively

$$\widehat{\eta}_{\alpha/2} = h - c_{(1-\alpha/2)}, \quad \widehat{\eta}_{(1-\alpha/2)} = h - c_{\alpha/2}, \tag{3.1}$$

where h is an estimate of the parameter η (or a realisation of H) and the

constants c are the appropriate quantiles of the distribution of H. We use the maximum likelihood estimates, $\widehat{\psi}$ and $\widehat{p}$, for h. Since the distribution of $H - \eta$ is unknown we consider approximate methods for estimating its quantiles.

3.2.1 Normal approximation

The simplest approach to constructing confidence limits applies a normal approximation $N(0, v)$ for $H - \eta$. The resulting lower and upper approximate confidence limits are, respectively,

$$\widehat{\eta}_{\alpha/2} = h - v^{1/2} z_{1-\alpha/2}, \quad \widehat{\eta}_{(1-\alpha/2)} = h + v^{1/2} z_{1-\alpha/2},$$

where the estimate h (in our case $\widehat{\psi}$ or $\widehat{p}$) is obtained from the original data, as is the asymptotic variance estimator v. The $z_{(1-\alpha/2)}$ critical value provides the $(1-\alpha/2)$ quantile from the inverse standard normal distribution, i.e. $z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$. This approach for estimating limits is potentially problematic when the parameter estimates are on, or near, the boundaries of the parameter space, which in this context are zero and one (Lebreton et al., 1992). This issue may be ameliorated to some extent by transforming the estimates onto an alternative scale that is not bounded between 0 and 1, e.g., logit or arcsin. The normal approximation, for a large-sample estimator, is based on asymptotic normality of the estimator $\widehat{\eta}$ (Lebreton et al., 1992; Davison and Hinkley, 1997, p.198). Consequently, we do not expect it to have good small-sample properties unless the sampling distribution of the parameter estimate is sufficiently close to normal, and it is likely to result in coverages that are inadequate. Given that we are dealing with detection histories characterised by asymmetry and sparseness, a common feature of clustered and rare populations, we do not expect to observe adequate coverages from the normal method. Restrictions of the normal method might further be explained by its degree of accuracy, which is to first-order, and its coverage, which is $\alpha + O(n^{-1/2})$.

3.2.2 Basic bootstrap method

The basic bootstrap method involves approximating the confidence limits by the quantiles of the empirical distribution formed by the R bootstrap replicates. Let H^* represent a random variable drawn from the empirical distribution function $\widehat{F}$ generated from the bootstrap samples. The random variable

H^* is the bootstrap equivalent of H, and t^* corresponds to the bootstrap estimate of h. Then, we aim to approximate the distribution of $H - \eta$ by the distribution of $H^* - h$. The quantiles $c_{\alpha/2}$ and $c_{(1-\alpha/2)}$ of $H - \eta$ are estimated by the corresponding quantiles of $H^* - h$. Let h^* represent a realisation of H^*. The m^{th} quantile is estimated by the $(R+1)m^{\text{th}}$ ordered value of $h^* - h$: $h^*_{((R+1)m)} - h$. Choose R such that $(R+1)(\alpha/2)$ is an integer. So, the $\alpha/2$ quantile of $H - \eta$ will be the $(R+1)(\alpha/2)^{\text{th}}$ ordered value of $h^* - h$, that is, $h^*_{((R+1)\alpha/2)} - h$. Similarly, the $(1-\alpha/2)$ quantile of $H - \eta$ corresponds to $h^*_{((R+1)(1-\alpha/2))} - h$.

The lower and upper limits of the estimated $(1-\alpha)$ equi-tailed confidence interval for the parameter, η, are

$$\widehat{\eta}_{\alpha/2} = 2h - h^*_{((R+1)(1-\alpha/2))}, \quad \widehat{\eta}_{(1-\alpha/2)} = 2h - h^*_{((R+1)\alpha/2)},$$

where $h = \widehat{\psi}$ or $\widehat{p}$, is the maximum likelihood estimate of ψ or p, respectively, produced from the original data set. Accordingly h^* represents a realization of the bootstrap estimates of ψ or p. So, for example, the approximate confidence interval for ψ is obtained from the quantiles of $h^* = \left(\widehat{\psi}^*_{(1)}, \widehat{\psi}^*_{(2)}, \ldots, \widehat{\psi}^*_{(R)}\right)$. In the same way, the quantiles may be obtained from the corresponding distribution of ordered bootstrap estimates for the probability of detection p.

3.2.3 Studentised bootstrap method

The studentised bootstrap method is also known as the 'bootstrap-t' method and was first proposed by Efron (1982). The confidence limits are obtained by replacing the N(0, 1) approximation by the studentised statistic $Z = (H - \eta)/V^{1/2}$. As the distribution of Z is unknown, it is approximated by its bootstrap counterpart Z^*.

Each bootstrap replicate will produce $h^* = \{\widehat{\psi}^*, \widehat{p}^*\}$ and an associated asymptotic variance estimate $v^* = \{\widehat{\text{var}}(\widehat{\psi}^*), \widehat{\text{var}}(\widehat{p}^*)\}$, resulting in the bootstrap pivot $z^* = (h^* - h)/v^{*1/2}$. The R bootstrap estimates of z^* are ordered and the $(R+1)m^{\text{th}}$ ordered value estimates the m^{th} quantile of Z. Then the studentised confidence limits are

$$\widehat{\eta}_{\alpha/2} = h - v^{1/2} z^*_{((R+1)(1-\alpha/2))}, \quad \widehat{\eta}_{(1-\alpha/2)} = h - v^{1/2} z^*_{((R+1)\alpha/2)}.$$

The accuracy of the estimated confidence limits relies on the number of boot-

straps, R, but also on the extent of agreement between the distributions of $H^* - h$ and $H - \eta$. Studentising improves the accuracy of the confidence limits by improving the level of agreement between the distributions of $H^* - h$ and $H - \eta$. It achieves stronger agreement by taking a *pivot*, since complete agreement occurs if the distribution of $H - \eta$ does not depend on any unknowns. It does this by eliminating the unknown standard deviation and replacing it with the sample estimate. In this way, it mimics the Student-t statistic for making inferences about a normal mean.

The studentised bootstrap method for calculating confidence intervals is inherently more accurate than the basic bootstrap and percentile method. The first offers second-order accuracy whose coverage is $\alpha + O(n^{-1})$ whereas the last two are both restricted to first-order accuracy, $\alpha + O(n^{-1/2})$. However, the basic and percentile limits are still better than the normal approximation limits (Davison and Hinkley, 1997, pp. 212-4).

3.2.4 Percentile method

The studentised bootstrap method is an example of how the transformation of a statistic, in this case Student's-t transform, improves the accuracy of confidence limits. Another bootstrap method for approximating confidence limits based on a transformation is the percentile method. Unlike the studentised method, the percentile method implicitly assumes the existence of a good transformation, but does not require that a good transformation be found (Davison and Hinkley, 1997, Section 5.3, p. 202). It is based on the premise that the resulting distribution for some unknown transformation, $g(H)$, will be symmetric. This premise of symmetry is used to derive confidence limits independent of the transformation $g(H)$. The confidence interval of Equation (3.1) is rewritten with $c_{\alpha/2} = -c_{(1-\alpha/2)}$. Then by applying the basic bootstrap method (Section 3.2.2) and after back-transformation the percentile confidence limits on the original scale are

$$\widehat{\eta}_{\alpha/2} = h^*_{((R+1)\alpha/2)}, \quad \widehat{\eta}_{(1-\alpha/2)} = h^*_{((R+1)(1-\alpha/2))},$$

where these new limits are independent of the transformation $g(H)$. The h^* are the bootstrap estimates of occupancy and detectability, ψ and p respectively.

3.3 Simulation and Case study

3.3.1 Simulation study

We used a simulation study to assess the statistical properties of the interval estimators under a range of sample scenarios. To simultaneously investigate the effects of the number of sites sampled (or sample size), N, and number of survey occasions (T visitations to a site), we considered three sample sizes ($N = 10, 30, 50$) across two levels of survey occasions, ($T = 5, 10$). The six combinations allow for a comparison with the results from the case study we present, as well as with results presented in MacKenzie et al. (2002).

Simulations for the first study were performed as follows. A site occupation matrix was constructed, row-wise, by comparing a uniform random variate for each simulated site with the set probability of occupation. Hence the site occupation matrix comprised rows (i.e. sites) of all 0s or all 1s. A detection matrix of the same dimension was constructed cell-wise by comparing $N \times T$ uniform random variates with the set detection probability. The detection matrix also comprised all 0s or all 1s. The simulated detection (history) matrix was then constructed as the elementwise product of the site occupation and the detection matrices.

We resampled 100 or 250 bootstrap replicates (R), respectively, for each sample size (N) by survey occasion (T) combination to assess the effect of bootstrap replication number on the accuracy of the interval estimator. We computed coverage probabilities and average confidence interval widths from each scenario. The expression of variance for a binomial proportion was used to estimate the standard error of coverage. Furthermore, the simulation study covered a range of combinations for occupancy, $\psi = \{0.4, 0.5 \ldots, 0.8\}$, and detectability, $p = \{0.4, 0.5 \ldots, 0.8\}$, resulting in $3 \times 2 \times 2 \times 5 \times 5 = 300$ scenarios for which $s = 1000$ simulations each were generated.

3.3.2 Case study

For a second study, we used data from a study of occupancy and detectability for the endangered growling grass frog (GGF) (*Litoria raniformis*) in southern Victoria, Australia (Heard et al., 2006). The parameters estimated from these data were used to guide the design of the simulation experiment. Details of the sampling are given in Heard et al. (2006). Briefly, the study area encompassed

the catchments of the Merri Yuroke and Darebin Creeks on the northern outskirts of the Melbourne metropolitan area. Spotlight surveys were conducted and the detection histories were recorded for adult frogs and tadpoles detected from the 2002–2003 season (Table 2.2, Section 2.7.2 in Chapter 2).

Alternatively, a logistic model may be fit, such as that presented in MacKenzie et al. (2002). We explore this option in Chapter 4.

In total, our case study comprised $N = 27$ measurement sites, $T = 4$ survey occasions, and $R = 199$ bootstrap replicates drawn from the observed GGF history matrix. We applied the four interval estimators to the bootstraps of the detection histories of the GGF to obtain confidence limits and corresponding interval widths.

During the analysis, we noted that the sample variance of the data (3.35) was high compared to the theoretical Binomial variance (0.387; the dispersion ratio was 8.65). This is strong evidence of overdispersion, which suggests that the data are clustered.

In order to generate interval coverages that would be comparable to those computed from the GGF data requires reproducing the overdispersion inherent in the GGF detection matrix. Thus we conducted a second simulation study assuming overdispersion. We incorporated a degree of overdispersion similar to the GGF data by randomly selecting the same number of sites (rows), with replacement, from the GGF matrix. We obtained 2000 simulated detection matrices from which we then obtained bootstrap interval estimates. For each of the 2000 simulated matrices we generated 250 bootstrap replicates, subsequently computing coverage rates and average interval widths. We chose 2000 simulations to deal with the large (up to 40%) failure rate for maximisation of the likelihood. The conditions under which the MLE equations will not always be valid were studied and addressed in Chapter 2 (see the edge solutions given in Section 2.3.2), as well as in (MacKenzie et al., 2002; Wintle et al., 2004; Guillera-Arroita et al., 2010). In Chapter 4 we address this using a partial likelihood approach, for models that also include covariates.

The difference between the first simulation study and the GGF simulation study lies in the method used for generating the simulated detection matrices, from which we subsequently took the bootstrap samples. The latter study, which allows for overdispersion, randomly selected sites (rows) with replacement directly from the GGF detection matrix, followed by bootstrapping.

In evaluating the coverage provided by an interval estimation method, we adopt the approach of Newcombe (1998), who suggested that if $1-\alpha$ is construed as an average coverage rate, the mean CP (coverage probability) should ideally be a little over $1-\alpha$ and the minimum CP a little under it. This provides a sensible aim when constructing non-conservative CIs; it recognises that if the nominal level cannot be achieved exactly, an almost universal choice would be to be above it rather than to be a similar amount below it (Newcombe, 1998; Hepworth, 2004).

To assist in interpreting the plots of the coverages, we added a dashed line to each plot representing the lowest possible value of the estimate that would still contain 0.95 within its 95% confidence interval with 1000 samples. We ignore the simulation failures in this calculation hence this is an approximation. We assume the number of simulations to be $s = 1000$. We solve the following equation, for λ, to obtain the approximate upper limit for the CI of the nominal $(= 0.95)$ coverage level:

$$\begin{aligned} 1-\alpha &= \lambda + c_{\alpha/2}\sqrt{\frac{\lambda(1-\lambda)}{s}} \\ \text{i.e. } 0.95 &= \lambda + 1.96\sqrt{\frac{\lambda(1-\lambda)}{1000}} \\ \text{therefore } \lambda &= 0.934 \end{aligned}$$

All estimates of ψ, or p, that are above the dashed line will, according to this construction, include 0.95 in their 95% confidence interval. We have ignored the possibility that the estimate is too high as we are less concerned with this condition.

3.4 Results

3.4.1 First study

Based on the values that we chose to compare, we found that the number of sites sampled, N, had the largest impact on coverage for each of the interval estimators (Figures 3.1a and 3.1b)[1]. Note that the full suite of results included interval coverage rates and widths for each parameter conditional on specific values of the other. Here we report only the marginal results. We examined

[1]Labelled Figure 1. and 2., respectively, in Karavarsamis et al. (2013).

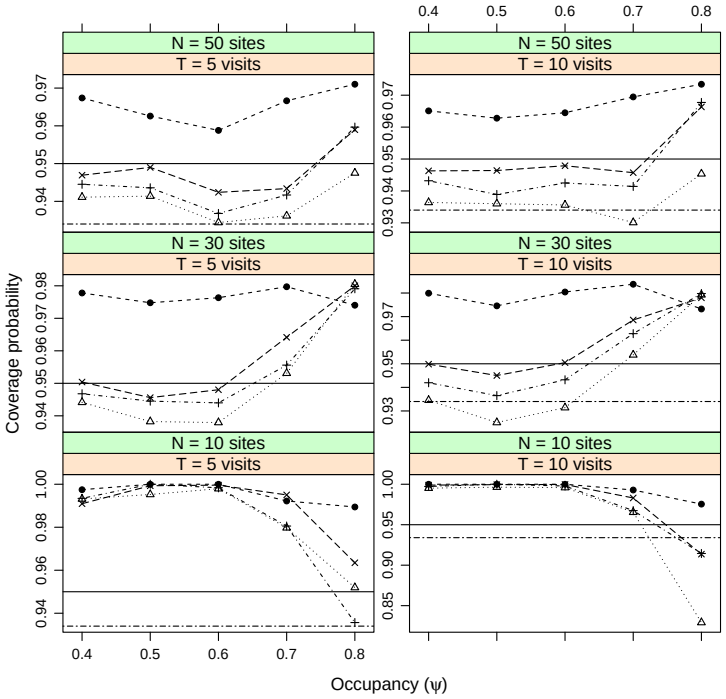
some conditional results (not shown here) and found no evidence that averaging would provide a misleading summary.

Figure 3.1a shows coverage probabilities for occupancy ψ produced from the first simulation study, assuming there is no overdispersion, were averaged over p, for R ($= 100$ or 250) replicates and s ($\leqslant 1000$) simulations. Similarly, Figure 3.1b shows coverages for detectability p averaged over ψ. The panels display combinations of sample size N by the number of survey occasions T. The solid horizontal line indicates the 0.95 nominal confidence level. The dashed line denotes the line ($= 0.934$) below which any point would not include 0.95 in its normal approximation based confidence interval. The y-axis scale for the plots differs between rows.

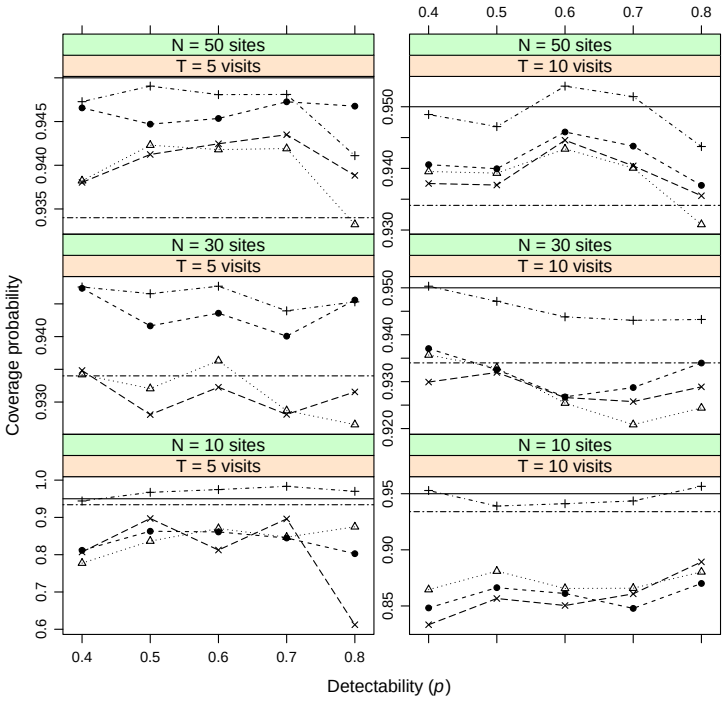
Table 3.1 shows coverage probabilities and average interval widths of confidence interval for occupancy ψ and detectability p, at the nominal 0.95 level of confidence. Interval widths, which lie between 0 and 1, and corresponding coverages are obtained from the first simulation study (Section 3.3.1), assuming there is no overdispersion in the data. Widths and coverages are averaged over all combinations of ψ and p, for R ($= 100$ or 250) replicates and s ($\leqslant 1000$ per scenario) simulations.

An increase in N can result in an increase in observed detections, thus improving convergence of the optimization of the likelihood function and therefore also the number of successful simulations. In general, our results were unreliable when $N = 10$ as it caused many simulations (20%) to fail. Given this, we focus on those results produced by larger values of N.

When estimating occupancy, the studentised method produced the most consistent (least variable) estimated coverage probabilities of the estimators (Figure 3.1a and Table 3.1). It had the fewest fluctuations, coverages were below nominal 0.95 as well as above $\lambda(= 0.934)$, and the fewest outliers. The remaining three estimators produced coverages that at times were well below the nominal coverage; for example the basic method produced a coverage of 0.55 when $\psi = 0.8, p = 0.8, N = 10$, $T = 10$ and $R = 250$ (not shown here). Overall, the basic method for approximating the confidence limits for occupancy was the least reliable, with its coverages generally being further below nominal than the other methods (Figure 3.1a). The percentile, basic and normal approximation results were parallel with each other across most of the parameter space within the scope of the study (Figure 3.1a). Overall, the percentile method outperformed the normal approximation when estimat-



(a) Coverage probabilities of occupancy ψ.



(b) Coverage probabilities for detectability p.

Figure 3.1: Coverage probabilities for occupancy ψ and detectability p. Solid horizontal line indicates the 0.95 nominal confidence level and dashed line denotes $\lambda(= 0.934)$ the approximate upper limit for the CI below which any point would not include 0.95 in its normal approximation based confidence interval. Legend: • studentised bootstrap; + normal approximation; △ basic bootstrap; and × percentile.

ing occupancy (Figure 3.1a). As the probability of presence increased, the coverage improved likewise, for most combinations of N and T.

Interval Estimator	**Occupancy** ψ		**Detectability** p	
	coverage	interval width	coverage	interval width
Basic				
mean	0.9521	0.4124	0.9084	0.1949
sd	0.0406	0.1513	0.0487	0.0695
Studentised				
mean	0.9795	0.4875	0.9097	0.1939
sd	0.0147	0.2214	0.0551	0.0685
Percentile				
mean	0.9639	0.4124	0.9004	0.1949
sd	0.0254	0.1513	0.0739	0.0695
Normal				
mean	0.9593	0.4097	0.9503	0.2106
sd	0.0273	0.1559	0.0158	0.0933

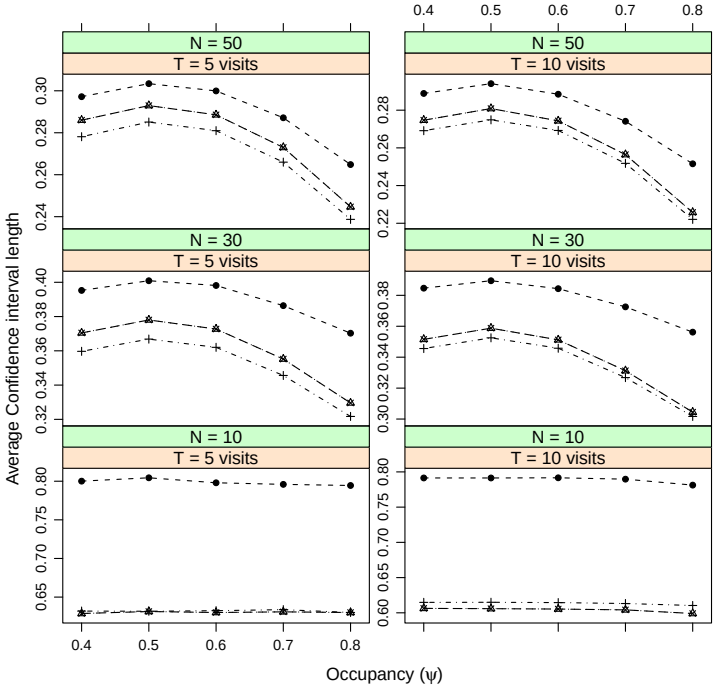
Table 3.1: Coverage probabilities and average widths of confidence interval for occupancy ψ and detectability p, at the nominal 0.95 level of confidence.

Figure 3.2a shows average confidence interval widths (which lie between 0 and 1) of occupancy ψ produced from the first simulation study, assuming there is no overdispersion, averaged over p for R ($= 100$ or 250) replicates and $s(\leqslant 1000)$ simulations. The panels display combinations of sample size N by the number of survey occasions T. To make the differences visible between estimators for each combination of N and T, the scale on the y-axis is unique to each plot. Similarly, Figure 3.2a shows average confidence interval widths of detectability p averaged over ψ.

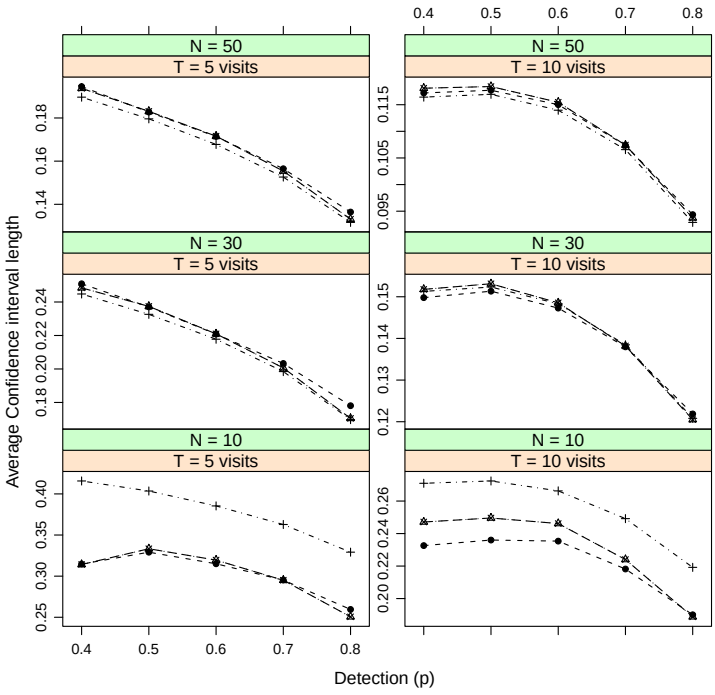
In general, average confidence interval widths were narrower for detectability (p) compared with occupancy (ψ) (Figures 3.2a and 3.2b, and Table 3.1)[2].

The coverages for detectability were below nominal (Table 3.1). Coverages of the normal approximation intervals for detectability were closest to nominal, and these intervals were also generally wider than those for the other methods.

[2]Labelled Figures 3 and 4, and Table 2, respectively, in Karavarsamis et al. (2013)



(a) Average interval widths for occupancy ψ.



(b) Average interval widths for detectability p.

Figure 3.2: Average confidence interval widths for occupancy and detectability. Legend: • studentised bootstrap; + normal approximation; △ basic bootstrap; and × percentile.

3.4.2 Second study

Of the 2000 simulated GGF (Growling grass frog) detection matrices, 1155 returned MLEs for all 250 bootstrap-created detection matrices. Our goal was to reproduce the overdispersion inherent in the GGF matrix. The simulated matrices were generated with an average overdispersion of 8.16 compared to 8.45 present in the GGF detection matrix. Dispersion factors amongst simulated matrices ranged from 5.062 to 10.95. The GGF dispersion factor was adequately preserved during the bootstrapping. Averaged over simulations, these ranged from 4.93 to 10.85 with a mean of 8.04.

Table 3.2 shows the Growling grass frog occupancy ψ and detectability p confidence limits, and widths, at the nominal 0.95 level of confidence. Interval limits and corresponding widths, which lie between 0 and 1, were generated from $R = 199$ bootstrap replicates drawn from the (27×4) GGF detection matrix. Coverages are obtained from the second simulation study assuming overdispersion: 250 bootstrap replicates drawn from the GGF detection matrix for each of the 1155 out of 2000 successful simulations.

Interval Estimator	**Occupancy** $\widehat{\psi} = 0.56$				**Detectability** $\widehat{p} = 0.78$			
	lower limit	upper limit	interval width	cp	lower limit	upper limit	interval width	cp
Basic	0.371	0.777	0.406	0.943	0.651	0.932	0.281	0.978
Studentised	0.347	0.787	0.441	0.989	0.577	0.904	0.327	0.977
Percentile	0.337	0.742	0.406	0.964	0.631	0.912	0.281	0.967
Normal	0.368	0.744	0.376	0.946	0.675	0.888	0.212	0.886

Table 3.2: Growling grass frog (GGF) occupancy ψ and detectability p confidence limits, widths and coverages.

We applied the four interval estimators to the probability that the growling grass frog was present (ψ), and to the probability of detecting a frog (p). Consistent with findings from the first simulation study (Figure 3.2a), the GGF occupancy and detectability interval widths are equal for the basic bootstrap and percentile methods (Table 3.2[3]). From the GGF study, the studentised estimator resulted in a marginally wider interval for occupancy than did the basic or percentile methods, and a substantially wider interval for detectabil-

[3]Labelled Table 3 in Karavarsamis et al. (2013)

ity overall. However, corresponding normal coverages, particularly for p, were below nominal based on the GGF simulation study, suggesting the normal approximation-based interval estimates were too narrow (Table 3.2). In particular, Table 3.2 shows that the normal approximation CIs did not work for detectability, the coverage probability is below 0.9. This may be due to overdispersion that is not accounted for in the model. In Chapter 2 we showed that the asymptotic variance estimator underestimates the actual (or exact) variance, for occupancy.

Overall, the estimated interval widths for the probability p of detecting frogs at a site were considerably narrower than those corresponding to the probability ψ of frogs occupying a site, as expected (Table 3.2). This implies that we know with more certainty the chance of detecting frogs at a site, than the chance with which the GGF actually occupies a site. To verify this conclusion, associated coverages obtained from the simulation study should be considered. The studentised estimator produced above nominal (0.95) and consistent coverages over the parameter space [0,1] for ψ. It achieved these results for simulations conducted in both the presence, and absence, of overdispersion (Table 3.2 and Figure 3.1a, respectively). When overdispersion is not assumed and when $\psi = 0.56$, $N = 30$ and $T = 5$, the percentile method produces coverages closest to nominal (Figure 3.1a). However, over the entire range of possible values of occupancy (0 to 1), the studentised estimator does not fall below nominal coverage. It provides more consistent, albeit sometimes more conservative, coverages.

3.5 Conclusions

Overall, we found that the studentised estimator outperformed the others when estimating occupancy, in that it gave coverages which were stable and remained above nominal (Figure 3.1a), but the same did not hold when estimating detectability. In addition, the performance of the estimator depends on whether overdispersion is present in the detection matrix. We showed in Chapter 2 that the asymptotic variance for occupancy underestimates the actual (or exact) variance (Section 2.5.4). Hence normal intervals are probably too narrow, and the studentised are closer to the actual widths i.e. closer to widths that they should be. In addition, studentised intervals are the most consistent overall (closest, or above, nominal level and above λ), albeit wider and possibly conservative on occasion.

For detectability p, and for data that are not overdispersed, we found that on the whole the normal approximation provided coverages which approached the nominal level from below, as either N or T increased. The discrepancy in the performance between estimators depends on how closely the empirical distribution agrees with the probability distribution (Davison and Hinkley, 1997, p. 29). Agreement between the two distributions ensures that the coverage of the interval estimator converges faster to $1 - \alpha$.

For some estimators, the distribution of bootstrapped estimates directly contributes to the limits via the $\alpha/2$ and $1 - \alpha/2$ percentiles of z^* or h^*. This applies to the studentised, the basic and the percentile methods. Thus when parameter estimation is problematic (for example, when estimating some values of p close to the boundaries i.e. 0 and 1), the corresponding empirical distribution generated from the bootstrapped estimates does not closely agree with the population distribution, leading to poor coverages for these three estimators.

In contrast, the critical value associated with the normal approximation is not recovered from the distribution of the estimates (for example from the distribution of h^*), but is obtained from the inverse cumulative distribution function for the (standard) normal distribution. However, the normal approximation outperforms all others when estimating p. We examined the distribution of the bootstrap estimates of p from the GGF data. The assumption of normality of the bootstrap distribution appeared to be reasonable in this case. This would imply that, on average, we would expect the normal CI to do very well, certainly better than the basic and percentile methods.

The poor estimation of p from our model may be attributed to the assumption of constant detectability across sites and survey occasions. The assumption of constant detectability is unlikely to be true for many species, so this would pose a problem for a great host of ecological studies. A model that allows for heterogeneous detections between survey occasions, and even amongst sites, may lend itself to better estimates for p. A number of studies have developed models, such as a logistic regression (MacKenzie et al., 2002) or finite mixture models (Royle and Nichols, 2003), which allow for heterogeneity in p. In Chapter 4, we propose models for these scenarios using a partial likelihood approach, and introduce a model based on a conditional likelihood function.

For the overdispersed data resulting from resampling the GGF detection history matrix, the studentised and percentile estimators work best for ψ and p alike. The reasons are similar to those which apply to the large simulation study. Specifically, pivotal quantities are less influenced by overdispersion because they are scale-free. Thus for the overdispersed data in particular, the empirical distributions derived from pivotal quantities will more closely agree with the true probability distribution compared to, say, the normal approximation which assumes an underlying normal distribution. Consequently the basic method produced inconsistent coverages which were below nominal for ψ and above for p. This is possibly because it relies solely on the distribution of the bootstrap estimates and it does not include the variation from within the detection matrix in its estimation process. Referring to our approach for evaluating the interval estimators described earlier (Section 3.3.2), the studentised estimator outperformed the percentile when the data were overdispersed because it includes the variability of the detection matrix in the pivots.

We expected that an increase of information would produce coverages closer to the nominal level and narrower interval widths. Coverages for ψ approached the nominal level and average interval widths narrowed as we accrued information about occupancy, as the probability of presence approached one. In contrast, we found for p that as the number of detections increased, i.e., as the probability of detection improved, the average interval width narrowed (Figures 3.1b and 3.2b).

We observed these patterns when exploring interval widths and associated coverages for the simulations not assuming overdispersion. This is not unlike the case study, where we found that interval widths produced for detectability, irrespective of interval estimator, were considerably narrower than for occupancy. We also found via simulations without overdispersion that the narrower

interval widths corresponding to detectability, in contrast to occupancy, are misleading in that they gave coverages for detectability which are substantially below the nominal level, even below 0.9 in some instances. Contrary to what we expected, the decline in coverage as detectability improved could be indicating a difficulty in accurately estimating detectability. When the data were overdispersed, the coverage for p was above nominal for all except the normal method. It should be noted though that the estimate for p was high for the GGF data ($\widehat{p} = 0.78$), whereas coverages for the non-overdispersed study were averaged over values of p ranging from 0.4 to 0.8.

When N is adequate ($\geqslant 10$), we recommend using the studentised estimator for ψ as it provides more consistent coverage than the other estimators, generally generating coverages at, or above, nominal, and that will be above $\lambda(= 0.934)$ i.e. the approximate upper limit that will include 0.95 in their 95% CI. It provided intervals which appropriately reflect the small sample size for N. We have demonstrated that the basic method is not appropriate for this likelihood. The normal approximation gives a better result for occupancy compared to detectability. Importantly, the normal approximation appears to perform best when estimating p and second worst, after the basic, when estimating ψ for the non-overdispersed study. The studentised method is reliable, in that it produced the most consistent coverages along the range of values for ψ and p, for a moderate sample size ($N \geqslant 25$) even when there are few survey occasions ($T = 4, 5$ or 10), but the same does not apply for small sample sizes ($N \leqslant 10$).

The studentised pivot incorporates the spread of the empirical distribution by using an estimate of the variance in constructing confidence limits, a difference from its percentile counterpart. As a result the intervals from the studentised estimator were wider in this study. In Chapter 2 we found that the asymptotic variance underestimates the actual variance. Also that summaries in this chapter are averaged of the entire range for ψ and p which may have affected results, as seen in Chapter 2, that plausibility for estimation and accuracy of estimation are affected by the level for ψ and p (for example see Section 2.3.3). For example, the expectation for occupancy is not always monotone increasing (Section 2.5.3). Plus, the boundary problems when directly maximising the full likelihood prevents convergence. This is investigated and evaluated in detail in Chapter 4. This, coupled with the fact that its coverages are consistently closest to nominal levels, suggests that the studentised estimator is correctly adjusting for the effects of bias and nonconstant variance inherent in the estimators of ψ and p, and asymmetry in the underlying distributions. Such an

observation is consistent with the studentised method providing second-order accuracy, unlike the percentile, basic and normal approximation, which provide first-order accuracy (Davison and Hinkley, 1997, pp. 212-4). Consequently, coverages from the studentised method did not fall below the nominal level, although they did for the other three estimators in the simulation study.

In the simulation study, the percentile method closely followed the studentised estimator with respect to coverages approaching nominal levels. Of the estimators we explored, the studentised and percentile estimators outperformed the others. The percentile method coverages were not as conservative as those resulting from the studentised method, but did fall below nominal, and its intervals were generally narrower. In other instances, when N and s were small (10 and $\ll$ 1000 respectively), interval estimates could not be produced for the studentised and normal approximation. We could not elicit the asymptotic variances from the hessian matrix since it was not invertible in these instances. This occurred for scenarios of the simulation study and for the GGF case studies. However, the percentile method always produced interval estimates.

From our simulation studies, we found that the magnitude of N has a different effect on occupancy than it has on detectability. An increase in N increases the amount of information provided to the likelihood. A larger N is needed for an improvement in accuracy when estimating detectability p than is required for estimating ψ. This is possibly because for a fixed number of individuals at a site, or a fixed density of presence of a species on a site, we detect a fraction of these. Consequently, more information is known for the patterns of presence than detection for a fixed size of N.

Our simulation results do not align with the asymptotic variance estimators as provided by Guillera-Arroita et al. (2010), who showed that, asymptotically, N should have the same effect upon p as it does on ψ. However, it is possible that the sample sizes that we considered operationally realistic are too small to reflect the asymptotic results.

Our results were unreliable for a small sample size i.e., when we sampled $N = 10$. For a large number of the bootstrap replicates the likelihood did not converge so there were not enough simulated confidence intervals from which to accurately estimate coverage; the number of successful simulations was small. In summary, s affects the accuracy of the interval estimator, or coverage, and N affects the accuracy of the occupancy (or detectability) estimator, as reflected

by the width of the confidence interval. Our results for small N are consistent with findings from other studies, in which the estimator properties depart from those predicted by large-sample approximations (Wintle et al., 2004; Guillera-Arroita et al., 2010).

We also found problems with convergence for values of ψ and p close to their boundaries (results not shown here). It is known that for some data sets, maximisation of this likelihood function may result in estimates of occupancy that are greater than 1 (MacKenzie et al., 2006; Guillera-Arroita et al., 2010, p. 95). The likelihood does not implicitly constrain the estimates of the probabilities to fall between 0 and 1 (MacKenzie et al., 2006, p. 95). Consequently, our results include only those estimates for ψ and p which happen to lie within the boundaries. Issues of nonconvergence need to be overcome for more definitive comparisons between these interval estimators. As mentioned previously, a transformation onto an alternative scale not bounded between 0 and 1 may help overcome this issue (see Section 3.2.1). Alternatively, models could be explored which are appropriate for presence-absence data (Wintle et al., 2004; Holzmann et al., 2006; Dorazio, 2007). Another option includes making adjustments directly to the likelihood function we have used here. We explore some models in the next chapter and extend existing models using partial likelihoods, incorporating covariates.

Convergence rates may also have been compromised by the choice of the maximisation process. Here, we used a Newton-Raphson type algorithm for maximising the likelihood, which may account for some problems with convergence (Nelder and Mead, 1965). This algorithm works best when the gradient of the profile likelihood is sufficiently steep. The likelihood function we used could be quite flat for some parameter combinations. It might be useful to explore alternative maximisation algorithms for this likelihood function, although that is beyond the scope of this study. For example, the L-BFGS-B maximisation algorithm of Byrd et al. (1995), a modified BFGS quasi-Newton method, has been used by Haas and Stokes (1998) and Hwang and Shen (2010) to overcome problems with convergence.

Chapter 4

A partial likelihood approach for modelling occupancy with covariates

4.1 Overview

In this chapter we focus on using partial likelihoods for occupancy and detectability to analyse complex data structures using standard modelling techniques.

In Chapters 2 and 3 we identified and explored in detail some problems concerning the full likelihood approach to the occupancy model. These are known challenges and have been explored by, for example, Wintle et al. (2004); Guillera-Arroita et al. (2010); Welsh et al. (2013); Karavarsamis et al. (2013). In particular, Welsh et al. (2013) and in Chapter 3 of this thesis (the work published in Karavarsamis et al. (2013)) illustrate that there are many problems associated with using methods based on the full likelihood to fit occupancy models. For example, there are issues of identifiability when the estimates exceed 1 or are near the boundary of the parameter space, as were addressed in Chapters 2 and 3. And, these issues persist if we wish to include covariates (Welsh et al., 2013). We show that by using partial likelihood and basing inference for the detection probabilities on the redetections we may overcome many of these problems. The resulting models can include both parametric and nonparametric covariates using standard GLM methods.

Partial likelihood techniques are often used to simplify complex likelihoods.

Here we exploit this approach to partition the likelihood into two partial likelihoods. We then use the first to estimate what are essentially nuisance parameters (detectability) in the second and finally estimate the quantities of interest, which here are the occupancy probabilities. We show how sophisticated models can be more easily applied using a partial likelihood approach. Whilst with the basic occupancy model it is not computationally simple to estimate occupancy ψ and detection p when covariates and complex models are involved we show that this can be easily done using a partial likelihood approach.

By using a partial likelihood approach we are able to directly and fully explore, and model, characteristics of the data. For example, we can make use of GLMs and GAMs to fit parametric and nonparametric functional relations. The use of standard GLM methods allows the use of iteratively reweighted least squares, rather than nonlinear optimisation, which yields computationally efficient stable estimates.

A particular advantage is that we can first focus on the modelling of the detection probabilities then, once this is complete, move to modelling the occupancy. As we shall see in the two-stage procedure that we propose, estimation of the detection probabilities is through a standard GLM and hence the full range of GLM modelling methods are available. This is discussed in Section 4.4.2.

We begin the chapter with a review of existing methods and provide a motivation for the methods we go on to develop (Section 4.2).

In Sections 4.3, 4.4, 4.5, and 4.6, we introduce some new theory. We evaluate this with a range of simulation studies and with several real applications. We apply our methods to the motivating frog data and two sets of fish data, using standard software.

In Section 4.3 we first outline the theory of partial likelihood for the homogeneous case that assumes constant ψ and p among sites and for all survey occasions. We compare the homogeneous partial likelihood approach to direct maximisation of the full likelihood (as shown in Chapter 3) via simulations and real data. We first compare these to the MLE equations for the full likelihood derived in Chapter 2. We use simulations to evaluate the efficiency for these approaches.

In Section 4.4 the homogeneous partial likelihood function is extended to allow for site inhomogeneity. The site inhomogeneous likelihood function permits

site occupancy and detectability to vary among sites, but still remain constant within sites i.e. over survey occasions. We consider three approaches to estimate occupancy for the site inhomogenous likelihood. The first is a direct maximisation of the partial likelihood, the second is via an iterative approach that involves an offset (in the likelihood), and the third, is through a ratio estimator for ψ. We compare these approaches with a number of simulation studies and applications to real data. Again we compare our approach to existing approaches.

Then, in Section 4.5 we develop a partial likelihood for time dependent covariates for estimating detection probabilities. Here detection is modelled as a function of covariates that vary with sites and time occasions. Occupancy is supposed constant within sites but may vary among sites.

For all estimators for occupancy and detectability that we propose, we also derive approximate standard errors. We exploit asymptotic theory for these derivations, and examine their consistency and efficiency using simulations and real data. We also assess their performance against some typical variance approximations that involve bootstrapping or are based on Bayesian methods, discussed in Section 4.2.

For the homogeneous case (Section 4.3), we show that our variance estimator for occupancy perform well compared to the asymptotic estimator for the full likelihood. In Chapter 2 we showed that the asymptotic standard error for ψ from the basic occupancy model is unreliable and seriously underestimates the actual variance. Comparisons in Section 2.5.4 of exact and asymptotic variance, and MSE, are shown in Figures 2.9 and 2.10, for example. And, in Chapter 3 we compared the asymptotic estimator to bootstrap based methods (this work is published in Karavarsamis et al. (2013)).

For the site inhomogeneous case (Section 4.4), we derive approximate standard errors for use with GLMs and adapt these for use with GAMs. We show these approximations to be reliable and to perform at least as well, if not better under certain conditions, than those for the full likelihood.

The advantage of the estimator we propose in this chapter, is that it relies on the data i.e. on the redetections that are recorded, whereas the score equation of the full likelihood requires an approximation for $\widehat{\theta}$, that is typically a numerical approximation. For example, the `Solver` function in Excel or `uniroot` in R may be used (MacKenzie and Hines, 2009; Fiske and Chandler, 2011, etc.).

Recall from Chapter 2 that $\widehat{p}/\widehat{\theta} = y/\{(S - f_0)\tau\}$[1], where $\widehat{\theta} = 1 - (1 - \widehat{p})^{\tau}$) where we used both computer functions. In addition, our methods will work easily and readily in R, thus on many computer platforms, whereas the standalone software `PRESENCE` by (MacKenzie and Hines, 2009) is not so easily run on Apple Computers.

[Note: In this chapter we adopt alternative notation to easily express equations for complex covariate structures. We believe this change will make the notation concise and neat. It will be made clear throughout the chapter when the notation is altered.]

4.2 Review of methods and motivation

Modern statisticians have become accustomed to using methods such as Generalized Additive Models (GAMs; Hastie and Tibshirani (1990); Wood (2006)) as a matter of common practice. These are readily implemented in the generalized linear model (GLM) framework. However, their extension to settings such as occupancy models is difficult and can require extensive computer programming.

The basic occupancy model (or full likelihood) simultaneously estimates occupancy and detectability, which is considered a nuisance parameter (e.g. Hall, 2000; MacKenzie et al., 2002; Tyre et al., 2003). A standard method of handling nuisance parameters is the profile likelihood but as with the full likelihood this would require the writing of specialist programs that do not take advantage of available methods for generalized linear models. In particular, fitting GAMs to both detection and occupancy probabilities would become a fearsome computational task. We note that in occupancy models there are repeated observations at each site so that there is more information on the detection probabilities than on the occupation probabilities. Thus, it may be advantageous to gain computational efficiency at the cost of some small loss of efficiency in estimating the detection probabilities.

In GAMs there may be a large number of parameters to estimate given there are two functions that must be simultaneously estimated. Partial likelihoods are particularly appealing because they result in a simple binomial model that can be fitted with the common logistic function, the `logit` link which is commonplace in all modelling software packages.

[1] $S \equiv N$, $\tau \equiv T$ and $f_0 = N - k$

Partial likelihood was initially developed in Cox (1975) and the theory is well developed (e.g. Wong, 1986; Gill, 1992). In order to use GLMs in both stages of estimation we adopt this approach to occupancy models by using the redetections after the first detection to estimate the parameters in the model for the detection probabilities. This is the first stage of the analysis and the second partial likelihood. These practices have been adopted for some time now, for example in Yip et al. (1996); Pledger et al. (2003). Stoklosa et al. (2011) showed they are useful in capture-recapture models with covariates, where they are readily implemented.

We may also conduct inference on the detection probabilities by conditioning on at least one detection at a site. We refer to this as the conditional likelihood and call the approach based on redetections the partial likelihood. The conditional likelihood can be implemented as vector generalized linear or additive models (VGLM VGAM) but these are not yet as well developed as GLM and GAM methods and we do not emphasise this approach here.

In some cases, we can compare our methods to corresponding models using the full likelihood via the `unmarked` R package, and also to the stand alone `PRESENCE` software for occupancy models (Fiske and Chandler, 2011; MacKenzie and Hines, 2009; R Development Core Team, 2009). Both these products may fit a range of similar occupancy models. In particular we use the `occu` procedure in `unmarked` which fits the 'single season occupancy model' by maximising the full likelihood and obtains standard errors from the observed information matrix, as described in MacKenzie et al. (2002, 2006); Royle and Dorazio (2008); Fiske and Chandler (2011).

4.3 Homogeneous case

4.3.1 Estimating detection and occupancy

We begin by considering the homogenous case as this allows us to examine the efficiency of the two-stage approach. Consider S sites labelled $s = 1, \ldots, S$ and τ occasions at each site where the presence of a species may be observed. Let ψ_s be the probability that site s is occupied and let p_s be the probability the species is observed on a given occasion given it is present. For the homogeneous case we consider that occupancy and detection are constant, $\psi_s \equiv \psi$ and $p_s \equiv p$.

Denote by y_s[2] the number of occasions upon which the species was detected at site s. Let $s = 1, \ldots, O$ denote the sites at which at least one detection occurred and let a_s be the first time upon which a detection occurred. Then $b_s = \tau - a_s$ is the number of occasions after the first detection and $b = \sum_{s=1}^{O} b_s$ is the total number of occasions upon which species could be redetected, over the O sites with any detections. Finally, let f_0 be the number of sites at which no detections occurred and so $S - f_0$ is the number of sites with any detections[3].

The full likelihood[4] may be re-written,

$$\begin{aligned} L(\psi, p) &\propto (1 - \psi + \psi(1-p)^{\tau})^{f_0} \prod_{s=1}^{O} \psi p^{y_s} (1-p)^{\tau - y_s} \\ &= (1 - \psi + \psi(1-p)^{\tau})^{f_0} \psi^{S-f_0} \prod_{s=1}^{O} (1-p)^{a_s - 1} p \prod_{s=1}^{O} p^{y_s - 1} (1-p)^{b_s - y_s + 1} \\ &= L_1(\psi, p) L_2(p). \end{aligned} \tag{4.1}$$

where

$$\begin{aligned} L_1(\psi, p) &= (1 - \psi + \psi(1-p)^{\tau})^{f_0} \psi^{S-f_0} \prod_{s=1}^{O} (1-p)^{a_s - 1} p \\ &= (1 - \psi\theta)^{f_0} \psi^{S-f_0} \prod_{s=1}^{O} (1-p)^{a_s - 1} p, \end{aligned} \tag{4.2}$$

and

$$L_2(p) = \prod_{s=1}^{O} p^{y_s - 1} (1-p)^{b_s - y_s + 1} = p^{y-O} (1-p)^{b-(y-O)}. \tag{4.3}$$

Here $y = \sum_{s}^{O} y_s$ is the number of total detections for the history matrix[5] and, as usual, $\theta = 1 - (1-p)^{\tau}$ is the probability of at least one detection. Occupancy, ψ, is only involved in the first component $L_1(\psi, p)$. The first component $L_1(\psi, p)$ also includes the probability of first detection. The second component $L_2(p)$ only involves the detection probability p.

[2] Previously defined as $x_{i\bullet}$, i.e. $x_{i\bullet} \equiv y_s$.

[3] Changes to the notation from earlier chapters: $N - k = f_0$ where $S \equiv N$ and $O \equiv k$.

[4] Referred to also as the basic occupancy-detectability (BOD) model. A full derivation was given in Section 2.2

[5] $y \equiv x$.

Thus, maximising $L_2(p)$ yields

$$\widehat{p} = \frac{y - O}{b}, \tag{4.4}$$

the ratio of the number of redetections over the number of remaining visits after the first detection. The estimator for p given here differs to that derived from the full likelihood (see Equation (2.10) of Section 2.3.1 in Chapter 2) since this version is estimated from the redetections, whereas the full likelihood also includes the first detection. The advantage of the current estimator, is that it relies on the data i.e. on the redetections that are recorded, whereas the score equation of the full likelihood requires an approximation for $\widehat{\theta}$, that is typically a numerical approximation.

To estimate ψ we maximise $L_1(\psi, \widehat{p})$ as a function of ψ. This yields

$$\widehat{\psi} = \frac{S - f_0}{S\widehat{\theta}} \tag{4.5}$$

which is equivalent to the score equation of the basic occupancy model ($\widehat{\psi} = k/N\widehat{\theta}$) given in Equation 2.10 of Section 2.3.1 in Chapter 2. Thus in the homogeneous case, the estimators have simple closed forms. By construction, $0 \leqslant \widehat{p} \leqslant 1$ and to have $\widehat{\psi} < 1$ we require the simple condition that $f_0 > S(1-\widehat{p})^{\tau}$. Note that $\widehat{\psi}$ given by (4.5) is a ratio estimator i.e. $\dfrac{(S - f_0)/S}{\widehat{\theta}}$ etc..

Furthermore, note that this estimator ignores any information on p from first detections. The advantage is that we have two simple closed form estimators.

4.3.2 Standard error for detection

The variance of $\widehat{p}$ is derived from the partial likelihood $L_2(p)$ in Equation (4.3). The distribution for the number of detections conditional on the first detection is $\left(y - O \,\middle|\, b\right) \overset{\mathrm{d}}{=} \text{Binomial}(b, p)$ so that $E\left(\widehat{p} \,\middle|\, b\right) = E\left\{(y - O)/b \,\middle|\, b\right\} = E\left(1/b \,\middle|\, b\right) E\left(r \,\middle|\, b\right) = p$ with variance $\text{Var}\left(\widehat{p} \,\middle|\, b\right) = p(1-p)/b$. The variance for detectability is given by the standard property $\text{Var}(\widehat{p}\,) = E\left\{\text{Var}(\widehat{p} \,\middle|\, b)\right\} + \text{Var}\left\{E\left(\widehat{p} \,\middle|\, b\right)\right\}$. Now, $\text{Var}\left\{E\left(\widehat{p} \,\middle|\, b\right)\right\} = \text{Var}(p) = 0$, and $E\left\{\text{Var}(\widehat{p} \,\middle|\, b)\right\} = E\left\{p(1-p)/b\right\} = p(1-p)\,E\left(1/b\right)$ which gives $\text{Var}\left(\widehat{p}\right) = p(1-p)\,E\left(1/b\right)$. Thus we estimate the variance of $\widehat{p}$ by $S_p^2 = \widehat{p}\left(1 - \widehat{p}\right)/b$.

We verify by simulations in Section 4.3.5 this expression for the variance of $\widehat{p}$ that we derived here.

4.3.3 Standard error for occupancy

Now, $f_0 \sim \text{Binomial}(S, 1-\psi\theta)$ [6] where $E(f_0) = S(1-\psi\theta)$, $\widetilde{\psi} = \widehat{\psi}(a) = (S-f_0)/(S\theta)$ so that if p and hence θ are known $E(\widetilde{\psi}) = \psi$ and $\text{Var}(\widetilde{\psi}) = \psi(1-\psi\theta)/(S\theta)$.

We use a Taylor expansion for $\widehat{\psi} = (S-f_0)/S\widehat{\theta} = \widetilde{\psi}\theta/\widehat{\theta}$. Firstly,

$$\begin{aligned}\widehat{\theta} &\approx \widehat{\theta}(p) + \widehat{\theta}'(p)(\widehat{p}-p)\\ &= \theta + \tau(1-p)^{\tau-1}(\widehat{p}-p)\end{aligned} \tag{4.6}$$

where $\widehat{\theta}(p) = 1-(1-p)^{\tau} = \theta$ and $\widehat{\theta}'(p) = \partial\widehat{\theta}/\partial p = \tau(1-p)^{\tau-1}$. Then,

$$\begin{aligned}\widehat{\psi} &\approx \widehat{\psi}(\theta) + \widehat{\psi}'(\theta)(\widehat{\theta}-\theta)\\ &= \frac{S-f_0}{S\theta} - \frac{(S-f_0)}{S\theta^2}(\widehat{\theta}-\theta)\\ &= \widetilde{\psi}\left\{1 - \frac{(\widehat{\theta}-\theta)}{\theta}\right\}\\ &\approx \widetilde{\psi}\left\{1 - \frac{\tau(1-p)^{\tau-1}(\widehat{p}-p)}{\theta}\right\}\end{aligned} \tag{4.7}$$

where $\widehat{\psi}'(\theta) = \partial\widehat{\psi}/\partial\theta = -(S-f_0)/S\theta^2$. Alternatively, the expansion for $\widehat{\psi}$ may be found directly in terms of p, where $\widehat{\psi} \approx \widehat{\psi}(p) + \widehat{\psi}'(p)(\widehat{p}-p)$, and $\widehat{\psi}'(p) = (S-f_0)\theta'/S\theta^2$.

Then noting that $E\left(\widehat{\psi} \,\middle|\, b, S-f_0\right) \approx \widetilde{\psi}$ and $E(\widetilde{\psi}^2) = \psi(1-\psi\theta)/(S\theta) + \psi^2$ ((4.30) and (4.31), see Appendix 4.9.1 for detailed derivations) yields

$$\begin{aligned}\text{Var}(\widehat{\psi} \,|\, b, S-f_0) &\approx E(\widehat{\psi}^2 \,|\, b, S-f_0) - \left\{E(\widehat{\psi} \,|\, b, S-f_0)\right\}^2\\ &\approx \left(\frac{\psi(1-\psi\theta)}{S\theta} + \psi^2\right)\left\{1 - \frac{1}{\theta^2}\text{Var}(\widehat{\theta} \,|\, b, S-f_0)\right\} - \widetilde{\psi}^2\\ &\approx \frac{\widetilde{\psi}^2\tau^2(1-p)^{2(\tau-1)}}{\theta^2}\text{Var}(\widehat{p} \,|\, b, S-f_0).\end{aligned} \tag{4.8}$$

The derivation for the conditional squared expectation, $E(\widehat{\psi}^2 \,|\, b, f_0)$, and for $\text{Var}(\widehat{\theta} \,|\, b, S-f_0)$ can be found in Appendix 4.9.2.

To find an approximation to the variance of $\widehat{\psi}$ we condition on b (= the number of remaining visits after the time of first detection) and $S-f_0$ (= the number

[6] Note the similarity to $K \overset{\text{d}}{=} \text{Bi}(N, \psi\theta)$ as derived in Chapter 2 (Equation (2.31), Section 2.5.2), and can be re-written $N-K \overset{\text{d}}{=} \text{Bi}(N, 1-\psi\theta)$, where $N \equiv S$ and $K = S-f_0$.

of detected sites). Then

$$\begin{aligned}\text{Var}(\widehat{\psi}) &= \text{Var}\left\{E\left(\widehat{\psi}\,|\,b, S-f_0\right)\right\} + E\left\{\text{Var}\left(\widehat{\psi}\,|\,b, S-f_0\right)\right\} \\ &\approx \frac{\psi(1-\psi\theta)}{S\theta} + \left(\frac{\psi(1-\psi\theta)}{S\theta} + \psi^2\right)\frac{\tau^2(1-p)^{2(\tau-1)}}{\theta^2}\frac{p(1-p)}{b},\end{aligned} \tag{4.9}$$

where $\text{Var}(\tilde{\psi})$ was derived above and a detailed derivation for $\text{Var}(\widehat{\psi})$ is given in Appendix 4.9.3

Recall the asymptotic variance given by MacKenzie et al. (2002) (Equation (2.28), Section 2.4 of Chapter 2) as well as the exact variance that we derive in Chapter 2 (Equation (2.44), Section 2.5.4 of Chapter 2). We compare these asymptotic variances to that of the partial likelihood in the following sections.

4.3.4 Comparisons

Here, we conduct a small simulation study to compare four methods for estimating occupancy and detectability for the site homogenous case. We use the partial likelihood, and four methods for the full likelihood. These include, direct maximisation of the full likelihood with: 1) the `optim` procedure in R with our methods from Chapter 3, 2) the `occu` procedure in the R `unmarked` package that also implements `optim`. Then, with 3) the score equations (Equations (2.10)) or Edge solutions (Section 2.3.2) of the full likelihood given in Chapter 2, and 4) the 'single-season-constant-p' model in `PRESENCE` (MacKenzie and Hines, 2009). We simulate from two populations for $S = 28$ sites over $\tau = 10$ survey occasions. The results are presented in Table 4.1. As usual, the full likelihood based estimators are not in closed form and require a numerical approximation for $\widehat{\theta}$ to obtain $\widehat{p}$ and $\widehat{\psi}$. We adopt `uniroot` in R to use with `optim`.

4.3.5 Efficiency

To examine the efficiency of the partial likelihood approach, we simulate an experiment and compute the full and partial likelihood estimates. We use `optim` slightly differently from Section 4.3.4 for direct maximisation of the full likelihood. This method uses the partial likelihood estimates as starting values for $\widehat{\psi}$ and $\widehat{p}$ in the full likelihood direct maximisation. The standard errors of

	$\tau = 10$	
	$\widehat{\psi}$	$\widehat{p}$
Partial	0.794	0.530
`optim`	0.794	0.526
Edge solutions	0.794	0.526
`PRESENCE`	0.794	0.526
`unmarked`	0.794	0.526

Table 4.1: Comparison of the homogeneous partial likelihood estimates to estimates from three methods for the full likelihood.

the full maximum likelihood estimates were computed using the numerically computed observed information matrix. We conducted 1000 simulations at each step. Results are shown in Table 4.2. We report the medians of the estimated values and the associated standard errors and the median absolute deviation (MAD) of the estimates, given by MAD $= c \times$ median $(|x_i - x_M|)$, where $c = 1/\Phi^{(-1)}(3/4)$, $\Phi^{(-1)}(3/4)$ is the third quartile of the inverse standard normal distribution, and x_M is the median of x_i. The efficiency is the usual ratio of variances. In neither method is there any indication of bias, both estimated standard errors appear reliable and the efficiency of the partial likelihood method for ψ is above 90%. In settings with smaller values of p or S, the stability of the full likelihood estimators was an issue so we do not consider them further in the simulations.

	Partial		Full		Partial		Full	
	$\widehat{p}$	$\widehat{\psi}$	$\widehat{p}$	$\widehat{\psi}$	$\widehat{p}$	$\widehat{\psi}$	$\widehat{p}$	$\widehat{\psi}$
$S = 1000$, $\tau = 5$	0.100	0.400	0.100	0.400	0.050	0.400	0.050	0.400
Median estimate	0.100	0.401	0.100	0.402	0.049	0.411	0.049	0.413
Median s.e.	0.016	0.058	0.015	0.057	0.016	0.128	0.015	0.123
MAD	0.015	0.060	0.015	0.057	0.016	0.128	0.016	0.127
Efficiency	0.915	0.925			0.957	0.991		
$S = 100$, $\tau = 5$	0.200	0.400	0.200	0.400	0.200	0.600	0.200	0.600
Median estimate	0.199	0.405	0.197	0.411	0.198	0.609	0.197	0.609
Median s.e.	0.049	0.091	0.045	0.088	0.040	0.106	0.037	0.101
MAD	0.051	0.093	0.045	0.089	0.039	0.103	0.036	0.101
Efficiency	0.801	0.914			0.843	0.909		

Table 4.2: Simulation results to compare the partial likelihood and full likelihood approaches.

To examine the effects of small probabilities and different numbers of occasions, we took $S = 100$, $p = 0.05$, $\psi = 0.6$ and $\tau = 5$ or $\tau = 10$ for 1000 simulations. Results are summarised in Table 4.3 where we only report for simulations that satisfy $f_0 > S(1 - \widehat{p})^\tau$ since this ensures that $\widehat{\psi} < 1$ by construction of the estimator (Equation (4.5)). There is evidence of some bias for small values of p and small τ which are consistent with findings from Chapters 2 and 3, and from findings of published studies as discussed previously (e.g. Wintle et al., 2004; Guillera-Arroita et al., 2010).

	$\tau = 5$		$\tau = 10$	
	$\widehat{p}$	$\widehat{\psi}$	$\widehat{p}$	$\widehat{\psi}$
Value	0.050	0.600	0.050	0.600
Median estimate	0.067	0.475	0.052	0.587
Median s.e.	0.046	0.326	0.020	0.208
MAD	0.031	0.227	0.017	0.182

Table 4.3: Simulations of the interative estimator with $S = 100$ sites and small values of p ($p = 0.05, \psi = 0.6$).

Finally, we took $S = 27$, $\tau = 4$, $\psi = 0.6$ and $p = 0.6$, which is similar to the values in the frog application (Section 4.3.6). The results are reported in Table 4.4. In this setting, the estimator again performs well.

	Partial		Full	
	$\widehat{p}$	$\widehat{\psi}$	$\widehat{p}$	$\widehat{\psi}$
Value	0.600	0.600	0.600	0.600
Median estimate	0.600	0.604	0.600	0.604
Median s.e.	0.078	0.097	0.066	0.097
MAD	0.078	0.104	0.065	0.105
Efficiency	0.709	0.991		

Table 4.4: Simulations with a small number of sites ($S = 27$), small number of occasions ($\tau = 4$) and large values of p are comparable to those in the example ($p = 0.6, \psi = 0.6$).

The conclusion from our small simulation study is that the two-stage partial likelihood approach has quite good efficiency to estimate ψ compared with the full maximum likelihood approach, and as there are analytic forms for the estimators they are more stable than the full maximum likelihood estimators.

4.3.6 Applications

We apply the partial likelihood approach to the frog data and compare this to the full likelihood approach.

The partial likelihood equations give $\widehat{p} = (y-O)/b = (47-15)/36 = 0.889$ with standard error 0.052 and $\widehat{\psi} = (S - f_0)/(S\widehat{\theta}) \approx 15/27 = 0.556$ with standard error 0.096. Here $\widehat{\theta}$ is close to 1. Recall that in Section 2.3.1 of Chapter 2 we found $\widehat{\theta} = 1 - (1 - \widehat{p})^{\tau} = 0.997$.

The full likelihood yields estimates $\widehat{p} = 0.782$ with standard error 0.054 and $\widehat{\psi} = 0.557$ with standard error 0.096. These were generated in the usual way, direct maximisation of the likelihood with the `optim` procedure in R (described in Chapter 3) and starting values from the partial likelihood.

Some of the difference in the estimates of $\widehat{p}$ may be explained by the estimators themselves. The partial likelihood considers redetections, whereas the full likelihood incorporates information from the first detection. So it may be that there is some difference between the probability of detection between the first, and subsequent detections. Moreover, it is possible that a more accurate estimate for detection could be reached if detection probability was not assumed constant between visitations, or among sites. We will explore these scenarios in the following sections.

An alternative approach to estimating the detection probabilities is to condition on at least one detection at a site. This may be implemented in VGLM and when applied to the sample data gave $\widehat{p} = 0.783$, which agrees more closely with the full likelihood estimate. We examine the conditional likelihood in Section 4.6.

When applied to the Coosa bass data, the two-stage method yielded $\widehat{p} = 0.83$ with estimated standard error 0.03 and $\widehat{\psi} = 0.78$ with estimated standard error 0.06. The full likelihood yielded $\widehat{p} = 0.80$ with estimated standard error 0.03 and $\widehat{\psi} = 0.78$ with estimated standard error 0.06. Implementing conditional likelihood via VGLM gave $\widehat{p} = 0.80$ when applied to these data.

When applied to the brook trout data, we obtained $\widehat{p} = 0.54$, with estimated standard error 0.07 and $\widehat{\psi} = 0.46$ with estimated standard error 0.07. The full likelihood yielded $\widehat{p} = 0.53$ with estimated standard error 0.06 and $\widehat{\psi} = 0.46$ with estimated standard error 0.07. Implementing conditional likelihood via VGLM gave $\widehat{p} = 0.59$ when applied to these data.

These three applications indicate good agreement between the partial likelihood and full likelihood estimates.

4.4 Site inhomogeneity

The homogeneous model assumes that ψ and p remain constant across sites and survey occasions. Of more interest, and relevance to practical situations, is when occupancy and detectability are allowed to vary between sites according to some covariates.

Thus, we extend the theory of the basic (homogeneous) partial likelihood function to allow the occupancy and detection probabilities to vary among sites according to observable covariates. It is possible to compute a full likelihood over both occupancy and detection probabilities, but, particularly if semiparametric models are used, the number of parameters in the likelihood could become quite large. As we observed in the simple homogeneous model that the full likelihood can be numerically unstable, and taking into account the observations of Welsh et al. (2013), here we extend the two–stage approach.

As mentioned earlier we consider three approaches to estimating occupancy in the presence of site inhomogeneity. In all approaches the detection probabilities are estimated in the first stage using GLM or GAM. A direct maximisation of the partial likelihood $L_1(\psi, p)$ of the second stage is presented in Section 4.4.3. An iterative approach is developed in Section 4.4.3 and finally a ratio estimator is developed in Section 4.4.4. In Section 4.4.5 we show how the iterative approach may be extended to incorporate GAMs using the R `mgcv` package.

4.4.1 Notation and the likelihood

Now let $\psi_s = h(x_s^T\alpha)$ where h is for example the logistic function $h(x) = (1+\exp(-x))^{-1}$ for a vector of covariates $x_s \in \mathbb{R}^{q_\psi}$ and a vector of parameters $\alpha \in \mathbb{R}^{q_\psi}$. We also allow the detection probabilities to depend on covariates, $p_s = p(u_s, \beta) = h(u_s^T\beta)$ for a possibly different vector of covariates $u_s \in \mathbb{R}^{q_p}$ and parameters $\beta \in \mathbb{R}^{q_p}$. Recall y_s is the number of occasions upon which the species was detected at site s, $z_s = I(y_s = 0)$[7] is the indicator of no detections at site s and for $y_s > 0$ let $r_s = y_s - 1$ be the number of redetections at site s. Further, recall we have labelled the sites where at least one individual

[7] In chapter 2 we defined $u_i = \mathrm{I}(x_{i\bullet} \geqslant 1) \Rightarrow z_s = 1 - u_i$ $(i \equiv s)$

was observed $1, \ldots, O$. In the heterogeneous case, (4.1)-(4.3) are modified to incorporate heterogeneity and now the contribution to the full likelihood of site s is

$$\begin{aligned} L_s(\psi_s, p_s) &= (1 - \psi_s + \psi_s(1-p_s)^\tau)^{z_s} \left\{ \binom{\tau}{y_s} \psi_s p_s^{y_s} (1-p_s)^{\tau - y_s} \right\}^{1-z_s} \\ &\propto (1 - \psi_s\theta_s)^{z_s} \psi_s^{1-z_s} \left\{ p_s(1-p_s)^{(a_s-1)} \right\}^{(1-z_s)} \left\{ p_s^{(y_s-1)} (1-p_s)^{b_s - y_s + 1} \right\}^{(1-z_s)} \\ &= L_{1s}(\psi_s, p_s) L_{2s}(p_s) \end{aligned} \tag{4.10}$$

where

$$L_{2s}(p_s) = \left\{ p_s^{(y_s-1)} (1-p_s)^{b_s - y_s + 1} \right\}^{(1-z_s)} \tag{4.11}$$

and

$$\begin{aligned} L_{1s}(\psi_s, p_s) &= (1 - \psi_s + \psi_s(1-p_s)^\tau)^{z_s} \psi_s^{1-z_s} \left\{ p_s(1-p_s)^{(a_s-1)} \right\}^{(1-z_s)} \\ &= (1 - \psi_s\theta_s)^{z_s} \psi_s^{1-z_s} \left\{ p_s(1-p_s)^{(a_s-1)} \right\}^{(1-z_s)} \end{aligned} \tag{4.12}$$

4.4.2 Estimating detection

The partial likelihood for the parameters β, associated with the redetections r_s is

$$L_2(\beta) = \prod_{s=1}^{O} p(u_s, \beta)^{r_s} (1 - p(u_s, \beta))^{b_s - r_s} \tag{4.13}$$

where $s = 1, \ldots, O$ denotes the sites where at least one individual was observed. Then $L_2(\beta)$ corresponds to a binomial model so that β may be estimated by considering redetections over the sites where at least one individual was observed. We denote its estimate by $\hat{\beta}$ and its estimated covariance matrix by $\widehat{V}_\beta$.

In particular, we may fit generalized linear models (GLMs) or generalized additive models (GAMS) to the redetections using standard software.

It is convenient to note here that if there were at least one detection at a site, the score equations for β would involve the differences $r_s - b_s p(u_s, \beta)$, which have mean zero given at least one detection. If there were no detections at a site, then both $b_s = 0$ and $r_s = 0$, so $E(r_s - b_s p(u_s, \beta) | z_s) = 0$ and

the partial score equations for the detection probabilities and the occupancy probabilities are hence uncorrelated. A similar argument applies to inferences based on a likelihood conditional on at least one detection at a site. That is, the conditional score function is also uncorrelated with the partial likelihood for occupancy and hence their development is a relatively minor extension of the present approach. In the literature if this occurs for the full likelihood these parameters are referred to as orthogonal (Cox and Reid, 2004). We raise this in the conclusion (Section 4.8).

4.4.3 Estimating occupancy using partial likelihood

Suppose we have obtained estimates of $\theta_s = 1 - (1 - p_s)^\tau$ for site s, where typically the p_s are estimated using the partial likelihood Section 4.4.2. To estimate ψ_s we estimate the coefficients α by maximising the partial likelihood $\prod_{s=1}^{S} L_{1s}(\psi_s, \widehat{p}_s)$.

4.4.3.1 Likelihood and standard error

Let, $p_s = p_s(\beta)$ be the detection probabilities p_s. Then the portion of the partial likelihood that involves the ψ_s is

$$L(\alpha, \beta) = \prod_{s=1}^{S} L_{1s}(\psi_s, \beta) \approx \prod_{s=1}^{S} (1-\psi_s+\psi_s(1-p_s)^\tau)^{z_s} \psi_s^{1-z_s} = \prod_{s=1}^{S} (1-\psi_s\theta_s)^{z_s} \psi_s^{1-z_s}.$$

Let $\widehat{\alpha}(\beta)$ be the estimator of α arising from solving $Q(\alpha, \beta) = 0$ for a given β. Denote the partial score function by $Q(\alpha, \beta) = \partial \ell(\alpha, \beta)/\partial \alpha^T$; the second partial derivative with respect to α by $I(\alpha, \beta) = \partial Q(\alpha, \beta)/\partial \alpha^T$; and the partial derivative with respect to β by $\widetilde{B}(\alpha, \beta) = \partial Q(\alpha, \beta)/\partial \beta^T$. We use these quantities to derive the expressions for the asymptotic variance of our estimators in the linear logistic case. Specifically, we will derive the variance for $\widehat{\alpha}(\widehat{\beta})$ under mild regularity conditions.

Let α_0 and β_0 be the true values of α and β respectively. Suppose that $\widehat{\beta}$ is a consistent estimator for β arising from the first stage. We suppose that for some $q_s \times q_s$ matrix $B(\alpha, \beta)$ and for convergence in probability, or almost surely,

$$S^{-1} \frac{\partial Q(\alpha_0, \beta_0)}{\partial \beta_0^T} \to B(\alpha_0, \beta_0)$$

and we further suppose that for some $q_o \times q_o$ matrix $A(\alpha, \beta)$,

$$S^{-1} I(\alpha_0, \beta_0) = -S^{-1} \frac{\partial Q(\alpha_0, \beta_0)}{\partial \alpha_0^T} \to A(\alpha_0, \beta_0),$$

where $I(\alpha, \beta) = -\partial Q(\alpha, \beta)/\partial \alpha^T$. Moreover, we suppose that the central limit theorem (CLT) is applicable to $Q(\alpha_0, \beta_0)$ and

$$S^{-1/2} Q(\alpha_0, \beta_0) \xrightarrow{d} N(0, \Sigma_Q).$$

This will often follow if the sites are independent (see Appendix 4.10 for a detailed derivation). We also suppose that the estimate from the first stage satisfies

$$S^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Sigma_\beta).$$

This will usually follow from the GLM procedure applied in the first stage. Finally, we suppose that the partial score functions for α are uncorrelated with those for β. We have noted in Section 4.4.2 that this holds for our partial likelihood and for the likelihood conditional on at least one detection at a site.

The log-partial likelihood $\sum_{s=1}^{S} \log L_{1s}(\psi_s, p_s)$ is equal to

$$\ell(\alpha, \beta) = \sum_{s=1}^{S} \{(1 - w_s) \log(1 - \psi_s(\alpha)\theta_s) + w \log(\psi_s(\alpha))\}$$

so that

$$Q(\alpha, \beta) \approx \frac{\partial \ell(\alpha, \beta)}{\partial \alpha^T} = \sum_{s=1}^{S} x_s \frac{\{w_s - \psi_s(\alpha)\theta_s\}(1 - \psi_s(\alpha))}{1 - \psi_s(\alpha)\theta_s},$$

where a detailed derivation is given in Appendix 4.10.1.

The second derivative, or the derivative of $Q(\alpha, \beta)$ with respect to α is found by the chain rule for $\dfrac{\partial Q}{\partial \alpha^T} = \dfrac{\partial Q}{\partial \psi_s} \dfrac{\partial \psi_s}{\partial \alpha^T}$ to be

$$\begin{aligned} I(\alpha, \beta) &= -\frac{\partial Q(\alpha, \beta)}{\partial \alpha^T} \\ &= \sum_{s=1}^{S} x_s x_s^T \left\{ \frac{\theta_s - 2\psi_s(\alpha)\theta_s + \psi_s(\alpha)^2 \theta_s^2 + w_s(1 - \theta_s)}{\{(1 - \psi_s(\alpha)\theta_s\}^2} \right\} \psi_s(\alpha)\{1 - \psi_s(\alpha)\}. \end{aligned} \tag{4.14}$$

Then, to obtain the estimators $\widehat{\alpha}(\widehat{\beta})$ we solve $Q(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}) = 0$. Now, a Taylor expansion for Q about α_0 gives

$$0 = Q(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}) \approx Q(\alpha_0, \widehat{\beta}) + I(\alpha_0, \widehat{\beta}) \left(\widehat{\alpha}(\widehat{\beta}) - \alpha_0 \right),$$

where $I(\alpha_0, \widehat{\beta}) = -\partial Q(\alpha_0, \widehat{\beta})/\partial \alpha_0^T$ so that

$$\widehat{\alpha}(\widehat{\beta}) - \alpha_0 = -I(\alpha_0, \widehat{\beta})^{-1} Q(\alpha_0, \widehat{\beta}).$$

Multiply the above equation through by $S^{1/2}$ to get

$$S^{1/2} \left(\widehat{\alpha}(\widehat{\beta}) - \alpha_0 \right) = - \left\{ S^{-1} I(\alpha_0, \widehat{\beta}) \right\}^{-1} S^{-1/2} Q(\alpha_0, \widehat{\beta}). \tag{4.15}$$

Using the consistency of $\widehat{\beta}$ we see that

$$S^{-1} I(\alpha_0, \widehat{\beta}) = S^{-1} I(\alpha_0, \beta_0) + O(\widehat{\beta} - \beta_0) \approx S^{-1} I(\alpha_0, \beta_0).$$

Next an expansion for $\widehat{\beta}$ leads to

$$S^{-1/2} Q(\alpha_0, \widehat{\beta}) \approx S^{-1/2} Q(\alpha_0, \beta_0) + S^{-1} \frac{\partial Q(\alpha_0, \beta_0)}{\partial \beta_0} S^{1/2} (\widehat{\beta} - \beta_0). \tag{4.16}$$

To determine $\widetilde{B}(\alpha, \beta) = \partial Q(\alpha, \beta)/\partial \beta^T$ note that

$$\frac{\partial Q(\alpha, \beta)}{\partial \theta_s} = -\frac{x_s \psi_s (1 - \psi_s)(1 - w_s)}{(1 - \psi_s \theta_s)^2} \tag{4.17}$$

and recall that $\dfrac{\partial \theta_s}{\partial p_s} = \tau (1 - p_s)^{\tau - 1}$ so that using the chain rule $\dfrac{\partial Q}{\partial p_s} = \dfrac{\partial Q}{\partial \theta_s} \dfrac{\partial \theta_s}{\partial p_s}$,

$$q_s = \frac{\partial Q(\alpha, \beta)}{\partial p_s} = -\frac{x_s \psi_s (1 - \psi_s)(1 - w_s) \tau (1 - p_s)^{\tau - 1}}{(1 - \psi_s \theta_s)^2}.$$

As $\partial p_s / \partial \beta^T = p_s (1 - p_s) u_s$ and $\dfrac{\partial Q}{\partial \beta^T} = \dfrac{\partial Q}{\partial p_s} \dfrac{\partial p_s}{\partial \beta^T}$ we then see that

$$\widetilde{B}(\alpha, \beta) = \frac{\partial Q(\alpha, \beta)}{\partial \beta^T} = -\sum_{s=1}^{S} \frac{x_s u_s^T \psi_s (1 - \psi_s)(1 - w_s) \tau (1 - p_s)^{\tau} p_s}{(1 - \psi_s \theta_s)^2}. \tag{4.18}$$

Recall that $S^{-1} I(\alpha_0, \beta_0) \approx A(\alpha_0, \beta_0)$ and $S^{-1} \partial Q(\alpha_0, \beta_0)/\partial \beta_0^T \approx B(\alpha_0, \beta_0)$ so that substituting these, together with equation (4.16), (back) into equation

(4.15) gives

$$S^{1/2}(\widehat{\alpha}(\widehat{\beta}) - \alpha_0) = -\left\{S^{-1}I(\alpha_0, \widehat{\beta})\right\}^{-1} S^{-1/2}Q(\alpha_0, \widehat{\beta})$$
$$\approx -A(\alpha, \beta)^{-1}\left\{S^{-1/2}Q(\alpha_0, \beta_0) + B(\alpha_0, \beta_0)S^{1/2}(\hat{\beta} - \beta)\right\}.$$

Then

$$S^{1/2}(\widehat{\alpha}(\widehat{\beta}) - \alpha_0) \sim N\big(0, A(\alpha_0, \beta_0)^{-1}\left\{\Sigma_Q + B(\alpha_0, \beta_0)\Sigma_\beta B(\alpha_0, \beta_0)^T\right\} A(\alpha_0, \beta_0)^{-T}\big),$$

So, $\mathrm{Var}(S^{1/2}\{\widehat{\alpha}(\widehat{\beta}) - \alpha_0)\} = S\ \mathrm{Var}\{\widehat{\alpha}(\widehat{\beta})\}$, or

$$\mathrm{Var}\{\widehat{\alpha}(\widehat{\beta})\} = S^{-1}\left\{A(\alpha_0, \beta_0)^{-1} + A(\alpha_0, \beta_0)^{-1}B(\alpha_0, \beta_0)\Sigma_\beta B(\alpha_0, \beta_0)^T A(\alpha_0, \beta_0)^{-T}\right\}.$$

Note that $\Sigma_\beta = S\mathrm{Var}(\widehat{\beta}) = SV_\beta$ where V_β is the covariance matrix of $\widehat{\beta}$. And, that $A(\alpha_0, \beta_0)^{-1}\Sigma_Q A(\alpha_0, \beta_0)^{-T} = A(\alpha_0, \beta_0)^{-1}$, which implies that $\Sigma_Q = A(\alpha_0, \beta 0)^T$. Thus we estimate $\mathrm{Var}\{\widehat{\alpha}(\widehat{\beta})\}$ by

$$\widehat{\mathrm{Var}}\{\widehat{\alpha}(\widehat{\beta})\} = S^{-1}\Big[\{S^{-1}I(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})\}^{-1}$$
$$+ \{S^{-1}I(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})\}^{-1}S^{-1}\widetilde{B}(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})S\widehat{V}_\beta S^{-1}\widetilde{B}(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})^T\{S^{-1}I(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})\}^{-1}\Big]$$

which reduces to

$$\widehat{\mathrm{Var}}\{\widehat{\alpha}(\widehat{\beta})\} = I\{(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})\}^{-1}$$
$$+ I\{(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})\}^{-1}\widetilde{B}\{(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})\}\widehat{V}_B\widetilde{B}\{(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})\}^T I\{(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})\}^{-1}. \tag{4.19}$$

As noted above, $\widehat{V}_\beta$ is computed by the GLM in the first stage, $I(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})$ may be computed from (4.14) and $\widetilde{B}(\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})$ from (4.18).

We use $\widehat{\beta}$ obtained from the first stage, based on the $S - f_0$ sites with redetections, to predict $\widehat{\theta}_s$ for the remainder f_0 sites of the detection matrix i.e. for sites without redetections.

Then, the estimated occupancy probability for site s is

$$\widehat{\psi}_s = [1 + \exp\{-x_s^T\widehat{\alpha}(\widehat{\beta})\}]^{-1},$$

which has estimated variance

$$\widehat{\mathrm{Var}}(\widehat{\psi}_s) = \{\widehat{\psi}_s(1 - \widehat{\psi}_s)\}^2 x_s^T\widehat{\mathrm{Var}}\{\widehat{\alpha}(\widehat{\beta})\}x_s.$$

We give the detailed derivations in the appendices (Section 4.10).

4.4.3.2 Computing the estimates - offset with the iterative procedure

The second stage estimates may be computed by maximising the partial likelihood. However, this direct approach does not allow us to use GLM methods.

Let $w_s = 1 - z_s$ be an indicator of presence for site s. Suppose we have estimates of θ_s for site s, typically estimated from the partial likelihood of the redetections. Suppose the sites are labelled $s = 1, \ldots, S$. Write the contribution to the likelihood of a single site by

$$L_s = (1 - \psi_s\theta_s)^{z_s}\psi_s^{1-z_s} \propto (1 - \psi_s\theta_s)^{z_s}(\psi_s\theta_s)^{1-z_s} = (1 - \psi_s\theta_s)^{1-w_s}(\psi_s\theta_s)^{w_s}, \tag{4.20}$$

which is proportional to a binomial likelihood with probability $\psi_s\theta_s$.

Consider a single site. Now under the logistic model,

$$\psi_s(x) = \frac{\exp(\alpha^T x_s)}{1 + \exp(\alpha^T x_s)}$$

so that

$$\psi_s(x)\theta_s = \frac{\exp(\alpha^T x_s + \log(\theta_s))}{1 + \exp(\alpha^T x_s)}$$

If we let

$$a_s(x) = \log(\theta_s) - \log\{1 + \exp(\alpha^T x_s)(1 - \theta_s)\},$$

we have

$$\psi_s(x)\widehat{\theta}_s = \frac{\exp(\alpha^T x_s + a_s(x))}{1 + \exp(\alpha^T x_s + a_s(x))}$$

so that $a_s(x)$ has the appearance of an offset. However, it is a function of the linear predictor $\alpha^T x_s$.

This suggests a simple iterative approach to estimation in the second stage. Let $\widehat{\alpha}_0$ be an initial estimate for α which may be obtained by fitting a logistic model for w_s without any offset.

1. If $\widehat{\alpha}_{i-1}$ is the estimate of α from the previous step, compute

$$a_s^{(i)}(x) = \log(\widehat{\theta}_s) - \log\{1 + \exp(\alpha_{i-1}^T x_s)(1 - \widehat{\theta}_s)\}.$$

2. Fit a binomial generalised linear model to w_s with the usual logit link and offset $a_s^{(i)}(x)$ to produce a new estimate $\widehat{\alpha}_i$ of α.

Repeat steps 1 and 2 until convergence.

Then we obtain estimates for ψ_s from

$$\widehat{\psi}_s(x) = \frac{\exp(\widehat{\alpha}_i^T x_s)}{1 + \exp(\widehat{\alpha}_i^T x_s)}.$$

4.4.3.3 Small simulation study

To assess the performance of the iterative estimator and the approximate standard error we conducted a small simulation study. We considered two independent covariates x_1 and x_2 each with $N(0, 2)$ distributions and took $\alpha_\psi = (1, 0.5, 0.5)$ and $\beta_p = (-1, 0.5, 0.5)$. In each set of simulations we simulated the covariates once then simulated 1000 detection experiments. We only report the results for α associated with the sites and give these in Table 4.5. These indicate that the estimators are relatively unbiased and the formula for the standard errors is reasonable.

	α_0	α_1	α_2
$S = 1000$, $\tau = 4$			
Actual	1.000	0.500	0.500
Median	1.016	0.506	0.499
MAD	0.197	0.094	0.097
Med s.e.	0.202	0.096	0.096
$S = 1000$, $\tau = 6$			
Actual	1.000	0.500	0.500
Median	1.006	0.504	0.499
MAD	0.136	0.075	0.085
Med s.e.	0.136	0.075	0.078

Table 4.5: Results of simulations for the covariate model.

4.4.3.4 Simulations - iterative and `unmarked`

We compare estimates from the iterative procedure to the full likelihood as implemented with the `occu` procedure from the R `unmarked` package. The `occu` procedure fits the single-season model of MacKenzie et al. (2002), and uses a direct maximisation of the likelihood function for simultaneous estimation of ψ and p.

We begin with a null model without any covariates. We simulate a single detection matrix from a population where it is presumed $\psi = 0.6, p = 0.7$, and $S = 100, \tau = 5$. Estimates for ψ and p as well as the estimated intercepts for the same model are shown in Table 4.6. Significance of the estimates are indicated by the t-statistic for occupancy coefficients (t_α) and for detectability coefficients (t_β). Estimates for ψ and p are similar for the iterative method and `occu` procedure.

Then we fit a model with an intercept and two covariates. To generate the population we used starting parameter coefficient values $\beta_p = (0, 0.5, 0)$ and $\alpha_\psi = (0, 0, 0.5)$ for an intercept and two covariates. Then we simulated covariate coefficients from the population and fit the same model with the iterative and with the `unmarked` methods. Table 4.7 shows the coefficient estimates. The results for the two methods are still similar.

To assess the efficiency of the iterative method compared to the `unmarked` software we ran simulations for a variety of S and τ (Table 4.8). We considered the same model for occupancy ψ and detection p. We ran this for moderate (≈ 0.5) values for ψ and p, and for uncorrelated ψ and p.

Table 4.8 summarises for simulations for the occupancy covariate coefficients (α). Specifically, the summaries include: the median (Med.), the median absolute deviations (MAD) and the standard error of the medians (Med se). It reveals that the partial likelihood estimators compare quite well to the full likelihood, where there appears to be a gain in efficiency for $S = 30$. This is unexpected and may be due to either simulation error or convergence problems in the `unmarked` procedure.

We ran 1000 simulations for several scenarios for S and τ (Table 4.8). To generate the population we used starting parameter coefficient values $\beta_p = (0, 0.5, 0)$ and $\alpha_\psi = (0, 0, 0.5)$ for an intercept and two covariates. Then we simulated covariate coefficients from the population and fit the same model with the iterative and with the `unmarked` methods. Specifically, we fit `occu` $= (\sim x + x.1 \sim x + x.1,$ `data`$)$ with the `unmarked` package. We fit the same

covariates with the iterative method.

	$\widehat{p}$	s.e. $\widehat{p}$		$\widehat{\psi}$	s.e. $\widehat{\psi}$	
Actual	0.700			0.600		
iterative	0.691	0.033		0.542	0.050	
`unmarked`	0.687	0.029		0.542	0.050	
	$\widehat{\beta}$	s.e. $\widehat{\beta}$	t_β	$\widehat{\alpha}$	s.e. $\widehat{\alpha}$	t_α
iterative	0.805	0.157	5.142	0.166	0.201	0.827
`unmarked`	0.785	0.133	5.893	0.167	0.201	0.829

Table 4.6: Estimates for a null model using both the iterative and the `occu` procedure for a single simulated history matrix, for S= 100, $\tau = 5$. t-statistics are displayed for occupancy (t_α) and detectability (t_β) coefficients.

		$\widehat{\beta}$	s.e. $\widehat{\beta}$	$t_{\widehat{\beta}}$	$\widehat{\alpha}$	s.e. $\widehat{\alpha}$	$t_{\widehat{\alpha}}$
iterative	Intercept	-0.273	0.177	-1.540	0.206	0.229	0.900
	x	0.442	0.169	2.620	-0.075	0.224	-0.335
	x.1	0.283	0.158	1.798	0.623	0.208	2.995
`unmarked`	Intercept	-0.228	0.153	-1.489	0.198	0.224	0.882
	x	0.459	0.142	3.221	-0.081	0.221	-0.368
	x.1	0.085	0.126	0.673	0.710	0.210	3.380

Table 4.7: Estimates for a single simulated data set for $S = 150$ and $\tau = 4$.

		$\tau = 4$			$\tau = 8$			$\tau = 16$		
		α_0	α_1	α_2	α_0	α_1	α_2	α_0	α_1	α_2
S	Actual	0	0	0.5	0	0	0.5	0	0	0.5
30	$\bar{\psi}_s$, $\bar{p}_s$, ρ	0.521	0.501	-0.215	0.502	0.480	0.029	0.462	0.459	0.165
Iter.	Med.	0.076	0.029	0.607	0.055	-0.024	0.559	-0.027	0.008	0.544
	MAD	0.496	0.522	0.55	0.385	0.487	0.562	0.459	0.469	0.467
	Med se	0.507	0.519	0.542	0.406	0.467	0.534	0.430	0.436	0.435
Unm.	Med.	0.201	-0.077	0.576	0.026	-0.044	0.628	0.024	-0.006	0.564
	MAD	0.594	0.642	0.655	0.409	0.519	0.586	0.446	0.471	0.470
	Med se	0.534	0.539	0.554	0.408	0.470	0.541	0.431	0.440	0.442
150	$\bar{\psi}_s$, $\bar{p}_s$, ρ	0.493	0.523	0.027	0.494	0.487	-0.085	0.503	0.493	0.120
Iter.	Med.	0.029	-0.022	0.526	-0.005	-0.009	0.512	0.001	-0.001	0.505
	MAD	0.215	0.219	0.200	0.181	0.164	0.172	0.161	0.176	0.189
	Med se	0.208	0.212	0.193	0.175	0.173	0.172	0.170	0.168	0.175
Unm.	Med.	0.025	-0.020	0.525	-0.008	-0.009	0.515	0.001	0.000	0.505
	MAD	0.196	0.216	0.191	0.182	0.165	0.177	0.162	0.175	0.189
	Med se	0.204	0.209	0.191	0.174	0.172	0.172	0.170	0.168	0.175
400	$\bar{\psi}_s$, $\bar{p}_s$, ρ	0.508	0.494	-0.047	0.498	0.505	-0.024	0.502	0.503	-0.063
Iter.	Med.	0.013	-0.015	0.515	0.006	-0.005	0.501	0.004	-0.006	0.504
	MAD	0.129	0.129	0.126	0.103	0.101	0.119	0.108	0.101	0.110
	Med se	0.122	0.124	0.123	0.105	0.105	0.117	0.103	0.100	0.109
Unm.	Med.	0.008	-0.018	0.515	0.006	-0.004	0.503	0.004	-0.005	0.505
	MAD	0.125	0.124	0.116	0.104	0.100	0.120	0.109	0.101	0.110
	Med se	0.120	0.122	0.121	0.105	0.105	0.117	0.103	0.100	0.109
1000	$\bar{\psi}_s$, $\bar{p}_s$, ρ	0.508	0.505	0.036	0.502	0.505	0.004	0.504	0.492	-0.022
Iter.	Med.	0.004	0.002	0.507	-0.001	-0.003	0.502	-0.005	0.003	0.500
	MAD	0.078	0.077	0.086	0.072	0.070	0.068	0.068	0.065	0.066
	Med se	0.076	0.079	0.080	0.066	0.069	0.069	0.065	0.065	0.068
Unm.	Med.	0.004	0.002	0.508	0.000	-0.003	0.501	-0.005	0.002	0.500
	MAD	0.076	0.078	0.084	0.071	0.069	0.068	0.068	0.065	0.066
	Med se	0.075	0.078	0.079	0.066	0.069	0.068	0.065	0.065	0.068

Table 4.8: Summaries for occupancy coefficient estimates for 1000 simulations for the iterative GLM and `occu` procedures. Detection starting values were $\beta_p = (0, 0.5, 0)$.

4.4.3.5 Applications

The Coosa Bass data included two covariates, stream link magnitude (MAG) and the coefficient of variation of stream flow (CV). Fitting both terms as linear effects yielded Table 4.9 a). We would conclude from the table that detection is not related to either of the covariates but occupancy is related to MAG. For the brook trout data we fitted the single covariate Elevation (Ele.) as a linear term. See Table 4.9 b). This term is related to occupancy but not detection. A variety of covariates were available for the data of Table 4.9 c). We considered area and a measure of water vegetation (WV). Neither were related to occupancy or detection. An intercept (Int.) was included in all models.

		Detection			Occupancy		
		$\widehat{\beta}$	s.e. $\widehat{\beta}$	t_β	$\widehat{\alpha}$	s.e. $\widehat{\alpha}$	t_α
a) Coosa bass							
iterative	Int.	1.144	0.538	2.126	4.106	1.843	2.228
	MAG	-0.195	0.275	-0.710	3.707	1.353	2.741
	CV	0.393	0.354	1.110	-0.155	0.693	-0.223
`unmarked`	Int.	1.252	0.440	2.846	3.942	1.751	2.251
	MAG	-0.269	0.213	-1.265	3.570	1.277	2.797
	CV	0.219	0.270	0.812	-0.143	0.680	-0.210
b) Brook trout							
iterative	Int.	-0.1031	1.4304	-0.0721	-4.1534	1.1773	-3.5278
	Ele.	0.0001	0.0004	0.1982	0.0014	0.0004	3.5578
`unmarked`	Int.	-2.207	0.205	-10.792	-3.338	1.533	-2.178
	Elev.	0.000	0.000	7.739	0.003	0.001	4.031
c) Frogs							
iterative	Int.	-0.052	1.310	-0.040	-0.494	0.796	-0.621
	Area	0.000	0.000	0.864	0.000	0.000	0.965
	WV	0.049	0.037	1.310	0.014	0.019	0.747
`unmarked`	Int.	-0.0006	0.5356	-0.0012	-0.0005	0.7826	-0.0007
	Area	0.0001	0.0001	1.0582	0.0002	0.0002	0.9414
	WV	0.0037	0.0128	0.2878	-0.0081	0.0193	-0.4212

Table 4.9: iterative GLM and `unmarked` with linear effects a) for the Coosa bass data, b) for the Brook trout data, c) for the reduced frog data (GGF).

4.4.4 Ratio estimator

From (4.5) an alternative and simpler approach to the second stage involves a ratio estimator where we estimate the probability η_s that a species is observed at site s. This gives a natural ratio estimator for occupancy since $\eta_s = \psi_s\theta_s$ gives $\psi_s = \eta_s/\theta_s$.

We will see that this estimator works well under certain conditions. For example, it is suitable when p is large, but not for small p. However, obtaining analytic standard errors is not trivial. Lastly, we will show that the ratio estimator does not work well in the GAMs family, unless p is large.

We show next how to obtain estimates for ψ_s.

4.4.4.1 Computing the estimates

Let $\hat{p}_s$ be the estimate of p_s from $L_{2s}(p_s)$ as outlined in Section 4.4.2, the first stage of estimation. Then the second stage of estimating occupancy, proceeds as follows. Recall that we could re-write the likelihood as in Equation (4.20), where

$$\eta_s = \psi_s\theta_s, \tag{4.21}$$

is the probability there has been at least one sighting (θ_s) i.e. $P(w_s = 1) = \eta_s = \psi_s\theta_s$ and the likelihood is

$$L_{1s}(\eta_s, p_s) = (1-\eta_s)^{z_s}\eta_s^{1-z_s} = (1-\eta_s)^{(1-w_s)}\eta_s^{w_s}.$$

Now, we use a logistic regression to fit the model

$$\eta_s(x) = \frac{\exp(\delta^T x_s)}{1+\exp(\delta^T x_s)}, \tag{4.22}$$

and estimates for ψ are obtained by,

$$\breve{\psi}_s(x) = \frac{\exp(\widehat{\delta}^T x_s)}{\left(1+\exp(\widehat{\delta}^T x_s)\right)\,\widehat{\theta}_s}, \tag{4.23}$$

since $\widehat{\eta}_s = \breve{\psi}_s\widehat{\theta}_s$ (where $\widehat{\theta}_s = 1-(1-\widehat{p}_s)^\tau$). Note that as long as τ is large, $\psi_s \approx \eta_s$.

4.4.4.2 Standard error

We may replace w_s by its expectation $\widehat{w}_s$, $\psi_s\theta_s$, in $\widetilde{B}(\alpha,\beta) = \partial Q(\alpha,\beta)/\partial\beta^T$ (Equation (4.18)) and in $I(\alpha,\beta) = \partial Q(\alpha,\beta)/\partial\alpha^T$ (Equation (4.18)). Then an approximate estimated variance is

$$\widehat{\mathrm{Var}}(\breve{\psi}_s) = \{\breve{\psi}_s(1-\breve{\psi}_s)\}^2 x_s^T \widehat{\mathrm{Var}}\{\widehat{\alpha}(\widehat{\beta})\}x_s. \tag{4.24}$$

4.4.4.3 Simulations

To compare the iterative and ratio methods of estimation for occupancy, ψ, for the site inhomogeneous case we conducted a simulation study. We examine the consistency of the estimators and the standard error (s.e.) formulation for the iterative method.

We begin with a study for a small number of sites to evaluate and compare the variation of the estimates resulting from both iterative and ratio methods (Section 4.4.4.4). We illustrate that estimates for the ratio method are considerably more variable than those produced from the iterative method.

Then, in Section 4.4.4.5 we conduct a simulation study for a large number of sites to examine the two methods of estimating ψ for positive, negative, and zero correlation between ψ and p according to the model covariates. Results are reported in Tables 4.10, 4.11 and 4.12. Positive correlation results from coefficients with the same sign in models. Then, a negative correlation is achieved with opposite signs for coefficients in models. We also consider dependent and independent models for ψ and p. Dependent models occur when ψ and p are functions of the same covariates, conversely, independent models are obtained when ψ and p are functions of dissimilar covariates. In addition, for these situations we examine methods for a high, moderate and low level of detectability.

Section 4.4.4.6 explores the estimators for a variety of S and τ. Simulation results, reported in Table 4.13, were unstable for small S and τ and in many of these cases the estimates either did not exist or were clearly inadequate i.e. strongly biased. This was related to small values of some $\widehat{p}_s$, parameters not being able to be estimated and some $\widehat{p}_s$ being estimated as one or zero. In the reported simulations for $S = 30$, $S = 50$ and $\tau = 4$ we only reported cases where $min(\widehat{p}_s) > 0.1$, all parameters could be estimated and none of $\widehat{p}_s$ were 1, and continued simulations until we had 1000 estimates. We note that in

practice, one would interactively change the model to hopefully obtain sensible estimates but this is not feasible in a large number of simulations. With $S = 30$ and $\tau = 4$ there were 400 failed simulations (out of 1400 conducted) and with $S = 30$ and $\tau = 4$ there were 237 failed simulations (out of 1237 conducted). We note some slight bias when $S = 30$ that may be due to the failed estimates.

4.4.4.4 Simulations: small number of sites

To compare the iterative and ratio estimates we generate a population for the 100 sites from a population with mean $p_s = 0.3$ and mean $\psi_s = 0.7$ and with correlation $\rho(\psi_s, p_s) = 0.95$ (Figure 4.1).

A simulated population was generated for two covariates and intercept from $N(0, 1)$ with parameters $\alpha_\psi = (-1, 0.5, 0.5)$ and $\beta_p = (1, 0.5, 0.5)$, and 1000 simulated detection matrices were produced.

Results from this simulation study show that the variability of the estimates from the ratio method are considerably higher than those produced by the iterative method (Figure 4.2 b). Results such as these discourage us from deriving standard errors for $\breve{\psi}$ from the model for $\widehat{\eta}$ particularly when this is not straightforward. This is particularly so when coefficient estimates and standard errors from the GLM software are not readily useable.

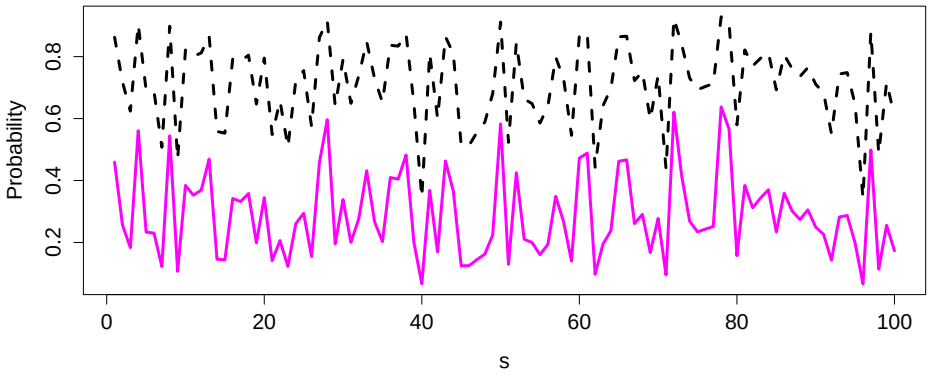
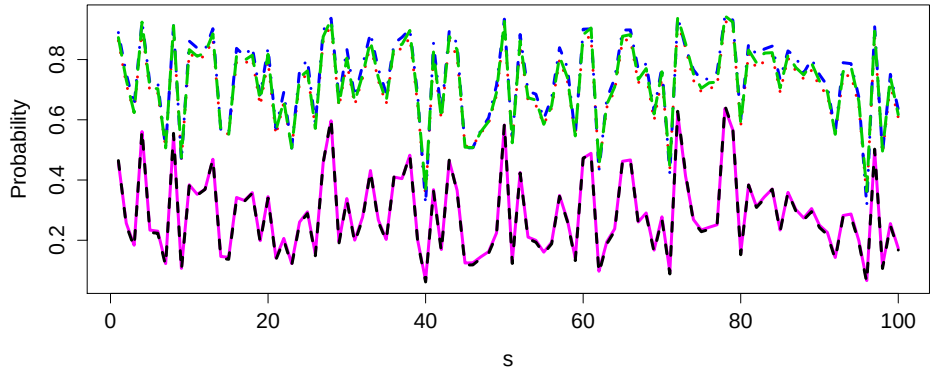
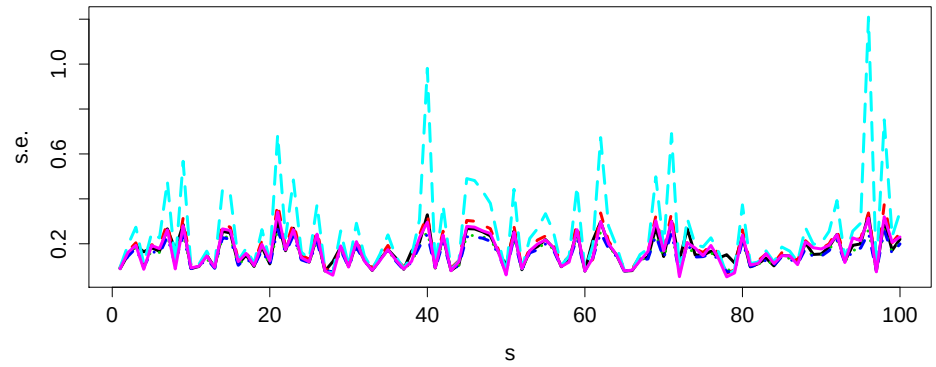


Figure 4.1: Values of p_s and ψ_s used in simulations for $S = 100$ and $\tau = 4$, $\rho = 0.946$. The solid pink line shows p_s and the dotted black line shows ψ_s.



a)



b)

Figure 4.2: Results for the iterative and ratio methods of estimation for $S = 100, \tau = 4$, $\rho = 0.946$ for 1000 simulations. a) median values for $\widehat{\psi}_s$ (iterative method) are blue; green for the ratio estimator $\breve{\psi}$ and red for the actual ψ_s. Median $\widehat{p}_s$ is pink and the actual p_s in black. And b) standard deviation (s.d.) for $\breve{\psi}_s$ (light blue), and median absolute deviation (MAD) for $\breve{\psi}_s$ (magenta), as well as s.d. $\widehat{\psi}_s$ (black), MAD $\widehat{\psi}_s$ (red), the mean (green) and medians (dark blue) of the estimated standard errors for the iterative method.

4.4.4.5 Simulations: large number of sites

The simulation study here examines the impact of varying the level of detectability. We explore, a high, moderate and low detectability for a number of models for the scenarios described above.

We examine the consistency of the iterative and ratio estimators and the standard error (s.e.) formulation for the iterative method. Results from the small sample simulation study above (Section 4.4.4.4) showed that the variability of the estimates from the ratio method were considerably higher than those produced by the iterative method (Figure 4.2 b). Given results favoured the iterative method we did not derive an expression for the standard error for ψ arising from the ratio method. Here we illustrate that even though the two methods give similar results for high detectability, the iterative method is superior to the ratio method for moderate or low detectability regardless of the number of sites sampled e.g. $S = 1000$. In this case the number of visits is small i.e. $\tau = 4$.

For simplicity we only considered an intercept term and up to two covariates,

$$p_s = \beta_0 + \beta_1 x_1 + \beta_2 x_2, \quad \psi_s = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2, \quad s = 1, \ldots, S.$$

We examined independent and dependent models for high positive, high negative and zero correlation (ρ). We conducted 1000 simulated experiments at each combination. Results are shown separately for high, moderate and low detectability, in Tables 4.10, 4.11 and 4.12, respectively. For all tables, simulations are conducted for a moderate level $\psi_s = 0.5$. In each table, four models are considered, and summary statistics for occupancy coefficient estimates displayed. The means for the simulated parameters are shown, i.e. $\bar{\psi}_s$, $\bar{p}_s$, as well as the actual correlation ρ.

Although the program was set to run 1000 simulations for each case, not all simulations were able to produce results, thus for some cases the number of simulations is < 1000.

To evaluate simulations we compute a number of summary statistics. These are shown in detail for the iterative method. For the ratio estimator the root mean square is shown.

The average of the root mean square (RMS). That is, RMS is the average of $\sqrt{S^{-1}\sum_{s=1}^{S}(\widehat{\psi}_s - \psi_s)^2}$ over the simulations, for the iterative method (RMS

iter.) and the ratio method (RMS ratio) of estimation. Similarly we computed RMS s.e. as the average of $\sqrt{S^{-1}\sum_{s=1}^{S}(\widehat{s.e.}_s - sd_s)^2}$ where sd_s is the median of the estimates of ψ_s and $\widehat{s.e.}_s$ is the estimated standard error at s. Note that there is no RMS s.e. to report for the ratio estimator $\bar{\psi}$ since there is no analytic standard error for it.

Results here verify that the iterative method is more stable than the ratio method. It is more likely to converge to an estimate whereas the ratio method will not, in many cases, and that the ratio method tends to overestimate occupancy. Moreover, we could find no reliable estimates of the standard errors for the ratio occupancy estimator.

	Intercept	x_1	x_2	Intercept	x_1	x_2
Actual ψ	0.000	0.000	0.500	0.000	0.500	0.000
p	1.000	0.500	0.000	1.000	-0.500	0.000
$\bar{\psi}_s$, $\bar{p}_s$, ρ	0.499	0.713	0.047	0.494	0.726	-0.990
	Intercept	α_1	α_2	Intercept	α_1	α_2
Median iterative	-0.001	0.002	0.498	-0.002	0.501	-0.001
MAD iterative	0.064	0.073	0.068	0.067	0.071	0.066
Med iterative s.e.	0.066	0.069	0.070	0.066	0.073	0.066
RMS: iter., s.e., ratio	0.025	0.001	0.026	0.024	0.002	0.026
	Intercept	x_1	x_2	Intercept	x_1	x_2
Actual ψ	0.000	0.500	0.000	0.000	0.250	0.250
p	1.000	0.500	0.000	1.000	0.250	0.250
$\bar{\psi}_s$, $\bar{p}_s$, ρ	0.500	0.720	0.991	0.502	0.727	0.995
	Intercept	α_1	α_2	Intercept	α_1	α_2
Median iterative	0.001	0.503	0.004	0.002	0.250	0.252
MAD iterative	0.064	0.071	0.066	0.067	0.067	0.066
Med iterative s.e.	0.066	0.072	0.067	0.065	0.067	0.064
RMS: iter., s.e., ratio	0.024	0.001	0.025	0.025	0.001	0.025

Table 4.10: $S = 1000, \tau = 4$ for 1000 simulations for high detectability ($p \approx 0.7$). Comparisons of the iterative and ratio method of estimation for ψ, including the standard error (s.e.) of the RMS for the iterative estimates.

	Intercept	x_1	x_2	Intercept	x_1	x_2
Actual ψ	0.000	0.000	0.500	0.000	0.500	0.000
p	0.000	0.500	0.000	0.000	-0.500	0.000
$\bar{\psi}_s$, $\bar{p}_s$, ρ	0.504	0.501	-0.015	0.510	0.491	-1.000
	Intercept	α_1	α_2	Intercept	α_1	α_2
Median iterative	0.003	-0.006	0.500	-0.001	0.504	0.003
MAD iterative	0.078	0.079	0.083	0.074	0.101	0.078
Med iterative s.e.	0.077	0.078	0.080	0.078	0.090	0.074
RMS: iter., s.e., ratio	0.029	0.002	0.039	0.030	0.002	0.061
	Intercept	x_1	x_2	Intercept	x_1	x_2
Actual ψ	0.000	0.500	0.000	0.000	0.250	0.250
p	0.000	0.500	0.000	0.000	0.250	0.250
$\bar{\psi}_s$, $\bar{p}_s$, ρ	0.506	0.506	1.000	0.503	0.503	1.000
	Intercept	α_1	α_2	Intercept	α_1	α_2
Median iterative	0.006	0.503	-0.004	0.009	0.249	0.255
MAD iterative	0.075	0.082	0.069	0.075	0.076	0.074
Med iterative s.e.	0.074	0.082	0.070	0.072	0.074	0.073
RMS: iter., s.e., ratio	0.027	0.002	0.029	0.028	0.002	0.029

Table 4.11: $S = 1000, \tau = 4$ for 1000 simulations for moderate detectability ($p \approx 0.5$). Comparisons of the iterative (iter.) and ratio method of estimation for ψ, including the standard error (s.e.) of the RMS for the iterative estimates.

	Intercept	x_1	x_2	Intercept	x_1	x_2
Actual ψ	0.000	0.000	0.500	0.000	0.500	0.000
p	-1.000	0.500	0.000	-1.000	-0.500	0.000
$\bar{\psi}_s$, $\bar{p}_s$, ρ	0.500	0.279	0.053	0.501	0.279	-0.990
	Intercept	α_1	α_2	Intercept	α_1	α_2
Median iterative	0.014	0.001	0.503	0.013	0.505	-0.000
MAD iterative	0.142	0.129	0.128	0.150	0.160	0.123
Med iterative s.e.	0.134	0.126	0.125	0.142	0.141	0.117
RMS: iter., s.e., ratio	0.047	0.006	0.060	0.051	0.007	0.088
	Intercept	x_1	x_2	Intercept	x_1	x_2
Actual ψ	0.000	0.500	0.000	0.000	0.250	0.250
p	-1.000	0.500	0.000	-1.000	0.250	0.250
$\bar{\psi}_s$, $\bar{p}_s$, ρ	0.502	0.281	0.990	0.501	0.275	0.995
	Intercept	α_1	α_2	Intercept	α_1	α_2
Median iterative	0.010	0.508	-0.007	0.020	0.254	0.243
MAD iterative	0.122	0.120	0.111	0.126	0.109	0.117
Med iterative s.e.	0.128	0.122	0.108	0.123	0.112	0.119
RMS: iter., s.e., ratio	0.043	0.006	0.046	0.045	0.006	0.046

Table 4.12: $S = 1000, \tau = 4$ for 1000 simulations for low detectability ($p \approx 0.3$). Comparisons of the iterative (iter.) and ratio method of estimation for ψ, including the standard error (s.e.) of the RMS for the iterative estimates.

4.4.4.6 Varying number of sites and visits

We ran 1000 simulations for each combination for a moderate level of detection and occupancy. And, we ran these for independent models for ψ and p, with starting parameters $\alpha_\psi = (0, 0, 0.5)$ and $\beta_p = (0, 0.5, 0)$. The RMS for the iterative and ratio methods are shown in Table 4.13.

Table 4.14 shows the number of estimates greater than 1 returned by the ratio estimator. All estimates returned by the iterative method were constrained to between 0 and 1. Simulations were run until 1000 valid estimates were obtained. The simulations for $S \leqslant 30$ and $S = 50$ for a few visitations, $\tau = 4$, returned a failure rate of about 90% and 20%, respectively. We found similar failure rates for non-convergence in Karavarsamis et al. (2013) (Chapter 3).

S	RMS	$\tau = 4$	$\tau = 8$	$\tau = 16$
30	iterative	0.178	0.148	0.143
	s.e.	0.140	0.033	0.032
	ratio	0.183	0.154	0.143
50	iterative	0.145	0.109	0.112
	s.e.	0.056	0.015	0.017
	ratio	0.185	0.110	0.116
150	iterative	0.076	0.066	0.062
	s.e.	0.011	0.008	0.006
	ratio	0.085	0.068	0.062
400	iterative	0.047	0.038	0.038
	s.e.	0.005	0.003	0.003
	ratio	0.059	0.041	0.038
1000	iterative	0.028	0.025	0.024
	s.e.	0.002	0.001	0.001
	ratio	0.037	0.027	0.024

Table 4.13: Root mean square (RMS) comparing the iterative and ratio methods of estimation of ψ, including the standard error (s.e.) of the RMS for the iterative estimates.

S	$\tau = 4$	$\tau = 8$	$\tau = 16$
30	790	219	5
50	1537	8	180
150	661	42	0
400	1585	260	0
1000	687	16	0

Table 4.14: Number of $\widehat{\psi} > 1$ from the ratio method. Simulations were run until 1000 valid estimates for ψ were found.

4.4.5 Estimating occupancy using GAMs

The MGCV R package readily allows us to fit generalized linear additive models (GAMs) through a penalised likelihood. Development of our partial likelihood approach and the offset method of estimation developed in Section 4.4.3.2 allows a straightforward implementation of GAMs methods to model occupancy data. As it is a minor change to the GLM formulae, we briefly outline the approach to demonstrate the derivation of the associated standard error formulae.

Let u_s be the covariates associated with detectability whose effects will be modelled parametrically and $v_{s1}, \ldots, v_{sK}$ those that will be modelled nonparametrically. Then the GAM for the linear predictor κ_s is

$$\kappa_s = u_s^T \alpha + f_1(v_{s1}) + \cdots + f_K(v_{sK}).$$

Similarly, for occupancy let x_s be the parametric covariates and $r_{s1}, \ldots, r_{sJ}$ the nonparametric components. Then the GAM for the linear predictor is

$$\nu_s = x_s^T \alpha + g_1(rv_{s1}) + \cdots + g_J(r_{sJ}).$$

These are fitted through penalised partial likelihoods as follows in two stages, where in the second stage we use the iterative method to obtain estimates for occupancy.

1. Fit a GAM to the redetection data. This yields estimates $\widehat{p}_s$ of p_s for all sites as functions of u_s and $v_{s1}, \ldots, v_{sK}$. In particular we may compute $\widehat{\theta}_s$ for each s.

2. Fit a GAMS to the indicators of presence w_s with an offset computed as in Section 4.4.3.2 using the iterative procedure described there. That is at the ith step the offset is

   $$a_s^{(i)} = \log(\theta_s) - \log\{1 + \exp(\nu_s^{(i-1)})(1 - \theta_s)\}.$$

   where $\nu_s^{(i-1)}$ is the predictor arising at the $(i-1)$th step.

We use penalised cubic B-splines coupled with the GCV criterion to find the optimal smoothing parameter for the spline (de Boor, 1978; Wood, 2006, p. 177). The GAM penalised likelihood is maximised, and coefficients estimated,

by penalised iteratively re-weighted least squares (P-IRLS) (Wood, 2006, p. 169).

4.4.5.1 Standard errors

We next outline the derivation of the standard error.

Firstly, we may retrieve an estimate V_β of the variance matrix of the parameters in the model for the detection probabilities from the output of the `gam` command. Write $\alpha = (\alpha_1^T, \alpha_2^T)^T$ where α_1 are the parameters in the parametric component of the model and α_2 those in the nonparametric component.

For a given λ_S and penalty matrix $\mathcal{P}$, the penalised partial log-likelihood is

$$\ell_\lambda(\alpha, \beta) = \sum_{s=1}^{S} \{(1 - w_s) \log(1 - \psi_s(\alpha)\theta_s) + w_s \log(\psi_s(\alpha))\} - \frac{1}{2}\lambda_S \alpha_2^T \mathcal{P} \alpha_2$$

so that the penalised partial score function is

$$Q_\lambda(\alpha, \beta) = \frac{\partial \ell(\alpha, \beta)}{\partial \alpha^T} = \sum_{s=1}^{S} x_s \frac{\{w_s - \psi_s(\alpha)\theta_s\}(1 - \psi_s(\alpha))}{1 - \psi_s(\alpha)\theta_s} - \lambda_S \mathcal{P}^* \alpha_2.$$

Here $\mathcal{P}^*$ is a penalty matrix that penalises parameters in the nonparametric component of the model. It has zero columns initially corresponding to α_1. Let $\widehat{\alpha}_\lambda(\beta)$ be the solution of $Q_\lambda(\alpha, \beta) = 0$. Then

$$0 = Q_\lambda(\widehat{\alpha}_\lambda(\widehat{\beta}), \widehat{\beta}) \approx Q_\lambda(\alpha, \widehat{\beta}) + I_\lambda(\alpha, \widehat{\beta}) \left(\widehat{\alpha}_\lambda(\widehat{\beta}) - \alpha\right)$$

where

$$I_\lambda(\alpha, \widehat{\beta}) = I(\alpha, \widehat{\beta}) + \lambda_S \mathcal{P}^*.$$

A minor modification of the derivations leading to the variance expression for the iterative method (4.19) yields (4.25). Specifically, a simple substitution of $I(\alpha, \beta)$ with $I_\lambda(\alpha, \beta)$ will yield (4.25).

Following the derivations leading to (4.19) we see that

$$\begin{aligned}\widehat{\mathrm{Var}}(\widehat{\alpha}_\lambda(\widehat{\beta})) \approx\ & I_\lambda\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}\}^{-1} I\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}\} I_\lambda\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}\}^{-1} \\ & + I_\lambda\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}\}^{-1} \widetilde{B}\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta})\} \widehat{V}_\beta \widetilde{B}\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}\}^T I_\lambda\{\widehat{\alpha}(\widehat{\beta}), \widehat{\beta}\}^{-1}.\end{aligned} \tag{4.25}$$

Again the estimated occupancy probability is $\widehat{\psi}_s^{(\lambda)} = [1 + \exp\{-x_s^T \widehat{\alpha}_\lambda(\widehat{\beta}\,)\}]^{-1}$, which has estimated variance

$$\widehat{\mathrm{Var}}(\widehat{\psi}_s^{(\lambda)}) = \{\widehat{\psi}_s(1-\widehat{\psi}_s)\}^2 x_s^T \widehat{\mathrm{Var}}\{\widehat{\alpha}_\lambda(\widehat{\beta}\,)\} x_s.$$

Remark For our theory to apply we need the penalty matrix used in the GAM fit. These can be extracted from the R `gam` object.

4.4.5.2 Simulations

To examine the consistency of the estimators and the above standard error formulation we again conducted a simulation study. For simplicity we only considered an intercept term and a single covariate. We considered a single covariate s corresponding to the index of the site and took

$$p_s = \beta_0 + \beta_1 \sin(8s/S), \quad \psi_s = \alpha_0 + \alpha_1 \sin(8s/S), \quad s = 1, \ldots, S$$

For $S = 1000$, $\tau = 4$, $\beta_0 = \beta_1 = 0.5$ and $\alpha_0 = 0$, $\alpha_1 = 0.5$ we plot the values of p_s and ψ_s in Figure 4.3. These high values of detectability and occupancy are perhaps not that unusual in practice and allow us to consider relatively small S without too many complications. We conducted 1000 simulated experiments at each combination. For the first simulations we plot the actual values and the medians of the estimates arising from the simulations (Figure 4.3) as well as their standard deviation (s.d.) and median absolute deviation (MAD) and the mean (av s.e.) and median (med s.e.) of the estimates of the standard error estimator (Figure 4.4). This gives us confidence that our asymptotic results are reasonable.

S	τ	RMS	s.e. RMS
1000	4	0.037	0.005
1000	6	0.036	0.005
1000	20	0.037	0.005
500	4	0.051	0.009
500	6	0.051	0.008
500	20	0.051	0.008
400	4	0.057	0.008
400	6	0.056	0.008
400	20	0.056	0.010
100	4	0.107	0.028
100	6	0.103	0.026
100	20	0.108	0.028
50	4	0.511	0.006
50	6	0.141	0.039
50	20	0.144	0.044

Table 4.15: Average root mean square, of the estimates and their standard errors. Simulations (1000 per case) for the GAM model with a single covariate.

We considered $S = 1000, 500, 400, 100, 50$ and $\tau = 4, 6, 20$ and conducted 1000

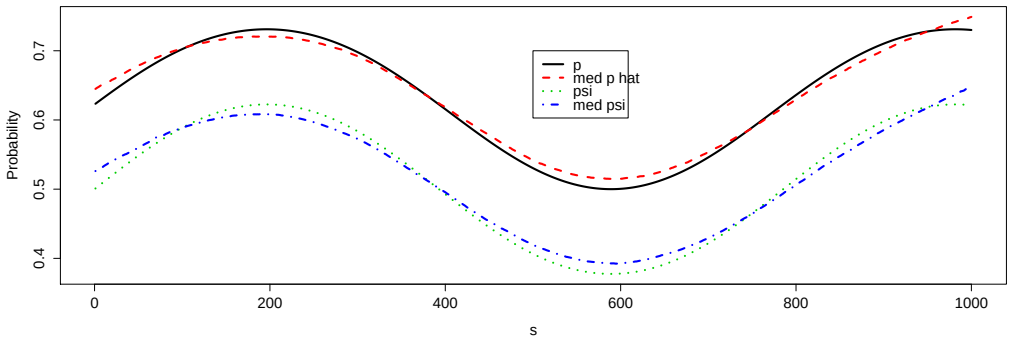


Figure 4.3: Median values of p_s and ψ_s used in the first set of simulations: $S = 1000$, $\tau = 4$, $\beta_0 = \beta_1 = 0.5$ and $\alpha_0 = 0$, $\alpha_1 = 0.5$.

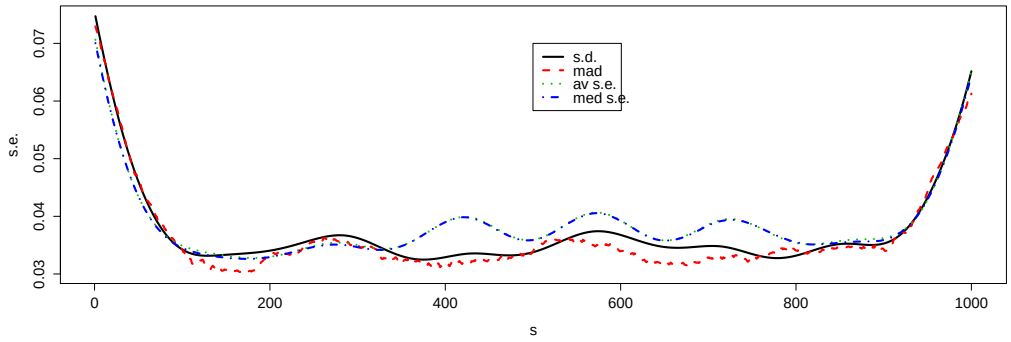


Figure 4.4: Standard errors for values of p_s and ψ_s used in the first set of simulations: $S = 1000$, $\tau = 4$, $\beta_0 = \beta_1 = 0.5$ and $\alpha_0 = 0$, $\alpha_1 = 0.5$.

simulated experiments at each combination; results are summarised in Table 4.15.

Our simulations show that for a small number of sites and low detection probabilities, the standard error estimator of the estimated occupancies do not perform well.

To fit GAMs requires a substantial amount of data. The practical examples considered in this thesis were not suitable for GAMs. Practical applications are deferred for future work.

4.5 Time dependent covariates

In some instances it is possible that there are time dependent covariates that we may then relate to detection, for example in the brook trout data described in Section 4.3.6. The models considered here suppose a constant occupancy but it is straightforward to allow the detection probabilities to vary with covariates.

Let y_{sj} take the value 1 if an individual was detected at site s on occasion j and zero otherwise. Let p_{sj} be the probability of detection at site s on occasion j if site s is occupied. When we allow the detection probabilities to depend on time dependent covariates, the partial likelihood is now

$$\prod_{j=a_s+1}^{\tau} p_{sj}^{y_{sj}}(1-p_{sj})^{1-y_{sj}},$$

where $p_{sj} = h(u_{sj}^T\beta)$ and h is for example the logistic function $h(u) = (1+\exp(-u))^{-1}$ for a vector of covariates $u_{sj} \in \mathbb{R}^{q_p}$ and a vector of parameters $\beta \in \mathbb{R}^{q_p}$. That is, with time dependent covariates we see that the redetections form a sequence of independent Bernoulli trials. To do this, we construct a 0-1 response variable for each occasion at which redetections were possible and further construct a matrix of covariates.

Let $\widehat{\beta}$ again denote the estimator of β and let $\widehat{V}_\beta$ be its estimated variance matrix. Now we let

$$\widetilde{B}(\alpha,\beta) = -\sum_{s=1}^{S} \frac{x_s\psi_s(1-\psi_s)(1-w_s)(1-\theta_s)\sum_{j=1}^{\tau} p_{sj}u_{sj}^T}{(1-\psi_s\theta_s)^2}$$

and with these changes, the expression (4.19) for the variance of $\widehat{\alpha}(\widehat{\beta})$ still holds.

Recall $\widetilde{B}(\alpha,\beta) = \dfrac{\partial Q(\alpha,\beta)}{\partial\beta^T} = \dfrac{\partial Q(\alpha,\beta)}{\partial\theta_s}\dfrac{\partial\theta_s}{\partial\beta^T}$. In this case $\theta_s = 1-\prod_{j=1}^{\tau}(1-p_{sj})$ and hence for the logistic model,

$$\begin{aligned}\frac{\partial\theta_s}{\partial\beta^T} &= \sum_{j=1}^{\tau}\prod_{k\neq j}(1-p_{sk})\frac{\partial p_{sj}}{\partial\beta^T}\\ &= \sum_{j=1}^{\tau}\prod_{k=1}^{\tau}(1-p_{sk})p_{sj}u_{sj} \qquad (4.26)\\ &= (1-\theta_s)\sum_{j=1}^{\tau}p_{sj}u_{sj}.\end{aligned}$$

Now $\partial Q(\alpha,\beta)/\partial\theta_s$ is unchanged from (4.17) and with (4.26) this yields the

version of the expression (4.19) for the standard errors given in Section 4.5. Note that if $p_{sj} \equiv p_s$ and $u_{sj} \equiv u_s$ then $\sum_{j=1}^{\tau} p_{sj} u_{sj}^T = \tau p_s u_s$ and we recover our previous expression. Also as before we can replace w_s by $\psi_s \theta_s$. In preliminary simulations of the time dependent case, we observed that the estimated standard errors computed in this fashion were more stable so we used these in the application and simulations below.

Remark 1 To fit separate effects for each occasion, including the first, we need to use a conditional likelihood rather than the partial likelihood applied to the redetections. By only considering redetection, the partial likelihood is unable to estimate the probability of detection on the first occasion. This is of course possible but for simplicity we defer a consideration of this. However, as noted before, this only requires a minor change to the variance formula in that $\widehat{V}_\beta$ is different.

Remark 2 Time dependent occupancy where occupancy can vary from occasion to occasion does not permit use of partial and conditional likelihoods.

4.5.1 Application

Fitting a GLM to the redetections of the brook trout with Elevation and stream mean cross-sectional area (CSA) yielded Table 4.16. Thus, neither covariate was significant in a linear model.

In Table 4.16 we allow the detection probabilities to depend on both elevation and the time varying covariate CSA and the occupancy probabilities depend on elevation. Whilst none of the covariates were significantly related to detection, we retained them in the model to estimate the occupancies. The results in Table 4.16 relating to occupancy are similar to those in Table 4.9 b).

	Estimate	Std. Error	z-value	p-value
Detectability				
Intercept	0.2007	1.4648	0.137	0.891
Elevation	0.0003	0.0004	0.690	0.490
CSA	-0.6657	0.4252	-1.566	0.117
Occupancy				
Intercept	-3.9413	1.2374	-3.1867	0.0014
Elevation	0.0013	0.0004	3.3231	0.0009

Table 4.16: GLM fitting time dependent covariates to the detection probabilities of the brook trout and the resulting estimates of occupancy.

4.5.2 Simulations

Again to verify the method we conducted a small simulation study that mimics the brook trout data. Results are shown in Table 4.17. We took $\beta = 0.2, 0.0003, -0.66$, $\alpha = -4.0, 0.0013$, simulated elevation from a normal distribution with $\mu_e = 2860$ and $\sigma_e = 1140$, and CSA from a normal distribution with $\mu_c = 1.6$ and $\sigma_c = 0.5$. We only report the results for α as the detection probabilities are once again fitted using a standard GLM. We considered $\tau = 4$ and $S = 1000, 400, 150$ and 50. These results show there is little evident bias and the estimated standard errors are reasonable.

	$S = 1000$	$\tau = 4$	$S = 400$	$\tau = 4$
α	-4.0000	0.0013	-4.0000	0.0013
Median $\hat{\alpha}$	-4.0031	0.0013	-3.9985	0.0013
MAD $\hat{\alpha}$	0.3174	0.0001	0.5490	0.0002
Median s.e. $\hat{\alpha}$	0.3120	0.0001	0.4998	0.0002
	$S = 150$	$\tau = 4$	$S = 50$	$\tau = 4$
α	-4.0000	0.0013	-4.0000	0.0013
Median $\hat{\alpha}$	-4.0721	0.0013	-4.1052	0.0013
MAD $\hat{\alpha}$	0.8001	0.0003	1.5028	0.0005
Median s.e. $\hat{\alpha}$	0.8204	0.0003	1.4641	0.0004

Table 4.17: Simulation results for the model for the occupation probabilities with time dependent detection probabilities.

4.6 The basic conditional likelihood

As mentioned earlier, yet another alternative approach to estimating the detection probabilities is to condition on at least one detection at a site rather than the first detection. Here we restrict the discussion to an introduction of the conditional likelihood approach as its implementation with any software (e.g. VGAM) is more complicated than the software available for GLMs and GAMs.

From the full likelihood given for site inhomogeneity (4.10), we may also write

$$L_s(\psi_s, p_s) = (1 - \psi_s\theta_s)^{z_s}(\psi_s\theta_s)^{1-z_s}\left(\frac{\binom{\tau}{y_s}p_s^{y_s}(1-p_s)^{\tau-y_s}}{\theta_s}\right)^{1-z_s}, \qquad (4.27)$$

the conditional likelihood given at least one detection at a sight which is based on the second term in this formulation i.e. $\psi_s\theta_s$. This conditional likelihood constitutes the positive binomial distribution and can be modelled in VGAM which offers the positive binomial link family.

The two components of the likelihood to be considered at each stage of estimation separately for ψ_s and p_s are, for detection

$$L_{1s}(p_s) = \left(\frac{\binom{\tau}{y_s}p_s^{y_s}(1-p_s)^{\tau-y_s}}{\theta_s}\right)^{1-z_s} \qquad (4.28)$$

and at the second stage, the likelihood for occupancy is

$$L_{2s}(\psi_s, \theta_s) = (1 - \psi_s\theta_s)^{z_s}(\psi_s\theta_s)^{1-z_s}. \qquad (4.29)$$

With this formulation we use posbinomial family in VGAM to model the number of detections for detected sites i.e. for sites with at least one detection, to obtain estimates for detection p_s. Then use the presence indicator $w_s = 1 - z_s$ to estimate occupancy according to the iterative method for GLM or GAM, for example. We have reported some results using this formulation in the above sections and will not consider this further in this thesis.

Alternatively, for estimating detection we could use a formulation from previous sections which considers redetections. In this way we ignore the contribution to the likelihood related to first detections, which is

$$L_{1s}(p_s) = \left\{ \frac{p_s(1-p_s)^{(a_s-1)}}{\theta_s} \right\}^{(1-z_s)},$$

where this is the probability of being detected for the first time at a_s given detected at least once. That is, our two-stage process using the partial likelihood ignores the information on the detection probabilities in the first detections of each detected site (site with at least one detection).

4.7 Crossbill application

Here we compare our iterative method to the direct maximisation of the full likelihood fit with the `occu` function of `unmarked` package.

We apply the two-stage iterative GLM methods to data on the European crossbill (Loxia curvirostra) collected in 267 1-km^2 sample quadrats in Switzerland, 1999 (Schmidt, 2004). We analyse the data from the example presented in Fiske and Chandler (2011) that involves the site covariates elevation and forest.

We fit two models for the crossbill data: a model with linear effects for elevation and forest (Table 4.18) and another that includes a squared term for elevation (Table 4.19).

Results given here comparing the two methods are consistent with simulation results given in Section 4.4.3.4, in that there is less agreement in estimates between the two methods as more terms are added to the model. See for example Table 4.6 compared to Table 4.7. In the former table, for the null model (i.e. no covariates) estimates between the iterative and `occu` are similar. In contrast, in Table 4.7 for a model that includes two linear covariates the two methods give estimates which differ more than the null model. Similarly, in the crossbill application here, estimates for the two methods agree more in the simpler model (Table 4.18) than the complicated model that includes the squared term (Table 4.19).

There is not much difference in the results between the two methods for the simple model in Table 4.18, if we consider that standard errors from `occu` are known to be an underestimate, and that there is some loss of efficiency with the iterative method in estimating detectability, $\widehat{\beta}$. Observations here are verified from simulations in Tables 4.6, 4.7 and 4.8.

Detections from the first year of the study, in 1999, were used in the analysis, $\tau = 3$. To make comparison of the two methods possible, sites with any missing detections were removed prior to fitting models, $S = 201$.

In the example presented in Fiske and Chandler (2011) the covariates are standardised to make it possible to find the maximum likelihood estimates with the direct maximisation of the full likelihood of MacKenzie et al. (2002). It is recommended that covariates are standardised to help stabilise the optimisation algorithm used in `unmarked` (see p. 2 of the R vignette Fiske and Chandler

(2014)[8]). In our applications and simulations with `unmarked` we observed this to be true.

Whilst the stability of our iterative approach is unaffected by standardisation we also use the standardised covariates. We encountered difficulties when the R `predict` procedure was used on `unmarked` data with standardised covariates. This is a known problem with the methods[9]. Recall though that it is recommended that `unmarked` covariates are standardised to help stabilise the maximisation process.

More complex models such as GAMs could not be fit successfully to this data because there are an insufficient number of redetections for reasonable estimation.

		$\widehat{\beta}$	s.e. $\widehat{\beta}$	$t_{\widehat{\beta}}$	$\widehat{\alpha}$	s.e. $\widehat{\alpha}$	$t_{\widehat{\alpha}}$
Iterative	Intercept	-1.448	0.477	-3.038	0.329	0.787	0.418
	Elevation	1.490	0.399	3.733	0.240	0.559	0.428
	Forest	0.102	0.337	0.302	0.050	0.401	0.124
`unmarked`	Intercept	-1.545	0.415	-3.727	0.272	0.680	0.400
	Elevation	0.964	0.413	2.335	0.614	0.724	0.848
	Forest	0.223	0.332	0.671	0.011	0.569	0.020

Table 4.18: Iterative GLM and `unmarked` with linear effects for the crossbill data, $S = 201$ and $\tau = 3$.

		$\widehat{\beta}$	s.e. $\widehat{\beta}$	$t_{\widehat{\beta}}$	$\widehat{\alpha}$	s.e. $\widehat{\alpha}$	$t_{\widehat{\alpha}}$
Iterative	Intercept	-1.468	0.514	-2.853	1.504	1.164	1.292
	Elevation	2.301	1.184	1.943	-3.296	2.203	-1.497
	Elevation^2	-0.486	0.635	-0.765	1.652	0.961	1.718
	Forest	0.076	0.343	0.222	0.087	0.403	0.215
`unmarked`	Intercept	-1.890	0.240	-7.870	3.132	1.739	1.801
	Elevation	0.292	0.483	0.604	3.346	1.281	2.612
	Elevation^2	0.450	0.269	1.672	-2.147	0.927	-2.316
	Forest	0.123	0.176	0.702	0.574	0.694	0.826

Table 4.19: Iterative GLM and `unmarked` with linear effects including a squared term for the crossbill data, $S = 201$ and $\tau = 3$.

In the next section we explore and compare these methods with simulations.

[8] cran.at.r-project.org/web/packages/unmarked/vignettes/unmarked.pdf
[9] https://sites.google.com/site/unmarkedinfo/home/known-bugs

4.7.1 Simulations

To evaluate the use of our method on the crossbill data we ran 1000 simulations that mimic this example. In all models we simulated samples for $S = 250$ and $\tau = 3$. Simulations for a selection of scenarios compare the iterative two-stage partial and conditional likelihood methods, and the direct maximisation of the full likelihood using the `occu` procedure. The conditional likelihood was fitted using the `vglm` function of the VGAM procedure under the positive binomial link function.

Table 4.20 shows simulation results for a model similar to the crossbill application that includes an intercept x_0 and two terms x_1, x_2, each generated once from $N(0, 1)$, plus a squared term x_1^2. Then 1000 simulated detection experiments were generated for $\beta_p = (1, 1, -0.5, 0.5)$ and $\alpha_\psi = (1, 0.5, 0.5, 0.45)$. In all simulations, mean $p = 0.800$ and median $p = 0.787$. And, mean $\psi = 0.624$, median $\psi = 0.694$ and correlation between p and ψ was $\rho = 0.286$. The partial and conditional likelihoods gave similar results which means that ignoring first detections in the partial likelihood did not cause any great loss of efficiency. Next, estimates produced by our methods are not substantially different from `unmarked` and our methods offer full use of GLM and VGAM functionalities. These results verify our methods even though estimates varied between methods when this model was applied to the crossbill data (results shown in Table 4.19).

To explore differences in estimates produced by our methods and `unmarked` for the crossbill data we ran 1000 simulations for the same model where x_1 and x_2 are generated once from $N(0, 1)$. Estimates from the crossbill data (Table 4.19) obtained for the model including the square for elevation were used here as starting values for simulations. Specifically, we simulated from a population with $\beta_p = (-1.89, 0.292, 0.45, 0.123)$ and $\alpha_\psi = (3.132, 3.346, -2.147, 0.574)$ given by `unmarked` from the crossbill data. In all simulations, mean $p = 0.212$ and median $p = 0.157$. And, mean $\psi = 0.680$, median $\psi = 0.907$ and correlation between p and ψ was $\rho = -0.216$. Results are shown in Table 4.21. Estimates produced by the partial and conditional likelihood approaches are considerably more accurate for occupancy compared to `unmarked`. All three methods give similar estimates for detectability.

We used the covariates from the crossbill data then simulated 1000 detection experiments based on starting values $\beta_p = (-1.89, 0.292, 0.45, 0.123)$ and $\alpha_\psi = (3.132, 3.346, -2.147, 0.574)$ obtained from `unmarked` on the crossbill

data. Results are shown in Table 4.22. For simulated covariates generated from crossbill data, mean $p = 0.210$ and median $p = 0.156$. And, mean $\psi = 0.881$, median $\psi = 0.881$ and correlation between p and ψ was $\rho = 0.300$.

Results here verify findings from the previous section of the crossbill application.

	ψ				p			
	x_0	x_1	x_1^2	x_2	x_0	x_1	x_1^2	x_2
Actual	1.000	1.000	-0.500	0.500	1.000	0.500	0.500	0.450
Partial								
Median	1.004	1.023	-0.504	0.495	1.011	0.507	0.530	0.450
MAD	0.206	0.181	0.145	0.160	0.210	0.265	0.278	0.158
Med. s.e.	0.204	0.180	0.151	0.162	0.202	0.256	0.260	0.165
Conditional								
Median	1.006	1.025	-0.505	0.496	1.000	0.508	0.512	0.450
MAD	0.206	0.184	0.146	0.158	0.174	0.200	0.227	0.135
Med. s.e.	0.203	0.180	0.150	0.161	0.164	0.205	0.209	0.134
`unmarked`								
Median	1.006	1.026	-0.505	0.497	0.998	0.510	0.511	0.450
MAD	0.206	0.183	0.147	0.161	0.173	0.197	0.220	0.134
Med. s.e.	0.203	0.180	0.150	0.161	0.164	0.205	0.209	0.133

Table 4.20: Simulations (=1000) for the Partial, Conditional and `unmarked` methods (S =250, $\tau = 3$).

	ψ				p			
	x_0	x_1	x_1^2	x_2	x_0	x_1	x_1^2	x_2
Actual	3.132	3.346	-2.147	0.574	-1.890	0.292	0.450	0.123
Partial								
Median	2.279	2.478	-1.602	0.211	-1.934	0.104	0.566	0.119
MAD	0.963	0.835	0.499	0.514	0.397	0.531	0.374	0.249
Med. s.e.	1.734	1.140	0.811	0.618	0.413	0.710	0.472	0.246
Conditional								
Median	2.260	2.494	-1.623	0.213	-1.916	0.099	0.553	0.116
MAD	0.879	0.763	0.484	0.489	0.357	0.486	0.326	0.219
Med. s.e.	1.656	1.098	0.786	0.605	0.383	0.643	0.413	0.217
`unmarked`								
Median	6.387	6.014	-4.088	0.919	-1.929	0.272	0.481	0.123
MAD	6.849	5.127	3.714	1.905	0.216	0.442	0.276	0.129
Med. s.e.	4.245	3.149	2.438	1.241	0.171	0.367	0.256	0.119

Table 4.21: Simulations (=1000) for the Partial, Conditional and `unmarked` methods, from a population with parameters from `unmarked` estimates for crossbill data (S =250, $\tau = 3$).

	ψ				p			
	Int.	Elev.	Elev.^{2}	Forest	Int.	Elev.	Elev.^{2}	Forest
Actual	3.132	3.346	-2.147	0.574	-1.890	0.292	0.450	0.123
Partial								
Median	2.075	2.512	-1.415	0.306	-1.887	0.117	0.548	0.098
MAD	0.995	0.913	0.636	0.661	0.449	0.683	0.396	0.325
Med. s.e.	1.793	1.218	0.933	0.698	0.493	0.765	0.439	0.340
Conditional								
Median	2.118	2.506	-1.453	0.279	-1.911	0.147	0.546	0.115
MAD	0.951	0.860	0.594	0.595	0.440	0.585	0.342	0.283
Med. s.e.	1.745	1.160	0.904	0.671	0.456	0.688	0.387	0.301
`unmarked`								
Median	4.601	4.749	-2.838	0.791	-1.864	0.261	0.465	0.125
MAD	4.423	3.205	2.263	1.259	0.273	0.465	0.275	0.175
Med. s.e.	2.738	2.019	1.519	0.864	0.233	0.415	0.247	0.173

Table 4.22: Simulations (=1000) from crossbill data for the Partial, Conditional and `unmarked` methods (S =201, $\tau = 3$).

4.8 Conclusions

Our approach using partial likelihood allows the use of standard GLM methods to estimate the model parameters, greatly simplifying the programming required to fit complex models. It compares well to the full likelihood in simulation studies. An advantage of our methods is that we now obtain analytic standard errors for both $\widehat{\psi}$ and $\widehat{p}$, that perform well, whereas for the full likelihood there are no closed form solutions and asymptotic standard errors tend to be an underestimate (e.g results produced by `unmarked` in Tables 4.20, 4.21 and 4.22.

We may also use conditional likelihood to estimate the detection probabilities. Note that a conditional likelihood is a special case of a partial likelihood and our theory is readily transferable. Whilst more difficult to implement, the conditional likelihood can be fitted using the VGAM package. However, VGAM is not yet as sophisticated as GAMs, so our current focus has been on the partial likelihood. Further, this is a minor extension of our methods and we do not expand on this in detail. As noted in our example, the partial likelihood can have some disadvantages — for example if there is a "learning effect" in the detection probabilities. That is, after the initial visit the investigator may find it easier to detect individuals at a given site. We defer consideration of this and our concern here has been with time homogeneous detection probabilities.

In developing our estimator we also considered a ratio estimator. This approach has the advantage that both sets of covariate coefficients, for p and ψ, may be simply estimated using standard GLM methods and in particular GAMS. Our simulations showed that these estimators tended to be biased and simple bias corrections were not sufficient. As the offset approach of Section 4.4.3.2 readily allows the use of GLM techniques we did not pursue this further.

4.9 Appendix 1: Proofs for Homogeneous case

4.9.1 Expectation for $\widehat{\psi}$

Here we show $E\left(\widehat{\psi}\,|\,b, S-f_0\right) \approx \widetilde{\psi}$. Recall the quantities from Section 4.3.3, $\widetilde{\psi} = \widehat{\psi}(a) = (S-f_0)/(S\theta)$ and Equation (4.7) $\widehat{\psi} \approx \widetilde{\psi}\left\{1 - \dfrac{\tau(1-p)^{\tau-1}(\widehat{p}-p)}{\theta}\right\} \approx \widetilde{\psi}\left\{1 - \dfrac{1}{\theta}(\widehat{\theta}-\theta)\right\}$, then

$$\begin{aligned} E\left(\widehat{\psi}\,|\,b, S-f_0\right) &= E\left\{\widetilde{\psi}\left(1-\frac{1}{\theta}(\widehat{\theta}-\theta)\right)\,|\,b, S-f_0\right\} \\ &= E\left\{\widetilde{\psi}\,|\,b, S-f_0\right\} - \frac{1}{\theta}E\left\{\widetilde{\psi}(\widehat{\theta}-\theta)\,|\,b, S-f_0\right\} \\ &= \widetilde{\psi} - \frac{1}{\theta}\widetilde{\psi}\,E\left\{(\widehat{\theta}-\theta)\,|\,b, S-f_0\right\} \\ &\approx \widetilde{\psi}, \end{aligned} \tag{4.30}$$

using the consistency of $\widehat{\theta}$, the dominant convergence theorem and asymptotic independence. We showed that the partial score equations for the detection and occupancy probabilities are uncorrelated, so ψ and p are partially orthogonal. Thus $\widehat{\theta}$, a function of redetections, is asymptotically independent of b and the number of non detected sites $S-f_0$ (see Sections 4.4.2 and 4.4.3.1 for more details). Then taking the expectation gives $E\left(\widehat{\psi}\right) \approx E\left(\widetilde{\psi}\right) \approx \psi$.

Now we find $E\left(\widetilde{\psi}^2\right)$. Then together with $\text{Var}(\widetilde{\psi})$ (see Section 4.3.3),

$$\begin{aligned} E\left(\widetilde{\psi}^2\right) &= \text{Var}\left(\widetilde{\psi}\right) + \left(E\left(\widetilde{\psi}\right)\right)^2 \\ &= \frac{\psi(1-\psi\theta)}{S\theta} + \psi^2. \end{aligned} \tag{4.31}$$

4.9.2 Conditional expectation for $\widehat{\psi}^2$

Here we show the proof for the conditional expectation for $\widehat{\psi}^2$.

$$\begin{aligned} E\left(\widehat{\psi}^2 \mid b, S-f_0\right) &= E\left\{\widetilde{\psi}^2\{1-\frac{\tau(1-p)^{\tau-1}}{\theta}(\widehat{p}-p)\}^2 \mid b, S-f_0\right\} \\ &= \widetilde{\psi}^2\, E\left\{(\widehat{p}-p)^2 \mid b, S-f_0\right\} \frac{\tau^2(1-p)^{2(\tau-1)}}{\theta^2} \\ &= \widetilde{\psi}^2\left\{1-\frac{\tau^2(1-p)^{2(\tau-1)}}{\theta^2}\,\mathrm{Var}\,(\widehat{p})\right\}, \end{aligned}$$

or in terms of $\widehat{\theta}$,

$$\begin{aligned} &= E\left\{\widetilde{\psi}^2(1-\frac{1}{\theta^2}(\widehat{\theta}-\theta))^2 \mid b, S-f_0\right\} \\ &= \widetilde{\psi}^2\left\{1-\frac{1}{\theta^2}E[(\widehat{\theta}-\theta)^2 \mid b, S-f_0]\right\} \\ &= \widetilde{\psi}^2\left\{1-\frac{1}{\theta^2}\,\mathrm{Var}\left(\widehat{\theta}\right)\right\}, \end{aligned} \tag{4.32}$$

where $\mathrm{Var}\left(\widehat{p}\,\middle|\, b, S-f_0\right) \approx \mathrm{Var}(\widehat{p})$ and $\mathrm{Var}\left(\widehat{\theta}\,\middle|\, b, S-f_0\right) \approx \mathrm{Var}(\widehat{\theta})$, based on the assumption of asymptotic independence. The variance for $\widehat{\theta}$ is derived in the next section.

4.9.3 Variances

An approximate expression for $\mathrm{Var}(\widehat{\theta})$ is based on a Taylor expansion for $\widehat{\theta}$ about p, $\widehat{\theta} \approx \theta + \tau(1-p)^{\tau-1}(\widehat{p}-p)$ (see Equation (4.6)). Then $\mathrm{Var}(\widehat{\theta}) \approx \mathrm{Var}\,(\theta+\tau(1-p)^{\tau-1}(\widehat{p}-p)) = \tau^2(1-p)^{2(\tau-1)}\,\mathrm{Var}\,(\widehat{p})$.

Thus, an approximate variance for conditional $\widehat{\theta}$ is

$$\begin{aligned} \mathrm{Var}\left(\widehat{\theta} \mid b, S-f_0\right) &\approx \mathrm{Var}\left\{\tau(1-p)^{\tau-1}\widehat{p} \mid b, S-f_0\right\} \\ &= \tau^2(1-p)^{2(\tau-1)}\,\mathrm{Var}\left(\widehat{p} \mid b, S-f_0\right). \end{aligned} \tag{4.33}$$

Estimates are used in place of $\mathrm{Var}(\widehat{p})$ and may be obtained from a fitted model for p, $\mathrm{Var}\,(\widehat{p}) \approx \widehat{\mathrm{Var}}\,(\widehat{p})$.

To find an approximation for the variance for $\widehat{\psi}$ we condition on b (= the number of remaining visits after the time of first detection) and $S-f_0$ (= the

number of detected sites),

$$\text{Var}(\widehat{\psi}\,) = \text{Var}\left\{E\left(\widehat{\psi}\;\middle|\, b, S - f_0\right)\right\} + E\left\{\text{Var}\left(\widehat{\psi}\;\middle|\, b, S - f_0\right)\right\}. \tag{4.34}$$

Now,

$$\text{Var}\left(\widehat{\psi}\;\middle|\, b, S - f_0\right) \approx E\left(\widehat{\psi}^2\;\middle|\, b, S - f_0\right) - \left\{E\left(\widehat{\psi}\;\middle|\, b, S - f_0\right)\right\}^2, \tag{4.35}$$

which is found from substituting Equations (4.30) and (4.32). Then, together with $\text{Var}(\widehat{p}\,) \approx p(1-p)/b$ (Section 4.3.2),

$$\begin{aligned} E\left\{\text{Var}(\widehat{\psi}\;\middle|\, b, S - f_0)\right\} &\approx E\left\{\frac{\widetilde{\psi}^2\tau^2(1-p)^{2(\tau-1)}\,\text{Var}(\widehat{p}\,)}{\theta^2}\right\} \\ &\approx E\left(\widetilde{\psi}^2\right)\frac{\tau^2(1-p)^{2(\tau-1)}}{\theta^2}E\left(\frac{p(1-p)}{b}\right) \\ &\approx \left\{\frac{\psi(1-\psi\theta)}{S\theta} + \psi^2\right\}\frac{\tau^2(1-p)^{2(\tau-1)}}{\theta^2}\frac{p(1-p)}{b}. \end{aligned} \tag{4.36}$$

Finally,

$$\begin{aligned} \text{Var}\left\{E\left(\widehat{\psi}\;\middle|\, b, S - f_0\right)\right\} &= \text{Var}(\widetilde{\psi}\,) \\ &= \frac{\psi(1-\psi\theta)}{S\theta}, \end{aligned} \tag{4.37}$$

where from before we found that $E\left(\widehat{\psi}\;\middle|\, b, S - f_0\right) = \widetilde{\psi}$.

4.10 Appendix 2: Proofs for site inhomogeneity

4.10.1 Basic quantities

Here we derive some basic quantities used in derivations of Section 4.4.

We denote $\psi_s(\alpha) = (1+\exp(-\alpha^T x_s))^{-1} = \exp(\alpha^T x_s)/(1+\exp(\alpha^T x_s))$. Note, $1-\psi_s(\alpha) = \exp(-\alpha^T x_s)/(1+\exp(-\alpha^T x_s))$.

The partial log-likelihood is

$$\ell(\alpha,\beta) = \sum_{s=1}^{S}(1-w_s)\log(1-\psi_s(\alpha)\widehat{\theta}_s) + w_s \log\psi_s(\alpha).$$

Let $f(\alpha) = 1-\psi_s(\alpha)\widehat{\theta}_s$ and $g(\alpha) = \psi_s(\alpha)$ so that

$$f'(\alpha) = \partial(1-\psi_s(\alpha)\widehat{\theta}_s)/\partial\alpha = \theta_s x_s \exp(-\alpha^T x_s)/\psi_s(\alpha)^2$$

and

$$g'(\alpha) = x_s \exp(-\alpha^T x_s)/\psi_s(\alpha)^2 = x_s/(1+\exp(-\alpha^T x_s) \quad = x_s\psi(1-\psi).$$

This gives

$$\begin{aligned}
Q(\alpha,\beta) &= \frac{\partial\ell(\alpha,\beta)}{\partial\alpha^T} \\
&= (1-w_s)\frac{f'(a)}{f(a)} + w_s\frac{g'(a)}{g(a)} \\
&= \frac{(1-w_s)\left\{-\theta_s x_s \exp(-\alpha^T x_s)/\psi_s(\alpha)^2\right\}}{1-\psi_s(\alpha)\theta_s} \\
&\quad + \frac{w_s x_s \exp(-\alpha^T x_s)/\psi_s(\alpha)^2}{\psi_s(\alpha)} \\
&= \frac{\left\{-\theta_s\psi_s + w_s\right\}\psi_s x_s\left\{1-\psi_s(\alpha)\right\}}{\psi_s(\alpha)\left\{1-\psi_s(\alpha)\theta_s\right\}} \\
&= \frac{x_s\left\{w_s - \theta_s\psi_s(\alpha)\right\}\left\{1-\psi_s(\alpha)\right\}}{1-\psi_s\theta_s}.
\end{aligned}$$

4.10.2 Variance for $\widehat{\alpha}$

Here we show the variance for $\widehat{\alpha}$ evaluated at $\widehat{\beta}$ used in derivations for the approximate variance of occupancy (Sections: 4.4.3.1, 4.4.4.2 and 4.4.5.1). $Q(\alpha_0, \beta_0)$ is the sum of independent random variables, thus under mild regularity conditions the central limit theorem holds so that

$$S^{-1/2} Q(\alpha_0, \beta_0) \xrightarrow{d} N(0, \Sigma_Q).$$

Recall that

$$Q(\alpha_0, \beta_0) = \frac{\partial \ell(\alpha_0, \beta_0)}{\partial \alpha^T} = \frac{\partial \log L(\alpha_0, \beta_0)}{\partial \alpha^T}.$$

and

$$I(\alpha_0, \beta_0) = -\frac{\partial Q(\alpha_0, \beta_0)}{\partial \alpha^T} = -\frac{\partial \ell^2(\alpha_0, \beta_0)}{(\partial \alpha^T)^2}.$$

Now we show that $\Sigma_Q = A(\alpha_0, \beta_0)^T$ from $A(\alpha_0, \beta_0)^{-1} \Sigma_Q A(\alpha_0, \beta_0)^{-T} = A(\alpha_0, \beta_0)^{-1}$ which implies that $\Sigma_Q = A(\alpha_0, \beta_0)^T$. To show this we find

$$\begin{aligned}
\text{Var}\{Q(\alpha_0, \beta_0)\} &= E\left\{Q(\alpha_0, \beta_0)^2\right\} - E\left\{Q(\alpha_0, \beta_0)\right\}^2 \\
&= E\left\{Q(\alpha_0, \beta_0)^2\right\} \\
&= E\left\{\left(\frac{\partial \ell(\alpha_0, \beta_0)}{\partial \alpha^T}\right)^2\right\} \\
&= -E\left\{\frac{\partial^2 \ell(\alpha_0, \beta_0)}{(\partial \alpha^T)^2}\right\} \\
&= -E\left\{-I(\alpha_0, \beta_0)\right\} \\
&= I(\alpha_0, \beta_0) \\
&= -\frac{\partial Q(\alpha_0, \beta_0)}{\partial \alpha^T}.
\end{aligned}$$

Alternatively, we can use $\text{Var}\big\{Q(\alpha_0, \beta_0)\big\} = \Sigma q_s q_s^T$ and

$$\begin{aligned}
\text{Var}(Q(\alpha_0, \beta_0)) &= \Sigma q_s q_s^T \\
&= \Sigma E(q_s q_s^T) \\
&= \Sigma E\left\{\frac{\partial^2 \ell(\alpha_0, \beta_0)}{(\partial \alpha^T)^2}\right\} \\
&= I(\alpha_0, \beta_0) \\
&= -\frac{\partial Q(\alpha_0, \beta_0)}{\partial \alpha^T},
\end{aligned}$$

where $E\left\{Q(\alpha_0, \beta_0)\right\} = E\left\{\dfrac{\partial \log L(\alpha_0, \beta_0)}{\partial \alpha^T}\right\} = 0$ by definition. Thus,

$$\begin{aligned}
\Sigma_Q &= \text{Var}\{S^{-1/2} Q(\alpha_0, \beta_0)\} \\
&= S^{-1}\text{Var}\{Q(\alpha_0, \beta_0)\} \\
&= S^{-1} I(\alpha_0, \beta_0) \\
&= -S^{-1}\frac{\partial Q(\alpha_0, \beta_0)}{\partial \alpha^T} \\
&\approx A(\alpha_0, \beta_0)^T.
\end{aligned}$$

Now we can use this to show that

$$\begin{aligned}
A(\alpha_0, \beta_0)^{-1} \Sigma_Q A(\alpha_0, \beta_0)^{-T} &= A(\alpha_0, \beta_0)^{-1} A(\alpha_0, \beta_0)^T A(\alpha_0, \beta_0)^{-T} \\
&= A(\alpha_0, \beta_0)^{-1}.
\end{aligned}$$

We assume that $\widehat{\beta}$ has been constructed from a method such as our partial or conditional likelihood so that

$$S^{1/2}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, \Sigma_\beta)$$

holds and

$$\text{Var}\left\{S^{1/2}(\widehat{\beta} - \beta)\right\} = S\text{Var}\{\widehat{\beta}\} = SV_\beta = \Sigma_\beta,$$

and V_β is the covariance matrix for $\widehat{\beta}$.

We go on to find the variance for $\widehat{\alpha}$ evaluated at $\widehat{\beta}$. Given that $\text{Var}\left\{S^{-1/2} Q(\alpha_0, \beta_0)\right\} = \Sigma_Q$, $\text{Var}\left\{S^{1/2}(\widehat{\beta} - \beta)\right\} = \Sigma_\beta$, and assuming independence between α_0 and β_0

which implies that the covariance is 0,

$$\begin{aligned}
&\text{Var}\left\{S^{1/2}(\widehat{\alpha}(\widehat{\beta}) - \alpha_0)\right\} \\
&= \text{Var}\left\{-A(\alpha_0,\beta_0)^{-1}\left\{S^{-1/2}Q(\alpha_0,\beta_0) + B(\alpha_0,\beta_0)S^{1/2}(\widehat{\beta}-\beta)\right\}\right\} \\
&= A(\alpha_0,\beta_0)^{-1}\Big\{\text{Var}\left\{S^{-1/2}Q(\alpha_0,\beta_0) + B(\alpha_0,\beta_0)S^{1/2}(\widehat{\beta}-\beta)\right\}\Big\}A(\alpha_0,\beta_0)^{-T} \\
&= A(\alpha_0,\beta_0)^{-1}\Big\{\text{Var}\left\{S^{-1/2}Q(\alpha_0,\beta_0)\right\} \\
&\quad + \text{Var}\left\{B(\alpha_0,\beta_0)S^{1/2}(\widehat{\beta}-\beta)\right\}\Big\}A(\alpha_0,\beta_0)^{-T} \\
&= A(\alpha_0,\beta_0)^{-1}\left\{\Sigma_Q + B(\alpha_0,\beta_0)\Sigma_\beta B(\alpha_0,\beta_0)^T\right\}A(\alpha_0,\beta_0)^{-T}.
\end{aligned}$$

Next, using $S\widehat{V}_B = \widehat{\Sigma}_\beta$, $\Sigma_Q = A(\alpha_0,\beta_0)^T$, $S^{-1}I(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta}) \approx (\widehat{\alpha},\widehat{\beta})$ and $S^{-1}B(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta}) \approx B(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta})$, we derive

$$\begin{aligned}
\widehat{\text{Var}}\left\{\widehat{\alpha}(\widehat{\beta})\right\} &= S^{-1}\Bigg[\left\{S^{-1}I(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta})\right\}^{-1} \\
&\quad + \left\{S^{-1}I(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta})\right\}^{-1}S^{-1}\widetilde{B}(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta})S\widehat{V}_B S^{-1}\widetilde{B}(\widehat{\alpha},\widehat{\beta})\left\{S^{-1}I(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta})\right\}^{-1}\Bigg] \\
&= I(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta})^{-1} + I(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta})^{-1}\widetilde{B}(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta})\widehat{V}_B\widetilde{B}(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta})^T I(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta})^{-1},
\end{aligned}$$

where $\widetilde{B}(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta}) = \partial Q(\widehat{\alpha}(\widehat{\beta}),\widehat{\beta})/\partial\beta^T$, as shown in Equation (4.18) of Section 4.4.3.1.

Chapter 5

Conclusions, discussion and future work

Investigations in Chapters 2, 3 and 4, of the full likelihood function in the homogeneous case, where ψ and p are assumed constant, reveal several practical limitations. These were found in theoretical results as well as applied examples considered in these chapters. In Chapter 2 we considered small, moderate and large sized studies, in Chapter 4 we considered also larger studies. And the extension to penalised likelihood for the full likelihood is not trivial, and possibly unnecessary given the more than satisfactory performance of our partial likelihood approach. Broadly speaking, these (limitations) may be grouped into: 1) boundary estimates; 2) nonconvergence of the likelihood; 3) inadequate asymptotic standard error; and 4) computational issues in implementing modern GLM methods. These limitations extend through to the site inhomogeneous case (nonconstant site ψ and p) especially when covariates are considered.

We addressed these limitations and provided resolutions. Initially this thesis considered the homogenous case that does not include any covariates.

The first limitation regarding boundary estimates (probabilities of zero or one) was addressed in Chapter 2. It is possible that estimates are equal to zero or greater than one, which means that the score equations are not always the solutions for the ML estimates. We resolved this by giving the correct expressions for the MLEs for ψ and p that apply to the three boundaries of the sample space, i.e. the 'Edge solutions' (Section 2.3.2).

For small values of detectability, we showed that estimation of occupancy is

unfeasible, in that the ML estimates do not exist everywhere in the parameter space. We produced a rule specifying a region where estimation of (ψ, p) is 'plausible': a) the MLEs always exist; and b) estimates are less biased. The region ensures that the expectation function for occupancy will always be monotone increasing. The amount of bias will depend on how densely populated is the plausible region.

Knowing the edge solutions means that boundary estimates are no longer an issue because given the sufficient statistics of an experiment the correct expression (either the score equations or the edge solutions) for the MLEs always can be determined.

Next, consideration was given to the precision of the estimates (Chapter 2). The asymptotic variance for the full likelihood for the homogeneous case does not provide a closed form solution for the estimates. This makes it unreliable, the information matrix is not always invertible so that values cannot always be calculated. When values are produced, it is suspected that these may underestimate the actual variance especially for small samples (and bootstrap methods are instead recommended) (e.g. MacKenzie et al., 2002).

To evaluate the ML estimator of ψ, we derived the expression for the joint probability mass function of the sufficient statistics; the number of detected sites and the total number of detections for a study. With this we were able to obtain expressions for, and evaluate, the exact expectation, the exact variance, and the exact bias, and to explore bias corrections for occupancy.

The asymptotic variance was found to be less than satisfactory; it seriously underestimates the variance, unless N and T are large. This was illustrated by comparisons of the exact and asymptotic variances, and the MSE in Section 2.5.4, Chapter 2 (for example, see Figures 2.9 and 2.10). This will lead to confidence intervals for $\widehat{\psi}$ that are too narrow, for example the normal approximation interval width, and coverage, including Figure 3.4 and Table 3.2. A possible correction for the variance estimates an overdispersion factor to inflate them (MacKenzie and Bailey, 2004). Their estimate was based on a simple Pearson chi-square statistic over the average test statistics obtained from parametric bootstraps.

A number of alternative interval estimators were explored in Chapter 3. These results have been published in Karavarsamis et al. (2013). We compared three bootstrap-based interval estimators against the normal approximation. Bootstrap estimators were selected on the basis that these are recommended and

commonly used (for example, see MacKenzie et al., 2002). It was necessary to determine which performs best. Overall, we found that when detectability, the number of sites, and the number of survey-occasions of the study are not too small, the studentised interval estimator gives the most consistent interval estimates over the range of p and ψ. Alternatively, a better approximation than the current asymptotic estimator should be devised for the full likelihood. We propose better variance approximations for the partial likelihood.

Next, the bias of occupancy was evaluated. We tried bias corrections on a conditional and unconditional expectation for $\widehat{\psi}$ within the plausible region, i.e. the region of the parameter space where MLEs always exist and are less biased. However, no corrections were effective for few sites or survey occasions, or for small detection probability. And, bias corrections are unnecessary when all these quantities are large.

There are far greater issues with the full likelihood which prohibit any reasonable bias correction to be adequate and/or effective. These sentiments are shared by other studies. For example, Welsh et al. (2013) found that when data are sparse and when detection depends on abundance (or occupancy) estimators cannot be corrected for bias. They showed that to ignore the possibility of nondetection (to assume that the site is unoccupied) results in a similar bias, and smaller variance, than when it is not ignored. They concluded that because detection error is not directly observable its presence (or magnitude) cannot be determined, which means that it is not possible to adjust for bias.

Given the discussion thus far, it is no surprise that our investigations confirmed that the full likelihood is affected by nonconvergence and identifiability (results shown in Chapter 3 and published in Karavarsamis et al. (2013)). Overall, our work is well anchored within the literature for occupancy models, for example MacKenzie et al. (2002); Wintle et al. (2004); MacKenzie et al. (2009); Guillera-Arroita et al. (2010); Welsh et al. (2013) .

In the final part of this thesis we developed a basic methodology for occupancy models with partial likelihoods. Thus we were able to overcome many of the issues of the full likelihood. Fundamentally, we devised a two-stage process (by exploiting redetections at sites) to estimate occupancy and detectability separately, that gives full access to GLM and GAM machinery in standard software. Thus our methods are easily extendable to allow for more possibilities; for example sliced inverse regression (SIR) is now accessible.

Throughout the development of methods in Chapter 4, we avoided issues to

do with identifiability, which are known to affect the full likelihood for the homogeneous case, previously referred to as the basic occupancy model. In this chapter we demonstrated that our partial likelihood two-stage approach gives estimators that are less affected by identifiability and that it is almost always possible to obtain estimates. We showed our estimators to be consistent, efficient and more robust to boundary estimates and identifiability. The analytic forms of our estimators means that estimates are more stable than those generated from the full likelihood.

We also developed a better approximation to the variance for occupancy and detectability for GLMs and GAMs by simple extension to the case of time varying covariates for detection. The analytic forms of the variances guarantee estimates are always available. We found our methods to be at least comparable or better than existing methods, for example, when compared to asymptotic and bootstrap approximations (used by `PRESENCE` and `unmarked` software, for example). In addition, our methods are readily applied with existing software, such as R, which may be run easily on many computer platforms (e.g Apple, Linux, Windows etc.). Whereas other software is not so easily adaptable (e.g `PRSENCE` which runs readily and easily on Windows but no so on Apple Computers, for example.)

This thesis focussed on covariates that are constant over time for occupancy and detectability, then introduced time varying covariates for detectability. A natural extension would be to consider when occupancy and detectability vary in time.

When occupancy and detectability both vary in time the partial likelihood is not appropriate. We could consider extensions to a composite likelihood, models that may include seasonality, but it is beyond the scope of this thesis. Briefly, the model would estimate overdispersion. And, these models would still have the benefits of the GLM framework. Comparative and alternative methods include random-effects models, and more generally mixed models (or hierarchical models). These would be natural extensions to models considered in this thesis. We note that hierarchical models, as well as multi-season models, have been considered extensively for the full likelihood case (see Royle and Dorazio, 2008, for example).

Robust designs founded in capture-recapture methods would be appropriate for our models (for example, Pollock, 1982; Pollock et al., 1990). These allow for sampling at two different time scales: a primary and secondary sampling

phase. Time scales are selected such that within each (sampling phase) probabilities are assumed constant. We may choose to visit sites within the same season over a number of years. In this way we assume that occupancy and detectability are constant at each time scale but not across time scales; at the first phase probabilities are constant between years but can vary within years; at the second phase probabilities are constant within seasons but may vary between seasons. The primary sampling phase takes place over years and the secondary sampling phase occurs within years, at the season level. This idea has been implemented in dynamic occupancy models (MacKenzie et al., 2003; McClintock et al., 2010).

Currently, our methods assume time constant models for both occupancy and detection probability. These cannot be easily extended to a nonconstant occupancy. A next step would be to consider seasonal models, with a conditional likelihood for season.

Our methods have the possibility to be extended to account for misidentification, or false-positive detections (Miller et al., 2011). These occur when detection of a species is incorrectly recorded when the site actually is unoccupied. In the site inhomogeneous likelihood, detection probabilities could be estimated from redetections after the initial positive detection, where we assume false positives do not occur. However, a likelihood conditional on at least one positive detection at a site might be better. We may require a Bayes probability approach for uncertain detections.

Finally we mention that extensions to our work would include combining a variety of data structures. For example, we could use capture-recapture data recorded for a species from a small study to augment presence-absences recorded from a larger study for the species.

In closing, in this thesis we resolved some existing problems and extended the basic methodology for the analysis of occupancy models, an active research area which continues to grow.

Bibliography

Aing, C., Halls, S., Oken, K., Dobrow, R., and Fieberg, J. (2011). A bayesian hierarchical occupancy model for track surveys conducted in a series of linear, spatially correlated, sites. *Journal of Applied Ecology*, 48:1508–1517.

Azuma, D., Baldwin, J., and Noon, B. (1990). Estimating the occupancy of spotted owl habitat areas by sampling and adjusting bias. *USDA Forest Service General Technical Report. PSW-124*.

Bartlett, M. (1955). Approximate confidence intervals, iii. a bias correction. *Biometrika*, 42:201–204.

Bayley, N. and Peterson, J. (2001). An approach to estimate probability of presence and richness of fish species. *Transactions of the American Fisheries Society*, 130:620–633.

Binns, M., Nyrop, J., and van der Werf, W. (2000). *Sampling and monitoring in crop protection: the theoretial basis for developing practical decision guides*. CAB International Publishing, Wallingford, Oxon, UK.

Brent, R. P. (1973). *Algorithms for minimization without derivatives*. Courier Dover Publications.

Burnham, K., Anderson, D., White, G., Brownie, C., and Pollock, K. (1987). *Design and Analysis Methods for Fish Survival Experiments based on Release-Recapture*. American Fisheries Society Monographs 5, Bethesda, Maryland.

Burnham, K. and Overton, W. (1978). Estimation of the size of a closed population when capture proabilities vary among animals. *Biometrika*, 65:625–633.

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

Byrd, R., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Scientific Computing*, 16(5):1190–1208.

Canty, A. and Ripley, B. (2009). *boot: Bootstrap R (S-Plus) Functions*. R Foundation for Statistical Computing. Version 1.2-41, Vienna, Austria. ISBN 3-900051-07-0.

Casella, G. and Berger, R. (2002). *Statistical Inference*. Duxbury, Pacific Grove, CA, USA, 2nd edition edition.

Chao, A. (2001). An overview of closed capture-recapture models. *Journal of Agricultural, Biological, and Environmental Statistics*, 6(2):158–175.

Cormack, R. (1964). Estimates of survival from sightings of marked animals. *Biometrika*, 51:455–506.

Cox, D. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

Cox, D. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3):729–737.

Davison, A. and Hinkley, D. (1997). *Bootstrap methods and their application*. Cambridge University Press, Cambridge, UK.

de Boor, C. (1978). *A Practical Guide to Splines*. Cambridge University Press, Cambridge, UK.

Dorazio, R. M. (2007). On the choice of statistical models for estimating occurrence and extinction from animal surveys. *Ecology*, 88(11):2773–2782.

Efron, B. (1982). *The Jacknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia.

Fiske, I. and Chandler, R. (2014). *Overview of Unmarked: An R Package for the Analysis of Data from Unmarked Animals*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Fiske, I. J. and Chandler, R. B. (2011). unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, 43(10):1–23.

Geissler, P. and Fuller, M. (1987). Estimation of the proportion of area occupied by an animal species. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 1986:533–538.

Gill, R. (1992). Marginal partial likelihood. *Scandinavian Journal of Statistics*, 19(2):133–137.

Gimenez, O., Bonner, S., King, R., Parker, R., Brooks, S., Jamieson, L., Grosbois, V., Morgan, B., and Thomas, L. (2009). Winbugs for population ecologists: Bayesian modeling using markov chain monte carlo methods. In Thomson, D. L., Cooch, E. G., and Conroy, M. J., editors, *Modeling Demographic Processes In Marked Populations*, volume 3 of *Environmental and Ecological Statistics*, pages 883–915. Springer US.

Gimenez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H., Vila, J., and Pradel, R. (2007). State–space modelling of data on marked individuals. *Ecological Modelling*, 206(431–438).

Guillera-Arroita, G., Lahoz-Monfort, J., MacKenzie, D., Wintle, B., and McCarthy, M. (2014). Ignoring imperfect detection in biological surveys is dangerous: A response to 'Fitting and interpreting occupancy models'. PLoS ONE, 9(7):e99571. doi:10.1371/journal.pone.0099571.

Guillera-Arroita, G., Ridout, M., and Morgan, B. (2010). Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution*, 1(2):131–139.

Guillera-Arroita, G., Ridout, M., Morgan, B., and Linkie, M. (2011). Models for species-detection data collected along transects in the presence of abundance-induced heterogeneity and clustering in the detection process. *Methods in Ecology and Evolution*, 1(2):131–139.

Haas, P. and Stokes, L. (1998). Estimating the number of classes in a finite population. *Journal of the American Statistical Association*, 93(444):1475–1487.

Hall, D. B. (2000). Zero-inflated poisson and binomial regression with random effects: a case study. *Biometrics*, 56:1030–1039.

Hanski, I. (1994). Patch-occupancy dynamics in fragmented landscapes. *Trends in Ecology & Evolution*, 9(4):131–135.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models (with discussion). *Statistical Science*, 1:297–318.

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*. Chapman & Hall.

Heard, G., Robertson, P., and Scroggie, M. (2006). Assessing detection probabilities for the endangeared growling grass frog in Southern Victoria. *Wildlife Research*, 33:557–564.

Hepworth, G. (2004). Mid-P confidence intervals based on the likelihood ratio for proportions estimated by group testing. *Australian and New Zealand Journal of Statistics*, 46:391–405.

Hepworth, G. and Watson, R. (2009). Debiased estimation of proportions in group testing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1):105–121.

Hoffman, J., Aguilar-Amuchastegui, N., and Tyre, A. (2010). Use of simulated data from a oricess-based habitat model to evaliate methods for predicting species occurrence. *Ecography*, 33:656–666.

Holzmann, H., Munk, A., and Zucchini, W. (2006). On identifiability in capture-recapture models. *Biometrics*, 62:934–939.

Hu, M. and Lachin, J. (2003). Corrections for bias in maximum likelihood parameter estimates due to nuissance parameters. *Communications in Statistics*, 32(3):619–639.

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

Huggins, R. M. and Hwang, W. (2007). Non-parametric estimation of population size from capture-recapture data when the capture probability depends on a covariate. *Applied Statistics*, 56(4):429–443.

Hui, C., Foxcroft, L. C., Richardson, D. M., and MacFadyen, S. (2011). Defining optimal sampling effort for large-scale monitoring of invasive alien plants: a bayesian method for estimating abundance and distribution. *Journal of Applied Ecology*, 48(3):768–776.

Hwang, W. and Shen, T. (2010). Small-sample estimation of species richness applied to forest communities. *Biometrics*, 66(4):1052–60.

Jolly, G. (1965). Explicit estimates from capture-recapture data with both death and immigration: Stochastic model. *Biometrics*, 52:225–247.

Karavarsamis, N., Robinson, A. P., Hepworth, G., Hamilton, A., and Heard, G. (2013). Comparison of four bootstrap-based interval estimators of species

occupancy and detection probabilities. *Australian and New Zealand Journal of Statistics*, 55(3):235–252.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34:1–14.

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

Levin, B. and Kong, F. (1990). Bartlett's bias correction to the profile score function is a saddlepoint correction. *Biometrika*, 77(1):219–221.

Lindenmayer, D. and Burgman, M. (2005). *Practical Conservation Biology*. CSIRO Publishing, Collingwood, Victoria, Australia.

Lindsay, B. (1982). Conditional score functions: some optimality results. *Biometrika*, 69:503–512.

Lunn, D., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs – a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325–337.

MacKenzie, D. and Bailey, L. (2004). Assessing the fit of site-occupancy models. *Journal of Agricultural Biological and Environmental Statistics*, 9(3):300–318.

MacKenzie, D. and Hines, J. (2009). *PRESENCE ver 6.1*. Proteus Research and Consulting, and USGS, ARMI.

MacKenzie, D., Nichols, J., Hines, J., Knutson, M., and Franklin, A. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, 84(8):2200–2207.

MacKenzie, D. I., Bailey, L. L., and Hines, J. E. (2011). An integrated model of habitat and species occurrence dynamics. *Methods in Ecology and Evolution*, 2(6):612–622.

MacKenzie, D. I., Nichols, J., Royle, J., Pollock, K., Bailey, L., and Hines, J. (2006). *Occupancy Estimation and Modeling Inferring Patterns and Dynamics of Species Occurrence*. Elsevier, San Diego, CA.

MacKenzie, D. I., Nichols, J., Seamans, M. E., and Gutiërrez, R. J. (2009). Modeling species occurrence dynamics with multiple states and imperfect detection. *Ecology*, 90(3):823–835.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J., and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8):2248–2255.

Martin, J., Royle, J., MacKenzie, D., Edwards, H., Kéry, M., and Gardner, B. (2011). Accounting for non-independent detection when estimating abundance of organisms with a Bayesian approach. *Methods in Ecology and Evolution*, 2(6):595–601.

McClintock, B., Nichols, J. D., Bailey, L. L., MacKenzie, D. I., Kendall, W. L., and Franklin, A. B. (2010). Seeking a second opinion: uncertainty in disease ecology. *Ecology Letters*, 13:659–674.

McCullagh, P. and Nelder, J. (1989). *Generalised linear models*. Chapman and Hall, London, England, United Kingdom, Second edition.

Miller, D., Nichols, J., McClintock, B., Grant, E., Bailey, L., and Weir, L. (2011). Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology*, 92(7):1422–1428.

Milne, B., Johnston, K., and Forman, R. (1989). Scale-dependent proximity of wildlife habitat in a spatially-neutral Bayesian model. *Landscape Ecology*, 2:101–110.

Nelder, J. and Mead, R. (1965). A simplex method for function minimisation. *The Computer Journal*, 7(4):308–313.

Newcombe, R. (1998). Two-sided confidence intervals for the single proportion. *Statistics in Medicine*, 17:857–872.

Nichols, J., Hines, J. E., MacKenzie, D. I., Seamans, M., and Gutiérrez, R. (2007). Occupancy estimation and modeling with multiple states and state uncertainty. *Ecology*, 88(6):1395–1400.

Nichols, J. and Karanth, K. (2002). Statistical concepts: assesing spatial distributions. In Karanth, K. and Nichols, J., editors, *Monitoring Tigers and Their Prey: A Manual for Researchers, Managers, and Conservationists in Tropical Asia*, pages 29–38. Centre for Widlife Studies, Bangalore, India.

Pereira, J. M. C. and Itami, R. M. (1991). GIS-based modelling using logistic multiple regression: a case study of the Mt Graham red squirrel. *Photogrammetric Engineering and Remote Sensing*, 57:1475–1486.

Pledger, S., Pollock, K., and Norris, J. (2003). Open capture-recapture models with heterogeneity: I. Cormack-Jolly-Seber model. *Biometrics*, 59:786–794.

Pollock, K., Nichols, J., Brownie, C., and Hines, J. (1990). Statistical inference for capture–recapture experiments. *Wildlife Monographs*, 107.

Pollock, K. H. (1982). A capture–recapture design robust to unequal probability of capture. *Journal of Wildlife Management*, 46:757–760.

Pollock, K. H. (2002). The use of auxiliary variables in capture-recapture modeling: an overview. *Journal of Applied Statistics*, 29:85–102.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ridout, M., Demétrio, C., and Hinde, J. (1998). Models for count data with many zeros. *Proceedings of the XIXth International Biometric Conference*, 19:179–192.

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

Royle, J. A. and Dorazio, R. (2008). *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations, and Communities*. Academic Press, San Diego, CA.

Royle, J. A. and Nichols, J. D. (2003). Estimating abundance from repeated presence-absence data or point counts. *Ecology*, 84(3):777–790.

Schmidt, B. (2004). Declining amphibian populations: The pitfalls of count data in the study of diversity, distributions, dynamics, and demography. *Herpetological Journal*, 14(4):167–174.

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

Stoklosa, J., Hwang, W.-H., Wu, S.-H., and Huggins, R. (2011). Heterogeneous capture-recapture models with covariates: A partial likelihood approach for closed populations. *Biometrics*, 67(4):1659–1665.

Tyre, A., Tenhumberg, B., Field, S., and Niejalke, D. (2003). Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications*, 13(6):1790 –1801.

Welsh, A. H., Cunningham, R., Donnelly, C., and Lindenmayer, D. B. (1996). Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological modelling*, 88:297–308.

Welsh, A. H., Lindenmayer, D. B., and Donnelly, C. F. (2013). Fitting and interpreting occupancy models. *PLoS ONE*, 8(1):e52015. doi:10.1371/journal.pone.0052015.s001.

Wintle, B., Kavanagh, R. P., Mccarthy, M. A., and Burgman, M. (2005). Estimating and dealing with detectability in occupancy surveys for forest owls and arboreal marsupials. *The Journal of Wildlife Management*, 69(3):905–917.

Wintle, B., McCarthy, M., Parris, K., and Burgman, M. (2004). Precision and bias of methods for estimating point survey detection probabilities. *Ecological Applications*, 14:703–712.

Wintle, B., McCarthy, M., Volinsky, C., and Kavanagh, R. (2003). The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology*, 17(6):1579–1590.

Wong, W. (1986). Theory of partial likelihood. *The Annals of Statistics*, 14(1):88–123.

Wood, S. N. (2006). *Generalized additive models: An introduction with R.* Chapman & Hall/CRC, Florida, US.

Yee, T. W. and Wild, C. J. (1996). Vector generalized additive models. *Journal of the Royal Statistical Society, Series B, Methodological*, 58:481–493.

Yip, P. S. F., Huggins, R. M., and Lin, D. Y. (1996). Inference for capture-recapture experiments in continuous time with variable capture rates. *Biometrika*, 83:477–483.

Author/s:
KARAVARSAMIS, NATALIE

Title:
Methods for estimating occupancy

Date:
2014

Persistent Link:
http://hdl.handle.net/11343/44235

File Description:
Methods for estimating occupancy