University of Southampton

Faculty of Social Sciences
School of Mathematical Sciences

Sampling Effort and Uncertainty Assessment in Capture Recapture Studies

by

Su Na Chin
ORCiD: 0000-0002-6826-266X

A thesis for the degree of
Doctor of Philosophy

September 2025

University of Southampton

Abstract

Faculty of Social Sciences
School of Mathematical Sciences

Doctor of Philosophy

Sampling Effort and Uncertainty Assessment in Capture Recapture Studies

by Su Na Chin

This thesis examines the optimal allocation of sampling effort in capture-recapture studies, with the aim of improving population size estimation. Sampling effort is a critical component of study design, yet it is often determined without a systematic approach. To address this gap, a simple, practical framework is developed using theoretical insights, basic mathematical computations, and computer simulations. The analysis begins with the Lincoln–Petersen model, which involves two primary capture occasions, each comprising multiple sub-occasions. When capture probabilities are either constant or follow a consistent pattern, distributing sub-occasions equally across primary occasions yields the most accurate estimates. However, when capture probabilities vary and are known in advance, optimal allocation is better achieved through numerical methods such as the Newton–Raphson algorithm and simple estimation techniques that incorporate prior information. These targeted strategies outperform equal allocation, particularly when capture probabilities fluctuate significantly over time. The investigation then extends to the Schnabel model, which focuses on determining the optimal number of capture occasions. To account for unobserved individuals, the model incorporates zero-truncated count data. In cases where closed-form solutions are unavailable, the Expectation–Maximisation algorithm is employed to estimate parameters. The hierarchical structure is expanded to scenarios involving multiple capture occasions. When capture probabilities remain stable or change predictably, uniform sampling effort remains effective. However, in contexts where capture probabilities decline over time, allocating greater effort in later occasions leads to more accurate population estimates by compensating for reduced detectability. The thesis also provides practical recommendations for real-world applications where resources are limited. The proposed methods support informed decisions about sampling effort, avoiding reliance on arbitrary or overly conservative designs. By clarifying the relationship between detectability, study design, and estimation precision, the framework enables more efficient planning.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Su Na Chin, declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been submitted for publication as:
    - Chin, S. N. and Böhning, D. (2025). Choice of sampling effort in a Schnabel census for accurate population size estimates. *Environmental and Ecological Statistics*, 32(2):753–769
    - Chin, S. N., Overstall, A., and Böhning, D. (2026). Sampling effort and its allocation in the Lincoln–Petersen experiment: A hierarchical approach. *Journal of Statistical Planning and Inference*, 241:106330

Signed:..... Date:.....

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Dankmar Böhning for his consistent support, guidance, and encouragement throughout this research journey. His deep understanding and helpful feedback helped shape the ideas in this thesis. I am also thankful to my co-supervisor, Dr Antony Overstall, for his helpful suggestions and guidance, which made a big difference in the progress of this work.

I extend my appreciation to the School of Mathematical Sciences and Statistical Sciences Research Institute (S3RI) for providing a supportive academic environment. Special thanks go to Chuanyang, Disha, and Ziyan for their friendship, conversations, and shared moments that made this journey more enjoyable.

This research was made possible through the financial support of the Ministry of Higher Education Malaysia and Universiti Malaysia Sabah, for which I am sincerely thankful. Their support has enabled me to pursue this academic journey and contribute to the field.

My heartfelt thanks go to my family. To my husband, Jack, thank you for your steadfast support and for being by my side throughout this journey in the UK. Your belief in me made this journey possible. To my children, Isaac and Isabelle, your happiness and strength gave me the courage to keep going. I am also grateful to my parents and sisters for their continued emotional support and love from afar.

Thanks also go to Southampton Chinese Christian Church for providing a strong community of faith during my time here. A special thank you to Denise and Erik, for their prayers and kindness, which uplifted me during challenging times.

Above all, I give thanks to God, whose grace, strength, and guidance have carried me through this journey.

To Jack, Isaac and Isabelle

Chapter 1

Introduction

1.1 Background Study

1.1.1 Overview of Capture-Recapture

Capture-recapture methodology has emerged as a vital statistical tool for population estimation across diverse disciplines including ecology, epidemiology, criminology, and social sciences (King and McCrea, 2019). In most real-world contexts, observing all individuals in a population is not possible. This approach addresses the challenge of incomplete population enumeration by offering robust methods for indirect estimation.

The core methodology involves capturing, uniquely marking, and releasing individuals, followed by subsequent recapture sessions. Population size is then estimated by analysing the proportion of marked individuals in recapture samples (Seber and Schofield, 2023; Amstrup et al., 2006). First developed for wildlife population assessment (Petersen, 1896; Lincoln, 1930), these methods have demonstrated remarkable versatility, finding applications in human population studies (Sekar and Deming, 1949), linguistics (Williams et al., 2014), and digital technology research (Lu and Li, 2010).

In public health, capture-recapture offers a cost-effective solution for disease surveillance, particularly for estimating prevalence of under-reported conditions (Ramos et al., 2020; Böhning et al., 2020). Its ability to quantify hidden populations and uncover concealed phenomena has made it an indispensable analytical tool. The method's strength lies in its capacity to integrate data from multiple independent sources, thereby enhancing the evidence base for health policy decisions (Bailly et al., 2019). The broad use of capture-recapture methods across diverse fields highlights their adaptability and real-world usefulness.

1.1.2 Uncertainty Assessment

Providing only a point estimate without showing its reliability can be misleading in population studies (Williams et al., 2002). This is particularly true for capture-recapture studies, where researchers work with incomplete data to estimate population sizes (Borchers et al., 2002; Seber, 1982). Since these estimates rely on indirect observations, they carry more potential for error, making it crucial to measure and report their uncertainty.

Uncertainty represents the possible error in population estimates. Researchers typically measure it using confidence intervals, standard errors, and coefficients of variation (Buckland and Garthwaite, 1991). These measures help determine if the data quality supports strong conclusions. For instance, wide confidence intervals indicate low precision, suggesting a need for more data or improved methodology.

Proper uncertainty assessment has real-world importance. In conservation, healthcare, or policy-making, decisions based on uncertain data may lead to poor outcomes (Williams et al., 2002). Transparent reporting of uncertainty helps decision-makers understand the limitations of the data, leading to better-informed choices.

1.1.3 The Challenges of Sampling Effort

Sampling effort refers to the amount of resources, such as the number of capture occasions or the intensity of sampling activities, allocated toward capturing and recapturing individuals from a target population (Xi et al., 2008). The magnitude of sampling effort directly affects the precision of population size estimates, as it governs the amount of information available for analysis.

Designing an effective capture-recapture study involves a balancing resource constraints against the desired level of estimation precision (Robson and Regier, 1964; Thompson et al., 1998). Practical constrains such as limited funding, labour and time often restrict sampling intensity, potentially resulting in imprecise or biased population estimates. Generally, greater sampling effort leads to more precise estimates by reducing uncertainty. However, this relationship follows a law of diminishing returns. While initial increases in effort produce substantial gains in precision, further additions yield progressively smaller improvements (Kordjazi et al., 2016).

The acceptable level of uncertainty in a study directly influences the required sampling effort (Robson and Regier, 1964). Studies that can tolerate broader confidence intervals may proceed with fewer sampling occasions, conserving resources. Conversely, achieving highly precise estimates necessitates more sampling effort, increasing the demand for time, labour, and funding.

Determining sampling effort for a study is a self-starting process. The accuracy of estimates associated with a given sampling effort depends on the population size being

sampled. Since the study's goal is often to estimate this unknown population size, predicting accuracy in advance is difficult. Essentially, estimating the study's precision is like trying to determine the population size before the experiment, which makes it hard to define the necessary sampling efforts.

The solution to this dilemma can be found in a pilot study, or by making an informed judgement about the population size based on prior knowledge and experience. To avoid planning the experiment entirely in the dark, it is necessary to make educated guesses, at the very least about the order of magnitude of the population.The purpose of the experiment can then be regarded as the objective confirmation or refinement of these earlier estimates (Robson and Regier, 1964). Prior empirical work or expert knowledge can be incorporated into simulation-based methods for the purpose of determining the required sampling efforts (Bröder et al., 2020; Paterson et al., 2019)

1.2 Problem Statement

Although capture-recapture strategies are widely utilized, insufficient consideration is given to the level of sampling effort. Quantifying the necessary sampling effort for accurate estimations in a capture-recapture study is challenging, particularly when the sample size depends on multiple aspects, including the detectability of the animals, desired level of precision, and scope of the investigation (Schorr et al., 2014).

Choosing the right sample intensity and length to achieve a desired degree of precision is a difficult task that might introduce bias or impact estimations if there is insufficient data (Kordjazi et al., 2016). One of the most critical considerations to make when designing a capture-recapture technique is the quantity of resources to be put in the experiment in terms of personnel, time, and money. The higher the investment, the more accurate the population size estimate will be if resources are allocated properly (Robson and Regier, 1964). A good planning requires (1) marking enough animals to in-sure recapture in other populations, (2) sampling enough populations to detect the range of dispersal, and (3) sampling for a duration sufficient to allow movements among populations.

Previous researchers have emphasised the necessity of reporting measures of uncertainty alongside population size estimates (Borchers et al., 2002; Seber and Schofield, 2023), yet practical guidelines on balancing this uncertainty against available sampling resources remain limited. Without a clear framework for assessing the relationship between sampling effort, detectability, and uncertainty, researchers may resort to arbitrary or suboptimal study design. This gap in methodology is particularly critical in cases involving heterogeneity or elusive species, where capturabilities can vary considerably across sampling occasions (Pollock et al., 1990; Borchers et al., 2002).

Thus, this study addresses the need to develop and validate a structured framework for systematically determining optimal sampling effort allocation and explicitly assessing uncertainty in capture-recapture studies.

The research questions for this study include: -

1. How often should one plan for recapture occasions in capture-recapture method?
2. How precise is the population size estimation with the optimum design?

1.3 Aim and Objectives

This study aims to investigate how the sampling effort can be effectively determined in capture-recapture studies to improve the precision of population size estimates.

The specific objectives of this study includes:

1. To develop a theoretical framework for identifying the optimal allocation of sampling effort in capture-recapture studies.
2. To examine the application of a hierarchical structure within the Lincoln-Petersen design .
3. To determine the optimal sampling effort allocation for population estimation in Schnabel census.
4. To extend the hierarchical sampling effort framework to Schnabel census models.

1.4 Underlying Assumptions

The validity of the capture-recapture method relies heavily on a set of core assumptions underlying the data collection process (Gerritse et al., 2015).

1. The population is assumed to be closed, meaning no birth, death, immigration, or emigration occurs during the sampling period.
2. Captures are assumed to be independent events. Each individual must have the same probability of being captured on each occasion, regardless of past captures.
3. Homogeneous capture probability is assumed across all individuals, implying equal likelihood of detection.
4. No tag loss or misidentification occurs, ensuring that individuals can be accurately matched between capture events.

This thesis primarily considered models in which these assumptions are fully satisfied. However, to reflect more realistic conditions, some models introduced in later chapters

incorporate heterogeneity in capture probabilities, allowing for individual or time variation in detectability.

1.5 Report Organisation

Chapter One provides an an overview of capture-recapture method, highlights the importance of uncertainty assessment, and discusses the challenges in determining sampling effort for reliable and efficient study design. The chapter also presents the problem statement and outlines the aim and objectives set to be achieved. Additionally, the scope of the study and the underlying assumptions are defined in this chapter.

Chapter Two provides review of the existing literature and relevant studies related to the capture-recapture method. This includes an explanation of the basic concepts and assumptions of capture-recapture studies, as well as the estimation and analysis methods. The chapter also highlights the importance of sampling efforts in capture-recapture studies.

Chapter Three focuses on the Lincoln-Petersen experiment with a hierarchical design. This chapter proposes the sampling strategies for this type of study and demonstrates the impact of different levels of sampling effort on the accuracy and precision of population size estimates.

Chapter Four examines sampling efforts within the context of a Schnabel census, employing the framework of the zero-truncated binomial distribution. This chapter explores how the application of this specific method can enhance the accuracy and precision of population estimates through careful consideration of sampling efforts.

Chapter Five extends the hierarchical Lincoln-Petersen method to the hierarchical Schnabel census approach. It explores how different strategies for distributing effort across sub-occasions influence the precision of population size estimates.

Chapter Six summarises the main findings of the thesis, highlighting key insights on optimal sampling effort and uncertainty in capture-recapture studies. It also discusses limitations of the current work, and directions for future research.

1.5.1 Notation and Definition

The statistical methods for capture-recapture study in this thesis involves some parameters and statistics. In this section, some general notations and definitions are explained here so that readers could easily understand the context.

N	actual population size of the target population.
$\hat{N}$	estimator of population size.
T	total number of capture occasions.
n	number of individuals captured at least once during the study.
θ	individual capture probability on a single capture occasion.
f_x	frequency of capturing an individual for exactly x times.
f_0	frequency of unobserved individuals.
p_0	probability of non- detection across the study.
Y_{ij}	indicator variable of the i-th individual being identified on the j-th occasion,

$$Y_{ij} = \begin{cases} 1 & \text{if the } i\text{-th individual is captured on the } j\text{-th occasion.} \\ 0 & \text{otherwise.} \end{cases}$$

Chapter 2

Literature Reviews

2.1 Introduction

The capture-recapture method is a widely used technique for estimating the size of animal or human populations. Applying this method requires careful consideration of various factors, such as the sampling design, the assumptions, and the estimation and analysis methods. In this chapter, a comprehensive review of the existing literature and relevant studies on the capture-recapture method are presented. This chapter highlights the importance of sampling efforts in capture-recapture studies, and how they affect the accuracy and precision of the estimates. This chapter also describes the basic concepts and assumptions of capture-recapture studies, as well as the different estimation and analysis methods that have been developed over the years.

2.2 A Brief History of Capture-Recapture

Historically, capture-recapture method can be traced back to Graunt's use in estimating England's population in 1662 (Graunt, 1939; Cochran, 1978) and Laplace's similar work in France during 1783 (Laplace, 1786). Nevertheless, the formal conceptualization of the capture-recapture technique is attributed to Petersen (1896), who initially applied it to assess fish populations. Subsequently, Lincoln (1930) independently rediscovered the method to estimate waterfowl abundance in North America. Since then, capture-recapture techniques have widely been adopted to study animal abundance (Pollock et al., 1990; Hickey and Sollmann, 2018), survival probabilities (Lebreton et al., 1992; Reinke et al., 2020), mortality rates (Kordjazi et al., 2016; Jiménez-Ruiz et al., 2023), and migratory patterns (Matechou and Caron, 2017; Matechou and Argiento, 2023).

In human population research, an early significant use was by Sekar and Deming (1949), who employed capture-recapture methods to estimate birth and death rates as well as the completeness of civil registration systems. This study marked the emergence of capture-recapture as an important technique in public health and epidemiology, providing

an alternative to traditional prevalence research methods such as cross-sectional surveys and case counting. The strength of capture-recapture in epidemiological contexts lies in its ability to estimate undetected cases by combining data from independent registers. Initial epidemiological applications appeared in studies estimating the incidence of birth defects (Wittes and Sidel, 1968) and hospital infections (Lewis and Hassanein, 1969). Since then, its scope in health research has broadened considerably.

Capture-recapture methods are frequently employed to assess the prevalence and incidence of birth defects (Akkaya-Hocagil et al., 2017; Egeland et al., 1995), drug abuse (Plettinckx et al., 2021; Raag et al., 2019), infectious diseases (Rocchetti et al., 2020; Straetemans et al., 2020), cancer cases (Ghojazadeh et al., 2013; Plouvier et al., 2019), and dementia prevalence (Bailly et al., 2019; Sanderson et al., 2003). Beyond health sciences, the capture-recapture approach has demonstrated versatility in fields such as criminology (Tajuddin et al., 2022; Charette and van Koppen, 2016), information and communication technology (Lu and Li, 2010; Petersson et al., 2004), and linguistics (Williams et al., 2014; Alderete and Davies, 2019). The profound significance of the capture-recapture method in the domain of data analysis is substantiated by the recent emergence of extensive list of books, which provide comprehensive introductions and elaborate discussions pertaining to this methodology (Böhning et al., 2018; Borchers et al., 2002; McCrea and Morgan, 2015; Seber and Schofield, 2023).

2.3 Importance of Sampling Efforts

Sample size and sampling effort influence precision and accuracy of estimates. This was shown quite early on by Robson and Regier (1964), who investigated how the performance of the Lincoln-Petersen estimator is affected by varying sample sizes. Likewise, Otis et al. (1978) stressed that live-capture work needs to have a high number of unique animals caught as well as maximum numbers of recaptures. Applying 26 years of slider turtle capture-recapture data, Burke et al. (1995) examine the impact on accuracy in estimating meta-population size through sample size and study duration.

Reducing sample size or effort can result in reduced precision in estimates. According to McKelvey and Pearson (2001), 98 percent of samples in small animal populations were not enough to measure population density. Similarly, Howe et al. (2013) revealed that 15–25 traps were insufficient for calculating female black bear density. In reality, greater sample size is often desirable to get a more precise estimate, but that is not feasible at all times based on constraints in terms of logistics or finance. Kordjazi et al. (2016) illustrates that precision improves marginally with increased level of effort in sampling. This suggests that there is some point at which sampling effort and precision could both be optimised, where high-quality data can be collected at non-excessive costs in terms of effort and resources. At the end, it is a trade-off between desired precision on one hand and study feasibility on another (Conner et al., 2015).

Previous research has suggested minimum sampling requirements to minimize bias. Gaskell and George (1972) mentioned that less than ten recaptures tend to lead to bad estimation. Similarly, Seber (1982) and Krebs (2014) advised that for a Lincoln-Petersen design, product of sample sizes on both occasions must be higher than that for the population size with at least seven recaptures of marked animals. Robson and Regier (1964) presented an even more conservative rule, suggesting that $n_1 \times n_2 > 4N$ is needed to achieve minimum estimation bias under random sampling. Supporting this, Greenwood and Robinson (2006) demonstrated that the estimates in Lincoln-Petersen studies become biased when fewer than eight marked individuals are recaptured in the second sample. Xi et al. (2008) tested minimum proportion to be captured in order to achieve reliable estimates for population size under multiple capture-recapture models, including discrete-time models that can be used for the Schnabel census.

Despite these recommendations, sampling efficacy can be limited by biological, environmental, and logistical constraints. Catchability is dependent on individuals in capture-recapture experiments, which is subject to seasonal fluctuations, behaviour, or environmental factors (Kordjazi et al., 2016). Catchability in animals such as lobsters (Frusher et al., 2003), crabs (Williams and Hill, 1982), and crayfish (Somers and Stechey, 1986) is smaller in colder months. Physiological processes such as moulting and copulation may decrease chances of being caught, thus compromising precision and validity in estimates of survival rates (Kelly et al., 1999; Ziegler et al., 2004).

Capture probability is another primary determinant influencing sampling efficiency. Simulation experiments with different sample sizes (ranging from 1,000 to 100,000 marked items per year) by Schorr et al. (2014) have shown that making individuals more detectable improves the precision in estimates considerably. This is in agreement with Burnham et al. (1987), who emphasized that detectability is an important determinant for the required sample size. Furthermore, Papadatou et al. (2012) provided evidence that enhancing detectability combined with using contemporary analytical methodology reduces sample numbers while making estimates stronger.

2.4 Concept of Capture-Recapture

The fundamental concept of capture-recapture involves sampling or capturing individuals, marking them, releasing them back into the population, and then conducting a second survey to recapture them, count them, and mark them again. Subsequently, a tally is taken of the number of individuals recorded in the subsequent surveys that have been previously marked. The survey can be conducted for a total of T iterations. The number of individuals being identified from all the surveys is recorded as the capture-recapture histories. The observed frequencies derived from the capture-recapture histories are utilized to estimate the size of the targeted population, or the count of individuals that are either missing or have not been observed. In ecological studies, models are often

applied to data collected from several capture events. However, in research involving human populations, the data typically comes from multiple lists or repeated observations on a single list instead of multiple capture occasions (Bird and King, 2018).

In typical capture-recapture studies, each individual can be uniquely identified during every capture event. This allows researchers to create detailed capture histories for each individual, documenting when they were captured or missed throughout the study. Traditionally, unique identification is achieved by marking the individual during its first observation. New technologies enable researchers to uniquely identify individual, such as the use of DNA and motion sensor trap. Marking using these technologies has several benefits as they do not affect the subjects by the additional markings and such marking are generally not prone to being lost. However, the downside is that uniquely identifying individuals can become more challenging, potentially leading to increased uncertainty in the capture histories (King and McCrea, 2019).

2.4.1 Fundamental Assumptions

The results of the capture-recapture method are highly dependent on several assumptions underlying the data (Gerritse et al., 2015). These assumptions include:

- Closed population - The population must remain constant throughout the study period, with no immigration, emigration, births, or deaths. For example, if individuals appear in one sample but are not present in others, this will alter the capture probability across samples and lead to either an underestimation or overestimation of the population size N.
- Independence - Each capture event must be independent of the others. Positive dependence (trap-happy behaviour) increases recaptures and can underestimate population size, whereas negative dependence (trap-shy behaviour) decreases recaptures and can overestimate population size.
- Equal catchability - Every individual in the population should have an equal chance of being captured. If some individuals have a lower capture probability than others, N is likely to be underestimated.
- Reliable marks or tags - No marks or tags should be lost, ensuring that individuals can be accurately matched from initial capture to recapture. Missing true matches will falsely reduce the number of recaptured individuals, leading to an overestimation of N.

Violation of assumptions may affect the accuracy of the estimate. To address this issue, researchers have developed various variants of models that accommodate the violation of these assumptions (Gaskell and George, 1972; Skalski and Robson, 1982; Wolter, 1990; Seber et al., 2000).

2.5 Two-Occasion Model

The Lincoln-Petersen estimator, named after biologists F.C. Lincoln and C.J.G. Petersen, is a fundamental capture-recapture method that involves a single marking occasion followed by a single recapture event. This method is widely used to estimate population sizes in both animal and human populations, with the latter often referred to as the dual-list method. While the Lincoln-Petersen estimator is relatively simple, it is important to recognize that its results are heavily dependent on the assumptions outlined in Section 2.4.1. It is worth noting that capture probabilities can vary between occasions, with θ_1 representing the probability on the first occasion and θ_2 on the second.

An individual's capture history during the study can be denoted as (y_1, y_2), where y_j takes the value 0 if the individual was not observed and 1 if observed at sampling occasion j, for $j = 1, 2$. There are three possible observed capture history patterns:

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

The pattern $(1,1)$ means the individual was detected in both sampling events, $(1,0)$ indicates detection only in the first event, and $(0,1)$ represents detection only in the second. There is also an unobserved case, $(0,0)$, where the individual was never detected.

The data can be summarized in a contingency table, as shown in Table 2.1. The observed data consists of m, n_{10}, and n_{01}, representing the number of individuals with each specific combination of observations at the capture occasions. The count of individuals with the capture history $(0,0)$, denoted by n_{00}, is unknown. The total number of observed individuals is $n = m + n_{10} + n_{01}$, while the actual population size, which is the key parameter of interest, is given by $N = n_{00} + n$.

Table 2.1: Capture-recapture data with two occasions model.

		Occasion 2		
		1	0	
Occasion 1	1	m	n_{10}	n_1
	0	n_{01}	n_{00}	
		n_2		N

2.5.1 Lincoln-Petersen Estimator

Assuming a population of size N, a sample of $n_1 = n_{10} + m$ individuals was captured and tagged during the first capture occasion and then released. Consequently, the proportion of the population that is marked is given by:

$$\frac{n_1}{N}.$$

During the second capture occasion, n_2 individuals were captured, of which m were already tagged. The proportion of tagged individuals in this second sample is $\frac{m}{n_2}$. Assuming that this proportion matches the tagged proportion in the population, we have

$$\frac{m}{n_2} \approx \frac{n_1}{N},$$

which yields the estimate of the population size:

$$\hat{N}_{LP} = \frac{n_1 n_2}{m}. \tag{2.1}$$

The approximate variance of $\hat{N}_{LP}$ is given by:

$$\text{Var}(\hat{N}_{LP}) = \frac{n_1 n_2 (n_1 - m)(n_2 - m)}{m^3}.$$

Further explanation and derivation of this variance formula can be found in Section 3.4.

Various sampling models have been proposed to justify the Lincoln-Petersen estimator and to estimate its standard error. One commonly used model assumes a multinomial distribution, which presumes that each individual has an identical capture probability at each occasion (Chao and Hugginns, 2006). Under this multinomial model, the number of captures during each occasion is treated as a stochastic variable. In this framework, there are three distinct capture patterns (10, 01, and 11) and three parameters (N, π_1, and π_2), where π_1 and π_2 denote the detection probabilities for the initial and subsequent occasions.

It is also notable that the Lincoln-Petersen estimator serves as an approximate maximum likelihood estimator (MLE) for N in this context (Seber, 1982; Bailey, 1951). Applying large sample theory, where n_1, n_2, and m tend to infinity while maintaining constant ratios, the likelihood function is given by:

$$L = \binom{n_2}{m} \left(\frac{n_1}{N}\right)^m \left(\frac{N - n_1}{N}\right)^{n_2 - m}.$$

Therefore,

$$\ell = \log\binom{n_2}{m} + m\log(n_1) + (n_2 - m)\log(N - n_1) - n_2\log(N). \tag{2.2}$$

Differentiate Equation (2.2) with respect to N gives:

$$\frac{\partial \ell}{\partial N} = \frac{n_2 - m}{N - n_1} - \frac{n_2}{N}.$$

Hence the MLE of N is:

$$\hat{N} = \frac{n_1 n_2}{m}.$$

An alternative method uses the hypergeometric model, treating the sample sizes n_1 and n_2 as constants. This model involves only one parameter, N, and one stochastic variable, m. Assuming uniform catchability, the MLE for N is the integer value of the Lincoln-Petersen estimator (Chao and Hugginns, 2006).

2.5.2 Chapman Estimator

When $n_1 + n_2 \geq N$, Chapman (1951) proposed an unbiased estimator:

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

The rationale behind the effectiveness of this estimator in reducing bias lies in addressing the bias inherent in the Lincoln-Petersen estimator, which is particularly pronounced when m is small, especially in the scenario where $m = 0$. When $m = 0$, indicating that no individuals are captured twice, the estimate from Equation (2.1) becomes infinite. To address this issue, an additional individual is added to the count of those captured on both occasions, thereby ensuring that m is at least one. Consequently, the values of n_1, n_2, and N are each incremented by one.

Chapman's estimator is based on the insight that $\frac{(n_1+1)(n_2+1)}{m+1}$ provides a valid estimate for $N+1$. In cases where $n_1 + n_2 < N$, the bias of Chapman's estimator is approximately given by $-N \exp\left[-\frac{(n_1+1)(n_2+1)}{N}\right]$ (Chao and Hugginns, 2006). Additionally, Seber (1982) derived the variance for Chapman's estimator:

$$\mathrm{Var}(\hat{N}_C) = \frac{(n_1+1)(n_2+1)(n_1-m)(n_2-m)}{(m+1)^2(m+2)}.$$

2.6 Multiple Occasions Model

The extension of two-occasion capture-recapture models into multiple occasions was introduced by Schnabel (1938) and further refined by Darroch (1958). In this extended framework, individuals are captured, examined for tags, marked if untagged, and released across successive capture occasions. Each sampling event contributes to a cumulative dataset from which population size is inferred. The Schnabel method assumes a closed population and is widely used for estimating abundance when repeated sampling is possible (McCrea and Morgan, 2015).

Capture-recapture data in this context are typically recorded as individual capture histories, where each entry is a binary indicator variable. Let Y_{ij} denote whether individual i is captured ($Y_{ij} = 1$) or not ($Y_{ij} = 0$) on capture occasion j, with $i = 1, 2, \ldots, N$

and $j = 1, 2, \ldots, T$. These capture histories form the basis for estimating the total population size. The total number of captures for individual i is defined as $Y_i = \sum_{j=1}^{T} Y_{ij}$. Individuals with $Y_i = 0$ are never observed and therefore absent from the data. The observed distribution is considered zero-truncated. Estimating the number of unobserved individuals requires statistical models that account for this truncation.

The summarized datasets in Table 2.3 and Table 2.4 (presented later on page 23) illustrate the count distribution from the Schnabel census. In particular, Table 2.4 demonstrates the zero-truncated characteristic of capture-recapture data, where only individuals detected at least once are included, while those never observed are omitted.

2.7 Likelihood Analysis

Since the late 1960s, the standard approach for analysing capture-recapture data has involved using explicitly defined probability models such as Binomial and Poisson distributions, with mathematically specified likelihood functions, together with MLE to estimate the unknown parameters (Manly et al., 2005).

2.7.1 Likelihood Based on Binomial Distribution

When assuming homogeneous catchability throughout the study, the number of captures X for a given individual across T capture occasions can be modelled using a binomial distribution. The probability mass function (pmf) is given by:

$$p_x = P(X = x) = \binom{T}{x} \theta^x (1-\theta)^{T-x}, \quad x = 0, 1, 2, \ldots, T.$$

Here, $p_0 = (1-\theta)^T$, T represents the number of trapping occasions, and θ is the capture probability of each individual at each trapping occasion. Let f_x denote the number of individuals captured exactly x times, for $x = 1, 2, \ldots, T$. Since only individuals with $X > 0$ are observed, parameter estimation is carried out by maximising the likelihood function based on the zero-truncated Binomial distribution:

$$L(\theta) = \prod_{x=1}^{T} \left[\frac{1}{1-(1-\theta)^T} \binom{T}{x} \theta^x (1-\theta)^{T-x} \right]^{f_x},$$

with respect to θ.

2.7.2 Likelihood Based on Poisson Distribution

Assuming the capture count X follows a Poisson distribution, the pmf is:

$$p_x = P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

Here, $p_0 = e^{-\lambda}$. Parameter estimation is performed by maximising the likelihood function based on the zero-truncated Poisson density:

$$L(\lambda) = \prod_{x=1}^{T} \left[\frac{\lambda^x}{x!(e^{-\lambda} - 1)} \right]^{f_x},$$

with respect to λ.

2.7.3 Likelihood Based on Beta-Binomial Distribution

Assuming the capture count X follows a Binomial distribution

$$X \sim \text{Binomial}(T, \theta),$$

while the capture probability θ follows a Beta distribution

$$\theta \sim \text{Beta.}(\alpha, \beta).$$

The pmf is

$$P(X = x) = \binom{T}{x} \frac{B(\alpha + x, \beta + T - x)}{B(\alpha, \beta)}, \quad x = 0, 1, 2, \ldots, T,$$

where $B(\cdot, \cdot)$ denotes the Beta function. The likelihood function based on the zero-truncated Beta-Binomial density may be written in the form of:

$$L(\alpha, \beta) = \prod_{x=1}^{T} \left[\binom{T}{x} \frac{B(\alpha + x, \beta + T - x)}{B(\alpha, \beta) - B(\alpha, \beta + T)} \right]^{f_x}.$$

2.7.4 Likelihood Based on Binomial mixture Distribution

Assume the capture count X follows a Binomial mixture distribution with k components of the Binomial distribution, the pmf is:

$$P(X = x) = \sum_{j=1}^{k} w_j \, \text{Bino}(x|T, \theta_j), \quad x = 0, 1, 2, \ldots, T,$$

where

$$\text{Bino}(x|T, \theta_j) = \binom{T}{x} \theta_j^x (1 - \theta_j)^{T-x}, \quad j = 1, 2, \ldots, k,$$

with $\sum_{j=1}^{k} w_j = 1$. The likelihood based on the zero-truncated Binomial mixture model may be written as:

$$L(\boldsymbol{\theta}, \mathbf{w}) = \prod_{x=1}^{T} \left[\frac{\sum_{j=1}^{k} w_j \binom{T}{x} \theta_j^x (1 - \theta_j)^{T-x}}{1 - \sum_{j=1}^{k} w_j (1 - \theta_j)^T} \right]^{f_x},$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ and $\mathbf{w} = (w_1, \ldots, w_k)$ are the component-specific parameters and weights, respectively. Since both $\boldsymbol{\theta}$ and $\mathbf{w}$ are unknown, the likelihood function must be maximised jointly with respect to both sets of parameters, subject to the constraint $\sum_{j=1}^{k} w_j = 1$.

2.7.5 Model Selection

Choosing an appropriate model for a dataset requires balancing goodness-of-fit against complexity. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) provide quantitative measures for this comparison. The formulas are given by:

$$\text{AIC} = -2 \log L + 2K,$$

$$\text{BIC} = -2 \log L + K \log n,$$

where L is the model's maximum likelihood, K the parameter count, and n the number of observations. The model with lowest AIC or BIC values represents the optimal trade-off between fit and simplicity.

2.8 Estimating Population Size

While two-occasion estimation is straightforward (see Equation (2.1)), analysing Schnabel census data (≥ 3 occasions) requires more sophisticated approaches. MLE is typically employed, where numerical optimization identifies parameter estimates that maximize the likelihood function. These estimates then inform the Horvitz-Thompson estimator to derive population size.

2.8.1 Horvitz-Thompson Estimator

Let p_0 denote the probability of an individual remaining undetected. Assuming homogeneous detection probabilities, the population size N decomposes into observed n, and unobserved Np_0 components:

$$N = n + Np_0.$$

Rearranging yields the Horvitz-Thompson estimator:

$$\hat{N} = \frac{n}{1 - \hat{p}_0}.$$

Two approaches exist for estimating p_0 under homogeneity: the Good-Turing estimator and EM algorithm-based MLE.

2.8.2 Good-Turing Estimator

Developed by Good (1953), this method estimates undetected individuals using capture frequencies. Let f_x denote the number of individuals observed exactly x times across T capture occasions. Then, the total number of observed individuals is $n = \sum_{x=1}^{T} f_x$, and the total number of captures is $C = \sum_{x=1}^{T} x f_x$.Based on these quantities, the Good–Turing approach provides an estimate of p_0 that an individual was never captured.

For example, the Binomial model estimates

$$\hat{p}_0 = \left(\frac{f_1}{C}\right)^{T/(T-1)},$$

yielding the population estimator:

$$\hat{N}_{GT} = \frac{n}{1 - (f_1/C)^{T/(T-1)}}.$$

For large values of T, the approximation $\dfrac{T}{(T-1)} \approx 1$ holds, simplifying the estimator to:

$$\hat{N}_{GT} = \frac{n}{1 - (f_1/C)}.$$

Under Poisson model, also simplifies to $\hat{p}_0 = f_1/C$. As noted by Böhning et al. (2018), this approach proves particularly useful for large samples.

2.8.3 Expectation-Maximisation Algorithm

The EM algorithm is an iterative computational technique commonly used to estimate parameters when data are incomplete or partially observed. The method assumes that individual capture counts follow a discrete distribution, and only those with $X > 0$ are observed. The capture events are considered independent, and individuals are assumed to have identically distributed capture probabilities unless stated otherwise.

Let f_x be the number of individuals captured exactly x times, for $x = 1, 2, \ldots, T$, and let $n = \sum_{x=1}^{T} f_x$. Since f_0, the count of individuals with zero captures, is unknown, it is treated as missing data. The complete-data likelihood is:

$$L_{\text{complete}}(\theta) = \prod_{x=0}^{T} p_x(\theta)^{f_x},$$

with corresponding log-likelihood:

$$\ell_{\text{complete}} = \sum_{x=0}^{T} f_x \log p_x(\theta).$$

In the E-step, the conditional expectation of f_0 given the observed frequencies is computed using the current parameter $\hat{\theta}$:

$$\hat{f}_0 = n \cdot \frac{p_0(\hat{\theta})}{1 - p_0(\hat{\theta})}.$$

This conditional expected $\hat{f}_0$ is then substituted into the complete-data log-likelihood.

In the M-step, the parameter estimate $\hat{\theta}$ is updated by maximizing the completed log-likelihood ℓ_{complete} with respect to θ. The EM steps are iterated until convergence. The final $\hat{\theta}$ is used to estimate p_0, which is then incorporated into the Horvitz-Thompson estimator to obtain $\hat{N}$.

2.9 Numerical Optimization Methods

Parameter estimation in capture-recapture model is usually done by maximising complex likelihoods for which closed-form solutions cannot be determined. Two widely used numerical approaches are employed in this thesis, the Newton-Raphson method, which is a traditional approach to utilising both first- and second-order derivatives, and the quasi-Newton methods, which use gradient evaluations to approximate second-order information in an efficient manner, suitable for large-scale computational effort.

2.9.1 Newton-Raphson Method

The Newton–Raphson method is a well-established technique in numerical optimization, frequently used to find the maximum or minimum of a differentiable function. It operates by iteratively updating parameter values to locate a point at which the gradient of the function equals zero, which is the function's root or stationary point. The updated parameter is obtained using:

$$\boldsymbol{\theta}^{(r+1)} = \boldsymbol{\theta}^{(r)} - \left[\mathbf{H}\left(\boldsymbol{\theta}^{(r)}\right)\right]^{-1} \nabla f\left(\boldsymbol{\theta}^{(r)}\right).$$

In this expression, $\boldsymbol{\theta}^{(r)}$ represents the estimate at the r-th iteration, $\nabla f(\boldsymbol{\theta}^{(r)})$ is the gradient (first derivative), and $\mathbf{H}\left(\boldsymbol{\theta}^{(r)}\right)$ denotes the Hessian matrix (second derivative) of the considered function at that point.

One of the main advantages of the Newton–Raphson method is its quadratic rate of convergence under regularity conditions and with a good initial guess, it can converge very rapidly to the true optimum (Ypma, 1995). This efficiency makes it popular in many statistical contexts, particularly in maximum likelihood estimation.

However, the method is not without challenges. Its performance depends critically on the accuracy of the gradient and Hessian. In practice, the Hessian matrix may be difficult to compute or invert, especially in high-dimensional settings or with poorly conditioned

problems. These limitations can lead to instability or slow convergence, particularly if the starting value is far from the optimum or the likelihood surface is flat or irregular (Lange, 2010).

Despite these issues, the Newton–Raphson method remains a powerful and elegant tool when applied under appropriate conditions. Its speed and theoretical properties make it valuable for many small- to medium-scale problems, though alternative methods may be preferred for large-scale or non-convex optimizations (Lange, 2010).

2.9.2 Quasi-Newton Method

Quasi-Newton methods are a family of optimization algorithms that offer a computationally efficient alternative to the Newton-Raphson method. Rather than calculating the exact Hessian matrix of second derivatives at each iteration, quasi-Newton methods build up an approximation to the inverse Hessian using only the gradient information. This greatly reduces the computational burden, especially in high-dimensional problems where evaluating or inverting the Hessian can be costly or unstable.

A widely used member of this family is the BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm, which updates the inverse Hessian approximation using the following rule,

$$\mathbf{B}_{n+1}^{-1} = \left(\mathbf{I} - \frac{\mathbf{s}_n \mathbf{y}_n^\top}{\mathbf{y}_n^\top \mathbf{s}_n}\right) \mathbf{B}_n^{-1} \left(\mathbf{I} - \frac{\mathbf{y}_n \mathbf{s}_n^\top}{\mathbf{y}_n^\top \mathbf{s}_n}\right) + \frac{\mathbf{s}_n \mathbf{s}_n^\top}{\mathbf{y}_n^\top \mathbf{s}_n},$$

where
$\mathbf{s}_n = \boldsymbol{\theta}_{n+1} - \boldsymbol{\theta}_n$ is the change in parameter estimates,
$\mathbf{y}_n = \nabla f(\boldsymbol{\theta}_{n+1}) - \nabla f(\boldsymbol{\theta}_n)$ is the change in gradients,
$\mathbf{B}_n^{-1}$ is the approximation to the inverse Hessian matrix at iteration n.

In this thesis, the optimization algorithm applied is the L-BFGS-B (Limited-memory BFGS with Box constraints), which is a modification of BFGS designed for large-scale problems with boundary constraints on parameters. The "limited-memory" approach avoids storing the full matrix $\mathbf{B}_n^{-1}$ by keeping just a small number of recent update vectors $\mathbf{s}_n$ and $\mathbf{y}_n$, thus making the algorithm computationally efficient in high dimensions (Byrd et al., 1995).The approach further enables easy introduction of simple Box constraints (lower and upper bounds) on parameters, an advantage in applied statistical models to keep estimates within relevant or interpretable bounds.

2.10 Uncertainty Assessment Methods

Uncertainty is a good indicator in determining the optimal sampling efforts for a capture-recapture study, as it reflects the precision and accuracy of the population estimate.

Several approaches can be considered to estimate the variances of the estimate, such as using bootstrap methods, and using conditional moment.

2.10.1 Conditional Moments

Considering the population size estimator $\hat{N} = n + \hat{f}_0$, the variance of this estimator originates from two factors: one associated with the random variable n and the other one from the estimator $\hat{f}_0$ (Böhning et al., 2018). Leveraging the Law of Total Variances, a straightforward formula for the variance of the population size estimator is provided as follows:

$$\text{Var}(\hat{N}) = \text{Var}_n\{\mathbb{E}(\hat{N}|n)\} + \mathbb{E}_n\{\text{Var}(\hat{N}|n)\},$$

where $\mathbb{E}_n$ and Var_n refer to the first and second moments of the marginal distribution of n.

2.10.2 The Delta Method

The delta method is a standard statistical tool for approximating the variance of an estimator that cannot be expressed as a simple sum of observations (Hosmer et al., 2008). At its core, the method linearises a non-linear function using a first-order Taylor series expansion around the mean of the underlying random variable. For the delta method to apply, the function must be continuously differentiable without sharp discontinuities.

Consider a smooth function $f(X)$ of a random variable X. The first-order Taylor approximation near the mean μ of X is:

$$f(X) \approx f(\mu) + (X - \mu)f'(\mu),$$

where $f'(\mu)$ is the derivative of f evaluated at μ. Using this approximation, the variance of $f(X)$ simplifies to:

$$\begin{aligned}\text{Var}\left[f(X)\right] &\approx \text{Var}(X - \mu)\left[f'(\mu)\right]^2 \\ &\approx \sigma^2\left[f'(\mu)\right]^2,\end{aligned}$$

with σ^2 denoting the variance of X. The delta method estimator replaces μ and σ^2 with their sample counterparts, yielding:

$$\widehat{\text{Var}}\left[f(X)\right] \approx \hat{\sigma}^2\left[f'(\hat{\mu})\right]^2.$$

2.10.3 Bootstrap Method

Application of the bootstrap or re-sampling approach combined with robust percentile interval has been studied by Buckland and Garthwaite (1991) for purpose of variance

estimation as well as for construction of confidence intervals for estimation in populations. Bootstrap techniques have an easy implementation irrespective of the particular model being analysed. The bootstrap algorithm for use in variance estimation is as below.

The bootstrap method is designed based on the assumption that the capture-recapture history can be defined by multinomial likelihood with $T+2$ parameters $(N, p_0, p_1, p_2, \cdots, p_T)$ (Anan et al., 2017). The pmf is given by:

$$\binom{N}{f_0, f_1, f_2, \cdots, f_T} p_0^{f_0} p_1^{f_1} p_2^{f_2} \cdots p_T^{f_T}.$$

For simulation studies where the true population size N and the full probability model $(p_0, p_1, \ldots, p_T)$ are assumed known, a parametric bootstrap can be used for estimating the variance of the population size estimator. Each individual is drawn from the multinomial distribution with full probability model $(N, p_0, p_1, p_2, \cdots, p_T)$, and the parameters are estimated. The bootstrap algorithm for variance estimation is as follow:

Bootstrap Algorithm for Variance Estimation of Population Size

1. Use $\hat{N}$ as described in the previous sections, or use the true N if it's available. This provides an estimate of f_0.

2. The estimates of capture-recapture probabilities $\hat{\mathbf{p}}$ are obtained using the relative frequencies:
$$\hat{\mathbf{p}} = \left(\frac{f_0}{\hat{N}}, \frac{f_1}{\hat{N}}, \frac{f_2}{\hat{N}}, \cdots, \frac{f_T}{\hat{N}}\right).$$

3. Re-sample N or $\hat{N}$ capture histories under multinomial distribution with parameter estimates $(\hat{N}, \hat{\mathbf{p}})$, and count the associated frequencies (f^*). This can be done using the function `rmultinorm()` in `R`.

4. For each sample generated, estimate $\hat{N}$. In the case when true N is available, the true f_0 is ignored. If $\hat{N}$ is not an integer, Buckland and Garthwaite (1991) suggest to round it to the nearest integer value.

5. Repeat step 2 and step 3 for B times and compute the following statistics:
a) mean of population size
$$\hat{\mathbb{E}}(\hat{N}) = \frac{1}{B}\sum_{b=1}^{B} \hat{N}^{(b)}.$$
b) bootstrap variance of population size
$$\widehat{\mathrm{Var}}(\hat{N}) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{N}^{(b)} - \hat{\mathbb{E}}(\hat{N})\right)^2.$$

2.11 Application of Capture-Recapture

This section presents several datasets derived from capture-recapture studies. These real examples illustrate the structure of capture-recapture data and provide insight into the sampling efforts involved. Examining these datasets offers understanding of how data is collected and organized in capture-recapture studies.

2.11.1 Alzheimer's Disease in Alpes-Maritimes

Bailly et al. (2019) used two data sources to estimate the incidence of Alzheimer's disease in France in the department of Alpes-Maritimes: the French National Alzheimer Database (BNA) and the health insurance cohort (HIC). In 2010 or 2011, residents who visited a specialist in memory clinics, were exempt from co-payment for Alzheimer's, or were prescribed medication for Alzheimer's were included in the study.

TABLE 2.2: Contingency table of Alzheimer's disease cases identified by the French National Alzheimer Database (BNA) and the Health Insurance Cohort (HIC) during 2010–2011. n_{00} represents the number of cases missed by both sources and is unobserved.

		HIC		
		1	0	
BNA	1	856	1738	2594
	0	2968	n_{00}	
		3824		N

An overall total of $n = 5562$ was detected, 3824 from HIC, 2594 from BNA, and 856 reported in common. The overlap between sources is indicated in the contingency table in Table 2.2. In the table, n_{00} denotes the number of individuals missed by both sources and is not directly observed. Under an assumption of independence between systems, Lincoln-Petersen estimator yielded a population size estimate of $\hat{N}_{LP} = 11588(SE = 285.621)$, while the Chapman estimator produced $\hat{N}_C = 11581(SE = 285.047)$.

2.11.2 Golf Tees in St. Andrew

Table 2.3 presents data from a controlled capture-recapture experiment conducted in St. Andrews, Scotland, and described in Borchers et al. (2002). In this study, a total of 250 golf tee clusters were distributed within a fixed area, simulating a population of stationary individuals. The goal was to evaluate how well capture-recapture models could estimate population size under known conditions.

Eight observers independently searched the area once each, mimicking $T = 8$ capture occasions. A cluster was considered "captured" if it was detected during a particular search. Table 2.3 shows the frequency distribution of clusters by the number of times they were detected. Of the 250 golf tee clusters, 162 were detected at least once, while 88 remained undetected ($f_0 = 88$).

TABLE 2.3: Number of identifications per golf tee for the capture-recapture experiment of recovering 250 golf tees in St. Andrews.

Number of Identifications (x)	**Frequency** (f_x)
0	88
1	46
2	28
3	21
4	13
5	23
6	14
7	6
8	11

2.11.3 Snowshoe Hare in Alaska

The snowshoe hare data were originally collected by Cushwa and Burnham (1974) in 1972, north of Fairbanks, Alaska, and later presented by Otis et al. (1978). Over $T = 6$ trapping occasions, $n = 68$ distinct hares were captured. The number of captures for each of the six occasions was 16, 28, 20, 26, 23, and 32, respectively. The capture count distribution is summarized in Table 2.4, which details the frequency of how many times a hare was captured across the six trapping occasions. For instance, $f_1 = 25$ hares were captured exactly once, while $f_2 = 22$ hares were captured exactly twice, and so on. However, since not all hares in the population were captured, the list is incomplete, and the number of hares that were never captured, f_0, remains unknown.

TABLE 2.4: Frequency distribution of snowshoe hares by number of captures over $T = 6$ trapping occasions.

Number of Captures (x)	**Frequency** (f_x)
1	25
2	22
3	13
4	5
5	1
6	2

2.11.4 Cottontail Rabbits in Tennessee

The study by McWherter (1991) aims to estimate the density of cottontail rabbits in a 25-hectare area at TVA's Land Between the Lakes, Stewart County, Tennessee. Live trapping was conducted over three distinct periods: March, August, and December 1985. Each period consisted of 15 consecutive days, resulting in a total of $T = 45$ trapping occasions. During these periods, $n = 114$ rabbits were captured across 241 trap events. Specifically, the number of individual rabbits captured was 40 in March, 46 in August,

and 66 in December. Table 2.5 summarises the capture data from the three 15-day trapping sessions.

Table 2.5: Capture data for cottontail rabbits across three 15-Day trapping sessions.

15-day trapping month	Rabbits captured	New rabbits	Recaptures*	Total captures
March	40	40	19	59
August	46	37	18	64
December	66	37	52	118
Total	152	114	89	241

* Number of recaptures within 15-day trapping period.

2.11.5 Cottontail Rabbits in Ohio

The study by Edwards and Eberhardt (1967), later republished by Chao (1987), involved an investigation on a restricted population of known size using live-trapping techniques. The research was conducted in a 4-acre rabbit-proof enclosure at the Olentangy Wildlife Experiment Station, Ohio, in October 1961.

Table 2.6: Frequency of capture counts for cottontail rabbits across 18 trapping occasions. Captures beyond 7 omitted due to zero counts.

Number of Captures (x)	**Frequency (f_x)**
1	43
2	16
3	8
4	6
5	0
6	2
7	1

$N = 135$ wild cottontail rabbits were released into the enclosure. Over $T = 18$ consecutive nights, $n = 76$ of these rabbits were captured. Specifically, $f_1 = 43$ rabbits were captured once, $f_2 = 16$ twice, $f_3 = 8$ three times, $f_4 = 6$ four times, $f_6 = 2$ six times, and $f_7 = 1$ seven times, totaling 142 captures. The known true population size allowed for the calculation of uncaptured rabbits as $f_0 = 135$ - $76 = 59$. The capture data is summarized in Table 2.6.

2.11.6 Bangkok Heroin Users

The study by Lanumteang (2011) investigated patterns of heroin user contacts in Bangkok, Thailand, during the year 2001. Data were collected by the Office of the Narcotics Control Board (ONCB), in collaboration with the Drug Abuse Prevention and Treatment Division,

the Health Department, and the Medical Service Department of the Bangkok Metropolitan Administration. These data were drawn from 61 private and public treatment centres across the Bangkok metropolitan area.

The dataset, summarized in Table 2.7, comprises records from a total of $n = 5515$ heroin users. Each user's treatment history is represented by the number of treatment episodes they experienced during the year, ranging from 1 to 11. Specifically, $f_1 = 3137$ individuals were observed only once, while $f_2 = 1129$ were observed twice. The remainder were recorded with between three and eleven treatment episodes. The number of unobserved or hidden drug users, denoted as f_0, remains unknown. In this dataset, the number of capture occasions, T, is determined by the maximum number of recorded treatment episodes. Accordingly, $T = 11$ represents the highest count of episodes observed in the dataset.

TABLE 2.7: Frequency distribution of heroin users in Bangkok (2001) by number of treatment episodes.

Number of Episodes (x)	**Frequency** (f_x)
1	3137
2	1129
3	528
4	314
5	185
6	127
7	76
8	12
9	6
10	0
11	1

Chapter 3

Sampling Effort in Hierarchical Lincoln-Petersen Models

3.1 Introduction

This chapter focuses on how to optimise sampling effort in hierarchical Lincoln-Petersen models. It begins by introducing the key concepts, including the estimator of the population size and how the variance of the estimate is calculated.

The chapter then explains how sampling effort can be structured within a hierarchical design. The chapter first explores scenarios where the capture probabilities are fixed and known. Two common situations are considered: when capture probabilities are equal and when they are proportional across occasions.

The second part of the chapter deals with more complex settings where capture probabilities are not known. A pseudo-Bayesian approach is introduced to handle this uncertainty. The final sections apply the hierarchical design to time-window data.

3.2 Hierarchical Structure in Lincoln-Petersen Model

Lincoln-Petersen method is widely used to estimate population size in both animal and human population. While Lincoln-Petersen is relatively simple, inefficient designs risk underestimating the population size, especially of the elusive or rare species, which exhibit patchy distributions and low detectability, resulting in datasets dominated by non-detections (Thompson, 2004). In epidemiology, zero-inflated data from incomplete registries obscure true disease prevalence. The under-reporting of COVID-19 cases, for instance, led to inaccurate fatality rate estimates and impaired response efforts. (Böhning et al., 2020).

Hence, it is often practical to employ multi-stage sampling strategies in a capture-recapture study. Each stage involves sampling within a specific time window, which can

be continuous or discrete. For example, a researcher may conduct repeated trapping sessions over several nights during the first capture occasion, followed by a pause before conducting additional nights of trapping during the subsequent capture occasion. Two capture occasions in a Lincoln-Petersen model can be hierarchically levelled into a number of sub-occasions which totals a number T, as illustrated in Figure 3.1. Traditional capture-recapture method treat capture occasions as single stages, ignoring the hierarchical potential to subdivide effort into sub-occasions that adaptively boost detectability. By framing sampling effort as a divisible resource, this hierarchical approach addresses the 'too few recaptures' issues in studies of rare or elusive populations.

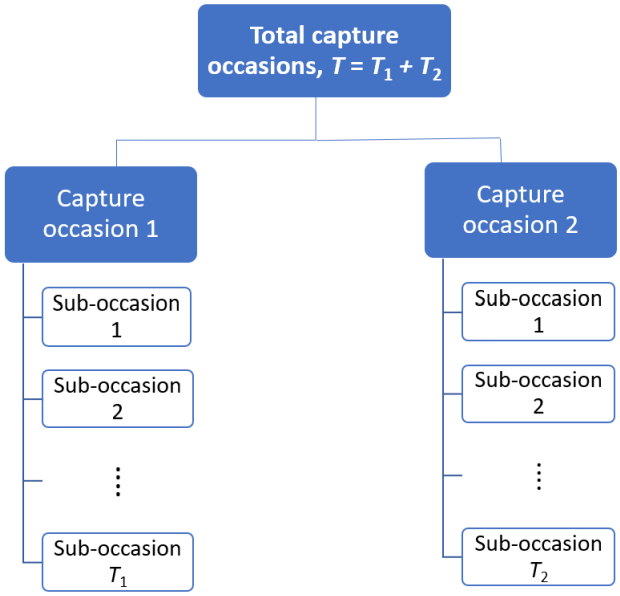


FIGURE 3.1: A hierarchical structure for a Lincoln-Petersen model.

A key limitation in the previous studies is the assumption that sampling effort is evenly distributed across capture occasions, regardless of changes in detectability. They did not explore how effort could be strategically shifted between sub-occasions to maximise detection efficiency. This issue is particularly relevant for rare or spatially scattered populations, where zero-inflated data and logistical fieldwork constraints make efficient sampling strategies essential. Current methods fail to provide a systematic way to optimize effort allocation in hierarchical designs, leaving a gap in the field.

3.3 Lincoln-Petersen Estimator

Let N represents the population size, n_1 the count of subjects captured and tagged on the first occasion, n_2 the count of subjects captured on the second occasion, and m the count of subjects identified on both capture occasions. Observations from a Lincoln-Petersen model may be presented in a table like in Table 3.1. In the table, n_{00} denotes the number of individuals missed by both sources and is not directly observed.

TABLE 3.1: Observed frequencies from a Lincoln-Petersen model.

		Occasion 2		
		1	0	
Occasion 1	1	m	$n_1 - m$	n_1
	0	$n_2 - m$	n_{00}	
		n_2		N

Let π_i denotes the probability of detecting a subject at occasion i $(i = 1, 2)$. Table 3.2 presents the join distribution of a Lincoln-Petersen model, assuming the occasions are independent.

TABLE 3.2: Joint distribution of identifying a subject in a Lincoln-Petersen model.

		Occasion 2	
		1	0
Occasion 1	1	$\pi_1\pi_2$	$(1-\pi_1)\pi_2$
	0	$\pi_1(1-\pi_2)$	$(1-\pi_1)(1-\pi_2)$

The frequency of the missing subjects, n_{00}, is unknown. Hence, the population size $N = n_1 + n_2 - m + n_{00}$ remains unknown and become the target of the inference. In this context, the population size can be estimated using the Lincoln-Petersen estimator:

$$\hat{N} = \frac{n_1 n_2}{m}.$$

3.4 Variance of Lincoln-Petersen Estimator

Let the random variable M represent the number of individuals captured in both the first and second occasions, with respective detection probabilities π_1 and π_2. Under a hypergeometric model, the pmf of M is given by:

$$P(M = m) = \frac{\binom{n_1}{m}\binom{N - n_1}{n_2 - m}}{\binom{N}{n_2}},$$

where n_1 and n_2 are the numbers of individuals captured during the first and second occasions, respectively (as defined in Table 3.1). In what follows these counts are assumed to be fixed. Since individuals not captured in either occasion remain unobserved, the aim is to estimate the total population size N.

Under the hypergeometric model, the expected value and variance of M are:

$$\mathbb{E}(M) = \frac{n_1 n_2}{N} = N\pi_1\pi_2,$$

and

$$\mathrm{Var}(M) = n_2 \left(\frac{n_1}{N}\right)\left(\frac{N-n_1}{N}\right)\left(\frac{N-n_2}{N-1}\right),$$

respectively.

For large N, the approximation $N-1 \approx N$ is applied. Defining the fractions $\pi_1 = n_1/N$ and $\pi_2 = n_2/N$, the variance simplifies to:

$$\begin{aligned}\mathrm{Var}(M) &\approx \left(\frac{n_1 n_2}{N}\right)\left(1-\frac{n_1}{N}\right)\left(\frac{N-n_2}{N}\right)\\ &= N\pi_1\pi_2(1-\pi_1)(1-\pi_2).\end{aligned}$$

The Lincoln-Petersen estimator for the population size is given by:

$$\hat{N} = \frac{n_1 n_2}{M}. \tag{3.1}$$

Its variance can be approximated as:

$$\mathrm{Var}(\hat{N}) \approx n_1^2\ n_2^2\ \mathrm{Var}(\frac{1}{M}). \tag{3.2}$$

The delta method (Sekar and Deming, 1949) is used to approximate the variance of the reciprocal of M. This method relies on a first-order Taylor expansion of a smooth function around its expected value. For a function $f(X)$, this expansion is:

$$f(X) \approx f(\mu) + (X-\mu)f'(\mu),$$

where $\mu = \mathbb{E}(X)$ and $f'(\mu)$ is the derivative of f evaluated at μ. The variance is then approximated by:

$$\mathrm{Var}\left(f(X)\right) \approx \mathrm{Var}(X)\cdot[f'(\mu)]^2.$$

Applying this to $f(M) = 1/M$ gives:

$$\mathrm{Var}\left(\frac{1}{M}\right) \approx \left[-\frac{1}{\mathbb{E}(M)^2}\right]^2 \mathrm{Var}(M).$$

Substituting into the variance of $\hat{N}$ in (3.2) yields:

$$\begin{aligned}\mathrm{Var}(\hat{N}) &\approx n_1^2\, n_2^2 \left[\frac{1}{\mathbb{E}(M)^4}\right] \mathrm{Var}(M) \\ &\approx n_1^2\, n_2^2 \left[\frac{1}{\mathbb{E}(M)^4}\right] N\pi_1\pi_2(1-\pi_1)(1-\pi_2).\end{aligned}$$

To express this variance in observable quantities, the detection probabilities are estimated using:

$$\hat{\pi}_1 = \frac{n_1}{\hat{N}} = \frac{M}{n_2}, \quad \hat{\pi}_2 = \frac{n_2}{\hat{N}} = \frac{M}{n_1}.$$

The expected value $\mathbb{E}(M)$ and population size N are approximated by their respective observed values m and $\hat{N}$. This leads to the following variance estimator:

$$\begin{aligned}\widehat{\mathrm{Var}}(\hat{N}) &\approx n_1^2\, n_2^2 \left(\frac{1}{m^4}\right) \hat{N}\, \hat{\pi}_1\, \hat{\pi}_2\, (1-\hat{\pi}_1)\, (1-\hat{\pi}_2) \\ &\approx n_1^2\, n_2^2 \left(\frac{1}{m^4}\right) \hat{N} \left(\frac{m}{n_2}\right)\left(\frac{m}{n_1}\right)\left(1-\frac{m}{n_2}\right)\left(1-\frac{m}{n_1}\right) \\ &\approx n_1^2\, n_2^2 \left(\frac{1}{m^4}\right) \hat{N} \left(\frac{m}{n_2}\right)\left(\frac{m}{n_1}\right)\left(\frac{n_2-m}{n_2}\right)\left(\frac{n_1-m}{n_1}\right) \\ &\approx \hat{N}\, \frac{n_2-m}{m}\, \frac{n_1-m}{m}. \end{aligned} \tag{3.3}$$

This derivation shows that the variance of $\hat{N}$ decreases as the overlap m increases, particularly when m approaches n_1 and n_2. Given that $\mathbb{E}(M) = N\pi_1\pi_2$, improving the reliability of the Lincoln-Petersen estimator involves increasing the product of detection probabilities $\pi_1\pi_2$.

3.5 Optimising Sampling Effort

Since $\mathbb{E}(M) = N\pi_1\pi_2$, it is apparent that on average, the value of M will not increase unless either π_1 or π_2 increases. Increasing the target population size N is generally impractical in most field settings. However, certain studies incorporate repeated sampling within a single capture occasion. For instance, in live-trapping surveys, animals are often trapped over multiple nights within the same occasion.

Consider a scenario in which capture occasion 1 consists of T_1 repeated identification efforts, and capture occasion 2 consists of T_2 repetitions. The probability of not detecting an individual during capture occasion 1 is $1-\pi_1 = (1-\theta_1)^{T_1}$, where θ_1 is the per-sub-occasion capture probability. Similarly, the probability of not detecting an individual during capture occasion 2 is $1-\pi_2 = (1-\theta_2)^{T_2}$. Consequently, the marginal detection

probabilities are

$$\pi_1 = 1 - (1 - \theta_1)^{T_1},$$
$$\pi_2 = 1 - (1 - \theta_2)^{T_2}.$$

To maximise the expected number of jointly captured individuals, m, it is necessary to maximise the product $\pi_1\pi_2$, subject to a fixed total sampling effort $T = T_1 + T_2$.

Define $\phi_1 = 1 - \theta_1$ and $\phi_2 = 1 - \theta_2$, the product of detection probabilities becomes

$$\pi_1\pi_2 = \left(1 - \phi_1^{T_1}\right)\left(1 - \phi_2^{T_2}\right).$$

To formalise this optimisation, let

$$f(t; \phi_1, \phi_2) = (1 - \phi_1^t)(1 - \phi_2^{T-t}), \tag{3.4}$$

where $t = 1, 2, ..., (T-1)$, representing the number of sub-occasions allocated to capture occasion 1. The objective is to determine the optimal allocation t^* that maximises $f(t; \phi_1, \phi_2)$. The following sections investigate this optimisation under various scenarios, aiming to identify the most effective allocation of sampling effort between the two caption occasions.

3.6 Optimising Sampling Effort with Fixed Catchabilities

3.6.1 Scenario 1: Equal Catchabilities $\boldsymbol{\phi_1 = \phi_2}$

A general solution to the problem of optimising the function in Equation (3.4) can be obtained when $\phi_1 = \phi_2 = \phi$. Under this condition, the function simplifies to

$$\begin{aligned} f(t; \phi) &= \left(1 - \phi^t\right)\left(1 - \phi^{T-t}\right) \\ &= 1 - \phi^t - \phi^{T-t} + \phi^T. \end{aligned}$$

Hence, maximising $f(t; \phi)$ is equivalent to minimising the simpler function $\phi^t + \phi^{T-t}$ with respect to t.

Theorem 3.1. *Let $f(t; \phi) = (1 - \phi^t)(1 - \phi^{T-t})$ where $\phi \in (0, 1)$. Then the function $f(t; \phi)$ achieves its maximum at*

$$t^* = \begin{cases} \dfrac{T}{2}, & \text{if } T \text{ is even;} \\ \dfrac{T-1}{2} \text{ or } \dfrac{T+1}{2}, & \text{if } T \text{ is odd,} \end{cases}$$

where $t = 1, 2, \ldots, T-1$.

The result from Theorem 3.1 provides a clear practical guideline for allocating sampling effort between two capture occasions. Specifically, when the total number of sub-occasions T is even, the optimal allocation evenly divides sampling effort into two sets of $T/2$ sub-occasions each. Conversely, when T is odd, one capture-occasion receives $(T+1)/2$ sub-occasions while the other receives $(T-1)/2$. It is noteworthy that the optimal allocation identified here is **independent of both the value of $\boldsymbol{\phi}$ and the individual detection probability $\boldsymbol{\theta}$**.

Proof of Theorem 3.1 Consider the function $g(t) = \phi^t + \phi^{T-t}$ for $\phi \in (0,1)$. Observed that $g(t)$ is symmetric about t, satisfying $g(t) = g(T-t)$. To determine the minimum of $g(t)$, consider the stepwise decreasing behaviour of the function.

For $t = T-1$,

$$g(T-1) = \phi + \phi^{T-1} \geq g(T-2) = \phi^2 + \phi^{T-2},$$

as

$$\phi(1-\phi) \geq \phi^{T-2}(1-\phi)$$

or equivalently,

$$\phi \geq \phi^{T-2},$$

holds true because $T-2 \geq 1$ and $\phi \in (0,1)$.

Similarly, for $t = T-2$,

$$g(T-2) = \phi^2 + \phi^{T-2} \geq g(T-3) = \phi^3 + \phi^{T-3},$$

as

$$\phi^2(1-\phi) \geq \phi^{T-3}(1-\phi)$$

or equivalently,

$$\phi^2 \geq \phi^{T-3},$$

which holds since $T-3 \geq 1$ and $\phi \in (0,1)$.

Generalising this argument for $t \leq T-t$, it follows that

$$\phi^t \geq \phi^{T-(t+1)},$$

leading to

$$g(t) \geq g(t+1).$$

Thus, $g(t)$ decreases until reaching the midpoint $t = T/2$. Consequently,

- If T is even, the function $g(t)$ is minimised uniquely at $t^* = T/2$.

- If T is odd, due to symmetry, $g(t)$ attains its minimum at two consecutive points $t^* = (T-1)/2$ and $t^* = (T+1)/2$.

This completes the proof.

3.6.2 Golf Tees in St. Andrew

Example in this section describes a dataset from a golf tee experiment carried out in St. Andrews. In the study, 250 golf tee clusters were distributed across a 1680-square-meter area. Eight students then surveyed the area, aiming to locate as many clusters as possible. Additional information is available in Borchers et al. (2002). Table 3.3 shows the number of golf tee clusters detected by each surveyor in the St. Andrews field experiment. In total, the surveyors detected only 162 clusters, while 88 remained undiscovered.

TABLE 3.3: Number of golf tee clusters detected by each surveyor in St. Andrew field experiment.

Surveyor	1	2	3	4	5	6	7	8
Number of Detections	64	67	73	59	69	58	63	93

To test the assumption of equal catchability across surveyors, a chi-square goodness-of-fit test was performed. The result ($\chi^2 = 12.799$, $df = 7$, $p = 0.077$) indicated no significant differences in catchability among the surveyors. This suggests that the assumption of equal capture probabilities is plausible.

For demonstration purposes, the total number of surveyors ($T = 8$) was divided into two groups, with T_1 in the first group and T_2 in the second. Group 1 was treated as the first capture event, and Group 2 as the second. Seven different allocations (T_1, T_2) were considered, as shown in Table 3.4. For each configuration, the quantities n_1 and n_2 were determined, and the Lincoln–Petersen estimate $\hat{N}$ and its variance $\widehat{\text{Var}}(\hat{N})$ were computed using Equation (3.1) and Equation (3.3), respectively.

The results in Table 3.4 demonstrate that the optimal design, $t^* = T/2 = 4$, yields the smallest variance among all scenarios considered. This outcome confirms the theoretical expectation regarding optimal effort allocation under equal catchability.

3.6.3 Scenario 2: Proportional Catchabilities $\boldsymbol{\phi_1 = k\phi_2}$

When $\phi_1 \neq \phi_2$, Theorem 3.1 no longer applies. Instead, optimal sampling effort allocation can be determined by finding the value t^* that maximises the function $f(t; \phi_1, \phi_2)$ defined in Equation (3.4). Consider the scenario where the catchabilities are proportional, i.e.

TABLE 3.4: Lincoln-Petersen estimates and associated variance under different splits of eight surveyors into two groups.

T_1	T_2	n_1	n_2	m	$\hat{N}$	$\text{Var}(\hat{N})$
1	7	64	158	60	169	18.351
2	6	97	153	88	169	12.740
3	5	115	147	100	169	11.918
4	4	124	142	104	169	**11.897**
5	3	135	127	100	171	16.202
6	2	139	116	93	173	21.208
7	1	144	93	75	179	39.426

$\phi_1 = k\phi_2$, with k as a constant. In this case, the function to maximise becomes

$$\begin{aligned} f(t;\phi,k) &= (1-(k\phi)^t)(1-\phi^{T-t}) \\ &= 1-(k\phi)^t-\phi^{T-t}+k^t\phi^T. \end{aligned} \tag{3.5}$$

Maximising $f(t;\phi,k)$ in Equation (3.5) is equivalent to minimising the function

$$h(t;\phi,k) = (k\phi)^t+\phi^{T-t}-k^t\phi^T. \tag{3.6}$$

The first and second derivatives of $h(t;\phi,k)$ are:

$$h'(t) = (k\phi)^t\log(k\phi)-\phi^{T-t}\log(\phi)-\phi^T k^t\log(k),$$

and

$$h''(t) = (k\phi)^t[\log(k\phi)]^2-\phi^{T-t}[\log(\phi)]^2-\phi^T k^t[\log(k)]^2,$$

respectively. The optimisation of function in Equation (3.6) has no closed-form solution for t. However, numerical methods such as the Newton-Raphson approach can be employed to identify the optimal solution t^*. The following algorithm outlines the steps involved in finding the optimal value of t.

Step 1. Set an initial value $t^{(0)}$ for the optimal solution t^*.

Step 2. Iteratively update $t^{(r)}$ using the Newton-Raphson formula:

$$\begin{aligned} t^{(r+1)} &= t^{(r)}-\frac{h'(t^{(r)})}{h''(t^{(r)})} \\ &= t^{(r)}-\frac{(k\phi)^{t^{(r)}}\log(k\phi)-\phi^{T-t^{(r)}}\log(\phi)-\phi^T k^{t^{(r)}}\log(k)}{(k\phi)^{t^{(r)}}[\log(k\phi)]^2-\phi^{T-t^{(r)}}[\log(\phi)]^2-\phi^T k^{t^{(r)}}[\log(k)]^2}. \end{aligned}$$

Step 3. Repeat Step 2 until convergence is achieved.

To reduce the number of iterations required, it is recommended to initialise the procedure with $t^{(0)} = T/2$.

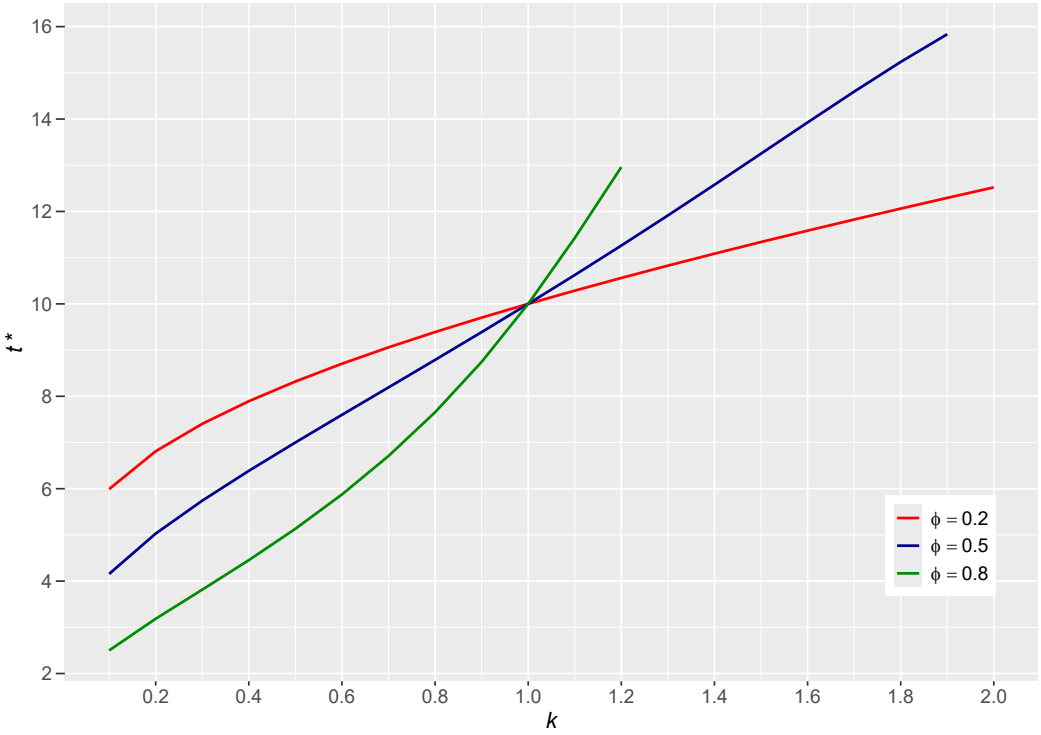


FIGURE 3.2: Optimal allocation t^* for various values of k and ϕ when T=20, with constrain $k\phi \in (0, 1)$.

Figure 3.2 illustrates how the optimal allocation t^* changes with varying values of k and ϕ for a total of $T = 20$ occasions. When $k = 1$, which corresponds to the previously discussed case $\phi_1 = \phi_2$, the optimal strategy evenly divides the sampling effort between two occasions. As k increases beyond 1, indicating lower detectability in capture occasion 1 (i.e. higher ϕ_1), it becomes advantageous to allocate a greater sampling effort to the first capture occasion. This allocation strategy optimises the precision of population size estimates.

3.6.3.1 Simulation

In this study, the Bernoulli distribution was used to generate simulated capture-recapture data. The simulation assumed two capture probabilities: θ_1 for capture occasion 1 and θ_2 for the capture occasion 2. For each trial, θ_1 was assigned to the first half of the capture occasions $t = 1, 2, ..., \lceil T/2 \rceil$, and θ_2 to the remaining occasions. Based on these probabilities, binary outcomes (1, 0) were generated for a population of N individuals across a total of $T = T_1 + T_2$ capture occasions. A value of 1 indicated that an animal was captured, while a value of 0 indicated no capture. Table 3.5 summarises the parameter configurations used in the simulation study.

TABLE 3.5: Parameter settings used in the simulation study.

N	T	$(\theta_1\ ,\ \theta_2)$
1000	20	(0.40 , 0.20)
400	11	(0.40 , 0.04)
200	9	(0.20 , 0.30)
100	6	(0.20 , 0.20)

The simulation consisted of 64 parameter combinations derived from four case population sizes, N={1000, 400, 200, 100}; four levels of total captures occasions, T={20, 11, 9, 6}; and four pairs of sub-occasions capture probabilities, i.e.: (θ_1, θ_2) = {(0.4,0.2), (0.4,0.04), (0.2,0.3), (0.2, 0.2)}, representing different ratio of $\theta_1 : \theta_2$. Each combinations was simulated with $B = 50000$ replicate datasets. For each replicate, detection at sub-occasion j within capture occasion i followed

$$X_{ijk} \sim \text{Bernoulli}(\theta_i) \quad \text{for} \quad i = 1, 2; \quad j = 1, \ldots, T_i; \quad k = 1, ..., N.$$

TABLE 3.6: Simulation results of Lincoln-Petersen estimation procedure applied on data generated with B=50000 replicates for N=1000, T=20, $\theta_1 = 0.4$, $\theta_2 = 0.2$.

t	$\mathbf{mean}(\hat{N})$	$\mathbf{Var}(\hat{N})$	$\pi_1 \times \pi_2$
1	1000.002	1.61412	0.400
2	999.997	1.02507	0.639
3	1000.005	0.82488	0.782
4	1000.003	0.74839	0.866
5	999.988	0.71788	0.915
6	999.995	0.69075	0.940
7	999.997	**0.69002**	**0.949**
8	1000.002	0.69101	0.945
9	1000.005	0.69394	0.926
10	1000.005	0.72879	0.887
11	1000.000	0.74907	0.862
12	1000.004	0.78166	0.829
13	999.992	0.83059	0.788
14	999.993	0.88458	0.736
15	1000.001	0.95868	0.671
16	999.997	1.10304	0.589
17	1000.001	1.34057	0.487
18	1000.004	1.79398	0.360
19	1000.003	3.25199	0.200

For each simulated dataset, the population size estimate $\hat{N}$ was computed using values of m, n_1, and n_2. The Chapman estimator was applied to mitigate bias, particularly in scenarios where m is small or equal to zero. The mean and variance of $\hat{N}$ across all replications were recorded. The allocation t that minimised $\text{Var}(\hat{N})$ was identified as the optimal allocation allocation, denoted by t^*.

Table 3.6 presents an example of simulation results for N=1000, T=20, θ_1=0.4, and θ_2=0.2. It illustrates how the variance of $\hat{N}$ changes across different values of t. The optimal allocation, $t^* = 7$, corresponds to assigning 7 sub-occasions to capture occasion 1 and 13 to capture occasion 2. This configuration yields the smallest variance and, therefore, the most precise estimate. The improvement is also reflected in the maximised product of marginal capture probabilities, $\pi_1 \times \pi_2$. Full simulation results for all 64 allocation combinations are provided in Chapter B.

Table 3.7 compares the optimal t^* values obtained from the simulation with those derived using the Newton-Raphson optimization. The results show that the optimal allocation is largely insensitive to changes in population size N, remaining stable across different values. However, changes in T or in the capture probabilities (θ_1, θ_2) do affect the optimal t. The close agreement between simulation and Newton-Raphson outcome, where each differing less than one unit, validates the robustness of the findings. Furthermore, the results in Chapter B indicate that selecting $T > 10$ is advisable in study design to reduce estimator variance.

3.7 Optimizing Sampling Effort in the Presence of Unknown Catchabilities

Solutions presented in Section 3.6 require prior knowledge of detection probabilities for each capture occasion. When these detection probabilities are unknown, a Pseudo-Bayesian approach can be used by assigning prior distributions to ϕ_1 and ϕ_2. The central idea of this pseudo-Bayesian method is to integrate the prior distribution into the objective function. By accounting for all potential values of ϕ_1 and ϕ_2, this approach explores the objective function's behaviour. The resulting expectation, denoted as $g(t)$, provides an average representation of the objective function, thereby reducing the influence of extreme or outlier values.

The first step is to calculate the expected value of $f(t; \phi_1, \phi_2)$ with respect to the distribution ϕ_1 and ϕ_2:

$$g(t) = \mathbb{E}\Big[f(t; \phi_1, \phi_2)\Big]. \tag{3.7}$$

The following subsections discuss three different scenarios, assuming ϕ_1 and ϕ_2 follow uniform distributions.

3.7.1 Scenario 1: Uninformative Priors for Catchabilities

When no prior information is available regarding ϕ_1 and ϕ_2, it is reasonable to assume independent uninformative priors, specifically $\phi_1 \sim$ Uniform (0, 1) and $\phi_2 \sim$ Uniform (0, 1). These two uniform distributions are considered independent, reflecting the assumption that the detection probabilities at each capture occasion vary independently in the absence of prior knowledge.

TABLE 3.7: Comparison of optimal allocation values t^* determined by Newton-Raphson (NR) optimisation and simulation methods, across various combinations of total capture occasions T, capture probabilities θ_1 and θ_2, and population sizes N.

T	θ_1	θ_2	**Optimal t^***				
			NR	$N = 1000$	$N = 400$	$N = 200$	$N = 100$
20	0.4	0.2	7.16	7	7	7	7
	0.4	0.04	4.82	6	5	5	5
	0.2	0.3	11.54	13	12	12	12
	0.2	0.2	10.00	9	9	9	10
11	0.4	0.2	4.29	4	4	4	4
	0.4	0.04	3.36	3	3	3	3
	0.2	0.3	6.13	7	7	7	7
	0.2	0.2	5.50	6	6	6	5
9	0.4	0.2	3.61	3	3	3	3
	0.4	0.04	2.95	3	3	3	3
	0.2	0.3	4.96	5	5	6	6
	0.2	0.2	4.50	4	4	4	5
6	0.4	0.2	2.54	2	2	2	2
	0.4	0.04	2.21	2	2	2	2
	0.2	0.3	3.23	4	4	4	4
	0.2	0.2	3.00	3	3	3	3

Recall that for a standard uniform distribution $\phi \sim \text{Uniform}(0,1)$, the t-th moment is given by:

$$\mathbb{E}(\phi^t) = \int_0^1 \phi^t \, d\phi = \frac{1}{t+1}.$$

Under this assumption, the expectation defined in Equation (3.7) becomes:

$$\begin{aligned} g(t) &= \mathbb{E}\left[(1-\phi_1^t)(1-\phi_2^{T-t})\right] \\ &= 1 - \mathbb{E}(\phi_1^t) - \mathbb{E}(\phi_2^{T-t}) + \mathbb{E}(\phi_1^t)\mathbb{E}(\phi_2^{T-t}) \\ &= 1 - \frac{1}{t+1} - \frac{1}{T-t+1} + \left(\frac{1}{t+1}\right)\left(\frac{1}{T-t+1}\right) \\ &= \frac{Tt - t^2}{Tt - t^2 + T + 1}. \end{aligned}$$

The derivative of $g(t)$ with respect to t is:

$$g'(t) = \frac{(T-2t)(T+1)}{(Tt - t^2 + T + 1)^2}. \tag{3.8}$$

Setting Equation (3.8) equal to zero yields the optimal solution $t^* = T/2$. Figure 3.3 illustrates the behaviour of $g(t)$ when $T = 20$, clearly showing that the optimal allocation occurs at $t^* = 10$.

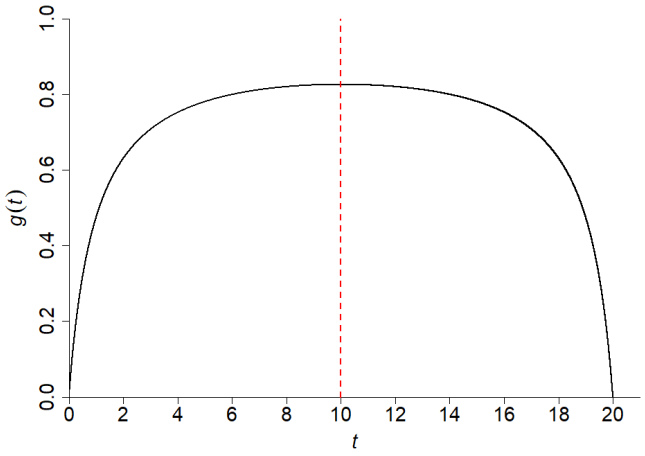


FIGURE 3.3: $g(t)$ function when $T = 20$. Red line marks the optimal value of t.

3.7.2 Scenario 2: Higher Re-catchability $\phi_2 < \phi_1$

In this scenario, the detection probabilities are modelled such that $\phi_1 \sim$ Uniform $(0,1)$, and conditional on ϕ_1, the second detection probability follows $\phi_2 \mid \phi_1 \sim$ Uniform $(0, \phi_1)$. The conditional density of $\phi_2 \mid \phi_1$ is therefore given by:

$$f(\phi_2 \mid \phi_1) = \frac{1}{\phi_1}, \quad 0 \le \phi_1 \le 1 \quad \text{and} \quad 0 \le \phi_2 \le \phi_1.$$

The marginal distribution of ϕ_1 is uniform over [0,1], i.e.,

$$f(\phi_1) = 1, \quad 0 \leq \phi_1 \leq 1.$$

Combining these yields the joint probability density function:

$$f(\phi_1, \phi_2) = f(\phi_2 \mid \phi_1) f(\phi_1) = \frac{1}{\phi_1}, \quad 0 \leq \phi_2 \leq \phi_1 \leq 1.$$

The marginal density of ϕ_1 is confirmed via integration:

$$f(\phi_1) = \int_0^{\phi_1} f(\phi_1, \phi_2) d\phi_2 = \int_0^{\phi_1} \frac{1}{\phi_1} d\phi_2 = 1.$$

The marginal density of ϕ_2 is obtained as:

$$f(\phi_2) = \int_{\phi_2}^1 f(\phi_1, \phi_2) d\phi_1 = \int_{\phi_2}^1 \frac{1}{\phi_1} = -\log(\phi_2), \quad 0 < \phi_2 \leq 1.$$

To calculate the expectation of $f(t; \phi_1, \phi_2)$ from Equation (3.4) with respect to ϕ_1 and ϕ_2, the law of total expectation is applied, giving

$$\begin{aligned}
g(t) &= \mathbb{E}[f(t; \phi_1, \phi_2)] \\
&= \mathbb{E}\Big\{\mathbb{E}[f(t; \phi_1, \phi_2)|\phi_1]\Big\} \\
&= \mathbb{E}\Big\{\mathbb{E}[1 - \phi_1^t - \phi_2^{T-t} + \phi_1^t \phi_2^{T-t}|\phi_1]\Big\} \\
&= \mathbb{E}\Big\{1 - \phi_1^t - \mathbb{E}(\phi_2^{T-t}|\phi_1) + \phi_1^t \mathbb{E}(\phi_2^{T-t}|\phi_1)\Big\} \\
&= \mathbb{E}\left\{1 - \phi_1^t - \frac{\phi_1^{T-t}}{(T-t+1)} + \frac{\phi_1^T}{(T-t+1)}\right\} \\
&= 1 - \mathbb{E}(\phi_1^t) - \frac{\mathbb{E}(\phi_1^{T-t})}{(T-t+1)} + \frac{\mathbb{E}(\phi_1^T)}{(T-t+1)} \\
&= 1 - \frac{1}{t+1} - \frac{1}{(T-t+1)^2} + \frac{1}{(T-t+1)(T+1)}.
\end{aligned} \tag{3.9}$$

The first and second derivatives of $g(t)$ are:

$$g'(t) = \frac{1}{(t+1)^2} + \frac{1}{(T+1)(T-t+1)^2} - \frac{2}{(T-t+1)^3},$$

and

$$g''(t) = -\frac{2}{(t+1)^3} + \frac{2}{(T+1)(T-t+1)^3} - \frac{6}{(T-t+1)^4},$$

respectively. The function $g(t)$ in Equation (3.9) does not have a closed-form solution for its maximum. Therefore, numerical method such as Newton-Raphson approach are employed. The steps for optimising $g(t)$ are as follows:

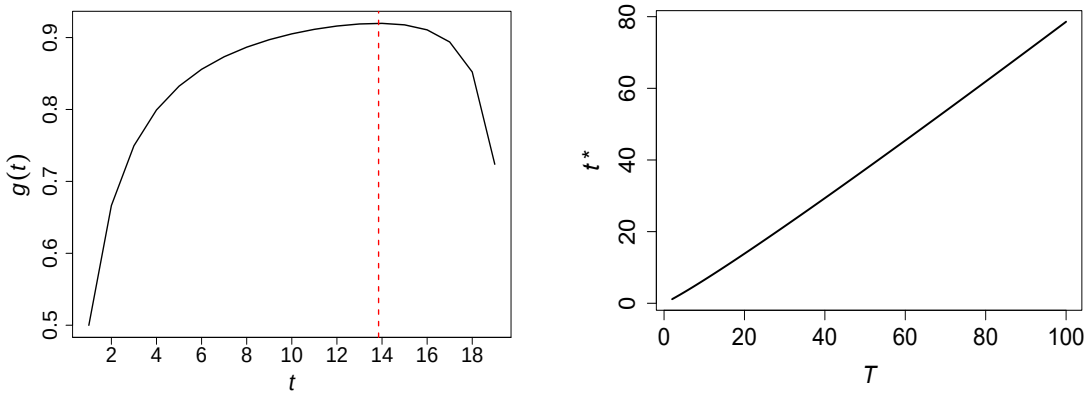
Step 1. Set an initial guess for the optimal t^*, denoted as $t^{(0)}$.

Step 2. Update $t^{(r)}$ iteratively using the formula

$$t^{(r+1)} = t^{(r)} - \frac{g'(t^{(r)})}{g''(t^{(r)})}.$$

Step 3. Repeat Step 2 until convergence is achieved.

Figure 3.4(A) illustrates an example of the function $g(t)$ when $\phi_1 \sim$ Uniform (0, 1), $\phi_2 < \phi_1$, and $T = 20$. Using the Newton-Raphson approach, the optimal allocation is determined as $t^* = 13.848$. Figure 3.4(B) further depicts the relationship between the total sampling effort T and the optimal allocation t^* across values of T ranging from 2 to 100.



(A) (B)

FIGURE 3.4: Optimal allocation t^* for the scenario $\phi_2 < \phi_1$, where $\phi_1 \sim$ Uniform (0, 1). (A) Function $g(t)$ with $T = 20$; the vertical red line marks the optimal value of t. (B) Relationship between optimal allocation t^* and total sampling effort T.

3.7.3 Scenario 3: Lower Re-catchability $\phi_1 < \phi_2$

In the third scenario, capture probability on the second occasion is assumed lower than on the first occasion ($\phi_1 < \phi_2$). Specifically, $\phi_2 \sim$ Uniform (0, 1) and $\phi_1 \mid \phi_2 \sim$ Uniform (0, ϕ_2). Under these conditions, the joint and marginal probability density functions are

defined as:

$$f(\phi_1, \phi_2) = \frac{1}{\phi_2}, \quad 0 \leq \phi_1 \leq \phi_2 \leq 1,$$
$$f(\phi_1) = -\log(\phi_1), \quad 0 \leq \phi_1 \leq 1,$$
$$f(\phi_2) = 1, \quad 0 \leq \phi_2 \leq 1.$$

The expectation of $f(t; \phi_1, \phi_2)$, defined in Equation (3.4), with respect to ϕ_1 and ϕ_2 is calculated using the law of total expectation.

$$\begin{aligned}
g(t) &= \mathbb{E}\left[f(t; \phi_1, \phi_2)\right] \\
&= \mathbb{E}\Big\{\mathbb{E}[f(t; \phi_1, \phi_2)|\phi_2]\Big\} \\
&= \mathbb{E}\Big\{\mathbb{E}[1 - \phi_1^t - \phi_2^{T-t} + \phi_1^t\phi_2^{T-t} \mid \phi_2]\Big\} \\
&= \mathbb{E}\Big[1 - \mathbb{E}(\phi_1^t \mid \phi_2) - \phi_2^{T-t} + \phi_2^{T-t}\ \mathbb{E}(\phi_1^t \mid \phi_2)\Big] \\
&= \mathbb{E}\left[1 - \frac{\phi_2^t}{t+1} - \phi_2^{T-t} + \frac{\phi_2^T}{t+1}\right] \\
&= 1 - \frac{\mathbb{E}\left(\phi_2^t\right)}{t+1} - \mathbb{E}\left(\phi_2^{T-t}\right) + \frac{\mathbb{E}\left(\phi_2^T\right)}{t+1} \\
&= 1 - \frac{1}{(t+1)^2} - \frac{1}{(T-t+1)} + \frac{1}{(T+1)(t+1)}.
\end{aligned}$$

The first and second derivatives of $g(t)$ are:

$$g'(t) = \frac{2}{(t+1)^3} - \frac{1}{(T-t+1)^2} - \frac{1}{(T+1)(t+1)^2},$$

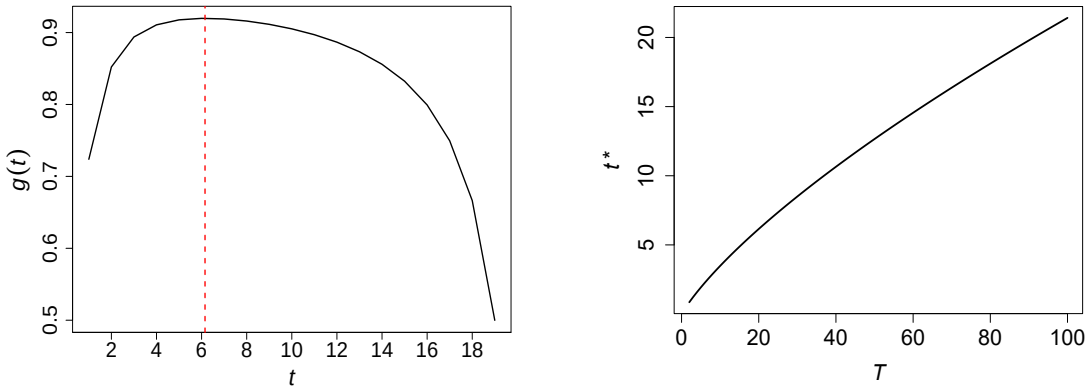
and

$$g''(t) = -\frac{6}{(t+1)^4} - \frac{2}{(T-t+1)^3} + \frac{2}{(T+1)(t+1)^3},$$

respectively. As no closed-form solution exists for maximising $g(t)$, numerical methods such as the Newton-Raphson approach are employed. Figure 3.5(A) shows the function $g(t)$ for $T = 20$. The Newton-Raphson method yields an optimal allocation $t^* = 6.152$. Figure 3.5(B) illustrates the optimal allocation t^* across varying total sampling effort T ranging from 2 to 100.

3.7.4 Cottontail Rabbits in Tennesses

A practical example of hierarchical design in capture-recapture studies is demonstrated by McWherter (1991). The study investigated the capture of cottontail rabbits (*Sylvilagus floridanus*) at TVA's Land Between the Lakes, Stewart County, Tennessee. Rabbits were



(A) (B)

FIGURE 3.5: Optimal allocation t^* when $\phi_2 \sim$ Uniform (0, 1), and $\phi_1 < \phi_2$. (A) Function $g(t)$ for $T = 20$; the vertical red line indicates the optimal allocation t^*. (B) Relationship between optimal allocation t^* and total sampling effort T.

captured over 15 consecutive days during March, August, and December of 1985. For illustrative purposes, only data from August and December are analysed here.

TABLE 3.8: Cottontail rabbit captures organized in a hierarchical Lincoln-Petersen design. August capture events represent capture occasion 1, and December capture events represent capture occasion 2. n_{00} denotes the number of rabbits missed during both occasions and is not directly observed.

		Occasion 2 (December)		
		Captured	Not Captured	Total
Occasion 1	Captured	17	31	48
(August)	Not Captured	46	n_{00}	
		63		N

In the hierarchical Lincoln–Petersen design, August captures form Occasion 1 with $T_1 = 15$ sub-occasions, and December captures form Occasion 2 with $T_2 = 15$. The capture data are summarised in Table 3.8. Using the Lincoln–Petersen estimator, the estimated population size is $\hat{N} = 178$, with an estimated variance 877.717.

$$\hat{N} = \frac{n_1 n_2}{m} = 178.$$

$$\widehat{\mathrm{Var}}(\hat{N}) = \frac{n_1\, n_2 (n_1 - m)(n_2 - m)}{m^3} = 877.717.$$

Based on the data, catchability was higher in December than in August. A greater proportion of individuals were detected during the December trapping period, suggesting that environmental or behavioural conditions at that time were more favourable for capture. This observed difference in detectability has important implications for study

TABLE 3.9: Estimation results for (a) Scenario 1 with uniform catchabilities ($\phi_1 \sim$ Uniform(0,1) and $\phi_2 \sim$ Uniform(0,1)), (b) Scenario 2 with higher re-catchabilities ($\phi_1 \sim$ Uniform(0,1) and $\phi_2 \mid \phi_1 \sim$ Uniform(0,ϕ_1)), and (c) Scenario 3 with lower re-catchabilities ($\phi_1 \mid \phi_2 \sim$ Uniform(0,ϕ_2) and $\phi_2 \sim$ Uniform(0,1)) - N=1000, T=20.

	(a) Scenario 1: $\phi_1 \sim$ Uniform(0,1) and $\phi_2 \sim$ Uniform(0,1)				(b) Scenario 2: $\phi_1 \sim$ Uniform(0,1) and $\phi_2 < \phi_1$				(c) Scenario 3: $\phi_2 \sim$ Uniform(0,1) and $\phi_2 > \phi_1$			
t	**mean($\hat{N}$)**	**over-estimate (%)**	**under-estimate (%)**	**accurate (%)**	**mean($\hat{N}$)**	**over-estimate (%)**	**under-estimate(%)**	**accurate (%)**	**mean($\hat{N}$)**	**over-estimate (%)**	**under-estimate (%)**	**accurate (%)**
1	1000.0	8.1	11.0	80.9	999.6	4.0	7.4	88.6	1000.0	5.3	5.5	89.2
2	999.8	9.5	10.8	79.7	999.5	4.8	6.8	88.4	999.9	6.5	6.1	87.4
3	999.7	12.2	10.5	77.2	999.6	5.6	6.7	87.7	999.9	8.6	5.7	85.7
4	999.6	14.3	10.7	75.0	999.9	6.3	6.5	87.2	999.9	10.7	5.8	83.5
5	1000.0	17.3	10.1	72.5	1000.0	7.5	6.2	86.4	1000.1	13.9	5.6	80.5
6	1000.0	18.5	10.5	70.9	999.8	8.2	6.3	85.5	1000.0	15.6	5.7	78.7
7	999.7	18.5	10.9	70.6	999.9	8.4	6.6	85.0	999.9	15.6	6.0	78.4
8	999.9	18.7	10.5	70.8	999.8	9.7	6.5	83.9	1000.0	14.3	6.1	79.6
9	1000.1	17.9	10.5	71.6	999.8	10.6	6.5	82.8	1000.0	13.2	5.7	81.1
10	1000.1	17.4	10.2	72.4	1000.0	11.5	6.4	82.1	1000.0	10.9	5.9	83.3
11	999.7	18.2	10.2	71.5	1000.0	13.9	6.1	80.0	999.9	10.1	5.8	84.2
12	1000.0	18.7	10.3	71.0	999.9	15.4	6.3	78.3	999.9	8.9	6.0	85.1
13	999.9	18.8	10.5	70.7	1000.0	16.6	5.9	77.4	1000.0	7.9	5.7	86.4
14	1000.0	19.0	10.2	70.8	1000.0	16.5	6.1	77.3	999.9	7.1	6.1	86.9
15	999.8	17.2	10.2	72.6	999.9	14.0	6.6	79.4	999.8	6.3	5.9	87.7
16	999.9	14.2	10.4	75.3	1000.1	12.1	5.8	82.1	999.8	5.6	6.0	88.4
17	999.6	11.9	10.7	77.4	999.8	9.4	6.2	84.4	999.9	4.8	6.2	89.0
18	1000.1	10.2	10.5	79.3	999.9	7.1	6.6	86.4	999.7	4.5	6.2	89.3
19	999.5	7.9	10.9	81.2	1000.1	5.9	6.1	88.0	1000.0	3.4	6.7	89.9

design. Instead of distributing sampling effort equally across both months, it is more efficient to allocate additional effort to August, where catchability was lower.

Using the pseudo-Bayesian allocation method described in Section 3.7.2, the optimal sampling configuration assigns $t^* = 22$ sub-occasions to August and 8 to December, keeping the total $T = 30$.

3.7.5 Addressing Potential Biases in Estimation

This section investigates whether assuming a uniform prior distribution for catchabilities introduces bias, particularly by inducing heterogeneity across capture events that could skew population size estimates. To assess this, simulations were performed for all three scenarios, with a fixed population size of $N = 1000$ and $T = 20$ sampling sub-occasions. Each scenario was replicated $B = 10000$ times for robust inference.

Table 3.9 summarises the proportion of simulations where the estimates are overestimated ($\hat{N} > N$), underestimated ($\hat{N} < N$), or approximately equal to the true population size ($\hat{N} = N$). In practice, since $\hat{N}$ is a continuous quantity, the evaluation of $\hat{N} = N$ was based on the rounded estimate, i.e., $\hat{N}$ was rounded to the nearest integer before comparison with the true value $N = 1000$.

These metrics provide a quantitative measure of estimator performance under different catchability assumptions. The results demonstrate that the method yields highly accurate estimates, with minimal systematic over- or underestimation. This consistency suggests that the uniform prior assumption does not substantially bias estimates, supporting the reliability of the approach for population size inference.

3.8 Hierarchical Lincoln-Petersen Approach with Time Window

In some studies, capture–recapture results are available in the form of a two-dimensional table, where each entry n_{ij} represents the number of individuals captured i times during the first occasion and j times during the second occasion (see Table 3.10). The indices take values $i = 0, 1, \ldots, I$ and $j = 0, 1, \ldots, J$, where I and J denote the maximum number of times an individual is observed in each respective occasion.

In the case study by Lerdsuwansri and Böhning (2014), the 12-month observation period was divided into two segments, each treated as a distinct capture source, as shown in Table 3.10. Each source comprises several consecutive time windows, measured in months. The first segment represents the initial months, while the second segment includes the remaining months. This division introduces a hierarchical structure into the Lincoln-Petersen design, as illustrated in Figure 3.6. Throughout the observation period, individuals could be identified repeatedly across both sources.

TABLE 3.10: Identification frequency of heroin user in Bangkok, year 2001.

		2nd half year							
		0	1	2	3	4	5	6	
	0	-	1401	369	98	23	1	1	1893
	1	1736	315	129	50	26	1	0	2257
1st	2	445	137	105	53	20	4	0	764
half	3	164	89	75	49	30	1	2	410
year	4	47	25	48	34	8	0	0	162
	5	5	7	8	2	3	0	0	25
	6	1	0	1	1	0	0	0	3
	8	0	0	0	1	0	0	0	1
		2398	1974	735	288	110	7	3	5515

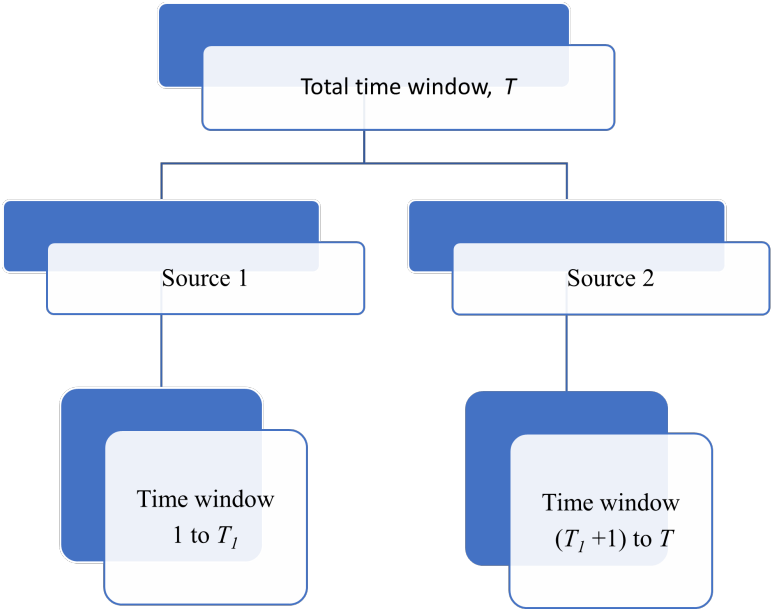


FIGURE 3.6: A hierarchical structure within a time window.

3.8.1 Maximum Likelihood Estimation

Assume that for each individual $d = 1, 2, \ldots, N$, the observed pair of counts $y_d = (y_{d1}, y_{d2})$ follows a product of two independent Poisson distributions:

$$\begin{aligned} f(y_1, y_2; \lambda_1, \lambda_2) &= \text{Poi}(y_1; \lambda_1)\text{Poi}(y_2; \lambda_2) \\ &= \left(\frac{\lambda_1^{y_{d1}}}{y_{d1}!}\, e^{-\lambda_1}\right)\left(\frac{\lambda_2^{y_{d2}}}{y_{d2}!}\, e^{-\lambda_2}\right). \end{aligned}$$

The joint likelihood function for the complete data is then:

$$L(y; \lambda_1, \lambda_2) = \prod_{d=1}^{N} f(y_d; \lambda_1, \lambda_2) = \prod_{d=1}^{N} \left(\frac{\lambda_1^{y_{d1}}}{y_{d1}!}\, e^{-\lambda_1}\right)\left(\frac{\lambda_2^{y_{d2}}}{y_{d2}!}\, e^{-\lambda_2}\right).$$

Using the observed frequencies n_{ij} from Table 3.10, the likelihood simplifies to:

$$L(n;\lambda_1,\lambda_2) = \prod_{i=0}^{I}\prod_{ij=0}^{J}\left[\left(\frac{\lambda_1^i}{i!}\ e^{-\lambda_1}\right)\ \left(\frac{\lambda_2^j}{j!}\ e^{-\lambda_2}\right)\right]^{n_{ij}}.$$

This likelihood function is constructed from the observed frequencies n_{ij} for $i+j \geq 1$, along with a single unobserved frequency n_{00}. Together, these define the complete data: an observed component n_{ij} with $i+j \geq 1$, and a missing component n_{00}. The total observed sample size is given by $n = \sum\sum_{i+j\geq 1} n_{ij}$.

To address the missing data, the EM algorithm is applied. This method replaces the unobserved count n_{00} with its conditional expectation, given the observed frequencies. Since the observed data arise from a zero-truncated Poisson distribution, the conditional expectation of n_{00} is:

$$\begin{aligned}\tilde{n}_{00} &= \mathbb{E}(n_{00} \mid n_{ij}, i+j \geq 1)\\ &= n\ \frac{\text{Poi}(0;\lambda_1)\text{Poi}(0;\lambda_2)}{1-\text{Poi}(0;\lambda_1)\text{Poi}(0;\lambda_2)}\\ &= n\left[\frac{e^{-\lambda_1-\lambda_2}}{1-e^{-\lambda_1-\lambda_2}}\right].\end{aligned}$$

This complete the E-step of the EM algorithm. For the subsequent M-step, the goal is to maximise the joint log-likelihood function, now augmented by the imputed value $\tilde{n}_{00}$. The log-likelihood becomes:

$$\ell(n;\lambda_1,\lambda_2) = \sum_{i=0}^{I}\sum_{j=0}^{J}\ n_{ij}\log\left[\left(\frac{\lambda_1^i}{i!}\ e^{-\lambda_1}\right)\ \left(\frac{\lambda_2^j}{j!}\ e^{-\lambda_2}\right)\right],$$

in which n_{00} is replaced by $\tilde{n}_{00}$, yielding:

$$\ell(n \mid n_{ij} i+j \geq 1;\lambda_1,\lambda_2) = \tilde{n}_{00}(-\lambda_1-\lambda_2) + \sum_{i+j\geq 1}\sum\ n_{ij}\log\left[\left(\frac{\lambda_1^i}{i!}\ e^{-\lambda_1}\right)\ \left(\frac{\lambda_2^j}{j!}\ e^{-\lambda_2}\right)\right].$$

This is the function we have to maximise with respect to the parameter λ_1 and λ_2. To do this, we calculate the first derivative (ignoring constants)

To estimate λ_1 and λ_2, take the partial derivatives of the log-likelihood and set them to zero. Ignoring constants, the derivative with respect to λ_1 is:

$$\begin{aligned}\frac{\partial \ell}{\partial \lambda_1} &= -\tilde{n}_{00} + \sum\sum_{i+j\geq 1} n_{ij}\left(\frac{i}{\lambda_1} - 1\right) \\ &= -\tilde{N} + \frac{1}{\lambda_1}\sum\sum_{i+j\geq 1} i n_{ij}. \end{aligned} \tag{3.10}$$

Let $\tilde{N} = \tilde{n}_{00} + \sum\sum_{i+j\geq 1} n_{ij}$, solving the derivative equation in Equation (3.10) yields:

$$\hat{\lambda}_1 = \frac{1}{\tilde{N}}\sum\sum_{i+j\geq 1} i n_{ij}.$$

Applying the same procedure with respect to λ_2 yields:

$$\hat{\lambda}_2 = \frac{1}{\tilde{N}}\sum\sum_{i+j\geq 1} j n_{ij}.$$

These expressions form the basis of the EM algorithm used to compute the maximum likelihood estimates.

EM Algorithm for MLE with Zero-Truncated Two-Poisson Distribution

Step 0. **Initialisation:** Choose starting values $\hat{\lambda}_1^{(0)}$ and $\hat{\lambda}_2^{(0)}$. A practical approach is to initialise using the observed marginal means:

$$\hat{\lambda}_1^{(0)} = \frac{\sum\sum_{i+j\geq 1} i n_{ij}}{n}, \quad \hat{\lambda}_2^{(0)} = \frac{\sum\sum_{i+j\geq 1} j n_{ij}}{n},$$

where $n = \sum\sum_{i+j\geq 1} n_{ij}$. Alternatively, both values can be initialised at 1.

Step 1. **E-step:** Compute the expected number of unobserved individuals:

$$\hat{n}_{00} = n\left[\frac{e^{-\hat{\lambda}_1-\hat{\lambda}_2}}{1-e^{-\hat{\lambda}_1-\hat{\lambda}_2}}\right].$$

Step 2. **M-step:** Use the imputed value $\hat{n}_{00}$ to update the parameter estimates:

$$\hat{\lambda}_1 = \frac{1}{(n+\hat{n}_{00})}\sum\sum_{i+j\geq 1} i n_{ij}, \quad \hat{\lambda}_2 = \frac{1}{(n+\hat{n}_{00})}\sum\sum_{i+j\geq 1} j n_{ij}.$$

Step 3. Iterate steps 2 and 3 until convergence.

3.8.2 Sampling Efforts Determination

In contrast to traditional capture-recapture methods, the maximum number of captures in such design is a random variable. Assuming independence and homogeneity of capture events, the capture count during the first half of the time window is modelled as $Y_1 \sim \text{Po}(\lambda_1, T_1)$, and for the second half, as $Y_2 \sim \text{Po}(\lambda_2, T_2)$, with $T_1 + T_2 = T$ and T represent the fixed total observation period (e.g. 12 months). Here, λ_j and T_j represents the mean captures and the number of time units within each time window $j \in \{1, 2\}$, respectively.

The probability of not detecting a subject during capture occasion 1 is $1 - \pi_1 = e^{-\lambda_1 T_1}$. Similarly, the probability of not detecting a subject on capture occasion 2 is $1 - \pi_2 = e^{-\lambda_2 T_2}$. To optimize sampling effort allocation, the goal is to determine the value of T_1 (and implicitly $T_2 = T - T_1$) that maximises the joint detection probability

$$\pi_1 \pi_2 = (1 - e^{-\lambda_1 T_1})(1 - e^{-\lambda_2 T_2}).$$

By defining $\phi_1 = e^{-\lambda_1}$, $\phi_2 = e^{-\lambda_2}$, and $t = T_1$, the objective function can be re-expressed as:

$$f(t; \phi_1, \phi_2) = (1 - \phi_1^t)(1 - \phi_2^{T-t}),$$

which aligns with the general form given in Equation (3.4).

Using the observed repeated count data from the Bangkok heroin user case study presented in Table 3.10, the estimates of the Poisson rates are $\hat{\lambda}_1 = 0.817$ and $\hat{\lambda}_2 = 0.675$. Applying these estimates within the optimisation procedure described in Section 3.6.3 yields an optimal t^*=6. This indicates that allocating 6 months each in source 1 and source 2 results in the most efficient population size estimates.

3.9 Discussion and Conclusion

This chapter has examined the optimisation of sampling effort in the hierarchical Lincoln-Petersen framework, emphasizing the importance of maximising the joint detection probabilities across both capture occasions. Theoretical results and simulation studies consistently show that optimal allocation of sub-occasions improves the precision of population size estimates.

When detection probabilities are equal across capture occasions, the optimal design is achieved by allocating the sampling effort equally, that is, setting $t = T/2$. This approach does not require prior knowledge of individual capture probabilities θ and is particularly relevant when capture events occur over a short period. As highlighted by Robson and Regier (1964) and Greenwood and Robinson (2006), an equal allocation of resources between both occasions is optimal for minimizing errors when funding is limited, making it cost-effective to allocate the same sampling effort during each occasion.

In contrast, when catchability differs across occasions, asymmetrical allocations that reflect this heterogeneity yield lower estimator variance, as demonstrated by Efford et al. (2013). In cases where detection probabilities are unknown, this study introduced a pseudo-Bayesian approach. Simulation results confirmed that the method delivers robust and unbiased estimates under uniform prior assumptions. While the method introduced here simplifies the full Bayesian approach, its credibility is strengthened by the consistent findings from fully Bayesian studies such as those by Basu and Ebrahimi (2001) and Wang et al. (2015). These studies show that Bayesian methods are generally robust to uncertainty in parameter estimates, which supports the value of adopting simplified pseudo-Bayesian approaches. Such methods retain many of the core strengths of Bayesian analysis, such as incorporating prior information and managing uncertainty, while offering a more practical and computationally efficient option for researchers.

In conclusion, the hierarchical Lincoln–Petersen method offers a systematic and flexible strategy for improving the efficiency of capture–recapture studies. By explicitly linking sampling effort, detectability, and uncertainty, it provides researchers with a practical foundation for optimising study design in resource-constrained setting.

Chapter 4

Sampling Efforts in Schnabel Census

4.1 Introduction

The Schnabel census method (Schnabel, 1938) is commonly used in capture-recapture studies. This method takes multiple independent random samples from a closed population. Each unmarked individual is given a unique mark when captured and then released. A key assumption is that every individual has the same capture probability (Seber, 1982).

Determining the right sampling effort is important in Schnabel census studies. The number of sampling occasions affects accuracy, especially when resources like time, labour, and funding are limited. Sample size and effort directly influence the reliability of estimates.

This chapter examines the required sampling effort in Schnabel census studies. It considers different capture success rates and population heterogeneity. First, the relationship between sampling effort, capture probability, and success rate is analysed using a simple Binomial model. Then, more flexible models, including Beta-Binomial and Binomial mixture, are introduced to account for detectability differences.

4.2 Capture Histories from Schnabel Census

In a Schnabel census, captured individuals are checked for tags, marked if new, and released. This process repeats in each sampling event. The method relies on a series of independent random samples from a closed population. All unmarked individuals are tagged except in the final sampling (Schnabel, 1938; McCrea and Morgan, 2015).

The raw data consist of capture histories for observed individuals. The capture-recapture scenario is shown in Table 4.1. Here, Y_{ij} represents whether individual i was captured

on occasion j: $Y_{ij} = 1$ means captured, and $Y_{ij} = 0$ means not captured. The indices $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, T$ refer to individuals and sampling occasions. The total captures for individual i is $Y_{i.} = \sum_{j=1}^{T} Y_{ij}$. If $Y_{i.} = 0$, the individual was never captured.

The data include two parts: observed counts $Y_1., Y_2., \ldots, Y_n.$ for sampled individuals, and unobserved counts $Y_{n+1}., Y_{n+2}., \ldots, Y_N.$ for missed individuals. This separation helps distinguish between the full population counts and the zero-truncated sample. Capture-recapture methods estimate the number of missed individuals, providing an estimate of the total population size N.

TABLE 4.1: Structure of capture histories from a Schnabel census. Each entry Y_{ij} indicates whether individual i was detected ($Y_{ij} = 1$) or not ($Y_{ij} = 0$) during occasion j. The observed data include capture counts for individuals $i = 1, 2, \ldots, n$. The remaining individuals $i = n + 1, \ldots, N$, who were never detected, contribute unobserved zero histories and are not present in the recorded sample.

Individual	Occasion j				
i	1	2	...	T	
1	$Y_{1,1}$	$Y_{1,2}$	...	$Y_{1,T}$	$Y_1.$
2	$Y_{2,1}$	$Y_{2,2}$	...	$Y_{2,T}$	$Y_2.$
3	$Y_{3,1}$	$Y_{3,2}$	...	$Y_{3,T}$	$Y_3.$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	$Y_{n,1}$	$Y_{n,2}$	...	$Y_{n,T}$	$Y_n.$
$n+1$	$Y_{n+1,1}$	$Y_{n+1,2}$	...	$Y_{n+1,T}$	$Y_{n+1}.$
$n+2$	$Y_{n+2,1}$	$Y_{n+2,2}$	...	$Y_{n+2,T}$	$Y_{n+2}.$
$n+3$	$Y_{n+3,1}$	$Y_{n+3,2}$	...	$Y_{n+3,T}$	$Y_{n+3}.$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$N-1$	$Y_{N-1,1}$	$Y_{N-1,2}$	...	$Y_{N-1,T}$	$Y_{N-1}.$
N	$Y_{N,1}$	$Y_{N,2}$	...	$Y_{N,T}$	$Y_N.$

4.2.1 Counting Distribution

From the capture-recapture histories, a count distribution is formed by creating a frequency table, such as Table 4.2, which summarizes how frequently each unit was identified. In this context, f_x indicates the number of individuals captured exactly x times during the study period, while $n = \sum_{x=1}^{T} f_x$ denotes the number of individuals captured at least once, ,where $x = 1, 2, \ldots, T$.

TABLE 4.2: Frequency table for capture counts with T capture occasions/sources.

$\boldsymbol{x}$	0	1	2	...	T
$\boldsymbol{f_x}$	f_0	f_1	f_2	...	f_T

Individuals who were never captured contribute to the zero count, f_0, which is absent from the data and is known as zero-truncated count data. Statistically, this entails working with a zero-truncated count distribution. Suppose $P(x)$ is a suitable distribution

model for the capture counts of each individual in the population. Here, $p_0 = P(X = 0)$ denotes the probability that an individual is not detected across all T capture occasions. This parameter can be interpreted as the proportion of the population that remains unobserved and is a key component in estimating the number of unseen individuals.

To estimate the number of missing observation f_0, the Horvitz-Thompson estimator is utilized:

$$\hat{N} = \frac{n}{1 - p_0}, \tag{4.1}$$

which further leads to

$$\hat{f}_0 = n\frac{p_0}{1 - p_0}.$$

It is important to note that

$$\begin{aligned}\mathbb{E}(\hat{N}) &= \frac{1}{1 - p_0}\mathbb{E}(n) \\ &= \frac{1}{1 - p_0}N(1 - p_0) \\ &= N.\end{aligned}$$

This holds under the assumption that p_0 is known. However, in most cases, p_x is unknown and needs to be estimated. To achieve this, a distributional model is assumed, introducing parameter θ, such that $p_x = p_x(\theta)$.

4.2.2 Zero-Truncated Counting Distribution

In capture-recapture studies, the total population size N and the frequency of undetected individuals f_0 are unknown, as individuals not captured in any occasion remain unobserved. Consequently, the data consist of a zero-truncated count of X, with $x = 1, \ldots, T$.

Assume that X follows a discrete distribution with pmf $p_x(\theta) = P(X = x \mid \theta)$, where θ is the model parameter. Because observations with $X = 0$ are unobserved, the observed data likelihood is based on the zero-truncated distribution:

$$p_x^+(\theta) = \frac{p_x(\theta)}{1 - p_0(\theta)}, \quad x = 1, \ldots, T$$

where $p_0(\theta)$ is the probability of zero detection. The observed data likelihood is then:

$$L(\theta) = \prod_{x=1}^{T}\left[\frac{p_x(\theta)}{1 - p_0(\theta)}\right]^{f_x},$$

which depends only on the observed frequencies $f_1, f_2, \ldots, f_T$. Closed-form solutions for MLE of zero-truncated count distributions are generally unavailable. However, these estimators can be derived using the EM algorithm, as outlined in the works of Böhning et al. (2018) and Dempster et al. (1977).

EM Algorithm for Zero-Truncated Counting Distribution

Step 0. **Initialisation:** Initialize the parameter θ to some arbitrarily chosen value $\hat{\theta}$.

Step 1. **E-step:** Compute the expected frequency of undetected individuals (i.e., those with $X = 0$), conditional on the observed data and current parameter estimate:

$$\hat{f}_0 = \mathbb{E}(f_0 \mid f_1, \ldots, f_T, \hat{\theta}) = n \frac{p_0(\hat{\theta})}{1 - p_0(\hat{\theta})}.$$

Step 2. **M-step:** Construct the complete frequency table $(\hat{f}_0, f_1, \ldots, f_T)$, and compute the MLE $\hat{\theta}$ by maximising the complete-data likelihood:

$$\ell_{\text{complete}}(\theta) = \sum_{x=0}^{T} f_x \log p_x(\theta).$$

Step 3. Iterate E- and M-steps until convergence.

This algorithm aims to compute $\hat{\theta}$ that optimises the likelihood of truncated densities:

$$\prod_{x=1}^{T} \left[\frac{p_x(\theta)}{1 - p_0(\theta)} \right]^{f_x}.$$

It is important to note that the specific computation of the parameter updates in the M-step depends on the chosen model and will be described in the following sections.

4.3 Likelihood Based on Binomial Distribution

In scenarios where capture probability is constant across individuals and capture occasions, the number of captures per individual, denoted X, can be modelled using a Binomial distribution. Specifically,

$$X \sim \text{Bin}(T, \theta).$$

The Binomial pmf is given by:

$$p_x = P(X = x) = \binom{T}{x} \theta^x (1 - \theta)^{T-x}, \quad x = 0, 1, 2, \ldots, T, \tag{4.2}$$

where T is the total number of capture occasions and θ is the homogeneous capture probability. Given the nature of capture-recapture studies, only individuals that are captured at least once ($X > 0$) are observed, resulting in a zero-truncated Binomial distribution. As a closed-form solution for MLE of this distribution is unavailable, the estimators can be obtained using the EM algorithm.

The full-data likelihood is given by:

$$L(\theta) = \prod_{x=0}^{T} \left[\binom{T}{x} \theta^x (1-\theta)^{T-x} \right]^{f_x}.$$

Taking the natural logarithm yields the complete-data log-likelihood function:

$$\begin{aligned}
\ell(\theta) &= \log \left\{ \prod_{x=0}^{T} \left[\binom{T}{x} \theta^x (1-\theta)^{T-x} \right]^{f_x} \right\} \\
&= \sum_{x=0}^{T} f_x \log \left[\binom{T}{x} \theta^x (1-\theta)^{T-x} \right] \\
&= \sum_{x=0}^{T} f_x \log \binom{T}{x} + \sum_{x=0}^{T} x f_x \log \theta + \sum_{x=0}^{T} (T-x) f_x \log(1-\theta) \\
&= \sum_{x=0}^{T} f_x \log \binom{T}{x} + \sum_{x=0}^{T} x f_x \log \theta + NT \log(1-\theta) - \sum_{x=0}^{T} x f_x \log(1-\theta). \qquad (4.3)
\end{aligned}$$

Differentiating Equation (4.3) with respect to θ and solving for zero gives the MLE:

$$\begin{aligned}
\hat{\theta} &= \frac{\sum_{x=0}^{T} x f_x}{NT} \\
&= \frac{1}{(n+f_0)T} \sum_{x=0}^{T} x f_x. \qquad (4.4)
\end{aligned}$$

The probability of non-detection under the Binomial pmf is:

$$\begin{aligned}
P(X=0) = p_0 &= \binom{T}{0} \theta^0 (1-\theta)^T \\
p_0 &= (1-\theta)^T.
\end{aligned}$$

Given $\hat{\theta}$, the estimated number of undetected individuals is:

$$\hat{f}_0 = n \frac{(1-\hat{\theta})^T}{1-(1-\hat{\theta})^T}.$$

This leads to the set-up of the following EM algorithm to iteratively estimate θ based on a zero-truncated Binomial distribution.

EM Algorithm for MLE with Zero-Truncated Binomial Distribution

Step 0. **Initialisation:** Initialise the parameter θ to some arbitrarily chosen values. A practical choice is:

$$\hat{\theta} = \frac{\sum_{x=1}^{T} x f_x}{n\ T},$$

which is based on the observed sample mean.

Step 1. **E-step:** Compute the expected number of unobserved individuals:

$$\hat{f}_0 = n\frac{(1-\hat{\theta})^T}{1-(1-\hat{\theta})^T}.$$

Step 2. **M-step:** Use the imputed full frequency table $(\hat{f}_0, f_1, \ldots, f_T)$ to update the parameter estimate:

$$\hat{\theta} = \frac{1}{(n+\hat{f}_0)T}\sum_{x=0}^{T} x f_x.$$

Step 3. Iterate E- and M-steps until convergence.

4.4 The Idea of Sampling Effort

To study the sampling effort required in a Schnabel census, let n denote the number of observed units out of unknown N. Let η be the random variable representing the number of observed units in the sample. Assuming each unit is independently observed with probability $1-p_0$, η follows a Binomial distribution:

$$\eta \sim \text{Bin}(N, 1-p_0),$$

and the pmf and variance are given by:

$$P(\eta = n) = \binom{N}{n} p_0^{N-n}(1-p_0)^n,$$

and

$$\text{Var}(\eta) = Np_0(1-p_0),$$

respectively. Applying the Horvitz–Thompson estimator in Equation (4.1), $\text{Var}(\hat{N})$ is derived as:

$$\begin{aligned}\text{Var}(\hat{N}) &= \frac{1}{(1-p_0)^2}\text{Var}(\eta) \\ &= \frac{1}{(1-p_0)^2}N(1-p_0)p_0 \\ &= N\frac{p_0}{1-p_0}. \end{aligned} \tag{4.5}$$

From Equation (4.5), the variance of the estimated population size is equal to the product of the actual N and the odds of missing observation. Hence, the uncertainty in $\hat{N}$ can be controlled by adjusting the width of the $(1-\alpha)\%$ confidence interval for N,

$$\hat{N} \pm z_{\alpha/2}\sqrt{N\frac{p_0}{1-p_0}},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

For illustration, consider a case where $N = 100$, and $p_0 = 0.5$. Based on the variance formula, then $\text{Var}(\hat{N}) = N$, implying that the standard deviation of $\hat{N}$ is $\sqrt{N}$. This suggests that, under this setting, a 95% confidence interval for N based on $\hat{N}$ would be approximately $\hat{N} \pm 1.96\sqrt{N} \approx (80, 120)$, which seems reasonable, according to the rough rule of thumb suggested by Pollock et al. (1990).

Generalizing, assume the researcher is willing to accommodate variations in population size within $\kappa \times 100\%$ above/below the true N at $(1-\alpha)$ confidence level. The margin of the confidence interval will be equal to:

$$z_{\alpha/2}\sqrt{N\frac{p_0}{1-p_0}} = \kappa N,$$

hence

$$p_0 = \frac{\left(\kappa/z_{\alpha/2}\right)^2 N}{1+\left(\kappa/z_{\alpha/2}\right)^2 N}.$$

Table 4.3 lists down the associated p_0 for various N and κ while Figure 4.1 depicts the relationship between p_0 and κ for various N when $1-\alpha$=0.95. The graph demonstrates that keeping p_0 below 0.5 limits the uncertainty of the estimate to within 20% above or below the true N for $N \geq 100$. Thus, p_0 is an effective tool for controlling uncertainty. Replacing $p_0 = (1-\theta)^T$ in Equation (4.5) gives

$$\text{Var}(\hat{N}) = N\frac{(1-\theta)^T}{1-(1-\theta)^T}. \tag{4.6}$$

Reducing the population size N to lower the variance in Equation (4.6) is not possible as we have no information on the value of N. However, adjusting T to a large value can cause $(1-\theta)^T \to 0$ for a positive $\theta \in (0,1)$, thereby reducing estimation variance. Since

$$p_0 = (1-\theta)^T, \tag{4.7}$$

solving for T in Equation (4.7) yields

$$T = \frac{\log(p_0)}{\log(1-\theta)}. \tag{4.8}$$

TABLE 4.3: Proportion of undetected individuals (p_0) for various population sizes (N) and relative margins of error (κ) at a 95% confidence level ($1 - \alpha = 0.95$).

N	κ	p_0
50	0.50	0.76
	0.40	0.68
	0.30	0.54
	0.25	0.45
	0.20	0.34
	0.10	0.12
	0.05	0.03
100	0.50	0.87
	0.40	0.81
	0.30	0.70
	0.25	0.62
	0.20	0.51
	0.10	0.21
	0.05	0.06
200	0.50	0.93
	0.40	0.89
	0.30	0.83
	0.25	0.76
	0.20	0.68
	0.10	0.34
	0.05	0.12
400	0.50	0.96
	0.40	0.94
	0.30	0.90
	0.25	0.87
	0.20	0.81
	0.10	0.51
	0.05	0.21
1000	0.50	0.98
	0.40	0.98
	0.30	0.96
	0.25	0.94
	0.20	0.91
	0.10	0.72
	0.05	0.39

Hence, T depends on p_0 and θ. Researchers can leverage this property to choose the appropriate T by specifying the desired capture success rate, i.e., $1 - {p_0}^*$. For instance, when $1 - {p_0}^* = 0.5$, the required T to achieved this desired capture success rate for various θ is as shown in Table 4.4. The relationship between T, p_0, and θ are presented in Equation (4.8) and illustrated in Figure 4.2.

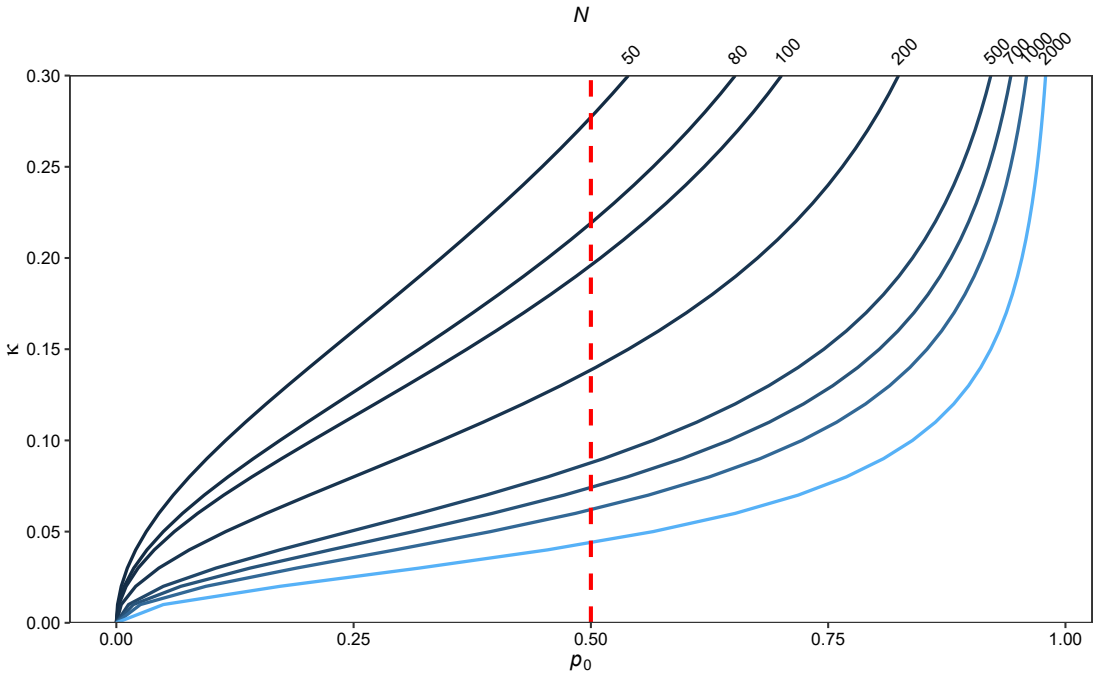


FIGURE 4.1: Relationship between population size N, relative margin of error κ, and the proportion of undetected individuals p_0 at a 95% confidence level. The red dashed line indicates the reference point where $p_0 = 0.5$.

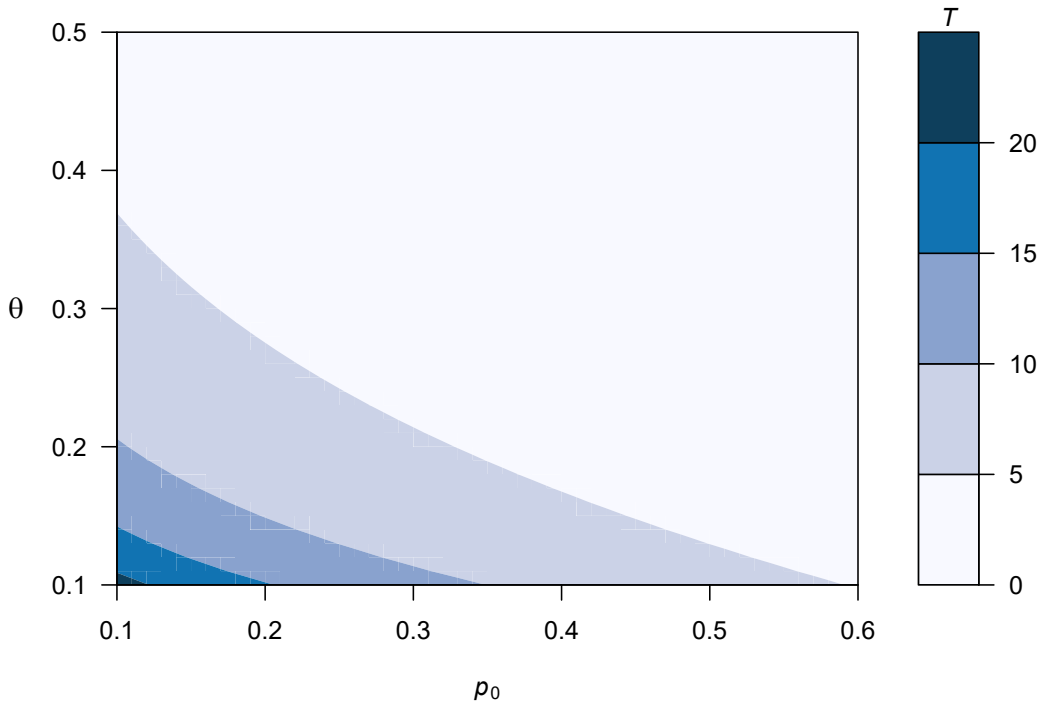


FIGURE 4.2: Contour plot of the required number of capture occasions T in relation to the proportion of undetected individuals p_0 and the capture probability θ.

TABLE 4.4: Required sampling occasions T for different capture probabilities θ to achieve a 50% of capture success rate.

θ	0.1	0.2	0.3	0.4	0.5
T	7	3	2	1	1

4.5 Allowing Heterogeneity: Mixture Models

The previous section employed a simple Binomial model to analyse capture counts. This model assumes that the observations are independent and the parameters are homogeneous. This section extends the framework to more flexible models using the Beta-Binomial distribution and the Binomial mixture distribution.

4.5.1 Beta-Binomial Distribution

4.5.1.1 Maximum Likelihood Estimator

In situations where the capture probability varies across individuals, the capture counts X are suitably modelled using a Beta-Binomial distribution. In this model, the individual capture probabilities θ are assumed to vary randomly across the population according to a Beta(α, β) distribution.

The pmf is for a Beta-Binomial distribution is given by:

$$p_x = P(X = x) = \binom{T}{x} \frac{B(\alpha + x, T + \beta - x)}{B(\alpha, \beta)}, \quad x = 0, 1, \ldots, T.$$

where $B(\cdot, \cdot)$ is the Beta function, defined in term of Gamma function as:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

The probability of non-detection is given by:

$$p_0 = \frac{B(\alpha, T + \beta)}{B(\alpha, \beta)},$$

leading to the estimate

$$\hat{f}_0 = n \left[\frac{B(\alpha, T + \beta)}{B(\alpha, \beta) - B(\alpha, T + \beta)} \right].$$

The complete data likelihood is given by:

$$L(\alpha, \beta) = \left[\frac{B(\alpha, T + \beta)}{B(\alpha, \beta)} \right]^{f_0} \times \prod_{x=1}^{T} \left[\binom{T}{x} \frac{B(\alpha + x, T + \beta - x)}{B(\alpha, \beta)} \right]^{f_x},$$

and the log-likelihood

$$\begin{aligned}\ell(\alpha,\beta) &= f_0 \log\left[\frac{B(\alpha,T+\beta)}{B(\alpha,\beta)}\right] + \sum_{x=1}^{T} f_x \log\left[\binom{T}{x}\frac{B(\alpha+x,T+\beta-x)}{B(\alpha,\beta)}\right] \\ &= f_0\left[\log B(\alpha,T+\beta) - \log B(\alpha,\beta)\right] \\ &\quad + \sum_{x=1}^{T} f_x\left[\log\binom{T}{x} + \log B(\alpha+x,T+\beta-x) - \log B(\alpha,\beta)\right]. \end{aligned} \tag{4.9}$$

To maximise the log-likelihood in Equation (4.9), a quasi-Newton method with L-BFGS-B algorithm is run using the `optim` function in `R`. This sets up the EM algorithm to find the MLE for the zero-truncated Beta-Binomial distribution.

EM Algorithm for MLE with Zero-Truncated Beta-Binomial Distribution

Step 0. **Initialisation:** Initialise the parameters α and β to some arbitrarily chosen values $\hat{\alpha}$ and $\hat{\beta}$. Reasonable initial estimates can be chosen based on the method of moments from the sample mean and variance. If no prior knowledge is available, a default of $\hat{\alpha} = \hat{\beta} = 1$ can be used.

Step 1. **E-step:** Compute the expected number of unobserved individuals f_0 given the current parameters:

$$\hat{f}_0 = n\left[\frac{B(\hat{\alpha},T+\hat{\beta})}{B(\hat{\alpha},\hat{\beta}) - B(\hat{\alpha},T+\hat{\beta})}\right].$$

Step 2. **M-step:** Maximise the complete-data log-likelihood Equation (4.9) using to obtain updated estimates:

$$(\hat{\alpha},\hat{\beta}) = \arg\max_{\alpha,\beta} \ell(\alpha,\beta).$$

Step 3. Iterate E- and M-step until convergence is achieved.

4.5.1.2 Sampling Effort Determination

If the capture counts X follows a BetaBinomial(T,α,β) distribution, the optimal sampling effort T can be determined using the relationship

$$p_0 = \frac{B(\alpha,T+\beta)}{B(\alpha,\beta)}, \tag{4.10}$$

based on the desired capture success rate, $1-{p_0}^*$. No closed-form solution exists for T in Equation (4.10). However, it can be solved using numerical methods such as

Newton-Raphson. Starting with an initial guess $T^{(0)}$, iteratively apply

$$T^{(r+1)} = T^{(r)} - \frac{h(T^{(r)})}{h'(T^{(r)})},$$

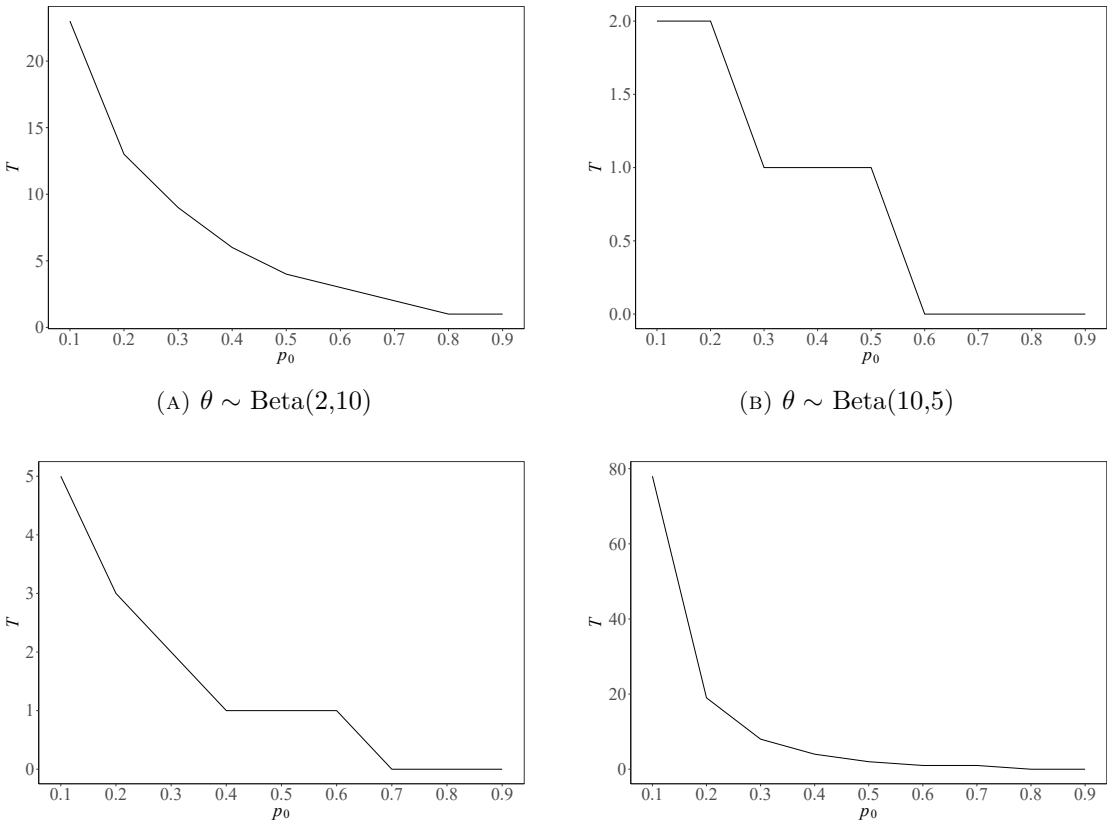
where

$$h(T) = \frac{B(\alpha, T+\beta)}{B(\alpha,\beta)} - {p_0}^*,$$

and

$$h'(T) = \frac{B(\alpha, T+\beta)}{B(\alpha,\beta)} \left[\psi(T+\beta) - \psi(\alpha+T+\beta)\right].$$

Here, $\psi(\cdot)$ denotes a digamma function. Figure 4.3 displays the required sampling efforts for various combinations of α and β values.



(A) $\theta \sim$ Beta(2,10)

(B) $\theta \sim$ Beta(10,5)

(C) $\theta \sim$ Beta(2,2)

(D) $\theta \sim$ Beta(0.5,1)

FIGURE 4.3: Required sampling effort T under different assumptions for the distribution of capture probability $\theta \sim \text{Beta}(\alpha, \beta)$.

4.5.2 Binomial Mixture Distribution

4.5.2.1 Maximum Likelihood Estimator

In many practical situations, the capture probability varies among individuals in a more complex manner. To account for this heterogeneity, the number of captures, X per individual can be modelled as a mixture of Binomial distributions with k heterogeneous Binomial components:

$$p_x = P(X = x) = \sum_{j=1}^{k} w_j \, \text{Bino}(x \mid T, \theta_j),$$

where

$$\text{Bino}(x|T, \theta_j) = \binom{T}{x} \theta_j^x (1 - \theta_j)^{T-x}, \quad j = 1, 2, \ldots, k.$$

The mixing distribution, $\Theta = \{w_1, w_2, \ldots, w_k, \ \theta_1, \theta_2, \ldots, \theta_k\}$, assigns non-negative weights w_j to θ_j, with $\sum_{j=1}^{k} w_j = 1$. Similar to previous models, only individuals capture at least once ($X > 0$) are observed, resulting in zero-truncated data.

The probability of non-detection is given by:

$$p_0 = \sum_{j=1}^{k} w_j (1 - \theta_j)^T,$$

leading to the estimate of the unobserved individuals:

$$\hat{f}_0 = n \left[\frac{\sum_{j=1}^{k} w_j \, (1 - \theta_j)^T}{1 - \sum_{j=1}^{k} w_j \, (1 - \theta_j)^T} \right].$$

To fit a zero-truncated Binomial mixture distribution to the capture data, a nested EM algorithm adapted from Böhning et al. (2005) to estimate the parameters Θ. The following steps outline the algorithm:

EM Algorithm for MLE with Zero-Truncated Binomial Mixture Distribution

Step 0. **Initialisation:** Set initial values for the full parameter vector with

$$\hat{\Theta} = (\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_k, \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k).$$

The recommended choices is to initialise $\hat{w}_j = 1/k$ and set $\hat{\theta}_j$ evenly spaced over $(0.1, 0.9)$.

Step 1. **E-step:** Compute the expected number of unobserved individuals:

$$\hat{f}_0 = n \left[\frac{\sum_{j=1}^{k} \hat{w}_j \, (1 - \hat{\theta}_j)^T}{1 - \sum_{j=1}^{k} \hat{w}_j \, (1 - \hat{\theta}_j)^T} \right].$$

Define the expected value of observation i belongs to component j:

$$e_{ij} = \frac{w_j \, \text{Bino}(i|T, \theta_j)}{\sum_{j=1}^{k} w_j \, \text{Bino}(i|T, \theta_j)}.$$

Step 2. **M-step:** Use the complete frequency table $\hat{f}_0, f_1, f_2, \ldots, f_T$ to maximise the complete-data log-likelihood:

$$\ell(\Theta) = \sum_{i=0}^{T} \sum_{j=1}^{k} f_i \, e_{ij} \, \log \left[w_j \binom{T}{i} \theta_j^i (1 - \theta_j)^{T-i} \right],$$

which leads to the following estimates updates:

$$\hat{w}_j = \frac{\sum_{i=0}^{T} f_i \, \hat{e}_{ij}}{\sum_{i=0}^{T} f_i},$$
$$\hat{\theta}_j = \frac{\sum_{i=0}^{T} i \, f_i \, \hat{e}_{ij}}{\sum_{i=0}^{T} T \, f_i \, \hat{e}_{ij}}.$$

Step 3. Repeat E- and M- steps until convergence is achieved.

The EM framework is structured in two levels:

- The main EM loop alternate between the E-step (which estimates $\hat{f}_0$ and updates the full data) and the M-step (which maximises the complete-data log-likelihood),
- An inner EM step within the main M-step performs component-wise updates of the latent class parameters (w_j, θ_j) using the non-parametric maximum likelihood estimator (NPMLE) (McLachlan and Peel, 2000).

The inner EM updates for the mixture parameters w_j and θ_j are executed once per iteration of the main EM algorithm. After this inner update, the algorithm returns to the main E-step, and the overall EM procedure continues iterate between E- and M-steps until convergence (Böhning et al., 2005).

4.5.2.2 Sampling Effort Determination

If the capture count follows a Binomial mixture distribution, the sampling effort, T, can be determined using the relationship:

$$p_0 = \sum_{j=1}^{k} w_j (1-\theta_j)^T. \tag{4.11}$$

By setting the desired capture success rate, $1 - {p_0}^*$, the required number of capture occasions, T, can be determined by finding the root T of Equation (4.11), which provides no closed-form solution. However, it can be solved using numerical methods such as Newton-Raphson. Starting with an initial guess $T^{(0)}$, iteratively apply the formula:

$$T^{(r+1)} = T^{(r)} - \frac{h(T^{(r)})}{h'(T^{(r)})},$$

where

$$h(T) = \sum_{j=1}^{k} w_j (1-\theta_j)^T - {p_0}^*,$$

and

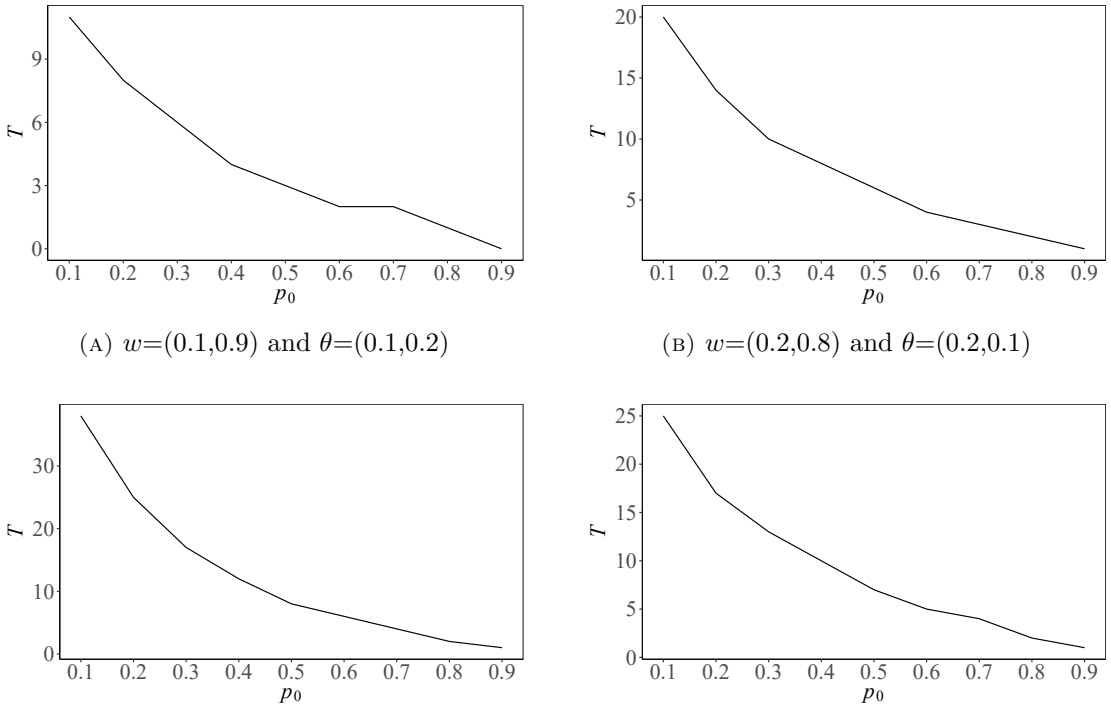
$$h'(T) = \sum_{j=1}^{k} \log(1-\theta_j) w_j (1-\theta_j)^T.$$

Repeat this iteration until the change between successive values of T is minimal, indicating that an approximate root has been found.

To illustrate the effect of population heterogeneity on required sampling effort, four representative scenarios are presented in Figure 4.4, each based on a two-component Binomial mixture model. These scenarios reflect different structures of heterogeneity, defined by specific combinations of weights $\boldsymbol{w}$ and capture probabilities $\boldsymbol{\theta}$.

Figure 4.4(A) captures a population with a small elusive subgroup (10% of individuals with low detectability $\theta = 0.1$) and a dominant group with moderate detectability ($\theta = 0.2$). This reflects situations where a minority, such as juveniles or trap-averse individuals, is harder to detect, a phenomenon widely reported in both wildlife studies and clinical applications. In Figure 4.4(B), 20% of the population is moderately detectable ($\theta = 0.2$) while 80% is harder to capture ($\theta = 0.1$). This reversed structure creates a skew toward low detectability in the population, possibly representing conditions like widespread trap shyness, or where behavioural or environmental factors reduce capture likelihood in the majority group.

Figure 4.4(C) illustrates a stronger degree of heterogeneity, where 30% of the population has moderate detectability ($\theta = 0.2$) and 70% are highly elusive ($\theta = 0.05$). The large gap in detectability between the two groups models populations with strong trap avoidance, such as those observed in rare species or in heavily disturbed settings. Figure 4.4(D) shows a more balanced scenario, with equal proportions of individuals having slightly different detectabilities ($\theta = 0.1$ and $\theta = 0.08$). Such subtle variation is common in



(A) w=(0.1,0.9) and θ=(0.1,0.2)

(B) w=(0.2,0.8) and θ=(0.2,0.1)

(C) w=(0.3,0.7) and θ=(0.2,0.05)

(D) w=(0.5,0.5) and θ=(0.1,0.08)

FIGURE 4.4: Required sampling efforts, T, for different combination of w and θ.

structured populations (e.g., adults vs. subadults) where slight differences in behaviour or exposure influence detection probabilities.

Together, these scenarios demonstrate how varying levels and forms of heterogeneity influence the sampling effort required for reliable population estimation, as displayed in Figure 4.4.

4.6 Cottontail Rabbit in Ohio

The study by Edwards and Eberhardt (1967), later republished by Chao (1987), involved an investigation on a restricted population of known size using live-trapping techniques. Within a 4-acre rabbit-proof enclosure, 135 wild cottontail rabbits were subject to live-trapping over 18 consecutive nights. Among them, 76 were captured at least once. The recorded capture frequencies (f_1 to f_7) were as follows:

$$43,16,8,6,0,2,1$$

Four zero-truncated models were fitted on the data, namely the Binomial (ZTB), Beta-Binomial (ZTBB), two-component Binomial mixture (ZTMB2), and three-component Binomial mixture (ZTMB3). Table 4.5 presents the estimated population sizes and the results of the model fittings to the data.

TABLE 4.5: Observed and expected frequencies of capture counts (f_x and $\hat{f}_x$) for cottontail rabbit data under different zero-truncated models: ZTB (Binomial), ZTBB (Beta-Binomial), ZTMB2 (Two-component Binomial mixture), and ZTMB3 (Three-component Binomial mixture). Also shown are estimated population size $\hat{N}$, degrees of freedom (df), and p-values from goodness-of-fit tests.

x	f_x	$\hat{f}_{x\{ZTB\}}$	$\hat{f}_{x\{ZTBB\}}$	$\hat{f}_{x\{ZTMB2\}}$	$\hat{f}_{x\{ZTMB3\}}$
1	43	34	43	43	43
2	16	25	17	16	16
3	8	12	8	8	9
4	6	4	4	5	4
5	0	1	2	3	2
6	2	0	1	1	1
7	1	0	1	0	1
$\hat{N}$	135	98	316	143	219
df		3	2	1	-
p		0.011	0.569	0.375	-

4.6.1 Zero-Truncated Binomial

Assuming the capture counts follow a Binomial distribution with homogeneous capture probability across all individuals, a zero-truncated Binomial distribution was fitted to the observed data. An EM algorithm was used to obtain the MLE for the capture probability θ and the unobserved frequency f_0. The algorithm was initiated with an initial value of $\hat{\theta}_0 = 0.5$.

The resulting estimates were $\hat{\theta} = 0.081$ and $\hat{N} = 98$, which is notably lower than the actual population size of $N = 135$. The Chi-square goodness-of-fit test yielded a p-value of 0.011 ($\chi^2 = 11.093, df = 3$), indicating a poor fit between the model and the observed capture frequencies. To ensure the validity of the test, several cells from f_4 to f_{18} were combined due to to small expected values (< 5).

4.6.2 Zero-Truncated Beta-Binomial

Assuming that the capture probability θ follows a beta distribution, an EM algorithm was employed to obtain the MLE of the parameters α and β for the zero-truncated Beta-Binomial distribution. The resulting parameter estimates were $\hat{\alpha} = 0.273$ and $\hat{\beta} = 10.631$. The Chi-square goodness-of-fit test produced a p-value of 0.569 ($\chi^2 = 1.126, df = 2$), indicating a good fit to the observed capture data. However, the estimated population size $\hat{N} = 316$ was substantially higher than the actual size of $N = 135$, suggesting potential overestimation.

4.6.3 Zero-Truncated Binomial mixture

Fitting a two-component Binomial mixture model on the cottontail rabbits data resulting in parameter estimates of $\hat{w}_1 = 0.840$, $\hat{w}_2 = 0.160$, $\hat{\theta}_1 = 0.033$, and $\hat{\theta}_2 = 0.175$. The Chi-square goodness-of-fit test resulted in a p-value of 0.375 ($\chi^2 = 0.786, df = 1$), indicating a good fit to the data.

Meanwhile, fitting a three-component Binomial mixture model results in $\hat{w}_1 = 0.766$, $\hat{w}_2 = 0.203$, $\hat{w}_3 = 0.031$, $\hat{\theta}_1 = 0.012$, $\hat{\theta}_2 = 0.096$, $\hat{\theta}_3 = 0.243$. However, the Chi-square goodness-of-fit test could not be performed, as several frequency cells (f_5 to f_{18}) were merged due to small expected values, resulting in zero degrees of freedom.

4.6.4 Model Evaluation

Table 4.6 compares the fit of four zero-truncated models to cottontail rabbit capture-recapture data using log-likelihood, AIC, BIC, and the resulting population estimates. Among the four models, Binomial models has the highest AIC and severely underestimates the population size ($\hat{N} = 98$ compared to the true $N = 135$), indicating that the simple Binomial model fails to capture heterogeneity. Beta-Binomial achieves the lowest AIC and BIC values but greatly overestimates the population size ($\hat{N} = 316$), suggesting over-dispersion. Both the two-component and three-component Binomial mixture models produce estimates closer to the true value ($\hat{N} = 143$ and $\hat{N} = 219$, respectively). However, the two-component model is preferred because it attains lower AIC and BIC than three-component, avoiding unnecessary model complexity. Overall, two-component Binomial mixture model provides the best balance between goodness-of-fit, model parsimony, and accurate estimation of the true population size.

TABLE 4.6: Comparison of zero-truncated models fitted to cottontail rabbit data: ZTB (Binomial), ZTBB (Beta-Binomial), ZTMB2 (Two-component Binomial mixture), and ZTMB3 (Three-component Binomial mixture).

distribution	$\hat{N}$	**log-likelihood**	**AIC**	**BIC**
ZTB	98	-105.109	212.218	214.549
ZTBB	316	-97.763	199.525	204.187
ZTMB2	143	-97.748	201.496	208.489
ZTMB3	219	-97.524	205.049	216.702

4.6.5 Sampling Efforts Determination

To determine the required number of capture occasions T for the zero-truncated two-component Binomial mixture model, the Newton-Raphson method was employed as described in Section 4.5.2. The parameter estimates used in the calculation were $\hat{w}_1 =$

0.840, $\hat{w}_2 = 0.160$, $\hat{\theta}_1 = 0.033$, and $\hat{\theta}_2 = 0.175$. Table 4.7 presents the required values of T for achieving various levels of capture success rates, $1 - {p_0}^*$.

The 25th and 75th percentiles for T were derived from bootstrap samples based on cottontail rabbit capture–recapture data. Starting with the true value $N = 135$, f_0 was calculated to be 59. Capture count probabilities $\hat{p}_x = \dfrac{f_x}{N}$ for $x = 0, 1, 2, \ldots, 7$ were then computed. Subsequently, 5000 bootstrap samples were generated by drawing capture counts from a multinomial distribution with probabilities $(\hat{p_0}, \hat{p_1}, \hat{p_2}, \cdots, \hat{p_7})$. For each sample, the zero count f_0 was excluded, and the EM algorithm was applied to the observed frequencies $(f_1, f_2, \ldots, f_7)$ to estimate the parameters $(w_1, w_2, \theta_1, \theta_2)$. These parameter estimates were then used to calculate the required sampling efforts T.

The resulting distribution of T values from the bootstrap replicates was summarised to obtain the 25th and 75th percentiles, which provide an indication of the uncertainty in estimating the required sampling effort. These percentiles are reported alongside the point estimates in Table 4.7.

TABLE 4.7: Required number of capture occasions T for different levels of capture success rate $1 - {p_0}^*$, along with the 25th - 75th percentile range derived from bootstrap estimates.

Desired capture success rate, $1 - {p_0}^*$	Requires sampling effort, T	25% - 75% percentile
0.4	11	9 - 23
0.5	**16**	**12 - 35**
0.6	22	17 - 50
0.7	31	23 - 72
0.8	43	32 - 98
0.9	63	47 - 154

For a relatively high ${p_0}^*$ value of 0.6, indicating a lower desired success rate of 40%, 11 capture occasions are sufficient. As the success rate requirement increases, the number of capture occasions also increases significantly. For a success rate of 50% (${p_0}^* = 0.5$), 16 capture occasions are required. This corresponds roughly to the value of T actually used in the cottontail rabbits study. Further increasing the success rate to 60% (${p_0}^* = 0.4$) requires 22 capture occasions. For a 70% success rate (${p_0}^* = 0.3$), 31 capture occasions are required. An 80% success rate (${p_0}^* = 0.2$) needs 43 capture occasions. Finally, a 90% success rate (${p_0}^* = 0.1$) requires 63 capture occasions.

4.7 Discussion and Conclusion

This chapter examines how to choose sampling effort in Schnabel studies for accurate population size estimates. The analysis started with a simple Binomial model. It showed a clear connection between capture occasions, capture probability, and success rate. More flexible models, like the Beta-Binomial and Binomial mixture, were then used to account for differences in detection among individuals.

A common guideline suggests that population size estimate should be within 20% of the true size (Pollock et al., 1990). For populations of 100 or more, this accuracy is achievable with a 50% success rate ($1 - p_0 = 0.5$). Researchers can set a target success rate ($1 - p_0^*$) and determine the required number of sampling occasions.

The results show a drastic increase in required capture occasions as the target success rate goes up. For example, 11 occasions give a 40% success rate, but 43 are needed for 80% in the cottontail rabbits study. This means that additional sampling provides diminishing returns. It highlights the trade-off between better estimation and higher costs. These findings agree with past research (Kordjazi et al., 2016), which also found that more effort leads to smaller improvements in precision.

Model comparisons using AIC and BIC scores in Table 4.6 show that both Beta-Binomial and Binomial mixture models fit the cottontail rabbit data well. However, the Beta-Binomial model can perform poorly when zero counts are excluded (Böhning, 2015). It may overestimate missed individuals because its assumptions do not always match real-world variation. This can lead to unreliable results.

Besides model choice, capture probability also affects uncertainty and required sampling effort (Burnham et al., 1987; Sanderlin et al., 2014). Lower catchability means more capture occasions are needed for the same success rate. Improving detection methods can reduce effort and increase precision (Papadatou et al., 2012).

Both this study and Xi et al. (2008) highlight the importance of capture success rate. Xi et al. determine the minimum success rate by setting a limit on variance, but their method requires knowing the population size in advance. This study instead treats p_0 as a design parameter and calculates the needed capture occasions T directly. This allows planning based on realistic detection assumptions, including heterogeneity

In summary, this chapter offers practical insights into the relationship between capture success rate and sampling effort in Schnabel census designs. The proposed framework provides a structured and adaptable method for determining the optimal number of capture occasions, T, based on a target capture success rate. The relationship $p_0 = (1 - \theta)^T$, along with its extensions for heterogeneous models, offers researchers a clear and accessible way to align sampling effort with the level of uncertainty they are willing to accept.

Chapter 5

Multiple Captures Model with Hierarchical Structure

5.1 Introduction

This chapter extends the hierarchical Lincoln–Petersen framework to the Schnabel census by introducing a hierarchical sampling design tailored to studies involving multiple capture occasions. In this approach, the total sampling effort is organized into a nested structure: each primary capture occasion comprises several sub-occasions, such as individual trapping days or trap deployments. These primary occasions are separated by periods with no sampling, allowing for individual movement between capture events. This nested arrangement forms a hierarchical structure, with sub-occasions embedded within primary occasions, as illustrated in Figure 5.1.

Adopting this hierarchical design enables more efficient allocation of effort across occasions, with the aim of improving the precision of population estimates in Schnabel-type capture–recapture studies. The subsequent sections explore how different strategies for sub-occasion allocation influence estimation outcomes within this framework.

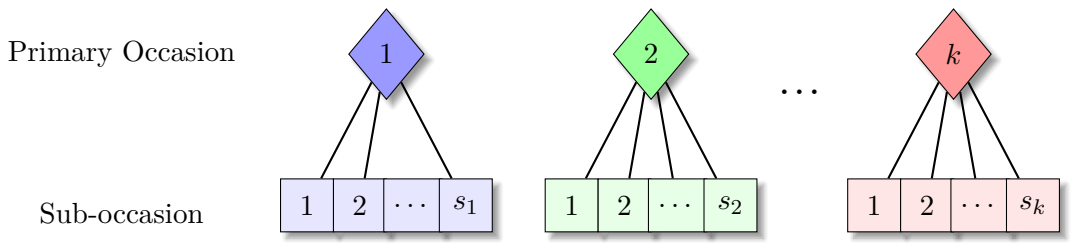


FIGURE 5.1: Hierarchical structure of a Schnabel census. Each of the k primary occasions represents a distinct capture period, typically separated by intervals without sampling. Within each primary occasion j $(j = 1, 2, \ldots, k)$, s_j sub-occasions correspond to individual trapping days or the number of traps deployed.

5.2 Likelihood Analysis

Consider a Schnabel census with k capture occasions, each with a distinct sampling effort s_j for $j = 1, 2, \ldots, k$, and $\sum_{j=1}^{k} s_j = T$. Assuming consistent capture probability, θ, across all sub-occasions, the detection probability for each primary capture occasion p_j depends on the number of sub-occasions allocated to it, as described by

$$p_j = 1 - (1 - \theta)^{s_j}, \tag{5.1}$$

where θ is the per-sub-occasion capture probability (hereafter referred to as capture probability), and p_j is the primary occasion detection probability (hereafter referred to as detection rate). Let $q_j = 1 - p_j$, n_j represent the number of individuals detected in the jth occasion, and n the total number of unique individuals detected throughout the experiment. This model, termed as the time-varying model M_t, has its likelihood expressed as

$$L(N, p_j) \propto \frac{N!}{(N-n)!} \prod_{j=1}^{k} \left\{ p_j^{n_j} q_j^{N-n_j} \right\}.$$

Darroch (1958) demonstrated that the MLE $\hat{N}$ for the population size is the unique root, exceeding n, of the polynomial of degree $k - 1$ given by:

$$\left(1 - \frac{n}{N}\right) = \prod_{j=1}^{k} \left(1 - \frac{n_j}{N}\right).$$

The asymptotic variance of the population size estimate, according to Seber and Schofield (2023), is

$$\mathrm{Var}(\hat{N}) \approx N \left[\frac{1}{\prod_{j=1}^{k} q_j} + k - 1 - \sum_{j=1}^{k} \frac{1}{q_j} \right]^{-1}. \tag{5.2}$$

5.3 Allocating Sampling Efforts Across Capture Occasions

To obtain an accurate estimate of N, it is essential to minimise the variance outlined in Equation (5.2). While N remains unknown and the first term in the brackets of Equation (5.2) is constant for any allocation strategy due to the constraint $\sum_{j=1}^{k} s_j = T$, minimizing the variance primarily depends on the last term, $\sum_{j=1}^{k} \frac{1}{q_j}$. This can be achieved by optimizing the distribution of secondary sampling sub-occasions s_j across primary occasions. The detection probability p_j, given in Equation (5.1), directly influences the calculation of q_j. By rearranging the terms, q_j can be expressed as a function of s_j and θ

$$\frac{1}{q_j} = \frac{1}{(1-\theta)^{s_j}}.$$

By defining

$$\begin{aligned} f(s) &= \frac{1}{(1-\theta)^s} \\ &= \left(\frac{1}{1-\theta}\right)^s \\ &= \exp\left[s \log\left(\frac{1}{1-\theta}\right)\right], \end{aligned} \tag{5.3}$$

the second derivative of the function in Equation (5.3) with respect to s confirms its convexity, since

$$f''(s) = \left[\log\left(\frac{1}{1-\theta}\right)\right]^2 \exp\left[s \log\left(\frac{1}{1-\theta}\right)\right] \geq 0.$$

This is because the first squared term, as well as the second exponential term are always positive for $\theta \in (0, 1)$. Thus, the function $f(s)$ is convex.

The objective is to minimise the sum

$$\sum_{j=1}^{k} \frac{1}{q_j} = \sum_{j=1}^{k} \frac{1}{(1-\theta)^{s_j}},$$

subject to the constraint

$$\sum_{j=1}^{k} s_j = T,$$

where $s_j \geq 0$ and $\theta \in (0, 1)$. $f(s)$ has been shown to be convex because its second derivative is non-negative. According to Jensen's inequality, for any convex function, the function's value at the mean is less than or equal to the mean of the function's values. Mathematically, this relationship can be expressed as

$$\begin{aligned} f\left(\frac{1}{k}\sum_{j=1}^{k} s_j\right) &\leq \frac{1}{k}\sum_{j=1}^{k} f(s) \\ \frac{1}{(1-\theta)^{T/k}} &\leq \frac{1}{k}\sum_{j=1}^{k} \frac{1}{(1-\theta)^{s_j}}. \end{aligned}$$

This inequality implies that an equal allocation of $s_j = T/k$ minimises $\sum_{j=1}^{k} \frac{1}{(1-\theta)^{s_j}}$, leading to a reduction in the variance of the population size estimate.

5.4 Simulation Study

To validate the proposed method for allocating sampling effort, capture-recapture experiments with varying allocation strategies were simulated. Three distinct allocation

strategies for sampling effort are considered to explore their impact on the estimation of population size and variance. The "even allocation" strategy distributes the total sampling effort equally across all primary capture occasions. The "skewed allocation" strategy concentrates more effort at the beginning of the study period. Lastly, the "random allocation" strategy uses a multinomial distribution to assign the effort. The simulation study includes the following steps.

1. Distribute the total sampling effort, T across k primary occasions using three different strategies: even allocation, skewed allocation, and random allocation.

2. For each individual in the population N, perform a Bernoulli trial to determine capture histories based on the probability p_j, which is calculated from θ and the allocated s_j for each occasion. Record the number captured n_j in each occasion and identify the total number of unique individuals n captured across all occasions.

3. Calculate the MLE $\hat{N}$ by solving the equation

$$\left(1 - \frac{n}{N}\right) = \prod_{j=1}^{k} \left(1 - \frac{n_j}{N}\right).$$

4. Determine asymptotic variance of $\hat{N}$ using the formula

$$\mathrm{Var}(\hat{N}) \approx N \left[\frac{1}{\prod_{j=1}^{k} q_j} + k - 1 - \sum_{j=1}^{k} \frac{1}{q_j} \right]^{-1}.$$

5. Conduct the simulation for B iterations to calculate the average $\hat{N}$, empirical variance, and asymptotic variance for each allocation strategy.

The results, as presented in Table 5.1, detail the effectiveness of each strategy when $\theta = 0.1$, demonstrating how different allocation strategies influence the variance of the estimated population size under various scenarios. The simulation was conducted with $B = 10000$ repetitions. It is shown that the MLE method accurately estimates the population size N, closely approximating the actual N. Furthermore, both empirical and asymptotic variances are minimised when using the even sampling allocation strategy, compared to the outcomes with skewed and random allocations. It is also noted that variances are larger when fewer sub-occasions are utilized (small T), and decrease as T is increased, highlighting the direct impact of increased sampling effort on enhancing the precision of variance estimates. This observation emphasizes the significance of sufficient sampling effort in achieving more accurate population size estimations.

TABLE 5.1: Simulation results for different allocation strategies when $\theta = 0.1$, presenting the estimated population size ($\hat{N}$), empirical variance, and asymptotic variance under varying combinations of population size (N), number of primary capture occasions (k), and total sampling effort (T). Each strategy (even, skewed, and random) specifies the sub-occasion allocation (s_j) across the k occasions. Bold values indicate the lowest variance in each scenario, representing the most precise allocation strategy.

N	k	T	Allocation	$\{s_j\}$	mean $\hat{N}$	Empirical Variance	Asymptotic Variance
100	3	9	even	3,3,3	101.893	**282.051**	**214.616**
			skewed	4,3,2	102.034	282.938	221.874
			random	5,3,1	102.767	331.198	247.028
100	3	15	even	5,5,5	100.429	**60.275**	**56.293**
			skewed	7,4,4	100.611	61.815	58.212
			random	2,9,4	100.664	73.638	65.918
100	3	30	even	10,10,10	100.038	**6.153**	**5.887**
			skewed	15,8,7	99.995	6.272	6.128
			random	6,14,10	100.081	6.274	6.072
100	5	15	even	3,3,3,3,3	100.335	**52.212**	**50.044**
			skewed	7,2,2,2,2	100.430	56.700	54.707
			random	1,2,6,1,5	100.342	58.674	54.796
100	5	30	even	6,6,6,6,6	100.057	**5.605**	**5.500**
			skewed	15,4,4,4,3	100.021	5.953	5.956
			random	4,7,9,1,9	100.031	5.789	5.649

Continued on next page

Table 5.1 – continued from previous page

N	k	T	Allocation	$\{s_j\}$	mean $\hat{N}$	Empirical Variance	Asymptotic Variance
			even	10,10,10,10,10	100.557	**0.484**	**0.544**
100	5	50	skewed	25,7,6,6,6	100.582	0.521	0.567
			random	13,2,15,10,10	100.679	0.543	0.563
			even	3,3,3,3,3,3,3,3,3,3	100.008	**5.379**	**5.299**
100	10	30	skewed	15,2,2,2,2,2,2,1,1,1	100.004	5.846	5.885
			random	3,5,1,5,2,1,4,2,3,4	100.030	5.511	5.342
			even	5,5,5,5,5,5,5,5,5,5	100.681	**0.369**	**0.537**
100	10	50	skewed	25,3,3,3,3,3,3,3,2,2	100.704	0.465	0.565
			random	3,1,8,4,4,6,3,4,9,8	100.574	0.377	0.539
			even	3,3,3	501.686	**1137.012**	**1073.080**
500	3	9	skewed	4,3,2	501.305	1115.431	1109.368
			random	1,2,6	502.188	1447.106	1413.150
			even	5,5,5	500.390	**280.034**	**281.467**
500	3	15	skewed	7,4,4	500.593	293.887	291.058
			random	1,2,12	501.130	526.096	515.172
			even	10,10,10	500.022	**29.006**	**29.436**
500	3	30	skewed	15,8,7	500.037	30.725	30.639
			random	9,7,14	500.046	30.773	30.218

Continued on next page

Table 5.1 – continued from previous page

N	k	T	Allocation	$\{s_j\}$	mean $\hat{N}$	Empirical Variance	Asymptotic Variance
500	5	15	even	3,3,3,3,3	500.486	**255.627**	**250.222**
			skewed	7,2,2,2,2	500.414	272.171	273.536
			random	3,5,3,2,2	500.460	261.512	256.304
500	5	30	even	6,6,6,6,6	499.997	**27.792**	**27.501**
			skewed	15,4,4,4,3	500.035	29.206	29.782
			random	16,1,3,3,7	500.112	30.303	30.773
500	5	50	even	10,10,10,10,10	500.012	**2.658**	**2.722**
			skewed	25,7,6,6,6	499.980	2.831	2.835
			random	17,7,13,4,9	500.006	2.768	2.749
500	10	30	even	3,3,3,3,3,3,3,3,3,3	500.020	**26.528**	**26.494**
			skewed	15,2,2,2,2,2,2,1,1,1	500.043	29.162	29.425
			random	1,1,2,5,2,4,9,2,3,1	500.048	27.104	27.200
500	10	50	even	5,5,5,5,5,5,5,5,5,5	499.973	**2.741**	**2.687**
			skewed	25,3,3,3,3,3,3,3,2,2	499.973	2.881	2.824
			random	5,4,4,13,7,2,1,4,5,5	499.996	2.803	2.703

Continued on next page

Table 5.1 – continued from previous page

N	k	T	Allocation	$\{s_j\}$	mean $\hat{N}$	Empirical Variance	Asymptotic Variance
1000	3	9	even	3,3,3	1002.294	**2168.335**	**2146.160**
			skewed	4,3,2	1002.312	2218.461	2218.737
			random	3,2,4	1001.548	2243.036	2218.737
1000	3	15	even	5,5,5	1000.771	**575.091**	**562.933**
			skewed	7,4,4	1000.772	577.016	582.117
			random	3,6,6	1000.525	575.784	580.774
1000	3	30	even	10,10,10	1000.147	**57.767**	**58.872**
			skewed	15,8,7	1000.046	60.649	61.278
			random	5,12,13	1000.054	61.502	60.896
1000	5	15	even	3,3,3,3,3	1000.341	**516.512**	**500.444**
			skewed	7,2,2,2,2	1000.471	546.867	547.072
			random	4,2,5,1,3	1000.314	519.277	520.333
1000	5	30	even	6,6,6,6,6	1000.005	**53.063**	**55.001**
			skewed	15,4,4,4,3	1000.211	60.224	59.564
			random	6,4,4,11,5	1000.024	56.065	56.244
1000	5	50	even	10,10,10,10,10	1000.022	**5.444**	**5.444**
			skewed	25,7,6,6,6	999.961	5.700	5.670
			random	1,12,21,7,9	1000.005	5.499	5.569

Continued on next page

Table 5.1 – continued from previous page

N	k	T	Allocation	$\{s_j\}$	mean $\hat{N}$	Empirical Variance	Asymptotic Variance
			even	3,3,3,3,3,3,3,3,3,3	1000.070	**52.866**	**52.987**
1000	10	30	skewed	15,2,2,2,2,2,2,1,1,1	1000.072	58.658	58.851
			random	4,5,1,2,2,8,2,2,3,1	999.774	54.287	54.005
			even	5,5,5,5,5,5,5,5,5,5	999.963	**5.354**	**5.374**
1000	10	50	skewed	25,3,3,3,3,3,3,3,2,2	1000.026	5.526	5.649
			random	4,5,6,1,4,3,2,10,9,6	999.999	5.410	5.395

5.5 Varying Detection Rates Across Primary Occasions

The hierarchical Schnabel census model adapts to accommodate varying detection rates across primary occasions. Let θ_j represent the per-sub-occasion capture probability for the jth primary occasion. This model allows variability in θ across different occasions. Consequently, the detection probability for each primary capture occasion, denoted p_j, is defined by:

$$p_j = 1 - (1 - \theta_j)^{s_j},$$

where s_j denotes the number of sub-occasions allocated to the jth primary occasion. Accordingly, the probability q_j that an individual remains undetected during occasion j is

$$q_j = (1 - \theta_j)^{s_j},$$

facilitating a modified likelihood function

$$L(N, \mathbf{p}) \propto \frac{N!}{(N-n)!} \prod_{j=1}^{k} \{p_j^{n_j} q_j^{N-n_j}\},$$

where n_j represents detections in the jth occasion, n totals unique detections, and

$$\mathbf{p} = (p_1, p_2, \ldots, p_k).$$

The asymptotic variance of the population estimate $\hat{N}$, now influenced by variable θ_j, is

$$\mathrm{Var}(\hat{N}) \approx N \left[\frac{1}{\prod_{j=1}^{k} q_j} + k - 1 - \sum_{j=1}^{k} \frac{1}{q_j} \right]^{-1}.$$

With $q_j = (1 - \theta_j)^{s_j}$, the variance refines to:

$$\mathrm{Var}(\hat{N}) \approx N \left[\frac{1}{\prod_{j=1}^{k} (1 - \theta_j)^{s_j}} + k - 1 - \sum_{j=1}^{k} \frac{1}{(1 - \theta_j)^{s_j}} \right]^{-1}.$$

Optimizing sub-occasion allocations s_j minimises $\mathrm{Var}(\hat{N})$, subject to the total sub-occasion constraint $s_1 + s_2 + \cdots + s_k = T$. The objective focuses on minimizing

$$\sum_{j=1}^{k} \frac{1}{(1 - \theta_j)^{s_j}}.$$

The specific values of θ_j determine the optimal allocation, leading to the formulation:

$$\min_{s_1,s_2,\ldots,s_k} \sum_{j=1}^{k} \frac{1}{(1-\theta_j)^{s_j}},$$

subject to:

$$s_1 + s_2 + \cdots + s_k = T.$$

5.5.1 Uniform Distributed Catchability

If θ is considered to follow a uniform distribution,

$$\theta_j \sim \text{Uniform}(a, b), \quad 0 \leq a < b < 1.$$

Since θ_j is an unobserved random variable, a pseudo-Bayesian approach is applied based on the prior distribution of θ_j. The optimization problem now involves minimizing the expectation of the key term in the asymptotic variance:

$$\min_{s_1,s_2,\ldots,s_k} \sum_{j=1}^{k} \mathbb{E}[(1-\theta_j)^{-s_j}], \tag{5.4}$$

subject to:

$$s_1 + s_2 + \cdots + s_k = T, \quad s_j > 1. \tag{5.5}$$

The expectation $\mathbb{E}[(1-\theta_j)^{-s_j}]$ is derived from the uniform distribution of θ_j as:

$$\mathbb{E}[(1-\theta_j)^{-s_j}] = \frac{1}{b-a}\left[\frac{(1-a)^{1-s_j} - (1-b)^{1-s_j}}{1-s_j}\right], \quad s_j > 1. \tag{5.6}$$

Designating the right hand side (RHS) of (5.6) as $g(s_j)$, the optimization target in (5.4) simplifies to:

$$G(s_1, s_2, \ldots, s_k) = \sum_{j=1}^{k} g(s_j),$$

with the normalisation constraint in Equation (5.5).

One approach to solving this optimization problem is the Newton-Raphson method, an iterative algorithm that leverages both the gradient (first derivative) and the Hessian (second derivative) to find the minimum of a function. The update rule is given by:

$$\mathbf{S}_{\text{new}} = \mathbf{S}_{\text{old}} - \mathbf{H}^{-1} \cdot \nabla G(\mathbf{S}), \tag{5.7}$$

where

$\mathbf{S}$ represents the vector of s_j values,

$\nabla G(\mathbf{S})$ is the gradient of the objective function $G(s_1, s_2, \ldots, s_k)$, and

$\mathbf{H}$ is the Hessian matrix, a diagonal matrix with elements $\dfrac{\partial^2 G}{\partial s_j^2}$.

The objective function to be minimised is:

$$G(s_1, s_2, \ldots, s_k) = \sum_{j=1}^{k} \frac{1}{b-a} \left[\frac{(1-a)^{1-s_j} - (1-b)^{1-s_j}}{1-s_j} \right].$$

The gradient $\nabla G(\mathbf{S})$ consists of the first derivative elements

$$g'(s_j) = \frac{1}{b-a} \left[\frac{-(1-a)^{1-s_j} \log(1-a) + (1-b)^{1-s_j} \log(1-b)}{1-s_j} + \frac{(1-a)^{1-s_j} - (1-b)^{1-s_j}}{(1-s_j)^2} \right].$$

The Hessian matrix $\mathbf{H}$ is a diagonal matrix with second derivative elements given by:

$$g''(s_j) = \frac{1}{b-a} \cdot \left[\frac{-(1-a)^{1-s_j} (\log(1-a))^2 + (1-b)^{1-s_j} (\log(1-b))^2}{1-s_j} + \frac{2\left[(1-a)^{1-s_j} \log(1-a) - (1-b)^{1-s_j} \log(1-b)\right]}{(1-s_j)^2} - \frac{2\left[(1-a)^{1-s_j} - (1-b)^{1-s_j}\right]}{(1-s_j)^3} \right]. \quad (5.8)$$

Using the update rule in Equation (5.7), the vector $\mathbf{S}$ is iteratively updated until convergence is achieved. The final S_j values are then projected onto the feasible region, ensuring $s_j > 1$, and normalized to satisfy the constraint $\sum_{j=1}^{k} s_j = T$

$$s_j \leftarrow s_j \cdot \frac{T}{\sum_{j=1}^{k} s_j}.$$

An equal allocation of $s_j = T/k$ for all $j = 1, 2, \ldots, k$ is recommended as the initial setting. This choice satisfies the total sampling effort constraint, avoids boundary issues (e.g., $s_j < 1$), and often leads to faster convergence, particularly under symmetric or uniform detectability assumptions.

Table 5.2 presents the optimal allocation results for various values of a, b, k, and T. These results are obtained by minimizing the expected key term $\sum_{j=1}^{k} \mathbb{E}\left[(1-\theta_j)^{-s_j}\right]$ using the Newton-Raphson method. The results indicate that the optimal allocation is evenly distributed across all primary occasions.

TABLE 5.2: Optimal allocation of sub-occasions s_j across k primary capture occasions for various values of the uniform distribution parameters a and b, where $\theta_j \sim \text{Uniform}(a, b)$, subject to the constraint $\sum_{j=1}^{k} s_j = T$.

$\theta \sim$ Uniform (a,b)		k	T	Optimal $\{s_j\}$
a	b			
0.1	0.9	3	9	3, 3, 3
0.3	0.7	5	20	4, 4, 4, 4, 4
0.05	0.5	8	40	5, 5, 5, 5, 5, 5, 5, 5
0.01	0.4	10	100	10, 10, 10, 10, 10, 10, 10, 10, 10, 10
0.005	0.3	20	220	11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11

The finding that equal allocation of sampling effort minimises variance when capture probabilities follow a uniform distribution is a result of the combined effects of convexity and symmetry in the optimization problem. The goal is to minimise the key term in the asymptotic variance of the population size estimator in Equation (5.4), where θ_j follows a uniform distribution over the interval (a, b). Taking the expectation results in the integral in Equation (5.6). A key property of this function is its convexity in s_j. The second derivative of this function in Equation (5.8) is positive for all $s_j > 1$, and $0 \leq a < b < 1$, establishing the convexity of the function. This convexity property enables the application of Jensen's inequality, which state that for any convex function f,

$$\mathbb{E}\left[f(s_j)\right] \geq f\left(\mathbb{E}[s_j]\right).$$

Additionally, the assumption that θ_j are uniformly distributed introduces an important form of symmetry. Since all θ_j are drawn from the same distribution, no occasion has an inherently higher or lower capture probability than another. As a result, the expected contribution to variance is identical across occasions when effort is equally distributed. If effort were instead distributed unevenly, the convexity of the variance function would cause the total variance to increase. More formally, the sum of expected values satisfies

$$\sum_{j=1}^{k} \mathbb{E}\left[(1-\theta_j)^{-s_j}\right] \geq k \cdot \mathbb{E}\left[(1-\theta_j)^{-T/k}\right],$$

with equality holding only when each s_j is equal to T/k, which confirms that equal allocation minimises variance.

The assumption of a uniform distribution for θ_j plays a crucial role in this result. If capture probabilities varied systematically across occasions, such as being consistently

higher or lower during specific periods, it might be beneficial to adjust the sampling effort to reflect those patterns. However, under a symmetric distribution of capture probabilities, no such directional trend exists, making equal allocation the best strategy. The next sub-section examines sampling strategies under a declining pattern in capture probability across occasions.

5.5.2 Monotonically Decreasing Catchabilities

If θ_j decreases monotonically from $j = 1$ to $j = k$, it follows an ordered uniform distribution such that

$$\theta_1 \sim \text{Uniform}(a, b);\ \theta_2 \mid \theta_1 \sim \text{Uniform}(a, \theta_1); \dots\ \theta_k \mid \theta_{k-1} \sim \text{Uniform}(a, \theta_{k-1}), \quad (5.9)$$

where $0 < a < b < 1$, and $s_j \geq 1$ for $j = 1, 2, \dots, k$.

Under this assumption, the per-sub-occasion capture probabilities vary, with later primary occasions exhibiting lower baseline capture probability. This assumption is not only mathematically convenient but also highly realistic in practice, as it reflects the behavioural response of individuals becoming increasingly trap-shy over time. This phenomenon is often referred to as a learning effect, where animals avoid recapture after initial capture experience.

This structure implies that the expected capture probability decreases as j increases. For instance, given $\mathbb{E}[\theta_1] = 1/2$ and, conditional on θ_1, $\mathbb{E}[\theta_2 \mid \theta_1] = \theta_1/2$, it follows that $\mathbb{E}[\theta_2] = 1/4$. Extending this pattern, the expected value of θ_j approximates $(1/2)^j$, indicating that later primary occasions are, on average, less effective at capturing individuals per-sub-occasion.

In the previous analysis with constant θ, an equal allocation of sub-occasions minimised $\sum_{j=1}^{k}(1-\theta)^{-s_j}$ via Jensen's inequality. However, with θ_j varying and trending downward, this allocation may no longer be optimal. Intuitively, to counterbalance the declining capture efficiency, increasing s_j where θ_j is lower may be beneficial. The objective is to determine the optimal allocation of sub-occasions $s_1, s_2, \dots, s_k$ by minimising

$$F(s_1, \dots, s_k) = \sum_{j=1}^{k} f_j(s_j),$$

subject to the constraint

$$\sum_{j=1}^{k} s_j = T, \quad s_j \geq 1 \quad \text{for all } j.$$

Each $f_j(s_j)$ is given by:

$$f_j(s_j) = \int \frac{1}{(1-\theta_j)^{s_j}}\, dP(\theta_j) = \mathbb{E}\left[(1-\theta_j)^{-s_j}\right], \tag{5.10}$$

where $P(\theta_j)$ is the conditional distribution defined by the ordered uniform priors in Equation (5.9).

Using a Lagrange multiplier Λ for the constraint, the Lagrangian function is

$$\mathcal{L}(s_1, \ldots, s_k, \Lambda) = \sum_{j=1}^{k} f_j(s_j) + \Lambda \left(\sum_{j=1}^{k} s_j - T \right).$$

Differentiating with respect to s_j yields

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s_j} &= f_j'(s_j) + \Lambda = 0 \\ \frac{\partial}{\partial s_j} \mathbb{E}\big[(1-\theta_j)^{-s_j}\big] + \Lambda &= 0 \\ -\mathbb{E}\big[(1-\theta_j)^{-s_j} \log(1-\theta_j)\big] + \Lambda &= 0. \end{aligned}$$

Thus, the optimal set of allocations $\{s_j^*\}$ must satisfy

$$\mathbb{E}\Big[(1-\theta_j)^{-s_j^*} \log(1-\theta_j)\Big] = \Lambda, \tag{5.11}$$

for each j, with the constraint $\sum_{j=1}^{k} s_j^* = T$.

Since the expectation in each equation inherits the ordered structure of θ_j (where θ_2 depends on θ_1, and so forth), solving these equations in closed form is unlikely. Instead, the expectations in Equation (5.10) can be estimated numerically using Monte Carlo simulation based on the ordered priors. A numerical solver can then be applied to determine s_j^* that satisfy Equation (5.11) under the given constrains.

Because the objective function is estimated via Monte Carlo sampling, function evaluations can be noisy. This noise complicates the accurate computation of the gradient and Hessian, making Newton–Raphson methods challenging to apply. In contrast, the `optim` function with the `"L-BFGS-B"` method does not require explicit Hessian computations and tends to be more robust when handling Monte Carlo noise.

The following algorithm outlines the procedure for determining the optimal allocation when the detection rate decreases monotonically.

Monte Carlo Optimal Allocation for Monotonically Decreasing Detection Rates

Step 1. Generate Monte Carlo samples by drawing $\theta_1 \sim \text{Uniform}(a, b)$, and

$$\theta_j \sim \text{Uniform}(a, \theta_{j-1})$$

for the subsequent $j = 2, \ldots, k$.

Step 2. Assign initial values to s_j. To ensure $s_j \geq 1$ and $\sum_{j=1}^{k} s_j = T$, re-parametrise using

$$s'_j = 1 + (T - k)\frac{e^{s_j - M}}{\sum_{j=1}^{k} e^{s_j - M}}.$$

where $M = \max(\mathbf{S})$.

Step 3. Using Monte Carlo samples from Step 1, and re-parametrised values from Step 2, evaluate the objective function

$$\sum_{j=1}^{k} \mathbb{E}\left[\frac{1}{(1-\theta_j)^{s'_j}}\right].$$

Step 4. Apply the `"L-BFGS-B"` method via `optim` function in `R` to minimise the objective function and determine the optimum allocation s_j^*.

Table 5.3 compares the performance of three allocation strategies across varying parameter settings (a, b, k, and T). The optimal strategy is identified by minimizing the objective function, defined as the sum of the expected values of $(1-\theta_j)^{-s_j}$ across all primary occasions. Three strategies are considered:

1. Optimal Allocation: Minimizes the objective value.
2. Equal Allocation: Distributes the total effort evenly across all primary occasions.
3. Random Allocation: Assigns effort randomly without specific structure.

Lower objective values reflect higher efficiency. The results indicate that equal allocation is suboptimal when detection probabilities decrease over time. In such cases, concentrating more effort on later occasions helps balance detection efficiency, thereby reducing variance in the population estimate.

5.6 Discussion and Conclusion

This study examined the optimal allocation of sampling effort in a hierarchical Schnabel census and demonstrated how different sub-occasion allocation strategies influence the accuracy and precision of population estimates.

In the first scenario, the per-sub-occasion capture probability θ is known to be constant across all k primary occasions. Under this condition, equal allocation of sub-occasions is

TABLE 5.3: Comparison of Allocation Strategies for Different Values of a, b, k, and T. The objective value, which is the sum of the expected values of $(1-\theta_j)^{-s_j}$, indicates the effectiveness of each strategy, with lower values representing better accuracy.

a	b	k	T	Allocation	$\{s_j\}$	Objective value
0.10	0.90	3	9	optimal	1, 3, 5	**24.93**
				equal	3, 3, 3	69.42
				random	4, 3, 2	406.61
0.30	0.70	5	20	optimal	1, 3, 4, 5, 7	**35.95**
				equal	4, 4, 4, 4, 4	54.31
				random	2, 7, 2, 3, 6	118.26
0.05	0.50	8	40	optimal	1, 1, 1, 2, 5, 8, 10, 12	**11.61**
				equal	5, 5, 5, 5, 5, 5, 5, 5	20.20
				random	6, 4, 4, 5, 6, 5, 9, 1	24.64
0.01	0.10	10	50	optimal	1, 1, 1, 1, 1, 1, 7, 11, 13, 13	**10.64**
				equal	5, 5, 5, 5, 5, 5, 5, 5, 5, 5	11.06
				random	4, 4, 3, 4, 8, 6, 6, 2, 5, 8	10.94
0.01	0.05	12	120	optimal	1, 1, 1, 1, 1, 1, 9, 17, 20, 22, 23, 23	**12.69**
				equal	10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10	13.15
				random	7, 10, 13, 9, 9, 6, 12, 9, 12, 13, 9, 11	13.07

shown to be the optimal design. In the second scenario, where θ varies following a uniform distribution, equal allocation remains the most effective strategy. These outcomes align with classical sampling theory, where evenly distributing resources is generally optimal under constant or uniform capture probabilities, due to the convex nature of the key term (Cochran, 1977). These results offer practical guidance: since optimality holds for any value of θ, equal allocation remains valid even when θ is unknown but assumed to follow a constant or symmetric distribution. This conclusion also extends to scenarios where θ follows a beta distribution with identical parameters across primary occasions. In such cases, the symmetry of the distribution and the convexity of the key variance term support the optimality of equal allocation, as established by Jensen's inequality.

The third scenario considers a setting in which θ follows a monotonically decreasing trend across primary occasions. In this context, applying a pseudo-Bayesian framework is not feasible, as the θ values are not independent but instead depend on values from preceding occasions. This nested dependency complicates the direct computation of expectations required for optimisation. Instead, Monte Carlo simulations are used to generate samples of θ values, from which the expected values of key term are computed. Optimal sampling effort is then derived from these expectations. This simulation-based approach proves robust in handling complex dependencies among capture probabilities across occasions (Seber and Schofield, 2023). Under this decreasing trend, equal allocation is no longer optimal, as it increases the variance of the population size estimate. Instead, adaptive strategies that allocate more effort to later occasions provide improved precision. Such strategies are particularly relevant in real-world studies where catchability may decline over time due to behavioural responses or changing environmental conditions (Pollock et al., 2006; Thompson et al., 1998).

The hierarchical framework presented shares structural similarities with Pollock (1982)'s robust design, which organises sampling into primary periods (open population) and secondary occasions (closed within periods). However, a key difference lies in the population closure assumption: the hierarchical design assumes that the population remains closed throughout the study, whereas the robust design explicitly models demographic events such as births, deaths, and migration between primary periods. This closure assumption simplifies the model, enabling precise allocation of effort across sub-occasions without the need to account for demographic changes.

In summary, this chapter provides valuable insights into the design of hierarchical Schnabel censuses. It highlights how the structure and variability of capture probabilities influence optimal sampling strategies and offers a principled framework for allocating sampling effort to achieve improved estimation performance.

Chapter 6

Conclusion and Future Work

6.1 General Discussion and Conclusion

This thesis has examined optimisation of sampling effort in capture-recapture studies in three structured models, including Hierarchical Lincoln-Petersen model, Schnabel census, and Hierarchical Schnabel census. A unifying conclusion across these models is that planned sampling effort can greatly improve estimates' accuracy and precision. This is substantiated by theoretical work and simulation exercises, which produce practical recommendations that can apply to most field situations.

This work contributes by offering a clear framework for planning capture-recapture studies when resources are limited. The methods help researchers choose sampling effort in a smart and efficient way. This reduces the need to rely on guesswork. By linking capture probabilities, sampling effort, and estimate precision, the framework supports better planning. It also helps make the best use of time, manpower, and other resources. This is especially useful in fields like ecology and health, where it is often hard to collect good data and resources are limited.

6.1.1 Equal Effort Allocation Under Homogeneous Conditions

When capture probabilities are constant across sampling occasions, distributing effort equally leads to optimal precision. This was demonstrated across both the Hierarchical Lincoln-Petersen and Hierarchical Schnabel census frameworks. In Chapter 3, equal allocation of effort ($t = T/2$) minimises variance without requiring prior knowledge of the capture probabilities θ, echoing recommendations by Robson and Regier (1964) and Pollock et al. (1990). Similarly, in Chapter 5, the use of Jensen's inequality demonstrates that when θ is constant, the optimal strategy is to distribute sub-occasions evenly across k primary occasions, i.e. $s_j = T/k$. Simulation confirmed that this strategy consistently reduces estimator variance by 2 - 25 % compared to unequal allocations.

These findings are consistent with principles in classical sampling theory (Cochran, 1977), which advocates for balances sampling when underlying probabilities are stable. From a practical perspective, equal allocation provides a simple yet effective design strategy, particularly in situations where capture probabilities are not well understood and ease of implementation is a priority.

6.1.2 Adaptive Effort Allocation Under Heterogeneity

When capture probabilities vary across occasions, optimal sampling design becomes more complex. If the heterogeneity in θ is known or can be estimated, adaptive allocations strategies which assign more effort to periods of lower capturabilities offer advantages. Chapter 3 through Chapter 5 demonstrated how numerical techniques, including Newton-Raphson and Quasi-Newton optimisation, can be employed to determine effort distributions that reduce estimator variance.

For scenarios where the exact detection probabilities are unknown, a pseudo-Bayesian framework using uniform or structured priors is introduced. This approach, although not fully Bayesian, leverages expected values of θ to inform sampling design and has shown robust performance in simulation studies. These findings align with previous studies on Bayesian estimation (Basu and Ebrahimi, 2001; Wang et al., 2015), which emphasise the adaptability of Bayesian and pseudo-Bayesian framework in handling uncertainty effectively.

6.1.3 Modelling via Zero-Truncated and EM algorithm

This thesis has demonstrated the use of zero-truncated models to count for undetected individuals in capture-recapture experiments. Zero-truncated models focus on observed count distributions in order to make an inference on the complete population that consists of both observed and unobserved individuals. This is especially important in research on elusive or hard-to-sample populations, in which under-reporting is usual (Böhning and Kuhnert, 2006).

The application of the EM algorithm adds to the suitability of zero-truncated models. The EM algorithm solves parameter estimation challenges presented by incomplete data or complicated likelihood functions. The EM algorithm iteratively estimates missing data (E-step) and maximises the expectation of log-likelihood (M-step) to converge to stable parameter estimates. This iterative computation is desirable for use in capture-recapture data when the closed-form solution is not readily available (Dempster et al., 1977).

6.1.4 Pilot Study as a Design Tool

Pilot studies are an integral part in designing capture-recapture studies by suggesting initial estimates for capture rates and informing sampling effort allocation choices. It is impossible to define an optimal effort level, since it is based on the size of the population,

which is an unobserved quantity that one is attempting to estimate. This is inherently a circular problem in planning. Pilot data can break this impediment, according to Bröder et al. (2020), by providing some early guidance that aids simulation-based design as well as parameter calibration.

Even small-scale pilot studies have value in adding to model development and enhancing estimates of precision for populations. Robson and Regier (1964) argued that planning at an early stage does not necessarily have to depend on guesswork, but that reasoned assumptions, on an expert basis or from available data, can define an initial working point. In conjunction with contemporary simulation methods (Paterson et al., 2019), such early observations allow researchers to trade off study aims with practical limitations to produce more efficient and impactful sampling plans.

6.1.5 Detectability Enhancements

Improving detectability provides an efficient substitute for increasing sampling effort in capture-recapture analyses. In situations where capture probabilities are low, variance in estimates can increase, necessitating large samples to achieve precision. Rationing resources to capture large numbers of individuals can be replaced by enhancing probability of recapture by using sophisticated technologies, including camera traps, GPS collars, and RFID tags. According to Burnham et al. (1987), even modest enhancements in detectability can substantially cut sampling occasions. Such were also the conclusions by Papadatou et al. (2012) as well as by Schorr et al. (2014), who emphasized advantages in using technology combined with enhanced analytical techniques in terms of accessing populations that elude, or remain under-reported.

In addition to tools based on technology, field-based enhancements like optimised observer timing, observer training, and habitat-specific techniques can improve catchability at constant effort. Such refinements have been demonstrated to improve data accuracy and estimation credibility, even at limited budgets (Conner et al., 2015). Together, enhancing detectability offers an inexpensive means to achieve precision and practicality, particularly at resource-restricted surveys or in populations that prove tough to sample.

6.2 Limitations

Although this thesis presents an organized approach to optimising sampling effort in capture-recapture experiments, some limitations should be noted. Firstly, the methods assume closed populations, meaning there is neither birth, death, immigration, nor emigration throughout the study. Though making analysis easy, this assumption might not apply to long-term study or dynamic ecological communities.

Secondly, the methods are limited to single-species studies. Multi-species or community capture-recapture studies have extra levels of heterogeneity that could influence efficiency in sampling. In such contexts, variations in species-specific detectability, behaviour,

as well as habitat use could make allocation of sampling effort more complicated and necessitate more adaptable methods.

Third, there is no allowance for spatial considerations. Several studies in real-world situations have geographically distributed populations, for which individuals movement and trap location may have impact on detectability. Not allowing for spatial structure could decrease precision when making estimates in such contexts.

6.3 Future Work

There are several ways this work can be extended. One possible way is to include open-population models that permit demographic change over years, such as the Jolly-Seber or Cormack-Jolly-Seber models. This could make the methods more relevant to monitoring over longer periods or to investigations involving transient or migratory populations.

Second, the methods could be combined with spatial capture-recapture (SCR) models to allow researchers to account for spatial variation in detection. By considering trap location and animal movement, integrating with SCR models could improve the precision and accuracy of estimates at spatially structured sites.

Finally, to make the methods more practically relevant, embedding financial and logistical limitations directly in them would be beneficial. By making the trade-off between cost, availability of manpower, and precision desired explicit, researchers as well as decision-makers could design studies more effectively within practical operational constraints.

Appendix A

Codes Availability

The R scripts used in this thesis including model fittings, sampling effort calculations and simulations have been compiled and uploaded to GitHub. The repository is accessible at: https://github.com/snchin17/sampling-effort-in-cr

Appendix B

TABLE B.1: Simulation results for various combinations of population size N, total sub-occasions T, and capture probabilities θ_1 and θ_2. Variance of the population size estimates $\hat{N}$ is shown across different values of t, the number of sub-occasions allocated to capture occasion 1. $\pi_1 \times \pi_2$ represents the joint detectability under each allocation.

N	θ_1	θ_2	T	t	$\hat{N}$	$\text{Var}(\hat{N})$	$\pi_1 \times \pi_2$
1000	0.40	0.20	20	1	1000.002	1.61412	0.400
				2	999.997	1.02507	0.639
				3	1000.005	0.82488	0.782
				4	1000.003	0.74839	0.866
				5	999.988	0.71788	0.915
				6	999.995	0.69075	0.940
				7	999.997	**0.69002**	**0.949**
				8	1000.002	0.69101	0.945
				9	1000.005	0.69394	0.926
				10	1000.005	0.72879	0.887
				11	1000.000	0.74907	0.862
				12	1000.004	0.78166	0.829
				13	999.992	0.83059	0.788
				14	999.993	0.88458	0.736
				15	1000.001	0.95868	0.671
				16	999.997	1.10304	0.589
				17	1000.001	1.34057	0.487
				18	1000.004	1.79398	0.360
				19	1000.003	3.25199	0.200
400	0.40	0.20	20	1	399.998	0.64826	0.400
				2	400.005	0.40687	0.639
				3	400.002	0.33011	0.782
				4	399.998	0.30093	0.866
				5	399.994	0.29125	0.915
				6	399.999	0.27955	0.940
				7	400.002	**0.27243**	**0.949**
				8	400.001	0.27426	0.945
				9	400.000	0.27947	0.926
				10	399.998	0.29618	0.887

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
400	0.40	0.20	20	11	400.005	0.29838	0.862
				12	400.001	0.31782	0.829
				13	399.995	0.33210	0.788
				14	400.002	0.35356	0.736
				15	400.000	0.38646	0.671
				16	400.001	0.44672	0.589
				17	399.999	0.53587	0.487
				18	400.000	0.72905	0.360
				19	400.004	1.34237	0.200
200	0.40	0.20	20	1	200.003	0.33145	0.400
				2	200.002	0.19914	0.639
				3	200.001	0.16615	0.782
				4	200.001	0.14818	0.866
				5	200.002	0.13988	0.915
				6	199.999	0.13952	0.940
				7	200.000	**0.13739**	**0.949**
				8	199.999	0.13922	0.945
				9	199.998	0.14184	0.926
				10	200.000	0.14599	0.887
				11	200.000	0.15128	0.862
				12	199.996	0.16183	0.829
				13	200.001	0.16666	0.788
				14	199.998	0.18059	0.736
				15	200.002	0.19273	0.671
				16	199.996	0.22166	0.589
				17	200.000	0.27053	0.487
				18	200.002	0.37606	0.360
				19	200.004	0.68291	0.200
100	0.40	0.20	20	1	99.998	0.16127	0.400
				2	99.998	0.10190	0.639
				3	99.999	0.08374	0.782
				4	100.001	0.07608	0.866
				5	100.001	0.07268	0.915
				6	100.000	0.07025	0.940
				7	100.001	**0.06847**	**0.949**
				8	100.000	0.07054	0.945
				9	100.001	0.07009	0.926
				10	100.000	0.07388	0.887
				11	99.998	0.07762	0.862

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	$\mathrm{Var}(\hat{N})$	$\pi_1 \times \pi_2$
100	0.40	0.20	20	12	99.999	0.08063	0.829
				13	100.004	0.08092	0.788
				14	100.000	0.08723	0.736
				15	99.999	0.09772	0.671
				16	100.001	0.11061	0.589
				17	99.998	0.13681	0.487
				18	99.998	0.18570	0.360
				19	99.996	0.32651	0.200
1000	0.40	0.20	11	1	1000.019	39.84736	0.390
				2	999.975	24.98227	0.613
				3	1000.027	21.04894	0.729
				4	999.997	**20.03410**	**0.768**
				5	999.987	20.59317	0.741
				6	999.962	23.81535	0.641
				7	999.984	26.85745	0.568
				8	1000.049	32.42120	0.473
				9	999.987	43.68427	0.351
				10	999.943	78.54668	0.196
400	0.40	0.20	11	1	399.973	15.87043	0.390
				2	400.026	10.05821	0.613
				3	399.989	8.42737	0.729
				4	399.982	**7.96212**	**0.768**
				5	399.997	8.31759	0.741
				6	400.010	9.63558	0.641
				7	400.011	10.82248	0.568
				8	399.987	13.17537	0.473
				9	399.995	17.59126	0.351
				10	400.023	31.95277	0.196
200	0.40	0.20	11	1	200.020	7.97465	0.390
				2	199.987	5.10416	0.613
				3	199.997	4.25128	0.729
				4	199.989	**4.00744**	**0.768**
				5	200.000	4.20080	0.741
				6	200.008	4.80543	0.641
				7	200.012	5.36740	0.568
				8	199.998	6.41656	0.473
				9	199.989	8.83306	0.351
				10	199.998	15.86001	0.196

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
100	0.40	0.20	11	1	100.004	4.05168	0.390
				2	100.004	2.54486	0.613
				3	99.994	2.11171	0.729
				4	99.992	**2.02134**	**0.768**
				5	100.006	2.10949	0.741
				6	100.003	2.41748	0.641
				7	100.008	2.74798	0.568
				8	99.992	3.29612	0.473
				9	99.986	4.51417	0.351
				10	99.980	8.11384	0.196
1000	0.40	0.20	9	1	1000.033	85.09817	0.379
				2	1000.002	54.58878	0.583
				3	999.972	**47.80553**	**0.668**
				4	999.980	48.33574	0.656
				5	1000.023	58.69180	0.544
				6	1000.055	69.48011	0.458
				7	1000.040	93.29367	0.342
				8	1000.096	166.74571	0.192
400	0.40	0.20	9	1	399.967	33.94290	0.379
				2	400.049	22.15873	0.583
				3	399.980	**18.93364**	**0.668**
				4	399.954	19.30434	0.656
				5	399.980	23.58513	0.544
				6	400.007	27.84115	0.458
				7	399.989	37.26586	0.342
				8	399.970	67.24473	0.192
200	0.40	0.20	9	1	200.019	16.96110	0.379
				2	200.001	10.98783	0.583
				3	200.006	**9.59422**	**0.668**
				4	199.976	9.77997	0.656
				5	200.003	11.90724	0.544
				6	199.990	14.21688	0.458
				7	199.994	18.89231	0.342
				8	199.974	34.10754	0.192
100	0.40	0.20	9	1	100.008	8.71122	0.379
				2	99.993	5.55723	0.583
				3	99.999	**4.80709**	**0.668**
				4	100.000	4.95729	0.656
				5	99.992	5.92919	0.544

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
				6	100.000	7.04506	0.458
				7	99.996	9.63569	0.342
				8	100.033	17.68656	0.192
				1	1000.045	337.54914	0.326
				2	999.964	**250.80606**	**0.443**
1000	0.40	0.20	6	3	1000.111	289.18672	0.383
				4	999.908	373.07805	0.298
				5	1000.129	641.18028	0.172
				1	399.948	136.18842	0.326
				2	400.055	**101.30991**	**0.443**
400	0.40	0.20	6	3	400.031	116.22270	0.383
				4	400.028	149.19497	0.298
				5	399.887	262.35257	0.172
				1	200.018	68.47299	0.326
				2	200.001	**50.61695**	**0.443**
200	0.40	0.20	6	3	200.015	58.33231	0.383
				4	200.043	75.46981	0.298
				5	200.047	133.35814	0.172
				1	99.999	34.92987	0.326
				2	100.006	**25.61743**	**0.443**
100	0.40	0.20	6	3	100.005	29.97232	0.383
				4	100.011	39.10938	0.298
				5	99.991	69.14689	0.172
				1	1000.008	10.17658	0.397
				2	999.977	6.44075	0.633
				3	1000.016	5.27905	0.769
				4	1000.004	4.78985	0.843
				5	999.991	4.67459	**0.875**
				6	999.992	**4.65767**	0.871
				7	999.994	4.82338	0.832
				8	999.997	5.38991	0.748
1000	0.40	0.04	20	9	1000.001	6.78132	0.595
				10	999.994	12.05748	0.333
				11	1000.009	12.97420	0.306
				12	1000.000	14.58426	0.277
				13	1000.005	16.29605	0.247
				14	1000.003	18.79028	0.216
				15	1000.002	22.02738	0.184
				16	999.995	27.05554	0.150

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
1000	0.4	0.04	20	17	1000.020	35.83042	0.115
				18	1000.025	51.41786	0.078
				19	1000.005	104.97994	0.040
400	0.40	0.04	20	1	399.994	4.05688	0.397
				2	400.022	2.54910	0.633
				3	400.005	2.07404	0.769
				4	400.001	1.89545	0.843
				5	400.003	**1.84330**	**0.875**
				6	399.996	1.87552	0.871
				7	400.008	1.95673	0.832
				8	400.001	2.13282	0.748
				9	400.002	2.71936	0.595
				10	400.013	4.91493	0.333
				11	399.996	5.30131	0.306
				12	399.994	5.91001	0.277
				13	400.000	6.45882	0.247
				14	399.993	7.52883	0.216
				15	400.008	8.79508	0.184
				16	399.988	10.90968	0.150
				17	399.994	14.23187	0.115
				18	400.007	21.27610	0.078
				19	399.957	42.13856	0.040
200	0.40	0.04	20	1	200.006	2.05477	0.397
				2	200.002	1.28716	0.633
				3	199.995	1.05665	0.769
				4	200.000	0.95498	0.843
				5	200.005	**0.91472**	**0.875**
				6	200.003	0.92777	0.871
				7	200.004	0.95989	0.832
				8	199.994	1.07914	0.748
				9	200.000	1.36110	0.595
				10	200.004	2.44176	0.333
				11	200.000	2.66244	0.306
				12	200.009	2.95805	0.277
				13	200.004	3.30263	0.247
				14	200.020	3.85647	0.216
				15	200.011	4.50684	0.184
				16	199.991	5.53798	0.150
				17	200.011	7.37784	0.115

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
200	0.40	0.04	20	18	200.013	11.44155	0.078
				19	200.057	25.47595	0.040
100	0.40	0.04	20	1	100.003	1.02324	0.397
				2	99.997	0.64569	0.633
				3	100.001	0.52660	0.769
				4	100.000	0.48149	0.843
				5	100.003	0.46347	**0.875**
				6	100.003	**0.46277**	0.871
				7	100.004	0.48616	0.832
				8	99.997	0.54773	0.748
				9	100.007	0.68075	0.595
				10	99.994	1.23514	0.333
				11	100.011	1.36794	0.306
				12	100.006	1.50805	0.277
				13	99.991	1.70111	0.247
				14	99.998	1.88711	0.216
				15	100.009	2.36482	0.184
				16	99.997	2.83676	0.150
				17	99.990	3.76911	0.115
				18	100.005	6.09567	0.078
				19	99.977	12.71870	0.040
1000	0.40	0.04	11	1	1000.015	101.81531	0.375
				2	999.992	66.28857	0.572
				3	999.973	**59.33562**	**0.646**
				4	999.971	61.87069	0.615
				5	1000.013	80.76675	0.471
				6	1000.135	218.95865	0.176
				7	1000.080	266.13141	0.144
				8	1000.114	347.13378	0.110
				9	999.939	514.02562	0.075
				10	1000.035	1015.40640	0.038
400	0.40	0.04	11	1	399.952	40.91505	0.375
				2	400.060	27.01138	0.572
				3	400.003	**23.43184**	**0.646**
				4	399.959	24.65851	0.615
				5	399.981	32.30659	0.471
				6	399.959	88.86733	0.176
				7	399.944	107.20853	0.144
				8	399.928	139.83693	0.110

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
				9	399.991	212.28328	0.075
				10	400.040	424.00587	0.038
200	0.40	0.04	11	1	200.016	20.42160	0.375
				2	200.003	13.45071	0.572
				3	200.024	**11.78262**	**0.646**
				4	199.978	12.48160	0.615
				5	200.007	16.38337	0.471
				6	199.952	44.56189	0.176
				7	199.979	54.75488	0.144
				8	199.940	71.76161	0.110
				9	200.062	113.16873	0.075
				10	200.086	248.57429	0.038
100	0.40	0.04	11	1	100.010	10.50531	0.375
				2	99.999	6.76747	0.572
				3	100.003	**5.91764**	**0.646**
				4	99.995	6.32527	0.615
				5	99.995	8.28599	0.471
				6	99.985	22.84014	0.176
				7	99.956	28.67953	0.144
				8	100.078	38.74779	0.110
				9	100.018	61.09105	0.075
				10	99.895	131.00861	0.038
1000	0.40	0.04	9	1	1000.013	186.03592	0.356
				2	999.975	126.46995	0.523
				3	999.981	**122.47537**	**0.544**
				4	999.990	155.12161	0.427
				5	1000.119	475.62498	0.139
				6	1000.031	621.51216	0.107
				7	1000.216	938.09866	0.073
				8	1000.212	1817.39475	0.037
400	0.40	0.04	9	1	399.935	75.24986	0.356
				2	400.076	51.41765	0.523
				3	400.005	**48.34958**	**0.544**
				4	400.036	62.08107	0.427
				5	400.041	195.96256	0.139
				6	400.034	255.56284	0.107
				7	400.020	376.59993	0.073
				8	399.973	778.90276	0.037

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
200	0.40	0.04	9	1	200.028	37.39335	0.356
				2	199.979	25.60675	0.523
				3	199.982	**24.07146**	**0.544**
				4	199.990	31.55119	0.427
				5	199.991	100.62498	0.139
				6	199.986	128.03501	0.107
				7	199.924	196.46916	0.073
				8	199.841	423.03924	0.037
100	0.40	0.04	9	1	100.018	19.22178	0.356
				2	100.002	12.83212	0.523
				3	99.992	**12.35083**	**0.544**
				4	99.998	16.10005	0.427
				5	100.016	52.96268	0.139
				6	99.992	70.13979	0.107
				7	99.937	107.30220	0.073
				8	99.993	240.08157	0.037
1000	0.40	0.04	6	1	999.866	701.03610	0.273
				2	1000.150	**632.68915**	**0.300**
				3	999.952	2145.40145	0.090
				4	1000.081	3132.18517	0.062
				5	1000.006	6223.96122	0.032
400	0.40	0.04	6	1	400.021	281.38436	0.273
				2	399.949	**257.35220**	**0.300**
				3	400.024	864.12864	0.090
				4	400.149	1288.27374	0.062
				5	399.956	2692.12528	0.032
200	0.40	0.04	6	1	200.035	144.17675	0.273
				2	200.004	**129.19782**	**0.300**
				3	199.937	458.36149	0.090
				4	200.122	700.94095	0.062
				5	199.701	1496.09011	0.032
100	0.40	0.04	6	1	99.990	73.01039	0.273
				2	99.973	**66.60945**	**0.300**
				3	100.049	250.31070	0.090
				4	100.044	396.38946	0.062
				5	99.452	741.17675	0.032

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
1000	0.20	0.30	20	1	1000.014	15.26957	0.199
				2	1000.017	8.46379	0.358
				3	1000.006	6.29728	0.485
				4	1000.009	5.21344	0.586
				5	999.999	4.56433	0.666
				6	1000.005	4.13049	0.729
				7	999.992	3.90014	0.779
				8	999.997	3.72282	0.817
				9	1000.004	3.59890	0.846
				10	999.999	3.51447	0.867
				11	1000.001	3.40374	0.888
				12	1000.001	3.40471	**0.893**
				13	1000.012	**3.37075**	0.884
				14	999.995	3.52558	0.860
				15	1000.007	3.71326	0.817
				16	1000.019	4.01491	0.750
				17	999.990	4.69302	0.651
				18	999.986	6.04208	0.507
				19	1000.022	10.29022	0.299
400	0.20	0.30	20	1	400.008	6.20283	0.199
				2	399.983	3.38080	0.358
				3	399.992	2.47532	0.485
				4	399.999	2.07870	0.586
				5	399.992	1.81612	0.666
				6	400.005	1.66145	0.729
				7	400.004	1.56210	0.779
				8	400.000	1.49455	0.817
				9	400.005	1.44853	0.846
				10	399.998	1.41164	0.867
				11	399.999	1.37313	0.888
				12	400.002	**1.35091**	**0.893**
				13	400.006	1.37652	0.884
				14	400.007	1.40325	0.860
				15	399.997	1.49657	0.817
				16	400.003	1.60754	0.750
				17	399.995	1.88992	0.651
				18	400.008	2.39611	0.507
				19	399.989	4.11279	0.299

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
200	0.20	0.30	20	1	200.000	3.14634	0.199
				2	199.993	1.71248	0.358
				3	199.997	1.26715	0.485
				4	199.996	1.03997	0.586
				5	199.998	0.92867	0.666
				6	199.997	0.82871	0.729
				7	199.996	0.77950	0.779
				8	200.009	0.73719	0.817
				9	200.002	0.72143	0.846
				10	200.004	0.69874	0.867
				11	199.993	0.69883	0.888
				12	200.003	**0.68003**	**0.893**
				13	199.998	0.68031	0.884
				14	199.997	0.70732	0.860
				15	199.997	0.74144	0.817
				16	199.998	0.82277	0.750
				17	200.002	0.94225	0.651
				18	199.991	1.22679	0.507
				19	200.001	2.10092	0.299
100	0.20	0.30	20	1	99.998	1.57657	0.199
				2	99.999	0.86136	0.358
				3	99.995	0.64962	0.485
				4	99.997	0.53097	0.586
				5	100.002	0.46477	0.666
				6	100.004	0.42165	0.729
				7	100.003	0.38948	0.779
				8	99.999	0.37849	0.817
				9	99.999	0.36655	0.846
				10	99.997	0.35703	0.867
				11	100.002	0.34984	0.888
				12	100.003	**0.34263**	**0.893**
				13	100.002	0.34618	0.884
				14	100.001	0.35381	0.860
				15	99.998	0.37726	0.817
				16	100.000	0.40875	0.750
				17	100.003	0.47286	0.651
				18	100.000	0.60835	0.507
				19	99.997	1.04666	0.299

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
1000	0.20	0.30	11	1	1000.035	234.74663	0.189
				2	1000.041	131.72948	0.335
				3	999.926	98.75575	0.446
				4	1000.008	83.92004	0.527
				5	999.976	76.33431	0.582
				6	999.954	71.54065	0.614
				7	999.934	**71.26267**	**0.620**
				8	999.990	76.72793	0.573
				9	1000.002	94.24763	0.464
				10	999.935	157.59457	0.281
400	0.20	0.30	11	1	399.991	94.53089	0.189
				2	399.941	53.44939	0.335
				3	399.990	39.74985	0.446
				4	400.011	33.73059	0.527
				5	400.028	30.13335	0.582
				6	399.992	28.70424	0.614
				7	400.008	**28.43448**	**0.620**
				8	399.974	30.68011	0.573
				9	400.014	38.35742	0.464
				10	400.046	62.96347	0.281
200	0.20	0.30	11	1	199.975	48.01513	0.189
				2	199.988	26.92612	0.335
				3	199.995	19.89610	0.446
				4	199.971	16.97683	0.527
				5	199.989	15.45606	0.582
				6	200.007	14.43744	0.614
				7	200.016	**14.29707**	**0.620**
				8	200.014	15.58594	0.573
				9	199.999	19.29449	0.464
				10	199.991	32.16498	0.281
100	0.20	0.30	11	1	100.015	24.64565	0.189
				2	100.002	13.51635	0.335
				3	100.011	10.09100	0.446
				4	99.975	8.50821	0.527
				5	100.001	7.76646	0.582
				6	100.015	7.35786	0.614
				7	100.019	**7.24066**	**0.620**
				8	99.984	7.80146	0.573
				9	100.006	9.83161	0.464

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
				10	99.971	16.19698	0.281
1000	0.20	0.30	9	1	1000.010	439.70144	0.180
				2	1000.050	249.90893	0.316
				3	999.994	190.99047	0.413
				4	1000.005	165.33367	0.477
				5	999.987	**155.15733**	**0.511**
				6	999.893	155.64554	0.506
				7	999.932	184.92036	0.428
				8	1000.094	296.91160	0.266
400	0.20	0.30	9	1	399.952	177.95142	0.180
				2	399.885	102.26529	0.316
				3	400.036	76.84311	0.413
				4	400.037	66.31521	0.477
				5	400.033	**62.24731**	**0.511**
				6	400.022	62.74837	0.506
				7	400.043	73.37413	0.428
				8	400.047	118.38983	0.266
200	0.20	0.30	9	1	199.997	89.06173	0.180
				2	199.993	50.69520	0.316
				3	200.003	38.01509	0.413
				4	199.964	33.46795	0.477
				5	199.992	31.28852	**0.511**
				6	200.014	**31.24086**	0.506
				7	200.024	37.45130	0.428
				8	199.977	60.37311	0.266
100	0.20	0.30	9	1	99.993	45.85341	0.180
				2	99.997	25.67065	0.316
				3	100.019	19.43870	0.413
				4	100.002	17.07727	0.477
				5	100.004	15.79147	**0.511**
				6	100.018	**15.76262**	0.506
				7	100.019	18.97691	0.428
				8	100.040	31.36693	0.266
1000	0.20	0.30	6	1	1000.074	1139.75533	0.156
				2	1000.041	675.34651	0.261
				3	1000.051	551.12955	0.321
				4	1000.015	**544.74590**	**0.327**
				5	1000.084	792.04127	0.225

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
400	0.20	0.30	6	1	399.958	459.97981	0.156
				2	399.885	275.33325	0.261
				3	399.978	221.46595	0.321
				4	400.081	**217.18555**	**0.327**
				5	399.879	317.49138	0.225
200	0.20	0.30	6	1	199.955	231.63097	0.156
				2	200.025	138.21490	0.261
				3	200.023	110.29669	0.321
				4	199.978	**108.60884**	**0.327**
				5	200.021	160.91161	0.225
100	0.20	0.30	6	1	99.950	121.81119	0.156
				2	99.957	70.12269	0.261
				3	100.034	57.21304	0.321
				4	100.028	**56.06053**	**0.327**
				5	100.011	83.34439	0.225
1000	0.20	0.20	20	1	1000.030	58.52430	0.197
				2	1000.039	32.82862	0.354
				3	999.979	24.47909	0.477
				4	1000.000	20.11527	0.574
				5	1000.017	17.79618	0.649
				6	999.995	16.33161	0.705
				7	999.977	15.33543	0.747
				8	999.997	14.83202	0.775
				9	999.996	**14.52483**	0.791
				10	1000.008	14.54079	**0.797**
				11	999.976	14.54795	0.791
				12	1000.013	14.75667	0.775
				13	1000.024	15.38108	0.747
				14	1000.022	16.25866	0.705
				15	1000.006	17.78856	0.649
				16	1000.007	20.17270	0.574
				17	1000.024	24.48242	0.477
				18	1000.004	32.65594	0.354
				19	1000.000	59.54632	0.197
400	0.20	0.20	20	1	399.988	23.75930	0.197
				2	399.964	13.18055	0.354
				3	399.992	9.59444	0.477
				4	400.018	8.08307	0.574
				5	399.998	7.09768	0.649

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
400	0.20	0.20	20	6	400.001	6.45807	0.705
				7	400.014	6.19439	0.747
				8	400.014	5.98236	0.775
				9	400.017	**5.79329**	0.791
				10	400.001	5.80569	**0.797**
				11	400.021	5.84169	0.791
				12	399.993	5.97018	0.775
				13	399.997	6.27686	0.747
				14	400.011	6.49205	0.705
				15	400.014	7.21365	0.649
				16	399.991	8.10064	0.574
				17	400.003	9.77112	0.477
				18	399.973	13.08119	0.354
				19	399.989	23.51883	0.197
200	0.20	0.20	20	1	200.005	11.99455	0.197
				2	200.010	6.73770	0.354
				3	200.007	4.90508	0.477
				4	199.987	4.04614	0.574
				5	200.005	3.62285	0.649
				6	199.993	3.27424	0.705
				7	200.003	3.08832	0.747
				8	200.020	2.99533	0.775
				9	199.992	**2.92758**	0.791
				10	199.991	2.95325	**0.797**
				11	199.994	2.96866	0.791
				12	200.009	2.99711	0.775
				13	200.004	3.09738	0.747
				14	200.005	3.28713	0.705
				15	199.987	3.58421	0.649
				16	199.991	4.13288	0.574
				17	200.001	4.87042	0.477
				18	200.012	6.67900	0.354
				19	200.023	12.21611	0.197
100	0.20	0.20	20	1	99.998	6.01710	0.197
				2	100.004	3.36582	0.354
				3	99.996	2.49414	0.477
				4	99.985	2.05373	0.574
				5	99.997	1.81509	0.649
				6	100.009	1.63037	0.705

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
100	0.20	0.20	20	7	100.007	1.55642	0.747
				8	99.991	1.51284	0.775
				9	99.996	1.47323	0.791
				10	100.000	**1.45288**	**0.797**
				11	100.004	1.46591	0.791
				12	100.001	1.50713	0.775
				13	100.007	1.55027	0.747
				14	100.009	1.63094	0.705
				15	99.996	1.81978	0.649
				16	99.992	2.04158	0.574
				17	100.005	2.51453	0.477
				18	99.996	3.34815	0.354
				19	99.984	6.15262	0.197
1000	0.20	0.20	11	1	1000.003	487.80940	0.179
				2	1000.025	275.69795	0.312
				3	1000.005	212.88121	0.406
				4	1000.023	185.04754	0.467
				5	999.966	174.61199	0.496
				6	999.895	**173.93271**	**0.496**
				7	999.981	185.76239	0.467
				8	999.959	211.58431	0.406
				9	1000.054	276.34476	0.312
				10	999.862	482.51660	0.179
400	0.20	0.20	11	1	399.950	195.95300	0.179
				2	399.886	113.08298	0.312
				3	400.023	85.41825	0.406
				4	400.048	73.98121	0.467
				5	400.041	70.13034	0.496
				6	399.978	**70.01883**	**0.496**
				7	400.032	73.21792	0.467
				8	400.060	85.61335	0.406
				9	399.955	111.05639	0.312
				10	399.977	194.28128	0.179
200	0.20	0.20	11	1	199.997	98.14625	0.179
				2	200.009	56.24801	0.312
				3	199.993	42.17733	0.406
				4	199.965	37.41929	0.467
				5	200.002	35.18198	0.496
				6	200.036	**34.99074**	**0.496**

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
200	0.20	0.20	11	7	200.027	36.85606	0.467
				8	199.971	42.99717	0.406
				9	200.024	56.75058	0.312
				10	200.008	99.91992	0.179
100	0.20	0.20	11	1	99.993	50.83936	0.179
				2	100.008	28.36625	0.312
				3	100.013	21.53170	0.406
				4	99.995	19.13093	0.467
				5	99.996	**17.74922**	**0.496**
				6	100.032	17.77946	0.496
				7	100.017	19.02649	0.467
				8	100.029	21.81404	0.406
				9	100.008	28.83897	0.312
				10	99.968	50.58676	0.179
1000	0.20	0.20	9	1	1000.048	814.86809	0.166
				2	1000.042	472.15831	0.285
				3	1000.029	376.13107	0.360
				4	999.993	**340.82486**	**0.397**
				5	999.916	342.79984	0.397
				6	1000.142	374.31619	0.360
				7	1000.063	476.03566	0.285
				8	1000.003	808.44415	0.166
400	0.20	0.20	9	1	399.934	327.64490	0.166
				2	399.879	191.95925	0.285
				3	400.012	149.82985	0.360
				4	400.012	**136.10755**	**0.397**
				5	400.059	135.93427	0.397
				6	400.042	151.44799	0.360
				7	399.942	188.99205	0.285
				8	399.971	326.39290	0.166
200	0.20	0.20	9	1	199.996	164.91882	0.166
				2	200.034	96.01413	0.285
				3	200.036	74.86640	0.360
				4	199.979	**68.16321**	**0.397**
				5	199.971	68.63594	0.397
				6	199.959	75.98330	0.360
				7	199.978	96.41995	0.285
				8	199.926	166.67589	0.166

Table B.1 – continued from previous page

N	θ_1	θ_2	T	t	$\hat{N}$	Var($\hat{N}$)	$\pi_1 \times \pi_2$
100	0.20	0.20	9	1	99.932	86.77309	0.166
				2	99.956	48.49649	0.285
				3	100.030	38.52192	0.360
				4	100.006	35.44022	0.397
				5	100.000	**34.79634**	**0.397**
				6	99.998	38.76851	0.360
				7	99.983	49.46244	0.285
				8	100.095	87.46232	0.166
1000	0.20	0.20	6	1	999.825	1968.03248	0.134
				2	1000.079	1239.45606	0.213
				3	999.932	**1109.51589**	**0.238**
				4	1000.006	1246.42894	0.213
				5	1000.199	1986.55606	0.134
400	0.20	0.20	6	1	399.967	804.39065	0.134
				2	399.864	502.29215	0.213
				3	399.928	**443.48720**	**0.238**
				4	399.970	503.18223	0.213
				5	399.945	803.98924	0.134
200	0.20	0.20	6	1	200.020	407.30970	0.134
				2	200.037	255.11974	0.213
				3	199.965	**225.04819**	**0.238**
				4	200.016	254.28274	0.213
				5	200.077	414.34798	0.134
100	0.20	0.20	6	1	99.941	217.76051	0.134
				2	99.970	130.43723	0.213
				3	99.991	**117.19450**	**0.238**
				4	100.034	135.50141	0.213
				5	99.995	218.07184	0.134

References

Akkaya-Hocagil, T., Hsu, W. H., Sommerhalter, K., McGarry, C., and Van Zutphen, A. (2017). Utility of capture–recapture methodology to estimate prevalence of congenital heart defects among adolescents in 11 New York state counties: 2008 to 2010. *Birth Defects Research*, 109(18):1423–1429.

Alderete, J. and Davies, M. (2019). Investigating perceptual biases, data reliability, and data discovery in a methodology for collecting speech errors from audio recordings. *Language and Speech*, 62(2):281–317.

Amstrup, S. C., Manly, B. F. J., and McDonald, T. L., editors (2006). *Handbook of Capture-Recapture Analysis*. Princeton University Press, Princeton, N.J.

Anan, O., Böhning, D., and Maruotti, A. (2017). Uncertainty estimation in heterogeneous capture–recapture count data. *Journal of Statistical Computation and Simulation*, 87(10):2094–2114.

Bailey, N. T. J. (1951). On estimating the size of mobile populations from recapture data. *Biometrika*, 38:293–306.

Bailly, L., David, R., Chevrier, R., Grebet, J., Moncada, M., Fuch, A., Sciortino, V., Robert, P., and Pradier, C. (2019). Alzheimer's disease: Estimating its prevalence rate in a French geographical unit using the National Alzheimer Data Bank and national health insurance information systems. *PLoS ONE*, 14(5):1–11.

Basu, S. and Ebrahimi, N. (2001). Bayesian capture-recapture methods for error detection and estimation of population size: Heterogeneity and dependence. *Biometrika*, 88(1):269–279.

Bird, S. M. and King, R. (2018). Multiple systems estimation (or capture-recapture estimation) to inform public policy. *Annual Review of Statistics and Its Application*, 5(1):95–118.

Borchers, D. L., Buckland, S. T., and Zucchini, W. (2002). *Estimating Animal Abundance: Closed Populations*. Statistics for Biology and Health. Springer, London.

Bröder, L., Tatin, L., Hochkirch, A., Schuld, A., Pabst, L., and Besnard, A. (2020). Optimization of capture–recapture monitoring of elusive species illustrated with a threatened grasshopper. *Conservation Biology*, 34(3):743–753.

Buckland, S. T. and Garthwaite, P. H. (1991). Quantifying precision of mark-recapture estimates using the bootstrap and related Methods. *Biometrics*, 47(1):255–268.

Burke, V. J., Greene, J. L., and Gibbons, J. W. (1995). The effect of sample size and study duration on metapopulation estimates for slider turtles (Trachemys scripta). *Herpetologica*, 51(4,):451–456.

Burnham, K. P., Anderson, D. R., White, G. C., Brownie, C., and Pollock, K. H. (1987). *Design and Analysis Methods for Fish Survival Experiments Based on Release-Recapture*. Number 5 in American Fisheries Society Monograph. American Fisheries Society, Bethesda, Md.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

Böhning, D. (2015). Power series mixtures and the ratio plot with applications to zero-truncated count distribution modelling. *METRON*, 73(2):201–216.

Böhning, D., Dietz, E., Kuhnert, R., and Schön, D. (2005). Mixture models for capture-recapture count data. *Statistical Methods and Applications*, 14(1):29–43.

Böhning, D. and Kuhnert, R. (2006). Equivalence of truncated count mixture distributions and mixtures of truncated count distributions. *Biometrics*, 62(4):1207–1215.

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

Böhning, D., van der Heijden, P. G., and Bunge, J. (2018). *Capture-Recapture Methods for the Social and Medical Sciences*. Chapman and Hall/CRC, Boca Raton, FL.

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, 43(4):783–791.

Chao, A. and Hugginns, R. M. (2006). Classical closed-population capture–recapture models. In *Handbook of Capture-Recapture Analysis*, pages 21–35. Princeton University Press, New Jersey.

Chapman, D. (1951). *Some Properties of The Hypergeometric Distribution with Applications to Zoological Sample Censuses*. University of California publications in statistics. University of California Press.

Charette, Y. and van Koppen, V. (2016). A capture-recapture model to estimate the effects of extra-legal disparities on crime funnel selectivity and punishment avoidance. *Security Journal*, 29(4):561–583.

Chin, S. N. and Böhning, D. (2025). Choice of sampling effort in a Schnabel census for accurate population size estimates. *Environmental and Ecological Statistics*, 32(2):753–769.

Chin, S. N., Overstall, A., and Böhning, D. (2026). Sampling effort and its allocation in the Lincoln–Petersen experiment: A hierarchical approach. *Journal of Statistical Planning and Inference*, 241:106330.

Cochran, W. G. (1977). *Sampling Techniques*. Wiley Series in Probability and Mathematical statistics. J. Wiley & sons, New York, 3rd edition.

Cochran, W. G. (1978). Laplace's ratio estimator. In David, H., editor, *Contributions to Survey Sampling and Applied Statistics*, pages 3–10. Academic Press.

Conner, M. M., Bennett, S. N., Saunders, W. C., and Bouwes, N. (2015). Comparison of tributary survival estimates of steelhead using Cormack–Jolly–Seber and barker models: implications for sampling efforts and designs. *Transactions of the American Fisheries Society*, 144(1):34–47.

Cushwa, C. T. and Burnham, K. P. (1974). An inexpensive live trap for snowshoe hares. *The Journal of Wildlife Management*, 38(4):939–941.

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

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Edwards, W. R. . and Eberhardt, L. (1967). Estimating cottontail abundance from livetrapping Data. *The Journal of Wildlife Management*, 31(1):87–96.

Efford, M. G., Borchers, D. L., and Mowat, G. (2013). Varying effort in capture-recapture studies. *Methods in Ecology and Evolution*, 4(7):629–636.

Egeland, G. M., Perham-hester, K. A., and Hook, E. B. (1995). Use of capture-recapture analyses in fetal alcohol syndrome surveillance in Alaska. *American Journal of Epidemiology*, 141(4):335–341.

Frusher, S. D., Johnson, C. R., and Ziegler, P. E. (2003). Space-time variation in catchability of southern rock lobster Jasus edwardsii in Tasmania explained by environmental, physiological and density-dependent processes. *Fisheries Research*, 61(1-3):107–123.

Gaskell, T. J. and George, B. J. (1972). A Bayesian modification of the Lincoln index. *The Journal of Applied Ecology*, 9(2):377.

Gerritse, S. C., van der Heijden, P. G., and Bakker, B. F. (2015). Sensitivity of population size estimation for violating parametric assumptions in log-linear models. *Journal of Official Statistics*, 31(3):357–379.

Ghojazadeh, M., Mohammadi, M., Azami-Aghdash, S., Sadighi, A., Piri, R., and Naghavi-Behzad, M. (2013). Estimation of cancer cases using capture-recapture method in northwest iran. *Asian Pacific Journal of Cancer Prevention*, 14(5):3237–3241.

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, 40(3/4):237–264.

Graunt, J. (1939). *Natural and Political Observations Made Upon the Bills of Mortality.* Baltimore The Johns Hopkins Press.

Greenwood, J. J. D. and Robinson, R. A. (2006). General census methods. In Sutherland, W. J., editor, *Ecological Census Techniques*, pages 87–185. Cambridge University Press, 2nd edition.

Hickey, J. R. and Sollmann, R. (2018). A new mark-recapture approach for abundance estimation of social species. *PLoS ONE*, 13(12):1–14.

Hosmer, D. W., Lemeshow, S., and May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data.* Wiley Series in Probability and Statistics. Wiley, 1 edition.

Howe, E. J., Obbard, M. E., and Kyle, C. J. (2013). Combining data from 43 standardized surveys to estimate densities of female American black bears by spatially explicit capture-recapture. *Population Ecology*, 55(4):595–607.

Jiménez-Ruiz, S., Rafael, M., Coelho, J., Pacheco, H., Fernandes, M., Alves, P. C., and Santos, N. (2023). High mortality of wild european rabbits during a natural outbreak of rabbit haemorrhagic disease GI.2 revealed by a capture-mark-recapture study. *Transboundary and Emerging Diseases*, 2023:1–9.

Kelly, S., MacDiarmid, A. B., and Babcock, R. C. (1999). Characteristics of spiny lobster, Jasus edwardsii, aggregations in exposed reef and sandy areas. *Marine and Freshwater Research*, 50(5):409–416.

King, R. and McCrea, R. (2019). Capture–recapture methods and models: estimating population size. In *Handbook of Statistics*, volume 40, pages 33–83. Elsevier.

Kordjazi, Z., Frusher, S., Buxton, C., Gardner, C., and Bird, T. (2016). The influence of mark-recapture sampling effort on estimates of rock lobster survival. *PLoS ONE*, 11(3):e0151683.

Krebs, C. (2014). Ecological methodology part one: estimating abundance in animal and plant populations. *Ecological Methodology*, pages 24–77.

Lange, K. (2010). *Numerical Analysis for Statisticians*. Statistics and Computing. Springer, New York.

Lanumteang, K. (2011). *Estimation of the size of a target population using capture-recapture methods based upon multiple sources and continuous time experiments*. PhD thesis, University of Reading, School of Mathematical and Physical Sciences.

Laplace, P. S. (1786). Sur les naissances, les mariages et les mortsa Paris depuis 1771 jusqu'a 1784 et dans toute l'étendue de la France, pendant les années 1781 et 1782. *Mémoires de l'Académie Royale des Sciences présentés par divers savans*, pages 35–46.

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

Lerdsuwansri, R. and Böhning, D. (2014). A new estimator of population size based upon the conditional independent Poisson Mixture Model. In *Proceedings of the International Conference on Applied Statistics*, pages 1–8, Khon Kaen, Thailand.

Lewis, C. E. and Hassanein, K. M. (1969). The relative effectiveness of different approaches to the surveillance of infections among hospitalized patients. *Medical Care*, 7(5):379–384.

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

Lu, J. and Li, D. (2010). Estimating deep web data source size by capture–recapture method. *Information Retrieval*, 13(1):70–95.

Manly, B. F. J., McDonald, T. L., and Amstrup, S. C. (2005). Introduction to the handbook. In *Handbook of Capture-Recapture Analysis*, pages 1–21. Princeton University Press, New Jersey.

Matechou, E. and Argiento, R. (2023). Capture-recapture models with heterogeneous temporary emigration. *Journal of the American Statistical Association*, 118(541):56–69.

Matechou, E. and Caron, F. (2017). Modelling individual migration patterns using a Bayesian nonparametric approach for capture–recapture data. *Annals of Applied Statistics*, 11(1):21–40.

McCrea, R. S. and Morgan, B. J. (2015). *Analysis of Capture-Recapture Data*. CRC Press, Boca Raton, FL.

McKelvey, K. S. and Pearson, D. E. (2001). Population estimation with sparse data: the role of estimators versus indices revisited. *Canadian Journal of Zoology*, 79(10):1754–1765.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. Wiley Series in Probability and Statistics. Applied Probability and Statistics Section. Wiley, New York.

McWherter, G. R. (1991). Estimating abundance of cottontail rabbits using live trapping and visual surveys. Master's thesis, University of Tennessee.

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

Papadatou, E., Pradel, R., Schaub, M., Dolch, D., Geiger, H., Ibañez, C., Kerth, G., Popa-Lisseanu, A., Schorcht, W., Teubner, J., and Gimenez, O. (2012). Comparing survival among species with imperfect detection using multilevel analysis of mark-recapture data: A case study on bats. *Ecography*, 35(2):153–161.

Paterson, J. T., Proffitt, K., Jimenez, B., Rotella, J., and Garrott, R. (2019). Simulation-based validation of spatial capture-recapture models: A case study using mountain lions. *PLoS ONE*, 14(4):e0215458.

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

Petersson, H., Thelin, T., Runeson, P., and Wohlin, C. (2004). Capture–recapture in software inspections after 10 years research—theory, evaluation and application. *Journal of Systems and Software*, 72(2):249–264.

Plettinckx, E., Crawford, F. W., Antoine, J., Gremeaux, L., and Van Baelen, L. (2021). Estimates of people who injected drugs within the last 12 months in Belgium based on a capture-recapture and multiplier method. *Drug and Alcohol Dependence*, 219.

Plouvier, S. D., Bernillon, P., Ligier, K., Theis, D., Miquel, P. H., Pasquier, D., and Rivest, L. P. (2019). Completeness of a newly implemented general cancer registry in northern France: Application of a three-source capture-recapture method. *Revue d'Epidemiologie et de Sante Publique*, 67(4):239–245.

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

Pollock, K. H., Marsh, H. D., Lawler, I. R., and Alldredge, M. W. (2006). Estimating animal abundance in heterogeneous environments: an application to aerial surveys for dugongs. *Journal of Wildlife Management*, 70(1):255–262.

Pollock, K. H., Nichols, J. D., Brownie, C., and Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife Monograph*, 107:1–97.

Raag, M., Vorobjov, S., and Uusküla, A. (2019). Prevalence of injecting drug use in Estonia 2010-2015: A capture-recapture study. *Harm Reduction Journal*, 16(1):1–8.

Ramos, P. L., Sousa, I., Santana, R., Morgan, W. H., Gordon, K., Crewe, J., Rocha-Sousa, A., and Macedo, A. F. (2020). A review of capture-recapture methods and its possibilities in ophthalmology and vision sciences. *Ophthalmic Epidemiology*, 27(4):310–324.

Reinke, B. A., Hoekstra, L., Bronikowski, A. M., Janzen, F. J., and Miller, D. (2020). Joint estimation of growth and survival from mark–recapture data to improve estimates of senescence in wild populations. *Ecology*, 101(1):1–7.

Robson, D. S. and Regier, H. A. (1964). Sample size in Petersen mark-recapture experiments. *Transactions of the American Fisheries Society*, 93(3):215–226.

Rocchetti, I., Böhning, D., Holling, H., and Maruotti, A. (2020). Estimating the size of undetected cases of the COVID-19 outbreak in Europe: an upper bound estimator. *Epidemiologic Methods*, 9(s1):20200024.

Sanderlin, J. S., Block, W. M., and Ganey, J. L. (2014). Optimizing study design for multi-species avian monitoring programmes. *Journal of Applied Ecology*, 51(4):860–870.

Sanderson, M., Benjamin, J. T., Lane, M. J., Cornman, C. B., and Davis, D. R. (2003). Application of capture-recapture methodology to estimate the prevalence of dementia in South Carolina. *Annals of Epidemiology*, 13(7):518–524.

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

Schorr, R. A., Ellison, L. E., and Lukacs, P. M. (2014). Estimating sample size for landscape-scale mark-recapture studies of north american migratory tree bats. *Acta Chiropterologica*, 16(1):231–239.

Seber, G. (1982). *The Estimation of Animal Abundance and Related Parameters*. Griffin Publishing Company, London.

Seber, G. A. F., Huakau, J. T., and Simmons, D. (2000). Capture-recapture, epidemiology, and list mismatches: two lists. *Biometrics*, 56(4):1227–1232.

Seber, G. A. F. and Schofield, M. R. (2023). *Estimating Presence and Abundance of Closed Populations*. Statistics for Biology and Health. Springer International Publishing, Cham.

Sekar, C. C. and Deming, W. E. (1949). On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association*, 44(245):101–115.

Skalski, J. R. and Robson, D. S. (1982). A mark and removal field procedure for estimating population abundance. *The Journal of Wildlife Management*, 46(3):741.

Somers, K. M. . and Stechey, D. P. . M. . (1986). Variable trappability of crayfish associated with bait type , water temperature and lunar phase. *The American Midland Naturalist*, 116(1):36–44.

Straetemans, M., Bakker, M. I., Alba, S., Mergenthaler, C., Rood, E., Andersen, P. H., Schimmel, H., Simunovic, A., Svetina, P., Carvalho, C., Lyytikäinen, O., Abubakar, I., Harris, R. J., Ködmön, C., Van Der Werf, M. J., and Van Hest, R. (2020). Completeness of tuberculosis (TB) notification: inventory studies and capture-recapture analyses, six european union countries, 2014 to 2016. *Eurosurveillance*, 25(12):1900568–1900568.

Tajuddin, R. R. M., Ismail, N., and Ibrahim, K. (2022). Estimating population size of criminals: a new Horvitz–Thompson estimator under one-inflated positive Poisson–Lindley model. *Crime & Delinquency*, 68(6-7):1004–1034.

Thompson, W. (2004). *Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters*. Island Press, Washington.

Thompson, W. L., Thompson, W. L., White, G. C., and Gowan, C. (1998). *Monitoring Vertebrate Populations*. Academic Press, San Diego.

Wang, X., He, Z., and Sun, D. (2015). Bayesian estimation of population size via capture-recapture model with time variation and behavioral response. *Open Journal of Ecology*, 05(01):1–13.

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and Management of Animal Populations*. Academic Press, 1st edition.

Williams, J., Segalowitz, N., and Leclair, T. (2014). Estimating second language productive vocabulary size: A Capture-Recapture approach. *The Mental Lexicon*, 9(1):23–47.

Williams, M. J. and Hill, B. J. (1982). Factors influencing pot catches and population estimates of the portunid crab Scylla serrata. *Marine Biology*, 71(2):187–192.

Wittes, J. and Sidel, V. W. (1968). A generalization of the simple capture-recapture model with applications to epidemiological research. *Journal of Chronic Diseases*, 21(5):287–301.

Wolter, K. M. (1990). Capture-recapture estimation in the presence of a known sex ratio. *Biometrics*, 46(1):157–162.

Xi, L., Watson, R., and Yip, P. S. F. (2008). The minimum capture proportion for reliable estimation in capture–recapture models. *Biometrics*, 64(1):242–249.

Ypma, T. J. (1995). Historical development of the Newton–Raphson method. *SIAM Review*, 37(4):531–551.

Ziegler, P. E., Haddon, M., Frusher, S. D., and Johnson, C. R. (2004). Modelling seasonal catchability of the southern rock lobster Jasus edwardsii by water temperature, moulting, and mating. *Marine Biology*, 145(1):179–190.