Inference from Single Occasion Capture Experiments Using Genetic Markers

Chathurika Kumari Hettiarachchi Hettiarachchige

School of Mathematics and Statistics,
The University of Melbourne

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

June, 2016
Abstract

Accurate estimation of the size of animal populations is an important task in ecological science. However, it is still a challenge in the case of estimation for many endangered species. Classical capture-recapture models have not been able to address this issue to complete satisfaction. Estimating the population size using genetic data also has sometimes led to estimates that differ markedly from each other and also from classical capture-recapture estimates. We use genetic markers to estimate the number of mothers in a population using a single occasion capture experiment. Our work is based on the idea that the degree of biological relatedness between individuals in a sample contributes to the estimation of the size of a population, especially when the population is closed, following Skaug [33]. We mainly construct our estimators using method of moments. First we show that it is feasible to estimate the population size using mother-daughter pairs in the sample, identified using genetic data, observed from a population consisting of two generations. Then we incorporate other relationships into the estimators and extend our method to populations consisting of more than two generations. We show that our estimators are statistically consistent and estimate their standard errors. We also consider the maximum likelihood method in this setting and discover problems that preclude its use. However, we show the method of moments produces reliable estimates. Finally, we study the impact of the misclassification of relationships on the proposed estimators. We illustrate our methods on small subsets of data collected at several sites at Mt Hotham, Victoria, Australia between 2010-2012.
Declaration

This is to certify that:

1. the thesis comprises only my original work towards the PhD;

2. due acknowledgement has been made in the text to all other material used; and

3. the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Chathurika Kumari Hettiarachchi Hettiarachchige
Acknowledgements

I would first like to express my sincerest gratitude to my supervisor, Prof. Richard Huggins; you have been a tremendous mentor for me. I’m thankful for giving me the opportunity to work with you, for your time, patience and encouragement. Your excellent guidance, insights and knowledge have been exceptionally valuable while your relaxed nature made this journey smooth and enjoyable.

I’m grateful to Dr. Andrew Weeks who provided the data on the mountain pygmy possum, and also for his expert biological advice. My sincere thanks to Dr. Guoqi Qian, the chair of my advisory committee, for his valuable discussions and ideas. I would like to express my deep gratitude to Dr. Davide Ferrari for his insightful discussions and suggestions which improved my thesis. I’m grateful to A/Prof Rachel Fewster for her personal communication on the research.

I would like to extend my sincere appreciation to Kirsten Hoak, Ben Hess and other professional staff of the School of Mathematics and Statistics and the staff of Mathematics and Statistics Learning Centre for the assistance and support they provided during my PhD.

My time at the University of Melbourne was made enjoyable in large part due to the many friends and groups that became a part of my life. Thanks a lot to my colleagues in the department and members of the Sri Lankan Graduates’ Society for helpful discussions and encouragements. The time spent with you, various discussions and encouragements were undoubtedly helpful. I’m grateful to Liz Atkinson, Sabrina Rodrigues and Kasun Rathnayake for reading my thesis.

Most importantly, none of this would have been possible without the love and patience of my family. I would like to thank my parents, brother, my sisters, sister and
brothers-in-law, nephews and nieces. Your love, encouragement, trust and prayers always bring me strength and happiness. I could not have finished this work without you and your dreams on me.
Contents

List of Figures xiii
List of Tables xv

1 Introduction 1

1.1 Classical capture-recapture analysis 2

1.1.1 Closed population 3

1.1.2 Open population 5

1.1.3 Computer programs for estimating the population size 7

1.2 Genetic capture-recapture analysis 8

1.2.1 Paternity analysis 9

1.2.2 Genetic distribution 12

1.2.3 Allele sharing methods 14

1.3 Related work on mountain pygmy possums 16

1.3.1 Simulation studies 18

1.4 Thesis overview 19

2 Estimating the number of mothers in a population with a simple structure 21

2.1 Background 21

2.2 Notation and assumptions 23

2.3 Estimators 25

2.3.1 Moment equalities 25

2.3.2 Mean number of daughters per mother is known 28

2.3.3 Mean number of daughters per mother is unknown 29

2.4 Consistency 31

2.5 Estimator of the standard errors 33

2.5.1 Theoretical estimate of the standard error 34

2.5.2 Bootstrap estimate of the standard error 42

2.6 Simulation studies 44

2.6.1 Performance of the proposed estimators $\hat{N}$, $\hat{\mu}$ and $\hat{p}$ 45

2.6.2 The effect of capture probability on the estimators 53
## Contents

2.6.3 Estimating the total population ........................................ 56
2.7 Estimating N when the number of mother-daughter pairs in the sample is unknown .......................................................... 58
2.8 Application ................................................................................. 60
2.9 Populations with mean number of daughters per mother less than one 62
  2.9.1 High probability of observing an equal number of mothers and daughters .......................................................... 62
  2.9.2 Sensitivity of the estimators to the observed number of mothers and daughters .......................................................... 68
2.10 Discussion ................................................................................. 70

3 Estimating the number of mothers in a population with a simple structure using maximum likelihood 73
  3.1 Background ............................................................................... 73
  3.2 Notation and assumptions ............................................................ 74
  3.3 The estimator of the number of mothers in a population ................. 75
  3.4 Limitations of the maximum likelihood estimators ......................... 78
    3.4.1 Boundary solutions ................................................................ 78
    3.4.2 Profile likelihood ................................................................... 80
    3.4.3 Non-unique solution ............................................................... 82
    3.4.4 Computational problems .......................................................... 83
    3.4.5 Variance .............................................................................. 87
  3.5 Application ................................................................................. 90
  3.6 Discussion ................................................................................. 90

4 Estimating the size of a population with three generations 93
  4.1 Background ............................................................................... 93
  4.2 Notation and assumptions ............................................................ 94
  4.3 Estimators .................................................................................. 95
    4.3.1 Moment equalities ................................................................. 96
    4.3.2 Mean number of daughters per mother is known ...................... 98
    4.3.3 Mean number of daughters per mother is unknown .................... 100
  4.4 Estimating N when the number of mother-daughter pairs in the sample is unknown .......................................................... 102
  4.5 Estimator of the standard error ..................................................... 103
  4.6 Simulation studies ....................................................................... 104
    4.6.1 Comparison of the two estimators of N .................................... 104
    4.6.2 Performance of $\hat{N}$ ............................................................ 106
  4.7 Application ................................................................................. 113
  4.8 Discussion ................................................................................. 114
  4.A Consistency of the estimators ..................................................... 116
## 5 More precise estimator of the number of mothers in a population with more complicated population structure

5.1 Background ................................................. 119

5.2 Two generations ........................................ 120

5.2.1 Notation and assumptions ............................ 120

5.2.2 Estimators ............................................. 121

5.2.2.1 Moment equalities ............................... 122

5.2.2.2 Mean number of daughters per mother is known ... 124

5.2.2.3 Mean number of daughters per mother is unknown . 125

5.2.3 Estimator of the standard errors ...................... 127

5.2.4 Simulation studies ...................................... 128

5.3 Three generations ........................................ 133

5.3.1 Notation and assumptions ............................ 133

5.3.2 Estimators ............................................. 134

5.3.2.1 Moment equalities ............................... 134

5.3.2.2 Mean number of daughters per mother is known ... 136

5.3.2.3 Mean number of daughters per mother is unknown . 137

5.3.3 Estimator of the standard errors ...................... 139

5.3.4 Simulation studies ...................................... 139

5.4 Discussion ................................................ 143

5.A Consistency of the estimators for two generations .... 144

5.B Consistency of the estimators for three generations .... 147

## 6 The effect of misclassification of relationships

6.1 Background ................................................. 151

6.1.1 False discovery rate .................................. 151

6.2 Notation and assumptions ............................... 152

6.3 Simulation studies ........................................ 153

6.4 Limiting distributions ...................................... 156

6.5 Misclassification error .................................... 159

6.5.1 Misclassification error in \( \hat{d} \) ...................... 161

6.5.2 Misclassification error in \( \hat{\mu p} \) ..................... 166

6.5.3 Misclassification error in \( \hat{\mu} \) ....................... 168

6.5.4 Misclassification error in \( \hat{N} \) ....................... 173

6.6 Discussion ................................................ 179

## 7 Conclusions and future work

181
List of Figures

2.1 Distribution of \( \hat{N} \) ................................................................. 50
2.2 Distribution of \( \hat{\mu} \) ................................................................. 51
2.3 Distribution of \( \hat{\rho} \) ................................................................. 52
2.4 Theoretical standard errors of \( \hat{N} \) and \( \hat{N}(\mu) \) against \( p \) ............... 55
2.5 Percentage bias of \( \hat{N} + \hat{D} \) ....................................................... 56
2.6 Allele combinations of parents and their children ................................. 59
2.7 Map of \textit{Burramys parvus} collections sites in Mount Hotham, Australia 60
2.8 The probability \( d = x \) when \( p = 0.3 \) and \( \mu \leq 1 \) ....................... 64
2.9 The probability \( d = x \) when \( p = 0.3 \) and \( 0 \leq \mu \leq 5 \) ....................... 65
2.10 \( Pr(d = x|N, \mu) \) against \( \mu \) for different values of \( N \) \( (p = 0.3) \) ............. 66
2.11 \( Pr(d = x|N, \mu) \) against \( N \) for different values of \( \mu \) \( (p = 0.3) \) ............. 66
2.12 \( Pr(d = x|N, \mu) \) against \( p \) for different values of \( \mu \) \( (N = 100) \) ............. 67
2.13 Rate of change of \( \hat{\mu}p \) against \( \bar{d} \) ........................................... 69
3.1 Likelihood surface for a sample with \( n=43, d=5 \) and \( x=4 \) when \( N=38 \) 80
3.2 Profile likelihood surface for a sample with \( n=43, d=5 \) and \( x=4 \) ....... 82
3.3 Likelihood surface for \( N = 40, 50, 60, 68 \) (moment estimate for this sample), 100 and 150 ......................................................... 86
4.1 Bias of different estimators of \( N \) ..................................................... 105
4.2 Distribution of \( \hat{N} \) ................................................................. 109
4.3 Distribution of \( \hat{\mu} \) ................................................................. 110
4.4 Distribution of \( \hat{\rho} \) ................................................................. 111
4.5 Percentage bias of \( \hat{N} + \hat{D} \) ....................................................... 113
5.1 Distribution of \( \hat{N} \) ................................................................. 131
5.2 Distribution of \( \hat{\mu} \) ................................................................. 142
6.1 Estimates of \( \hat{N} \) against \( \beta \) ......................................................... 158
6.2 Estimates of \( \hat{\mu} \) against \( \beta \) ......................................................... 158
6.3 \( \frac{d}{\beta} \) against \( \beta \) ................................................................. 165
6.4 \( \hat{\mu}p_o - \hat{\mu}p \) against \( \beta \) ......................................................... 167
6.5 \( \hat{\mu}_o - \hat{\mu} \) against \( \beta \) ......................................................... 172
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.6</td>
<td>$\frac{\hat{N} - \tilde{N}}{N}$ against $\beta$</td>
<td>175</td>
</tr>
<tr>
<td>6.7</td>
<td>Percentage of error of $\hat{N}$ caused by misclassification against $\tilde{N}$ when $\beta = 0.05$</td>
<td>176</td>
</tr>
<tr>
<td>6.8</td>
<td>Percentage of error of $\hat{N}$ caused by misclassification against $\tilde{N}$ when $\beta = 0.1, \beta = 0.15$</td>
<td>178</td>
</tr>
</tbody>
</table>
## List of Tables

1.1 Estimates of size of different populations presented in Mitrovski et al. [20] ................................................. 17
1.2 Simulation results for the estimators used by Mitrovski et al. [20] ............................................................... 19

2.1 Notation .......................................................................................................................... 24
2.2 Averages of 1000 estimates of the population parameters ......................................................... 46
2.3 Theoretical estimates of standard errors .................................................................................. 47
2.4 Bootstrap estimates of standard errors .................................................................................... 49
2.5 Simulation results illustrating the increased precision at higher $p$ ($p=0.5$) ......................................................... 53
2.6 Simulation results with $p=0.15$, 0.2 and 0.7 .............................................................................. 54
2.7 Standard errors of $\hat{N} + \hat{D}$ ............................................................................................. 57
2.8 The application data .............................................................................................................. 61
2.9 Parameter estimates for the three sites .................................................................................... 61
2.10 Examples of change in estimates from small changes in $d$ and $x$ .................................................... 70

3.1 Notation .......................................................................................................................... 75
3.2 Estimates from the maximum likelihood approach for simulated samples producing only finite values during the optimisation process ($p=0.3$) ......................................................... 88
3.3 Estimates from the maximum likelihood approach for simulated samples producing only finite values during the optimisation process and non-negative diagonal values in the inverted Hessian matrix ($p = 0.3$) ......................................................... 89
3.4 Parameter estimates for the three sites .................................................................................... 90

4.1 Notation .......................................................................................................................... 95
4.2 Mean squared errors of $\hat{N}$ and $\hat{N}_g$ .................................................................................. 106
4.3 Averages of 1000 estimates of the population parameters ......................................................... 107
4.4 Estimated standard errors ..................................................................................................... 108
4.5 Increased precision at higher $p$ ($p=0.5$) .................................................................................. 112
4.6 Parameter estimates for the three sites .................................................................................... 114

5.1 Notation .......................................................................................................................... 121
5.2 Averages of 1000 estimates of the population parameters ......................................................... 129
5.3 Estimated standard errors ..................................................................................................... 130
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4</td>
<td>Increased precision at higher $p$ ($p=0.5$)</td>
<td>132</td>
</tr>
<tr>
<td>5.5</td>
<td>Notation</td>
<td>134</td>
</tr>
<tr>
<td>5.6</td>
<td>Averages of 1000 estimates of the population parameters</td>
<td>140</td>
</tr>
<tr>
<td>5.7</td>
<td>Estimated standard errors</td>
<td>141</td>
</tr>
<tr>
<td>5.8</td>
<td>Increased precision at higher $p$ ($p=0.5$)</td>
<td>143</td>
</tr>
<tr>
<td>6.1</td>
<td>Notation</td>
<td>154</td>
</tr>
</tbody>
</table>
Estimating population demographics is of immense importance to many specialists such as ecologists and biologists. The population size plays a vital role among other population demographics. Observing the entire population is not feasible for most natural populations. This raises the necessity of estimating the population size. Precise estimation is still a difficult challenge, especially for endangered species where collecting data may be demanding and expensive. Furthermore, the disturbance to individuals or their natural habitats is not desirable, hence these interactions should be minimal.

Estimation of population size has a long history. Classical capture-recapture analysis is one of the widely used techniques from the early stages of population size estimation. Many modifications to the classical capture-recapture model were developed over time to suit different population settings. Concurrently, recent advances in the field of molecular genetics have raised the possibility of applying genetic markers to the estimation of the demographics of wildlife populations. Using genetic markers instead of physical tags was an initial step towards the use of genetic markers in estimating population demographics, which with time lead to many different techniques. Thus the related literature has been growing rapidly. However, estimating the population size using genetic data has sometimes led to estimates that differ markedly from each other and also from classical capture-recapture estimates. The resolution of these discrepancies requires more sophisticated models than those that are currently available.
Our work is based on the idea that the degree of biological relatedness between the individuals in the sample reflects the size of a population which is presumably closed. The higher the degree of relatedness between the individuals in the sample, the lower the population size, and vice versa. We use the genetic markers to determine the degree of biological relatedness between the individuals. This gives a ‘one occasion capture experiment’. The identified relationships allow us to treat some of the captures as recaptures, hence this relates to a capture-recapture experiment. Nevertheless, our method uses one-occasion capture data, maintaining the disturbance to the natural population at a minimal level. We primarily use the method of moments for estimation. We also consider maximum likelihood in this setting and discover problems that largely preclude its use. In this work we consider only the size of the female population. Our application is on mountain pygmy-possum (*Burramys parvus*) populations. The mountain pygmy-possum is an endangered marsupial restricted to the alpine regions of Victoria and New South Wales, Australia. Female pygmy possums inhabit the prime breeding locations where there is high food availability, while the males disperse after breeding. Therefore the population structure we consider here is relevant to populations of pygmy possums.

This chapter presents the literature related to our work. It begins with introducing the capture-recapture analysis in Section 1.1. In Section 1.2 we present the work on genetic capture-recapture analysis, and how those works differ from what we do. Section 1.3 describes work related to pygmy possums. Finally, an overview of the remaining chapters is given in Section 1.4.

### 1.1 Classical capture-recapture analysis

Capture-recapture analysis, also known as mark-recapture analysis, is a method commonly used in ecology to estimate the size of a population when it is not practical to count all of the individuals in the population. This method is further used for other objectives; for determining unbiased estimates of other important demographic parameters such as population density, survival, recruitment and migration rates, and for testing biological hypotheses in ecology such as temporal and age variation, effects of population density and climate effects.
In this method, a portion of the population is captured, marked and released. Later, another sample of individuals is captured and the number of marked individuals in the captured sample is counted. The basic concept of mark-recapture analysis is that the number of marked individuals within the second sample should be proportional to the number of marked individuals in the whole population. Therefore an estimate of the total population size can be obtained by dividing the number of marked individuals by the proportion of marked individuals in the second sample.

Pierre Simon Laplace (1749-1827) is known as the pioneer of capture-recapture analysis. In September 1802, he used a capture-recapture type of approach to estimate the size of the human population of France. Although Laplace is commonly thought of as the first to use capture-recapture idea, he was preceded, by almost 200 years, by John Graunt in his attempts to use similar methods to estimate the effect of plague and the size of populations in England in the early 1600s. Consequently, the theories and applications of capture-recapture have moved far beyond the concepts of John Graunt and Pierre Laplace. However, current methods also share the basic concept of ratios between captured sample and population, that guided those pioneers [1]. Different capture-recapture methods are used for different population settings and different sampling methods etc. Capture-recapture techniques may be applied for open or closed populations and require different methods.

1.1.1 Closed population

A closed population is one in which the total number of individuals does not change through births, deaths, immigration, or emigration during the period of study. Even though most natural populations do not meet the requirements of a closed population in the longer term, the methods developed for these populations are widely used in the short term when the closure assumption is reasonable. The Petersen method and the Schnabel method are two widely used methods for estimating the size of a closed population.
Lincoln–Petersen method

The Petersen method is the simplest of the mark and recapture methods. The procedure involves capturing a number of individuals, marking them, releasing them back to the population and then capturing another sample and counting the number of marked individuals in that sample. All these need to be done over a relatively short period of time. Thus, this method involves only a single marking step.

In this method the population size is estimated by

\[ \hat{N} = \frac{CM}{R} \]

Here,

- \( N \) - Population size
- \( M \) - The number of individuals marked in the first sample
- \( C \) - The total number of individuals captured in the second sample
- \( R \) - The number of individuals in the second sample that are marked

This formula is widely used because it is intuitively clear. But this formula can produce a biased population size estimate, especially when the sample size is small. Therefore a modified version with less bias was originally developed by Chapman [6] for fisheries data.

Chapman’s modification of the Petersen estimator:

\[ \hat{N} = \frac{(M + 1)(C + 1)}{R + 1} - 1 \]

Schnabel method

The Schnabel method is an extension of the Petersen method to a series of samples [30]. Individuals caught at each sample are first examined for marks, then marked and released. Only a single type of mark need to be used because we just need to distinguish two types of individuals: marked, caught in one or more prior samples;
and, unmarked, never caught before. Thus, this method involves multiple marking steps.

For each sample $t$, the following data are recorded:

- $C_t$ - Total number of individuals caught in sample $t$
- $R_t$ - Number of individuals already marked (recaptures) when caught in sample $t$
- $U_t$ - Number of individuals marked for the first time and released in sample $t$

Marks accumulate over time. Therefore,

- $M_t$ - Number of marked individuals in the population just before the $t^{th}$ sample is taken

or as a formula:

$$M_t = \sum_{i=1}^{t-1} U_i$$

where $t$ is the number of samples in the series.

Schnabel [29] proposed the following weighted average of Petersen estimates as an estimate of population density.

$$\hat{N} = \frac{\sum_{i=1}^{t} C_i M_i}{\sum_{i=1}^{t} R_i}$$

### 1.1.2 Open population

An open population is one that is changing during the course of the study, because of any combination of births, deaths, immigration, or emigration. Most natural populations are open populations. The Jolly-Seber method is one of the most famous methods used for estimating the size of an open population.

**Jolly-Seber method**

The Jolly-Seber method extends the mark and recapture method to open populations. The biggest change in the sampling procedures over the Schnabel method is
the inclusion of the information of when a marked individual was last captured. In this method, the time interval between samples need not be constant; any number of samples (at least three) can be accommodated so that a series of data extending over many years can be used in the method. An estimator for the size of the population at any given time is presented in Krebs [16] based on the Cormack-Jolly-Seber model using the following notation:

- $m_t$ - Number of marked individuals caught in sample $t$
- $u_t$ - Number of unmarked individuals caught in sample $t$
- $n_t$ - Total number of individuals caught in sample $t$, i.e. $n_t = (m_t + u_t)$
- $s_t$ - Total number of individuals released after sample $t$, i.e. $s_t = (n_t - \text{accidental deaths or removals})$
- $m_{rt}$ - Number of marked individuals caught in sample $t$ last caught in sample $r$
- $R_t$ - Number of the $s_t$ individuals released at sample $t$ and caught again in some later sample
- $Z_t$ - Number of individuals marked before sample $t$, not caught in sample $t$, but caught in some sample after sample $t$

The estimators proposed in Krebs [16] are,

\[
\hat{N}_t = \frac{\hat{M}_t}{\hat{\alpha}_t},
\]

with

\[
\hat{M}_t = \frac{(s_t + 1)Z_t}{R_t + 1} + m_t,
\]

and

\[
\hat{\alpha}_t = \frac{m_t + 1}{n_t + 1}.
\]

Here, $\hat{\alpha}_t$ is an estimator for the proportion of the population that is marked at time $t$, $\hat{M}_t$ is an estimator for the marked population just before sample time $t$ and $\hat{N}_t$ is an estimator for the population size at time $t$. 
Note that in these traditional methods for both closed and open populations identifying the recaptured individuals is required.

1.1.3 Computer programs for estimating the population size

Various computer programs have been developed for estimating the population the models mentioned above. CAPTURE and MARK are two popular programs used for this purpose [16].

Program CAPTURE

This program is used for estimating the size of closed populations. Both the Petersen method and the Schnabel method rely upon the assumption that all individuals in the population are equally likely to be captured in each sample. Violation of the assumption of equal capture probabilities in mark-recapture studies can lead to unreliable estimates of population size. Program CAPTURE allows unequal capture probabilities. It uses different estimators (models) to account for different sources of variation in capture probabilities:

- $M_0$ (Equal catchability model/Null model) - Assumes every individual in the population has the same capture probability for each sampling period in the study.
- $M_h$ (The heterogeneity model) - Assumes that each individual has a unique capture probability that remains constant over all trapping occasions.
- $M_b$ (The trap response model) - Adjusts for a change in capture probabilities caused by a response to trapping. An assumption of the $M_b$ model is that the initial capture probability for all individuals is the same (equal catchability).
- $M_{bh}$ (The heterogeneity and trap response model) - Based on the assumption that each individual has its own unique pair of potential capture probabilities, $p_j$ and $c_j$ ($j = 1, \ldots, N$ individuals in the population), where $p_j$ is the initial capture probability and $c_j$ is the recapture probability.
• $M_t$ (The time variation (Schnabel) model) - Based on the assumption that every individual in the population has the same capture probability for a given sampling occasion, but capture probabilities can vary at each sampling time.

• Other time-dependent models: $M_{th}$, $M_{tb}$ and $M_{tbh}$

Program $MARK$

Program $MARK$ is a Windows-based software application, developed by Gary White (Colorado State University) for the analysis of data from marked individuals. The primary motivation for developing this program was to bring a common interface to the problem of parameter estimation. Program $MARK$ is a very comprehensive and widely used software application currently available for the analysis of data from marked individuals (hence the name $MARK$). $MARK$ is a very flexible and powerful program, which has many options, and a high level technical and theoretical sophistication. Program $MARK$ encompasses virtually all currently used methods for the analysis of marked individuals for closed population models, open population models and mixture models.

Program $MARK$ provides parameter estimates from marked individuals when they are re-encountered at a later time. One major advantage of this program is that the time intervals between re-encounters do not have to be equal. Time intervals are assumed to be one time unit unless specified. Program $MARK$ computes the estimates of model parameters via numerical maximum likelihood techniques.

1.2 Genetic capture-recapture analysis

The estimation of the size of a population from genetic data has a long history. Genetic markers were used instead of physical tags as an initial enhancement of the traditional capture-recapture experiments. The work we discuss in this section uses the genetic markers in a more advanced form than solely being a replacement for physical tags. Genetic capture-recapture methods for estimating the population size often depend on temporal changes in allele frequencies. Apart from estimation of population size, genetic data has been widely used for paternity analysis and estimation of the number of successfully breeding males.
1.2.1 Paternity analysis

A method using genetic data was applied to a population of painted turtles on the Mississippi river by Pearse et al. [25] to estimate the number of successfully breeding males. Three datasets were used in this study utilising three different approaches. ‘Data set 1’ involved male turtles physically marked and recaptured in two sampling periods and then analysed using an approach analogous to a traditional mark-recapture method. ‘Data set 2’ involved data collected on 25 trapping days for an analysis based on a direct genetic method. In this method, DNA was collected from the individuals, and the multilocus genotypes of individuals were considered to be the ‘marks’. ‘Data set 3’ encompassed the parentage-analysis approach, wherein for four sampling periods the marks and recaptures were of males whose genotypes were deduced as the sires of assayed clutches. Estimates of population size based on data set 1 were calculated by hand using the Petersen method. The computer program CAPTURE, which assumes a closed population, was employed in the analysis of both the other data sets. In CAPTURE, the models $M_h$-jackknife and $M_h$-Chao were employed because both permit heterogeneity in trapping probabilities and are fairly robust to low recapture rates. For the genetic parentage data (data set 3), the program MARK was utilised to estimate population size under the Jolly-Seber open population model. This is appropriate given that these data were collected over a four year period. Even though the ‘Data set 3’ involves a parentage analysis approach, this is different from our work. Here, Pearse et al. [25] identify the sires of assayed clutches and considers the recaptures of the sires for estimation, whereas our work considers the parent-offspring pairs for estimation.

In the biological world there are many polygynous (or polyandrous) species in which groups of full siblings and one parent are easily captured, whereas the other parent is difficult or impossible to sample directly. The above genetic parentage-based approach is especially well suited to such situations. Pearse et al. [25] show that it is also possible to use the parentage data as both marks and recaptures, provided that the correct time-series of samples is obtained. While they do not specify what the correct time-series of samples is, it depends on the species. For example, Pearse et al. [25] use data from four sampling periods during 1995-1998 for their study. They consider only one clutch from each female per year since female painted
turtles typically store and utilise sperm between clutches within a year, and, hence, their progeny would not represent independent samples of sire genotypes over this time frame. The resulting data may be appropriate for multiple sample capture-recapture estimators, such as the Schnabel or Jolly-Seber methods for closed or open populations, respectively. This approach has the advantage that there is never a need to trap individuals of one sex (in this case, males), because both the capture and recapture data on sires (in this case) are reconstructed from the other parent and the known progeny arrays. The only limitation is that enough young (generally at least five) must be available from each clutch to allow reliable reconstruction of the genotypes of the non-sampled parents. Overall, higher estimates of population size (and wider 95% confidence limits) have emerged from the genetic parentage-based approach than those emerged from traditional mark-recapture approach and the direct genetic approach mentioned above. This is somewhat surprising given that the physical samples included some sub-adult males, whereas the genetic samples could only include reproductively mature individuals. Pearse et al. [25] explained that in acquiring mates, female painted turtles may have travelled beyond the local population or otherwise mixed with resident males of neighbouring breeding areas. It should also be noted that in the genetic parentage-based approach, the genetic data is used to estimate the effective number of breeding males, and this value will be influenced by factors such as the variance in male reproductive success or the migration of mate-searching females to neighbouring populations. Thus, this study demonstrates the potential of this approach to provide additional biological information about natural populations beyond the estimation of population size, although it has not yet been explored in detail.

Nielsen et al. [22] have presented an approach to paternity analysis in natural populations based on a Bayesian methodology with an application to the North Atlantic humpback whales. The method made inferences directly on the basis of the calculated parentage probabilities, which was the posterior probability that a particular putative father is the actual father. For estimating the posterior probability that a particular individual might be the father of a known offspring, they assume that the prior probability that a particular male is the father is 1/N, where N is the number of potentially breeding males in the breeding area. Since this method required information regarding the number of potentially breeding males in the population, they
have separately treated cases where such information is not available by suggesting an explicit decision-theoretic approach, i.e. they defined a specific loss function, which provides the “loss” incurred if a wrong classification is made. By minimising the expectation of the loss (the risk), they establish an appropriate decision rule that determines the classification of parent-offspring relationships. They have assessed the performance through determining the necessary number of loci by computer simulations. Multiple data sets were generated, for which the population allele frequencies were estimated from the observed allele frequencies. Then paternity analysis has been performed on them which clearly results that the variability of the locus is a major determining factor of the performance of the method. Their approach takes the incomplete sampling of putative fathers into account. Through computer simulations, Nielsen et al. [22] have assessed how important the problem of ignoring family structure and variation in fertility versus the importance of ignoring incomplete sampling is, and determined that the most critical model assumptions are the assumptions regarding complete sampling and the number of breeding males. Subsequently, they developed a method for testing hypotheses regarding the reproductive success of different ecologically or behaviourally defined groups using likelihood function for the ratio of the reproductive success of the two groups, assuming that the probabilities of an individual male siring two offspring are independent. The critical value was obtained using the chi-square distribution since they have determined that the chi-square approximation works well, even for the limited sample size, using simulations. Further, they used maximum likelihood estimation to estimate the number of breeding males.

Garrigue et al. [13] have presented capture-recapture estimates of the abundance of humpback whales on the wintering grounds of New Caledonia using photo identification, microsatellite genotyping and molecular identification of sex. They have used the genotypes of individual males in the population to identify the likely paternity of calves. Their purpose was to evaluate the reproductive autonomy of this wintering ground by comparing the gametic recapture estimate to the organismal recapture estimate, i.e. whether the proportion of fathers captured by the paternity analysis is consistent with the estimated size of the local male population. The probability of identity, which represents the average probability that two unrelated individuals share the same genotype by chance alone, has been calculated in accordance with
the methodology used by Paetkau and Strobeck [24]. Paternity analysis was conducted with a computer program, CERVUS 2.0, using the frequency of alleles at each locus to calculate the probability of non-exclusion for each male relative to each cow-calf pair (i.e. the probability of a non-father matching by chance) and the likelihood of paternity for a particular male, relative to the likelihood of paternity for a randomly selected male in the population [19]. This likelihood was expressed as the natural log of the likelihood ratios (referred to as the LOD score) summed across all (presumably unlinked) loci. For each cow-calf pair, CERVUS ranks all candidate males according to their LOD score, allowing for correction of a specified level of genotyping error (default value of 0.01; 19, 35). CERVUS calculated the difference between the LOD score of the most likely male and the second most likely male in the sample to judge the statistical confidence of inferred paternities. The critical value of this difference was generated by simulations of parental and offspring genotypes based on the observed allele frequencies, a specified genotype error-rate parameter, and the estimated size of the candidate male population, including unsampled males. Organismal population abundance was estimated using 2 models: Petersen’s weighted mean [3] based on comparisons between each year and the sum of all previous years; and Chapman’s modification of the Petersen 2-sample model [31] based on the comparison of the 2001 samples to all previous years. An alternative estimate of male abundance was derived by ‘gametic’ capture-recapture analysis using Chapman’s modification to the Petersen 2-sample model based on paternity assignment [25].

1.2.2 Genetic distribution

Census population size ($N_c$) is the actual number of living individuals in a population. This was traditionally used in ecology for conservation and management purposes over a long time. However, the effective size of a population ($N_e$) has now become one of most fundamental parameters in evolutionary and conservation biology. The effective size of a population is the number of breeding individuals in an idealised population that would show the same amount of dispersion of allele frequencies under random genetic drift or the same amount of inbreeding as the population under consideration. It determines the rate of change in the composition
of a population caused by genetic drift, which is the random sampling of genetic variants in a finite population. $N_e$ is crucial in determining the level of variability in a population, and the effectiveness of selection relative to drift [7].

A review of the estimators of DNA-based census ($N_c$) and effective ($N_e$) population sizes has been performed by Schwartz et al. [30]. Two advances in molecular genetics, 1. development of highly polymorphic DNA markers, and 2. the ability to amplify these markers with the polymerase chain reaction (PCR) from low-quality, low-quantity DNA samples, hold great promise for the application of genetic markers to the estimation of $N_e$ or $N_c$ of wildlife populations. Schwartz et al. [30] discussed three methods for estimating effective population size: 1. the gametic disequilibrium method, 2. the heterozygote excess method, 3. the temporal allele method. Each of the estimators of the effective population size ($N_e$) was based on the assumption that genetic drift increases when $N_e$ decreases.

In the first method, i.e. gametic disequilibrium ($D$) method, also known as linkage disequilibrium method, Schwartz et al. [30] assume that observed $D$ was produced by drift in a small population among unlinked loci and focus on how this relationship can be exploited to estimate $N_e$. One major advantage of this method was that it can be used to estimate $N_e$ with a single sample. But this method was problematic for endangered species with extremely small population sizes, and species dispersed over large geographic areas or rugged terrain that require extensive sampling efforts.

The second method (heterozygote excess method) also required only one sample to estimate $N_e$. This method was based on the observation that small populations of dioecious species differ in genotype frequencies between sexes and deviate from overall Hardy-Weinberg (HW) proportions [11]. Unfortunately, this method may only be valid when a species has a breeding system where there is a random union of gametes such as a broadcast spawner [26].

The temporal method, which required data from two samples, was based on changes in allele frequencies that occur between two samples taken from a small population. This method requires samples of approximately 30 individuals in successive generations and more than five polymorphic microsatellite loci to achieve precise estimates of $N_e$, when $N_e$ is very small (i.e. four). The method requires 45 individuals and 10 loci when $N_e$ is 10-20 [17]. Further, this method produces biased estimates of $N_e$.
when \( N_e \) is small [28]. However this bias was negligible unless \( N_e \) was less than approximately 6-10 or time in generations between samples is large (> 5; 17). One of the major weaknesses of this method is that it requires two or more samples, which can be expensive and time consuming. Yet, this is the most promising method for estimating \( N_e \) as it provides increasingly precise estimates as the number of alleles per locus increases. However, it is difficult to compare the relative usefulness of the three methods because their performance has never been compared under the same conditions. Further, Schwartz et al. [30] reviewed two methods for estimating the census population size, the first using one sample and the second using two or more samples. The authors reveal an important weakness of the methods using genetic samples; that is while they may be non-invasive, they will not provide potentially important demographic information, such as age structure, individual health and sexual condition, that can be obtained via conventional survey methods. Nevertheless, the use of genetic data may be the only way to obtain sufficient demographic information for some species, and may be the least risky way to acquire these data. Additionally, mark loss and estimator bias associated with behavioural responses to ‘trapping’ may be minimised with genetic data.

1.2.3 Allele sharing methods

Allele sharing methods were discussed by Skaug [33]. This work was based on the idea that the degree of biological relationship between individuals in the sample reflects the size of the population and that DNA profiles provide information about relatedness as biologically related individuals share more alleles than biologically unrelated individuals. The work by Schwartz et al. [30] is suitable only for small populations since it is based on genetic drift. But the work [33] does not have that limitation. Further, unlike the traditional two-sample mark-recapture method, this one sample method is applicable to situations where the sampling method is lethal. This method depends upon three major assumptions of random sampling, Hardy-Weinberg and linkage equilibriums and random mating. The approach of Skaug [33] consists of two parts. In the first step, the probability of a given combination of haplotypes in two individuals was computed as a function of the relationship between
individuals measured through the kinship coefficient. In the second step, the probability of the particular relationship was computed. The combination of the two steps allowed the construction of a pseudo-likelihood function. The dependency between the DNA profiles of the individuals cannot be ignored completely. Therefore, the joint distribution of \( D_1, \ldots, D_n \), where \( D_i \) is the DNA profile of the \( i \)\textsuperscript{th} individual, is likely to be complicated. Hence a full likelihood approach for estimation of \( N \) does not seem feasible. Therefore, as a compromise, a pseudolikelihood approach was taken, involving only the distribution of pairs \((D_i, D_j)\) of DNA profiles. The estimator of the population size is obtained by maximising the pseudo-log-likelihood function through an iterative method. This method requires a large sample to be useful in estimation.

The determination of kinship from haplotypes has also been extensively studied [34]. One major problem associated with the search for pairs (dyads) of related individuals in large databases of DNA-profiles is that many, partly dependent, pairwise comparisons introduce statistical issues. Sequential Bonferroni correction, probably the most widely known procedure used to correct for such multiple comparisons, guards against even a single false positive, thus causes an increase in the number of false negatives, i.e. true dyads in the sample that are undetected. This study is focused on the situation where only a small fraction of the population has been sampled leading to a low probability of sampling both members of dyads of related individuals, therefore, the chance that a true dyad goes undetected is quite high. In this approach, the suitability of the False Discovery Rate (FDR) procedure to control the proportion of false positives in situations consisting of a large number of pairwise comparisons is discussed. The likelihood ratio (LOD score) considers the hypothesis of unrelatedness versus an alternative, such as that of half siblings, and is calculated for all pairwise comparisons. The corresponding p-values are calculated using a permutation approach for testing above hypotheses. The p-values are arranged in the increasing order \( P_1 \leq P_2 \leq \ldots \leq P_m \) and the first ‘\( R \)’ pairs are identified as related individuals. This threshold ‘\( R \)’ is defined as the largest value of \( r \) for which \( P_r \leq r/m \cdot q \) (where \( q \) is the target FDR). The FDR procedure is applied to the data set in two stages. In the first stage, related dyads are identified using the LOD scores of all pairs and a limited number of loci at a higher FDR with simulation-based p-values. In the second stage, this identification is done using
the LOD scores of the pairs which were determined to be related in the first stage by considering some extra loci at a relatively low FDR. This approach has many advantages such as the choice of the FDR level which is very much dependent upon the objective of identification of dyads of related individuals, and the loci considered in the second stage can be determined after the first stage has been completed, so that the number of individuals to be genotyped is known, which is cost effective.

1.3 Related work on mountain pygmy possums

In this section we describe work related to a species that motivated our work. The mountain pygmy possum (*Burramys parvus*) is an endangered marsupial restricted to the alpine region of Victoria and New South Wales, Australia. The level of genetic variation in endangered mammals is often low, thus there exists genetic divergence among populations.

Mitrovski et al. [20] have studied the population structure of *B. parvus* in detail examining microsatellite markers and scoring these on more than 700 individuals obtained throughout the range of *B. parvus*. Their study had several main objectives: 1. examine genetic diversity in *B. parvus* since low genetic diversity may be expected in *B. parvus* as they were thought to have been separated into three small, isolated populations since the end of the last Pleistocene glacial period [18]; 2. determine if there was genetic structure within the three separate regions occupied by *B. parvus*; 3. determine whether fragmentation had led to a loss of genetic diversity and increased homozygosity in any populations. DNA were extracted at eight microsatellite loci from hair samples of individual *B. parvus* collected from 13 populations across its entire range.

In this work Mitrovski et al. [20], estimated the effective population size (*N*<sub>e</sub>). They used three methods to calculate the effective population size (*N*<sub>e</sub>) for all populations sampled in the study. First, they estimated *N*<sub>e</sub> directly using the formula

\[ N_e^{(dir)} = \frac{4N_{males}N_{females}}{N_{males} + N_{females}} \]  

[5]. They also used two indirect methods to calculate *N*<sub>e</sub> based on average microsatellite heterozygosity and mutation rates of microsatellite loci. The first method calculates the effective population size

\[ N_e^{(IAM)} = \frac{H_E}{4\nu(1 - H_E)} \]

and follows the assumptions of the infinite alleles
Introduction

model (IAM) of microsatellite evolution [15]. The second method is based on the stepwise mutation model (SMM) [23] and calculates the effective population size as 

\[ N_e(SMM) = \frac{(1/(1 - H_E)^2) - 1}{8\nu} \]

For each method, \( H_E \) is the average expected heterozygosity per population and \( \nu \) is the mutation rate of microsatellite loci (estimates of \( 10^{-3} \) were used here; 9, 36). These three methods resulted in vastly different estimates of the effective population size for all populations of \( B. parvus \) (Table 1.1). The adult population count for each population in this table has been estimated from trapping data collected from 1999 to 2003 using the intensive trapping scheme described in Mansergh [18] and averaged over this period.

Table 1.1: Estimates of size of different populations presented in Mitrovski et al. [20]

<table>
<thead>
<tr>
<th>Region</th>
<th>Population</th>
<th>( N )</th>
<th>( N_e(dir) )</th>
<th>( N_e(IAM) )</th>
<th>( N_e(SMM) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern</td>
<td>Mount Buller</td>
<td>150</td>
<td>54</td>
<td>107.7</td>
<td>130.8</td>
</tr>
<tr>
<td></td>
<td>Mount Bogong</td>
<td>100</td>
<td>75</td>
<td>384.5</td>
<td>680.2</td>
</tr>
<tr>
<td></td>
<td>Timm Spur</td>
<td>120</td>
<td>90</td>
<td>476.7</td>
<td>931.3</td>
</tr>
<tr>
<td></td>
<td>Falls Creek</td>
<td>30</td>
<td>22.5</td>
<td>505.3</td>
<td>1015.9</td>
</tr>
<tr>
<td></td>
<td>Pretty Valley West</td>
<td>50</td>
<td>37.5</td>
<td>543.7</td>
<td>1134.8</td>
</tr>
<tr>
<td></td>
<td>Bundara</td>
<td>120</td>
<td>90</td>
<td>446.4</td>
<td>844.9</td>
</tr>
<tr>
<td></td>
<td>Mount Loch</td>
<td>400</td>
<td>300</td>
<td>505.3</td>
<td>1015.9</td>
</tr>
<tr>
<td></td>
<td>Mount Higginbotham L</td>
<td>50</td>
<td>37.5</td>
<td>489.6</td>
<td>969.1</td>
</tr>
<tr>
<td></td>
<td>Mount Higginbotham U</td>
<td>50</td>
<td>37.5</td>
<td>551.3</td>
<td>1159.1</td>
</tr>
<tr>
<td></td>
<td>Mount Higginbotham W</td>
<td>250</td>
<td>187.5</td>
<td>516.9</td>
<td>1051.2</td>
</tr>
<tr>
<td>Northern</td>
<td>Summit Road</td>
<td>25</td>
<td>24</td>
<td>387.8</td>
<td>688.5</td>
</tr>
<tr>
<td></td>
<td>Charlottes Pass</td>
<td>45</td>
<td>43</td>
<td>413.1</td>
<td>754.5</td>
</tr>
<tr>
<td></td>
<td>Paralyser</td>
<td>22</td>
<td>21</td>
<td>416.7</td>
<td>763.9</td>
</tr>
</tbody>
</table>

Here \( N \) is the total adult population size and \( N_e(dir) \) are direct estimates. Indirect estimates of \( N_e \) were calculated under the assumptions of an infinite allele model (IAM) and stepwise mutation model (SMM), using mutation rates of \( 10^{-3} \).

1.3.1 Simulation studies

We conducted simulation studies to assess the performance of the three estimators considered by Mitrovski et al. [20] for a population of pygmy possums (*B. parvus*). To do this, we generated a population genetic model possessing the following properties, considering DNA profiles with 20 genetic markers.

- The population evolves in discrete time steps.
- The individuals born at a particular time constitute a generation.
- Only individuals belonging to two generations live at a particular time point.
- An individual has equal probability of being a male or a female.
- In generating the offspring population, a male and a female is selected completely randomly and considered as a parent pair of an offspring. The DNA profile of the offspring is simulated using the DNA profiles of the randomly picked mother and father.

Initially a population of 500 males and 500 females was simulated independently from each other such that the population is in Hardy-Weinberg equilibrium and the loci of an individual is in linkage equilibrium. The second generation was simulated assuming that the number of offspring of given parents are independent and have a Poisson(2.5) distribution. It was considered that the probability that each offspring receives an allele from his mother and father at each locus is exactly the same. The first population was totally terminated when the third offspring generation was simulated, to allow only two generations exist at a particular time point. Also, 20% of the second generation was removed randomly to avoid the population size growing exponentially. Generations for five time steps were simulated correspondingly.

The population was sampled so that each individual in the population has a capture probability of 0.3 in the fifth time step, yielding individuals belonging to fifth and sixth generations. The proposed three estimators were calculated for each sample. In calculating the estimates, \( \nu \) (mutation rate of microsatellite loci) was taken as
The average expected heterozygosity ($H_E$) per population was calculated by,

$$H_E = \frac{1}{m} \sum_{i=1}^{m} \frac{2n (1 - \sum_{i=1}^{k} x_i^2)}{2n - 1},$$

where $x_i$ is the sample allele frequency of the $i^{th}$ of $k$ alleles and the $m$ is the number of loci [21].

One thousand simulations were performed to assess the estimators. The simulation study produced estimates which were vastly different from each other, and also, from the true population size of 2100. Particularly, the estimates produced by the first two methods were much lower than the true population size, whereas the third method resulted in a much larger value. (Table 1.2)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_e(dir) = 4N_{males}N_{females}/(N_{males} + N_{females})$</td>
<td>629.73</td>
<td>21.32</td>
</tr>
<tr>
<td>$N_e(IAM) = H_E/4\nu(1 - H_E)$</td>
<td>372.40</td>
<td>6.32</td>
</tr>
<tr>
<td>$N_e(SMM) = [(1/(1 - H_E)^2) - 1]/8\nu$</td>
<td>3146.78</td>
<td>100.51</td>
</tr>
</tbody>
</table>

1.4 Thesis overview

This thesis consists of an Introduction, five research chapters and a conclusion, which also offers suggestions for future work. Initially, our models in Chapters 2 to 5 were constructed supposing that the biological relationships are perfectly identifiable.

In Chapter 2 we develop estimators for the number of mothers in a population using mother-daughter pairs in the sample. As an initial approach, we restrict the population structure to a population with two generations. In this chapter we use a method of moments to develop our estimators of the initial number of mothers, the mean number of daughters per mother and the capture probability. The approach
we adapt here using estimating equations is simple, resulting in explicit estimators for our population parameters.

Chapter 3 considers the same population structure as in Chapter 2. In this chapter we examine maximum likelihood estimation for estimating the number of mothers instead of method of moments, and discover problems that preclude its use in practice. Here also we use only mother-daughter pairs.

In Chapter 4 we extend the method of moment estimators suggested in Chapter 2 to a population consisting of three generations. We introduce two different estimators for the size of the first generation in the population, which we refer to as mother generation, first, using only the mother-daughter pairs among the three generations, and second, using the mother-daughter-granddaughter triple along with the mother-daughter pairs. The two models are then compared. We propose a generalisation of the estimator for the number of females in the first generation of any population with any known number of generations.

We incorporate sister pairs into the estimators in Chapter 5. This allows us to increase the precision of the estimators proposed in Chapters 2 and 4. The estimators are constructed for populations consisting of two and three generations, and propose a generalisation for a population with any known number of generations.

In Chapters 2 to 5, the models are constructed supposing that the biological relationships are perfectly identifiable. However, the perfect identification of these relationships is not possible for most natural populations. In our work we adopt the approach using genotype data to estimate the number of parent-offspring pairs as in Skaug et al. [34]. There may be a possibility of misclassification of relationships when they are being estimated. In Chapter 6 we study the impact of this misclassification of relationships on the estimator proposed in Chapter 2.

A discussion for each of these chapters is given in the respective chapters, followed by an overall conclusion along with suggestions for future work in Chapter 6.
Chapter 2

Estimating the number of mothers in a population with a simple structure

2.1 Background

In Chapter 2 and 3 we use the mother-daughter pairs detected in the sample to develop estimators for the number of mothers in a population. Initially, our models in these two chapters are constructed supposing mother-daughter pairs are perfectly identifiable. Note that we consider only females in the population in our work.

Female pygmy possums inhabit the prime breeding locations where there is high food availability while the males disperse after breeding. Therefore, the population structure we consider here is relevant to populations of pygmy possums. While the actual population will be a mix of mothers from different generations, as a starting point, we show that statistically consistent estimators of population size may be obtained in a simpler population structure. The population structure we consider treats the individuals from the first generation as the mothers, and the individuals from the second generation as the daughters. In other words, each female in the population is a female from the mother generation or the daughter generation, where the
two generations are mutually exclusive and collectively exhaustive. Subsequently, no female can be both a mother and a daughter.

We suppose that the observed population consists of an initial number of mothers and their daughters. In observing a sample from the population we suppose that mother-daughter pairs can be identified. Note that this setting is appropriate for a population of pygmy possums, as described above.

Here we demonstrate that it is possible to estimate the size of a closed population using genotype data from a single capture occasion. To understand our approach, recall that the simplest capture-recapture experiments are the Lincoln-Petersen capture-recapture experiments with two capture occasions (Section 1.1.1). On the first occasion, individuals are captured, marked and released. On the second occasion it is noted which of the captured individuals had been marked on the first occasion. Population size is then estimated based on the concept that the number of marked individuals in the second sample should be proportional to the number of marked individuals in the whole population. Intuitively, the use of genetic markers would allow us to treat the capture of an individual as a capture and the offspring of captured individuals as recaptures, allowing us to estimate the population size from one capture occasion. Here, a difference from the classical estimates is that the recapture probability depends on the number of daughters of an individual.

To cast this problem in the Lincoln–Petersen setting, suppose that all family groups consist of a mother and $K$ daughters. If the probability of capture is $p$, then the probability at least one daughter in a family group is captured is $(1 - (1 - p)^K)$ and is the same for all family groups. Suppose that the family groups and mothers and daughters can be distinguished even if only one individual in a family group is captured. In the notation of Lincoln–Petersen method in Section 1.1.1, $M$ is the number of family groups where the mother was captured, $C$ is the number of family groups with at least one individual captured, $R$ is the number of family groups with the mother and at least one daughter captured. Then the Lincoln–Petersen estimator of $\hat{N} = CM/R$ may be used. There are several problems with implementing this. Firstly, if only one individual from a family is captured we usually cannot determine if the individual is a mother or a daughter. Secondly, we would also need to identify sisters from family groups where the mother was not captured. We discuss the
difficulties in identifying sister pairs in Section 2.7. In general, of course, the sizes of the family groups vary so the recapture probabilities are not homogeneous and we would need to apply heterogeneous capture approaches.

Here we show that for a female population consisting of two generations, which we label as mothers and daughters, it is indeed possible to estimate the population size. The estimation procedure introduces the mean number of daughters per mother as a nuisance parameter, along with the capture probability.

In this chapter, we use a method of moments approach to develop our estimators of the initial number of mothers, which we refer to as the population size from here onwards, the mean number of daughters per mother and the capture probability, i.e. the probability a given individual is included in the sample. The approach we adapt here, using estimating equations, is simple, resulting in explicit estimators for our population parameters.

### 2.2 Notation and assumptions

We construct estimators for the size of a closed population using a single sample from a two generation population consisting only of mothers and their daughters where we can identify mother-daughter pairs. There are \( N \) mothers in the population and \( D \) daughters. \( \mu \) is the mean number of daughters per mother. Each individual in the population has the same probability \( p \) of being captured. We treat \( N, \mu \) and \( p \) as parameters, where \( N \) is our parameter of interest while \( \mu \) and \( p \) are nuisance parameters, that arise during the estimation procedure. A single sample is observed as a result of a capture experiment. The data consist of \( n \) individuals captured in total; \( x \) observed mothers, each of which has at least one captured daughter; and \( d_1, \ldots, d_x \), which are the numbers of caught daughters of the \( x \) observed mothers. Thus, a total of \( d \) daughters is observed, where \( d = \sum_{i=1}^{x} d_i \). Note that at least one of the daughters of a mother needs to be captured in order to identify that mother and, hence, be treated as an observed mother. A daughter also becomes an observed daughter only if her mother is observed. Our notation is summarised in Table 2.1.

We assume that:
Table 2.1: Notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Number of captured individuals</td>
</tr>
<tr>
<td>$x$</td>
<td>Number of observed mothers with at least one daughter observed</td>
</tr>
<tr>
<td>$d$</td>
<td>Number of observed mother-daughter pairs (i.e. Number of observed daughters)</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of mothers in the population</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Mean number of daughters per mother</td>
</tr>
<tr>
<td>$p$</td>
<td>Capture probability</td>
</tr>
<tr>
<td>$D$</td>
<td>Number of daughters in the population</td>
</tr>
</tbody>
</table>

1. The number of daughters of a given mother has a Poisson($\mu$) distribution where $\mu \geq 1$ and the number of daughters is independent for different mothers.

2. Mother-daughter pairs can be reliably determined.

3. Each individual in the population has the same probability $p$ of being captured.

4. Individuals are captured independently of each other.

5. The population considered is presumably closed.

The first assumption on $\mu$ seems natural. However, there may be populations where the maternal generation is bigger in size than the offspring generation. Nevertheless, we introduce this assumption mostly for technical reasons. We discuss the assumption of $\mu \geq 1$ extensively in Section 2.9. The second assumption requires observations on sufficient genotypes for identification to be reliable. Note that the latter two assumptions are those required for the simple homogeneity model for a capture-recapture experiment. Finally, the population we consider requires to be presumably closed to use our method.
2.3 Estimators

Our primary objective is to estimate the size of the mother population, \( N \). We also estimate the two nuisance parameters \( \mu \) (mean number of daughters per mother) and \( p \) (capture probability). We construct the estimators in two steps. Firstly, we develop the estimator for \( N \), assuming \( \mu \) is known. Secondly, we develop an estimator for \( \mu \), which is a nuisance parameter. The parameter \( p \) is also estimated in this second step. In Section 2.3.1, we derive the moment equalities required to derive the aforementioned estimators.

2.3.1 Moment equalities

For random variables \( X \) and \( Y \), let \( X|Y \) denote the conditional distribution of \( X \) given \( Y \). To derive our estimators we require the following moments:

**Proposition 2.1.**

\[
\begin{align*}
E(D) &= N\mu \\
E(n) &= N(1 + \mu)p \\
E(d|x) &= \frac{x\mu p}{1 - e^{-\mu p}} \\
E(x) &= Np(1 - e^{-\mu p}) \\
E(d) &= N\mu p^2 \\
E(x|n) &= \frac{n}{1 + \mu}(1 - e^{-\mu p})
\end{align*}
\]

**Proof.**

**Proof of (2.1)**

Let \( D_i \) be the number of daughters of mother \( i \), \( i = 1, \ldots, N \). Assumption 1 yields that \( D_i \sim \text{Poisson}(\mu) \) and hence letting \( D = \sum_{i=1}^{N} D_i \), \( D \sim \text{Poisson}(N\mu) \), giving (2.1).

**Proof of (2.2)**
To obtain (2.2) note that $n|D \sim \text{bin}(N + D, p)$ and hence $E(n) = E(E(n|D)) = E((N + D)p)$. (2.2) follows from this and (2.1).

**Proof of (2.3)**

Let $d_i$ be the number of observed daughters of a particular mother. Then for $i = 1, 2, ..., N$, $d_i \sim \text{Poisson}(\mu p)$, and hence the number of observed daughters of $i^{th}$ mother given at least one daughter is observed follows a zero truncated Poisson (ZTP) distribution, i.e. $d_i|d_i > 0 \sim \text{ZTP}(\mu p)$. Given that there are $x$ mothers from whom at least one daughter is observed in the sample, the total number of observed daughters with their mothers being observed will be the sum of $d_i$ over these mothers: $d = \sum_{i=1}^{x} d_i$, where we note we must have $d_i > 0$ for a mother to be identified. Thus, the distribution of $d$ given $x$ is that of a sum of independent zero truncated Poisson random variables, giving (2.3).

**Proof of (2.4)**

For a mother to be identified she must be captured along with at least one of her daughters. Recall that the number of observed daughters of a particular mother follows a Poisson($\mu p$) distribution. Hence the probability a mother from $N$ mothers is captured along with at least one of her daughters is $p(1 - e^{-\mu p})$. Thus $x \sim \text{bin}(N, p(1 - e^{-\mu p}))$ giving (2.4)

**Proof of (2.5)**

It is known that, $E(d) = E(E(d|x))$. This gives (2.5).

**Proof of (2.6)**

For this, consider three events; $I$, the event an individual is identified as a mother; $M$, the event an individual is a mother; $C$, the event an individual is captured. Then the probability an individual is a mother and identified given that the individual is captured is

$$Pr(I \cap M|C) = \frac{Pr(I \cap M \cap C)}{Pr(C)} = \frac{Pr(I|M \cap C)Pr(M \cap C)}{Pr(C)}$$
But the events that an individual is a mother and an individual is captured, are independent and hence $\Pr(M \cap C) = \Pr(M)\Pr(C)$. Now,

$$\Pr(I \cap M|C) = \Pr(I|M \cap C)\Pr(M)$$

Similarly

$$\Pr(I \cap M|C, D) = \Pr(I|M \cap C, D)\Pr(M|D)$$

Now, the expected value of $x$ identified mothers, given that $n$ individuals are captured and $D$ number of daughters are in the population is

$$E(x|n, D) = \sum_{i=1}^{n} \Pr(i^{th} \text{ individual is a captured mother}|i^{th} \text{ individual is captured})$$

$$= \sum_{i=1}^{n} \Pr(I|M \cap C, D)\Pr(M|D)$$

The probability that a captured mother is identified, $\Pr(I|M \cap C)$ is the same as the probability that at least one of her daughters is captured. This probability is $(1 - e^{-\mu p})$ since $d_i$ follows a Poisson$(\mu p)$ distribution. Then, $\Pr(I|M \cap C, D)$ is the same. Also, $\Pr(M|D) = N/(N + D)$. Now,

$$E(x|n, D) = n(1 - e^{-\mu p})\frac{N}{N + D}$$

Further, $E(x|n) = E(E(x|n, D))$ and hence,

$$E(x|n) = E \left( n(1 - e^{-\mu p})\frac{N}{N + D} \right)$$

$$= n(1 - e^{-\mu p})E \left( \frac{N}{N + D} \right).$$

Substituting (2.1) here gives (2.6).
2.3.2 Mean number of daughters per mother is known

The mean number of daughters per mother is usually unknown for natural populations. In the first step we assume this is known, and construct an estimator \( \tilde{N}(\mu) \) for \( N \). Recall that we assume that \( d \) and \( x \) are perfectly identified.

**Theorem 2.2.** Under the assumptions stated in Section 2.2, the number of mothers in the population \( (N) \), whose mean number of daughters \( (\mu) \) is known can be estimated by,

\[
\tilde{N}(\mu) = \frac{\mu n(n-1)}{d(1+\mu)^2}.
\]

(2.7)

Here it is assumed \( d \) and \( x \) is perfectly identified in the sample of size \( n \).

**Proof.** The total number of possible pairs in the population is \( \binom{N+D}{2} \). The number of mother-daughter pairs is equal to \( D \), the number of daughters in the population. Therefore, the proportion of all possible pairs in the population that are mother-daughter pairs is,

\[
\frac{D}{\binom{N+D}{2}} = \frac{2D}{(N+D)(N+D-1)}.
\]

(2.8)

Now,

\[
\frac{2D}{(N+D)(N+D-1)} - \frac{2\mu}{N(1+\mu)^2} = \frac{2ND(1+\mu)^2 - 2\mu(N+D)(N+D-1)}{(N+D)(N+D-1)N(1+\mu)^2} \approx \frac{2ND(1+\mu)^2 - 2\mu(N+D)^2}{(N+D)^2N(1+\mu)^2} \text{; for large } N
\]

With the law of large numbers and from (2.1), \( D \approx N\mu \). Therefore,

\[
\frac{2D}{(N+D)(N+D-1)} - \frac{2\mu}{N(1+\mu)^2} \xrightarrow{p} \frac{2N^2\mu(1+\mu)^2 - 2N^2\mu(1+\mu)^2}{N^3(1+\mu)^3} = 0.
\]

Next, the proportion of all possible pairs in the sample that are mother-daughter pairs is,

\[
\frac{d}{\binom{n}{2}} = \frac{2d}{n(n-1)}.
\]

Then by setting,

\[
\frac{2\mu}{N(1+\mu)^2} = \frac{2d}{n(n-1)}
\]

we obtain the estimator (2.7).
Thus, if the mean number of daughters per mother is known and the number of mother-daughter pairs in the sample can be identified, an estimate for the number of mothers in the population can be determined.

### 2.3.3 Mean number of daughters per mother is unknown

The estimator in (2.7) requires knowledge of the mean number of daughters per mother to estimate the number of mothers in the population. In practice, this quantity is not known and, therefore, needs to be estimated.

In developing our estimator we first treat $\mu p$ as a single parameter and estimate it using the method of moments. For this we use the expected value of the mean number of observed daughters for each identified mother. The estimating equation is solved using the principal branch of Lambert’s $W$ function [8]. In the second step $\mu$ is estimated considering the expected value of $n/x$. Subsequently, in the third step we derive $\hat{N}$ following $\tilde{N}(\mu)$ given by (2.7).

Note that the Lambert $W$ function is defined as the multivalued inverse of a function $W(z)$, where $W(z)$ is defined as the function satisfying $z = W(z)e^{W(z)}$. If $x$ is real, then the branch satisfying $W(x) \geq -1$ is referred to as the principal branch of the $W$ function and is denoted by $W_0(x)$, or just $W(x)$, and has a single real solution. This function is a continuous function which is increasing and differentiable for all $x > -1/e$. For further details on the Lambert’s $W$ function, see [8].

**Theorem 2.3.** Under the assumptions stated in Section 2.2, the number of mothers in the population ($N$), whose mean number of daughters ($\mu$) is unknown can be estimated by,

$$
\hat{N} = \frac{\hat{\mu}n(n-1)}{d(1+\hat{\mu})^2},
$$

(2.9)

Here $\hat{\mu}$ is,

$$
\hat{\mu} = \begin{cases} 
\frac{n(1-e^{-\tilde{\mu}p})}{x} - 1, & \text{if } n(1-e^{-\tilde{\mu}p}) > 2x, \\
1, & \text{otherwise},
\end{cases}
$$

(2.10)

where

$$
\tilde{\mu}p = \tilde{d} + \text{Lambert}\left(\frac{-\tilde{d}}{e^{\tilde{d}}}\right)
$$

(2.11)
with ‘Lambert’ being (the principal branch of) Lambert’s W function and \( \bar{d} = d/x \) if \( d > x \) and \( \bar{d} = d/x + 0.1 \) if \( d = x \). Then the capture probability, \( p \), can be estimated by
\[
\hat{p} = \frac{\hat{\mu}p}{\hat{\mu}}.
\]

**Proof.** Recall that \( \bar{d} = d/x \), so that (2.3) yields
\[
E(\bar{d}|x) = \frac{\mu p}{(1 - e^{-\mu p})}
\]
and therefore,
\[
E(\bar{d}) = \frac{\mu p}{(1 - e^{-\mu p})}.
\]
(2.13)

Let us first treat \( \mu p \) as a single parameter. Then a moment estimator of \( \mu p \) solves,
\[
\bar{d} = \frac{\mu p}{(1 - e^{-\mu p})}.
\]
Denote the solution by \( \hat{\mu}p \). Then, as \( \bar{d}/e^{\bar{d}} \leq 1/e \) for \( \bar{d} > 1 \), we have (2.11).

Applying the method of moments on (2.6) yields that,
\[
\frac{x}{n} \approx \frac{1 - e^{-\mu p}}{1 + \mu}
\]
which motivates us to take
\[
\hat{\mu} = \frac{n(1 - e^{-\hat{\mu}p})}{x} - 1.
\]
Note that when \( n(1 - e^{-\hat{\mu}p}) < 2x \), \( \hat{\mu} \) takes values less than 1, which are sometimes negative. This arises if there is a small number of daughters per mother within the selected sample (i.e. \( \bar{d} \approx 1 \)) leading to a very small estimate \( \hat{\mu}p \). We have assumed that \( \mu \geq 1 \) in assumption 1. Therefore, in this situation we take \( \hat{\mu} \) to be one, giving (2.10).

Substituting (2.10) in (2.7) gives (2.9). Further (2.10) and (2.11) results in (2.12). \( \square \)
If $\bar{d} = 1$, $\hat{\mu} p$ (2.11) becomes zero which results in $\hat{\mu} = -1$ (2.10). This is unrealistic. Therefore in (2.11) we do an adjustment of 0.1 when $d = x$ and define $\bar{d} = d/x$ if $d > x$ and $\bar{d} = d/x + 0.1$ if $d = x$.

Finally using (2.1), the size of the total female population, i.e. the total number of mothers and daughters in the population, can estimated by,

$$\hat{N} + \hat{D} = \hat{N}(1 + \hat{\mu}), \quad (2.14)$$

where $\hat{D}$ is estimated by $\hat{N}\hat{\mu}$.

### 2.4 Consistency

In this section we show that the two estimators we proposed for $N$, (2.7) and (2.9), the estimator of $\mu$, (2.10), and the estimator of $p$, (2.12), are statistically consistent.

**Theorem 2.4.** Consider a sequence $j = 1, 2, \ldots$ of populations of size $N^{(j)}$ with $N^{(j)} \to \infty$. Let $D^{(j)}$, $n^{(j)}$ and $d^{(j)}$ be $D$, $n$ and $d$ defined for the $j^{th}$ population and similarly for $\tilde{N}^{(j)}(\mu)$ and $\hat{N}^{(j)}$. Suppose that $\mu$ and $p$ do not depend on $j$. Under the assumptions stated in Section 2.2, for known $d$ and unknown $\mu$, as $j \to \infty$,

\begin{align*}
\tilde{N}^{(j)}(\mu) - N^{(j)} &= o_p(N^{(j)}), \quad (2.15) \\
\hat{\mu}^{(j)} &\xrightarrow{p} \mu, \quad (2.16) \\
\hat{N}^{(j)} - N^{(j)} &= o_p(N^{(j)}), \quad (2.17) \\
\text{and} \\
\hat{p}^{(j)} &\xrightarrow{p} p. \quad (2.18)
\end{align*}

**Proof.** Firstly, for given $D^{(j)}$, $n^{(j)}$ follows a $\text{bin}(N^{(j)} + D^{(j)}, p)$ distribution. So that,

$$(N^{(j)})^{-1}n^{(j)} = (N^{(j)})^{-1}(N^{(j)} + D^{(j)})(N^{(j)} + D^{(j)})^{-1}n^{(j)} \xrightarrow{p} (1 + \mu)p,$$

and hence

$$(N^{(j)})^{-2}n^{(j)}(n^{(j)} - 1) \xrightarrow{p} (1 + \mu)^2p^2.$$
The law of large numbers yields

\[(N^{(j)})^{-1}d^{(j)} \xrightarrow{p} \mu p^2.\]

Then (2.15) follows as

\[
(N^{(j)})^{-1} \tilde{N}^{(j)}(\mu) = \frac{\mu(N^{(j)})^{-2}n^{(j)}(n^{(j)} - 1)}{(N^{(j)})^{-1}d^{(j)}(1 + \mu)^2}
\]

\[
\xrightarrow{p} \frac{\mu(1 + \mu)^2p^2}{\mu p^2(1 + \mu)^2}
\]

\[= 1.\]

Next, it is straightforward that

\[
\bar{d} = \frac{d^{(j)}}{x^{(j)}}
\]

\[
= \frac{(N^{(j)})^{-1}d^{(j)}}{(N^{(j)})^{-1}x^{(j)}}
\]

\[
\xrightarrow{p} \frac{\mu p^2}{p(1 - e^{-\mu p})}
\]

\[= \frac{\mu p}{1 - e^{-\mu p}}
\]

\[= g(\mu p).\]

Therefore,

\[\text{Lambert} \left( \frac{-\bar{d}}{e^{-d}} \right) \xrightarrow{p} \text{Lambert} \left( -\frac{g(\mu p)}{e^{g(\mu p)}} \right).\]

Hence

\[\bar{d} + \text{Lambert} \left( \frac{-\bar{d}}{e^{-d}} \right) \xrightarrow{p} g(\mu p) + \text{Lambert} \left( -\frac{g(\mu p)}{e^{g(\mu p)}} \right),\]

which is the solution \(a\) of

\[g(\mu p) = \frac{a}{1 - e^{-a}}.\]

But,

\[g(\mu p) = \frac{\mu p}{1 - e^{-\mu p}}.\]

Therefore \(a = \mu p\) and hence we conclude that

\[\hat{\mu} p^{(j)} = \bar{d}^{(j)} + \text{Lambert} \left( \frac{-\bar{d}^{(j)}}{e^{-\bar{d}^{(j)}}} \right) \xrightarrow{p} \mu p.\]  \hspace{1cm} (2.19)
Now,

\[
\frac{n^{(j)}}{x^{(j)}} = \frac{(N^{(j)})^{-1}n^{(j)}}{(N^{(j)})^{-1}x^{(j)}}
\]

\[
p \rightarrow \frac{(1 + \mu)p}{p(1 - e^{-\mu p})}
\]

\[
= \frac{(1 + \mu)}{1 - e^{-\mu p}}
\]

so that

\[
\hat{\mu}^{(j)} = \frac{n^{(j)}(1 - e^{-\hat{\mu}^{(j)}})}{x^{(j)}} - 1
\]

\[
p \rightarrow \frac{(1 + \mu)(1 - e^{-\mu p})}{1 - e^{-\mu p}} - 1
\]

\[
= \mu,
\]

giving (2.16).

Then,

\[
(N^{(j)})^{-1}\hat{N}^{(j)} = \frac{\hat{\mu}}{(1 + \hat{\mu})^2} \frac{(N^{(j)})^{-2}n^{(j)}(n^{(j)}) - 1}{(N^{(j)})^{-1}d^{(j)}}
\]

\[
p \rightarrow \frac{\mu}{(1 + \mu)^2} \frac{(1 + \mu)^2 p^2}{\mu p^2}
\]

\[
= 1,
\]

which yields (2.17). Finally (2.18) follows from (2.16) and (2.19).

Remark: It may be more realistic to let \( p \) depend on \( j \). That is, \( p_j \rightarrow 0 \) as \( j \rightarrow \infty \).

\section{2.5 Estimator of the standard errors}

In this section, we estimate the standard error of proposed estimators given by (2.7) and (2.9) initially using a theoretical approach and then a bootstrap approach in the next two subsections.
2.5.1 Theoretical estimate of the standard error

Our estimator (2.9) is a function of $d$, $n$, $x$ and $\hat{\mu}$. Therefore, to construct the theoretical estimate of the standard error of (2.9) we first derive, in Proposition 2.5, the asymptotic joint distribution of $d/N$, $n/N$ and $x/N$, using the central limit theorem. Then, in Proposition 2.6, we estimate the variance of $\hat{\mu}$. Finally, we utilise them in estimating the variance of $\hat{N}$ in Theorem 2.7.

**Proposition 2.5.** Let $X = (d/N, n/N, x/N)^T$. Then $X$ asymptotically has a multivariate normal distribution with the vector of means, $\mu_X$ and covariance matrix, $\Sigma_X$. Now,

$$N^{1/2}(X - \mu_X) \xrightarrow{d} N(0, \Sigma_X),$$

where

vector of means, $\mu_X = (\mu_d, \mu_n, \mu_x)^T$ and

covariance matrix, $\Sigma_X = \begin{pmatrix} \sigma_d^2 & \sigma_{dn} & \sigma_{dx} \\ \sigma_{dn} & \sigma_n^2 & \sigma_{nx} \\ \sigma_{dx} & \sigma_{nx} & \sigma_x^2 \end{pmatrix}$

Here $\mu_d$, $\mu_n$, $\mu_x$, $\sigma_d^2$, $\sigma_n^2$, $\sigma_x^2$, $\sigma_{dn}$, $\sigma_{dx}$ and $\sigma_{nx}$ are,

$$\begin{align*}
\mu_d &= \mu p^2 \\
\mu_n &= (1 + \mu)p \\
\mu_x &= p(1 - e^{-\mu p}) \\
\sigma_d^2 &= \mu p^2 + \mu^2 p^3 - \mu^2 p^4 \\
\sigma_n^2 &= p(1 - p) + \mu p \\
\sigma_x^2 &= p(1 - e^{-\mu p}) (1 - p(1 - e^{-\mu p})) \\
\sigma_{dn} &= \mu p^2 (2 - p) \\
\sigma_{dx} &= \mu p^2 (1 - p(1 - e^{-\mu p})) \\
\sigma_{nx} &= p(1 - e^{-\mu p}) + \mu p^2 - (1 + \mu)p^2 (1 - e^{-\mu p})
\end{align*}$$

(2.20)
Proof. We consider \( d = \sum_{i=1}^{N} d_i \), \( n = \sum_{i=1}^{N} n_i \) and \( x = \sum_{i=1}^{N} x_i \) and apply the central limit theorem to derive the asymptotic joint distribution of \( d/N \), \( n/N \) and \( x/N \). In this proof, for convenience, we refer the \( i^{th} \) mother and her daughters as \( i^{th} \) family. Now, suppose there are \( N \) families labelled \( i = 1, \ldots, N \). Let \( d_i \) be the number of observed daughters of family \( i \), \( x_i \) be 1 if the mother of family \( i \) is captured along with at least one of her daughters, and zero otherwise, and \( n_i \) be the total number of individuals captured from family \( i \) (irrespective of identified or not).

Then, to derive the means and variances of \( d/N \), \( n/N \) and \( x/N \) we define a dichotomous variable \( m_i \), which takes the value 1 if the mother of family \( i \) is captured (irrespective of whether identified or not), and zero otherwise. Also, let \( e_i \) be the number of captured daughters of family \( i \). Note that \( m_i \) and \( e_i \) are independent.

Now the number of identified daughters captured from family \( i \), is \( d_i = m_i e_i \). Note that \( n_i = m_i + e_i \), which is the number of all individuals captured from family \( i \), consisting of all captured mothers and daughters irrespective of whether they are identified or not. It can be seen that \( x_i \) is a dichotomous variable taking value 1 if the mother and at least one daughter of family \( i \) is captured, and zero otherwise; i.e. \( x_i = m_i I(e_i > 0) \).

We suppose the captures of families are independent. Now,

\[ m_i \sim \text{Bernoulli}(p) \]

and

\[ e_i \sim \text{Poisson}(\mu p). \]

It is easily seen that

\[ \mu_d = E(d_i) = E(m_i e_i) = \mu p^2, \]

\[ \mu_n = E(n_i) = E(m_i + e_i) = (1 + \mu)p, \]

and

\[ \mu_x = E(x_i) = E(m_i I(e_i > 0)) = p(1 - e^{-\mu p}). \]

Further, we show that

\[ E(m_i^2) = \text{Var}(m_i) + E^2(m_i) = p(1 - p) + p^2 = p \]
and
\[ E(e_i^2) = \text{Var}(e_i) + E^2(e_i) = \mu p + \mu^2 p^2. \]

Moreover,
\[
\sigma_d^2 = \text{Var}(d_i) = E(m_i^2 e_i^2) - E^2(m_i e_i) \\
= p(\mu p + \mu^2 p^2) - \mu^2 p^4,
\]
\[
\sigma_n^2 = \text{Var}(n_i) = \text{Var}(m_i + e_i) \\
= p(1 - p) + \mu p,
\]
and
\[
\sigma_x^2 = \text{Var}(x_i) = E(x_i^2) - E^2(x_i) \\
= E((m_i I(e_i > 0))^2) - E^2(m_i I(e_i > 0)) \\
= p(1 - e^{-\mu p}) \left(1 - p(1 - e^{-\mu p})\right).
\]

Now, to compute the covariance terms we have
\[
E(d_i n_i) = E(m_i e_i (m_i + e_i)) \\
= E(m_i^2 e_i) + E(m_i e_i^2).
\]
Here,
\[
E(m_i^2 e_i) = E(m_i^2) E(e_i) = \mu p^2,
\]
and
\[
E(m_i e_i^2) = E(m_i) E(e_i^2) = p \left(\mu p + \mu^2 p^2\right) = \mu p^2 (1 + \mu p).
\]
Therefore,
\[
E(d_i n_i) = \mu p^2 + \mu p^2 (1 + \mu p) = \mu p^2 (2 + \mu p).
\]
So that
\[
\sigma_{dn} = \text{Cov}(d_i, n_i) = \mu p^2 (2 + \mu p) - \mu p^3 (1 + \mu) \\
= \mu p^2 (2 - p).
\]
Next,

\[ E(d_i x_i) = E(m_i e_i m_i I(e_i > 0)) = E(m_i e_i) = E(d_i) = \mu p^2 \]

yielding

\[ \sigma_{dx} = \text{Cov}(d_i, x_i) = E(d_i x_i) - E(d_i) E(x_i) = \mu p^2 (1 - p(1 - e^{-\mu p})). \]

Moreover,

\[ E(n_i x_i) = E((m_i + e_i)m_i (I(e_i > 0))) = E(x_i) + E(d_i), \]

and hence

\[ \sigma_{nx} = \text{Cov}(n_i, x_i) = E(n_i x_i) - E(n_i) E(x_i) = p(1 - e^{-\mu p}) + \mu p^2 - (1 + \mu)p^2 (1 - e^{-\mu p}). \]

The central limit theorem now yields that for large \( N \)

\[ N^{1/2} \left( \mathbf{X} - \mu_X \right) \xrightarrow{d} N(0, \Sigma_X) \]

with vector of means \( \mu_X \) and covariance matrix \( \Sigma_X \).

Now in Proposition 2.6, to derive the variance of \( \hat{\mu} \), which is a function of \( n, d \) and \( x \), we use the asymptotic joint distribution of \( d/N, n/N \) and \( x/N \), we derived in Proposition 2.5.

**Proposition 2.6.** Consider, \( \mathbf{A} = (A_1, A_2, A_3) \) where

\[
A = e^{-\mu p} \frac{(1 - e^{-\mu p})^2}{1 - e^{-\mu p} - \mu p e^{-\mu p}},
\]

\[ A_1 = \frac{1}{p}, \]

\[ A_2 = A \left( \frac{(1 + \mu)}{p(1 - e^{-\mu p})^2} \right), \]

\[ A_3 = \left( A \frac{(1 + \mu) p}{(1 - e^{-\mu p})^2 + \mu + 1} \right) \frac{1}{p(1 - e^{-\mu p})}. \]
Then
\[ N^{1/2}(\hat{\mu} - \mu) \xrightarrow{d} N(0, A^T\Sigma_X A), \]
so that
\[ \text{Var}(\hat{\mu}) \approx A^T\Sigma_X A, \]
where \( \Sigma_X \) is as defined in Proposition 2.5.

Proof. Let
\[ S(\mu p) = \bar{d} - \frac{\mu p}{1 - e^{-\mu p}}. \]
Solving \( S(\mu p) = 0 \) gives \( \hat{\mu} \). Also, the second order Taylor expansion of \( S(\hat{\mu}) \) around \( \hat{\mu} = \mu p \) gives
\[ S(\hat{\mu}) \approx S(\mu p) + S'(\mu p)(\hat{\mu} - \mu p), \]
and letting \( S(\hat{\mu}) = 0 \) we get
\[
\hat{\mu} - \mu p \approx -S'(\mu p)^{-1}S(\mu p) \\
= \left( \frac{-1}{1 - e^{-\mu p}} + \frac{\mu pe^{-\mu p}}{(1 - e^{-\mu p})^2} \right)^{-1} \left( \bar{d} - \frac{\mu p}{1 - e^{-\mu p}} \right) \\
= \frac{(1 - e^{-\mu p})^2}{1 - e^{-\mu p} - \mu pe^{-\mu p}} \left( \bar{d} - \frac{\mu p}{1 - e^{-\mu p}} \right). \tag{2.21}
\]
Moreover,
\[ \hat{\mu} - \mu = \frac{n}{x}(1 - e^{-\hat{\mu} p}) - (\mu + 1). \]
The second order Taylor expansion of \( e^{-\hat{\mu} p} \) around \( \hat{\mu} p = \mu p \) gives
\[ e^{-\hat{\mu} p} \approx e^{-\mu p} - e^{-\mu p}(\hat{\mu} p - \mu p) + O((\hat{\mu} p - \mu p)^2) \]
Now, as $O((\hat{\mu}p - \mu p)^2)$ is small and negligible, applying this and (2.21) in the above, we obtain

\[
\hat{\mu} - \mu \approx \frac{n}{x}(1 - e^{-\mu p}) + \frac{n}{x}e^{-\mu p}(\hat{\mu}p - \mu p) - (\mu + 1)
\approx \frac{n}{x}(1 - e^{-\mu p}) + \frac{n}{x}e^{-\mu p}\left(\frac{1 - e^{-\mu p}}{1 - e^{-\mu p}} - \mu p\right) - (\mu + 1)
= \frac{n}{x}(1 - e^{-\mu p}) - x(\mu + 1) + e^{-\mu p}\left(\frac{1 - e^{-\mu p}}{1 - e^{-\mu p}} - (\mu + 1)\right).
\]

The first term is

\[
\frac{n(1 - e^{-\mu p}) - x(\mu + 1)}{x} = 1 - e^{-\mu p} - x(\mu + 1) - \frac{\mu + 1}{x}N - p(1 - e^{-\mu p}),
\]

and the second involves

\[
\frac{d}{x} - \frac{\mu p}{1 - e^{-\mu p}} = 1 - \frac{\mu p}{N} - p(1 - \mu p) + A \frac{\mu p}{N} - p(1 - e^{-\mu p}).
\]

Let

\[
A = e^{-\mu p} \frac{(1 - e^{-\mu p})^2}{1 - e^{-\mu p} - \mu p e^{-\mu p}}.
\]

Then

\[
\hat{\mu} - \mu = \frac{N}{x}(1 - e^{-\mu p}) \left(\frac{n}{N} - p(1 + \mu)\right) + A \left(\frac{n}{x}\right) \frac{N}{x} \frac{\mu p}{1 - e^{-\mu p}} - (\mu + 1) \frac{N}{x} - p(1 - e^{-\mu p}).
\]

Now, letting $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, where,

\[
\alpha_1 = \frac{N}{x}(1 - e^{-\mu p}),
\]

\[
\alpha_2 = A \left(\frac{n}{x}\right) \frac{N}{x},
\]

and $\alpha_3 = \left(A \left(\frac{n}{x}\right) \frac{\mu p}{1 - e^{-\mu p}} + \mu + 1\right) \frac{N}{x}$

gives

\[
N^{1/2}(\hat{\mu} - \mu) = \alpha^T N^{1/2} (X - \mu X).
\]
Recall that \( x/N \xrightarrow{p} p(1-e^{-\mu p}) \) and \( n/x \xrightarrow{p} (1+\mu)/(1-e^{-\mu p}) \). Therefore,

\[
\begin{align*}
\alpha_1 & \xrightarrow{p} \frac{1}{p} = A_1, \\
\alpha_2 & \xrightarrow{p} A \left( \frac{(1+\mu)}{p(1-e^{-\mu p})^2} \right) = A_2, \\
\text{and} \quad \alpha_2 & \xrightarrow{p} \left( A \left( \frac{(1+\mu)\mu p}{(1-e^{-\mu p})^2} + \mu + 1 \right) \frac{1}{p(1-e^{-\mu p})} \right) = A_3.
\end{align*}
\]

Now it can be seen that \( \alpha \xrightarrow{p} A \), and from Proposition 2.5, \( N^{1/2} (X - \mu X) \xrightarrow{d} N(0, \Sigma_X) \). Therefore, following Slutsky’s theorem,

\[
N^{1/2}(\hat{\mu} - \mu) \xrightarrow{d} N(0, A^T \Sigma_A A)
\]

yielding Proposition 2.6.

Finally, we use the asymptotic joint distribution of \( d/N, n/N \) and \( x/N \) from Proposition 2.5 and the estimator of the variance of \( \hat{\mu} \) from Proposition 2.6 to derive estimators for the variances of \( \tilde{N} \) and \( \hat{N} \).

**Theorem 2.7.** The variances of \( \tilde{N} \) and \( \hat{N} \) are given by

\[
\text{Var}(\tilde{N}) = B^T \Sigma_X B N \quad (2.22)
\]

and

\[
\text{Var}(\hat{N}) = (B - CA)^T \Sigma_X (B - CA) N, \quad (2.23)
\]

where, \( B = (B_1, B_2, 0) \) with

\[
B_1 = \frac{-1}{\mu p^2}, \quad B_2 = \frac{1}{p(1+\mu)} \quad \text{and} \quad C = \frac{\mu - 1}{\mu(1+\mu)}.
\]
Proof. First, it is easily seen that

\[ \tilde{N} - N = \frac{\mu n(n - 1)}{d(1 + \mu)^2} - N \]

\[ = \frac{\mu n(n - 1) - Nd(1 + \mu)^2}{d(1 + \mu)^2} \]

\[ = \frac{\mu N^{-2}n(n - 1) - N^{-1}d(1 + \mu)^2}{N^{-2}d(1 + \mu)^2} \]

\[ = \frac{\mu N^{-1}(n - 1)}{N^{-2}d(1 + \mu)^2} \left( \frac{n}{N} - (1 + \mu)p \right) - \frac{1}{N^{-2}d} \left( N^{-1}d - \mu p^2 \right) \]

Now, letting \( \beta = (\beta_1, \beta_2, 0) \), where,

\[ \beta_1 = \frac{1}{N^{-1}d}, \]

and \( \beta_2 = \frac{\mu N^{-1}(n - 1)}{N^{-1}d(1 + \mu)^2} \)

gives

\[ \tilde{N} - N = \beta^T (X - \mu_X) N. \]  \hspace{1cm} (2.24)

Now, from (2.2), (2.5) and for large \( N \), we obtain \( N^{-1}(n - 1) \approx N^{-1}n \xrightarrow{p} (1 + \mu)p \)

and \( N^{-1}d \xrightarrow{p} \mu p^2 \). Therefore,

\[ \beta_1 \xrightarrow{p} \frac{-1}{\mu p^2} = B_1 \]

and \( \beta_2 \xrightarrow{p} \frac{1}{p(1 + \mu)} = B_2. \)

Now it can be seen that \( \beta \xrightarrow{p} B \), and from Proposition 2.5, \( N^{1/2} (X - \mu_X) \xrightarrow{d} N(0, \Sigma_X) \). Therefore, following Slutsky’s theorem,

\[ \tilde{N} - N \xrightarrow{d} N(0, B^T \Sigma_X B N) \]

yielding (2.22).

Now, recall from (2.9) that,

\[ \hat{N} = \frac{\hat{\mu} n(n - 1)}{d(1 + \hat{\mu})^2}. \]
From the second order Taylor expansion of \( \hat{N} \) around \( \hat{\mu} = \mu \),

\[
\hat{N} = \tilde{N} + \left( \frac{1}{\hat{\mu}} - \frac{-2}{1 + \mu} \right) \tilde{N}(\hat{\mu} - \mu) + O((\hat{\mu} - \mu)^2)
\]

\[
\approx \tilde{N} - \frac{\mu - 1}{\mu(1 + \mu)} \tilde{N}(\hat{\mu} - \mu) ; \text{as } O((\hat{\mu} - \mu)^2) \text{ is small}
\]

\[
= \tilde{N} - \frac{(\mu - 1)n(n - 1)}{d(1 + \mu)^3}(\hat{\mu} - \mu).
\]

Again recall that \( N^{-1}(n) = N^{-1}(n - 1) = (1 + \mu)p \) and \( N^{-1}d = \mu p^2 \) for large \( N \). Therefore,

\[
\tilde{N} - N \approx \tilde{N} - \frac{(\mu - 1)n(n - 1)}{d(1 + \mu)^3}(\hat{\mu} - \mu) - N
\]

\[
= \tilde{N} - N - \frac{(\mu - 1)n(n - 1)}{d(1 + \mu)^3}(\hat{\mu} - \mu)
\]

\[
= \tilde{N} - N - N^{-2}(\mu - 1)n(n - 1) \frac{N^{-1}d(1 + \mu)^3}{N^{-1}d(1 + \mu)^3} N(\hat{\mu} - \mu)
\]

\[
= \tilde{N} - N - \frac{(\mu - 1)}{\mu(1 + \mu)} N(\hat{\mu} - \mu)
\]

\[
= \tilde{N} - N - CN(\hat{\mu} - \mu)
\]

\[
= B^T(X - \mu_X)N - CA^T(X - \mu_X)N
\]

\[
= (B - CA)^T N (X - \mu_X),
\]

where,

\[
C = \frac{(\mu - 1)}{\mu(1 + \mu)}.
\]

Therefore,

\[
\text{Var}(\hat{N}) = (B - CA)^T \Sigma_X (B - CA)N,
\]

which completes the proof. \( \Box \)

### 2.5.2 Bootstrap estimate of the standard error

The bootstrap method was used as an alternate method to estimate the standard error of the proposed estimator for the number of mothers in the population.
Bootstrap methods depend on the notion of a bootstrap sample. A bootstrap sample is a sample of size $n$ drawn with replacement from a population of $n$ independent objects, all of which are equally likely. The first step of a bootstrap algorithm for estimating the standard errors is selecting $B$ independent bootstrap samples $x^{*1}, x^{*2}, \ldots, x^{*B}$, each consisting of $n$ data values drawn with replacement from the $x$ ($= x_1, x_2, \ldots, x_n$). Then the bootstrap replication corresponding to each bootstrap sample is evaluated and the standard error is estimated by the sample standard deviation of the $B$ replications. But in our case, drawing samples with replacement leads to complications when determining the number of mother-daughter pairs. Therefore, a non-parametric bootstrap method is not appropriate for estimating the standard error in this case.

Alternatively, we adopt a parametric bootstrap approach [10] to determine the standard errors. Suppose the distribution $P$ belongs to a parametric family of distributions $P_{\theta}$ with densities $p(x|\theta)$. If $\hat{\theta}$ is an estimate of the true parameter, $\theta_0$ say, an obvious estimate of $P$ is the distribution $\hat{P} = P_{\hat{\theta}}$ with density $p(x|\hat{\theta})$. In this case, we can still use the bootstrapping method to obtain an estimate of the sampling distribution of $\hat{\theta}$ (or any statistic $g(\hat{\theta})$). Our knowledge about $P$ is incorporated into the bootstrap algorithm by substituting the parametric distribution $P_{\hat{\theta}}$ for the empirical distribution. This is called the parametric bootstrap. $B$ independent bootstrap samples $x^{*1}, x^{*2}, \ldots, x^{*B}$ are drawn from $P_{\hat{\theta}}$, each consisting of $n$ data values. Then the bootstrap replication corresponding to each sample is computed. The standard error is estimated by the sample standard deviation of the $B$ replications.

Our parametric bootstrap procedure begins with computing $\hat{N}$, $\hat{\mu}$ and $\hat{p}$ using the proposed estimators for the captured sample. Then a population with $\hat{N}$ number of mothers is simulated such that the mean number of daughters is equal to $\hat{\mu}$. One thousand bootstrap samples are drawn with a capture probability $\hat{p}$ from this simulated population and the estimators $\hat{N}_B$, $\hat{\mu}_B$ and $\hat{p}_B$ are computed for the 1000 bootstrap samples. The standard deviations of the estimates are calculated as the bootstrap estimate of the standard errors of the estimates.

We compute the confidence intervals (95%) of the estimates, with bootstrap estimates of the standard deviation and normal quantiles. For example, 95% confidence interval of $\hat{N}$ is computed as $\hat{N} \pm 1.96\hat{\sigma}_{\hat{N}(B)}$, where $\hat{\sigma}_{\hat{N}(B)}$ is the bootstrap estimate of the standard deviation of $\hat{N}$. 


2.6 Simulation studies

We examined the performance of the estimators in a simulation study. To do this we generated a population assuming the following properties.

- The population evolves at discrete time steps.
- The individuals born at a particular time constitute a generation.
- Only individuals belonging to the two most recent generations live at a particular point in time.

Simulation studies were conducted in line with the assumptions mentioned in Section 2.2, except for the second assumption. In particular, we assumed that mother-daughter pairs are perfectly identifiable (instead of reliably identifiable) in the simulation studies.

The populations were generated according to the following guidelines.

1. Only the female population was generated as we consider only females in the study.

2. We simulate only two generations at a time to comply with the population structure we consider.

3. The number of mothers in the population ($N$) was selected arbitrarily to consider a wide scope of population size from small to big to examine the performance of the estimator in different settings. We considered $N=50, 100, 200$ and 500.

4. The daughter population was generated assuming that the number of daughters of a given mother are independent and have a Poisson($\mu$) distribution. Again this $\mu$ was determined arbitrarily for the same reason as above. We considered $\mu = 1.25, 2, 5$ and 10.
5. Individuals were labelled such that mother-daughter pairs can be perfectly identified.

Note that we have considered 16 different populations with varying $N$ and $\mu$. One thousand samples were drawn from each population. The samples were drawn such that each individual has the same probability ($p$) of being captured.

### 2.6.1 Performance of the proposed estimators $\hat{N}$, $\hat{\mu}$ and $\hat{p}$

We considered that the capture probability, $p$ is 0.3 in the initial studies. We selected $p=0.3$ for the simulation studies since the standard errors of our estimators starts behaving favourably at this value. We discuss this extensively in Section 2.6.2, where smaller values of $p$ are considered.

$\hat{N}$, $\hat{\mu}$ and $\hat{p}$ was computed for 1000 samples using (2.9), (2.10) and (2.12) respectively.

We present our estimates of $N$, $\mu$ and $p$, which are the means of 1000 estimates (using 1000 samples) computed for a single population in Table 2.2. The estimates of $N$ are very close to the true size of the population over all range, of $N$ and $\mu$ considered in the simulation study. There is some evidence of positive bias in the estimates of $\mu$ when it is small, for example, when $\mu=1.25$. The estimates of $p$ are very close to the true $p$ except slightly underestimated for small $\mu$.

In Table 2.3 we present the true standard deviations of the estimates of $N$ and their theoretical estimates. Recall that $\tilde{N}(\mu)$ is the estimate of $N$ if $\mu$ is known.

In the first line of each row we have the true standard deviation of $\tilde{N}(\mu)$, i.e. the standard deviation of 1000 $\tilde{N}(\mu)$ estimates from 1000 samples. Then we compute the theoretical standard error using (2.22) with true $N$, $\mu$ and $p$ for 1000 samples, and its mean is presented in the second line of each row. In the third line of the row we have the mean of 1000 estimates computed using (2.22) with $\tilde{N}(\mu)$, $\mu$ and $\hat{p}$. We have the standard deviation of 1000 $\hat{N}$’s from 1000 samples in the fourth line of the row. The mean of 1000 estimates computed using (2.23) with $N$, $\mu$ and $p$ is in the fifth line of each row, while the same computed with $\tilde{N}$, $\hat{\mu}$ and $\hat{p}$ is in the last line of each row. The estimated standard errors of $\tilde{N}(\mu)$ are very close to the
true standard deviation when computed with $N$, $\mu$ and $p$. The estimates are also reasonable in general when computed with $\tilde{N}(\mu)$, $\mu$ and $\hat{p}$. The estimated standard errors of $\hat{N}$ are also sensible in most cases, except in some cases where the standard deviation is over estimated, specifically when $\mu$ is small. The distribution of our estimates, and hence also the variance of $\hat{N}$, is positively skewed when $\mu$ is small, resulting in a higher estimate of the standard error. Considering the median of the estimates of the square root of variance of $\hat{N}$ as the theoretical estimate of standard error of $\hat{N}$ instead of the mean delivers less bias. For example, when $N = 50$, $\mu = 2$ and $p = 0.3$, the median of the estimates is 26.45, whereas the mean is 30.76 and
the true standard deviation is 25.9, and when \( N = 200, \mu = 1.25 \) and \( p = 0.3 \), the median of the estimates is 49.85, where as the mean is 60.07 and the true standard deviation is 39.94.

Table 2.3: Theoretical estimates of standard errors

<table>
<thead>
<tr>
<th>( N )</th>
<th>( \mu = 1.25 )</th>
<th>( \mu = 2 )</th>
<th>( \mu = 5 )</th>
<th>( \mu = 10 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>( \sigma_{\hat{N}(\mu)} )</td>
<td>20.55</td>
<td>18.10</td>
<td>14.28</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(\mu)(1)}) )</td>
<td>19.33</td>
<td>15.99</td>
<td>12.64</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(\mu)(2)}) )</td>
<td>28.72</td>
<td>22.80</td>
<td>19.37</td>
</tr>
<tr>
<td></td>
<td>( \sigma_{\hat{N}} )</td>
<td>20.37</td>
<td>19.53</td>
<td>15.83</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(T)(1)}) )</td>
<td>21.42</td>
<td>20.79</td>
<td>12.57</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(T)(2)}) )</td>
<td>31.18</td>
<td>25.92</td>
<td>14.93</td>
</tr>
<tr>
<td>100</td>
<td>( \sigma_{\hat{N}(\mu)} )</td>
<td>27.10</td>
<td>22.57</td>
<td>18.95</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(\mu)(1)}) )</td>
<td>27.33</td>
<td>22.61</td>
<td>17.87</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(\mu)(2)}) )</td>
<td>37.48</td>
<td>26.37</td>
<td>18.83</td>
</tr>
<tr>
<td></td>
<td>( \sigma_{\hat{N}} )</td>
<td>24.80</td>
<td>25.90</td>
<td>18.01</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(T)(1)}) )</td>
<td>30.29</td>
<td>29.39</td>
<td>17.78</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(T)(2)}) )</td>
<td>40.44</td>
<td>30.76</td>
<td>18.03</td>
</tr>
<tr>
<td>200</td>
<td>( \sigma_{\hat{N}(\mu)} )</td>
<td>39.61</td>
<td>31.68</td>
<td>24.59</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(\mu)(1)}) )</td>
<td>38.65</td>
<td>31.97</td>
<td>25.28</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(\mu)(2)}) )</td>
<td>55.48</td>
<td>34.20</td>
<td>25.67</td>
</tr>
<tr>
<td></td>
<td>( \sigma_{\hat{N}} )</td>
<td>39.94</td>
<td>38.13</td>
<td>24.02</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(T)(1)}) )</td>
<td>42.83</td>
<td>41.57</td>
<td>25.14</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(T)(2)}) )</td>
<td>60.07</td>
<td>41.10</td>
<td>25.29</td>
</tr>
<tr>
<td>500</td>
<td>( \sigma_{\hat{N}(\mu)} )</td>
<td>58.63</td>
<td>47.37</td>
<td>38.49</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(\mu)(1)}) )</td>
<td>61.11</td>
<td>50.55</td>
<td>39.97</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(\mu)(2)}) )</td>
<td>80.73</td>
<td>51.66</td>
<td>40.66</td>
</tr>
<tr>
<td></td>
<td>( \sigma_{\hat{N}} )</td>
<td>60.00</td>
<td>58.61</td>
<td>36.35</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(T)(1)}) )</td>
<td>67.72</td>
<td>65.73</td>
<td>39.75</td>
</tr>
<tr>
<td></td>
<td>( mean(\hat{\sigma}_{\hat{N}(T)(2)}) )</td>
<td>87.57</td>
<td>64.85</td>
<td>39.85</td>
</tr>
</tbody>
</table>

The standard deviation of the 1000 \( \hat{N}(\mu) \) is denoted by \( \sigma_{\hat{N}(\mu)} \). The mean of the 1000 theoretical estimates of standard deviations of \( \hat{N}(\mu) \), computed with true \( N, \mu \) and \( p \) is denoted by \( mean(\hat{\sigma}_{\hat{N}(\mu)(1)}) \) and the mean of the same computed with \( \hat{N}(\mu), \mu \) and \( \hat{p} \) is denoted by \( mean(\hat{\sigma}_{\hat{N}(\mu)(2)}) \). The standard deviation of the 1000 \( \hat{N} \) is denoted by \( \sigma_{\hat{N}(T)} \). The mean of the 1000 theoretical estimates of standard deviations of \( \hat{N} \), computed with true \( N, \mu \) and \( p \) is denoted by \( mean(\hat{\sigma}_{\hat{N}(T)(1)}) \) and the mean of the same computed with \( \hat{N}, \mu \) and \( \hat{p} \) is denoted by \( mean(\hat{\sigma}_{\hat{N}(T)(2)}) \).

In Table 2.4 we present the true standard deviations of \( \hat{N} \), \( \hat{\mu} \) and \( \hat{p} \) and their bootstrap estimates. True standard deviation is the standard deviation of 1000 estimates from 1000 samples. A population was simulated with \( \hat{N}, \hat{\mu} \) and \( \hat{p} \) of each sample.
One thousand bootstrap samples were drawn from this population and the standard deviation of those estimates was computed as the bootstrap estimate of the standard deviation for the sample. The mean of 1000 bootstrap estimates of standard deviation is then considered as the estimated standard deviation. The estimated standard deviation of $\hat{N}$ is in general, reasonable. However, it is slightly overestimating the standard deviation of $\hat{N}$ for small $\mu$, especially when $N$ is large. The positive skewness of the distribution of estimates is the reason for this. Hence, considering the median of the bootstrap estimates of standard deviation of $\hat{N}$ for all samples as the bootstrap estimate of the standard deviation of $\hat{N}$ instead of the mean delivers less bias. For example, when $N = 100$, $\mu = 2$ and $p = 0.3$, the median of the estimates is 28.88, whereas the mean is 31.46 and the true standard error is 25.9, and when $N = 500$, $\mu = 1.25$ and $p = 0.3$, the median of the estimates is 75.76, whereas the mean is 80.39 and the true standard error is 60.0.

In comparison with the estimates from the theoretical approach presented in Table 2.3 the estimates from the bootstrap approach are, in general, better in estimating the standard deviation of $\hat{N}$. Therefore, from this point forward we use the bootstrap approach for estimating the standard deviation of all estimators.

Figure 2.1 shows the distribution of $\hat{N}$ for populations of different sizes with varying $\mu$. The dotted line shows the true $N$ value. In the figure we also report the percentage bias. In general, the bias of $\hat{N}$ is small. The numbers below each boxplot show the mean of the parametric bootstrap estimates of the standard errors. Confidence intervals (95%) of the estimates were computed with bootstrap estimates of the standard deviation and normal quantiles. Then the coverage probability, i.e. the proportion of the confidence intervals that contains the true value of the parameter was calculated. We present these coverage probability of nominal 95% confidence intervals for $N$ in the figure, denoted by $\phi$.

The same information for $\hat{\mu}$ and $\hat{p}$ is shown respectively in Figure 2.2 and Figure 2.3. It can be seen that, $\mu$ is estimated satisfactorily, except a considerable bias for populations of size 50 and 100 when $\mu = 1.25$. Also, the distribution of $\hat{p}$ is satisfying, except a considerable bias at $\mu=1.25$. 
Table 2.4: Bootstrap estimates of standard errors

<table>
<thead>
<tr>
<th>N</th>
<th>$\sigma_N$</th>
<th>$\mu = 1.25$</th>
<th>$\mu = 2$</th>
<th>$\mu = 5$</th>
<th>$\mu = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>20.37</td>
<td>19.53</td>
<td>15.83</td>
<td>7.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean($\hat{\sigma}_N(B)$)</td>
<td>20.98</td>
<td>22.47</td>
<td>21.62</td>
<td>9.70</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\mu$</td>
<td>1.71</td>
<td>1.65</td>
<td>1.82</td>
<td>2.70</td>
</tr>
<tr>
<td></td>
<td>mean($\hat{\sigma}_\mu(B)$)</td>
<td>1.81</td>
<td>1.92</td>
<td>2.37</td>
<td>3.32</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\rho$</td>
<td>0.09</td>
<td>0.10</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>mean($\hat{\sigma}_\rho(B)$)</td>
<td>0.08</td>
<td>0.09</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>100</td>
<td>24.80</td>
<td>25.90</td>
<td>18.01</td>
<td>10.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean($\hat{\sigma}_N(B)$)</td>
<td>36.36</td>
<td>31.46</td>
<td>19.41</td>
<td>10.72</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\mu$</td>
<td>0.95</td>
<td>1.16</td>
<td>1.26</td>
<td>1.79</td>
</tr>
<tr>
<td></td>
<td>mean($\hat{\sigma}_\mu(B)$)</td>
<td>1.82</td>
<td>1.48</td>
<td>1.38</td>
<td>1.93</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\rho$</td>
<td>0.09</td>
<td>0.08</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>mean($\hat{\sigma}_\rho(B)$)</td>
<td>0.07</td>
<td>0.08</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>200</td>
<td>39.94</td>
<td>38.13</td>
<td>24.02</td>
<td>14.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean($\hat{\sigma}_N(B)$)</td>
<td>50.99</td>
<td>40.14</td>
<td>24.62</td>
<td>14.30</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\mu$</td>
<td>0.71</td>
<td>0.86</td>
<td>0.82</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>mean($\hat{\sigma}_\mu(B)$)</td>
<td>1.36</td>
<td>0.89</td>
<td>0.87</td>
<td>1.28</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\rho$</td>
<td>0.08</td>
<td>0.07</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>mean($\hat{\sigma}_\rho(B)$)</td>
<td>0.07</td>
<td>0.07</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>500</td>
<td>60.00</td>
<td>58.61</td>
<td>36.35</td>
<td>23.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean($\hat{\sigma}_N(B)$)</td>
<td>80.39</td>
<td>60.37</td>
<td>37.18</td>
<td>22.29</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\mu$</td>
<td>0.45</td>
<td>0.53</td>
<td>0.51</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>mean($\hat{\sigma}_\mu(B)$)</td>
<td>0.78</td>
<td>0.54</td>
<td>0.53</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\rho$</td>
<td>0.06</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>mean($\hat{\sigma}_\rho(B)$)</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
</tr>
</tbody>
</table>

The standard deviation of the 1000 estimates of $N$ is denoted by $\sigma_N$, and the mean of the 1000 bootstrap estimates of standard deviations is denoted by $mean(\hat{\sigma}_N(B))$, and so forth for $\mu$ and $p$. 
Figure 2.1: Distribution of $\hat{N}$

These are based on the 1000 simulations under scenarios (a) $N=50$ (b) $N=100$ (c) $N=200$ (d) $N=500$ with $p=0.3$. The numbers above each boxplot show percentage bias and the numbers below show the mean of standard errors determined using 1000 bootstrap samples ($E(\hat{\sigma})$) and the coverage probability of nominal 95% confidence intervals ($\Phi$).
These are based on the 1000 simulations under scenarios (a) $\mu=1.25$ (b) $\mu=2$ (c) $\mu=5$ (d) $\mu=10$ with $p=0.3$. The numbers above each boxplot show percentage bias and the numbers below show the mean of standard errors determined using 1000 bootstrap samples ($E(\hat{\sigma})$) and the coverage probability of nominal 95% confidence intervals ($\Phi$).
These are based on the 1000 simulations under scenarios (a) N=50 (b) N=100 (c) N=200 (d) N=500 with \( p = 0.3 \). The numbers above each boxplot show percentage bias and the numbers below show the mean of standard errors determined using 1000 bootstrap samples (\( E(\hat{\sigma}) \)) and the coverage probability of nominal 95% confidence intervals (\( \Phi \)).
2.6.2 The effect of capture probability on the estimators

We conducted some simulation studies with $p=0.5$ to observe the effect of capture probability. We re-ran the simulations for the 3 cases as below:

1. $N=50$ and $\mu=2$
2. $N=100$ and $\mu=5$
3. $N=200$ and $\mu=2$

We observed a marked increase in precision when $p=0.5$ compared with $p=0.3$ (Table 2.5). The standard errors are now much smaller with very good coverage probabilities.

Table 2.5: Simulation results illustrating the increased precision at higher $p$ ($p=0.5$)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$N=50$, $\mu=2$</th>
<th>$N=100$, $\mu=5$</th>
<th>$N=200$, $\mu=2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SD</td>
<td>SE</td>
</tr>
<tr>
<td>$N$</td>
<td>49.92</td>
<td>9.93</td>
<td>11.12</td>
</tr>
<tr>
<td>$\mu$</td>
<td>2.00</td>
<td>0.79</td>
<td>0.88</td>
</tr>
<tr>
<td>$p$</td>
<td>0.50</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>94%</td>
<td>95%</td>
<td>95%</td>
</tr>
</tbody>
</table>

Here $Estimate$ is the mean of 1000 estimates of the parameter, $SD$ is the standard deviation of the 1000 estimates and $SE$ is the mean of 1000 standard errors of the estimates.

In Table 2.6 we present results of simulation studies with $p=0.15$, 0.2 and 0.7 for the same 3 cases considered in Table 2.5. As seen in Table 2.6, with small $p$, particularly at $p=0.15$ and $p=0.2$, the standard deviations of the estimates are quite large. Also, the standard errors determined using the bootstrap method do not estimate the standard deviation satisfactorily. When the capture probability is as large as 0.7 the standard deviations of the estimates are very small while the bootstrap method estimates them excellently.
Table 2.6: Simulation results with $p=0.15$, 0.2 and 0.7

<table>
<thead>
<tr>
<th>$p$</th>
<th>Parameter</th>
<th>$N=50$, $\mu=2$</th>
<th></th>
<th></th>
<th>$N=100$, $\mu=5$</th>
<th></th>
<th></th>
<th>$N=200$, $\mu=2$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SD</td>
<td>SE</td>
<td></td>
<td>Estimate</td>
<td>SD</td>
<td>SE</td>
<td></td>
<td>Estimate</td>
</tr>
<tr>
<td>0.15</td>
<td>$N$</td>
<td>47.35</td>
<td>23.59</td>
<td>19.99</td>
<td></td>
<td>122.89</td>
<td>65.03</td>
<td>104.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>2.10</td>
<td>1.93</td>
<td>1.81</td>
<td></td>
<td>5.01</td>
<td>3.66</td>
<td>3.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p$</td>
<td>0.20</td>
<td>0.06</td>
<td>0.13</td>
<td></td>
<td>0.16</td>
<td>0.06</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>$N$</td>
<td>50.64</td>
<td>24.12</td>
<td>23.04</td>
<td></td>
<td>105.98</td>
<td>36.56</td>
<td>58.58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>2.15</td>
<td>2.07</td>
<td>1.94</td>
<td></td>
<td>5.84</td>
<td>2.59</td>
<td>2.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p$</td>
<td>0.21</td>
<td>0.07</td>
<td>0.13</td>
<td></td>
<td>0.21</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>$N$</td>
<td>50.44</td>
<td>5.27</td>
<td>4.89</td>
<td></td>
<td>100.68</td>
<td>3.72</td>
<td>3.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>1.92</td>
<td>0.41</td>
<td>0.41</td>
<td></td>
<td>5.07</td>
<td>0.39</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p$</td>
<td>0.72</td>
<td>0.10</td>
<td>0.09</td>
<td></td>
<td>0.70</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

Here *Estimate* is the mean of 1000 estimates of the parameter, *SD* is the standard deviation of the 1000 estimates and *SE* is the mean of 1000 standard errors of the estimates.

The theoretical estimates of standard error of $\hat{N}$ and $\hat{N}(\mu)$ are plotted against the capture probability in Figure 2.4. We have considered a population with $N=100$ and $\mu=2$. The theoretical estimates are computed with true $N$, $\mu$ and $p$. The standard errors of $\hat{N}$ and $\hat{N}(\mu)$ are quite large for $p < 0.3$. The standard errors start behaving favourably after $p=0.3$ and perform better as $p$ increases. Therefore, in our simulations we have used $p=0.3$ and we recommend that capture probability is at least 0.3 to use this method. The theoretical estimates of standard error of $\hat{N}$ and $\hat{N}(\mu)$ for other values of $N$ and $\mu$ we have considered in our simulations studies behave in the same manner, which we do not report here.
Here we have considered a population with \( N=100 \) and \( \mu=2 \). \( \hat{\sigma}_N \) and \( \hat{\sigma}_{\tilde{N}(\mu)} \) are the theoretical estimates of standard error of \( \tilde{N} \) and \( \tilde{N}(\mu) \) computed with true \( N \), \( \mu \) and \( p \) respectively.

Figure 2.4: Theoretical standard errors of \( \tilde{N} \) and \( \tilde{N}(\mu) \) against \( p \)
2.6.3 Estimating the total population

The total female population size, $E(N + D)$, was computed following (2.14). As shown in Figure 2.5 the percentage bias of $\hat{N} + \hat{D}$ is quite small, especially when either $N$ or $\mu$ is large this is negligible.

![Figure 2.5: Percentage bias of $\hat{N} + \hat{D}$](image)

We do not report the standard errors of $\hat{N} + \hat{D}$ in detail, as they are often too large to be useful in practice. We present several examples in Table 2.7. It can be seen that a higher capture probability reduces the standard errors of $\hat{N} + \hat{D}$. 
Table 2.7: Standard errors of $\hat{N} + \hat{D}$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$N + D$</th>
<th>$E(\hat{N} + \hat{D})$</th>
<th>$\sigma(\hat{N} + \hat{D})$</th>
<th>$mean(\hat{\sigma}(\hat{N} + \hat{D})(B))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>$N + D$</td>
<td>600</td>
<td>600</td>
<td>606.8</td>
</tr>
<tr>
<td></td>
<td>$E(\hat{N} + \hat{D})$</td>
<td>642.4</td>
<td>597.2</td>
<td>624.2</td>
</tr>
<tr>
<td></td>
<td>$\sigma(\hat{N} + \hat{D})$</td>
<td>127.6</td>
<td>131.1</td>
<td>71.1</td>
</tr>
<tr>
<td></td>
<td>$mean(\hat{\sigma}(\hat{N} + \hat{D})(B))$</td>
<td>142.5</td>
<td>139.4</td>
<td>142.5</td>
</tr>
</tbody>
</table>

The standard deviation of the 1000 estimates of $N + D$ is denoted by $\sigma(\hat{N} + \hat{D})$ and the mean of the 1000 bootstrap estimates of standard deviations is denoted by $mean(\hat{\sigma}(\hat{N} + \hat{D})(B))$. 
2.7 Estimating $N$ when the number of mother-daughter pairs in the sample is unknown

The number of mother-daughter pairs in the sample is required to estimate the population size, $N$, using the proposed estimators given by (2.9) and (2.7). In the construction of these two estimators, we considered that the mother-daughter pairs are perfectly identifiable. Nevertheless, we need to estimate this in most practical situations.

Here we adopt the approach using genotype data to identify parent-offspring pairs of Skaug et al. [34], and their R program available from http://folk.uib.no/hsk021/fdr_program.htm. This approach uses a false discovery rate (FDR) procedure based on the likelihood ratio (LOD score). The LOD score is computed using genetic markers based on an alternative hypothesis that the individuals are related with a specified kinship coefficient. The kinship coefficient for a parent-offspring pair is 0.25. Therefore the LOD score to identify mother-daughter pairs is computed with the hypothesis that the individuals are unrelated against the alternative of a relationship with kinship coefficient equal to 0.25. Hence a large LOD score implies a relationship with kinship coefficient of 0.25; i.e. a parent-offspring pair. To determine how ‘large’ the LOD score requires to be for a pair to be identified as having a relationship with kinship coefficient 0.25, a p-value is computed using a permutation approach. That is, a sample of the same size is simulated by distributing all alleles in a particular locus independently across loci, which results in a sample with the same allele frequencies as the original, yet with unrelated individuals. The LOD scores are computed for all $m$ pairs ($m = n(n - 1)/2$ for a sample of size $n$) in the simulated data set. The p-value for a pair is the proportion of simulated LOD scores exceeding that of the pair. These p-values for all pairs are then arranged in increasing order and the first $R$ pairs are considered as related. This approach is based on the concept of controlling the proportion of false positives, i.e. the proportion of unrelated pairs which are labelled as related. Hence Skaug et al. [34] define $R$ as the largest value of $r$ for which $P_r \leq r/m \cdot q$ such that the FDR is controlled at a nominated level $q$. Note that the kinship coefficient is 0.25 for both parent-offspring relationship and full siblings. But, when considering an alternative
hypothesis of a kinship coefficient equal to 0.25, the LOD score presented in Skaug et al. [34] (followed by Skaug [33]) counts as zero if the pair do not share an allele at each loci. Therefore, this method will not identify a pair if they do not share an allele at each locus even if they have a kinship coefficient of 0.25. We explain this further in Figure 2.6. Every child’s cell has a paternal allele paired with a maternal allele at each loci. Therefore, a mother and each of her daughters share an allele at each loci. But, two daughters do not necessarily share an allele at each loci. For example, Child 1 and Child 4 does not share any allele in the Figure 2.6.

![Figure 2.6: Allele combinations of parents and their children](image)

The probability that full sisters share at least one allele at all, \( k \) loci, is very low for large \( k \) (\( =0.75^k \), supposing that there is no inbreeding). Also the existence of full sisters, subsequently the probability of capturing such a pair, is not large for pygmy possums, being a non-monogamous species. Therefore, the chance that a full-sibling pair is identified using this procedure is very small. But this is not an issue for identifying parent-offspring pairs, as they necessarily share an allele at each locus. Hence, we ignore that full sisters exist among pairs identified using the procedure above in our application. In the application we considered genotype data at 21 loci.

The number of pairs identified in the above fashion is \( d \). The individuals involved with these \( d \) pairs is the collection of the daughters of whose mother is also captured (which is again equal to \( d \)), and mothers with at least one daughter captured (\( x \)) under the restriction of two generations. Thus, the difference between the number of
unique individuals in the $d$ pairs and $d$ gives $x$. Note that the explicit classification of mothers and daughters is not necessary as identification of mother-daughter pairs and unique individuals allow us to determine $n$, $d$ and $x$ as described above.

2.8 Application

We consider data on the female mountain pygmy possums (*Burramys parvus*) collected at several sites at Mount Hotham, Victoria, Australia between 2010 and 2012 during their breeding season in October and November. The captured possums were genotyped at 21 loci. ‘Site A’ consists of the possums captured in Mount Higginbotham West (H100 in Figure 2.7) and ‘Site B’ Mount Higginbotham (H52, H54 and H100 in Figure 2.7), whereas ‘Site C’ consists of those at the Mount Little Higginbotham site (L68 and L69 in Figure 2.7). The data is summarised in Table 2.8.

![Figure 2.7: Map of *Burramys parvus* collections sites in Mount Hotham, Australia](image-url)
Table 2.8: The application data

<table>
<thead>
<tr>
<th>Information</th>
<th>Site A</th>
<th>Site B</th>
<th>Site C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>43</td>
<td>70</td>
<td>57</td>
</tr>
<tr>
<td>$d$</td>
<td>5</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>$x$</td>
<td>4</td>
<td>7</td>
<td>9</td>
</tr>
</tbody>
</table>

Recall $n$ is the number of captured individuals, $d$ is the number of mother-daughter pairs and $x$ is the number of mothers with at least one daughter identified.

Here the number of mother-daughter pairs has to be estimated. We utilise the LOD score-based method described in Section 2.7. We controlled the proportion of false positives at a maximum of 0.05. Again, the standard errors of the estimators were computed using the parametric bootstrap method described in Section 2.5.2. The estimates arising from our approach are given in Table 2.9. The standard errors are quite high, reflecting the small estimated capture probabilities and the fairly small population sizes. However, these sample sizes are quite small. Note that Skaug [33] regarded 334 as a small sample size. Here, we use genetic makers to identify mother-daughter pairs, but we have not incorporated any variability that may arise from this classification into our estimates of standard errors. This will be discussed in Chapter 6.

Table 2.9: Parameter estimates for the three sites

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Site A</th>
<th>Site B</th>
<th>Site C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
<td>Estimate</td>
</tr>
<tr>
<td>$N$</td>
<td>67.81</td>
<td>50.48</td>
<td>146.90</td>
</tr>
<tr>
<td>$\mu$</td>
<td>2.99</td>
<td>3.29</td>
<td>1.39</td>
</tr>
<tr>
<td>$p$</td>
<td>0.16</td>
<td>0.11</td>
<td>0.20</td>
</tr>
</tbody>
</table>
2.9 Populations with mean number of daughters per mother less than one

In Section 2.2, we mentioned that we assume the number of daughters of a given mother has a Poisson(\(\mu\)) distribution where \(\mu \geq 1\) and the number of daughters are independent for different mothers. We discuss the technical requirement behind this assumption and the issues raised when we use the estimators we proposed for a population with \(\mu < 1\).

2.9.1 High probability of observing an equal number of mothers and daughters

One major problem that occurs is that when \(\mu < 1\) the probability \(d = x\) is quite high. For example, when \(N = 50, \mu = 0.6\) and \(p = 0.3\) the probability \(d = x\) is 0.81 (we show the computation later in the section). Recall that our estimator of \(\hat{\mu}p\) (2.11) is 0 when \(\hat{d} = 1\), i.e. when \(d = x\). Therefore, we often need to apply an adjustment to \(\hat{d}\) resulting in biased estimates. The problem is that mothers are only identified through capture of them along with their daughters. Therefore, for \(d/x > 1\) we need to observe one mother with multiple daughters. If \(\mu < 1\) this may be rare, and in this case we expect a high probability that \(d = x\).

Recall that \(d_i\) is the number of observed daughters of a particular mother. Given that there are \(x\) mothers from whom at least one daughter is observed in the sample the total number of observed daughters with their mother being observed will be the sum of \(d_i\) over these mothers: \(d = \sum_{i=1}^{x} d_i\).

For the number of observed mothers \(x\) to be equal to the number of observed daughters \(d\), the number of observed daughters of a mother has to be 1 for all mothers.

\[
Pr(d = x|N, \mu, x) = Pr(d_i = 1 \text{ for all } i = 1, \ldots, x|N, \mu, x)
\]
The number of observed daughters per each observed mother is independent from each other. Therefore,

\[ Pr(d = x | N, \mu, x) = (Pr(d_i = 1 | N, \mu))^x. \]

Then as shown in the proof of (2.3),

\[ d_i | d_i > 0 \sim ZTP (\mu p). \]

Therefore,

\[ Pr(d = x | N, \mu, x) = \left( \frac{\mu pe^{-\mu p}}{1 - e^{-\mu p}} \right)^x. \]

Now,

\[ Pr(d = x | N, \mu) = E(Pr(d = x | N, \mu, x)) \]
\[ = E \left( \left( \frac{\mu pe^{-\mu p}}{1 - e^{-\mu p}} \right)^x \right) \]
\[ = G \left( \frac{\mu pe^{-\mu p}}{1 - e^{-\mu p}} \right). \]

Here \( G \) is the probability generating function of \( x \).

Recall that,

\[ x \sim \text{bin}(N, p(1 - e^{-\mu p})) \]

Hence, using the probability generation function of binomially distributed random variable we get,

\[ Pr(d = x | N, \mu) = \left( 1 - p(1 - e^{-\mu p}) + \frac{\mu pe^{-\mu p}}{1 - e^{-\mu p}} p(1 - e^{-\mu p}) \right)^N \]
\[ = (1 - p + pe^{-\mu p} + \mu p^2 e^{-\mu p})^N \quad (2.25) \]

We plot the \( Pr(d = x | N, \mu) \) for \( \mu \leq 1 \) and \( N = 50, 100, \ldots 500 \) when the capture probability is 0.3 in Figure 2.8.
Figure 2.8: The probability $d = x$ when $p = 0.3$ and $\mu \leq 1$

Curves are for $N = 50, 100, \ldots, 500$ from black to blue.

For example, consider a population where $N = 50$ and $\mu = 0.6$. If $p = 0.3$ the probability $d = x$ from (2.25) is,

$$Pr(d = x|N = 50, \mu = 0.6) = 0.81$$

We have confirmed these probabilities with simulation studies considering different $N$ and $\mu$ values, which we do not report here.

The first derivative of (2.25) with respect to $\mu$ is negative for all values of $N$, $\mu$ and $p$.

$$\frac{\partial}{\partial \mu} Pr(d = x|N, \mu) = -N \mu p^3 e^{-\mu p}(1 - p + pe^{-\mu p} + \mu p^2 e^{-\mu p})^{N-1}$$
Therefore, \( Pr(d = x|N, \mu) \) is a decreasing function of \( \mu \) for a given value \( N \) and \( p \). We further demonstrate this in Figure 2.9 where we plot the \( Pr(d = x|N, \mu) \) for \( 0 \leq \mu \leq 5 \) and \( N = 50, 100, \ldots, 500 \) when the capture probability is 0.3. As seen in the figure the probability becomes very small as \( \mu \) increases. Figure 2.10 also exhibits this.

Also, the first derivative of (2.25) with respect to \( N \) is negative for all values of \( N \), \( \mu \) and \( p \).

\[
\frac{\partial}{\partial N} Pr(d = x|N, \mu) = (1 - p + pe^{-\mu p} + \mu p^2 e^{-\mu p})^N \ln(1 - p + pe^{-\mu p} + \mu p^2 e^{-\mu p})
\]

Therefore, \( Pr(d = x|N, \mu) \) decreases with \( N \) for a given value of \( \mu \) and \( p \). We illustrate this in Figure 2.11 in which we plot \( Pr(d = x|N, \mu) \) against \( N \) for different values of \( \mu \). This also can be seen in Figures 2.8 and 2.9.
Further, we examined the effect of $p$ on $Pr(d = x|N, \mu)$. Again, first derivative of (2.25) with respect to $N$ is negative for all values of $N$, $\mu$ and $p$.

$$\frac{\partial}{\partial p} Pr(d = x|N, \mu) = -N \left(1 - e^{-\mu p} - pe^{-\mu p} + \mu^2 p^2 e^{-\mu p} \right) \left(1 - p + pe^{-\mu p} + \mu p^2 e^{-\mu p} \right)^{-1}$$
This implies $Pr(d = x|N, \mu)$ decreases with $p$ for a given value $N$ and $\mu$. We graphically show this in Figure 2.12.

Figure 2.12: $Pr(d = x|N, \mu)$ against $p$ for different values of $\mu$ ($N = 100$)
2.9.2 Sensitivity of the estimators to the observed number of mothers and daughters

As we mentioned in the previous section $\bar{d}$ is close to one when $\mu < 1$. When $\bar{d}$ is close to one our estimators are very sensitive to changes in $\bar{d}$.

Recall our estimator of $\hat{\mu}_p$ given in (2.11).

$$\hat{\mu}_p = \bar{d} + \text{Lambert}\left(\frac{-d}{e^d}\right)$$

Now the rate of change of $\hat{\mu}_p$ with $\bar{d}$ is,

$$\frac{d\hat{\mu}_p}{d\bar{d}} = (1 - \bar{d})\text{Lambert}(-\bar{d}e^{-\bar{d}}) + 1$$

We plot this rate of change of $\hat{\mu}_p$ against $\bar{d}$ in Figure 2.13. As seen in the figure this rate is the highest, in fact, approximately 2 when $\bar{d}$ is close to 1. Therefore, even a small difference in $\bar{d}$ makes a bigger change in $\hat{\mu}_p$.

Now recall our estimator of $\hat{\mu}$ given in (2.10).

$$\hat{\mu} = \begin{cases} \frac{n(1-e^{-\hat{\mu}_p})}{x} - 1, & \text{if } n(1-e^{-\hat{\mu}_p}) > 2x, \\ 1, & \text{otherwise}, \end{cases}$$

When the mean number of daughters per mother is small, the number of observed mother-daughter pairs is also small. Therefore $n/x$ tends to be large. This magnifies the change of $\hat{\mu}_p$, which is already magnified by the exponential function.

Thus the changes in $\bar{d}$ cause bigger changes in our estimates. This effect takes place in two ways. One is the adjustment to $\bar{d}$ we perform when we observe equal number of daughters and mothers as we described in Section 2.9.1. The other is the natural fluctuations in $\bar{d}$. This can be due to natural variation of the variables around their mean or misclassification that we discuss in Chapter 6. Note that when having small $\bar{d}$ and $x$ when $\mu$ is small, an additional daughter or mother can cause a big impact on $\bar{d}$.
We further show this with a few examples in Table 2.10. We consider two values of \(N\) (=50 and 500) four values of \(\mu\) (=0.4, 0.8, 2, 5) and \(p=0.3\). In the first set of rows of the table we have the expected values of \(n, d, x\) and \(\bar{d}\) computed using (2.2), (2.5), (2.4) and (2.13) respectively. In the second set of rows we round the expected values and present \(n, d\) and \(x\) values that we use for calculating our estimates. In the next two set of rows we compute the estimates adding a one and two daughters respectively, where as in the last set of rows we compute the estimates adding a mother. It can be seen that the change in \(\bar{d}\) and the estimates is much larger for a population with small \(\mu\) when we make a small alteration in \(d\) or \(x\), than we do the same on a population with larger \(\mu\). This is also larger for a population with small \(N\) than a large \(N\). The addition of a mother or a daughter does not make a considerable difference in the estimates when \(\mu\) and \(N\) are large.
Simple population structure

Table 2.10: Examples of change in estimates from small changes in $d$ and $x$

<table>
<thead>
<tr>
<th></th>
<th>N=50</th>
<th>N=500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\mu}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4 0.8 2 5</td>
<td>0.4 0.8 2 5</td>
</tr>
<tr>
<td>$E(n)$</td>
<td>21.0 27.0 45.0 90.0</td>
<td>210.0 270.0 450.0 900.0</td>
</tr>
<tr>
<td>$E(d)$</td>
<td>1.8 3.6 9.0 22.5</td>
<td>18.0 36.0 90.0 225.0</td>
</tr>
<tr>
<td>$E(x)$</td>
<td>1.7 3.2 6.8 11.7</td>
<td>17.0 32.0 67.7 116.5</td>
</tr>
<tr>
<td>$E(\bar{d})$</td>
<td>1.1 1.1 1.3 1.9</td>
<td>1.1 1.1 1.3 1.9</td>
</tr>
<tr>
<td>$n$</td>
<td>21 27 45 90</td>
<td>210 270 450 900</td>
</tr>
<tr>
<td>$d$</td>
<td>2 4 9 23 7</td>
<td>18 36 90 225</td>
</tr>
<tr>
<td>$x$</td>
<td>2 3 7 12</td>
<td>17 32 68 117</td>
</tr>
<tr>
<td>With original $n$, $d$, $x$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tilde{d}$</td>
<td>$\tilde{\mu}$</td>
</tr>
<tr>
<td></td>
<td>1.00 1.33 1.29 1.92</td>
<td>1.06 1.13 1.32 1.92</td>
</tr>
<tr>
<td>$\tilde{\mu}$</td>
<td>0.19 0.61 0.53 1.48</td>
<td>0.12 0.24 0.59 1.49</td>
</tr>
<tr>
<td>$\tilde{\mu}$</td>
<td>0.85 3.09 1.63 4.79</td>
<td>0.35 0.80 1.95 4.96</td>
</tr>
<tr>
<td>$\tilde{N}$</td>
<td>52.2 32.4 51.9 49.7</td>
<td>466.1 498.3 503.3 502.3</td>
</tr>
<tr>
<td>Adding 1 to $d$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{d}$</td>
<td>$\hat{\mu}$</td>
</tr>
<tr>
<td></td>
<td>1.50 1.67 1.43 2.00</td>
<td>1.12 1.16 1.34 1.93</td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>0.87 1.33 0.76 1.59</td>
<td>0.23 0.30 0.61 1.50</td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>5.12 5.08 2.43 4.98</td>
<td>1.51 1.17 2.04 4.98</td>
</tr>
<tr>
<td>$\hat{N}$</td>
<td>19.1 19.3 40.9 46.5</td>
<td>554.0 487.6 490.4 498.7</td>
</tr>
<tr>
<td>Adding 2 to $d$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{d}$</td>
<td>$\hat{\mu}$</td>
</tr>
<tr>
<td></td>
<td>2.00 2.00 1.57 2.08</td>
<td>1.18 1.19 1.35 1.94</td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>1.59 1.59 0.98 1.70</td>
<td>0.33 0.35 0.64 1.51</td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>7.37 6.17 3.03 5.14</td>
<td>2.51 1.52 2.12 5.00</td>
</tr>
<tr>
<td>$\hat{N}$</td>
<td>11.0 14.0 33.6 43.7</td>
<td>447.0 457.7 478.1 495.2</td>
</tr>
<tr>
<td>Adding 1 to $x$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tilde{d}$</td>
<td>$\tilde{x}$</td>
</tr>
<tr>
<td></td>
<td>1.00 1.13 1.77 1.00</td>
<td>1.00 1.09 1.30 1.91</td>
</tr>
<tr>
<td>$\tilde{\mu}$</td>
<td>0.19 0.24 1.71 0.19</td>
<td>0.19 0.18 0.56 1.47</td>
</tr>
<tr>
<td>$\tilde{\mu}$</td>
<td>0.19 0.20 5.71 1.05</td>
<td>0.32 1.79 4.87</td>
</tr>
<tr>
<td>$\tilde{N}$</td>
<td>23.5 30.7 44.2 609.1</td>
<td>373.3 516.6 508.4</td>
</tr>
</tbody>
</table>

The capture probability, $p$, is 0.3 in all calculations.

2.10 Discussion

In this chapter we examine the feasibility of estimating the population size using genotype data from a single capture occasion and show that it can be successfully achieved. We show that for a population consisting of two generations, which we label mothers and daughters, it is indeed possible to estimate the population size. In this process we introduce the mean number of daughters per mother as a nuisance parameter, along with the capture probability. We show that these estimators are statistically consistent and determine the asymptotic variances of the estimators.
We illustrate with simulation studies that our estimator for the primary variable of interest, the number of mothers in the population and the estimators for nuisance parameters, the mean number of daughters per mother and capture probability performs well. The bias is slight and the variances can be well estimated using the parametric bootstrap approach. Consequently, the coverage probability of nominal 95% confidence intervals are satisfying. The standard error of the estimator for the number of mothers in the population can be further reduced by increasing the capture probability. This standard error is considerably large for $p < 0.3$. Therefore we recommend that capture probability is at least 0.3 to use this method.

In our application, a female pygmy possum can reproduce from an age of one year to an age of twelve years with up to four offspring at one time. Breeding usually occurs annually for many consecutive years. Therefore, in a natural population of pygmy possums the mean number of daughters per mother will not be very small. Therefore in this case the assumption $\mu > 1$ seems appropriate. However, this assumption is mostly for the technical reasons we discussed in Section 2.9. Also, for the same reasons, there is high chance the sample consists of mother-daughter pairs from more than two generations. The purpose of the application here is to illustrate the use of the method for a natural data set with genetic information and not to provide a realistic estimate of the size of the *Burramys parvus* populations. There are other populations, such as the North Atlantic Minke Whales population [33], where it can be assumed that no more than two generations are coexisting in the population.

It is not possible to use the method if no mother-daughter pair is captured in the sample. However, as mentioned above, the mean number of daughters per mother can not be very small for pygmy possums. Hence there is very small possibility that no mother-daughter pair is captured in the sample, providing the sample is not too small.

Note that $\hat{\mu}p$ becomes zero when $\bar{d} = 1$ (see (2.11)). Therefore when the number $x$ of mothers for whom at least one daughter is observed is equal to $d$, which is the total number of captured daughters of these mothers, we took $\bar{d} = d/x + 0.1$ in estimating procedure. But as the mean number of daughters per mother is not small for pygmy possums, in our application this adjustment was not required.
We use a false discovery rate (FDR) procedure based on the likelihood ratio (LOD score) [34] to determine the number of mother-daughter pairs. This method uses genetic markers for identification. Whilst this method can involve measurement errors, we consider that mother-daughter pairs are perfectly identifiable in this chapter. The error that can be caused by misclassification is discussed extensively in Chapter 6.

The distribution of $\hat{N}$ is exhibits slight positive skewness with a few larger estimates. The skewness is particularly high when $\mu$ and $p$ is small. Therefore, the standard deviations of the estimators are quite large when $N$, $\mu$ or $p$ is small. The estimates of $\hat{N} + \hat{D}$ have large standard errors, when the capture probability is small. Therefore, we recommend the capture probability is at least $p = 0.5$, for the estimates of $\hat{N} + \hat{D}$ to be useful in practice.
Chapter 3

Estimating the number of mothers in a population with a simple structure using maximum likelihood

3.1 Background

In this chapter we examine the maximum likelihood approach to estimate the number of mothers in the population. We continue to restrict the population structure to two generations and, again, we consider a single sex. In this chapter, we also assume that mother-daughter pairs in the sample are perfectly identifiable.

As in Chapter 2, here also we develop methods to estimate the population size using genotype data from a single capture occasion. The genetic markers would allow us to identify any captured offspring of captured individuals. We treat the captures of the offspring of captured individuals as recaptures and hence estimate the population size from one occasion. In this chapter we do not derive explicit estimators for our parameters, but rather construct a likelihood function, of which the logarithm can be maximised to compute the estimates of the parameters.
I would like to thank A/Prof Rachel Fewster for suggesting the idea of the maximum likelihood estimation for this problem in a personal communication.

### 3.2 Notation and assumptions

We consider a likelihood approach to estimate the size of a population using a single sample from a two generation population consisting only of mothers and their daughters. Whilst we reveal problems with the likelihood approach it is informative and interesting to examine this approach. There are \( N \) mothers in the population and \( \mu \) is the mean number of daughters per mother. Each individual has probability \( p \) of being captured. We treat \( N \), \( \mu \) and \( p \) as parameters. A single sample is observed as a result of a capture experiment. The data consist of the \( n \) individuals captured in total; \( x \) observed mothers each of which has at least one captured daughter; and \( d_1, \ldots, d_x \) which are the numbers of caught daughters of the \( x \) observed mothers. Thus a total of \( d \) daughters is observed, where \( d = \sum_{i=1}^{x} d_i \). Table 3.1 contains a summary of the notation used within this chapter.

We assume that:

1. The number of daughters of a given mother has a Poisson(\( \mu \)) distribution and the number of daughters are independent for different mothers.

2. Mother-daughter pairs can be reliably determined.

3. Each individual in the population has the same probability \( p \) of being captured.

4. Individuals are captured independently of each other.

5. The population considered is presumably closed.
### Table 3.1: Notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Number of captured individuals</td>
</tr>
<tr>
<td>$x$</td>
<td>Number of observed mothers with at least one daughter observed</td>
</tr>
<tr>
<td>$d_i$</td>
<td>Number of caught daughters of $i$th observed mother</td>
</tr>
<tr>
<td>$d$</td>
<td>Number of observed mother-daughter pairs (i.e. Number of observed daughters)</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of mothers in the population</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Mean number of daughters per mother</td>
</tr>
<tr>
<td>$p$</td>
<td>Capture probability</td>
</tr>
<tr>
<td>$C$</td>
<td>Number of the $N$ mothers that were captured (Includes both identified and unidentified mothers)</td>
</tr>
</tbody>
</table>

### 3.3 The estimator of the number of mothers in a population

To construct a likelihood, we introduce a latent variable $C$, which is the number of the $N$ mothers that were captured, i.e. all captured mothers, irrespective of their daughters being captured or not. For a given $C$ there are $x$ observed mothers, $d$ observed daughters of those $x$ mothers, and $n - C - d$ other individuals, who are daughters of the $N - C$ uncaptured mothers. Recall that the observed mothers are the mothers in the sample from whom at least one daughter is captured. The number of captured mothers $C$ is necessarily greater than or equal to $x$, which is the number of captured mothers with at least one daughter captured. The maximum value $C$ can take is the minimum of $N$, the number of mothers in the population and $n - d$, the number of individuals not identified as daughters in the sample. In other words, $x \leq C \leq \min(N, n - d)$. 
Theorem 3.1. The full likelihood function of $N$, $\mu$ and $p$ for the observed data $n$, $x$, $d_1$, $\ldots$, $d_x$, $d$ is

$$L(\mu, p, N) = \sum_{C=x}^{\min(N, n-d)} \binom{N}{C} p^C (1-p)^{N-C} L_C(\mu, p, N),$$

(3.1)

where $L_C$ is the conditional likelihood function for a given $C$ and

$$L_C(\mu, p, N) = \begin{cases} e^{-N \mu p} \frac{(\mu p)^d}{\prod_{i=1}^{C} d_i!} \frac{(N-C)_{N-p}^{n-C-d}}{(n-C-d)!}, & \text{if } x \leq C \leq n-d, \\ 0, & \text{otherwise.} \end{cases}$$

(3.2)

Proof. In this proof $Pr(\cdot|\cdot)$ represents a generic probability function. The full likelihood function of $N$, $\mu$ and $p$ for the observed data $n$, $d_1$, $\ldots$, $d_x$, $x$ is denoted by $L(\mu, p, N)$ and can be obtained using a conditioning argument as follows:

$$L(\mu, p, N) = \sum_{C=0}^{N} Pr(n, d_1, \ldots, d_x, x|C, N, \mu, p) Pr(C|N, p).$$

(3.3)

Here we denote the joint probability function of $(n, d_1, \ldots, d_x, x)$ conditioned on $C, N, \mu$ and $p$ by $Pr(n, d_1, \ldots, d_x, x|C, N, \mu, p)$. This can be derived from

$$Pr(n, d_1, \ldots, d_x, x|C, N, \mu, p) = Pr(n|d_1, \ldots, d_x, x, C, N, \mu, p) Pr(d_1, \ldots, d_x|x, C, N, \mu, p) Pr(x|C, N, \mu, p).$$

(3.4)

Note that, we use $Pr(n, d_1, \ldots, d_x, x|C, N)$ to express $Pr(n = n^0, d_1 = d_1^0, \ldots, d_x = d_x^0, x = x^0|C, N)$, $Pr(n|d_1, \ldots, d_x, x, C, N)$ to express $Pr(n = n^0|d_1 = d_1^0, \ldots, d_x = d_x^0, x = x^0, C, N)$ and so forth.

Let us now derive the three probability functions which constitutes (3.4).

Recall from the proof of (2.3) that the number of observed daughters of a particular mother follows a Poisson($\mu p$) distribution. It can be seen that $n - C - d$ captured individuals must be daughters of the $N-C$ uncaptured mothers. Therefore $n - C - d$ follows a Poisson($((N-C)\mu p)$) distribution conditioned on $d_1, \ldots, d_x, x, C, N, \mu$ and
Maximum likelihood estimators

This gives the first term of (3.4) as

$$Pr(n|d_1, \ldots, d_x, x, C, N, \mu, p) = Pr(n, x, C, N, \mu, p)$$

$$= \begin{cases} Pr(n - C - d|x, C, N, \mu, p), & \text{if } n \geq C + d, \\ 0, & \text{otherwise,} \end{cases}$$

$$= \begin{cases} \frac{(N-C)^{n-C-d}}{(n-C-d)!} e^{-(N-C)\mu p}, & \text{if } n \geq C + d, \\ 0, & \text{otherwise.} \end{cases} \quad (3.5)$$

For the second term of (3.4), note that given $x, C, N, \mu$ and $p$, we have independent observations $d_1, \ldots, d_x$ from a zero truncated Poisson($\mu p$) distribution and hence

$$Pr(d_1, \ldots, d_x|x, C, N, \mu, p) = (Pr(d_i|x, C, N, \mu, p))^x$$

$$= \prod_{i=1}^{x} \frac{(\mu p)^{d_i} e^{-\mu p}}{d_i!(1 - e^{-\mu p})}$$

$$= \frac{(\mu p)^{d_x} e^{-\mu p}}{(1 - e^{-\mu p})^x \prod_{i=1}^{x} d_i!}. \quad (3.6)$$

The probability that a captured mother has at least one daughter captured is the probability of an observation from Poisson($\mu p$) distribution being non-zero, which is $(1 - e^{-\mu p})$. Therefore, $(x|C, N, \mu, p)$ follows a bin($C, 1 - e^{-\mu p}$) distribution. Now, the third term of (3.4) is

$$Pr(x|C, N, \mu, p) = \begin{cases} \binom{C}{x}(1 - e^{-\mu p})^x e^{-\mu p(C-x)}, & \text{if } x \leq C, \\ 0, & \text{otherwise.} \end{cases} \quad (3.7)$$

Recall that $d = \sum_{i=1}^{x} d_i$. Substituting (3.5), (3.6) and (3.7) in (3.4) gives

$$Pr(n, d_1, \ldots, d_x, x|C, N, \mu, p)$$

$$= \begin{cases} e^{-N\mu p} \frac{(\mu p)^d}{\prod_{i=1}^{x} d_i!} \binom{C}{x} \frac{(N-C)^n}{(n-C-d)!} e^{-(N-C)\mu p} \prod_{i=1}^{x} d_i!, & \text{if } (n, d_1, \ldots, d_x, x) \text{ satisfies } x \leq C \leq n - d, \\ 0, & \text{otherwise.} \end{cases} \quad (3.8)$$
Also, the conditional likelihood function for a given $C$ is

$$L_C(\mu, p, N) = Pr(n, d_1, \ldots, d_x, x|C, N),$$

and hence we get (3.2).

Further, $C|N$ follows a bin$(N, p)$ distribution and hence

$$Pr(C|N, p) = \begin{cases} \binom{N}{C} p^C (1 - p)^{N-C}, & \text{if } C \leq N, \\ 0, & \text{otherwise.} \end{cases} \quad (3.9)$$

Now, as we show in (3.3), the summation of the product of (3.8) and (3.9) gives $L(\mu, p, N)$.

This gives the full likelihood stated in (3.1).

The aim is then to estimate $N, \mu$ and $p$ by maximising the log-likelihood obtained from (3.1). The explicit estimators for the parameters can not be determined. We use R software to maximise the log-likelihood function.

### 3.4 Limitations of the maximum likelihood estimators

The likelihood function of $N, \mu$ and $p$, $L(\mu, p, N)$, derived in the previous section has several issues. This motivated us to develop the moment estimators which we presented in Chapter 2. In this section we analyse and discuss these impediments in detail.

#### 3.4.1 Boundary solutions

The likelihood function $L(\mu, p, N)$ is undefined when the capture probability is 1. To see this, consider (3.1). If $p = 1$, then $c = N$. This results in a term $0^0$ in (3.1) which
is not mathematically defined. Thus there are points in the parameter space where the likelihood is not defined. Note that \( C = N \) corresponds to all mothers being captured and is a possible outcome of the experiment and so should be included in the likelihood.

To illustrate the effect of this, we apply a real dataset to the estimators. We consider data on the female mountain pygmy possums (\textit{Burramys parvus}) collected at Mt Higginbotham West, Victoria, Australia between 2010 and 2012. We described this dataset in Section 2.8. This sample of 43 pygmy possums \((n)\) consists of 5 mother-daughter pairs \((d)\) and 4 mothers who had at least one captured daughter \((x)\). The maximum likelihood estimate, computed using the genetic algorithm \texttt{genoud} in R gave \( \hat{N} = 38 \) \((= 43 - 5 = n - d)\), \( \hat{\mu} = 0.13 \) \((= 5/38 = d/(n - d))\) and \( \hat{p} = 1 \). That is, without constraints, the maximum likelihood estimate corresponds to capturing the entire population.

The likelihood surface in this case, i.e. for a sample with \( n=43, d=5 \) and \( x=4 \) when \( N = 38 \) is shown in Figure 3.1. A feature of this figure is the apparent peak that occurs when \( p = 1 \).

Substituting \( C = N \) and \( p = 1 \) in (3.2) gives

\[
L_C(\mu, 1, N) = e^{-N\mu} \frac{\mu^d}{\prod_{i=1}^x d_i} \left( \frac{N}{x} \right) \frac{(N-N)\mu^{n-N-d}}{(n-N-d)!}.
\]

Again since all mothers and daughters are caught, \( N (= c = x) \) will be equal to the number of non-daughters (i.e. \( n - d \)) as well. At this point \((1 - p)^{N-c} \) is \( 0^0 \), which is mathematically undefined. This ultimately leads to an undefined likelihood function at particular values of the parameters. Unfortunately, the \texttt{R} software sets \( 0^0 = 1 \) and computes this contribution to the likelihood which leads to boundary solutions. Taking \( 0^0 = 1 \) the likelihood given by (3.1) when \( p = 1 \) is,

\[
L(\mu, 1, N) = e^{-N\mu} \frac{\mu^d}{\prod_{i=1}^x d_i}.
\]

Taking \( p = 1 \) and \( N = n - d \) and optimising (3.10) gives \( \hat{\mu} = d/(n - d) \). Thus the maximum likelihood estimators at this situation are \( \hat{p} = 1 \), \( \hat{N} = n - d \) and
$\hat{\mu} = d/(n - d)$. Even though we do not show, analytically, that this is a maximum, the data and simulations provide empirical evidence for this. Constraining $\mu$ to be larger than 1 does not resolve this problem and indicates that the likelihood still involves the 0$^0$ terms.

### 3.4.2 Profile likelihood

We employ the profile likelihood technique to visualise the likelihood function to help understand its behaviour. Barndorff-Nielsen and Cox [2] defines the profile
Maximum likelihood estimators

The profile likelihood may be used to perceive and exhibit various features of a full likelihood surface. For example, by plotting the profile likelihood for a variety of choices of the parameter function \( \psi \) the behaviour of the full likelihood can be revealed. Above all, the profile likelihood can be considered as if it was a genuine likelihood to a considerable extent. Especially, the estimate of the maximum profile likelihood coincides with that of the maximum likelihood [2].

In Figure 3.2 we plot the profile likelihood for the dataset discussed above \((n=43, d=5, x=4)\) for \(N\) in 4, \ldots, 50. This profile likelihood function is not smooth and has an apparent jump at \(N = 38\), literally at the value resulted as the maximum likelihood estimate of \(N\). At this point \(\hat{p} = 1\) and \(\hat{\mu} = 0.13\), again the point corresponding to a boundary solution.
We then compute the maximum likelihood estimates for the same sample (n=43, d=5, x=4), under the constraint $\mu > 1$. This gives more reasonable estimates of $\hat{N}=70.8$, $\hat{\mu}=2.77$ and $\hat{p}=0.16$. Recall that for this sample the moment estimates are $\hat{N}=67.8$, $\hat{\mu}=2.99$ and $\hat{p}=0.16$. However, there is no obvious reason to constrain $\mu > 1$. We discuss this in detail in Section 3.4.3.

### 3.4.3 Non-unique solution

The likelihood function, $\mathcal{L}(\theta)$ is maximised at the true value of $\theta$ with probability going to 1, therefore we consider the value of $\theta$ which maximises $\mathcal{L}(\theta)$ as its maximum likelihood estimate [14]. Here $\theta$ needs to be an interior point of the parameter space for the usual theory to apply. But, in this section we show that our likelihood function does not comprise a single maxima and hence fails to produce a unique estimate(s) for the parameter(s).
We consider the same application data set \( (n=43, \, d=5, \, x=4) \) and plot the full likelihood surface for different values of \( N \) to show that this likelihood function has two peaks (Figure 3.3). Here, it is seen that for small values of \( N \), the likelihood surface has two peaks (Figure 3.3 (a) to (d)). The surface for \( \mu < 1 \) it is plotted in black and for \( \mu \geq 1 \) it is plotted in red. It is seen that one of the peaks occur when \( \mu < 1 \) and the other when \( \mu \geq 1 \). However, it is also seen in Figure 3.3 (e) and (f), that for the same data set there is only one peak when \( N \) is large. A solution to this problem could be to constrain \( \mu \geq 1 \) in determining the maximum likelihood estimates. But still the 0th problem remains. Note that \( C = N \) corresponds to capturing all mothers and as this is a feasible outcome of the sampling it seems unreasonable to dismiss this possibility.

### 3.4.4 Computational problems

Computational problems are also associated with this likelihood function. The likelihood function takes a numerically infinitely large value, when either the sample is large (large \( n \)) or when the initial \( N \) or \( \mu \) used for optimising the likelihood function is large or all of them are large. Therefore packages like \( R \), for example, fail to optimise the likelihood function. This is because this likelihood function involves computing factorials and exponents which are huge for large values of \( N \) and \( n \). For instance, the \( R \) computes the likelihood as infinity when \( n=191, \, d=33 \) and \( x=22 \).

We used two different functions \texttt{optim} and \texttt{genoud} in \( R \) software to maximise the log-likelihood function. The \texttt{optim} function often tends to find a local solution near the starting values we provided, hence the estimates often depend on the starting values, whereas the \texttt{genoud} function, when \( \mu \) is unconstrained finds the boundary solutions.
Maximum likelihood estimators

(a) N=40

(b) N=50
Maximum likelihood estimators

(c) $N=60$

(d) $N=68$
Figure 3.3: Likelihood surface for $N = 40, 50, 60, 68$ (moment estimate for this sample), 100 and 150
### 3.4.5 Variance

Simulation studies were conducted to study the standard errors of the maximum likelihood estimates, computed allowing $0^0 = 1$ in the algorithms and constraining $\mu \geq 1$ in the maximisation process. We used the same simulated samples that we used in Section 2.6 to compute the moment estimates. Recall that we have 16 different populations consisting of different numbers of mothers, and different mean numbers of daughters per mother. The maximum likelihood estimates for $\hat{\mu}$, $\hat{p}$ and $\hat{N}$ were calculated for each of the 1000 samples from each of the populations. Also the standard deviation of the estimates was calculated. Their standard errors were estimated by inverting the estimated Hessian matrix at the maximum likelihood estimates, then considering its square-root. Nevertheless, the standard errors computed by inverting the estimated Hessian matrix at the maximum likelihood estimates fail to estimate the true standard deviations universally. For some simulated samples the maximum likelihood estimates could not be computed, since the likelihood function took infinitely large values as described in Section 3.4.4. In Table 3.2, we have considered only the samples which did not have this problem. Note that we do not have estimates for populations with $N = 500$ in Table 3.2, as all 1000 samples we considered resulted in infinitely large values for the likelihood function. Also note that only 3 samples drawn from the population simulated with $N = 200$ and $\mu = 5$ gave a finite value for the likelihood function. From among these samples, some produced Hessian matrices that had negative diagonal values when inverted. But a variance cannot be negative. When this problem occurs, $\hat{\mu}$ is 1, i.e. at this point $\hat{\mu}$ takes the boundary value of $\mu$. Thus, for these samples the standard errors could not be estimated. Note that this does not happen every time $\hat{\mu} = 1$. In Table 3.3 we present only the estimates of the samples for which the standard error could be estimated using the Hessian matrix. The number of simulated samples used for computing the estimates in each setting is mentioned in the first row of each setting as $N_{\text{sim}}$. Table 3.2 and 3.3 display the mean of the estimates, standard deviation of the estimates ($\sigma$) and the mean standard error of estimates ($E(\hat{\sigma})$) for all the populations.

There is a considerable discrepancy between $E(\hat{\sigma})$ and $\sigma$ in all cases, which is particularly noticeable for small values of $\mu$ and $N$. This indicates the failure of the
Hessian matrix approach to successfully estimate the standard deviations of the maximum likelihood estimates. We also considered the median of the estimated standard deviations using Hessian matrix, where it was non-negative in Table 3.3. The medians are more favourable than the means when $N$ and $\mu$ is small, the means are closer otherwise. Further, we compare the standard errors of the maximum likelihood estimates with that of the moment estimates considering the same samples which we present in Chapter 2 (Table 2.4). The variance of estimates computed with the maximum likelihood method are quite close to the variance of estimates computed with the method of moments, sometimes even slightly larger.

Table 3.2: Estimates from the maximum likelihood approach for simulated samples producing only finite values during the optimisation process ($p=0.3$)

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\mu = 1.25$</th>
<th>$\mu = 2$</th>
<th>$\mu = 5$</th>
<th>$\mu = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{sim}$</td>
<td>1000</td>
<td>999</td>
<td>998</td>
<td>993</td>
</tr>
<tr>
<td>$E(\hat{N})$</td>
<td>51.20</td>
<td>51.83</td>
<td>51.96</td>
<td>50.84</td>
</tr>
<tr>
<td>$E(\hat{\mu})$</td>
<td>1.66</td>
<td>1.92</td>
<td>5.02</td>
<td>10.32</td>
</tr>
<tr>
<td>$E(\hat{p})$</td>
<td>0.32</td>
<td>0.35</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>$\sigma_{\hat{N}}$</td>
<td>20.91</td>
<td>19.85</td>
<td>16.17</td>
<td>7.65</td>
</tr>
<tr>
<td>$\sigma_{\hat{\mu}}$</td>
<td>1.68</td>
<td>1.62</td>
<td>1.83</td>
<td>2.71</td>
</tr>
<tr>
<td>$\sigma_{\hat{p}}$</td>
<td>0.13</td>
<td>0.13</td>
<td>0.08</td>
<td>0.07</td>
</tr>
</tbody>
</table>

| $N_{sim}$ | 994 | 999 | 977 | 448 |
| $E(\hat{N})$ | 101.38 | 101.89 | 101.64 | 95.96 |
| $E(\hat{\mu})$ | 1.45 | 2.17 | 5.43 | 9.16 |
| $E(\hat{p})$ | 0.33 | 0.33 | 0.30 | 0.34 |
| $\sigma_{\hat{N}}$ | 24.71 | 25.90 | 18.26 | 7.49 |
| $\sigma_{\hat{\mu}}$ | 0.92 | 1.17 | 1.26 | 1.19 |
| $\sigma_{\hat{p}}$ | 0.10 | 0.09 | 0.06 | 0.04 |

| $N_{sim}$ | 994 | 966 | 3 | 0 |
| $E(\hat{N})$ | 200.88 | 201.73 | 171.67 |
| $E(\hat{\mu})$ | 1.41 | 2.02 | 3.90 |
| $E(\hat{p})$ | 0.31 | 0.32 | 0.43 |
| $\sigma_{\hat{N}}$ | 39.79 | 38.08 | 5.45 |
| $\sigma_{\hat{\mu}}$ | 0.69 | 0.85 | 0.28 |
| $\sigma_{\hat{p}}$ | 0.07 | 0.07 | 0.01 |

$N_{sim}$ is the number of samples producing only finite values during the optimisation process out of 1000 samples. Mean of estimates of $\hat{N}$ is denoted by $E(\hat{N})$, standard deviation of the $\hat{N}$ by $\sigma_{\hat{N}}$ and so forth for $\hat{\mu}$ and $\hat{p}$. 
Table 3.3: Estimates from the maximum likelihood approach for simulated samples producing only finite values during the optimisation process and non-negative diagonal values in the inverted Hessian matrix ($p = 0.3$)

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\mu = 1.25$</th>
<th>$\mu = 2$</th>
<th>$\mu = 5$</th>
<th>$\mu = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{sim}$</td>
<td>294</td>
<td>567</td>
<td>997</td>
<td>993</td>
</tr>
<tr>
<td>$E(\hat{N})$</td>
<td>40.19</td>
<td>46.08</td>
<td>51.86</td>
<td>50.84</td>
</tr>
<tr>
<td>$E(\hat{\mu})$</td>
<td>3.25</td>
<td>2.61</td>
<td>5.03</td>
<td>10.32</td>
</tr>
<tr>
<td>$E(\hat{p})$</td>
<td>0.26</td>
<td>0.32</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>$\sigma_{\hat{N}}$</td>
<td>11.62</td>
<td>13.55</td>
<td>15.86</td>
<td>7.65</td>
</tr>
<tr>
<td>$E(\hat{\sigma}_{\hat{N}})$</td>
<td>22.18</td>
<td>19.42</td>
<td>13.86</td>
<td>7.34</td>
</tr>
<tr>
<td>$M(\hat{\sigma}_{\hat{N}})$</td>
<td>19.63</td>
<td>15.68</td>
<td>11.68</td>
<td>6.77</td>
</tr>
<tr>
<td>$\sigma_{\hat{\mu}}$</td>
<td>2.45</td>
<td>1.87</td>
<td>1.83</td>
<td>2.71</td>
</tr>
<tr>
<td>$E(\hat{\sigma}_{\hat{\mu}})$</td>
<td>4.37</td>
<td>3.57</td>
<td>1.90</td>
<td>2.62</td>
</tr>
<tr>
<td>$M(\hat{\sigma}_{\hat{\mu}})$</td>
<td>3.15</td>
<td>2.47</td>
<td>1.70</td>
<td>2.34</td>
</tr>
<tr>
<td>$\sigma_{\hat{p}}$</td>
<td>0.11</td>
<td>0.12</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>$E(\hat{\sigma}_{\hat{p}})$</td>
<td>0.42</td>
<td>0.46</td>
<td>0.11</td>
<td>0.09</td>
</tr>
<tr>
<td>$M(\hat{\sigma}_{\hat{p}})$</td>
<td>0.20</td>
<td>0.18</td>
<td>0.10</td>
<td>0.09</td>
</tr>
</tbody>
</table>

$N_{sim}$ is the number of samples producing only finite values during the optimisation process and non-negative diagonal values in the inverted Hessian matrix at the maximum likelihood estimates out of 1000 samples. Mean and standard deviation of estimates of $N$ is denoted by $E(\hat{N})$ and $\sigma_{\hat{N}}$, mean and median of the estimated standard deviations (using Hessian matrix) by $E(\hat{\sigma}_{\hat{N}})$ and $M(\hat{\sigma}_{\hat{N}})$ and so forth for $\mu$ and $p$. 
3.5 Application

Here we compute the estimates for the same sites described in Section 2.8. We presented the data in Table 2.8.

The estimates \( \hat{N} \), \( \hat{\mu} \) and \( \hat{p} \) were computed by optimising the logarithm of the likelihood given in (3.1) using the genetic algorithm, genoud in R software. The variances were calculated by inverting the estimated Hessian matrix at the maximum likelihood estimates. The estimates arising from our approach are given in Table 3.4. Here we constrained \( \mu \) to be greater than 1 avoiding the non-unique solutions problem. The value of the likelihood function for the three datasets were finite as \( N \), \( n \) and \( \mu \) were not large. Recall the estimates computed using method of moments (Table 2.9). The estimates from both the methods are quite close to each other.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Site 1 Estimate</th>
<th>SE</th>
<th>Site 2 Estimate</th>
<th>SE</th>
<th>Site 3 Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>70.80</td>
<td>49.78</td>
<td>150.27</td>
<td>46.87</td>
<td>44.45</td>
<td>13.63</td>
</tr>
<tr>
<td>( \mu )</td>
<td>2.77</td>
<td>3.58</td>
<td>1.04</td>
<td>2.81</td>
<td>2.83</td>
<td>1.50</td>
</tr>
<tr>
<td>( p )</td>
<td>0.16</td>
<td>0.10</td>
<td>0.23</td>
<td>0.31</td>
<td>0.33</td>
<td>0.11</td>
</tr>
</tbody>
</table>

3.6 Discussion

We examine the use of the maximum likelihood approach in this chapter. Here also we consider exactly the same setting we considered in Chapter 2. In this chapter also we assume the mother-daughter pairs are perfectly identifiable.

Even though, maximum likelihood estimators are widely used due to their many theoretical advantages, it does not perform well in this problem as we discovered in Section 3.4. The theoretical and computational problems associated with the likelihood approach limit its use in practice. Maximum likelihood estimators usually
have the small variances, but in this problem the moment estimators had similar or sometimes even smaller variances.

Since, moment estimators works better than maximum likelihood estimators in this problem, we do not continue to study the maximum likelihood estimators and use the method of moment for estimating the parameters, in more complicated population structures in Chapters 4 and 5.
Chapter 4

Estimating the size of a population with three generations

4.1 Background

In this chapter we extend our method of moments approach to estimate the number of mothers in the population, from a population of two generations to three generations. As in Chapter 2, we use a method of moments to develop our model resulting in explicit estimators for our population parameters. In this chapter, also, we consider only single sex. We suppose that the observed population consists of an initial number of mothers, their daughters and granddaughters. In other words, each female in the population is a female from the mother generation, daughter generation or granddaughter generation, where the three generations are mutually exclusive and collectively exhaustive. We initially suppose that the mother-daughter pairs and mother-daughter-granddaughter triples are perfectly identifiable in developing our model.

We develop simple method of moments type estimators of the initial number of mothers, the mean number of daughters per mother and the capture probability, i.e. the probability a given individual is included in the sample.
4.2 Notation and assumptions

Our objective is to construct estimators for the size of a population using a single sample from a three generation population consisting only of mothers, their daughters and granddaughters. There are $N$ mothers in the population who have $D$ daughters and $\mu$ is the mean number of daughters per mother. These $D$ daughters become mothers to the next generation having $G$ daughters, i.e. $G$ granddaughters of $N$ mothers. We consider $\mu$ as the mean number of daughters per mother for this generation too. Note that we refer to the first generation of size $N$ as mothers, the second generation of size $D$ as daughters and the third generation of size $G$ as granddaughters. Each individual has probability $p$ of being captured.

We treat $N$, $\mu$ and $p$ as parameters. A single sample is observed as a result of a capture experiment. The data consist of $n$ individuals captured in total; $x$ observed mothers, each of which has at least one captured daughter; $d_1, \ldots, d_x$ which are the numbers of caught daughters of the $x$ observed mothers. Note that the $x$ mothers can either be from the mother generation or the daughter generation and $d$ daughters can either be from the daughter generation or the granddaughter generation. Thus, a total of $d$ daughters is observed, where $d = \sum_{i=1}^{x} d_i$. $g$ is the number of granddaughters whose mother and grandmother is also observed. More specifically, $d$ is the number of observed mother-daughter pairs in the sample and $g$ is the number of observed mother-daughter-granddaughter triples. Our notation is summarised in Table 4.1.

The primary aim is to estimate the parameter $N$ (number of mothers), followed by estimating the two nuisance parameters $\mu$ (mean number of daughters per mother) and $p$ (capture probability).

We assume that:

1. The number of daughters of a given mother has a Poisson($\mu$) distribution, where $\mu \geq 1$ and the number of daughters are independent for different mothers.

2. The mean number of daughters per mother ($\mu$) remains constant over the two generations of mothers.
### Table 4.1: Notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Number of captured individuals</td>
</tr>
<tr>
<td>$x$</td>
<td>Number of observed mothers with at least one daughter observed</td>
</tr>
<tr>
<td>$d$</td>
<td>Number of observed mother-daughter pairs (i.e. Number of observed daughters)</td>
</tr>
<tr>
<td>$g$</td>
<td>Number of observed mother-daughter-granddaughter triples</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of mothers in the population</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Mean number of daughters per mother</td>
</tr>
<tr>
<td>$p$</td>
<td>Capture probability</td>
</tr>
<tr>
<td>$D$</td>
<td>Number of daughters in the population</td>
</tr>
<tr>
<td>$G$</td>
<td>Number of granddaughters in the population</td>
</tr>
</tbody>
</table>

3. Mother-daughter pairs and mother-daughter-granddaughter triples can be reliably determined.

4. Each individual in the population has the same probability $p$ of being captured.

5. Individuals are captured independently of each other.

6. The population considered is presumably closed.

We explain these assumptions in Section 2.2, with the exception of the second assumption. The second assumption seems appropriate for most natural populations. We assume this mostly for technical reasons.

### 4.3 Estimators

We develop models for estimating our parameters using the method of moments approach in this section. Initially we estimate $N$, our primary parameter of interest.
assuming $\mu$ is known. Then we construct estimators for our nuisance parameters $\mu$ and $p$. Prior to showing the construction of these estimators, in Section 4.3.1 we present the derivation of our moment equalities required for the model.

### 4.3.1 Moment equalities

We use the following moments to derive our estimators in the subsequent sections. For random variables $X$ and $Y$, let $X|Y$ denote the conditional distribution of $X$ given $Y$.

**Proposition 4.1.**

\[
\begin{align*}
E(D) &= N\mu \\
E(G) &= N\mu^2 \\
E(n) &= N(1 + \mu + \mu^2)p \\
E(d|x) &= \frac{x\mu p}{1 - e^{-\mu p}} \\
E(x) &= Np(1 + \mu)(1 - e^{-\mu p}) \\
E(d) &= N\mu(1 + \mu)p^2 \\
E(x|n) &= \frac{n(1 + \mu)}{1 + \mu + \mu^2}(1 - e^{-\mu p})
\end{align*}
\]

**Proof.**

Proofs of (4.1) and (4.4) are similar to the proofs of (2.1) and (2.3) respectively, and hence not presented here.

**Proof of (4.2)**

Let $G_i$ be the number of daughters of mother $i$, $i = 1, \ldots, D$. Assumption 1 yields that $G_i \sim \text{Poisson}(\mu)$ and hence letting $G = \sum_{i=1}^{D} G_i$, $G|D \sim \text{Poisson}(D\mu)$, giving $E(G|D) = D\mu$. $E(G) = E(E(G|D))$ resulting (4.2).

**Proof of (4.3)**

To obtain (4.3) note that $n|(D,G) \sim \text{bin}(N + D + G, p)$. Therefore, $E(n) = E(E(n|(D,G)) = E((N + D + G)p)$. Now, (4.3) follows from this, (4.1) and (4.2).
Proof of (4.5)
For a mother to be identified she must be captured along with at least one of her daughters. Recall that the number of observed daughters of a particular mother follows a Poisson($\mu p$) distribution. Hence, the probability a mother from $N + D$ mothers is captured along with at least one of her daughters is $p(1 - e^{-\mu p})$. Thus $x \sim \text{bin}(N + D, p(1 - e^{-\mu p}))$. Using (4.1) this can be approximated by $x \sim \text{bin}(N(1 + \mu), p(1 - e^{-\mu p}))$ giving (4.5).

Proof of (4.6)
It is known that, $E(d) = E(E(d|x))$. This gives (4.6).

Proof of (4.7)
For this, consider three events; $I$, the event an individual is identified as a mother; $M$, the event an individual is a mother; $C$, the event an individual is captured. Then the probability an individual is a mother and identified given that the individual is captured is

$$Pr(I \cap M | C) = \frac{Pr(I \cap M \cap C)}{Pr(C)} = \frac{Pr(I | M \cap C)Pr(M \cap C)}{Pr(C)}$$

But the events that an individual is a mother and an individual is captured, are independent and hence $Pr(M \cap C) = Pr(M)Pr(C)$. Now,

$$Pr(I \cap M | C) = Pr(I | M \cap C)Pr(M)$$

Now, the $x$ identified mothers, given that $n$ individuals are captured is

$$E(x|n) = \sum_{i=1}^{n} Pr(i^{th} \text{ individual is a captured mother}|i^{th} \text{ individual is captured})$$

$$= \sum_{i=1}^{n} Pr(I | M \cap C)Pr(M)$$
Three generations

Same argument can be used to get an expression for $E(x|n, D, G)$ as $\sum_{i=1}^{n} Pr(I|M \cap C, n, D, G)Pr(M|n, D, G)$. Therefore,

$$E(x|n) = E(E(x|n, D, G))$$
$$= E \left( \sum_{i=1}^{n} Pr(I|M \cap C, n, D, G)Pr(M|n, D, G) \right).$$

The probability that a captured mother is identified, $Pr(I|M \cap C, n, D, G)$ is the same as the probability that at least one of her daughters is captured. This probability is $(1 - e^{-\mu p})$ since $d_i$ follows a Poisson($\mu p$) distribution. Also, $Pr(M|n, D, G) = (N + D)/(N + D + G)$. Therefore,

$$E(x|n) = E \left( n(1 - e^{-\mu p}) \frac{N + D}{N + D + G} \right)$$
$$= n(1 - e^{-\mu p})E \left( \frac{N + D}{N + D + G} \right).$$

Substituting (4.1) and (4.2) here gives (4.7).

4.3.2 Mean number of daughters per mother is known

In the first phase of developing the model we assume that the mean number daughters per mother is known. We construct two estimators for $N$: $\tilde{N}(\mu)$ considering only the number of mother-daughter pairs in the sample and $\tilde{N}_g(\mu)$ considering the mother-daughter-granddaughter triples. Recall that we assume that $d, x$ and $g$ are perfectly identified.

**Theorem 4.2.** Under the assumptions stated in Section 4.2, the number of mothers in the population ($N$), whose mean number of daughters ($\mu$) is known can be estimated by,

$$\tilde{N}(\mu) = \frac{n(n-1)(\mu + \mu^2)}{d(1 + \mu + \mu^2)^2} \quad (4.8)$$

and

$$\tilde{N}_g(\mu) = \sqrt{\frac{n(n-1)(n-2)\mu^2}{g(1 + \mu + \mu^2)^3}}. \quad (4.9)$$

Here it is assumed that $d, x$ and $g$ are perfectly identified in the sample of size $n$. 
Proof. The total number of possible pairs in the population is \( \binom{N+D+G}{2} \). The total mother-daughter pairs in the population is comprised of the mother-daughter and daughter-granddaughter pairs. Therefore, the number of mother-daughter pairs in the population is equal to the sum of the number of daughters and granddaughters in the population \((D + G)\). Hence the proportion of all possible pairs that are mother-daughter pairs is

\[
\frac{D + G}{\binom{N+D+G}{2}} = \frac{2(D + G)}{(N + D + G)(N + D + G - 1)}
\approx \frac{2(D + G)}{(N + D + G)^2}; \text{ for large } N
\]

Using (4.1), (4.2) and law of large numbers,

\[
\frac{2(D + G)}{(N + D + G)^2} - \frac{2\mu(1 + \mu)}{N(1 + \mu + \mu^2)^2} \xrightarrow{p} 0.
\]

Next, the proportion of all possible pairs in the sample that are mother-daughter pairs is

\[
\frac{d}{\binom{n}{2}} = \frac{2d}{n(n-1)}
\]

Then, setting

\[
\frac{2\mu(1 + \mu)}{N(1 + \mu + \mu^2)^2} = \frac{2d}{n(n-1)}
\]

yields (4.8).

Moreover, the total number of possible triples in the population is \( \binom{N+D+G}{3} \). Also, there are \( G \) number of mother-daughter-granddaughter triples in the population. Hence, the proportion of all possible triples that are mother-daughter-granddaughter triples is

\[
\frac{6G}{\binom{N+D+G}{3}} = \frac{6G}{(N + D + G)(N + D + G - 1)(N + D + G - 2)}
\approx \frac{6G}{(N + D + G)^3}; \text{ for large } N.
\]
With (4.1) and (4.2) and law of large numbers, \( D \approx N\mu \) and \( G \approx N\mu (1 + \mu) \). Therefore,

\[
\frac{6G}{(N + D + G)^3} - \frac{6\mu^2}{N^2(1 + \mu + \mu^2)^3} \xrightarrow{p} 0.
\]

Next, the proportion of all possible triples in the sample that are mother-daughter-granddaughter triples is

\[
\frac{g}{\binom{n}{3}} = \frac{6g}{n(n-1)(n-2)}.
\]

Then by setting

\[
\frac{6\mu^2}{N^2(1 + \mu + \mu^2)^3} = \frac{6g}{n(n-1)(n-2)}
\]

we obtain the estimator (4.9).

\[\square\]

### 4.3.3 Mean number of daughters per mother is unknown

The mean number of daughters per mother is unknown for most natural populations. The estimator we developed in Section 4.3.2 requires knowledge of \( \mu \). In this section, following Section 2.3.3, we adopt an estimator for the nuisance parameter \( \mu \) using method of moments and subsequently derive \( \hat{N} \) and \( \hat{N}_g \) following (4.8) and (4.9).

**Theorem 4.3.** Under the assumptions stated in Section 4.2, the number of mothers in the population \( N \), whose mean number of daughters (\( \mu \)) is unknown can be estimated by

\[
\hat{N} = \frac{n(n-1)(\hat{\mu} + \hat{\mu}^2)}{d(1 + \hat{\mu} + \hat{\mu}^2)^2}, \quad \text{(4.10)}
\]

and

\[
\hat{N}_g = \frac{n(n-1)(n-2)\hat{\mu}^2}{g(1 + \hat{\mu} + \hat{\mu}^2)^3}. \quad \text{(4.11)}
\]

Here

\[
\hat{\mu} = \begin{cases} 
\frac{n(1-e^{-\hat{\mu}p}) - x + \sqrt{(n(1-e^{-\hat{\mu}p}) - x)(n(1-e^{-\hat{\mu}p}) + 3x)}}{2x}, & \text{if } n(1-e^{-\hat{\mu}p}) \geq 1.5x, \\
1, & \text{otherwise}.
\end{cases} \quad \text{(4.12)}
\]
where,
\[ \hat{\mu}p = \bar{d} + \text{Lambert} \left( \frac{-\bar{d}}{e^d} \right) \]  
(4.13)

with ‘Lambert’ being (the principal branch of) Lambert’s W function and \( \bar{d} = d/x \) if \( d > x \) and \( \bar{d} = d/x + 0.1 \) if \( d = x \). Then the capture probability, \( p \), can be estimated by
\[ \hat{p} = \frac{\hat{\mu}p}{\hat{\mu}}. \]  
(4.14)

Note that it is assumed \( d, x \) and \( y \) are perfectly identified in the sample of size \( n \).

Proof. We derive \( \hat{\mu}p \) similar to the Section 2.3.3.

Now, applying the method of moments to (4.7) yields the equation,
\[ \frac{x}{n} = \frac{1 + \mu}{1 + \mu + \mu^2}(1 - e^{-\hat{\mu}p}), \]
which motivates us to take
\[ \hat{\mu} = \frac{n(1 - e^{-\hat{\mu}p}) - x + \sqrt{(n(1 - e^{-\hat{\mu}p}) - x)(n(1 - e^{-\hat{\mu}p}) + 3x)}}{2x}. \]  
(4.15)

Note that when \( n(1 - e^{-\hat{\mu}p}) < 1.5x \), \( \hat{\mu} \) takes a value less than 1. This arises if there is a small number of daughters per mother within the selected sample (i.e. \( \bar{d} \approx 1 \)) leading to a very small estimate \( \hat{\mu}p \). We have assumed that \( \mu \geq 1 \) in assumption 1. Therefore, in this situation we take \( \hat{\mu} \) to be 1, giving (4.12).

Substituting (4.12) in (4.8) and (4.9) gives (4.10) and (4.11). Further (4.12) and (4.13) results in (4.14).

Finally using (4.1) and (4.2), the size of the total female population, i.e. the total number of mothers, daughters and granddaughters in the population can be estimated by,
\[ \hat{N} + \hat{D} + \hat{G} = \hat{N}(1 + \hat{\mu} + \hat{\mu}^2). \]  
(4.16)
Remark: (4.10) suggests a generalisation of an estimator of the number of mothers in a population known to have $z(\geq 2)$ generations of

$$\hat{N}_z = \frac{n(n-1)\sum_{j=1}^{z-1} \hat{\mu}^j}{d(1 + \sum_{j=1}^{z-1} \hat{\mu})^2}.$$  

Here, $\hat{\mu}$ can be obtained by solving,

$$\frac{x}{n} = \frac{\sum_{j=1}^{z-1} \hat{\mu}^j - 1}{\sum_{j=1}^{z} \hat{\mu}^j - 1} (1 - e^{-\hat{\mu}p})$$

where,

$$\hat{\mu}p = \bar{d} + \text{Lambert} \left( \frac{-\bar{d}}{e^x} \right).$$

### 4.4 Estimating $N$ when the number of mother-daughter pairs in the sample is unknown

In developing estimators (4.10), (4.11), (4.8) and (4.9) we have considered that the mother-daughter pairs and mother-daughter-granddaughter triples in the sample are perfectly identifiable. This is not usually known and hence needs to be estimated in most practical situations. In Chapter 2 we explain the method we used for determining the number of mother-daughter pairs in a sample following Skaug et al. [34].

Identifying the mother-daughter-granddaughter triples in a sample is intricate. Recognition of two mother-daughter pairs with one individual in common raises the need to distinguish between whether they are two daughters of one mother or a mother-daughter-granddaughter triple. We use a method based on a kinship coefficient for determining the related pairs [34] in the sample using the genetic markers. The kinship coefficient takes the value $1/8$ for both the grandmother-granddaughter and half sister relationships making it difficult to distinguish between whether they are two daughters of one mother or a mother-daughter-granddaughter triple. Therefore, additional information is required to properly identify mother-daughter-granddaughter
triples in the sample. Note that the procedure we use for identifying mother-daughter pairs does not allow us to particularly distinguish between mothers and daughters, rather than identifying them as pair. This can be resolved if other variables such as age is available for the individuals, producing more information on the relationships.

This makes the use of (4.11) restricted. For this reason and justifications we show in Section 4.6.1 that we recommend using (4.10) to estimate the size of a population consisting of three generations. Hence, we show the consistency and conduct most of the simulation studies, except the comparison of (4.11) and (4.10) is done only for (4.10).

The statistical consistency of our estimators follows as in Section 2.4, and is presented in Appendix 4.A.

4.5 Estimator of the standard error

A parametric bootstrap method as described in Section 2.5.2 was, again, used for estimating the standard error of the proposed estimator for the number of mothers in the population.

Our parametric bootstrap procedure begins by computing \( \hat{N}, \hat{\mu} \) and \( \hat{p} \) using the proposed estimators (4.10), (4.12) and (4.14) for the captured sample. Then a population with \( \hat{N} \) number of mothers is simulated such that the mean number of daughters per mother is equal to \( \hat{\mu} \). One thousand bootstrap samples are drawn with a capture probability \( \hat{p} \) from this simulated population and the estimators \( \hat{N}_B, \hat{\mu}_B \) and \( \hat{p}_B \) are computed for the 1000 bootstrap samples. The standard deviations of the estimates are calculated as the bootstrap estimates of the standard deviations of the estimates.

We compute the confidence intervals (95%) of the estimates, with bootstrap estimates of the standard deviation and normal quantiles. For example, 95% confidence interval of \( \hat{N} \) is computed as \( \hat{N} \pm 1.96\hat{\sigma}_{\hat{N}(B)} \), where \( \hat{\sigma}_{\hat{N}(B)} \) is the bootstrap estimates of the standard deviation of \( \hat{N} \).
4.6 Simulation studies

Simulation studies were conducted such that the assumptions mentioned in Section 4.2 are satisfied, except for the third assumption. In particular, we assumed that mother-daughter pairs and mother-daughter-granddaughter triples are perfectly identifiable (instead of reliably identifiable) in the simulation studies.

We follow the approach of Section 2.6 with the obvious changes in simulating the populations:

1. We simulate only three generations at a time to comply with the population structure we consider.

2. The daughter population was generated assuming that the number of daughters of a given mother are independent and have a Poisson($\mu$) distribution. The granddaughter population was generated considering the daughter population as mothers and then assuming that the number of daughters of each of them is independent and has a Poisson($\mu$) distribution where this $\mu$ is the same as the $\mu$ in their mother’s generation.

4.6.1 Comparison of the two estimators of $N$

In Section 4.3.3, we introduced two estimators (4.10) and (4.11) for $N$. Here, we compare the two estimators with simulated populations consisting of three generations, of which the mother-daughter pairs and mother-daughter-granddaughter triples can be perfectly identified.

We considered populations consisting of $N=50$ and $N=200$ mothers, with $\mu = 1.25, 2, 5$ and $10$ for comparison. Then, 1000 samples were drawn with a capture probability of $0.3$, and $\hat{N}$ and $\hat{N}_g$ were computed for all samples using (4.10) and (4.11). The mean of the 1000 estimates were computed as $E(\hat{N})$ and $E(\hat{N}_g)$ respectively. These are plotted in Figure 4.1. In terms of bias, $E(\hat{N})$ showed better performance than $E(\hat{N}_g)$. 
The two estimators \( \hat{N} \) and \( \hat{N}_g \) were then compared based on mean squared error. Mean squared error was computed as the summation of squared bias and variance of the estimator.

\[
MSE(\hat{N}) = Bias(\hat{N})^2 + Var(\hat{N})
\]

Simulation studies show that \( \hat{N}_g \) has a higher mean squared error compared to \( \hat{N} \) (Table 4.2).

Recall that we discussed the limitation of using (4.11) in practice due to the difficulty in determining the number of mother-daughter-granddaughter triples in the sample using genotype data in Section 4.4. For convenience of securing necessary information and for better performance, we continue our inferences with \( \hat{N} \) given in (4.10).
Table 4.2: Mean squared errors of $\hat{N}$ and $\hat{N}_g$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\mu = 1.25$</th>
<th>$\mu = 2$</th>
<th>$\mu = 5$</th>
<th>$\mu = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>506.4</td>
<td>648.9</td>
<td>70.6</td>
<td>15.6</td>
</tr>
<tr>
<td></td>
<td>673.7</td>
<td>1081.6</td>
<td>141.5</td>
<td>75.2</td>
</tr>
<tr>
<td>200</td>
<td>2633.5</td>
<td>2948.3</td>
<td>294.3</td>
<td>66.5</td>
</tr>
<tr>
<td></td>
<td>4210.7</td>
<td>3645.4</td>
<td>491.4</td>
<td>191.3</td>
</tr>
</tbody>
</table>

4.6.2 Performance of $\hat{N}$

Recall that we have considered 16 different populations with varying $N$ and $\mu$ as described at the start of Section 4.6. One thousand samples were drawn from each population. Then, our parameter estimates, $\hat{N}$, $\hat{\mu}$ and $\hat{p}$ were computed for 1000 samples using (4.10), (4.12) and (4.14) respectively. We present our estimates of $N$, $\mu$ and $p$, which are the means of 1000 estimates (using 1000 samples) computed for a single population in Table 4.3.

In Table 4.4 we present the true and estimated standard deviations of $\hat{N}$, $\hat{\mu}$ and $\hat{p}$. True standard deviation is the standard deviation of 1000 estimates resulting from 1000 samples. A population was simulated with $\hat{N}$, $\hat{\mu}$ and $\hat{p}$ of each sample. One thousand bootstrap samples were drawn from this population. The standard deviation of the 1000 bootstrap estimates of $N$ was computed as the bootstrap estimate of the standard deviation. The mean of 1000 bootstrap estimates of the standard deviation is considered as the estimated standard deviation of $\hat{N}$. The estimated standard deviations of $\hat{N}$ are, in general, reasonable if perhaps slightly conservative. The standard deviations of $\hat{\mu}$ and $\hat{p}$ were also estimated similarly. These estimates were very close to the true standard deviations, except slightly overestimating the standard deviation for small values of $N$ and $\mu$.

Figure 4.2 shows the distribution of $\hat{N}$ for populations of different sizes with varying $\mu$. The dotted line shows the true value of $N$. In Figure 4.2 we also report the percentage bias. In general, the bias of $\hat{N}$ is small. The numbers below each boxplot show the mean of the parametric bootstrap estimates of the standard errors and normal quantiles. Confidence intervals (95%) of the estimates were computed with
Three generations

Table 4.3: Averages of 1000 estimates of the population parameters

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\mu = 1.25$</th>
<th>$\mu = 2$</th>
<th>$\mu = 5$</th>
<th>$\mu = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\text{mean}(\hat{N})$</td>
<td>49.55</td>
<td>52.12</td>
<td>49.58</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{\mu})$</td>
<td>1.56</td>
<td>2.15</td>
<td>5.22</td>
</tr>
<tr>
<td>50</td>
<td>$\text{mean}(\hat{p})$</td>
<td>0.22</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{N} + \hat{D} + \hat{G})$</td>
<td>209.51</td>
<td>344.05</td>
<td>1628.33</td>
</tr>
<tr>
<td></td>
<td>$E(N + D + G)$</td>
<td>190.625</td>
<td>350</td>
<td>1550</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{N})$</td>
<td>101.42</td>
<td>101.51</td>
<td>100.61</td>
</tr>
<tr>
<td>100</td>
<td>$\text{mean}(\hat{\mu})$</td>
<td>1.38</td>
<td>2.15</td>
<td>5.07</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{p})$</td>
<td>0.26</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{N} + \hat{D} + \hat{G})$</td>
<td>403.00</td>
<td>723.64</td>
<td>3174.11</td>
</tr>
<tr>
<td></td>
<td>$E(N + D + G)$</td>
<td>381.25</td>
<td>700</td>
<td>3100</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{N})$</td>
<td>200.25</td>
<td>200.57</td>
<td>201.30</td>
</tr>
<tr>
<td>200</td>
<td>$\text{mean}(\hat{\mu})$</td>
<td>1.39</td>
<td>2.03</td>
<td>5.05</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{p})$</td>
<td>0.28</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{N} + \hat{D} + \hat{G})$</td>
<td>821.46</td>
<td>1365.75</td>
<td>6314.09</td>
</tr>
<tr>
<td></td>
<td>$E(N + D + G)$</td>
<td>762.5</td>
<td>1400</td>
<td>6200</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{N})$</td>
<td>497.84</td>
<td>498.27</td>
<td>500.00</td>
</tr>
<tr>
<td>500</td>
<td>$\text{mean}(\hat{\mu})$</td>
<td>1.26</td>
<td>2.06</td>
<td>5.02</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{p})$</td>
<td>0.29</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{N} + \hat{D} + \hat{G})$</td>
<td>1880.46</td>
<td>3566.80</td>
<td>15559.57</td>
</tr>
<tr>
<td></td>
<td>$E(N + D + G)$</td>
<td>1906.25</td>
<td>3500</td>
<td>15500</td>
</tr>
</tbody>
</table>

The true capture probability, $p$, is 0.3 here. The true values of $N$ and $\mu$ are as given in the row and column headers. The bootstrap estimates of the standard deviation. Then the coverage probability, i.e. the proportion of the occurrences that the confidence interval contains the true value of the parameter was calculated. We present these coverage probability of nominal 95% confidence intervals for $N$ in Figure 4.2 below the boxplot.

The same information for the distribution of $\hat{\mu}$ and $\hat{p}$ is shown respectively in Figure 4.3 and Figure 4.4. It can be seen that $\mu$ is estimated satisfactorily, except there is a considerable bias for populations of size 50 and 100 when $\mu = 1.25$. Further, $\hat{p}$ is also satisfying, except there is considerable bias at $\mu=1.25$. 


Table 4.4: Estimated standard errors

<table>
<thead>
<tr>
<th>N</th>
<th>$\mu = 1.25$</th>
<th>$\mu = 2$</th>
<th>$\mu = 5$</th>
<th>$\mu = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma_N$</td>
<td>22.50</td>
<td>25.38</td>
<td>8.39</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{\sigma}_N(B))$</td>
<td>24.97</td>
<td>29.03</td>
<td>8.63</td>
</tr>
<tr>
<td>50</td>
<td>$\sigma_\mu$</td>
<td>0.90</td>
<td>0.87</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{\sigma}_\mu(B))$</td>
<td>1.31</td>
<td>1.06</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>$\sigma_p$</td>
<td>0.12</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{\sigma}_p(B))$</td>
<td>0.10</td>
<td>0.08</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>$\sigma_N$</td>
<td>32.34</td>
<td>36.56</td>
<td>11.74</td>
</tr>
<tr>
<td>100</td>
<td>$\text{mean}(\hat{\sigma}_N(B))$</td>
<td>43.18</td>
<td>37.17</td>
<td>12.15</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\mu$</td>
<td>0.59</td>
<td>0.59</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{\sigma}_\mu(B))$</td>
<td>1.05</td>
<td>0.63</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>$\sigma_p$</td>
<td>0.08</td>
<td>0.05</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{\sigma}_p(B))$</td>
<td>0.08</td>
<td>0.05</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>$\sigma_N$</td>
<td>51.32</td>
<td>54.30</td>
<td>17.11</td>
</tr>
<tr>
<td>200</td>
<td>$\text{mean}(\hat{\sigma}_N(B))$</td>
<td>57.92</td>
<td>53.73</td>
<td>17.09</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\mu$</td>
<td>0.46</td>
<td>0.46</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{\sigma}_\mu(B))$</td>
<td>0.58</td>
<td>0.44</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>$\sigma_p$</td>
<td>0.05</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{\sigma}_p(B))$</td>
<td>0.06</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>$\sigma_N$</td>
<td>87.32</td>
<td>80.02</td>
<td>26.46</td>
</tr>
<tr>
<td>500</td>
<td>$\text{mean}(\hat{\sigma}_N(B))$</td>
<td>93.05</td>
<td>84.81</td>
<td>26.62</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\mu$</td>
<td>0.29</td>
<td>0.27</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{\sigma}_\mu(B))$</td>
<td>0.33</td>
<td>0.28</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>$\sigma_p$</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>$\text{mean}(\hat{\sigma}_p(B))$</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
</tr>
</tbody>
</table>

The standard deviation of the 1000 estimates of $N$ is denoted by $\sigma_N$ and the mean of the 1000 bootstrap estimates of standard deviations is denoted by $\text{mean}(\hat{\sigma}_N(B))$ and so forth for $\mu$ and $p$. 
These are based on the 1000 simulations under scenarios (a) $N=50$ (b) $N=100$ (c) $N=200$ (d) $N=500$ with $p=0.3$. The numbers above each boxplot show percentage bias and the numbers below show the mean of standard errors determined using 1000 bootstrap samples ($E(\hat{\sigma})$) and the coverage probability of nominal 95% confidence intervals ($\Phi$).
These are based on the 1000 simulations under scenarios (a) $\mu=1.25$ (b) $\mu=2$ (c) $\mu=5$ (d) $\mu=10$ with $p=0.3$. The numbers above each boxplot show percentage bias and the numbers below show the mean of standard errors determined using 1000 bootstrap samples ($E(\hat{\sigma})$) and the coverage probability of nominal 95% confidence intervals ($\Phi$).
These are based on the 1000 simulations under scenarios (a) $N=50$ (b) $N=100$ (c) $N=200$ (d) $N=500$ with $p=0.3$. The numbers above each boxplot show percentage bias and the numbers below show the mean of standard errors determined using 1000 bootstrap samples ($E(\hat{\sigma})$) and the coverage probability of nominal 95% confidence intervals ($\Phi$).
The effect of capture probability was observed with simulation studies with higher capture probability ($p=0.5$). Simulations were re-run for the three population settings below:

1. $N=50$ and $\mu=2$
2. $N=100$ and $\mu=5$
3. $N=200$ and $\mu=2$

The precision was markedly increased when $p=0.5$, compared with $p=0.3$ (Table 4.5). The standard errors are now reduced significantly and deliver good coverage probabilities.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$N=50$, $\mu=2$</th>
<th>$N=100$, $\mu=5$</th>
<th>$N=200$, $\mu=2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>50.59 12.23</td>
<td>100.05 5.41</td>
<td>200.84 22.56</td>
</tr>
<tr>
<td>$\mu$</td>
<td>2.00 0.40</td>
<td>5.04 0.22</td>
<td>2.05 0.19</td>
</tr>
<tr>
<td>$p$</td>
<td>0.50 0.06</td>
<td>0.50 0.02</td>
<td>0.50 0.03</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>93%</td>
<td>94%</td>
<td>94%</td>
</tr>
</tbody>
</table>

The expected total female population, $E(N+D+G)$, was computed following (4.16). Figure 4.5 shows that the bias percentage of $\hat{N}+\hat{D}+\hat{G}$ is quite small. When either $N$ or $\mu$ is large, this is almost negligible. The standard errors of $\hat{N}+\hat{D}+\hat{G}$ are often too large to be useful in practice, hence we do not report them in detail here. For example, with $N=50$ and $\mu=2$ and $p=0.3$, the standard deviation of $\hat{N}+\hat{D}+\hat{G}$ was 84.7 with a bootstrap estimate of 96.8 (the average of the estimates $\hat{N}+\hat{D}+\hat{G}$ was 344) and with $N=200$, $\mu=2$ and $p=0.3$ the standard deviation of $\hat{N}+\hat{D}+\hat{G}$ was 159.4 with a bootstrap estimate of 131.5 (the average of the estimates $\hat{N}+\hat{D}+\hat{G}$ was 1365.7). A higher capture probability reduces the standard errors of $\hat{N}+\hat{D}+\hat{G}$. We considered the same two populations, but with $p=0.5$. For the population of $N=50$, the standard deviation of $\hat{N}+\hat{D}+\hat{G}$ was 40.7 with a bootstrap estimate of 43.8 (the average of the estimates $\hat{N}+\hat{D}+\hat{G}$ was 340) and for the population of
\[ N = 200 \] the standard deviation of \( \hat{N} + \hat{D} + \hat{G} \) was 85.0 with a bootstrap estimate of 84.3 (the average of the estimates \( \hat{N} + \hat{D} + \hat{G} \) was 1441).

![Graph showing percentage bias of \( \hat{N} + \hat{D} \) for different values of \( \mu \).](image)

**Figure 4.5:** Percentage bias of \( \hat{N} + \hat{D} \)

### 4.7 Application

We apply the developed model to data on the female mountain pygmy possums (*Burramus parvus*) collected at several sites at Mt Hotham, Victoria, Australia between 2010 and 2012 during their breeding season in October and November. This is the same dataset that we used in Chapter 2, which was described in Section 2.8. Here we assume that the population consist of three generations. The captured possums were genotyped at 21 loci. ‘Site A’ consists of the possums captured in Mt Higginbotham West, ‘Site B’ Mt Higginbotham (both East and West), whereas ‘Site C’ consists of those at the Mt Little Higginbotham site. We presented the data in Table 2.8.

Here the number of mother-daughter pairs in the sample has to be determined in order to use the proposed estimators to estimate \( N \). We utilise the LOD score
based method described in Section 4.4. We control the proportion of false positives at a maximum of 0.05. The standard errors of the estimators are computed using the parametric bootstrap method described in Section 4.5. The estimates resulting from our approach are given in Table 4.6. The standard errors are quite high, reflecting the small estimated capture probabilities and the fairly small population sizes. However, these sample sizes are quite small. Note that Skaug [33] regards 334 as a small sample size.

Table 4.6: Parameter estimates for the three sites

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Site 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate   SE</td>
<td>Estimate   SE</td>
<td>Estimate   SE</td>
</tr>
<tr>
<td>(N)</td>
<td>17.92     47.50</td>
<td>71.14     84.29</td>
<td>11.75     23.33</td>
</tr>
<tr>
<td>(\mu)</td>
<td>3.78      3.02</td>
<td>2.06      2.64</td>
<td>3.70      2.01</td>
</tr>
<tr>
<td>(p)</td>
<td>0.12      0.09</td>
<td>0.13      0.08</td>
<td>0.26      0.09</td>
</tr>
</tbody>
</table>

### 4.8 Discussion

In this chapter we show that the method of moment estimator we proposed in Chapter 2 for a population with two generations can be extended to a population with more than two generations. We illustrate, with simulation studies, that this works well for a population with three generations. We also propose a generalisation for a population with any known number of generations. The standard errors of the estimator of number of mothers in the population are estimated with a parametric bootstrap approach. Theses estimates are, in general, reasonable if perhaps slightly conservative. The coverage probability of nominal 95% confidence intervals are also satisfying.

For a population with three generations we propose two estimators: first using only mother-daughter pairs and the second using mother-daughter-granddaughter triples. For the reasons we mention in Section 4.4, deducing mother-daughter-granddaughter triples is intricate. Therefore, the use of estimator, which requires the number of mother-daughter-granddaughter triples, is limited. Hence, and also for better
performance as we discussed in Section 4.6.1, we recommend using the estimator that requires only mother-daughter pairs unless additional information is available to identify mother-daughter-granddaughter triples.

We have assumed that $\mu > 1$ in developing our model. As described in Chapter 2, this assumption seems appropriate for most natural populations. Particularly, a female pygmy possum can reproduce from the age of one year to twelve years of age with up to four offspring at a time. Breeding usually occurs annually for many consecutive years. Therefore, in a natural population of pygmy possums the mean number of daughters per mother can not be very small. However, as described in Section 2.9, we make this assumption mainly due to technical requirements.

Similar to Chapter 2, we need to observe at least one mother-daughter pair to use these estimators. In practice, however, as the mean number of daughters per mother can not be very small for pygmy possums, it is highly probable that at least one mother-daughter pair is captured in the sample.

As seen in (4.13), $\hat{\mu}p$ becomes zero when $\bar{d} = 1$. Therefore, as in the previous chapter, we consider $\bar{d} = d/x + 0.1$ instead of $\bar{d} = d/x$ when $d=x$ in estimating procedure. However, the chances are low that $d=x$ when $\mu$ and $p$ is not very small.

We use a false discovery rate (FDR) procedure based on the likelihood ratio (LOD score) [34] to determine the number of mother-daughter pairs. The accuracy of this procedure is discussed in Chapter 6.

The estimators has considerably large standard deviations, especially with small $N$, small $\mu$ and small $p$. The distribution of $\hat{N}$ is positively skewed with few very large estimates. The skewness is particularly high when $\mu$ and $p$ is small, leading to higher standard errors.
Appendix

4. A Consistency of the estimators

In this section we show that the two estimators we proposed for $N$, (4.8) and (4.10), the estimator of $\mu$, (4.12), and the estimator of $p$, (4.14), are statistically consistent.

Theorem 4.4. Consider a sequence $j = 1, 2, \ldots$ of populations of size $N^{(j)}$ with $N^{(j)} \to \infty$. Let $n^{(j)}$ and $d^{(j)}$ be $n$ and $d$ defined for the $j$th population and similarly for $\tilde{N}^{(j)}(\mu)$ and $\hat{N}^{(j)}$. Suppose that $\mu$ and $p$ do not depend on $j$. Under the assumptions stated in Section 4.2, for known $d$ and unknown $\mu$, as $j \to \infty$,

$$\tilde{N}^{(j)}(\mu) - N^{(j)} = o_p(N^{(j)}), \quad (4.17)$$

$$\hat{\mu}^{(j)} \xrightarrow{p} \mu, \quad (4.18)$$

$$\hat{N}^{(j)} - N^{(j)} = o_p(N^{(j)}), \quad (4.19)$$

and

$$\hat{p}^{(j)} \xrightarrow{p} p. \quad (4.20)$$

Proof. Firstly, given $D^{(j)}$ and $G^{(j)}$, $n^{(j)}$ follows a bin$(N^{(j)} + D^{(j)} + G^{(j)}, p)$ distribution.

So that

$$(N^{(j)})^{-1}n^{(j)} = (N^{(j)})^{-1}(N^{(j)} + D^{(j)} + G^{(j)})(N^{(j)} + D^{(j)} + G^{(j)})^{-1}n^{(j)} \xrightarrow{p} (1 + \mu + \mu^2)p,$$

and hence

$$(N^{(j)})^{-2}n^{(j)}(n^{(j)} - 1) \xrightarrow{p} (1 + \mu + \mu^2)^2p^2.$$

The law of large numbers yields

$$(N^{(j)})^{-1}d^{(j)} \xrightarrow{p} (1 + \mu)p^2.$$
Then \((4.17)\) follows as

\[
(N^{(j)})^{-1} \tilde{N}^{(j)}(\mu) = \frac{(N^{(j)})^{-2} n^{(j)}(n^{(j)} - 1)(\mu + \mu^2)}{(N^{(j)})^{-1} d^{(j)}(1 + \mu + \mu^2)^2} \xrightarrow{p} \frac{(1 + \mu + \mu^2)^2 p^2 \mu(1 + \mu)}{(1 + \mu) \mu p^2 (1 + \mu + \mu^2)^2} = 1.
\]

Next, it is straightforward that

\[
\bar{d} = \frac{d^{(j)}}{x^{(j)}} = \frac{(N^{(j)})^{-1} d^{(j)}}{(N^{(j)})^{-1} x^{(j)}} \xrightarrow{p} \frac{(1 + \mu) \mu p^2}{\mu p} = \frac{1}{1 - e^{-\mu p}} = g(\mu p).
\]

Therefore,

\[
\text{Lambert} \left( \frac{-\bar{d}}{e^{\bar{d}}} \right) \xrightarrow{p} \text{Lambert} \left( -\frac{g(\mu p)}{e^{g(\mu p)}} \right).
\]

Hence

\[
\bar{d} + \text{Lambert} \left( \frac{-\bar{d}}{e^{\bar{d}}} \right) \xrightarrow{p} g(\mu p) + \text{Lambert} \left( -\frac{g(\mu p)}{e^{g(\mu p)}} \right),
\]

which is the solution \(a\) of

\[
g(\mu p) = \frac{a}{1 - e^{-a}}.
\]

But

\[
g(\mu p) = \frac{\mu p}{1 - e^{-\mu p}}.
\]

Therefore \(a = \mu p\) and hence we conclude that

\[
\hat{\mu p}^{(j)} = \bar{d}^{(j)} + \text{Lambert} \left( \frac{-\bar{d}^{(j)}}{e^{\bar{d}^{(j)}}} \right) \xrightarrow{p} \mu p.
\] (4.21)
Now,
\[
\hat{\mu}_j = \frac{n^{(j)} - x^{(j)} - n^{(j)}e^{-\hat{\mu}^{(j)}} + \sqrt{-3(x^{(j)})^2 + 2x^{(j)}n^{(j)}(1 - e^{-\hat{\mu}^{(j)}}) + (n^{(j)})^2(1 - e^{-\hat{\mu}^{(j)}})^2}}}{2x^{(j)}}.
\] (4.22)

Substituting \( n^{(j)} = N^{(j)}(1 + \mu + \mu^2)p \) and \( d^{(j)} = N^{(j)}(1 + \mu)p^2 \) in (4.22) yields (4.18).

Then
\[
(N^{(j)})^{-1}\hat{N}^{(j)} = \frac{\hat{\mu}^{(j)}(1 + \hat{\mu}^{(j)})}{(1 + \hat{\mu}^{(j)} + (\hat{\mu}^{(j)})^2)} \left( \frac{(N^{(j)})^{-2}n^{(j)}(n^{(j)} - 1)}{(N^{(j)})^{-1}d^{(j)}} \right)
\]
\[
\rightarrow \frac{\mu(1 + \mu)}{(1 + \mu + \mu^2)^2} \frac{(1 + \mu + \mu^2)p^2}{(1 + \mu)(1 + \mu)p^2}
\]
\[
= 1,
\]
which yields (4.19). Finally (4.20) follows from (4.18) and (4.21). \( \square \)
Chapter 5

More precise estimator of the number of mothers in a population with more complicated population structure

5.1 Background

In Chapter 2 and 4 we estimate the number of mothers in a population using method of moments considering the mother-daughter pairs in the sample. In this chapter we incorporate the sister pairs in the sample to estimate the number of mothers in the population. This is an attempt to develop a more precise estimator by reducing its variance. The estimators are constructed for populations consisting of two and three generations and we propose a generalisation for a population with any known number of generations.

Here, also, we consider a single sex. Further, we suppose that the observed population consists of an initial number of mothers and their daughters. As in our previous chapters, here, also, we use genotype data from a single capture occasion to develop our methods to estimate the population size. We develop simple method of moments type estimators for the number of mothers in the population, which is
our primary parameter of interest. We also develop estimators for the mean number of daughters per mother and the capture probability.

5.2 Two generations

We restrict the population structure to two generations in this section. We consider that mother-daughter pairs and the sister pairs among daughters in the sample are perfectly identifiable. We consider only the sisters in the second generation for our estimator, i.e. a sister pair is a pair of daughters of the same mother in the population.

5.2.1 Notation and assumptions

The primary objective of this section is to construct estimators for the number of mothers in a population with two generations. There are $N$ mothers in the population who have $D$ daughters and $\mu$ is the mean number of daughters per mother. These $D$ daughters produce $S$ sister pairs. Each individual has probability $p$ of being captured. We treat $N$, $\mu$ and $p$ as parameters, where $N$ is our primary parameter of interest while $\mu$ and $p$ are nuisance parameters. Then we estimate $D$ and $S$ with $N$, $\mu$ and $p$ as we explain in Section 5.2.2.1.

A single sample is observed as a result of a capture experiment. The data consist of $n$ individuals captured in total; $x$ observed mothers, each of which has at least one captured daughter; $d_1, \ldots, d_x$, which are the numbers of caught daughters of the $x$ observed mothers; and $s$ pairs of sisters. Thus a total of $d$ daughters is observed, where $d = \sum_{i=1}^{x} d_i$. Note that $s$ pairs of sisters arise from the $d$ observed daughters and hence $s = \binom{d}{2}$. For the convenience of developing estimators we introduce two other variables $c^*$, which is the number of sister clusters whose mother is uncaptured, i.e. groups of captured daughters of $c^*$ uncaptured mothers; and $s_{i1}^*, \ldots, s_{ic^*}$ number of sister pairs, comprised of daughters of these $c^*$ mothers. Hence, there exist a total of $s^*$ sister pairs whose mother is not captured, i.e. $s^* = \sum_{i=1}^{c^*} s_{i1}^*$. Our notation is summarised in Table 5.1.
We make the same assumptions as in Section 2.2 except, additionally, we suppose that the sister pairs among daughters can be reliably determined.

### 5.2.2 Estimators

We present the models we develop for estimating our parameters using method of moments in this section. We initially estimate our primary parameter $N$ assuming $\mu$ is known. Then we estimate our nuisance parameters $\mu$ and $p$. In Section 5.2.2.1 we present the moment equalities required to derive the aforementioned estimators.

#### Table 5.1: Notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Number of captured individuals</td>
</tr>
<tr>
<td>$x$</td>
<td>Number of observed mothers with at least one daughter observed</td>
</tr>
<tr>
<td>$d$</td>
<td>Number of observed mother-daughter pairs (i.e. Number of observed daughters)</td>
</tr>
<tr>
<td>$s$</td>
<td>Number of observed sister pairs among daughters</td>
</tr>
<tr>
<td>$c^*$</td>
<td>Number of observed sister clusters whose mother is unobserved</td>
</tr>
<tr>
<td>$s^*$</td>
<td>Number of observed sister pairs whose mother is unobserved</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of mothers in the population</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Mean number of daughters per mother</td>
</tr>
<tr>
<td>$p$</td>
<td>Capture probability</td>
</tr>
<tr>
<td>$D$</td>
<td>Number of daughters in the population</td>
</tr>
<tr>
<td>$S$</td>
<td>Number of sister pairs among daughters in the population</td>
</tr>
</tbody>
</table>
5.2.2.1 Moment equalities

Note that for random variables \( X \) and \( Y \), let \( X|Y \) denote the conditional distribution of \( X \) given \( Y \). To derive our estimators we require the following moments:

**Proposition 5.1.**

\[
E(D) = N\mu \quad (5.1)
\]
\[
E(S) = \frac{N\mu^2}{2} \quad (5.2)
\]
\[
E(n) = N(1+\mu)p \quad (5.3)
\]
\[
E(d|x) = \frac{x\mu p}{1-e^{-\mu p}} \quad (5.4)
\]
\[
E(s^*|c^*) = \frac{c^*\mu^2 p^2}{2(1-e^{-\mu p} - \mu p e^{-\mu p})} \quad (5.5)
\]
\[
E(x) = Np(1-e^{-\mu p}) \quad (5.6)
\]
\[
E(d) = N\mu p^2 \quad (5.7)
\]
\[
E(x|n) = \frac{n}{1+\mu}(1-e^{-\mu p}) \quad (5.8)
\]

**Proof.**

Proofs of (5.1), (5.3), (5.4), (5.6), (5.7) and (5.8) are similar to the proofs of (2.1), (2.2), (2.3), (2.4), (2.5) and (2.6) respectively, and hence we do not present them here.

**Proof of (5.2)**

Given that \( D_i \) be the number of daughters of mother \( i \) and \( S_i \) is the number of sister pairs made of those \( D_i \) daughters,

\[
S_i = \binom{D_i}{2} = \frac{D_i(D_i-1)}{2}.
\]

Under the assumption that the number of daughters of a given mother\( (D_i) \) has a Poisson\( (\mu) \) distribution,

\[
E(S_i) = E\left(\frac{1}{2}(D_i^2 - D_i)\right) = \frac{\mu^2}{2}
\]
Now, $S = \sum_{i=1}^{N} S_i$ gives

$$E(S) = \sum_{i=1}^{N} E(S_i).$$

and thus yields (5.2).

**Proof of (5.5)**

Recall that $s^*$ is the number of sister pairs of whom mother is not captured and $c^*$ is the number of sister clusters made of above $s^*$ sisters. Note that one such cluster is made of daughters of a single mother who is not captured. Let $s_i^*$ be the number of sister pairs who are daughters of uncaught mother $i$ and $d_i^*$ the daughters of mother $i$ making these $s_i^*$ sister pairs.

$$E(s_i^*) = E\left(\frac{d_i^*(d_i^* - 1)}{2}\right) = \frac{1}{2} \left( E(d_i^{*2}) - E(d_i^*) \right)$$

Note that at least two daughters of a single mother should be caught to make a sister pair.

$$E(d_i^*) = E(d_i|d_i \geq 2)$$

$$= \frac{\mu p(1 - e^{-\mu p})}{1 - e^{-\mu p} - \mu pe^{-\mu p}}$$

and

$$E(d_i^{*2}) = E(d_i^2|d_i \geq 2)$$

$$= \frac{\mu p + (\mu p)^2 - \mu pe^{-\mu p}}{1 - e^{-\mu p} - \mu pe^{-\mu p}}.$$ 

Therefore,

$$E(s_i^*) = \frac{\mu^2 p^2}{2(1 - e^{-\mu p} - \mu pe^{-\mu p})}.$$ 

Also, $s^* = \sum_{i=1}^{c^*} s_i^*$ yielding (5.5). 

\square
5.2.2.2 Mean number of daughters per mother is known

The mean number of daughters per mother is usually unknown for natural populations. In the first step we assume it is known and construct an estimator \( \hat{N}(\mu) \) for \( N \). Recall that we assume that \( d \), \( s \) and \( x \) are perfectly identifiable.

**Theorem 5.2.** Supposing that the number of mother-daughter pairs \( (d) \) and the number of sister pairs \( (s) \) in the sample of size \( n \) is perfectly identifiable, and the mean number of daughters per mother \( (\mu) \) is known, the method of moment estimator of \( N \) under the assumptions stated in Section 5.2.1 is

\[
\hat{N}(\mu) = \frac{n(n-1)(\mu + \frac{\mu^2}{2})}{(d+s)(1+\mu)^2}.
\]  

(5.9)

**Proof.** The total number of possible pairs in the population is \( \binom{N+D}{2} \). The number of mother-daughter pairs is equal to the number, \( D \), of daughters in the population and the number of sister pairs is \( S \). Therefore, the proportion of all possible pairs that are either mother-daughter pairs or sister pairs is

\[
\frac{D+S}{\binom{N+D}{2}} = \frac{2(D+S)}{(N+D)(N+D-1)}.
\]

Then

\[
\frac{2(D+S)}{(N+D)(N+D-1)} - \frac{2(\mu + \frac{\mu^2}{2})}{N(1+\mu)^2} = \frac{2N(D+S)(1+\mu)^2 - 2(\mu + \frac{\mu^2}{2})(N+D)(N+D-1)}{(N+D)(N+D-1)N(1+\mu)^2}
\]

\[
\approx \frac{2N(D+S)(1+\mu)^2 - 2(\mu + \frac{\mu^2}{2})(N+D)^2}{(N+D)^2N(1+\mu)^2}; \text{for large } N.
\]

With the law of large numbers and from (5.1) and (5.2), \( D \approx N\mu \) and \( S \approx N\mu^2/2 \). Therefore,

\[
\frac{2(D+S)}{(N+D)(N+D-1)} - \frac{2(\mu + \frac{\mu^2}{2})}{N(1+\mu)^2} \overset{p}{\to} 0.
\]
Next, the proportion of all possible pairs in the sample that are either mother-daughter pairs or sister pairs is

$$\frac{(d + s)}{\binom{n}{2}} = \frac{2(d + s)}{n(n - 1)}.$$

Then (5.9) arises from getting

$$\frac{2(\mu + \frac{\mu^2}{2})}{N(1 + \mu)^2} = \frac{2(d + s)}{n(n - 1)}.$$

\[\square\]

5.2.2.3 Mean number of daughters per mother is unknown

The mean number of daughters per mother, $\mu$ is necessary for estimating the number of mothers in the population size using $\hat{N}(\mu)$ (5.9). In this section, we derive estimators for $\hat{\mu}_p$, $\hat{\mu}$ and $\hat{N}$ as in Section 2.3.3. We propose two estimators for $\hat{\mu}_p$; first using the expected value of the mean number of observed daughters for each identified mother; second using the above, the expected value of the mean number of observed daughters for each identified mother, along with the expected value of the mean number of observed sister pairs of whom the mother is unobserved for each of their mothers. The first estimating equation is solved using the principal branch of Lambert’s W function [8]. The second does not have an explicit solution, hence needs to be solved for $\mu p$ using a computer software (for example R) to find $\hat{\mu}_p$.

**Theorem 5.3.** Supposing that the number of mother-daughter pairs ($d$) and number of sister pairs ($s$) in the sample of size $n$ is perfectly identifiable, and the mean number of daughters per mother ($\mu$) is unknown, the method of moment estimator of $N$ under the assumptions stated in Section 5.2.1 is

$$\hat{N} = \frac{n(n - 1)(\hat{\mu} + \frac{\hat{\mu}^2}{2})}{(d + s)(1 + \hat{\mu})^2} \tag{5.10}$$

where $\hat{\mu}$ is,

$$\hat{\mu} = \begin{cases} \frac{n(1 - e^{-\bar{\mu}_p})}{\bar{d}} - 1, & \text{if } n(1 - e^{-\bar{\mu}_p}) > 2x, \\ 1, & \text{otherwise,} \end{cases} \tag{5.11}$$
More complicated population structure

Here, $\mu_p$ can be estimated by,

$$\hat{\mu_p} = \bar{d} + Lambert\left(\frac{-\bar{d}}{e^{\bar{d}}}\right)$$  \hspace{1cm} (5.12)

with ‘Lambert’ being (the principal branch of) Lambert’s W function and $\bar{d} = d/x$ if $d > x$ and $\bar{d} = d/x + 0.1$ if $d = x$, or by solving

$$\bar{d} + \bar{s}^* = \frac{\mu_p}{1 - e^{-\mu_p}} + \frac{\mu^2 p^2}{2(1 - e^{-\mu_p} - \mu pe^{-\mu_p})}$$  \hspace{1cm} (5.13)

with $\bar{s}^* = s^*/c^*$ for $\mu_p$.

Then the capture probability, $p$, can be estimated by

$$\hat{p} = \frac{\hat{\mu_p}}{\hat{\mu}}.$$  \hspace{1cm} (5.14)

Proof. We first treat $\mu_p$ as a single parameter. The proof of (5.12) is similar to the proof of (2.11) in Section 2.3.3.

For (5.13), recall that $\bar{d} = d/x$, so that (5.4) yields

$$E(d|x) = \frac{\mu_p}{(1 - e^{-\mu_p})}.$$  \hspace{1cm} (5.15)

Also, recall that $\bar{s}^* = s^*/c^*$, so that (5.5) yields

$$E(\bar{s}^*|c^*) = \frac{\mu^2 p^2}{2(1 - e^{-\mu_p} - \mu pe^{-\mu_p})}.$$

By the method of moments,

$$\bar{s}^* = \frac{\mu^2 p^2}{2(1 - e^{-\mu_p} - \mu pe^{-\mu_p})}.$$  \hspace{1cm} (5.15)

This and (5.15) gives (5.13).

The estimator $\hat{\mu}$, (5.11), is derived similarly to (2.10) in Section 2.3.3. Substituting (5.11) in (5.9) gives (5.10). Further (5.11) and (5.12) results in (5.14).
Finally using (5.1), the size of the total female population, i.e. the total number of mothers and daughters in the population, is estimated by

$$\hat{N} + \hat{D} = \hat{N}(1 + \hat{\mu})$$  \hspace{1cm} (5.16)

where $\hat{D}$ is estimated by $\hat{N}\hat{\mu}$.

The second estimator we suggested for $\mu_p$ in (5.13) raises issues when $\mu$ is small. When $\bar{d} = \hat{s^*} = 1$, solving (5.13) for $\mu_p$ results in $\hat{\mu_p} = 0$. There is a high probability that $\bar{d} = \hat{s^*} = 1$ when either $\mu$, $p$ or $n$ is small. Further, since (5.13) does not give an explicit estimator for $\mu_p$ the use of it in practice is also not convenient. Hence we recommend using (5.12) for estimating $\mu_p$. In our simulation studies in Section 5.2.4 also, we use (5.12) for estimating $\mu_p$. However, (5.13) can be used with sufficiently large $\mu$, $p$ and $n$.

The statistical consistency of our estimators follows as in Section 2.4, and is presented in Appendix 5.A.

**5.2.3 Estimator of the standard errors**

Following Chapter 2 and 4, a bootstrap method was used to estimate the standard error of the proposed estimator for the number of mothers in the population. In Chapter 2 we described the inappropriateness of non-parametric bootstrap methods in estimating the standard error in our case. Hence, here also, we adopt the parametric bootstrap approach described in Section 2.5.2 [10] to determine the standard errors.

We compute $\hat{N}$, $\hat{\mu}$ and $\hat{p}$ using the proposed estimators for the captured sample as the first step of our parametric bootstrap procedure. Then we simulate a population with $\hat{N}$ number of mothers whose mean number of daughters is equal to $\hat{\mu}$. One thousand bootstrap samples are drawn with a capture probability $\hat{p}$ from this simulated population. For each of the 1000 bootstrap samples, the estimators $\hat{N}_B$, $\hat{\mu}_B$ and $\hat{p}_B$ are computed. The standard deviations of the estimates are calculated as the bootstrap estimate of the standard errors of the estimates. We compute the
confidence intervals (95%) of the estimates, with bootstrap estimates of the standard deviation and normal quantiles. For example, 95% confidence interval of $\hat{N}$ is computed as $\hat{N} \pm 1.96\hat{\sigma}_{\hat{N}(B)}$, where $\hat{\sigma}_{\hat{N}(B)}$ is the bootstrap estimates of the standard deviation of $N$.

5.2.4 Simulation studies

Simulation studies were conducted following the approach of Section 2.6, except we assumed that the sister pairs, in addition to the mother-daughter pairs, can be perfectly identified.

We have considered 16 different populations in our simulations with varying $N$ and $\mu$ from each of which we draw 1000 samples with capture probability 0.3. $\hat{N}$, $\hat{\mu}$ and $\hat{p}$ were computed for the 1000 samples using (5.10), (5.11) and (5.14) respectively. Our estimates of $N$, $\mu$ and $p$, which are the means of 1000 estimates (using 1000 samples), computed for a single population are presented in Table 5.2.

We present the true and estimated standard deviations of $N$, $\mu$ and $p$ in Table 5.3. The standard deviation of 1000 estimates from 1000 samples is referred to as true standard deviation. A population was simulated with $\hat{N}$, $\hat{\mu}$ and $\hat{p}$ of each sample. Then 1000 bootstrap samples were drawn from this population. The standard deviation of the estimates of these 1000 samples was computed as the bootstrap estimate of the standard deviation. The mean of 1000 bootstrap estimates of standard deviation is presented as the estimated standard deviation in Table 5.3. The estimated standard deviation of $\hat{N}$ is, in general, reasonable, if perhaps slightly conservative. Comparing the standard deviation of $\hat{N}$ with the estimates presented in Table 2.4, it can be seen that they are now much less than when it was estimated with only the mother-daughter pairs in Chapter 2.

The distribution of $\hat{N}$ for populations of different sizes with varying $\mu$ is presented in Figure 5.1. The dotted line represents the true $N$ value. Further, we report the percentage bias in the figure. As seen in the figure the bias of $\hat{N}$ is small. Also, the mean of the parametric bootstrap estimates of the standard errors is reported below each boxplot. Confidence intervals (95%) of the estimates were computed with bootstrap estimates of the standard deviation and normal quantiles. Then,
### Table 5.2: Averages of 1000 estimates of the population parameters

<table>
<thead>
<tr>
<th>N</th>
<th>(\mu = 1.25)</th>
<th>(\mu = 2)</th>
<th>(\mu = 5)</th>
<th>(\mu = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean((\hat{N}))</td>
<td>51.80</td>
<td>51.24</td>
<td>50.11</td>
</tr>
<tr>
<td></td>
<td>mean((\hat{\mu}))</td>
<td>1.92</td>
<td>2.05</td>
<td>4.99</td>
</tr>
<tr>
<td>50</td>
<td>mean((\hat{p}))</td>
<td>0.25</td>
<td>0.27</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>(\text{mean}(\hat{N} + \hat{D}))</td>
<td>153.74</td>
<td>159.02</td>
<td>298.74</td>
</tr>
<tr>
<td></td>
<td>(E(\hat{N} + \hat{D}))</td>
<td>112.5</td>
<td>150</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>mean((\hat{N}))</td>
<td>100.69</td>
<td>100.08</td>
<td>100.12</td>
</tr>
<tr>
<td></td>
<td>mean((\hat{\mu}))</td>
<td>1.80</td>
<td>2.05</td>
<td>5.08</td>
</tr>
<tr>
<td>100</td>
<td>mean((\hat{p}))</td>
<td>0.27</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>(\text{mean}(\hat{N} + \hat{D}))</td>
<td>286.62</td>
<td>308.87</td>
<td>607.61</td>
</tr>
<tr>
<td></td>
<td>(E(\hat{N} + \hat{D}))</td>
<td>225</td>
<td>300</td>
<td>600</td>
</tr>
<tr>
<td></td>
<td>mean((\hat{N}))</td>
<td>199.14</td>
<td>199.75</td>
<td>199.95</td>
</tr>
<tr>
<td></td>
<td>mean((\hat{\mu}))</td>
<td>1.67</td>
<td>2.00</td>
<td>5.09</td>
</tr>
<tr>
<td>200</td>
<td>mean((\hat{p}))</td>
<td>0.27</td>
<td>0.31</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>(\text{mean}(\hat{N} + \hat{D}))</td>
<td>537.67</td>
<td>602.80</td>
<td>1215.51</td>
</tr>
<tr>
<td></td>
<td>(E(\hat{N} + \hat{D}))</td>
<td>450</td>
<td>600</td>
<td>1200</td>
</tr>
<tr>
<td></td>
<td>mean((\hat{N}))</td>
<td>505.44</td>
<td>500.60</td>
<td>502.03</td>
</tr>
<tr>
<td></td>
<td>mean((\hat{\mu}))</td>
<td>1.38</td>
<td>1.96</td>
<td>5.05</td>
</tr>
<tr>
<td>500</td>
<td>mean((\hat{p}))</td>
<td>0.28</td>
<td>0.31</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>(\text{mean}(\hat{N} + \hat{D}))</td>
<td>1174.01</td>
<td>1487.60</td>
<td>3034.75</td>
</tr>
<tr>
<td></td>
<td>(E(\hat{N} + \hat{D}))</td>
<td>1125</td>
<td>1500</td>
<td>3000</td>
</tr>
</tbody>
</table>

The true capture probability, \(p\), is 0.3 here. The true values of \(N\) and \(\mu\) are as given in the row and column headers.

The proportion of the occurrences that the confidence interval contains the true value of the parameter, i.e. the coverage probability was calculated. These coverage probabilities of nominal 95\% confidence intervals for \(N\) are also presented in the figure. The decrease of the standard errors of the estimates by incorporating sister pairs into the estimators can again be seen by comparing Figure 5.1 with Figure 2.1.
The standard deviation of the 1000 estimates of $N$ is denoted by $\sigma_N$ and the mean of the 1000 bootstrap estimates of standard deviations is denoted by $\text{mean}(\hat{\sigma}_N(B))$ and so forth for $\mu$ and $p$.
These are based on the 1000 simulations under scenarios (a) $N=50$ (b) $N=100$ (c) $N=200$ (d) $N=500$ with $p=0.3$. The numbers above each boxplot show percentage bias and numbers below show the mean of standard errors determined using 1000 bootstrap samples ($E(\hat{\sigma})$) and the coverage probability of nominal 95% confidence intervals ($\Phi$).
We observed the effect of the capture probability by re-conducting the simulation studies with $p=0.5$. For this, we considered only three cases.

1. $N=50$ and $\mu=2$
2. $N=100$ and $\mu=5$
3. $N=200$ and $\mu=2$

A marked increase in precision was apparent when $p=0.5$ compared with $p=0.3$ (Table 5.4). The standard errors are now much smaller with good coverage probabilities.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$N=50$, $\mu=2$</th>
<th>$N=100$, $\mu=5$</th>
<th>$N=200$, $\mu=2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate SD mean(SE)</td>
<td>Estimate SD mean(SE)</td>
<td>Estimate SD mean(SE)</td>
</tr>
<tr>
<td>$N$</td>
<td>50.74 4.49 5.74</td>
<td>100.03 3.76 3.79</td>
<td>200.10 9.98 10.42</td>
</tr>
<tr>
<td>$\mu$</td>
<td>2.05 0.74 0.81</td>
<td>5.09 0.63 0.65</td>
<td>2.05 0.39 0.39</td>
</tr>
<tr>
<td>$p$</td>
<td>0.52 0.11 0.11</td>
<td>0.50 0.06 0.06</td>
<td>0.50 0.06 0.06</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>95.8%</td>
<td>94.4%</td>
<td>94.3%</td>
</tr>
</tbody>
</table>

We do not report the distribution information of $\hat{\mu}$ and $\hat{p}$ here, since we use the same estimators as in Chapter 2.
5.3 Three generations

In this section we restrict the population structure to three generations. We consider that all mother-daughter pairs and the sister pairs among daughters and granddaughters in the sample are perfectly identifiable. We consider the sisters only in the second and third generations in developing our estimator.

5.3.1 Notation and assumptions

There are  \( N \) mothers in the population which have \( D \) daughters and \( \mu \) is the mean number of daughters per mother. These \( D \) daughters become mothers to the next generation having \( G \) daughters, i.e. \( G \) granddaughters of \( N \) mothers. We consider \( \mu \) as the mean number of daughters per mother for this generation too. Among the second and third generation there are \( S \) numbers of sister pairs. Mark that we refer to first generation of size \( N \) as mothers, second generation of size \( D \) as daughters and third generation of size \( G \) as granddaughters. Each individual has probability \( p \) of being captured. We treat \( N \), \( \mu \) and \( p \) as parameters and then estimate \( D \), \( G \) and \( S \) with \( N \), \( \mu \) and \( p \) as shown in Section 5.3.2.1.

A single sample is observed as a result of a capture experiment. The data consist of \( n \) individuals captured in total; \( x \) observed mothers each of which has at least one captured daughter; \( d_1, \ldots, d_x \) which are the numbers of caught daughters of the \( x \) observed mothers; and \( s \) pairs of sisters. Thus a total of \( d \) daughters is observed, where \( d = \sum_{i=1}^{x} d_i \). More specifically \( d \) is the number of mother-daughter pairs in the sample. Note that \( x \) mothers can either be from the mother generation or the daughter generation and \( d \) daughters can either be from the daughter generation or the granddaughter generation. Also note that \( s \) pairs of sisters are made of \( d \) daughters and hence \( s = \binom{d}{2} \). Our notation is summarised in Table 5.5.

The primary aim is to estimate the parameter \( N \) (number of mothers), followed by estimating the two nuisance parameters \( \mu \) (mean number of daughters per mother) and \( p \) (capture probability).

We have the same assumptions as in Section 4.2, except we also suppose that the sister pairs can also be reliably identified.
More complicated population structure

Table 5.5: Notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>Number of captured individuals</td>
</tr>
<tr>
<td>( x )</td>
<td>Number of observed mothers with at least one daughter observed</td>
</tr>
<tr>
<td>( d )</td>
<td>Number of observed mother-daughter pairs (i.e. Number of observed daughters)</td>
</tr>
<tr>
<td>( s )</td>
<td>Number of observed sister pairs among daughters and granddaughters</td>
</tr>
<tr>
<td>( N )</td>
<td>Number of mothers in the population</td>
</tr>
<tr>
<td>( \mu )</td>
<td>Mean number of daughters per mother</td>
</tr>
<tr>
<td>( p )</td>
<td>Capture probability</td>
</tr>
<tr>
<td>( D )</td>
<td>Number of daughters in the population</td>
</tr>
<tr>
<td>( G )</td>
<td>Number of granddaughters in the population</td>
</tr>
<tr>
<td>( S )</td>
<td>Number of sister pairs among daughters and granddaughters in the population</td>
</tr>
</tbody>
</table>

5.3.2 Estimators

In this section we develop models for estimating parameters using method of moments. Firstly, we estimate our primary parameter \( N \) assuming \( \mu \) is known. Secondly, we estimate our nuisance parameters \( \mu \) and \( p \). In Section 5.3.2.1 we derive the moment equalities required to determine the above mentioned estimators.

5.3.2.1 Moment equalities

For random variables \( X \) and \( Y \), let \( X|Y \) denote the conditional distribution of \( X \) given \( Y \). We use the following moments equations for deriving our estimators:
Proposition 5.4.

\[ E(D) = N\mu \]  
\[ E(G) = N\mu^2 \]  
\[ E(S) = \frac{N\mu^2}{2}(1 + \mu) \]  
\[ E(n) = N(1 + \mu + \mu^2)p \]

Proof.

Proofs of (5.17), (5.18) and (5.20) are similar to the proofs of (2.1), (4.2) and (4.3) respectively, and hence we do not present them here.

Proof of (5.19)

Let \( S^{(1)} \) be the number of sister pairs among the daughters and \( S^{(2)} \) be the number of sister pairs among the granddaughters.

Given that \( D_i \) be the number of daughters of mother \( i \) and \( S^{(1)}_i \) is the number of sister pairs made of those \( D_i \) daughters,

\[ S^{(1)}_i = \binom{D_i}{2} = \frac{D_i(D_i - 1)}{2}. \]

From (5.3) we get,

\[ S^{(1)} = \frac{N\mu^2}{2}. \]

Given that \( G_i \) is the number of daughters of mother \( i \) and \( S^{(2)}_i \) is the number of sister pairs made of those \( G_i \) daughters,

\[ S^{(2)}_i = \binom{G_i}{2} = \frac{G_i(G_i - 1)}{2}. \]

Under the assumption that the number of daughters of a given mother \( D_i \) has a Poisson(\( \mu \)) distribution,

\[ E(S^{(2)}_i) = E\left( \frac{1}{2} \sum_{i=1}^{N} (G_i^2 - G_i) \right) = \frac{\mu^2}{2}. \]
Also,
\[ S^{(2)} = \sum_{i=1}^{D} S_i^{(2)}. \]

Therefore,
\[ E(S^{(2)}) = \frac{D\mu^2}{2}. \]

Applying (5.17) we get
\[ E(S^{(2)}) = \frac{N\mu^3}{2}. \]

Now, \( S = S^{(1)} + S^{(2)} \) giving (5.19).

5.3.2.2 Mean number of daughters per mother is known

In the first step we assume that the mean number of daughters per mother is known and construct an estimator \( \tilde{N}(\mu) \) for \( N \). Recall that we assume that \( d, s \) and \( x \) are perfectly identifiable.

**Theorem 5.5.** Supposing that the number of mother-daughter pairs \( (d) \) and number of sister pairs \( (s) \) in the sample of size \( n \) is perfectly identifiable, and the mean number of daughters per mother \( (\mu) \) is known, the method of moment estimator of \( N \) under the assumptions stated in Section 5.3.1 is,

\[ \tilde{N}(\mu) = \frac{n(n-1)(\mu + \frac{\mu^2}{2})(1 + \mu)}{2(d + s)(1 + \mu + \mu^2)^2} \]  

(5.21)

**Proof.** The total number of possible pairs in the population is \( \binom{N+D+G}{2} \). The number of mother-daughter pairs is equal to the number, \( D + G \), of daughters in the population and the number of sister pairs is \( S \). Therefore, the proportion of all possible pairs that are either mother-daughter pairs or sister pairs is

\[ \frac{D + S}{\binom{N+D+G}{2}} = \frac{2(D + S)}{(N + D + G)(N + D + G - 1)}. \]
Then
\[
\frac{2(D + S)}{(N + D + G)(N + D + G - 1)} - \frac{\mu(2 + \mu)(1 + \mu)}{N(1 + \mu + \mu^2)^2} = \frac{2N(D + S)(1 + \mu + \mu^2)^2 - \mu(2 + \mu)(1 + \mu)(N + D + G)(N + D + G - 1)}{(N + D + G)(N + D + G - 1)N(1 + \mu + \mu^2)^2} \approx \frac{2N(D + S)(1 + \mu + \mu^2)^2 - \mu(2 + \mu)(1 + \mu)(N + D + G)^2}{(N + D + G)^2N(1 + \mu + \mu^2)^2} \text{; for large } N.
\]

With the law of large numbers and from (5.17), (5.18) and (5.19), \( D \approx N\mu, G \approx N\mu(1 + \mu) \) and \( S \approx \frac{N\mu^2}{2}(1 + \mu) \). Therefore,
\[
\frac{2(D + S)}{(N + D + G)(N + D + G - 1)} - \frac{\mu(2 + \mu)(1 + \mu)}{N(1 + \mu + \mu^2)^2} \xrightarrow{p} 0.
\]

Next, the proportion of all possible pairs in the sample that are either mother-daughter pairs or sister pairs is
\[
\frac{(d + s)}{\binom{n}{2}} = \frac{2(d + s)}{n(n - 1)}.
\]

Then (5.9) arises from getting
\[
\frac{\mu(2 + \mu)(1 + \mu)}{N(1 + \mu + \mu^2)^2} = \frac{2(d + s)}{n(n - 1)}.
\]

\[\square\]

### 5.3.2.3 Mean number of daughters per mother is unknown

Knowledge about \( \mu \) is required to use (5.21) for estimating the population size. In this section we present an estimator that can be used when \( \mu \) is unknown. In fact, we use the same \( \hat{\mu}p \) and \( \hat{\mu} \) we developed in Chapter 4, following Section 5.2.

**Theorem 5.6.** Under the assumptions stated in Section 5.3.1, the number of mothers in the population \( (N) \), whose mean number of daughters \( (\mu) \) is unknown can be estimated by
\[
\hat{N} = \frac{n(n - 1)(\hat{\mu} + \frac{\hat{\mu}^2}{2})(1 + \hat{\mu})}{2(d + s)(1 + \hat{\mu} + \hat{\mu}^2)^2} \quad (5.22)
\]
More complicated population structure

Here

\[
\hat{\mu} = \begin{cases} 
\frac{n(1-e^{-\hat{\mu} p}) - x + \sqrt{(n(1-e^{-\hat{\mu} p}) - x)(n(1-e^{-\hat{\mu} p}) + 3x)}}{2x}, & \text{if } n(1 - e^{-\hat{\mu} p}) \geq 1.5x, \\
1, & \text{otherwise.}
\end{cases}
\tag{5.23}
\]

where,

\[
\hat{\mu} p = \bar{d} + \text{Lambert}\left(\frac{-\bar{d}}{\text{e}^{\bar{d}}}\right)
\tag{5.24}
\]

with ‘Lambert’ being (the principal branch of) Lambert’s W function and \(\bar{d} = d/x\) if \(d > x\) and \(\bar{d} = d/x + 0.1\) if \(d = x\). Then the capture probability, \(p\), can be estimated by

\[
\hat{p} = \frac{\hat{\mu} p}{\bar{\mu}}.
\tag{5.25}
\]

We obtain this estimator (5.22) from substituting \(\mu\) with \(\hat{\mu}\) in (5.21). As we use the same \(\hat{\mu} p\) and \(\hat{\mu}\) we developed in Chapter 4 ((4.13) and (4.12) respectively), we do not present their proof here.

**Remark:** Following (5.10) and (5.22), a generalisation of an estimator of the number of mothers in a population known to have \(z (\geq 2)\) generations is suggested as follows.

\[
\hat{N}_z = \frac{n(n - 1) \hat{\mu}(2 + \hat{\mu})}{2(d + s)} \frac{\sum_{i=1}^{z-1} \hat{\mu}^{i-1}}{\left(\sum_{i=1}^{z} \hat{\mu}^{i-1}\right)^2}.
\]

Here, \(\hat{\mu}\) can be obtained by solving

\[
\frac{x}{n} = \frac{\sum_{j=1}^{z-1} \hat{\mu}^{j-1}}{\sum_{j=1}^{z} \hat{\mu}^{j-1}} (1 - e^{-\hat{\mu} p})
\]

where,

\[
\hat{\mu} p = \bar{d} + \text{Lambert}\left(\frac{-\bar{d}}{\text{e}^{\bar{d}}}\right).
\]

The statistical consistency of our estimators follows as in Section 2.4, and is presented in Appendix 5.B.
5.3.3 Estimator of the standard errors

The standard error of the proposed estimator for the number of mothers in the population was determined following the parametric bootstrap procedure similar to that described in Section 5.2.3.

5.3.4 Simulation studies

We conducted simulation studies to observe the performance of the estimators here as well. We followed the same approach that was described in Section 2.6 with the following obvious exceptions: We follow the approach of Section 2.6 with the obvious changes:

1. We simulate only three generations at a time to comply with the population structure we consider.

2. The daughter population was generated assuming that the number of daughters of a given mother are independent and have a Poisson(\(\mu\)) distribution. The granddaughter population was generated considering the daughter population as mothers and then assuming that the number of daughters of each of them is independent and has a Poisson(\(\mu\)) distribution where this \(\mu\) is the same as the \(\mu\) in their mother’s generation.

3. The individuals were labelled such that the mother-daughter pairs and the sister pairs among daughters and granddaughters are perfectly identifiable.

The samples were drawn such that each individual had the same probability (\(p\)) of being captured, which we considered as 0.3 in the initial studies. One thousand samples were drawn from each of the 16 different populations with varying \(N\) and \(\mu\). \(\hat{N}\), \(\hat{\mu}\) and \(\hat{p}\) were computed for 1000 samples using (5.10), (5.23) and (5.25) respectively. We present our estimates of \(N\), \(\mu\) and \(p\) which are the means of 1000 estimates (using 1000 samples) computed for a single population in Table 5.6.

The true and estimated standard deviations of \(N\), \(\mu\) and \(p\) are presented in Table 5.7. The standard deviation of 1000 estimates from 1000 samples is referred to as true
Table 5.6: Averages of 1000 estimates of the population parameters

<table>
<thead>
<tr>
<th>(N)</th>
<th>(\mu = 1.25)</th>
<th>(\mu = 2)</th>
<th>(\mu = 5)</th>
<th>(\mu = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{mean}(\hat{N}))</td>
<td>50.73</td>
<td>51.16</td>
<td>50.76</td>
<td>50.59</td>
</tr>
<tr>
<td>(\text{mean}(\hat{\mu}))</td>
<td>1.49</td>
<td>2.31</td>
<td>5.09</td>
<td>9.96</td>
</tr>
<tr>
<td>(\text{mean}(\hat{p}))</td>
<td>0.24</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>(\text{mean}(\hat{N} + \hat{D} + \hat{G}))</td>
<td>225.03</td>
<td>409.56</td>
<td>1605.11</td>
<td>5550.37</td>
</tr>
<tr>
<td>(E(\hat{N} + \hat{D} + \hat{G}))</td>
<td>190.625</td>
<td>350</td>
<td>1550</td>
<td>5550</td>
</tr>
<tr>
<td>(\text{mean}(\hat{N}))</td>
<td>100.83</td>
<td>100.23</td>
<td>99.24</td>
<td>100.56</td>
</tr>
<tr>
<td>(\text{mean}(\hat{\mu}))</td>
<td>1.37</td>
<td>2.05</td>
<td>5.06</td>
<td>9.96</td>
</tr>
<tr>
<td>(\text{mean}(\hat{p}))</td>
<td>0.24</td>
<td>0.31</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>(\text{mean}(\hat{N} + \hat{D} + \hat{G}))</td>
<td>413.70</td>
<td>699.04</td>
<td>3121.60</td>
<td>11055.45</td>
</tr>
<tr>
<td>(E(\hat{N} + \hat{D} + \hat{G}))</td>
<td>381.25</td>
<td>700</td>
<td>3100</td>
<td>11100</td>
</tr>
<tr>
<td>(\text{mean}(\hat{N}))</td>
<td>199.30</td>
<td>200.46</td>
<td>201.50</td>
<td>200.59</td>
</tr>
<tr>
<td>(\text{mean}(\hat{\mu}))</td>
<td>1.33</td>
<td>2.04</td>
<td>5.02</td>
<td>10.11</td>
</tr>
<tr>
<td>(\text{mean}(\hat{p}))</td>
<td>0.27</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>(\text{mean}(\hat{N} + \hat{D} + \hat{G}))</td>
<td>802.00</td>
<td>1412.26</td>
<td>6269.61</td>
<td>22701.83</td>
</tr>
<tr>
<td>(E(\hat{N} + \hat{D} + \hat{G}))</td>
<td>762.5</td>
<td>1400</td>
<td>6200</td>
<td>22200</td>
</tr>
<tr>
<td>(\text{mean}(\hat{N}))</td>
<td>501.50</td>
<td>503.10</td>
<td>499.77</td>
<td>499.12</td>
</tr>
<tr>
<td>(\text{mean}(\hat{\mu}))</td>
<td>1.26</td>
<td>1.96</td>
<td>5.04</td>
<td>10.08</td>
</tr>
<tr>
<td>(\text{mean}(\hat{p}))</td>
<td>0.29</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>(\text{mean}(\hat{N} + \hat{D} + \hat{G}))</td>
<td>1925.62</td>
<td>3403.56</td>
<td>15712.22</td>
<td>56180.53</td>
</tr>
<tr>
<td>(E(\hat{N} + \hat{D} + \hat{G}))</td>
<td>1906.25</td>
<td>3500</td>
<td>15500</td>
<td>55500</td>
</tr>
</tbody>
</table>

The true capture probability, \(p\), is 0.3 here. The true values of \(N\) and \(\mu\) are as given in the row and column headers.

standard deviation. A population was simulated with \(\hat{N}\), \(\hat{\mu}\) and \(\hat{p}\) of each sample. Then 1000 bootstrap samples were drawn from this population. The bootstrap estimate of the standard deviation for the sample is the standard deviation of the 1000 estimates of these samples. The mean of these 1000 bootstrap estimates of the standard deviation is presented as the estimated standard deviation in Table 5.7. The estimated standard deviations of \(\hat{N}\) are, in general, reasonable. Comparing the standard deviation of \(\hat{N}\) with the estimates presented in Table 4.4, it can be seen that they are now much less than when the standard deviation was estimated with only the mother-daughter pairs in Chapter 4.
The standard deviation of the 1000 estimates of $N$ is denoted by $\sigma_N$ and the mean of the 1000 bootstrap estimates of standard deviations is denoted by $\text{mean}(\hat{\sigma}_N(B))$ and so forth for $\mu$ and $p$.

Figure 5.2 shows the distribution of $\hat{N}$ for populations of different sizes with varying $\mu$. The dotted line is for the true value of $N$. We also report the percentage bias in the figure. According to the figure, the bias of $\hat{N}$ is small. The mean of the parametric bootstrap estimates of the standard deviations are also reported below each boxplot. We computed the 95% confidence intervals of the estimates with bootstrap estimates of the standard deviation and normal quantiles. Then, the proportion of the occurrences that the confidence interval contains the true value of the parameter, i.e. the coverage probability was calculated. These coverage probabilities of nominal 95% confidence intervals for $N$ are also presented in the figure and are very
satisfying. The reduction of the standard errors of \( \hat{N} \) by incorporating sister pairs into the estimators can further be seen by comparing this figure with Figure 4.2.

![Boxplots of \( \hat{N} \) for different \( N \) and \( \mu \) values](image)

Figure 5.2: Distribution of \( \hat{N} \)

These are based on the 1000 simulations under scenarios (a) \( N=50 \) (b) \( N=100 \) (c) \( N=200 \) (d) \( N=500 \) with \( p=0.3 \). The numbers above each boxplot show percentage bias and numbers below show the mean of standard errors determined using 1000 bootstrap samples (\( E(\hat{\sigma}) \)) and the coverage probability of nominal 95% confidence intervals (\( \Phi \)).

We observed the effect of the capture probability by re-conducting the simulation studies with \( p=0.5 \). For this, we considered only three cases.

1. \( N=50 \) and \( \mu=2 \)
2. \( N=100 \) and \( \mu=5 \)
3. \( N=200 \) and \( \mu=2 \)

A marked increase in precision was apparent when \( p=0.5 \) compared with \( p=0.3 \) (Table 5.8). The standard errors are now significantly reduced with good coverage probabilities.

We do not report the distribution information of \( \hat{\mu} \) and \( \hat{p} \) here, since we used the same estimators as in Chapter 4.
5.4 Discussion

In this chapter we extend our method to use more information from the sample we capture. In particular, we incorporate sister pairs in addition to the mother-daughter pairs into the estimator. Thereby, we can substantially decrease the standard error of the estimator for the number of mothers in the population.

We construct estimators using method of moments, incorporating sister pairs in the sample for the number of mothers in the population. Here we consider populations with two generations and three generations and propose a generalisation for a population with any known number of generations. We illustrate that the estimators perform well for populations with two generations and three generations with simulation studies. The simulations exhibit a significant drop in the standard error of $\hat{N}$ than when only mother-daughter pairs were used for estimation. This can be reduced further by increasing the capture probability. In this chapter, also, the standard deviation of the estimator of number of mothers in the population is estimated with a parametric bootstrap approach. These estimates are reasonable in general. Consequently the coverage probability of nominal 95% confidence intervals is also satisfying.

Sister pairs we consider here are the pairs of daughters which have the same mother. They are not necessarily full siblings. Recognising these pairs using genetic markers is not discussed in the current work and we leave it for future work. The objective
of this chapter is to examine the possibility of improving the precision of the estimators through reducing the variability by incorporating more relationships into the estimator.

5.A Consistency of the estimators for two generations

Here we show that the two estimators we proposed for $N$, (5.9) and (5.10), the estimator of $\mu$, (5.11), and estimator of $p$, (5.14), are statistically consistent. Note that as the estimator for $\hat{\mu} p$ we consider only (5.12) since it is the estimator we use in the simulation studies.

Theorem 5.7. Consider a sequence $j = 1, 2, \ldots$ of populations of size $N^{(j)}$ with $N^{(j)} \to \infty$. Let $D^{(j)}, n^{(j)}, d^{(j)}$ and $s^{(j)}$ be $D$, $n$, $d$ and $s$ defined for the $j^{th}$ population and similarly for $\tilde{N}^{(j)}(\mu)$ and $\hat{N}^{(j)}$. Suppose that $\mu$ and $p$ do not depend on $j$. Under the assumptions stated in Section 5.2.1, for known $d$, $s$ and unknown $\mu$, as $j \to \infty$,

$$\tilde{N}^{(j)}(\mu) - N^{(j)} = o_p(N^{(j)}),$$  \hspace{1cm} (5.26)

$$\hat{\mu}^{(j)} \xrightarrow{p} \mu,$$  \hspace{1cm} (5.27)

$$\hat{N}^{(j)} - N^{(j)} = o_p(N^{(j)}),$$  \hspace{1cm} (5.28)

and

$$\hat{p}^{(j)} \xrightarrow{p} p.$$  \hspace{1cm} (5.29)

Proof. Firstly, given $D^{(j)}, n^{(j)}$ follows a bin($N^{(j)} + D^{(j)}, p$) distribution.

So that,

$$(N^{(j)})^{-1}n^{(j)} = (N^{(j)})^{-1}(N^{(j)} + D^{(j)})(N^{(j)} + D^{(j)})^{-1}n^{(j)} \xrightarrow{p} (1 + \mu)p,$$

and hence

$$(N^{(j)})^{-2}n^{(j)}(n^{(j)} - 1) \xrightarrow{p} (1 + \mu)^2p^2.$$
The law of large numbers yields

\[ (N^{(j)})^{-1}d^{(j)} \xrightarrow{p} \mu p^2 \]

and

\[ (N^{(j)})^{-1}s^{(j)} \xrightarrow{p} \frac{\mu^2 p^2}{2} \]

Then (5.26) follows as

\[
(N^{(j)})^{-1}\tilde{N}^{(j)}(\mu) = \frac{(\mu + \frac{\mu^2}{2})(N^{(j)})^{-2}n^{(j)}(n^{(j)} - 1)}{(N^{(j)})^{-1}(d^{(j)} + s^{(j)})(1 + \mu)^2} \xrightarrow{p} \frac{(\mu + \frac{\mu^2}{2})(1 + \mu)^2 p^2}{(\mu p^2 + \frac{\mu^2 p^2}{2})(1 + \mu)^2} = 1.
\]

Next, it is straightforward that

\[
\tilde{d} = \frac{d^{(j)}}{x^{(j)}} = (N^{(j)})^{-1}d^{(j)} (N^{(j)})^{-1}x^{(j)} \xrightarrow{p} \frac{\mu p^2}{p(1 - e^{-\mu p})} = \frac{\mu p}{1 - e^{-\mu p}} = g(\mu p).
\]

Therefore,

\[
\text{Lambert} \left( \frac{-d}{e^d} \right) \xrightarrow{p} \text{Lambert} \left( -\frac{g(\mu p)}{e^{g(\mu p)}} \right).
\]

Hence

\[
\tilde{d} + \text{Lambert} \left( \frac{-d}{e^d} \right) \xrightarrow{p} g(\mu p) + \text{Lambert} \left( -\frac{g(\mu p)}{e^{g(\mu p)}} \right),
\]

which is the solution \( a \) of

\[
g(\mu p) = \frac{a}{1 - e^{-a}}.
\]

But

\[
g(\mu p) = \frac{\mu p}{1 - e^{-\mu p}}.
\]
Therefore $a = \mu p$ and hence we conclude that

$$\hat{\mu p}^{(j)} = \bar{d}^{(j)} + \text{Lambert}\left(\frac{-d^{(j)}}{e^{\bar{d}^{(j)}}}\right) \xrightarrow{p} \mu p. \quad (5.30)$$

Now,

$$n^{(j)} \frac{x^{(j)}}{(N^{(j)})^{-1} x^{(j)}} \xrightarrow{p} \frac{(1 + \mu)p}{p(1 - e^{-\mu p})} = \frac{(1 + \mu)}{1 - e^{-\mu p}}.$$ 

So that

$$\hat{\mu}^{(j)} = n^{(j)} \frac{(1 - e^{-\hat{\mu p}^{(j)}})}{x^{(j)}} - 1 \xrightarrow{p} \frac{(1 + \mu)(1 - e^{-\mu p})}{1 - e^{-\mu p}} - 1 = \mu,$$

giving (5.27).

Then,

$$(N^{(j)})^{-1} \hat{N}^{(j)} = \frac{(\hat{\mu} + \frac{\mu^2}{2}) (N^{(j)})^{-2} n^{(j)}(n^{(j)} - 1)}{(1 + \hat{\mu})^2 (N^{(j)})^{-1}(d^{(j)} + s^{(j)})} \xrightarrow{p} \frac{(\mu + \frac{\mu^2}{2}) \mu^2}{(1 + \mu)^2 (\mu p^2 + \frac{\mu^2 p^2}{2})} = 1,$$

which yields (5.28). Finally (5.29) follows from (5.27) and (5.30).

**Remark:** It may be more realistic to let $p$ depend on $j$. That is, $p_j \to 0$ as $j \to \infty$. \qed
5.B Consistency of the estimators for three generations

Here we show that the two estimators we proposed for $N$, (5.21) and (5.22), the estimator of $\mu$, (5.23), and the estimator of $p$, (5.25), are statistically consistent.

**Theorem 5.8.** Consider a sequence $j = 1, 2, \ldots$ of populations of size $N^{(j)}$ with $N^{(j)} \to \infty$. Let $n^{(j)}$ and $d^{(j)}$ be $n$ and $d$ defined for the $j$th population and similarly for $\tilde{N}^{(j)}(\mu)$ and $\hat{N}^{(j)}$. Suppose that $\mu$ and $p$ do not depend on $j$. Under the assumptions stated in Section 5.3.1, for known $d$, $s$ and unknown $\mu$, as $j \to \infty$,

\[
\tilde{N}^{(j)}(\mu) - N^{(j)} = o_p(N^{(j)}), \tag{5.31}
\]

\[
\hat{\mu}^{(j)} \xrightarrow{p} \mu, \tag{5.32}
\]

\[
\hat{N}^{(j)} - N^{(j)} = o_p(N^{(j)}), \tag{5.33}
\]

and

\[
\hat{p}^{(j)} \xrightarrow{p} p. \tag{5.34}
\]

**Proof.** Firstly, given $D^{(j)}$ and $G^{(j)}$, $n^{(j)}$ follows a bin($N^{(j)} + D^{(j)} + G^{(j)}$, $p$) distribution.

So that

\[
(N^{(j)})^{-1}n^{(j)} = (N^{(j)})^{-1}(N^{(j)} + D^{(j)} + G^{(j)})(N^{(j)} + D^{(j)} + G^{(j)})^{-1} n^{(j)} \xrightarrow{p} (1 + \mu + \mu^2)p,
\]

and hence $(N^{(j)})^{-2}n^{(j)}(n^{(j)} - 1) \xrightarrow{p} (1 + \mu + \mu^2)^2 p^2$.

The law of large numbers yields

\[
(N^{(j)})^{-1}d^{(j)} \xrightarrow{p} (1 + \mu)\mu p^2
\]

and

\[
(N^{(j)})^{-1}s^{(j)} \xrightarrow{p} \frac{\mu^2 p^2}{2}(1 + \mu).
\]
Then (5.31) follows as

\[
(N^{(j)})^{-1} \tilde{N}^{(j)}(\mu) = \frac{(N^{(j)})^{-2} n^{(j)}(n^{(j)} - 1)(\mu + \frac{\mu^2}{2})(1 + \mu)}{(N^{(j)})^{-1}(d^{(j)} + s^{(j)})(1 + \mu + \mu^2)^2} - \frac{p}{(1 + \mu + \mu^2)^2}p^2(\mu + \frac{\mu^2}{2})(1 + \mu)
\]

\[
\frac{1 + \mu + \mu^2}{(1 + \frac{\mu}{2})(1 + \mu)\mu p^2(1 + \mu + \mu^2)^2} = 1.
\]

Next, it is straightforward that

\[
\bar{d} = \frac{d^{(j)}}{x^{(j)}} = \frac{(N^{(j)})^{-1} d^{(j)}}{(N^{(j)})^{-1}x^{(j)}} - \frac{p}{(1 + \mu)\mu p^2} = \frac{(1 + \mu)\mu p^2}{\mu p} = 1 - e^{-\mu p} = g(\mu p).
\]

Therefore,

\[
\text{Lambert}\left(\frac{-\bar{d}}{e^\bar{d}}\right) \overset{p}{\rightarrow} \text{Lambert}\left(\frac{-g(\mu p)}{e^{g(\mu p)}}\right).
\]

Hence

\[
\bar{d} + \text{Lambert}\left(\frac{-\bar{d}}{e^\bar{d}}\right) \overset{p}{\rightarrow} g(\mu p) + \text{Lambert}\left(\frac{-g(\mu p)}{e^{g(\mu p)}}\right),
\]

which is the solution \(a\) of

\[
g(\mu p) = \frac{a}{1 - e^{-a}}.
\]

But

\[
g(\mu p) = \frac{\mu p}{1 - e^{-\mu p}}.
\]

Therefore \(a = \mu p\) and hence we conclude that

\[
\hat{\mu}p^{(j)} = \tilde{d}^{(j)} + \text{Lambert}\left(\frac{-\tilde{d}^{(j)}}{e^{\tilde{d}^{(j)}}}\right) \overset{p}{\rightarrow} \mu p.
\] (5.35)
Now,
\[
\hat{\mu}_j = \frac{n^{(j)} - x^{(j)} - n^{(j)}e^{-\hat{\mu}_p^{(j)}} + \sqrt{-3(x^{(j)})^2 + 2x^{(j)}n^{(j)}(1 - e^{-\hat{\mu}_p^{(j)}})} + (n^{(j)})^2(1 - e^{-\hat{\mu}_p^{(j)}})^2}{2x^{(j)}} \tag{5.36}
\]
Replacing \(n^{(j)} = N^{(j)}(1 + \mu + \mu^2)p\) and \(d^{(j)} = N^{(j)}(1 + \mu)\mu p^2\) in (5.36) yields (5.32).

Then,
\[
(N^{(j)})^{-1}\hat{N}^{(j)} = \frac{\hat{\mu}^{(j)}(1 + \hat{\mu}^{(j)})}{(1 + \hat{\mu}^{(j)} + (\hat{\mu}^{(j)})^2)^2} \left( \frac{(N^{(j)})^{-2}n^{(j)}((n^{(j)}) - 1)}{(N^{(j)})^{-1}d^{(j)}} \right)
\]
\[
\rightarrow \frac{\mu(1 + \mu)}{(1 + \mu + \mu^2)^2} \frac{(1 + \mu + \mu^2)^2p^2}{(1 + \mu)\mu p^2}
\]
\[
= 1,
\]
which yields (5.33). Finally (5.34) follows from (5.32) and (5.35).
Chapter 6

The effect of misclassification of relationships

6.1 Background

Our approach requires us to know the number of mother-daughter pairs \((d)\) and the number of mothers with at least one daughter being observed \((x)\) in the captured sample in order to estimate the population size. It is not possible to reveal \(d\) and \(x\) directly from the captured animals for most natural populations and therefore we need an alternate approach.

Recall from Section 2.8 that we use genotype data at 21 loci collected from female mountain pygmy possums \((Burramys parvus)\) to identify parent-offspring pairs. In order to do this, we use a false discovery rate procedure based on the likelihood ratio following Skaug et al. [34]. Their R program is available from \(http://folk.uib.no/hsk021/fdr_program.htm\). We determine \(d\) using this approach and subsequently deduce \(x\).

In this approach the relationship between a pair of individuals is determined using the log likelihood ratio (LOD score) based on genotypes. Determining the number of related dyads using this method involves multiple simultaneous comparisons, raising some statistical issues. The most famous method for compensating these
multiple comparison issues is Sequential Bonferroni correction [27] which guards against even a single false positive. This conservative criterion increases the number of false negatives, i.e. the undetected but related dyads. The method of Skaug et al. [34] accepts a small proportion of false positives, which keeps the number of false negatives at a minimum. Their approach uses the false discovery rate (FDR) to control the proportion of false positives, i.e. the proportion of dyads consisting of unrelated individuals which would have been labelled as related.

LOD scores are computed by comparing the hypothesis of unrelatedness with the alternative of a parent-offspring relationship. A large LOD score implies a parent-offspring pair. To determine how ‘large’ the LOD score needs to be for a pair to be identified as a parent-offspring pair, a p-value is computed using a permutation approach. That is, a sample of the same size is simulated by distributing all alleles in a particular locus independently across loci, which results in a sample with same allele frequencies as the original, yet with unrelated, individuals. The LOD scores are computed for all pairs in the simulated data set. The p-value for a pair is the proportion of simulated LOD scores exceeding that of the pair. Then, the p-values for all pairs are arranged in an increasing order and the first \( R \) pairs are considered as related. Skaug et al. [34] define \( R \) as the largest value of \( r \) for which \( P_r \leq r/m \cdot q \) so that the FDR is controlled at a nominated level \( q \). Here \( m \) is the total number of LOD scores (\( m = n(n-1)/2 \)).

The number of identified parent-offspring pairs above is \( d \). All the unique individuals involved with these \( d \) pairs are the collection of the daughters whose mother is also captured (which is again equal to \( d \)) and mothers whose at least one daughter is captured \( (x) \). Thus the difference between the number of unique individuals in the \( d \) pairs and \( d \) gives \( x \). Note that this approach does not identify mothers and daughters explicitly but rather identify the mother-daughter pairs we use to deduce the number of daughters whose mother is also captured and the number of mothers from whom at least one daughter is captured.

6.1.1 False discovery rate

FDR-controlling procedures control the expected proportion of rejected null hypotheses that were incorrect rejections, i.e. false discoveries [4]. Unlike familywise
error rate (FWER) controlling procedures, such as the Bonferroni correction procedure which controls the probability of at least one Type I error, FDR-controlling procedures are less stringent in controlling individual Type I errors. Therefore, FDR-controlling procedures have greater power, at the cost of increased rates of Type I errors [32].

In our study, with a certain expected proportion of false positives being permitted, the detected number of mother-daughter pairs could be contaminated. This may arise through misclassified daughters, as well as misclassified mothers, which affects our estimators. Further, in the FDR procedure the number of false negatives is expected to be very small and considered negligible in our work.

In this chapter we consider only the two generations setting where we use only mother-daughter pairs for estimation. We developed these estimators with the method of moment in Chapter 2. First, we show that this misclassification of mothers and daughters causes a bias in our estimators \( \hat{\mu} \) and \( \hat{N} \) with simulation studies. Then, we quantify the error caused in our estimator due to misclassification theoretically and finally illustrate it with simulations.

### 6.2 Notation and assumptions

As noted above, the observed number of mother-daughter pairs and observed number of mothers with at least one daughter captured can be different from the true statistics. The notations we use is summarised in Table 6.1.

We denote the false discovery proportion by \( \beta \) and the number of observed mother-daughter pairs by \( d_o \). Hence,

\[
\beta = \frac{d_o - d}{d_o}.
\]

(6.1)

This gives,

\[
d_o = \frac{d}{1 - \beta}.
\]

(6.2)

Let the false discovery rate be \( q \) and this is defined by,

\[
q = E(\beta)
\]
Table 6.1: Notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Number of captured individuals</td>
</tr>
<tr>
<td>$x_o$</td>
<td>Number of observed mothers with at least one daughter observed</td>
</tr>
<tr>
<td>$d_o$</td>
<td>Number of observed mother-daughter pairs (i.e. Number of observed daughters)</td>
</tr>
<tr>
<td>$x$</td>
<td>Number of mothers in the sample with at least one truly related daughter observed</td>
</tr>
<tr>
<td>$d$</td>
<td>True number of mother-daughter pairs in the sample (i.e. Number of daughters of whom the true mother is captured.)</td>
</tr>
</tbody>
</table>

Note that we control the false discovery rate at an upper limit of $q$ and $\beta$ is the proportion of false discoveries which is unobserved.

Given that there are $d_o$ observed mother-daughter pairs, the number of misclassified pairs, i.e. $d_o - d$ will have a binomial distribution. (Recall that, in our work, we have considered that the number of false negatives negligible.)

$$d_o - d|d_o \sim \text{bin}(d_o, q) \quad (6.3)$$

Here we assume the trials are independent since in the FDR procedure the selection of a pair as related has no impact on the next pair to be selected.

One of the following two things may occur when an unrelated pair is labelled as related.

1. Neither of the individuals in this pair has been included in a truly related pair. This means that none of the individuals in the pair is truly related to any individual in the sample. Observing a pair like this contaminates both $d$ and $x$. This can be considered as observing a new spurious mother and a new spurious daughter.
2. One of the individuals in this pair is already included in a truly related pair. The one who is not included in a truly related pair has to be considered as a daughter, since even though one mother can have several daughters, a daughter cannot have more than one mother. This means we consider these three individuals as one mother and her two daughters. Hence, observing a pair like this contaminates only $d$, but not $x$. This can be considered as observing a new spurious daughter of a mother who also has a true daughter in the sample.

This explains that the number of misclassified mothers, $x_o - x$ is dependent on the number of misclassified daughters, $d - d_o$, and they are not necessarily the same.

A misclassified pair can also include an individual who is involved in another misclassified pair. However, we ignore this scenario since the probability of its occurrence is very small. This probability can be approximated as follows. Each individual in a pair is involved in $n - 1$ pairs among $n(n - 1)/2 - d$ pairs. Therefore, the probability an individual is involved in another pair by chance is $(n - 1)/(n(n - 1)/2 - d) = 2/n$. Now the probability either of individuals in a pair is involved in another pair is bounded by $4/n$. Therefore, the probability an individual in a misclassified pair is also involved in another misclassified pair is bounded by,

$$1 - \left(\frac{n - 4}{n}\right)^{d_o - d - 1} = 1 - \left(\frac{n - 4}{n}\right)^{\frac{d_o - d}{n} - 1}$$

from (6.1) and (6.2)

which is small.

The probability an individual is misclassified as a mother, given that individual is misclassified as related, is the probability that none of the individuals in the misclassified pair is included in a truly related pair, i.e. $(1 - (d + x)/n)^2$. Therefore, given that there are $d_o - d$ misclassified pairs, (i.e. misclassified daughters), the number of misclassified mothers will have approximately a binomial distribution.

$$x_o - x | x, d_o, d \sim \text{bin} \left( d_o - d, \left(1 - \frac{d + x}{n}\right)^2 \right)$$

(6.4)

Event though these $d_o - d$ misclassified pairs are drawn without replacement from a total of $n(n - 1)/2 - d$ unrelated pairs, suggesting a hypergeometric distribution, we
argue that this can be approximated by a binomial distribution as \((d_o - d)(n(n - 1)/2 - d)\) is small.

\[
\frac{d_o - d}{n(n-1)/2 - d} \approx \frac{2d\beta}{(1 - \beta) (n(n - 1) - 2d)} \quad \text{from (6.1) and (6.2)}
\]

When \(N\) is large, from (6.6) and (6.9),

\[
\frac{2d\beta}{(1 - \beta) (n(n - 1) - 2d)} \approx \frac{2\beta}{(1 - \beta) (N(1 + \mu)p)^2 - 2N\mu p^2} \xrightarrow{p} 0
\]

Note that the probability an individual in a misclassified pair is also involved in another misclassified pair (bounded by \(1 - ((n - 4)/n)^{d\beta(1 - \beta)^{-1}}\)), which is close to zero when \(n\) is large, is much smaller than the probability that none of the individuals in the misclassified pair is included in a truly related pair \(((1 - (d + x)/n)^2)\). Therefore we consider that ignoring the scenario that a misclassified pair can have an individual who is involved in another misclassified pair is justifiable.

### 6.3 Simulation studies

Initially we conduct simulation studies to show the effect of misclassification on our estimates.

1. First a population with known mothers and daughters is generated.

2. A sample of size \(n\) is drawn from the population and the number of truly related dyads \((d)\) is determined. This is possible since the relationships are known for the generated population.

3. \(d_o/\beta\) pairs are drawn randomly from among all \(n(n - 1)/2 - d\) pairs which are not truly related.

4. Considering \(d\) number of truly related pairs and newly selected \(d_o/\beta\) unrelated (but labelled as related) pairs, the number of mothers with at least one daughter captured \((x_o)\) is determined. In fact, \(x_o - x\) is the number of misclassified
pairs in which both the individuals in the pair are not truly related to any individual in the sample. Now, the number of observed mother-daughter pairs is $d_o$ and the sample size still remains as $n$.

5. $\hat{N}$ and $\hat{\mu}$ is computed as a function of $n$, $d_o$ and $x_o$. We also compute $\bar{d}$ and $\hat{\mu}_p$ to illustrate the lemmas presented later in the chapter (Section 6.5).

6. Steps 2-5 are repeated 1000 times and the averages of $\hat{\mu}$ and $\hat{N}$ is calculated as $E(\hat{\mu})$ and $E(\hat{N})$ respectively. (Similarly $E(\bar{d})$ and $E(\hat{\mu}_p)$ is also computed to be used in Section 6.5)

7. The procedure is repeated varying the proportion of false positives ($\beta$).

8. $E(\hat{N})$ and $E(\hat{\mu})$ are plotted against $\beta$.

Figure 6.1 and 6.2 show that the bias in $\hat{N}$ and $\hat{\mu}$ caused by misclassification increases with the false discovery proportion. The bias is small when the false discovery proportion is as small as 0.05. As the next step, we quantify this bias theoretically and show that it is very small with large $N$.
Figure 6.1: Estimates of $\hat{N}$ against $\beta$

Figure 6.2: Estimates of $\hat{\mu}$ against $\beta$
6.4 Limiting distributions

In this section we derive the limiting distributions required to quantify the bias due to misclassification.

Recall that in Section 2.3.1 we established the following moment equalities.

\[
E(D) = N\mu \\
E(n) = N(1 + \mu)p \\
E(d|x) = \frac{x\mu p}{1 - e^{-\mu p}} \\
E(x) = Np(1 - e^{-\mu p}) \\
E(d) = N\mu p^2
\]

Also in Section 2.4 we showed that

\[
\tilde{d} \xrightarrow{p} \mu p \left(1 - e^{-\mu p}\right) \quad (6.10) \\
\hat{\mu}_p^{(j)} \xrightarrow{p} \mu p \quad (6.11) \\
\hat{\mu}^{(j)} \xrightarrow{p} \mu \quad (6.12)
\]

Similarly we can show that

\[
\frac{d}{n} \xrightarrow{p} \frac{\mu p}{1 + \mu} \quad (6.13) \\
\frac{x}{n} \xrightarrow{p} \frac{1 - e^{-\mu p}}{1 + \mu} \quad (6.14) \\
\frac{d + x}{n} \xrightarrow{p} \frac{1 + \mu p - e^{-\mu p}}{1 + \mu} \quad (6.15)
\]

**Proposition 6.1.** Consider the false discovery proportion \(\beta\) and the false discovery rate \(q\). When \(N\) is large,

\[
\beta \xrightarrow{p} q. \quad (6.16)
\]
Proof. From (6.3),
\[ E(d_o - d|d_o) = d_o q. \]

Taking the expected value of both sides yields
\[ E(d_o - d) = E(d_o q) = qE(d_o). \]

Therefore,
\[ E(d_o) (1 - q) = E(d), \]

and hence
\[ E(d_o) = \frac{E(d)}{1 - q}. \]

Substituting (6.9) gives
\[ E(d_o) = \frac{N \mu p^2}{1 - q}. \tag{6.17} \]

Consider a sequence \( j = 1, 2, \ldots \) of populations of size \( N^{(j)} \) with \( N^{(j)} \to \infty \). Let \( d^{(j)} \) and \( d_o^{(j)} \) be \( d \) and \( d_o \) defined for the \( j \)th population.

Now the limiting distribution of \((d_o - d)/d_o\).
\[
\frac{d_o - d}{d_o} = \frac{d_o^{(j)} - d^{(j)}}{d_o^{(j)}}
= \frac{(N^{(j)})^{-1}d_o^{(j)} - (N^{(j)})^{-1}d^{(j)}}{(N^{(j)})^{-1}d_o^{(j)}}
\]

Applying (6.9) and (6.17) yields
\[
\frac{d_o - d}{d_o} \overset{p}{\rightarrow} \frac{\mu p^2}{1 - q} = q
\]
gives (6.16).
6.5 Misclassification error

In this section we derive the error in our estimator due to misclassified mother-daughter pairs. This is done in steps and presented as lemmas in following subsections. In each subsection we prove the lemmas and then confirm them with the simulated results obtained in Section 6.3.

Let $\bar{d}_o = d_o/x_o$, $E_{x_o} = E(x_o|x, d_o, d)$ and $\sigma^2_{x_o} = Var(x_o|x, d_o, d)$.

From (6.4),

$$E_{x_o} = E(x_0|x, d_o, d)$$

$$= x + (d_o - d) \left(1 - \frac{d + x}{n}\right)^2.$$

Using (6.1) and (6.2) we get

$$E_{x_o} = x + \left(\frac{\beta d}{1 - \beta}\right) \left(1 - \frac{d + x}{n}\right)^2. \quad (6.18)$$

Similarly,

$$\sigma^2_{x_o} = (d_o - d) \left(1 - \frac{d + x}{n}\right)^2 \left(1 - \left(1 - \frac{d + x}{n}\right)^2\right)$$

$$= (d_o - d) \left(\frac{d + x}{n}\right) \left(2 - \frac{d + x}{n}\right) \left(1 - \frac{d + x}{n}\right)^2$$

$$= \left(\frac{\beta d}{1 - \beta}\right) \left(\frac{d + x}{n}\right) \left(2 - \frac{d + x}{n}\right) \left(1 - \frac{d + x}{n}\right)^2. \quad (6.19)$$

6.5.1 Misclassification error in $\bar{d}$

Recall that $\bar{d} = \frac{d}{x}$.

Lemma 6.2.

$$\bar{d}_o \approx \bar{d}\delta_1$$
Where,
\[ \delta_1 = \frac{1}{1 - \beta + \beta \left( \frac{\mu p}{1 - e^{-\mu p}} \right)(1 - \frac{1 + \mu p - e^{-\mu p}}{1 + \mu})^2}. \]

**Proof.** We first consider the third order Taylor expansion of \( \bar{d}_o \) around \( x_o = E_{x_o} \).
Then we show that the second and third terms of the expansion are negligible and hence \( \bar{d}_o \) can be approximated by the first term of the expansion. The third order Taylor expansion of \( \bar{d}_o \) around \( x_o = E_{x_o} \) gives

\[ \bar{d}_o = \frac{d_o}{E_{x_o}} - \frac{d_o}{E_{x_o}^2} (x_o - E_{x_o}) + \frac{d_o}{E_{x_o}^3} (x_o - E_{x_o})^2 + O((x_o - E_{x_o})^3) \quad (6.20) \]

To show that the second term in (6.20) is negligible, we consider

\[ \Pr \left( \frac{d_o}{E_{x_o}^2} | x_o - E_{x_o} | \geq \epsilon \bigg| x, d_o, d \right) = \Pr \left( | x_o - E_{x_o} | \geq \epsilon \frac{E_{x_o}^2}{d_o} \bigg| x, d_o, d \right) \leq \frac{\sigma_{x_o}^2 d_o^2}{\epsilon^2 E_{x_o}^4} \text{ from Chebyshev’s inequality.} \]

Substituting (6.2), (6.18) and (6.19),

\[ \frac{\sigma_{x_o}^2 d_o^2}{\epsilon^2 E_{x_o}^4} = \frac{\beta d^3 \left( \frac{d+1}{n} \right)(2 - \frac{d+1}{n})(1 - \frac{d+1}{n})^2}{\epsilon^2 (1 - \beta)^3 \left( x + \left( \frac{\beta d}{1 - \beta} \right) \left( 1 - \frac{d+1}{n} \right)^2 \right)^4} \]

\[ = \frac{\beta \left( \frac{d}{n} \right)^3 \left( \frac{d+1}{n} \right)(2 - \frac{d+1}{n})(1 - \frac{d+1}{n})^2}{\epsilon^2 (1 - \beta)^3 n \left( \frac{x}{n} + \left( \frac{\beta}{1 - \beta} \right) \left( 1 - \frac{d+1}{n} \right)^2 \right)^4} \]

\[ = \frac{\beta \left( \frac{d}{n} \right)^2 \left( \frac{d+1}{n} \right)(2 - \frac{d+1}{n})(1 - \frac{d+1}{n})^2}{\epsilon^2 (1 - \beta)^3 n \left( 1 + \left( \frac{\beta}{1 - \beta} \right) \left( \frac{d}{n} \right) \left( 1 - \frac{d+1}{n} \right)^2 \right)^4}. \]

Applying the limits (6.10), (6.13), (6.15) and \( n \xrightarrow{p} N(1 + \mu)p \),

\[ \frac{\sigma_{x_o}^2 d_o^2}{\epsilon^2 E_{x_o}^4} \xrightarrow{p} \frac{\beta \left( \frac{\mu p}{1 + \mu} \right)^2 \left( \frac{\mu p}{1 - e^{-\mu p}} \right)^2 (2 - \frac{1 + \mu p - e^{-\mu p}}{1 + \mu})^2 \left( 1 - \frac{1 + \mu p - e^{-\mu p}}{1 + \mu} \right)^2}{\epsilon^2 (1 - \beta)^3 N(1 + \mu)p \left( 1 + \left( \frac{\beta}{1 - \beta} \right) \left( \frac{\mu p}{1 - e^{-\mu p}} \right) \left( 1 - \frac{1 + \mu p - e^{-\mu p}}{1 + \mu} \right)^2 \right)^4}. \]
As $N$ becomes large, $\sigma^2_{d_o} \epsilon E_{X_o}^3$ approaches zero. Therefore, the second term of (6.20) is very small and hence we ignore it here.

Now, to show the third term in (6.20), we consider

$$Pr\left(\frac{d_o}{E_{X_o}^3}(x_o - E_{x_o})^2 \geq \epsilon \mid x, d_o, d\right) = Pr\left(\sqrt{\frac{d_o}{E_{X_o}^3}}|x_o - E_{x_o}| \geq \sqrt{\epsilon} \mid x, d_o, d\right)$$

$$= Pr\left(|x_o - E_{x_o}| \geq \sqrt{\frac{E_{X_o}^3 \epsilon}{d_o}} \mid x, d_o, d\right)$$

$$\leq \frac{\sigma^2_{d_o} d_o}{\epsilon E_{X_o}^3}$$

again from Chebyshev’s inequality.

Again substituting (6.2), (6.18) and (6.19),

$$\frac{\sigma^2_{d_o} d_o}{\epsilon E_{X_o}^3} = \frac{\beta d_o \left(\frac{d + x}{n}\right) \left(2 - \frac{d + x}{n}\right) \left(1 - \frac{d + x}{n}\right)^2}{\epsilon (1 - \beta)^2 \left(x + \left(\frac{\beta d}{1 - \beta}\right) \left(1 - \frac{d + x}{n}\right)^2\right)^3}$$

$$= \frac{\beta \left(\frac{d}{n}\right)^2 \left(\frac{d + x}{n}\right) \left(2 - \frac{d + x}{n}\right) \left(1 - \frac{d + x}{n}\right)^2}{\epsilon (1 - \beta)^2 \left(n + \left(\frac{\beta}{1 - \beta}\right) \left(1 - \frac{d + x}{n}\right)\right)^3}$$

$$= \frac{\beta \left(\frac{d}{n}\right)^2 \left(\frac{d + x}{n}\right) \left(2 - \frac{d + x}{n}\right) \left(1 - \frac{d + x}{n}\right)^2}{\epsilon (1 - \beta)^2 \left(n + \left(\frac{\beta}{1 - \beta}\right) \left(1 - \frac{d + x}{n}\right)\right)^3}.$$ 

Applying the limits (6.10), (6.13), (6.15) and $n \overset{p}{\to} N(1 + \mu)p,$

$$\frac{\sigma^2_{d_o} d_o}{\epsilon E_{X_o}^3} \overset{p}{\to} \frac{\beta \left(\frac{\mu p}{1 + \mu}\right) \left(\frac{\mu p}{1 - e^{-\mu p}}\right) \left(1 + \mu p - e^{-\mu p}\right) \left(2 - \frac{1 + \mu p - e^{-\mu p}}{1 + \mu}\right) \left(1 - \frac{1 + \mu p - e^{-\mu p}}{1 + \mu}\right)^2}{\epsilon (1 - \beta)^2 \left(n + \left(\frac{\beta}{1 - \beta}\right) \left(1 - \frac{\mu p}{1 - e^{-\mu p}}\right)\right)^3}.$$ 

As $N$ becomes large, $\frac{\sigma^2_{d_o} d_o}{\epsilon E_{X_o}^3}$ approaches zero. Therefore the third term of (6.20) is very small and hence we also ignore it here.
Now from (6.20) we get

\[
\bar{d}_o \approx \frac{d_o}{E(x_o|x, d_o, d)}
\]

\[
= \frac{d}{x + \beta \frac{d}{1-\beta} \left(1 - \frac{d+x}{n}\right)^2}
\]

\[
= \frac{d}{x} \left( \frac{1}{1 - \beta + \frac{\beta d}{x}(1 - \frac{d+x}{n})^2} \right)
\]

\[
= \bar{d} \left( \frac{1}{1 - \beta + \beta d(1 - \frac{d+x}{n})^2} \right).
\]  \(6.21\)

Applying the limits (6.10) and (6.15) in (6.21) gives

\[
\bar{d}_o \approx \bar{d} \left( \frac{1}{1 - \beta + \beta \frac{\mu p}{1+\mu e^{-\mu p}} \left(1 - \frac{1+\mu p-e^{-\mu p}}{1+\mu} \right)^2} \right)
\]

resulting in Lemma 6.2.

Figure 6.3 plots \(\bar{d}_o/\bar{d}\) against false discovery proportion, \(\beta\). The blue dotted line represents the theoretical results, i.e. \(\delta_1\). The red solid line represents the simulated results. The simulated results are closely aligned with the theoretical results except when \(N\) is small and \(\mu\) is large. This illustrates Lemma 6.2. Our results are asymptotic and hence deviate slightly from the simulated results for small \(N\). This deviation is very small for small \(\beta\). However, note that the plots in Figure 6.3 for different \(N\) and \(\mu\) are in different scales, otherwise this deviation is small in magnitude.
Figure 6.3: $\frac{\delta_d}{d}$ against $\beta$

Columns 1-4 have the plots for $N=50, 100, 200$ and $500$ respectively. Rows 1-4 have the plots for $\mu=1.25, 2, 5$ and $10$ respectively. The blue dotted line represents the theoretical results and the red solid line represents the simulated results.
6.5.2 Misclassification error in $\hat{\mu p}$

Let $\hat{\mu p}_o$ be the $\hat{\mu p}$ computed with $d_o$ and $x_o$.

**Lemma 6.3.**

$$\hat{\mu p}_o \approx \hat{\mu p} + \delta_2$$

Here,

$$\delta_2 = \frac{\mu p}{1 - e^{-\mu p}} \delta_1 + \text{Lambert}\left(\frac{-\mu p}{1 - e^{-\mu p}} \delta_1 e^{\frac{-\mu p}{1 - e^{-\mu p}} \delta_1}\right) - \mu p$$

and $\delta_1$ as defined in Lemma 6.2.

**Proof.** Recall the estimator (2.11) for $\mu p$. Then,

$$\hat{\mu p}_o = \tilde{d}_o + \text{Lambert}(-\tilde{d}_o e^{-\tilde{d}_o}) \approx \tilde{d} \delta_1 + \text{Lambert}(-\tilde{d} \delta_1 e^{-\tilde{d} \delta_1})$$

So that

$$\hat{\mu p}_o - \hat{\mu p} \approx \tilde{d} \delta_1 + \text{Lambert}(-\tilde{d} \delta_1 e^{-\tilde{d} \delta_1}) - \hat{\mu p}. \quad (6.22)$$

Applying the limits (6.15) and (6.11) on (6.22) we get

$$\hat{\mu p}_o - \hat{\mu p} \approx \frac{\mu p}{1 - e^{-\mu p}} \delta_1 + \text{Lambert}\left(\frac{-\mu p}{1 - e^{-\mu p}} \delta_1 e^{\frac{-\mu p}{1 - e^{-\mu p}} \delta_1}\right) - \mu p$$

resulting in Lemma 6.3.

As in Lemma 6.2, here again we confirm Lemma 6.3 with simulations. Figure 6.4 plots $\hat{\mu p}_o - \hat{\mu p}$ against false discovery proportion, $\beta$. The blue dotted line represents the theoretical results, i.e. $\delta_2$. The red solid line represents the simulated results. Again the simulated results align well with the theoretical results except when $N$ is small and $\mu$ is very large providing further evidence to Lemma 6.3. Recall that our results are asymptotic and hence deviate slightly from simulated results for small $N$ and the plots in Figure 6.4 for different $N$ and $\mu$ are in different scales, otherwise this deviation is small in magnitude.
Figure 6.4: $\hat{\mu}_p - \hat{\mu}$ against $\beta$

Columns 1-4 have the plots for $N=50, 100, 200$ and $500$ respectively. Rows 1-4 have the plots for $\mu=1.25, 2, 5$ and $10$ respectively. The blue dotted line represents the theoretical results and the red solid line represents the simulated results.
6.5.3 Misclassification error in \( \hat{\mu} \)

Let \( \hat{\mu}_o \) be the \( \hat{\mu} \) computed with \( d_o \) and \( x_o \).

**Lemma 6.4.**

\[ \hat{\mu}_o \approx \hat{\mu} + \delta_3 \]

Here,

\[ \delta_3 = (1 + \mu) \left( \frac{(1 - e^{-(\mu p + \delta_2)})(1 - \beta)}{(1 - \beta)(1 - e^{-\mu p}) + \mu p \beta (1 - \frac{1 + \mu p - e^{-\mu p}}{1 + \mu})^2} - 1 \right), \]

\( \delta_1 \) as defined in Lemma 6.2 and \( \delta_2 \) as defined in Lemma 6.3.

**Proof.** Recall the estimator (2.10) for \( \mu \). Now,

\[ \hat{\mu}_o = \frac{n(1 - e^{-\tilde{\mu}_p o})}{x_o} - 1. \] (6.23)

We first consider the third order Taylor expansion of \( \hat{\mu}_o \) around \( x_o = E(x_o|x, d_o, d) \) (i.e. \( E_{x_o} \)). Then we show that the second and third terms of the expansion are negligible and hence \( \hat{\mu}_o \) can be approximated by the first term of the expansion.

The third order Taylor expansion of \( \hat{\mu}_o \) around \( E_{x_o} \) gives,

\[ \tilde{\mu}_o = \frac{n(1 - e^{-\tilde{\mu}_p o})}{E_{x_o}} - 1 - \frac{n(1 - e^{-\tilde{\mu}_p o})}{E_{x_o}^2} (x_o - E_{x_o}) + \frac{n(1 - e^{-\tilde{\mu}_p o})}{E_{x_o}^3} (x_o - E_{x_o})^2 + O((x_o - E_{x_o})^3). \] (6.24)

To show the second term in (6.24), we consider

\[ Pr \left( \frac{n(1 - e^{-\tilde{\mu}_p o})}{E_{x_o}^2} |x_o - E_{x_o}| \geq \epsilon \mid x_o, d_o, d \right) = Pr \left( |x_o - E_{x_o}| \geq \epsilon n \frac{E_{x_o}^2}{n(1 - e^{-\tilde{\mu}_p o})} \mid x_o, d_o, d \right) \]

\[ \leq \frac{\sigma_{x_o}^2 n^2 (1 - e^{-\tilde{\mu}_p o})^2}{\epsilon^2 E_{x_o}^4} \]

from Chebyshev’s inequality.

(Here, note that \( \tilde{\mu}_p o \) is a constant for given \( x, d_o, d \).)
Substituting (6.18), (6.19) and using Lemma 6.3 gives

\[
\frac{\sigma^2_{\bar{x}_o} n^2 (1 - e^{-\bar{\mu}p})^2}{e^2 E_{\bar{x}_o}^4} = \beta d \left( \frac{d+\bar{x}}{n} \right) (2 - \frac{d+\bar{x}}{n}) (1 - \frac{d+\bar{x}}{n})^2 \frac{(1 - e^{-\bar{\mu}p})^2}{e^2 (1 - \beta) \left( x + (\beta d \frac{d+\bar{x}}{n}) (1 - \frac{d+\bar{x}}{n})^2 \right)^4} = \beta d \frac{d+\bar{x}}{n} \frac{d+\bar{x}}{n} (2 - \frac{d+\bar{x}}{n}) (1 - \frac{d+\bar{x}}{n})^2 (1 - e^{-\bar{\mu}p})^2 \frac{\epsilon^2}{e^2 (1 - \beta) \left( x + (\beta d \frac{d+\bar{x}}{n}) (1 - \frac{d+\bar{x}}{n})^2 \right)^4}.
\]

Applying the limits (6.13), (6.14), (6.15), (6.11) and \( n \to N(1 + \mu)p \),

\[
\frac{\sigma^2_{\bar{x}_o} n^2 (1 - e^{-\bar{\mu}p})^2}{e^2 E_{\bar{x}_o}^4} \xrightarrow{p} \frac{\beta}{e^2 (1 - \beta)} \frac{(\mu p \frac{1+\mu p}{1+\mu}) (1+\mu p-e^{-\mu p}) (2 - 1+\mu p-e^{-\mu p}) (1 - 1+\mu p-e^{-\mu p})^2 (1 - e^{-\mu p})^2}{N(1 + \mu)p \left( 1+\mu p \frac{1+\mu p}{1+\mu} + (\mu p \frac{1+\mu p}{1+\mu}) (1 - 1+\mu p-e^{-\mu p})^2 \right)^4}.
\]

As \( N \) becomes large, \( \sigma^2_{\bar{x}_o} n^2 (1 - e^{-\bar{\mu}p})^2 / e^2 E_{\bar{x}_o}^4 \) approaches zero. Therefore the second term of (6.24) is very small and hence we ignore it here.

Also, to show the third term in (6.24), we consider

\[
Pr \left( \frac{n(1 - e^{-\bar{\mu}p})}{E_{\bar{x}_o}^3} (x_o - E_{\bar{x}_o})^2 \geq \epsilon \mid x, d_o, d \right) = Pr \left( \sqrt{\frac{n(1 - e^{-\bar{\mu}p})}{E_{\bar{x}_o}^3}} |x_o - E_{\bar{x}_o}| \geq \sqrt{\epsilon} \mid x, d_o, d \right) = Pr \left( |x_o - E_{\bar{x}_o}| \geq \sqrt{\frac{E_{\bar{x}_o}^3 \epsilon}{n(1 - e^{-\bar{\mu}p})}} \mid x, d_o, d \right) \leq \frac{\sigma^2_{\bar{x}_o} n(1 - e^{-\bar{\mu}p})}{\epsilon E_{\bar{x}_o}^3}
\]

again from Chebyshev's inequality.
Substituting (6.18), (6.19) and Lemma 6.3,

\[
\frac{\sigma_{x_o}^2 n(1 - e^{-\widehat{\mu} p_o})}{\epsilon E_{x_o}^3} = \frac{\beta d(\frac{d + x}{n})(2 - \frac{d + x}{n})(1 - \frac{d + x}{n})^2 n(1 - e^{-\widehat{\mu} + \delta_2})}{\epsilon(1 - \beta) \left( x + (\frac{\beta d}{1 - \beta}) \left( 1 - \frac{d + x}{n} \right)^2 \right)^3} = \frac{\beta \left( \frac{d + x}{n} \right) (2 - \frac{d + x}{n})(1 - \frac{d + x}{n})^2 (1 - e^{-\widehat{\mu} + \delta_2})}{\epsilon(1 - \beta) n \left( \frac{z}{n} + (\frac{d}{n}) (\frac{\beta}{1 - \beta}) \left( 1 - \frac{d + x}{n} \right)^2 \right)^3}.
\]

Applying the limits (6.13), (6.14), (6.15), (6.11) and \( n \xrightarrow{p} N(1 + \mu)p \),

\[
\frac{\sigma_{x_o}^2 n(1 - e^{-\widehat{\mu} p_o})}{\epsilon E_{x_o}^3} \xrightarrow{p} \frac{\beta \beta \left( \frac{d + x}{n} \right) (2 - \frac{d + x}{n})(1 - \frac{d + x}{n})^2 (1 - e^{-\widehat{\mu} + \delta_2})}{\epsilon(1 - \beta) \left( (\frac{\mu}{1 + \mu}) \left( 1+\mu p-e^{-\mu p} \right)(2 - \frac{1+\mu p-e^{-\mu p}}{1+\mu})(1 - \frac{1+\mu p-e^{-\mu p}}{1+\mu})^2 \left( 1 - e^{-\mu p - \delta_2} \right) \right)^3 \frac{1 - (\frac{\mu}{1 + \mu}) + (\frac{\mu}{1 + \mu})}{1 - (\frac{1+\mu p-e^{-\mu p}}{1+\mu})^3}}.
\]

As \( N \) becomes large, \( \sigma_{x_o}^2 n(1 - e^{-\widehat{\mu} p_o})/\epsilon E_{x_o}^3 \) approaches zero. Therefore the third term of (6.24) is very small and hence we ignore it here.

Now, from (6.24) we get

\[
\hat{\mu}_o \approx \frac{n(1 - e^{-\widehat{\mu} p_o})}{E_{x_o}} - 1 = \frac{n(1 - e^{-\widehat{\mu} + \delta_2})}{x + \beta \left( \frac{\beta}{1 - \beta} \right) \left( 1 - \frac{d + x}{n} \right)^2} - 1 = \frac{\frac{z}{n} + \frac{d}{n} (\frac{\beta}{1 - \beta}) (1 - \frac{d + x}{n})^2}{1 - 1 - e^{-\widehat{\mu} + \delta_2}} - 1.
\]

Hence,

\[
\hat{\mu}_o - \hat{\mu} \approx \frac{1 - e^{-\widehat{\mu} + \delta_2}}{\frac{z}{n} + \frac{d}{n} (\frac{\beta}{1 - \beta}) (1 - \frac{d + x}{n})^2} - 1 - \hat{\mu}.
\]
Applying the limits (6.11), (6.12), (6.13), (6.14) and (6.15) on (6.26) gives,

\[
\hat{\mu}_o - \hat{\mu} \approx \frac{1 - e^{-(\mu p + \delta_2)}}{1 + \mu} + \frac{\mu p}{1 + \mu} \left(1 - \frac{1 + \mu p - e^{-\mu p}}{1 + \mu}\right)^2 - 1 - \mu
\]

\[
= (1 + \mu) \left(\frac{1 - e^{-(\mu p + \delta_2)}}{(1 - \beta)(1 - e^{-\mu p}) + \mu p \beta(1 - \frac{1 + \mu p - e^{-\mu p}}{1 + \mu})^2 - 1}\right)
\]

(6.27)

Figure 6.5 plots \(\hat{\mu}_o - \hat{\mu}\) against \(\beta\). The blue dotted line represents the theoretical results, i.e. \(\delta_3\). The red solid line represents the simulated results. The simulated results overlap the theoretical results except when \(N\) and \(\mu\) are small providing further evidence of Lemma 6.4.
Figure 6.5: $\hat{\mu}_o - \hat{\mu}$ against $\beta$

Columns 1-4 have the plots for $N=50, 100, 200$ and $500$ respectively. Rows 1-4 have the plots for $\mu=1.25, 2, 5$ and $10$ respectively. The blue dotted line represents the theoretical results and the red solid line represents the simulated results.
6.5.4 Misclassification error in $\hat{N}$

**Theorem 6.5.** The proportion of error on $\hat{N}$ due to misclassification when the false discovery proportion is $\beta$,

$$\frac{\hat{N}_o - \hat{N}}{\hat{N}} \approx \frac{(1 + \mu)^2(\mu + \delta_3)(1 - \beta)}{\mu(1 + \mu + \delta_3)^2} - 1.$$

Here, $\delta_1$ as defined in Lemma 6.2, $\delta_2$ as defined in Lemma 6.3 and $\delta_3$ as defined in Lemma 6.4.

**Proof.** Recall the estimator (2.9) for $\mu$. Now,

$$\hat{N}_o = \frac{\hat{\mu}_o n(n - 1)}{d_o(1 + \hat{\mu}_o)^2}. \tag{6.28}$$

Applying (6.2) and Lemma 6.4 in (6.28),

$$\hat{N}_o \approx \frac{(\hat{\mu} + \delta_3)n(n - 1)(1 - \beta)}{d(1 + \hat{\mu} + \delta_3)^2}. \tag{6.29}$$

Now,

$$\hat{N}_o - \hat{N} \approx \frac{(\hat{\mu} + \delta_3)n(n - 1)(1 - \beta)}{d(1 + \hat{\mu} + \delta_3)^2} - \hat{N} = \frac{(\hat{\mu} + \delta_3)(1 - \beta)}{(1 + \hat{\mu} + \delta_3)^2} \frac{\hat{N}(1 + \hat{\mu})^2}{\hat{\mu}} - \hat{N}. \tag{6.29}$$

Applying the limit (6.12) on (6.29) gives

$$\frac{\hat{N}_o - \hat{N}}{\hat{N}} \approx \hat{N} \left( \frac{(1 + \mu)^2(\mu + \delta_3)(1 - \beta)}{\mu(1 + \mu + \delta_3)^2} - 1 \right),$$

$$\frac{\hat{N}_o - \hat{N}}{\hat{N}} \approx \frac{(1 + \mu)^2(\mu + \delta_3)(1 - \beta)}{\mu(1 + \mu + \delta_3)^2} - 1,$$

$$\frac{\hat{N}_o - \hat{N}}{\hat{N}} \% \approx \left( \frac{(1 + \mu)^2(\mu + \delta_3)(1 - \beta)}{\mu(1 + \mu + \delta_3)^2} - 1 \right) \times 100.$$
Figure 6.6 plots the percentage of error on $\hat{N}$ due to misclassification, i.e. $((\hat{N}_o - \hat{N})/\hat{N})\%$, against the false discovery proportion $\beta$. The blue dotted line represents the theoretical results, i.e. $((1 + \mu)^2(\mu + \delta)(1 - \beta)/\mu(1 + \mu + \delta)^2 - 1) \times 100$. The red solid line represents the simulated results. The simulated results fairly coincide with the theoretical results providing evidence for Theorem 6.5.
Figure 6.6: $\frac{\hat{N}_o - \hat{N}}{N}$% against $\beta$

Columns 1-4 have the plots for $N=50, 100, 200$ and $500$ respectively. Rows 1-4 have the plots for $\mu=1.25, 2, 5$ and $10$ respectively. The blue dotted line represents the theoretical results and the red solid line represents the simulated results.
In our study we control the false discovery rate at 0.05. Recall that this is the expected value of false discovery proportion ($\beta$). Therefore we plot the percentage of bias caused in $\hat{N}$, i.e. $((\hat{N}_o - \hat{N})/\hat{N})\%$ due to misclassification, against $\hat{N}$ when the false discovery proportion ($\beta$) is 0.05 in Figure 6.7. The solid line represents the theoretical results and dotted line represents the simulated results. It can be seen that the proportion of error caused by misclassification is very small. The theoretical model seems to describe the simulated data well except when $N$ and $\mu$ are both small. In fact, the theoretical model appears to provide an upper bound for the percentage bias. Even when $N$ and $\mu$ are both small, the discrepancy is very small.

![Figure 6.7: Percentage of error of $\hat{N}$ caused by misclassification against $\hat{N}$ when $\beta = 0.05$]

The solid line represents the theoretical results and the dotted line represents the simulated results. the different colours represent values of $\mu$ as given in the legend.

Then, again, we plot the percentage bias of $\hat{N}$, i.e. $((\hat{N}_o - \hat{N})/\hat{N})\%$, when $\beta$ is larger ($\beta=0.1$ and 0.15) in Figure 6.8. Again, it can be seen that this bias increases as the false discovery proportion increases and is large for small $N$. Again, the
misclassification seems to describe the simulated data well. Recall that our model is asymptotic and hence deviates slightly for small $N$. 
Figure 6.8: Percentage of error of $\hat{N}$ caused by misclassification against $\hat{N}$ when $\beta = 0.1, \beta = 0.15$

The solid line represents the theoretical results and the dotted line represents the simulated results. The different colours represent values of $\mu$ as given in the legend.
6.6 Discussion

In this chapter we study the effect of misclassification of relationships on our estimators. The misclassification we consider here is the one that may occur when the genotype data are used for determining the relationship. Here we consider only false positives, i.e. the unrelated pairs that are labelled as related, assuming that the number of false negatives are small and hence negligible. In our work we adopt the approach using genotype data to identify parent-offspring pairs of Skaug et al. [34]. Here we only consider the moment estimators for a population with two generations derived in Chapter 2.

Based on the simulations, we show that the effect on the estimates increases as the false discovery proportion increases. Then we model this effect based on the false discovery rate concept and limiting distributions. We do this in steps, modelling the effect on \( \bar{d}, \hat{\mu}_p, \hat{\mu} \) and \( \hat{N} \). These theoretical models well describe the simulated results. Our results are asymptotic and hence deviate slightly from simulated results for small \( N \). However, this deviation is small in magnitude.

Further, we show that this effect is very small when the false discovery proportion is small (0.05). When the false discovery rate is 0.05, the population size is underestimated slightly for small \( \mu \) (e.g. \( \mu \leq 5 \)) and overestimated slightly for large \( \mu \) (e.g. \( \mu = 10 \)). With the false discovery rates 0.1 and 0.15, we observe a similar pattern, with an increased bias.

This can be extended to model the effect of the misclassification of relationships on the other estimators we have proposed in Chapters 4 and 5.
Chapter 7

Conclusions and future work

In this thesis we proposed methods for inference from a single occasion capture experiments using genetic markers. Even though this is a single occasion capture experiment, the relationships identified using genetic markers allowed us to treat the experiment as a capture-recapture experiment. Throughout the thesis we considered only the female population.

We considered a population with only two generations in Chapters 2 and 3. We used the mother-daughter relationships only in the sample for inference in these two chapters. In Chapter 2 we estimated the number of mothers in the population along with two nuisance parameters, namely the mean number of daughters per mother and the capture probability using the method of moments. In Chapter 3 we estimated our parameters using the method of maximum likelihood. We showed some problems with the likelihood approach limiting its use for estimation and suggest the method of moments as a better approach in this scenario. Therefore, the method of moments was used in the remainder of the thesis for estimation. A population with three generations was considered in Chapter 4. We proposed two estimators for the size of the first female generation, which we referred to as mothers. The first estimator used only the mother-daughter relationship, while the second estimator used the mother-daughter-granddaughter relationship as well. Due to the practical difficulty of identifying mother-daughter-granddaughter triples and better performance of the first estimator, we recommended it for use. Further, in this chapter we proposed a generalisation for the estimator of the number of females in the first generation for
Conclusions

In Chapter 5, we further modified the estimators proposed in Chapters 2 and 4 by incorporating the sister pairs to the estimator. With this modification we managed to substantially reduce the standard errors of the estimators. In developing all estimators in Chapters 2, 3, 4 and 5 we assumed the relationships are perfectly identifiable. In our application, we adopted an approach using genetic markers to identify the biological relationships. In Chapter 6 we showed that the impact of any possible misclassification of this approach on our estimator proposed in Chapter 2 is small. We further proposed an approach to quantify the impact of this misclassification error.

In our work we considered a female population only. This population structure is well suited to the pygmy possums, where the females inhabit the prime breeding locations, where there is high food availability while the males disperse after the breeding season. Extending this method to a population with both males and females would be of interest for future research.

One important assumption we made throughout the thesis is that \( \mu \geq 1 \). This seems natural for most populations where the maternal generation is smaller in size than the offspring generation. A female pygmy possum can reproduce annually, with four offspring at a time from an age of one year to an age of twelve years, thus this assumption is valid for a pygmy possum population. However, we made this assumption mostly for technical reasons. In Section 2.9 we presented the problems that can occur when \( \mu < 1 \). In fact, the estimators worked better for large \( \mu \). A modification to suit populations with \( \mu < 1 \) would be interesting.

Also, our work is based on the assumption that the number of daughters of a given mother has a Poisson(\( \mu \)) distribution where \( \mu \geq 1 \). A natural generalisation is the Poisson-gamma mixture model [12]. This can be used to extend the proposed method to a more practical framework while relaxing our basic assumption on the distribution of the number of daughters per mother. Fisher’s [12] original model was formulated for sampling butterfly populations, but it can be directly adapted to the context of genetics. That is, the generalised model assumes that the number of daughters of the given \( i \)th mother follows a Poisson(\( \mu_i \)) distribution, and \( \mu_1, \mu_2, \ldots, \mu_N \) can be regarded as a random sample from a two-parameter gamma distribution. The unconditional number of daughters then follows a negative binomial
distribution. With including one additional parameter, the Poisson framework can be extended to a more practical and more general model. In addition, the restrictive assumption, that \( \mu \geq 1 \), is not needed in the Poisson-gamma model.

In Chapter 4 we presented estimators for a population with three generations and then propose a generalisation for a population with any known number of generations. However, determining the number of generations is beyond the current work and is an area for future work. This requires additional information on individuals such as the age of the observed individual and their reproductive behaviour. Determining the number of generations in a population would be an interesting yet challenging research problem.

Further, we used only mother-daughter pairs in our estimators in Chapters 2, 3 and 4 and incorporated sister pairs in Chapter 5. Our estimators can be used only if at least one related pair is identified. It is recommended that we capture at least one mother with a few daughters for the estimator to work well. In some widely spread populations a sample may not consist of at least a single mother-daughter or sister pair, making our estimator redundant. Incorporating other relationships, such as cousins, will increase the usability of the estimators, while improving the precision. This in return raises the problem of identifying the distant relationships using genetic markers.

Another restrictive assumption in the proposed model is that “equal-catchability” assumption, i.e. each individual in the population has the same probability of being captured. It would be interesting to see the impact of the heterogeneity of capture probability.

In Section 2.4 we show that our method of moments estimator of \( N \) is consistent as the population size tends to infinity. However, the population size does not tend to infinity in practice, especially in presumably closed populations. It would be interesting to determine a guideline regarding how large the population size should be for the validity of consistency. Further, we do not discuss the consistency of our maximum likelihood estimator we proposed in Chapter 3. This is another problem of interest for future work. It would be interesting also to examine whether the two types of estimators are asymptotically equivalent.
Bibliography


