A FLEXIBLE, DISCRETE AND SMOOTH CAPTURE-RECAPTURE MODEL BASED UPON COUNTS OF REPEATED IDENTIFICATIONS USING VALIDATION SAMPLES

BY

MISS PARAWAN PIJITRATTANA

A DISSERTATION SUBMITTED IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY (STATISTICS)
DEPARTMENT OF MATHEMATICS AND STATISTICS
FACULTY OF SCIENCE AND TECHNOLOGY
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2018

A FLEXIBLE, DISCRETE AND SMOOTH CAPTURE-RECAPTURE MODEL BASED UPON COUNTS OF REPEATED IDENTIFICATIONS USING VALIDATION SAMPLES

BY

MISS PARAWAN PIJITRATTANA

A DISSERTATION SUBMITTED IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY (STATISTICS)
DEPARTMENT OF MATHEMATICS AND STATISTICS
FACULTY OF SCIENCE AND TECHNOLOGY
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2018

THAMMASAT UNIVERSITY
FACULTY OF SCIENCE AND TECHNOLOGY

DISSERTATION

BY

MISS PARAWAN PIJITRATTANA

ENTITLED

A FLEXIBLE, DISCRETE AND SMOOTH CAPTURE-RECAPTURE MODEL BASED UPON COUNTS OF REPEATED IDENTIFICATIONS USING VALIDATION SAMPLES

was approved as partial fulfillment of the requirements for
the degree of Doctor of Philosophy (Statistics)

on July 24 , 2019

Chairman Chukiat Viwatwongkasem
(Associate Professor Chukiat Viwatwongkasem, Ph.D.)

Member and Advisor Rattana Lerdsuwansri
(Assistant Professor Rattana Lerdsuwansri, Ph.D.)

Member Wararit Panichkitkosolkul
(Associate Professor Wararit Panichkitkosolkul, Ph.D.)

Member Patchanok Srisuradetchai
(Assistant Professor Patchanok Srisuradetchai, Ph.D.)

Member Monthira Duangsaphon
(Monthira Duangsaphon, Ph.D.)

Dean
(Associate Professor Pakorn Sermsuk, M.Sc.)

Thesis Title	A FLEXIBLE, DISCRETE AND SMOOTH CAPTURE-RECAPTURE MODEL BASED UPON COUNTS OF REPEATED IDENTIFICATIONS USING VALIDATION SAMPLES
Author	Miss Parawan Pijitrattana
Degree	Doctor of Philosophy (Statistics)
Department/Faculty/University	Mathematics and Statistics Faculty of Science and Technology Thammasat University
Thesis Advisor	Assistant Professor Rattana Lerdsuwansri, Ph.D.
Academic Year	2018

ABSTRACT

Capture-recapture is a powerful method for estimating the size of an elusive target population. A new estimator of the population size is proposed by allowing the Normal mixing density leading to the Hermite distribution. The EM algorithm is discussed for maximum likelihood estimation.

The discrete mixture of a Hermite distribution is adopted to model the heterogeneity of an unobserved population. The parameters of zero-truncated count mixture of Hermite distributions are estimated by using a PMLE with the EM algorithm. The penalized maximum likelihood estimator is proposed based on zero-truncated Hermite distribution through the Horvitz-Thomson approach.

The development of estimation to include validation information in the capture-recapture modeling is introduced, to increase the accuracy and efficiency of population size estimation. The nonparametric maximum likelihood estimator is developed based on zero-truncated Hermite distribution, which includes validation information.

With regard to making inference about the unknown size N of the population, confidence interval estimations are proposed. The profile mixture likelihood is used to construct the confidence intervals for the population size N. A simulation study was conducted to compare the performance of the proposed estimators with the maximum likelihood Poisson, Turing, Chao and Censored estimators. The proposed methods were applied to estimate the number of heroin users in Chiang Mai from 2016 to 2018.

Keywords: Capture-recapture, Poisson-Normal Mixture, EM algorithm, Mixture of truncated Hermite distributions

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and gratitude to my advisor Assistant Professor Dr. Rattana Lerdsuwansri for her patience and guidance. I am very lucky to have this opportunity. It is an incredible experience to work with her. I have to specially thank Professor Dr. Dankmar Böhning for his inspiration and guidance throughout this research. Also, I would like to thank Assistant Professor Dr. Krisana Lanumteang for the data of heroin users in Chiang Mai, Thailand.

I am also grateful to my committee members, Associate Professor Dr.Chukiat Viwatwongkasem, the chairman, Associate Professor Dr. Wararit Panichkitkosolkul, Assistant Professor Dr. Patchanok Srisuradetchai, and Dr. Monthira Duangsaphon for their constructive comments and suggestions.

I am very thankful to all those who helped me, including the faculties and staffs members of the Department of Mathematics and Statistics, Thammasat University for their assistance and encouragement.I gratefully acknowledge the financial support provided by Science Achievement Scholarship of Thailand (SAST).

Last but not least, I would like to express my deep gratitude to my father and my mother to encourage and understand me in Ph.D. study. They always let me chose my own way and encourage me. Without their great care, I could never have completed this thesis.

Miss Parawan Pijitrattana

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviations	**Terms**
MLE	Maximum Likelihood Estimator
PMLE	Penalized Maximum Likelihood Estimator
NPMLE	Nonparametric Maximum Likelihood Estimator
EM algorithm	Expectation-Maximization Algorithm
BIC	Bayesian Information Criterion
Rbias	Relative Bias
RRMSE	Relative Root Mean Square Error
pgf	Probability Generating Function
CI	Confidence Interval
CP	Coverage Probability
AL	Average Length

LIST OF SYMBOLS

N	population size
n	sample size
$\hat{N}$	population size estimator
$Se(\hat{N})$	the standard error of population size estimator
$Herm$	Hermite distribiton
ν	the mean of Hermite distribiton
d	the coefficient of dispersion of Hermite distribiton
Poi	Poisson distribution
λ	the mean of Poisson distribiton
$\hat{N}_{HT}$	the well-known Horvitz-Thompson estimator
$\hat{N}_{Turing}$	Turing's estimator
$\hat{N}_{MLE}$	the maximum likelihood estimator
$\hat{N}_{Chao}$	Chao's lower bound estimator
$\hat{N}_{Censored}$	the censored estimator
$\hat{N}_{Herm}$	the maximum likelihood estimator based on Hermite distribution
$\hat{N}_{Profile}$	the profile maximum likelihood estimator based on Hermite distribution
$\hat{N}_{CMLE}$	the maximum likelihood estimator based on discrete mixtures of Hermite distribution
$\hat{N}_{PMLE}$	the penalized maximum likelihood estimator based on discrete mixtures of Hermite distribution
$\hat{N}_{profileMix}$	the profile nonparametric maximum likelihood estimator based on discrete mixture of Hermite distribution

$\hat{N}_{Valid}$	the maximum likelihood estimator based on discrete mixtures of Hermite distribution with validation information
$\hat{N}_{profileValid}$	the profile nonparametric maximum likelihood estimator based on discrete mixture of Hermite distribution with validation information
$\widehat{Var}(\hat{N}_{Turing})$	The variance of Turing's estimator
$\widehat{Var}(\hat{N}_{MLE})$	The variance of the maximum likelihood estimator
$\widehat{Var}(\hat{N}_{Chao})$	The variance of Chao's lower bound estimator
$\widehat{Var}(\hat{N}_{Censored})$	The variance ofthe censored estimator

CHAPTER 1

INTRODUCTION

1.1 Overview

A capture-recapture approach aims to estimate an elusive target population size. A large number of fields have adopted capture-recapture to estimate unobserved units as well as the total population size. Capture-Recapture methods have been applied in many areas, including biology, sociology, criminology, public health and epidemiology, for estimating the number of female grizzly bears in the Yellowstone ecosystem (Chao, 2005), the number of transport injuries in a developing country (Tercero & Andersson, 2004), the size of a criminal population (Van Der Heijden, Cruyff, & Van Houwelingen, 2003), the number of injecting drug users in Scotland (McDonald et al., 2014), or infectious disease incidence (Van Hest et al., 2008).

Capture-recapture surveys assume that the target population N is closed, with no births, deaths, or migration over the observational period. Some identification mechanisms, such as registration or trapping systems, are used to identify observed units. The identifications provide a count Y_j of the number of times the j^{th} individual was identified during the observational period, for $j = 1, 2, ..., N$. This is called positive sample. The frequency of units identified exactly $0, 1, ..., m$ times is denoted by $f_0, f_1, ..., f_m$ respectively, where m is the largest observed count. As all units in a population cannot be captured. The number of zero-count data, f_0, is unknown. An example of capture-recapture data is presented in Table 1.1. As a consequence, the number of observed units $n = f_1 + f_2 + ... + f_m$ and $N = n + f_0$. Estimation of f_0 is required to estimate the population size N.

Table 1.1: Frequencies of capture-recapture data.

i	0	1	2	3	...	m
f_i	?	f_1	f_2	f_3	...	f_m

A count distribution arises when we summarize how often a unit was identified. Let $p_i = P(Y = i)$ denote the probability of identifying a unit i times. Accordingly, p_0 is the probability that a unit is not identified. The unobserved f_0 might be replaced by the expected value Np_0. If p_0 is known then we can solve for $N = n + f_0$. The fact that p_0 is unknown but is required for estimation produces the well-known Horvitz-Thompson estimator:

$$\hat{N}_{HT} = \frac{n}{1 - \hat{p}_0}. \tag{1.1}$$

In general, count data are modeled as a Poisson distribution having parameter λ. Under the homogeneous Poisson model, the probability of each unit being identified exactly i times is $p_i = \frac{e^{-\lambda}\lambda^i}{i!}$. Several estimators have been developed to estimate the hidden population size based on a homogeneous case. These include maximum likelihood estimation and Turing's estimator (Good, 1953). Due to the fact that, in practice, λ is not identical for all units in the population it is more reasonable to assume that the target population may consist of a set of subgroups. An alternative model incorporating heterogeneity of the Poisson parameter might be more realistic. Introducing such heterogeneity into the model has been proposed in the literature (Chao, 1987; Zelterman, 1988; Niwitpong et al., 2013). The probability under heterogeneity Poisson with density $f(\lambda)$ is given by

$$p_i = \int_0^\infty \frac{e^{-\lambda}\lambda^i}{i!} f(\lambda) d\lambda. \tag{1.2}$$

The more general case is to allow gamma-mixing, which leads to a Poisson-Gamma mixture or negative binomial. For example, Rocchetti et al. (2011) and Böhning et al. (2013) proposed population size estimators under a zero-truncated Poisson-Gamma mixture model. However, this is problematic due to the boundary problem (Böhning, 2015), which results in a spurious estimate for the size N of a population, as demonstrated in Kuhnert and Böhning (2009). We propose the use of $f(\lambda)$ as the normal density. A new estimator is proposed, based on the zero-truncated Poisson-Normal distribution.

A mixture model is a flexible approach to cope with data from a population which is suspected to contain sub-populations. Böhning and Schön (2005) proposed nonparametric maximum likelihood estimation (NPMLE) of a population of size N based on the binomial or Poisson distributions. Viwatwongkasem et al. (2013) applied the NPMLE for estimating the mixture of zero-trucated Poisson distributions, when estimating the number of drug users in Thailand 2005-2007. With this motivation, the discrete mixtures of Poisson-Normal distributions is adopted to model the heterogeneity of an unobserved population, and the EM algorithm is employed for maximum likelihood estimation.

Sometimes addition information on the observed units is available from another sub-sample of the target population, called a validation sample. Such samples are smaller and contain no hidden cases, so that zero counts are observed. Let $g_0, g_1, ..., g_m$ be the frequency of units identified exactly $0, 1, 2, ..., m$ times. Note that g_0 is known. Table 1.2 provides example data of a positive sample and validation sample.

Table 1.2: Frequencies in positive sample and validation sample.

i	0	1	2	3	...	m
f_i	?	f_1	f_2	f_3	...	f_m
g_i	g_0	g_1	g_2	g_3	...	g_m

Böhning et al. (2016) incorporated information from the validation sample into the capture-recapture model. Arnold, Böhning, and Azevedo (2017) demonstrated that the use of a validation sample increases the estimation efficiency and reduces the bias. In this research, the extension of estimation including validation information is proposed, to increase the accuracy and efficiency of population size estimation.

1.2 Objectives

1. To explore a more appropriate alternative distribution for zero-truncated count data.
2. To develop a new estimator of the population size N, based on the Poisson-Normal distribution.
3. To develop a new estimator of the population size N, based on discrete mixtures of the Poisson-Normal distributions.

4. To develop a new estimator of the population size N, based on discrete mixtures of the Poisson-Normal distributions with validation samples.
5. To evaluate the efficiency of the population size estimators under various conditions.

1.3 Benefits of the research

1. To derive efficient population size estimators.
2. To apply the proposed estimators to real situations.

1.4 Thesis outline

Chapter 2 introduces capture-recapture methodology, and the research methodology of the dissertation.

In Chapter 3, Laplace's method is used to derive the Poisson-Normal mixture model and the EM algorithm is employed for maximum likelihood estimation of the parameters. The performance of the new estimator is compared with those of other well-known estimators in different cases. For making inferences about the unknown size N of a closed population, this project uses confidence interval estimation.

In Chapter 4, we propose the inclusion of discrete mixtures to increase generality. We propose an estimator when the counts are assumed to come from a zero-truncated Poisson-Normal distribution. In addition, we propose an estimator that combines maximum likelihood estimation for zero-truncated Poisson-Normal distributions with validation information. The performance of the new estimators compared with those of other well-known estimators under various conditions. Confidence intervals are constructed based upon the profile mixture likelihood.

In Chapter 5, data of heroin users in Chiang Mai (Thailand) are provided as illustrations of practical application.

The thesis ends with a chapter of conclusions and future work.

CHAPTER 2

REVIEW OF METHODOLOGY

This chapter focuses on the capture-recapture methodology and two types of data set used in capture-recapture studies. Well-known estimators based on homogeneous and heterogeneous Poisson models are provided. The methodology of this dissertation is introduced.

2.1 Capture-recapture methodology

Capture-recapture methods have been widely used to estimate the size of a population in many field including ecology, criminology, social science, public health, and epidemiology. Examples include estimating the number of females with cubs-of-the-year in the Yellowstone grizzly bear population (Keating et al., 2002), the size of the hidden scrapie population in Great Britain (Böhning & Vilas, 2008), the number of illegal immigrants living in the Netherlands from Middle East countries (Heijden et al., 2012), or the number of adult cannabinoid users in Italy (Farcomeni et al., 2013). Studies in Thailand that have used the capture-recapture methodology for estimating population sizes N include Mastro et al. (1994), who estimated the number of HIV-infected injection drug users in Bangkok. Böhning et al. (2004), who estimated the number of drug users in Bangkok in 2001, and Viwatwongkasem et al. (2013), who estimated the number of drug users in Thailand 2005-2007.

Capture-recapture surveys assume that the target population N is closed with no births, deaths, or migration over the observational period. As all units in a population cannot be collected, separate mechanisms are used to identify the observed units. A registration system can be used as the identification mechanism in human populations, and a trapping system in wildlife populations.

Intuitively, some units of the population are observed and some units are unobserved. As a consequence, the number of missing units must be estimated. Suppose that every unit has the same probability of identification given by $1 - p_0$. Since the population size N consists of an observed and unobserved part,

$$N = N(1 - p_0) + Np_0 = n + Np_0. \tag{2.1}$$

The expected number of identification units $N(1 - p_0)$ can be estimated from the number of observed units n. Equation (2.1) can be solved for N to provide the Horvitz Thompson estimator:

$$\hat{N}_{HT} = \frac{n}{1 - \hat{p}_0}. \tag{2.2}$$

To model probability identification and estimate p_0, two types of capture-recapture data with different sources and repeated counting data are taken into account.

A unit is identified at different occasions, periods, or sources of identification. Capture-recapture data with different sources was introduced by Böhning and Schön (2005). Let the units be indexed as $1, 2, ..., N$ with m trapping occasions indexed as $1, 2, ..., m$. Data are tagged as present (1) or absent (0). The data of all units are provided as an m-vector consisting of 0s and 1s. An example of capture-recapture data, for 38 deer mice with six capture occasions (Amstrup et al., 2010), is shown in Table 2.1. Let Y_j be the number of times that the j^{th} unit is identified during a period of study with m trapping occasions, $j = 1, 2, ..., N$. From Table 2.1, the 1^{st} unit was observed in all trappings, giving $Y_1 = 6$. The 2^{nd} unit was trapped in the 1^{st}, 4^{th}, 5^{th}, and 6^{th} trappings, so that $Y_2 = 4$. The vector [0, 0, 0, 0, 0, 0], indicating a case which has never been identified, is not recorded, and the number of units with $Y = 0$ is unknown. Let f_i be the frequency of units identified exactly i times, $i = 1, 2, ..., m$. The frequency counts for the 38 deer mice are shown in Table 2.2. As $Y = 0$ is not observed, the frequency of unobserved units, f_0, is unknown and becomes a crucial part of the estimation.

A unit may also be identified repeatedly by the same mechanism during the observational period. This leads to repeated counting data (Böhning & Schön, 2005). These represent how often each unit is identified. Let Y be the number of times that a unit is identified over the observational period, and let f_i be the frequency of units identified exactly i times, $i = 1, 2, ..., m$ where m is the largest observed count. Examples of repeated counting data include the number of illegal immigrants in four large cities in the Netherlands (Van Der Heijden, Bustami, et al., 2003) and the number of

Table 2.1: Individual capture history of 38 deer mice with capture occasions.

unit	Occasion						Y_j
j	1	2	3	4	5	6	
1	1	1	1	1	1	1	6
2	1	0	0	1	1	1	4
3	1	1	0	0	1	1	4
4	1	1	0	1	1	1	5
5	1	1	1	1	1	1	6
6	1	1	0	1	1	1	5
7	1	1	1	1	1	0	5
8	1	1	1	0	0	1	4
9	1	1	1	1	1	1	6
10	1	1	0	1	1	1	5
11	1	1	0	1	1	1	5
12	1	1	1	0	1	1	5
13	1	1	1	1	1	1	6
14	1	0	1	1	1	0	4
15	1	0	0	1	0	0	2
16	0	1	0	0	1	0	2
17	0	1	1	0	0	1	3
18	0	1	0	0	0	1	2
19	0	1	0	1	0	1	3
20	0	1	1	0	1	0	3
21	0	1	0	1	0	1	3
22	0	1	0	0	0	1	2
23	0	1	0	0	1	1	3
24	0	0	1	0	0	0	1
25	0	0	1	1	1	1	4
26	0	0	1	0	1	1	3
27	0	0	1	1	1	1	4
28	0	0	1	0	1	0	2
29	0	0	1	0	0	0	1
30	0	0	0	1	0	0	1
31	0	0	0	1	1	1	3
32	0	0	0	1	1	0	2
33	0	0	0	0	1	0	1
34	0	0	0	0	1	0	1
35	0	0	0	0	1	0	1
36	0	0	0	0	0	1	1
37	0	0	0	0	0	1	1
38	0	0	0	0	0	1	1

Table 2.2: Frequency count of 38 deer mice.

y	1	2	3	4	5	6
f_y	9	6	7	6	6	4

methamphetamine users for each count of treatment episodes from 61 health treatment centres in Bangkok (Böhning et al., 2004). The frequency counts of illegal immigrants in the Netherlands and methamphetamine users in Bangkok are shown in Tables 2.3 and 2.4, respectively. As can be seen from Table 2.3, $f_1 = 1,645$ illegal immigrants were observed once, $f_2 = 183$ were observed twice, $f_3 = 37$ were observed three times, $f_4 = 13$ were observed four times, $f_5 = 1$ were observed five times, and $f_6 = 1$ were observed six times. The number of illegal immigrants that were not seen, f_0, was unknown.

Table 2.3: Frequency count of the illegal immigrant population in Netherlands.

y	1	2	3	4	5	6
f_y	1645	183	37	13	1	1

Table 2.4: Frequency count of methamphetamine users for each count of treatment episode from 61 health treatment centres in Bangkok.

y	1	2	3	4	5	6	7	8	9	10
f_y	3114	163	23	20	9	3	3	3	4	3

In both types of capture-recapture data, Y is the number of times that a unit is identified over the study period. Here, $Y = 1, 2, ..., m$, where m is the largest observed count and f_i the frequency of units identified exactly i times for $i = 1, 2, ..., m$. The number of observed units is given by $n = f_1 + f_2 + ... + f_m$. As zero counts are not observed, f_0 is unknown. Since the population size N consists of an observed and unobserved part, $N = n + f_0$. Therefore, f_0 must be estimated to derive the estimate of population size N. Also, modeling and estimating of p_0 is required for estimating the population size N, as $\hat{N} = \frac{n}{1-\hat{p}_0}$.

2.2 Estimation of population size N

Several estimators have been proposed for estimating the population size N. This section focuses on well-known estimators based on homogeneous and heterogeneous Poisson models. The maximum likelihood estimation under a Poisson model and Turing's estimator are used as estimators in the homogeneous case. Two estimators for the heterogeneous case, Chao's lower bound estimator and the Censored estimator, are also considered.

2.2.1 Horvitz-Thomson Approach

Horvitz and Thompson (1952) proposed a general approach to estimating the population size. From the identification mechanism, identification is assumed to occur independently for each unit with probability $1-p_0$. The population consists of N units. Let Z_j be the indicator identifying the j^{th} unit in the population for $j = 1, 2, ..., N$ with

$$Z_j = \begin{cases} 1 & \text{if} \quad j^{th} \quad \text{unit is identified} \\ 0 & \text{otherwise.} \end{cases}$$

Then, $E(\sum_{j=1}^{N} Z_j) = \sum_{j=1}^{N} E(Z_j) = \sum_{j=1}^{N} P(Z_j) = N(1-p_0)$. Equating this expected value to the observed number of cases $n = N(1-p_0)$ yeilds

$$\hat{N}_{HT} = \frac{n}{1-\hat{p}_0}, \tag{2.3}$$

which is well-known as the Horvitz-Thompson estimator.

2.2.2 Turing's estimator

Let Y be the number of times that a unit is identified over the observational period, and let f_i be the frequency of units identified exactly i times, $i = 1, 2, ..., m$ where m is the largest observed count. Also let p_i be the probability that a unit is identified exactly i times. Under the assumption that Y has a homogeneous Poisson distribution with parameter λ,

$$p_0 = e^{-\lambda} = \frac{\lambda e^{-\lambda}}{\lambda} = \frac{p_1}{E(Y)} = \frac{E(f_1)/N}{E(S)/N} = \frac{E(f_1)}{E(S)},$$

where $S = \sum_{i=1}^{m} if_i$. Replacing the expected value by the observed frequencies yields:

$$\hat{p}_0 = \frac{f_1}{S}.$$

Then, Turing's estimator (Good, 1953) is given by

$$\hat{N}_{Turing} = \frac{n}{1 - f_1/S}. \tag{2.4}$$

A simple variant of Turing's estimator (Lerdsuwansri, 2012) is given by

$$\widehat{Var}(\hat{N}_{Turing}) = \frac{n\frac{f_1}{S}}{(1+\frac{f_1}{S})^2} + \frac{n^2}{(1+\frac{f_1}{S})^4}\left[\frac{f_1(1-\frac{f_1}{\hat{N}})}{S^2} + \frac{f_1^2}{S^3}\right]. \tag{2.5}$$

The benefits of Turing's estimator are that it is easy to calculate, its value can be obtained in a straightforward way, iteration is not required.

2.2.3 Maximum Likelihood Estimation

Let Y be the number of times that a unit is identified over the study period. The frequency of units identified exactly i times is given as f_i where $i = 1, 2, ..., m$ and m is the largest observed count. Since count data from capture-recapture studies have non-zero counts, they can be modeled as a zero-truncated Poisson distribution:

$$p(y; \lambda) = \frac{e^{-\lambda}\lambda^y/y!}{1 - e^{-\lambda}}.$$

In statistics, maximum likelihood estimation is a well-known method for estimating parameters, by finding the parameter value that maximizes the likelihood function. Consider the likelihood function of count data Y :

$$L(\lambda) = \prod_{i=1}^{m}\left(\frac{e^{-\lambda}\lambda^i/i!}{1 - e^{-\lambda}}\right)^{f_i}.$$

The log-likelihood function is

$$\log L(\lambda) = -n\lambda + \log\lambda\sum_{i=1}^{m} if_i - \sum_{i=1}^{m} f_i \log i! - n\log(1 - e^{-\lambda}). \tag{2.6}$$

To find the maximum likelihood estimator (MLE) of parameter λ, (2.6) is differentiated with respect to λ and is equated to 0. The MLE of parameter λ is given by

$$\hat{\lambda} = \frac{S}{n}(1 - e^{-\hat{\lambda}}) = \frac{S}{\hat{N}},$$

where $S = \sum_{i=1}^{m} if_i$. Since there is no closed-form solution, the EM algorithm is used to find the MLE by iteratively applying these two steps:

$$\begin{aligned} (i) \quad \hat{N}^{(t+1)} &= \frac{n}{1-e^{-\hat{\lambda}^t}} \\ (ii) \quad \hat{\lambda}^{(t+1)} &= \frac{S}{\hat{N}^{(t+1)}}. \end{aligned}$$

The steps are repeated until convergence, to find $\hat{\lambda}_{MLE}$. Thus, the population size estimator is of the form

$$\hat{N}_{MLE} = \frac{n}{1-e^{-\hat{\lambda}_{MLE}}}. \tag{2.7}$$

The variance of (2.7) is given as

$$\widehat{Var}(\hat{N}_{MLE}) = \frac{\hat{N}_{MLE}}{e^{\frac{S}{\hat{N}_{MLE}}} - \frac{S}{\hat{N}_{MLE}} - 1}, \tag{2.8}$$

see (Böhning, 2008) for more details.

2.2.4 Chao's lower bound estimator

Chao (1987, 1989) proposed an alternative estimator of population size for a heterogeneous Poisson model. Count data are assumed to be generated from a mixed Poisson model with arbitrary density $g(\lambda) : p_i = \int_0^\infty \frac{e^{-\lambda}\lambda^i}{i!} g(\lambda) d\lambda$, where $i = 0, 1, 2, ...$ Chao's estimator is derived based on the Cauchy-Schwarz inequality:

$$E[UV]^2 = \left|E[UV]\right|^2 \leqslant E[U]^2 E[V]^2.$$

Then

$$(\int UV)^2 \leqslant \int U^2 \int V^2.$$

Let $U(\lambda) = \sqrt{e^{-\lambda}}$ and $V(\lambda) = \lambda\sqrt{e^{-\lambda}}$. This yields

$$\begin{aligned} (\int_0^\infty \lambda e^{-\lambda} d\lambda)^2 &\leqslant \int_0^\infty e^{-\lambda} d\lambda \int_0^\infty \lambda^2 e^{-\lambda} d\lambda \\ p_1^2 &\leqslant p_0 \times 2p_2. \end{aligned}$$

The lower bound for p_0 is derived by

$$p_0 \leqslant \frac{p_1^2}{2p_2}.$$

Multipling inequality by N leads to

$$\begin{aligned} Np_0 &\leqslant \frac{(Np_1)^2}{2Np_2} \\ f_0 &\leqslant \frac{f_1^2}{2f_2}. \end{aligned}$$

Hence, the lower bound estimator is $\hat{f}_0 = \frac{f_1^2}{2f_2}$ and Chao's lower bound estimator is

$$\hat{N}_{Chao} = n + \frac{f_1^2}{2f_2}. \tag{2.9}$$

The estimated variance of Chao's estimator is

$$\widehat{Var}(\hat{N}_{Chao}) = \frac{1}{4}\frac{{f_1}^4}{{f_2}^3} + \frac{{f_1}^3}{{f_2}^2} + \frac{1}{2}\frac{{f_1}^2}{f_2} - \frac{1}{4}\frac{{f_1}^4}{n{f_2}^2} - \frac{1}{2}\frac{{f_1}^4}{f_2(2nf_2 + {f_1}^2)}, \tag{2.10}$$

see (Böhning, 2008) for more details.

2.2.5 The censored estimator

Niwitpong et al.(2013) proposed an estimator under a geometric distribution with $k_y(p) = p(1-p)^y$ for $y = 0, 1, 2, ...$ Consider the conventional zero-truncated geometric $k_y^+ = \frac{p(1-p)^y}{1-p} = p(1-p)^{y-1}$ for $y = 1, 2, ...$ Then $P(Y = 1) = k_1^+ = p$ and $P(Y > 1) = \sum_{y=2}^{\infty} k_y^+ = 1 - p$. The log-likelihood is

$$l(p) = f_1 \log p + (n - f_1)\log(1-p). \tag{2.11}$$

To find the MLE of parameter p, (2.11) is differentiated with respect to p and is equated to 0

$$\begin{aligned} \frac{\partial}{\partial p} l(p) &= 0 \\ \frac{f_1}{p} - \frac{n - f_1}{1 - p} &= 0 \\ \hat{p} &= \frac{f_1}{n}. \end{aligned}$$

Hence, the MLE of parameter p is $\hat{p} = \frac{f_1}{n}$. From $e_0 = E(f_0|p) = \frac{np}{1-p}$, $\hat{e}_0 = \frac{n\frac{f_1}{n}}{1-\frac{f_1}{n}} = \frac{f_1}{1-\frac{f_1}{n}}$ and the censored estimator is

$$\hat{N}_{Censored} = n + \frac{f_1}{1 - \frac{f_1}{n}} = \frac{n^2}{n - f_1}. \tag{2.12}$$

An estimate of the variance of the censored estimator is given by

$$\widehat{Var}(\hat{N}_{Censored}) = \frac{f_1}{(1-\frac{f_1}{n})^2} \times \frac{2n-f_1}{n-f_1}. \tag{2.13}$$

2.3 Mixture models for estimating population size

This section reviews mixture models introduced into the framework of capture-recapture. In general, count data are often modeled by a Poisson distribution having parameter λ. As λ is not identical for all units in the population in practice, it is more reasonable to assume that the target population may consist of a multiple subgroups. Mixture models are more flexible than simple models, which require specific assumptions, and can cope with data from a heterogeneous population which is known or suspected to contain dissimilar parts. Suppose that Y is a random variable that arises from s sub-populations or components. Each component has the same density but different parameter values, i.e. $\lambda_1, \lambda_2, ..., \lambda_s$. Let $p(y;\lambda_k)$ be the component density of k for $k = 1, 2, ..., s$. Then, the finite mixture distribution

$$f(y;Q) = \sum_{k=1}^{s} q_k p(y;\lambda_k) \tag{2.14}$$

arises as the marginal distribution with respect to some latent variable Z with distribution Q. The mixing distribution $Q = \begin{pmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_s \\ q_1 & q_2 & \cdots & q_s \end{pmatrix}$ gives a non-negative weight q_k to parameter λ_k such that $\sum_{k=1}^{s} q_k = 1$.

2.3.1 Zero-truncated count mixture distributions

Böhning et al. (2005) proposed the NPMLE of population size based on the zero-truncated mixture Poisson distribution.

Let Y be the number of times that a unit was identified over the study period. Assuming that Y is mixture of densities $p(y;\lambda_k)$, then the marginal distribution of Y is

$$f(y;Q) = \sum_{k=1}^{s} q_k p(y|\lambda_k), \tag{2.15}$$

where $p(y|\lambda_k) = \frac{e^{-\lambda_k}\lambda_k^y}{y!}$ and the mixing distribution $Q = \begin{pmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_s \\ q_1 & q_2 & \cdots & q_s \end{pmatrix}$.

Assume f_i be the number of units identified exactly i times, where $i = 1, 2, ..., m$ and m is the largest occurring count. As $Y = 0$ is not observed, f_0 is unknown. The observed, incomplete data likelihood is of the form

$$L(Q) = \prod_{i=1}^{m} \left(\frac{\sum_{k=1}^{s} q_k p(i|\lambda_k)}{1 - \sum_{k=1}^{s} q_k p(0|\lambda_k)} \right)^{f_i}. \tag{2.16}$$

The incomplete data log-likelihood is

$$l(Q) = \sum_{i=1}^{m} f_i \log \left(\frac{\sum_{k=1}^{s} q_k p(i|\lambda_k)}{1 - \sum_{k=1}^{s} q_k p(0|\lambda_k)} \right). \tag{2.17}$$

An estimate of Q can be found by maximizing (2.17) leading to the NPMLE. It has become very common to use the EM algorithm (Dempster et al., 1977) for maximum likelihood estimation in a mixture model. To apply the EM algorithm, the complete data log-likelihood is needed.

At the E-step, the unobserved frequency f_0 is replaced by its expected value given observed frequencies and current values of Q. Let the expected value of f_0, denoted by $\hat{f}_0$, be written as

$$\begin{aligned} \hat{f}_0 &= E(f_0|\text{observed data}; Q) \\ &= \frac{nf(0; Q)}{1 - f(0; Q)} \\ &= \frac{n \sum_{k=1}^{s} q_k p(0|\nu_k, d)}{1 - \sum_{k=1}^{s} q_k p(0|\nu_k, d)}. \end{aligned} \tag{2.18}$$

The log-likelihood for the complete data is given as

$$l_{cd}(Q) = \sum_{i=0}^{m} f_i \sum_{k=1}^{s} z_{ik} \log q_k + \sum_{i=0}^{m} f_i \sum_{k=1}^{s} z_{ik} \log p(i|\lambda_k), \tag{2.19}$$

where the unobserved covariate z_{ik} is 1 if i belongs to component k and 0 otherwise.

In the E-step, the unobserved covariates z_{ik} are replaced by their expected values, e_{ik}, conditional upon the observed data and current values of λ_k and q_k for $k = 1, 2, ..., s$. This gives

$$e_{ik} = E(z_{ik}|\text{observed data}; \lambda_k, q_k) = \frac{q_k p(i|\lambda_k)}{\sum_{k=1}^{s} q_k p(i|\lambda_k)}. \tag{2.20}$$

The expected log-likelihood of the complete data is

$$E[l_{cd}(Q)] = \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log q_k + \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\lambda_k). \tag{2.21}$$

M-step: in mathematical optimization, the method of Lagrange multipliers is a strategy for finding the local maximum and minimum of a function, subject to equality constraints. To maximize (2.21) subject to the constraint $\sum_{k=1}^{s} q_k = 1$,

$$L(Q,\gamma) = \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log q_k + \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\lambda_k) + \gamma(1 - \sum_{k=1}^{s} q_k). \tag{2.22}$$

The MLEs of $q_k, \lambda_1, \lambda_2, ..., \lambda_s$ are found by solving the equations of derivative (2.22) with respect to γ, q_k, and λ_k. The new estimators of q_k and λ_k are

$$\hat{q}_k = \frac{1}{n + \hat{f}_0} \sum_{i=0}^{m} f_i e_{ik} \qquad \text{for} \quad k = 1, 2, ..., s \tag{2.23}$$

$$\hat{\lambda}_k = \frac{\sum_{i=0}^{m} i f_i e_{ik}}{\sum_{i=0}^{m} f_i e_{ik}} (1 - e^{-\hat{\lambda}_k}) \qquad \text{for} \quad k = 1, 2, ..., s. \tag{2.24}$$

To estimate an unknown population of size N, the Horvitz-Thompson estimate under the zero-truncated mixture of Poisson models is

$$\hat{N}_{TMIX} = \frac{n}{1 - \sum_{k=1}^{s} \hat{q}_k e^{-\hat{\lambda}_k}}. \tag{2.25}$$

Note that (2.24) does not provide a closed form solution of $\hat{\lambda}_k$. Therefore, the following algorithm is used to compute the population size estimator:

Step 0: Choose some initial values of $\hat{q}_k^{(0)}$ and $\hat{\lambda}_k^{(0)}$ for $\hat{Q}^{(0)}$, and set $t = 0$.

Step 1: Compute $\hat{f}_0^{(t+1)} = \frac{n \sum_{k=1}^{s} \hat{q}_k^{(t)} e^{-\hat{\lambda}_k^{(t)}}}{1 - \sum_{k=1}^{s} \hat{q}_k^{(t)} e^{-\hat{\lambda}_k^{(t)}}}$.

Step 2: Use complete data $\hat{f}_0^{(t+1)}, f_1, f_2, ..., f_m$ to compute the new MLEs of

$$\begin{aligned} \hat{q}_k^{(t+1)} &= \frac{\sum_{i=0}^{m} f_i e_{ik}^{(t)}}{n + \hat{f}_0^{(t+1)}} \quad \text{and} \\ \hat{\lambda}_k^{(t+1)} &= \frac{\sum_{i=0}^{m} i f_i e_{ik}^{(t)}}{\sum_{i=0}^{m} f_i e_{ik}^{(t)}} (1 - e^{-\hat{\lambda}_k^{(t)}}) \end{aligned}$$

for $\hat{Q}^{(t+1)}$, where $e_{ik}^{(t)} = \frac{\hat{q}_k^{(t)} p(i|\hat{\lambda}_k^{(t)})}{\sum_{k=1}^{s} \hat{q}_k^{(t)} p(i|\hat{\lambda}_k^{(t)})}$.

Step 3: Set $t = t + 1$ and repeat Step 1.

Steps 1 and 2 are repeated until convergence to a constant with an acceptable error.

2.3.2 Mixtures of zero-truncated count distributions

Viwatwongkasem et al. (2013) adopted the NPMLE for estimating the mixture parameters of zero-truncated Poisson distributions. The method was used to estimate the number of drug user in Thailand 2005-2007, using surveillance data on drug users identified by treatment episodes in over 1,140 health treatment centers in Thailand (Bureau of Health Service System Development, Ministry of Public Health).

Let Y be the number of times that a unit was identified over the study period. Assuming Y is mixture of zero-truncated Poisson distributions with densities $p_+(y;\lambda_k)$, then the marginal distribution of Y is

$$f_+(y;Q) = \sum_{k=1}^{s} q_k p_+(y|\lambda_k), \tag{2.26}$$

where $p_+(y;\lambda_k) = \frac{p(y|\lambda_k)}{1-p(0|\lambda_k)} = \frac{e^{-\lambda_k}\lambda_k^y/y!}{1-e^{-\lambda_k}}$and the mixing distribution is

$$Q = \begin{pmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_s \\ q_1 & q_2 & \cdots & q_s \end{pmatrix}.$$

Assume f_i to be the number of units identified exactly i times, where $i = 1, 2, ..., m$ and m is the largest occurring count. As $Y = 0$ is not observed, f_0 is unknown and must be estimated. The observed, incomplete data likelihood is of the form

$$\begin{aligned} L(Q) &= \prod_{i=1}^{m} \Big(f_+(i;Q) \Big)^{f_i} \\ &= \prod_{i=1}^{m} \Big(\sum_{k=1}^{s} q_k p_+(i|\lambda_k) \Big)^{f_i}. \end{aligned} \tag{2.27}$$

The log-likelihood for the mixture of zero-truncated count densities is

$$\begin{aligned} l(Q) &= \sum_{i=1}^{m} f_i \log f_+(i;Q) \\ &= \sum_{i=1}^{m} f_i \log \Bigg[\sum_{k=1}^{s} q_k p_+(i|\lambda_k) \Bigg]. \end{aligned} \tag{2.28}$$

The NPMLE of Q can be found by maximizing (2.28). To apply the EM algorithm, the complete data log-likelihood is needed, given in this case by

$$l_{cd}(Q) = \sum_{i=1}^{m} f_i \sum_{k=1}^{s} z_{ik} \log q_k + \sum_{i=1}^{m} f_i \sum_{k=1}^{s} z_{ik} \log p_+(i|\lambda_k). \tag{2.29}$$

Here, the unobserved covariate z_{ik} is 1 if i belongs to component k and 0 otherwise.

In the E-step, the unobserved covariates z_{ik} replaced by their expected values, e_{ik}, conditional upon the observed data and current values of λ_k and q_k for $k = 1, 2, ..., s$. This gives

$$e_{ik} = E(z_{ik}|\text{observed data}; \lambda_k, q_k) = \frac{q_k p_+(i|\lambda_k)}{\sum_{k=1}^{s} q_k p_+(i|\lambda_k)}. \tag{2.30}$$

The expected log-likelihood of the complete data is

$$E[l_{cd}(Q)] = \sum_{i=1}^{m} f_i \sum_{k=1}^{s} e_{ik} \log q_k + \sum_{i=1}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p_+(i|\lambda_k). \tag{2.31}$$

In the M-step, Lagrange multipliers is applied to find the local maximum of complete data log-likelihood (2.31), subject to the constraint $\sum_{k=1}^{s} q_k = 1$,

$$L(Q, \gamma) = \sum_{i=1}^{m} f_i \sum_{k=1}^{s} e_{ik} \log q_k + \sum_{i=1}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p_+(i|\lambda_k) + \gamma(1 - \sum_{k=1}^{s} q_k). \tag{2.32}$$

The MLEs of weights q_k are found by solving the equations of derivative (2.32) with respect to λ and q_k and is equated to 0. The new estimators of weights q_k are

$$\hat{q}_k = \frac{1}{n} \sum_{i=1}^{m} f_i e_{ik}, \qquad \text{for} \quad k = 1, 2, ..., s. \tag{2.33}$$

Similarly, the MLEs of the unknown parameters $\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_s$ are found by solving the equations of derivatives (2.32) with respect to λ_k, given by

$$\hat{\lambda}_k = \frac{\sum_{i=1}^{m} i f_i e_{ik}}{\sum_{i=1}^{m} f_i e_{ik}} (1 - e^{-\hat{\lambda}_k}) \qquad \text{for} \quad k = 1, 2, ..., s. \tag{2.34}$$

To estimate an unknown population size N, the Horvitz-Thompson estimate under mixtures of zero-truncated Poisson models is

$$\hat{N}_{MIXT} = \frac{n}{1 - \sum_{k=1}^{s} \hat{q}_k e^{-\hat{\lambda}_k}}. \tag{2.35}$$

Note that (2.34) does not provide a closed-form solution of $\hat{\lambda}_k$; iteration is applied until the desired accuracy is achieved. The following algorithm is used to compute the population size estimator:

Step 0: Choose some initial values of $\hat{q}_k^{(0)}$ and $\hat{\lambda}_k^{(0)}$ for $\hat{Q}^{(0)}$, and set $t = 0$.

Step 1: Use observed data $f_1, f_2, ..., f_m$ to compute the new MLE

$$\begin{aligned}\hat{q}_k^{(t+1)} &= \frac{1}{n}\sum_{i=1}^{m} f_i e_{ik}^{(t)} \quad \text{and} \\ \hat{\lambda}_k^{(t+1)} &= \frac{\sum_{i=1}^{m} i f_i e_{ik}^{(t)}}{\sum_{i=1}^{m} f_i e_{ik}^{(t)}}(1-e^{-\hat{\lambda}_k^{(t)}})\end{aligned}$$

for $\hat{Q}^{(t+1)}$, where $e_{ik}^{(t)} = \frac{\hat{q}_k^{(t)} p_+(i|\hat{\lambda}_k^{(t)})}{\sum_{k=1}^{s} \hat{q}_k^{(t)} p_+(i|\hat{\lambda}_k^{(t)})}$.

Step 2: Set $t = t + 1$ and repeat Step 1.

Steps 1 and 2 are repeated until convergence to a constant with an acceptable error.

Böhning and Kuhnert (2006) showed the equivalence of the zero-truncated count mixture distributions and the mixtures of zero-truncated count distributions. They observed that, for any mixing distribution of the truncated mixture, a different mixing distribution of the mixture could be found. This implied that the likelihood surfaces of the two models agree, and in this sense the models are equivalent. Consequently, estimating the population size N of two estimators associated with the two models provides equal values.

2.4 Unconditional nonparametric maximum likelihood estimator

Let Y be the number of times that a unit was identified over the study period. Assuming Y is mixture of densities $p(y; \lambda_k)$, then the marginal distribution of Y is

$$f(y; Q) = \sum_{k=1}^{s} q_k p(y|\lambda_k)$$

where mixing distribution $Q = \begin{pmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_s \\ q_1 & q_2 & \cdots & q_s \end{pmatrix}$. For capture-recapture modeling, two likelihood methods are possible. One is based upon the full, unconditional likelihood

$$L(N, Q) = \frac{N!}{f_0! f_1! ... f_m!} \prod_{i=0}^{m} f(i; Q)^{f_i}. \tag{2.36}$$

This may be rewritten in two parts (Sanathanan, 1977)

$$L(N, Q) = L_b(N, Q) \times L_c(Q)$$

where

$$\begin{aligned} L_b(N,Q) &= \frac{N!}{f_0!(N-f_0)!} f(0;Q)^{f_0}(1-f(0;Q))^{N-f_0}, \\ L_c(Q) &= \frac{n!}{f_1!f_2!...f_m!} \prod_{i=1}^{m} \left(\frac{f(i;Q)}{1-f(0;Q)} \right)^{f_i}. \end{aligned}$$

Two estimates of N arise naturally. The first is maximum likelihood estimation, denoted as $\hat{N}_u$ and defined by the condition that there exists a value $\hat{Q}_u$ such that $(\hat{N}_u, \hat{Q}_u)$ maximizes $L(N,Q)$ over all admissible values of (N,Q). The second estimate, and called the conditional maximum likelihood estimate, is denoted by $\hat{N}_c$ and defined by the condition that $\hat{N}_c$ maximizes $L_b(N,Q)$. This yields $\hat{Q}_c$ the value of Q that maximizes $L_c(Q)$, as discussed in the previous section.

The conditional NPMLE faces two major problems: the boundary problem and the lack of identifiability in mixture models. The boundary problem deals with the circumstance in which the mixing distribution component parameters converge to 0^+ with a positive weight (J.-P. Z. Wang & Lindsay, 2005; J.-P. Wang & Lindsay, 2008). Kuhnert et al. (2008) demonstrated that the boundary problem results in overestimation of population size N. Nonidentifiability also affects the inference of population size (Link, 2003). Different models that provide different estimates of N might have identical distributions. To avoid these problems, the unconditional maximum likelihood for inferring the unknown population size N was suggested by Pawitan (2001). The resulting likelihood is called a profile mixture likelihood.

The population size N and mixing distribution Q are unknown, where N is the parameter of interest and Q is a nuisance parameter. We need to estimate both N and Q, but are interested lies only in the parameter N. To achieve this, the nuisance parameter is often profiled out. Since $N = n + f_0$, estimating the population size N is equivalent to estimating f_0. The full likelihood (2.36) can be rewritten as

$$L(n+f_0, Q) = \frac{(n+f_0)!}{f_0!f_1!...f_m!} \prod_{i=0}^{m} f(i;Q)^{f_i}. \tag{2.37}$$

Given a fixed f_0, the log-likelihood function of (2.37), with the constant terms omitted, takes the form

$$l(Q|f_0) = \sum_{i=0}^{m} f_i \log \sum_{k=1}^{s} q_k p(i|\lambda_k). \tag{2.38}$$

To obtain the MLE of Q for a given fixed f_0, (2.38) is maximized with respect to Q, leading to

$$\hat{Q}(f_0) = \text{arg}\max_{Q} l(Q(f_0)|f_0). \tag{2.39}$$

For each fixed f_0 we can evaluate $\hat{Q}(f_0)$ and derive a new curve $l(\hat{Q}(f_0)|f_0)$, f_0. The maximum over all these curves is the MLE of f_0:

$$\hat{f}_0 = \text{arg}\max_{f_0} l(\hat{Q}(f_0)|f_0). \tag{2.40}$$

Consequently, the MLEs of f_0 and Q are $\hat{f}_0$ and $Q(\hat{f}_0)$, respectively. The estimator for population size becomes $\hat{N} = n+\hat{f}_0$. This is called the profile nonparametric maximum likelihood estimator (profile NPMLE).

2.5 Confidence interval estimation for population size N based upon the profile likelihood

In capture-recapture studies, derivation of $Var(\hat{N})$ to form the confidence interval of N is not easy. Profile likelihood can be used to construct the confidence intervals of population size N (Norris III & Pollock, 1996; Norris & Pollock, 1998) as follows.

Let $\hat{N} = n + \hat{f}_0$ be the profile NPMLE. The log-likelihood ratio statistic is given by

$$2[l(n + \hat{f}_0, Q(\hat{f}_0)) - l(n + f_0, \hat{Q}(f_0))] \sim \chi^2(1).$$

Using the log-likelihood ratio statistic, all Ns corresponding to

$$2[l(n + \hat{f}_0, Q(\hat{f}_0)) - l(n + f_0, \hat{Q}(f_0))] \leq (z_{1-\alpha/2})^2$$

form the $100(1-\alpha)\%$ confidence set for N. Therefore, the 95% confidence interval for N is the range of N that satisfies

$$2[l(n + \hat{f}_0, Q(\hat{f}_0)) - l(n + f_0, \hat{Q}(f_0))] - (1.96)^2 \leq 0. \tag{2.41}$$

2.6 Confidence interval of the population size N under normal approximation

Based on the population size estimators $\hat{N}$ and the estimated variance $\widehat{Var}(\hat{N})$, a confidence interval of the population size $\hat{N}$ can be constructed by the normal approximation approach. The $100(1-\alpha)\%$ confidence interval for $\hat{N}$ is given as follows:

$$\hat{N} \pm z_{1-\frac{\alpha}{2}} \widehat{Se}(\hat{N}), \tag{2.42}$$

where $z_{1-\frac{\alpha}{2}}$ is the $\left(1-\frac{\alpha}{2}\right)^{th}$ percentile of the standard normal distribution, $\widehat{Se}(\hat{N})$ denotes the standard error of $\hat{N}$, approximated by the asymptotic standard error, and $\widehat{Se}(\hat{N}) = \sqrt{\widehat{Var}(\hat{N})}$.

2.7 Expectation-Maximization(EM) algorithm

The Expectation-Maximization (EM) algorithm is an efficient iterative procedure for computing the MLE. It can be used for solving the problem of complete and incomplete data.

Let $\mathbf{y} = (y_1, ..., y_n)'$ be an incomplete observed data vector of size n from the population function $f(y; \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)'$ is a vector of p unknown parameters. Let $\mathbf{z}$ denote the vector containing unobservable or missing data and let $\mathbf{x}$ denote the complete data vector $\mathbf{x} = (\mathbf{y}', \mathbf{z}')'$.

Let $g_c(\mathbf{x}; \boldsymbol{\theta})$ denote the probability function of the random vector $\mathbf{X}$ corresponding to the complete data vector $\mathbf{x}$. The complete-data log-likelihood function is given by

$$l(\mathbf{x}; \boldsymbol{\theta}) = \log L_c(\mathbf{x}; \boldsymbol{\theta}) = \log g_c(\mathbf{x}; \boldsymbol{\theta}).$$

The EM algorithm finds the MLE by iteratively applying these two steps:

Expectation step (E step): This calculates the expected value of the complete log-likelihood $l(\mathbf{x}; \boldsymbol{\theta})$ with respect to the conditional distribution of $\mathbf{z}$ given the observed data vector $\mathbf{y}$ and the current estimate of the parameter vector $\boldsymbol{\theta}^{(k-1)}$at the $(k-1)^{th}$ iteration:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{k-1}) = E[l(\mathbf{x}; \boldsymbol{\theta})|\mathbf{y}; \boldsymbol{\theta}^{k-1}].$$

Maximization step (M step): This maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{k-1})$ with respect to $\boldsymbol{\theta}$ to give updated values $\boldsymbol{\theta}^{k}$ until convergence with an acceptable error.

2.8 Model selection criteria

Selection of the appropriate number of components is important in the finite mixture model. Several selection criteria have been suggested. Schwarz et al. (1978) provide the Bayesian Information Criterion (BIC) for model selection in the context of mixture models. The BIC is defined as

$$BIC = -2\log L(\hat{Q}_s) + k\log n, \tag{2.43}$$

where $\log L(\hat{Q}_s)$ is the maximum log-likelihood of the model with s components and k is the number of parameters estimated by the model. The model with the lowest BIC is preferred.

2.9 Laplace approximation

An alternative approach to the approximation of integrals is to use Laplace's method. The Laplace method approximates integrals of the form

$$\int_a^b e^{h(x)}dx \cong \sqrt{2\pi}e^{h(\tilde{x})}\left|\left.\frac{\partial^2 h(x)}{\partial x^2}\right|_{x=\tilde{x}}\right|^{-\frac{1}{2}}, \tag{2.44}$$

where a and b may be finite or infinite and $\tilde{x}$ denotes the value of x that maximizes $h(x)$.

2.10 Ratio plot for identify a distribution

Graphical statistics is a simple and quick method for identifying a distribution. In the capture-recapture method a graphical device, named the ratio plot, was developed by Böhning et al.(2013) for investigating homogeneous and heterogeneous Poisson models. The ratio plot for untruncated probability is defined as

$$r_x = (x+1)\frac{p_{x+1}}{p_x}. \tag{2.45}$$

In capture-recapture studies, the observed data arise from the zero-truncated distribution with probability

$$p_+(x) = \frac{p_x}{1-p_0}. \tag{2.46}$$

Then, the ratio plot for the zero-truncated probability is expressed as

$$r_x = (x+1)\frac{p_{x+1}/(1-p_0)}{p_x/(1-p_0)}. \tag{2.47}$$

Since

$$r_x = (x+1)\frac{p_{x+1}}{p_x} = (x+1)\frac{p_{x+1}/(1-p_0)}{p_x/(1-p_0)}, \tag{2.48}$$

the ratio plots for untruncated and truncated distributions are identical. As a consequence, the ratio plot $r_x = (x+1)\frac{p_{x+1}}{p_x}$ can be used for both zero-truncated and completed count distributions.

The probability function of Poisson distribution is $p_x = \frac{e^{-\lambda}\lambda^x}{x!}$. The ratio plot can be calculated by

$$\begin{aligned} r_x &= (x+1)\frac{p_{x+1}}{p_x} \\ &= (x+1)\frac{\frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!}}{\frac{e^{-\lambda}\lambda^x}{x!}} \\ &= \lambda. \end{aligned} \tag{2.49}$$

The ratio plot r_x of Poisson distribution is constant with varying count x. Therefore, plotting r_x against x can be used to identify the model of count data. If the ratio plot produces a pattern of a horizontal line, this can be taken as evidence for the presence of a Poisson distribution. If the ratio plot produces a straight line with a positive slope, this is indicative of structured heterogeneity. In practice, the ratio plot is estimated by

$$r_x^* = (x+1)\frac{\hat{p}_{x+1}N}{\hat{p}_x N} = (x+1)\frac{f_{x+1}/N}{f_x/N} = (x+1)\frac{f_{x+1}}{f_x}, \tag{2.50}$$

where f_x is the frequency of count x.

CHAPTER 3

MAXIMUM LIKELIHOOD ESTIMATION OF POSSION-NORMAL MIXTURE

In general, count data are modeled by a Poisson distribution with parameter λ. Under the homogeneous Poisson model, the probability of each unit identified exactly i times is $p_i = \frac{e^{-\lambda}\lambda^i}{i!}$. Due to the fact that, in practice, λ is not identical for all units in the population it is more reasonable to assume that the target population may consist of a set of subgroups. An alternative model incorporating heterogeneity of Poisson parameter might be more realistic. The probability under heterogeneity Poisson with density $f(\lambda)$ is given by

$$p_i = \int_0^\infty \frac{e^{-\lambda}\lambda^i}{i!} f(\lambda) d\lambda. \tag{3.1}$$

For example, if $f(\lambda)$ is the gamma distribution with a shape parameter k and a scale parameter $\theta = \frac{1-\pi}{\pi}$. The probability function is

$$f(\lambda; k, \theta) = \frac{\theta^{-k}\lambda^{k-1}e^{\frac{-\lambda}{\theta}}}{\Gamma(k)}, \tag{3.2}$$

where $\lambda, \theta, k > 0$. The probability that a unit identifies exactly i times is

$$\begin{aligned} p_i &= \int_0^\infty \frac{e^{-\lambda}\lambda^i}{i!} f(\lambda) d\lambda \\ &= \int_0^\infty \frac{e^{-\lambda}\lambda^i}{i!} \times \frac{\theta^{-k}\lambda^{k-1}e^{\frac{-\lambda}{\theta}}}{\Gamma(k)} d\lambda \\ &= \frac{\theta^{-k}}{\Gamma(i+1)\Gamma(k)} \int_0^\infty e^{-\lambda(1+\frac{1}{\theta})}\lambda^{i+k-1} d\lambda. \end{aligned} \tag{3.3}$$

Let $w = \lambda(1 + \frac{1}{\theta}) = \lambda(\frac{\theta+1}{\theta})$, so $\lambda = \frac{w\theta}{\theta+1}$ and $d\lambda = \frac{\theta}{\theta+1}dw$. Then

$$\begin{aligned} p_i &= \frac{\theta^{-k}}{\Gamma(i+1)\Gamma(k)} \int_0^\infty e^{-w} \left(\frac{w\theta}{\theta+1}\right)^{i+k-1} \left(\frac{\theta}{\theta+1}\right) dw \\ &= \frac{\theta^{-k}}{\Gamma(i+1)\Gamma(k)} \left(\frac{\theta}{\theta+1}\right)^{i+k} \int_0^\infty e^{-w} w^{i+k-1} dw. \end{aligned} \tag{3.4}$$

Since $\int_0^\infty e^{-w} w^{i+k-1} dw = \Gamma(i+k)$,

$$\begin{aligned} p_i &= \frac{\Gamma(i+k)}{\Gamma(i+1)\Gamma(k)} \theta^{-k} \left(\frac{\theta}{\theta+1}\right)^{i+k} \qquad (3.5) \\ &= \frac{\Gamma(i+k)}{\Gamma(i+1)\Gamma(k)} \theta^{-k} \left(\frac{\theta}{\theta+1}\right)^{i} \left(\frac{1}{\theta+1}\right)^{k}. \end{aligned}$$

As $\theta = \frac{1-\pi}{\pi}$, $\theta + 1 = \frac{1}{\pi}$, and $\frac{\theta}{\theta+1} = 1 - \pi$,

$$p_i = \frac{\Gamma(i+k)}{\Gamma(i+1)\Gamma(k)} \pi^k (1-\pi)^i \qquad (3.6)$$

which is the probability function of negative binomial distribution with parameter π and k. Modeling and estimating p_0 by negative binomial distribution leads to $\hat{N} = \frac{n}{1-\hat{\pi}^{\hat{k}}}$. The failure of a dispersion parameter estimation in negative binomial affects a spurious estimate for the population size N. This is called boundary problem (Böhning, 2015), as demonstrated in Kuhnert et al. (2008). To explore an alternative distribution, we propose to use $f(\lambda)$ as the normal distribution and discuss in this chapter. The EM algorithm is used for estimating MLEs of parameters.

3.1 Poisson-Normal mixture model

A discrete distribution commonly used in practice is also a compound Poisson distribution, a Poisson random variable with parameter λ, where parameter Λ is a random variable with density $f(\lambda)$. In such a case, the probability generating function (pgf) be written as

$$G(s) = \int_{-\infty}^{\infty} e^{\lambda(s-1)} f(\lambda) d\lambda. \qquad (3.7)$$

Let Λ follows a normal distribution with mean μ and variance σ^2. It is not possible be compound a Poisson with a normal distribution, because the normally distributed random variable can be negative. In order for the normal distribution taking positive values with probability near 1, it must hold that $\mu - 3\sigma > 0$, or μ is much greater than σ^2. From (3.7) the pgf of compound Poisson with a normal distribution is

$$\begin{aligned} G(s) &= \int_0^\infty e^{\lambda(s-1)} f(\lambda) d\lambda \\ &= \int_0^\infty e^{\lambda(s-1)} (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{(\lambda-\mu)^2}{2\sigma^2}} d\lambda \\ &= (2\pi\sigma^2)^{-\frac{1}{2}} \int_0^\infty e^{\lambda(s-1) - \frac{(\lambda-\mu)^2}{2\sigma^2}} d\lambda. \qquad (3.8) \end{aligned}$$

To evaluate Equation (3.8), approximation of $\int_0^\infty e^{\lambda(s-1)-\frac{(\lambda-\mu)^2}{2\sigma^2}} d\lambda$ is needed.

Let

$$h(\lambda) = \lambda(s-1) - \frac{(\lambda-\mu)^2}{2\sigma^2}. \tag{3.9}$$

The integral is

$$\int_0^\infty e^{\lambda(s-1)-\frac{(\lambda-\mu)^2}{2\sigma^2}} d\lambda = \int_0^\infty e^{h(\lambda)} d\lambda. \tag{3.10}$$

Laplace's method is applied to approximate the integral in Equation (3.10)

$$\int_0^\infty e^{h(\lambda)} d\lambda \cong (2\pi)^{\frac{1}{2}} e^{h(\tilde{\lambda})} \left| \frac{\partial^2 h(\lambda)}{\partial \lambda^2}\bigg|_{\lambda=\tilde{\lambda}} \right|^{-\frac{1}{2}},$$

where $\tilde{\lambda}$ denotes the value of λ that maximizes $h(\lambda)$. The first derivative of $h(\lambda)$ is

$$h'(\lambda) = (s-1) - \frac{(\lambda-\mu)}{\sigma^2}. \tag{3.11}$$

The value of λ that maximizes $h(\lambda)$ is found by solving

$$\begin{aligned} h'(\lambda) &= 0 \\ (s-1) - \frac{(\lambda-\mu)}{\sigma^2} &= 0 \\ (s-1) &= \frac{(\lambda-\mu)}{\sigma^2}. \end{aligned}$$

The values that maximizes $h(\lambda)$ is

$$\tilde{\lambda} = \sigma^2(s-1) + \mu. \tag{3.12}$$

Since $\sigma^2 > 0$, second derivative of $h(\lambda)$ is $h''(\lambda) = -\frac{1}{\sigma^2} < 0$. Then, $\tilde{\lambda} = \sigma^2(s-1) + \mu$ maximizes $h(\lambda)$. The integral approximation of (3.10) is

$$\begin{aligned} \int_0^\infty e^{\lambda(s-1)-\frac{(\lambda-\mu)^2}{2\sigma^2}} d\lambda &= \int_0^\infty e^{h(\lambda)} d\lambda \\ &= (2\pi)^{\frac{1}{2}} e^{h(\tilde{\lambda})} \left| h''(\tilde{\lambda}) \right|^{-\frac{1}{2}} \\ &= (2\pi)^{\frac{1}{2}} e^{[\sigma^2(s-1)+\mu][s-1]-\frac{\sigma^4(s-1)^2}{2\sigma^2}} \left| -\frac{1}{\sigma^2} \right|^{-\frac{1}{2}}. \end{aligned} \tag{3.13}$$

Substituting (3.13) in Equation (3.8) the pgf of compound Poisson with a normal distribution is

$$
\begin{aligned}
G(s) &= (2\pi\sigma^2)^{-\frac{1}{2}} \int_0^\infty e^{\lambda(s-1)-\frac{(\lambda-\mu)^2}{2\sigma^2}} d\lambda \\
&= (2\pi\sigma^2)^{-\frac{1}{2}} (2\pi)^{\frac{1}{2}} e^{[\sigma^2(s-1)+\mu][s-1]-\frac{\sigma^4(s-1)^2}{2\sigma^2}} \left| -\frac{1}{\sigma^2} \right|^{-\frac{1}{2}} \\
&= e^{[\sigma^2(s-1)+\mu][s-1]-\frac{\sigma^2(s-1)^2}{2}} \\
&= e^{\sigma^2(s-1)^2+\mu(s-1)-\frac{\sigma^2(s-1)^2}{2}} \\
&= e^{\frac{\sigma^2(s-1)^2}{2}+\mu(s-1)-\sigma^2(s-1)+\sigma^2(s-1)} \\
&= e^{(\mu-\sigma^2)(s-1)+\sigma^2(s-1)+\frac{\sigma^2(s-1)^2}{2}} \\
&= e^{(\mu-\sigma^2)(s-1)+\frac{\sigma^2(s-1)}{2}(2+s-1)} \\
&= e^{(\mu-\sigma^2)(s-1)+\frac{\sigma^2(s-1)}{2}(s+1)} \\
&= e^{(\mu-\sigma^2)(s-1)+\frac{\sigma^2(s^2-1)}{2}} \qquad (3.14)
\end{aligned}
$$

which is the pgf of Hermite distribution.

The probability function of Hermite distribution, in terms of the parameters $a_1 = \mu - \sigma^2$ and $a_2 = \frac{\sigma^2}{2}$, has the expression

$$p(y|a_1, a_2) = e^{-(a_1+a_2)} \sum_{j=0}^{[y/2]} \frac{a_1^{y-2j} a_2^j}{(y-2j)!j!}. \qquad (3.15)$$

Here, $y = 0, 1, 2, \ldots ; a_1, a_2 > 0$ and $[y/2]$ is the integer part of $\frac{y}{2}$. The mean and variance of Hermite distribution are $\nu = a_1 + 2a_2$ and $\tilde{\sigma}^2 = a_1 + 4a_2$, respectively. The coefficient of dispersion $d = 1 + \frac{2a_2}{a_1+2a_2}$ allows a moderate overdispersion between 1 and 2. Reparameterize the probability function by ν and d gives the probability function as

$$p(y|\nu, d) = e^{\nu(\frac{d-3}{2})} [\nu(2-d)]^y \sum_{j=0}^{[y/2]} \left(\frac{d-1}{2\nu(2-d)^2} \right)^j \frac{1}{(y-2j)!j!}, \qquad (3.16)$$

where $y = 0, 1, 2, \ldots$; $\nu > 0$ and $1 < d < 2$.

The probability function of Hermite distribution for some mean ν and dispersion d is shown in Figure 3.1.

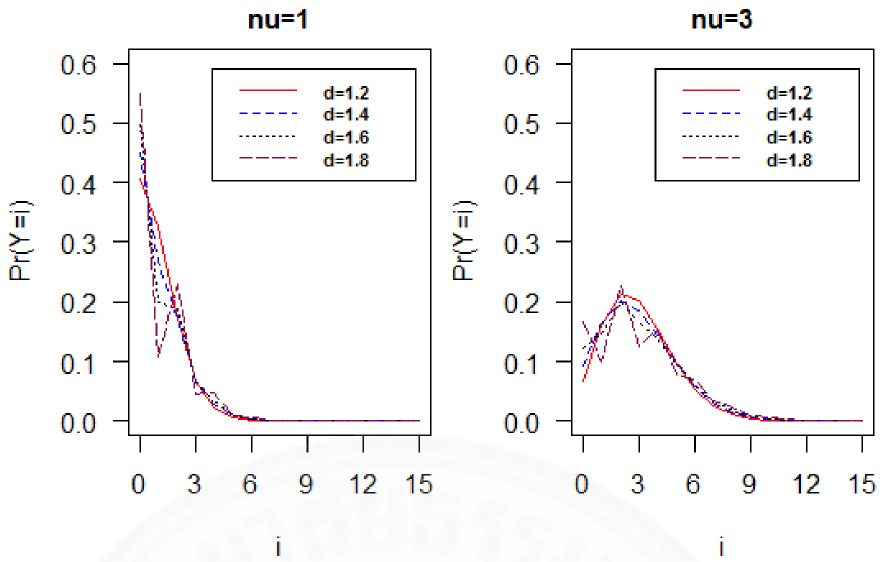


Figure 3.1: The probability function of Hermite distribution.

3.2 Maximum likelihood estimation of Poisson-Normal mixture

Let Y be the number of times that a unit was identified over the study period. Count Y is modeled with a Hermite distribution having probability function

$$p(y|\nu, d) = e^{\nu(\frac{d-3}{2})}[\nu(2-d)]^y \sum_{j=0}^{[y/2]} \left(\frac{d-1}{2\nu(2-d)^2}\right)^j \frac{1}{(y-2j)!j!}, \tag{3.17}$$

where $y = 0, 1, 2, ...; \nu > 0$ and $1 < d < 2$. Since the observed sample from capture-recapture contains only non-zero counts, the associated probability function becomes a zero-truncated Hermite. The incompletely observed likelihood relative to zero truncated count frequencies is

$$L(\nu, d) = \prod_{i=1}^{m} p_i'^{f_i}, \tag{3.18}$$

where $p_i' = \frac{p(i|\nu,d)}{1-p(0|\nu,d)}$.
The MLEs of ν and d can be found by maximizing (3.18). It has become common to use EM algorithm. To apply the EM algorithm, the complete data is required.

In the E-step, suppose that unknown missing frequency f_0 is replaced by its conditional expectation e_0 given the observed frequencies $f_1, f_2, ..., f_m$ and the current values of ν, d. The expected value e_0 under the Hermite density is obtained as

$$\begin{aligned} e_0 &= E(f_0|f_1, ..., f_m; \nu, d,) \\ &= Np_0 \\ &= (e_0 + n)p_0 \\ &= \frac{np_0}{1 - p_0}. \end{aligned} \tag{3.19}$$

The complete data likelihood with density p_i is

$$\begin{aligned} L_{cd}(\nu, d) &= \prod_{i=0}^{m} p_i^{f_i} \\ &= p_0^{e_0} p_1^{f_1} ... p_m^{f_m}. \end{aligned} \tag{3.20}$$

Here, $p_i = p(i|\nu, d) = e^{\nu(\frac{d-3}{2})}[\nu(2-d)]^i \sum_{j=0}^{[i/2]} \left(\frac{d-1}{2\nu(2-d)^2}\right)^j \frac{1}{(i-2j)!j!}$. The log-likelihood of complete data is

$$l_{cd}(\nu, d) = e_0 \log p_0 + \sum_{i=1}^{m} f_i \log p_i. \tag{3.21}$$

The complete log-likelihood (3.21) with the expectation and observed data is rewritten as

$$\begin{aligned} l_{cd}(\nu, d) &= e_0 \log p_0 + \sum_{i=1}^{m} f_i \log p_i \\ &= \left\{ \begin{array}{c} e_0 log\left(e^{\nu(\frac{d-3}{2})}\right) \\ + \sum_{i=1}^{m} f_i \log \left[e^{\nu(\frac{d-3}{2})}[\nu(2-d)]^i \sum_{j=0}^{[i/2]} \left(\frac{d-1}{2\nu(2-d)^2}\right)^j \frac{1}{(i-2j)!j!} \right] \end{array} \right\} \\ &= \left\{ \begin{array}{c} e_0 \nu\left(\frac{d-3}{2}\right) \\ \sum_{i=1}^{m} f_i \left[\nu\left(\frac{d-3}{2}\right) + i \log \nu(2-d) + \log \sum_{j=0}^{[i/2]} \left(\frac{d-1}{2\nu(2-d)^2}\right)^j \frac{1}{(i-2j)!j!} \right] \end{array} \right\}. \end{aligned} \tag{3.22}$$

Let $q_i(\theta) = \sum_{j=0}^{[i/2]} \frac{\theta^j}{(i-2j)!j!}$ and $\theta = \frac{d-1}{2\nu(2-d)^2}$. The complete log-likelihood (3.22),

$$\begin{aligned} l_{cd}(\nu, d) &= e_0 \nu\left(\frac{d-3}{2}\right) + \sum_{i=1}^{m} f_i \left[\nu\left(\frac{d-3}{2}\right) + i \log \nu(2-d) + \log q_i(\theta) \right] \\ &= e_0 \nu\left(\frac{d-3}{2}\right) + \sum_{i=1}^{m} f_i \nu\left(\frac{d-3}{2}\right) + \sum_{i=1}^{m} i f_i \log \nu(2-d) + \sum_{i=1}^{m} f_i \log q_i(\theta). \end{aligned} \tag{3.23}$$

In the M-step taking derivative of (3.23) with respect to ν and d, and setting the result to 0 yields

$$\begin{aligned}
\frac{\partial}{\partial \nu} l_{cd}(\nu, d) &= 0 \\
\left\{ \begin{array}{c} e_0 \left(\frac{d-3}{2} \right) + \sum_{i=1}^{m} f_i \left(\frac{d-3}{2} \right) \\ + \frac{1}{\nu} \sum_{i=1}^{m} i f_i + \frac{\partial}{\partial \nu} \left[\sum_{i=1}^{m} f_i \log q_i(\theta) \right] \end{array} \right\} &= 0 \\
\left\{ \begin{array}{c} \left(\frac{d-3}{2} \right) (e_0 + \sum_{i=1}^{m} f_i) + \frac{1}{\nu} \sum_{i=1}^{m} i f_i \\ + \frac{\partial}{\partial \nu} \left[\sum_{i=1}^{m} f_i \log q_i(\theta) \right] \end{array} \right\} &= 0. \qquad (3.24)
\end{aligned}$$

Consider

$$\begin{aligned}
\frac{\partial}{\partial \nu} \left[\sum_{i=1}^{m} f_i \log q_i(\theta) \right] &= \sum_{i=1}^{m} f_i \left[\frac{\partial}{\partial \nu} \log q_i(\theta) \right] \\
&= \sum_{i=1}^{m} f_i \left[\frac{1}{q_i(\theta)} \cdot \frac{\partial}{\partial \theta} q_i(\theta) \cdot \frac{\partial \theta}{\partial \nu} \right] \\
&= \sum_{i=1}^{m} f_i \left[\frac{1}{q_i(\theta)} \cdot q_i'(\theta) \cdot \frac{\partial}{\partial \nu} \left(\frac{d-1}{2\nu(2-d)^2} \right) \right] \\
&= \sum_{i=1}^{m} f_i \left[\frac{1}{q_i(\theta)} \cdot q_i'(\theta) \cdot \left(- \frac{d-1}{2\nu^2(2-d)^2} \right) \right] \\
&= - \frac{d-1}{2\nu^2(2-d)^2} \sum_{i=1}^{m} \frac{f_i q_i'(\theta)}{q_i(\theta)}, \qquad (3.25)
\end{aligned}$$

where $q_i'(\theta) = \frac{\partial}{\partial \theta} q_i(\theta)$. Substituting (3.25) in (3.24),

$$\begin{aligned}
\frac{\partial}{\partial \nu} l_{cd}(\nu, d) &= 0 \\
\left\{ \begin{array}{c} \left(\frac{d-3}{2} \right) (e_0 + \sum_{i=1}^{m} f_i) + \frac{1}{\nu} \sum_{i=1}^{m} i f_i \\ - \frac{d-1}{2\nu^2(2-d)^2} \sum_{i=1}^{m} \frac{f_i q_i'(\theta)}{q_i(\theta)} \end{array} \right\} &= 0. \qquad (3.26)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial d} l_{cd}(\nu, d) &= 0 \\
\left\{\begin{matrix} \frac{e_0\nu}{2} + \frac{\nu}{2}\sum_{i=1}^{m} f_i - \frac{1}{2-d}\sum_{i=1}^{m} i f_i \\ +\frac{\partial}{\partial d}\Big[\sum_{i=1}^{m} f_i \log q_i(\theta)\Big] \end{matrix}\right\} &= 0 \\
\left\{\begin{matrix} \frac{\nu}{2}(e_0 + \sum_{i=1}^{m} f_i) - \frac{1}{2-d}\sum_{i=1}^{m} i f_i \\ +\frac{\partial}{\partial d}\Big[\sum_{i=1}^{m} f_i \log q_i(\theta)\Big] \end{matrix}\right\} &= 0. \qquad (3.27)
\end{aligned}$$

Consider

$$\begin{aligned}
\frac{\partial}{\partial d}\Big[\sum_{i=1}^{m} f_i \log q_i(\theta)\Big] &= \sum_{i=1}^{m} f_i \Big[\frac{\partial}{\partial d}\log q_i(\theta)\Big] \\
&= \sum_{i=1}^{m} f_i \Big[\frac{1}{q_i(\theta)} \cdot \frac{\partial}{\partial \theta} q_i(\theta) \cdot \frac{\partial \theta}{\partial d}\Big] \\
&= \sum_{i=1}^{m} f_i \Big[\frac{1}{q_i(\theta)} \cdot q_i'(\theta) \cdot \frac{\partial}{\partial d}\Big(\frac{d-1}{2\nu(2-d)^2}\Big)\Big] \\
&= \sum_{i=1}^{m} f_i \Big[\frac{1}{q_i(\theta)} \cdot q_i'(\theta) \cdot \Big(\frac{1}{2\nu(2-d)^2} + \frac{d-1}{\nu(2-d)^3}\Big)\Big] \\
&= \frac{1}{2\nu(2-d)^2}\sum_{i=1}^{m} \frac{f_i q_i'(\theta)}{q_i(\theta)} + \frac{d-1}{\nu(2-d)^3}\sum_{i=1}^{m} \frac{f_i q_i'(\theta)}{q_i(\theta)} \\
&= \frac{d}{2\nu(2-d)^3}\sum_{i=1}^{m} \frac{f_i q_i'(\theta)}{q_i(\theta)}, \qquad (3.28)
\end{aligned}$$

where $q_i'(\theta) = \frac{\partial}{\partial \theta} q_i(\theta)$. Substituting (3.28) in (3.27),

$$\begin{aligned}
\frac{\partial}{\partial d} l_{cd}(\nu, d) &= 0 \\
\left\{\begin{matrix} \frac{\nu}{2}(e_0 + \sum_{i=1}^{m} f_i) - \frac{1}{2-d}\sum_{i=1}^{m} i f_i \\ +\frac{d}{2\nu(2-d)^3}\sum_{i=1}^{m} \frac{f_i q_i'(\theta)}{q_i(\theta)} \end{matrix}\right\} &= 0. \qquad (3.29)
\end{aligned}$$

Rewrite (3.29) as

$$\begin{aligned}
\sum_{i=1}^{m} \frac{f_i q_i'(\theta)}{q_i(\theta)} &= \frac{2\nu(2-d)^3}{d}\Big[\frac{1}{2-d}\sum_{i=1}^{m} i f_i - \frac{\nu}{2}(e_0 + \sum_{i=1}^{m} f_i)\Big] \\
&= \frac{2\nu(2-d)^2}{d}\sum_{i=1}^{m} i f_i - \frac{\nu^2(2-d)^3}{d}(e_0 + \sum_{i=1}^{m} f_i). \qquad (3.30)
\end{aligned}$$

Substituting (3.30) in (3.26),

$$\begin{aligned}
\left\{\begin{array}{c} \left(\frac{d-3}{2}\right)(e_0+\sum_{i=1}^{m}f_i)+\frac{1}{\nu}\sum_{i=1}^{m}if_i \\ +\frac{(d-1)(2-d)}{2d}(e_0+\sum_{i=1}^{m}f_i)-\frac{d-1}{\nu d}\sum_{i=1}^{m}if_i \end{array}\right\} &= 0\\
\frac{1}{\nu d}\sum_{i=1}^{m}if_i-\frac{1}{d}(e_0+\sum_{i=1}^{m}f_i) &= 0\\
\frac{1}{\nu d}\sum_{i=1}^{m}if_i &= \frac{1}{d}(e_0+\sum_{i=1}^{m}f_i)\\
\hat{\nu} &= \frac{\sum_{i=1}^{m}if_i}{e_0+\sum_{i=1}^{m}f_i}. \qquad (3.31)
\end{aligned}$$

Substituting (3.31) in (3.30),

$$\begin{aligned}
\sum_{i=1}^{m}\frac{f_iq_i'(\hat{\theta})}{q_i(\hat{\theta})} &= \frac{2\hat{\nu}(2-d)^2}{d}\sum_{i=1}^{m}if_i-\frac{\hat{\nu}^2(2-d)^3}{d}(e_0+\sum_{i=1}^{m}f_i)\\
&= \frac{2\hat{\nu}^2(2-d)^2}{d}(e_0+\sum_{i=1}^{m}f_i)-\frac{\hat{\nu}^2(2-d)^3}{d}(e_0+\sum_{i=1}^{m}f_i)\\
&= [\hat{\nu}(2-d)]^2(e_0+\sum_{i=1}^{m}f_i). \qquad (3.32)
\end{aligned}$$

The MLE of parameter d be found by solving (3.32). The package $hermite$ in program R allows to estimate parameter d given an univariate sample by means of the function $glm.hermite$, see (Moriña, Higueras, Puig, & Oliveira, 2015) for more details.

The population size estimator based on Hermite distribution through the Horvitz-Thomson approach is

$$\hat{N}_{Herm}=\frac{n}{1-e^{\hat{\nu}\left(\frac{\hat{d}-3}{2}\right)}}. \qquad (3.33)$$

3.2.1 EM algorithm

In practical terms, the EM algorithm is implemented as follows.

Step 0: Choose initial value $\hat{\nu}^{(0)}, \hat{d}^{(0)}$, set $t=0$.

Step 1: Compute $e_0^{(t+1)}=\frac{ne^{\hat{\nu}^{(t)}(\frac{\hat{d}^{(t)}-3}{2})}}{1-e^{\hat{\nu}^{(t)}(\frac{\hat{d}^{(t)}-3}{2})}}$.

Step 2: Use complete data $e_0^{(t+1)}, f_1, f_2, ..., f_m$ to compute the new MLEs of

$$\begin{aligned}
\hat{\nu}^{(t+1)} &= \frac{\sum_{i=0}^{m}if_i}{e_0^{(t+1)}+\sum_{i=1}^{m}f_i}\\
\hat{d}^{(t+1)} &= glm.hermite(e_0^{(t+1)}, f_1, ..., f_m \sim 1, link="log", start=NULL, m^\star=2).
\end{aligned}$$

Step 3: Set $t = t + 1$ and repeat to step 1. Steps 1 and 2 are repeated until:

$$\left|\hat{\nu}^{(t+1)} - \hat{\nu}^{(t)}\right| \leq 10^{-4} \text{ and } \left|\hat{d}^{(t+1)} - \hat{d}^{(t)}\right| \leq 10^{-4}.$$

3.3 Confidence interval estimation for population size N based upon the profile likelihood

In this section focuses on inferring the unknown size N of a closed population. In capture-recapture studies, derivation of $Var(\hat{N})$ to form the confidence interval of N is not easy. Bootstrapping approach can be used to construct confidence intervals. Alternatively, profile likelihood could be used to construct confidence intervals (Norris III & Pollock, 1996; Norris & Pollock, 1998). We construct confidence interval estimation of N based upon the profile likelihood as follows.

3.3.1 Profile likelihood of Poisson-Normal mixture models

Let Y be the number of times that a unit was identified over the observational period. Assume count Y is modeled with a Hermite distribution having probability function

$$p(y|\nu, d) = e^{\nu(\frac{d-3}{2})}[\nu(2-d)]^y \sum_{j=0}^{[y/2]} \left(\frac{d-1}{2\nu(2-d)^2}\right)^j \frac{1}{(y-2j)!j!}, \tag{3.34}$$

where $y = 0, 1, 2, ...$; $\nu > 0$ and $1 < d < 2$. The unconditional likelihood function is given by

$$L(N, \nu, d) = \frac{N!}{f_0!f_1!...f_m!} \prod_{i=0}^{m} \left(p(i|\nu, d)\right)^{f_i}. \tag{3.35}$$

The unconditional likelihood is described by full parameter (N, ν, d), but we are interested only in N. Nuisance parameters ν and d are eliminated by replacing they with their MLEs at each fixed value of N (Pawitan, 2001). The resulting is called a profile likelihood.

Since $N = n + f_0$, finding a profile likelihood of N is equivalent to finding a profile likelihood of f_0. The full likelihood (3.35) can be rewritten as

$$L(n + f_0, \nu, d) = \frac{(n + f_0)!}{f_0!f_1!...f_m!} \prod_{i=0}^{m} \left(p(i|\nu, d)\right)^{f_i}. \tag{3.36}$$

Given fixed f_0, the log-likelihood function of (3.36), with the constant terms are omitted, takes the form

$$\begin{aligned} l(\nu, d|f_0) &= \sum_{i=0}^{m} f_i \log p(i|\nu, d) \\ &= \sum_{i=0}^{m} f_i \log \left[e^{\nu\left(\frac{d-3}{2}\right)} [\nu(2-d)]^i \sum_{j=0}^{[i/2]} \left(\frac{d-1}{2\nu(2-d)^2} \right)^j \frac{1}{(i-2j)!j!} \right] \\ &= \sum_{i=0}^{m} f_i \left[\nu\left(\frac{d-3}{2} \right) + i \log \nu(2-d) + \log q_i(\theta) \right] \\ &= \sum_{i=0}^{m} f_i \nu\left(\frac{d-3}{2} \right) + \sum_{i=0}^{m} i f_i \log \nu(2-d) + \sum_{i=0}^{m} f_i \log q_i(\theta). \qquad (3.37) \end{aligned}$$

Here, $q_i(\theta) = \sum_{j=0}^{[i/2]} \frac{\theta^j}{(i-2j)!j!}$ and $\theta = \frac{d-1}{2\nu(2-d)^2}$. MLEs of unknown parameters ν and d for any fixed f_0 be found by differentiating (3.37) with respect to ν and d, and setting the result to 0:

$$\begin{aligned} \frac{\partial}{\partial \nu} l(\nu, d|f_0) &= 0 \\ \sum_{i=0}^{m} f_i \left(\frac{d-3}{2} \right) + \frac{1}{\nu} \sum_{i=0}^{m} i f_i + \frac{\partial}{\partial \nu} \left[\sum_{i=1}^{m} f_i \log q_i(\theta) \right] &= 0. \qquad (3.38) \end{aligned}$$

Substituting (3.25) in (3.38),

$$\begin{aligned} \frac{\partial}{\partial \nu} l(\nu, d|f_0) &= 0 \\ \sum_{i=0}^{m} f_i \left(\frac{d-3}{2} \right) + \frac{1}{\nu} \sum_{i=0}^{m} i f_i - \frac{d-1}{2\nu^2(2-d)^2} \sum_{i=0}^{m} \frac{f_i q_i'(\theta)}{q_i(\theta)} &= 0 \\ \left(\frac{d-3}{2} \right) \sum_{i=0}^{m} f_i + \frac{1}{\nu} \sum_{i=0}^{m} i f_i - \frac{d-1}{2\nu^2(2-d)^2} \sum_{i=0}^{m} \frac{f_i q_i'(\theta)}{q_i(\theta)} &= 0. \qquad (3.39) \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial d} l(\nu, d|f_0) &= 0 \\ \frac{\nu}{2} \sum_{i=0}^{m} f_i - \frac{1}{2-d} \sum_{i=0}^{m} i f_i + \frac{\partial}{\partial d} \left[\sum_{i=1}^{m} f_i \log q_i(\theta) \right] &= 0. \qquad (3.40) \end{aligned}$$

Substituting (3.28) in (3.40),

$$\begin{aligned} \frac{\partial}{\partial d} l(\nu, d|f_0) &= 0 \\ \frac{\nu}{2} \sum_{i=0}^{m} f_i - \frac{1}{2-d} \sum_{i=0}^{m} i f_i + \frac{d}{2\nu(2-d)^3} \sum_{i=0}^{m} \frac{f_i q_i'(\theta)}{q_i(\theta)} &= 0. \qquad (3.41) \end{aligned}$$

Rewrite (3.41) as

$$\begin{aligned}\sum_{i=0}^{m}\frac{f_i q_i'(\theta)}{q_i(\theta)} &= \frac{2\nu(2-d)^3}{d}\left[\frac{1}{2-d}\sum_{i=0}^{m} if_i - \frac{\nu}{2}\sum_{i=0}^{m} f_i\right]\\ &= \frac{2\nu(2-d)^2}{d}\sum_{i=0}^{m} if_i - \frac{\nu^2(2-d)^3}{d}\sum_{i=0}^{m} f_i. \qquad (3.42)\end{aligned}$$

Substituting (3.42) in (3.39),

$$\begin{aligned}\left(\frac{d-3}{2}\right)\sum_{i=0}^{m} f_i + \frac{1}{\nu}\sum_{i=0}^{m} if_i + \frac{(d-1)(2-d)}{2d}\sum_{i=0}^{m} f_i - \frac{d-1}{\nu d}\sum_{i=0}^{m} if_i &= 0\\ \frac{1}{\nu d}\sum_{i=0}^{m} if_i - \frac{1}{d}\sum_{i=0}^{m} f_0 &= 0\\ \frac{1}{\nu d}\sum_{i=0}^{m} if_i &= \frac{1}{d}\sum_{i=0}^{m} f_i\\ \hat{\nu} &= \frac{\sum_{i=0}^{m} if_i}{\sum_{i=0}^{m} f_i}. \qquad (3.43)\end{aligned}$$

Substituting (3.43) in (3.42),

$$\begin{aligned}\sum_{i=0}^{m}\frac{f_i q_i'(\hat{\theta})}{q_i(\hat{\theta})} &= \frac{2\hat{\nu}(2-d)^2}{d}\sum_{i=0}^{m} if_i - \frac{\hat{\nu}^2(2-d)^3}{d}\sum_{i=0}^{m} f_i\\ &= \frac{2\hat{\nu}^2(2-d)^2}{d}\sum_{i=1}^{m} f_i - \frac{\hat{\nu}^2(2-d)^3}{d}\sum_{i=0}^{m} f_i\\ &= [\hat{\nu}(2-d)]^2\sum_{i=0}^{m} f_i. \qquad (3.44)\end{aligned}$$

MLE of parameter d be found by solving (3.44). The package $hermite$ in program R allows to estimate parameter d given an univariate sample by means of the function $glm.hermite$, see (Moriña et al., 2015) for more details. The profile MLEs of ν and d for any fixed f_0 are

$$\hat{\nu}(f_0) = \frac{\sum_{i=0}^{m} if_i}{\sum_{i=0}^{m} f_i} \qquad (3.45)$$

$$\hat{d}(f_0) = glm.hermite(f_0, f_1, f_2, ..., f_m \sim 1, link = "log", start = NULL, m^\star = 2). \qquad (3.46)$$

The profile log-likelihood for any fixed f_0 be achieved as

$$l\Big(f_0, \hat{\nu}(f_0), \hat{d}(f_0)\Big) = \log\Gamma(n+f_0+1) - \sum_{i=0}^{m}\log\Gamma(f_i+1) + \sum_{i=0}^{m} f_i \log p\Big(i \mid \hat{\nu}(f_0), \hat{d}(f_0)\Big). \qquad (3.47)$$

The population size estimator under Hermite distribution based on profile likelihood is

$$\hat{N}_{Profile} = n + \hat{f}_0 \tag{3.48}$$

which $\hat{f}_0$ is the value of f_0 that maximizes (3.47).

3.3.2 Confidence interval estimation for population size N

Let $\hat{N}_{Profile}$ be profile maximum likelihood estimator (profile MLE) based on Hermite distribution. Since $\hat{N}_{Profile} = n + \hat{f}_0$, the likelihood ratio is defined by

$$2\big[l\big(n + \hat{f}_0, \nu(\hat{f}_0), d(\hat{f}_0)\big) - l\big(n + f_0, \hat{\nu}(f_0), \hat{d}(f_0)\big)\big] \sim \chi^2(1).$$

Using the log-likelihood ratio statistic, all Ns corresponding to

$$2\big[l\big(n + \hat{f}_0, \nu(\hat{f}_0), d(\hat{f}_0)\big) - l\big(n + f_0, \hat{\nu}(f_0), \hat{d}(f_0)\big)\big] \leq (z_{1-\alpha/2})^2$$

form the $100(1-\alpha)\%$ confidence set for $\hat{N}_{Profile}$. Therefore, the 95% confidence interval for $\hat{N}_{Profile}$ is the range of $\hat{N}_{Profile}$ that satisfies

$$2\big[l\big(n + \hat{f}_0, \nu(\hat{f}_0), d(\hat{f}_0)\big) - l\big(n + f_0, \hat{\nu}(f_0), \hat{d}(f_0)\big)\big] - (1.96)^2 \leq 0. \tag{3.49}$$

3.4 Simulation plan

3.4.1 Point estimation part

A simulation was used to study the performance of proposed estimator $(\hat{N}_{Herm})$ and to compare with those of other well-known estimators based upon homogeneous and heterogeneous case. The maximum likelihood $(\hat{N}_{MLE})$ and Turing's estimator $(\hat{N}_{Turing})$ were used as estimators in homogeneous case. Two population estimators Chao's lower bound estimator $(\hat{N}_{Chao})$ and Censored estimator $(\hat{N}_{Censored})$ were used as estimators in heterogeneous case. The count data were generated by Monte Carlo technique using program R. The population size was $N = 100, 500$, and $1,000$ with following distribution:

i. The Poisson distribution with parameters $\lambda \in \{1, 2, 3\}$.

ii. The Hermite distribution with

$$p_i = e^{\nu(-1+\frac{d-1}{2})}[\nu(2-d)]^i \sum_{j=0}^{[i/2]}(\tfrac{d-1}{2\nu(2-d)^2})^j \tfrac{1}{(i-2j)!j!}.$$

Here, parameters $\nu \in \{1, 2, 3, 4\}$ and dispersion parameters $d \in \{1.2, 1.4, 1.6, 1.8\}$.

iii. The negative binomial distribution with

$$p_i = \tfrac{\Gamma(k+i)}{\Gamma(i+1)\Gamma(k)} p^k (1-p)^i,$$

where parameters $p \in \{0.6, 0.7, 0.8\}$ and dispersion parameters $k \in \{3, 4, 5\}$. Here, expected value and variance are given as $\frac{k(1-p)}{p}$ and $\frac{k(1-p)}{p^2}$, respectively.

iv. The geometric distribution with

$$p_i = p(1-p)^i,$$

where parameters $p \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$.

The criteria of comparing the performance of estimators were relative bias (Rbias) and relative root mean square error (RRMSE) from 1,000 repeated times:

$$Rbias = \frac{E(\hat{N}) - N}{N}$$

$$RRMSE = \frac{1}{N}\sqrt{Var(\hat{N}) + (E(\hat{N}) - N)^2}.$$

Here, $E(\hat{N}) = \frac{1}{1{,}000}\sum_{t=1}^{1{,}000} \hat{N}_{(t)}$, $Var(\hat{N}) = \frac{1}{999}\sum_{t=1}^{1{,}000}\{\hat{N}_{(t)} - E(\hat{N})\}^2$, and $\hat{N}_{(t)}$ denotes the estimated values of the population size at replication t.

3.4.2 Interval estimation part

A simulation was used to investigate performance of confidence intervals based on the profile likelihood. Further simulation was conducted to compare confidence intervals obtained from the proposed one and several estimators including $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, $\hat{N}_{Chao}$, and $\hat{N}_{Censored}$ that were done by means of normal approximation, $\hat{N} \pm z_{1-\frac{\alpha}{2}}\widehat{Se}(\hat{N})$. The count data were generated by Monte Carlo technique using program R. The population size was $N = 100$ and $1,000$. Count data were generated from the following distribution:

i. The Poisson distribution with parameters $\lambda \in \{1, 3\}$.

ii. The Hermite distribution with

$$p_i = e^{\nu(-1+\frac{d-1}{2})}[\nu(2-d)]^i \sum_{j=0}^{[i/2]} (\tfrac{d-1}{2\nu(2-d)^2})^j \tfrac{1}{(i-2j)!j!}.$$

Here, parameters $\nu \in \{2, 4\}$, and dispersion parameters $d \in \{1.2, 1.4, 1.6, 1.8\}$.

The percentage of 100 simulated data in which the 95% confidence interval covered the true N called coverage probability (CP) and average lengths (AL) of achieved confidence intervals were the criteria for comparing the performance of estimation. The CP and AL be calculated as

$$CP = \frac{\sum_{t=1}^{100} C_{(t)}}{100}, \tag{3.50}$$

where $C_{(t)}$ equal to 1 if the true population size N contain in the confidence interval, and 0 otherwise.

$$AL = \frac{\sum_{t=1}^{100} (\hat{N}_{U_{(t)}} - \hat{N}_{L_{(t)}})}{100}, \tag{3.51}$$

where $\hat{N}_{U_{(t)}}$ and $\hat{N}_{L_{(t)}}$ are the upper and lower estimation of N at replication t, respectively.

3.5 Simulation results

3.5.1 Point estimation part

3.5.1.1 Simulation results based on the Poisson distribution

From the simulation results showed in Table 3.1, the Rbias of $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, $\hat{N}_{Chao}$, and $\hat{N}_{Herm}$ converged to zero when the population size increases, they were asymptotically unbiased with respect to the population size. $\hat{N}_{MLE}$ and $\hat{N}_{Turing}$ were derived on the basis of homogeneity Poisson model, they might be expected to be an appropriated choice. They gave the smallest Rbias among the other estimators. If parameter λ and population size N increased, the Rbias of all estimators decreased . $\hat{N}_{Censored}$ provided severe overestimation for all population size, especially for small λ.

RRMSE of all estimators were presented in Table 3.2, revealed that $\hat{N}_{MLE}$ and $\hat{N}_{Turing}$ produced the smallest RRMSE for all cases. $\hat{N}_{Herm}$ provided smaller RRMSE than $\hat{N}_{Chao}$ and $\hat{N}_{Censored}$ when parameter $\lambda = 3$ for all population sizes. $\hat{N}_{Censored}$ provided the largest RRMSE for all population size, especially for small λ. Therefore, $\hat{N}_{Herm}$ is an efficient estimator for estimating the population size N based on Poisson distribution when $N \geqslant 500$ and $\lambda \geqslant 2$.

Table 3.1: The relative bias of estimators with different parameters in the Poisson distribution.

λ	MLE	Turing	Chao	Censored	Herm
N=100					
1	0.0223	0.0216	0.0427	0.5621	0.1222
2	0.0053	0.0058	0.0154	0.2721	0.0347
3	0.0023	0.0019	0.0075	0.1312	0.0116
N=500					
1	0.0020	0.0019	0.0049	0.5179	0.0614
2	0.0020	0.0016	0.0024	0.2605	0.0124
3	0.0001	0.0000	0.0009	0.1280	0.0031
N=1,000					
1	0.0007	0.0009	0.0035	0.5150	0.0379
2	-0.0003	-0.0007	-0.0006	0.2569	0.0055
3	0.0002	0.0000	0.0003	0.1270	0.0021

Table 3.2: The relative root mean square error of estimators with different parameters in the Poisson distribution.

λ	MLE	Turing	Chao	Censored	Herm
N=100					
1	0.1295	0.1343	0.1927	0.6262	0.2650
2	0.0483	0.0517	0.0732	0.2916	0.0878
3	0.0259	0.0279	0.0401	0.1432	0.0362
N=500					
1	0.0489	0.0504	0.0674	0.5284	0.1429
2	0.0214	0.0231	0.0318	0.2647	0.0368
3	0.0110	0.0122	0.0167	0.1306	0.0127
N=1,000					
1	0.0383	0.0395	0.0516	0.5213	0.0944
2	0.0154	0.0165	0.0219	0.2591	0.0194
3	0.0079	0.0086	0.0114	0.1282	0.0089

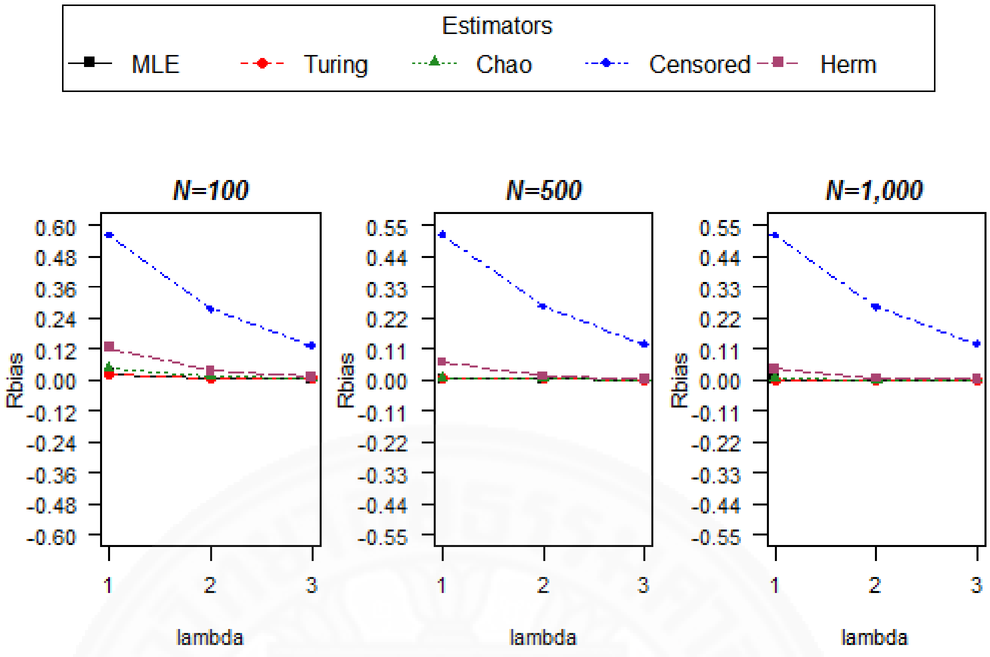


Figure 3.2: The relative bias of estimators with different parameters in the Poisson distribution.

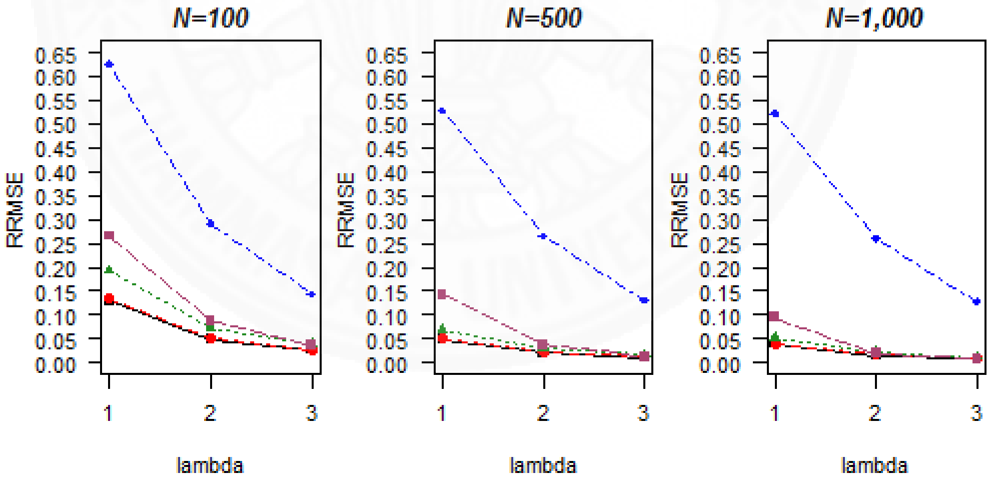


Figure 3.3: The relative root mean square error of estimators with different parameters in the Poisson distribution.

3.5.1.2 Simulation results based on the Hermite distribution

$\hat{N}_{Herm}$ was derived on the basis of Hermite distribution, it might be expected to be an appropriated choice. From Table 3.3, $\hat{N}_{Herm}$ gave the smallest Rbias among the other estimators for all cases. $\hat{N}_{Herm}$ converged to zero when the population sizes increases, it was asymptotically unbiased with respect to the population size. $\hat{N}_{Censored}$ provided severe overestimation when dispersion parameter $d = 1.2$, and severe underestimation when dispersion parameter $d = 1.8$ for all population size, especially for small ν. Moreover, $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, and $\hat{N}_{Chao}$ provided severe underestimation when dispersion parameter $d = 1.6$ and $d = 1.8$ for all population size, especially for small ν.

From Table 3.4, $\hat{N}_{Herm}$ produced the smallest RRMSE in almost all cases. $\hat{N}_{Censored}$ provided the largest RRMSE when dispersion parameter $d = 1.2$ for all population sizes and ν. Moreover, $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, and $\hat{N}_{Chao}$ provided larger RRMSEs than $\hat{N}_{Censored}$ and $\hat{N}_{Herm}$ when dispersion parameter $d = 1.6$ and $d = 1.8$ for all population size, especially for small ν. Therefore, $\hat{N}_{Herm}$ is the best performing estimator for estimating the population size N based on Hermite distribution.

Table 3.3: The relative bias of estimators with different parameters in the Hermite distribution.

ν	d	MLE	Turing	Chao	Censored	Herm
N=100						
1	1.2	-0.1183	-0.1087	-0.0695	0.3382	0.0453
	1.4	-0.2454	-0.2377	-0.2108	0.1004	-0.0061
	1.6	-0.3617	-0.3693	-0.3810	-0.1596	0.0250
	1.8	-0.4649	-0.4918	-0.5193	-0.3985	0.0620
2	1.2	-0.0466	-0.0361	-0.0110	0.2283	0.0262
	1.4	-0.1049	-0.0907	-0.0611	0.1511	-0.0004
	1.6	-0.1701	-0.1616	-0.1510	0.0273	-0.0180
	1.8	-0.2487	-0.2576	-0.2724	-0.1559	0.0040
3	1.2	-0.0223	-0.0153	-0.0013	0.1277	0.0086
	1.4	-0.0492	-0.0377	-0.0173	0.1113	0.0056
	1.6	-0.0885	-0.0775	-0.0619	0.0556	-0.0130
	1.8	-0.1410	-0.1393	-0.1435	-0.0551	-0.0120
4	1.2	-0.0104	-0.0062	0.0024	0.0694	0.0034
	1.4	-0.0240	-0.0163	-0.0020	0.0700	0.0056
	1.6	-0.0480	-0.0390	-0.0257	0.0471	-0.0043
	1.8	-0.0771	-0.0725	-0.0714	-0.0099	-0.0093
N=500						
1	1.2	-0.1283	-0.1162	-0.0894	0.3213	0.0125
	1.4	-0.2521	-0.2444	-0.2329	0.0833	0.0034
	1.6	-0.3635	-0.3707	-0.3900	-0.1662	0.0108
	1.8	-0.4657	-0.4923	-0.5218	-0.4024	0.0165
2	1.2	-0.0491	-0.0382	-0.0213	0.2226	0.0153
	1.4	-0.1061	-0.0921	-0.0718	0.1453	-0.0017
	1.6	-0.1722	-0.1642	-0.1606	0.0201	-0.0042
	1.8	-0.2494	-0.2572	-0.2741	-0.1564	0.0007
3	1.2	-0.0206	-0.0132	-0.0039	0.1298	0.0037
	1.4	-0.0506	-0.0389	-0.0240	0.1076	0.0040
	1.6	-0.0885	-0.0776	-0.0678	0.0532	-0.0073
	1.8	-0.1386	-0.1367	-0.1431	-0.0523	0.0011
4	1.2	-0.0094	-0.0051	-0.0005	0.0698	0.0014
	1.4	-0.0241	-0.0162	-0.0073	0.0689	0.0033
	1.6	-0.0460	-0.0371	-0.0284	0.0480	-0.0010
	1.8	-0.0793	-0.0747	-0.0758	-0.0129	-0.0011
N=1,000						
1	1.2	-0.1316	-0.1195	-0.0943	0.3163	0.0112
	1.4	-0.2527	-0.2442	-0.2321	0.0836	-0.0101
	1.6	-0.3644	-0.3712	-0.3910	-0.1668	0.0062
	1.8	-0.4661	-0.4924	-0.5218	-0.4028	0.0084
2	1.2	-0.0483	-0.0373	-0.0206	0.2236	0.0099
	1.4	-0.1064	-0.0919	-0.0716	0.1462	-0.0027
	1.6	-0.1730	-0.1645	-0.1608	0.0205	-0.0035
	1.8	-0.2496	-0.2574	-0.2745	-0.1565	0.0002
3	1.2	-0.0215	-0.0143	-0.0062	0.1274	0.0012
	1.4	-0.0502	-0.0382	-0.0238	0.1088	0.0046
	1.6	-0.0882	-0.0769	-0.0668	0.0551	-0.0048
	1.8	-0.1388	-0.1368	-0.1435	-0.0526	0.0003
4	1.2	-0.0098	-0.0056	-0.0017	0.0683	0.0002
	1.4	-0.0246	-0.0168	-0.0088	0.0678	0.0015
	1.6	-0.0464	-0.0374	-0.0288	0.0479	-0.0011
	1.8	-0.0785	-0.0737	-0.0751	-0.0115	0.0000

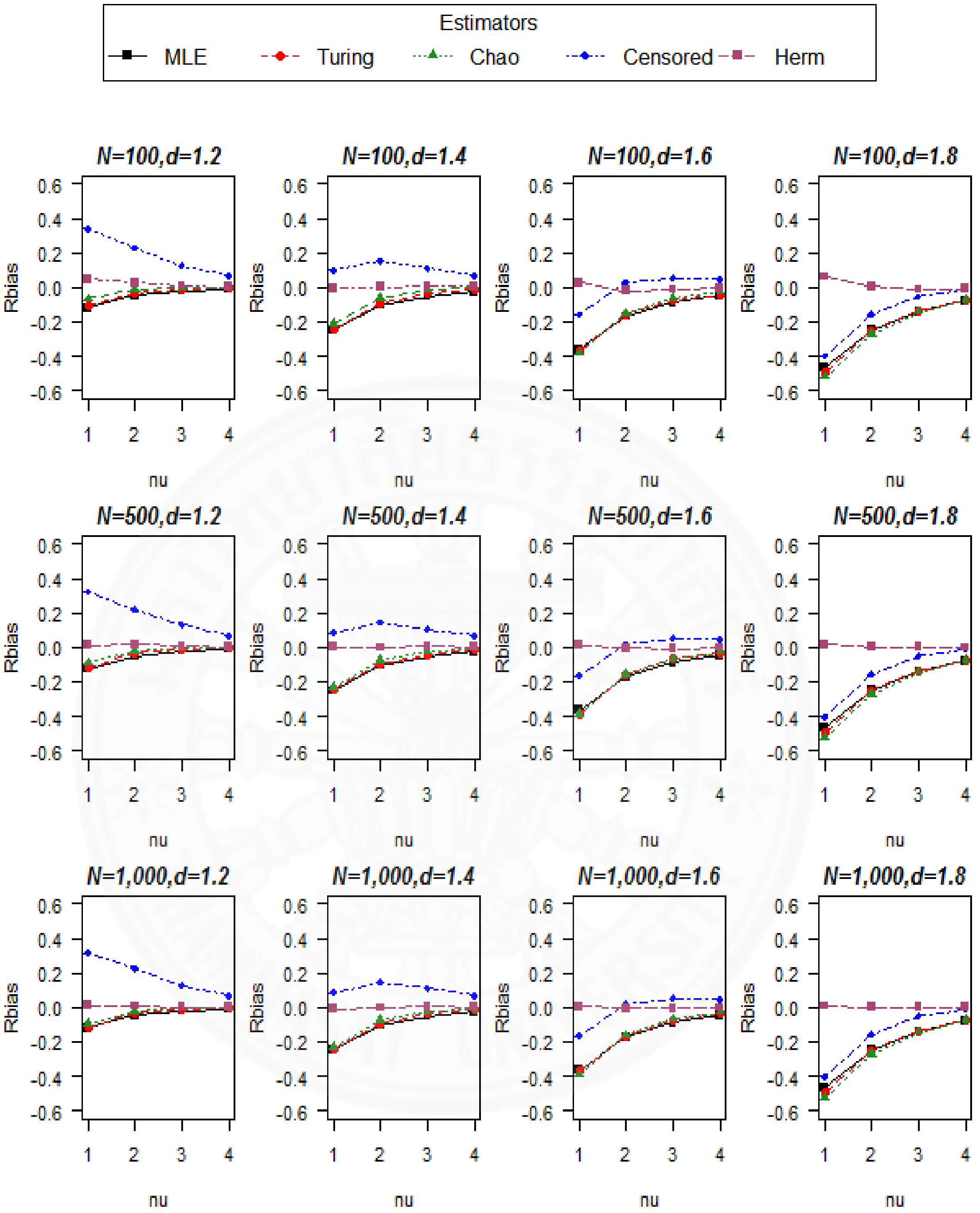


Figure 3.4: The relative bias of estimators with different parameters in the Hermite distribution.

Table 3.4: The relative root mean square error of estimators with different parameters in the Hermite distribution.

ν	d	MLE	Turing	Chao	Censored	Herm
N=100						
1	1.2	0.1576	0.1540	0.1776	0.4068	0.2215
	1.4	0.2610	0.2550	0.2475	0.2086	0.2357
	1.6	0.3696	0.3772	0.3909	0.2088	0.2775
	1.8	0.4692	0.4956	0.5225	0.4079	0.3004
2	1.2	0.0688	0.0649	0.0827	0.2527	0.1015
	1.4	0.1167	0.1061	0.1013	0.1848	0.0978
	1.6	0.1776	0.1703	0.1656	0.0971	0.1046
	1.8	0.2538	0.2625	0.2773	0.1724	0.1071
3	1.2	0.0365	0.0346	0.0456	0.1421	0.0486
	1.4	0.0587	0.0511	0.0545	0.1297	0.0561
	1.6	0.0949	0.0857	0.0799	0.0857	0.0580
	1.8	0.1464	0.1449	0.1494	0.0793	0.0614
4	1.2	0.0206	0.0199	0.0285	0.0795	0.0247
	1.4	0.0318	0.0274	0.0357	0.0824	0.0323
	1.6	0.0537	0.0466	0.0442	0.0653	0.0357
	1.8	0.0827	0.0787	0.0787	0.0450	0.0418
N=500						
1	1.2	0.1367	0.1261	0.1117	0.3365	0.1354
	1.4	0.2550	0.2477	0.2391	0.1154	0.1497
	1.6	0.3650	0.3721	0.3917	0.1756	0.1177
	1.8	0.4665	0.4930	0.5223	0.4041	0.1073
2	1.2	0.0540	0.0449	0.0391	0.2273	0.0649
	1.4	0.1086	0.0953	0.0793	0.1525	0.0644
	1.6	0.1737	0.1659	0.1632	0.0462	0.0597
	1.8	0.2504	0.2582	0.2751	0.1597	0.0447
3	1.2	0.0239	0.0186	0.0192	0.1326	0.0227
	1.4	0.0525	0.0417	0.0318	0.1113	0.0335
	1.6	0.0899	0.0794	0.0709	0.0604	0.0366
	1.8	0.1397	0.1379	0.1443	0.0587	0.0276
4	1.2	0.0121	0.0097	0.0114	0.0720	0.0106
	1.4	0.0258	0.0189	0.0154	0.0714	0.0166
	1.6	0.0473	0.0388	0.0318	0.0520	0.0208
	1.8	0.0804	0.0758	0.0771	0.0236	0.0181
N=1,000						
1	1.2	0.1355	0.1242	0.1057	0.3241	0.1184
	1.4	0.2542	0.2459	0.2354	0.1012	0.1188
	1.6	0.3651	0.3719	0.3918	0.1714	0.0791
	1.8	0.4666	0.4928	0.5221	0.4037	0.0763
2	1.2	0.0507	0.0408	0.0314	0.2261	0.0485
	1.4	0.1075	0.0934	0.0753	0.1497	0.0517
	1.6	0.1738	0.1654	0.1621	0.0353	0.0443
	1.8	0.2501	0.2579	0.2750	0.1582	0.0329
3	1.2	0.0234	0.0174	0.0150	0.1290	0.0152
	1.4	0.0512	0.0397	0.0282	0.1108	0.0266
	1.6	0.0889	0.0778	0.0685	0.0587	0.0274
	1.8	0.1394	0.1374	0.1441	0.0557	0.0189
4	1.2	0.0112	0.0081	0.0082	0.0694	0.0068
	1.4	0.0255	0.0182	0.0129	0.0691	0.0124
	1.6	0.0471	0.0383	0.0307	0.0499	0.0151
	1.8	0.0791	0.0743	0.0758	0.0178	0.0123

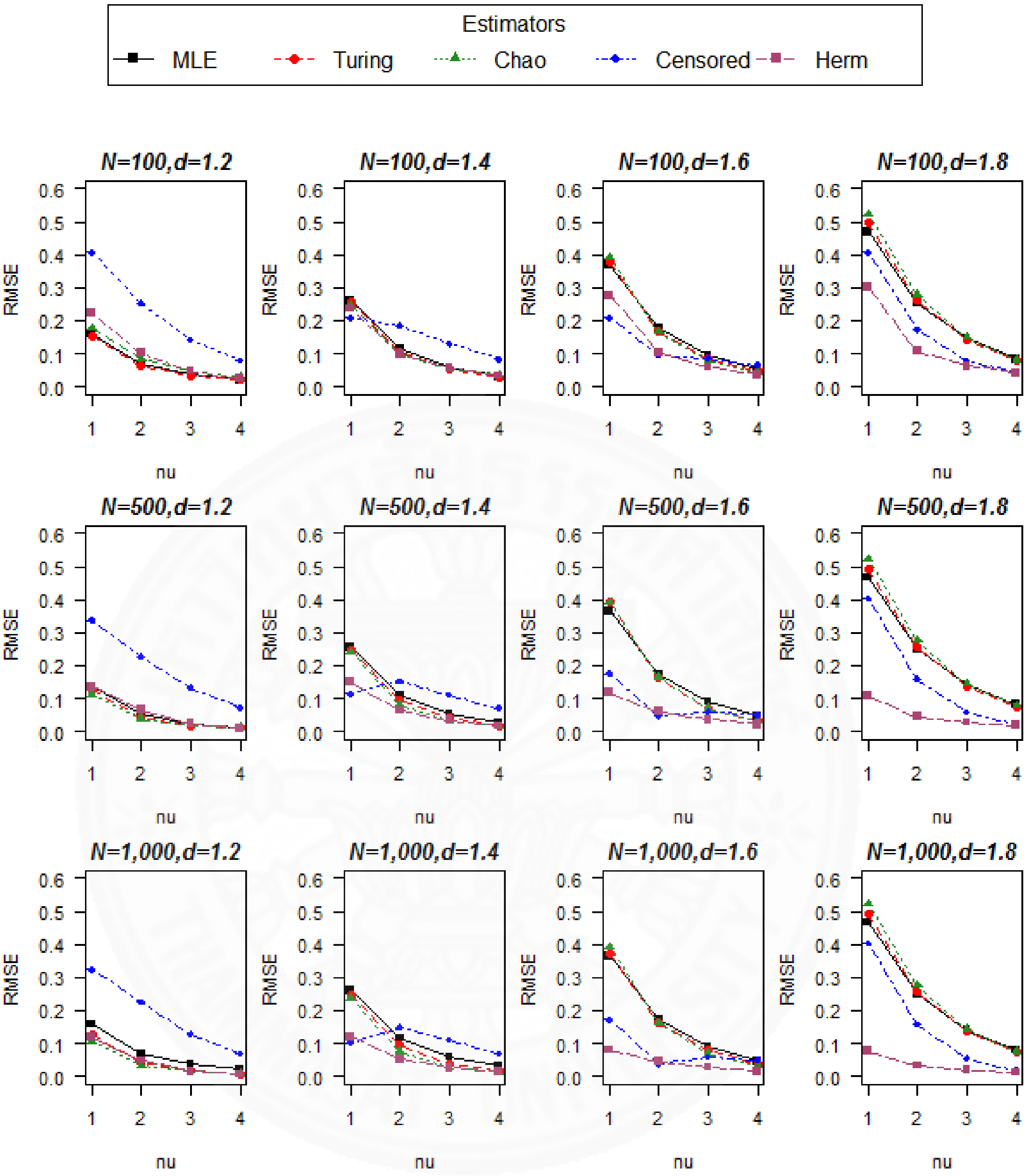


Figure 3.5: The relative root mean square error of estimators with different parameters in the Hermite distribution.

3.5.1.3 Simulation results based on the geometric distribution

$\hat{N}_{Censored}$ was derived on the basis of geometric distribution, it might be expected to be an appropriated choice. From Table 3.5, $\hat{N}_{Censored}$ gave the smallest Rbias among the other estimators for all cases. $\hat{N}_{Censored}$ converged to zero when the population sizes increases, it was asymptotically unbiased with respect to the population size. $\hat{N}_{MLE}$ and $\hat{N}_{Turing}$ provided severe underestimation for all population size, especially for large p. Moreover, $\hat{N}_{Chao}$ and $\hat{N}_{Herm}$ provided underestimation for all population size, especially for large p.

From Table 3.6, $\hat{N}_{Censored}$ produced the smallest RRMSE for all cases. $\hat{N}_{MLE}$ and $\hat{N}_{Turing}$ provided the largest RRMSE for all population size, especially for large p. Moreover, $\hat{N}_{Chao}$ and $\hat{N}_{Herm}$ provided the large RRMSE for all population size, especially for large p. Therefore, $\hat{N}_{Censored}$ is the best performing estimator for estimating the population size N based on geometric distribution.

Table 3.5: The relative bias of estimators with different parameters in the geometric distribution.

p	MLE	Turing	Chao	Censored	Herm
N=100					
0.1	-0.1012	-0.0921	-0.0354	-0.0010	-0.0973
0.2	-0.1958	-0.1684	-0.0865	-0.0011	-0.1571
0.3	-0.2677	-0.2293	-0.1339	0.0033	-0.1873
0.4	-0.3231	-0.2824	-0.1767	0.0078	-0.2051
0.5	-0.3598	-0.3217	-0.2110	0.0282	-0.1974
N=500					
0.1	-0.0995	-0.0904	-0.0474	0.0005	-0.0959
0.2	-0.1941	-0.1664	-0.0969	0.0011	-0.1566
0.3	-0.2687	-0.2291	-0.1433	0.0042	-0.1918
0.4	-0.3260	-0.2842	-0.1954	0.0029	-0.2179
0.5	-0.3713	-0.3326	-0.2449	0.0023	-0.2401
N=1,000					
0.1	-0.0994	-0.0903	-0.0479	0.0009	-0.0959
0.2	-0.1940	-0.1663	-0.0986	0.0003	-0.1566
0.3	-0.2690	-0.2296	-0.1465	0.0021	-0.1925
0.4	-0.3270	-0.2850	-0.1971	0.0016	-0.2206
0.5	-0.3722	-0.3333	-0.2482	0.0007	-0.2434

Table 3.6: The relative root mean square error of estimators with different parameters in the geometric distribution.

p	MLE	Turing	Chao	Censored	Herm
N=100					
0.1	0.1059	0.0974	0.0834	0.0499	0.1022
0.2	0.1999	0.1737	0.1240	0.0764	0.1633
0.3	0.2720	0.2353	0.1697	0.1020	0.1964
0.4	0.3289	0.2906	0.2248	0.1396	0.2237
0.5	0.3694	0.3341	0.2692	0.1895	0.2503
N=500					
0.1	0.1004	0.0915	0.0525	0.0219	0.0969
0.2	0.1950	0.1674	0.1021	0.0334	0.1579
0.3	0.2695	0.2302	0.1489	0.0433	0.1936
0.4	0.3271	0.2858	0.2027	0.0612	0.2206
0.5	0.3729	0.3347	0.2532	0.0789	0.2455
N=1,000					
0.1	0.0999	0.0909	0.0504	0.0156	0.0964
0.2	0.1944	0.1668	0.1011	0.0233	0.1573
0.3	0.2695	0.2302	0.1495	0.0327	0.1934
0.4	0.3276	0.2858	0.2006	0.0439	0.2220
0.5	0.3729	0.3342	0.2520	0.0537	0.2457

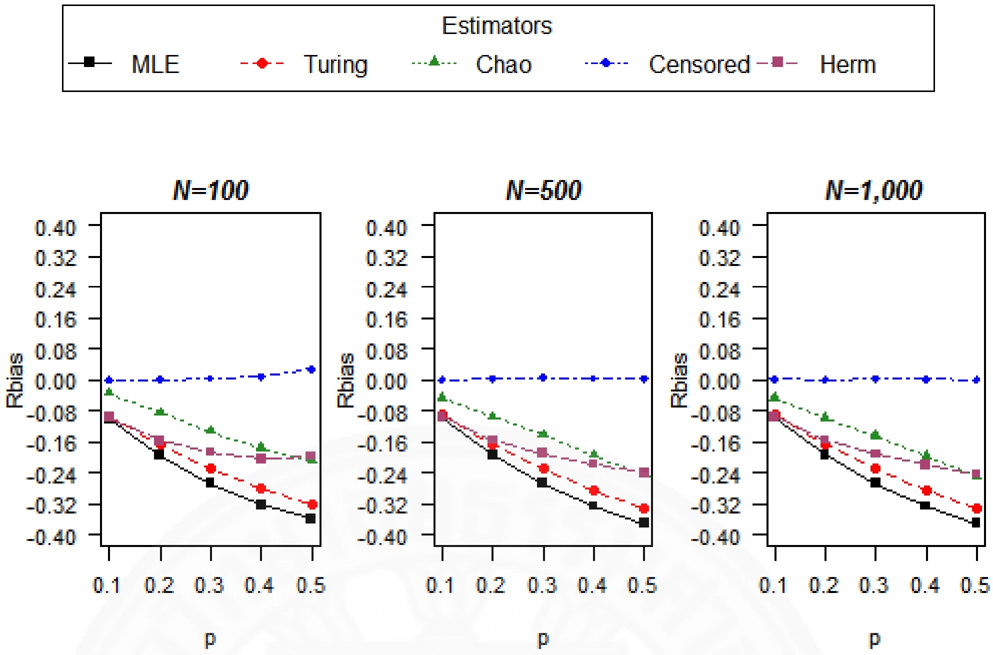


Figure 3.6: The relative bias of estimators with different parameters in the geometric distribution.

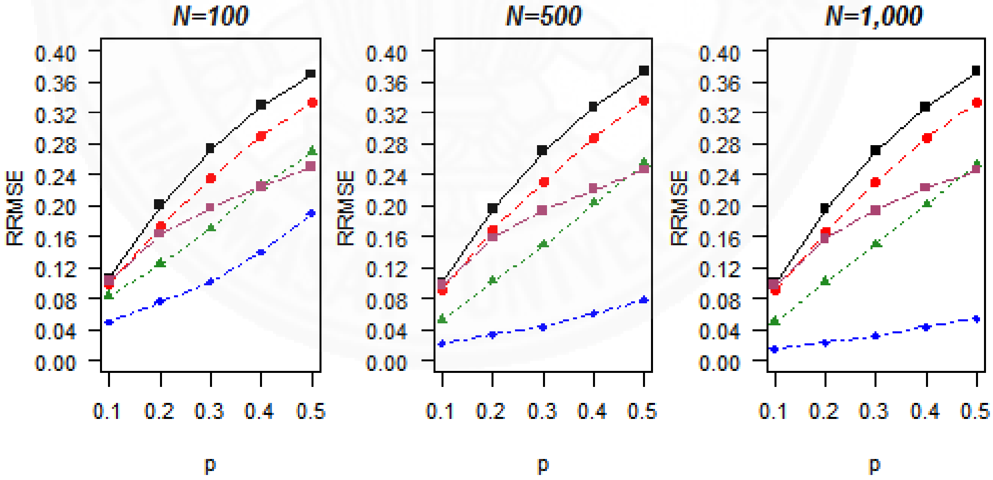


Figure 3.7: The relative root mean square error of estimators with different parameters in the geometric distribution.

3.5.1.4 Simulation results based on the negative binomial distribution

From Table 3.7, $\hat{N}_{Herm}$ produced the smallest Rbias among the other estimators for $k = 3$ and $k = 4$ when $N = 500$ and $N = 1,000$. $\hat{N}_{Herm}$ gave the smallest Rbias among the other estimators for $k = 3$ and $N = 100$. $\hat{N}_{Chao}$ gave the smallest Rbias for $k = 5$ in all population size. $\hat{N}_{MLE}$ and $\hat{N}_{Turing}$ provided underestimation for all population size. $\hat{N}_{Censored}$ provided severe overestimation for all population size, especially for large p. Moreover, Rbias of all estimators decreased when parameter p decreased.

From Table 3.8, $\hat{N}_{Herm}$ produced the smallest RRMSE among the other estimators for $k = 3$ when $N = 500$ and $N = 1,000$. $\hat{N}_{Chao}$ gave the smallest RRMSE for $k = 4$ and $k = 5$ when $N = 500$ and $N = 1,000$. Moreover, $\hat{N}_{Turing}$ provided the smallest RRMSE in almost all cases for $N = 100$. $\hat{N}_{Censored}$ provided the largest RRMSE for all cases, especially for large p. Therefore, $\hat{N}_{Herm}$ is an efficient estimator for estimating the population size N based on negative binomial distribution when the population size greater than 500.

Table 3.7: The relative bias of estimators with different parameters in the negative binomial distribution.

k	p	MLE	Turing	Chao	Censored	Herm
N=100						
3	0.6	-0.1240	-0.0960	-0.0382	0.1798	0.0096
	0.7	-0.1574	-0.1315	-0.0633	0.2538	0.0142
	0.8	-0.1678	-0.1493	-0.0755	0.3733	0.0754
4	0.6	-0.0769	-0.0552	-0.0168	0.1462	0.0277
	0.7	-0.1047	-0.0824	-0.0315	0.2328	0.0320
	0.8	-0.1231	-0.1060	-0.0524	0.3581	0.0678
5	0.6	-0.0485	-0.0324	-0.0062	0.1103	0.0242
	0.7	-0.0726	-0.0539	-0.0161	0.1979	0.0353
	0.8	-0.0926	-0.0769	-0.0310	0.3318	0.0515
N=500						
3	0.6	-0.1267	-0.0983	-0.0514	0.1734	0.0075
	0.7	-0.1620	-0.1345	-0.0801	0.2429	0.0108
	0.8	-0.1905	-0.1696	0.1192	0.3271	0.0295
4	0.6	-0.0785	-0.0563	-0.0246	0.1427	0.0230
	0.7	-0.1099	-0.0862	-0.0452	0.2248	0.0399
	0.8	-0.1379	-0.1185	-0.0773	0.3325	0.0490
5	0.6	-0.0488	-0.0323	-0.0114	0.1098	0.0273
	0.7	-0.0769	-0.0565	-0.0254	0.1952	0.0465
	0.8	-0.1059	-0.0874	-0.0511	0.3145	0.0577
N=1,000						
3	0.6	-0.1282	-0.0995	-0.0529	0.1714	0.0016
	0.7	-0.1624	-0.1351	-0.0831	0.2408	0.0099
	0.8	-0.1942	-0.1733	-0.1262	0.3181	0.0205
4	0.6	-0.0789	-0.0566	-0.0256	0.1425	0.0233
	0.7	-0.1111	-0.0870	-0.0465	0.2243	0.0379
	0.8	-0.1406	-0.1206	-0.0791	0.3295	0.0517
5	0.6	-0.0492	-0.0326	-0.0121	0.1097	0.0267
	0.7	-0.0776	-0.0571	-0.0269	0.1943	0.0473
	0.8	-0.1078	-0.0888	-0.0531	0.3119	0.0599

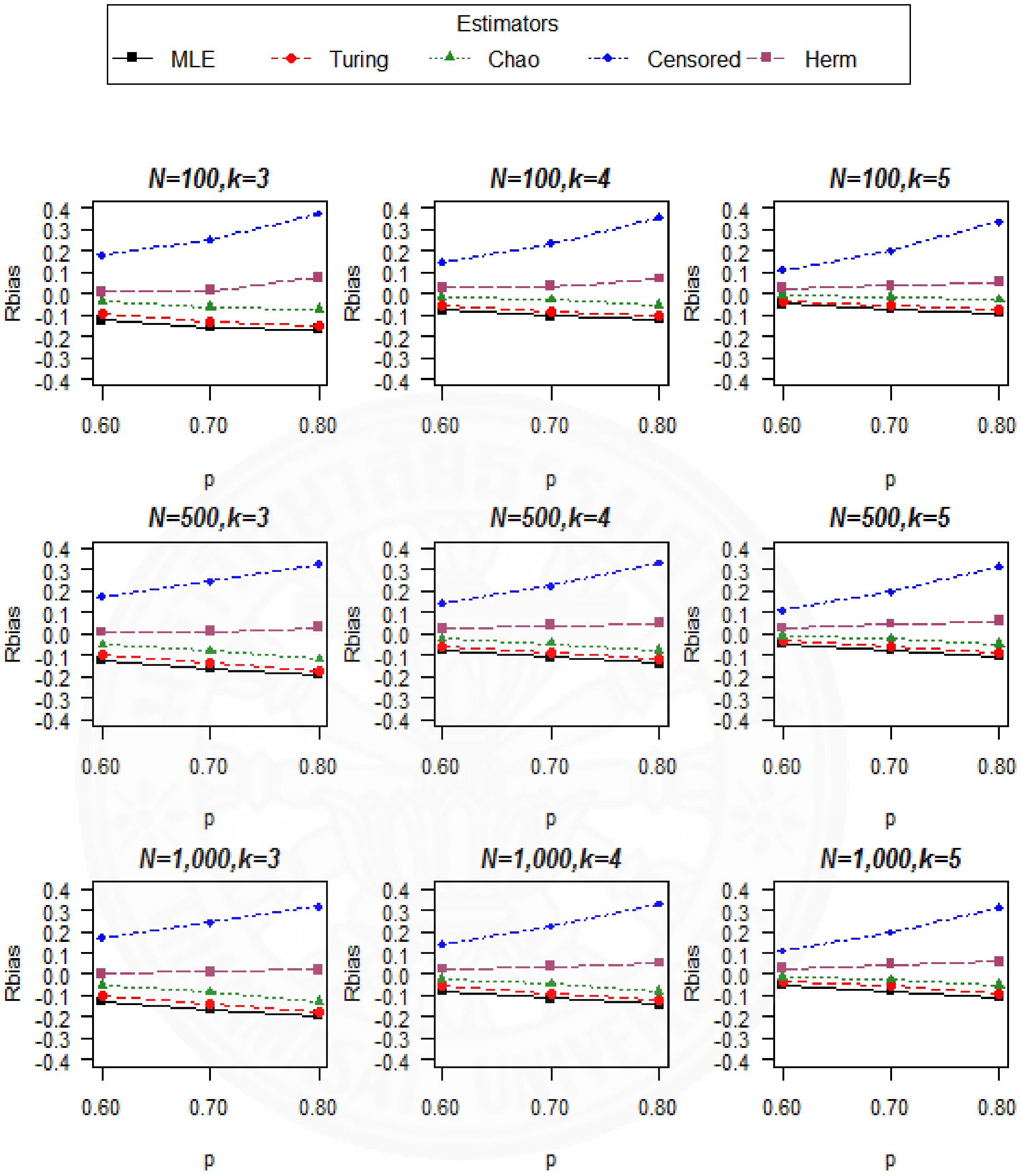


Figure 3.8: The relative bias of estimators with different parameters in the negative binomial distribution.

Table 3.8: The relative root mean square error of estimators with different parameters in the negative binomial distribution.

k	p	MLE	Turing	Chao	Censored	Herm
N=100						
3	0.6	0.1345	0.1113	0.1021	0.2126	0.0850
	0.7	0.1754	0.1547	0.1508	0.3055	0.1551
	0.8	0.2210	0.2109	0.2527	0.4904	0.3321
4	0.6	0.0853	0.0685	0.0685	0.1677	0.0689
	0.7	0.1222	0.1065	0.1115	0.2708	0.1179
	0.8	0.1674	0.1573	0.1799	0.4303	0.2303
5	0.6	0.0567	0.0451	0.0492	0.1259	0.0510
	0.7	0.0858	0.0734	0.0845	0.2240	0.0949
	0.8	0.1273	0.1197	0.1431	0.3808	0.1795
N=500						
3	0.6	0.1287	0.1013	0.0652	0.1807	0.0385
	0.7	0.1657	0.1397	0.0985	0.2549	0.0731
	0.8	0.1993	0.1806	0.1501	0.3526	0.1554
4	0.6	0.0804	0.0594	0.0371	0.1474	0.0366
	0.7	0.1131	0.0911	0.0641	0.2326	0.0729
	0.8	0.1458	0.1284	0.1041	0.3475	0.1373
5	0.6	0.0505	0.0352	0.0236	0.1133	0.0346
	0.7	0.0798	0.0611	0.0438	0.2009	0.0674
	0.8	0.1125	0.0962	0.0777	0.3250	0.1234
N=1,000						
3	0.6	0.1293	0.1011	0.0596	0.1749	0.0264
	0.7	0.1643	0.1377	0.0928	0.2470	0.0494
	0.8	0.1983	0.1785	0.1407	0.3309	0.1121
4	0.6	0.0798	0.0581	0.0323	0.1449	0.0307
	0.7	0.1128	0.0894	0.0560	0.2282	0.0576
	0.8	0.1445	0.1254	0.0935	0.3373	0.1096
5	0.6	0.0501	0.0341	0.0191	0.1114	0.0305
	0.7	0.0791	0.0593	0.0365	0.1970	0.0586
	0.8	0.1110	0.0932	0.0675	0.3173	0.1018

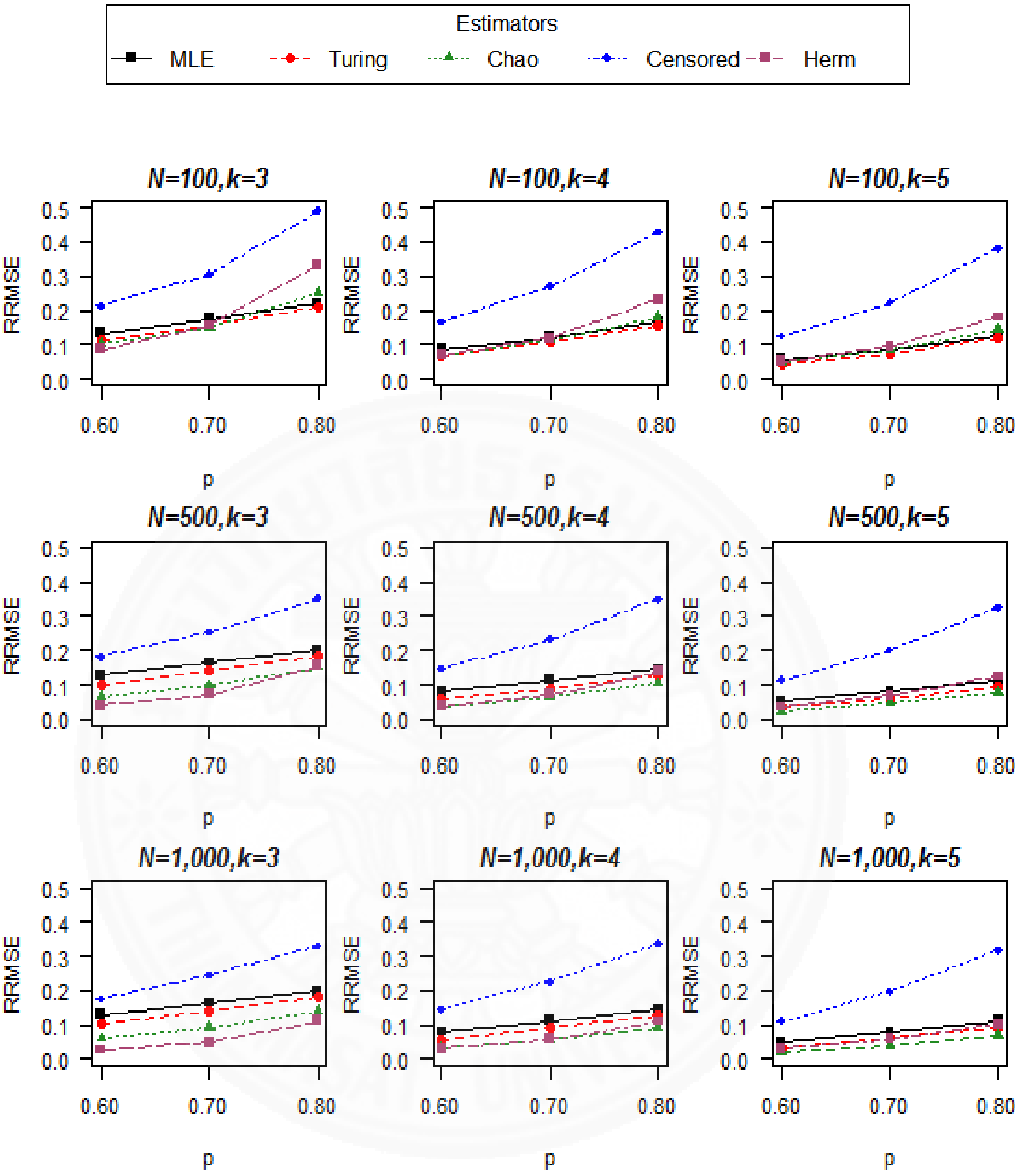


Figure 3.9: The relative root mean square error of estimators with different parameters in the negative binomial distribution.

3.5.2 Interval estimation part

3.5.2.1 Simulation results based on the Poisson distribution

The coverage probabilities of the 95% CIs based on the Poisson distribution were presented in Table 3.9. A comparison of the estimators showed that $\hat{N}_{Herm}$ provided overestimation for small parameter $\lambda = 1$ and small population size $N = 100$. $\hat{N}_{Censored}$ provided severe overestimation for all cases, especially for small λ.

The CPs of proposed CI gave the highest CP for all cases. CIs of $\hat{N}_{Chao}$ performed the second best in which the CP increased with increasing N. CIs of $\hat{N}_{Censored}$ provided low CP for all cases, and did not cover the true population size for $N = 1,000$. CIs of $\hat{N}_{MLE}$ and $\hat{N}_{Turing}$ provided low CP for small parameter $\lambda = 1$, and increased to the nominal level when the parameter λ increase.

Since ALs of proposed CI were quite shorter than those of $\hat{N}_{Chao}$, the proposed CI is the best choice for estimating the CI of population size N based on Poisson distribution, especially for small parameter $\lambda = 1$ and population size $N = 100$ and for large parameter $\lambda = 3$ and population size $N = 1,000$.

Table 3.9: Comparison of various estimators and 95% CI of N with different parameters in the Poisson distribution.

	Estimators	Average $\hat{N}$	Average $\widehat{se}(\hat{N})$	Coverage probability	Average length
N=100			$Poi(1)$		
	MLE	101.41	6.14	0.80	24.06
	Turing	101.90	4.08	0.49	15.98
	Chao	105.72	17.41	0.91	68.26
	Censored	156.61	28.57	0.50	111.97
	Herm	114.91	-	-	-
	Profile	103.22	-	1.00	30.62
			$Poi(3)$		
	MLE	100.41	2.22	0.95	8.69
	Turing	100.30	2.27	0.92	8.90
	Chao	100.71	3.58	0.92	14.03
	Censored	112.77	6.64	0.59	26.04
	Herm	101.40	-	-	-
	Profile	101.20	-	0.99	16.73
N=1,000			$Poi(1)$		
	MLE	997.46	19.13	0.71	74.99
	Turing	997.79	12.87	0.51	50.44
	Chao	1,000.08	47.46	0.95	186.02
	Censored	1,509.51	84.54	0.00	331.38
	Herm	1,029.80	-	-	-
	Profile	1,027.85	-	1.00	196.00
			$Poi(3)$		
	MLE	999.88	7.06	0.90	27.66
	Turing	1,000.25	7.37	0.87	28.88
	Chao	1,001.41	11.28	0.93	44.22
	Censored	1,128.99	21.56	0.00	84.52
	Herm	1,002.48	-	-	-
	Profile	1,001.75	-	0.99	33.21

3.5.2.2 Simulation results based on the Hermite distribution

The coverage probabilities of the 95% CIs based on the Hermite distribution were presented in Tables 3.10 and 3.11. A comparison of the estimators showed that $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, and $\hat{N}_{Chao}$ provided severe underestimation for small parameter $\nu = 2$, especially for $d = 1.6$ and $d = 1.8$, in all population size. $\hat{N}_{Censored}$ provided overestimation for small parameter $\nu = 2$ for $d = 1.2$ and $d = 1.4$, in all population size.

For $N = 100$, the proposed CI provided the highest CP in all cases. The CIs of $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, $\hat{N}_{Chao}$, and $\hat{N}_{Censored}$ provided low CP in almost all cases. The CIs of $\hat{N}_{Turing}$, $\hat{N}_{Chao}$, and proposed CI were close to the nominal level when the count data was generated from $Herm(4, 1.2)$. The CP of $\hat{N}_{Censored}$ was similar to the proposed CI when the count data were generated from $Herm(2, 1.6)$ and $Herm(4, 1.6)$, but the ALs of the proposed CI were shorter than those of $\hat{N}_{Censored}$.

For $N = 1,000$, the proposed CI provides the highest CP for all cases. The CIs of $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, $\hat{N}_{Chao}$, and $\hat{N}_{Censored}$ do not cover the true population size N in almost all cases. The CP of $\hat{N}_{Censored}$ is higher than the proposed CI when the count data were generated from $Herm(2, 1.6)$. The CI of $\hat{N}_{Censored}$ gives a smaller AL in this case.

The proposed CI provided the highest CP and quite short AL for all cases. Therefore, the proposed CI is the best choice for estimating the CI of population size N based on Hermite distribution.

Table 3.10: Comparison of various estimators and 95% CI of N for $N = 100$ with different parameters in the Hermite distribution .

Estimators	Average $\hat{N}$	Average $\widehat{se}(\hat{N})$	Coverage probability	Average length	Average $\hat{N}$	Average $\widehat{se}(\hat{N})$	Coverage probability	Average length
		$Herm(2, 1.2)$				$Herm(4, 1.2)$		
MLE	95.07	3.40	0.61	13.34	99.14	1.34	0.88	5.27
Turing	96.26	3.30	0.66	12.92	99.62	1.60	0.94	6.26
Chao	98.79	7.42	0.86	29.08	100.74	2.92	0.96	11.43
Censord	122.73	11.97	0.59	46.92	107.52	4.83	0.80	18.93
Herm	102.50	-	-	-	100.73	-	-	-
Profile	102.41	-	0.96	40.24	100.68	-	0.99	11.93
		$Herm(2, 1.4)$				$Herm(4, 1.4)$		
MLE	89.69	3.06	0.25	12.00	97.65	1.26	0.57	4.94
Turing	91.21	3.09	0.34	12.12	98.46	1.62	0.75	6.34
Chao	94.76	7.32	0.70	28.70	100.15	3.28	0.87	12.85
Censored	115.36	11.11	0.83	43.54	107.12	5.05	0.82	19.80
Herm	100.16	-	-	-	100.78	-	-	-
Profile	97.69	-	0.97	39.23	99.79	-	0.97	12.26
		$Herm(2, 1.6)$				$Herm(4, 1.6)$		
MLE	82.62	2.69	0.04	10.55	95.17	1.18	0.25	4.63
Turing	83.46	2.75	0.04	10.78	96.04	1.58	0.41	6.18
Chao	84.57	5.27	0.23	20.66	97.37	3.01	0.68	11.78
Censored	101.97	9.20	0.93	36.07	104.43	4.96	0.94	19.44
Herm	98.16	-	-	-	99.22	-	-	-
Profile	94.98	-	0.91	32.08	97.88	-	0.93	13.12
		$Herm(2, 1.8)$				$Herm(4, 1.8)$		
MLE	75.56	2.29	0.00	8.96	92.51	1.09	0.02	4.26
Turing	74.77	2.15	0.00	8.44	92.98	1.35	0.07	5.31
Chao	73.30	2.39	0.00	9.37	93.09	1.80	0.19	7.07
Censored	84.99	6.22	0.39	24.39	99.18	4.17	0.88	16.33
Herm	100.33	-	-	-	99.53	-	-	-
Profile	99.27	-	0.96	36.25	98.59	-	0.95	14.27

Table 3.11: Comparison of various estimators and 95% CI of N for $N = 1,000$ with different parameters in the Hermite distribution .

Estimators	Average $\hat{N}$	Average $\widehat{se}(\hat{N})$	Coverage probability	Average length	Average $\hat{N}$	Average $\widehat{se}(\hat{N})$	Coverage probability	Average length
	$Herm(2, 1.2)$				$Herm(4, 1.2)$			
MLE	950.04	10.77	0.00	42.21	989.98	4.18	0.34	16.40
Turing	961.47	10.49	0.12	41.12	994.31	4.98	0.73	19.54
Chao	979.39	22.46	0.78	88.02	998.34	7.69	0.86	30.13
Censored	1,222.35	37.47	0.00	146.86	1,068.98	14.91	0.00	58.45
Herm	1,007.43	-	-	-	1,000.15	-	-	-
Profile	1,008.40	-	0.93	176.35	999.36	-	0.94	26.62
	$Herm(2, 1.4)$				$Herm(4, 1.4)$			
MLE	893.26	9.78	0.00	38.33	974.74	4.02	0.01	15.76
Turing	907.99	9.89	0.00	38.78	982.63	5.20	0.16	20.37
Chao	930.04	21.43	0.14	83.99	991.06	8.93	0.75	35.00
Censored	1,146.70	34.99	0.00	137.15	1,067.63	16.01	0.00	62.75
Herm	987.74	-	-	-	1,000.77	-	-	-
Profile	983.08	-	0.97	189.93	1,000.55	-	0.96	52.04
	$Herm(2, 1.6)$				$Herm(4, 1.6)$			
MLE	825.46	8.58	0.00	33.64	953.37	3.80	0.00	14.91
Turing	834.04	8.82	0.00	34.59	962.45	5.13	0.00	20.12
Chao	837.76	15.82	0.00	62.00	970.95	8.84	0.16	34.64
Censored	1,018.50	29.18	0.99	114.37	1,048.05	16.04	0.14	62.88
Herm	992.91	-	-	-	996.58	-	-	-
Profile	972.39	-	0.96	159.42	994.33	-	0.93	55.98
	$Herm(2, 1.8)$				$Herm(4, 1.8)$			
MLE	750.09	7.23	0.00	28.33	921.15	3.47	0.00	13.61
Turing	742.42	6.90	0.00	27.04	926.11	4.40	0.00	17.26
Chao	725.44	7.31	0.00	28.66	924.75	5.44	0.00	21.32
Censored	842.95	19.66	0.00	77.08	989.06	13.46	0.87	52.78
Herm	1,000.03	-	-	-	1,000.29	-	-	-
Profile	998.92	-	0.96	121.51	999.52	-	0.95	46.67

3.6 Conclusion

The basic model for the count data is homogeneous Poisson model, but it rarely occurs in real situations. It is more reasonable to assume that the population may consist of a set of subgroups. Heterogeneous Poisson model might be more realistic. The negative binomial and geometric distributions have been commonly used as a model of capture-recapture data. The failure of a dispersion parameter estimation in negative binomial distributions which results in a spurious estimate for the population size N have been demonstrated in many studies. Therefore, the Poisson-Normal distribution is proposed in this study.

The new population size $\hat{N}_{Herm}$ based on the Poisson-Normal mixture or Hermite distribution is derived using the maximum likelihood estimation. The simulation study is used to consider the performance of the proposed estimator under homogeneous and heterogeneous count data. The simulation results reveal that $\hat{N}_{Herm}$ is an asymptotic estimator under Poisson and Hermite distributions. For Hermite distribution, $\hat{N}_{Herm}$ works very well compared with others. Its performance is close to the $\hat{N}_{MLE}$ and $\hat{N}_{Turing}$ for $N \geqslant 500$ and $\lambda \geqslant 2$ under Poisson distribution. In addition, $\hat{N}_{Herm}$ is an efficient estimator for estimating the population size N based on negative binomial distribution when $N \geqslant 500$.

The profile likelihood is used to construct the confidence intervals for the population size N. The proposed CI provided CPs close to the nominal level at 95% under Poisson distribution. The average lengths of proposed CI were quite shorter than those of others, which gave CPs close to the nominal level in Poisson distribution. Therefore, proposed CI is an appropriate choice for estimating the CI of population size N based on Poisson distribution. Based on Hermite distribution, the proposed CI provided CPs close to the nominal level at 95% while the others gave low CPs. It is clear that the proposed CI is the best choice for estimating the CI of population size N based on Hermite distribution.

CHAPTER 4

DISCRETE MIXTURES OF HERMITE DISTRIBUTIONS

A mixture model is a flexible approach to cope with data from a population which is known or suspected to contain contaminated parts. Mixture model has been widely used in many fields. For example, Böhning et al. (2005) proposed the NPMLE of population size based on the zero-truncated count mixture model. Viwatwongkasem, Kuhnert, and Satitvipawee (2008) proposed the estimator of population size based on the mixtures of zero-truncated Poisson distributions. Böhning and Kuhnert (2006) showed the equivalence of the zero-truncated count mixture distributions and the mixtures of zero-truncated count distributions. Therefore, estimating the population size N of two estimators associated with the two models provides equal values. Lerdsuwansri and Böhning (2017) proposed discrete mixtures of bivariate, conditional independent Poisson model to estimate an unknown population size in two-source situation. The results for estimating the parameters of zero-truncated count mixtures of Hermite distributions implying a unique estimator for the population size N are proposed in this study.

4.1 Maximum likelihood estimation of discrete mixtures of Hermite distributions

Let Y be the number of times that a unit was identified over the study period. Assume Y is mixture of the Hermite distribution that consists of s components with component specific mean ν_k and weight q_k. Furthermore, all components have the same dispersion d. The marginal distribution of Y is

$$f(y;Q) = \sum_{k=1}^{s} q_k p(y|\nu_k, d). \tag{4.1}$$

Here, $p(y|\nu_k, d) = e^{\nu_k(\frac{d-3}{2})}[\nu_k(2-d)]^y q_{yk}(\theta_k)$, $q_{yk}(\theta_k) = \sum_{j=0}^{[y/2]} \frac{\theta_k^j}{(y-2j)!j!}$, and $\theta_k = \frac{d-1}{2\nu_k(2-d)^2}$. The finite mixture distribution arises as the marginal distribution with respect to the unobserved variables Z with distribution Q. A discrete mixing distribution $Q = \begin{pmatrix} \nu_1 & \nu_2 & \cdots & \nu_s \\ d & d & \cdots & d \\ q_1 & q_2 & \cdots & q_s \end{pmatrix}$ gives weight q_k to parameters ν_k and d for $k = 1, 2, ..., s$, where s is the number of unobserved components. Note that $q_k \geq 0$ and $\sum_{k=1}^{s} q_k = 1$. Assume $Y_1, Y_2, ..., Y_n$ are observed and drawn from mixture density. The incomplete data likelihood is

$$L(Q) = \prod_{i=1}^{m} \left(\frac{\sum_{k=1}^{s} q_k p(i|\nu_k, d)}{1 - \sum_{k=1}^{s} q_k p(0|\nu_k, d)} \right)^{f_i}.$$

The incomplete data log-likelihood is

$$l(Q) = \sum_{i=1}^{m} f_i \log \left(\frac{\sum_{k=1}^{s} q_k p(i|\nu_k, d)}{1 - \sum_{k=1}^{s} q_k p(0|\nu_k, d)} \right). \tag{4.2}$$

An estimate of Q be achieved by maximizing (4.2), giving the NPMLE. The EM algorithm has become common for maximum likelihood estimation in mixture models. To carry on the EM algorithm, the complete data log-likelihood is required.

At the E-step, the unobserved frequency f_0 is replaced by its expected value given observed frequencies and current values of Q. Let the expected value of f_0 denoted by $\hat{f}_0$ be written as

$$\begin{aligned} \hat{f}_0 &= E(f_0|\text{observed data}; Q) \\ &= \frac{nf(0;Q)}{1 - f(0;Q)} \\ &= \frac{n \sum_{k=1}^{s} q_k p(0|\nu_k, d)}{1 - \sum_{k=1}^{s} q_k p(0|\nu_k, d)}. \end{aligned} \tag{4.3}$$

The log-likelihood for the complete data is given by

$$l_{cd}(Q) = \hat{f}_0 \log \sum_{k=1}^{s} q_k p(0|\nu_k, d) + \sum_{i=1}^{m} f_i \log \sum_{k=1}^{s} q_k p(i|\nu_k, d). \tag{4.4}$$

To manipulate the MLE of Q, the log-likelihood is maximized by applying the EM algorithm as well. In this case, a variable indicating component to which the count i belongs is ignored. Let z_{ik} be indicator variables defined as:

$$z_{ik} = \begin{cases} 1 & \text{if count } i \text{ arose from component } k \\ 0 & \text{otherwise.} \end{cases}$$

If z_{ik} are observable, the complete data log-likelihood is

$$l_{cd}(Q) = \sum_{i=0}^{m} f_i \sum_{k=1}^{s} z_{ik} \log q_k + \sum_{i=0}^{m} f_i \sum_{k=1}^{s} z_{ik} \log p(i|\nu_k, d). \tag{4.5}$$

At the E-step, the unobserved indicator z_{ik} is replaced by e_{ik}, its expected value conditioning on the observed data and current values of Q

$$e_{ik} = E(z_{ik}|\text{observed data}; Q) = \frac{q_k p(i|\nu_k, d)}{\sum_{k=1}^{s} q_k p(i|\nu_k, d)}. \tag{4.6}$$

Substituting e_{ik} into (4.5), the expected of the complete data log-likelihood is given by

$$E[l_{cd}(Q)] = \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log q_k + \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k, d). \tag{4.7}$$

M-step: in mathematical optimization, Lagrange's method is a strategy for finding the local maximum of a function subject to equality constraints. To maximize (4.7) subject to the constraint $\sum_{k=1}^{s} q_k = 1$, the Lagrange function is

$$L(Q, \gamma) = \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log q_k + \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k, d) + \gamma(1 - \sum_{k=1}^{s} q_k). \tag{4.8}$$

Differentiating (4.8) with respect to γ, q_k, ν_k, and d, and setting the result to 0:

$$\begin{aligned} \frac{\partial}{\partial \gamma} L(Q, \gamma) &= 0 \\ 1 - \sum_{k=1}^{s} q_k &= 0 \\ \sum_{k=1}^{s} q_k &= 1. \end{aligned} \tag{4.9}$$

$$\begin{aligned} \frac{\partial}{\partial q_k} L(Q, \gamma) &= 0 \\ \frac{\sum_{i=0}^{m} f_i e_{ik}}{q_k} - \gamma &= 0 \\ q_k &= \frac{\sum_{i=0}^{m} f_i e_{ik}}{\gamma} \\ \sum_{k=1}^{s} q_k &= \frac{\sum_{i=0}^{m} \sum_{k=1}^{s} f_i e_{ik}}{\gamma}. \end{aligned} \tag{4.10}$$

Solving equation (4.9) and (4.10) provides

$$\gamma = \sum_{i=0}^{m} \sum_{k=1}^{s} f_i e_{ik}.$$

Therefore,

$$\begin{aligned}\hat{q}_k &= \frac{\sum_{i=0}^{m} f_i e_{ik}}{\sum_{i=0}^{m}\sum_{k=1}^{s} f_i e_{ik}}\\ &= \frac{\sum_{i=0}^{m} f_i e_{ik}}{\hat{N}}.\end{aligned} \tag{4.11}$$

$$\begin{aligned}\frac{\partial}{\partial \nu_k}L(Q,\gamma) &= 0\\ \frac{\partial}{\partial \nu_k}\Bigg(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik}\log p(i|\nu_k,d)\Bigg) &= 0\\ \frac{\partial}{\partial \nu_k}\Bigg(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik}\log\Big[e^{\nu_k(\frac{d-3}{2})}[\nu_k(2-d)]^i q_{ik}(\theta_k)\Big]\Bigg) &= 0\\ \frac{\partial}{\partial \nu_k}\Bigg(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik}\Big[\nu_k(\frac{d-3}{2}) + i\log[\nu_k(2-d)] + \log q_{ik}(\theta_k)\Big]\Bigg) &= 0\\ \frac{d-3}{2}\sum_{i=0}^{m} f_i e_{ik} + \frac{1}{\nu_k}\sum_{i=0}^{m} i f_i e_{ik} + \frac{\partial}{\partial \nu_k}\Bigg[\sum_{i=1}^{m} f_i \sum_{k=1}^{s} e_{ik}\log q_{ik}(\theta_k)\Bigg] &= 0.\end{aligned} \tag{4.12}$$

Consider

$$\begin{aligned}\frac{\partial}{\partial \nu_k}\Bigg[\sum_{i=1}^{m} f_i \sum_{k=1}^{s} e_{ik}\log q_{ik}(\theta_k)\Bigg] &= \sum_{i=1}^{m} f_i \frac{\partial}{\partial \nu_k}\Bigg[\sum_{k=1}^{s} e_{ik}\log q_{ik}(\theta_k)\Bigg]\\ &= \sum_{i=1}^{m} f_i e_{ik}\frac{\partial}{\partial \nu_k}\Big[\log q_{ik}(\theta_k)\Big]\\ &= \sum_{i=1}^{m} f_i e_{ik}\Bigg[\frac{1}{q_{ik}(\theta_k)}\cdot\frac{\partial}{\partial \theta_k}q_{ik}(\theta_k)\cdot\frac{\partial \theta_k}{\partial \nu_k}\Bigg]\\ &= \sum_{i=1}^{m} f_i e_{ik}\Bigg[\frac{1}{q_{ik}(\theta_k)}\cdot q'_{ik}(\theta_k)\cdot\frac{\partial}{\partial \nu_k}\Bigg(\frac{d-1}{2\nu_k(2-d)^2}\Bigg)\Bigg]\\ &= \sum_{i=1}^{m} f_i e_{ik}\Bigg[\frac{1}{q_{ik}(\theta_k)}\cdot q'_{ik}(\theta_k)\cdot\Bigg(-\frac{d-1}{2\nu_k^2(2-d)^2}\Bigg)\Bigg]\\ &= -\frac{d-1}{2(2-d)^2{\nu_k}^2}\sum_{i=0}^{m}\frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)},\end{aligned} \tag{4.13}$$

where $q'_{ik}(\theta_k) = \frac{\partial}{\partial \theta_k}q_{ik}(\theta_k)$. Substituting (4.13) in (4.12),

$$
\begin{aligned}
\frac{\partial}{\partial \nu_k} L(Q,\gamma) &= 0 \\
\frac{\partial}{\partial \nu_k}\Bigg(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k, d) \Bigg) &= 0 \\
\frac{d-3}{2} \sum_{i=0}^{m} f_i e_{ik} + \frac{1}{\nu_k} \sum_{i=0}^{m} i f_i e_{ik} - \frac{d-1}{2(2-d)^2 {\nu_k}^2} \sum_{i=0}^{m} \frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} &= 0.
\end{aligned}
\tag{4.14}
$$

$$
\begin{aligned}
\frac{\partial}{\partial d} L(Q,\gamma) &= 0 \\
\frac{\partial}{\partial d}\Bigg(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k, d) \Bigg) &= 0 \\
\frac{\partial}{\partial d}\Bigg(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log \Big[e^{\nu_k(\frac{d-3}{2})} [\nu_k(2-d)]^i q_{ik}(\theta_k) \Big] \Bigg) &= 0 \\
\frac{\partial}{\partial d}\Bigg(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \Big[\nu_k (\frac{d-3}{2}) + i \log[\nu_k(2-d)] + \log q_{ik}(\theta_k) \Big] \Bigg) &= 0 \\
\frac{1}{2} \sum_{i=0}^{m} f_i e_{ik} \nu_k - \frac{1}{2-d} \sum_{i=0}^{m} i f_i e_{ik} + \frac{\partial}{\partial d} \Big[\sum_{i=1}^{m} f_i \sum_{k=1}^{s} e_{ik} \log q_{ik}(\theta_k) \Big] &= 0.
\end{aligned}
\tag{4.15}
$$

Consider

$$
\begin{aligned}
\frac{\partial}{\partial d} \Big[\sum_{i=1}^{m} f_i \sum_{k=1}^{s} e_{ik} \log q_{ik}(\theta_k) \Big] &= \sum_{i=1}^{m} f_i \frac{\partial}{\partial d} \Big[\sum_{k=1}^{s} e_{ik} \log q_{ik}(\theta_k) \Big] \\
&= \sum_{i=1}^{m} f_i e_{ik} \frac{\partial}{\partial d} \Big[\log q_{ik}(\theta_k) \Big] \\
&= \sum_{i=1}^{m} f_i e_{ik} \Big[\frac{1}{q_{ik}(\theta_k)} \cdot \frac{\partial}{\partial \theta_k} q_{ik}(\theta_k) \cdot \frac{\partial \theta_k}{\partial d} \Big] \\
&= \sum_{i=1}^{m} f_i e_{ik} \Big[\frac{1}{q_{ik}(\theta_k)} \cdot q'_{ik}(\theta_k) \cdot \frac{\partial}{\partial d} \Big(\frac{d-1}{2\nu_k(2-d)^2} \Big) \Big] \\
&= \sum_{i=1}^{m} f_i e_{ik} \Big[\frac{q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} \cdot \Big(\frac{1}{2\nu_k(2-d)^2} + \frac{d-1}{\nu_k(2-d)^3} \Big) \Big] \\
&= \frac{d}{2\nu_k(2-d)^3} \sum_{i=0}^{m} \frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)},
\end{aligned}
\tag{4.16}
$$

where $q'_{ik}(\theta_k) = \frac{\partial}{\partial\theta_k} q_{ik}(\theta_k)$. Substituting (4.16) in (4.15),

$$\begin{aligned} \frac{\partial}{\partial d} L(Q,\gamma) &= 0 \\ \frac{\partial}{\partial d}\left(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k, d)\right) &= 0 \\ \frac{\nu_k}{2}\sum_{i=0}^{m} f_i e_{ik} - \frac{1}{2-d}\sum_{i=0}^{m} i f_i e_{ik} + \frac{d}{2\nu_k(2-d)^3}\sum_{i=0}^{m} \frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} &= 0. \end{aligned} \tag{4.17}$$

Rewrite (4.17) as

$$\begin{aligned} \sum_{i=0}^{m} \frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} &= \frac{2(2-d)^3\nu_k}{d}\left(\frac{1}{2-d}\sum_{i=0}^{m} i f_i e_{ik} - \frac{\nu_k}{2}\sum_{i=0}^{m} f_i e_{ik}\right) \\ &= \frac{2(2-d)^2\nu_k}{d}\sum_{i=0}^{m} i f_i e_{ik} - \frac{(2-d)^3\nu_k^2}{d}\sum_{i=0}^{m} f_i e_{ik}. \end{aligned} \tag{4.18}$$

Substituting (4.18) in (4.14),

$$\begin{aligned} \left\{\begin{array}{c} \frac{d-3}{2}\sum_{i=0}^{m} f_i e_{ik} + \frac{1}{\nu_k}\sum_{i=0}^{m} i f_i e_{ik} \\ -\frac{d-1}{2(2-d)^2{\nu_k}^2}\left[\frac{2(2-d)^2\nu_k}{d}\sum_{i=0}^{m} i f_i e_{ik} - \frac{(2-d)^3\nu_k^2}{d}\sum_{i=0}^{m} f_i e_{ik}\right] \end{array}\right\} &= 0 \\ \left(\frac{d-3}{2} + \frac{(d-1)(2-d)}{2d}\right)\sum_{i=0}^{m} f_i e_{ik} + \left(\frac{1}{\nu_k} - \frac{d-1}{\nu_k d}\right)\sum_{i=0}^{m} i f_i e_{ik} &= 0 \\ -\frac{1}{d}\sum_{i=0}^{m} f_i e_{ik} + \frac{1}{\nu_k d}\sum_{i=0}^{m} i f_i e_{ik} &= 0 \\ \frac{1}{\nu_k d}\sum_{i=0}^{m} i f_i e_{ik} &= \frac{1}{d}\sum_{i=0}^{m} f_i e_{ik} \\ \hat{\nu}_k &= \frac{\sum_{i=0}^{m} i f_i e_{ik}}{\sum_{i=0}^{m} f_i e_{ik}}. \end{aligned} \tag{4.19}$$

Substituting (4.19) in (4.18),

$$\begin{aligned} \sum_{i=0}^{m} \frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} &= \frac{2(2-d)^2\hat{\nu}_k^2}{d}\sum_{i=0}^{m} f_i e_{ik} - \frac{(2-d)^3\hat{\nu}_k^2}{d}\sum_{i=0}^{m} f_i e_{ik} \\ &= [(2-d)\hat{\nu}_k]^2\sum_{i=0}^{m} f_i e_{ik}. \end{aligned} \tag{4.20}$$

Parameter d be found by solving (4.20). The package $hermite$ in program R allows to estimate the parameter d given an univariate sample by means of the function $glm.hermite$, see (Moriña et al., 2015) for more details.

The population size estimator based on discrete mixtures of Hermite distributions through the Horvitz-Thomson approach is

$$\hat{N}_{CMLE} = \frac{n}{1 - \sum_{k=1}^{s} \hat{q}_k e^{\hat{\nu}_k(\frac{\hat{d}-3}{2})}}. \tag{4.21}$$

The EM algorithm used to compute the estimate of population size is given as follows.

4.1.1 EM algorithm

Step 0: Choose initial value $\hat{Q}^{(0)} = \begin{pmatrix} \hat{\nu}_1^{(0)} & \hat{\nu}_2^{(0)} & \cdots & \hat{\nu}_s^{(0)} \\ \hat{d}^{(0)} & \hat{d}^{(0)} & \cdots & \hat{d}^{(0)} \\ \hat{q}_1^{(0)} & \hat{q}_2^{(0)} & \cdots & \hat{q}_s^{(0)} \end{pmatrix}$, set $t = 0$.

Step 1: Compute

$$\begin{aligned} \hat{f}_0^{(t+1)} &= \frac{nf(0;\hat{Q}^{(t)})}{1 - f(0;\hat{Q}^{(t)})} \\ e_{ik}^{(t)} &= \frac{q_k^{(t)} p(i|\nu_k^{(t)}, d^{(t)})}{\sum_{k=1}^{s} q_k^{(t)} p(i|\nu_k^{(t)}, d^{(t)})}, \qquad \text{for} \quad k = 1, 2, ..., s. \end{aligned}$$

Step 2: Use complete data $\hat{f}_0^{(t+1)}, f_1, f_2, ..., f_m$ to compute the new MLEs of

$$\begin{aligned} \hat{q}_k^{(t+1)} &= \frac{\sum_{i=0}^{m} f_i e_{ik}^{(t)}}{n + \hat{f}_0^{(t+1)}}, \qquad \text{for} \quad k = 1, 2, ..., s \\ \hat{\nu}_k^{(t+1)} &= \frac{\sum_{i=0}^{m} i f_i e_{ik}^{(t)}}{\sum_{i=0}^{m} f_i e_{ik}^{(t)}} \\ \hat{d}_k^{(t+1)} &= glm.hermite(\hat{f}_0^{(t+1)}, f_1, ..., f_m \sim 1, link = "log", start = NULL, m^{\star} = 2). \end{aligned}$$

Step 3: Set $t = t + 1$ and repeat Step 1. Steps 1 and 2 are repeated until:

$$\left|\hat{q}^{(t+1)} - \hat{q}^{(t)}\right| \leq 10^{-4}, \quad \left|\hat{\nu}^{(t+1)} - \hat{\nu}^{(t)}\right| \leq 10^{-4}, \text{ and } \quad \left|\hat{d}^{(t+1)} - \hat{d}^{(t)}\right| \leq 10^{-4}.$$

4.2 Mixture model and the boundary problem

The maximum likelihood estimation discussed above faces two major problems, a boundary problem and a lack of identification. The boundary problem deals with the circumstance in which the mixing distribution equates component parameters nearly 0^+ with positive weight (J.-P. Wang & Lindsay, 2008). Kuhnert et al. (2008) demonstrated that mixtures also suffer under the boundary problem, results in overestimation of population size N. The lack of identifiability problem also affects the inference of population size N (Link, 2003). Different models providing different estimates of N might have identical distributions.

To illustrate the boundary problem of maximum likelihood estimation discussed above, the example data were generated from two-components Hermite mixture $0.5Herm(1, 1.6) + 0.5Herm(4, 1.6)$ with $N = 500$. Counts are shown in Table 4.1 which observed data $n = 369$. The MLEs of Q for each s are computed from $s = 1, 2, 3, ...$ until the log-likelihood stops increasing. Details of the likelihood analysis for $s = 1$ to $s = 3$ components are presented in Table 4.2. It can be seen that the differences in the log-likelihood for models with $s = 2$ and $s = 3$ components are minor, which is clearly evident when the Bayesian Information Criterion (BIC) is considered:

$$BIC = -2 \log L(\hat{Q}_s) + k \log n, \tag{4.22}$$

where $\log L(\hat{Q}_s)$ is the maximum log-likelihood of the model with s components and k is the number of parameters estimated by the model. For $s = 1$, ν and d are parameter estimates. If $s = 2$, the number of parameters estimated by the model is 4 including q_1, ν_1, ν_2, and d. For $s = 3$, the estimated parameters consist of $q_1, q_2, \nu_1, \nu_2, \nu_3$, and d. Therefore, the number of parameters estimated by the model with s components is $2s$. Models are selected on the basis of small BIC-values. From the result show in Table 4.2, the appropriate model is the two-component model. The mixing distribution equates component parameters ν_k nearly 0^+ with positive weight. The boundary problem occurs, results in overestimation of population size N. Consequently, we suggest to improve the estimation by using penalized maximum likelihood.

Table 4.1: Examples data from $0.5Herm(1, 1.6) + 0.5Herm(4, 1.6)$ with $N = 500$.

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
f_i	131	76	92	60	52	26	25	18	13	1	3	2	0	0	1

Table 4.2: Mixture likelihood analysis for the example data in Table 4.1.

s	$\hat{q}_k$	$\hat{\nu}_k$	$\hat{d}$	$\log L(\hat{Q}_k)$	BIC	$\hat{f}_0$	$\hat{N}_{CMLE}$
1	1.0000	3.1151	1.5720	-743.1164	1,497.9657	43	412
2	0.5048	0.0006	1.8198	-734.8085	1,493.0829	558	927
	0.4952	2.6376					
3	0.3119	0.0000	1.8286	-736.8547	1,508.9082	808	1,177
	0.2972	0.0007					
	0.3909	2.6231					

4.2.1 Penalized maximum likelihood estimation of discrete mixtures of Hermite distributions

Let Y be the number of times that a unit was identified over the study period. Assume Y is mixture of the Hermite distribution that consists of s components with component specific mean ν_k and weight q_k. Furthermore, all components have the same dispersion d. The marginal distribution of Y is

$$f(y;Q) = \sum_{k=1}^{s} q_k p(y|\nu_k, d). \tag{4.23}$$

Here, $p(y|\nu_k, d) = e^{\nu_k(\frac{d-3}{2})}[\nu_k(2-d)]^y q_{yk}(\theta_k)$, $q_{yk}(\theta_k) = \sum_{j=0}^{[y/2]} \frac{\theta_k^j}{(y-2j)!j!}$, and $\theta_k = \frac{d-1}{2\nu_k(2-d)^2}$ with respect to the unobserved variables Z having distribution Q. A discrete mixing distribution $Q = \begin{pmatrix} \nu_1 & \nu_2 & \cdots & \nu_s \\ d & d & \cdots & d \\ q_1 & q_2 & \cdots & q_s \end{pmatrix}$ gives weight q_k to parameters ν_k and d for $k = 1, 2, ..., s$, where s is the number of unobserved components. Note that $q_k \geq 0$ and $\sum_{k=1}^{s} q_k = 1$. Assume $Y_1, Y_2, ..., Y_n$ are observed and drawn from mixture density. The incomplete data likelihood is

$$L(Q) = \prod_{i=1}^{m} \left(\frac{\sum_{k=1}^{s} q_k p(i|\nu_k, d)}{1 - \sum_{k=1}^{s} q_k p(0|\nu_k, d)} \right)^{f_i}.$$

The incomplete data log-likelihood is

$$l(Q) = \sum_{i=1}^{m} f_i \log \left(\frac{\sum_{k=1}^{s} q_k p(i|\nu_k, d)}{1 - \sum_{k=1}^{s} q_k p(0|\nu_k, d)} \right). \tag{4.24}$$

An estimate of Q can be achieved by maximizing (4.24) giving the NPMLE. The EM algorithm is used for maximum likelihood estimation in mixture models. To carry on the EM algorithm, the complete data log-likelihood is required.

At the E-step, the unobserved frequency f_0 is replaced by its expected value given observed frequencies and current values of Q. Let the expected value of f_0 denoted by $\hat{f}_0$ be written as

$$\begin{aligned} \hat{f}_0 &= E(f_0|\text{observed data}; Q) \\ &= \frac{nf(0;Q)}{1-f(0;Q)} \\ &= \frac{n\sum_{k=1}^{s} q_k p(0|\nu_k, d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k, d)}. \end{aligned} \tag{4.25}$$

The log-likelihood for the complete data is given by

$$l_{cd}(Q) = \hat{f}_0 \log \sum_{k=1}^{s} q_k p(0|\nu_k, d) + \sum_{i=1}^{m} f_i \log \sum_{k=1}^{s} q_k p(i|\nu_k, d). \tag{4.26}$$

To manipulate the MLE of Q, the log-likelihood is maximized by applying the EM algorithm as well. In this case, a variable indicating component to which the count i belongs is ignored. Let z_{ik} be indicator variables defined as following

$$z_{ik} = \begin{cases} 1 & \text{if count } i \text{ arose from component } k \\ 0 & \text{otherwise.} \end{cases}$$

If z_{ik} are observable, the complete data log-likelihood is

$$l_{cd}(Q) = \sum_{i=0}^{m} f_i \sum_{k=1}^{s} z_{ik} \log q_k + \sum_{i=0}^{m} f_i \sum_{k=1}^{s} z_{ik} \log p(i|\nu_k, d). \tag{4.27}$$

At the E-step, the unobserved indicators z_{ik} are replaced by e_{ik}, the expected values conditioning on the observed data and current values of Q

$$e_{ik} = E(z_{ik}|\text{observed data}; Q) = \frac{q_k p(i|\nu_k, d)}{\sum_{k=1}^{s} q_k p(i|\nu_k, d)}. \tag{4.28}$$

The expected log-likelihood is given by

$$E[l_{cd}(Q)] = \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log q_k + \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k, d). \tag{4.29}$$

Boundary problem can be found in mixture models. As a result, the plug-in estimator of the parameters involving the probability at the boundary points are usually biased.

Wang and Lindsay (J.-P. Wang & Lindsay, 2008) used a partial prior for Q and simplified into a penalized likelihood to improve the estimation. The expected log-likelihood under the exponential partial prior is written as

$$E_p[l_{cd}(Q)] = \left\{ \begin{array}{c} \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log q_k + \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k, d) \\ -\log\left(\frac{\sum_{k=1}^{s} q_k p(0|\nu_k, d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k, d)} \right) \end{array} \right\}. \quad (4.30)$$

M-step: In mathemathical optimization, the method of Lagrange multipliers is a strategy for finding the maximum or minimum of a function subject to equality constraints. To maximize (4.30) subject to the constraint $\sum_{k=1}^{s} q_k = 1$, the Lagrange function is defined by

$$L_p(Q, \gamma) = \left\{ \begin{array}{c} \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log q_k + \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k, d) \\ -\log\left(\frac{\sum_{k=1}^{s} q_k p(0|\nu_k, d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k, d)} \right) + \gamma(1 - \sum_{k=1}^{s} q_k) \end{array} \right\}. \quad (4.31)$$

Differentiating (4.31) with respect to γ , q_k, ν_k, and d, and setting the result to 0:

$$\begin{aligned} \frac{\partial}{\partial \gamma} L_p(Q, \gamma) &= 0 \\ 1 - \sum_{k=1}^{s} q_k &= 0 \\ \sum_{k=1}^{s} q_k &= 1. \end{aligned} \quad (4.32)$$

$$
\begin{aligned}
\frac{\partial}{\partial q_k} L_p(Q,\gamma) &= 0 \\
\frac{\partial}{\partial q_k}\left\{\begin{array}{c} \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik}\log q_k \\ -\log\left(\frac{\sum_{k=1}^{s} q_k p(0|\nu_k,d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k,d)}\right) + \gamma(1-\sum_{k=1}^{s} q_k) \end{array}\right\} &= 0 \\
\frac{\partial}{\partial q_k}\left\{\begin{array}{c} \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik}\log q_k - \log\sum_{k=1}^{s} q_k p(0|\nu_k,d) \\ +\log(1-\sum_{k=1}^{s} q_k p(0|\nu_k,d)) + \gamma(1-\sum_{k=1}^{s} q_k) \end{array}\right\} &= 0 \\
\frac{\sum_{i=0}^{m} f_i e_{ik}}{q_k} - \frac{p(0|\nu_k,d)}{\sum_{k=1}^{s} q_k p(0|\nu_k,d)} - \frac{p(0|\nu_k,d)}{(1-\sum_{k=1}^{s} q_k p(0|\nu_k,d))} - \gamma &= 0 \\
\frac{\sum_{i=0}^{m} f_i e_{ik}}{q_k} - \frac{p(0|\nu_k,d)}{\sum_{k=1}^{s} q_k p(0|\nu_k,d)(1-\sum_{k=1}^{s} q_k p(0|\nu_k,d))} - \gamma &= 0 \\
\frac{\sum_{i=0}^{m} f_i e_{ik}}{q_k} - \frac{p(0|\nu_k,d)}{\sum_{k=1}^{s} q_k p(0|\nu_k,d)(1-\sum_{k=1}^{s} q_k p(0|\nu_k,d))} &= \gamma \\
\sum_{i=0}^{m} f_i e_{ik} - \frac{q_k p(0|\nu_k,d)}{\sum_{k=1}^{s} q_k p(0|\nu_k,d)(1-\sum_{k=1}^{s} q_k p(0|\nu_k,d))} &= q_k\gamma.
\end{aligned}
\tag{4.33}
$$

$$
\therefore \hat{q}_k = \frac{\sum_{i=0}^{m} f_i e_{ik} - \frac{q_k p(0|\nu_k,d)}{\sum_{k=1}^{s} q_k p(0|\nu_k,d)(1-\sum_{k=1}^{s} q_k p(0|\nu_k,d))}}{\gamma}. \tag{4.34}
$$

From (4.33), it follows that

$$
\sum_{k=1}^{s}\sum_{i=0}^{m} f_i e_{ik} - \frac{\sum_{k=1}^{s} q_k p(0|\nu_k,d)}{\sum_{k=1}^{s} q_k p(0|\nu_k,d)(1-\sum_{k=1}^{s} q_k p(0|\nu_k,d))} = \gamma\sum_{k=1}^{s} q_k.
$$

Since $\sum_{k=1}^{s} q_k = 1$,

$$
\begin{aligned}
\gamma &= \sum_{k=1}^{s}\sum_{i=0}^{m} f_i e_{ik} - \frac{\sum_{k=1}^{s} q_k p(0|\nu_k,d)}{\sum_{k=1}^{s} q_k p(0|\nu_k,d)(1-\sum_{k=1}^{s} q_k p(0|\nu_k,d))} \\
&= \sum_{k=1}^{s}\sum_{i=0}^{m} f_i e_{ik} - \frac{1}{1-\sum_{k=1}^{s} q_k p(0|\nu_k,d)}.
\end{aligned}
\tag{4.35}
$$

Substituting (4.35) in (4.34) provides

$$
\begin{aligned}
\hat{q}_k &= \frac{\sum_{i=0}^{m} f_i e_{ik} - \frac{q_k p(0|\nu_k,d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k,d)}}{\sum_{k=1}^{s}\sum_{i=0}^{m} f_i e_{ik} - \frac{1}{1-\sum_{k=1}^{s} q_k p(0|\nu_k,d)}} \\
&= \frac{\sum_{i=0}^{m} f_i e_{ik} - P_1}{\sum_{i=0}^{m}\sum_{k=1}^{s} f_i e_{ik} - P_2}.
\end{aligned}
\tag{4.36}
$$

Here, $P_1 = \frac{q_k p(0|\nu_k,d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k,d)}$ and $P_2 = \frac{1}{1-\sum_{k=1}^{s} q_k p(0|\nu_k,d)}$.

$$\begin{aligned}
\frac{\partial}{\partial \nu_k} L_p(Q,\gamma) &= 0 \\
\frac{\partial}{\partial \nu_k}\Bigg(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k,d) - \log\left[\frac{\sum_{k=1}^{s} q_k p(0|\nu_k,d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k,d)}\right]\Bigg) &= 0 \\
\frac{\partial}{\partial \nu_k}\Bigg(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k,d)\Bigg) - \frac{\partial}{\partial \nu_k}\log\left[\frac{\sum_{k=1}^{s} q_k p(0|\nu_k,d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k,d)}\right] &= 0.
\end{aligned} \tag{4.37}$$

From (4.14),

$$\frac{\partial}{\partial \nu_k}\Bigg(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k,d)\Bigg) = \left\{\begin{array}{c} \frac{d-3}{2}\sum_{i=0}^{m} f_i e_{ik} + \frac{1}{\nu_k}\sum_{i=0}^{m} i f_i e_{ik} \\ -\frac{d-1}{2(2-d)^2{\nu_k}^2}\sum_{i=0}^{m}\frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} \end{array}\right\}. \tag{4.38}$$

Consider

$$\frac{\partial}{\partial \nu_k}\log\left[\frac{\sum_{k=1}^{s} q_k p(0|\nu_k,d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k,d)}\right] = \left\{\begin{array}{c} \frac{\partial}{\partial \nu_k}\log \sum_{k=1}^{s} q_k p(0|\nu_k,d) \\ -\frac{\partial}{\partial \nu_k}\log\big(1-\sum_{k=1}^{s} q_k p(0|\nu_k,d)\big) \end{array}\right\}. \tag{4.39}$$

$$\begin{aligned}
\frac{\partial}{\partial \nu_k}\log \sum_{k=1}^{s} q_k p(0|\nu_k,d) &= \frac{\partial}{\partial \nu_k}\log \sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)} \\
&= \frac{1}{\sum_{k=1}^{s} q_k e^{\nu_k(\frac{d-3}{2})}}\cdot q_k e^{\nu_k(\frac{d-3}{2})}\cdot\left(\frac{d-3}{2}\right) \\
&= \left(\frac{d-3}{2}\right)\frac{q_k e^{\nu_k(\frac{d-3}{2})}}{\sum_{k=1}^{s} q_k e^{\nu_k(\frac{d-3}{2})}}.
\end{aligned} \tag{4.40}$$

$$\begin{aligned}
\frac{\partial}{\partial \nu_k}\log\big(1-\sum_{k=1}^{s} q_k p(0|\nu_k,d)\big) &= \frac{\partial}{\partial \nu_k}\log\big(1-\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}\big) \\
&= \frac{1}{1-\sum_{k=1}^{s} q_k e^{\nu_k(\frac{d-3}{2})}}\cdot\big(-q_k e^{\nu_k(\frac{d-3}{2})}\big)\cdot\left(\frac{d-3}{2}\right) \\
&= -\left(\frac{d-3}{2}\right)\frac{q_k e^{\nu_k(\frac{d-3}{2})}}{1-\sum_{k=1}^{s} q_k e^{\nu_k(\frac{d-3}{2})}}.
\end{aligned} \tag{4.41}$$

Substituting (4.40) and (4.41) in (4.39),

$$\begin{aligned}\frac{\partial}{\partial \nu_k}\log\left[\frac{\sum_{k=1}^{s} q_k p(0|\nu_k, d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k, d)}\right] &= \left\{\begin{array}{c}\left(\frac{d-3}{2}\right)\frac{q_k e^{\nu_k(\frac{d-3}{2})}}{\sum_{k=1}^{s} q_k e^{\nu_k(\frac{d-3}{2})}} \\ +\left(\frac{d-3}{2}\right)\frac{q_k e^{\nu_k(\frac{d-3}{2})}}{1-\sum_{k=1}^{s} q_k e^{\nu_k(\frac{d-3}{2})}}\end{array}\right\} \\ &= \left(\frac{d-3}{2}\right)\frac{q_k e^{\nu_k(\frac{d-3}{2})}}{\sum_{k=1}^{s} q_k e^{\nu_k(\frac{d-3}{2})}\left(1-\sum_{k=1}^{s} q_k e^{\nu_k(\frac{d-3}{2})}\right)}.\end{aligned} \tag{4.42}$$

Substituting (4.38) and (4.42) in (4.37),

$$\begin{aligned}\frac{\partial}{\partial \nu_k}L_p(Q,\gamma) &= 0 \\ \left\{\begin{array}{c}\left(\frac{d-3}{2}\right)\sum_{i=0}^{m} f_i e_{ik} + \frac{1}{\nu_k}\sum_{i=0}^{m} i f_i e_{ik} - \frac{d-1}{2\nu_k^2(2-d)^2}\sum_{i=0}^{m}\frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} \\ -\left(\frac{d-3}{2}\right)\frac{q_k e^{\nu_k(\frac{d-3}{2})}}{\sum_{k=1}^{s} q_k e^{\nu_k(\frac{d-3}{2})}\left(1-\sum_{k=1}^{s} q_k e^{\nu_k(\frac{d-3}{2})}\right)}\end{array}\right\} &= 0.\end{aligned} \tag{4.43}$$

$$\begin{aligned}\frac{\partial}{\partial d}L_p(Q,\gamma) &= 0 \\ \frac{\partial}{\partial d}\left(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik}\log p(i|\nu_k, d) - \log\left[\frac{\sum_{k=1}^{s} q_k p(0|\nu_k, d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k, d)}\right]\right) &= 0 \\ \frac{\partial}{\partial d}\left(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik}\log p(i|\nu_k, d)\right) - \frac{\partial}{\partial d}\log\left[\frac{\sum_{k=1}^{s} q_k p(0|\nu_k, d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k, d)}\right] &= 0.\end{aligned} \tag{4.44}$$

From (4.17),

$$\frac{\partial}{\partial d}\left(\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik}\log p(i|\nu_k, d)\right) = \left\{\begin{array}{c}\frac{\nu_k}{2}\sum_{i=0}^{m} f_i e_{ik} - \frac{1}{2-d}\sum_{i=0}^{m} i f_i e_{ik} \\ +\frac{d}{2\nu_k(2-d)^3}\sum_{i=0}^{m}\frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)}\end{array}\right\}. \tag{4.45}$$

Consider

$$\frac{\partial}{\partial d}\log\left[\frac{\sum_{k=1}^{s} q_k p(0|\nu_k, d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k, d)}\right] = \left\{\begin{array}{c}\frac{\partial}{\partial d}\log\sum_{k=1}^{s} q_k p(0|\nu_k, d) \\ -\frac{\partial}{\partial d}\log\left(1-\sum_{k=1}^{s} q_k p(0|\nu_k, d)\right)\end{array}\right\}. \tag{4.46}$$

$$\begin{aligned}
\frac{\partial}{\partial d}\log\sum_{k=1}^{s} q_k p(0|\nu_k, d) &= \frac{\partial}{\partial d}\log\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)} \\
&= \frac{1}{\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}\cdot q_k e^{\nu_k\left(\frac{d-3}{2}\right)}\cdot\left(\frac{\nu_k}{2}\right) \\
&= \left(\frac{\nu_k}{2}\right)\frac{q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}{\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}. \qquad (4.47)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial d}\log\big(1-\sum_{k=1}^{s} q_k p(0|\nu_k, d)\big) &= \frac{\partial}{\partial d}\log\big(1-\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}\big) \\
&= \frac{1}{1-\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}\cdot\big(-q_k e^{\nu_k\left(\frac{d-3}{2}\right)}\big)\cdot\left(\frac{\nu_k}{2}\right) \\
&= -\left(\frac{\nu_k}{2}\right)\frac{q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}{1-\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}. \qquad (4.48)
\end{aligned}$$

Substituting (4.47) and (4.48) in (4.46),

$$\begin{aligned}
\frac{\partial}{\partial d}\log\left[\frac{\sum_{k=1}^{s} q_k p(0|\nu_k, d)}{1-\sum_{k=1}^{s} q_k p(0|\nu_k, d)}\right] &= \left\{\begin{array}{c}\left(\frac{\nu_k}{2}\right)\frac{q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}{\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}} \\ +\left(\frac{\nu_k}{2}\right)\frac{q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}{1-\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}\end{array}\right\} \\
&= \frac{\nu_k}{2}\left(\frac{q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}{1-\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}\right). \qquad (4.49)
\end{aligned}$$

Substituting (4.45) and (4.49) in (4.44),

$$\begin{aligned}
\frac{\partial}{\partial d}L_p(Q,\gamma) &= 0 \\
\left\{\begin{array}{c}\frac{\nu_k}{2}\sum_{i=0}^{m} f_i e_{ik} - \frac{1}{2-d}\sum_{i=0}^{m} i f_i e_{ik} \\ +\frac{d}{2\nu_k(2-d)^3}\sum_{i=0}^{m}\frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} - \frac{\nu_k}{2}\left(\frac{q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}{1-\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}\right)\end{array}\right\} &= 0. \qquad (4.50)
\end{aligned}$$

Rewrite (4.50) as

$$\sum_{i=0}^{m}\frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} = \left\{\begin{array}{c}\frac{2\nu_k(2-d)^2}{d}\sum_{i=0}^{m} i f_i e_{ik} - \frac{\nu_k^2(2-d)^3}{d}\sum_{i=0}^{m} f_i e_{ik} \\ +\frac{\nu_k^2(2-d)^3}{d}\frac{q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}{1-\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}\end{array}\right\}. \qquad (4.51)$$

Substituting (4.51) in (4.43),

$$\begin{aligned}
\left\{\begin{array}{c}
\left(\frac{d-3}{2}\right)\sum_{i=0}^{m} f_i e_{ik} + \frac{1}{\nu_k}\sum_{i=0}^{m} i f_i e_{ik} \\
-\frac{d-1}{\nu_k d}\sum_{i=0}^{m} i f_i e_{ik} - \frac{(d-1)(2-d) q_k e^{\nu_k(\frac{d-3}{2})}}{2d\left(1-\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}\right)} \\
+\frac{(d-1)(2-d)}{2d}\sum_{i=0}^{m} f_i e_{ik} - \left(\frac{d-3}{2}\right)\frac{q_k e^{\nu_k(\frac{d-3}{2})}}{1-\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}}
\end{array}\right\} &= 0 \\
-\frac{1}{d}\sum_{i=0}^{m} f_i e_{ik} + \frac{1}{\nu_k d}\sum_{i=0}^{m} i f_i e_{ik} + \frac{q_k e^{\nu_k(\frac{d-3}{2})}}{d\left(1-\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}\right)} &= 0 \\
-\sum_{i=0}^{m} f_i e_{ik} + \frac{1}{\nu_k}\sum_{i=0}^{m} i f_i e_{ik} + \frac{q_k e^{\nu_k(\frac{d-3}{2})}}{1-\sum_{k=1}^{s} q_k e^{\nu_k\left(\frac{d-3}{2}\right)}} &= 0 \\
\frac{1}{\nu_k} = \frac{\sum_{i=0}^{m} f_i e_{ik} - \frac{q_k e^{\nu_k(\frac{d-3}{2})}}{1-\sum_{k=1}^{s} q_k e^{\nu_k(\frac{d-3}{2})}}}{\sum_{i=0}^{m} i f_i e_{ik}}
\end{aligned}$$

$$\hat{\nu}_k = \frac{\sum_{i=0}^{m} i f_i e_{ik}}{\sum_{i=0}^{m} f_i e_{ik} - \frac{q_k e^{\nu_k(\frac{d-3}{2})}}{1-\sum_{k=1}^{s} q_k e^{\nu_k(\frac{d-3}{2})}}}. \tag{4.52}$$

Substituting (4.52) in (4.51),

$$\begin{aligned}
\sum_{i=0}^{m}\frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} &= \left\{\begin{array}{c}
\frac{2\hat{\nu}_k(2-d)^2}{d}\sum_{i=1}^{m} i f_i e_{ik} - \frac{\hat{\nu}_k^2(2-d)^3}{d}\sum_{i=0}^{m} f_i e_{ik} \\
+\frac{\hat{\nu}_k^2(2-d)^3}{d}\frac{q_k e^{\nu_k(\frac{d-3}{2})}}{1-\sum_{k=1}^{s} q_k e^{\hat{\nu}_k(\frac{d-3}{2})}}.
\end{array}\right\} \\
&= [\hat{\nu}_k(2-d)]^2\left[\sum_{i=0}^{m} f_i e_{ik} - \frac{q_k e^{\nu_k(\frac{d-3}{2})}}{1-\sum_{k=1}^{s} q_k e^{\hat{\nu}_k(\frac{d-3}{2})}}\right].
\end{aligned} \tag{4.53}$$

Parameter d be found by solving (4.53). The package $hermite$ in program R allows to estimate the parameter d given an univariate sample by means of the function $glm.hermite$.

The penalized maximum likelihood estimator (PMLE) based on discrete mixtures of Hermite distributions through the Horvitz-Thomson approach is

$$\hat{N}_{PMLE} = \frac{n}{1-\sum_{k=1}^{s}\hat{q}_k e^{\hat{\nu}_k(\frac{\hat{d}-3}{2})}}. \tag{4.54}$$

The EM algorithm used to compute the estimate of population size is given as follows.

4.2.1.1 EM algorithm

Step 0: Choose initial value $\hat{Q}^{(0)} = \begin{pmatrix} \hat{\nu}_1^{(0)} & \hat{\nu}_2^{(0)} & \cdots & \hat{\nu}_s^{(0)} \\ \hat{d}^{(0)} & \hat{d}^{(0)} & \cdots & \hat{d}^{(0)} \\ \hat{q}_1^{(0)} & \hat{q}_2^{(0)} & \cdots & \hat{q}_s^{(0)} \end{pmatrix}$, set $t = 0$.

Step 1: Compute

$$\begin{aligned} \hat{f}_0^{(t+1)} &= \frac{nf(0;\hat{Q}^{(t)})}{1-f(0;\hat{Q}^{(t)})} \\ e_{ik}^{(t)} &= \frac{q_k^{(t)}p(i|\nu_k^{(t)},d^{(t)})}{\sum_{k=1}^{s} q_k^{(t)}p(i|\nu_k^{(t)},d^{(t)})}, \qquad \text{for} \quad k=1,2,...,s \\ P_1^{(t)} &= \frac{e_{0k}^{(t)}}{1-\sum_{k=1}^{s} q_k^{(t)}p(0|\nu_k^{(t)},d^{(t)})} \\ P_2^{(t)} &= \frac{1}{1-\sum_{k=1}^{s} q_k^{(t)}p(0|\nu_k^{(t)},d^{(t)})}. \end{aligned}$$

Step 2: Use complete data $\hat{f}_0^{(t+1)}, f_1, f_2, ..., f_m$ to compute the new MLEs of

$$\begin{aligned} \hat{q}_k^{(t+1)} &= \frac{\sum_{i=0}^{m} f_i e_{ik}^{(t)} - P_1^{(t)}}{n+\hat{f}_0^{(t+1)} - P_2^{(t)}}, \qquad \text{for} \quad k=1,2,...,s \\ \hat{\nu}_k^{(t+1)} &= \frac{\sum_{i=0}^{m} i f_i e_{ik}^{(t)}}{\sum_{i=0}^{m} f_i e_{ik}^{(t)} - P_1^{(t)}} \\ \hat{d}_k^{(t+1)} &= glm.hermite(\hat{f}_0^{(t+1)}, f_1, ..., f_m \sim 1, link = "log", start = NULL, m^{\star} = 2). \end{aligned}$$

Step 3: Set $t = t + 1$ and repeat Step 1. Steps 1 and 2 are repeated until:

$$\left|\hat{q}^{(t+1)} - \hat{q}^{(t)}\right| \leq 10^{-4}, \quad \left|\hat{\nu}^{(t+1)} - \hat{\nu}^{(t)}\right| \leq 10^{-4}, \text{ and } \quad \left|\hat{d}^{(t+1)} - \hat{d}^{(t)}\right| \leq 10^{-4}.$$

To illustrate that the penalized likelihood improves the boundary problem, we look at the example data in Table 4.1 again. Table 4.3 provides the results of mixture model based on penalized likelihood analysis. There is no difference between the log-likelihoods for models with $s = 2$ and $s = 3$ components. On the basis of selection criteria, the model that give smaller BIC is better. As a consequence, the appropriate model is two components model leading to an estimate of 506 for the population size. Evidently, PMLE can improve the problem and provides the estimation close to the population size N.

Table 4.3: Penalized likelihood analysis for the example data in Table 4.1.

s	$\hat{q}_k$	$\hat{\nu}_k$	$\hat{d}$	log$L(\hat{Q}_k)$	BIC	$\hat{f}_0$	$\hat{N}_{CMLE}$
1	1.0000	3.1324	1.5655	-743.3534	1,498.4397	42	411
2	0.1350 0.8650	0.1675 2.7871	1.7519	-726.0088	1,475.4835	137	506
3	0.0342 0.1008 0.8650	0.1675 0.1675 2.7871	1.7519	-726.0090	1,487.2168	137	506

4.2.2 Confidence interval estimation for population size N based on the profile mixture likelihood

Inferring the unknown size N of closed population is discussed in this section. Chao (1989) pointed out that derivation of $Var(\hat{N})$ to form the confidence interval of N is not easy task in capture-recapture studies. A bootstrapping approach can be used to construct confidence intervals. Alternatively, profile likelihood can be used to construct confidence intervals (Norris III & Pollock, 1996; Norris & Pollock, 1998). Confidence interval estimations of N based on the profile mixture likelihood of Hermite distribution can be constructed as follows.

4.2.2.1 Profile mixture likelihood of Poisson-Normal mixture models

The full likelihood function is given by

$$L(N,Q) = \frac{N!}{f_0!f_1!...f_m!}\left(\sum_{k=1}^{s} q_k p(0|\nu_k,d)\right)^{N-n}\prod_{i=1}^{m}\left(\sum_{k=1}^{s} q_k p(i|\nu_k,d)\right)^{f_i}. \quad (4.55)$$

The likelihood is described by full parameter (N, Q), but we are interested only in N. Consequently, a nuisance parameter Q is eliminated by replacing it with its MLE at each fixed value of N.

Since $N = n + f_0$, finding a profile likelihood of N is equivalent to finding a profile likelihood of f_0. The full likelihood function (4.55) can be rewritten as

$$L(f_0, Q) = \frac{(n+f_0)!}{f_0!f_1!...f_m!}\left(\sum_{k=1}^{s} q_k p(0|\nu_k, d)\right)^{f_0} \prod_{i=1}^{m}\left(\sum_{k=1}^{s} q_k p(i|\nu_k, d)\right)^{f_i}. \tag{4.56}$$

Given a fixed f_0, the log-likelihood function of (4.56), with the constant terms are omit, takes the form

$$l(Q|f_0) = \sum_{i=0}^{m} f_i \log \sum_{k=1}^{s} q_k p(i|\nu_k, d). \tag{4.57}$$

Here, $p(i|\nu_k, d) = e^{\nu_k(\frac{d-3}{2})}[\nu_k(2-d)]^i q_{ik}(\theta_k)$, $q_{yk}(\theta_k) = \sum_{j=0}^{[y/2]} \frac{\theta_k^j}{(y-2j)!j!}$, and $\theta_k = \frac{d-1}{2\nu_k(2-d)^2}$ with respect to the unobserved variables Z having distribution Q. A discrete mixing distribution $Q = \begin{pmatrix} \nu_1 & \nu_2 & \cdots & \nu_s \\ d & d & \cdots & d \\ q_1 & q_2 & \cdots & q_s \end{pmatrix}$ gives weight q_k to parameters ν_k and d for $k = 1, 2, ..., s$, where s is the number of unobserved components. Note that $q_k \geq 0$ and $\sum_{k=1}^{s} q_k = 1$.

To find the MLE of Q for any fixed f_0, (4.57) is maximized by using the EM algorithm as well. Let z_{ik} be indicator variables that defined as following

$$z_{ik} = \begin{cases} 1 & \text{if count } i \text{ arose from component } k \\ 0 & \text{otherwise.} \end{cases}$$

If z_{ik} are observable, the log-likelihood is

$$l(Q|f_0) = \sum_{i=0}^{m} f_i \sum_{k=1}^{s} z_{ik} \log q_k + \sum_{i=0}^{m} f_i \sum_{k=1}^{s} z_{ik} \log p(i|\nu_k, d). \tag{4.58}$$

At the E-step, the unobserved indicator z_{ik} is replaced by e_{ik}, its expected value conditional given the observed data and current values of Q

$$e_{ik} = E(z_{ik}|\text{observed data}; Q) = \frac{q_k p(i|\nu_k, d)}{\sum_{k=1}^{s} q_k p(i|\nu_k, d)}. \tag{4.59}$$

Substituting e_{ik} into (4.58), the expected log-likelihood is given as

$$l(Q|f_0) = \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log q_k + \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k, d). \tag{4.60}$$

M-step: In mathematical optimization, the method of Lagrange multipliers is a strategy for finding the local maximum or minimum of function subject to equality constraints. To maximize (4.60) subject to the constraint $\sum_{k=1}^{s} q_k = 1$, the Lagrange function is given by

$$L(Q|f_0, \lambda) = \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log q_k + \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k, d) + \gamma(1 - \sum_{k=1}^{s} q_k). \quad (4.61)$$

Differentiating (4.61) with respect to γ, q_k, ν_k, and d, and setting the result to 0:

$$\begin{aligned} \frac{\partial}{\partial \gamma} L(Q|f_0, \gamma) &= 0 \\ 1 - \sum_{k=1}^{s} q_k &= 0 \\ \sum_{k=1}^{s} q_k &= 1. \end{aligned} \quad (4.62)$$

$$\begin{aligned} \frac{\partial}{\partial q_k} L(Q|f_0, \gamma) &= 0 \\ \frac{\sum_{i=0}^{m} f_i e_{ik}}{q_k} - \gamma &= 0 \\ q_k &= \frac{\sum_{i=0}^{m} f_i e_{ik}}{\gamma} \\ \sum_{k=1}^{s} q_k &= \frac{\sum_{i=0}^{m} \sum_{k=1}^{s} f_i e_{ik}}{\gamma}. \end{aligned} \quad (4.63)$$

Solving equation (4.62) and (4.63) provides

$$\gamma = \sum_{i=0}^{m} \sum_{k=1}^{s} f_i e_{ik}.$$

Therefore,

$$\begin{aligned} \hat{q}_k &= \frac{\sum_{i=0}^{m} f_i e_{ik}}{\sum_{i=0}^{m} \sum_{k=1}^{s} f_i e_{ik}} \\ &= \frac{\sum_{i=0}^{m} f_i e_{ik}}{\hat{N}}. \end{aligned} \quad (4.64)$$

$$\begin{aligned} \frac{\partial}{\partial \nu_k} L(Q|f_0, \gamma) &= 0 \\ \frac{\partial}{\partial \nu_k} \sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik} \log p(i|\nu_k, d) &= 0. \end{aligned} \quad (4.65)$$

Substituting (4.38) in (4.65),

$$\begin{aligned} \frac{\partial}{\partial \nu_k} L(Q|f_0, \gamma) &= 0 \\ \frac{d-3}{2} \sum_{i=0}^{m} f_i e_{ik} + \frac{1}{\nu_k} \sum_{i=0}^{m} i f_i e_{ik} - \frac{d-1}{2(2-d)^2 {\nu_k}^2} \sum_{i=0}^{m} \frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} &= 0. \end{aligned} \quad (4.66)$$

$$\begin{aligned} \frac{\partial}{\partial d} L(Q|f_0,\gamma) &= 0 \\ \frac{\partial}{\partial d}\sum_{i=0}^{m} f_i \sum_{k=1}^{s} e_{ik}\log p(i|\nu_k,d) &= 0. \end{aligned} \tag{4.67}$$

Substituting (4.45) in (4.67),

$$\begin{aligned} \frac{\partial}{\partial d} L(Q|f_0,\gamma) &= 0 \\ \frac{1}{2}\sum_{i=0}^{m} f_i e_{ik}\nu_k - \frac{1}{2-d}\sum_{i=0}^{m} i f_i e_{ik} + \frac{d}{2(2-d)^3\nu_k}\sum_{i=0}^{m}\frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} &= 0. \end{aligned} \tag{4.68}$$

Rewrite (4.68) as

$$\begin{aligned} \sum_{i=0}^{m}\frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} &= \frac{2(2-d)^3\nu_k}{d}\left(\frac{1}{2-d}\sum_{i=0}^{m} i f_i e_{ik} - \frac{\nu_k}{2}\sum_{i=0}^{m} f_i e_{ik}\right) \\ &= \frac{2(2-d)^2\nu_k}{d}\sum_{i=0}^{m} i f_i e_{ik} - \frac{(2-d)^3\nu_k^2}{d}\sum_{i=0}^{m} f_i e_{ik}. \end{aligned} \tag{4.69}$$

Substituting (4.69) in (4.66),

$$\begin{aligned} \left(\frac{d-3}{2} + \frac{(d-1)(2-d)}{2d}\right)\sum_{i=0}^{m} f_i e_{ik} + \left(\frac{1}{\nu_k} - \frac{d-1}{\nu_k d}\right)\sum_{i=0}^{m} i f_i e_{ik} &= 0 \\ -\frac{1}{d}\sum_{i=0}^{m} f_i e_{ik} + \frac{1}{\nu_k d}\sum_{i=0}^{m} i f_i e_{ik} &= 0 \\ \frac{1}{\nu_k d}\sum_{i=0}^{m} i f_i e_{ik} &= \frac{1}{d}\sum_{i=0}^{m} f_i e_{ik} \\ \hat{\nu}_k &= \frac{\sum_{i=0}^{m} i f_i e_{ik}}{\sum_{i=0}^{m} f_i e_{ik}}. \end{aligned} \tag{4.70}$$

Substituting (4.70) in (4.69),

$$\begin{aligned} \sum_{i=0}^{m}\frac{f_i e_{ik} q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} &= \frac{2(2-d)^2\hat{\nu}_k^2}{d}\sum_{i=0}^{m} f_i e_{ik} - \frac{(2-d)^3\hat{\nu}_k^2}{d}\sum_{i=0}^{m} f_i e_{ik} \\ &= [(2-d)\hat{\nu}_k]^2\sum_{i=0}^{m} f_i e_{ik}. \end{aligned} \tag{4.71}$$

Parameter d be found by solving (4.71). The package $hermite$ in program R allows to estimate the parameter d given an univariate sample by means of the function $glm.hermite$, see (Moriña et al., 2015) for more details. The profile MLE of q_k, ν_k, and d for any fixed f_0 are

$$\hat{q}_k(f_0) = \frac{\sum_{i=0}^{m} f_i e_{ik}}{\sum_{i=0}^{m}\sum_{k=1}^{s} f_i e_{ik}} \tag{4.72}$$

$$\hat{\nu}_k(f_0) = \frac{\sum_{i=0}^{m} i f_i e_{ik}}{\sum_{i=0}^{m} f_i e_{ik}} \tag{4.73}$$

$$\hat{d}(f_0) = glm.hermite(f_0, f_1, f_2, ..., f_m \sim 1, link = \text{"}log\text{"}, start = NULL, m^{\star} = 2). \tag{4.74}$$

The profile log-likelihood for any fixed f_0 under discrete mixture of Hermite distribution can be achieved as

$$l\big(f_0, \hat{Q}(f_0)\big) = \left\{ \begin{array}{c} \log \Gamma(n + f_0 + 1) - \sum_{i=0}^{m} \log \Gamma(f_i + 1) \\ + \sum_{i=0}^{m} f_i \log \left(\sum_{k=1}^{s} \hat{q}_k(f_0) p\big(i | \hat{\nu}_k(f_0), \hat{d}(f_0)\big) \right) \end{array} \right\}. \tag{4.75}$$

The population size estimator under discrete mixtures of Hermite distributions based on profile mixture likelihood is

$$\hat{N}_{profileMix} = n + \hat{f}_0 \tag{4.76}$$

which $\hat{f}_0$ is the value of f_0 that maximizes (4.75).

4.2.2.2 Confidence interval estimation for population size N

Let $\hat{N}_{profileMix}$ be profile nonparametric maximum likelihood estimator (profile NPMLE) based on discrete mixture of Hermite distribution. Since $\hat{N}_{profileMix} = n + \hat{f}_0$, the likelihood ratio given by

$$2\big[l\big(n + \hat{f}_0, Q(\hat{f}_0)\big) - l\big(n + f_0, \hat{Q}(f_0)\big)\big] \sim \chi^2(1).$$

Using the log-likelihood ratio statistic, all Ns corresponding to

$$2\big[l\big(n + \hat{f}_0, Q(\hat{f}_0)\big) - l\big(n + f_0, \hat{Q}(f_0)\big)\big] \leq (z_{1-\alpha/2})^2$$

form the $100(1-\alpha)\%$ confidence set for $\hat{N}_{profileMix}$. Therefore, the 95% confidence interval for $\hat{N}_{profileMix}$ is the range of $\hat{N}_{profileMix}$ that satisfies

$$2\big[l\big(n + \hat{f}_0, Q(\hat{f}_0)\big) - l\big(n + f_0, \hat{Q}(f_0)\big)\big] - (1.96)^2 \leq 0. \tag{4.77}$$

Table 4.4: Frequencies of test-positives for subjects with the disease.

i	0	1	2	3	...	m
f_i	?	f_1	f_2	f_3	...	f_m

4.3 Mixture model with validation information

In the medical field, several screening test are applied to detect a clinical disease. Due to low sensitivity of the test and human error, any screening test is not 100% accurate. It is possible to find people with negative test but they have the disease so that there are hidden cases. Assume that we analyze a specific disease which status can be tested at m occasions. Also let the count i denotes the number of times the screening test is positive, called a positive sample as show in Table 4.4. If the test is negative at all m times, the true status of the person is unknown. Estimation of f_0 is needed.

Table 4.5: Frequencies of test-positives for a repeated diagnostic testing subjects with the disease.

i	0	1	2	3	...	m
g_i	g_0	g_1	g_2	g_3	$\cdots$	g_m

Sometimes addition information on the observed units is available from another sub-sample of the target population, called a validation sample. In this sample, the size are usually smaller than positive sample, and contain no hidden cases, so that zero counts are observed. It is possible to add the information from validation sample into the model to decrease the bias of estimation of N. Also let $g_0, g_1, ..., g_m$ be the frequency of units identified exactly $0, 1, 2, ..., m$ times. Note that g_0 is known. Table 4.5 provides the example data of the validation sample.

Real data example is surveillance data from Durusoy and Karababa (2010) on Syphilis in Izmir (Turkey) between 21/01/2003 and 25/03/2005 as shown in Table 4.6. Data were collected by one of the two university hospitals, or one of the other six public hospitals. In addition, 133 serology laboratories participated in the study with cases frequently identified by multiple laboratories. As can be seen in Table 4.6 the frequency distribution counts of positive sample from each serology laboratory in Izmir. There are 73 subjects who tested positively once, 52 subjects who tested positively twice and so on. Data were checked for duplicates and matched with the cases that identified by hospitals. Data from hospitals sample showed that there were 18 persons with Syphilis which the diagnostic test was negative at all times.

Table 4.6: Frequencies of Syphilis cases by hospital and count of laboratory identifications.

i	laboratory							Total
	0	1	2	3	4	5	6	
hospital								
0	-	73	52	17	6	1	0	-
1	18	25	22	10	9	1	1	86
	-	98	74	27	15	2	1	235

Böhning et al. (2016) mentioned the capture-recapture modeling using validation sample in extension of generic ratio regression approach. In addition, Arnold et al. (2017) demonstrated that the use of validation sample not only substantially increases the estimation efficiency but also reduces the bias considerably. With this motivation the development of estimation to include validation information in the capture-recapture modeling are proposed in this study, to increase the accuracy and efficiency of population size estimation.

4.3.1 Conditional nonparametric maximum likelihood estimation with validation information

Let Y be the number of times that a unit was identified over the study period. Let Y is mixture of the Hermite distribution. The marginal distribution of Y is

$$f(y;Q)=\sum_{k=1}^{s} q_k p(y|\nu_k,d), \tag{4.78}$$

where $p(y|\nu_k,d) = e^{\nu_k(\frac{d-3}{2})}[\nu_k(2-d)]^y q_{yk}(\theta_k)$, $q_{yk}(\theta_k) = \sum_{j=0}^{[y/2]} \frac{\theta_k^j}{(y-2j)!j!}$, and $\theta_k = \frac{d-1}{2\nu_k(2-d)^2}$ with respect to the unobserved variables Z having distribution Q. A discrete mixing distribution $Q = \begin{pmatrix} \nu_1 & \nu_2 & \cdots & \nu_s \\ d & d & \cdots & d \\ q_1 & q_2 & \cdots & q_s \end{pmatrix}$ gives weight q_k to parameters ν_k and d for $k = 1, 2, ..., s$, where s is the number of unobserved components. Note that $q_k \geq 0$ and $\sum_{k=1}^{s} q_k = 1$. Assume the mixing distribution Q is valid for both positive and validation samples.

Assume $Y_1, Y_2, ..., Y_n$ are observed and drawn from mixture density. The incomplete data likelihood is

$$L(Q) = \prod_{i=1}^{m} \left(\frac{\sum_{k=1}^{s} q_k p(i|\nu_k, d)}{1 - \sum_{k=1}^{s} q_k p(0|\nu_k, d)} \right)^{f_i} \times \prod_{i=0}^{m} \big(\sum_{k=1}^{s} q_k p(i|\nu_k, d) \big)^{g_i}.$$

The incomplete data log-likelihood is

$$l(Q) = \sum_{i=1}^{m} f_i \log \left(\frac{\sum_{k=1}^{s} q_k p(i|\nu_k, d)}{1 - \sum_{k=1}^{s} q_k p(0|\nu_k, d)} \right) + \sum_{i=0}^{m} g_i \log \big(\sum_{k=1}^{s} q_k p(i|\nu_k, d) \big). \quad (4.79)$$

An estimate of Q be achieved by maximizing (4.79) leading to the NPMLE. The EM algorithm has become common for maximum likelihood estimation in mixture models. To apply the EM algorithm, the complete data log-likelihood is required.

At the E-step, the unobserved frequency f_0 is replaced by its expected value given observed frequencies and current values of Q. Let the expected value of f_0, denoted by $\hat{f}_0$, be written as

$$\begin{aligned} \hat{f}_0 &= E(f_0 | \text{observed data}; Q) \\ &= \frac{n f(0; Q)}{1 - f(0; Q)} \\ &= \frac{n \sum_{k=1}^{s} q_k p(0|\nu_k, d)}{1 - \sum_{k=1}^{s} q_k p(0|\nu_k, d)}. \end{aligned} \quad (4.80)$$

The log-likelihood for the complete data is given by

$$\begin{aligned} l_{cd}(Q) &= \sum_{i=0}^{m} f_i \log \big(\sum_{k=1}^{s} q_k p(i|\nu_k, d) \big) + \sum_{i=0}^{m} g_i \log \big(\sum_{k=1}^{s} q_k p(i|\nu_k, d) \big) \\ &= \sum_{i=0}^{m} (f_i + g_i) \log \big(\sum_{k=1}^{s} q_k p(i|\nu_k, d) \big). \end{aligned} \quad (4.81)$$

To manipulate the MLE of $\hat{Q}$, (4.81) is maximized by applying the EM algorithm as well. In this case, a variable indicating component to which the count i

belongs is ignored. Let z_{ik} be indicator variables defined as following

$$z_{ik} = \begin{cases} 1 & \text{if count } i \text{ arose from component } k \\ 0 & \text{otherwise.} \end{cases}$$

If z_{ik} are observable, the complete data log-likelihood is

$$l_{cd}(Q) = \sum_{i=0}^{m}(f_i + g_i)\sum_{k=1}^{s} z_{ik}\log q_k + \sum_{i=0}^{m}(f_i + g_i)\sum_{k=1}^{s} z_{ik}\log p(i|\nu_k, d). \qquad (4.82)$$

At the E-step, the unobserved indicator z_{ik} is replaced by e_{ik}, its expected value conditional given the observed data and current values of Q

$$e_{ik} = E(z_{ik}|\text{observed data}; Q) = \frac{q_k p(i|\nu_k, d)}{\sum_{k=1}^{s} q_k p(i|\nu_k, d)}. \qquad (4.83)$$

The expected log-likelihood is defined as

$$E_V[l_{cd}(Q)] = \sum_{i=0}^{m}(f_i + g_i)\sum_{k=1}^{s} e_{ik}\log q_k + \sum_{i=0}^{m}(f_i + g_i)\sum_{k=1}^{s} e_{ik}\log p(i|\nu_k, d). \qquad (4.84)$$

M-step: In mathematical optimization, the method of Lagrange multipliers is a strategy for finding the local maximum or minimum of a function subject to equality constraints. To maximize (4.84) subject to the constraint $\sum_{k=1}^{s} q_k = 1$, the Lagrange function is given by

$$L_V(Q,\gamma) = \sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log q_k + \sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log p(i|\nu_k, d) + \gamma(1-\sum_{k=1}^{s} q_k). \qquad (4.85)$$

Differentiating (4.85) with respect to parameters γ , q_k, ν_k, and d, and setting the result to 0 yield

$$\begin{aligned} \frac{\partial}{\partial\gamma} L_V(Q,\gamma) &= 0 \\ 1 - \sum_{k=1}^{s} q_k &= 0 \\ \sum_{k=1}^{s} q_k &= 1. \end{aligned} \qquad (4.86)$$

$$\begin{aligned}
\frac{\partial}{\partial q_k} L_V(Q,\gamma) &= 0 \\
\frac{\sum_{i=0}^{m}(f_i+g_i)e_{ik}}{q_k} - \gamma &= 0 \\
q_k &= \frac{\sum_{i=0}^{m}(f_i+g_i)e_{ik}}{\gamma} \\
\sum_{k=1}^{s} q_k &= \frac{\sum_{i=0}^{m}\sum_{k=1}^{s}(f_i+g_i)e_{ik}}{\gamma}.
\end{aligned} \tag{4.87}$$

Solving equation (4.86) and (4.87) provides

$$\begin{aligned}
\gamma &= \sum_{i=0}^{m}\sum_{k=1}^{s}(f_i+g_i)e_{ik} \\
\hat{q}_k &= \frac{\sum_{i=0}^{m}(f_i+g_i)e_{ik}}{\sum_{i=0}^{m}\sum_{k=1}^{s}(f_i+g_i)e_{ik}} \\
\hat{q}_k &= \frac{\sum_{i=0}^{m}(f_i+g_i)e_{ik}}{\hat{N}}.
\end{aligned} \tag{4.88}$$

$$\begin{aligned}
\frac{\partial}{\partial \nu_k} L_V(Q,\gamma) &= 0 \\
\frac{\partial}{\partial \nu_k}\left(\sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log p(i|\nu_k,d)\right) &= 0 \\
\frac{\partial}{\partial \nu_k}\left(\sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log\left[e^{\nu_k(\frac{d-3}{2})}[\nu_k(2-d)]^i q_{ik}(\theta_k)\right]\right) &= 0 \\
\frac{\partial}{\partial \nu_k}\left(\sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\left[\nu_k(\frac{d-3}{2}) + i\log[\nu_k(2-d)] + \log q_{ik}(\theta_k)\right]\right) &= 0 \\
\left\{\begin{array}{l} \frac{d-3}{2}\sum_{i=0}^{m}(f_i+g_i)e_{ik} + \frac{1}{\nu_k}\sum_{i=0}^{m} i(f_i+g_i)e_{ik} \\ +\frac{\partial}{\partial \nu_k}\left[\sum_{i=1}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log q_{ik}(\theta_k)\right] \end{array}\right\} &= 0.
\end{aligned}$$

(4.89)

Consider

$$\begin{aligned}
\frac{\partial}{\partial \nu_k}\left[\sum_{i=1}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log q_{ik}(\theta_k)\right] &= \sum_{i=1}^{m}(f_i+g_i)\frac{\partial}{\partial \nu_k}\left[\sum_{k=1}^{s} e_{ik}\log q_{ik}(\theta_k)\right]\\
&= \sum_{i=1}^{m}(f_i+g_i)e_{ik}\frac{\partial}{\partial \nu_k}\left[\log q_{ik}(\theta_k)\right]\\
&= \sum_{i=1}^{m}(f_i+g_i)e_{ik}\left[\frac{1}{q_{ik}(\theta_k)}\cdot\frac{\partial}{\partial \theta_k}q_{ik}(\theta_k)\cdot\frac{\partial \theta_k}{\partial \nu_k}\right]\\
&= \sum_{i=1}^{m}(f_i+g_i)e_{ik}\left[\frac{q'_{ik}(\theta_k)}{q_{ik}(\theta_k)}\cdot\frac{\partial}{\partial \nu_k}\left(\frac{d-1}{2\nu_k(2-d)^2}\right)\right]\\
&= \sum_{i=1}^{m}(f_i+g_i)e_{ik}\left[\frac{q'_{ik}(\theta_k)}{q_{ik}(\theta_k)}\cdot\left(-\frac{d-1}{2\nu_k^2(2-d)^2}\right)\right]\\
&= -\frac{d-1}{2(2-d)^2{\nu_k}^2}\sum_{i=0}^{m}\frac{(f_i+g_i)e_{ik}q'_{ik}(\theta_k)}{q_{ik}(\theta_k)},
\end{aligned} \tag{4.90}$$

where $q'_{ik}(\theta_k)=\frac{\partial}{\partial \theta_k}q_{ik}(\theta_k)$. Substituting (4.90) in (4.89),

$$\begin{aligned}
\frac{\partial}{\partial \nu_k}L_V(Q,\gamma) &= 0\\
\frac{\partial}{\partial \nu_k}\left(\sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log p(i|\nu_k,d)\right) &= 0\\
\left\{\begin{array}{c}\frac{d-3}{2}\sum_{i=0}^{m}(f_i+g_i)e_{ik}+\frac{1}{\nu_k}\sum_{i=0}^{m}i(f_i+g_i)e_{ik}\\ -\frac{d-1}{2(2-d)^2{\nu_k}^2}\sum_{i=0}^{m}\frac{(f_i+g_i)e_{ik}q'_{ik}(\theta_k)}{q_{ik}(\theta_k)}\end{array}\right\} &= 0.
\end{aligned} \tag{4.91}$$

$$\begin{aligned}
\frac{\partial}{\partial d}L_V(Q,\gamma) &= 0\\
\frac{\partial}{\partial d}\left(\sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log p(i|\nu_k,d)\right) &= 0\\
\frac{\partial}{\partial d}\left(\sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log\left[e^{\nu_k(\frac{d-3}{2})}[\nu_k(2-d)]^i q_{ik}(\theta_k)\right]\right) &= 0\\
\frac{\partial}{\partial d}\left(\sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\left[\nu_k(\frac{d-3}{2})+i\log[\nu_k(2-d)]+\log q_{ik}(\theta_k)\right]\right) &= 0\\
\left\{\begin{array}{c}\frac{1}{2}\sum_{i=0}^{m}(f_i+g_i)e_{ik}\nu_k-\frac{1}{2-d}\sum_{i=0}^{m}i(f_i+g_i)e_{ik}\\ +\frac{\partial}{\partial d}\left[\sum_{i=1}^{m}(f_i+g_i)\sum_{k=1}^{s}e_{ik}\log q_{ik}(\theta_k)\right]\end{array}\right\} &= 0.
\end{aligned} \tag{4.92}$$

Consider

$$\begin{aligned}
\frac{\partial}{\partial d}\left[\sum_{i=1}^{m}(f_i+g_i)\sum_{k=1}^{s}e_{ik}\log q_{ik}(\theta_k)\right] &= \sum_{i=1}^{m}(f_i+g_i)\frac{\partial}{\partial d}\left[\sum_{k=1}^{s}e_{ik}\log q_{ik}(\theta_k)\right]\\
&= \sum_{i=1}^{m}(f_i+g_i)e_{ik}\frac{\partial}{\partial d}\left[\log q_{ik}(\theta_k)\right]\\
&= \sum_{i=1}^{m}(f_i+g_i)e_{ik}\left[\frac{1}{q_{ik}(\theta_k)}\cdot\frac{\partial}{\partial \theta_k}q_{ik}(\theta_k)\cdot\frac{\partial \theta_k}{\partial d}\right]\\
&= \sum_{i=1}^{m}(f_i+g_i)e_{ik}\left[\frac{q'_{ik}(\theta_k)}{q_{ik}(\theta_k)}\cdot\frac{\partial}{\partial d}\left(\frac{d-1}{2\nu_k(2-d)^2}\right)\right]\\
&= \frac{d}{2\nu_k(2-d)^3}\sum_{i=0}^{m}\frac{(f_i+g_i)e_{ik}q'_{ik}(\theta_k)}{q_{ik}(\theta_k)}, \qquad (4.93)
\end{aligned}$$

where $q'_{ik}(\theta_k)=\frac{\partial}{\partial\theta_k}q_{ik}(\theta_k)$. Substituting (4.93) in (4.92),

$$\begin{aligned}
\frac{\partial}{\partial d}L_V(Q,\gamma) &= 0\\
\frac{\partial}{\partial d}\left(\sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s}e_{ik}\log p(i|\nu_k,d)\right) &= 0\\
\left\{\begin{array}{c}\frac{1}{2}\sum_{i=0}^{m}(f_i+g_i)e_{ik}\nu_k-\frac{1}{2-d}\sum_{i=0}^{m}i(f_i+g_i)e_{ik}\\ +\frac{d}{2(2-d)^3\nu_k}\sum_{i=0}^{m}\frac{(f_i+g_i)e_{ik}q'_{ik}(\theta_k)}{q_{ik}(\theta_k)}\end{array}\right\} &= 0. \qquad (4.94)
\end{aligned}$$

Rewrite (4.94) as

$$\begin{aligned}
\sum_{i=0}^{m}\frac{(f_i+g_i)e_{ik}q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} &= \frac{2(2-d)^3\nu_k}{d}\left(\frac{1}{2-d}\sum_{i=0}^{m}i(f_i+g_i)e_{ik}-\frac{\nu_k}{2}\sum_{i=0}^{m}(f_i+g_i)e_{ik}\right)\\
&= \frac{2(2-d)^2\nu_k}{d}\sum_{i=0}^{m}i(f_i+g_i)e_{ik}-\frac{(2-d)^3\nu_k^2}{d}\sum_{i=0}^{m}(f_i+g_i)e_{ik}.
\end{aligned}$$

(4.95)

Substituting (4.95) in (4.91),

$$\begin{aligned}
\left\{ \begin{array}{c} \left(\frac{d-3}{2}+\frac{(d-1)(2-d)}{2d}\right)\sum_{i=0}^{m}(f_i+g_i)e_{ik} \\ +\left(\frac{1}{\nu_k}-\frac{d-1}{\nu_k d}\right)\sum_{i=0}^{m} i(f_i+g_i)e_{ik} \end{array} \right\} &= 0 \\
-\frac{1}{d}\sum_{i=0}^{m}(f_i+g_i)e_{ik}+\frac{1}{\nu_k d}\sum_{i=0}^{m} i(f_i+g_i)e_{ik} &= 0 \\
\frac{1}{\nu_k d}\sum_{i=0}^{m} i(f_i+g_i)e_{ik} &= \frac{1}{d}\sum_{i=0}^{m}(f_i+g_i)e_{ik} \\
\hat{\nu}_k &= \frac{\sum_{i=0}^{m} i(f_i+g_i)e_{ik}}{\sum_{i=0}^{m}(f_i+g_i)e_{ik}}.
\end{aligned} \tag{4.96}$$

Substituting (4.96) in (4.95),

$$\begin{aligned}
\sum_{i=0}^{m}\frac{(f_i+g_i)e_{ik}q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} &= \frac{2(2-d)^2\hat{\nu}_k^2}{d}\sum_{i=0}^{m}(f_i+g_i)e_{ik}-\frac{(2-d)^3\hat{\nu}_k^2}{d}\sum_{i=0}^{m}(f_i+g_i)e_{ik} \\
&= [(2-d)\hat{\nu}_k]^2\sum_{i=0}^{m}(f_i+g_i)e_{ik}. \qquad (4.97)
\end{aligned}$$

Parameter d be found by solving (4.97). The package $hermite$ in program R allows to estimate the parameter d given an univariate sample by means of the function $glm.hermite$.

The population size estimator based on discrete mixtures of Hermite distribution with validation information through the Horvitz-Thomson approach is

$$\hat{N}_{Valid}=\frac{n}{1-\sum_{k=1}^{s}\hat{q}_k e^{\hat{\nu}_k(\frac{\hat{d}-3}{2})}}. \tag{4.98}$$

4.3.1.1 EM algorithm

In practical terms, the EM algorithm is implemented as follows.

Step 0: Choose initial value $\hat{Q}^{(0)}=\begin{pmatrix}\hat{\nu}_1^{(0)} & \hat{\nu}_2^{(0)} & \cdots & \hat{\nu}_s^{(0)} \\ \hat{d}^{(0)} & \hat{d}^{(0)} & \cdots & \hat{d}^{(0)} \\ \hat{q}_1^{(0)} & \hat{q}_2^{(0)} & \cdots & \hat{q}_s^{(0)}\end{pmatrix}$, set $t=0$.

Step 1: Compute

$$\begin{aligned}
\hat{f}_0^{(t+1)} &= \frac{nf(0;\hat{Q}^{(t)})}{1-f(0;\hat{Q}^{(t)})} \\
e_{ik}^{(t)} &= \frac{q_k^{(t)}p(i|\nu_k^{(t)},d^{(t)})}{\sum_{k=1}^{s}q_k^{(t)}p(i|\nu_k^{(t)},d^{(t)})}, \qquad \text{for} \quad k=1,2,...,s.
\end{aligned}$$

Step 2: Use complete data $\hat{f}_0^{(t+1)}, f_1, f_2, ..., f_m$ and validation data $g_0, g_1, ..., g_m$ to compute the new MLEs of

$$\begin{aligned}
\hat{q}_k^{(t+1)} &= \frac{\sum_{i=0}^{m}(f_i+g_i)e_{ik}^{(t)}}{\sum_{i=0}^{m}\sum_{k=1}^{s}(f_i+g_i)e_{ik}^{(t)}}, \qquad \text{for} \quad k = 1, 2, ..., s. \\
\hat{\nu}_k^{(t+1)} &= \frac{\sum_{i=0}^{m} i(f_i+g_i)e_{ik}^{(t)}}{\sum_{i=0}^{m}(f_i+g_i)e_{ik}^{(t)}} \\
\hat{d}_k^{(t+1)} &= glm.hermite(\hat{f}_0^{(t+1)}, ..., f_m, g_0, ..., g_m \sim 1, link = "log", start = NULL, m^{\star} = 2)
\end{aligned}$$

Step 3: Set $t = t + 1$ and repeat Step 1. Steps 1 and 2 are repeated until:

$$\left|\hat{q}^{(t+1)} - \hat{q}^{(t)}\right| \leq 10^{-4}, \quad \left|\hat{\nu}^{(t+1)} - \hat{\nu}^{(t)}\right| \leq 10^{-4}, \text{ and } \quad \left|\hat{d}^{(t+1)} - \hat{d}^{(t)}\right| \leq 10^{-4}.$$

4.3.2 Confidence interval estimation for population size N based on the profile mixture likelihood with validation information

In this section, focuses on inferring the unknown population size N. Derivation of $Var(\hat{N})$ to form the confidence interval of N is not easy in capture-recapture studies. The profile mixture likelihood is used to construct confidence intervals estimation of N as follows.

4.3.2.1 Profile mixture likelihood of Poisson-Normal mixture models with validation information

The unconditional nonparametric likelihood is

$$L(N,Q) = \frac{N!}{f_0!f_1!...f_m!g_0!g_1!...g_m!}\left(\sum_{k=1}^{s} q_k p(0|\nu_k, d)\right)^{N-n+g_0}\prod_{i=1}^{m}\left(\sum_{k=1}^{s} q_k p(i|\nu_k, d)\right)^{f_i+g_i}. \tag{4.99}$$

The likelihood is described by full parameter (N, Q), but we are interested only in N. A nuisance parameter Q is eliminated and replaced by its MLE at each fixed value of N (Pawitan, 2001). This is called a profile mixture likelihood.

Since $N = n + f_0$, finding a profile likelihood of N is equivalent to finding a profile likelihood of f_0. The full likelihood function (4.99) can be rewritten as

$$L(f_0, Q) = \frac{(n+f_0)!}{f_0!f_1!...f_m!g_0!g_1!...g_m!}\Bigg(\sum_{k=1}^{s} q_k p(0|\nu_k, d)\Bigg)^{f_0+g_0}\prod_{i=1}^{m}\Bigg(\sum_{k=1}^{s} q_k p(i|\nu_k, d)\Bigg)^{f_i+g_i}. \tag{4.100}$$

Given fixed f_0, the log-likelihood function which the constants term are omit takes the form

$$l(Q|f_0) = \sum_{i=0}^{m}(f_i+g_i)\log\big(\sum_{k=1}^{s} q_k p(i|\nu_k, d)\big). \tag{4.101}$$

To obtain the MLEs of Q at given fixed f_0, we uses the EM algorithm similar to one proposed previously. Let z_{ik} be indicator variables defined as following

$$z_{ik} = \begin{cases} 1 & \text{if count } i \text{ arose from component } k \\ 0 & \text{otherwise.} \end{cases}$$

If z_{ik} were observed, the log-likelihood for the complete data is given by

$$l(Q|f_0) = \sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} z_{ik}\log q_k + \sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} z_{ik}\log p(i|\nu_k, d). \tag{4.102}$$

At the E-step, the unobserved indicator z_{ik} is replaced by e_{ik}, its expected value conditional given the observed data and current values of Q

$$e_{ik} = E(z_{ik}|\text{observed data}; Q) = \frac{q_k p(i|\nu_k, d)}{\sum_{k=1}^{s} q_k p(i|\nu_k, d)}. \tag{4.103}$$

The expected log-likelihood is given by

$$E_V[l(Q|f_0)] = \sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log q_k + \sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log p(i|\nu_k, d). \tag{4.104}$$

M-step: In mathematical optimization, the method of Lagrange multipliers is a strategy for finding the local maximum or minimum of a function subject to equality constraints. To maximize (4.104) subject to the constraint $\sum_{k=1}^{s} q_k = 1$, the Lagrange function is

$$L_V(Q|f_0, \gamma) = \sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log q_k + \sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log p(i|\nu_k, d) + \gamma(1-\sum_{k=1}^{s} q_k). \tag{4.105}$$

Differentiating (4.105) with respect to γ , q_k, ν_k, and d , and setting the result to 0:

$$\begin{aligned} \frac{\partial}{\partial \gamma} L_V(Q|f_0,\gamma) &= 0 \\ 1 - \sum_{k=1}^{s} q_k &= 0 \\ \sum_{k=1}^{s} q_k &= 1. \end{aligned} \tag{4.106}$$

$$\begin{aligned} \frac{\partial}{\partial q_k} L_V(Q|f_0,\gamma) &= 0 \\ \frac{\sum_{i=0}^{m}(f_i+g_i)e_{ik}}{q_k} - \gamma &= 0 \\ q_k &= \frac{\sum_{i=0}^{m}(f_i+g_i)e_{ik}}{\gamma} \\ \sum_{k=1}^{s} q_k &= \frac{\sum_{i=0}^{m}\sum_{k=1}^{s}(f_i+g_i)e_{ik}}{\gamma}. \end{aligned} \tag{4.107}$$

Solving equation (4.106) and (4.107) provides

$$\begin{aligned} \gamma &= \sum_{i=0}^{m}\sum_{k=1}^{s}(f_i+g_i)e_{ik} \\ \hat{q}_k &= \frac{\sum_{i=0}^{m}(f_i+g_i)e_{ik}}{\sum_{i=0}^{m}\sum_{k=1}^{s}(f_i+g_i)e_{ik}} \\ \hat{q}_k &= \frac{\sum_{i=0}^{m}(f_i+g_i)e_{ik}}{\hat{N}}. \end{aligned} \tag{4.108}$$

$$\begin{aligned} \frac{\partial}{\partial \nu_k} L_V(Q|f_0,\gamma) &= 0 \\ \frac{\partial}{\partial \nu_k} \sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log p(i|\nu_k,d) &= 0. \end{aligned} \tag{4.109}$$

Substituting (4.91) in (4.109),

$$\begin{aligned} \frac{\partial}{\partial \nu_k} L_V(Q|f_0,\gamma) &= 0 \\ \left\{ \begin{array}{c} \frac{d-3}{2}\sum_{i=0}^{m}(f_i+g_i)e_{ik} + \frac{1}{\nu_k}\sum_{i=0}^{m} i(f_i+g_i)e_{ik} \\ -\frac{d-1}{2(2-d)^2{\nu_k}^2}\sum_{i=0}^{m}\frac{(f_i+g_i)e_{ik}q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} \end{array} \right\} &= 0. \end{aligned} \tag{4.110}$$

$$\begin{aligned} \frac{\partial}{\partial d} L_V(Q|f_0,\gamma) &= 0 \\ \frac{\partial}{\partial d} \sum_{i=0}^{m}(f_i+g_i)\sum_{k=1}^{s} e_{ik}\log p(i|\nu_k,d) &= 0. \end{aligned} \tag{4.111}$$

Substituting (4.94) in (4.111),

$$\begin{aligned} \frac{\partial}{\partial d} L_V(Q|f_0,\gamma) &= 0 \\ \left\{\begin{array}{c} \frac{1}{2}\sum_{i=0}^{m}(f_i+g_i)e_{ik}\nu_k - \frac{1}{2-d}\sum_{i=0}^{m} i(f_i+g_i)e_{ik} \\ +\frac{d}{2(2-d)^3\nu_k}\sum_{i=0}^{m}\frac{(f_i+g_i)e_{ik}q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} \end{array}\right\} &= 0. \end{aligned} \tag{4.112}$$

Rewrite (4.112) as

$$\begin{aligned} \sum_{i=0}^{m}\frac{(f_i+g_i)e_{ik}q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} &= \frac{2(2-d)^3\nu_k}{d}\left(\frac{1}{2-d}\sum_{i=0}^{m} i(f_i+g_i)e_{ik} - \frac{\nu_k}{2}\sum_{i=0}^{m}(f_i+g_i)e_{ik}\right) \\ &= \frac{2(2-d)^2\nu_k}{d}\sum_{i=0}^{m} i(f_i+g_i)e_{ik} - \frac{(2-d)^3\nu_k^2}{d}\sum_{i=0}^{m}(f_i+g_i)e_{ik}. \end{aligned} \tag{4.113}$$

Substituting (4.113) in (4.110),

$$\begin{aligned} \left\{\begin{array}{c} \left(\frac{d-3}{2}+\frac{(d-1)(2-d)}{2d}\right)\sum_{i=0}^{m}(f_i+g_i)e_{ik} \\ +\left(\frac{1}{\nu_k}-\frac{d-1}{\nu_k d}\right)\sum_{i=0}^{m} i(f_i+g_i)e_{ik} \end{array}\right\} &= 0 \\ -\frac{1}{d}\sum_{i=0}^{m}(f_i+g_i)e_{ik} + \frac{1}{\nu_k d}\sum_{i=0}^{m} i(f_i+g_i)e_{ik} &= 0 \\ \frac{1}{\nu_k d}\sum_{i=0}^{m} i(f_i+g_i)e_{ik} &= \frac{1}{d}\sum_{i=0}^{m}(f_i+g_i)e_{ik} \\ \hat{\nu}_k &= \frac{\sum_{i=0}^{m} i(f_i+g_i)e_{ik}}{\sum_{i=0}^{m}(f_i+g_i)e_{ik}}. \end{aligned} \tag{4.114}$$

Substituting (4.114) in (4.113),

$$\begin{aligned} \sum_{i=0}^{m}\frac{(f_i+g_i)e_{ik}q'_{ik}(\theta_k)}{q_{ik}(\theta_k)} &= \frac{2(2-d)^2\hat{\nu}_k^2}{d}\sum_{i=0}^{m}(f_i+g_i)e_{ik} - \frac{(2-d)^3\hat{\nu}_k^2}{d}\sum_{i=0}^{m}(f_i+g_i)e_{ik} \\ &= [(2-d)\hat{\nu}_k]^2\sum_{i=0}^{m}(f_i+g_i)e_{ik}. \end{aligned} \tag{4.115}$$

Parameter d be found by solving (4.115). The package $hermite$ in program R allows to estimate the parameter d given an univariate sample by means of the function $glm.hermite$. The profile MLEs of q_k, ν_k, and d for any fixed f_0 are

$$\hat{q}_k(f_0) = \frac{\sum_{i=0}^{m}(f_i+g_i)e_{ik}}{\sum_{i=0}^{m}\sum_{k=1}^{s}(f_i+g_i)e_{ik}} \tag{4.116}$$

$$\hat{\nu}_k(f_0) = \frac{\sum_{i=0}^{m} i(f_i + g_i)e_{ik}}{\sum_{i=0}^{m}(f_i + g_i)e_{ik}} \tag{4.117}$$

$$\hat{d}(f_0) = glm.hermite(f_0, ..., f_m, g_0, ..., g_m \sim 1, link = "log", start = NULL, m^{\star} = 2). \tag{4.118}$$

The profile log-likelihood for any fixed f_0 under discrete mixture of Hermite distribution with validation information be achieved as

$$\begin{aligned} l_V\big(f_0, \hat{Q}(f_0)\big) &= \log\Gamma(n + f_0 + 1) - \sum_{i=0}^{m} \log\Gamma(f_i + 1) - \sum_{i=0}^{m} \log\Gamma(g_i + 1) \\ &+ \sum_{i=0}^{m}(f_i + g_i)\log\left(\sum_{k=1}^{s} \hat{q}_k(f_0)p\big(i|\hat{\nu}_k(f_0), \hat{d}(f_0)\big)\right). \end{aligned} \tag{4.119}$$

The population size estimator under discrete mixture of Hermite distribution with validation information based on profile mixture likelihood is

$$\hat{N}_{profileValid} = n + \hat{f}_0 \tag{4.120}$$

which $\hat{f}_0$ is the value of f_0 that maximizes (4.119).

4.3.2.2 Confidence interval estimation for population size N

Let $\hat{N}_{profileValid}$ be profile NPMLE based upon discrete mixture of Hermite distribution with validation information.

Since $\hat{N}_{profileValid} = n + \hat{f}_0$, the likelihood ratio given by

$$2\big[l_V\big(n + \hat{f}_0, Q(\hat{f}_0)\big) - l_V\big(n + f_0, \hat{Q}(f_0)\big)\big] \sim \chi^2(1).$$

Using the log-likelihood ratio statistic, all Ns corresponding to

$$2\big[l_V\big(n + \hat{f}_0, Q(\hat{f}_0)\big) - l_V\big(n + f_0, \hat{Q}(f_0)\big)\big] \leq (z_{1-\alpha/2})^2$$

form the $100(1 - \alpha)\%$ confidence set for $\hat{N}_{profileValid}$. Therefore, the 95% confidence interval for $\hat{N}_{profileValid}$ is the range of $\hat{N}_{profileValid}$ that satisfies

$$2\big[l_V\big(n + \hat{f}_0, Q(\hat{f}_0)\big) - l_V\big(n + f_0, \hat{Q}(f_0)\big)\big] - (1.96)^2 \leq 0. \tag{4.121}$$

4.4 Simulation plan

4.4.1 Point estimation part

A simulation was used to study the performance of proposed estimators, $\hat{N}_{PMLE}$ and $\hat{N}_{Valid}$, and to compare with those of other well-known estimators based on homogeneous and heterogeneous case. $\hat{N}_{MLE}$ and $\hat{N}_{Turing}$ were used as estimators in homogeneous case. Two population size estimators $\hat{N}_{Chao}$ and $\hat{N}_{Censored}$ were used as estimators in heterogeneous case. The count data were generated by Monte Carlo technique using program R. The population size was $N = 100, 500$, and $1,000$ with 25% and 50% validation sample as following distribution:

i. The two-components Poisson mixture model with equal weight

$$p_i = 0.5Poi(1) + 0.5Poi(\lambda)\ .$$

Here, parameters $\lambda \in \{2, 3, 4\}$ indicating weak, moderate, and strong heterogeneity, respectively.

ii. The two-components Hermite mixture model with equal weight

$$p_i = 0.5Herm(1, d) + 0.5Herm(\nu, d)$$

Here, parameters $\nu \in \{2, 3, 4\}$ indicating weak, moderate, and strong heterogeneity, respectively, and dispersion parameters $d \in \{1.2, 1.4, 1.6, 1.8\}$.

The criteria of comparing the performance of estimators were relative bias (Rbias) and relative root mean square error (RRMSE) from 1,000 repeated times as follow:

$$Rbias = \frac{E(\hat{N}) - N}{N}$$

$$RRMSE = \frac{1}{N}\sqrt{Var(\hat{N}) + (E(\hat{N}) - N)^2}.$$

Here, $E(\hat{N}) = \frac{1}{1,000}\sum_{t=1}^{1,000}\hat{N}_{(t)}$, $Var(\hat{N}) = \frac{1}{999}\sum_{t=1}^{1,000}\{\hat{N}_{(t)} - E(\hat{N})\}^2$, and $\hat{N}_{(t)}$ denotes the estimated values of the population size at replication t.

4.4.2 Interval estimation part

A simulation was used to investigate the performance of proposed confidence intervals. The CI based on the profile mixture likelihood without validation information called CI ProfileMix. The CI based on the profile mixture likelihood with validation information called CI ProfileValid. Further simulation was conducted to compare the proposed CIs and several estimators including $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, $\hat{N}_{Chao}$, and $\hat{N}_{Censored}$ that were done by means of normal approximation, $\hat{N} \pm z_{1-\frac{\alpha}{2}}\widehat{Se}(\hat{N})$. The count data were generated by Monte Carlo technique using program R. The population size was $N = 100$ and $1,000$ with 25% and 50% validation sample. Count data were generated from the following distribution:

i. The two-components Poisson mixture model with equal weight

$$p_i = 0.5Poi(1) + 0.5Poi(\lambda)\ .$$

Here, parameters $\lambda \in \{2, 4\}$ indicating weak and strong heterogeneity, respectively.

ii. The two-components Hermite mixture model with equal weight

$$p_i = 0.5Herm(1, d) + 0.5Herm(\nu, d)$$

Here, parameters $\nu \in \{2, 4\}$ indicating weak and strong heterogeneity, respectively, and dispersion parameters $d \in \{1.2, 1.4, 1.6, 1.8\}$.

The percentage of 100 simulated data in which the 95% confidence interval covered the true N called coverage probability (CP) and average lengths (AL) of achieved confidence intervals were the criteria for comparing the performance of estimators. The CP and AL be calculated as

$$CP = \frac{\sum_{t=1}^{100} C_{(t)}}{100}, \tag{4.122}$$

where $C_{(t)}$ equal to 1 if the true population size N contain in the confidence interval, and 0 otherwise.

$$AL = \frac{\sum_{t=1}^{100}(\hat{N}_{U_{(t)}} - \hat{N}_{L_{(t)}})}{100}, \tag{4.123}$$

where $\hat{N}_{U_{(t)}}$ and $\hat{N}_{L_{(t)}}$ are the upper and lower estimation of N at replication t, respectively.

4.5 Simulation results

4.5.1 Point estimation part

4.5.1.1 Simulation results based on the two-components Poisson mixture model

Table 4.7 provided the Rbias of all estimators, which used validation information and did not used validation information. $\hat{N}_{Valid}$ with 50% validation sample provided the smallest Rbias for weak and strong heterogeneity when $N = 500, 1,000$. $\hat{N}_{Valid}$ with 25% validation sample performed the second best in these cases. $\hat{N}_{PMLE}$ gave the smallest Rbias for moderate heterogeneity in all population size. In addition, $\hat{N}_{Valid}$ with 50% validation sample performed the second best for these cases. For weak and strong heterogeneity in $N = 100$, $\hat{N}_{Chao}$ produced the smallest Rbias. $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, and $\hat{N}_{Chao}$ provided underestimation for all cases. $\hat{N}_{Censored}$ gave severe overestimation for all cases, especially for weak heterogeneity.

From the simulation results showed in Table 4.8, $\hat{N}_{Valid}$ with 50% validation sample provided the smallest RRMSE in almost all cases. $\hat{N}_{Valid}$ with 25% validation sample and $\hat{N}_{PMLE}$ performed the second best. $\hat{N}_{PMLE}$ gave the smallest RRMSE for moderate heterogeneity when $N = 1,000$. $\hat{N}_{Censored}$ provided high RRMSE, especially for weak and moderate heterogeneity. $\hat{N}_{MLE}$ and $\hat{N}_{Turing}$ produced high RRMSE, especially for strong heterogeneity. In addition, RRMSE of $\hat{N}_{Chao}$ decreased when the population size N increased.

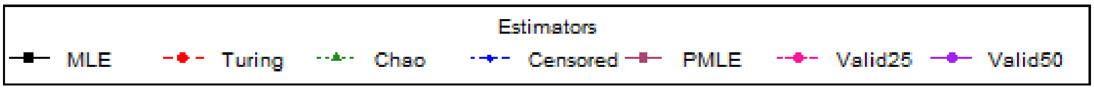
Simulation results indicated that $\hat{N}_{Valid}$ tends to perform reasonably well. It can be an appropriate method for population size estimation for $N \geqslant 500$ in weak and strong heterogeneity. The larger validation sample size, the more efficient estimation. $\hat{N}_{PMLE}$ is an appropriate estimator for population size estimation for moderate heterogeneity in all population size.

Table 4.7: The relative bias of estimators with different parameters in the two-components Poisson mixture model $p_i = 0.5Poi(1) + 0.5Poi(\lambda)$ with 25% and 50% validation sample.

λ	Without validation					With validation	
	MLE	Turing	Chao	Censored	PMLE	Valid	
						25%	50%
				N=100			
2	-0.0610	-0.0460	-0.0050	0.3227	0.0371	0.0511	0.0252
3	-0.1129	-0.0895	-0.0377	0.1736	0.0139	0.0486	0.0387
4	-0.1443	-0.1129	-0.0345	0.1106	-0.0692	-0.0428	-0.0369
				N=500			
2	-0.0600	-0.0483	-0.0265	0.3058	0.0560	0.0106	0.0085
3	-0.1149	-0.0904	-0.0460	0.1739	-0.0070	0.0365	0.0282
4	-0.1448	-0.1127	-0.0441	0.1119	-0.0644	-0.0452	-0.0414
				N=1,000			
2	-0.0628	-0.0508	-0.0297	0.3012	0.0642	0.0125	0.0058
3	-0.1179	-0.0922	-0.0446	0.1738	-0.0131	0.0386	0.0272
4	-0.1475	-0.1149	-0.0453	0.1104	-0.0672	-0.0486	0.0251

Table 4.8: The relative root mean square error of estimators with different parameters in the two-components Poisson mixture model $p_i = 0.5Poi(1) + 0.5Poi(\lambda)$ with 25% and 50% validation sample.

λ	Without validation					With validation	
	MLE	Turing	Chao	Censored	PMLE	Valid	
						25%	50%
N=100							
2	0.0933	0.0877	0.1177	0.3583	0.1510	0.1053	0.0619
3	0.1276	0.1097	0.0989	0.2130	0.0867	0.0860	0.0652
4	0.1523	0.1247	0.1034	0.1504	0.0920	0.0626	0.0512
N=500							
2	0.0684	0.0590	0.0529	0.3131	0.1199	0.0368	0.0285
3	0.1171	0.0933	0.0566	0.1789	0.0462	0.0511	0.0368
4	0.1460	0.1144	0.0526	0.1169	0.0693	0.0493	0.0445
N=1,000							
2	0.0666	0.0556	0.0434	0.3047	0.1220	0.0316	0.0191
3	0.1188	0.0934	0.0506	0.1766	0.0260	0.0459	0.0328
4	0.1480	0.1156	0.0515	0.1138	0.0691	0.0501	0.0286



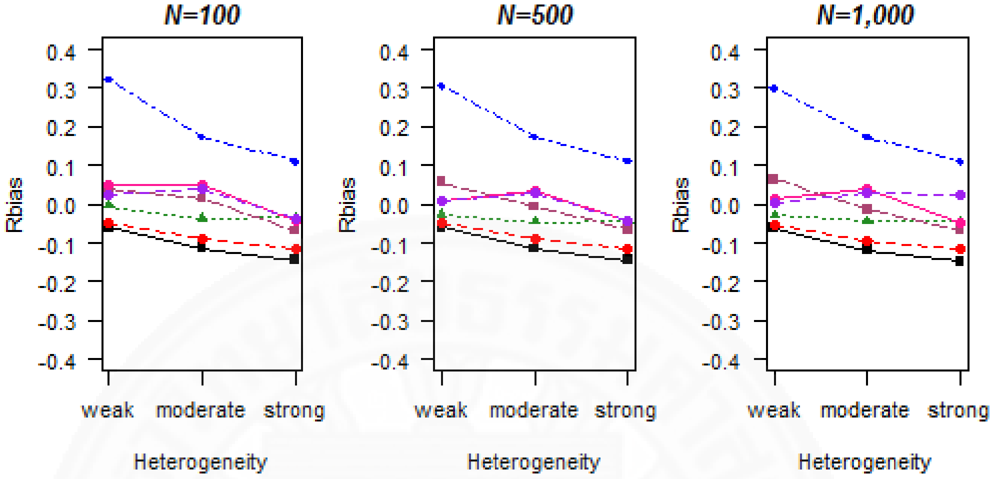


Figure 4.1: The relative bias of estimators with different parameters in the two-components Poisson mixture model.

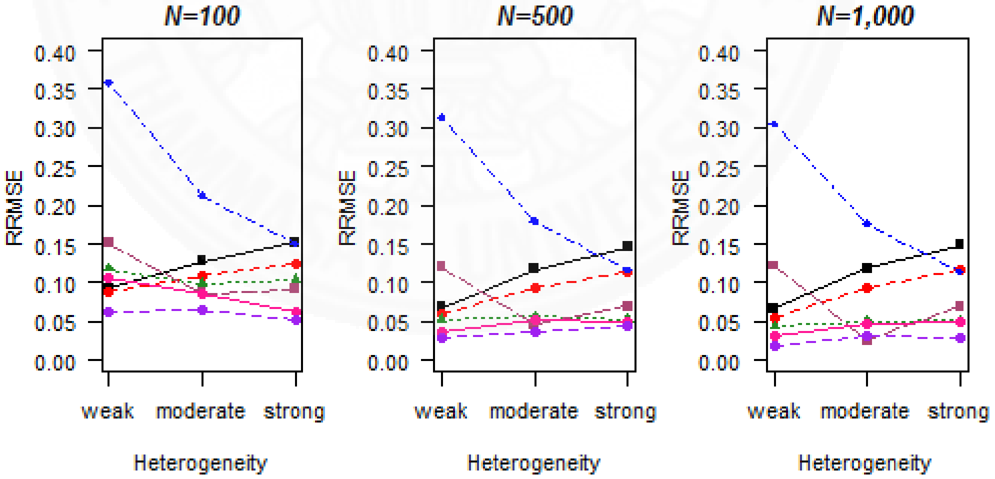


Figure 4.2: The relative root mean square error of estimators with different parameters in the two-components Poisson mixture model.

4.5.1.2 Simulation results based on the two-components Hermite mixture model

Table 4.9 presented the Rbias of estimators, which used validation information and did not used validation information. The count data were generated based on two-components Hermite distribution with $N = 100$. The results showed that $\hat{N}_{Valid}$ with 50% validation sample provided the smallest Rbias in almost all cases. $\hat{N}_{Valid}$ with 25% validation sample performed the second best in almost all cases. $\hat{N}_{PMLE}$ gave the smallest Rbias in the case of weak heterogeneity with $d = 1.8$ and moderate heterogeneity with $d = 1.2$. $\hat{N}_{Censored}$ gave the smallest Rbias in the case of strong heterogeneity with $d = 1.4$. Based on the estimators which did not used validation information, $\hat{N}_{PMLE}$ gave the smallest Rbias in the case of weak, moderate, and strong heterogeneity with $d = 1.2$ and $d = 1.8$. In addition, $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, and $\hat{N}_{Chao}$ provided severe underestimation for all cases, and Rbias increased with increasing d.

Table 4.10 presented the RRMSE of estimators for count data based on two-components Hermite distribution when $N = 100$. The results showed that $\hat{N}_{Valid}$ with 50% validation sample provided the smallest RRMSE in all cases. $\hat{N}_{Valid}$ with 25% validation sample performed the second best in all cases. Others gave quite high RRMSE, and RRMSE increased with increasing d. Furthermore, $\hat{N}_{PMLE}$ provided the smallest RRMSE compare to other estimators which did not used validation information in the case of moderate heterogeneity with $d = 1.8$ and strong heterogeneity with $d = 1.2$.

Tables 4.11 and 4.13 presented the Rbias of estimators for count data based on two-components Hermite distribution when $N = 500$ and $N = 1,000$, respectively. The results showed that $\hat{N}_{Valid}$ with 50% validation sample provided the smallest Rbias in almost all cases. $\hat{N}_{Valid}$ with 25% validation sample performed the second best. $\hat{N}_{PMLE}$ produced the smallest Rbias in the case of weak heterogeneity with $d = 1.2$. Based on strong heterogeneity with $d = 1.4$, $\hat{N}_{Censored}$ gave the smallest Rbias. $\hat{N}_{PMLE}$ and $\hat{N}_{Censored}$ gave underestimation in almost all cases. $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, and $\hat{N}_{Chao}$ provided severe underestimation for all cases, especially for $d = 1.6$ and $d = 1.8$. Based on the without validation estimators, $\hat{N}_{PMLE}$ gave the smallest Rbias in the case of weak heterogeneity.

Tables 4.12 and 4.14 presented the RRMSE of estimators for count data based on two-components Hermite distribution when $N = 500$ and $N = 1,000$, respectively. The results showed that $\hat{N}_{Valid}$ with 50% validation sample provided the smallest RRMSE in all cases. $\hat{N}_{Valid}$ with 25% validation sample performed the second best in almost all cases. $\hat{N}_{PMLE}$ provided the smallest RRMSE of all estimators which did not used validation information in the case of weak and moderate heterogeneity. $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, and $\hat{N}_{Chao}$ gave quite high RRMSE.

Simulation results indicated that $\hat{N}_{Valid}$ tends to perform reasonably well. It can be an appropriate method for estimating population size based on the two-components Hermite mixture model. The larger validation sample size, the more efficient estimation. $\hat{N}_{PMLE}$ is an appropriate estimator for population size estimation in the case of weak and moderate heterogeneity.

Table 4.9: The relative bias of estimators with different parameters in the two-components Hermite mixture model $p_i = 0.5Herm(1, d) + 0.5Herm(\nu, d)$ with 25% and 50% validation sample for $N = 100$.

ν	d	Without validation					With validation	
		MLE	Turing	Chao	Censored	PMLE	Valid	
							25%	50%
2	1.2	-0.1128	-0.0961	-0.0534	0.2360	0.0381	0.0488	0.0270
	1.4	-0.2012	-0.1868	-0.1548	0.0819	-0.0419	0.0366	0.0338
	1.6	-0.2728	-0.2647	-0.2577	-0.0565	0.0698	0.0713	0.0406
	1.8	-0.3703	-0.3832	-0.4023	-0.2833	0.0250	0.0593	0.0335
3	1.2	-0.1588	-0.1336	-0.0774	0.1119	-0.0085	0.0530	0.0406
	1.4	-0.1914	-0.1685	-0.1251	0.0492	-0.0948	0.0522	0.0355
	1.6	-0.2562	-0.2411	-0.2221	-0.0746	-0.1433	0.0565	0.0232
	1.8	-0.3086	-0.3087	-0.3169	-0.2142	-0.1420	0.0426	0.0252
4	1.2	-0.1678	-0.1383	-0.0701	0.0716	0.0470	0.0600	0.0414
	1.4	-0.2087	-0.1814	-0.1202	0.0082	-0.1380	0.0764	0.0416
	1.6	-0.2440	-0.2258	-0.2005	-0.0842	0.3334	0.0777	0.0358
	1.8	-0.3040	-0.2979	-0.2991	-0.2162	0.1412	0.0592	0.0227

Table 4.10: The relative root mean square error of estimators with different parameters in the two-components Hermite mixture model $p_i = 0.5Herm(1, d) + 0.5Herm(\nu, d)$ with 25% and 50% validation sample for $N = 100$.

ν	d	Without validation					With validation	
		MLE	Turing	Chao	Censored	PMLE	Valid	
							25%	50%
2	1.2	0.1398	0.1282	0.1262	0.2869	0.2318	0.0995	0.0611
	1.4	0.2102	0.1965	0.1741	0.1400	0.2294	0.0955	0.0621
	1.6	0.2805	0.2746	0.2739	0.1420	0.7151	0.1199	0.0783
	1.8	0.3744	0.3870	0.4058	0.2929	0.3475	0.1017	0.0698
3	1.2	0.1653	0.1428	0.1182	0.1507	0.1580	0.1003	0.0769
	1.4	0.1972	0.1761	0.1468	0.1056	0.1159	0.1080	0.0655
	1.6	0.2619	0.2477	0.2342	0.1203	0.1661	0.0956	0.0534
	1.8	0.3122	0.3126	0.3210	0.2258	0.1625	0.0982	0.0525
4	1.2	0.1724	0.1453	0.1087	0.1137	0.1023	0.0925	0.0685
	1.4	0.2153	0.1899	0.1474	0.0953	0.1505	0.1220	0.0651
	1.6	0.2486	0.2312	0.2112	0.1159	0.1993	0.1147	0.0660
	1.8	0.3083	0.3026	0.3042	0.2287	0.8122	0.0899	0.0413

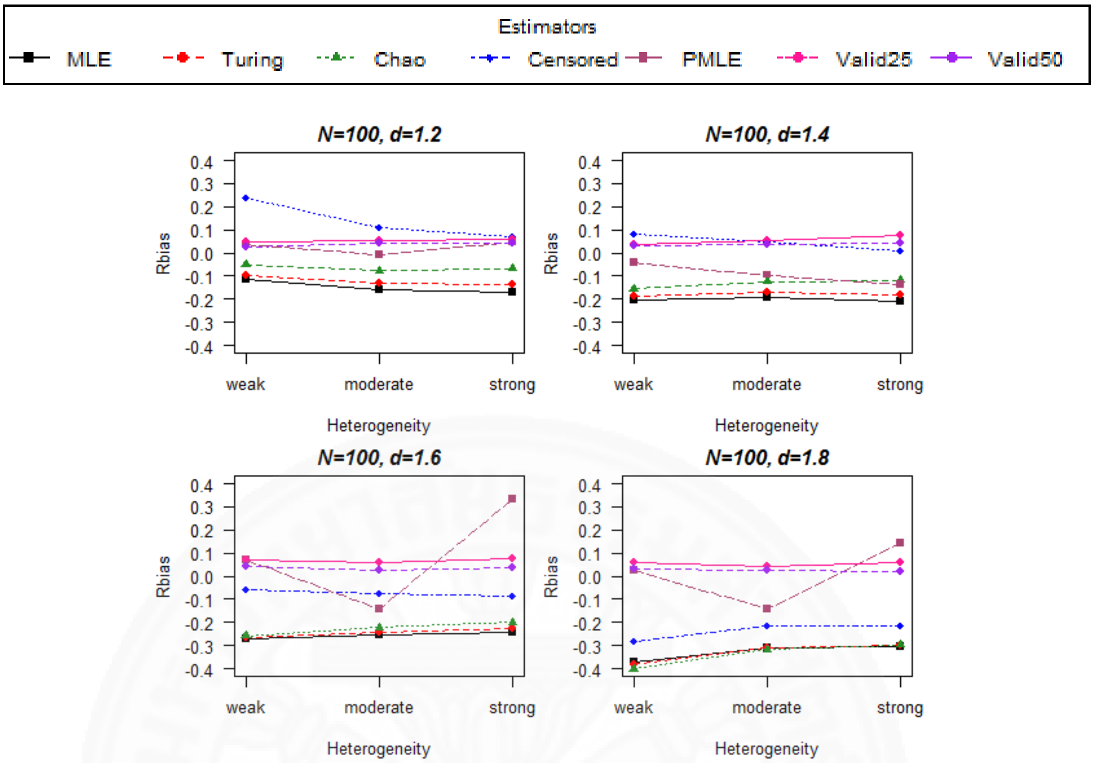


Figure 4.3: The relative bias of estimators with different parameters in the two-components Hermite mixture model for $N = 100$.

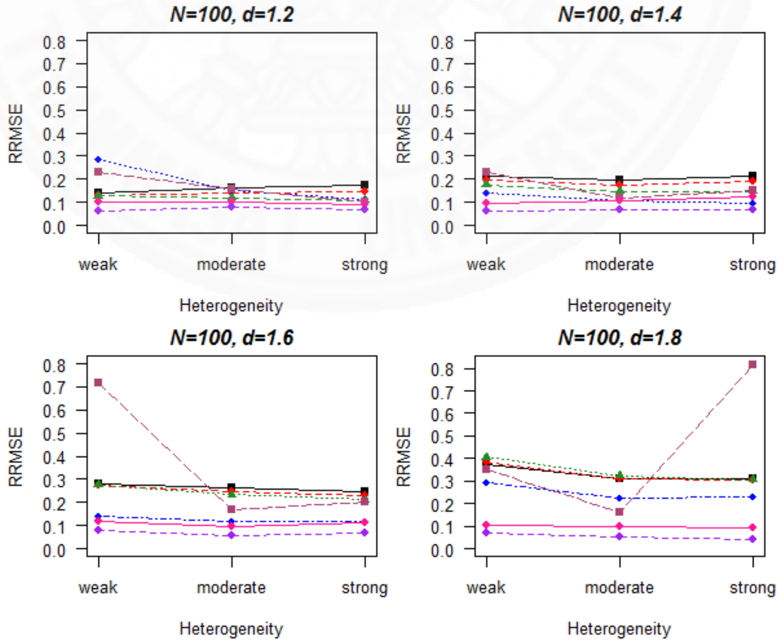


Figure 4.4: The relative root mean square error of estimators with different parameters in the two-components Hermite mixture model for $N = 100$.

Table 4.11: The relative bias of estimators with different parameters in the two-components Hermite mixture model $p_i = 0.5Herm(1, d) + 0.5Herm(\nu, d)$ with 25% and 50% validation sample for $N = 500$.

ν	d	Without validation					With validation	
		MLE	Turing	Chao	Censored	PMLE	Valid	
							25%	50%
2	1.2	-0.1329	-0.1128	-0.0767	0.2142	0.0203	0.0442	0.0283
	1.4	-0.2043	-0.1860	-0.1585	0.0903	-0.0510	0.0444	0.0251
	1.6	-0.2856	-0.2778	-0.2764	-0.0821	-0.0914	0.0267	0.0158
	1.8	-0.3662	-0.3776	-0.3979	-0.2796	-0.0823	0.0206	0.0117
3	1.2	-0.1578	-0.1303	-0.0809	0.1218	-0.0503	0.0440	0.0290
	1.4	-0.2066	-0.1820	-0.1430	0.0371	-0.1035	0.0306	0.0212
	1.6	-0.2598	-0.2449	-0.2332	-0.0825	-0.1375	0.0257	0.0150
	1.8	-0.3198	-0.3202	-0.3312	-0.2306	-0.1494	0.0274	0.0195
4	1.2	-0.1791	-0.1474	-0.0802	0.0650	-0.1018	0.0452	0.0332
	1.4	-0.2153	-0.1880	-0.1390	-0.0032	-0.1403	0.0493	0.0290
	1.6	-0.2566	-0.2376	-0.2173	-0.0972	-0.1754	0.0508	0.0267
	1.8	-0.3025	-0.2963	-0.3007	-0.2165	-0.1975	0.0485	0.0260

Table 4.12: The relative root mean square error of estimators with different parameters in the two-components Hermite mixture model $p_i = 0.5Herm(1, d) + 0.5Herm(\nu, d)$ with 25% and 50% validation sample for $N = 500$.

ν	d	Without validation					With validation	
		MLE	Turing	Chao	Censored	PMLE	Valid	
							25%	50%
2	1.2	0.1360	0.1173	0.0910	0.2237	0.0744	0.0592	0.0406
	1.4	0.2059	0.1880	0.1633	0.1040	0.0827	0.0569	0.0357
	1.6	0.2868	0.2792	0.2782	0.0940	0.1155	0.0453	0.0284
	1.8	0.3670	0.3783	0.3985	0.2811	0.1017	0.0413	0.0305
3	1.2	0.1596	0.1330	0.0902	0.1317	0.0603	0.0583	0.0391
	1.4	0.2079	0.1838	0.1480	0.0612	0.1080	0.0432	0.0313
	1.6	0.2605	0.2457	0.2344	0.0888	0.1428	0.0351	0.0225
	1.8	0.3206	0.3210	0.3319	0.2326	0.1547	0.0348	0.0289
4	1.2	0.1802	0.1491	0.0890	0.0779	0.1052	0.0568	0.0451
	1.4	0.2163	0.1893	0.1431	0.0388	0.1429	0.0647	0.0377
	1.6	0.2576	0.2388	0.2191	0.1035	0.1779	0.0630	0.0339
	1.8	0.3035	0.2973	0.3016	0.2185	0.2004	0.0564	0.0313

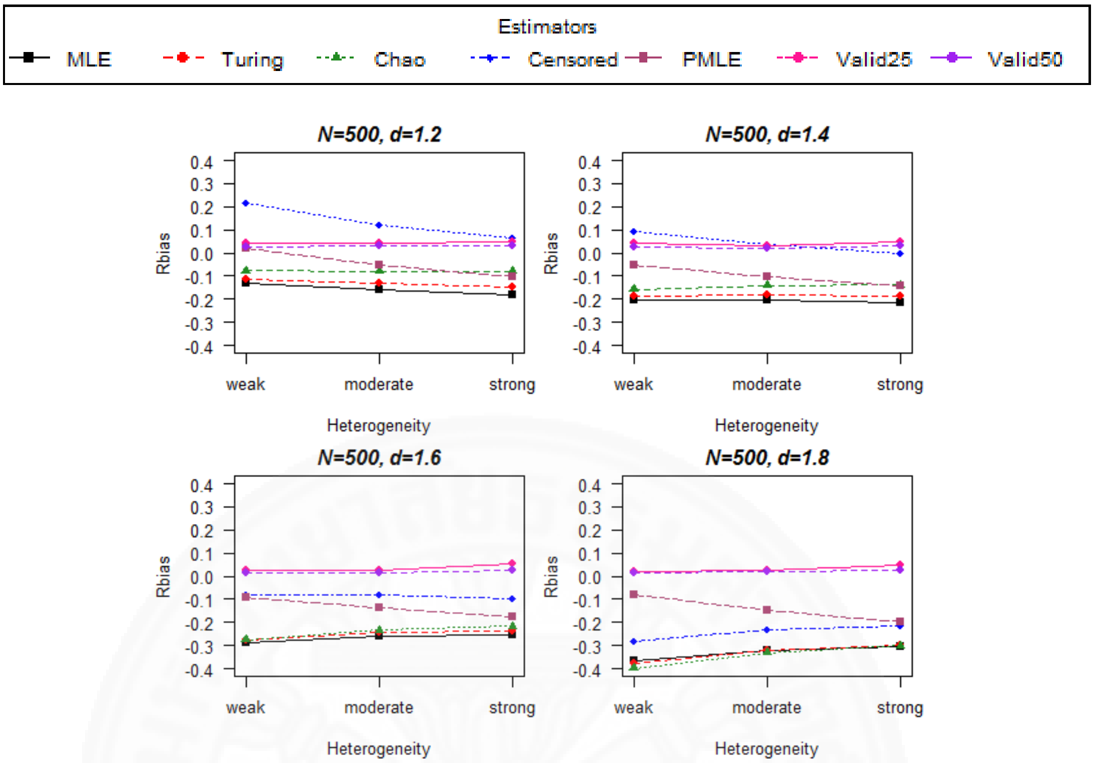


Figure 4.5: The relative bias of estimators with different parameters in the two-components Hermite mixture model for $N = 500$.

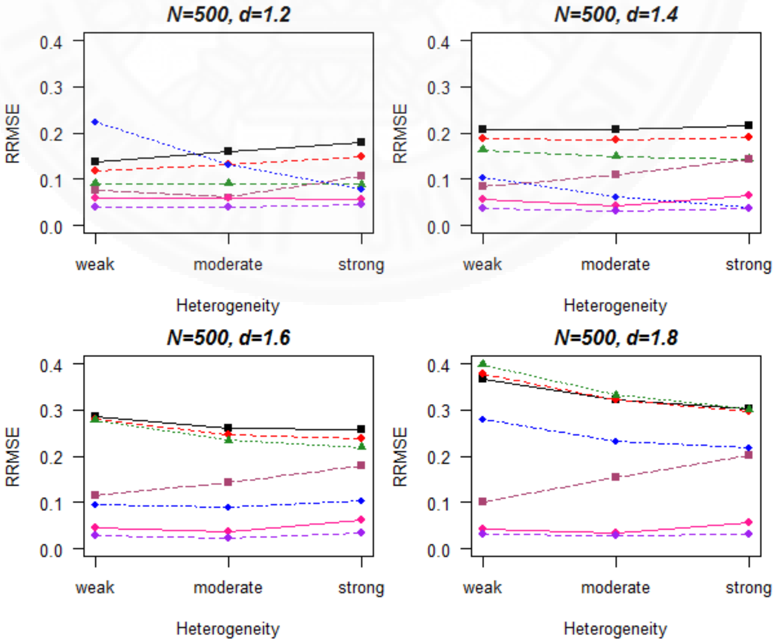


Figure 4.6: The relative root mean square error of estimators with different parameters in the two-components Hermite mixture model for $N = 500$.

Table 4.13: The relative bias of estimators with different parameters in the two-components Hermite mixture model $p_i = 0.5Herm(1, d) + 0.5Herm(\nu, d)$ with 25% and 50% validation sample for $N = 1,000$.

ν	d	Without validation					With validation	
		MLE	Turing	Chao	Censored	PMLE	Valid	
							25%	50%
2	1.2	-0.1307	-0.1100	-0.0729	0.2190	0.0136	0.0388	0.0241
	1.4	-0.2061	-0.1872	-0.1578	0.0891	-0.0671	0.0397	0.0240
	1.6	-0.2807	-0.2731	-0.2727	-0.0748	-0.0763	0.0276	0.0159
	1.8	-0.3661	-0.3783	-0.3994	-0.2811	-0.0756	0.0164	0.0091
3	1.2	-0.1609	-0.1341	-0.0878	0.1147	-0.0521	0.0377	0.0256
	1.4	-0.2069	-0.1819	-0.1427	0.0388	-0.1017	0.0268	0.0137
	1.6	-0.2551	-0.2392	-0.2260	-0.0701	-0.1323	0.0216	0.0110
	1.8	-0.3203	-0.3206	-0.3324	-0.2306	-0.1450	0.0317	0.0154
4	1.2	-0.1773	-0.1460	-0.0822	0.0656	-0.0995	0.0408	0.0265
	1.4	-0.2114	-0.1844	-0.1377	-0.0012	-0.1368	0.0423	0.0249
	1.6	-0.2538	-0.2349	-0.2151	-0.0945	-0.1747	0.0433	0.0290
	1.8	-0.3009	-0.2950	-0.3001	-0.2161	-0.1929	0.0431	0.0259

Table 4.14: The relative root mean square error of estimators with different parameters in the two-components Hermite mixture model $p_i = 0.5Herm(1, d) + 0.5Herm(\nu, d)$ with 25% and 50% validation sample for $N = 1,000$.

ν	d	Without validation					With validation	
		MLE	Turing	Chao	Censored	PMLE	Valid	
							25%	50%
2	1.2	0.1324	0.1122	0.0796	0.2233	0.0495	0.0494	0.0323
	1.4	0.2070	0.1884	0.1607	0.0987	0.0854	0.0508	0.0332
	1.6	0.2813	0.2738	0.2737	0.0815	0.0891	0.0363	0.0238
	1.8	0.3665	0.3787	0.3997	0.2821	0.0833	0.0340	0.0193
3	1.2	0.1618	0.1352	0.0910	0.1191	0.0580	0.0456	0.0325
	1.4	0.2076	0.1828	0.1452	0.0517	0.1046	0.0352	0.0190
	1.6	0.2554	0.2398	0.2271	0.0763	0.1345	0.0274	0.0146
	1.8	0.3207	0.3210	0.3328	0.2316	0.1475	0.0376	0.0193
4	1.2	0.1779	0.1467	0.0854	0.0704	0.1009	0.0469	0.0309
	1.4	0.2118	0.1850	0.1394	0.0251	0.1378	0.0509	0.0308
	1.6	0.2542	0.2354	0.2159	0.0975	0.1758	0.0500	0.0343
	1.8	0.3013	0.2954	0.3005	0.2171	0.1938	0.0470	0.0288

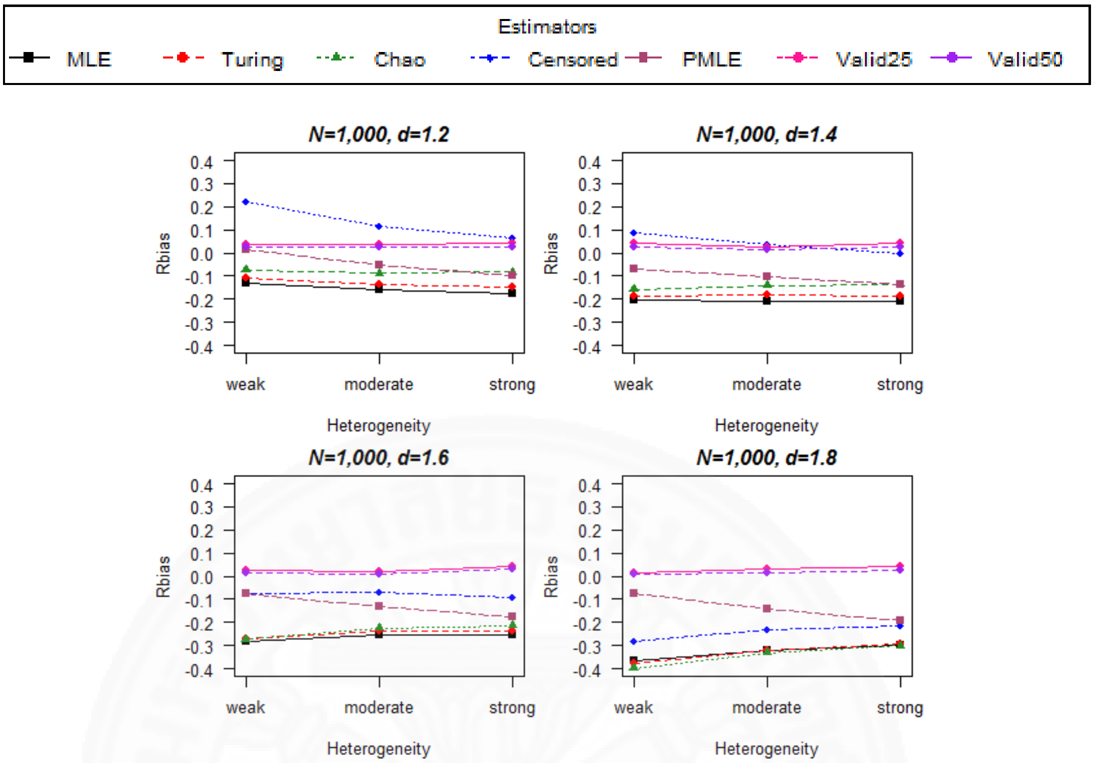


Figure 4.7: The relative bias of estimators with different parameters in the two-components Hermite mixture model for $N = 1,000$.

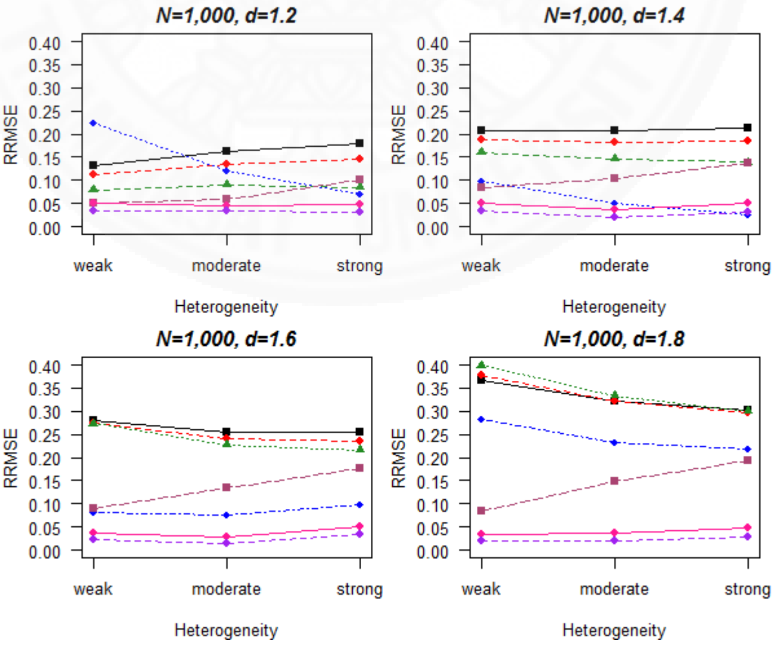


Figure 4.8: The relative root mean square error of estimators with different parameters in the two-components Hermite mixture model for $N = 1,000$.

4.5.2 Interval estimation part

4.5.2.1 Simulation results based on the two-components Poisson mixture model

The coverage probabilities of the 95% CIs based on the two-components Poisson mixture model were presented in Table 4.15. A comparison of the estimators showed that $\hat{N}_{Censored}$ provided severe overestimation for all cases, especially for weak heterogeneity. $\hat{N}_{MLE}$ and $\hat{N}_{Turing}$ provided severe underestimation for all cases, especially for strong heterogeneity. $\hat{N}_{Chao}$ provided underestimation for all cases, especially for strong heterogeneity with $N = 1,000$.

The CI ProfileValid with 25% and 50% validation information gave the highest CP for all cases. The average lengths of CI ProfileValid with 50% validation information were shorter than those of 25% validation information for all cases. CI ProfileMix performed the second best for weak heterogeneity. CI ProfileMix produced low CP in strong heterogeneity for $N = 1,000$. CIs of $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, $\hat{N}_{Chao}$, and $\hat{N}_{Censored}$ provided low CP for all cases. CIs of $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, and $\hat{N}_{Censored}$ did not cover the true population size in weak and strong heterogeneity for $N = 1,000$.

Simulation results indicated that CI ProfileValid tends to perform reasonably well. It can be an appropriate method for estimating CI of population size N based on the two-components Poisson mixture model. The larger validation sample size, the more efficient estimation. In addition, CI ProfileMix is an appropriate method for estimating CI of population size N in the case of weak heterogeneity.

Table 4.15: Comparison of various estimators and 95% CI of N in the two-components Poisson mixture model $p_i = 0.5Poi(1) + 0.5Poi(\lambda)$ with 25% and 50% validation sample.

λ		Without validation					With validation	
		MLE	Turing	Chao	Censored	ProfileMix	ProfileValid	
							25%	50%
	N=100							
2	Average $\hat{N}$	95.25	96.15	98.97	132.34	101.48	99.66	99.54
	Average $\widehat{se}(\hat{N})$	4.47	3.76	10.33	16.74	-	-	-
	Coverage probability	0.67	0.61	0.88	0.53	0.99	1.00	1.00
	Average length	17.51	14.75	40.50	65.60	40.87	38.12	31.87
4	Average $\hat{N}$	85.57	88.71	96.55	111.06	92.27	97.26	97.87
	Average $\widehat{se}(\hat{N})$	2.13	2.68	8.29	9.89	-	-	-
	Coverage probability	0.04	0.15	0.76	0.88	0.84	1.00	1.00
	Average length	8.34	10.50	32.49	38.75	26.36	32.44	27.40
	N=1,000							
2	Average $\hat{N}$	938.11	949.25	970.29	1,301.22	1,053.20	998.56	998.68
	Average $\widehat{se}(\hat{N})$	13.76	11.82	30.71	51.35	-	-	-
	Coverage probability	0.06	0.15	0.80	0.00	0.96	1.00	1.00
	Average length	53.93	46.35	120.36	201.29	201.78	140.67	114.65
4	Average $\hat{N}$	852.53	885.15	954.71	1,110.35	931.94	979.58	984.55
	Average $\widehat{se}(\hat{N})$	6.73	8.58	24.97	31.47	-	-	-
	Coverage probability	0.00	0.00	0.51	0.01	0.11	1.00	1.00
	Average length	26.36	33.62	97.87	123.37	75.25	94.58	83.76

4.5.2.2 Simulation results based on the two-components Hermite mixture model

The coverage probabilities of the 95% CIs for $N = 100$ and $N = 1,000$ based on the Hermite distribution were presented in Tables 4.16 and 4.17, respectively. A comparison of the estimators showed that $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, and $\hat{N}_{Chao}$ provided severe underestimation for all case, especially for large d in all population size. $\hat{N}_{Censored}$ provided overestimation in weak heterogeneity for $d = 1.2$, and gave severe underestimation in weak and strong heterogeneity for $d = 1.8$ in all population size . $\hat{N}_{ProfileMix}$ produced severe under estimation for strong heterogeneity.

For $N = 100$, CI ProfileValid with 25% and 50% validation information gave the highest CP for all cases. The average lengths of CI ProfileValid with 50% validation information were shorter than those of 25% validation information for all cases. CI ProfileMix performed the second best for weak heterogeneity, and ALs of CI ProfileMix were slightly larger than those of CI ProfileValid. On the other hand, CI ProfileMix produced low CP in strong heterogeneity. CIs of $\hat{N}_{MLE}$ and $\hat{N}_{Turing}$ provided low CP for all cases, and did not cover the true population size in almost all cases. CPs of CI of $\hat{N}_{Chao}$ were low for all cases. CPs of CI of $\hat{N}_{Censored}$ were close to the nominal level in weak and strong heterogeneity for $d = 1.2$ and 1.4, but they produced the largest ALs compare to the proposed CIs.

For $N = 1,000$, CI ProfileValid with 25% and 50% validation information gave the highest CP in almost all cases. The average lengths of CI ProfileValid with 50% validation information were shorter than those of 25% validation information for all cases. CI ProfileMix performed the second best for weak heterogeneity, but ALs of CI ProfileMix were larger than those of CI ProfileValid. On the other hand, CI ProfileMix did not cover the true population size in strong heterogeneity. CIs of $\hat{N}_{MLE}$, $\hat{N}_{Turing}$, and $\hat{N}_{Chao}$ did not cover the true population size for all cases. CPs of CI of $\hat{N}_{Censored}$ were low in almost all cases. CI of $\hat{N}_{Censored}$ gave the highest CP in strong heterogeneity for $d = 1.4$, but it produced the largest ALs compare to the proposed CIs.

Simulation results indicated that CI ProfileValid tends to perform reasonably well. It can be an appropriate method for estimating CI of population size N based on the two-components Hermite mixture model. The larger validation sample size, the more efficient estimation. In addition, CI ProfileMix is an appropriate method for estimating CI of population size N in the case of weak heterogeneity.

Table 4.16: Comparison of various estimators and 95% CI of N for $N = 100$ in the two-components Hermite mixture model $p_i = 0.5Herm(1, d) + 0.5Herm(\nu, d)$ with 25% and 50% validation sample .

d		Without validation					With validation	
		MLE	Turing	Chao	Censored	ProfileMix	ProfileValid	
							25%	50%
				$\nu = 2$				
1.2	Average $\hat{N}$	87.96	89.75	94.18	122.91	97.23	99.21	99.93
	Average $\widehat{se}(\hat{N})$	3.99	3.55	10.08	15.41	-	-	-
	Coverage probability	0.32	0.35	0.83	0.78	1.00	1.00	1.00
	Average length	15.62	13.91	39.53	60.42	40.82	38.16	36.67
1.4	Average $\hat{N}$	80.12	81.88	85.60	109.70	91.23	97.97	98.78
	Average $\widehat{se}(\hat{N})$	3.48	3.26	8.63	13.32	-	-	-
	Coverage probability	0.04	0.06	0.52	0.95	0.99	0.98	1.00
	Average length	13.64	12.78	33.84	52.21	39.83	39.41	35.54
1.6	Average $\hat{N}$	72.01	72.81	73.79	92.96	85.10	97.64	98.18
	Average $\widehat{se}(\hat{N})$	2.99	2.87	5.98	10.47	-	-	-
	Coverage probability	0.00	0.00	0.13	0.77	0.86	0.98	0.99
	Average length	11.71	11.26	23.43	41.03	41.45	35.12	32.85
1.8	Average $\hat{N}$	64.14	62.97	61.13	73.22	90.72	101.00	100.03
	Average $\widehat{se}(\hat{N})$	2.48	2.24	2.54	6.60	-	-	-
	Coverage probability	0.00	0.00	0.00	0.07	0.86	0.94	0.95
	Average length	9.74	8.78	9.95	25.86	36.64	30.77	28.32
				$\nu = 4$				
1.2	Average $\hat{N}$	82.49	85.59	93.76	106.73	88.87	94.89	95.58
	Average $\widehat{se}(\hat{N})$	1.96	2.56	8.30	9.48	-	-	-
	Coverage probability	0.00	0.04	0.69	0.97	0.66	1.00	1.00
	Average length	7.70	10.05	32.55	37.18	22.39	34.14	30.40
1.4	Average $\hat{N}$	78.74	81.38	87.28	99.72	85.15	93.23	95.07
	Average $\widehat{se}(\hat{N})$	1.80	2.39	6.87	8.61	-	-	-
	Coverage probability	0.00	0.01	0.49	0.95	0.45	0.98	0.98
	Average length	7.04	9.37	26.93	33.75	21.19	36.42	33.22
1.6	Average $\hat{N}$	74.79	76.60	79.22	90.50	81.91	93.02	95.25
	Average $\widehat{se}(\hat{N})$	1.62	2.12	4.68	7.18	-	-	-
	Coverage probability	0.00	0.00	0.14	0.66	0.34	0.96	0.99
	Average length	6.34	8.31	18.35	28.14	19.62	38.54	34.5
1.8	Average $\hat{N}$	70.63	71.28	71.13	79.74	80.41	96.55	97.56
	Average $\widehat{se}(\hat{N})$	1.43	1.70	2.26	5.25	-	-	-
	Coverage probability	0.00	0.00	0.01	0.14	0.24	1.00	0.98
	Average length	5.61	6.66	8.84	20.59	19.60	37.78	35.2

Table 4.17: Comparison of various estimators and 95% CI of N for $N = 1,000$ in the two-components Hermite mixture model $p_i = 0.5Herm(1, d) + 0.5Herm(\nu, d)$ with 25% and 50% validation sample .

d		Without validation					With validation	
		MLE	Turing	Chao	Censored	ProfileMix	ProfileValid	
							25%	50%
				$\nu = 2$				
1.2	Average $\hat{N}$	878.67	889.88	923.02	1,215.18	1,012.69	1,002.06	1,000.51
	Average $\widehat{se}(\hat{N})$	12.58	11.22	29.79	47.81	-	-	-
	Coverage probability	0.00	0.00	0.26	0.00	1.00	1.00	1.00
	Average length	49.32	43.97	116.79	187.41	266.37	161.56	131.23
1.4	Average $\hat{N}$	792.52	810.02	837.07	1,081.42	932.82	987.06	990.57
	Average $\widehat{se}(\hat{N})$	10.93	10.32	25.57	41.41	-	-	-
	Coverage probability	0.00	0.00	0.01	0.57	1.00	1.00	1.00
	Average length	42.86	40.46	100.22	162.31	236.86	165.10	137.00
1.6	Average $\hat{N}$	713.80	719.58	717.46	908.74	919.77	980.42	986.77
	Average $\widehat{se}(\hat{N})$	9.35	9.03	16.67	31.72	-	-	-
	Coverage probability	0.00	0.00	0.00	0.20	0.85	1.00	1.00
	Average length	36.66	35.38	65.33	124.34	189.11	172.10	147.94
1.8	Average $\hat{N}$	634.44	621.73	600.84	715.74	925.68	994.26	995.86
	Average $\widehat{se}(\hat{N})$	7.70	6.95	7.05	19.82	-	-	-
	Coverage probability	0.00	0.00	0.00	0.00	0.72	1.00	1.00
	Average length	30.18	27.24	27.62	77.71	156.42	172.71	156.29
				$\nu = 4$				
1.2	Average $\hat{N}$	822.20	853.03	914.33	1,061.96	898.77	969.44	976.27
	Average $\widehat{se}(\hat{N})$	6.27	8.20	23.10	29.87	-	-	-
	Coverage probability	0.00	0.00	0.07	0.42	0.00	0.93	0.94
	Average length	24.58	32.14	90.54	117.10	72.49	94.77	88.47
1.4	Average $\hat{N}$	787.48	814.12	860.32	996.09	860.99	954.30	964.27
	Average $\widehat{se}(\hat{N})$	5.73	7.64	19.78	27.19	-	-	-
	Coverage probability	0.00	0.00	0.00	0.96	0.00	0.67	0.70
	Average length	22.45	29.95	77.55	106.57	66.97	96.90	85.60
1.6	Average $\hat{N}$	747.97	766.39	784.92	904.94	827.63	945.80	958.60
	Average $\widehat{se}(\hat{N})$	5.12	6.77	13.59	22.74	-	-	-
	Coverage probability	0.00	0.00	0.00	0.02	0.00	0.72	0.72
	Average length	20.05	26.53	53.25	89.16	68.67	113.31	95.17
1.8	Average $\hat{N}$	701.64	707.13	701.75	784.53	810.35	973.93	982.03
	Average $\widehat{se}(\hat{N})$	4.42	5.25	6.11	15.90	-	-	-
	Coverage probability	0.00	0.00	0.00	0.00	0.00	1.00	1.00
	Average length	17.32	20.60	23.96	62.32	64.58	159.84	133.54

4.6 Conclusion

Since mixture model is a flexible approach to cope with data from a population which contain contaminated parts, it has been widely used in many fields. With this motivation, the discrete mixtures of Poisson-Normal or Hermite distributions is adopted to model the heterogeneity of an unobserved population. The results for estimating the parameters of zero-truncated count mixtures of Hermite distributions implying a unique estimator for the population size N are proposed in this study.

The penalized maximum likelihood estimator, $\hat{N}_{PMLE}$, based on discrete mixtures of Hermite distributions is proposed. The simulation study was used to consider the performance of the proposed estimator under two-components Poisson and two-components Hermite mixture models. Based on two-components Poisson mixture model, $\hat{N}_{PMLE}$ is an appropriate estimator for population size estimation for moderate heterogeneity. Moreover, $\hat{N}_{PMLE}$ is an appropriate estimator for population size estimation in the case of weak and moderate heterogeneity in two-components Hermite mixture model.

The development of estimation to include validation information in the capture-recapture model is proposed in this research, to increase the accuracy and efficiency of population size estimation. The simulation study was used to consider the performance of the proposed estimator, $\hat{N}_{Valid}$, under two-components Poisson and two-components Hermite mixture models. In the case of two-components Poisson mixture model, $\hat{N}_{Valid}$ can be an appropriate method for estimate the population size when $N \geqslant 500$ in weak and strong heterogeneity. It is tends to perform reasonably well, and is the best method for estimating population size based on the two-components Hermite mixture model. The simulation results revealed that the larger validation sample size, the more efficient estimation.

The profile mixture likelihood is used to construct the confidence intervals for the population size N. Based on the two-components Poisson and two-components Hermite mixture models, CI ProfileMix is an appropriate method for estimating CI of population size N in the case of weak heterogeneity.

In addition, the profile mixture likelihood which include validation information is proposed. Simulation results indicated that CI ProfileValid tends to perform reasonably well. It can be an appropriate method for estimating CI of population size N based on the two-components Poisson and two-components Hermite mixture models. The larger validation sample size, the more efficient estimation.

CHAPTER 5

REAL DATA EXAMPLES

Real data example was applied in this chapter. The proposed estimators were used to estimate the number of heroin users in Chiang Mai, Thailand (Sanfan, Chan-ngam, & Chaiarporn, 2019). The number of heroin users that contacts the treatment centers were presented in Table 5.1. Data of heroin users were collected by a hospital and a health treatment center in Chiang Mai from 1 October 2016 to 30 September 2017. The observed frequencies were $f_1 = 316$, $f_2 = 57$, $f_3 = 12$, $f_4 = 2$, $f_5 = 0$, $f_6 = 2$, and the observed number of heroin users $n = 389$. Using the graphical approach to identify a distribution, the frequencies distribution and the ratio plot with weighted regression line of heroin users in Chiang Mai were shown in Figures 5.1 and 5.2, respectively. Since the ratio plot showed a straight line with a positive slope, the heterogeneous Poisson model seemed to appropriate for this data.

Table 5.1: Frequencies of heroin users in Chiang Mai.

i	1	2	3	4	5	6
f_i	316	57	12	2	0	2

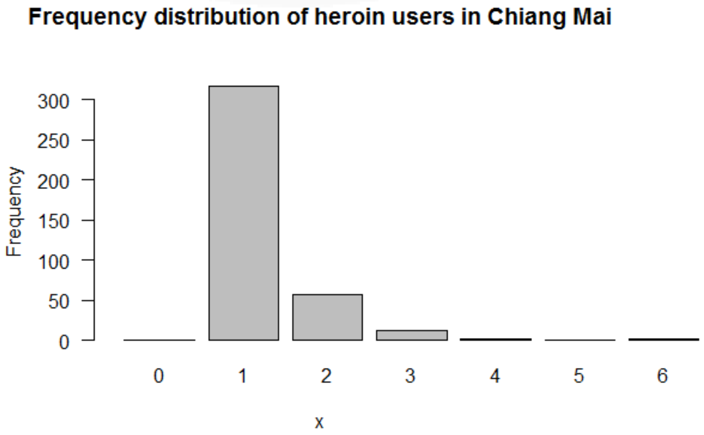


Figure 5.1: Frequency distribution of heroin users in Chiang Mai.

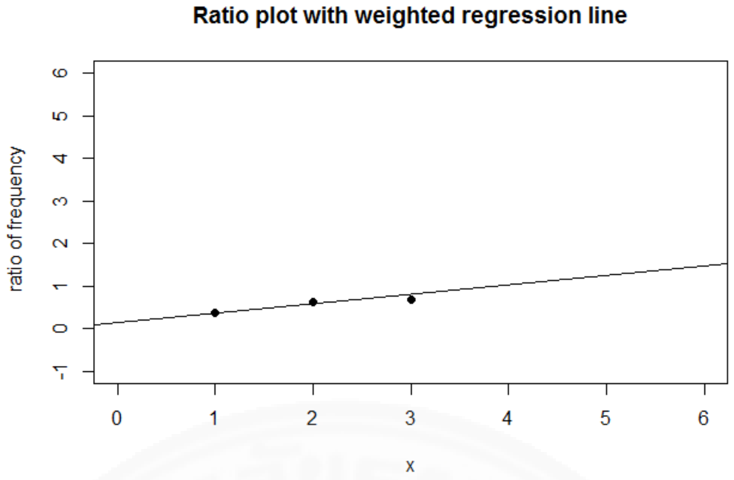


Figure 5.2: Ratio plot with weighted regression line of heroin users in Chiang Mai.

The proposed and other well-known estimators were used to estimate the number of heroin users in Chiang Mai. Results were showed in Table 5.2. Since the ratio plot revealed that the model of heterogeneity was appropriate than the Poisson model, $\hat{N}_{MLE}$ and $\hat{N}_{Turing}$ were not valid. $\hat{N}_{Chao}$, $\hat{N}_{Censored}$, $\hat{N}_{Herm}$, and $\hat{N}_{PMLE}$ were appropriate for estimating the number of heroin users in Chiang Mai.

From Table 5.2, $\hat{N}_{MLE}$ provided the lowest estimate 1,049 with 95% of CI (999, 1,099). $\hat{N}_{Turing}$ provided a low estimate 1,112 with 95% of CI (1,090, 1,134). The $\hat{N}_{Censored}$ gave the largest estimate 2,073 with 95% of CI (1,606, 2,540). The number of heroin users that estimated by $\hat{N}_{Herm}$ were 1,390. The Chao lower bound estimator, $\hat{N}_{Chao}$, estimated 1,265 which smaller than $\hat{N}_{Herm}$. Analysis of $\hat{N}_{PMLE}$ was presented in Table 5.3. The results showed that one component of the zero-truncated Hermite model was the best fit with the highest log-likelihood and smallest BIC. The estimated number of heroin users was 1,324 which close to $\hat{N}_{Herm}$. Analysis of $\hat{N}_{ProfileMix}$ was presented in Table 5.4. The results showed that one component of the zero-truncated Hermite model was the best fit. The profileMix provided the estimate 1,388 for the number of heroin users with 95% of CI (994, 1,889).

Table 5.5, compared frequency distribution of heroin users among the observed counts, zero-truncated Poisson, and zero-truncated Hermite with $\nu = 0.35$ and $\nu = 0.37$. The differences between the observed frequencies and fitted frequencies of the zero-truncated Hermite distributions are minor. This was supported by Chi-square goodness-of-fit test with $\chi^2 = 4.44$ ($p - value = 0.0351$) and $\chi^2 = 4.41$ ($p - value = 0.0357$), df=1. At a 0.025 level of significance, there was sufficient evidence to conclude that the distribution of heroin users was not different from the zero-truncated Hermite distribution with parameters $\nu = 0.35$ and $\nu = 0.37$ for $d = 1.12$. The homogeneous Poisson model gave the poorest goodness-of-fit compared with the zero-truncated Hermite distribution. In addition, Figure 5.3 supported that the zero-truncated Hermite distribution fitted than the zero-truncated Poisson distribution.

Table 5.2: Estimated total number of heroin users in Chiang Mai.

Estimator	$\hat{N}$	$\hat{Se}(\hat{N})$	95% Confidence Interval	
			Approximate normal	Profile likelihood
MLE ($\hat{\lambda}$ =0.4634)	1,049	25.69	(999, 1,099)	-
Turing	1,112	11.41	(1,090, 1,134)	-
Chao	1,265	146.53	(978, 1,552)	-
Censored	2,073	238.30	(1,606, 2,540)	-
Herm ($\hat{\nu}$ =0.35, $\hat{d}$ =1.12)	1,390	-	-	-
PMLE ($\hat{\nu}$ =0.37, $\hat{d}$ =1.12)	1,324	-	-	-
ProfileMix	1,388	-	-	(994, 1,889)

Table 5.3: $\hat{N}_{PMLE}$ analysis for heroin users in Chiang Mai.

s	$\hat{N}_{PMLE}$	$\hat{q}_k$	$\hat{\nu}_k$	$\hat{d}$	log$L(\hat{Q}_k)$	BIC
1	1,324	1.0000	0.3681	1.1099	-246.5136	504.9544
2	7,225	0.5688	0.0002	1.2958	-251.9496	527.7535
		0.4312	0.1561			
3	12,868	0.3837	0.0000	1.3092	-253.3989	542.5793
		0.3428	0.0000			
		0.2736	0.1382			

Table 5.4: $\hat{N}_{ProfileMix}$ analysis for heroin users in Chiang Mai.

s	$\hat{N}_{ProfileMix}$	$\hat{q}_k$	$\hat{\nu}_k$	$\hat{d}$	$\log L(\hat{Q}_k)$	BIC
1	1,388	1.0000	0.3501	1.1223	-19.8613	51.6498
2	2,389	0.1753	0.0000	1.2226	332.9126	-641.9710
		0.8247	0.2466			
3	2,389	0.4108	0.1480	1.2226	337.9245	-640.0675
		0.2618	0.2831			
		0.3274	0.2094			

Table 5.5: Observed data and fitted frequencies based on zero-truncated Poisson and zero-truncated Hermite of heroin users in Chiang Mai.

x	Obseved data	$\hat{f}_x$ (MLE)	$\hat{f}_x$ (Herm)	$\hat{f}_x$ (PMLE)
1	316	306	307	306
2	57	71	69	69
3	12	11	11	12
4	4	1	2	2
	χ^2	12.18	4.44	4.41

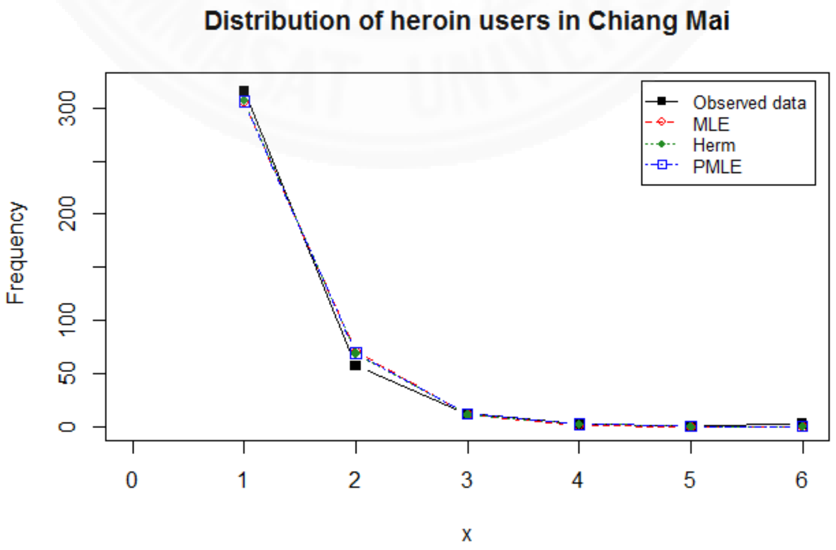


Figure 5.3: Frequency distribution of heroin users among the observed counts, zero-truncated Poisson, and zero-truncated Hermite .

Addition information on the observed of heroin users was available from another observational period. Data of heroin users were collected by a hospital and a health treatment center in Chiang Mai from 1 October 2017 to 30 September 2018. Data were checked for duplicates and matched with the cases that identified from 1 October 2016 to 30 September 2017. The number of heroin users that contacts the treatment centers were presented in Table 5.6. The observed frequencies were $g_0 = 304$, $g_1 = 59$, $g_2 = 19$, $g_3 = 2$, $g_4 = 2$, $g_5 = 2$, $g_6 = 1$. Using the graphical approach to identify a distribution, the frequency distribution of validation data and the ratio plots of positive and validation sample with weighted regression lines were shown in Figures 5.4 and 5.5, respectively. Both ratio plots were similar, and showed a straight line with a positive slope. Therefore, the heterogeneous Poisson model seemed to appropriate for this data.

Table 5.6: Frequencies in positive sample and validation sample of heroin users in Chiang Mai.

x	0	1	2	3	4	5	6
f_x	-	316	57	12	2	0	2
g_x	304	59	19	2	2	2	1

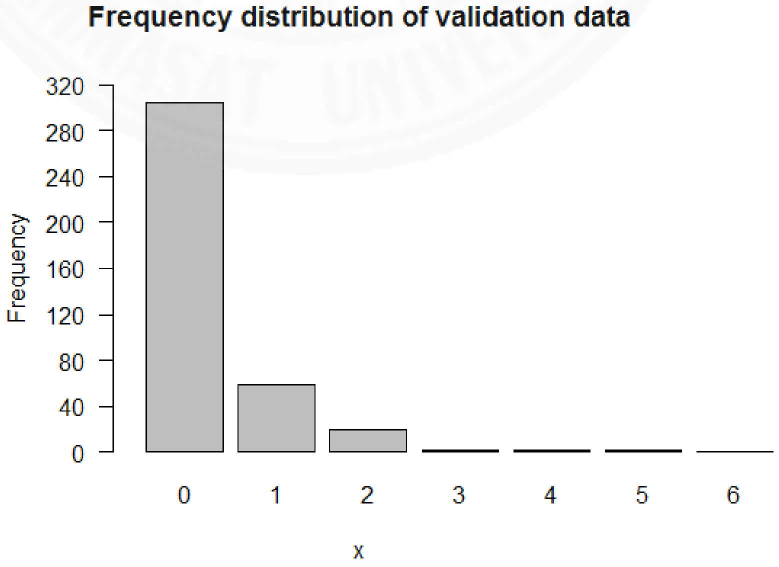


Figure 5.4: Frequency distribution of the validation data.

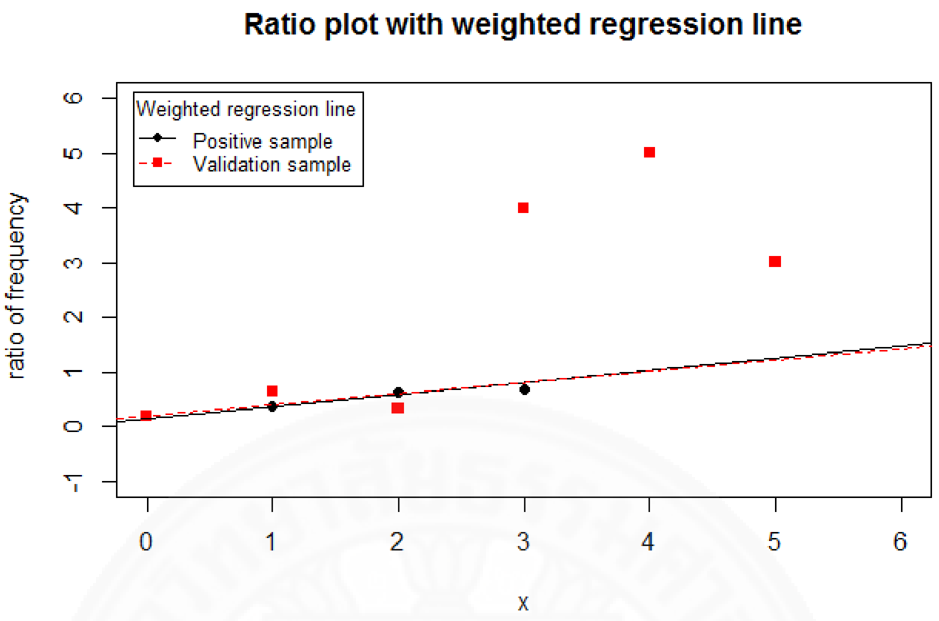


Figure 5.5: Ratio plot of positive and validation sample with weighted regression line of heroin users in Chiang Mai

The proposed estimators were used to estimate the number of heroin users in Chiang Mai. Results were showed in Table 5.7. Analysis of $\hat{N}_{Valid}$ was presented in Table 5.8. The results showed that one component of the zero-truncated Hermite model was the best fit with the highest log-likelihood and smallest BIC. The estimated number of heroin users was 1,664. Analysis of $\hat{N}_{ProfileValid}$ was presented in Table 5.9. The results showed that two component mixture of the zero-truncated Hermite model with the mixing distribution $Q = \begin{pmatrix} 0.2175 & 0.4121 \\ 1.2189 & 1.2189 \\ 0.6003 & 0.3997 \end{pmatrix}$ was the best fit. The ProfileValid provided the estimate 1,687 for the number of heroin users with 95% of CI (1,405, 2,044).

Table 5.10 and Figure 5.6, compared frequency distribution of heroin users among the observed counts, zero-truncated Poisson, and zero-truncated Hermite with and without validation information. Chi-square goodness-of-fit test of zero-truncated Hermite with validation information was $\chi^2 = 7.84$ ($p - value = 0.00511$), df=1. At a 0.005 level of significance, there was sufficient evidence to conclude that the distribution of heroin users was not different from the zero-truncated Hermite distribution with parameters $\nu = 0.30$ and $d = 1.22$. Comparing to the previous $\hat{N}_{PMLE}$ and $\hat{N}_{Herm}$, the predicted frequencies from $\hat{N}_{PMLE}$ provided the best fit to the observed data. $\hat{N}_{Herm}$ performed the second best.

Table 5.7: Estimated total number of heroin users in Chiang Mai, Thailand.

Estimator	$\hat{N}$	$\hat{Se}(\hat{N})$	95% Confidence Interval	
			Approximate normal	Profile likelihood
MLE ($\hat{\lambda}$ =0.4634)	1,049	25.69	(999, 1,099)	-
Turing	1,112	11.41	(1,090, 1,134)	-
Chao	1,265	146.53	(978, 1,552)	-
Censored	2,073	238.30	(1,606, 2,540)	-
Herm ($\hat{\nu}$ =0.35, $\hat{d}$ =1.12)	1,390	-	-	-
PMLE($\hat{\nu}$ =0.37, $\hat{d}$ =1.12)	1,324	-	-	-
Valid ($\hat{\nu}$ =0.30, $\hat{d}$ =1.22)	1,664	-	-	-
ProfileMix	1,388	-	-	(994, 1,889)
ProfileValid	1,687	-	-	(1,405, 2,044)

Table 5.8: $\hat{N}_{Valid}$ analysis for heroin users in Chiang Mai, Thailand.

s	$\hat{N}_{Valid}$	$\hat{q}_k$	$\hat{\nu}_k$	$\hat{d}$	log$L(\hat{Q}_k)$	BIC
1	1,664	1.0000	0.2985	1.2171	-912.5820	1,837.0910
2	1,810	0.1306	0.0031	1.2293	-937.9518	1,899.7580
		0.8694	0.3201			
3	1,810	0.1306	0.0031	1.2293	-940.2486	1,916.2790
		0.2618	0.2831			
		0.3274	0.2094			

Table 5.9: $\hat{N}_{ProfileValid}$ analysis for heroin users in Chiang Mai, Thailand.

s	$\hat{N}_{ProfileValid}$	$\hat{q}_k$	$\hat{\nu}_k$	$\hat{d}$	log$L(\hat{Q}_k)$	BIC
1	1,664	1.000	0.2985	1.2171	-1,977.6440	3,967.2150
2	1,687	0.6003	0.2175	1.2189	-1.976.2860	3,976.4250
		0.3997	0.4121			
3	1,678	0.4034	0.2146	1.2181	-1,976.6499	3,989.0810
		0.2683	0.4122			
		0.3283	0.4122			

Table 5.10: Observed data and fitted frequencies based on zero-truncated Poisson and zero-truncated Hermite with and without validation information.

x	Obseved data	$\hat{f}_x$ (MLE)	$\hat{f}_x$ (Herm)	$\hat{f}_x$ (PMLE)	$\hat{f}_x$ (Valid)
1	316	306	307	306	298
2	57	71	69	69	76
3	12	11	11	12	12
4	4	1	2	2	2
	χ^2	12.18	4.44	4.41	7.84

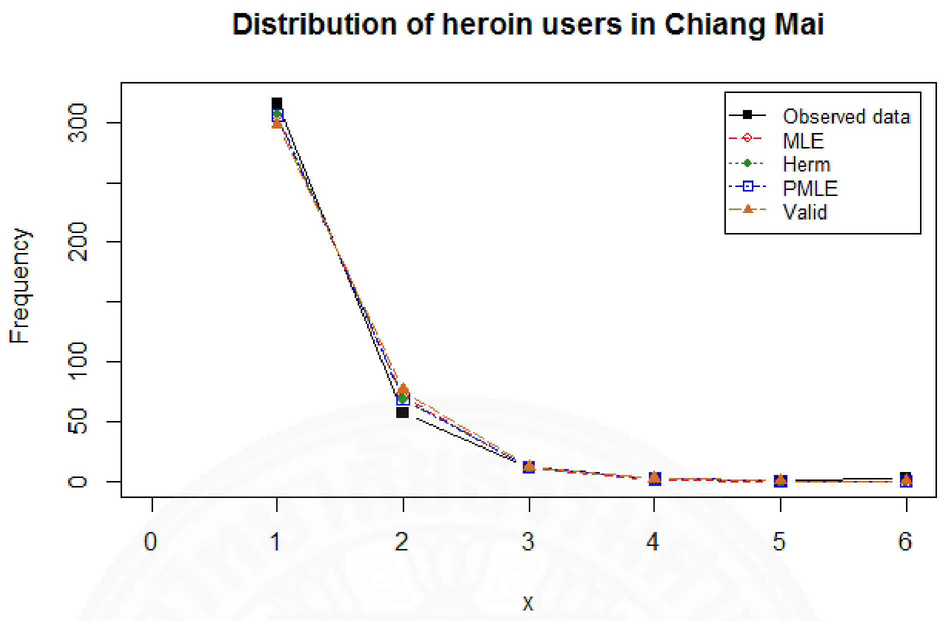


Figure 5.6: Observed frequencies and fitted frequencies of heroin users in Chiang Mai, Thailand.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

Capture-recapture is a powerful method for estimating the size of an elusive target population. Capture-recapture sampling provide the frequency count data that observed during the observational period. The basic model for the count data is homogeneous Poisson model, but it rarely occurs in real situations. It is more reasonable to assume that the population may consist of a set of subgroups. Heterogeneous Poisson model might be more realistic. The negative binomial distribution have been commonly used as a model of capture-recapture data. The failure of a dispersion parameter estimation in negative binomial distribution which results in a spurious estimate for the population size N have been demonstrated in many studies. Therefore, the Poisson-Normal distribution is proposed in this study.

The Laplace's method is applied to approximate the Poisson-Normal distribution. The resulting distribution is called Hermite distribution which has two parameters ν and d. It is also flexible in terms of its ability to allow a moderate over-dispersion in the data. The EM algorithm is used to estimate the MLE of two parameters. A new estimator, $\hat{N}_{Herm}$, is proposed based on zero-truncated Hermite distribution through the Horvitz-Thomson approach. The simulation results reveal that $\hat{N}_{Herm}$ is an asymptotic estimator under Poisson and Hermite distributions. For Hermite distribution, $\hat{N}_{Herm}$ works very well compared with others. Its performance is close to the $\hat{N}_{MLE}$ and $\hat{N}_{Turing}$ for $N \geqslant 500$ and $\lambda \geqslant 2$ under Poisson distribution. $\hat{N}_{Herm}$ is an efficient estimator for estimating the population size N based on negative binomial distribution when $N \geqslant 500$. In addition, the profile likelihood is used to construct the confidence intervals for the population size N. The proposed CI is an appropriate choice for estimating the CI of population size based on Poisson distribution, and is the best choice for Hermite distribution.

Mixture model is a flexible approach to cope with data from a population which contain contaminated parts. With this motivation, the discrete mixtures of a Hermite distributions is adopted to model the heterogeneity of an unobserved population. The results for estimating the parameters of zero-truncated count mixtures of Hermite distributions implying a unique estimator for the population size N are proposed in this study. The penalized maximum likelihood estimator, $\hat{N}_{PMLE}$, based on discrete mixtures of Hermite distribution is proposed. The simulation study was used to consider the performance of $\hat{N}_{PMLE}$. Based on two-components Poisson mixture model, $\hat{N}_{PMLE}$ is an appropriate estimator for population size estimation in moderate heterogeneity. Moreover, $\hat{N}_{PMLE}$ is an appropriate estimator for population size estimation in the case of weak and moderate heterogeneity in two-components Hermite mixture model. The profile mixture likelihood, CI ProfileMix, is used to construct the confidence intervals for the population size N. Based on the two-components Poisson and two-components Hermite mixture models, CI ProfileMix is an appropriate method for estimating CI of population size N in the case of weak heterogeneity.

Sometimes addition information on the observed units is available from another sub-sample of the target population, called a validation sample. Böhning et al. (2016) mentioned the capture-recapture modeling using validation sample in extension of generic ratio regression approach. In addition, Arnold et al. (2017) demonstrated that the use of validation sample not only substantially increases the estimation efficiency but also reduces the bias considerably. With this motivation the development of estimation to include validation information in the capture-recapture modeling is proposed in this study, to increase the accuracy and efficiency of population size estimation. The nonparametric maximum likelihood estimator is developed based on zero-truncated Hermite distribution which include validation information, denoted $\hat{N}_{Valid}$. The simulation study was used to consider the performance of $\hat{N}_{Valid}$. In the case of two-components Poisson mixture model, $\hat{N}_{Valid}$ can be an appropriate method for estimate the population size when $N \geqslant 500$ in weak and strong heterogeneity. It is tends to perform reasonably well, and is the best method for estimating population size based on the two-components Hermite mixture model. In addition, the simulation results indicated that the larger validation sample size, the more efficient estimation. The profile mixture likelihood which include validation information, CI ProfileValid, is proposed. Simulation results indicated that CI ProfileValid tends to perform reasonably well. It can be an appropriate method for estimating CI of population size N based on the two-components Poisson and two-components Hermite mixture models.

6.2 Future work

Although the proposed methods in this study tend to perform reasonably well, there are some important aspects that could be developed in the future.

1. This research proposed the penalized maximum likelihood estimator (PMLE) to improve the boundary problem. The PMLE can improve the problem and provides the estimation close to the population size N. Therefore, it might be valuable to include the PMLE with validation information for comparison.
2. The mixture model provides a reasonably well estimation, but derivation of its variance to form the confidence interval of N is not easy. The profile mixture likelihood was used to construct confidence interval estimation of N, which is computation-intensive task. Further study should focus on the estimation of the variance of mixture model.
3. In this study, the profile mixture likelihood is used to construct confidence intervals estimation of N. It might be beneficial to use bootstrap resampling technique to construct confidence intervals, and compare with the CI from the profile mixture likelihood .

REFERENCES

Amstrup, S. C., McDonald, T. L., & Manly, B. F. (2010). *Handbook of capture-recapture analysis*. Princeton University Press.

Arnold, M., Böhning, D., & Azevedo, C. (2017). A ratio regression approach to estimate the size of the salmonella-infected flock population using validation information. In *Capture-recapture methods for the social and medical sciences* (pp. 61–78). Chapman and Hall/CRC.

Böhning, D. (2008). A simple variance formula for population size estimators by conditioning. *Statistical Methodology*, *5*(5), 410–423.

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

Böhning, D., Baksh, M. F., Lerdsuwansri, R., & Gallagher, J. (2013). Use of the ratio plot in capture–recapture estimation. *Journal of Computational and Graphical Statistics*, *22*(1), 135–155.

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

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

Böhning, D., et al. (2016). Ratio plot and ratio regression with applications to social and medical sciences. *Statistical Science*, *31*(2), 205–218.

Böhning, D., & Schön, D. (2005). Nonparametric maximum likelihood estimation of population size based on the counting distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(4), 721–737.

Böhning, D., Suppawattanabodee, B., Kusolvisitkul, W., & Viwatwongkasem, C. (2004). Estimating the number of drug users in bangkok 2001: A capture–recapture approach using repeated entries in one list. *European Journal of Epidemiology*, *19*(12), 1075.

Böhning, D., & Vilas, V. J. D. R. (2008). Estimating the hidden number of scrapie affected holdings in great britain using a simple, truncated count model allowing for heterogeneity. *Journal of agricultural, biological, and environmental statistics*, *13*(1), 1.

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

Chao, A. (1989). Estimating population size for sparse data in capture-recapture experiments. *Biometrics*, 427–438.

Chao, A. (2005). Modern closed-population capture-recapture models. *Handbook of capture-recapture analysis*, 58–87.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1–38.

Durusoy, R., & Karababa, A. O. (2010). Completeness of hepatitis, brucellosis, syphilis, measles and hiv/aids surveillance in izmir, turkey. *BMC Public Health*, *10*(1), 71.

Farcomeni, A., Scacciatelli, D., et al. (2013). Heterogeneity and behavioral response in continuous time capture–recapture, with application to street cannabis use in italy. *The Annals of Applied Statistics*, *7*(4), 2293–2314.

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

Heijden, P. G. Van der, Whittaker, J., Cruyff, M., Bakker, B., Vliet, R. Van der, et al. (2012). People born in the middle east but residing in the netherlands: Invariant population size estimates and the role of active and passive covariates. *The Annals of Applied Statistics*, *6*(3), 831–852.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, *47*(260), 663–685.

Keating, K. A., Schwartz, C. C., Haroldson, M. A., & Moody, D. (2002). Estimating numbers of females with cubs-of-the-year in the yellowstone grizzly bear population. *Ursus*, 161–174.

Kuhnert, R., & Böhning, D. (2009). Camcr: Computer-assisted mixture model analysis for capture–recapture count data. *AStA Advances in Statistical Analysis*, *93*(1), 61–71.

Kuhnert, R., Del Rio Vilas, V. J., Gallagher, J., & Böhning, D. (2008). A bagging-based correction for the mixture model estimator of population size. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *50*(6), 993–1005.

Lerdsuwansri, R. (2012). *Generalisation of the lincoln-petersen approach to non-binary source variables*. Unpublished doctoral dissertation, University of Reading.

Lerdsuwansri, R., & Böhning, D. (2017). Extending the lincoln–petersen estimator when both sources are counts. In *Capture-recapture methods for the social and medical sciences* (pp. 341–360). Chapman and Hall/CRC.

Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics*, *59*(4), 1123–1130.

Mastro, T. D., Kitayaporn, D., Weniger, B. G., Vanichseni, S., Laosunthorn, V., Uneklabh, T., et al. (1994). Estimating the number of hiv-infected injection drug users in bangkok: a capture–recapture method. *American Journal of Public Health*, *84*(7), 1094–1099.

McDonald, S., Hutchinson, S., Schnier, C., McLeod, A., & Goldberg, D. (2014). Estimating the number of injecting drug users in scotland's hcv-diagnosed population using capture–recapture methods. *Epidemiology & Infection*, *142*(1), 200–207.

Moriña, D., Higueras, M., Puig, P., & Oliveira, M. (2015). Generalized hermite distribution modelling with the r package hermite. *R Journal*, *7*(2).

Niwitpong, S.-a., Böhning, D., Heijden, P. G. van der, & Holling, H. (2013). Capture–recapture estimation based upon the geometric distribution allowing for heterogeneity. *Metrika*, *76*(4), 495–519.

Norris, J. L., & Pollock, K. H. (1998). Non-parametric mle for poisson species abundance models allowing for heterogeneity between species. *Environmental and Ecological Statistics*, *5*(4), 391–402.

Norris III, J. L., & Pollock, K. H. (1996). Nonparametric mle under two closed capture-recapture models with heterogeneity. *Biometrics*, 639–649.

Pawitan, Y. (2001). *In all likelihood: statistical modelling and inference using likeli-*

hood. Oxford University Press.

Rocchetti, I., Bunge, J., Böhning, D., et al. (2011). Population size estimation based upon ratios of recapture probabilities. *The Annals of Applied Statistics*, *5*(2B), 1512–1533.

Sanathanan, L. (1977). Estimating the size of a truncated sample. *Journal of the American Statistical Association*, *72*(359), 669–672.

Sanfan, P., Channgam, T., & Chaiarporn, N. (2019). Risk factors influencing patients at a drug treatment center to return to drugs. In *Proceedings of the national undergraduate conference on statistics 2019* (pp. 1–10).

Schwarz, G., et al. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

Tercero, F., & Andersson, R. (2004). Measuring transport injuries in a developing country: an application of the capture–recapture method. *Accident Analysis & Prevention*, *36*(1), 13–20.

Van Der Heijden, P. G., Bustami, R., Cruyff, M. J., Engbersen, G., & Van Houwelingen, H. C. (2003). Point and interval estimation of the population size using the truncated poisson regression model. *Statistical Modelling*, *3*(4), 305–322.

Van Der Heijden, P. G., Cruyff, M., & Van Houwelingen, H. C. (2003). Estimating the size of a criminal population from police records using the truncated poisson regression model. *Statistica Neerlandica*, *57*(3), 289–304.

Van Hest, N., Grant, A., Smit, F., Story, A., & Richardus, J. H. (2008). Estimating infectious diseases incidence: validity of capture–recapture analysis and truncated models for incomplete count data. *Epidemiology & Infection*, *136*(1), 14–22.

Viwatwongkasem, C., Kuhnert, R., & Satitvipawee, P. (2008). A comparison of population size estimators under the truncated count model with and without allowance for contaminations. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *50*(6), 1006–1021.

Viwatwongkasem, C., Satitvipawee, P., Jareinpituk, S., & Soontornpipit, P. (2013). Mixture models for estimating the number of drug users in thailand 2005-2007. *Applied Mathematics*, *4*(09), 1242.

Wang, J.-P., & Lindsay, B. G. (2008). An exponential partial prior for improving nonparametric maximum likelihood estimation in mixture models. *Statistical Methodology*, *5*(1), 30–45.

Wang, J.-P. Z., & Lindsay, B. G. (2005). A penalized nonparametric maximum likelihood approach to species richness estimation. *Journal of the American Statistical Association*, *100*(471), 942–959.

Zelterman, D. (1988). Robust estimation in truncated discrete distributions with application to capture-recapture experiments. *Journal of statistical planning and inference*, *18*(2), 225–237.

BIOGRAPHY

Name	Miss Parawan Pijitrattana
Date of Birth	January 22, 1991
Educational Attainment	Academic Year 2014: M.Sc. in Applied Statistics, Thammasat University, Thailand Academic Year 2012: B.Sc. in Mathematics, Mahidol University, Thailand
Scholarship	2019-2009 Human Resource Development in Science Project (Science Achievement Scholarship of Thailand, SAST)
Work Experiences	2019 -2018, Mathematics teacher Matthayom Watnairong School 2018 - 2017, Mathematics teacher Satriwit3 School 2015 - 2014, Mathematics team Learn Corporation Co.,Ltd

Publications

Pijitrattana, P. and Lerdsuwansri, R. (2015). Confidence Interval for population size using the Generalized Turing estimator, *Proceedings of the International Conference on Applied Statistics*, 278-285. Pattaya, Thailand, 15-17 July 2015.