Some Contributions to the Analysis of Dual-record System for Estimating Human Population Size

Dissertation submitted to Indian Statistical Institute
in candidacy for the degree of
Doctor of Philosophy
in
Statistics

By

Kiranmoy Chatterjee
Department of Statistics
Bidhannagar College, Salt Lake City
Kolkata-700064, India

Research Lab
Interdisciplinary Statistical Research Unit
Indian Statistical Institute, Kolkata
November 2016

I dedicate this thesis

to

My Grandfather - Late Prodyot Kumar Halder,

My lovely Son - Arunalok†

&

Shibu‡

---

† on his 1st Birth Anniversary, 24th May 2016

‡ a blind man selling his own handmade *Dhupkathi* daily at Shyambazar Five Point Crossing, Kolkata, India

Acknowledgements

First of all, I would like to thank my Ph.D. supervisor, Dr. Diganta Mukherjee, Indian Statistical Institute, Kolkata, for his guidance. His accessibility, support and encouragement have been of great value during the progression of my research. Equally, I would like to thank him for giving me exposures to participate and deliver lectures in workshops to gain valuable knowledge on Government Statistics in India. Thanks also should go to him for giving me the opportunity to have experience in both a census and sample survey operation as a part of his project on *Handloom Workers*, which certainly strengthen my understanding on my thesis topic as well as Statistics, as a truly applied science. His faith in both my ability and the quality of my work provided great reassurance during periods of doubt. I would like to thank the reviewers of my Ph.D. dissertation and the members of Ph.D. & D.Sc. Committee (Statistics) of Indian Statistical Institute (ISI) for their valuable comments and suggestions.

During the course of my Ph.D. in ISI Kolkata, I was benefited by several professors either by their teaching or by constant assistance to make my environment convenient. In particular, I am greatly thankful to Prof. Ayanendranath Basu for helping me to continue my Ph.D. at a very dark time of my Ph.D. life. In that toughest time, being head of the department his support gave me enough confidence and in later days also, fortunately, that support and careful monitoring has never stopped. Thanks also to Dr. Sourabh Bhattacharya, Prof. S. K. Bhandari and Prof. Murari Mitra (from IIEST, Shibpur, Howrah) for their essential help and valuable discussions on my thesis work.

I have been intensively enriched by the teaching of several research and masters level courses by some eminent teachers in ISI at the beginning of my Ph.D. tenure. Courses taught by Prof. Prabal Chaudhury, Prof. Alok Goswami, Prof. S. C. Bagchi, Prof. Arup Bose, Prof. S. K. Bhandari, Prof. Tapas Samanta and Dr. Sourabh Bhattacharya helped me in due courses of my research. At the same time, I would also convey my gratitude to my four college (Maulana Azad College, Kolkata, India) teachers - Pipul Dutta, Biswajit Roy, Arup Kr. Hait and Partha Pal, from whom I learned undergraduate statistics and from their quality teaching gave me the flavour of statistics towards applications in real life. Arup Sir, being my present colleague (in the Department of Statistics, Bidhanngar College, Kolkata), still giving me continuous encouragement to do good research. I also like to thank my school-time teacher Mohit Kr. Banerjee and my brother Sourav Paul, who also continuously encourage me as my close well-wishers.

I would like to thank all professors and office staff of Interdisciplinary Statistical Research Unit (ISRU), formerly BIRU, in ISI, for giving me such a nice environment for pursuing research and also for providing such delicious snacks before every interesting seminar talks. I like to express my gratitude to the present and earlier Heads of ISRU for providing me a complete setup to carry out my Ph.D. as full-time research fellow as well as a part-time Ph.D. student after I left ISI for my present job. I also thank Dean of Studies, Library and Director of this institute.

About my period in ISI, I would also like to thank Prasenjit Da and Prajamitra Da, for all the time spent together and the endless discussions about statistics, research, politics, etc. Thanks for all the entertaining tea-time conversations and general friendliness. My time in ISI was also made enjoyable by many other friends, particularly the old gang of my department BIRU - Maji(Avijit) Da, Moumita Di, Suman, Hemant. I missed that time and familiarity, when we all were there, up to Mid of 2014. My sincere appreciation is extended to Minerva, Kaushik, Swagata, Reetabrata, Samba (Shiva Rao) Da, Avinandan and Jayant.

Lastly, I should appreciate the blessings and supports of my parents and my own brother. Very very special thanks goes to my wife cum friend - Nandita, whose constant encouragement, support and sacrifice helped me lot to complete this work smoothly and within time.

Data acknowledgements. Thanks to all those involved in the running and maintenance of the Enumeration of Handloom Workers project. Specifically, thanks go to Dr. Diganta Mukherjee for his kind permission to use the Handloom PES data described in section 1.8.1 and analysed in section 4.5.3.

Institutional funding acknowledgements. Thanks to CSIR (Council of Scientific and Industrial Research, India) and Indian Statistical Institute for funding of my Ph.D. research for three and a half years (Jan. 2010 - Jul. 2013) and one year (Aug. 2013 - Sep. 2014) respectively.

ISI, Kolkata, 24th May 2016 *Kiranmoy Chatterjee*

Contents

Bibliography **153**

Publications **163**

Conference/Workshop Presentations **165**

1 Introduction

1.1 Dual-record System (DRS)

The problem of population size estimation is a very important administrative and statistical concern which includes a vast area of application in the fields of epidemiology, demography and official statistics. Federal agencies are generally interested to know the actual size (say, N) of a specified population or any vital event that occurred in a specified area within a given time span. Census or civil registration system often fails to extract the true size of the population. The degree of inaccuracy depends on the actual size of the population, its diversity and of course, on the quality of the counting process. Any attempt to count all the individuals in a given moderate or large population is inevitably subject to error. As a remedy, the use of capture-recapture type experiment is being used for a long time. As per record, Laplace (1783 [60]) made the first implementation of such experiment in order to estimate the number of inhabitants in France. Thereafter, identical methods independently proposed by Petersen (1896 [73]) and Lincoln (1930 [63]) in order to estimate the size of an interest population became famous as *Lincoln-Petersen* estimate which is based only on one recapture operation after the first capture attempt. Individuals counted at the time of first capture are matched with the list of individuals prepared by the second attempt. In literature, this type of data structure with only two counting attempts covering the population is known as *Dual-record System* or simply, *Dual System*. A stabilized version of *Lincoln-Petersen* estimator developed by Chapman (1951 [24]) is still in use by numerous practitioners. Schnabel (1938 [81]) considered a multi-sample extension of the *Lincoln-Petersen* method, where each sample captured commencing from the second is examined for marked members and then every member of the sample is given another mark before being returned to the population. Later this multi-sample extension became very popular, especially for wildlife populations and commonly known as *capture-mark-recapture* (CMR) or simply *capture-recapture* experiment. Most of the advanced statistical models are developed after 1950's in order to efficiently describe several situations arising in the real world for dependency between capture and recapture operations, varying individual capture probabilities, etc. Excellent reviews by Seber (1982[82], 1986[83]), Otis et. al.(1978[71]) and Chao (2001[23]) are available on capture-recapture theory. Thus, *Dual-record System* (DRS), which is particularly planned for human population, is technically very close to capture-recapture system with only two sampling occasions.

In modern era, an application of the *capture-recapture* method is made in order to measure the extent of registration for vital events (ChandraSekar and Deming 1949 [17]; Ayhan, 2000 [1]) in the form of DRS. Apart from the estimation of the size of a general population or vital events, this method has an extensive use for population growth estimation (Marks et al., 1974 [66]; Krotki, 1978 [59]), illusive or hard-to-count populations (Jibasen et al., 2012 [58]; Dreyfus et al., 2014 [33]) and also in several epidemiological applications including under-ascertainment in traditional epidemiological surveillance (Chao et al., 2001 [23]). Dual Record System Estimation is also used for the application to the problem of estimating undercount in census (Wolter, 1986 [102]; Cressie, 1989 [31]; Ayhan and Ekni, 2003 [2]; Elliot and Little, 2005 [34]; Watcher, 2008 [97]). In this application, a survey is usually conducted independently after the census operation to estimate the number of events missed in the census count by matching the two available lists of individuals. However, for human population, nature of the capture-recapture experiment is purposefully different than that for wildlife population. We will discuss about this difference in the next section. In the context of counting events in a human population, this kind of data structure with matching information from more than three lists is known as Multiple-record system, which is equivalent to the capture-recapture system in biological studies (*see* Otis et al., 1978 [71] and Seber, 1986 [83]). However, in the context of human population, more than two sources of information is hardly found since establishment of more than two sources is operationally complicated, expensive and also, human capture probabilities are relatively higher than animal. Application of DRS in some other studies related to human population includes estimation of the size of victims due to war, natural calamity, accidents, etc.

1.2 DRS & Two-sample Capture-Mark-Recapture (CMR) Technique

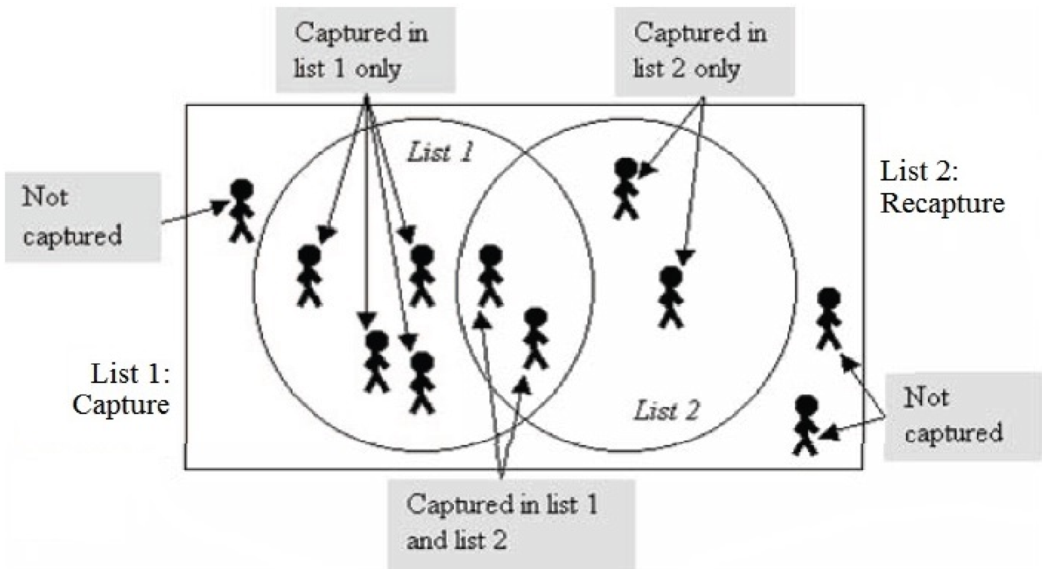
Any attempt to count all individuals belonging to a large or moderately large population is inevitably subject to errors. If the interest population is hard-to-reach (e.g. drug addicted persons, street-beggars, particular species of an animal, etc.), then also complete listing is quite impossible by any attempt. *Capture-recapture* technique deals with this problem by employing two or more independent attempts to count all individuals in each attempt and formulates a well-structured operational technique by matching all the available lists of individuals prepared from those attempts. Since a part (or, say, a sample) of the interest population is captured in each attempt, not all, hence these are usually called sampling occasions. Thus, in any capture-recapture experiment, T (say) no. of sampling occasions or counting attempts are performed, where $T \geq 2$. The first attempt is termed as *capture* and all subsequent attempts are called *recapture*. For details on *capture-recapture* technique, *see* Pollock (2000[74]) and Chao (2001a[22]).

Dual-record System (DRS) comprises of exactly one *recapture* attempt after the usual *capture* and therefore, there are only two sampling occasions. For human population, more than two attempts are seldom used. Very few evidence has been found yet where capture-recapture type experiment dealt with three attempts (Zaslavsky and Wolfgang, 1993[107]; Ruche et al., 2013[78]). Indeed, in order to study a human population, these two attempts (*capture* and one *recapture*) may not always be time-ordered. In several instances, two attempts for counting are carried out simultaneously by two different agencies.

Let us present some details on DRS in the following generic setup. Suppose, it is tried to count a population by two independent systems. Then individuals in the 1st list (made from first system) are matched one-by-one with the 2nd list of individuals (made from second system). Capturing status of an individual is denoted by any of the notations 11, 10, 01 and 00 representing respectively the category of persons (*i*) present in both the lists, (*ii*) present in List 1 but absent in List 2 and (*iii*) present in List 2 but absent in List 1 and (*iv*) absent in both the lists (*see* Figure 1.1).

Example 1.2.1 *Post Enumeration Survey (PES). This specialized survey is conducted independently within two to five months ahead from the usual census operation to estimate the size of possible undercount or overcount of people in a specified geographic region (say, whole country, or a state, or a county/district, etc). A sample of small administrative units are selected at the time of PES. Consider such an administrative unit U with the actual population size, say, N (without individuals living in institutions or homeless). Therefore, the census enumeration of population in that administrative unit comprises the List 1, whereas List 2 is implicitly made up of those people in the same administrative unit captured at the time of PES, which is called the P-sample. This sample is used to estimate the undercount in the original census for this block by matching persons in the census and P-sample lists. Another sample of individuals*

Figure 1.1: Venn Diagrammatic View of Dual-record System



from census population is made for quantifying the errors in census itself (such as duplications or erroneous enumerations) and their effect on census coverage. This sample is called the E-sample. *Clearly, in this example, two attempts (census and PES operations) are time-ordered as PES is carried out after census. More details about PES along with its operational strategies in several countries can also be found in Krotki (1978 [59]).*

Difference between DRS & Two-sample CMR System

Capture-Mark-Recapture (CMR) technique for wildlife populations and Multiple-record system (MRS) for human are broadly similar in concept. While the purpose of the CMR is estimation of population size, the estimation of the total number of events is usually of interest in the MRS. However, this difference is only in nomenclature, but the definition of population under consideration may be different for CMR and MRS (E1-Khorazaty et al., 1976 [35]). MRS is also used to estimate size of whole or a specific group of human population. Dissimilarities between the CMR and MRS techniques arise mainly because of their application to different types of populations. Instead of marking in animal populations, the characteristics (such as name, sex, occurrence date, address, etc.) of each event are recorded in the case of human population. Some writers believe that there is a wide divergence between these two techniques in the matching process (Marks, Seltzer and Krotki, 1974[66]). This is because the DRS, being a kind of MRS, is concerned essentially with the event recorded by different sources, and hence with a two-way match (which determines the matching status of all the events in both sources), while the CMR is concerned with determining the group to which an animal belongs (captured or uncaptured), and hence with a one-way match. But practically, both yields same result.

As stated earlier, DRS is the particularization of MRS when number of capture occasion is two. For human population capture probabilities are usually higher and conducting more than two counting attempts are generally difficult to implement, hence more than two are seldom used for human population. Animal requires relatively large number of captures, practically at least four, due to their very low capture probability and more mobility. Further, sampling occasions in CMR are usually time-ordered. But, there are some examples where DRS (or MRS) is constructed without any restrictions of time-ordered attempts, specially in the field of Epidemiology and the counting of hard-to-reach populations (Chao et al., 2001b[23]; Xu et al., 2014[103]). Marks et al. (1974[66]) presents a good relationship between the assumptions behind CMR and the equivalent assumptions for DRS technique.

1.3 DRS: Construction and Early Estimates

Prior to constructing a Dual-record system (DRS) originating from a given human population with unknown size N, let us state six main assumptions on the capture and recapture operations and individual's capture probabilities. Therefrom, we will notice that the different set of assumptions will lead to different models.

1.3.1 Assumptions

The general DRS is prepared based on the following first three key assumptions. Further, three assumptions which are commonly used for the analysis of human populations are also mentioned here.

A1. Closed Population. The population under consideration should be closed between the time of first and second counting attempts. No birth, death, immigration or emigration should occur between these two time points when capture attempts are exercised.

A2. Proper Matching. Ensuring careful matching of the census records with the records from survey, it is possible to make a determination without error that an individual is either captured only in List 1 or only in List 2 or in both.

Implication of this assumption is that each captured individual is classified correctly in any one of the three categories $(11, 10, 01)$, already mentioned in section 1.2. However, there is another cell which refers the individuals belongs neither to the List 1 nor the List 2. It is always thought that some of the N individuals belong to this fourth cell, that means it is non-empty but unknown also. This unknown cell makes the N unknown.

A3. Autonomous Independence. The two lists - List 1 and 2 are made as a result of N mutually independent trials. The position of each individual (i.e. in which category among 11, 10 or 01 s/he belongs) in the population is thought of as the realization of each trial.

This assumption has a relevance in statistical modelling as the capture status of i-th individual in a DRS $(11, 10, 01, 00)$ is distributed as the multinomial distribution with four cell probabilities $(p_{i11}, p_{i10}, p_{i01}, p_{i00})$ $\forall\, i = 1(1)N$.

A4. Causal Independence. Probability for an individual being included in List 1 is independent of his/her inclusion in List 2. Hence, the cross-product ratio $\theta_i = \frac{p_{i11}p_{i00}}{p_{i10}p_{i01}}$ is equal to 1 for i-th person, $i = 1(1)N$.

A5. Homogeneous Capture probabilities. The capture probabilities are same over all individuals separately for each of the two sources. Thus, $p_{i1\cdot} = p_{1\cdot}$ and $p_{i\cdot 1} = p_{\cdot 1}$ $\forall\, i$.

A6. Time-variation. This assumption is particularly relevant in CMR type experiment specially for human population. In DRS, it tells $p_{i1\cdot} \neq p_{i\cdot 1} \forall\, i$, that means different attempts have different capture probabilities. Thus, for homogeneous population, $p_{1\cdot} \neq p_{\cdot 1}$. For human population, this assumption is very reasonable.

The first two assumptions *A1* and *A2* are most basic and required for any *Closed Population* analysis of CMR data. Capture-recapture literature or the DRS literature discusses many of these assumptions. For further discussions, see Otis et al. (1978[71]), Seltzer and Adlakha (1974[84]), and Cowan and Bettin (1982[28]).

Remark: Most of the times populations are not homogeneous with respect to both capture probabilities. To ensure homogeneity, Chandrasekar and Deming (1949[17]) suggested forming poststrata dividing the population according to various cross-sectional age, race, sex and geographical groups so that people in each of those resultant post-strata are reasonably homogeneous, e.g. in 2000, US Census proposed about 12600 post-strata according to 50 States × 6 Races/Origin × 7 Age/Sex × 2 Tenures × 3 Geographic Regions. In recent times, Wang et al. (2006[98]) proposed a method to form homogeneous post-strata based on Bayesian treed capture-recapture model in the application of population size estimation for census undercount.

1.3.2 Construction of DRS

Consider a human population U of true size N. It is believed that any census counting fails to capture all individuals in U. To know the true extent of this coverage or equivalently the true N, one collects another list independently covering the same population. Therefore by matching the two lists, individuals in U can be classified according to a multinomial fashion based on the assumptions *A1*, *A2* and *A3*. As an example, Indian census uses a specialized survey, called Post Enumeration Survey (PES) and US Census Bureau uses their regular survey - Current Population Survey (CPS) as the second source. Proper matching generates four groups and each individual of U lies in any of this four groups depending upon his/her capture status. (*i*) Individuals who are present in both lists (x_{11}), (*ii*) Individuals who are present in first list only (x_{10}), (*iii*) Individuals who are present in second list only (x_{01}) and (*iv*) Individuals who are not present in any lists (x_{00}). Data structure is given in the left panel of Table 1.1. This particular data structure is known as Dual-record System (DRS) or capture-recapture data for two sampling occasions. $x_0 = x_{11} + x_{10} + x_{01}$ is the total number of distinct persons found by the two lists. The last cell (x_{00}), presenting the number of missed individuals by both the systems, makes the total population size $N(= x_{..})$ unknown. If we consider the population U having the characteristics of *homogeneity* (Assumption *A5*), expected cell probabilities for an individual are given in the right panel of Table 1.1. All these notation will be followed throughout in this thesis.

Table 1.1: 2×2 table for Dual-record-System Model

	Observed Cell Frequency			Expected Cell Proportions		
	List 2			List 2		
List 1	In	out	Total	In	out	Total
In	x_{11}	x_{10}	$x_{1\cdot}$	p_{11}	p_{10}	$p_{1\cdot}$
Out	x_{01}	x_{00}	$x_{0\cdot}$	p_{01}	p_{00}	$p_{0\cdot}$
Total	$x_{\cdot 1}$	$x_{\cdot 0}$	$x_{\cdot\cdot} = N$	$p_{\cdot 1}$	$p_{\cdot 0}$	$p_{\cdot\cdot} = 1$

1.3.3 Dual System Estimate (DSE)

Typical capture-recapture theory says that fraction recaptured among the 2nd sample estimates the fraction of the whole population caught the first time. This idea is equivalent to the assumption of *causal independence* (assumption *A4*) between capture probabilities of the two attempts. Thus, the probability to be captured in 2nd list has no connection with the capture probability for List 1. In other words, assumption of *causal independence* implies that the cross-product ratio for 2×2 table, $\theta_0 = \frac{p_{11}p_{00}}{p_{10}p_{01}}$, is 1. Therefore, $p_{00} = \frac{p_{10} \cdot p_{01}}{p_{11}}$. Hence, replacing p_{ij} by x_{ij}/N, we have $x_{00} = \frac{x_{10} \cdot x_{01}}{x_{11}}$. Thus, we have the *Dual System Estimator* (or DSE) for $(N, p_{1\cdot}, p_{\cdot 1})$ as:

$$\hat{N}_{DSE} = x_0 + \hat{x}_{00} = \left[\frac{x_{1\cdot}x_{\cdot 1}}{x_{11}}\right], \quad \hat{p}_{1\cdot,DSE} = \frac{x_{11}}{x_{\cdot 1}} \quad \text{and} \quad \hat{p}_{\cdot 1,DSE} = \frac{x_{11}}{x_{1\cdot}},$$

respectively, where $[u]$ refers to the largest integer not more than u. DSE for N is a traditional estimator used frequently in practice for human population and is equivalent to the well known *Lincoln-Petersen* estimator ($\hat{N}_{LP}$) (*see* section 1.4.1 for detail discussion). Note that the resulting conditional MLE $\hat{N}_t$ from model M_t, later in section 1.4, is also same as $\hat{N}_{DSE}$. As $\hat{N}_{DSE}$, $\hat{N}_{LP}$ and $\hat{N}_t$ are same and based on the independence assumption, henceforth, these are unanimously denoted as $\hat{N}_{ind}$ throughout the thesis and we call this DSE of N. Thus, DSE $\hat{N}_{ind}$ considers all the assumptions $A1 - A6$ stated in section 1.3.1, though assumption $A3$ is not required as $\hat{N}_{ind}$ does not build up on the assumption of statistical modelling.

$\hat{N}_{DSE}$ is also known as CD estimator (denoted as $\hat{N}^{(CD)}$), in honour of Chandrasekar and Deming (1949 [17]) for their pioneering work on human population in modern era. Chandrasekar and Deming (1949 [17]) proposed their model based on the assumption that the changes of an event being missed by the two data sources are independent of one another. In other words, the basic CD model explicitly assumes that the conditional probability of an event being caught by one system, given that it is caught by the second system, is equal to the conditional probability of an event being caught by the first system, given that it is missed by the second system. In practice, it is unlikely that these conditional probabilities are really equal (Seltzer and Adlakha, 1974 [84]).

Chandrasekar and Deming (1949 [17]) observed that it may be possible to reduce the bias that results from a lack of independence, by classifying the events into homogeneous groups, and making the estimate of events separately for each group (*see* the remark in section 1.3.1). This will be effective if the correlation for the contingency table for each grouping or stratum is near zero, but the correlation for the contingency table for all strata combined is not zero (Jabine and Bershad, 1968 [55]).

Both Chandrasekar and Deming, and Jabine and Bershad found evidence of positive correlation (between the events missed by the two sources) when the method was applied to subgroups as well as to the total data. While the method of subgrouping offers an improved estimate, it still suffers from the defect that independence within subgroups is assumed

(Greenfield, 1975 [45]).

Greenfield and Tam (1976 [46]) showed that $(x_{10}x_{01})^{1/2}$ is a good approximation of Greenfield's estimator $\hat{x}_{00}^{(G)}$, when $x_{11} \geq (x_{10}x_{01})^{1/2}$ and is exactly equal to $\hat{x}_{00}^{(G)}$ if $x_{11} < (x_{10}x_{01})^{1/2}$. When $x_{11} < (x_{10}x_{01})^{1/2} < 3x_{11}$, then $(x_{10}x_{01})^{1/2}$ gives a close approximation of $\hat{x}_{00}^{(G)}$.

In consequence, Chandrasekaran and Deming (1981 [18]) evaluated the earlier developments of the recent past in their paper. They emphasized that $\hat{x}_{00}^{(CD)} \neq (x_{10}x_{01}/x_{11})$ if the correlation is present. The associated correlation coefficient between the two lists is defined as,

$$\hat{\rho}^{(CD)} = \frac{x_{11}x_{00} - x_{10}x_{01}}{[(x_{11}+x_{10})(x_{11}+x_{01})(x_{10}+x_{00})(x_{01}+x_{00})]^{1/2}}.$$

Chandrasekaran and Deming (1981 [18]) has proposed that, the CD estimator (1949 [17]) $\hat{x}_{00}^{(CD)} = (x_{10}x_{01}/x_{11})$ if $\hat{\rho}^{(CD)} = 0$. If $\hat{\rho}^{(CD)} > 0$, then $\hat{x}_{00}^{(CD)} > (x_{10}x_{01}/x_{11})$ and CD method underestimates $\hat{x}_{00}^{(CD)}$. If $\hat{\rho}^{(CD)} < 0$, then $\hat{x}_{00}^{(CD)} < (x_{10}x_{01}/x_{11})$ and CD method overestimates $\hat{x}_{00}^{(CD)}$.

1.3.4 Some Modifications over DSE

Chapman's modified estimator. When the number of matches between two sources are found to be zero (0), we cannot compute $\hat{N}_{ind}$. To avoid this problem, a modification of this estimator due to Chapman (1951[24]) is given by

$$\hat{x}_{00} = \frac{x_{10}x_{01}}{x_{11}+1}.$$

Therefore, Chapman's modified estimator of N is given by

$$\hat{N}_{CPM} = \frac{(x_{1.}+1)(x_{.1}+1)}{x_{11}+1} - 1. \tag{1.1}$$

This estimate is less affected by zeros and is said to be less biased than the $\hat{N}_{ind}$ estimator.

However, DSE ($\hat{N}_{ind}$) often fails miserably either when x_{11} is found to be zero or very close to zero, or when the underlying independence assumption between capture probabilities is violated. Chandrasekar and Deming (1949[17]) addressed this problem and analyzed the extent of the lack of independence. Many methodologists (*see* Isaki et al., 1987[54]; El-Khorazaty, 2000[36]) and practitioners (Jarvis et al., 2000[56]; Tilling, 2001[94]; Xu et al., 2014[103]) argued that the independence assumption may not be justified often in reality. A nice review is done by Brittain and Böhning (2009[13]) on the various methods available by relaxing the independence assumption in DRS context. Now we present some alternative estimates.

Jabine-Bershad Estimate. Jabine and Bershad (1968 [55]) proposed the an estimator of N. They assume that both collections use the same procedure and therefore, the expected values of x_{10} and x_{01} are the same. Let us denote the same value as n^*, i.e. $x_{10} \approx x_{01} = n^*$. These are the numbers of vital events not included in one list but which are included in the other list.

$$\hat{N}^{(JB)} = x_{11} + 2n^* + \frac{(n^*)^2}{x_{11}} = \frac{(x_{11}+n^*)^2}{x_{11}}. \tag{1.2}$$

They also proposed the following expression for the bias of the estimator $\hat{N}^{(CD)}$.

$$B\left(\hat{N}^{(CD)}\right) = \hat{N}^{(CD)} - (x_{11} + 2n^* + x_{00}) = \frac{(n^*)^2}{x_{11}} - x_{00}.$$

The correlation between the two sets of observations was determined as,

$$\hat{\rho}^{JB} = \frac{x_{11}x_{00} - (n^*)^2}{(x_{11}+n*)(x_{00}+n*)}.$$

Chao's Lower Bound Estimate. Anne Chao (1987[19], 1989[20]) proposed an alternative estimator of population size by relaxing the assumption of *causal independence* (assumption $A4$). With the help of Cauchy-Schwartz inequality, Chao proposed lower bound estimate of x_{00} as

$$\hat{x}_{00} = \frac{(x_{10}+x_{01})^2}{4x_{11}}.$$

Therefore, Chao's estimator of N is given by

$$\hat{N}_{Chao} = x_0 + \frac{(x_{10}+x_{01})^2}{4x_{11}}. \tag{1.3}$$

Here, assumption $A6$ of *time-variation* is considered. If two marginal capture probabilities are same, i.e. $p_{1.} = p_{.1}$, we can see that Chao's estimator will be identical to $\hat{N}_{ind}$.

Zelterman Estimate. Let us consider the *Horvitz–Thompson* (H-T) estimator of population size given by

$$\hat{N}_{HT} = \frac{x_0}{1 - p_{00}}.$$

Zelterman (1988 [108]) proposed an estimate of p_{00} using only 1's and 2's from the zero-truncated count distribution. Their estimate performs well in terms of bias under the zero-truncated Poisson model. Moreover, this estimate works well if contaminations follow a Poisson mixture. The estimator of Zelterman (1988 [108]) is

$$\hat{N}_{Zelt} \quad = \quad \frac{x_0}{1 - exp[-2x_{11}/(x_{10} + x_{01})]}. \tag{1.4}$$

Ayhan's Estimator: Breakdown of Error Components. Ayhan (2000 [1]) has proposed an adjustment for the "not-reported cases" from both data sources. This provided a breakdown of the error components of the x_{00} cell of the contingency table, which has improved the outcome of the estimator. To achieve this, not reported cells was partitioned into "not reported" and "missed" cases. The unmatched cases from both data sources for the cells of the original layout were evaluated as "missed" as a result of undercoverage (UC), and "not reported" as a result of non-response (NR). For different cases, Ayhan (2000 [1]) has proposed several alternative estimators.

Several type of data sources (census, sample survey, or registry) can be utilized as a pair for evaluation. For using two sample cases for illustration, Ayhan (2000 [1]) has proposed the following estimators for the case of equal sample sizes, i.e. $n_1 = n_2$, where

$$n_1 = x_{1.} + NR(n_1), n_2 = x_{.1} + NR(n_2).$$

Now, the total number of events that should have been obtained from List 1 is n_1^*, with $\hat{n}_1^* = x_{1.} + NR(n_1) + UC(n_1^*)$. Similarly, the total number of events that should have been obtained from List 2 is n_2^*, with $\hat{n}_2^* = x_{.1} + NR(n_2) + UC(n_2^*)$.

Therefore, the estimator based on the List 1 was proposed by Ayhan (2000 [1]) as

$$\hat{N}_1^{(A)} \quad = \quad x_0 + \hat{x}_{00}^{(A)} = x_0 + \left[\left(n_1^* - x_{1.}\right) - x_{01}\right], \tag{1.5}$$

Again, the estimator based on the List 2 was proposed as

$$\hat{N}_2^{(A)} = x_0 + \hat{x}_{00}^{(A)} = x_0 + \big[\big(n_2^* - x_{.1}\big) - x_{10}\big]. \tag{1.6}$$

Details of their methodology are given in their paper.

Maximum Likelihood Estimator of p_{00} ***in*** $\hat{N}_{HT}$**.** Brittain and B$\ddot{o}$hning (2009[13]) derived a general form for the maximum likelihood estimator of p_{00} in the *Horvitz–Thompson* estimator $\hat{N}_{HT}$ mentioned above, based on the assumption that the observed data on the no. of times a captured individual is being enlisted follow a zero-truncated binomial distribution with trial parameter $m = 2$ and p is the underlying capture probability. Thus, $p_{00} = (1-p)^2$ and final estimate of N using the H-T estimator after plugged in the derived MLE for p is

$$\hat{N}_{Zelt} = \frac{x_0}{1-(1-\hat{p}_{MLE})^2}, \tag{1.7}$$

where $\hat{p}_{MLE} = 2x_{11}/(x_{10} + x_{01} + 2x_{11})$. From the construction of above two estimators, it is clear that both the assumptions *A6* of *time-variation* and *A4* of *causal independence* is not accounted for.

All of the above estimators are potentially relevant for wildlife populations, which is not of our interest in this present project. Except Chapman's adjusted and Chao's lower bound estimator, all other estimators do not consider the assumption *A6* of Time-variation which is very relevant for human population, especially for DRS (or, MRS) in social science and epidemiological applications.

Otis et al. (1978 [71]) specified eight models that incorporated potential sources of variation by modelling capture probabilities as dependent on time, behaviour, and/or heterogeneity, in the context of CMR technique where $T \geq 2$. In total, they presented eight models for different combination of the three factors (1) Time-variation, (2) Behavioral dependence and (3) Heterogeneity, depending on their presence. Here, we discuss some of those basic models briefly under DRS. Other models that are closely related to these models are actually under the broader class of basic capture-recapture models and not relevant here. In addition to the basic assumptions *A1* and *A2* pointed out earlier in section 1.3.1, here we consider another assumption *A3* which is responsible for the individuals to be classified in multinomial fashion. Thus, it helps to understand the different situations through well-structured parametric statistical modelling.

1.4 Time Variation Model (M_t)

1.4.1 Formulation and Lincoln-Petersen Estimate

A very common practice, across all fields, is to assume *casual independence* (assumption *A4*) for simplicity between two lists' probabilities. Let us consider one more assumption *A5*: individuals are *homogeneous* in terms of their capture probabilities, which can be achieved from Chandrasekar-Deming's suggestion of post-stratification. Assumption *A4* implies that the event of an individual being included in List 2 is independent of his/her inclusion in List 1. Hence, $p_{11} = p_{1\cdot}p_{\cdot 1}$. In addition to that, the assumption *A6* is accounted for this model i.e., two marginal capture probabilities satisfy $p_{1\cdot} \neq p_{\cdot 1}$. This model is well-known as M_t and associated likelihood for $N \geq x_0$ is

$$L_t(N, p_{1\cdot}, p_{\cdot 1}) \propto \frac{N!}{(N-x_0)!} p_{1\cdot}^{x_{1\cdot}} p_{\cdot 1}^{x_{\cdot 1}} (1-p_{1\cdot})^{N-x_{1\cdot}} (1-p_{\cdot 1})^{N-x_{\cdot 1}} \tag{1.8}$$

This model is one of the popular and useful in the arena of capture-recapture studies for human population, M_t model, which is similar to the one proposed by Lincoln and Petersen. Henceforth, it is also well known as *Lincoln-Petersen Model*. For fixed N, $t = (t_1, t_2)$ is sufficient, where $t_1 = x_{1\cdot}$ and $t_2 = x_{1\cdot} + x_{\cdot 1}$, so that a likelihood function for N may be based on the conditional distribution of x_{11}, x_{10}, x_{01} given t. The *conditional likelihood* obtained from (1.8) is

$$L_C(N) \propto \frac{\binom{N}{x_{11}, x_{10}, x_{01}}}{\binom{N}{t_1}\binom{N}{t_2 - t_1}} = \frac{(N-x_{1\cdot})!(N-x_{\cdot 1})!}{N!(N-x_0)!},$$

for $N \geq x_0$. Derivation of the above *conditional likelihood* can also be found in Severini (2000[87], pp. 281). Hence, for given observed data $\underline{\boldsymbol{x}} = (x_{1\cdot}, x_{\cdot 1}, x_{11})$, associated MLE for N and $p_{1\cdot}$ are:

$$\hat{N}_t = \left[\frac{x_{1\cdot}x_{\cdot 1}}{x_{1\cdot} + x_{\cdot 1} - x_0}\right] = \left[\frac{x_{1\cdot}x_{\cdot 1}}{x_{11}}\right], \tag{1.9}$$

$$\text{Therefore, } \hat{p}_{1\cdot;t} = x_{1\cdot}/\hat{N}_t \text{ and } \hat{p}_{\cdot 1;t} = x_{\cdot 1}/\hat{N}_t \tag{1.10}$$

where x_0 is total number of distinct captured individuals in two lists, $x_0 = x_{11} + x_{10} + x_{01}$. Thus, conditional MLE $\hat{N}_t$ becomes identical to the $\hat{N}_{ind}$ (*see* section 1.3.3).

Further, the unconditional MLE of N, $\hat{N}_{UMLE}$, can be obtained from the estimating equation $\sum_{j=1}^{x_0}(N-j+1)^{-1}+ln(1-x_{1.}/N)+ln(1-x_{.1}/N)=0$. If the root of the equation is N^* (say), then either $[N^*]$ or $[N^*+1]$ is the $\hat{N}_{UMLE}$ depending upon which has greater likelihood.

1.4.2 Other Estimates from M_t

Besides several likelihood and pseudo likelihood estimates (Darroch, 1958[32]; Otis et al., 1978[71] and Bolfarine et al., 1992[10]), some frequentist (Chao, 1987[19]; Zelterman, 1988[108]) methods have been proposed on this model. Bayesian approach was pioneered by Castledine (1981[15]), Smith (1988[90], 1991[91]) and later, by George and Robert (1992[42]) on a hierarchical Bayesian M_t model. Wang et al. (2007 [99]) opined that the choice of noninformative priors for model parameters depends on the number of sampling occasions only. Yang and Paul (2010[106]) comprehensively compare some popular estimates with their proposed empirical Bayes as well as interval estimators. Recently, Xu et al. (2014[103]) provide a nice objective prior in connection with this model.

1.5 Behavioral Response Model (M_b)

In case of human population, the assumption of *causal independence* has been highly criticized. Individual's probability of capture in List 2 may be changed in response to capture in List 1. This incidence can happen mainly due to individual's behavior or change in interviewer's behavior or operational strategies taken by the organization conducting field-work. Let us assume that the assumption of *time variation (assumption A6) in the capture probabilities* does not hold i.e. $p_{1.} = p_{.1}$ and assumption *A5* holds. In addition we denote

$$\begin{aligned}
Pr(\text{An individual is captured in List-2 | S/He is included in List-1}) &= \frac{p_{i11}}{p_{i1.}} = c, \\
Pr(\text{An individual is captured in List-2 | S/He is not included in List-1}) &= \frac{p_{i01}}{1-p_{i1.}} = p.
\end{aligned}$$

If $p < c$, then two lists are said to be positively associated and when p>c, then two lists are negatively associated. Hence, MLEs from the corresponding likelihood function

$$L_b(N,c,p) \propto \frac{N!}{(N-x_0)!} p^{x_0} c^{x_{11}} (1-c)^{x_{1.}-x_{11}} (1-p)^{2N-x_0-x_{1.}} \tag{1.11}$$

are as follows:

$$\hat{N}_b = \left[\frac{x_0}{1-\left(\frac{x_0-x_{1.}}{x_{1.}}\right)^2}, \right] \tag{1.12}$$

$$\hat{p}_b = \frac{2x_{1.}-x_0}{x_{1.}} \tag{1.13}$$

$$\text{and } \hat{c}_b = \frac{x_{11}}{x_{1.}}. \tag{1.14}$$

Remark: Though this model has some relevance for wildlife population, but for human it is not at all appropriate. We discuss this model here just as a basis for the next model.

1.6 Time-Behavioural Response Variation Model

1.6.1 Formulation & Likelihood Failure

Causal independence assumption (*A4*) is often criticised in surveys and censuses of human populations (Chandrasekar and Deming, 1949[17]). Often an individual who is captured by first attempt may have more chance to be included in the second list than the individual who has not been captured by first attempt. If it is, then the corresponding population is treated as *recapture prone*, otherwise, for reverse case, the population becomes *recapture averse*. This change in behavior of an individual at the time of second time capturing may occur due to different causes (*see* Wolter, 1986) and this feature is grossly known as behavioral response variation.

At first, we consider the assumption of homogeneity (*A5*) for a population as like last two models. Let us consider the notation of recapture probability c and other conditional probability p (same notation as stated in case of model M_b in section 1.5). Further, in addition to assumption *A6* used in model M_t, we further assume $c \neq p$ which refers a violation of *causal independence* (stated in assumption (*A4*) in DRS). Hence, the model likelihood function becomes

$$L_{tb}(\lambda) \propto \frac{N!}{(N-x_0)!} c^{x_{11}} p_{1\cdot}^{x_{1\cdot}} p^{x_{01}} (1-p_{1\cdot})^{N-x_{1\cdot}} (1-p)^{N-x_0} (1-c)^{x_{10}}, \tag{1.15}$$

where $\lambda = (N, p_{1\cdot}, p, c)$ with $\lambda \in \Lambda = \{(N, p_{1\cdot}, p, c) | N > x_0, 0 < p_{1\cdot}, p, c < 1\}$, and it consists of lesser number of sufficient statistics $(x_{11}, x_{1\cdot}, x_{\cdot 1})$ than the parameter λ. Thus, however, we have a model, (1.15), that makes intuitive sense of real scenario but is not identifiable. Next, we will see a re-parameterization in (1.15) that models the underlying causal dependence situation better but is still not identifiable.

Since, $c \neq p$, there always exists some constant ϕ (> 0) such that $c = \phi p$. This ϕ is termed as *behavioral response effect*. Now if we re-parameterize λ by $\lambda' = (N, p_{1\cdot}, p, \phi)$, then likelihood (1.15) can be reformed as

$$L_{tb}(\lambda') \propto \frac{N!}{(N-x_0)!} \phi^{x_{11}} p_{1\cdot}^{x_{1\cdot}} p^{x_{\cdot 1}} (1-p_{1\cdot})^{N-x_{1\cdot}} (1-p)^{N-x_0} (1-\phi p)^{x_{10}}. \tag{1.16}$$

Thus, when $\phi \neq 1$, the cross-product ratio for 2×2 table, $\theta_0 = \frac{p_{11} p_{00}}{p_{10} p_{01}}$, is not unity. A population is said to be *recapture prone* if and only if $\phi > 1$ or equivalently, $c > p$. Again, cross-product ratio, $\theta_0 < 1$ refers exactly the reverse picture, and therefore a population having this characteristic, is called *recapture averse*. Hence, recapture aversion is established if and only if $c < p$ or equivalently, $\phi < 1$. The parametric relations between $p_{1\cdot}$, p, c and ϕ in M_{tb} are as follows:

$$p = p_{01}/(1-p_{1\cdot}),\ c = p_{11}/p_{1\cdot}, \text{ and } c = \phi p. \tag{1.17}$$

However, inclusion of the new parameter ϕ through the re-parametrization does not help to reduce the dimension of parameter for the model in (1.15). Therefore, this current model suffers from a identifiability problem as ϕ (or p) is not identifiable but the product $\phi p = c$ is easily identifiable. The model M_{tb} acts as a generalization of the models M_t and M_b. Estimation of N in this M_{tb}-DRS context is one of our prime objects throughout all chapters in this thesis.

Remark: I. When $\phi = 1$, equivalently $c = p$, M_{tb} will be reduced to M_t. Therefore, conditional probability p will be identical to the marginal probability $p_{\cdot 1}$.
II. In other way, if $p_{1\cdot} = p_{\cdot 1}$ in M_{tb}, then, M_{tb} will be reduced to M_b.
III. In M_{tb}, ϕ is orthogonal to N (for details about parameter orthogonality, see Cox and Reid

(1987 [30]).

Likelihood (1.16) is still ill-behaved as (1.15). However, we consider likelihood function (1.16) for inferential purposes in forthcoming chapters since its underlying parameter ϕ is of interest in some instances. As per our knowledge, no successful method has been proposed for model M_{tb} in DRS, specially for the human population, except Nour (1982[68]). Nour(1982[68]) considered a population where causal independence does not hold in DRS but lists are assumed to be positively associated, that means, this situation is equivalent to the model M_{tb} with $\phi > 1$ (i.e., *recapture prone* population), as it is often likely in demographic study of human population. However, their approach is not model based and estimate of N was obtained as

$$\hat{N}_{Nour} \quad = \quad x_0 + \frac{2x_{11}x_{10}x_{01}}{(x_{11}^2 + x_{10}x_{01})}.$$

1.6.2 Mis-specification Analysis for DSE under M_{tb}

In the DRS case, the association between two sources is usually positive in demographic examples and the Chandrasekar-Deming estimate might be regarded reasonably as providing a lower limit. Greenfield (1975[45]) suggests an upper limit to the value of N while Nour (1982[68]) presents an estimate falling between these two estimates under natural assumptions related to demographic surveys. But for a population with sensitive characteristics, such as drug users, population with a disease like Common Congenital Anomaly, this association might become negative.

In the context of several real life applications on homogeneous human population, $\hat{N}_t$, equivalently $\hat{N}_{ind}$, derived from model M_t is commonly used due to its simplicity though appropriateness of M_{tb} is well-understood. Hence, a threat of model mis-specification naturally arises when independence assumption is believed to be violated. In this section we investigate how serious that threat could be. Suppose, the actual underlying model is M_{tb} with parametrization $(N, p_{1\cdot}, p, \phi)$. We have computed approximate bias and variance of the estimator $\hat{N}_{ind}$ and present it in the following theorem.

Theorem 1.6.1 *Large sample approximation to the bias and variance of* $\hat{N}_{ind} = \left(x_{1\cdot}x_{\cdot 1}/x_{11}\right)$

for estimating N are

$$\begin{aligned}
Bias\left(\hat{N}_{ind}\right)_{M_{tb}} &= N(1-p_{1\cdot})\frac{1-\phi}{\phi}+\frac{1}{\phi}\frac{(1-p_{1\cdot})(1-\phi p)}{p_{1\cdot}\phi p},\\
Var\left(\hat{N}_{ind}\right)_{M_{tb}} &= \frac{2p_{1\cdot}(1-p_{1\cdot})(\phi-1)}{\phi}+\frac{N(1-p_{1\cdot})}{\phi}\left[\frac{1-\phi p}{p p_{1\cdot}\phi}+(1-\phi)\right].
\end{aligned}$$

Proof. For the proof we will use the following Lemma.

Lemma 1.6.2 *(Raj, 1977[76], pp. 378.) Suppose x, y and z are three random variables with finite moments upto second order. Then, large sample approximation to the mean and variance of (xy/z) are*

$$\begin{aligned}
\mathrm{E}\left(\frac{xy}{z}\right) &\approx \frac{\mathrm{E}(x)\mathrm{E}(y)}{\mathrm{E}(z)}\left(1+\frac{Cov(x,y)}{E(x)E(y)}-\frac{Cov(x,z)}{E(x)E(z)}-\frac{Cov(y,z)}{E(y)E(z)}+\frac{Var(z)}{E^2(z)}\right),\\
\mathrm{Var}\left(\frac{xy}{z}\right) &\approx \frac{E^2(x)E^2(y)}{E^2(z)}\left(\frac{Var(x)}{E^2(x)}+\frac{Var(y)}{E^2(y)}+\frac{Var(z)}{E^2(z)}+2\frac{Cov(x,y)}{E(x)E(y)}-2\frac{Cov(x,z)}{E(x)E(z)}\right.\\
&\left.-2\frac{Cov(y,z)}{E(y)E(z)}\right). \qquad (1.18)
\end{aligned}$$

$$(1.19)$$

From multinomial setup of DRS, we have $E(x_{ab})=Np_{ab}$, $Var(x_{ab})=Np_{ab}(1-p_{ab})$ and $Cov(x_{ab},x_{cd})=-Np_{ab}p_{cd}$, $\forall\ a,b,c,d\in\{0,1,\cdot\}$ such that $(a,b)\neq(c,d)$. Then replacing x, y and z by $x_{1\cdot}$, $x_{\cdot 1}$ and x_{11} respectively in first result of Lemma 1.6.2, we have

$$E(\hat{N}_{ind})=Np_0+\frac{Np_{01}p_{10}}{p_{11}}\left(1+\frac{1}{N}+\frac{1-p_{11}}{Np_{11}}\right)=Np_0+\frac{Np_{01}p_{10}}{p_{11}}+\frac{p_{01}p_{10}}{p_{11}^2}.$$

Hence, Bias($\hat{N}_{ind}$)=$E(x_{1\cdot}x_{\cdot 1}/x_{11})-N=-N(1-p_0)+N(p_{01}p_{10}/p_{11})+(p_{01}p_{10}/p_{11}^2)$. Therefore, we found Bias($\hat{N}_{ind}$)=$N(1-p_{1\cdot})(1-\phi)/\phi+\frac{(1-p_{1\cdot})(1-\phi p)}{p_{1\cdot}\phi^2 p}$ based on the model parameters of M_{tb}, after some algebraic manipulation using (1.17).

Again by replacing x, y and z by $x_{1\cdot}$, $x_{\cdot 1}$ and x_{11} respectively in the result on the variance stated in Lemma 1.6.2, we have

$$V(\hat{N}_{ind})=\frac{N}{p_{11}}\left[\frac{p_{1\cdot}p_{\cdot 1}}{p_{11}}-p_{1\cdot}-p_{\cdot 1}+\frac{2}{N}\left(p_{11}p_{00}-p_{10}p_{01}\right)\right],$$

after series of algebraic simplifications. Then, using (1.17), $V(\hat{N}_{ind})$ is finally obtained in terms of the parameters of underlying model M_{tb} as

$$\begin{aligned} V(\hat{N}_{ind}) &= \frac{N(1-p_{1.})}{\phi}\left[(\phi-1)\left(\frac{2p_{1.}}{N}-1\right)\frac{1-\phi p}{p p_{1.}\phi}\right], \\ &= \frac{2p_{1.}(1-p_{1.})(\phi-1)}{\phi}+\frac{N(1-p_{1.})}{\phi}\left[\frac{1-\phi p}{p p_{1.}\phi}+(1-\phi)\right]. \end{aligned} \tag{1.20}$$

Hence, the proof of Theorem 1.6.1 is complete. ■

Clearly, when ϕ increases above one, second part of the right hand side in bias gradually goes down to 0 as $p_{1.}$ and $\phi p = c$ are expected to be more than 0.5. Hence, simple estimate $\hat{N}_{ind}$ underestimates N and its bias tends to $-N(1-p_{1.})$ as ϕ increases. Similarly, when ϕ decreases to 0, $\hat{N}_{ind}$ increasingly overestimates N. Moreover, $\hat{N}_t$ is not at all consistent if the underlying model deviates from *causal independence.* Thus, assumption of $\phi = 1$ may be risky and therefore, use of $\hat{N}_{ind}$ may lead to an inefficient estimate.

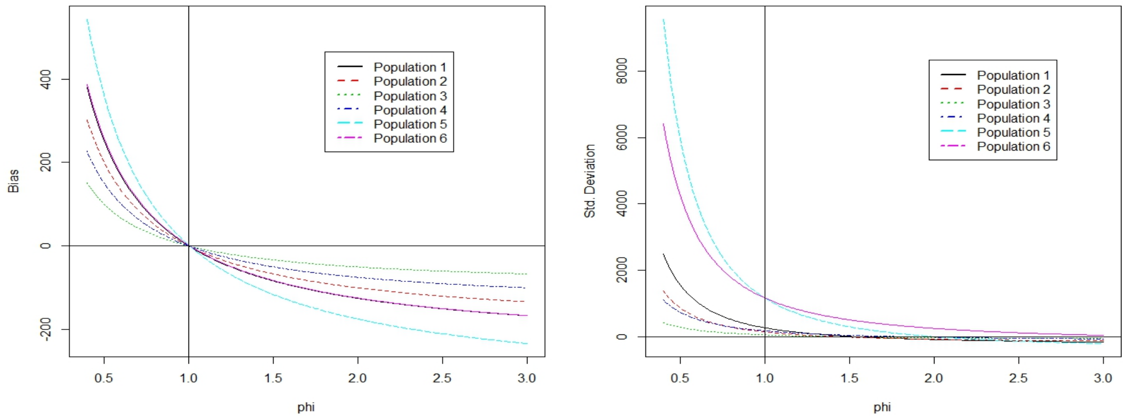
On the other hand, when ϕ is exactly 1 (i.e. *causal independence* case), bias and variance reduce to

$$\begin{aligned} Bias\left(\hat{N}_{ind}\right)_{M_t} &= \frac{(1-p_{1.})(1-p_{.1})}{p_{1.}p_{.1}}, \\ Var\left(\hat{N}_{ind}\right)_{M_{tb}} &= N\left[\frac{(1-p_{1.})(1-p_{.1})}{p_{1.}p_{.1}}\right] \end{aligned}$$

respectively, since $p = p_{.1}$ under independence. This results are also found identical to Wolter (1986[102], pp. 342). Therefore, bias will be negligible when $p_{1.}$ and $p_{.1}$ both are large.

Illustration. To present a graphical illustration of the extent of bias and variance of the DSE $\hat{N}_t$ under the model M_{tb}, here we simulate six populations with different pair of capture probabilities ($p_{1.}$, $p_{.1}$)={(0.50, 0.65), (0.60, 0.70), (0.80, 0.70), (0.70, 0.55), (0.30, 0.50), (0.50, 0.30)} characterising different plausible situations relevant for human population. We also consider two values for N, $N = 500$ and 2000. Figure 1.2 depicts the nature of bias incurred by the popular estimate $\hat{N}_t$ as well as its variance when behavioral response effect ϕ varies over positive real line. It is noticed that the variance will be negative for most of the populations when $\phi > 1.5$. Thus the above approximation for variance (in Theorem 1.6.1) does hold only up to a certain limit of ϕ; here say, that upper limit is 1.5. Thus, this approximate result is only useful to investigate the efficiency of $\hat{N}_t$ when model M_{tb} is appropriate but ϕ is not very far from 1.

Figure 1.2: Mis-specification analysis based on the bias and dispersion of $\hat{N}_t$ over ϕ for six simulated populations following model M_{tb} with $N = 500$



1.6.3 Estimation for Multiple CMR

Model M_{tb} has a strong relevance in practice for a group of homogeneous individuals when causal independence between the sample lists is not certain. Otis et al. (1978[71]) addressed the nonidentifiability problem related to the model M_{tb}. Several authors tried to solve the non-identifiability problem for homogeneous population when number of capturing occasions (T) is strictly more than two (i.e. $T \geq 3$) or three (i.e. $T \geq 4$), which are mainly focused for analysing the wildlife populations. Lloyd (1994[65]) used a martingle approach to solve the problem using an assumption that the recapture probabilities bear a constant relationship to the initial capture probabilities when number of capture occasions (T) is strictly greater than two. They also established the asymptotic equivalence of their proposed estimator and the MLEs for models M_t and M_b when the population size is large. Later, Chao et al. (2000[21]) extended this result for M_{tb} and they also established some exact and asymptotic equivalency results. Quasi-likelihood method by Chao et al. (2000[21]) and univariate Markovian approach proposed by Yang and Chao (2005[105]) also successfully solve the nonidentifiability for $T \geq 3$ and provide significant improvement over the classical solutions - unconditional and conditional MLE obtained from the popular assumption of Lloyd (1994[65]). But the identifiability problem persists in DRS since, in this case, Lloyd's assumption does not help to reduce the dimension of the model parameters. In Bayesian paradigm, Lee and Chen (1998[61]) applied the Gibbs sampling idea to the model M_{tb} but they did not use recapture data and estimates were unstable and prior sensitive. Later, Lee et al. (2003[62]) applied noninformative priors to all model parameters except ϕ, for which prior was chosen by a trial-and-error method. To discover a reasonable range for ϕ, they require large number of samples (i.e. $T \geq 4$) likely for animal population size estimation but unlikely for human. Finally, they came up with a fully Bayesian solution using MCMC, but their empirical study as well as real data application were exercised in the spirit of multiple lists ($T > 3$) problem. Wang et al. (2015[100]) also proposed a hierarchical Bayesian M_{tb} model for multiple lists with the assumption that the odds of recapture bears a constant relationship to the odds of initial capture. However, we think that the potential of the fully Bayesian method proposed by Lee et al. (2003[62]) should be investigated in this present complex DRS situation, which is not attempted earlier. In demographic context, usually $\phi > 1$ occurs which implies population is *recapture prone*. But for a population with sensitive characteristics, such as drug users, population with Common Congenital Anomaly disease etc., $\phi < 1$ and then one may call that population as *recapture averse*. When such information is available, one can hope that performance of any suitable method should improve.

1.7 Time-Heterogeneity Model (M_{th})

This model includes variation due to both time and heterogeneity in the capture probabilities. Hence, it considers only the assumptions *A4* and *A6*, but not *A5*. It can be shown that $\hat{N}_{ind}$ is not a consistent estimator in this context and approximate bias is

$$Bias(\hat{N}_{ind})_{M_{th}} = -N \frac{Cov(p_{1\cdot}, p_{\cdot 1})}{Cov(p_{1\cdot}, p_{\cdot 1}) + E(p_{1\cdot})E(p_{\cdot 1})} \tag{1.21}$$

where covariance is defined as $\frac{1}{N}\sum_i (p_{i1\cdot} - \bar{p}_{1\cdot})(p_{i\cdot 1} - \bar{p}_{\cdot 1})$, with $\bar{p}_{1\cdot} = E(p_{1\cdot}) = \frac{1}{N}\sum_i p_{i1\cdot}$ and $\bar{p}_{\cdot 1} = E(p_{\cdot 1}) = \frac{1}{N}\sum_i p_{i\cdot 1}$. When $Cov(p_{1\cdot}, p_{\cdot 1}) > 0$, estimator $\hat{N}_{ind}$ under this model shows downward bias and hence, N will be underestimated. This happens in most of human population size estimation problems in practice. Moreover, this estimator can be improved by considering one further assumption:

A7. Heterogeneous Independence. Two capture probabilities are uncorrelated in population, i.e., $Cov(p_{1\cdot}, p_{\cdot 1}) = 0$. A sufficient condition for heterogeneous independence is *homogeneity*, i.e., $p_{i1\cdot} = p_{1\cdot}$ or, $p_{i\cdot 1} = p_{\cdot 1}$ $\forall i$. So, homogeneity in at least one of the two capture probabilities can assure consistent estimation.

The above model has been outlined first time in Chandrasekar and Deming (1949 [17]).

Analysing the model M_{th} is a very tedious job in real life applications as it contains too many parameters. More number of parameters makes a model more complex and therefore, estimation will be rather challenging as nonidentifiability occurs due to high dimensional parametric model. In recent past, Link (2003[64]), Holzmann et al. (2006[50]) and Farcomeni and Tardella (2012[38]) worked on identifiability issues for heterogeneous population. However, this project is totally concentrated only on human population and in practice, most of the human populations are either homogeneous in nature or they are usually segregated into several homogeneous sub-populations, called post-strata. Moreover, assumption of Time-variation (*A6*) between two capture occasions is very much relevant in case of human population. Therefore, relevant homogeneous capture-recapture models (e.g. M_t, M_{tb}) can be applied for those post-strata and finally, we aggregate the estimates in order to obtain the estimate of size of bigger population. So, careful post-stratification has been recommended always (Wolter, 1986[102]). However, even after post-stratification, some unavoidable residual heterogeneity does matter. Gosky and Ghosh (2011[44]) found the model M_{tb} as the most robust model in estimating N based on comparative simulation study in Bayesian paradigm over all the models proposed in Otis et al. (1978[71]).

1.8 Motivating Vignettes

1.8.1 Real Data Sets and Analytical Issues

Let us explain briefly about the real data sets which are going to be analysed in order to illustrate our proposed methodologies in subsequent chapters. As per literature, the first two data sets are appropriate for model M_t, whereas other data sets demonstrate possible *causal dependence.*

Transmitted Tuberculosis Data. To estimate the number of transmitted Tuberculosis (TB) cases in three urban districts of Madrid during 1997-1999, Iñigo et al. (2003[53]) used conventional epidemiological data and the information on clustered cases obtained by DNA fingerprinting as independent Dual-record System. Using different covariates, they formed several stratifications in the population for the analysis. Here we consider the whole unstratified population and its stratification based on sex and age only to illustrate the independence model M_t.

Road Traffic Mortality Data. Another data on deaths from road traffic injuries (RTIs), available in Samuel et al. (2012[80]), is also considered in support of independence model. RTIs are responsible significantly for the preventable death and disability in developing countries and it is grossly under-reported. For that, police accident reports and a hospital-based trauma registry together build up an incomplete DRS and that is used to estimate the size of the Road Traffic Deaths separately for all inhabitants, men and women in the Lilongwe district of Malawi.

Malawi Death Data. Greenfield (1975[45]) reports a DRS data on birth, death and migration obtained from a Population Change Survey conducted by the National Statistical Office in Malawi between 1970 and 1972. The sample was stratified into five strata. To illustrate the application of the methods proposed in subsequent chapters we choose the data on death records only for two strata - (*i*) Lilongwe ($\hat{c} = 0.593$, $x_{10} > x_{01}$) and (*ii*) Other urban areas ($\hat{c} = 0.839$, $x_{10} < x_{01}$) due to its different $\hat{c}$ values and opposite nature of x_{10} and x_{01} values. Significantly lower $\hat{c}$ value helps to anticipate that the people of *Lilongwe* are less keen to give the information on deaths again in survey time than that of *Other urban areas* people.

Injection Drug user Data. Another example of DRS data is considered on injection drug user (IDU) of greater Victoria, British Columbia, Canada (Xu et al., 2014[104]). To track the changes in the prevalence of HIV and hepatitis C, the Public Health Agency of Canada developed the national, cross-sectional I-Track survey. With only two samples from the I-Track survey (phase I and phase II), some closed population mark-recapture models were implemented to estimate the number of IDUs in greater Victoria, BC. They found that estimate $\hat{N}_{ind}$ for the total number of injection drug users was 3329. They anticipated that $\hat{N}_{ind}$ might not be worthwhile for this situation and used Huggins (1989[52]) conditional likelihood approach to deal with plausible heterogeneity in the data and the estimate was 3342. Moreover, the time ordering of samples offers an opportunity to use model M_{tb}. Literature on epidemiological studies on such type of *hidden* or *hard to reach* population says that individual, who are listed in first survey, tries to avoid the listing operation in second survey. There is high possibility of *recapture-aversion* (i.e. $\phi < 1$). Low recapture rate $\hat{c} = 0.075$ strengthens this possibility.

Children Injury Data. In Epidemiological study, use of capture-recapture experiment is very popular but more than two lists are hardly ever found. The simple estimate $\hat{N}_{ind}$ assuming list-independence is widely employed in this domain, even sometimes without judging its relevancy. Here we consider a work by Jarvis et al. (2000[56]), in which authors illustrate the serious drawbacks in the use of this estimator specifically for injury related data. The problem was to enumerate those children under 15 years of age from addresses in Northumbria who were seriously injured in local Motor Vehicles Accidents (MVA) between 1 April, 1990 and 31 March, 1995. One source was *Stats19* data covering all road traffic accidents in Northumbria causing injuries to children that had been reported to the police and another was the *Hospital Episode* data (HES) covering admissions of children. The associated DRS data are presented in Table 4 of Jarvis et al. (2000[56], pp. 48) for three different classes - Cyclists, Passengers and Pedestrians. Jarvis et al. argued that children injured in MVAs as pedestrians or cyclists rarely enter insurance claims for which they have to inform police for case diary. Sometimes the police, in establishing whether an injury is serious, are recommended to contact the hospital to find out whether the child is admitted or not. It is noted that $\hat{N}_{ind}$'s are more than twice the total number of cases actually observed (x_0). Also, value of the estimate $\hat{c}$ for these three classes are 0.25, 0.40 and 0.59 respectively, which are substantially small. All these direct to the possibility of list dependency (indicating $\phi < 1$, due to very small amount of recapture) and this motivate us to include this example in our illustration.

Handloom PES Data. This is a new data from a survey aimed to estimate the undercount in the census of handloom workers residing at Gangarampur in South Dinajpur district of state West Bengal in India. The survey was post enumeration type (i.e. PES) and conducted in November, 2013 which is three months after the census operation. Handloom products have a rich tradition in this state. As an industrial and trade activity, Handloom Industry occupies a place second only to agriculture in providing livelihood to the people. The task was initiated to count all workers (master weavers and labours only) attached to Handloom Industry in West Bengal for the development of this industry. The present data on urban Gangarampur, which is going to be used here, is a part of the whole project. In the urban area, there are sixteen wards and out of them only two wards are selected randomly for PES. Sampled Ward no. 2 correctly counts 126 persons in main census operation, while PES counts 107 persons and 85 persons are matched correctly. Hence, total number of distinct captured individuals (x_0) is 148 which is very closer to the $\hat{N}_{ind} = 159$. Data related to another sampled Ward, no. 16, is as follows: correct census count 131, correct PES count 103 and matched 50 persons. Therefore, associated x_0 value is 184 but $\hat{N}_{ind}$ is 270, approximately. The nature of the data on two wards are surprisingly different except the similarity that both posses $x_{1\cdot} > x_{\cdot 1}$ and this is most probably because of temporary seasonal migration for outside work. Surveyors reported that workers in Ward 16, which is very close to town head-quarter, might be somewhat reluctant to enlist themselves in second time (i.e. at the time of PES). Moreover, most of them are working outside (other districts) and usually come home in particular seasons. That is why, Ward no. 16 results very low matches than Ward no. 2. Another reason may be that some people think that one-time enrollment at the time of census is enough.

So, underlying ϕ may be less than 1. These possibilities as well as the beliefs of the experts of Textile Directorate of Govt. drive the idea that the estimator $\hat{N}_{ind}$ is not suitable here as independence fails. Being quite certain about the homogeneity within wards from the experts of Textile Directorate, we apply the model M_{tb} to these data.

1.8.2 Methodological Issues

► M1. Estimation of Vital Statistics: Zero 4th Cell

Another estimate of N is used by considering the fourth unknown cell (in Table 1.1) to be Nil. Hence $\hat{N}$ will be just $x_{11} + x_{10} + x_{01} = x_0 = \hat{N}_{SRS}$. This estimator assuming $x_{00} = 0$ is usually practiced in Indian Sample Registration System (SRS) which is responsible for the count of vital events in India under the Office of the Registrar General, India. The rationale behind the assumption is that each and every event (i.e. new birth, death, marriage) is registered at least once by the continuous recording of vital events through Civil Registration system and the periodic retrospective survey (carried in every six months). The absolute value of the bias of this estimator is greater than the corresponding constant component of the bias of $\hat{N}_{ind}$. Raj (1977[76]) also empirically established the larger bias and almost equal variance than that of $\hat{N}_{ind}$. Besides this result, the assumption of complete capture by the dual system is not at all worthy. The general belief of the events omitted in one system are also likely to be omitted in the other results in some of the individuals remaining uncaptured and hence, x_{00} is likely to be non-zero.

► M2. Violation of Independence: Correlation Bias

In general, population consists different types of people with different capture probabilities. If anyone considers heterogeneity in his/her model then the calculation will be very much complex and cumbersome. Chandrasekar and Deming (1949[17]) first time addressed the bias in (1.21). Following their recommendation to reduce this bias-effect, post-stratification is made before the analysis of human coverage error including census undercount estimation. Wolter (1986[102]) suggested to use the simple model (M_t) or (M_b) within post-strata. But, some amount of residual heterogeneity probably remains and this residual part causes some bias in the estimate. In practice, US census bureau and some other census organizations use the formula for DSE which are not so simple as $\hat{N}_{ind}$ due to effect of weights and imputations. Each component of DSE may be subjected to bias. Moreover, another type of bias, called *correlation bias*, must play a role in DSE regardless of any other types of biases.

Actually, it originated from the failure of underlying independence assumption of DSE. This independence assumption can fail at least for any of the two following situations:

I. Heterogeneity. If heterogeneity is present in at least one of the capture probabilities (List 1 or 2), then $Cov(p_{\cdot 1}, p_{1\cdot}) > 0$ which leads to $\hat{N}_{ind}$ being underestimated from (1.21).

Chandrasekar and Deming(1949[17]) first time addressed this phenomenon. Later Otis et al. (1978[71]) modelled this factor along with *time-variation* assumption (*A6*) in Model M_{th} discussed in section 1.7.

II. Causal dependence. Existence of causal dependence either reflected by $c > p$ or $c < p$, where c and p are defined in earlier section. When $c > p$, $\hat{N}_{ind}$ will underestimate and for $c < p$, $\hat{N}_{ind}$ will overestimate.

When heterogeneity exists it is generally suspected to be of the form where persons more likely to be missed in the census are also more likely to be missed in the PES, then correlation bias is negative implying underestimation by the DSEs and in that way *Case I* implies *Case II*. While causal dependence can lead to either positive or negative biases in DSEs, generally the concern about correlation bias is that heterogeneity lead to underestimation (Bell, 2001[7]).

► M3. Bias due to a Newly Identified Source

Here, we introduce an idea on a possible source for bias in estimating N when independence model M_t is used. At the time of Post Enumeration Survey (PES) (or, second time capture occasion), more efficient and well-trained enumerators are appointed than census (or, first time capture occasion). Objective is to catch more and more persons, especially to capture those individuals who remain uncounted at the time of census (i.e. in List 1). But in some cases, this fact may raise the dependency between two systems. This causal dependence can be judged by deviation of the cross-product ratio, $\theta_0 = (p_{11}p_{00}/p_{01}p_{10})$, from 1. If in practice, more efforts are devoted to catch the missed persons in List 1, then p_{01} increases and p_{00} decreases. As a result θ_0 tends to zero, for fixed p_{11} and p_{10}. This phenomenon violates the causal independence assumption between census and PES.

In most of the countries, it has been found that correlation bias tends to lead to underestimation by DSE if persons missed in the census are less likely to be counted in the PES than those captured by the census. Hence, in human dual coverage system, $p < c$ is likely to occur. If enough training has been given to survey (i.e. PES) investigators, both p and c will increase. However, we always expect that after a careful training for survey, conducted for assessing census results, observed cell x_{01} exceeds x_{10} for which the necessary and sufficient condition is $p_{01} > p_{10}$.

► M4. Population Size Estimation in DRS: A Missing Data Analysis

Lists of individuals available from different sources on the same population through a capture-recapture type experiment are framed in a contingency table where one cell, referring absence in all lists, is always missing. Thus, the population size estimation, particularly from DRS, can also be viewed as a missing data estimation problem. Popularly a log-linear model is used for the estimation of the count in the empty cell. This model is estimated for the contingency table where the empty cell is treated as a structural zero. Once the model is found that describes the counts in the cell adequately, the parameter estimates of this model are projected onto the empty cell, yielding an estimate of the number of individuals missed by all lists (*see* Heijden et al., 2009[95]). However, in ecological models, usually direct estimation of total number of individuals is exercised, which is equivalent to the problem of estimation of the count of structurally missing cell. Thus, several relevant statistical tools which are popular in missing data analysis literature, viz. EM algorithm, Stochastic EM, Imputation, etc., can be applied.

► M5. Non-identifiability / Likelihood Failure

Inappropriateness of model M_t and practical sense of linkage between second and first time capture attempts indicates the existence of some behavioral dependence at the time of second attempts. Hence, model M_{tb} would be most relevant and certainly, appropriate choice for homogeneous closed population (*see* section 1.6.1 for detail discussion on M_{tb}). In DRS, primarily, model M_{tb} consists four unknown parameters $(N, p_{1\cdot}, p, c)$ in (1.15), whereas we have only three sufficient statistics available. However, inclusion of the new parameter ϕ through a re-parameterization does not help to reduce the dimension of parameter for model in (1.15). In (1.16), we cannot estimate both ϕ and p separately, because of the problem with identifiability.

Definition 1.8.1 *A parameter ξ for a family of distributions $\{f(x|\xi) : \xi \in \Xi\}$ is identifiable if for $\xi \neq \xi'$, $f(x|\xi)$, as a function of x, is not identical with $f(x|\xi')$.*

Therefore, this current model suffers from a identifiability problem which is a property of the model, not of an estimator. This results in *likelihood failure* and therefore difficulty arises in making inference about the underlying population size N.

In this circumstance, suitably adjusted or penalized likelihoods or pseudo-likelihoods, informative Bayesian techniques could be helpful. Inference from this ill-behaved likelihood is a challenging issue, which remains more or less untouched in the context of human population.

► M6. Knowledge on the Directional Nature of Behavioral Dependence

Earlier discussions advocate that the model M_{tb} would be more general and appropriate for closed homogeneous human population as list-independence is often violated for human populations. Moreover, M_{tb} suffers from the nonidentifiability problem. Literature (e.g., Nour, 1982[68]) suggests that right assumption on the direction of behavioral dependence might help in production of better result. If the correct knowledge on direction of the behavioral dependence is available, efficiency of estimation procedure for N would improve significantly, given the data. Therefore, researchers should give more emphasis on the identification of the possible nature of dependence, i.e. whether the given population is *recapture prone* or *averse*.

► M7. Asymptotics

Estimation of population size N is basically a statistical problem falling under Finite Population Sampling (FPS) but the relevant models from capture-recapture type experiments are not regular models. Consistency properties in regular statistical models are usually studied for eventually divergent number of observed units. In capture-recapture analyses most of the authors study the consistency property considering the inferential result as N diverges (Sen, 1985[85]). However, in this case N is an unknown parameter which makes the convergence problem more difficult to be conceived and addressed. Another aspect, often neglected in the literature, of evaluation for eventual behaviour of the inferential outcome is the one related to the accumulation of evidence gathered as the amount of trapping effort increases. Moreover, differently from N, the number T of trapping occasions is not a parameter and hence can be planned by researchers. Otherwise, consistency can be looked upon as any of the two capture probabilities $p_{1\cdot}$ and $p_{\cdot 1}$ converges to 1 or p_0 converges to 1. However, in this project, we follow the conventional approach of studying asymptotic results as $N \to \infty$.

1.9 Research Objectives & Overview

1.9.1 Objectives

Main goal of this project is to produce alternative and efficient estimates of the size N of a specified population based on only two sample capture-recapture type data-structure, called Dual-record System (DRS).

A popular model in this context, M_t, usually consists of the assumption of *causal independence* between two capture attempts. At first, we try to address some features and delimitations of this model M_t in the context of both methodological and application aspects.

Literature advocates that the assumption of causal independence does not work in reality. We encounter a problem of estimating the undercount rate in census as a regular practice of DRS by Census Bureaus of different countries. The problem of estimating undercount rate in census is equivalent to the problem of estimating N. We propose a more efficient undercount rate estimation rule by relaxing the usually practiced independence assumption. We will present this part of the work focusing on the Indian decennial census undercount rate estimation procedure.

As stated above that model M_t doesn't suit the reality, literature suggests best model for closed homogeneous human population is M_{tb}. Development of some efficient estimation procedures for N in the complex M_{tb}-DRS context, by avoiding the underlying model non-identifiability, is the prime object of this thesis.

In particular we propose new pseudo-likelihood based estimation methodologies for N which have a potential to produce a comparably efficient estimator among all other existing likelihood based estimates for model M_t. The newly proposed pseudo-likelihood based estimation methodologies are also applied for model M_{tb}. In rest of the current project, we confine ourselves to the construction of various methods in Bayesian paradigm to tackle the complex M_{tb} model with the presence of behavioral dependency among individuals at the time of second survey.

Our final aim is to develop classification strategies to identify the true nature of underlying possible behavioral dependency in individuals (discussed in 1.6). Given this identified knowledge, it is expected that more efficient inferential methodologies can be proposed for model M_{tb}.

1.9.2 Overview

In sections 1.4-1.7 of this chapter, we discussed different relevant models under Dual-record System (DRS) for closed and homogeneous human population. Model M_t, which is very popular in census undercount estimation & epidemiological studies, is presented along with different frequentist and likelihood estimates. But model M_t is often misleading as the assumption of independence does not work in many instances. Later, existence of *causal dependency* between two apparently independent data collection systems is examined through incorporation of an additional (i.e. third) data collection system. Thus, a brief idea is sketched out on Triple-record System (TRS). Thereafter, the extended model, termed as M_{tb}, in the presence of *behavioral dependency* at the time of second survey is discussed in detail. We present the likelihood and discuss the associated parameter non-identifiability problem. Extent of inaccuracy of the popular estimator $\hat{N}_t$ (or, equivalently, $\hat{N}_{ind}$) under the possible presence of *behavioral dependency* is calculated in order to understand the performance of the estimator $\hat{N}_{ind}$ when we deviate from the independence between two systems. We have studied the robustness of $\hat{N}_{ind}$ against possible departures from the basic causal independence assumptions. The estimate ($\hat{N}_t$) from independent capture-recapture model M_t is widely used in this context though appropriateness of the behavioral dependence model M_{tb} is unanimously acknowledged. Literature suggests that model M_{tb} can be free from parameter non-identifiability problem when available data collection systems is three or more. We briefly review some of the popular works done on the model M_{tb} especially for wild-life populations where more than three data collection systems are usually applied, because capture probabilities for animal are very small. Indeed, more than two systems are seldom used for human population. In M_{tb}, parameter ϕ is not estimable. This ϕ has two directions, either it is greater than 1 or it is less than 1. Nour (1982 [68]) presented an estimator in an equivalent platform assuming positive dependence (which is similar to $\phi > 1$ in M_{tb}) between two systems. This work motivates us to propose estimation strategies for N with known directional knowledge on ϕ. We believe that if the directional knowledge is correctly available, then proposal of more efficient estimation rules (for N) can be formulated successfully.

Every large census operation should undergo evaluation programs to find the sources and extent of inherent coverage errors. In chapter 2 (which is based on *Chatterjee and Mukherjee, 2016b[26]*), we briefly discuss the statistical methodology to estimate the undercount rate in Indian census based on DRS. We explicitly study the *correlation bias* (discussed in section 1.8.2) involved in the estimate, its extent, and consequences. A new potential source of bias in the estimate is identified and discussed. During the survey, more efficient enumerators compared to the census operations are appointed, and this fact may inflate the dependency between two lists and lead to a significant bias. Some examples are given to demonstrate this argument in various plausible situations. We suggest one simple and flexible approach which can control this bias. Our proposed estimator can efficiently overcome the potential bias by achieving the desired degree of accuracy (almost unbiased) with relatively higher efficiency. Overall improvements in the results are explored through simulation study on different populations.

Motivated by various applications, chapter 3 investigates the usage of a pseudo-likelihood method - profile-likelihood, explicitly for both the models M_t and M_{tb}. Therefore, an adjustment over profile likelihood is proposed for model M_{tb}. The proposed method is evaluated in terms of performance and compared with available Bayes estimate (Lee et al., 2003[62]) and $\hat{N}_t$ through extensive simulation study. Finally two real life examples with different characteristics are presented for illustration.

In chapter 4 (which is partially based on *Chatterjee and Mukherjee, 2016a[25]*), we discuss another important pseudo-likelihood function, called integrated likelihood, in the context of population size (N) estimation under DRS for both the models M_t and M_{tb}. At first an improved integrated likelihood is formulated from model M_t based on a suitably constructed weight function with the help of a novel idea by Severini (2007 [88]) using non-informative priors only. A comparative ordering is established among several likelihood and pseudo-likelihood based estimates from M_t. The resulting likelihood has several desirable properties.

For model M_{tb}, available and proposed methods are mostly developed in Bayesian paradigm due to the non-identifiability of the model M_{tb} under DRS. In chapter 4, our next contribution is in developing a non-Bayesian estimation strategy for model M_{tb} through the same improved integrated likelihood method (which is applied for model M_t) using informative priors depending upon the availability of the directional behavioral knowledge. By such construction, proposed integrated likelihood also possess several desirable properties including negligible prior sensitiveness. Simulation studies are carried out to explore the performance of the proposed method for both the models. Empirical results demonstrating efficiency and usefulness are reported. Finally, illustration based on relevant real life data sets (epidemiological and economic census) are presented separately for both the models.

Dual-record system (DRS) model with time and behavioral response variation has attracted much attention specifically in the domain of official statistics and epidemiology, as the assumption of list independence often fails. The relevant model suffers from parameter identifiability problem. One advantage of the Bayesian approaches is that even nonidentifiable models can be treated without any restrictions on parameters (Chao, 2001a[22]). Successful implementations in Lee and Chen (1998[61]), Lee et al. (2003[62]) etc. motivate us to take up Bayesian techniques in order to analyse the model M_{tb}. Complicated calculations in Bayesian approaches can now be handled by computer-intensive algorithms through the use of Gibbs sampling, a Markov chain Monte Carlo method.

Therefore, in chapter 5 (which is based on *Chatterjee and Mukherjee, 2016c[27]*) some problems in full Bayes method with flat non-informative prior are addressed, particularly in M_{tb}-DRS context. We formulate the population size estimation in DRS as a missing data problem. Two empirical Bayes approaches are developed along with a reformulation of an existing Bayes treatment, under the common roof of missing data analysis. Some features and associated posterior convergence for these methods are mentioned. Investigation through an extensive simulation study finds that our proposed approaches are comparably favourable to the existing Bayes approach for this complex model depending upon the availability of directional nature of underlying behavioral response effect. A real-data example is given to illustrate the methods.

In chapter 6, another empirical Bayes approach is proposed based on a functionally dependent informative prior in order to draw inference on N under M_{tb}-DRS setup using very simple Gibbs sampling strategy. Inference are drawn from resulting posterior for different loss function. We explore the features of this proposed method and its usages depending on the availability (or non-availability) of the information on the directional nature of behavioral response effect. Extensive simulation studies are carried out to evaluate their performance and compare with few available approaches. Finally, a real data application is provided for the model and the methods.

Problem of estimating human population size from dependent dual-record system (DRS) is a very challenging task due to the non-identifiability of M_{tb} model under DRS. In section 1.8.2, we already discussed about the possible benefit of the available knowledge on the directional nature of dependence, i.e. whether the given population is *recapture prone* or *averse.* Our contribution in chapter 7 lies in the construction of some competing strategies to identify the directional nature of underlying behavioral dependency of individuals (i.e. whether the population is *recapture prone* or *averse*). This classification strategies would be quite appealing in order to improve the inference as evident from the contemporary literature and the empirical evaluation studies of the proposed methods in chapters 5 and 6. Comparative simulation study and application to all the different real life data sets, used in other different chapters, are carried out to explore the performance of these strategies.

Chapter 8 concludes with our overall findings towards the aim of this whole project. We indicate a future research agenda generated during the course of this thesis work. However, these future works also include some of the interesting issues in this context, which are not addressed in section 1.8.

2 Census Coverage Error Estimation: With Particular Reference to India[1]

2.1 Introduction and Motivation

Census is the only primary and complete source of data which is used as an input for many of the socio-economic policies and planning by any government. Hence, knowing the exact size of population is very much essential for effective policy formulation and implementation. Moreover, census also gives extensive quantitative information to the researchers across many fields directly related to human life at a specified time canvassing each and every household in a country regardless of its size and all types of operational hazards. Aim of a census is to count and collect required information on every resident but, inevitably, some errors do exist in the census results due to a host of causes that includes non-response, duplication, erroneous enumeration, deficiency in collection strategy, its implementation and some other factors. The errors in counting are classified broadly as *coverage error*. *Coverage error* refers to either an undercount or overcount of the population. Indeed, vastness of the census undertaking might itself be responsible for existence of such errors. So serious census officials should evaluate their census operation to find the extent of different types of such errors and investigate their sources. Estimation of different types of coverage errors is also an integral part of census operation. Several authors are active in this domain of research since more than three decades (*see* Wolter, 1986 [102]; Cressie, 1989 [31]; Elliot and Little, 2005 [34]; Watcher, 2008 [97]). One important part of coverage error evaluation is to deal with methodologies to estimate coverage errors and then adjust the existing result using it. Necessity to have correct

[1] This chapter is based on *Chatterjee, K., Mukherjee, D. (2016b[26])*. An improved estimator of omission rate for census count: With particular reference to India. *Communications in Statistics - Theory and Methods* **45**, 1047-1062.

information regarding population characteristics is becoming more and more important for census officials and scientists. Dual System Estimation (DSE) [discussed in section 1.3.3] is a very simple and well known tool widely used for estimation of *coverage errors.* But it is being criticized since the last three decades by several statisticians and demographers. India, being one of the pioneers to start census evaluation program, uses this method since 1951 and as per our knowledge, no constructive discussion has been made yet on its methodological issues in Indian context. Assessment of methodological and operational aspects of coverage error estimation for Indian census is important and has enough scope to strengthen the evaluation process to elucidate better quality information from census.

Borrowing the idea from *capture-recapture* theory, DSE estimates the quality of census results. Chandrasekar and Deming (1949 [17]) first introduced this method to evaluate the performance of human vital record system. In chapter 1 we have noticed that DSE might be affected by sampling error and various types of non-sampling errors. Breiman (1994[12]) discussed some sources of general non-sampling errors in DSE from operational issues in US Census. Bias due to the presence of heterogeneity and/or behavioral dependence between capture-recapture probabilities are discussed by several researchers (*see* Chandrasekar and Deming (1949[17]), Wolter (1986[102]), Freedman et al. (1994[39]) and Stark (1998[92]). To ensure homogeneity, Chandrasekar and Deming (1949 [17]) suggested to form post-strata dividing the whole population according to various cross-sectional groups so that people in each of those post-strata are quite homogeneous. An estimate of coverage error in each post-stratum is calculated and then the coverage error for each block group (or larger administrative unit such as state, zone etc.) in the country is estimated from those post-strata level estimates according to the fraction of each post-stratum it contains. The current chapter deals with estimation methodologies of census omission rate for one arbitrary post-stratum.

A specialized survey, known as Post Enumeration Survey (PES), is conducted within three months of Indian census enumeration. This PES is used to estimate the coverage error in Indian census in terms of omission rate only for national and zonal levels. The estimates of omission rate (at per thousand individuals) in Indian census for 1991 and 2001 based on DSE are stated in Table 2.1. In section 2.2.2, we discuss how omission rate in Indian census

Table 2.1: Net Omission rate (at per thousand individuals) by residence in Indian census for 1991 and 2001.

	Rural	Urban	Total
1991	16.8	19.8	17.6
2001	16.8	39.8	23.3

Source: ORGI (2006[69]).

count through dual systems is estimated from PES. The most interesting counterpart in the bias of any Dual-record system estimate based on *causal independence* assumption is called *correlation bias* (briefly discussed in section 1.8.2). Assessment of the performance of DSE ($\hat{N}_{ind}$) in terms of overall bias, which is highly dominated by correlation bias, is made in this chapter. We mentioned a new type of source in section 1.8.2, which possibly inflates the correlation bias factor especially for a particular type of population. This is analyzed in detail in section 2.3. In section 2.4 of this chapter, a new estimator of omission rate has been proposed. It is shown that the potential source, which is responsible for increasing the correlation bias in DSE estimator, cannot affect our proposed estimator significantly. This proposed new approach increases the extent of accuracy as well as efficiency up to a certain level. Finally, in section 2.5, simulation results support this improvement over all other existing estimators.

2.2 Preliminaries

One of the pernicious features of coverage error is that an assessment of its extent cannot be made from the census data itself (Wolter, 1986 [102]). The enumerative check which was thought to be essential to evaluate the census results should be independent as otherwise it will give us strongly biased estimate. Thus, PES is done independently from census and selected households in the PES sample are checked against the census to estimate coverage errors. However, some countries use their regular survey as the second source in lieu of PES, e.g. US Census Bureau currently uses the Current Population Survey (CPS). Indeed, estimation of the size of omitted persons in census is equivalent to the estimation of the true population size. If N is that true size of a given population and C is the expected number of people counted by census, then $(N-C)$ would be the expected size of omitted people. Thus, census coverage error can be defined as

$$\begin{aligned} \text{Net Omission Rate, r} &= \frac{N-C}{C} \text{or} \\ \text{Undercount Rate, u} &= \frac{N-C}{N}. \end{aligned}$$

India estimates r whereas US census bureau measures u to report the extent of *coverage errors* in their respective censuses.

2.2.1 Post Enumeration Survey: A Dual-record System & Some Estimates

In DSE, the foremost assumption is *independence* between two lists. But if the population is not homogeneous enough then it makes the two lists correlated and serious bias can affect the DSE $\hat{N}_{ind}$. To ensure homogeneity assumption, target population is divided into several mutually exclusive post-strata (*see* Remark in section 1.3.1) and then DSE is calculated for each of those post-strata.

Let us consider a post-stratum U with size N and all individuals within are homogeneous with respect to capture probabilities. We also assume each individual in U is an inhabitant of exactly one of the M administrative or geographic clusters covering the whole population. Each cluster has on an average T individuals under U so that $MT = N$. Random sample of m clusters is selected to perform a Post Enumeration Survey (PES) independently after the census and every individual in those sampled clusters is tried to be captured. Therefore, for each post-stratum U, Census and PES act here as capture and recapture operations respectively. Therefore, individuals ($\in U$) in 1st list (made from census) are matched one-by-one with the list of individuals ($\in U$) in PES. Thus, the present data structure for each post-strata is similar to the DRS in Table 1.1, discussed in section 1.3.2. Following DSE (in section 1.3.3), here, the estimate of N will be

$$\hat{N} = \frac{M}{m}\frac{x_{.1}.x_{1.}}{x_{11}}. \tag{2.1}$$

Unfortunately, each of the ingredients used to calculate $\hat{N}$, in (2.1), is subject to error (Stark, 1998[92]). Heterogeneity within post-strata may be quite large (Freedman et al., 1994[39]). Dividing the population into many relatively small post-strata can increase within strata homogeneity. However, small strata can have high sampling variance and ratio bias (Hogan, 2001[49]). Positive dependence between two lists leads the population size to be under estimated. Such major drawback is due to a bias, termed as *correlation bias* which may occur due to failure of *causal independence* and *homogeneity* in the capture probabilities within post-strata.

For the population coverage error estimation, Ayhan and Ekni (2003 [2]) has proposed three alternative estimators, by using the Dual Record System estimation. They also proposed *Census Coverage Rate*, *Census Discrepancy Rate* and *Census Discrepancy* as coverage error measures for the censuses.

Ayhan and Ekni (2003 [2]) has proposed the following population total estimate which was based on the dual record system estimate of the Chandrasekar and Deming (1949 [17]) which was also based on the Population Census versus Sample Survey, as the two data sources. Let $N_h^{(1)}$ and $n_h^{(1)}$ respectively denote the projected population size and the selected sample size from the h^{th} region. Therefore, for the regional estimates, Ayhan and Ekni (2003 [2]) proposed the estimator for region h as

$$\hat{N}_h^{(1)} = F_h^{(1)} n_h^{(CD)},$$

where, the expansion factor

$$F_h^{(1)} = N_h^{(1)} / n_h^{(1)}$$

and $n_h^{(CD)}$ is the unweighted dual record system estimate developed by Chandrasekar and Deming (1949 [17]). Further details of this methodology can be found in their paper.

Ayhan (2000 [1]) has also proposed population total estimators in this context which are based on his Adjusted Dual Record System Estimator. The total number of events from the data in List 1 can be estimated as:

$$\hat{N}_1 = \frac{M}{m}\left[x_{1.} + (n_1 - x_{1.}) + (n_1^* - n_1).\right]$$

The total number of events from the data in List 2 can be estimated as:

$$\hat{N}_2 = \frac{M}{m}\left[x_{.1} + (n_2 - x_{.1}) + (n_2^* - n_2).\right]$$

Notations for the above estimators developed by Ayhan (2000 [1]) are clarified in section 1.3.4.

2.2.2 Coverage Error Estimation in Indian Census

Now we pay attention to the measurement of coverage error in the Indian context. India evaluates her census performance by estimating omission rate in census count. Studying the properties of this estimator is one important concern of this article.

In the census of 2001, every state was divided into three strata - rural, semi-urban and urban. Required number of Enumeration Blocks (EB) are selected linear systematically within each of three strata. EB is an administrative cluster unit equivalent to one Houselisting Block with population up to 800 (*see* Circular No. 20, Census of India, 2011 [70]). Each household in every sampled EB is visited and asked about current living status, census day residency etc. Then a 2×2 table for each EB under each stratum is prepared as in Table 1.1. Whole PES design, sampling and interviews are done independent of census operation.

Let z_{hj} and y_{hj} respectively denote the enumerated and omitted population of the j-th block of the h-th stratum (for rural and urban in each state). Total number of enumerated and omitted persons in the h-th stratum are estimated as $\hat{Z}_h = w_h \sum_{j=1}^{n_h} z_{hj}$ and $\hat{Y}_h = w_h \sum_{j=1}^{n_h} y_{hj}$ respectively. Where $w_h = \frac{N_h}{n_h}$ denotes weight for the h-th stratum, N_h= the total number of EBs which contain h-th stratum and n_h= number of sample EBs which contain h-th stratum. Thus, $\sum_h N_h = M$ and $\sum_h n_h = m$. $\hat{Z}_h$ obtained from $\hat{C} = w_h x_{1.}$ and hence, $\hat{Y}_h$ is estimated by $\hat{N}_{ind} - \hat{C} = w_h \left[(x_{1.} x_{.1}/x_{11}) - x_{1.}\right]$ as constructed for h-th stratum by summing over all sampled block level data. India evaluates her census counting by estimating the omission rate corresponds to h-th post-stratum as,

$$\hat{r}_h = \frac{\hat{N}_{ind} - \hat{C}}{\hat{C}} = \frac{\hat{Y}_h}{\hat{Z}_h}, \tag{2.2}$$

with estimated variance of omission rate, $v\hat{a}r(\hat{r}_h) = \frac{n_h(1-f_h)}{\hat{X}_h^2} \sum_{j=1}^{n_h} (y_{hj} - \hat{r}_h z_{hj})^2$. Since, $\hat{r}_h$ is based on DSE $\hat{N}_{ind}$, we call $\hat{r}_h$ as DSE of r_h. Detailed accounts of this coverage error estimation method along with other dual system methods are revisited in Rao et. al. (2009[77]). Finally, India publishes the omission rates only at national and zonal levels by age, sex and residence. One can find this in the Report on Post Enumeration Survey 2001, published by ORGI, Govt. of India. India produces the estimates on every sex-residence cross-sections for each of 13 age groups. However, details of the estimation procedure for zonal level estimates are not provided.

2.2.3 Extent of Correlation Bias

Correlation bias (CB) plays a vital role in DSE regardless of any other type of biases. This occurs due to failure of a general independence assumption that underlies the DSEs (Griffin, 2008[47]).

While causal dependence can lead to either positive or negative biases in DSEs, generally the concern about CB is heterogeneity leading to underestimation (Bell, 2001[7]). CB in $\hat{r}$ based on DSE can be modeled as $r^* - r$, where $r^* = \frac{N^* - C}{C}$ and N^*=$N + \frac{N_{10}N_{01}}{N_{11}} - N_{00}$, denotes the population total assuming both the lists are completely independent and $C = N_{1.} = N_{11} + N_{10}$ where N_{ij} bears similar meaning as the quantity x_{ij} in Table 1.1, for all $i, j \in \{1, 2\}$.

We will try to decompose the bias in $\hat{r}_h$ for h-th post-stratum into two parts. The first part is caused due to sampling and the second represents the extent of loss due to the assumption of complete independence. This second part refers to the correlation bias (CB), which is a completely non-sampling bias. Hence

$$\begin{aligned} Bias(\hat{r}_h) &= E(\hat{r}_h) - r_h \\ &= (E(\hat{r}_h) - r_h^*) + (r_h^* - r_h) \\ &= \text{Sampling Bias + Correlation Bias.} \end{aligned} \tag{2.3}$$

Now we calculate the bias of $\hat{r}$ and express it in the form of (2.3) in order to extract the extent due to CB. Let us consider new notation for the DSE $\hat{r}$ as $\hat{r}_{(1)}$ for the sake of notational clarity as the first estimator of r under consideration.

Theorem 2.2.1 *The large sample bias and variance of $\hat{r}_{(1)}$ up to the first order approximation are respectively given by*

$$Bias(\hat{r}_{(1)}) \approx \tilde{b}_1 = \frac{p_{01}}{np_{11}^2} - \frac{p_{01}p_{10}}{p_{1\cdot}p_{11}}(\theta_0 - 1), \tag{2.4}$$

$$Var(\hat{r}_{(1)}) \approx \tilde{v}_1 = \frac{p_{\cdot 1}p_{01}}{np_{11}^3}. \tag{2.5}$$

Proof. Let us consider the setup discussed in section 2.2.2. All notation are kept unchanged. Further, let a_j=$\sum_{k=1}^{T} I_{ajk}$ and a=$\sum_{j=1}^{m} a_j$, where the indicator I_{ajk} takes the value 1 with probability p_{11} when k-th individual of the j-th sample cluster is captured in both the census and PES lists, otherwise $I_{jk} = 0$. Similarly, b_j=$\sum_{k=1}^{T} I_{bjk}$ and $Pr(I_{bjk} = 1)$=p_{01}. $I_{bjk} = 1$ if k-th individual of the j-th sample EB is captured by PES only. Hence, $E(a_j)$=Tp_{11}, $E(b_j)$=Tp_{01}, $V(a_j)$= $Tp_{11}(1 - p_{11})$, $V(b_j)$= $Tp_{01}(1 - p_{01})$ and $Cov(a_j, b_{j'})$= - $Tp_{11}p_{01}$ if $j = j'$, otherwise 0. Now, $\hat{C} = \frac{M}{m}x_{1.}$ and from (2.1), we have $\hat{N} = w_h \frac{x_{1.}x_{.1}}{x_{11}}$. After some simplification, $\hat{r}_{(1)} = x_{01}/x_{11}$.

$$\begin{aligned} E(\text{a}) &= \text{mT}p_{11}, & E(\text{b}) &= \text{mT}p_{01}, \\ V(\text{a}) &= \text{mT}p_{11}(1 - p_{11}), & V(\text{b}) &= \text{mT}p_{01}(1 - p_{01}), \\ Cov(\text{a,b}) &= \textstyle\sum_{j=1}^{m} Cov(a_j, b_j) = -mTp_{11}p_{01}. \end{aligned}$$

Putting $a = x_{11}$ and $b = x_{01}$ and taking large sample approximation to the $E(\hat{r}_{(1)})$ and $V(\hat{r}_{(1)})$ with the help from Taylor's expansion (Raj, 1968 [75]), we can write

$$\begin{aligned}
E(\hat{r}_{(1)}) &= E\left(\frac{b}{a}\right) \\
&\approx \frac{E(b)}{E(a)}\left[1 - \frac{Cov(a,b)}{E(a)E(b)} + \frac{V(a)}{[E(a)]^2}\right] \\
&= \frac{p_{01}}{p_{11}} + \frac{p_{01}}{np_{11}^2}, \qquad (2.6)\\
V(\hat{r}_{(1)}) &\approx \frac{[E(b)]^2}{[E(a)]^2}\left[\frac{V(a)}{[E(a)]^2} + \frac{V(b)}{[E(b)]^2} - 2\frac{Cov(a,b)}{E(a)E(b)}\right] \\
&= \frac{p_{\cdot 1}p_{01}}{np_{11}^3}. \qquad (2.7)
\end{aligned}$$

So, $Bias(\hat{r}_{(1)}) \approx \frac{p_{01}}{np_{11}^2} - \frac{1}{p_{1\cdot}}\left(p_{00} - \frac{p_{10}p_{01}}{p_{11}}\right)$, where mT=n(say), total number of individuals in the sample to be interviewed. Correlation bias, $r^* - r = \frac{p_{01}}{p_{11}} - \frac{1}{p_{1\cdot}} + 1 = -\frac{1}{p_{1\cdot}}\left(p_{00} - \frac{p_{10}p_{01}}{p_{11}}\right)$. Hence, replacing p_{00} in terms of θ_0, we have the clear decomposed form of the $Bias(\hat{r}_{(1)})$ as in (2.4), according to (2.3). We also notice that $Bias(\hat{r}_1) \to (r^* - r) = CB(\hat{r})$ as $n \to \infty$. ■

Estimation of CB is another objective that has been tried by several practitioners, specially in the domain of coverage error or undercount rate estimation of census (*see* Bell, 1993 [6]; Wachter and Freedman, 2000 [96]; Shores and Sands, 2003 [89]; Griffin, 2008 [47]). Bell (1993 [6]) estimated the θ_0 from demographic analysis at national level. However, in this article we will not attempt to estimate this CB. We will just demonstrate how the estimate of omission rate based on DSE can be affected by the correlation bias factor in various situations.

Remark:Correlation bias depends on the population only. We can simply say that it is nothing but the difference between the actual population and the population where complete independence between two capturing is assumed.

2.3 A New Potential Source for Bias

At the time of PES, more efficient and well-trained enumerators (than census time) are appointed. Objective is to catch more and more persons, especially to capture those individuals who had small chance to be captured in the census (i.e. in List 1). But in some cases, this fact may raise the dependency between two systems. This causal dependence can be judged by deviation from cross-product ratio, $\theta_0 = (p_{11}p_{00}/p_{01}p_{10})$. Let us define c as the probability of the event that a person is detected at PES time when he/she was already captured by census and p as the probability that a person is detected at PES time when he/she was not captured by census enumerator. These two conditional probabilities characterize the behavioral dependency in individual (*see* model M_{tb} in section 1.6). Then from Theorem 2.2.1, CB in $\hat{r}$ can be expressed as

$$\begin{aligned} CB(\hat{r}_{(1)}) &= -\frac{p_{01}p_{10}}{p_{11}p_{1\cdot}}(\theta_0 - 1) & (2.8)\\ &= \frac{1-p_{1\cdot}}{p_{1\cdot}}\left(\frac{p}{c} - 1\right), & (2.9)\end{aligned}$$

using (1.17). In most of the countries, it has been found that correlation bias leads to underestimation of omission rate if persons missed in the census are less likely to be counted in PES than those captured by the census. Hence, in human dual coverage system, $p < c$ is likely to occur. If enough training has been given to survey investigators, both p and c will increase. However, we always expect that after a careful training for PES, conducted for assessing census results, observed $x_{\cdot 1}$ exceeds $x_{1\cdot}$ for which the necessary and sufficient condition is $p_{\cdot 1} > p_{1\cdot}$. But in most of the cases, this fact may raise the dependency between the two lists. If in practice, more efforts are devoted to catch the missed persons in List 1, then p_{01} increases and simultaneously, p_{00} decreases. As a result, θ_0 tends to 0, for fixed p_{11} and p_{10}. This phenomenon violates the causal independence assumption (*A4*) between census and PES. We addressed this issue in *M3* in section 1.8.2.

Example. Now, we consider some hypothetical populations in order to illustrate the above fact in various situations. We assume a post-strata with average 30 individuals per EB (or cluster) i.e. $T = 30$ and m no. of such EBs are selected with equal probabilities. For moderate and small sample, we consider $m = 30$ and 5. We consider two different populations representing different nature of dependence.

Population X. We consider a population where census captured people are more likely to be recaptured at PES time than the people not captured in census. So, $p < c$. So, we call it *recapture prone* population. A special training is given to survey investigators to capture more people. In normal sense, we can think c and p both will increase. We consider four alternative (c, p) values here. Empirically it is shown in Table 2.2 that correlation bias dominates the overall bias for both small and large sample sizes. Let us assume situation X2 or X3 occurs in lieu of situation X1, then result will be unsatisfactory. Negativity of CB increases. This apparent small increment can inflate the large population figure. When situation X3 occurs in lieu of situation X1 or X2, the result becomes little worse. But when situation X3 occurs in

Table 2.2: Four different situations of which all have $p < c$. Exact Correlation Bias (CB) and Gross Bias of $\hat{r}_{(1)}$ are shown in each case. Here T= 30.

Sl. No.	c	p	$p_{1\cdot}$	CB	Bias(m=30)	Bias(m=5)
X1	0.72	0.68	0.7	-0.0238	-0.0229	-0.0185
			0.85	-0.0098	-0.0095	-0.0080
X2	0.77	0.69	0.7	-0.0445	-0.0437	-0.0398
			0.85	-0.0183	-0.0181	-0.0167
X3	0.90	0.80	0.7	-0.0476	-0.0469	-0.0436
			0.85	-0.0196	-0.0194	-0.0182
X4	0.80	0.55	0.7	-0.1339	-0.1333	-0.1304
			0.85	-0.0551	-0.0549	-0.0539

place of situation X4, CB and bias will change significantly in favor. These scenarios are same for both the value of census capture probability $p_{1\cdot} = 0.70$ and 0.85.

Population Y. For some section of the population, people are less interested to be captured second time. This type of population is demostrating *recapture aversion*. So, for this case $p > c$. Moreover, better training at the survey time also helps $p > c$ to hold. Table 2.3 shows

Table 2.3: Four different situations of which all have $p > c$. Exact Correlation Bias (CB) and Gross Bias of $\hat{r}$ are shown in each case. Here T= 30.

Sl. No.	c	p	$p_{1\cdot}$	CB	Bias(m=30)	Bias(m=5)
Y1	0.65	0.82	0.7	0.1121	0.1134	0.1200
			0.85	0.0462	0.0466	0.0488
Y2	0.65	0.90	0.7	0.1648	0.1663	0.1735
			0.85	0.0679	0.0684	0.0708
Y3	0.68	0.76	0.7	0.0504	0.0515	0.0571
			0.85	0.0208	0.0211	0.0230
Y4	0.70	0.90	0.7	0.1224	0.1237	0.1299
			0.85	0.0504	0.0508	0.0530

that situation Y1 has equal c but greater p with respect to situation Y2. We see bias will change here from 26% to 39%. Bias is also increased significantly if parameter $p_{1\cdot}$=0.85 is assumed. In another example, if situation Y4 occurs in place of situation Y3, then estimate $\hat{r}$ will be seriously affected by correlation bias. Then bias will be approximately 28.6% from 11.7%. It is noted that people for whom $p > c$ holds might be affected more due to this kind of reason. If we consider T= 150, then CB will remain unchanged then we will certainly observe that bias is closer to CB as $mT = n$ increases. For both of the recapture averse or prone population, CB is acting as a major component of the bias.

It is clear that conducting much more efficient enumeration at PES time (than census) may sometimes distort the estimate under the assumption of *list-independence* or equivalently *causal independence*. If $p_{.1}$ increases due to increment of p_{01}, then p_{00} tends to 0. Hence θ_0 also tends to 0 and appropriateness of $\hat{r}_{(1)}$ is lost. Recapture averse population (Population-Y) may be affected more by this incident. It is observed that the estimate of omission rate in India can be seriously affected by correlation bias. As the true value of cross-product ratio (θ_0) moves far from 1, $\hat{r}_{(1)}$ will be more biased. However, estimator $\hat{r}_{(1)}$ itself sometimes do not perform well in both types of population when correlation bias is not negligible.

Remark I. For all situations under *Population X*, θ_0 is always greater than 1 and for *Population Y*, θ_0 is less than 1.

2.4 DSE-type Estimators and Proposed Affine Combination

2.4.1 DSE-type Estimator

Now, we shall develop one approach to obtain an almost unbiased estimate of omission rate r which also has smaller variance within a broad class of estimators obtained from any dual system approach.

Cross-product ratio regulates the extent of dependency in DRS. Fixing the cross-product ratio (θ_0) at a known θ ($\in [0,\infty)$), C and N can be estimated as

$$\begin{aligned}\hat{C} &= (M/m)x_{1.} \text{ and}\\ \hat{N}_\theta &= (M/m)\left(\frac{x_{1.}x_{.1}}{x_{11}} + (\theta-1)\frac{x_{10}x_{01}}{x_{11}}\right)\end{aligned} \tag{2.10}$$

respectively, based on PES. Therefore,

$$\hat{r}_\theta = \left(\frac{x_{01}}{x_{11}} + (\theta-1)\frac{x_{10}x_{01}}{x_{11}}\right) \tag{2.11}$$

For given θ, we call this $\hat{r}_\theta$ as DSE-type estimator for r. This is the generalized version of DSE estimate of N. Once the value of θ is assumed or estimated, then one can calculate $\hat{N}_\theta$.

Proposition 2.4.1 *Suppose x, y, w and z are four random variables with finite moments upto* *second order. Then, large sample approximation to the mean of* $\frac{xy}{wz}$ *is*

$$E\left(\frac{xy}{wz}\right) \approx \frac{\mathrm{E}(x)\mathrm{E}(y)}{\mathrm{E}(w)\mathrm{E}(z)}\Big[1+\frac{C(x,y)}{E(x)E(y)}+\frac{C(w,z)}{E(w)E(z)}-\frac{C(x,z)}{E(x)E(z)}-\frac{C(x,w)}{E(x)E(w)} -\frac{C(y,z)}{E(y)E(z)}-\frac{C(y,w)}{E(y)E(w)}+\frac{V(w)}{E^2(w)}+\frac{V(z)}{E^2(z)}\Big] \tag{2.12}$$

Theorem 2.4.1 *The large sample approximation to the bias and variance of* $\hat{r}_\theta$ *are*

$$Bias(\hat{r}_\theta) \approx \tilde{b}_2+\theta\frac{p_{01}p_{10}}{p_{11}p_{1\cdot}}\left(1+\frac{1}{np_{01}}+\frac{1}{np_{11}}\right), \tag{2.13}$$

$$Var(\hat{r}_\theta) \approx \theta^2\tilde{v}_1+(1-\theta)^2\tilde{v}_2+2\theta(1-\theta)\frac{p_{01}^2}{p_{1\cdot}p_{11}}\left(\frac{1}{np_{01}}-\frac{1}{n^2p_{11}p_{1\cdot}}\right), \tag{2.14}$$

where $\tilde{v}_1=\frac{p_{\cdot1}p_{01}}{np_{11}^3}$ *and* $\tilde{v}_2=\frac{p_0p_{01}}{np_{1\cdot}^3}$. *Further,* $\hat{r}_\theta$ *has minimum variance among all the DSE-type omission rate estimators* $\in\mathfrak{D}$ *at* $\theta=max\{0,(\tilde{v}_2-\tilde{v}_{12})/(\tilde{v}_1+\tilde{v}_2-2\tilde{v}_{12})\}$, *where* $\tilde{v}_{12}\approx Cov(\hat{r}_{(1)},\hat{r}_{(2)})$.

Proof. Using the simplified form of large sample approximation to the mean of (x/z), given in Raj (1977 [76]) and from the relation $\frac{x_{01}}{x_{1\cdot}}+\frac{x_{01}x_{10}}{x_{11}x_{1\cdot}}=\frac{x_{01}}{x_{11}}$, we have the following result using Proposition 2.4.1 as

$$E\left(\frac{x_{01}x_{10}}{x_{11}x_{1\cdot}}\right)=E\left(\frac{x_{01}}{x_{11}}\right)-E\left(\frac{x_{01}}{x_{1\cdot}}\right)\approx\frac{p_{01}p_{10}}{p_{11}p_{1\cdot}}\left(1+\frac{1}{np_{1\cdot}}+\frac{1}{np_{11}}\right).$$

From this result, $Bias(\hat{r}_\theta)$ in Theorem 2.4.1 is well implied. Using the result in Proposition 2.4.1, which can be proved in this context by replacing four cell frequencies from x_{01}, x_{10}, x_{11} and $x_{1\cdot}$ in the places of x, y, w and z respectively, we have $Cov(\hat{r}_{(1)},\hat{r}_{(2)})\approx\tilde{v}_{12}=\frac{p_{01}^2}{p_{1\cdot}p_{11}}\left(\frac{1}{np_{01}}-\frac{1}{n^2p_{11}p_{1\cdot}}\right)$. Hence the proof for $Var(\hat{r}_\theta)$. In the expression of $Var(\hat{r}_\theta)$, the coefficient of θ^2 is $V(\hat{r}_1-\hat{r}_2)$, which is positive and the coefficient of θ will be nonnegative when $\left(\frac{p_{1\cdot}}{p_{11}}-\frac{p_{01}}{np_{11}^2}\right)\geq\frac{p_0}{p_{1\cdot}}$ and $\hat{r}_\theta$ will have minimum variance at θ=max$\{0,\left(\frac{\tilde{v}_2-\tilde{v}_{12}}{\tilde{v}_1+\tilde{v}_2-2\tilde{v}_{12}}\right)\}$. ■

In the coverage error estimation problem, bias may be much more important than the s.e. of estimate under specific dependency assumption. So, our idea is to combine two conventional estimators and create a new weighted estimator using them, such that it would be almost unbiased as well as variance could be controlled more effectively.

Remark: Note that θ_0 is the true cross-product ratio of the two systems whereas θ is the assumed or estimated value of the underlying cross-product ratio (θ_0) in the model.

Let us first consider the method of estimating r, that Indian PES does, following Chandrasekhar-Deming (1949[17]) approach. Independence is assumed between two capture probabilities which leads to cross-product ratio, $\theta = 1$. Hence from (2.10), $\hat{N}_{\theta=1} = (M/m)(x_{1\cdot}x_{\cdot 1}/x_{11})$ and estimate of r becomes $\hat{r}_{(1)} = (x_{01}/x_{11}) = \hat{r}_{\theta=1}$ (say). Thus, $\hat{r}$, $\hat{r}_{(1)}$ (introduced in section 2.2.3) and $\hat{r}_{\theta=1}$ are exactly same estimators. The second estimator we consider is for $\theta = 0$. Here it is assumed that there is no individual left in each of the m sampled clusters who were missed by both the census and PES. Hence $\hat{N}_{\theta=0} = (M/m)x_0$, where $x_0 = x_{11} + x_{01} + x_{10}$ and this leads to the estimator, $\hat{r}_{(2)} = (x_{01}/x_{1\cdot}) = \hat{r}_{\theta=0}$ (say). Indian SRS (Sample Registration System) uses this estimator to estimate number of vital events (Raj, 1977 [76]). Now we present the following theorem and establish some basic features of $\hat{r}_{\theta=1}$ and $\hat{r}_{\theta=0}$ therefrom.

Theorem 2.4.2 *The large sample bias and variance of* $\hat{r}_{(2)}(=\hat{r}_{\theta=0})$ *up to the first order approximation are respectively given by*

$$Bias(\hat{r}_{\theta=0}) \approx \tilde{b}_2 = \frac{p_{01}}{np_{1\cdot}^2} - \theta_0 \frac{p_{01}p_{10}}{p_{1\cdot}p_{11}}, \tag{2.15}$$

$$Var(\hat{r}_{\theta=0}) \approx \tilde{v}_2 = \frac{p_0 p_{01}}{np_{1\cdot}^3}. \tag{2.16}$$

Proof. Following same steps as in the proof of the Theorem 2.2.1, one can easily find the approximate bias and variance of $\hat{r}_{(2)}$ by replacing a and b with $x_{1\cdot}$ and x_{01} respectively. ■

We will discuss later the use of the biases for $\hat{r}_{(1)}$ and $\hat{r}_{(2)}$ in section 2.4.2. From (2.4) and (2.15) it is clear that $Bias(\hat{r}_{(1)}) > Bias(\hat{r}_{(2)})$. Small bias in the estimate may produce non-negligible figure on omitted persons for large population. So, bias is the key factor in coverage error estimation. Thus, in any dual system analysis, first step is to take a specific value of θ (viz. 1 or 0) and then maintain that dependency level across all clusters operationally, so that θ_0 would be close to the chosen θ. Alternatively, θ_0 is estimated using some third source at some aggregated level and then it is assumed that same level of dependence holds for all subpopulations. But both the strategies have serious drawbacks for the following reason. For any estimator in dual system context, the bias always has two parts like (2.3). (*i*) Sampling bias in both of the $\hat{r}_{(1)}$ and $\hat{r}_{(2)}$ is found as of $O(n^{-1})$, where $n = mT$. For reasonably large n, the sampling bias will be relatively small. In any ratio estimator in this context, the numerator (i.e. the estimator of (N-C)) is always dependent on the model assumption. (*ii*) The component due to CB in $\hat{r}_{(1)}$ is negative if two lists are positively associated whereas it would be positive when two lists are negatively associated. The CB is not at all dependent on sample size. For large sample size, overall bias tends to CB and then only CB matters.

Thus, even if we adopt more efficient ratio estimators, as Hartley-Ross unbiased estimator or Murthy-Nanjamma's estimator, this will help to get rid of the existing sampling (design) bias and leave the major part (CB) unaffected. So improvement in the estimation of omission rate will not be significant with this strategy. Therefore, we employ an alternative strategy in the next section whereby the variance is minimized subject to a fixed bound on the bias.

2.4.2 Construction of Estimator

Suppose $\mathfrak{D}$, defined as $\mathfrak{D} = \{\hat{r}_\theta | \theta \in [0, \infty)\}$, denotes a class of all possible DSE-type omission rate estimators, where $\hat{r}_\theta = (\hat{N}_\theta - \hat{C})/\hat{C}$ from (2.10). So, $\hat{r}_{(t)} \in \mathfrak{D}$, for $t = 1, 2$. We consider a linear combination of the working estimators $\hat{r}_{(1)}$ and $\hat{r}_{(2)}$ as

$$\hat{r}_u = \omega_n^* \hat{r}_{(1)} + (1 - \omega_n^*)\hat{r}_{(2)}, \tag{2.17}$$

where the weight $\omega_n^* \in \mathbb{R}$ is to be estimated.

Theorem 2.4.3 *The estimator $\hat{r}_u$, having the form as in (2.17), belongs to $\mathfrak{D}$ if and only if $\omega_n^* \in [0, \infty)$.*

Proof. For an arbitrary $\omega_n^* \in \mathbb{R}$, $\hat{r}_u = \omega_n^* \hat{r}_{(1)} + (1 - \omega_n^*)\hat{r}_{(2)} = \frac{x_{01}}{x_{1\cdot}} + \omega_n^* \frac{x_{01}x_{10}}{x_{11}x_{1\cdot}} = \frac{x_{01}}{x_{11}} + (\omega_n^* - 1)\frac{x_{01}x_{10}}{x_{11}x_{1\cdot}}$, obtained from the relation $\frac{x_{01}}{x_{1\cdot}} + \frac{x_{01}x_{10}}{x_{11}x_{1\cdot}} = \frac{x_{01}}{x_{11}}$ and so, from (2.10) $\hat{r}_u$ can be expressed as $\hat{r}_{\theta = \omega_n^*}$. Hence ω_n^* must belong to $[0, \infty)$ if $\hat{r}_u \in \mathfrak{D}$.

On the contrary, if $\omega_n^* \in [0, \infty)$, then $\hat{r}_u \in \mathfrak{D}$, by the definition of $\mathfrak{D}$. ■

Since, $\omega_n^* \in [0, \infty)$, $\hat{r}_u \in \mathfrak{D}$ is clearly an affine combination of $\hat{r}_{(1)}$ and $\hat{r}_{(2)}$ with the weight ω_n^*. The motivation behind the consideration of such combination is gained from the possibility of increment in correlation bias discussed in section 2.3. Different post-strata have different level of dependency structures at the time of conducting a relatively efficient PES than census. The proposed approach follows the true underlying process and that itself will decide the weightage that should be given to the working estimator $\hat{r}_{(1)}$. Thus, $\hat{r}_u$ is flexible and robust.

Now, we will try to understand the behavior of the general DSE-type ratio estimator $\hat{r}_\theta$ over θ graphically, obtained by *monte-carlo* approximation. Figure 2.1 shows that variance of $\hat{r}_\theta$

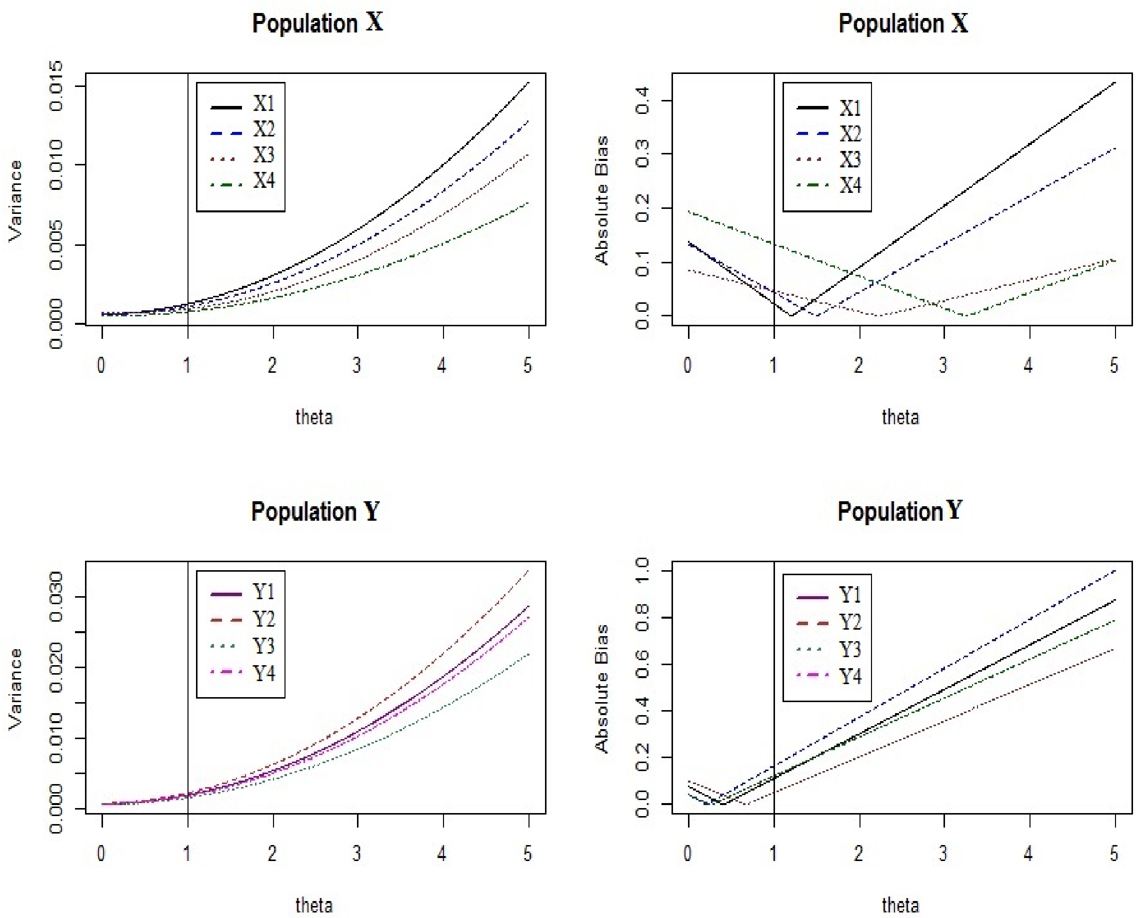


Figure 2.1: *Plot of variance and absolute bias of* $\hat{r}_\theta$ *for varying* θ *. Monte-Carlo variances and absolute biases are presented from 5000 simulations on each of four situations* - $1, 2, 3, 4$ *under* Population X *and under* Population Y.

has some increasing non-linear pattern and in the [0, 1] region, variance increases slowly. But absolute bias of $\hat{r}_\theta$ has totally reverse characteristic in [0, 1] for Population X whereas for Population Y, one can find certainly an $\hat{r}_\theta$ for $\theta \in (0, 1)$ which is better than $\hat{r}_{(1)}$ and $\hat{r}_{(2)}$ in terms of absolute bias and better than $\hat{r}_{(1)}$ in terms of variance. For Population X, there will be a trade-off between variance and absolute bias.

Estimation of $\boldsymbol{\theta_0}$**.** Bias($\hat{r}_u$) can be obtained from Bias($\hat{r}_{(1)}$) and Bias($\hat{r}_{(2)}$). Estimation of cross-product ratio, θ_0 (here, odds ratio OR) is needed for calculation of biases of $\hat{r}_{(1)}$ and $\hat{r}_{(2)}$ from complete 2×2 table. So one always wants to use as small a sample as possible to estimate θ_0. Along with three other cell values x_{11}, x_{01}, x_{10}, the last cell value x_{00} for those very small number of clusters can be calculated from authentic administrative records (if it exists) or otherwise from independent recounting and follow-up enumerations. Most common estimator for θ_0 is $\widehat{OR} = \frac{x_{11}x_{00}}{x_{01}x_{10}}$. But due to possibility of empty cell, a modified version of $\widehat{OR}$ was proposed by adding 0.5 to each cell (Haldane, 1955[48]). Parzen et al. (2002[72]) gave a median unbiased estimator of OR considering binomial distribution for each row of the 2×2 table and this method ensure that the estimate will be in $(0, \infty)$. Small sample adjustments for point estimation of OR in logarithmic scale produces very small bias and almost symmetric sampling distribution (*see* Gart and Zweifel, 1967[40]). But bias is not invariant under non-linear inverse transformation. A special adjustment over $\widehat{OR}$ was proposed by Jewell (1986[57]) for small sample estimation. This estimator posses significant reduction in bias and variance under the condition x_{01} and $x_{10} \neq 0$. The small sample adjustment by Haldane (1955[48]) on $\widehat{OR}$ is designed for bias reduction in estimating log(OR). As a result, its performance is poor in comparison to Jewell′s small sample adjusted estimator as a point estimate. Unconditionally this small sample estimator is essentially unbiased for each row (or column) sum of the 2×2 table larger than 20 (Jewell, 1986[57]). A monte

Table 2.4: Monte Carlo comparison between three methods for small sample adjusted OR estimation. Bias and RMSE are computed for each method over 5000 simulations for $T = 30$.

	Absolute Bias			RMSE		
sample size	*mmle*	*midp*	*small*	*mmle*	*midp*	*small*
	$c = 0.77$, $p = 0.69$, $p_{1.} = 0.70$, OR=1.5041					
2	0.0287	0.3286	0.0064	1.4934	1.4236	1.0586
4	0.0009	0.1618	0.0013	0.7960	0.7951	0.6994
	$c = 0.80$, $p = 0.80$, $p_{1.} = 0.70$, OR=1.0					
2	0.0755	0.2506	0.0089	1.1098	1.0431	0.7764
4	0.0313	0.1195	0.0051	0.5947	0.5956	0.5221
	$c = 0.68$, $p = 0.76$, $p_{1.} = 0.70$, OR=0.6711					
2	0.0486	0.1271	0.0043	0.5656	0.5612	0.4569
4	0.0183	0.0652	0.0039	0.3456	0.3481	0.3138
	$c = 0.80$, $p = 0.99$, $p_{1.} = 0.70$, OR=0.0404					
2	0.0867	0.2845	0.0024	0.1477	0.3383	0.1038
4	0.0319	0.1104	0.0009	0.0689	0.1266	0.0721

carlo comparison study between these three odds ratio estimation methods – (*i*) modified mle [*mmle*] due to Haldane (1955[48]) and Gart and Zweifel (1967[40]) in logarithmic scale, (*ii*) median unbiased estimator [*midp*] and (*iii*) small sample estimator [*small*] has been carried out via simulation over 5000 replications. We notice that if the data ($x_{11}, x_{01}, x_{10}, x_{00}$) from those small number of sub-sampled clusters is almost correctly known or counted or even x_{00} is little overcounted, then small sample estimator performs much better than *midp* and *mmle* for both $p_{1.} = 0.70$ and 0.85. Even for a sub-sample of size 2, the small sample estimator is almost unbiased and relatively efficient. In fact one can consider only 5% sub-sampling of m sampled clusters (for $m \geq 40$). Hence our natural aim would be to find accurate value of x_{00} for only those very small number of sub-sampled clusters such that no person having the characteristic (neither counted in census nor included in PES list) is omitted. For instance, we may select those clusters for which x_{00} is possibly found correctly. Thus as per the recommendation from empirical analysis, we adopt here the small sample estimate $\hat{\theta}_0^s$ due to Jewell (1986 [57]) for further analyses.

For fixed θ_0, the $mle's$ of $\tilde{b}_1$ and $\tilde{b}_2$ are given by

$$\hat{\tilde{b}}_1 = \frac{x_{01}}{x_{11}^2} - (\theta_0 - 1)\frac{x_{01}x_{10}}{x_{1\cdot}x_{11}}, \tag{2.18}$$

$$\hat{\tilde{b}}_2 = \frac{x_{01}}{x_{1\cdot}^2} - \theta_0\frac{x_{01}x_{10}}{x_{1\cdot}x_{11}}. \tag{2.19}$$

Finally we use jackknife bias reduction technique to the mle's $\hat{\tilde{b}}_1$ and $\hat{\tilde{b}}_2$ in (2.18) and (2.19) and then replace θ_0 by $\hat{\theta}_0^s$ in those expressions. But $Bias_{\omega_n^*}(\hat{r}_u)$ is still unknown due to ω_n^* only. Since, variance is not as much important as bias here, so, we shall fix an upper bound u_0 for the absolute bias of $\hat{r}_u$ at a desired level and then simply determine the optimal weight by minimizing the variance $V_{\omega_n^*}(\hat{r}_u)$ with respect ω_n^* over the domain $\Omega_{u_0} = \{\omega_n^* \in \mathbb{R}^+ : |B_{\omega_n^*}(\hat{r}_u)| \leq u_0\}$. we consider a computer intensive optimization technique. From (2.13) and (2.14), finally $B_{\omega_n^*}(\hat{r}_u)$ and $V_{\omega_n^*}(\hat{r}_u)$ are respectively estimated as $\widehat{B_{\omega_n^*}}(\hat{r}_u)$ and $\widehat{V_{\omega_n^*}}(\hat{r}_u)$ by using the optimal weight ω_n^*.

2.4.3 Consistency

From Theorems 2.4.1 and 2.4.3 says that $Var(\hat{r}_u)$ always decreases to zero, as $n \to \infty$, irrespective of the true value of cross-product ratio (θ_0). Now (2.13) and Theorem 2.4.3 together implies that the bias $Bias(\hat{r}_\theta)$ in Theorem 2.4.1 will vanish if θ is equal to the true θ_0 for large n. Thus we establish the following

Theorem 2.4.4 *The following statements are equivalent:*
(a) $\hat{r}_u$ is MSE consistent.
(b) Bias($\hat{r}_u$) tends to 0 as $\mathrm{n} \to \infty$.
(c) $\omega_n^* \to \theta_0$ as $\mathrm{n} \to \infty$.

2.5 Simulation Study: Evaluation

In this section we present some numerical evidence that illustrate the overall betterment of the proposed estimator $\hat{r}_u$ over any $\hat{r}_{(t)}$, $t \in \{1,2\}$. Finite sample behavior of the estimator of omission rate is taken into consideration for comparison with respect to MSE and relative absolute bias. We assume same number of sampled EBs (clusters) and average EB size under the given post-stratum as in section 2.3. We also consider same hypothetical situations as in section 2.3. For each of these situations, estimate of bias and variance of the estimators $\hat{r}_{(1)}$, $\hat{r}_{(2)}$ and the proposed $\hat{r}_u$ are presented in Tables 2.5 and 2.6.

Table 2.5: Comparison of proposed and classical DSE estimators for omission rate in *Population X* based on Monte Carlo estimates of bias, variance and MSE. Here $p \le c$ and $m = 30$, $T = 30$, $u_0 = 0.001$. All numerical figures are presented in the scale of 10^{-2}.

	$p_{1.} = 0.70$				$p_{1.} = 0.85$			
Estimator	X1	X2	X3	X4	X1	X2	X3	X4
	Bias				*Bias*			
$\hat{r}_{(1)}$	-2.292	-4.373	-4.467	-13.444	-1.073	-1.898	-1.761	-5.616
$\hat{r}_{(2)}$	-13.733	-13.289	-8.311	-19.382	-5.766	-5.556	-3.338	-8.061
$\hat{r}_u$	-0.099	-0.100	-0.097	-0.099	-0.098	-0.098	-0.084	-0.100
	Variance				Variance			
$\hat{r}_{(1)}$	0.1265	0.1104	0.0935	0.0762	0.0355	0.0312	0.0265	0.0224
$\hat{r}_{(2)}$	0.0600	0.0611	0.0735	0.0464	0.0176	0.0179	0.0212	0.0140
$\hat{r}_u$	0.1729	0.1931	0.2431	0.2937	0.0452	0.0472	0.0465	0.0577
	MSE				MSE			
$\hat{r}_{(1)}$	0.1790	0.3016	0.2930	1.8836	0.0470	0.0672	0.0575	0.3377
$\hat{r}_{(2)}$	1.9459	1.8270	0.7642	3.8029	0.3501	0.3266	0.1325	0.6637
$\hat{r}_u$	0.1730	0.1931	0.2431	0.2937	0.0453	0.0474	0.0465	0.0578

The estimator $\hat{r}_u$ achieves much higher accuracy than the two existing estimators $\hat{r}_{(1)}$ and $\hat{r}_{(2)}$. It is clear from Table 2.5 that for recapture prone population (*Population X*), $\hat{r}_{(1)}$ works better than $\hat{r}_{(2)}$ in terms of accuracy and efficiency (through MSE) when $p_{1.} = 0.70$ and 0.85. $\hat{r}_u$ has larger variance than that of $\hat{r}_{(1)}$ and $\hat{r}_{(2)}$ but it has least absolute bias.

In case of recapture averse population (*Population Y*) only, the optimal weight makes (2.17) as a convex combination. Appropriateness is not clear between $\hat{r}_{(1)}$ and $\hat{r}_{(2)}$. Results have clear agreement with natural inference that as θ_0 becomes close to 0, $\hat{r}_{(2)}$ is better than $\hat{r}_{(1)}$ otherwise $\hat{r}_{(1)}$ is better. However, our proposed affine combination approach performs significantly better than the working estimator $\hat{r}_{(1)}$ for *Population B* in terms of both accuracy and efficiency (through both the variance and MSE). In section 2.3 we already showed that

Table 2.6: Comparison of proposed and classical DSE estimators for omission rate in *Population Y* based on Monte Carlo estimates of bias, variance and MSE. Here $p > c$ and $m = 30$, $T = 30$, $u_0 = 0.001$. All numerical figures are presented in the scale of 10^{-2}.

	$p_{1\cdot} = 0.70$				$p_{1\cdot} = 0.85$			
Estimator	Y1	Y2	Y3	Y4	Y1	Y2	Y3	Y4
	Bias				*Bias*			
$\hat{r}_{(1)}$	11.266	16.606	5.065	12.392	4.671	6.833	2.002	5.082
$\hat{r}_{(2)}$	-7.749	-4.274	-10.343	-4.242	-3.170	-1.780	-4.348	-1.768
$\hat{r}_u$	-0.098	-0.089	-0.099	-0.087	-0.092	-0.071	-0.096	-0.071
	Variance				Variance			
$\hat{r}_{(1)}$	0.2044	0.2321	0.1662	0.1954	0.0550	0.0615	0.0456	0.0523
$\hat{r}_{(2)}$	0.0755	0.0849	0.0687	0.0854	0.0217	0.0241	0.0199	0.0241
$\hat{r}_u$	0.1078	0.0967	0.1269	0.0972	0.0325	0.0289	0.0371	0.0291
	MSE				MSE			
$\hat{r}_{(1)}$	1.4736	2.9896	0.4227	1.7311	0.2732	0.5285	0.0857	0.3106
$\hat{r}_{(2)}$	0.6759	0.2676	1.1383	0.2653	0.1221	0.0558	0.2090	0.0553
$\hat{r}_u$	0.1079	0.0968	0.1270	0.0973	0.0326	0.0290	0.0372	0.0291

the potential source may hamper the *Population Y* more. The proposed approach introduces the estimator $\hat{r}_u$ which is almost unbiased and more efficient than $\hat{r}_{(1)}$. In the current study emphasis is given on the bias and our approach is regulated by the parameter u_0. We consider $u_0 = 0.001$ here. One can reduce it more to have approximately unbiased estimate for all populations but in that case one might lose efficiency to some extent or Ω_{u_0} might become empty. This approach has a flexible nature that one can reduce the variance of $\hat{r}_u$ at higher order sacrificing the accuracy level little bit. Hence, $\hat{r}_u$ allows some trade-off between the levels of accuracy and efficiency between two classical dual system estimators. In Table 2.7, the basic descriptive statistics of the optimal estimates of weight ω_n^* over 1000 simulations is given. Clearly when the situations are under *Population X*, optimal weights are greater than 1. For *Population Y*, estimated ω_n^* chooses a appropriate balance factor between the working estimator $\hat{r}_{(1)}$ and the $\hat{r}_{(2)}$. Two additional populations S_X5 and S_Y5, corresponding to underlying θ_0 value as 1 and 0 respectively, are considered here for checking the internal consistency of the proposed approach. Result shows that when true model has the situation S_X5 and census capture probability is 0.7, the expected weight to the working dual system estimator $\hat{r}_{(1)}$ is 0.976, while for the situation S_Y5, the average optimal weight to $\hat{r}_{(1)}$ is 0.0062. That means for larger sample size, weights are getting closer to 1 and 0 for S_X5 and S_Y5 respectively since absolute bias has a very small upper bound. It proves that the data based approach to find optimum value of the weight function ω_n^* is consistent and also fulfill the *ICR inclusion* criterion advocated by Elliot and Little (2000 [37]). The *ICR (independence*

Table 2.7: Descriptive statistics for ω_n^* over 1000 simulated samples. $T = 30$ and $u_0 = 0.001$. For each situation with each value of $p_{1\cdot}$, upper and lower values correspond with sample size $m = 30$ and 60 respectively.

		Population X				*Population Y*			
$p_{1\cdot}$	m	Situation	Mean	Median	S.D.	Situation	Mean	Median	S.D.
0.70	30	X1	1.1963	1.0811	0.6326	Y1	0.4084	0.3565	0.2530
	60		1.1957	1.1319	0.4334		0.4053	0.3806	0.1738
0.85	30		1.1751	1.0130	0.7689		0.3988	0.3173	0.3588
	60		1.1925	1.1066	0.5608		0.3906	0.3579	0.2177
0.70	30	X2	1.4704	1.3218	0.7647	Y2	0.2035	0.1706	0.1557
	60		1.4841	1.4148	0.5444		0.2015	0.1855	0.1053
0.85	30		1.4257	1.2405	0.9138		0.2007	0.1466	0.2073
	60		1.4746	1.3705	0.6805		0.1942	0.1733	0.1375
0.70	30	X3	1.9229	1.7154	1.0863	Y3	0.6743	0.5918	0.3896
	60		2.0766	1.9306	0.8619		0.6685	0.6313	0.2657
0.85	30		1.7054	1.5125	1.1829		0.6706	0.5480	0.5139
	60		1.9446	1.7743	1.0027		0.6555	0.6018	0.3382
0.70	30	X4	2.7342	2.6534	1.0347	Y4	0.2533	0.2092	0.2017
	60		3.0292	2.9657	0.8727		0.2524	0.2272	0.1374
0.85	30		2.5020	2.3235	1.1491		0.2516	0.1825	0.2679
	60		2.8571	2.8044	0.9683		0.2423	0.2123	0.1729
0.70	30	$S_X 5$	0.9757	0.8438	0.6188	$S_Y 5$	0.0062	0.0061	0.0009
	60		0.9800	0.9196	0.4316		0.0030	0.0029	0.0007
0.85	30		0.9537	0.7943	0.7241		0.0001	0.0	0.0009
	60		0.9631	0.8799	0.5263		0.0	0.0	0.0003

of capture and re-capture) inclusion is a statistical parsimony principle that says without evidence of correlation bias, one should accept the independence model.

2.6 Conclusion

The present study is concerned with the methodological issues (*M2* and *M3* discussed in section 1.8.2) in the estimation of coverage error using omission rate estimator $\hat{r}_{\theta=1}$ (or $\hat{r}_{(1)}$) at post-stratum level. It is noticed that the classical estimator $\hat{r}_{(1)}$ will be biased and affected seriously by the dominating correlation bias factor if the underlying cross-product ratio is far from 1. It may produce a bad estimate under the threat of the possible new cause discussed in section 2.3, specially for the situations under recapture averse population. A class of DSE-type ratio estimators is defined and there is a clear trade-off between bias and variance for most of the estimators belonging to $\mathfrak{D}$. The proposed estimator is simple as well as flexible. This data based weighted linear combination actually follows the true underlying process. One can use $\hat{r}_u$ in lieu of $\hat{r}_{(1)}$ to improve the estimate of omission rate in census count. Our approach helps to get rid of the possible problem due to any kind of unwanted hike in the dependency or correlation bias. The affine combination approach has potency to produce an almost unbiased estimate. Even for small subsampling to estimate θ_0, one can achieve an unbiased estimate using Jewell's adjustment. The current affine combination approach in general helps to increase the efficiency by making a trade-off with accuracy level within its bound. Indeed, we develop an almost unbiased estimator for omission rate which also depends on the small sample estimation of θ_0 but the proposed estimator is much more robust. These results are well anticipated by the rigorous understanding of the nature of $\hat{r}_\theta$ for $\theta \in [0, \infty)$ and from the flexible construction of proposed $\hat{r}_u$. We have also established the consistency of our proposed estimator theoretically and empirically. The new estimator has a great improvement over usual DSE specially for the recapture averse population. For recapture prone population our estimator performs better in terms of accuracy. Thus we conclude that performance of the proposed weighted estimator is better than the estimator in practice $\hat{r}_{(1)}$ and other candidate estimator $\hat{r}_{(2)}$ according to both the criteria of accuracy and efficiency.

3 Profile Likelihood Method

3.1 Introduction and Motivation

In the introductory chapter 1, we have discussed that the problem of population size estimation for human is a very important statistical concern which includes a vast area of application in the fields of epidemiology, demography and official statistics. Application of capture-recapture type experiment in this regard along with special events, like war, natural calamity, etc., are also very popular in interdisciplinary platform. In the context of human population, Dual-record System (DRS) is often used. Among several models for DRS discussed in sections 1.4-1.7, model M_t has received much attention in practice for homogeneous population. But this M_t model is not appropriate in most of the situations for human population where the assumption of independence between capture probabilities fails. Discussion on the violation of such independence appears in *M2* in section 1.8.2. Hence, we would have a more complicated model M_{tb}, which is structurally most satisfactory for homogeneous population. Though the relevancy of the model M_{tb} is understood in many situations, but due to lack of identifiability in DRS, M_{tb} is seldom used for human population and model M_t becomes popular for its simplicity in both demographic and epidemiological studies. Hence the issue of model mis-specification can be raised. We have analysed the effect of model mis-specification in section 1.6.2 in detail based on the proposed result in Theorem 1.6.1. To overcome this non-identifiability issue of model M_{tb} (see M5 in section 1.8.2), several fully Bayesian techniques are postulated (*see* Lee and Chen, 1998 [61]; Lee et. al., 2003 [62]) with flat informative priors. However, in Bayesian paradigm, difficulty may arise as the resulting estimator for N may be very sensitive to the choice of prior. We will discuss this issue further in chapters 5 and 6 later. This possible threat for using Bayesian techniques motivate us to consider a non-Bayesian technique, say, pseudo-likelihood methods. In the present chapter we consider profile likelihood as a suitable choice of pseudo-likelihood method.

Profile likelihood approach replaces the nuisance parameter present in the model by its conditional MLE. Early discussions on the elimination of nuisance parameters and profile likelihood function are found in Cox (1975[29]) and Basu (1977[5]). Later, modified profile likelihood function was introduced by Barndorff-Nielsen (1983[3]); *see* also Barndorff-Nielsen and Cox (1994 [4], Ch. 8) to explore the properties of modified profile likelihood. Sometimes modified profile likelihood becomes difficult to calculate as it requires determination of a ancillary statistic. Therefore, adjusted profile likelihood function was developed by Cox and Reid (1987 [30]), though it has a limitation that it requires an orthogonal parametrization. In any capture-recapture type model, all the parameters except N are commonly regarded as nuisance parameters. In this context, profile and adjusted profile likelihood has been studied by Bolfarine et al. (1992 [10]) for independence model M_t. In this article, we confine ourselves to the profile likelihood that can summarize the set of likelihoods $\{L(N,\psi):\psi\in\Psi\}$ over Ψ, the domain of nuisance parameter ψ and some of its relevant modifications. We explicitly investigate the potentiality in application of these profile likelihood and related modifications for both the models M_t and M_{tb} in DRS context. We also propose a new adjustment to the profile likelihood for the generic model M_{tb} so that resulting pseudo-likelihood function can be used for estimating N efficiently. Our aim is to estimate the population size N from model M_{tb}-DRS based on these profile likelihood and related modifications in this chapter.

In the next section, the usefulness of profile and modified profile likelihood functions are investigated in connection to models of interest. Therefrom, we develop an adjustment to the profile likelihood for M_{tb} in section 3.3 to get rid of the identifiability problem. Evaluation and comparison of the proposed approach with the existing Bayes approach, developed by Lee et al. (2003), is carried out through an extensive simulation study. In addition to that, comparative graphical investigations on the performance and robustness of the proposed approach are carried out in this section, against the commonly used estimate $\hat{N}_{ind}$. Illustration of all the competing methods through two real datasets are presented in section 3.4. Finally in section 3.5, we summarize our findings and provide some comments about the usefulness of existing and our proposed profile likelihood based approaches.

3.2 Profile Likelihood Method and its Modifications

Let us consider a statistical model with likelihood function $L(\lambda)$ with $\lambda=(\theta,\psi)$, where θ is parameter of interest and ψ represents nuisance parameter, both may be vector valued.

Presence of more nuisance parameters in the model might affect the comparative inferential study based on the likelihood (Severini, 2000 [87]). Problem of eliminating the nuisance parameter can be handled in several ways. The problem of eliminating ψ is statistically equivalent to finding a function that can summarize the set of likelihoods $\mathscr{L}^* = \{L(\theta, \psi|\underline{\mathbf{x}}) : \psi \in \Psi\}$ over Ψ. The resulting function $L^*(\theta)$, as a function of θ, is used to some extent as a likelihood function as if the inference frame has θ as the full parameter. Although $L^*(\theta)$ might have many of the properties of a genuine likelihood function, in general it is not exactly a genuine likelihood function and, hence, inferences based on this assumption may be misleading, particularly when ψ is high-dimensional. Technically, such function $L^*(\theta)$ is referred to as pseudo-likelihood function of θ by summarizing $L(\theta, \psi|\underline{\mathbf{x}})$ over Ψ. This kind of pseudo-likelihood function includes profile likelihood and integrated likelihood, the former being of the interest in this chapter. There are several other kinds of pseudo-likelihood functions in the literature, such as marginal, conditional and partial likelihoods. Profile likelihood function† is one of the popular pseudo likelihood functions and here, nuisance parameter is replaced by its conditional MLE based on the interest parameter and data. Modified profile likelihoods (Barndorff-Nielsen, 1983 [3]) and adjusted profile likelihoods (Cox and Reid, 1987 [30]) are basically modifications to the profile likelihood function.

3.2.1 Profile Likelihood (PL) Method

Profile likelihood (PL) approach summarizes $\mathscr{L}^*$ at $\psi = \hat{\psi}_\theta$, the conditional *mle* of ψ for given θ. Thus, in profile likelihood, $L^*(\theta)$ becomes $L(\theta, \hat{\psi}_\theta) = L^P(\theta)$, say. Therefore, estimation of θ is obtained by maximizing $L^P(\theta)$ considering as a likelihood function of θ. In general, it is not a proper likelihood function. Thus, inference based on this assumption may be misleading, specifically when ψ is high-dimensional.

In the context of independent model, M_t, PL for interest parameter $\theta = N$ is given by

$$L_t^P(N) = L_t(N, \hat{p}_{1\cdot;N}, \hat{p}_{\cdot 1;N}) = \frac{N!}{(N-x_0)!}(N-x_{1\cdot})^{N-x_{1\cdot}}(N-x_{\cdot 1})^{N-x_{\cdot 1}} N^{-2N},$$

for $N \geq max(x_{1\cdot}, x_{1\cdot}, x_0) = x_0$. Here, as elsewhere in the paper, multiplicative terms not depending on N in likelihood function of N have been ignored. $L_t^P(N+1)/L_t^P(N) > 1$ implies $N < (x_{1\cdot}x_{\cdot 1}/x_{11}) - 1$. Thus, $L_t^P(N)$ is increasing in N for $N < (x_{1\cdot}x_{\cdot 1}/x_{11}) - 1$ and hence, when $(x_{1\cdot}x_{\cdot 1}/x_{11})$ is an integer, the corresponding *mle* $\hat{N}_t^P$ is $(x_{1\cdot}x_{\cdot 1}/x_{11}) - 1$. When $(x_{1\cdot}x_{\cdot 1}/x_{11})$ is not an integer, $\hat{N}_t^P$ is either $[x_{1\cdot}x_{\cdot 1}/x_{11}] - 1$ or $[x_{1\cdot}x_{\cdot 1}/x_{11}]$, according to which produces

† Integrated likelihood is another pseudo likelihood method which is the subject matter for the next chapter.

the maximum value of the profile likelihood, where $[u]$ denotes the greatest integer not greater than u, for $u \in \mathscr{R}$. $\hat{N}_t^P$ is finite iff $x_{11} > 0$. Maximum profile likelihood (PL) estimate can also be obtained by maximising $L_t^P(N)$ assuming N as a real number and using the formula for *digamma function* of any positive integer z (obtained from recursive relation), $\beta(z) = (\partial/\partial z) log(\Gamma(z)) = -\gamma + \Sigma_{a=1}^{z-1}(1/a)$, where γ is the *Euler-Mascheroni constant.*

For any parametrization of model M_{tb}, such as (1.15) or (1.16), the PL for N reduces to

$$L_{tb}^P(N) = L_{tb}(N, \hat{p}_{1\cdot;N}, \hat{p}, \hat{c}) = \frac{N!}{(N-x_0)!}(N-x_0)^{N-x_0} N^{-N},$$

for $N > x_0$, as PL is parametrization invariant. Clearly $L_{tb}^P(N)$ is decreasing for $N > x_0$ as $\prod_{i=1}^{x_0-1}(1-\frac{i}{N}) < (1-\frac{1}{N})^{x_0-1}$. It can be written that $L_{tb}^P(N) = (1-\frac{x_0}{N})^{N-x_0}\prod_{i=1}^{x_0-1}(1-\frac{i}{N}) < (1-\frac{1}{N})^{N-1}$. Now as $(1-\frac{1}{N})^{N-1} \downarrow N$, $L_{tb}^P(N)$ is a decreasing function in N for $N > x_0$. Hence, *mle* will be the lower bound of N i.e. $\hat{N}_{tb}^P = (x_0+1)$. It is clear that profile likelihood is not useful, as it stands, for estimating the population size N.

3.2.2 Modified Profile Likelihood (MPL) and Its Approximation (AMPL)

Since marginal and conditional likelihoods are not available for M_{tb}, the idea is to use a suitable modification to the profile likelihood. Several such modifications are suggested in the literature. PL cannot approximate a marginal or conditional likelihood function and that leads to poor performance. We now discuss a modification to the profile likelihood function. In general, modified profile likelihood (MPL) proposed by Barndorff-Nielsen (1983) is written as

$$L^{MP}(\theta) = D(\theta)|\hat{j}_{\psi\psi}(\theta,\hat{\psi}_\theta)|^{-1/2} L^P(\theta), \tag{3.1}$$

where $D(\theta) = |\frac{\partial \hat{\psi}_\theta}{\partial \hat{\psi}}|^{-1}$, the inverse of jacobian $J(\theta) = \partial \underline{\mathbf{x}}/\partial \hat{\psi}_\theta \propto \partial \hat{\psi}/\partial \hat{\psi}_\theta$ and $\hat{j}_{\psi\psi}$ is the observed Fisher information of ψ for fixed θ. The actual derivation of $L^{MP}(\theta)$ as an approximation to a conditional likelihood is sketched in Severini (2000) considering $(\hat{\psi}_\theta, a)$ as sufficient with θ held fixed and a is ancillary statistic. However, we can simply express the partial derivative factor in $L^{MP}(\theta)$ as follows.

Let us denote the logarithm of any likelihood $L(\cdot)$ as $\ell(\cdot)$. Then conditional *mle* $\hat{\psi}_\theta$ implies $\frac{\partial \ell(\theta,\psi|\hat{\theta},\hat{\psi},a)}{\partial \psi}|_{\psi=\hat{\psi}_\theta} = \ell_\psi(\theta,\hat{\psi}_\theta) = 0$, as sufficient statistics may be written in terms of $(\hat{\theta},\hat{\psi},a)$,

a being ancillary. Then, by differentiating with respect to $\hat{\psi}$ we have

$$\ell_{\psi;\psi}(\theta,\hat{\psi}_\theta)\frac{\partial\hat{\psi}_\theta}{\partial\hat{\psi}}+\ell_{\psi;\hat{\psi}}(\theta,\hat{\psi}_\theta)=0.$$

This implies $\frac{\partial\hat{\psi}_\theta}{\partial\hat{\psi}}=\hat{j}_{\psi\psi}(\theta,\hat{\psi}_\theta)^{-1}\ell_{\psi;\hat{\psi}}(\theta,\hat{\psi}_\theta)$, where $\hat{j}_{\psi\psi}(\theta,\hat{\psi}_\theta)=-\ell_{\psi;\psi}(\theta,\hat{\psi}_\theta)$. Hence, MPL in (3.1) may also be written in the following form

$$L^{MP}(\theta)=|\ell_{\psi;\hat{\psi}}(\theta,\hat{\psi}_\theta)|^{-1}|\hat{j}_{\psi\psi}(\theta,\hat{\psi}_\theta)|^{1/2}L^P(\theta), \tag{3.2}$$

and hence in (3.2), $D(\theta)=|\hat{j}_{\psi\psi}(\theta,\hat{\psi}_\theta)|/|\ell_{\psi;\hat{\psi}}(\theta,\hat{\psi}_\theta)|$, according to the form in (3.1).

There is an approximation to L^{MP} suggested by Severini (1998 [86]) in which $D(\theta)$ is taken as $|\hat{j}_{\psi\psi}(\theta,\hat{\psi}_\theta)|/|I(\theta,\hat{\psi}_\theta;\hat{\theta},\hat{\psi})|$, where Fisher's information

$$I(\theta,\psi;\theta_0,\psi_0)\equiv(\partial/\partial\psi_0)E\{\ell_\psi(\theta,\psi)|\theta_0,\psi_0\}$$

carries the parameters (θ,ψ) in the original score function, but the expectation part of random variables with the parameters (θ,ψ) is replaced by (θ_0,ψ_0) and $I(\theta,\psi;\theta_0,\psi_0)$ is an approximation to $\ell_{\psi;\psi_0}(\theta,\psi)$ as

$$E\{\ell_\psi(\theta,\psi)|\theta_0,\psi_0\}=\ell_\psi(\theta,\psi|\theta_0,\psi_0)+O(1) \text{ and}$$

$$\ell_{\psi;\psi_0}(\theta,\psi)=(\partial/\partial\psi_0)\ell_\psi(\theta,\psi|\theta_0,\psi_0).$$

Hence, approximated modified profile likelihood (AMPL) is

$$\widetilde{L}^{MP}(\theta)=|I(\theta,\hat{\psi}_\theta;\hat{\theta},\hat{\psi})|^{-1}|\hat{j}_{\psi\psi}(\theta,\hat{\psi}_\theta)|^{1/2}L^P(\theta). \tag{3.3}$$

Remark: Clearly, $L^{MP}(\theta)=\widetilde{L}^{MP}(\theta)$ *iff* $|\ell_{\psi;\hat{\psi}}(\theta,\hat{\psi}_\theta)|=|I(\theta,\hat{\psi}_\theta;\hat{\theta},\hat{\psi})|$, ignoring the terms not depending on θ.

Implementation to models M_t and M_{tb}:

Model M_t. The following result shows that MPL and AMPL are identical on the domain $N\geq x_0$ for model M_t. Severini (1998 [86]) only stated this result.

Theorem 3.2.1 *Both L^{MP} and $\widetilde{L}^{MP}$ are same for model M_t with $\theta = N$, $\psi = (p_{1\cdot}, p_{\cdot 1})$ and for $N \geq x_0$, it is given by*

$$\begin{aligned} L_t^{MP}(N) = \widetilde{L}_t^{MP}(N) &= \frac{N!(N - x_{1\cdot})^{N-x_{1\cdot}+1/2}(N - x_{\cdot 1})^{N-x_{\cdot 1}+1/2}}{(N - x_0)! N^{2N+1}} \\ &= L_t^P(N)(N - x_{1\cdot})^{1/2}(N - x_{\cdot 1})^{1/2} N^{-1}. \end{aligned}$$

Proof. According to parametrization $\theta = N$ and ψ=$(p_{1\cdot}, p_{\cdot 1})$, it is straightforward to show from log-likelihood $\ell^t(\theta, \psi)$ of model M_t that $(\partial/\partial\psi)\ell^t(\theta,\psi) = \ell^t_\psi(\theta,\psi) = \left(\frac{x_{1\cdot}}{p_{1\cdot}} - \frac{N-x_{1\cdot}}{1-p_{1\cdot}}, \frac{x_{\cdot 1}}{p_{\cdot 1}} - \frac{N-x_{\cdot 1}}{1-p_{\cdot 1}}\right)$ and

$$E\{\ell^t_\psi(\theta,\psi)|\theta_0,\psi_0\}\Big|_{\theta_0=\hat{\theta},\psi_0=\hat{\psi}} = \left(\frac{\hat{N}\hat{p}_{1\cdot}}{p_{1\cdot}} - \frac{N - \hat{N}\hat{p}_{1\cdot}}{1-p_{1\cdot}}, \frac{\hat{N}\hat{p}_{\cdot 1}}{p_{\cdot 1}} - \frac{N - \hat{N}\hat{p}_{\cdot 1}}{1-p_{\cdot 1}}\right).$$

Therefore, $|I^t(\theta, \hat{\psi}_\theta; \hat{\theta}, \hat{\psi})| = \frac{N^4}{(N-x_{1\cdot})(N-x_{\cdot 1})}$, since $\hat{\psi}_\theta = \left(\frac{x_{1\cdot}}{N}, \frac{x_{\cdot 1}}{N}\right)$ and

$$I^t(\theta,\psi;\hat{\theta},\hat{\psi}) = \frac{\partial}{\partial\hat{\psi}} E\{\ell^t_\psi(\theta,\psi);\theta_0,\psi_0\}\Big|_{\theta_0=\hat{\theta},\psi_0=\hat{\psi}}.$$

Again, from Severini (2000 [87]), we have $|\ell^t_{\psi;\hat{\psi}}(\theta,\hat{\psi}_\theta)| = \frac{N^4}{(N-x_{1\cdot})(N-x_{\cdot 1})}$, ignoring the terms not depending on data. Thus, $|\ell^t_{\psi;\hat{\psi}}(\theta,\hat{\psi}_\theta)| = |I^t(\theta,\hat{\psi}_\theta;\hat{\theta},\hat{\psi})|$. Therefore, from Remark 3.2.2, $L^{MP}(\theta) = \widetilde{L}^{MP}(\theta)$ for M_t and $\hat{j}^t_{\psi\psi}(\theta,\hat{\psi}_\theta) = -\ell_{\psi;\psi}(\theta,\hat{\psi}_\theta)$= $Diag\{\frac{N^3}{N-x_{1\cdot}}, \frac{N^3}{N-x_{\cdot 1}}\}$, which leads to the proof using equation (3.2). ■

An interesting relation between PL and MPL for the model M_t is formulated in the next theorem. Thereafter, Theorem 3.2.3 shows that MPL estimate is same as the ordinary likelihood estimate of N.

Theorem 3.2.2 *The maximum profile likelihood estimator, $\hat{N}_t^P$, is no greater than the maximum modified profile likelihood estimator $\hat{N}_t^{MP}$.*

Proof. Let us define $R_t^{MP}(N) = L_t^{MP}(N+1)/L_t^{MP}(N)$. Then we have

$$R_t^{MP}(N) = R_t^P(N) \times \frac{(N - x_{1\cdot} + 1)^{1/2}(N - x_{\cdot 1} + 1)^{1/2}}{(N - x_{1\cdot})^{1/2}(N - x_{\cdot 1})^{1/2}} \frac{N}{(N+1)},$$

where $R_t^P(N) = L_t^P(N+1)/L_t^P(N)$. Now, by some algebraic manipulation it can be shown that $\frac{(N-x_{1.}+1)^{1/2}(N-x_{.1}+1)^{1/2}}{(N-x_{1.})^{1/2}(N-x_{.1})^{1/2}} \frac{N}{(N+1)} \geq 1$ for all $N \geq \frac{2x_{1.}x_{.1}}{(x_{1.}+x_{.1})}$. Moreover, $\frac{2x_{1.}x_{.1}}{(x_{1.}+x_{.1})} < x_0$ always. So, $R_t^{MP}(N) \geq R_t^P(N) > 1$ for all $x_0 \leq N < (x_{1.}x_{.1}/x_{11}) - 1$. Therefore, the maximum profile likelihood estimate $\hat{N}_t^P$ is always less than or equal to the the maximum modified profile likelihood estimate $\hat{N}_t^{MP}$. ■

Theorem 3.2.3 *$L_t^{MP}(N)$ is increasing in N for $N < (x_{1.}x_{.1}/x_{11}) - 1$ and hence, the corresponding mle, $\hat{N}_t^{MP}$ is $[x_{1.}x_{.1}/x_{11}]$ if $(x_{1.}x_{.1}/x_{11})$ is not an integer; and is $(x_{1.}x_{.1}/x_{11}) - 1$, if $(x_{1.}x_{.1}/x_{11})$ is an integer.*

Proof. From Theorem 3.2.2, we have $R_t^{MP}(N) \geq R_t^P(N) > 1$ for all $N < (x_{1.}x_{.1}/x_{11}) - 1$. Now, if $L_t^P(N)$ is maximum at $N = \tilde{N}$ (say), then $R_t^P(\tilde{N}) \leq 1 < R_t^P(\tilde{N}-1) \leq R_t^{MP}(\tilde{N}-1)$ if $\tilde{N}-1 \geq x_0$. Since $R_t^P(\tilde{N}) \leq R_t^{MP}(\tilde{N})$ for $\tilde{N} \geq x_0$ i.e. $(x_{10}x_{01}/x_{11}) > 1$, one have to check whether $R_t^{MP}(\tilde{N}) > 1$ or not, for different possible $\tilde{N}$.

Now, it is clear that if $\tilde{N} = [x_{1.}x_{.1}/x_{11}] - 1$, $R_t^{MP}(\tilde{N}) > 1$ since $[x_{1.}x_{.1}/x_{11}] - 1 \leq (x_{1.}x_{.1}/x_{11}) - 1$, therefore $\hat{N}_t^{MP} = [x_{1.}x_{.1}/x_{11}]$.

If $\tilde{N} = [x_{1.}x_{.1}/x_{11}]$, $R_t^{MP}(\tilde{N}) < 1$ since $[x_{1.}x_{.1}/x_{11}] > (x_{1.}x_{.1}/x_{11})$–1, therefore $\hat{N}_t^{MP} = [x_{1.}x_{.1}/x_{11}]$.

When $(x_{1.}x_{.1}/x_{11})$ is integer, $\tilde{N} = (x_{1.}x_{.1}/x_{11}) - 1$, therefore $R_t^{MP}(\tilde{N}) < 1$, hence $\hat{N}_t^{MP} = (x_{1.}x_{.1}/x_{11}) - 1$.

Hence, associated *mle* $\hat{N}_t^{MP}$ is equal to $(x_{1.}x_{.1}/x_{11}) - 1$ if $(x_{1.}x_{.1}/x_{11})$ is an integer; otherwise $\hat{N}_t^{MP} = [x_{1.}x_{.1}/x_{11}]$. All estimates are finite iff $x_{11} > 0$. ■

Thus, for $(x_{1.}x_{.1}/x_{11}) \in \mathscr{Z}^+$, the set of positive integers, $\hat{N}_t^P = \hat{N}_t^{MP} = \tilde{N}_t^{MP} = (x_{1.}x_{.1}/x_{11}) - 1$ and for $(x_{1.}x_{.1}/x_{11})$ not $\in \mathscr{Z}^+$, $\hat{N}_t^{MP} = \tilde{N}_t^{MP} = [x_{1.}x_{.1}/x_{11}] \geq \hat{N}_t^P$.

Model M_{tb}. Now we present the computation of MPL and AMPL in the context of model M_{tb}. Let us consider at first the parametrization (1.15) and $\theta = N$, $\psi = (p_{1.}, p, c)$. So, by differentiating the log-likelihood with respect to ψ, we have $\ell_\psi^{tb}(\theta,\psi) = \left(\frac{x_{1.}}{p_{1.}} - \frac{N-x_{1.}}{1-p_{1.}}, \frac{x_{01}}{p} - \frac{N-x_0}{1-p}, \frac{x_{11}}{c} - \frac{x_{10}}{1-c}\right)$. Therefore,

$E\{\ell_\psi^{tb}(\theta,\psi)|\theta_0,\psi_0\}\Big|_{\theta_0=\hat{\theta},\psi_0=\hat{\psi}} =$
$\left(\frac{\hat{N}\hat{p}_{1.}}{p_{1.}} - \frac{N-\hat{N}\hat{p}_{1.}}{1-p_{1.}}, \frac{\hat{N}\hat{p}(1-\hat{p}_{1.})}{p} - \frac{N-\hat{N}\hat{p}(1-\hat{p}_{1.})-\hat{N}\hat{p}_{1.}}{1-p}, \frac{\hat{N}\hat{c}\hat{p}_{1.}}{c} - \frac{\hat{N}(1-\hat{c})\hat{p}_{1.}}{1-c}\right)$, since $E(x_0|\theta_0,\psi_0) = p_0(1-p_{1\cdot;0})+$

$p_{1\cdot;0}$. Now,

$$\begin{aligned} I^{tb}(\theta,\psi;\hat{\theta},\hat{\psi}) &= \frac{\partial}{\partial\hat{\psi}} E\{\ell_{\psi}^{tb}(\theta,\psi)|\theta_0,\psi_0\}\Big|_{\theta_0=\hat{\theta},\psi_0=\hat{\psi}} \\ \text{and } \hat{\psi}_\theta &= \left(\frac{x_{1\cdot}}{N}, \frac{x_{01}}{N-x_{1\cdot}}, \frac{x_{11}}{x_{1\cdot}}\right) \end{aligned}$$

Hence,

$$\begin{aligned} |I^{tb}(\theta,\hat{\psi}_\theta;\hat{\theta},\hat{\psi})| &\propto N^2(N-x_{1\cdot})/(N-x_0), \\ |\ell_{\psi;\hat{\psi}}(\theta,\hat{\psi}_\theta)| &= |\hat{j}_{\psi\psi}(\theta,\hat{\psi}_\theta)|\left|\frac{\partial\hat{\psi}_\theta}{\partial\hat{\psi}}\right|, \\ \left|\frac{\partial\hat{\psi}_\theta}{\partial\hat{\psi}}\right| &= N^{-1}(N-x_{1\cdot})^{-1} \\ \text{and } |\hat{j}_{\psi\psi}(\theta,\hat{\psi}_\theta)| &= N^3(N-x_{1\cdot})^2(N-x_0)^{-1}. \end{aligned}$$

So, $|\ell^{tb}_{\psi;\hat{\psi}}(\theta,\hat{\psi}_\theta)| \propto N^2(N-x_{1\cdot})/(N-x_0)$, where

$$\ell^{tb}_{\psi;\hat{\psi}}(\theta,\hat{\psi}_\theta) = \frac{\partial}{\partial\psi_0} E\{\ell_{\psi}^{tb}(\theta,\psi)|\theta_0,\psi_0\}\Big|_{\theta_0=\hat{\theta},\psi_0=\hat{\psi},\psi=\hat{\psi}_\theta}.$$

Hence, we have $|\ell^{tb}_{\psi;\hat{\psi}}(\theta,\hat{\psi}_\theta)| = |I^{tb}(\theta,\hat{\psi}_\theta;\hat{\theta},\hat{\psi})|$, ignoring the terms not depending on $\theta = N$. Therefore, from (3.2) and (3.3), $L^{MP}_{tb}(N) = \widetilde{L}^{MP}_{tb}(N)$ and hence, the following result.

Theorem 3.2.4 *For the model M_{tb} with $\theta = N$, $\psi = (p_{1\cdot}, p, c)$, both of L^{MP}_{tb} and $\widetilde{L}^{MP}_{tb}$ is equivalent to*

$$L_{tb}(N) = \frac{N!}{(N-x_0)!}(N-x_0)^{(N-x_0+1/2)}N^{-(N+1/2)} = L^P_{tb}(N)(1-x_0/N)^{1/2},$$

for $N > x_0$.

Now, $(\partial/\partial N)\ell^{MP}_{tb}(N) = (\partial/\partial N)\ell^P_{tb}(N) + \frac{1}{2(N-x_0)} - \frac{1}{2N}$. Using the asymptotic approximation of gamma function, $log(\Gamma(z+1)) = z\{log(z)-1\} + log(z)/2 + log(2\pi)/2 + O(z^{-1})$, we have $(\partial/\partial N)\ell^P_{tb}(N) = \frac{1}{2N} - \frac{1}{2(N-x_0)} + O(N^{-3}) = O(-N^{-2}) < 0$ for $N > x_0$. Therefore, $(\partial/\partial N)\ell^{MP}_{tb}(N) = (\partial/\partial N)\ell^P_{tb}(N) + \frac{x_0}{2N(N-x_0)} = O(N^{-3}) > 0$ for $N > x_0$. Hence clearly, L^{MP}_{tb} also does not give any finite maximum likelihood estimate.

So far we have understood that M_{tb} is the most suitable underlying model that a homogeneous capture-recapture system must follow but we also notice the failure of this model even in case of modified and approximate modified profile likelihoods. In the next section, we propose a suitable adjustment to the profile likelihood function for model M_{tb} so that reasonably good estimate is available. We also discuss the conditions under which the associated estimate of N exists. The adjustment is so designed as to preserve better frequentist and robust properties than $\hat{N}_{ind}$ even in a small neighbourhood around 1.

3.3 Proposed Methodology for M_{tb}

3.3.1 Adjustment to Profile Likelihood (AdPL) and Related Properties

Understanding the failure of PL and its two modifications - MPL and AMPL, for M_{tb} here we propose an adjusted version of the profile likelihood so that resulting likelihood is useful for non-Bayesian likelihood inference. Our proposed adjusted profile likelihood (AdPL) for generic model M_{tb} with adjustment coefficient δ $(\in \mathscr{R})$ is

$$\widehat{L}^{AP}(\theta) = \left|\frac{\partial \hat{\psi}_\theta}{\partial \hat{\psi}}\right|^{-\delta} |\hat{j}_{\psi\psi}(\theta, \hat{\psi}_\theta)|^{-1/2} L^P(\theta), \tag{3.4}$$

Note that in particular, when $\phi = 1$, $M_{tb} \Rightarrow M_t$ and therefore, $\widehat{L}^{AP}(\theta)$ will be same as $L^{MP}(\theta)$ in (3.1) iff the adjustment coefficient δ is fixed at 1. That means, for model M_t, our proposed AdPL reduces to the MPL, given in Theorem 3.2.1, if $\delta = 1$.

In the context of model M_{tb} with parametrization (1.15), $\left|\frac{\partial \hat{\psi}_\theta}{\partial \hat{\psi}}\right| = N^{-1}(N - x_{1.})^{-1}$. Hence we have the following result using (3.4) and $L^{MP}_{tb}(N)$.

Theorem 3.3.1 *For model M_{tb} with $\theta = N$, $\psi = (p_{1.}, p, c)$, the adjusted profile likelihood $\forall N > x_0$, according to (3.4), is given by*

$$\begin{aligned}
\widehat{L}^{AP}_{tb}(N) &= L^P_{tb}(N) N^{2(\delta-1)}(1 - x_{1.}/N)^{\delta-1}(1 - x_0/N)^{1/2} \\
&= L^{MP}_{tb}(N) N^{2(\delta-1)}(1 - x_{1.}/N)^{\delta-1} \\
&= \frac{N!}{(N - x_0)!} N^{\delta - N - 3/2}(N - x_{1.})^{\delta-1}(N - x_0)^{N - x_0 + 1/2}
\end{aligned}$$

Now the following theorem justifies the condition on the domain of δ in order to have a finite maxima for the adjusted profile likelihood for M_{tb}.

Theorem 3.3.2 *(a) Finite maximum adjusted profile likelihood estimate of N exists for the model M_{tb} only if $\delta < 1$.*
(b) For the model M_{tb}, $\exists$ some $\delta_0 (< 1) \ni \forall \delta < \delta_0$, $\widehat{L}^{AP}_{tb}(N) \downarrow N$ and hence, corresponding mle of N tend to the lower bound $(x_0 + 1)$.

Proof. (a) Let us define $(\partial/\partial N) log \widehat{L}^{AP}(N) = \widehat{\ell}'(N)$. We have $\widehat{\ell}'_{tb}(N) = \beta(N+1) - \beta(N - x_0 + 1) - logN + (\delta - 3/2 - N)/N + (\delta - 1)/(N - x_{1\cdot}) + log(N - x_0) + (N - x_0 + 1/2)/(N - x_0)$. After some algebraic simplification using the asymptotic approximation of *digamma function* $\beta(N) = O(N^{-1})$ we have, $\widehat{\ell}'_{tb}(N) = (\delta - 1)/N + (\delta - 1)/(N - x_{1\cdot}) + A_N$, where A_N is positive quantity decreases to zero and equivalent to $O(N^{-2})$, because $\beta'(N) = O(N^{-2})$. Clearly, if $\delta = 1$, $\widehat{\ell}'_{tb}(N) > 0$, for all $N > x_0$. When $\delta > 1$, $\widehat{\ell}'_{tb}(N) = O(N^{-1}) > 0$, for all $N > x_0$. Therefore, $\widehat{L}^{AP}(N)$ is strictly increasing for $N > x_0$ if $\delta \geq 1$ and hence, finite *mle*, $\hat{N}^{AP}_{tb}$, does not exist for $\delta \geq 1$. Again if $\delta < 1$, then $\widehat{\ell}'_{tb}(N) = A_N + B_N$, where $B_N = (\delta - 1)(2N - x_{1\cdot})/N(N - x_{1\cdot}) < 0$ is increases to zero. So, there may exist some N, for which $\widehat{\ell}_{tb}(N)$ has maxima. If B_N dominates A_N for all N, then maxima coincides with the lowest value, i.e. $(x_0 + 1)$. Hence we can certainly establish that, for any $\delta < 1$, $(x_0 + 1) \leq \hat{N}^{AP}_{tb} < \infty$. Thus, finite *mle* for M_{tb} exists only when $\delta < 1$.

(b) In case of model M_{tb}, as $L^P_{tb}(N) \downarrow N$ for $N \geq x_0$ and $L^{MP}_{tb}(N) \uparrow N$ for $N > x_0$, then from Result 3.2.4, we can say that $(1 - x_0/N)^{1/2}$ increases in N with a greater rate than the rate of decrement of $L^P_{tb}(N)$. Now, $N^{2(\delta-1)}(1 - x_{1\cdot}/N)^{\delta-1}$ decreases with N for $\delta < 1$. Therefore, from *Result 3.3.1* one can definitely say that there must exist some $\delta_0 < 1 \ni \forall \delta < \delta_0$, $\widehat{L}^{AP}_{tb}(N) \downarrow N$ and hence the proof. ■

Now we try to find a suitable δ (such that $\delta_0 < \delta < 1$), rather a class of suitable δ, in order to obtain a reasonable estimate of N. Considering N as real, we found the first derivative of adjusted profile log-likelihood as $(\partial/\partial N)\widehat{\ell}_{tb}(N) = (\delta - 1)/N + (\delta - 1)/(N - x_{1\cdot}) + A_N$, where sequence A_N is positive and equivalent to $O(N^{-2})$ for fixed data since *digamma function* $\beta(N) = O(N^{-1})$. Equating this to zero we have, $(1-\delta)O(N^{-1}) = A_N$ and this implies $\delta = 1 - B_N$, where B_N is positive sequence of N and equivalent to $O(N^{-1})$. In practice, one can choose a δ such that $\delta = 1 - O(N^{-1})$.

The expression for the maximum adjusted profile likelihood estimate of N terms out to be mathematically intractable and thus obtaining an explicit solution is not possible. We explore the *frequentist* and *robustness* properties of the estimator through computation in section 3.3.2.

Remark: If we apply the proposed adjusted profile likelihood to M_t, then

$$\widehat{L}_t^{AP}(N) = L_t^{MP}(N)N^{2(\delta-1)}, \text{ for all } N \geq x_0.$$

For model M_t, analogous to Theorem 3.3.2, we have following observations:
(a) $\exists$ some $\delta_0(<1) \ni \forall \delta < \delta_0$, $\widehat{L}_t^{AP}(N) \downarrow N$ and hence, corresponding *mle* of N tend to the lower bound x_0,
(b) $\exists$ some $\delta'(>1) \ni \forall \delta > \delta'$, $\widehat{L}_t^{AP}(N)$ does not have finite estimates.

3.3.2 Simulation Study

In this section, we consider some simulated populations reflecting different possible situations under M_{tb}, to illustrate the behaviour of our proposed estimate along with some competitive estimators in DRS. First, we simulated four populations for each behavioral dependence situation ($\phi = 1.25$ and $\phi = 0.80$ respectively represents the recapture proneness and aversion) that encompasses all possible combinations. Capture probabilities for those populations, each having size $N = 500$, are structurally presented in Table 3.1. In Indian PES each EB (Enumeration Block) is approximately $100-125$ households, which means approximately 450-600 individuals. Thus, we generically analyse simulated population of size 500 in this thesis (smaller sizes would be of little practical relevance). The expected number of distinct captured individuals ($E(x_0) = N(p_{11} + p_{01} + p_{10})$) for each population is cited in Table 3.1. The first two populations for each ϕ with $p_{1\cdot} < p_{\cdot 1}$ refers the usual situation in Post

Table 3.1: Populations with $N = 500$ considered for simulations study

Population	ϕ	$p_{1\cdot}$	$p_{\cdot 1}$	$E(x_0)$	Population	ϕ	$p_{1\cdot}$	$p_{\cdot 1}$	$E(x_0)$
P1	1.25	0.50	0.65	394	P5	0.80	0.50	0.65	430
P2	1.25	0.60	0.70	422	P6	0.80	0.60	0.70	459
P3	1.25	0.80	0.70	458	P7	0.80	0.80	0.70	483
P4	1.25	0.70	0.55	420	P8	0.80	0.70	0.55	446

Enumeration Survey (PES). The last two populations with $p_{1\cdot} > p_{\cdot 1}$ are just the opposite case which is observed often in a study of the estimation of injecting drug users (IDU). Now, 200 data sets $(x_{11}, x_{\cdot 1}, x_{1\cdot})$ are generated from each of the above eight populations.

We present the adjusted profile likelihood estimate (AdPL) for each situations for different reasonable δ values. To compare the performance of our proposed method with existing

Bayesian strategy, we compute the estimates by Lee et al. (2003[62]). However, Lee et al. (2003[62]) illustrated their approach in the context of animal capture-recapture experiment with a large number of sampling occasions. Details of their computation strategy, particularly for DRS, can be found in Chatterjee and Mukherjee (2016c[27]), on which the Chapter 5 is based on. In addition, $\hat{N}_{ind}$ is also computed to empirically measure the extent of bias due to model mis-specification discussed in section 1.6.2. Final estimates of N is obtained by averaging over 200 replications. Based on those 200 estimates, sample s.e., sample RMSE (Root Mean Square Error) and 95% bootstrap confidence interval (C.I.) are also presented in Table 3.2 for $\phi = 1.25$ representing *recapture-prone* situations and Table 3.3 for $\phi = 0.80$ representing *recapture-averse* situations. For Lee's Bayes estimates, 95% credible interval (C.I.) based on sample quantile of the marginal posterior distribution of N is presented.

Table 3.2: Summary results for populations P1-P4 (representing *recapture-prone* situations) when **No** directional information on ϕ is available. Lee et al. (2003[62])'s method is used with prior $\pi(\phi) = U(0.5, 2)$.

Method			P1	P2	P3	P4
$\hat{N}_{ind}$		$\hat{N}$(s.e.)	450(14.10)	460(11.23)	480(7.07)	469(12.01)
		RMSE	51.54	41.32	20.55	32.55
		C.I.	(425,480)	(438,481)	(465,493)	(444,491)
Lee		$\hat{N}$(s.e.)	468(20.56)	483(18.45)	485(6.61)	471(8.11)
		RMSE	37.94	24.97	16.97	30.61
		C.I.	(398,561)	(426,560)	(460,513)	(422,542)
AdPl	$\delta = 1-0.75N^{-1}$	$\hat{N}$(s.e.)	486(12.15)	513(10.61)	539(7.15)	499(9.74)
		RMSE	18.86	17.01	39.82	9.61
		C.I.	(461,507)	(491,532)	(525,552)	(578,516)
	$\delta = 1-1.25N^{-1}$	$\hat{N}$(s.e.)	461(11.47)	488(10.01)	515(6.78)	476(9.27)
		RMSE	40.32	15.54	16.32	25.68
		C.I.	(439,480)	(467,506)	(501,527)	(456,493)
	$\delta = 1-1.75N^{-1}$	$\hat{N}$(s.e.)	449(11.13)	476(9.77)	504(6.60)	466(9.02)
		RMSE	51.64	25.85	7.71	35.23
		C.I.	(428,469)	(455,493)	(491,516)	(446,482)

Table 3.2 says that as $\delta(<1)$ is chosen to be closer to 1, AdPL performs better for case of low capture probabilities (P1 and P4). In other situations (P2 and P3) where capture probabilities are high, efficient adjustment coefficient δ will be $(1-1.25N^{-1})$. In other words, we try to analyse the performance from the perspective of two kinds of populations where $x_{1\cdot} < x_{\cdot 1}$

and $x_{1.} > x_{.1}$. For both kind of situations $x_{1.} < x_{.1}$ (*i.e.* P1 and P2) and $x_{1.} >_{10}$ (*i.e.* P3 and P4), AdPL performs progressively better as $\delta(< 1)$ is chosen to be closer to 1. Except P3, AdPL shows more efficient result than Lee's method. In most of the recapture prone situations, $\hat{N}_{ind}$ misleads us, particularly for the cases where capture probabilities are low and/or when underlying ϕ is far from 1.

Table 3.3: Summary results for populations P1-P4 (representing *recapture-averse* situations) when **No** directional information on ϕ is available. Lee et al. (2003[62])'s method is used with prior $\pi(\phi) = U(0.5, 2)$.

Method			P5	P6	P7	P8
$\hat{N}_{ind}$		$\hat{N}$(s.e.)	563(23.15)	550(14.94)	526(8.08)	538(14.26)
		RMSE	67.21	52.48	27.09	40.44
		C.I.	(523,615)	(524,578)	(510,541)	(513,565)
Lee		$\hat{N}$(s.e.)	474(20.80)	512(15.76)	516(6.17)	517(13.02)
		RMSE	35.58	19.83	18.71	21.75
		C.I.	(431,566)	(461,575)	(486,553)	(451,615)
AdPl	$\delta = 1 - 0.75N^{-1}$	$\hat{N}$(s.e.)	533(9.53)	562(7.44)	574(5.70)	536(8.15)
		RMSE	34.57	63.05	74.25	36.88
		C.I.	(513,552)	(547,577)	(563,584)	(521,551)
	$\delta = 1 - 1.25N^{-1}$	$\hat{N}$(s.e.)	505(9.40)	534(6.98)	548(5.21)	510(7.75)
		RMSE	10.72	35.23	48.40	13.01
		C.I.	(487,524)	(519,547)	(537,557)	(497,525)
	$\delta = 1 - 1.75N^{-1}$	$\hat{N}$(s.e.)	492(9.18)	521(6.75)	535(5.00)	499(7.52)
		RMSE	12.45	22.04	35.88	9.65
		C.I.	(474,510)	(506,534)	(525,545)	(485,512)

Similarly, when we turn to analyse the hypothetical populations with recapture averseness, Table 3.3 shows that as δ is chosen to be relatively smaller at $(1 - 1.75N^{-1})$, AdPL performs reasonably better. In low capture situations (P5 and P8), AdPL shows more efficient result than Lee's method. Table 3.3 also shows that in any recapture averse situations, $\hat{N}_{ind}$ will highly overestimate N when ϕ is substantially different from 1.

Hence, in both situations of recapture aversion and proneness, performance of $\hat{N}_{ind}$ becomes worse particularly for the populations where $x_{1.} < x_{.1}$. Lee's Bayes estimate, with prior $\pi(\phi) = U(0.5, 2)$, generally underestimates for $\phi > 1$ and overestimates for $\phi < 1$ but use of their estimate is recommended than that of $\hat{N}_{ind}$ to avoid serious model mis-specification. However, we found that our proposed adjusted profile likelihood method, with suitably chosen value of δ, can be a good alternative to Lee et al. (2003 [62]) and it performs better than Lee's in most of the situations.

Variance of $\hat{N}_{tb}^{AP}$:

It is found in at the end of section 1.6.2 in Chapter 1 that s.e.$(\hat{N}_{ind})$ is $O(N^{1/2})$ when independence holds. We also found that estimator $\hat{N}_{tb}^{AP}$ is mathematically intractable for computing its variance. Hence, to study the nature of variability in $\hat{N}_{tb}^{AP}$, a graphical comparison of the extent of variability in $\hat{N}_{tb}^{AP}$ against $\hat{N}_t$, when underlying model is M_{tb}, is presented.

Let us consider four simulated populations P2, P4, P6 and P8 introduced earlier and rename them as S1, S2, S3 and S4 respectively. From each of these four population, 200 estimates from 200 generated data sets constitutes the sampling distributions of the estimator and hence we find the bootstrap s.e. of the estimate. Similar statistics are computed for the estimator $\hat{N}_{ind} = (x_{1\cdot}x_{\cdot 1}/x_{11})$ and finally, comparative behaviour of the $ln(s.e.)$ of both the estimators $\hat{N}_{ind}$ and $\hat{N}_{tb}^{AP}$ are plotted against $ln(N)$ in Figure 3.1. This figure shows that s.e.$(\hat{N}_{tb}^{AP})$ is less than s.e.$(\hat{N}_t)$ $\forall N$. Numerical investigations carried out above suggests that the proposed adjusted profile likelihood could be more helpful in the context of population size estimation (under the model M_{tb}) and it shows better efficiency than the usual DSE estimator $\hat{N}_{ind}$ in terms of s.e.

Now we examine some frequentist as well as robustness properties of the proposed adjusted profile-likelihood estimate $\hat{N}_{tb}^{AP}$ along with $\hat{N}_{ind} = (x_{1\cdot}x_{\cdot 1}/x_{11})$.

Frequentist Coverage Performance:

Firstly, under the mis-specification threat (*see* section 1.6.2), we graphically study the coverage performance of $\hat{N}_{ind}$ as N varies to compare with the $\hat{N}_{tb}^{AP}$. We consider all the artificial populations (under M_{tb}) simulated earlier in section 3.3.2. For moderately large population (say, $N > 100$), we found both the $\hat{N}_{ind}$ and $\hat{N}_{tb}^{AP}$ to be approximately normal. Figure 3.2 shows multiple plots of the 95% *relative UCL* ($= (\hat{N} + 1.96 s.e.(\hat{N}))/N$) and *LCL* ($= (\hat{N} - 1.96 s.e.(\hat{N}))/N$) corresponding to the estimators $\hat{N}_{ind}$ and $\hat{N}_{tb}^{AP}$ over several true N for all populations P1-P8. The *Relative LCL* and *relative UCL* contains 1 with 0.95 probability. Hence, we are able to compare how much the relative confidence limits deviate from 1 with gradually increasing true N (here, it ranges from 100 to 1000). Figure 3.2 shows that relative confidence bounds of $\hat{N}_{tb}^{AP}$ are tighter as well as closer to 1 in most of the situations compared to $\hat{N}_{ind}$ for all the populations for different N values.

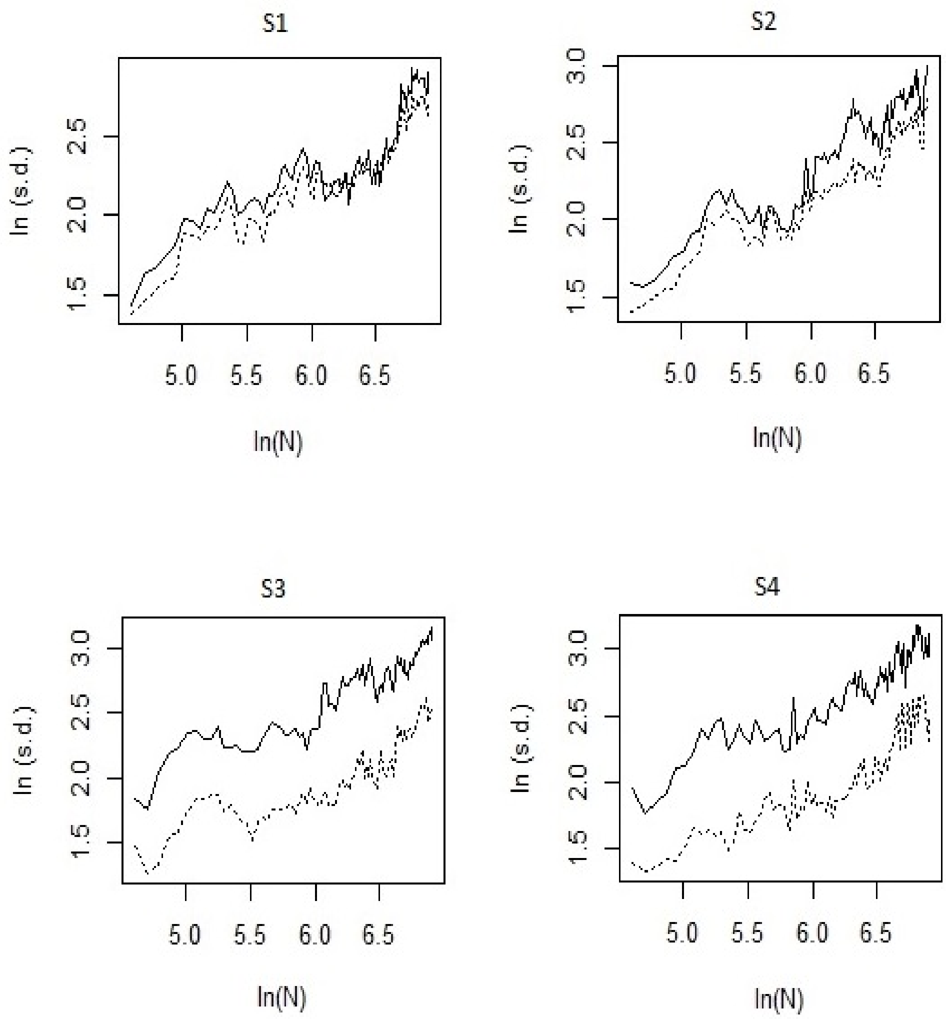


Figure 3.1: Comparative plots of $s.d.(\hat{N})$ for both the estimates $\hat{N}_{tb}^{AP}$ (dotted line) and $\hat{N}_{ind}$ (continuous line) are drawn against true N in a logarithmic scale for the simulated populations S1, S2, S3 and S4.

Robustness Consideration:

Our other interest lies on the robustness of the proposed estimator and $\hat{N}_{ind}$. Actually the model M_{tb} is driven by the unidentifiable behavioral response effect ϕ. An useful estimator for N should be robust to the underlying ϕ value and hence, in Figure 3.3, we present a comparative study on robustness for both the estimates against different ϕ. We fix true N at 500 and same four artificial situations of section 3.3.2 are assumed. ϕ is considered to vary between 0.50 and 3.00 for each of the four populations. Figure 3.3 depicts that $\hat{N}_{tb}^{AP}$ has better robustness w.r.t. ϕ than $\hat{N}_{ind}$ in all situations.

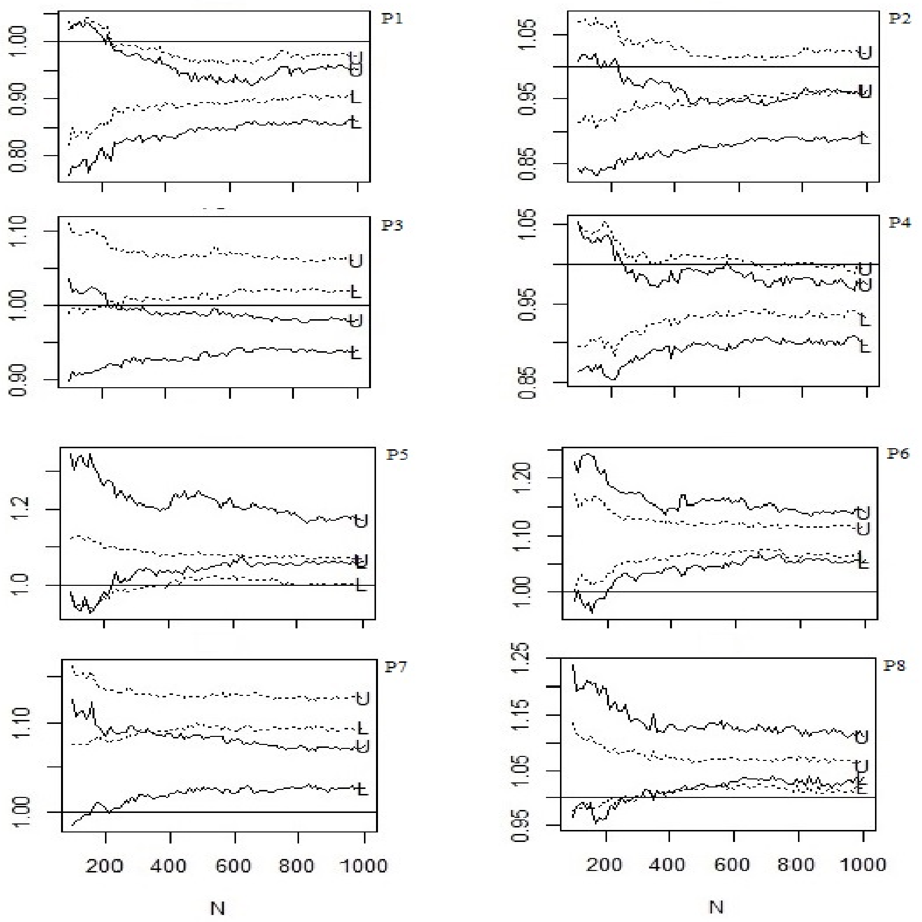


Figure 3.2: Comparative plots of confidence bands of $\hat{N}/N$ corresponding to both the estimates $\hat{N}_{tb}^{AP}$ (dotted line) and $\hat{N}_{ind}$ (continuous line) are drawn against true N for all the populations P1-P4 (recapture-prone cases) and P5-P8 (recapture-aversion cases). The targeted value of $\hat{N}/N$ is indicated at 1.0 (presenting unbiasedness).

3.4 Real Data Applications

3.4.1 Malawi Death Data

A DRS data is considered on death count obtained from a Population Change Survey conducted by the National Statistical Office in Malawi between 1970 and 1972 (for details, *see* Greenfield, 1975 [45]). Details on this data is presented in section 1.8.1 of Chapter 1. Only two strata, called Lilongwe ($\hat{c} = 0.593$, $x_{.1} > x_{1.}$) and Other urban areas ($\hat{c} = 0.839$, $x_{.1} < x_{1.}$), are selected to illustrate our approach observing different $\hat{c}$ values and opposite nature of $x_{.1}$ and $x_{1.}$.

Now, if anyone wishes to use the model M_t assuming list-independence and calculate the

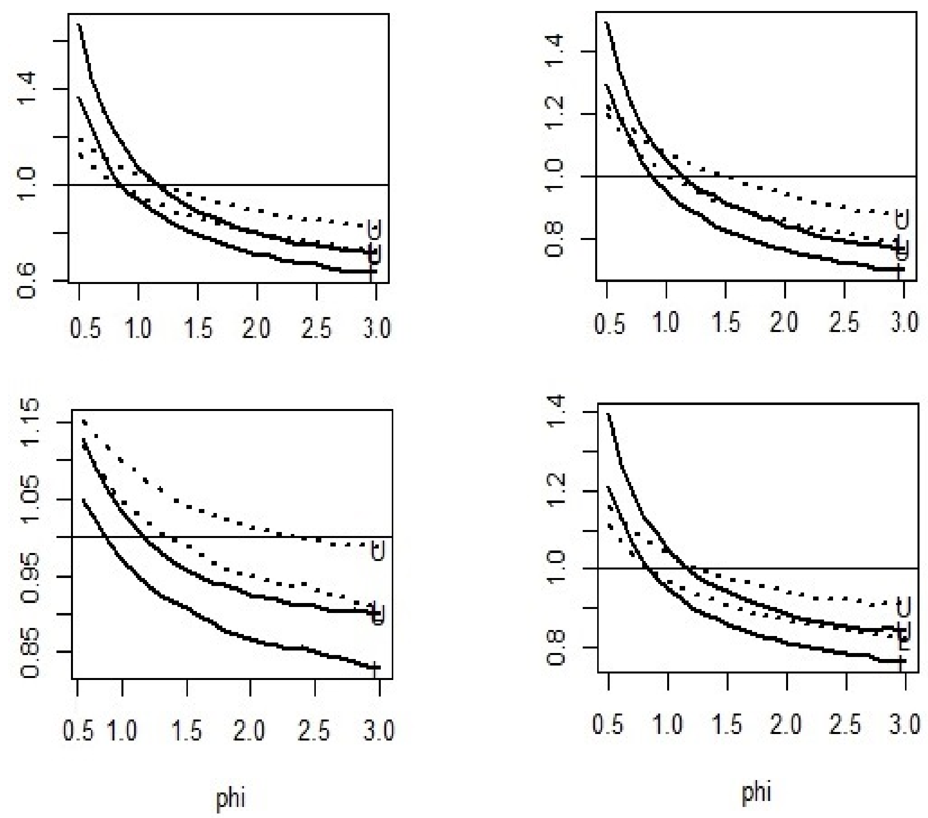


Figure 3.3: Comparative plots of confidence bands of $\hat{N}/N$ corresponding to both the estimates $\hat{N}_{tb}^{AP}$ (dotted line) and $\hat{N}_{ind}$ (continuous line) are drawn against different ϕ for four situations. The targeted value of $\hat{N}/N$ is indicated at 1.0 (presenting unbiasedness).

estimate $\hat{N}_{ind}$, he/she would find that 365 and 2920 deaths occurred in *Lilongwe* and *Other urban areas* respectively. Nour (1982 [68]) argued that the assumption of independent collection procedures is unacceptable in reality. Assuming the fact that two data sources are positively correlated (i.e. $\phi > 1$) in a human demographic study, they estimated death sizes as *378* (i.e. $\hat{\phi} = 1.33$) and *3046* (i.e. $\hat{\phi} = 1.13$) for *Lilongwe* and *Other urban areas* respectively. However, in this article we do not make any such assumptions on the directional nature of ϕ. Then, Lee et al.'s fully Bayes method with uniform prior $\pi(\phi) = U(0.1, 2)$ finds that 372 ($\hat{\phi} = 1.19$) and 3205 ($\hat{\phi} = 1.30$) deaths occured in *Lilongwe* and *Other urban areas* respectively. Our adjusted profile likelihood method estimates the death sizes as 378 ($\hat{\phi} = 1.33$) and 3428 ($\hat{\phi} = 1.53$) respectively, by suitably taking $\delta = 1 - 4(1 - \hat{c})N^{-1}$. Our estimates agree with Nour's for *Lilongwe* but Nour's estimate for *Other urban areas* is significantly smaller than Lee's estimate as well as our estimate.

3.4.2 Injection Drug user Data

Another example of DRS data is considered on injection drug user (IDU) of greater Victoria, British Columbia, Canada (Xu et al., 2014 [104]). To track the changes in the prevalence of HIV and hepatitis C, the Public Health Agency of Canada developed the national, cross-sectional I-Track survey. Details on this data is presented in section 1.8.1. Xu et al. (2014[104]) found that estimate $\hat{N}_{ind}$ for the total number of injection drug users was 3329. They anticipated that $\hat{N}_{ind}$ might not be worthwhile for this situation and used Huggins (1989 [52]) conditional likelihood approach to deal with plausible heterogeneity in the data and the estimate was 3342. Moreover, the time ordering of samples offers an opportunity to use model M_{tb}. Literature on epidemiological studies on such type of *hidden* or *hard to reach* population says that individual, who are listed in first survey, tries to avoid the listing operation in second survey. There is high possibility of *recapture-aversion* (i.e. $\phi < 1$). Low recapture rate $\hat{c} = 0.075$ which strengthens this possibility.

Considering the DRS data originated from model M_{tb} with $\phi > 0$, Lee's (2003[62]) fully Bayes method with prior $\pi(\phi) = U(0.01, 2)$ finds that 596 ($\hat{\phi} = 0.11$) number of drug users are in that population. As $\hat{c}$ is found very low, our adjusted profile likelihood method estimates the size of injection drug users as 584 ($\hat{\phi} = 0.09$) taking $\delta = 1 - 4(1 - \hat{c})N^{-1}$. Hence, Lee's method and our adjusted profile likelihood method says that if you consider the model M_{tb} as appropriate, then total number of injection drug user of greater Victoria is around 580 to 600, a much lower estimate than the estimate of drug users under independence.

3.5 Conclusions

We have considered the most general model M_{tb} for human population that allows the behaviour response effect to play a significant role along with time variation effect in estimating N. The model M_{tb} suffers from identifiability problem where evidence of potentiality to overcome that burden by suitable Bayesian methods is found in literature. However, in this article we have investigated the usefulness of pseudo likelihood approaches based on profiling the interest parameter N. Ordinary profile, modified profile and approximated modified profile likelihoods have been shown to be useless for model M_{tb}. An adjustment of profile likelihood (AdPL) is proposed tuned by an adjustment coefficient so that reasonably better solution can be obtained.

The proposed method depends on the choice of δ (close to $1 - N^{-1}$) using the knowledge of $\hat{c}$ and possible direction of ϕ. In real life situation, if ϕ is unknown, then uniform choice is possible. Full Bayes with uniform prior provides wider coverage than any other method but it also possesses lower efficiency than AdPL. Moreover, Lee's trial-and-error approach to discover a suitable range for uniform prior $\pi(\phi)$ may take a lot of computational time. Some other disadvantages are subjectiveness of the informative prior $\pi(\phi)$, highly dispersed conditional posterior of ϕ, etc. Thus, the adjusted profile method is useful to obtain an efficient estimate of population size (N) very quickly from this complex DRS. In addition to that, AdPL helps to produce more efficient alternatives specially in recapture prone situations.

4 Integrated Likelihood Approach[2]

4.1 Introduction and Motivation

Estimation of the size of a given homogeneous population based on DRS have wide range of applications from both the frequentist and Bayesian practice. Several estimation techniques from model M_t and a few for M_{tb} are available in literature, which are briefly mentioned in sections 1.4 and 1.6 respectively. In Bayesian paradigm, difficulty may arise as the resulting estimator for N, the population size, might be very sensitive to the choice of prior. However, when there is nuisance parameter in the model, inference is often based on a pseudo-likelihood function with properties similar to those of a likelihood function. Since the basic interest in any capture-recapture type experiment is to make inference on N only, all other model parameters are treated as nuisance parameters. Thus, non-Bayesian inferential techniques from such pseudo-likelihood functions are becoming very popular, specially for complex models like M_{tb}, M_{th}, etc. Usefulness of this kind of non-Bayesian techniques such as profile and adjusted profile likelihood (Bolfarine et al., 1992[10]), integrated likelihood with uniform and Jeffrey's prior (Salasar et al., 2014[79]) also have been investigated in the literature for model M_t. Advantage of integrated likelihood is that it always exists, unlike other pseudo-likelihoods - profile, marginal and conditional. Moreover, in small or moderate samples, it is well known that profile likelihood is not always effective. In integrated likelihood method, the main challenge is to choose a suitable weight function on the nuisance parameter. All the above merits of integrated likelihood method motivate us to take up the

[2] Part of this chapter is based on *Chatterjee, K., Mukherjee, D. (2016a[25])*. An improved integrated likelihood population size estimation in Dual-record System. *Statistics and Probability Letters* 110, 146-154.

issue of suggesting a new but comparably efficient integrated likelihood approach. In addition to that, a suitably constructed integrated likelihood could be a potential alternative to Bayesian method in order to overcome the non-identifiability burden that persists in the model M_{tb}. Severini (2007[88]) presents an efficient approach for selecting a weight function and resulting likelihood preserves some important statistical properties.

The present chapter has a two-fold aim. Firstly, explicit investigation of the applicability of our proposed integrated likelihood approach formulated in DRS, with the help of Severini's (2007[88]) technique, for the model M_t. Secondly, we extend our proposed integrated likelihood approach to the relatively complex M_{tb}-DRS context. In this chapter, we will further investigate the applicability of non-informative (such as uniform, Jeffrey's) priors as suitable weigh functions. If this is found to be ineffective, the challenge will be to overcome the non-identifiability problem of model M_{tb} successfully by choosing suitable informative priors.

In the next section, general idea of integrated likelihood method with uniform and Jeffrey's weights are presented under DRS in order to obtain an estimate of the population size N for both the models M_t and M_{tb}. Later we formulate an integrated likelihood approach with the help of Severini (2007[88]) in section 4.3. Implementation of this approach for both the models M_t and M_{tb} are explained in sections 4.4 and 4.5, respectively, along with the simulation studies and real data applications for each of the two said models. We also prove a comparative ordering among all the relevant likelihood and pseudo-likelihood estimates available for M_t in section 4.4.1. Finally we conclude this chapter in section 4.6.

4.2 Integrated Likelihood Method: Preliminaries

We draw special attention to the *integrated likelihood method* so that the resulting likelihood function is useful in this context and satisfies some desirable properties. Integrated likelihood summarizes $\mathscr{L}^*$ (discussed in the first paragraph of Section 3.2) by a weighted averaging of $\mathscr{L}^*$ over Ψ, with respect to a suitable nonnegative weight function $\pi(\psi|\theta)$, as

$$L^I(\theta) \quad = \quad \int_{\Psi} L(\theta, \psi|\underline{\mathbf{x}})\pi(\psi|\theta)\partial\psi. \tag{4.1}$$

One advantage of this pseudo-likelihood approach is that it is always possible to construct, unlike conditional or marginal likelihood. $\pi(\psi|\theta)$ need not be a proper density function. A drawback is the plausible subjectiveness in the choice of $\pi(\psi|\theta)$ and non-orthogonality of ψ to the parameter of interest θ. Uniform and Jeffrey's priors on ψ are two popular non-informative prior densities. Salasar et al. (2014[79]) considered these two priors for multiple capture-recapture experiment with model M_t.

Independence Model: M_t

In the context of model M_t, $\theta = N$ and $\psi = (p_{1\cdot}, p_{\cdot 1})$. Now, if the data is supposed to be generated from model M_t, and uniform prior is considered for both of $p_{1\cdot}$ and $p_{\cdot 1}$ independently, we obtain the integrated likelihood as

$$L_I^U(N) = \frac{\Gamma(N - x_{1\cdot} + 1)\Gamma(N - x_{\cdot 1} + 1)}{\Gamma(N - x_0 + 1)\Gamma(N + 2)(N + 1)}$$

and

$$R_I^U(N) = \frac{L_I^U(N+1)}{L_I^U(N)} = \frac{(N - x_{1\cdot} + 1)(N - x_{\cdot 1} + 1)(N + 1)}{(N - x_0 + 1)(N + 2)^2}.$$

Hence, MLE from L_U^I is $\hat{N}_I^U = [\{-b - \sqrt{(b^2 - 4ac)}\}/2a] + 1$, where $a = -(x_{11} + 2)$, $b = (x_{\cdot 1} x_{\cdot 1} - 2x_{11} + 2x_0 - 5)$ and $c = (x_{1\cdot} x_{\cdot 1} - x_{11} + 3x_0 - 3)$. On the other hand, Jeffrey's prior on ψ is $\pi_J(\psi) \propto \sqrt{|\mathbb{I}(\psi)|} = p_{1\cdot}^{-1/2} p_{\cdot 1}^{-1/2} (1 - p_{1\cdot})^{-1/2} (1 - p_{\cdot 1})^{-1/2}$, where $\mathbb{I}(\psi)$ is the Fisher's information matrix on ψ. The resulting integrated likelihood using $\pi_J(\psi)$ is

$$L_I^J(N) = \frac{\Gamma(N - x_{1\cdot} + 1/2)\Gamma(N - x_{\cdot 1} + 1/2)}{\Gamma(N + 1)\Gamma(N - x_0 + 1)}$$

and

$$R_I^J(N) = \frac{L_I^J(N+1)}{L_I^J(N)} = \frac{(N - x_{1\cdot} + 1/2)(N - x_{\cdot 1} + 1/2)}{(N - x_0 + 1)(N + 1)}.$$

Hence, associated MLE is

$$\hat{N}_I^J = [\{x_{1\cdot} x_{\cdot 1} + (x_{10} x_{01})/2 - 1\}/(x_{11} + 1)] + 1.$$

For any real u, $[u]$ denotes the greatest integer not more than u. Both estimates always exist, even for $x_{11} = 0$. The next theorem gives a comparative relation between these two basic integrated likelihoods.

Theorem 4.2.1 $\hat{N}_I^J \leq \hat{N}_I^U$*, provided* $x_{1\cdot}, x_{\cdot 1} \geq N/2$.

Proof. $R_I^J(N) = R_I^U(N)(N-x_{1.}+1/2)(N-x_{.1}+1/2)(N+2)^2(N-x_{1.}+1)^{-1}(N-x_{.1}+1)^{-1}(N+1)^{-2}$. Now, let $(N-x_{1.}+1/2) = a_1$ and $(N-x_{.1}+1/2) = a_2$. Then

$$\frac{(N-x_{1.}+1)(N-x_{.1}+1)}{(N-x_{1.}+1/2)(N-x_{.1}+1/2)} = \frac{(a_1+1/2)(a_2+1/2)}{a_1 a_2} = 1 + \frac{1}{2a_1} + \frac{1}{2a_2} + \frac{1}{4a_1 a_2}.$$

If we assume $x_{1.}, x_{.1} \geq N/2$, then $(N+1)^{-1} \leq (2a_1)^{-1}$ and $(N+1)^{-1} \leq (2a_2)^{-1}$ which establishes that $R_I^J(N) < R_I^U(N)$. ■

Behavioral Dependence Model: M_{tb}

In capture-recapture context, for fixed θ, Jeffrey's and uniform/constant priors are the two most popular non-informative prior densities on ψ (Salasar, 2014[79]). When $\theta = N$, the uniform prior, $\pi(\psi|N) \propto 1$, where $\psi = (p_{1.}, c, p)$ for likelihood (1.16), implies

$$L_U^I(N) = \int_\Psi L(N,\psi|\underline{\mathbf{x}})\partial\psi \propto (N+1)^{-1}(N-x_{1.}+1)^{-1}.$$

Again if we consider $\pi(\psi|N)$ as Jeffrey's prior, then $\pi(\psi|N)$ is proportional to $\sqrt{|\mathbb{I}_N(\psi)|}$, where $\mathscr{I}_N(\psi)$ is 3×3 Fisher's Information matrix for given N. Therefore,

$$\begin{aligned} \pi(\psi|N) &= \left| Diagonal\left(\frac{N}{p_{1.}(1-p_{1.})}, \frac{Np_{01}}{c(1-c)}, \frac{N(1-p_{1.})}{p(1-p)}\right)\right|^{1/2} \\ &= \{c(1-c)p(1-p)\}^{-1}. \end{aligned} \tag{4.2}$$

Hence,

$$L_J^I(N) = \int_\Psi L(\theta,\psi|\underline{\mathbf{x}})\pi(\psi|\theta)\partial\psi = \frac{(N-x_{1.})}{(N+1)(N-x_0)}.$$

Thus, both of the above pseudo likelihoods show failure of this integrated likelihood method when non-informative prior is used, because this prior could not include any extra information to the likelihood and that is why the resultant likelihood could not overcome the non-identifiability problem.

4.3 Proposed Integrated Likelihood Approach

In order to construct an integrated likelihood function to be useful for non-Bayesian inference, the resultant integrated likelihood should possess frequentist properties relevant for likelihood function. These properties should have implications for the selection of the conditional prior density $\pi(\psi|\theta)$. Severini (2007[88]) analysed some of these properties which suggest that $\pi(\psi|\theta)$ is to be chosen so that, for nuisance parameter γ, with the same dimension as ψ and *unrelated* to θ, γ and θ should be independent under $\pi(\psi|\theta)$ (at least approximately). Therefore the task is to find such parameter γ and then choose a prior density $\pi(\gamma)$ for γ that does not depend on θ. Here we take *unrelated* to mean that $\hat{\gamma}_\theta$, the maximum likelihood estimator of γ for fixed θ, is approximately constant as a function of θ (*see* Cox and Reid, 1987[30]). Hence, the integrated likelihood function for θ with respect to $\pi(\gamma)$ reduces to

$$\overline{L}_S(\theta) = \int_\Gamma L(\theta,\gamma|\underline{\mathbf{x}})\pi(\gamma)\partial\gamma. \tag{4.3}$$

Construction of such *unrelated* nuisance parameter γ is as follows:

If the model is re-parameterized by a nuisance parameter γ that is *strongly unrelated* to θ i.e. $\hat{\gamma}_\theta = \hat{\gamma} + O(n^{-1/2})O(|\theta-\hat{\theta}|)$ and if $\pi(\gamma|\theta)$ does not depend on θ, then the first two Bartlett identities hold for integrated likelihood approximately to $O(N^{-1})$. From the implicit equation we have

$$E\{\ell_\psi(\theta,\psi);\hat{\theta},\gamma\} \equiv E\{\ell_\psi(\theta,\psi);\theta_0,\gamma_0\}\,|_{(\theta_0=\hat{\theta},\gamma_0=\gamma)} = 0, \tag{4.4}$$

from which one can solve γ as $\gamma(\theta,\psi;\hat{\theta})$. Severini (2007 [88]) proved that $\hat{\gamma} = \hat{\psi}$ and γ is *strongly unrelated* to θ. Then solution $\gamma(\theta,\psi;\hat{\theta})$ from (4.4) is called zero-score-expectation parameter. Now, one can choose any suitable prior $\pi(\gamma)$ for γ as $\overline{L}^S$ does not heavily depend on the chosen prior whereas for orthogonal parameters, the proposed integrated likelihood may depend on the choice of prior.

One can find γ in a different way. The aim is to find a function $\gamma(\theta,\psi)$ such that $\hat{\gamma}_\theta = \hat{\gamma} + O(n^{-1/2})O(|\theta-\hat{\theta}|)$. Hence we find such a parameter $\gamma = g(\theta,\psi)$ which implies $\psi = h(\theta,\gamma)$ for some h, if such a function h exists, so that $\hat{\psi}_\theta = h(\theta,\hat{\gamma}_\theta) = h(\theta,\hat{\gamma}) + O(n^{-1/2})$. So, for any value of θ, $\hat{\psi}_\theta$ depends on the data only through $\hat{\gamma}$. In many situations, $\hat{\gamma}$ does not exist. So, we consider γ as a function of $\hat{\theta}$ in addition to (θ,ψ). It can be written as $\gamma = g(\theta,\psi;\hat{\theta})$, which implies $\psi = h(\theta, \gamma; \hat{\theta})$. As $\hat{\psi}_\theta = h(\theta, \hat{\gamma}_\theta; \hat{\theta})$, we must have

$$\hat{\psi}_\theta \quad = \quad h(\theta, \hat{\gamma}; \hat{\theta}) + O(n^{-1/2})O(|\theta - \hat{\theta}|), \tag{4.5}$$

where $\hat{\gamma}_\theta = \hat{\gamma} + O(n^{-1/2})O(|\theta - \hat{\theta}|)$. Hence, one may attempt to find such a function h so that (4.5) holds.

Now we discuss several desirable properties of $\overline{L}^S(\theta)$. γ is less related to θ than any parameter orthogonal to θ. Thus, $\pi(\gamma)$ is less sensitive to θ than $\pi(\psi|\theta)$. Moreover, the resulting integrated likelihood approximately satisfies the score unbiasedness and information unbiasedness properties, very less prior sensitivity and invariance with re-parametrization.

4.4 Analysis of Model M_t

4.4.1 Implementation & Associated Results

In order to apply the proposed approach to model M_t, at first we find the strongly unrelated nuisance parameter corresponding to $\psi = (\psi_1, \psi_2) = (p_{1\cdot}, p_{\cdot 1})$ and the resulting likelihood function is stated in the following theorem.

Theorem 4.4.1 *For model M_t, the strongly unrelated nuisance parameter is γ=(γ_1, γ_2), where $\gamma_1 = (N/\hat{N})p_{1\cdot}$ and $\gamma_2 = (N/\hat{N})p_{\cdot 1}$. Therefore, the likelihood given by (1.8) becomes*

$$L(N, \gamma_1, \gamma_2) \propto \frac{N!}{(N - x_0)!} N^{-(x_{1\cdot} + x_{\cdot 1})} \gamma_1^{x_{1\cdot}} \gamma_2^{x_{\cdot 1}} (1 - \gamma_1 \frac{\hat{N}}{N})^{N - x_{1\cdot}} (1 - \gamma_2 \frac{\hat{N}}{N})^{N - x_{\cdot 1}},$$

$N \geq x_0$ and $\gamma_1, \gamma_2 > 0$.

Proof. The log-likelihood of model M_t is $\ell(N, \psi_1, \psi_2) = \Sigma_{i=0}^{x_0 - 1} ln(N - i) + x_{1\cdot} ln(\psi_1) + (N - x_{1\cdot}) ln(1 - \psi_1) + x_{\cdot 1} ln(\psi_2) + (N - x_{\cdot 1}) ln(1 - \psi_2)$ and $\ell_{\psi_1}(N, \psi_1, \psi_2) = (x_{1\cdot}/\psi_1) - (N - x_{1\cdot})/(1 - \psi_1)$, $\ell_{\psi_2}(N, \psi_1, \psi_2) = (x_{\cdot 1}/\psi_2) - (N - x_{\cdot 1})/(1 - \psi_2)$. Therefore,

$$E(\ell_{\psi_1}(N, \psi_1, \psi_2) : N_0, \psi_{1,0}, \psi_{2,0}) = (N_0 \psi_{1,0}/\psi_1) - (N - N_0 \psi_{1,0})/(1 - \psi_1),$$

$$E(\ell_{\psi_2}(N, \psi_1, \psi_2) : N_0, \psi_{1,0}, \psi_{2,0}) = (N_0 \psi_{2,0}/\psi_2) - (N - N_0 \psi_{2,0})/(1 - \psi_2).$$

Using (4.4) we obtain $\frac{\hat{N}\gamma_1}{\psi_1(1-\psi_1)} = \frac{N}{1-\psi_1}$ and that implies

$$\gamma_1 = (N/\hat{N})\psi_1 = (N/\hat{N})p_{1\cdot}, \gamma_2 = (N/\hat{N})\psi_2 = (N/\hat{N})p_{\cdot 1}.$$

Thus, one may easily obtain the final likelihood after replacing $p_{1\cdot}$ and $p_{\cdot 1}$ with the corresponding unrelated nuisance parameters γ_1 and γ_2 respectively. ■

In order to choose a suitable prior $\pi(\gamma)$, for $\gamma_1, \gamma_2 > 0$, here we consider independent prior specification $\pi(\gamma) = \pi(\gamma_1)\pi(\gamma_2)$ and use only non-informative priors, such as uniform and Jeffrey's priors, as discussed in Section 4.2. Hence, we can easily find the likelihood function $\overline{L}_S(N)$ of N only, by integrating over the domain $\Gamma(= \mathbb{R}^+ \times \mathbb{R}^+)$ of γ, following (4.3).

Proposition 4.4.1 *Under the setup of Theorem 4.4.1, if improper uniform prior is used i.e.* $\pi(\gamma_1) \propto 1$ *and* $\pi(\gamma_2) \propto 1$, *then integrated likelihood of* $N(\geq x_0)$ *is given by*

$$\overline{L}_S^U(N) = \frac{\Gamma(N - x_{1\cdot} + 1)\Gamma(N - x_{\cdot 1} + 1)N^2}{\Gamma(N - x_0 + 1)\Gamma(N + 2)(N + 1)}. \tag{4.6}$$

It is clear that $\overline{L}_S^U(N)$ converges to 0 as $N \to \infty$. From uniform likelihood $\overline{L}_S^U(N)$, we obtain the ratio

$$R_S^U(N) = \frac{\overline{L}_S^U(N+1)}{\overline{L}_S^U(N)} = \frac{(N - x_{1\cdot} + 1)(N - x_{\cdot 1} + 1)(N + 1)^3}{(N - x_0 + 1)(N + 2)^2 N^2}$$

Hence, $\overline{L}_S^U(N)$ is increasing in N for $N < N_0$, where N_0 satisfies $N_0^4 x_{11} - N_0^3(x_{1\cdot}x_{\cdot 1} - 4x_{11} + 2) - N_0^2(3x_{1\cdot}x_{\cdot 1} - 2x_0 - 6x_{11} + 6) - N_0(3x_{1\cdot}x_{\cdot 1} - 4x_{1\cdot} - 4x_{\cdot 1} + 5) - (x_{1\cdot}x_{\cdot 1} - x_{1\cdot} - x_{\cdot 1}) = 0$. Hence one may have the following proposition.

Proposition 4.4.2 *The integrated likelihood estimate corresponding to (4.6) is* $\hat{N}_S^U = [N_0] + 1$, *if* N_0 *is not an integer and* $\hat{N}_S^U = N_0$, *if* N_0 *is an integer.*

Proposition 4.4.3 *Under the setup of Theorem 4.4.1, if the prior* $\pi(\gamma)$ *is Jeffrey's prior i.e.* $\pi(\gamma_1, \gamma_2) \propto \gamma_1^{-1/2}\gamma_2^{-1/2}(1 - \gamma_1\frac{\hat{N}}{N})^{-1/2}(1 - \gamma_2\frac{\hat{N}}{N})^{-1/2}$, *then integrated likelihood of* $N(\geq x_0)$ *is*

$$\overline{L}_S^J(N) = \frac{\Gamma(N - x_{1\cdot} + 1/2)\Gamma(N - x_{\cdot 1} + 1/2)}{\Gamma(N - x_0 + 1)\Gamma(N)}. \tag{4.7}$$

Again, $\overline{L}_S^J(N)$ converges to 0 as $N \to \infty$. From (4.7), we obtain the ratio

$$R_S^J(N) = \frac{\overline{L}_S^J(N+1)}{\overline{L}_S^J(N)} = \frac{(N - x_{1.} + 1/2)(N - x_{.1} + 1/2)}{N(N - x_0 + 1)}.$$

Thus, $\frac{\overline{L}_S^J(N+1)}{\overline{L}_S^J(N)} > 1$ implies $N < \frac{x_{1.}x_{.1}}{x_{11}} - \frac{x_{1.}+x_{.1}}{2x_{11}} + \frac{1}{4x_{11}} = N_1$ (say). Hence the following proposition.

Proposition 4.4.4 *The maximum integrated likelihood estimate obtained from (4.7) is $\hat{N}_S^J = [N_1]+1$, if N_1 is not an integer and $\hat{N}_S^J = N_1$, if N_1 is integer.*

Now we further establish some results on $\hat{N}_S^J$ and $\hat{N}_S^U$ along with some other pseudo-likelihood based estimates, such as, profile likelihood, conditional likelihood and basic integrated likelihood with non-informative priors. As it has been shown in section 1.4.1 of chapter 1 that conditional MLE $\hat{N}_C$ from the Lincoln-Petersen model is identical to the popular dual system estimate $\hat{N}_{ind}$.

Theorem 4.4.2 $\hat{N}_I^U \le \hat{N}_S^J$, *if and only if* $max(x_{1.}, x_{.1}) < 2N/3$.

Proof. $R_I^U(N) = (N - x_{1.} + 1)(N - x_{.1} + 1)(N+1)(N - x_0 + 1)^{-1}(N+2)^{-2}$. Let $a_1 = (N - x_{1.} + 1/2)$, $a_2 = (N - x_{.1} + 1/2)$ and $y = (N+1)$. If we assume $R_S^J > R_I^U$, then $\frac{a_1 a_2 (y+1)^2}{(a_1+1/2)(a_2+1/2)y(y-1)} > 1$

$$\begin{aligned}
\Leftrightarrow a_1 a_2 (3y+1) &> y(y-1)(a_1 + a_2 + 1/2)/2 \\
&= N(N+1)(N - (x_{1.} + x_{.1})/2 + 3/4) \\
&> N(N+1).min(a_1, a_2) \\
\Leftrightarrow a_2(3N+4) &> N(N+1), \text{(WLG, } a_1 < a_2\text{, i.e. } x_{.1} < x_{1.})
\end{aligned}$$

Therefore, neglecting the terms of $O(N^{-1})$, we find $2N + 3 - 3x_{.1} + 3/2 > 4x_{.1}/N$. Since $0 \le p_{.1} \le 1$, it is sufficient to check whether $2N + 3 - 3x_{.1} + 3/2 > 4$ or, $N > 3x_{.1}/2 - 1/4$. Thus, if $N > 3x_{.1}/2$, $R_S^J > R_I^U$. Similarly, when $a_1 > a_2$, $R_S^J > R_I^U$ only if $N > 3x_{1.}/2$. Hence, $R_S^J > R_I^U$ if and only if $max(x_{1.}, x_{.1}) \le 2N/3$. Otherwise, $R_S^J < R_I^U$. ■

If $(x_{1.} + x_{.1})/2x_{11}$ is close to 0, $\hat{N}_S^J$ is approximately equal to $\hat{N}_{ind}$, since $(1/4x_{11} = O_p(N^{-1}))$ is negligible for moderate or high N. It is easy to see that $\hat{N}_I^U \le \hat{N}_S^U$ and $\hat{N}_I^J \le \hat{N}_S^J$ from the expressions of $R_S^U(N)$, $R_I^U(N)$, $R_S^J(N)$ and $R_I^J(N)$. If we consider the case of *conditional MLE* $\hat{N}_C$, we have extensions of these results which are given as Theorems 4.4.3 and 4.4.5.

Theorem 4.4.3 *When $x_{11} > 0$,*
(a) $\hat{N}_I^U \le \hat{N}_C \le \hat{N}_S^U$. As all relevant estimates belong to subset of $\mathbb{Z}_+$, and relative difference between $R_S^U(N)$ and $R_C(N)$ is of $O(N^{-2})$, then $\hat{N}_S^U$ and $\hat{N}_C$ commonly coincide.
(b) $\hat{N}_I^J \le \hat{N}_S^J \le \hat{N}_C$.

Proof of Theorem 4.4.3(a). In association with likelihood (1.8) for model M_t, the conditional likelihood $L_C(N)$ is shown in Severini (2000, pp. 281). Then,

$$R_C(N) = \frac{L_C(N+1)}{L_C(N)} = \frac{(N - x_{1.} + 1)(N - x_{.1} + 1)}{(N - x_0 + 1)(N + 1)} \quad = \quad R_I^U(N)\frac{(N+2)^2}{(N+1)^2}.$$

Hence, $R_C(N) = R_I^U(N)(1 + A_N)$, where A_N is a positive quantity of order $O(N^{-1})$. Thus, $R_C(N) > R_I^U(N)$, which implies $\hat{N}_I^U \le \hat{N}_C$. Again, $R_C(N) = R_S^U(N)(N+2)^2N^2/(N+1)^4$. Using the result of A.M. > G.M, it can be easily shown that $R_C(N) < R_S^U(N)$, which implies $\hat{N}_S^U \ge \hat{N}_C$. Moreover, $R_C(N) = R_S^U(N)(1 - B_N)$, where B_N is a positive quantity of order $O(N^{-2})$. Hence, equality holds between $\hat{N}_S^U$ and $\hat{N}_C$ even for moderate N, as both of them are integer estimates.

In order to proof the second part of the above theorem, at first we consider $R_S^J(N)$ where $R_S^J(N) = (N - x_{1.} + 1/2)(N - x_{.1} + 1/2)(N - x_0 + 1)^{-1}N^{-1}$. Therefore,

$$R_C(N) = R_S^J(N) \times \frac{(N - x_{1.} + 1)(N - x_{.1} + 1)N}{(N - x_{1.} + 1/2)(N - x_{.1} + 1/2)(N + 1)}.$$

Now consider a_1 and a_2 as in the proof of Theorem 4.2.1. Then, $(a_1 + 1/2)(a_2 + 1/2)/a_1a_2 > 1 + min(a_1^{-1}, a_2^{-1}) + (a_1a_2)^{-1}/4 > 1 + N^{-1} = (N+1)/N$. So, $R_C(N) > R_S^J(N)$. Hence the proof using the result $\hat{N}_I^J \le \hat{N}_S^J$. ∎

$\hat{N}_C$ is almost identical to $\hat{N}_S^U$ which is due to the close relation between the integrated and conditional likelihood functions. The following theorem states yet another inequality relation between previously discussed estimates and the profile likelihood estimate $\hat{N}_P$, where $\hat{N}_P$ is just a DRS version of profile likelihood estimate discussed in Bolfarine et al. (1992, [10]). Let us consider the following Lemma in order to establish the next theorem.

Lemma 4.4.4 $\left(\frac{N}{N+1}\right)^{2N}\left(1 + \frac{1}{N - x_{1.}}\right)^{N - x_{1.}}\left(1 + \frac{1}{N - x_{.1}}\right)^{N - x_{.1}} < 1.$

Proof of Lemma. Since $N - x_{1.} < N$, $N - x_{.1} < N$ and $\ln(1+y)/y \downarrow y$, $\left(1 + \frac{1}{N - x_{1.}}\right)^{N - x_{1.}} < \left(1 + \frac{1}{N}\right)^N$ and $\left(1 + \frac{1}{N - x_{.1}}\right)^{N - x_{.1}} < \left(1 + \frac{1}{N}\right)^N$. ∎

Theorem 4.4.5 *(a)* $\hat{N}_P \le \hat{N}_C$,
(b) $\hat{N}_S^J \le \hat{N}_P$ *when* $x_{1.} + x_{.1} \ge N/2$,
(c) $\hat{N}_P \le \hat{N}_I^U$ *when* $x_{1.}, x_{.1} \ge 2N/3$ *and* $\hat{N}_P \ge \hat{N}_I^U$ *when* $x_{1.}, x_{.1} \le N/3$.

Proof. *Part (a).* We have $R_P(N) = \frac{L_P(N+1)}{L_P(N)}$

$$\begin{aligned} &= \frac{N^{2N}(N-x_{1.}+1)(N-x_{.1}+1)}{(N+1)^{2N+1}(N-x_0+1)}\left(1+\frac{1}{N-x_{1.}}\right)^{N-x_{1.}}\left(1+\frac{1}{N-x_{.1}}\right)^{N-x_{.1}} \\ &= R_C(N)\frac{N^{2N}}{(N+1)^{2N}}\left(1+\frac{1}{N-x_{1.}}\right)^{N-x_{1.}}\left(1+\frac{1}{N-x_{.1}}\right)^{N-x_{.1}}. \end{aligned}$$

Hence, using Lemma 4.4.4, the proof is complete from $R_P(N) < R_C(N)$.

Part (b). At first we find $R_P(N) = R_S^J(N)\frac{(N-x_{1.}+1)(N-x_{.1}+1)}{(N-x_{1.}+1/2)(N-x_{.1}+1/2)}$ $\times \frac{N}{N+1}\left(1+\frac{1}{N}\right)^{-2N}\left(1+\frac{1}{N-x_{1.}}\right)^{N-x_{1.}}\left(1+\frac{1}{N-x_{.1}}\right)^{N-x_{.1}}$. We assume $R_P(N) < R_S^J(N)$. Hence, using Lemma 4.4.4 we must have

$$\frac{(N-x_{1.}+1)(N-x_{.1}+1)}{(N-x_{1.}+1/2)(N-x_{.1}+1/2)}\frac{N}{N+1} < 1.$$

That implies numerator of the l.h.s. of the above expression, $Num = N(N-x_{1.})(N-x_{.1}) + N(N-x_{1.}) + N(N-x_{.1}) + N$ is less than the denominator of the l.h.s., $Denom = (N-x_{1.} + 1/2)(N-x_{.1}+1/2)(N+1)$. After some simple algebraic manipulations, one can write

$$Denom = Num + x_{1.}x_{.1} - (N+1)(x_{1.}+x_{.1}-1/2)/2$$

and thus we need $x_{1.}x_{.1} - (N+1)(x_{1.}+x_{.1}-1/2)/2 > 1$

$$\Leftrightarrow 2 \; > \; \left(\frac{1}{p_{1.}}+\frac{1}{p_{.1}}\right)+\frac{1}{N}\left(\frac{1}{p_{1.}}+\frac{1}{p_{.1}}-\frac{1}{2p_{1.}p_{.1}}\right)-\frac{1}{2N^2p_{1.}p_{.1}}, \tag{4.8}$$

since $x_{1.} = Np_{1.}$ and $x_{.1} = Np_{.1}$. If we ignore the terms of $O(N^{-2})$ for moderate or large N and assume $p_{1.} + p_{.1} > \frac{1}{2}$ or equivalently $x_{1.} + x_{.1} > \frac{N}{2}$, then inequality (4.8) does not hold. So, $R_P(N) > R_S^J(N)$ by contradiction and hence the proof of Theorem 4.4.5(b).

Part (c). We have
$R_P(N) = R_I^U(N)\left(1+\frac{1}{N}\right)^{-2N}\left(1+\frac{1}{N+1}\right)^2\left(1+\frac{1}{N-x_{1.}}\right)^{N-x_{1.}}\left(1+\frac{1}{N-x_{.1}}\right)^{N-x_{.1}}$.
Then, $\ln R_P(N) = \ln R_I^U(N) + M$, where M is the logarithm of the remaining terms of the

above expression. We consider second order approximation of M and after some routine simplification, we have

$$M \simeq \frac{1}{N} + \frac{2}{N+1} - \frac{1}{(N+1)^2} - \frac{1}{2(N-x_{1.})} - \frac{1}{2(N-x_{.1})}.$$

Now if $x_{1.}, x_{.1} \geq 2N/3$, then $\frac{1}{2(N-x_{1.})}, \frac{1}{2(N-x_{.1})} \geq 3/2N$. Hence, $M \leq \frac{1}{N} + \frac{2}{N+1} - \frac{1}{(N+1)^2} - \frac{3}{N} < 0$. Thus, $R_P(N) < R_I^U(N)$ under $x_{1.}, x_{.1} \geq 2N/3$. Again, if $x_{1.}, x_{.1} \leq N/3$, then $M \geq \frac{1}{N} + \frac{2}{N+1} - \frac{1}{(N+1)^2} - \frac{3}{2N} > 0$. Thus, $R_P(N) > R_I^U(N)$ under $x_{1.}, x_{.1} \leq N/3$. This completes the proof of Theorem 4.4.5(c). ■

Thus, summarizing all the above results, we have the theorem below.

Theorem 4.4.6 *For alternative capture-recapture propensity, we have the following ordering among relevant estimates:*
(a) $\hat{N}_I^J \leq \hat{N}_I^U \leq \hat{N}_S^J \leq \hat{N}_P \leq \hat{N}_C \leq \hat{N}_S^U$ *if* $x_{1.}, x_{.1} \in [\frac{N}{2}, \frac{2N}{3})$,
(b) $\hat{N}_I^J \leq \hat{N}_S^J \leq \hat{N}_P \leq \hat{N}_I^U \leq \hat{N}_C \leq \hat{N}_S^U$ *if* $x_{1.}, x_{.1} \geq \frac{2N}{3}$,
(c) If $min(x_{1.}, x_{.1}) \in [\frac{N}{2}, \frac{2N}{3})$ *and* $max(x_{1.}, x_{.1}) \geq \frac{2N}{3}$, *then either (b) holds or* $\hat{N}_I^J \leq \hat{N}_S^J \leq \hat{N}_I^U \leq \hat{N}_P \leq \hat{N}_C \leq \hat{N}_S^U$.

4.4.2 Simulation Study

In this section, performance of both the newly developed integrated likelihood estimates $\hat{N}_S^U$ and $\hat{N}_S^J$ are investigated along with *mle*, $\hat{N}_{UMLE}$, some other pseudo-likelihood based estimates - $\hat{N}_C$, profile likelihood $\hat{N}_P$ and other two general integrated likelihood estimates $\hat{N}_I^U$ and $\hat{N}_I^J$. First we simulate six hypothetical populations following model M_t, corresponding to six pairs of capture probabilities $(p_{1.}, p_{.1})$={(0.50, 0.65), (0.60, 0.70), (0.60, 0.80), (0.80, 0.70), (0.80, 0.60), (0.70, 0.55)} for each population of size 500 and 5000 (to investigate the size effect on estimator, if any). We denote these six populations as P1, P2, ..., P6 respectively and the associated results are presented in Table 4.1. 200 data sets on $(x_{1.}, x_{.1}, x_{11})$ are generated from each population and final estimate of N is obtained by averaging over 200 estimates. Now, based on those 200 estimates, sample s.e., sample RMSE (Root Mean Square Error) and 95% confidence interval (C.I.) are computed.

Empirical study confirms the findings of Theorem 4.4.3(a) that the proposed estimate $\hat{N}_S^U$ is exactly identical to the conditional *mle* $\hat{N}_C$. Also, $\hat{N}_C$ is identical to $\hat{N}_{ind}$. Thus only $\hat{N}_S^U$ is presented in Table 4.1. This table shows that the general integrated likelihood estimates $\hat{N}_I^U$ and $\hat{N}_I^J$ are almost same in DRS, while the performances of profile likelihood estimate $\hat{N}_P$

Table 4.1: Summary results of the proposed estimates $\hat{N}_S^U$ and $\hat{N}_S^J$ along with other existing integrated likelihoods, profile likelihood and unconditional *mle* for the populations P1-P6.

Population		$\hat{N}_{UMLE}$	$\hat{N}_P$	$\hat{N}_I^U$	$\hat{N}_I^J$	$\hat{N}_S^U$	$\hat{N}_S^J$
				$N = 500$			
P1	$\hat{N}$(s.e.)	499(15.78)	498(15.77)	499(15.67)	499(15.71)	500(15.87)	499(15.80)
	RMSE	15.82	16.74	15.84	15.92	15.87	16.85
	C.I.	(474,533)	(473,532)	(474,533)	(473,534)	(475,534)	(474,533)
P2	$\hat{N}$(s.e.)	499(11.79)	499(11.75)	499(11.70)	499(11.75)	500(11.80)	500(11.75)
	RMSE	11.84	11.80	11.74	11.78	11.80	11.75
	C.I.	(478,518)	(479,519)	(479,519)	(479,519)	(479,519)	(479,519)
P3	$\hat{N}$(s.e.)	499(9.23)	500(9.25)	499(8.95)	499(8.92)	501(9.23)	500(9.25)
	RMSE	9.29	9.25	9.00	8.98	10.21	9.25
	C.I.	(482,517)	(482,518)	(482,516)	(482,516)	(483,518)	(482,518)
P4	$\hat{N}$(s.e.)	499(8.05)	500(8.03)	500(7.95)	499(7.93)	500(8.10)	500(8.02)
	RMSE	8.12	8.03	7.95	7.95	8.10	8.02
	C.I.	(486,515)	(486,515)	(486,515)	(486,516)	(486,516)	(486,515)
P5	$\hat{N}$(s.e.)	499(8.56)	499(8.64)	500(8.61)	499(8.68)	500(8.65)	499(8.63)
	RMSE	8.63	8.70	8.61	8.73	8.65	8.68
	C.I.	(482,513)	(483,514)	(483,514)	(482,514)	(483,514)	(483,514)
P6	$\hat{N}$(s.e.)	498(12.98)	497(12.99)	498(12.93)	498(12.95)	499(13.03)	499(12.95)
	RMSE	13.13	14.24	14.35	14.36	13.07	12.05
	C.I.	(473,519)	(472,519)	(473,519)	(473,521)	(474,521)	(474,520)
				$N = 5000$			
P1	$\hat{N}$(s.e.)	5002(56.32)	4994(49.28)	5002(56.32)	5002(56.33)	5003(56.36)	5002(56.32)
	RMSE	56.35	49.65	56.35	56.37	56.44	56.35
	C.I.	(4900,5111)	(4906,5087)	(4902,5111)	(4901,5112)	(4903,5112)	(4902,5112)
P2	$\hat{N}$(s.e.)	4996(39.80)	4998(41.89)	4996(39.80)	4996(39.81)	4999(39.82)	4996(39.80)
	RMSE	40.00	41.95	40.00	40.01	39.85	40.00
	C.I.	(4905,5072)	(4917,5078)	(4906,5072)	(4906,5073)	(4908,5075)	(4906,5072)
P3	$\hat{N}$(s.e.)	4996(30.52)	4996(28.88)	4096(30.55)	4096(30.54)	4997(30.51)	4996(30.52)
	RMSE	30.78	28.99	30.83	30.81	30.73	30.78
	C.I.	(4940,5057)	(4948,5059)	(4941,5057)	(4941,5058)	(4941,5058)	(4940,5057)
P4	$\hat{N}$(s.e.)	5000(25.02)	4999(23.72)	5000(24.92)	4999(24.97)	5000(25.04)	4999(25.04)
	RMSE	25.02	23.74	24.92	25.01	25.04	25.08
	C.I.	(4957,5055)	(4951,5048)	(4957,5055)	(4957,5056)	(4957,5056)	(4957,5055)
P5	$\hat{N}$(s.e.)	4997(29.62)	4998(30.94)	4998(29.59)	4998(29.64)	4999(29.64)	4998(29.61)
	RMSE	29.77	30.98	29.64	29.71	29.68	29.68
	C.I.	(4940,5055)	(4942,5060)	(4941,5056)	(4941,5056)	(4941,5056)	(4941,5056)
P6	$\hat{N}$(s.e.)	4998(42.76)	5006(45.48)	4999(42.76)	4998(42.77)	4999(42.79)	4999(42.76)
	RMSE	42.81	45.87	42.77	42.82	42.80	42.77
	C.I.	(4908,5080)	(4919,5086)	(4909,5080)	(4909,5081)	(4910,5081)	(4909,5080)

and unconditional *mle* $\hat{N}_{UMLE}$ are quite similar. But $\hat{N}_P$ is less preferable due to its relatively small accuracy for low capture probabilities and significantly greater computation time. The present numerical results show that the proposed integrated likelihood estimate with Jeffrey's prior, $\hat{N}_S^J$, is slightly more efficient than $\hat{N}_S^U$, in most of the situations. Hence, it is found to be the best among all other competitive pseudo-likelihood estimates and slightly better than $\hat{N}_{UMLE}$. It is also found that $\hat{N}_S^J \le \hat{N}_S^U$, which is similar to the result for general integrated likelihood (*see* Theorem 4.2.1). Comparing the results for different population sizes, it is seen that the variation is roughly of order $\sqrt{N}$, which is consistent with the theory. Otherwise, the performance of alternative estimators show quantitatively similar ranking.

4.4.3 Real Data Applications

I. Transmitted Tuberculosis Data. To estimate the number of transmitted Tuberculosis (TB) cases in three urban districts of Madrid during 1997-1999, Iñigo et al. (2003, [53]) used conventional epidemiological data and the information on clustered cases obtained by DNA fingerprinting as independent Dual-record System. Using different covariates, they formed several stratifications in the population for the analysis. For illustration of our proposed methods, here we consider the whole unstratified population and its stratification based on sex and age only (see Table 4.2).

II. Road Traffic Mortality Data. Another data on deaths from road traffic injuries (RTIs), available in Samuel et al. (2012, [80]), is also considered for illustration purpose. RTIs are responsible significantly for the preventable death and disability in developing countries and it is grossly under-reported. For that, police accident reports and a hospital-based trauma registry together build up an incomplete DRS and that is used to estimate the size of the Road Traffic Deaths separately for all inhabitants, men and women in the Lilongwe district of Malawi (see Table 4.2).

Using the above two data sets we have computed the various estimates as put forward above, along with the performance measures, s.e. and 95% confidence interval (C.I.). Both of these measures are computed by usual parametric bootstrap method over a bootstrap sample of size 500. The ordering of the estimators in our real data illustrations are consistent with the theory and simulation results. The efficiency behaviour of $\hat{N}_S^J$, $\hat{N}_S^U$, $\hat{N}_I^J$ and $\hat{N}_I^U$ all turn out to be as expected. The variability is also well within the $\sqrt{N}$ order bounds as suggested by the numerical results.

4.5 Analysis of Model M_{tb}

4.5.1 Implementation & Associated Results

In this section we extend our investigation on the applicability of newly developed integrated likelihood approach (in section) for the time-behavioral dependence model. We construct the relevant unrelated nuisance parameter γ, then choose $\pi(\gamma)$ satisfying posterior unbi-

Table 4.2: Summary results of the proposed estimates $\hat{N}_S^U$ and $\hat{N}_S^J$ along with other estimates in two real data applications.

Population		$\hat{N}_{UMLE}$	$\hat{N}_P$	$\hat{N}_I^U$	$\hat{N}_I^J$	$\hat{N}_S^U$	$\hat{N}_S^J$
		TRANSMITTED TUBERCULOSIS DATA					
Total	$\hat{N}$(s.e.)	148(27.23)	148(27.15)	143(22.92)	144(24.73)	150(27.83)	148(0.76)
	C.I.	(109, 219)	(113, 235)	(108, 200)	(107, 207)	(112, 223)	(146, 149)
Male	$\hat{N}$(s.e.)	82(19.48)	83(20.93)	80(14.66)	80(16.26)	85(20.40)	83(19.52)
	C.I.	(59, 134)	(63, 147)	(60, 114)	(59, 119)	(61, 138)	(59, 135)
Female	$\hat{N}$(s.e.)	59(27.40)	60(26.95)	54(14.06)	56(17.47)	62(29.05)	60(27.27)
	C.I.	(37, 142)	(41, 148)	(37, 94)	(36, 110)	(38, 150)	(37, 143)
Lower Age (< 35 Years)	$\hat{N}$(s.e.)	106(70.28)	106(47.16)	92(28.02)	96(37.96)	111(74.81)	106(70.32)
	C.I.	(60, 308)	(65, 247)	(59, 171)	(58, 214)	(62, 327)	(61, 308)
Higher Age (≥ 35 Years)	$\hat{N}$(s.e.)	52(10.77)	52(10.45)	51(8.50)	51(9.25)	54(11.55)	52(10.83)
	C.I.	(38, 81)	(42, 85)	(38, 72)	(38, 75)	(39, 85)	(38, 81)
		ROAD TRAFFIC MORTALITY DATA					
Total	$\hat{N}$(s.e.)	382(47.03)	382(44.77)	375(43.76)	378(45.39)	385(46.70)	382(46.02)
	C.I.	(38, 81)	(321, 474)	(308, 477)	(308, 485)	(313, 497)	(309, 494)
Male	$\hat{N}$(s.e.)	289(34.67)	289(32.23)	284(31.92)	285(32.82)	292(35.01)	289(34.67)
	C.I.	(232, 371)	(241, 362)	(230, 357)	(231, 362)	(234, 374)	(233, 371)
Female	$\hat{N}$(s.e.)	68(39.85)	68(33.87)	60(17.49)	62(22.68)	71(42.60)	68(39.87)
	C.I.	(39, 194)	(43, 183)	(38, 106)	(38, 126)	(40, 209)	(39, 194)

asedness condition. We present the consequent theorems, results and properties for the parametrization (1.15) in chapter 1.

At first we consider the nuisance parameter $\psi = (p_{1\cdot}, c, p)$ along with the parameter of interest $\theta = N$. Following theorem finds strongly unrelated parameter corresponding to the nuisance parameter ψ.

Theorem 4.5.1 *Using (4.4), the strongly unrelated parameter for ψ is γ=(γ_1,γ_2,γ_3), where $\gamma_1 = (N/\hat{N}_{ind})p_{1\cdot}$, $\gamma_2 = c$ and $\gamma_3 = \frac{1-p_{1\cdot}}{(\hat{N}_{ind}/N)-p_{1\cdot}} p$.*

Since current model suffers from non-identifiability, so non-informative priors for resultant unrelated parameters (as in Theorem 4.5.1) would not work satisfactorily, as in the case of ordinary integrated likelihoods under uniform and Jeffrey's priors. Thus, we consider some informative prior for γ so that hyper-parameters satisfy some relations that lead to effective frequentist properties.

Theorem 4.5.2 *In connection with Theorem 4.5.1, if the prior $\pi(\gamma)$ is of the form $\pi(\gamma) = \pi(\gamma_1)\pi(\gamma_2)\pi(\gamma_3)$ and $\pi(\gamma_1) = GB1(b_1 = \frac{N}{\hat{N}_{ind}}, r_1, s_1)$, $\pi(\gamma_2) = Unif(0,1)$ and $\pi(\gamma_3) = GB1(b_2 =$*

$\frac{1-p_{1\cdot}}{(\hat{N}_{ind}/N)-p_{1\cdot}}$, r_2, s_2), *for any positive real numbers* r_2, s_2, r_1, s_1 *satisfying* $r_2+s_2=s_1$, *then integrated likelihood of* N *for model* M_{tb} *becomes*

$$\overline{L}^{I}_{tb}(N) \quad = \quad \frac{\Gamma(N-x_0+s_2)\Gamma(N+1)}{\Gamma(N+r_1+s_1)\Gamma(N-x_0+1)}, \tag{4.9}$$

for $N \geq x_0$*. Thus, for the given values of* r_1, s_1 *and* s_2, $\overline{L}^{I}_{tb}(N)$, *in (4.9), is non-decreasing in* N *for*

$$N \leq \frac{x_0(r_1+s_1-1)}{r_2+r_1} - 1 \tag{4.10}$$

and $\overline{L}^{I}_{tb}(N)$ *converges to 0 as* $N \to \infty$.

On the basis of the relationship between γ and ψ, mentioned in Theorem 4.5.1, we suggest the values of hyper-parameters r_2, s_2 and r_1 using posterior unbiasedness of γ (after considering γ as function of N). We suggest $r_2 = b\,x_{01}$ and $s_2 = b(N^* - x_0)$ for some b, where N^* is some other working estimate of N. So, $s_1 = b(N^* - x_{1\cdot})$. Now, for fixed N, $\gamma_1 = (N/\hat{N}_{ind})p_{1\cdot} = (N/\hat{N}_{ind})(x_{1\cdot}/N) = (x_{1\cdot}/\hat{N}_{ind})$. Hence, from the posterior unbiasedness condition regarding unrelated parameter γ_1, we have $(x_{1\cdot}/\hat{N}_{ind}) \simeq E_\pi(\gamma_1) = (N/\hat{N}_{ind})\{(r_1+x_{1\cdot})/(r_1+s_1+N)\}$ and this implies $r_1 = x_{1\cdot}s_1/(N-x_{1\cdot})$ after some algebraic manipulation. Thus r_1 depends on N and therefore, the right hand side of the condition in (4.10) also becomes dependent on N. So, $N \leq \frac{x_0\{r_1(N)+s_1-1\}}{\{r_2+r_1(N)\}} - 1 \Leftrightarrow \frac{x_0\{r_1(N)+s_1-1\}}{\{r_2+r_1(N)\}} - N \geq 1$ and $\frac{x_0\{r_1(N)+s_1-1\}}{\{r_2+r_1(N)\}} - N$ is non-increasing in N. Hence, the following theorem discusses the possibility of existence of the corresponding maximum likelihood estimate.

Theorem 4.5.3 *For* $r_1 = x_{1\cdot\cdot}s_1/(N-x_{1\cdot}) = r_1(N)$, $\exists$ *a real* $N > x_0$, *say* $N_0 = N_0(\underline{\boldsymbol{x}})$, *such that* $N \leq \frac{x_0\{r_1(N)+s_1-1\}}{\{r_2+r_1(N)\}} - 1 \Leftrightarrow \frac{x_0\{r_1(N)+s_1-1\}}{\{r_2+r_1(N)\}} - N \geq 1 \Leftrightarrow N \leq N_0 < \infty$.

Hence from Theorem 4.5.3, it can be said that $\overline{L}^{I}_{tb}(N)$ is increasing in N for $N \leq N_0$ and hence, $\hat{N}^{I}_{tb} = [N_0]+1$, if N_0 is not an integer and $\hat{N}^{I}_{tb} = [N_0]$ and $[N_0]+1$ if N_0 is an integer. The expression for N_0 terms out to be mathematically intractable and thus obtaining an explicit solution is not possible. We explore the frequentist properties of the estimator through computation. To implement the above prior specification, we suggest $N^* = (\hat{N}_{Nour} + \hat{N}_{ind})/2$ and $b = (1+(\hat{N}_{ind} - ((x_0+\hat{N}_{ind})/2) - 1)^{-1})/2$, when we don't know anything about the plausible direction of ϕ. On the other hand, if we know that $\phi > 1$ (recapture proneness), then $N^* =$

$\hat{N}_{Nour}$ and $b = 1$. If $phi < 1$ (recapture aversion), suggested $N^* = \hat{N}_{ind}$ and $b = (\hat{N}_{ind} - ((x_0 + \hat{N}_{ind})/2) - 1)^{-1}$. However, use of directional knowledge on ϕ can certainly improve the inference if the knowledge is correct. A classification strategy of the given population in terms of the behavioral nature is proposed in chapter 7 along with a discussion on how to apply it for real data.

Remark: If anyone is interested to consider $\theta = (N, \phi)$ or $\theta = (N, c, p)$, as $c = \phi p$, relevant for likelihood (1.15) then strongly unrelated parameters can be obtained from the relevant log-likelihood functions as before. Now, for $\psi = p_{1\cdot}$, if the prior is taken on the associated unrelated parameter γ as $\pi(\gamma) = GB1(b = \frac{N}{\hat{N}_{ind}}, r, s)$, for any positive real numbers r and s; then integrated likelihood for $N \geq x_0$ reduces to

$$\overline{L}^I_{M_{tb}}(N, c, p) = \frac{\Gamma(N - x_{1\cdot} + s)\Gamma(N+1)}{\Gamma(N + r + s)\Gamma(N - x_0 + 1)} c^{x_{11}} p^{x_{01}} (1-c)^{x_{10}} (1-p)^{N-x_0},$$

which fails to produce the *mle* of $\theta = (N, c, p)$ from the corresponding estimating equations. Perhaps the failure is due to the lack of enough information to make inference about the two parameters N and ϕ, which are actually orthogonal to each other (for details about parameter orthogonality, *see* Cox and Reid (1987[30]).

Analogous to the analysis of likelihood (1.15), one may like to consider the likelihood (1.16) associated to another parametrization to make inference about ϕ in addition to N. But, in lieu of including ϕ directly into θ, we include p in θ. Ultimately, estimate of ϕ can be obtained directly through the relation $\hat{c} = \phi p$. The theorem below followed by a result showing final integrated likelihood attached to parametrization in (1.16).

Theorem 4.5.4 *When $\theta = (N, p)$ and $\psi = (p_{1\cdot}, \phi)$ for model M_{tb} with parametrization (1.16), then strongly unrelated parameter is given by $\gamma = (\gamma_1, \gamma_2)$, where $\gamma_1 = (N/\hat{N}_{ind})p_{1\cdot}$ and $\gamma_2 = (p/\hat{p})\phi$.*

Theorem 4.5.5 *In connection with Theorem 4.5.4, if the prior $\pi(\gamma)$ is of the form $\pi(\gamma) = \pi(\gamma_1)\pi(\gamma_2)$ and $\pi(\gamma_1) = GB1(b_2 = \frac{N}{\hat{N}_{ind}}, r_2, s_2)$ and $\pi(\gamma_2) = GB1(b_1 = p^{-1}, r_1, s_1)$, for any positive real numbers r_2, s_2, r_1 and s_1, then integrated likelihood becomes*

$$\overline{L}^I_{tb}(N, p) = \frac{\Gamma(N - x_{1\cdot} + s_2)\Gamma(N+1)}{\Gamma(N + r_2 + s_2)\Gamma(N - x_0 + 1)} N^{p_1} p^{x_{01}} (1-p)^{N-x_0}. \tag{4.11}$$

for $N \geq x_0$ and r_1, s_1 are independent of N. However, we skip further development from (4.7), as our present aim is to estimate N only.

4.5.2 Simulation Study

In this section we conduct a simulation study to evaluate the performance of our proposed approach and understand its efficiency in applying the method using the available directional knowledge on ϕ. This study is designed as follows. Let us simulate eight hypothetical populations corresponding to four pairs of capture probabilities $(p_{1\cdot}, p_{\cdot 1})$={(0.50, 0.65), (0.60, 0.70), (0.80, 0.70), (0.70, 0.55)} for each case of *recapture prone* (represented by $\phi = 1.25$) and *recapture averse* (represented by $\phi = 0.80$) situations as like section 3.3.2 for simulation study in Chapter 3 (*see* Table 3.1). Here also, all the populations are truly of size $N = 500$. Results for truly *recapture prone* and *recapture averse* simulated populations are presented respectively in the top and bottom panels of Table 4.3. First two cases with $p_{1\cdot} < p_{\cdot 1}$, represent the usual situation in DRS from Post Enumeration Survey (PES) conducted for estimating census undercount estimation. 200 data sets on $(x_{1\cdot}, x_{\cdot 1}, x_{11})$ are generated from each of the eight populations. Our proposed integrated likelihood estimate have been obtained for each data sets. Finally, estimate of $\hat{N}^S_{tb}$ is obtained by averaging over 200 posterior means. Based on those 200 estimates, the bootstrap sample s.e., sample RMSE (Root Mean Square Error) and 95% confidence interval (C.I.) are computed. In addition to that, to compare the performance of our proposed method with a full Bayesian strategy developed by Lee et al. (2003[62]), we compute similar statistics for Lee's method. Detailed discussions on the computation strategy of Lee's method, particularly for DRS, can be found in Chatterjee and Mukherjee (2016c[27]) or in sections 5.2 and 5.4 of Chapter 5. A second benchmark for comparison, another pseudo-likelihood based method proposed in chapter 3, namely Adjusted Profile Likelihood (AdPl) based estimator, is also used. All the details of computation for AdPL are available in chapter 3. Summary results of our proposed integrated likelihood method as well as Lee's and AdPL (with $\delta = 1 - 1.25N^{-1}$) methods are presented in Table 4.3.

From Table 4.3 it can be clearly noticed that our proposed estimate is more efficient than Lee's and AdPL in most of the situations, in terms of RMSE and closeness to the true value. In all situations except P2, 95% credible intervals for Lee's estimate and AdPL are wider than the confidence intervals of $\hat{N}^S_{tb}$. Precisely, Lee > AdPL > $\hat{N}^S_{tb}$, in terms of length of the associated interval estimates.

4.5.3 Real Data Applications

To illustrate the method proposed in section 4.5.1, here we consider two real DRS datasets, I. Children Injury Data and II. Handloom PES Data. Details on these two datasets are already presented in section 1.8.1. Proposed estimates are obtained separately for known and unknown directional knowledge on underlying parameter ϕ for both of the data sets. $\hat{N}^S_{tb}$ is computed along with other comparative estimators include common DSE estimator $\hat{N}_{ind}$, Lee's(2003[62]) estimator and AdPL (*see* section 3.3). Besides estimate, s.e. and 95% confidence interval for N is also computed based on bootstrap technique for each estimator.

I. Children Injury Data. Let us consider the DRS data from Jarvis et al. (2000[56]), in which authors illustrate the serious drawbacks in the use of commonly DSE estimator $\hat{N}_{ind}$, specifically for injury related data. The problem was to enumerate those children under 15 years of age from addresses in Northumbria who were seriously injured in local Motor Vehicles Accidents (MVA) between 1 April, 1990 and 31 March, 1995. The common estimates under independence ($\hat{N}_{ind}$) for these three classes are shown in third column of Table 4.4. It is noted that $\hat{N}_{ind}$'s are more than twice the total number of cases actually observed (x_0). Also, value of the estimate $\hat{c}$ for these three classes are 0.25, 0.40 and 0.59 respectively, which are substantially small. All these direct to the possibility of list dependency indicating $\phi < 1$, due to very small amount of recapture and this motivate us to include this data in our illustration. These three classes have a common feature that $x_{1.} < x_{.1}$. We present summary of results in Table 4.4 for our proposed integrated likelihood method along with Lee et al's Bayes and AdPL estimate for comparison.

For the data of *Cyclists* and *Passengers* (in Table 4.4), all the estimators agree with the negative departure from independence. For *Pedestrians*, all estimators excluding AdPL indicate that this class has positive dependence between two available lists. Like previous example, here also Lee's estimates posses larger variation than all other estimates and hence its credible intervals are too wide. Estimate $\hat{N}^S_{tb}(WOA)$ has better efficiency than AdPL except for the *Pedestrians*. The proposed $\hat{N}^S_{tb}(WA)$ produces estimate closely to AdPL. Further, in most of the all cases, it has smaller variance and tighter confidence bounds, as expected.

II. Handloom PES Data. Let us consider a new data from a survey aimed to estimate the undercount in the census of Handloom workers residing at Gangarampur in South Dinajpur district of state West Bengal in India. Sampled Ward no. 2 correctly counts 126 persons in main census operation, while PES counts 107 persons and 85 persons are matched correctly. Hence, total number of distinct captured individuals (x_0) is 148 which is very closer to the

$\hat{N}_{ind} = 159$. Data related to another sampled Ward, no. 16, is as follows: correct census count 131, correct PES count 103 and matched 50 persons. Therefore, corresponding x_0 is 184 but $\hat{N}_{ind}$ is 270. The nature of the data on two wards are surprisingly different except the similarity that both posses $x_{1.} > x_{.1}$, which is opposite to the previous example of Children Injury Data. Reason behind $x_{1.} > x_{.1}$ is the temporary seasonal migration of local handloom worker to other districts in search of handloom work. Surveyors reported that workers in Ward 16, which is very close to town head-quarter, might be somewhat reluctant to enlist themselves in second time (i.e. at the time of PES). Moreover, most of them are working outside (other districts) and usually come home in particular seasons. That is why, Ward no. 16 results very low matches than Ward no. 2. Another reason may be that some people think that one-time enrollment at the time of census is enough. So, underlying ϕ may be less than 1. These possibilities as well as the beliefs of the experts of Textile Directorate of Govt. drive the idea that the estimator $\hat{N}_{ind}$ is not suitable here as independence fails. Being quite certain about the homogeneity within wards from the experts of Textile Directorate, we apply the model M_{tb} for these data and compute the estimates following our proposed integrated likelihood method . We also compute Bayes estimate proposed by Lee et al. (2003[62]) and AdPL, as earlier, for comparison in dependence situation. We also execute the summary results if list-independence is assumed in order to measure the extent of deviation of other dependent estimates from independence.

In Table 4.5, our proposed estimate for Ward no. 2, without assuming any directional nature, agrees with the estimate from independence assumption. Other the Bayesian estimates developed by Lee et al. (2003,[62]) and AdPL indicate small negative departure from independence. $\hat{N}^S_{tb}(WA)$ is around 164 assuming $\phi > 1$. For the other sampled Ward, if we incorporate the recapture aversion assumption in our proposed method, it implies that approximately 210 workers are residing in Ward no. 16, which is same as AdPL estimate.

4.6 Conclusion

The integrated likelihood method proposed here is demonstrated to be a potentially efficient alternative to the various maximum likelihood and pseudo-likelihood estimates available in the literature for the popular M_t model in the context of dual-record system (DRS). It is shown that the present integrated likelihood developed with the help of the idea introduced by Severini (2007[88]) is quite useful and yields efficient estimates for DRS or two-sample capture-recapture model. We prove a clear ordering among the relevant alternative estimates. Closeness between $\hat{N}_{ind}$ and N^U_S is also established. Extensive numerical illustration shows consistency with the theoretical findings on the efficiency of proposed estimates and ordering.

List-independence assumption does not hold satisfactorily in many instances. Various data from epidemiological studies and coverage error estimation motivate us to use a suitable model by relaxing the assumption of list-independence and build up efficient inferential strategy. As far as homogeneous human population size estimation is presumed throughout the project, two-sample capture-recapture experiment is appropriate along with M_{tb} modelling. Here we consider integrated likelihood method as a non-Bayesian strategy which has a potential to overcome the non-identifiability in $M_{tb} - DRS$ model. We have shown that general integrated likelihoods using common non-informative priors fail to produce estimates. To overcome this shortcoming, here we proposed an integrated likelihood approach with suitable prior on *unrelated* nuisance parameter. This pseudo-likelihood mechanism produces efficient estimates satisfying several frequentist properties, prior insensitiveness, invariance, etc. To deal with non-identifiability, suitable informative prior is used and the choice of some hyperparameters are suggested depending upon the availability of directional knowledge on ϕ, if any. Indeed, this project presents an efficient non-Bayesian strategy for the complex M_{tb}-DRS model.

Table 4.3: Summary results of the proposed integrated likelihood method along with full Bayes (Lee et al. (2003[62])) and non-Bayesian AdPL estimates for all the populations. Here, Populations P1-P8 correspond to model M_{tb}.

Population		Lee[a]	AdPL[b]	$\hat{N}^S_{tb}$
P1	$\hat{N}$(s.e.)	472(18.94)	461(11.47)	478(14.82)
	RMSE	34.11	40.32	26.75
	C.I.	(438,518)	(439,480)	(450,508)
P2	$\hat{N}$(s.e.)	478(11.04)	488(10.01)	486(13.15)
	RMSE	24.33	15.54	19.60
	C.I.	(451,519)	(467,506)	(460,511)
P3	$\hat{N}$(s.e.)	490(7.55)	515(6.78)	498(8.69)
	RMSE	12.02	16.32	9.05
	C.I.	(473,515)	(501,527)	(480,514)
P4	$\hat{N}$(s.e.)	488(12.38)	476(9.27)	497(13.51)
	RMSE	17.03	25.68	13.85
	C.I.	(457,531)	(456,493)	(471,521)
P5	$\hat{N}$(s.e.)	472(18.77)	505(9.40)	487(12.12)
	RMSE	35.42	10.72	17.81
	C.I.	(431,565)	(487,524)	(465,512)
P6	$\hat{N}$(s.e.)	490(10.91)	534(6.98)	499(8.28)
	RMSE	13.65	35.23	8.30
	C.I.	(460,560)	(519,547)	(484,515)
P7	$\hat{N}$(s.e.)	509(6.39)	548(5.21)	502(5.07)
	RMSE	10.34	48.40	5.40
	C.I.	(485,545)	(537,557)	(492,511)
P8	$\hat{N}$(s.e.)	481(9.39)	510(7.75)	481(8.17)
	RMSE	20.08	13.01	20.55
	C.I.	(448,545)	(497,525)	(467,495)

[a] Prior U(0.2, 1.5) is used for bottom panel populations as Lee et al.'s strategy fails to generate samples from U(0.2, 1)

[b] with $\delta = 1 - 1.25N^{-1}$

Table 4.4: Summary results of the proposed integrated likelihood estimates for **Children Injury Data**, with (*WA*) and without (*WOA*) the assumption on the directional knowledge on ϕ, are presented. Lee's full Bayes estimate and adjusted profile likelihood estimates (AdPL) with $\delta = 1-4(1-\hat{c})N^{-1}$ are also presented for comparison.

Class & (Estimate $\hat{c}$)		$\hat{N}_{ind}$	Lee	AdPL	$\hat{N}^S_{tb}$ (WOA)	$\hat{N}^S_{tb}$ (WA)
Cyclists	$\hat{N}$ (s.e.)	495 (69.68)	427 (72.72)	245 (6.81)	376 (34.81)	294 (18.52)[a]
(0.254)	C.I.	(359, 632)	(293, 522)	(231, 258)	(327, 463)	(259, 330)
Passengers	$\hat{N}$ (s.e.)	249 (24.05)	195 (21.40)	180 (5.41)	227 (13.89)	191 (8.93)[a]
(0.40)	C.I.	(202, 296)	(158, 225)	(170, 191)	(203, 255)	(175, 210)
Pedestrians	$\hat{N}$ (s.e.)	1323 (31.90)	1415 (284.5)	1198 (14.54)	1338 (28.18)	1159 (13.53)[b]
(0.592)	C.I.	(1260, 1385)	(1064, 1964)	(1166, 1223)	(1285, 1393)	(1135, 1186)

[a]Small $\hat{c}$ suggests *recapture aversion*
[b]*Recapture aversion* is assumed

Table 4.5: Summary results of the proposed integrated likelihood estimates for **Handloom PES Data**, with (*WA*) and without (*WOA*) the assumption on the directional knowledge on ϕ, are presented. Lee's full Bayes estimate and adjusted profile likelihood estimates (AdPL) with $\delta = 1-4(1-\hat{c})N^{-1}$ are also presented for comparison.

Population & (Estimate $\hat{c}$)		$\hat{N}_{ind}$	Lee	AdPL	$\hat{N}^S_{tb}$ (WOA)	$\hat{N}^S_{tb}$ (WA)
Ward 2	$\hat{N}$ (s.e.)	159 (4.50)	151 (4.02)	166 (4.66)	160 (5.30)	164 (6.46)[a]
(0.675)	C.I.	(150, 167)	(148, 162)	(158, 175)	(152, 172)	(154, 178)
Ward 16	$\hat{N}$ (s.e.)	270 (21.53)	245 (30.99)	206 (5.13)	250 (13.38)	210 (7.63)[b]
(0.382)	C.I.	(228, 312)	(203, 320))	(197, 216)	(223, 279)	(197, 226)

[a]moderately high $\hat{c}$ suggests *recapture proneness*
[b]small $\hat{c}$ suggests *recapture aversion*

5 Empirical Bayes Estimation: Missing Data Approaches[3]

5.1 Introduction and Motivation

Estimation of human population size or number of vital events occurred during a given time span is a very relevant statistical concern which includes a vast area of application in the fields of census coverage error estimation, population studies and epidemiology. In chapter 1 it is shown that capture-recapture type data structure, which is known as Multiple-record system, is formed in order to estimate the size of a given human population.

As stated in chapter 1, various frequentist and likelihood approaches under suitable assumptions are discussed in the literature (*see* Otis et al. (1978[71]), Seber (1986[83]), Chao et al. (2001a[22])) covering most of the basic and complex models developed for Capture-recapture system or equivalently, for Multiple-record system. In this context, a very simple and widely used capture-recapture model is M_t (also known as Lincoln-Petersen Model) which accounts for the time (t) variation effect (*see* section 1.4.1 for details). But this simple model often fails miserably when the underlying independence assumption between capture probabilities is violated. Many methodologists and practitioners (*see* El-Khorazaty, 2000[36]; Jarvis et al., 2000[56]) argued that the independence assumption may not be justified. A brief review is done by Brittain and Böhning (2009[13]) of the various methods available by relaxing the independence assumption in DRS context. Usually, independence is violated either due to dependent behavioral response at the time of second survey or due to heterogeneity among

[3] This chapter is based on *Chatterjee, K., Mukherjee, D. (2016c[27]).* On the estimation of homogeneous population size from a complex dual-record system. *Journal of Statistical Computation and Simulation* **86**, 3562-3581.

individuals. ChandraSekar and Deming (1949[17]) suggests to apply the model for homogeneous sub-groups constructed by stratifying the whole population and their recommendation is usually followed in practice for census undercount, epidemiological studies etc. Hence, the only factor responsible for violation of the independence assumption is the behavioral response variation dependent on the previous capturing. Thus, the model M_{tb} becomes most relevant which successfully characterizes the possible list dependency through a parameter ϕ ($\in R^{+}$) and which denotes *Behavioral Response Effect*. For detail discussion on model M_{tb}, we refer to section 1.6 of chapter 1.

Appropriateness of M_{tb} is discussed in section 1.6 but unfortunately this model suffers from identifiability problem (*see* Methodological issue no. *M5* in section 1.8.2) in DRS due to the unidentifiable ϕ. Different approaches using martingle theory (Lloyd, 1994), quasi-likelihood inference (Chao et al., 2000) are proposed to solve the problem for number of counting efforts, $T \geq 3$ with the help of an assumption, but that assumption can not solve the problem in DRS (i.e. when $T = 2$). Bayesian paradigm is found to be helpful to reasonably overcome this non-identifiability burden with a minimum subjective choice. Lee and Chen (1998) applied the Gibbs sampling idea to overcome the identifiability problem but they did not use recapture data and estimates became unstable and prior sensitive (*see* Lee et al. (2003)). Later, Lee et al. (2003) applied noninformative priors to all model parameters except ϕ, for which prior was chosen by a trial-and-error method. Finally, they came up with a fully Bayesian solution using MCMC, but, their empirical study as well as real data application were exercised in the spirit of multiple lists ($T > 3$) problem which is common in animal abundance, not for human population. Wang et al. (2015[100]) also proposed a hierarchical Bayesian M_{tb} model for multiple lists with the assumption that the odds of recapture bears a constant relationship to the odds of initial capture. However, we think that the potential of the fully Bayesian method proposed by Lee et al. (2003) should be investigated in this present complex DRS situation, which is not attempted earlier.

Lists of individuals available from different sources on the same population are framed in a contingency table where one cell, referring to absence in all lists, is always missing. Thus, the population size estimation from DRS can be viewed as missing data estimation (B$\ddot{o}$hning and Heijden, 2009). In section 1.8.2, Methodological issue no. *M4* sketches the present problem of N estimation in DRS as a missing data analysis. Hence, the present condition and scope motivate us to develop competing efficient Bayesian inferential strategies for estimating N in DRS framework under model M_{tb}. Two empirical Bayes approaches – EM-within-Gibbs and stochastic EM-within Gibbs are proposed in section 5.3 in addition to a discussion of existing fully Bayesian method due to Lee et al.(2003) (in 5.2.1). Advantages and disadvantages of the said approaches are discussed under the common roof of data augmentation strategies for missing data analysis. In demographic context, usually $\phi > 1$ occurs which implies population is *recapture prone*. But for a population with sensitive characteristics, such as drug users, population with Common Congenital Anomaly disease etc., $\phi < 1$ and then one may call that population as *recapture averse*. When such information on ϕ is available, one can hope that performance of any suitable method should improve. Advantages of the availability of *directional knowledge* is mentioned in methodological issue *M6* in section 1.8.2. Thus, we propose different sets of priors in our Bayesian methods to deal with the availability or non-availability of the prior *directional knowledge* on ϕ.

Extensive simulation study for different population sizes, capture probabilities and a real data application on death size estimation are presented in sections 5.4 and 5.5 respectively. Finally in section 5.6, we summarize our findings and enumerate the best possible estimation rule depending upon the availability of directional knowledge on ϕ.

5.2 Methods in Bayesian Framework: Preliminaries

Let us consider a typical missing data situation where the complete data can be classified into observed (U^{obs}) and missing (U^{mis}) components. We may parameterize a relevant model with interest parameter $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ and nuisance parameter ψ (may also be a vector). A subset θ^S of θ denotes missing quantity U^{mis} and the rest part $\theta_{-S} = \theta \setminus \theta^S$ is model parameter of interest. Here we reformulate the present N estimation problem as estimation of missing data. In this section, we discuss different versions of Gibbs sampler as a key algorithm for fitting hierarchical models with missing data. At first we discuss an existing Gibbs sampler strategy (*i*) Data Augmentation [DA] (Tanner and Wong, 1987[93]), which is a standard Bayesian approach in missing data context allowing priors to all unobserved quantity (i.e. nuisance parameter ψ is empty). Then, we propose ourselves an empirical Bayes approach, namely (*ii*) EM-within-Gibbs [EWiG], allowing priors to all unobserved quantity in the model except the nuisance parameter ψ. This method can be treated an expanded version of Monte Carlo Expectation-Maximization algorithm [MCEM] (Wei and Tanner, 1990[101]). We also sketch another alternative empirical Bayes approach, which we call (*iii*) *Stochastic EM-within-Gibbs* [SEMWiG], that possess certain advantages than EWiG. Each of the two proposed approaches (*ii*) and (*iii*) has its own posterior of interest depending upon whether non-empty ψ is considered to be random or not. We employ the Gibbs sampler strategy proposed by Lee et al. (2003[62]) exactly to workout this DA strategy and consider this method as a competing approach to our proposed methods. In the current article, our intension is not to account any external subjectiveness in prior selection so that posterior would not be sensitive to the choice of priors. Since, data generation in DRS falls under the finite population sampling, convergence of posterior density can be studied as the population size, $N \to \infty$ (refer methodological issue *M7* in section 1.8.2).

5.2.1 Data Augmentation [DA-Lee]

Data augmentation refers to strategies for constructing iterative optimization or sampling algorithms for all the unknown quantities in model. An existing Bayesian treatment in literature will be discussed here as a data augmentation method in a platform of missing data analysis, considering all the unknown quantities (θ, ψ) to be random. Posterior samples can be drawn iteratively through a suitable Gibbs sampling strategy. Tanner and Wong (1987[93]) developed this stochastic version of data augmentation (DA) to make simulation simple and straightforward. If $\pi(\psi|\theta, U^{obs})$ and $\pi(\theta_i|\theta_{-i}, \psi, U^{obs})$, for $i = 1, 2, \ldots, k$, are the resultant conditional posterior distributions, then Gibbs sampler is obtained from the following DA strategy:

Step 1	:	Set $t = 0$ and initialize $\theta^{(0)}$.
Step 2	:	Generate ψ and θ_i from $\pi(\psi|\theta^{(t)}, U^{obs})$ and $\pi(\theta_i|\hat{\theta}_{-i}^{(t)}, \psi, U^{obs})$.
Step 3	:	Update $\theta^{(t)}$ with $\theta^{(t+1)} = \{\theta_i; i = 1, 2, \ldots, k\}$.
Step 4	:	Repeat the last two steps until convergence of $\{\theta^{(t)}\}_{t\geq 0}$.

Finally, samples after burn-in period are considered to be generated from the targeted posterior $\pi(\theta|U^{obs})$. To implement this DA strategy, prior selection for all the unknown quantities (θ, ψ) and burn-in choice are made following Lee et al. (2003[62]) thoroughly. Henceforth, the above strategy is named as DA-Lee. This method has two-fold aim to be considered in this article. Firstly, we want to investigate the performance of this existing approach developed by Lee et al. (2003[62]) in particular to the present complex DRS model. Secondly, it will be used as a competing method in order to evaluate the two proposed methods, discussed in the next subsections.

Implementation. Following Lee et al. (2003[62]), we take $\theta^S = N$, $\theta_{-S} = (p_{1\cdot}, \phi, p)$ as ψ remains empty. In full Bayesian analysis, if priors are assigned independently i.e. $\pi(\theta) = \pi(\theta^S, \theta_{-S}) = \pi(N)\pi(p_{1\cdot})\pi(\phi)\pi(p)$, hence conditional posterior densities for given data $\underline{\mathbf{x}} = (x_{1\cdot}, x_{\cdot 1}, x_{11})$ are

$$\begin{aligned}
\pi(N - x_0|, p_{1\cdot}, p, \underline{\mathbf{x}}) &\propto \frac{N!}{(N-x_0)!}((1-p)(1-p_{1\cdot}))^N \pi(N), && (5.1)\\
\pi(p_{1\cdot}|N, \underline{\mathbf{x}}) &\propto p_{1\cdot}^{x_{1\cdot}}(1-p_{1\cdot})^{N-x_{1\cdot}}\pi(p_{1\cdot}), && (5.2)\\
\pi(\phi|N, p, \underline{\mathbf{x}}) &\propto \phi^{x_{11}}(1-\phi p)^{x_{10}}\pi(\phi), && (5.3)\\
\pi(p|N, \phi, \underline{\mathbf{x}}) &\propto p^{x_{\cdot 1}}(1-p)^{N-x_0}(1-\phi p)^{x_{10}}\pi(p). && (5.4)
\end{aligned}$$

All the prior specifications and required tools to construct posterior distribution through DA-Lee strategy is governed by Lee et al. (2003 [62]). It suggests flat non-informative priors for $p_{1\cdot}$ and p as $\pi(p_{1\cdot}) = \pi(p) \sim Unif(0,1)$ and Jeffrey's prior, $\pi(N) \propto N^{-1}$, which is improper. Again a flat prior $\pi(\phi) = Unif(\alpha, \beta)$ is chosen though this is no longer non-informative as specification of its range $[\alpha, \beta]$ is needed. Hence, the resulting joint posterior for $\theta = (N, p_{1\cdot}, \phi, p)$ is

$$\pi_{tb}(\theta|\underline{\mathbf{x}}) = \frac{(N-1)!}{(N-x_0)!(\beta-\alpha)}\phi^{x_{11}} p_{1\cdot}^{x_{1\cdot}} p^{x_{\cdot 1}}(1-p_{1\cdot})^{N-x_{1\cdot}}(1-p)^{N-x_0}(1-\phi p)^{x_{10}}. \tag{5.5}$$

In the following theorem, we explicitly establish the posterior properness of (5.5), which is required in order to use the resulting posterior for further Bayesian analysis.

Theorem 5.2.1 *Joint posterior in (5.5) is proper for any finite quantities α and β with $\alpha < \beta$.*

Proof. Joint posterior $\pi_{tb}(\theta|\underline{\mathbf{x}})$, in (5.5), can be rewritten as

$$\begin{aligned}
\pi_{tb}(\theta|\underline{\mathbf{x}}) &= \frac{(N-1)!}{(N-x_0)!(\beta-\alpha)}[(1-p_{1\cdot})(1-p)]^{N-x_0}\phi^{x_{11}} p_{1\cdot}^{x_{1\cdot}} p^{x_{\cdot 1}}(1-p_{1\cdot})^{x_{01}}(1-\phi p)^{x_{10}}\\
&\propto f(N-x_0|x_0,\mu)\frac{1}{(\beta-\alpha)}\phi^{x_{11}} p_{1\cdot}^{x_{1\cdot}} p^{x_{\cdot 1}}(1-p_{1\cdot})^{x_{01}}(1-\phi p)^{x_{10}}\mu^{-x_0}, && (5.6)
\end{aligned}$$

with $N \geq x_0$, $\alpha < \phi < \beta$, $0 < p_{1\cdot}, p < 1$ and $f(\cdot|\cdot)$ is a Negative Binomial pmf of $(N - x_0)$ with known parameter x_0 and unknown probability $\mu = 1-(1-p_{1\cdot})(1-p)$, for given the data $\underline{\mathbf{x}}$. Now, if we take the sum over $(N - x_0)$ on its domain $0, 1, 2, \ldots \infty$, then $\sum_{N \geq x_0} f(N - x_0|x_0, \mu) = 1$.

Therefore, in next step, integrating $\pi_{tb}(\phi, p_{1\cdot}, p|\underline{\mathbf{x}})$ w.r.t. ϕ we have

$$\begin{aligned}
\pi_{tb}(p_{1\cdot}, p|\underline{\mathbf{x}}) &= \int_{\alpha}^{\beta} \pi_{tb}(\phi, p_{1\cdot}, p|\underline{\mathbf{x}}) \partial\phi \\
&= \frac{p_{1\cdot}^{x_{1\cdot}} p^{x_{\cdot 1}} (1-p_{1\cdot})^{x_{01}}}{[1-(1-p_{1\cdot})(1-p)]^{x_0}(\beta-\alpha)} \int_{\alpha}^{\beta} \phi^{x_{11}}(1-\phi p)^{x_{10}} \partial\phi. \qquad (5.7)
\end{aligned}$$

It is clear that integrand in (5.7) is bounded above by $(\beta-\alpha)$. Thus, in order to prove that the joint posterior is proper for any pair of α and β, it is sufficient to prove that

$$\frac{p_{1\cdot}^{x_{1\cdot}} p^{x_{\cdot 1}} (1-p_{1\cdot})^{x_{01}}}{[1-(1-p_{1\cdot})(1-p)]^{x_0}}$$

is bounded above by some finite quantity.

Let us consider $\widetilde{p} = min(p_{1\cdot}, p) \in (0,1)$ and therefore, take M such that

$M > \widetilde{p}^{-x_0} \in (1, \infty)$, which implies

$\widetilde{p} > M^{-1/x_0} = \epsilon > 0.$

Given x_0, ϵ can be made small by taking M large enough. Therefore,

$$\begin{aligned}
(1-\widetilde{p}) &< 1-M^{-1/x_0} \\
\Rightarrow (1-p)(1-p_{1\cdot}) &< 1-M^{-1/x_0} \\
\Leftrightarrow [1-(1-p)(1-p_{1\cdot})]^{x_0} &> M^{-1} \\
\Rightarrow \frac{p_{1\cdot}^{x_{1\cdot}} p^{x_{\cdot 1}} (1-p_{1\cdot})^{x_{01}}}{[1-(1-p_{1\cdot})(1-p)]^{x_0}} &< M.
\end{aligned}$$

This completes the proof. ■

Henceforth, conditional posterior densities from (5.1), (5.2), (5.3) and (5.4) are as follows:

$$\begin{aligned}
\pi(N-x_0|, p_{1\cdot}, p, \underline{\mathbf{x}}) &\propto \text{Neg. Binomial}(x_0, \mu), \qquad (5.8) \\
\pi(p_{1\cdot}|N, \underline{\mathbf{x}}) &\propto Beta(x_{1\cdot}+1, N-x_{1\cdot}+1), \qquad (5.9) \\
\pi(p|N, \phi, \underline{\mathbf{x}}) &\propto p^{x_{\cdot 1}}(1-p)^{N-x_0}(1-\phi p)^{x_{10}}, \qquad (5.10) \\
\pi(\phi|N, p, \underline{\mathbf{x}}) &\propto \text{Gen.Beta-I}(x_{11}+1, x_{10}+1, 1, p) \times \mathscr{I}_{[\alpha,\beta]}(\phi), \qquad (5.11)
\end{aligned}$$

where $\mu = 1-((1-p)(1-p_{1\cdot}))$ and $\mathscr{I}_{[\alpha,\beta]}(\phi)$ is an indicator function for $\phi \in \Phi$, the domain of ϕ. Gibbs sampler provides samples on N and θ_{-S} using (5.8), (5.9), (5.11) and (5.10) at convergent posterior. But samples cannot be drawn directly from (5.10). As per Lee et al. (2003[62]) suggestion, adaptive rejection sampling technique is used to generate p from (5.10) and $\hat{R}^{1/2}$ technique, suggested by Gelman (1996[41]), for fixing the burn-in period. If no other information on ϕ is available, then choosing suitable (α, β) is not at all easy as the result of DA-Lee would be dependent on the choice of the prior distributions. We also adopt a *trial-and-error* procedure following Lee at al. (2003[62]) who opted for such α and β values for which the range of the posterior credible interval for ϕ is not too close to either side of the prior limits. Finally, iterations will ultimately yield draws from the true marginal posterior distributions after certain burn-in period. They admitted that this *trial-and-error* procedure works well when there is a large amount of recapture information and (α, β) can be chosen from the experts' experience. Usually for human population, moderate or high capture probabilities may be attained but number of samples is usually very small (not more than three). No investigation on the performance of Lee et al. (2003[62]) strategy has been carried out separately for given *recapture-averse* (i.e. $\phi < 1$) and *recapture-prone* (i.e. $\phi > 1$) populations. Here, we are also interested in the situation when directional knowledge on ϕ is available, so we use ($\alpha = 0$, $\beta = 1$) for recapture-averse and ($\alpha = 1$, $\beta = 2$ or 3) for recapture-prone population and compare the performance of DA-Lee as a competing method with the next two proposed approaches.

5.3 Proposed Methodologies

5.3.1 EM-within-Gibbs [EMWiG]

Instead of using full Bayesian method, if some model parameters are restricted to be updated by any consistent likelihood or frequentist method, then it may reduce the variance of the generated sequence. Boonstra et al. (2013[11]) presents a generic approach by expanding the MCEM (developed by Wei and Tanner (1990[101])) in the context of a prediction problem of fitting linear regression in the presence of an unknown nuisance hyperparameter. Considering the same idea, a novel empirical Bayes version is sketched out first-time in capture-recapture context where prior densities are assigned on the interest parameter θ only and nuisance parameter ψ (may also be a vector) remains to be estimated by non-Bayesian approach, such as *mle*. We also call this strategy as EM-within-Gibbs [EWiG]. Let us denote the likelihood and its logarithmic version by $[U^{obs}|.]$ and $\ell[U^{obs}|.]$ respectively, for given data U^{obs}. For a given point estimate $\hat{\psi}$, say MLE, Gibbs sampler generates B

draws from $\pi(\theta|\hat{\psi}, U^{obs})$. To obtain the MLE, computation of marginal $[U^{obs}|\psi]$ is required, which may include high-dimensional integration. However, one can write the marginal likelihood for ψ as $[U^{obs}|\psi] = [\theta, U^{obs}|\psi]/[\theta|\psi, U^{obs}]$. EM algorithm can automatically produce the marginal MLE of ψ after taking expectation on $\ell[U^{obs}|\psi]$ or equivalently, on $\ell[\theta, U^{obs}|\psi]$ with respect to $[\theta|\psi, U^{obs}]$. E-step nicely engaged the conditional posterior of θ, $\pi(\theta|\psi, U^{obs})$ and then at M-step, $E(\ell[\theta, U^{obs}|\psi])$ or its empirical version is maximized with respect to ψ. A monte carlo treatment for calculating the empirical EM is

$$\hat{\psi} = \underset{\psi\in\Psi}{\operatorname{argmax}} \left\{ B^{-1} \sum_{j=1}^{B} \ell(\theta^{(j)}, U^{obs}|\psi) \right\}. \tag{5.12}$$

Now we sketch the algorithm as follows to produce Gibbs sampler in order to obtain the posterior densities.

Step 1	:	Set $t = 0$ and initialize $\psi^{(0)}$.
Step 2	:	Generate $\{\theta^{(j)} = \left(\theta^{S(j)}, \theta_{-S}^{(j)}\right); j = 1(1)B\}$ by iteratively simulating from $\pi(\theta^S|\theta_{-S}, \psi^{(t)}, U^{obs})$ and $\pi(\theta_{-S}|\theta^S, \psi^{(t)}, U^{obs})$.
Step 3	:	*Expectation and Maximization.* Obtain $\psi^{(t+1)}$ by updating $\psi^{(t)}$ using (5.12).
Step 4	:	Repeat above two steps until the convergence of $\{\psi^{(t)}\}_{t\geq 0}$.

Since, the nuisance parameter ψ is estimated by EM, the sequence $\{\psi^{(t)}\}$ converges to the *mle* of ψ. At the convergent $\{\psi^{(t)}\}$ at $\hat{\psi}$, the final sample $\{\theta^{(j)} = (\theta^{S(j)}, \theta_{-S}^{(j)}); j = 1(1)B\}$ is drawn following Step 2. From this final sample we would thus obtain the estimate of posterior density $\pi(\theta_{-S}|\psi, U^{obs})$ and $\pi(\theta^S|\psi, U^{obs})$ respectively as

$$\hat{\pi}(\theta_{-S}|\hat{\psi}, U^{obs}) = B^{-1} \sum_{j=1}^{B} \pi(\theta_{S-}|\theta^{S(j)}, \hat{\psi}, U^{obs}) \tag{5.13}$$

$$\text{and } \hat{\pi}(\theta^S|\hat{\psi}, U^{obs}) = B^{-1} \textstyle\sum_{j=1}^{B} \pi(\theta^S|\theta_{-S}^{(j)}, \hat{\psi}, U^{obs}). \tag{5.14}$$

In any such empirical Bayes procedure, a fundamental concern is that how one can consider $\hat{\pi}(\theta_{-S}|\hat{\psi}, U^{obs})$ and $\hat{\pi}(\theta^S|\hat{\psi}, U^{obs})$ as estimates of $\pi(\theta_{-S}|\psi, U^{obs})$ and $\pi(\theta^S|\psi, U^{obs})$ respectively. The following theorem gives us the conditions under which the acceptability of $\hat{\pi}(\theta'|\hat{\psi}, U^{obs})$ as an estimate of $\pi(\theta'|\psi, U^{obs})$ for $\theta' = \{\theta_{-S}, \theta^S\}$ are established. The proof is explicitly stated in *Appendix*.

Theorem 5.3.1 *If $\hat{\psi}_N \overset{a.e.}{\to} \psi$ as $N \to \infty$ with respect to the marginal density $m(U^{obs}|\psi)$ and $\pi(\theta'|\psi, U^{obs})$ is continuous in ψ for $\theta' = \{\theta_{-S}, \theta^S\}$, then*

1. $\int_{\Theta_{-S}} | B^{-1} \sum_{j=1}^{B} \pi(\theta_{-S}|\theta^{S(j)}, \hat{\psi}, U^{obs}) - \pi(\theta_{-S}|\psi, U^{obs}) | \, d\theta_{-S} \overset{a.e.}{\to} 0,$
2. $\int_{\Theta^S} | B^{-1} \sum_{j=1}^{B} \pi(\theta^S|\theta_{-S}^{(j)}, \hat{\psi}, U^{obs}) - \pi(\theta^S|\psi, U^{obs}) | \, d\theta^S \overset{a.e.}{\to} 0,$

as $B, N \to \infty$.

Proof. Prior to prove this theorem, here we present two Lemmas in general form and associated preliminary setup with notations.

In the general Bayesian setup (section 5.2), our model of interest, treated as a missing data model, presents the hierarchy as

$$\begin{aligned} U^{obs} &\sim f(x|\theta, \psi), \theta = (\theta^S, \theta_{-S}), \\ \theta^S &\sim \pi(\theta^S|\psi, \gamma), \\ \theta_{-S} &\sim \pi(\theta_{-S}|\psi). \end{aligned} \tag{5.15}$$

In EWiG, the estimate of posterior density of θ_{-S} and θ^S are given in (5.13) and (5.14) respectively. Here, we are interested in the limiting behavior of these estimates as B and $N \to \infty$. Let us define

$$\begin{aligned} g(\psi; \psi') &= \int_{\Theta^S} \pi(\theta_{-S}|\theta^S, \psi, U^{obs}) \pi(\theta^S|\psi', U^{obs}) d\theta^S, \\ \hat{g}(\psi; \psi') &= B^{-1} \sum_{j=1}^{B} \pi(\theta_{-S}|\theta^{S(j)}, \psi, U^{obs}), \end{aligned}$$

where $\theta^{S(j)} \sim \pi(\theta^S|\theta_{-S}^{(j)}, \psi', U^{obs})$. Since, our present study belongs to finite population statistics, hence the estimator $\hat{\psi}_N$ of nuisance parameter ψ depends on N. Now, the following lemma presents the conditions on which almost everywhere (a.e.) convergence of $\hat{\pi}(\theta_{-S}|\hat{\psi}, U^{obs})$ to $\pi(\theta_{-S}|\psi, U^{obs})$ can be built.

Lemma 5.3.2 *Let us consider ψ_0 as a true value of ψ. Under the following assumptions along with ergodicity of EM generated Markov chain:*
A1. $\hat{\psi}_N \to \psi_0$ almost everywhere,
A2. $g(\psi;\psi')$ is continuous function of ψ and ψ',
A3. $\hat{g}(\psi;\psi')$ is continuous in ψ and stochastically equicontinuous in ψ',

$\exists$ a sequence $B_N \ni B_N \to \infty$ as $N \to \infty$, for which

$$|\hat{g}(\hat{\psi}_N;\hat{\psi}_N) - g(\psi_0;\psi_0)| \stackrel{a.e.}{\to} 0 \text{ as } N \to \infty. \tag{5.16}$$

Proof of the Lemma. By triangle inequality we can write

$$|\hat{g}(\hat{\psi}_N;\hat{\psi}_N) - g(\psi_0;\psi_0)| \leq |\hat{g}(\hat{\psi}_N;\hat{\psi}_N) - \hat{g}(\psi_0;\psi_0)| + |\hat{g}(\psi_0;\psi_0) - g(\psi_0;\psi_0)|.$$

By Ergodic property, second term on right hand side converges to 0. For any given N, choose $B_N^{(1)}$ so that $|\hat{g}(\psi_0;\psi_0) - g(\psi_0;\psi_0)| \leq \epsilon/3$. For first term, again by triangle inequality,

$$|\hat{g}(\hat{\psi}_N;\hat{\psi}_N) - \hat{g}(\psi_0;\psi_0)| \leq |\hat{g}(\hat{\psi}_N;\hat{\psi}_N) - \hat{g}(\psi_0;\hat{\psi}_N)| + |\hat{g}(\psi_0;\hat{\psi}_N) - \hat{g}(\psi_0;\psi_0)|.$$

From assumption A3, the first term on right hand side in the above inequality tends to 0 as $N \to \infty$. Since $\hat{\psi}_N \to \psi_0$ a.e. by assumption A1, we can choose N large enough so that $|\hat{\psi}_N \to \psi_0| < \delta$, except on a set with probability less than $\epsilon/2$. Assumption A3 also says that $\hat{g}(\psi;\psi')$ is stochastically equicontinuous in ψ' which means, for given $\epsilon > 0$, one can find a $\delta(>0) \ni |\hat{g}(\psi;\psi'_1) - \hat{g}(\psi;\psi'_2)| < \epsilon \; \forall \; |\psi'_1 - \psi'_2| < \delta$ except on a set with g-measure 0. Hence, we can choose $B_N^{(2)}$ so that $|\hat{g}(\psi_0;\psi') - \hat{g}(\psi_0;\psi_0)| \leq \epsilon/3 \; \forall \; |\psi' - \psi_0| < \delta$, except on a set with g-measure less than $\epsilon/2$. Hence, the second term in the right side is bounded as follows

$$|\hat{g}(\psi_0;\hat{\psi}_N) - \hat{g}(\psi_0;\psi_0)| \leq \sup_{\psi':|\psi'-\psi_0|<\delta} |\hat{g}(\psi_0;\psi') - \hat{g}(\psi_0;\psi_0)| \leq \epsilon/3.$$

Thus, for any arbitrary $\epsilon > 0$, we can choose a large N and $B_N = max(B_N^{(1)}, B_N^{(2)})$. Therefore, $|\hat{g}(\hat{\psi}_N;\hat{\psi}_N) - g(\psi_0;\psi_0)| \leq \epsilon$, except a set with probability less than $\epsilon/2 + \epsilon/2 = \epsilon$. ■

Lemma 5.3.3 *(Scheffé's Lemma.) If f_n is a sequence of integrable functions on a measure space (X, Ω, μ) that converges almost everywhere to another integrable function f, then,*

$$\int |f_n(y) - f(y)| \, d\mu \to 0 \text{ if and only if } \int |f_n| \, d\mu \to \int |f| \, d\mu \text{ for } n \to 0.$$

Proof of Theorem 5.3.1. Both the Lemma 5.3.1 and Lemma 5.3.3 together imply the *Theorem 5.3.1*(*1*) as $f_n = \hat{g}_N$ and $f = g$ both are density and hence positive. *Theorem 5.3.1*(*2*) can be proved following the same way just by exchanging $\hat{\pi}(\theta_{-S}|\hat{\psi}, U^{obs})$ and $\pi(\theta_{-S}|\psi, U^{obs})$ in their places and defining

$$
\begin{aligned}
g(\psi;\psi') &= \int_{\Theta_{-S}} \pi(\theta^S|\theta_{-S},\psi,U^{obs})\pi(\theta_{-S}|\psi',U^{obs})d\theta_{-S},\\
\hat{g}(\psi;\psi') &= B^{-1}\sum_{j=1}^{B}\pi(\theta^S|\theta_{-S}^{(j)},\psi,U^{obs}),
\end{aligned}
$$

where $\theta_{-S}^{(j)} \sim \pi(\theta_{-S}|\theta^{S(j)},\psi',U^{obs})$. ■

If θ refers to the missing and/or unobserved quantity only and ψ includes the rest of all the model parameters then our sketched EWiG reduces to original MCEM.

Implementation. Here we take $\theta^S = N$, $\theta_{-S} = (p_{1\cdot},\phi)$ and $\psi = p$. With this consideration of θ_{-S} and ψ, we call this method as EWiG-I. Priors on N and $p_{1\cdot}$ are assigned as stated earlier in case of DA-Lee (section 3.1). Same flat prior *Unif*(α,β) is chosen for ϕ, but here we give a plan to assign these prior limits strategically. When the population is recapture averse, $\alpha = c$ and $\beta = 1$ will be the automatic choice. Since $c = \phi p < \phi$ and $\phi p = c < 1 \Leftrightarrow \phi < p^{-1}$, we suggest α to 1 and $\beta = p^{-1}$ for recapture-prone population. So in both the cases, natural bounds are incorporated for the limits in *Unif*(α,β) depending upon the availability of directional knowledge on ϕ. Hence, step 2 is performed between the conditional posteriors (5.8), (5.9) and (5.11). In practice, c is replaced by its *mle* $\hat{c} = x_{11}/x_{1\cdot}$. Using updated sample $\{\theta^{(j)} = (N^{(j)}, p_{1\cdot}^{(j)}, \phi^{(j)}); j = 1(1)B\}$, p is estimated by maximizing the empirical average of log-likelihoods $\ell(p|\theta^{(j)},\underline{\mathbf{x}})$ for $j = 1(1)B$ following (11).

When no directional knowledge on ϕ is available, we use conjugate prior $\pi(\phi|p)$ = Generalised Beta Type-I (u, v, $rate = p$) which results

$$
\pi(\phi|p,\underline{\mathbf{x}}) \propto \text{GB-I}\ (x_{11}+u,\ x_{10}+v,\ 1,\ \text{rate} = \text{p}) \tag{5.17}
$$

and therefore, corresponding log-likelihood for p becomes $ln[p^{x_{\cdot 1}}(1-p)^{N-x_0}(1-\phi p)^{x_{10}+v-1}]$. If we take non-informative prior with $(u,\ v)$=(0, 0), that leads to eliminate the influential terms in (14) and conditional log-likelihood for p. If we take $(u,\ v)$=(1, 1), prior $\pi(\phi|p)$ simply reduces to constant density $\mathscr{I}_{(0,1/p)}(\phi)$ on natural domain $(0, p^{-1})$ of ϕ.

Another strategy can be adopted which allow to consider $\psi = \{p, \phi\}$, that means θ_{-S} includes $p_{1\cdot}$ only. Thus, unlikely to EWiG-I, here do not consider any prior on ϕ in addition to p. To update p, Step 3 is followed as usual but ϕ is updated by the relation $\hat{c}/p$, for given p. However, We name this new variant as EWiG-II and it can be noted that it is very close to the original MCEM approach as samples from posterior density (8) will eventually converge to its *mle* due to the flat prior $\pi(p_{1\cdot}|p) = Unif(0,1)$.

For both the EWiG-I and EWiG-II method, consistency conditions for p and ϕ in Theorem 5.3.1 holds as $\hat{p}_N = x_{01}/(N - x_{1\cdot}) \to p_{01}/(1-p_{1\cdot}) = p$ and $\hat{\phi}_N = \hat{c}/\hat{p}_N \to c/p = \phi$ for $N \to \infty$. However in this application, θ^S is equivalent to N which is also the index in the asymptotic result in finite population inference. Hence, the Theorem 5.3.1(2) does not hold here exactly. Therefore, we have following theorem (proved in *Appendix*) about the tail convergence of $\hat{\pi}(N|\hat{p}, \underline{\mathbf{x}})$ to $\pi(N|p_0, \underline{\mathbf{x}})$ for the true value p_0 of p which suffices for the posterior convergence on N when it is large enough.

Theorem 5.3.4 *Since $\hat{p}_N \overset{a.e.}{\to} p_0$ as $N \to \infty$ with respect to the marginal density $m(\underline{\boldsymbol{x}}|p)$ and $\pi(N|p, \underline{\boldsymbol{x}})$ is continuous in p, then $\hat{\pi}(N|\hat{p}, \underline{\boldsymbol{x}})$ is right tail equivalent to the marginal posterior $\pi(N|p_0, \underline{\boldsymbol{x}})$ for sufficiently large B.*

Proof. Two density functions F and G are said to be *right tail equivalent* if they have the same right endpoints ω $(\leq \infty)$ and $\lim\limits_{x\uparrow\omega} \frac{1-F(x)}{1-G(x)} = c$, for some constant $0 < c < \infty$.
Consider $\theta^S = N - x_0$, $\theta_{-S} = (p_{1\cdot}, \phi)$ and $\psi = p$, hence for large N and $M \to \infty$, Lemma produces

$$\hat{\pi}(N|\hat{p}, \underline{\mathbf{x}}) = B^{-1} \sum_{j=1}^{B} \pi(N|p_{1\cdot}^{(j)}, \phi^{(j)}, \hat{p}, \underline{\mathbf{x}}) \quad \overset{a.e.}{\to} \pi(N|p_0, \underline{\mathbf{x}}). \blacksquare$$

Note that Theorem 5.3.4 also holds for EWiG-II.

5.3.2 Stochastic EM-within-Gibbs [SEMWiG]

Constructing data augmentation schemes that result in both simple, quicker as well as efficient algorithms is a matter of art and also depends greatly on the underlying models. We now propose a very interesting alternative Gibbs sampling technique which is simpler, computationally easier and encompasses certain advantages over EWiG. This method is a stochastic extension of EM within Gibbs and hence, we name it as Stochastic EM-within-Gibbs

[SEMWiG]. If we put M=1 in original MCEM (Wei and Tanner, 1990 [101]), then we would have Stochastic Expectation-Maximization (SEM) procedure. Though the connection between SEM and MCEM seems very simple but their underlying philosophy is totally different. In SEM, stochastic imputation is done for the unobserved missing quantity θ^S and therefore, θ_{-S} is estimated from the complete data log-likelihood. However, MCEM replaces intractable computation of the conditional expectation of the log-likelihood of the complete data using a monte carlo approximation. Several evidences on the preference of SEM over EM are in the literature. SEM algorithm may accelerate the converegence (Celeux, 1985 [16]) and it is known to be more robust to poorly specified starting values (Gilks et al., 1996 [43]) than EM. SEM provides estimate of posterior density for θ^S but point estimates for other model parameters. However, SEMWiG is so designed with flexibility that one can have posterior density estimate also for θ_{-S} by suitably modifying the iteration steps. The whole Gibbs sampler including SEM step proceeds as follows:

Step 1	:	Set $t=0$ and initialize $\psi^{(0)}$ and $\theta_{-S}^{(0)}$.
Step 2	:	*Stochastic Imputation.* Generate pseudo-complete data by simulating θ^S from $\pi(\theta^S\|\theta_{-S}^{(t)},\psi^{(t)},U^{obs})$ and then simulate θ_{-S} from $\pi(\theta_{-S}\|\theta^S,\psi^{(t)},U^{obs})$.
Step 3	:	*Maximization.* Update $\psi^{(t)}$ by using $\psi^{(t+1)}=\underset{\psi\in\Psi}{\operatorname{argmax}}\left\{\ell(\psi\|\theta^S,\theta_{-S},U^{obs})\right\}$.
Step 4	:	Repeat above steps until the convergence of $\{\theta^{S(t)},\theta_{-S}^{(t)}\}_{t\geq 0}$.

Finally, $\hat{\psi}$ is calculated by averaging over a sufficient numbers $\psi^{(t)}$ after reaching its stationary regime. Equivalent result as in Theorem 5.3.1 can be proved also for SEMWiG under the assumption of ergodicity of SEM generated homogeneous Markov chain and consistency of $\hat{\psi}$. Application of SEM algorithm does not result in a single value for a parameter estimate. Instead, there is built-in variation induced by the simulated data around the estimate.

Implementation. To implement SEMWiG, here also we consider $\theta^S=N$, $\theta_{-S}=(\phi,p_{1\cdot})$ and $\psi=p$. Same priors are consider for N, $p_{1\cdot}$ and ϕ as previously discussed in EWiG-I, depending upon the availability of the directional knowledge on ϕ. The initial value of θ may come from a wide range of choices, so it might become unstable at the beginning of the process. We produce a Gibbs sequence $\{(p_{1\cdot}^{(t)}, \phi^{(t)}, N^{(t)});\ t=1,2,3,...\}$ and hence choose the burn-in period following the same $\hat{R}^{1/2}$ technique (Gelman, 1996 [41]) as like DA-Lee method.

5.4 Simulation Results

In this section we study the comparative performance of two proposed variants of Gibbs sampler (EWiG and SEMWiG) and one existing Gibbs sampler (DA-Lee) following Lee et al. (2003) in DRS under the model M_{tb}. We consider four artificial populations as in first paragraph of section 3.3.2 characterized by different values of capture probabilities ($p_{1\cdot}$, $p_{\cdot 1}$) and the value of $N = 500$ and $\phi = 1.25$ and 0.80. These for populations for each ϕ are presented in Table 3.1. 200 data sets on (x_{11}, x_{01}, x_{10}) were generated from each population.

To compute Bayes estimates proposed by Lee et al. (2003[62]), we follow Lee et al. (2003[62], pp. 482) to implement ARS for DA-Lee only. Let $S_n = \{a_1, a_2, \ldots, a_n\}$, where $a_1 < a_2 < \ldots < a_n$, denote a current set of abscissa in the range (0, 1). Considering $n = 5$, the initial value of a_3 (middle component of the set of abscissa, S_n for $n = 5$) is chosen as $(\hat{c}/\phi)$ from the relation $E(x_{11}) = p\phi E(x_{1\cdot})$, since $\hat{c} = p\phi$ and $\hat{c} = x_{11}/x_{1\cdot}$. DA-Lee and SEMWiG are designed in a way so that convergence of their Gibbs sampler can be examine through $\hat{R}^{1/2}$. To compute $\hat{R}^{1/2}$ in multiple chain method, we take 5 independent parallel chains and hence burn-in period is fixed at a value h for which $\hat{R}^{1/2}$ becomes smaller than 1.1. Convergence of EWiG methods is judged by plotting $N^{(h)}$ against h. After fixing the burn-in h, next h number of samples are used to construct the estimate of posterior densities for all approaches. Estimate of N is calculated as posterior mean. The whole estimation task is replicated 200 times and therefore, final estimate of N is presented just by averaging over 200 replicated estimates. Based on those 200 estimates, the sample RMSE (Root Mean Square Error) for the estimators are also calculated. Besides this, 95% credible interval (C.I.) is obtained based on sample quantile of the marginal posterior distribution of N for each replicate. Then final C.I. is obtained by averaging over those 200 replicated credible intervals. We also find the coverage probability of true N for each method. All the results for different variants of Gibbs samplers are summarized in Table 5.1 for $N = 500$. Based on availability of directional knowledge, priors on ϕ are assigned to all methods according to the respective recommendations, except for EWiG-II. In EWiG-II method, under the assumption of $\phi > 1$, p is maximised over the equivalent domain (0, $\hat{c}$) and for $\phi < 1$, p is maximised over ($\hat{c}$, 1). Computations of all the algorithms of four methods (considering EWiG-I and EWiG-II separately) are exercised through latest R-packages.

When the knowledge $\phi > 1$ is available (left-top panel of Table 5.1), all other methods perform relatively better than EWiG-II. It is also noted that overall SEMWiG estimates are most efficient in terms of rmse and coverage. DA-Lee with prior U(1,3) produces almost same result as U(1,2) but it has larger confidence length. Length of DA-Lee is too wide for small capture probabilities (P5 and P6). In larger capture probability situations (particularly, P2 and P3), DA-Lee, EWiG-I and SEMWiG are closely comparable. For small capture probabilities, EWiG-I performed consistently better.

When the correct knowledge of $\phi > 1$ is not available (left-bottom panel of Table 5.1), EWiG-I performed better in case of relatively low capture probability and $p_{1.} < p_{.1}$ (for P1 and P5). On the other hand, when $p_{1.} > p_{.1}$ and low capture probabilities, EWiG-II is better. In general, SEMWiG stands a little bit higher in terms of performance than EWiG-I and can be considered as the most efficient, except for coverage where Lee performs better, due to the too wide confidence interval.

In right-top panel of Table 5.1, when the knowledge $\phi < 1$ is available, SEMWiG works better than all other methods for all moderate and large capture probabilities (i.e. P7-P10) in terms of all criteria. Both the methods in EWiG perform satisfactorily better than DA-Lee, except population P9 and P11. It is noted that the EWiG methods overestimate the population size for all populations except the very low capture probabilities (P7-P10). Again here DA-Lee fails to generate Gibbs samples in most of the cases when available directional knowledge $\phi < 1$ is being used. Therefore, we follow the same remedy as done for $N = 200$.

When data is generated from the populations with $\phi < 1$ and the knowledge $\phi < 1$ is not available (i.e. right-bottom panel of Table 5.1), DA-Lee produces estimates with very wide confidence intervals like earlier situations and that results in high coverage except for low capture situation P11. In terms of coverage and rmse, SEMWiG works efficiently except for P12. SEMWiG may be the overall choice for its better efficiency, smaller rmse, more and substantial amount of coverage than EWiG methods.

Graphical Comparison between DA-Lee & SEMWiG

Figure 5.1 shows that $\hat{R}_k^{1/2}$ stabilizes very fast ($k \geq 150$) for all populations in case of DA-Lee. However, estimates are ultimately converging at far above (or distant from) the true value of N (= 500), approximately for h = 150, 200. This is because the non-informative priors for all model parameters could not help completely to get rid of the present complexity of the model in case of DA-Lee method. Moreover, use of directional knowledge on ϕ ($\phi < 1$, when actually it is), may mislead us, which is of course a negative feature of this method. Whereas convergence of SEMWiG estimates (from Figure 5.2) are better and less fluctuating except in the case of very low capture probabilities. Clearly, the figure shows a good parity of the SEMWiG estimates with the stabilization of $\hat{R}^{1/2}$ values. Reason behind this is SEMWiG avoids the identifiability problem, existing in likelihood, relatively better than DA-Lee as it successfully incorporates constant prior on the system-driven precise domain knowledge on ϕ.

5.5 Real Data Application: Malawi Death Data

Greenfield (1975) reports a DRS data on birth, death and migration obtained from a Population Change Survey conducted by the National Statistical Office in Malawi between 1970 and 1972. The sample was stratified into five strata. To illustrate the application of the two methods proposed in earlier sections and also to investigate the performance of the existing DA-Lee approach in literature, we choose the Malawi data on death records (*see* section 1.8.1 of chapter 1) only for two strata - Lilongwe ($\hat{c} = 0.593$, $x_{10} > x_{01}$) and Other urban areas ($\hat{c} = 0.839$, $x_{10} < x_{01}$) due to its different $\hat{c}$ values and opposite nature of x_{10} and x_{01} values. Significantly lower $\hat{c}$ value helps to anticipate that the people of *Lilongwe* are less keen to give the information on deaths again in survey time than that of *Other urban areas* people. Nour (1982) estimated the death sizes as *378* and *3046* for *Lilongwe* and *Other urban areas* respectively, considering the assumption that two data sources are positively correlated in human demographic study which is equivalent to $\phi > 1$. Table 5.2 summarizes the results from three methods discussed earlier. Under both the consideration of unrestricted ϕ (*i.e.* $\phi > 0$) and restricted ϕ (*i.e.* $\phi > 1$), 200 parallel chains for N are generated for different initial values. For DA-Lee, burn-in period k is fixed at 100 for *Lilongwe* and 2500 for *Other urban areas* after observing the plot of $\hat{R}^{1/2}$. Similarly for SEMWiG method, burn-in period is fixed at 1100 for *Lilongwe* and 9000 for *Other urban areas.*

For $\phi > 0$, DA-Lee says that the estimated number of deaths in *Lilongwe* is 360 (with 95% credible interval (348, 390)) and in *Other urban areas* is around 3280 (with 95% CI (2865, 3700)). DA-Lee with prior $U(0.5, 3)$ gives larger estimate but confidence intervals become too wide. SEMWiG produces the estimates as 356 and 2845. When recapture proneness is assumed (i.e. $\phi > 1$), DA-Lee estimates increases to 373 (with 95% CI (357, 397)) for *Lilongwe* and 3286 (with 95% CI (3005, 3647)) for *Other urban areas.* For large population, DA-Lee is larger than Nour's(1982[68]). EWiG-I says that 370 and 3200 deaths occurred in *Lilongwe* and *Other urban areas* respectively. When we also consider behavioral effect ϕ as another nuisance parameter, then the estimates from relevant method EWiG-II for *Lilongwe* and *Other urban areas* are around 365 and 3200 respectively. For both the populations, estimates from SEMWiG are slightly smaller than that of Nour (1982[68]) and gives around 372 and 2980 respectively.

In summary, for large population (*Other urban areas*), estimates from EWiG-I and EWiG-II are coincident. Use of directional knowledge improves the efficiency for all the estimates basically for small population. SEMWiG produces the lowest s.e. and hence smaller length of confidence interval when directional knowledge of recapture proneness is assumed. When we do not use the recapture proneness knowledge, all three methods agree that *Lilongwe* might not be recapture prone.

5.6 Conclusion

Motivated by the extensive use of Dual-record system (DRS) in various real life practices, we consider the problem of homogeneous population size estimation in M_{tb}-DRS framework based on Bayesian techniques. The present model M_{tb} suffers from non-identifiability of behavioral response effect, ϕ and hence, suitable Bayesian methods are thought to have the potential to overcome that burden to some extent. Moreover, we think that incorporation of correct prior knowledge on directional nature of ϕ, if available, helps in producing better estimates. Later in Chapter 7, we deal with the problem of classifying the directional nature of behavioral effect ϕ and propose an efficient workable strategy. In this chapter, we develop two empirical Bayes approaches conditionally and unconditionally on the directional knowledge available on ϕ under a common roof of missing data analysis and compare them with one of a few existing full Bayes methods in literature. Performance of the existing approach DA-Lee (Lee et al., 2003 [62]), which was designed in the spirit of multiple list problem, is investigated in DRS situation. As per our knowledge, it is the first attempt to make inference for this complex DRS situation via Bayesian methodologies. We have restricted ourselves to the use of non-informative or minimum informative constant priors so that subjectiveness can be reduced in the prior selection which makes the inference robust.

In general it has been found that SEMWiG is overall the most efficient method (both in terms of RMSE and s.e.). On the other hand, DA-Lee provides wide credible interval than any other method but also it possesses lower efficiency in most situations than SEMWiG. When applied to the present M_{tb}-DRS setup, using directional knowledge on ϕ, DA-Lee sometimes encounters problem of drawing samples from its highly dispersed conditional posterior densities. Moreover, the trial-and-error approach in DA-Lee may take a long time to discover a suitable range for uniform prior for ϕ and DA-Lee estimates ultimately may converge around a highly overestimated value in long run (*see* Figure 5.1). Coming to the

EWiG methods, EWiG-I performs better than EWiG-II in terms of efficiency when prior directional information on ϕ is known. Beside the advantage of less computational time and complexity, SEMWiG is easier than EWiGs to explain to the practitioners. Thus Bayesian analysis along with likelihood estimates (for nuisance parameters) helps to maintain small variability and converges around a reasonably good value. Thus after all kind of inspections, we find that empirical Bayes method is useful to construct an efficient strategy, such as SEMWiG, for population size estimation in this complex DRS depending upon the various possible practical situations. Though the proposed empirical Bayes methods consume little more computational time than the full Bayes DA-Lee (Lee et al., 2003 [62]) method, they also help successfully to get rid of the present identifiability problem efficiently and specifically, SEMWiG produce either comparable or more efficient estimates than the fully Bayes method.

Table 5.1: Summary results of all the discussed Gibbs sampling approaches applied to the populations P1-P8 with N=500.

Method & $\pi(\phi)$		P1	P2	P3	P4	P5	P6	P7	P8
When correct information on the directional nature of ϕ is available									
DA-Lee	$\hat{N}$(rmse)	472(34.11)	478(24.33)	490(12.02)	488(17.03)	472(35.42)	490(13.65)	509(10.34)	481(20.08)
U(1,2)	C.I.	(438,518)	(451,519)	(473,515)	(457,531)	(431,565)	(460,560)	(485,545)	(448,545)
	Coverage	76	92	92	97	99	99	100	100
EWiG-I	$\hat{N}$(rmse)	478(26.41)	490(15.30)	489(12.51)	477(26.53)	530(35.99)	529(31.93)	517(18.86)	520(24.04)
$U(1, p^{-1})$	C.I.	(457,503)	(471,510)	(477,505)	(458,500)	(506,557)	(510,550)	(504,531)	(499,544)
	Coverage	60	82	66	50	37	28	48	53
EWiG-II[a]	$\hat{N}$(rmse)	467(36.90)	472(30.70)	487(15.16)	485(20.82)	526(37.08)	514(27.39)	524(27.10)	522(26.17)
$\phi > 1$	C.I.	(447,489)	(456,490)	(475,501)	(464,509)	(495,558)	(522,565)	(511,542)	(499,548)
	Coverage	24	23	54	68	60	46	29	55
SEMWiG	$\hat{N}$(rmse)	484(23.92)	485(21.25)	495(9.87)	502(16.32)	478(25.27)	492(11.74)	498(5.28)	484(19.34)
$U(1, p^{-1})$	C.I.	(451,522)	(456,511)	(477,512)	(474,532)	(455,507)	(477,510)	(489,508)	(465,506)
	Coverage	98	86	88	100	60	96	75	71
When no information is available on ϕ									
DA-Lee	$\hat{N}$(rmse)	468(37.94)	483(24.97)	485(16.97)	471(30.61)	474(35.58)	512(19.83)	516(18.71)	517(21.75)
U(0.5,2)	C.I.	(398,561)	(426,560)	(460,513)	((422,542)	(431,566)	(461,575)	(486,553)	(451,615)
	Coverage	88	95	91	100	99	100	100	100
EWiG-I	$\hat{N}$(rmse)	488(33.61)	504(21.90)	507(12.81)	533(39.80)	495(23.70)	532(40.45)	540(43.55)	569(66.03)
$\pi(\phi\|p) \propto 1$[b]	C.I.	(460,520)	(480,532)	(489,529)	(492,568)	(463,530)	(500,569)	(512,571)	(519,615)
	Coverage	72	79	92	65	83	40	28	23
EWiG-II	$\hat{N}$(rmse)	456(46.84)	463(40.89)	480(21.40)	476(27.67)	465(39.87)	518(26.25)	528(29.42)	468(34.62)
$\phi > 0$	C.I.	(439,476)	(450,489)	(470,492)	(457,500)	(449,486)	(498,538)	(514,546)	(458,482)
	Coverage	19	40	48	55	22	39	26	19
SEMWiG[a]	$\hat{N}$(rmse)	459(43.57)	487(19.03)	496(8.81)	458(44.42)	474(27.75)	512(15.27)	510(12.60)	487(18.30)
$\pi(\phi\|p) \propto 1$	C.I.	(431,483)	(459,513)	(481,510)	(438,477)	(454,497)	(494,525)	(498,526)	(466,517)
	Coverage	75	93	87	30	54	98	62	84

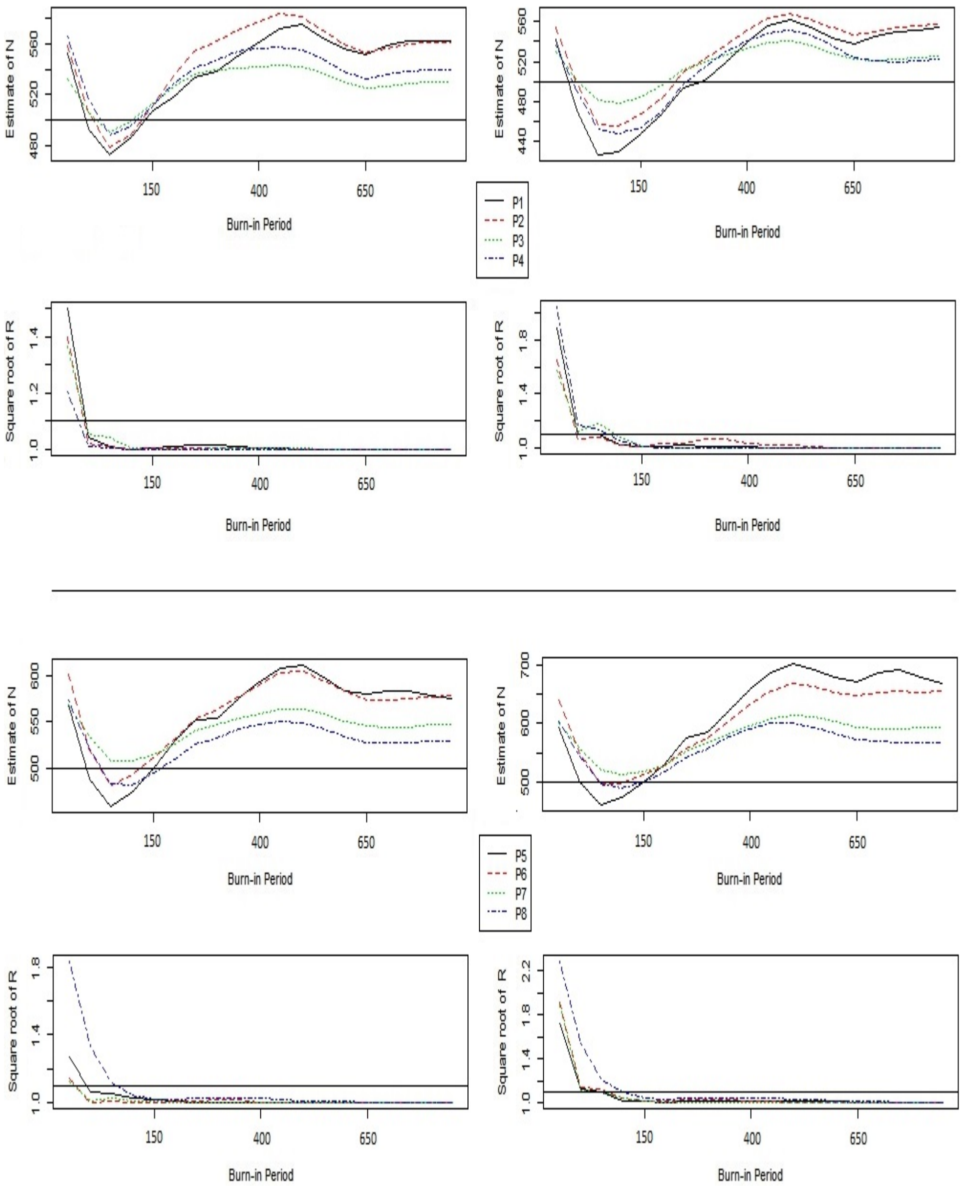


Figure 5.1: Plot of $\hat{N}$ (in 1st and 3rd row) and $\hat{R}^{1/2}$ (in 2nd and 4th row) against the index h for **DA-Lee** method. True value of N and suggested threshold value for $\hat{R}^{1/2}$ are indicated at 500 and 1.1 respectively. First two rows correspond populations **P1-P4** and last two rows for **P5-P8**. Plots in Left panel refers the situations of available directional knowledge on ϕ and right panel plots represent the situations with unavailable knowledge.

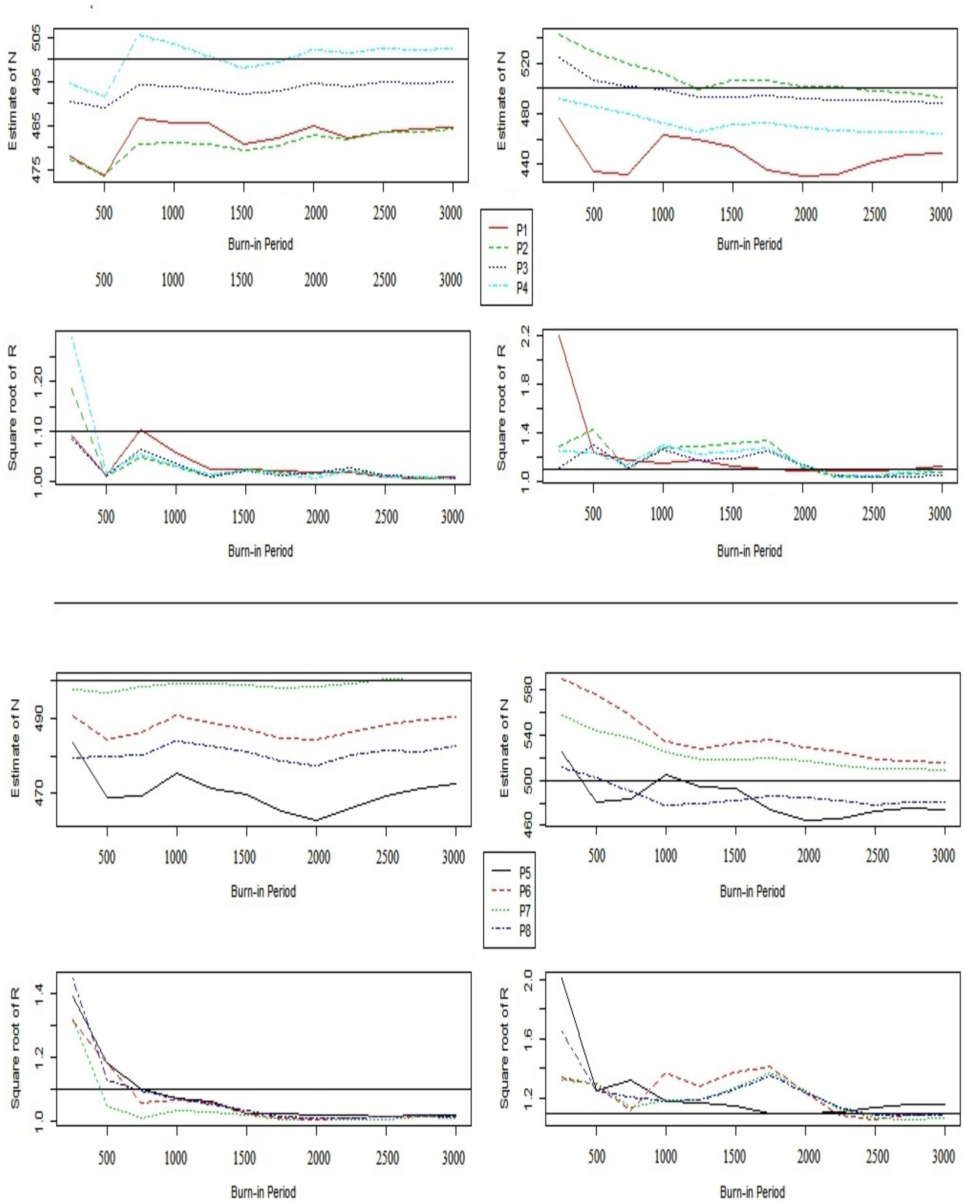


Figure 5.2: Plot of $\hat{N}$ (in 1st and 3rd row) and $\hat{R}^{1/2}$ (in 2nd and 4th row) against the index h for **SEMWiG** method. True value of N and suggested threshold value for $\hat{R}^{1/2}$ are indicated at 500 and 1.1 respectively. First two rows correspond populations **P1-P4** and last two rows for **P5-P8**. Plots in Left panel refers the situations of available directional knowledge on ϕ and right panel plots represent the situations with unavailable knowledge.

Table 5.2: Bayes and Empirical Bayes estimates with summary analyses for **Malawi Death data**, separately based on available or non-available directional knowledge on ϕ. s.e. is computed based on sample posterior distribution and the 95% posterior credible intervals for N and ϕ is determined based on percentile method.

Available knowledge	Method	$\pi(\phi)$	$\hat{N}$(s.e.)	95% CI of N	$\hat{\phi}$	95% CI of ϕ
			Lilongwe			
$\phi > 0$	DA-Lee[a]	$U(0.5, 2)$	359(12.51)	(348, 389)	0.86	(0.57, 1.62)
		$U(0.5, 3)$	362(12.84)	(348, 389)	0.92	(0.57, 1.62)
	EWiG-I	$\pi(\phi) \propto 1$	377(8.08)	(362, 394)	1.29	(1.11, 1.43)
	EWiG-II	—	351(2.03)	(348, 356)	0.67	-
	SEMWiG	$\pi(\phi) \propto 1$	356(4.08)	(352, 362)	0.79	(0.68, 0.92)
$\phi > 1$	DA-Lee[a]	$U(1, 2)$	373(10.98)	(357, 397)	1.22	(1.01, 1.59)
		$U(1, 3)$	373(9.65)	(357, 394)	1.22	(1.01, 1.64)
	EWiG-I	$U(1, 1/p)$	370(6.35)	(359, 383)	1.12	(1.01, 1.22)
	EWiG-II	—	365(5.22)	(356, 376)	1.01	-
	SEMWiG	$U(1, 1/p)$	372(1.30)	(370, 375)	1.16	(1.11, 1.20)
			Other urban areas			
$\phi > 0$	DA-Lee[a]	$U(0.5, 2)$	3276(218.75)	(2865, 3695)	1.37	(0.95, 1.79)
		$U(0.5, 3)$	3367(311.96)	(2898, 4184)	1.46	(0.98, 2.32)
	EWiG-I	$\pi(\phi) \propto 1$	3199(26.80)	(3149, 3251)	1.29	(1.26, 1.32)
	EWiG-II	—	3198(25.50)	(3150, 3249)	1.29	-
	SEMWiG	$\pi(\phi) \propto 1^a$	2847(26.03)	(2818, 2885)	0.93	(0.89, 0.96)
$\phi > 1$	DA-Lee[a]	$U(1, 2)$	3286(172.31)	(3005, 3647)	1.38	(1.09, 1.75)
		$U(1, 3)$	3360(254.19)	(3008, 3986)	1.46	(1.09, 2.10)
	EWiG-I	$U(1, 1/p)$	3199(25.56)	(3151, 3247)	1.29	(1.27, 1.32)
	EWiG-II	—	3196(25.44)	(3149, 3250)	1.29	-
	SEMWiG	$U(1, 1/p)$	2981(13.37)	(2970, 3001)	1.06	(1.05, 1.09)

6 Empirical Bayes method with Functionally Dependent Prior

6.1 Introduction and Motivation

DSE $\hat{N}_{ind}$ doesn't work satisfactorily as the violation of the underlying causal independence assumption often occurs between the capture probabilities. Chapter 1 shows that, for a homogeneous population, the capture probabilities might become correlated due to list dependence and it occurs when capture probability at the time of second survey depends on whether he/she is captured in first time. This behavioral dependency is driven by a parameter ϕ ($\in R^{+}$), called *Behavioral Response Effect*, incorporated in the model M_{tb}, in order to extend the model M_t. Thus, model M_{tb} has a strong relevance in practice and it acts as a generic model for a group of homogeneous individuals when the sample lists are not thought to be independent, or, at least when experimenter doesn't have assured knowledge that the given dual system is causally independent. A serious weakness of the model is that it suffers from nonidentifiability problem, which has been addressed in Otis et al. (1978[71]) not only for DRS, but also for the number of capture occasions more than two. For details about model M_{tb}, its nonidentifiability and associated review of few literature, readers are referred to sections 1.6 and 5.1. Keeping in mind the flexibility of Bayesian techniques and some successful Bayesian methods developed in this context, here also our basic aim is to develop some other suitable methodology for estimating N in M_{tb}-DRS context for homogeneous human population size (N) estimation as an alternative or supplement to the very few existing approaches where behavioral effect plays a key role along with time variation effect.

All of the few existing Bayesian techniques are proposed in the spirit of multiple capture-recapture model useful for wildlife population, where number of capture occasions are usually more than three. So, they require some assumptions, such as, recapture probabilities bear a constant relationship to initial capture probabilities or behavioral response effects (ϕ) remain same for all recapture stages. These assumptions purposefully avoid nonidentifiability problem and makes the model M_{tb} simpler. However, it is clear from the nature of the present M_{tb}-DRS model that no such assumptions is possible as here, only one recapture occasion is performed. Chapter 5 developed two strategic empirical Bayes approaches with minimum informative prior on the key parameter ϕ. Thus, from the nature of underlying model and its associated literature, it is clear that without the help of informative prior, it is not possible to have a reasonably good Bayesian solution. It is evident from chapter 5 that if the true directional knowledge on ϕ (i.e. whether $\phi > 1$ or < 1) is available, then expectedly, improved inference can be obtained. In a DRS framework, Nour (1982 [68]) proposed an estimator assuming the directional knowledge of recapture proneness (*i.e.* $\phi > 1$) in human demographic studies. Another possibility of recapture aversion (*i.e.* $\phi < 1$) might take place in some situations, e.g. drug abused population size estimation. In this chapter, we purposefully design a Gibbs sampling strategy using flat prior on suggested bound for ϕ when directional behavioral knowledge is available and a suitable informative prior on another nonidentifiable parameter p based on a functional constraint. On the other hand, by understanding a need to develop an approach when no such directional knowledge on ϕ is available (i.e. when $\phi > 0$), the proposed Gibbs sampling strategy is modified based on a informative conjugate prior on ϕ. In both situations, priors specification using the available domain knowledge, if available, preserve some characteristics of the underlying system and use minimal information so that efficient solutions could be obtained using simple Gibbs sampling techniques.

Both the strategies are presented along with their numerical evaluation over different simulated populations in sections 6.2 and 6.3. The full Bayes strategy as developed by Lee et al. (2003[62]) is also presented with or without the consideration of said directional knowledge. Section 6.4 examines a real DRS data on death records as an illustrative example. Finally in the last section, we summarize our findings and provide the best possible estimation rule depending upon the availability (or non-availability) of the directional knowledge on ϕ.

6.2 When Directional Knowledge on ϕ is Available

6.2.1 Proposed Methodology (AB-Flat)

In this section, we suggest a Gibbs sampling Bayesian strategy and propose associated computation technique so that the unidentifiability burden can be successfully overcome for estimating N with noninformative flat prior depending upon the available directional knowledge on ϕ. In (1.16), ϕ and p both are not identifiable but their product c $(= \phi p)$ is identifiable and well estimated by $\hat{c}_{mle} = (x_{11}/x_{1\cdot})$, where $E(\hat{c}_{mle}) \approx \frac{p_{11}}{p_{1\cdot}} = c$ even for moderately large N. Hence, for given c, we consider p as a function of ϕ only, i.e. $p = p(\phi|c) = c/\phi$. Thus, there is a functional constraint (i.e. $p = \phi/c$) that should be abided by any inferential strategy. We assign independent priors on $\Theta = (N, \phi, p_{1\cdot})$ as $\pi(\Theta) = \pi(p_{1\cdot})\pi(\phi)\pi(N)$ except p. For p, we assign a degenerated prior given ϕ, as $\pi(p|\phi) = 1$ for $p = \hat{c}/\phi$. The prime rationale behind the consideration of such informative prior is to maintain the functional dependence between two nonidentifiable parameters ϕ and p in the inferential process. Moreover, it helps to exercise the Bayesian strategy for such a complex model with a notable amount of lesser computational labour. Now, separate conditional posterior distributions for Θ are as follows

$$\pi(p_{1\cdot}|N) \propto p_{1\cdot}^{x_{1\cdot}}(1-p_{1\cdot})^{N-x_{1\cdot}}\pi(p_{1\cdot}), \tag{6.1}$$

$$\pi(\phi|N,p) \propto \phi^{x_{11}}(1-\phi p)^{x_{10}}\pi(\phi), \tag{6.2}$$

$$\pi(N-x_0|p_{1\cdot},p) \propto \frac{N!}{(N-x_0)!}((1-p_{1\cdot})(1-p))^N\pi(N), \tag{6.3}$$

We consider the noninformative prior for $p_{1\cdot}$ as $\pi(p_{1\cdot}) = Unif(0,1)$. Like Lee et al. (2003[62]), a flat prior density $\pi(\phi) = Unif(\alpha,\beta)$ is chosen for ϕ. It follows that (6.1) and (6.2) reduce to

$$\pi(p_{1\cdot}|N) \propto Beta(x_{1\cdot}+1, N-x_{1\cdot}+1), \tag{6.4}$$

$$\pi(\phi|p) \propto \text{GB-I}(x_{11}+1, x_{10}+1, 1, rate = p) \times \mathscr{I}_{[\alpha,\beta]}(\phi), \tag{6.5}$$

where $\mathscr{I}_{[\alpha,\beta]}(\phi)$ is an indicator function for $\phi \in [\alpha,\beta]$ and GB-I refers *Generalized Beta Type-I* density. Now, the hyperparameters α and β are to be chosen.

In addition to (6.1)-(6.3), Lee et al. (2003 [62]) also takes into account a prior on p; so they have identified known conditional posterior densities for N, $p_{1\cdot}$, ϕ and employed adaptive rejection sampling to generate p since explicit conditional posteriors for p was not available. For the choice of priors in (6.1)-(6.3) and also for p, readers are referred to section 5.2.1 of previous chapter.

If no other information on ϕ is available then choosing prior distribution is not at all easy. Lee et al. (2003[62]) proposed a trial-and-error procedure for this. They opt for such α and β for which the range of the posterior credible interval for ϕ is not too close to either side of the prior limits. They mentioned that this procedure seems to work well only when there is a large amount of recapture information. They also agreed that such kind of trial-and-error method has no theoretical justification and this judgement is highly subjective. For human population, high capture probabilities may be attained but number of samples is too small, usually, T=2. Now, we propose a model-driven prior limits for $\pi(\phi)$, or can say it is rather an automatic choice. Since $c = \phi p < \phi$, c is a good choice for the lower limit α. Hence, our selected objective prior is $\pi(\phi) = Unif(c, 1)$ when we know $\phi < 1$. If it is known that the population is recapture prone (i.e. $\phi > 1$), we recommend to set α to 1 and upper limit $\beta = 2$ or 3 is suitable for the analysis of human population. When directional knowledge is not at all available, $\alpha = c$ and $\beta = 2$ or 3 is a safe choice. Thus, $\pi(\phi)$ would be non-informative irrespective of the availability of directional knowledge on ϕ. Now, two different priors on N are considered as follows:

I. Poisson prior: $\pi(N) = Poi(\lambda)$, then conditional posterior (6.3) becomes

$$\pi(N - x_0|p_{1.}, p) \propto Poi(\lambda(1 - p_{1.})(1 - p)) \quad \text{and}$$

II. Jeffrey's prior: $\pi(N) \propto 1/N$, then conditional posterior (6.3) follows negative binomial distribution as

$$\pi(N - x_0|p_{1.}, p) \propto NB(x_0, \mu),$$

where, μ=$1 - (1 - p_{1.})(1 - p)$ and $p = c/\phi$, from definition of ϕ. For poisson prior, we can use empirical estimate of λ, as stated in George and Robert (1992 [42]). Here, we replace λ by $\hat{N}_{M_b}$, the likelihood estimate from model M_b (*see* section 1.5). Jeffrey's prior on N is also equivalent to the prior $\pi(N) = Poi(\lambda)$ with $\pi(\lambda) \propto 1/\lambda$. For the case $\phi > 1$, we also judge the performance of $\lambda = \hat{N}_{Nour}$, where $\hat{N}_{Nour}$ is the estimate of N based on Nour(1982 [68]).

At first we fix initial values $N^{(0)}$ and $\phi^{(0)}$. Then the initial value $p^{(0)}$ is obtained easily from the parametric relation $p = \hat{c}/\phi$ as $\hat{c}$ is consistent estimate for c. One can generate $p_{1.}^{(0)}$ from the conditional posterior (6.4) replacing N by $N^{(0)}$. Thereafter, Gibbs sampler proceeds to obtain a posterior adjusted by a functional dependence constraint, especially for N.

Thus, the above approach produces a Gibbs sequence $\{N^{(h)}, \phi^{(h)}, p^{(h)}, p_{1.}^{(h)}; h = 1, 2, 3,...\}$ by repeating this process $2k$ ($k > 0$, is to be specified) times. The initial value for Θ and therefore also for p comes from a wide range of choices, so it is generally unstable at beginning of the process. To avoid the influence of the starting value, we discard the first k iterative values as under the burn-in period and consider the remaining consecutive values to construct

Step 1	:	Simulate $\phi^{(1)}$ from $\pi(\phi\|p^{(0)})$, in (6.5).
Step 2	:	Simulate $N^{(1)}$ from $\pi(N-x_0\|p_{1\cdot}^{(0)},\phi^{(1)})$ and obtain $p^{(1)}=\hat{c}/\phi^{(1)}$.
Step 3	:	Generate $p_{1\cdot}^{(1)}$ from $\pi(p_{1\cdot}\|N^{(1)})$, in (6.4).
Step 4	:	Repeat the above three steps until the convergence.

posterior distributions for the model parameters. Here, conditioned on ϕ, the density $\pi(p|\phi)$ is considered to be the prior on p whose mean is empirically estimated by $\hat{c}/\phi$ and flat prior choice on ϕ is made based on available domain knowledge. Therefore, we call this strategy as empirical Bayes and denote it as *AB-Flat*. In addition to that, to maintain the functional relation between ϕ and p, we consider $\pi(p|\phi)$ as degenerate density (or equivalently, a point-mass prior) at $p=\hat{c}/\phi$, conditional on ϕ. We believe that such kind of functionally dependent restricted prior satisfying a structural relation, will help to get rid of the model complexity, especially when one tries to avoid subjective prior on the usual domain $(0,1)$ for p. Hence, The above approach is presented as a potential alternative to N estimation problem under the model M_{tb} when only two samples are available. Advantage of AB-Flat over Lee et al. (2003 [62]) is that the computational burden can be successfully overcome in order to generate Gibbs samples for p. Another advantage is that we don't need to setup prior limits for ϕ by trial-and-error method. Another relation, $p=x_{01}/(N-x_{1\cdot})$, suggested by Llyod (1994 [65]), can also be used in Step 2 in lieu of $p=\hat{c}/\phi^{(1)}$. Though, by this method we loose some information contained in the likelihood M_{tb}, but we can successfully overcome the model complexity and produce efficient estimates of N through a simpler computational involvement. Method *AB-Flat* can be implemented in practice for estimation with simple computation and maintenance of structural relation of underlying process. There are many methods available for diagnostic checking of posterior convergence in literature. In this study, we use the iterative simulation technique using multiple sequence method (*see* Gelman, 1996 [41]) and compute $\hat{R}^{1/2}$ exactly following Lee et al. (2003 [62], p.p. 483).

6.2.2 Numerical Illustrations

In this section we evaluate the performance of our approach proposed in last section and understand its efficiency in order to apply the method when directional knowledge on ϕ is available. Let us simulate eight hypothetical populations corresponding to four pairs of capture probabilities $(p_{1\cdot}, p_{\cdot 1})$={(0.50, 0.65), (0.60, 0.70), (0.80, 0.70), (0.70, 0.55)} in each case of *recapture prone* (represented by $\phi = 1.25$) and *recapture averse* (represented by $\phi = 0.80$) situations. All the populations are truly of size $N = 500$. Thus, here also we consider the identical populations incorporated in the earlier chapters. The results associated to the populations P1, P2, P3 and P4 are presented in Table 6.1, whereas associated results to other four populations P5, P6, P7 and P8 are shown in Table 6.2. The hyper-parameters α and β for $\pi(\phi) \propto U(\alpha, \beta)$ are taken as

(α, β)=$(1, 2)$ for prior domain knowledge $\phi > 1$ and

(α, β)=$(\hat{c}, 1)$ for prior domain knowledge $\phi < 1$.

The above priors are chosen depending upon the available true directional knowledge on ϕ. In addition to that, if no directional information on ϕ is available, another prior $U(\hat{c}, 2)$ is chosen assuming $\phi < 2$. 200 data sets on $(x_{1\cdot}, x_{\cdot 1}, x_{11})$ are generated from each of the eight populations. Our AB-Flat estimates have been obtained through simple Gibbs sampling from five independent parallel chains. Burn-in period is fixed at $k = 2000$ in general, after observing the performance of $\hat{R}^{1/2}$. Finally, estimate of N is obtained by averaging over 200 posterior means. Based on those 200 estimates, the bootstrap sample s.e. and sample RMSE (Root Mean Square Error) are calculated. We also compute the 95% credible interval (C.I.) based on sample quantile of the posterior distribution of N. In addition to that, we compute the similar statistics from the DA-Lee and SEMWiG methods. However, the DA-Lee method was illustrated originally in the context of animal capture-recapture experiment by Lee et al.(2003[62]) where large number of sampling occasions are typically considered. Their detailed computation strategy, particularly for DRS, can be found in section 5.2.1 of chapter 5. SEMWiG method and its computation details also can be found in chapter 5. Nour's (1982[68]) estimates are also calculated only for $\phi = 1.25$ cases as, Nour (1982[68]) deduced his approach for *recapture prone* situation. Their estimates as well as its S.E., RMSE, 95% confidence interval are computed over 200 generated datasets and present them as average estimate, sample SE, sample RMSE, 95% CI respectively in Table 6.1. When it is known that underlying $\phi > 1$, we also evaluate the performance of our AB-Flat approach with $\pi(N) \propto Poi(\lambda = \hat{N}_{Nour})$ (*see* section 1.6.1).

Population P1 and P2 in Table 6.1 demonstrate populations with $p_{1\cdot} < p_{\cdot 1}$. Here we evaluate the performance of our proposed approach with Jeffrey's prior (in first row), poisson prior with $\lambda = \hat{N}_{M_b}$ (in second row) on N corresponding to each of the four priors on ϕ previously mentioned. Estimates from priors $U(1,2)$ or $U(1,3)$ on ϕ are compared with the available Nour's (1982[68]) estimator. Results suggest that when one has the information that $\phi > 1$, the uniform priors $U(1,2)$ or $U(1,3)$ for ϕ is recommended for use. Another estimator is produced with prior Poi($\lambda = \hat{N}_{Nour}$) (in third row) for N, where $\hat{N}_{Nour}$ is the estimate of N due to Nour (1982 [68]). Tables show that the proposed approach from these two recommended priors are significantly better than the Nour's estimate based on RMSE and tighter confidence intervals around the true N. Results from Jeffrey's or Poi($\lambda = \hat{N}_{Nour}$) prior on N is better than that from Poi($\lambda = \hat{N}_{M_b}$). Overall, our proposed empirical Bayes solution, *AB-Flat*, based on flat uniform prior on ϕ and functionally dependent prior on p performs well and shows improvement over both the Lee's and Nour's approach, which is available only for $\phi > 1$.

Now we turn to the recapture averse cases and associated results are presented in Table 6.2. Population P5 and P6 in Table 6.2 demonstrates a case of recapture averse population with $p_{1.} < p_{.1}$. For P5, Bayes estimate corresponding to Jeffrey's prior performs moderately for $U(\hat{c},1)$ and if it is not known that ϕ is less than 1, then the other priors $U(\hat{c},2)$ overestimate the N due to low capture probabilities. Poisson prior with $\lambda = \hat{N}_{M_b}$ also misdirect the estimator due to same reason. In case of moderately high capture probabilities with $p_{1.} < p_{.1}$ in P6, our proposed strategy with $U(\hat{c},1)$ performs very well and other two estimates are also reasonably good. With the availability of the knowledge that ϕ is less than 1, Jeffrey's prior is relatively a better choice than poisson. Population P7 considers high capture probabilities with $p_{1.} > p_{.1}$. Estimate corresponding to the prior limit $(\hat{c},1)$ is better than other two priors. Though these other two estimates can be considered as good if we ignore their slight overestimation. For population P8 also, prior $U(\hat{c},1)$ with Jeffrey's prior on N produces reasonably good estimate whereas the other two priors are highly overestimates as the second capture probability is very small. For the situation when $p_{1.} > p_{.1}$, we recommend the use of poisson prior when no directional information on ϕ is available. Overall results from Table 6.2 indicate that our empirical Bayes estimate with prior $U(\hat{c},1)$ for ϕ works very well but one can use this range only when it is known that $\phi < 1$. Suppose the directional knowledge on the behavioral response effect is not available but we continue to use the AB-Flat method with the same uniform prior over the extended region $(\hat{c},2)$. These results are presented in the lower panels of both the Tables 6.1 and 6.2. For populations P1-P4, *AB-Flat* performs not well unless the probability of capture in List 1, $p_{1.}$, is very high. This happens as $U(\hat{c},2)$ extends the constant prior below 1 up to $\hat{c}$, so the estimates may become biased downwards. On the other hand, for populations P5-P8, other two priors can be employed with poisson prior for N. The results in these four tables also tell us that proposed approach, from other two prior limits, works

as results for these two cases are qualitatively similar, only present then for $U(1,2)$

Table 6.1: Summary results reflecting the performances of the proposed approach **AB-Flat** and other relevant methods applied to the populations P1-P4. Upper panel refers to the situation when the knowledge, $\phi > 1$, is available. Lower panel when nothing is available on ϕ.

Population (E(x_0))	Method: $\pi(\phi)$	Prior $\pi(N)$	Average Estimate	Sample SE	Sample RMSE	95%CI
		when it is known that $\phi > 1$				
P1(394)	Nour	-	479	14.84	25.49	(452, 509)
	Lee: U(1,2)	Jeffrey	472	18.94	34.11	(438, 518)
	SEMWiG: U($1, p^{-1}$)	Jeffrey	484	-	23.92	(451, 522)
	AB-Flat: U(1,2)	Jeffrey	497	20.51	20.89	(459, 539)
		Poi($\hat{N}_{M_b}$)	520	29.73	35.92	(469, 582)
		Poi($\hat{N}_{Nour}$)	489	17.99	21.04	(456, 526)
P2(422)	Nour	-	487	13.18	18.47	(461, 512)
	Lee: U(1,2)	Jeffrey	478	11.04	24.33	(451, 519)
	SEMWiG: U($1, p^{-1}$)	Jeffrey	485	-	21.25	(456, 511)
	AB-Flat: U(1,2)	Jeffrey	491	15.36	17.97	(460, 521)
		Poi($\hat{N}_{M_b}$)	492	15.84	17.77	(460, 524)
		Poi($\hat{N}_{Nour}$)	488	14.57	18.78	(459, 517)
P3(458)	Nour	-	499	8.74	8.76	(481, 516)
	Lee: U(1,2)	Jeffrey	490	7.55	12.02	(473, 515)
	SEMWiG: U($1, p^{-1}$)	Jeffrey	495	-	9.87	(477, 512)
	AB-Flat: U(1,2)	Jeffrey	499	9.06	9.08	(481, 517)
		Poi($\hat{N}_{M_b}$)	495	8.15	9.61	(478, 511)
		Poi($\hat{N}_{Nour}$)	499	8.86	8.98	(480, 516)
P4(420)	Nour	-	499	13.53	13.55	(473, 523)
	Lee: U(1,2)	Jeffrey	488	12.38	17.03	(457, 531)
	SEMWiG: U($1, p^{-1}$)	Jeffrey	502	-	16.32	(474, 532)
	AB-Flat: U(1,2)	Jeffrey	511	17.20	20.50	(481, 543)
		Poi($\hat{N}_{M_b}$)	489	12.81	16.84	(463, 511)
		Poi($\hat{N}_{Nour}$)	505	15.47	16.32	(478, 533)
		when no directional knowledge on ϕ is available				
P1	SEMWiG: $\pi(\phi\|p) \propto 1$	Jeffrey	459	-	43.57	(431, 483)
	AB-Flat: U($\hat{c}$,2)	Jeffrey	433	12.57	67.88	(410, 459)
		Poi($\hat{N}_{M_b}$)	447	16.33	55.68	(416, 481)
P2	SEMWiG: $\pi(\phi\|p) \propto 1$	Jeffrey	487	-	19.03	(459, 513)
	AB-Flat: U($\hat{c}$,2)	Jeffrey	448	10.24	53.38	(428, 466)
		Poi($\hat{N}_{M_b}$)	450	10.68	51.28	(428, 470)
P3	SEMWiG: $\pi(\phi\|p) \propto 1$	Jeffrey	496	-	8.81	(481, 510)
	AB-Flat: U($\hat{c}$,2)	Jeffrey	474	6.63	26.67	(460, 486)
		Poi($\hat{N}_{M_b}$)	473	6.52	27.64	(459, 485)
P4	SEMWiG: $\pi(\phi\|p) \propto 1$	Jeffrey	458	-	44.42	(438, 477)
	AB-Flat: U($\hat{c}$,2)	Jeffrey	458	10.56	43.66	(436, 477)
		Poi($\hat{N}_{M_b}$)	452	9.75	49.07	(432, 470)

reasonably good for high capture probabilities. It is also noted that estimates from prior Poi($\lambda = \hat{N}_{M_b}$) have smaller RMSE than that of Jeffrey's for high List-1 capture probability. Prior $\pi(\phi) = U(\hat{c}, 2)$ performs satisfactorily only for large $p_{1.}$, without considering the fact that $\phi < 1$ and in that situation, $\pi(N) = \text{Poi}(\hat{N}_{M_b})$ works better than Jeffrey's.

Table 6.2: Summary results reflecting the performances of the proposed approach **AB-Flat** and other relevant methods applied to the populations P5-P8. Upper panel refers to the situation when the knowledge, $\phi < 1$, is available. Lower panel refers when nothing is available on ϕ.

Population ($E(x_0)$)	Method: $\pi(\phi)$	Prior $\pi(N)$	Average Estimate	Sample SE	Sample RMSE	95%CI
		when it is known that $\phi < 1$				
P5(430)	Lee: U(0.2,1.4)[a]	Jeffrey	456	18.80	47.88	(432, 505)
	SEMWiG: U($\hat{c}$, 1)	Jeffrey	478	-	25.27	(455, 507)
	AB-Flat: U($\hat{c}$,1)	Jeffrey	482	12.39	21.62	(461, 508)
		Poi($\hat{N}_{M_b}$)	542	69.93	81.45	(488, 676)
P6(459)	Lee: U(0.2,1.4)[a]	Jeffrey	481	7.40	20.51	(460, 528)
	SEMWiG: U($\hat{c}$, 1)	Jeffrey	492	-	11.74	(477, 510)
	AB-Flat: U($\hat{c}$,1)	Jeffrey	495	8.38	9.98	(480, 510)
		Poi($\hat{N}_{M_b}$)	504	11.38	12.03	(485, 526)
P7(483)	Lee: U(0.2,1.4)[a]	Jeffrey	504	6.90	7.69	(484, 535)
	SEMWiG: U($\hat{c}$, 1)	Jeffrey	498	-	5.28	(489, 508)
	AB-Flat: U($\hat{c}$,1)	Jeffrey	500	5.00	5.01	(490, 508)
		Poi($\hat{N}_{M_b}$)	499	4.97	5.07	(489, 508)
P8(446)	Lee: U(0.2,1.4)[a]	Jeffrey	523	20.26	30.73	(466, 589)
	SEMWiG: U($\hat{c}$, 1)	Jeffrey	484	-	19.34	(465, 506)
	AB-Flat: U($\hat{c}$,1)	Jeffrey	483	8.72	19.31	(468, 498)
		Poi($\hat{N}_{M_b}$)	481	8.90	21.08	(466, 497)
		when no directional knowledge on ϕ is available				
P5	SEMWiG: $\pi(\phi\|p) \propto 1$	Jeffrey	474	-	27.75	(454, 497)
	AB-Flat: U($\hat{c}$,2)	Jeffrey	534	21.25	40.33	(497, 581)
		Poi($\hat{N}_{M_b}$)	609	108.33	153.77	(523, 823)
P6	SEMWiG: $\pi(\phi\|p) \propto 1$	Jeffrey	512	-	15.27	(494, 525)
	AB-Flat: U($\hat{c}$,2)	Jeffrey	526	13.56	28.98	(502, 550)
		Poi($\hat{N}_{M_b}$)	529	16.80	33.54	(495, 539)
P7	SEMWiG: $\pi(\phi\|p) \propto 1$	Jeffrey	510	-	12.60	(498, 526)
	AB-Flat: U($\hat{c}$,2)	Jeffrey	513	6.50	14.61	(501, 525)
		Poi($\hat{N}_{M_b}$)	507	5.92	9.29	(497, 518)
P8	SEMWiG: $\pi(\phi\|p) \propto 1$	Jeffrey	487	-	18.30	(466, 517)
	AB-Flat: U($\hat{c}$,2)	Jeffrey	522	20.17	29.77	(494, 562)
		Poi($\hat{N}_{M_b}$)	504	16.51	17.07	(481, 538)

[a] This prior is used as Lee et al.'s strategy fails to generate samples from U(0.2, 1), when $\phi > 1$

This analysis suggests that when directional information on ϕ is unavailable, it is not possible to have reasonably good estimate for N from flat objective prior on ϕ except in case of $x_{1.} > x_{.1}$ under $\phi < 1$. Moreover, subjectiveness makes an estimate prior sensitive which is not at all desirable. This motivates us to formulate a modified approach for unavailable directional knowledge of ϕ in the next section.

6.3 When Directional Knowledge on ϕ is Unavailable

6.3.1 Proposed Methodology (AB-Con)

In earlier section it is observed that when directional information on ϕ is not available, the reasonable choice $[\alpha,\beta]$=$[\hat{c},2]$ generally outperforms other methods for human population, if one uses constant or unform prior on ϕ. In practice, for human population, demographic or any other beneficiary type survey reflects a recapture prone nature among the individuals. But, if no such information is available, can we modify our earlier proposed strategy (in section 6.2.1) considering suitably minimum subjectiveness in prior selection on ϕ so that reasonably good estimate can be obtained? In this section we try to set a conjugate prior on ϕ and therefore investigate how our Bayes estimates, based on different potential loss functions, perform if hyperparameters are empirically estimated.

Unlike the previous case, we suggest that prior on ϕ is dependent on N and therefore, joint prior distribution becomes $\pi(\Theta) = \pi(p_{1\cdot})\pi(\phi|N)\pi(N)$. We support the argument made by Lee et al. (2003[62]) that, in practice, it is necessary to restrict the range of ϕ to be between some α and β. Let us restrict $\phi \in [\alpha,\beta]$ and consider a conjugate prior on ϕ as $\pi(\phi) =$ GB-I$(a,b,1,\text{rate}=1/\beta)$, for given a, b and β. Hence, $\pi(\phi|\alpha,\beta,a,b) \propto \phi^{a-1}(1-\phi/\beta)^{b-1} \times \mathscr{I}_{[\alpha,\beta]}(\phi)$. Since, $c = \phi p$ and $c < 1$, then p^{-1} might be thought as a good choice for the upper limit of ϕ, hence $\beta = p^{-1}$. In practice, $\hat{c}$ is taken as a good choice for α and p can be obtained using the relation $p = x_{01}/(N - x_{1\cdot})$, suggested by Llyod(1994[65]). This suggestion leads to a conditional posterior of ϕ as a well-known probability density function from which one can directly generate Gibbs samples. The conditional posterior density will be

$$\pi(\phi|N,a,b) \quad \propto \quad \text{GB-I}(x_{11}+a, x_{10}+b, 1, \text{rate}=1/\beta) \times \mathscr{I}_{[\hat{c},\beta]}(\phi), \tag{6.6}$$

where $\beta = p^{-1} = (N - x_{1\cdot})/x_{01}$ and a and b are chosen by equating $E_\pi(\phi|\alpha,\beta,a,b)$ with $c/p = c\beta$ to maintain the inter-relationships among model parameters in prior selection. Since, $E_\pi(\phi|\alpha,\beta,a,b) = \beta a(a+b)^{-1}$ which implies $a(a+b)^{-1} = c$. We choose $a = t(x_{11}/x_0)$ and $q = t(x_{10}/x_0)$ where t (> 0) is a tuning parameter that regulates the variance of the prior density such that $V_\pi(\phi) = O(t^{-1})$. Remaining parameters in Θ, i.e. $p_{1\cdot}$ and N, have same prior setup as mentioned in section 6.2.1. Hence, we can perform a simple Gibbs sampling MCMC technique with conditional posterior of ϕ in (6.6) and other two conditional posterior densities $\pi(N-x_0|\phi,p_{1\cdot})$ and $\pi(p_{1\cdot}|N)$ exactly same as in section 6.2.1. Here also the hyper-parameter λ is replaced by $\hat{N}_{M_b}$ as before for poisson prior. Firstly, we fix initial values $p_{1\cdot}^{(0)}$ and $p^{(0)}$ and the prior variance tuning parameter t. $\phi^{(0)}$ is simulated from conjugate prior GB-I with initial $\beta^{(0)}=1/p^{(0)}$. Then generate $N^{(0)}$ from its posterior $\pi(N-x_0|p_{1\cdot},p)$ replacing $p_{1\cdot}$ and p by $p_{1\cdot}^{(0)}$ and $p^{(0)}$ respectively. Then, subsequent steps in Gibbs sampling is carried out as follows.

Step 1	:	Simulate $p_{1\cdot}^{(1)}$ and $\phi^{(1)}$ from $\pi(p_{1\cdot}|N^{(0)})$ and $\pi(\phi|N^{(0)},a,b)$, in (6.6) respectively, where $\beta=1/p^{(0)}$.
Step 2	:	Obtain $p^{(1)}=\hat{c}/\phi^{(1)}$.
Step 3	:	Generate N from $\pi(N-x_0|p_{1\cdot}^{(1)},\phi^{(1)})$.
Step 4	:	Repeat the above three steps until the convergence is reached.

Hence, the values $\{N^{(h)}: k<h\leq 2k\}$, where k is the chosen burn-in period, are believed to be a very large sample from the resulting posterior distribution $\pi(N|\underline{\mathbf{x}})$. k is chosen based on the performance of $\hat{R}^{1/2}$ as stated earlier. To obtain estimate of true population size (N) from the resultant posterior, we consider some potential loss functions, such as squared error, absolute error and maximum a posteriori (MAP) loss functions which produce estimators respectively as posterior mean ($\hat{N}_{MEAN}$), median ($\hat{N}_{MED}$) and mode ($\hat{N}_{MAP}$). Casella (1986) suggested another estimate obtained by minimizing the squared relative error loss function $L(N,\hat{N})=\left(\frac{N-\hat{N}}{N}\right)^2$, and the corresponding estimate is $\hat{N}_{SRE}=E_{\pi(N|D)}(N^{-1})/E_{\pi(N|D)}(N^{-2})$. In practice, $\hat{N}_{SRE}$ is obtained using the ratio of $\sum 1/N^{(h)}$ and $\sum 1/[N^{(h)}]^2$ from posterior sample $\{N^{(h)}:k<h\leq 2k\}$. One feature of this setup is that though it uses informative prior but the hyper-parameters (a, b, λ) are taken as functions of data for given the variance tuning parameter t. In the next section, we numerically evaluate the performance of aforesaid MCMC algorithm in order to estimate N when directional knowledge on ϕ is absent, equivalently, domain knowledge is $\phi>0$. We call this method as *AB-Con* method since conjugated subjective prior choice is made for ϕ when directional knowledge on ϕ is *unavailable*.

6.3.2 Numerical Illustration

Let us consider all the simulated populations discussed in section 6.2.2 in order to illustrate the AB-Con method and also to suggest efficient priors under different loss functions considered. Prior belief on ϕ is considered with a reasonable value of $t = 20$. Hence, we observe the performance of $\hat{R}^{1/2}$ for all populations and fix a general k at 7000. Resulting posteriors from AB-Con method using Jeffrey's and Poisson priors on N are shown in Figure 6.1 along with posteriors from Lee et al. (2003[62]). Burn-in for Lee's(2003[62]) method is set at 150. Figure 6.1 shows that Lee's method produces tighter posteriors than AB-Con but it is bimodal in nature for almost all cases. Posterior from Jeffrey's prior in AB-Con method is almost similar with poisson prior across populations but it has larger variability than poisson. In some cases, Lee's(2003[62]) estimate is better in terms of squared error or maximum-a-posteriori loss, but for bimodal posteriors the higher mode is not close to the true value (as here, true N is 500).

Final Estimates of N are obtained by averaging over 200 posterior replications for each loss function. S.E. of each estimate is computed over 200 replicated estimates. It is clear from Table 6.3 that poisson and Jeffrey's prior performs better respectively in case of squared error and absolute error loss. Specifically when $x_{1.} < x_{.1}$, Jeffrey's prior performs better for both the loss functions. MAP-based estimator significantly underestimates N for recapture prone populations. Under this loss function, poisson prior is generally better than Jeffrey's. In particular for $x_{1.} > x_{.1}$, both the priors have similar performance. $\hat{N}_{SRE}$ would be effective in general with Jeffrey's prior for N. Indeed, $\phi > 1$ corresponds to the most likely case in human demographic studies. For example, a specialised survey is conducted following a large census count, e.g. Post Enumeration Survey (PES), $x_{1.} < x_{.1}$ is often experienced. For sensitive dual survey, e.g. estimation of drug abused population size, $x_{1.} > x_{.1}$ is usually observed for time ordered samples. Hence, in order to find an estimator when no information on ϕ is available, Poi($\lambda = \hat{N}_{M_b}$) is suggested as a reasonably good selection for $\pi(N)$ if maximum-a-posteriori loss function is the objective of choice. Otherwise, for posterior median and Casella (1986[14]) suggested loss function squared relative error (SRE), Jeffrey's non-informative prior on N is preferred.

6.4 Real Data Application: Malawi Death Data

To illustrate the methods developed in earlier sections of this current chapter, we consider the Malawi Death data (for details, *see* section 1.8.1 of chapter 1) again. Nour (1982[68]) estimated these death sizes as *378* and *3046* for *Lilongwe* and *Other urban areas* respectively

Table 6.3: Summary results for estimating N with associated s.e. in ( ) by the proposed **AB-Con** method from different loss functions and Lee's(2003[62]) method for all simulated populations, when directional knowledge on ϕ is **NOT** available.

Popln.	Prior $\pi(N)$	Lee[a]	$\hat{N}_{MEAN}$	$\hat{N}_{MAP}$	$\hat{N}_{MED}$	$\hat{N}_{SRE}$
P1	Jeffrey	468 (20.56)	483 (13.92)	410 (10.52)	467 (12.90)	467 (12.51)
	P($\hat{N}_{M_b}$)	-	499 (23.34)	419 (13.52)	494 (22.77)	483 (18.67)
P2	Jeffrey	483 (18.45)	493 (12.04)	433 (9.66)	481 (11.13)	483 (10.98)
	P($\hat{N}_{M_b}$)	-	489 (13.12)	438(13.22)	486 (12.87)	482 (12.02)
P3	Jeffrey	485 (6.61)	495 (7.40)	464 (6.45)	488 (6.90)	492 (7.06)
	P($\hat{N}_{M_b}$)	-	490 (6.94)	463 (6.17)	486 (6.81)	488 (6.69)
P4	Jeffrey	471 (8.11)	473 (10.81)	429 (8.61)	461 (9.85)	466 (9.94)
	P($\hat{N}_{M_b}$)	-	463 (9.91)	429 (8.62)	458 (9.69)	458 (9.57)
P5	Jeffrey	474 (20.80)	566 (14.85)	454 (10.33)	533 (12.13)	530 (11.30)
	P($\hat{N}_{M_b}$)	-	660 (18.50)	498(10.12)	646(17.36)	595(9.45)
P6	Jeffrey	512 (15.76)	558 (11.57)	478 (7.63)	544 (9.55)	544 (8.97)
	P($\hat{N}_{M_b}$)	-	570 (19.89)	490 (14.12)	561 (18.96)	553 (15.09)
P7	Jeffrey	516 (6.17)	539 (7.62)	492 (5.45)	526 (6.67)	532 (6.72)
	P($\hat{N}_{M_b}$)	-	528 (6.86)	493 (5.16)	524 (6.55)	525 (6.36)
P8	Jeffrey	517 (13.02)	525 (10.58)	460 (7.36)	505 (9.16)	510 (8.98)
	P($\hat{N}_{M_b}$)	-	507 (9.45)	462 (8.13)	501 (9.35)	501 (8.83)

[a]Prior $\pi(\phi) = U(0.5, 2)$ is chosen by the *trial-and-error* method discussed in Lee et al. (2003[62])

assuming the fact that two data sources are positively correlated (i.e. $\phi > 1$) in a human demographic study. To implement both our methods, 200 parallel chains are generated from different randomly selected starting points for all the competitive Gibbs samplers suggested in AB-Flat, AB-Con and Lee's(2003[62]) methods. Therefore, we compute $\hat{R}^{1/2}$ (with respect to N) to determine the burn-in period k. In Table 6.4, upper and lower panels respectively represent the results corresponding to *Lilongwe* and *Other Urban Area*.

At first, we analyse the data assuming that the two populations are recapture prone as in human demographic study, positive list-dependence occurs often. So, we also use the poisson prior with $\lambda = \hat{N}_{Nour}$ in addition to other two priors - Jeffrey's and Poi($\lambda = \hat{N}_{M_b}$), for method AB-Flat. In the first half of both panels of the Table 6.4, for the prior $U(1, 2)$ on ϕ, first, second and third row correspond to the Jeffrey's, Poi($\lambda = \hat{N}_{M_b}$) and Poi($\lambda = \hat{N}_{Nour}$) priors respectively. We fix burn-in k generally at 500 for Lilongwe and 3000 for Other urban areas (see Figure 6.2) and record the remaining k values in each of 200 chains. Proposed AB-Flat approach with any suitable prior gives very close and efficient results in comparison to Nour's(1982[68]). Moreover, it is also found that in *Lilongwe*, peoples are more keen to capture the death records again than that of survey time compared to *Other urban areas*. Our estimate with the assumption $\phi > 1$ indicates that around 380 deaths occurred in *Lilongwe* and around 3030 deaths occurred in *Other urban areas*. As N becomes large, we also notice that effect of a choice of larger β for $\pi(\phi)$ on the estimate becomes less.

Bottom half of both the panels in Table 6.4 present the results for both the strata when no information is available on the nature of behavioral response effect, i.e. AB-Con method. We fix burn-in k generally at 2000 for *Lilongwe* and 15000 for *Other urban areas* (*see* Figure 6.2). Figure 6.2 also suggests that burn-in period k in Lee's method is 100 and 3000 respectively for the two strata. For poisson prior on N, the hyper-parameter λ is replaced by $\hat{N}_{M_b}$. Prior $U(\hat{c}, 2)$ in AB-Flat method says that the estimated number of deaths in *Lilongwe* is 360 with a 95% credible interval (357, 365) and in *Other urban areas* is around 2856 with a 95% credible interval (2828, 2929), which are both underestimation than respective Nour's estimates. These results disagree with the recapture proneness of the population assumed by Nour (1982[68]). Without considering any directional knowledge on ϕ (i.e. AB-Con method) the posterior mean, median and SRE estimates suggest that number of deaths in *Lilongwe* is around 363 and contradict the assumption made by Nour(1982[68]) that this population is *recapture prone*. For *Other urban areas*, our analysis agrees that this population is *recapture prone* as ϕ is around 1.20 and corresponding estimate of N is nearly 3130, which is greater than Nour's. For both strata, our MAP based Bayes estimates provide lower estimates. Lee(2003) highly overestimates the death size in *Other urban areas* than all the proposed estimates including Nour's.

6.5 Conclusions

In this chapter, we have presented another two empirical Bayes approaches under a general framework for dual-record system (DRS) where behaviour response effect might play a significant role along with time variation effect. Here we suggest efficient Bayesian computation strategies conditionally and unconditionally on the directional knowledge available on ϕ. The first one is formulated with uniform prior on ϕ whereas the second one depends on subjective conjugate prior based on structural relationships among the underlying parameters. Both strategies implement the functional relation between two nonidentifiable parameters through a suitable empirically estimated prior. Some features of the first approach with uniform prior on behaviour effect (ϕ) are: noninformative prior for N and $p_{1.}$ is used and a reasonable range for ϕ is always available with or without the help of the available directional information on ϕ. When $\phi < 1$, specification of the lower bound of ϕ by $\hat{c}$ works successfully. But when $\phi > 1$, our study concludes that estimate with $\pi(\phi) \propto U(1,2)$ and $\pi(N) \propto N^{-1}$ is expected to be superior than Nour's estimate in terms of smaller RMSE and reasonably better CI. Moreover, the upper limit β is not at all influential if the nature of ϕ is correctly known. It is found that estimates from poisson prior with $\lambda = \hat{N}_{M_b}$ are less efficient than Jeffrey's. Hence, we conclude that the first empirical Bayes approach (discussed in section 6.2.1) performs very well based on the information on the possible range of underlying ϕ, when available. In practice, experts can usually judge whether the specified population is either recapture prone (i.e. $\phi > 1$) or averse (i.e. $\phi < 1$) from past studies. If so, our strategy with noninformative prior on ϕ has significant improvement over Nour (1982 [68]) in terms of efficiency.

An alternative Bayes approach (discussed in section 6.3.1) with informative generalised beta prior is also proposed when there is no reliable information available on ϕ. Some features of this empirical type Bayes approach are the following: For $\phi < 1$, MAP-based estimates are very efficient (compared to other loss functions) when the capture probabilities are high. In contrast, the other two estimates, $\hat{N}_{MED}$ and $\hat{N}_{SRE}$, with Jeffrey's prior or $\pi(N) \equiv$ Poi$(\lambda = \hat{N}_{Nour})$ perform relatively better than $\hat{N}_{MAP}$ for $\phi > 1$. When directional information is not available, $\hat{N}_{MAP}$ obtained from $\pi(N) \equiv \mathrm{P}(\lambda = \hat{N}_{M_b})$ would be a unique choice among the two approaches. It is also found that the second approach improves over the performance of the first approach for known information of $\phi > 1$. For recapture averse population ($\phi < 1$), the first approach is little better because of its relatively tighter prior domain. Hence, our proposed methods can be used to have a better and easily computable estimate of population size from this complex dual system. Apart from the computational advantage, these methods are transparent and relatively easy to explain to the practitioner. Though our methods incorporate subjective elements through the choice of priors, as necessary, but this subjectiveness helps the underlying model to successfully get rid of the identifiability problem.

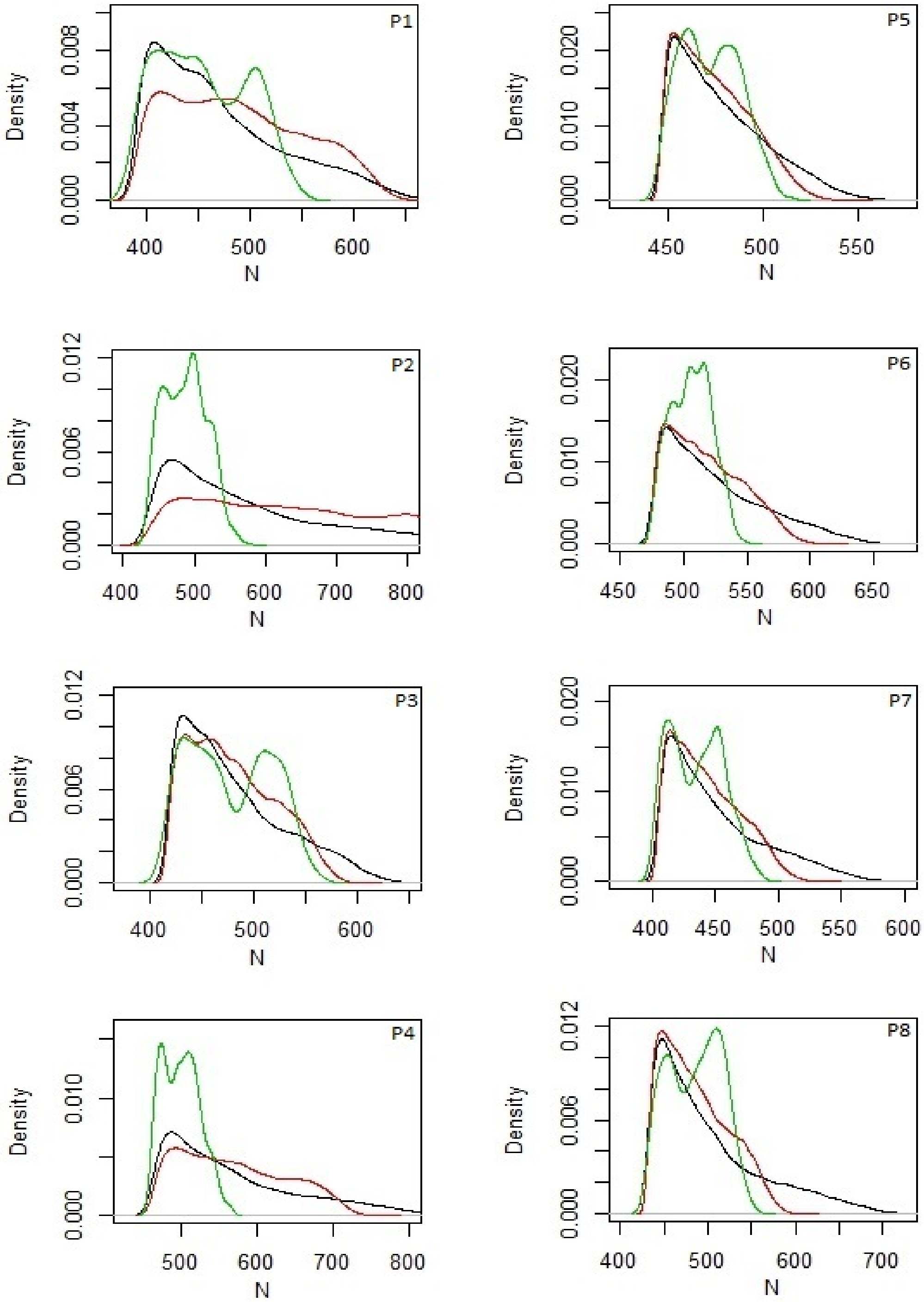


Figure 6.1: Posterior distributions of N based on AB-Con method (*Black* and *Red* lines respectively for Jeffrey's and Poi($\hat{N}_{M_b}$) priors for N) and Lee's method (*Green* line).

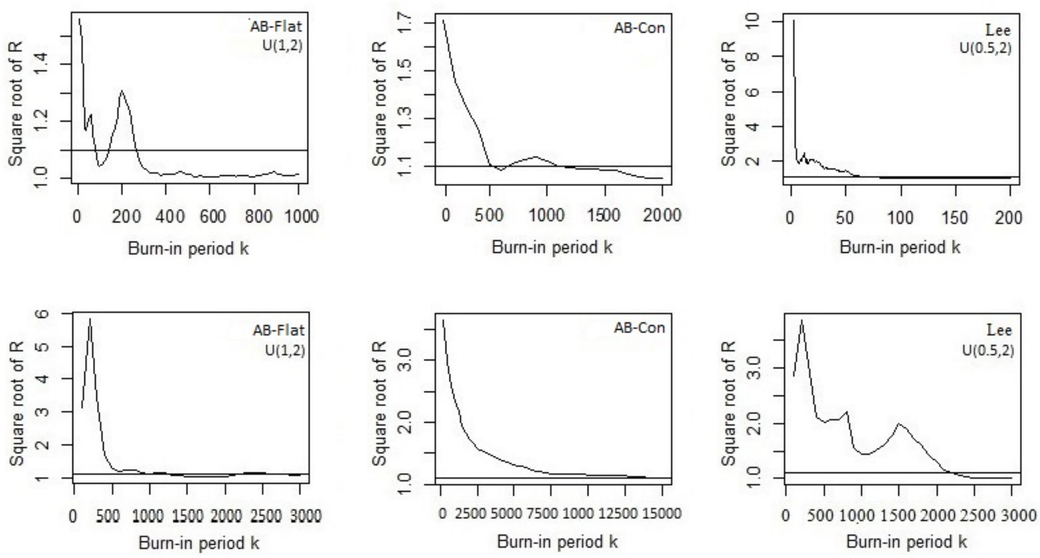


Figure 6.2: Plot of $\hat{R}^{1/2}$ against burn-in period k for MCMC with Jeffrey's prior for N in each of AB-Flat, AB-Con and Lee's methods. First and second rows are for *Lilongwe* and *Other urban areas* respectively. Horizontal line presents the threshold value 1.1 for $\hat{R}^{1/2}$.

Table 6.4: Bayesian estimates of total number of deaths using AB-Flat, AB-Con and Lee's methods. s.e. is computed based on sample posterior distribution and the 95% posterior credible intervals for N and ϕ is determined based on percentile method.

Method	$\pi(\phi)$	$\pi(N)$	$\hat{N}$	s.e.($\hat{N}$)	95% CI of N	$\hat{\phi}$	95% CI of ϕ
			Lilongwe				
Consider $\phi > 1$							
AB-Flat	U(1, 2)	Jeffrey	380	3.91	(373, 388)	1.35	(1.19, 1.54)
		P($\hat{N}_{M_b}$)	374	2.57	(370, 380)	1.35	(1.20, 1.53)
		P($\hat{N}_{Nour}$)	379	3.12	(373, 385)	1.35	(1.20, 1.54)
Consider $\phi > 0$[a]							
AB-Flat	U($\hat{c}$, 2)	Jeffrey	362	5.25	(355, 374)	0.94	(0.75, 1.22)
		P($\hat{N}_{M_b}$)	361	4.03	(354, 369)	0.94	(0.75, 1.19)
Lee	U(0.5, 2)[b]	Jeffrey	354	5.80	(348, 370)	0.74	(0.57, 1.14)
AB-Con	MEAN	Jeffrey	366	4.90	(357, 375)	1.03	(0.82, 1.24)
		P($\hat{N}_{M_b}$)	363	3.81	(357, 371)	1.03	(0.82, 1.24)
	MED	Jeffrey	362	5.37	(354, 372)	0.95	(0.75, 1.22)
		P($\hat{N}_{M_b}$)	361	4.61	(355, 371)	0.95	(0.75, 1.25)
	MAP	Jeffrey	352	5.25	(348, 366)	1.04	(0.62, 1.87)
		P($\hat{N}_{M_b}$)	353	5.93	(348, 372)	1.04	(0.62, 1.87)
	SRE	Jeffrey	365	4.56	(357, 373)	0.86	(0.75, 1.01)
		P($\hat{N}_{M_b}$)	363	3.65	(357, 370)	0.86	(0.75, 1.04)
			Other urban areas				
Consider $\phi > 1$							
AB-Flat	U(1, 2)	Jeffrey	3030	51	(2973, 3172)	1.12	(1.06, 1.26)
		P($\hat{N}_{M_b}$)	3056	46	(3000, 3180)	1.12	(1.06, 1.22)
		P($\hat{N}_{Nour}$)	3027	40	(2978, 3133)	1.16	(1.06, 1.28)
Consider $\phi > 0$[a]							
AB-Flat	U($\hat{c}$, 2)	Jeffrey	2860	41	(2816, 2967)	0.94	(0.89, 1.05)
		P($\hat{N}_{M_b}$)	2873	43	(2826, 2985)	0.94	(0.89, 1.05)
Lee	U(0.5, 2)[b]	Jeffrey	3455	223	(3096, 3870)	1.55	(1.19, 1.97)
AB-Con	MEAN	Jeffrey	3152	119	(2951, 3381)	1.24	(1.03, 1.49)
		P($\hat{N}_{M_b}$)	3146	101	(2972, 3348)	1.24	(1.04, 1.49)
	MED	Jeffrey	3109	149	(2907, 3441)	1.20	(0.99, 1.54)
		P($\hat{N}_{M_b}$)	3131	135	(2932, 3416)	1.20	(0.99, 1.55)
	MAP	Jeffrey	2953	215	(2774, 3491)	1.29	(0.86, 1.95)
		P($\hat{N}_{M_b}$)	3056	263	(2780, 3580)	1.28	(0.86, 1.94)
	SRE	Jeffrey	3109	109	(2935, 3344)	1.13	(1.00, 1.37)
		P($\hat{N}_{M_b}$)	3113	97	(2956, 3320)	1.14	(0.99, 1.37)

[a] $\phi > 0$ refers natural domain of ϕ, as no directional knowledge is considered
[b] This prior is chosen based on *trial-and-error* method discussed in Lee et al. (2003[62])

7 Classification of the Nature of Behavioural Dependence

7.1 Introduction and Motivation

Assumption of *causal independence* (*A4* in section 1.3.1) in the popular Lincoln-Petersen (or M_t) model may seriously mislead us in many situations for human population, specially when capture probabilities may vary with behavioral response (Chandrasekar and Deming, 1949[17]). Many methodologists and practitioners (*see* El-Khorazaty, 2000[36]; Jarvis et al., 2000[56]) argued that the independence assumption may not be justified in reality. When both the time variation effect (*t*) and behavior response effect (*b*) acts together, model M_{tb} is appropriate for homogeneous population. Moreover, this model can be treated as the most general and relevant statistical model for capture-recapture data under homogeneity. The underlying behavior response effect, ϕ (> 0) classifies a given population as recapture prone or recapture averse when $\phi > 1$ or $\phi < 1$, respectively. However, Otis et al. (1978[71]) addressed a non-identifiability problem related to this model as ϕ is not estimable in M_{tb}. Details on the model M_{tb} is found in section 1.6. Several authors tried to solve the non-identifiability problem for number of capturing occasions (T) strictly more than two (i.e. $T \geq 3$) or three (i.e. $T \geq 4$), which are mainly focused for analysing the wildlife populations. In recent past, some Bayesian approaches (Lee and Chen, 1998[61]; Lee et al., 2003[62]; Wang et al., 2015[100]) has been proposed and that can be applied when $T \geq 2$. But derivation of efficient estimate is hardly possible from this weak model likelihood unless some additional information is available. Particularly, if underlying ϕ for any population is correctly known to be greater than 1 or less than 1, then, uncertainty on the domain of ϕ will be reduced to $(1, \infty)$ or $(c, 1)$, respectively. Hence, any one can expect that inference would be better using that available knowledge. This issue has been addressed in *M6* of section 1.8.2.

Few methodologies are proposed in literature that estimate N better, presuming the directional knowledge on ϕ, i.e. either $\phi > 1$ or < 1. As evident from earlier literature (Nour, 1982[68]) and also from the work presented in chapters 5 and 6, one can say that the availability of knowledge on the nature of behavioral dependency certainly helps to infer N better in M_{tb}-DRS context. Apparently, one may suggest that a given population is *recapture prone* if $\hat{c}$ is very close to 1. On the other hand, if it is very close to 0, then associated population would be *recapture averse* with high probability. But giving idea about the possible direction of ϕ is always a challenging job if $\hat{c}$ is neither close to 1, nor close to 0. As per our knowledge, no strategy has been developed to understand the directional nature of ϕ from the given data alone. Therefore, motivated by the issue *M6* in section 1.8.2, our aim in this chapter is to develop some competing classification strategies for the given population in order to identify the underlying directional nature of ϕ that regulates the *causal dependency* between two sources in DRS.

In the next section we formulate three classification strategies and therefore, in section 7.3 we illustrate our proposed strategies and evaluate their performance through an extensive simulation study. In section 7.4 we apply our proposed strategies in order to classify the underlying dependency nature of the given populations associated with all real datasets (mentioned in section 1.8.1) considered for applications throughout the thesis. Finally, in concluding section, we discuss the implication, advantages and limitation of the proposed classification strategies.

7.2 Proposed Classification Strategies

In this section, we formulate a strategy for classification of the population in terms of the directional nature of ϕ, using observed data only. Let us assume that $p > 1/3$, which implies $3c - \phi > 0$ as $c = p\phi$. The *mle* $\hat{c} = (x_{11}/x_{1\cdot})$ is consistent and efficient estimate of c. Hence, in terms of data we can approximately write that $\phi = \hat{c}p^{-1}$ for sufficiently large N. Therefore,

$$\begin{aligned} \phi = \frac{x_{11}}{x_{1\cdot}}\frac{N - x_{1\cdot}}{x_{01}} &> \frac{x_{11}}{x_{1\cdot}} \\ or, (x_{1\cdot}\phi - x_{11}) &> 0, \end{aligned}$$

since $N - x_{1.} > x_{01}$. Thus, for constructing a classification strategy for ϕ, consider the inequality

$$\frac{(3\hat{c} - \phi)(x_{1.}\phi - x_{11})}{x_{.1}\phi} > k,$$

where k is some nonnegative real number. Hence the above inequality may be expressed as

$$\phi^2 + \phi\,\tfrac{kx_{.1} - 4x_{11}}{x_{1.}} + 3\hat{c}^2 \quad < 0 \tag{7.1}$$

$$or, \qquad (\phi - \phi_0)(\phi - \phi_1) \qquad < 0, \tag{7.2}$$

where ϕ_0 and ϕ_1 are two real roots of the quadratic equation (7.1) when equality holds and satisfy $\phi_0\phi_1 = 3\hat{c}^2$ and $\phi_0 + \phi_1 = (4x_{11} - kx_{10})/x_{1.}$ and $\phi_0 < \phi < \phi_1$. Since, A.M. $\geqslant$ G.M. for the two roots ϕ_0 and ϕ_1, then

$$k \leqslant (4 - 2\sqrt{3})x_{11}/x_{.1},$$

equality holds only when $\phi_0 = \phi_1 = \phi = \sqrt{3}\hat{c}$. In addition, as $k \geq 0$ holds under the assumption that $p > 1/3$, which usually holds for human population, the values of ϕ_0 and ϕ_1 corresponding to this lower bound are $\hat{c}$ and $3\hat{c}$ respectively. Furthermore, the root ϕ_0 is a monotonically increasing function, while ϕ_1 is a monotonically decreasing function, of k. This implies $\hat{c} \leq \phi_0 \leq \sqrt{3}\hat{c}$ and $\sqrt{3}\hat{c} \leq \phi_1 \leq 3\hat{c}$.

Now, for given ϕ_0 and ϕ_1, $\exists$ some $\xi \in (0,1)$ such that (7.2) may be written as $\phi = \xi\phi_1 + (1-\xi)\phi_0$, which implies $\xi = (\phi\phi_0 - \phi_0^2)/(3\hat{c}^2 - \phi_0^2)$ using the relation $\phi_0\phi_1 = 3\hat{c}^2$. Considering ξ as a function of ϕ_0, it is noted that for given ϕ and data, ϕ_0^* is a point at which ξ is maximum, where

$$\phi_0^* = \frac{3\hat{c}^2}{\phi} - \frac{3\hat{c}^2}{\phi}\sqrt{1 - \frac{\phi^2}{3\hat{c}^2}}$$

and that the corresponding value of ξ is ξ^* such that

$$\xi^* = \frac{1}{2}\left(1 - \sqrt{1 - \frac{\phi^2}{3\hat{c}^2}}\right)$$

and $\phi_1^* = 3\hat{c}^2/\phi_0^*$.

Now by definition, the value $\phi_0 = \phi_0^*$ is a lower bound for ϕ, i.e. $\phi > \phi_0^*$ which implies

$\phi < \sqrt{3}\hat{c}$. This is also a necessary condition for both ϕ_0^* and ξ^* to be real-valued.

It is in the nature of the current problem that additional information is required in order to obtain exact inference for ϕ. Making the additional assumption that for a fixed ϕ and a fixed $\hat{c}$, ϕ_0^* has the same range of values given to ϕ_0 under the present structure. Thus,

$$\hat{c} \le \phi_0^* \le \sqrt{3}\hat{c},$$

from which,

$$3\hat{c}/2 \le \phi \le \sqrt{3}\hat{c}. \tag{7.3}$$

Alternatively, the bound on ϕ_1^* (i.e. $\sqrt{3}\hat{c} \le \phi_1^* \le 3\hat{c}$) also leads to (7.3). Thus, we have a more tight bound for possible domain of ϕ under some mild assumptions. The assumption that ϕ_0^* has the same range of values as that assumed for ϕ_0, need to be validated. The inequality $\phi_0^* \le \sqrt{3}\hat{c}$ is always valid since all ϕ_0's, including ϕ_0^*, have $\sqrt{3}\hat{c}$ as the maximum. On the other hand, the inequality $\phi_0^* \ge \hat{c}$ does not necessarily hold for all ϕ_0's.

Now we present three classification rules for inferring about the type of behavioural dependency for a given population.

Rule I. Taking cue from Nour's argument (1982[68]), it is proposed that the lower bound in (7.3), which results from setting $\phi_0^* = \hat{c}$, be taken as a threshold for suggesting the type of behavioral nature. So, if $3\hat{c}/2$ is more than 1, i.e. if $\hat{c} > 0.667$, we say the population is *recapture prone*, otherwise it is *recapture averse*. We find that Nour's technique is rather conservative as it has a tendency towards *recapture aversion*.

Rule II. Admitting the conservativeness of the previous classification rule, a second rule is set at the mid-value of the range of ϕ in (7.3). If mean of the upper and lower limits in (7.3), i.e. $1.616\hat{c}$, is above 1 ($\hat{c} > 0.619$), then we call the population *recapture prone*, otherwise it will be *recapture averse*. This rule increases the chance of inferring a population as *recapture prone* and therefore, reduces the bias in *Rule I*.

Rule III. Here we propose a randomized rule to identify the direction of the underlying behavioural dependency. The probability function $\psi(\hat{c})$ behind the decision about the given population to be *recapture prone* is

$$\psi(\hat{c}) = \begin{cases} 1 & if \quad \frac{3\hat{c}}{2} \geq 1 \\ 0 & if \quad \sqrt{3}\hat{c} \leq 1 \\ \delta & if \quad \frac{1}{\sqrt{3}} < \hat{c} < \frac{2}{3}, \end{cases}$$

where $\delta = \left(\hat{c} - \frac{1}{\sqrt{3}}\right) \Big/ \left(\frac{2}{3} - \frac{1}{\sqrt{3}}\right)$. Thus, if observed $\hat{c} \in \left(\frac{1}{\sqrt{3}}, \frac{2}{3}\right)$, one has to perform a bernoulli experiment with probability of *recapture proneness* is equal to δ, in order to decide whether the given population is *recapture prone* or *averse*.

Numerical illustration for the above three classification rules is presented in the next two sections in terms of simulation and real data analyses.

7.3 Evaluation through Simulation

We consider 16 artificial populations characterized by different values of capture probabilities $(p_{1\cdot}, p_{\cdot 1})$ and the value $N = 500$; for one instance each of the two possible situations of behavioral effect - (i) *recapture prone* represented by $\phi = 1.50$ and (ii) *recapture averse* represented by $\phi = 0.60$. These 16 simulated populations for each of the two behavioral situations encompass all possible combinations that are presented in following Table 7.1. The true value of the parameter c is also presented for each population.

The following Table 7.2 presents the performance evaluation of the developed classification strategies in section 7.2, in terms of *correct classification rate* (CCR) of the underlying directional nature of ϕ. CCR is presented in percentage (%) after computing the number of correct classification out of 5000 replications for each simulated population.

Empirical evaluation of classification *Rule I* (columns 2 & 6 in Table 7.2) shows that this classification strategy works efficiently except for the truly *recapture prone* populations P5-P6, P10 and P15-P16 as well as for the truly *recapture averse* populations A3-A4, A8 and A13. Performance of the second strategy, presented in columns 3 & 7 in Table 7.2 is also efficient, except for a lesser number of situations (P6, P10, P15-P16 and A3-A4, A8, A13), where it fails. Indeed, *Rule II* produces improvements over *Rule I* towards the correct classification in almost all cases. So far it can be observed that both the two classification rules fail for those *recapture prone* populations which represents too small recapture probabilities $(p_{\cdot 1})$. On the other hand, both of them fail for those *recapture averse* populations in which recapture probabilities are large. Lastly, further improvements have been established in columns 4 & 8

Table 7.1: Hypothetical populations considered for simulation study for $N = 500$

Population	ϕ	$p_{1\cdot}$	$p_{\cdot 1}$	c	Population	ϕ	$p_{1\cdot}$	$p_{\cdot 1}$	c
P1	1.50	0.50	0.60	0.72	A1	0.60	0.50	0.60	0.45
P2	1.50	0.60	0.70	0.81	A2	0.60	0.60	0.70	0.55
P3	1.50	0.70	0.80	0.89	A3	0.60	0.70	0.80	0.67
P4	1.50	0.80	0.70	0.75	A4	0.60	0.80	0.70	0.62
P5	1.50	0.70	0.60	0.67	A5	0.60	0.70	0.60	0.50
P6	1.50	0.60	0.50	0.58	A6	0.60	0.60	0.50	0.39
P7	1.50	0.55	0.70	0.823	A7	0.60	0.55	0.70	0.538
P8	1.50	0.65	0.80	0.905	A8	0.60	0.65	0.80	0.649
P9	1.50	0.80	0.65	0.611	A9	0.60	0.80	0.65	0.458
P10	1.50	0.70	0.55	0.696	A10	0.60	0.70	0.55	0.574
P11	1.50	0.50	0.70	0.840	A11	0.60	0.50	0.70	0.525
P12	1.50	0.55	0.75	0.882	A12	0.60	0.55	0.75	0.577
P13	1.50	0.65	0.85	0.996	A13	0.60	0.65	0.85	0.689
P14	1.50	0.85	0.65	0.684	A14	0.60	0.85	0.65	0.590
P15	1.50	0.75	0.55	0.600	A15	0.60	0.75	0.55	0.471
P16	1.50	0.70	0.50	0.556	A16	0.60	0.70	0.50	0.417

of Table 7.2, as *Rule III* increases the rate of correct classification, specially for those situations where *Rules I* and *II* completely failed. These notable betterment helps us to make correct inference on classification for *recapture averse* populations (see results for populations A3-A4, A8 & A13).

7.4 Real Data Illustration

In this section we illustrate the three strategies proposed in section 7.2 through the application to the populations associated with all the real datasets discussed in section 1.8.1 of chapter 1. The following Table 7.3 presents the value of the key statistic $\hat{c} = x_{11}/x_{1\cdot}$ and the classification result (i.e. either recapture proneness or aversion) from all the three classification rules.

Note that Lee et al. (2003[62]) in addition with all the inferential methodologies proposed in Chapters 5 and 6, without any assumption on directional knowledge of ϕ, produce some idea about the possible direction on the behavioral nature of the given populations in real data analyses. Inference on the directional nature of the behavioral dependence drawn in earlier chapters for all the data sets may not match with the conclusions of this present classification strategy chapter. However, in the light of the current findings, it seems quite plausible that actually these data sets, except the population *Ward No. 2* and *Other Urban Areas*, are more

Table 7.2: Evaluation of the classification strategy of directional nature of ϕ in M_{tb}-DRS.

Population	Rule I CCR	Rule II CCR	Rule III CCR	Population	Rule I CCR	Rule II CCR	Rule III CCR
P1	96.80	100.00	99.56	A1	100.00	100.00	100.00
P2	100.00	100.00	100.00	A2	100.00	98.94	96.24
P3	100.00	100.00	100.00	A3	0.00	0.00	96.24
P4	99.98	100.00	100.00	A4	0.02	0.00	96.12
P5	49.44	97.26	98.50	A5	100.00	100.00	99.98
P6	00.02	7.12	12.74	A6	99.98	100.00	100.00
P7	100.00	100.00	100.00	A7	100.00	99.52	98.43
P8	100.00	100.00	100.00	A8	0.00	0.00	98.47
P9	89.40	99.92	98.78	A9	99.98	96.20	90.78
P10	1.46	38.14	39.42	A10	100.00	100.00	100.00
P11	100.00	100.00	100.00	A11	100.00	99.84	99.20
P12	100.00	100.00	100.00	A12	99.78	92.10	86.36
P13	100.00	100.00	100.00	A13	0.00	0.00	86.35
P14	78.08	99.72	96.64	A14	99.96	87.26	80.44
P15	0.32	23.30	28.34	A15	100.00	100.00	100.00
P16	0.00	0.74	3.66	A16	100.00	100.00	100.00

appropriately classified as *recapture-averse*. These other two populations have high chance to be *recapture-prone* for their relatively large value of $\hat{c}$. Inference for the population *Ward No. 2* drawn from *Rule I* is found to be opposite than that from *Rule II* and *III*.

7.5 Conclusion

From the extensive literature on capture-recapture data analysis on human population, it is quite clear that list-independence or assumption of *causal independence* does not hold satisfactorily in many instances. As far as homogeneous human population size estimation is concerned, two-sample capture-recapture experiment is appropriate along with M_{tb} modelling. But this model seriously suffers from the non-identifiability problem and analyses in previous chapters suggest that the availability of the knowledge on nature of behavioral dependency could improve the inference to a great extent. Eliciting such information is crucial, but such methodologies are absent in the literature. To address this gap, we develop three comparable strategies for classification of the given population (i.e. whether it is *recapture prone* or *averse*) under some mild and realistic assumptions in the context of human population. All the three classification rules are derived based on different threshold value for $\hat{c}$. The second (*Rule II*) and third (*Rule III*) classification strategies are quite appealing in order

Table 7.3: Application of proposed classification strategies for ϕ to four real datasets

Data	Population	$\hat{c}$	Nature Classified by Rule I	Rule II	Rule III
Malawi Death	Lilongwe	0.593	Averse	Averse	Averse*
	Other Urban Areas	0.839	Prone	Prone	Prone
Injection Drug User	Greater Victoria**	0.075	Averse	Averse	Averse
Children Injury	Cyclists	0.254	Averse	Averse	Averse
	Passengers	0.402	Averse	Averse	Averse
	Pedestrians	0.592	Averse	Averse	Averse*
Handloom	Ward No. 2	0.657	Averse	Prone	Prone*
	Ward No. 16	0.382	Averse	Averse	Averse

*decision taken by randomization
**in British Columbia, Canada

to develop more efficient inference in the context of M_{tb}-DRS. Moreover, third strategy is more accurate than second one and it produces nearly 100% success rate except for particular situations with too small recapture probabilities.

A limitation of the proposed classification strategy in order to identify the directional nature of ϕ is that one should have *time-ordered DRS*, that means List 1 is prepared before List 2 or vice versa. Details on the issue of *time-ordered DRS* can be found in chapter 8. Modifications by relaxing the assumptions and extension of this behavioral classification method may be possible for higher order capture occasions (i.e. when $T \geq 3$).

8 Overall Discussion & Future Works

8.1 Overview

Estimation of size of a given population is an important statistical issue which has a vast application in the field of Government statistics, demography and epidemiology. In practice, it is mostly impossible to count all the individuals in the population accurately by a census, specially when population is large enough and/or very hard to reach the individuals. As a remedy, more than one attempt is carried out independently, near to the census operation, and the population size (say, N) is estimated by matching the available lists (two or more) of information. This kind of data structure by matching lists is known as Multiple-record system and this is equivalent to the capture-recapture system, popularly relevant in biological or epidemiological studies.

In the context of closed human population, more than two sources of information is hardly found. When two attempts has been made to obtain N in capture-recapture format, then such data structure is known as Dual-record System (DRS). In section 1.3, details of DRS is described with all possible underlying assumptions. We have remained particular in the use of DRS for this whole project. Moreover, we have confined ourselves to the analysis of homogeneous human population. Different combinations of assumptions will lead to different models. Possible heterogeneity among individuals in the population can be factored out with the proposal given by Chandrasekar and Deming (1949[17]), which is widely implemented in practice. After forming several mutually exclusive post-strata following Chandrasekar and Deming (1949[17]), which are within homogeneous but between heterogeneous, statistical models relevant for homogeneous population (discussed in sections 1.4-1.6) can be analysed for each of those post-strata. Model M_t has received much attention in practice, greatly because of its model simplicity. But it is often misleading as the assumption of *causal independence* does not work in many instances. Dependence can be caused by the presence of variation in behavioral response at the time of preparing List 2. This feature is incorporated in model M_{tb} and that makes the model most general for this population. Relevancy as well as associated parameter non-identifiability problem for model M_{tb} motivated us to consider the inferential analysis of the model as a prime goal of this dissertation work. Extent of inaccuracy in the estimate $\hat{N}_{ind}$ as well as its robustness against possible departures from the basic *causal independence* assumption have also been studied.

Extent of coverage error in census can be obtained by estimating the omission rate which is equivalent to the problem of estimating size of the given population. In chapter 2, we briefly discuss the statistical methodology to estimate omission rate in Indian census. Extent of correlation bias (pointed out in section 1.8.2) and its consequences in estimating omission rate (equivalently, N) is illustrated. A new potential source for bias in the estimate is identified and demonstrated through examples. A flexible approach has been formulated so that it can achieve minimum variance under a controlled bias limits. Our proposed estimator can efficiently overcome the potential bias by achieving the desired degree of accuracy (almost unbiased) with relatively higher efficiency. Overall improvements in the results are explored through simulation study on different populations.

For model M_{tb}, available and proposed methods are mostly developed in Bayesian paradigm due to the non-identifiability of the model M_{tb} under DRS. Chapter 3 investigates the usage of profile likelihood, explicitly for both the models M_t and M_{tb}. Therefore, an adjustment over profile likelihood is proposed for model M_{tb}. The proposed method is evaluated in terms of performance and compared with available Bayes estimate and $\hat{N}_t$ through extensive simulation study. We also analyse the effect of possible model mis-specification, due to the use of model M_t, in terms of efficiency and robustness. Finally two real life examples with different characteristics are presented for illustration.

In chapter 4, an improved integrated likelihood has been formulated for model M_t based on a suitably constructed weight function using non-informative priors only. A comparative ordering is established among several likelihood and pseudo-likelihood based estimates from M_t. The resulting likelihood has several desirable properties. We have also implemented the same integrated likelihood approach to our interest model M_{tb} but informative prior on ϕ has been used depending upon the availability of the directional behavioral knowledge. Simulation studies have been carried out to explore the performance of the proposed method for both the models separately. Empirical results along with real data analysis demonstrating efficiency and usefulness have been reported.

As mentioned in the literature, parameter identifiability problem in model M_{tb} has been encountered with Bayesian treatments. Indeed, Bayesian inference might have the potential to conquer this problem simply by using suitable priors. In chapter 5, some problems in full Bayes method particularly in DRS, with flat non-informative prior by Lee et al. (2003[62]), are addressed. Considering N estimation in DRS as a missing data problem, two empirical Bayes approaches are proposed along with a reformulation of an existing Bayes treatment given by Lee et al. (2003[62]). Some features and associated posterior convergence for these methods have been discussed. Extensive simulation study established that our proposed approaches are comparably favourable to the existing Bayes approach. A real-data example has been given to illustrate the methods.

In chapter 6, we have proposed two other empirical Bayes approaches for the same problem but in different way where estimates have been obtained using very simple Gibbs sampling strategies through functional relation among parameters characterizing the model of interest M_{tb}. Two approaches correspond to two situations viz. when direction of behavioral dependence is available and when it is not. Simulation study has been carried out to evaluate their performance and compare with existing full Bayes approaches by Lee et al. (2003[62]) and the method proposed in chapter 5. Illustration for both the situations was given through a real data application.

Taking consideration of the issue $M6$ addressed in section 1.8.2 and the evaluation studies of the proposed methods in Chapters 5 and 6, we propose three strategies in Chapter 7 to identify the directional nature of underlying behavioral dependency of individuals. This classification strategy has been found appealing to improve the inference from this complex model. Simulation studies and application on all the real life data sets (thought to be fit for model M_{tb}) are carried out to explore the performance of this strategy.

Finally, in this current chapter, we give direction of some future research plans originating from this thesis dissertation. Now, we address some important interesting issues that either have been evolved from the previous chapters or remain interesting in the context of population size estimation based on homogeneous DRS.

8.2 Future Directions

8.2.1 Dependent Dual-record Systems

From several instances of census counting, it is usually observed that two or more mutually exclusive subgroups of the population, though residing in the same neighbourhood, may be correlated in terms of the probability of inclusion in census or any other counting attempts. Consider the following examples,

Example 1. Consider a locality where there are some houses, inhabited by either (*i*) *owners* or (*ii*) *rent payers*, along with some slum areas where we have, principally, a (*iii*) *squatter population.* If we conduct a post enumeration survey (PES) following a census operation here, the capture probabilities in both the systems for each of the sub-populations will be correlated, due to proximity and surveyor dependent factors. But there will also be some difference due to the status as mentioned. We believe that this situation can be properly modelled using a multivariate probability structure.

Example 2. In the data set named as *Handloom Data* described in section 1.8.1, the aim was to count all the persons attached with the Handloom industry in Gangarampur Town, South Dinajpur District, West Bengal, India. Here, two subgroups (*i*) *Loom-owner* and (*ii*) *Loom-worker* are supposed to be dependent due to same place of work, attachment to same business and also due to surveyor dependent factors but simultaneously, they are somehow different in terms of capture probabilities due to the different economic status, place of residence, etc.

Thus, the dependency between these two or more dual-record systems through suitable model is an important problem from statistical point of view. Estimating size of the population consisting of two or more such dependent subgroups is not a problem that has been tried by researchers. If the present underlying dependency is understood correctly and modelled properly, then resulting estimates for N would be more efficient. We will take up this interesting problem as one of our important future works.

8.2.2 Simpson's Paradox

In many countries like India, at first DRS is constructed for k-th small administrative unit (say, Enumeration Block) under h-th post-stratum (say, Rural-Female population) as the following 2×2 table (equivalent to the inner four cells of Table 1.1)

x_{11kh}	x_{10kh}
x_{01kh}	$\hat{x}_{00kh}$

where $\hat{x}_{00kh}$ is obtained from the relation (8.1) derived from the *causal independence* assumption *A4* in DRS

$$x_{11kh}.\hat{x}_{00kh} = x_{10kh}.x_{01kh} \forall k \& h. \tag{8.1}$$

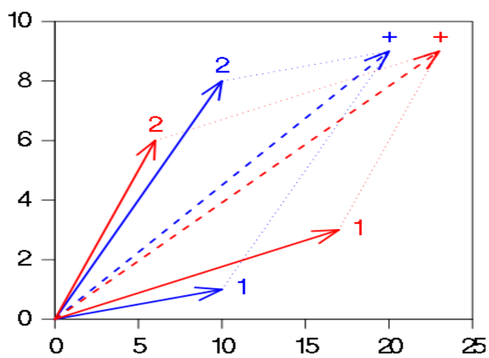
Therefore, the estimate for h-th post-stratum is obtained by adding the entries of 2×2 tables over all administrative units. The issue is that for the added entries over all administrative units, *causal independence* does not hold any more. This incident is called *Simpson's Paradox* and implication of this result is that if we apply *causal independence* at small administrative unit level, it does not work at the next post-stratum level. However, results are published at post-stratum levels.

Simpson's Paradox (Mittal, 1991[67]) occurs when and only when

$$\begin{aligned} x_{11h}.x_{00h} &= x_{10h}.x_{01h} \forall h \\ \text{or, } \theta_h = \frac{x_{11h}.x_{00h}}{x_{10h}.x_{01h}} &= 1 \ \forall h \qquad (8.2) \\ \text{But } \left(\sum_h x_{11h}\right)\left(\sum_h x_{00h}\right) &\neq \left(\sum_h x_{10h}\right)\left(\sum_h x_{01h}\right). \qquad (8.3) \end{aligned}$$

In Figure 8.1, Simpson's paradox says that even if a vector $\overrightarrow{b_1}$ (in blue in the figure) has a

Figure 8.1: Geometric View of Simpson's Paradox



smaller slope than another vector $\overrightarrow{r_1}$ (in red), and $\overrightarrow{b_2}$ has a smaller slope than $\overrightarrow{r_2}$, the sum of the two vectors $\overrightarrow{b_1} + \overrightarrow{b_2}$ (indicated by "+" in the figure) can still have a larger slope than the sum of the two vectors $\overrightarrow{r_1} + \overrightarrow{r_2}$.

This paradox warns us that if we stratify our population into several subpopulations and estimate the 4th unknown cell assuming *causal independence* for each subpopulation and then combine all the subpopulations to get result for overall population, the independence assumption may be violated. A foremost aim is to find out the necessary and sufficient condition for this paradox to occur in the context of DRS. Therefrom, try to suggest a remedy or give a thoughtful discussion to tackle this issue in practice.

8.2.3 Unordered Captures

Unordered captures means when the two capture attempts are not exercised one after another or in sequential manner. In Indian Sample Registration System (SRS), one list is made from the continuous recording system, called Civil Registration System and another list is made from the retrospective survey carried out every six months. This kind of DRS is clearly constructed from two *unordered capture attempts* and its implication lies in the usage of dataset. User can assume any of these two data collection systems as List 1. Therefore, natural interest is that whether any useable estimator is affected by the listing order. Technically, an estimator $T(x_{1.}, x_{.1}, x_{11})$ based on $(x_{1.}, x_{.1}, x_{11})$ should not differ if $x_{1.}$ (List 1 count) and $x_{.1}$ (List 1 count) interchange their positions in T. If we review the estimators $\hat{N}_{ind} = (x_{1.}x_{.1})/x_{11}$ and $\hat{N}_{SRS} = x_0$, both are unaffected by the listing order. But Nour's estimator (*see* last paragraph of section 1.6.1) does not. Thus, if the DRS data in hand is collected from two time-unordered attempts or systems, then estimators should be used with a caution. In wildlife application, this issue is not relevant as counting attempts are usually time-ordered; as in case of some instances for human population, like undercount estimation in census etc., dual listing is also made on the basis of time-ordered attempts.

As mentioned in section 7.5, classification problem of the directional nature of behavioral dependency in DRS would be more challenging if the capture instances are *time-unordered*. Therefore, we will try to develop an efficient strategy for this problem and that will be worthwhile for many real applications.

Bibliography

1. Ayhan, H. Ö. (2000). Estimators of vital events in dual-record systems. *Journal of Applied Statistics* **27**, 157-169.

2. Ayhan, H. Ö. and Ekni, S. (2003). Coverage error in population censuses: The case of Turkey. *Survey Methodology* **29**, 155-165.

3. Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343-365.

4. Barndorff-Nielsen, O. E. and Cox, D. R. (1994). Inference and Asymptotics. Chapman and Hall, London.

5. Basu, D. (1977). On the elimination of nuisance parameters. *Journal of American Statistical Association* **72**, 355-366.

6. Bell, W. R. (1993). Using Information from Demographic Analysis in Post-Enumeration Survey Estimation. *Journal of American Statistical Association* **88**, 1106-1118.

7. Bell, W. (2001), *Accuracy and Coverage Evaluation: Correlation Bias.* DSSD Census 2000 Procedures and Operations Memorandum Series B-12, U.S. Bureau of the Census.

8. Bishop, Y., Fienberg, S. and Holland, P.(1975). *Discrete Multivariate Analysis, Theory and Practice.* Cambridge, Massachusetts: MIT Press.

9. Böhning, D. and Heijden, P. V. D. (2009). Recent developments in life and social science applications of capture–recapture methods. *Advances in Statistical Analysis* **93**, 1-3.

10. Bolfarine, H., Leite, J.G. and Rodrigues, J. (1992). On the estimation of the size of a finite and closed population. *Biometrical Journal* **34**, 577-593.

11. Boonstra, P. S., Mukherjee, B. and Taylor, J. M. G. (2013). Bayesian Shrinkage Methods for Partially observed Data with Many Predictors. *The Annals of Applied Statistics* **7**, 2272-2292.

12. Breiman, L. (1994). The 1991 Census Adjustment: Undercount or bad data. *Statistical Science* **9**, 458-475.

13. Brittain, S. and Böhning, D. (2009). Estimators in capture-recapture studies with two sources. *Advances in Statistical Analysis* **93**, 23-47.

14. Casella, G. (1986). Stabilizing binomial n estimators. *Journal of American Statistical Association* **81**, 172-175.

15. Castledine, B. J. (1981). A Bayesian Analysis of Multiple Recapture. *Journal of American Statistical Association* **81**, 338-346.

16. Celeux, G. and Diebolt, J. (1985). The SEM Algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quaterly* **2**, 73-82.

17. ChandraSekar, C. and Deming, W.E. (1949). On a method of estimating birth and death rates and the extent of registration. *Journal of American Statistical Association* **44**, 101-115.

18. Chandrasekaran, C. and Deming, W.E. (1981). On the correlation bias in the application of Chandra-Deming method for estimating vital events. *Cairo Demographic Centre*, Working Paper No. 2, Cairo, Egypt. 31 pp.

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

20. Chao, A. (1989). Estimating population size for sparse data in capture–recapture experiments. *Biometrics* **45**, 427–438.

21. Chao, A., Chu, W. and Chiu, H. H. (2000). Capture-Recapture when Time and Behavioral Response Affect Capture Probabilities. *Biometrics* **56**, 427-433.

22. Chao, A. (2001a). An Overview of Closed Capture–Recapture Models. *Journal of Agricultural, Biological, and Environmental Statistics* **6**, 158–175.

23. Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. and Chao, D. Y. (2001b). Tutorial in Biostatistics: The Application of Capture-Recapture Models to Epidemiological Data. *Statistics in Medicine* **20**, 3123-3157.

24. Chapman, D. G. (1951). Some properties of the hypergeometric distribution with applications to zoological censuses. *University of California Publication in Statistics* **1**, 131–160.

25. Chatterjee, K. and Mukherjee, D. (2016a). An Improved integrated likelihood population size estimation in dual record system. *Statistics and Probability Letters* **110**, 146-154.

26. Chatterjee, K. and Mukherjee, D. (2016b). An Improved Estimator of Omission Rate for Census Count: With Particular Reference to India. *Communications in Statistics - Theory and Methods* **45**, 1047-1062.

27. Chatterjee, K. and Mukherjee, D. (2016c). On the Estimation of Homogeneous Population Size from a Complex Dual-record System. *Journal of Statistical Computation and Simulation* **86**, 3562–3581.

28. Cowan, C. and Bettin, P. (1982). *Estimates and Missing Data Problems in the Post Enumeration Program.* Unpublishd Technical Report, U.S. Bureau of the Census.

29. Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269-276

30. Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussions). *Journal of Royal Statistical Society B* **49**, 1-39.

31. Cressie, N. (1989). Empirical Bayes estimation of undercount in the decennial census. *Journal of American Statistical Association* **84**, 1033-1044.

32. Darroch, J. N. (1958). The multiple recapture census, I. Estimation of closed population. *Biometrika* **48**, 343-359.

33. Dreyfus, I., Chouquet, C., Ezzedine, K., Henner, S., Chiavérini, C., Maza, A., Pascal, S., Rodriguez, L., Vabres, P., Martin, L., Mallet, S., Barbarot, S., Dupuis, J. and Mazereeuw-Hautier, J. (2014). Prevalence of inherited ichthyosis in France: A study using capture-recapture method. *Orphanet Journal of Rare Diseases* **9**, 1-7.

34. Elliot, M. R. and Little, R. J. A. (2005). A Bayesian Approach to 2000 Census Evaluation Using ACE Survey Data and Demographic Analysis. *Journal of American Statistical Association* **100**, 380-388.

35. El-Khorazaty, M. N., Imrey, P. B., Koch, G. G. and Wells, H. B.(1976). *A Review of Methodological Strategies for Estimating the Total Number of Events with Data from Multiple-record Systems.* Technical Report. Institute of Statistics Mimeo Series No. 1095, November 1976.

36. El-Khorazaty, M. N. (2000). Dependent dual-record system estimation of number of events: a capture-mark-recapture strategy. *Environmetrics* **11**, 435-448.

37. Elliot, M. R. and Little, R. J. A. (2000). A Bayesian Approach to Combining Information from a Census, a Coverage Measurement Survey and Demographic Analysis. *Journal of American Statistical Association* **95**, 351-362.

38. Farcomeni, A. and Tardella, L. (2012). Identifiability and inferential issues in capture-recapture experiments with heterogeneous detection probabilities. *Electronic Journal of Statistics* **6**, 2602-2626.

39. Freedman, D. A. and Wachter, K. A. (1994). Heterogeneity and Census Adjustment for the Intercensal Base. *Statistical Science* **9**, 476-485.

40. Gart, J. J. and Zweifel, J. R. (1967). On the Bias of Various Estimators of the Logit and its Variance With Application to Quantal Bioassay. *Biometrika* **54**, 181-187.

41. Gelman, A. (1996). *Inference and Monitoring Convergence.* In *Markov Chain Monte Carlo in practice* (Edited by W. R. Gilks, S. Richardson and D. J. Spiegelhalter). Chapman and Hall, London.

42. George, E. I. and Robert, C. P. (1992). Capture-recapture estimation via Gibbs sampling. *Biometrika* **79**, 677-683.

43. Gilks, W. R., Richardson, S. and Speigelhalter, D. J. (1996). *Markov Chain Monte Carlo in practice.* First Edition, Chapman & Hall, London.

44. Gosky, R. and Ghosh, S. K. (2011). A Comparative Study of Bayes Estimators of Closed Population Size from Capture-Recapture Data. *Journal of Statistical Theory and Practice* **5**, 241-260.

45. Greenfield, C. C. (1975). On the estimation of a missing cell in a 2×2 contingency table. *Journal of Royal Statistical Society A* **138**, 51-61.

46. Greenfield, C. C. and Tam, S. M. (1976). A simple approximation for the upper limit to the value of a missing cell in 2×2 contingency table. *Journal of Royal Statistical Society A* **139**, 96-103.

47. Griffin, R. A. (2008). *Correlation Bias Adjustment by Individual Year of Age*, DSSD 2010 Census Coverage Measurement Memorandum Series #2010-E-19, U.S. Bureau of the Census.

48. Haldane, J. B. S. (1955). The estimation and significance of the logarithm of a ratio frequencies. *Annals of Human Genetics* **20**, 309-311.

49. Hogan, H. (2003). The Accuracy and Coverage Evaluation: Theory and Design. *Survey Methodology* **29**, 129-138.

50. Holzmann, H., Munk, A. and Zucchini, W. (2006). On Identifiability in Capture–Recapture Models. *Biometrics* **62**, 934-939.

51. Hook, E. B. and Regal, R. R. (1995). Capture-recapture methods in epidemiology: Methods and limitations. *Epidemiologic Review* **17**, 243-264.

52. Huggins, R. (1989). On the statistical analysis of capture-recapture experiments. *Biometrika* **76**, 133-140.

53. Iñigo, J., Arce, A., Martín-Moreno, JM., Herruzo, R., Palenque, E. and Chaves, F., (2003). Recent transmission of tuberculosis in Madrid: application of capture–recapture analysis to conventional and molecular epidemiology. *International Journal of Epidemiology* **32**, 763-769.

54. Isaki, C. T. and Schultz, L. K. (1987). The effect of correlation and matching error in dual system estimation. *Communications in Statistics: Theory and Methods* **16**, 2405-2427.

55. Jabine, T. B. and Bershad, M. A. (1968). Some comments on the Chandrasekaran-Deming technique for the measurement of population change. *Proceedings of the CENTO Symposium on demographic statistics*, 189-206.

56. Jarvis, S. N., Lowe, P. J., Avery, A., Levene, S. and Cormack, R. M. (2000). Children are not goldfish - mark/recapture techniques and their application to injury data. *Injury Prevention* **6**, 46-50.

57. Jewell, N. P. (1986). On the Bias of Commonly Used Measures of Association for 2×2 Tables. *Biometrics* **42**, 351-358.

58. Jibasen, D., Yahya, W. B. and Jolayemi, E. T. (2012). Capture-Recapture Estimation for Elusive Events with Two Lists. *Mathematical Theory and Modeling* **2**, 1-9.

59. Krotki, K. J.(ed.) (1978). *Developments in Dual System Estimation of Population Size and Growth.* The University of Alberta Press, Alberta, Canada.

60. Laplace, P. S. (1786). *Sur les Naissances, les Mariages et les Morts.* Histoire de L'Academie Royale des Sciences, 1783. Paris.

61. Lee, S. M. and Chen, C. W. S. (1998). Bayesian inference of Population Size for behavioral response models. *Statistica Sinica* **8**, 1233-1247.

62. Lee, S. M., Hwang, W. H. and Huang, L. H. (2003). Bayes estimation of Population Size from Capture-recapture Models with Time Variation and Behavior response. *Statistica Sinica* **13**, 477-494.

63. Lincoln, F. C. (1930). Calculating waterfowl abundance on the basis of banding returns. *USDA Circ.* **118**, 1–4.

64. Link, W. A. (2003). Nonidentifiability of Population Size from Capture-Recapture Data with Heterogeneous Detection Probabilities. *Biometrics* **59**, 1123-1130.

65. Lloyd, C. J. (1994). Efficiency of Martingle Methods in Recapture Studies. *Biometrika* **81**, 305-315.

66. Marks, E. S., Seltzer, W. and Krotki, K. R. (1974). *Population Growth Estimation.* The Population Council, New York.

67. Mittal, Y. (1991). Homogeneity of Subpopulations and Simpson's Paradox. *Journal of American Statistical Association* **86**, 167-172.

68. Nour, E. S. (1982). On the Estimation of the Total Number of Vital Events with Data from Dual-record Collection Systems. *Journal of Royal Statistical Society A* **145**, 106-116.

69. Office of the Registrar General and Census Commissioner, India. (2006). *Report on Post Enumeration Survey*, Census of India 2001.

70. Office of the Registrar General and Census Commissioner, India (2010). Circular No. 20 dated on July 14, 2010, Subject: *Preparation of revised Charge Register for Population Enumeration*, Census of India 2011.

71. Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978). Statistical Inference from Capture Data on Closed Animal Populations. *Wildlife Monographs* **62**, 1-135.

72. Parzen, M., Lipsitz, S., Ibrahim, J. and Klar, N. (2002). An Estimate of the Odds Ratio That Always Exists. *Journal of Computational and Graphical Statistics* **11**, 420–436.

73. Petersen, C. G. J. (1896). The yearly immigration of young plaice into the Limfjord from the German sea. *Rep. Dan. Biol. St.* **6**, 5–84.

74. Pollock, K. H. (2000). Capture-Recapture Models. *Journal of American Statistical Association* **95**, 293-296.

75. Raj, D. (1968). *Sampling Theory*, McGraw-Hill Book Company.

76. Raj, D. (1977). On Estimating the Number of Vital Events in Demographic Surveys. *Journal of American Statistical Association* **72**, 377-381.

77. Rao, A. S. R. S., Chatterjee, K., Bhattacharya, B. N. and Bose, A. (2009). *Post-Enumeration Survey for the Indian Census: A Methodological Perspective*, Project Report, Indian Statistical Institute, Kolkata.

78. Ruche et al. (2013). Capture–Recapture Method for Estimating Annual Incidence of Imported Dengue, France, 2007–2010. *Emerging Infectious Diseases* **19**, 1740-1748.

79. Salasar, L. E. B., Leite, J. G. and Louzada, F. (2014). On the integrated maximum likelihood estimators for a closed population capture–recapture model with unequal capture probabilities. *Statistics*. http://dx.doi.org/10.1080/02331888.2014.960870.

80. Samuel, J. C., Sankhulani, E., Qureshi, J. S., Baloyi, P., Thupi, C., Lee, C. N., Miller, W. C., Cairns and Charles, A. G. (2012). Under-Reporting of Road Traffic Mortality in Developing Countries: Application of a Capture-Recapture Statistical Model to Refine Mortality Estimates. *PLoS ONE* **7**, e31091.

81. Schnabel, Z. E. (1938). The estimation of the total fish population of a lake. *Amer. Math. Mon.* **45**, 348-352.

82. Seber, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters*. Second Edition. Charles W. Griffin, London.

83. Seber, G. A. F. (1986). A Review of Estimating Animal Abundance. *Biometrics* **42**, 267-292.

84. Seltzer, W. and Adlakha, A. (1974). *On the Effect of Errors in the Application of the Chandrasekaran-Deming Techniques*. University of North Carolina, Laboratory for Population Statistics. Reprint Series No. **14**.

85. Sen, P. K. (1985). *Asymptotics in Finite Population Sampling*. Technical Report. Institute of Statistics Mimeo Series No. 1483, March 1985.

86. Severini, T. A. (1998). An approximation to the modified profile likelihood function. *Biometrika* **85**, 403-411

87. Severini, T. A. (2000). *Likelihood Methods in Statistics*. Oxford University Press Inc., New York.

88. Severini, T.A. (2007). Integrated Likelihood Functions for Non-Bayesian Inference. *Biometrika* **94**, 529-542.

89. Shores, R. and Sands, R. (2003). *Correlation Bias Estimation in the Accuracy and Coverage Evaluation Revision II.* Technical Report presented at 2003 Joint Statistical Meetings - Section on Survey Research Methods.

90. Smith, P. J. (1988). Bayesian Analysis for multiple capture-recapture surveys. *Biometrics* **44**, 1177-1189.

91. Smith, P. J. (1991). Bayesian Analysis for a multiple capture-recapture model. *Biometrika* **78**, 399-407.

92. Stark, P. B. (1998). *A Statistician's Perspective on Census Adjustment.* Presented at the Berkeley Breakfast Club, 4 December 1998.

93. Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of American Statistical Association* **82**, 528–550.

94. Tilling, K. (2001). Capture-recapture methods — useful or misleading? *International Journal of Epidemiology* **30**, 12-14.

95. Van der Heijden, P. G. M., Zwane, E. and Hessen, D. (2009). Structurally missing data problems in multiple list capture–recapture data. *Advances in Statistical Analysis* **93**, 5-21.

96. Wachter, K. W. and Freedman, D. A. (2000). The Fifth Cell Correlation Bias in U.S. Census Adjustment. *Evaluation Review* **24**, 191–211.

97. Wachter, K. W. (2008). *The future of census coverage surveys.* Probability and Statistics: Essays in Honor of David A. Freedman. Institute of Mathematical Statistics, Vol. **2**, 234–245.

98. Wang, X., Lim, J. and Stokes, S. L. (2006). Forming post-strata via Bayesian treed capture-recapture models. *Biometrika* **93**, 861-876.

99. Wang, X., He, C. Z. and Sun, D. (2007). Bayesian population estimation for small sample capture-recapture data using noninformative priors. *Journal of Statistical Planning and Inference* **137**, 1099-1118.

100. Wang, X., He, C. Z. and Sun, D. (2015). Bayesian Estimation of Population Size via Capture-Recapture Model with Time Variation and Behavioral Response. *Journal of Ecology* **5**, 1-13.

101. Wei, G. C. G., Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of American Statistical Association* **85**, 699–704.

102. Wolter, K. M. (1986). Some Coverage Error Models for Census Data. *Journal of American Statistical Association* **81**, 338-346.

103. Xu, C., Sun, D. and He, C. (2014). Objective Bayesian analysis for a capture-recapture model. *Annals of the Institute of Statistical Mathematics* **66**, 245–278.

104. Xu, Y., Fyfe, M., Walker, L. and Cowen, L. L. (2014). Estimating the number of injection drug users in greater Victoria, Canada using capture-recapture methods. *Harm Reduction Journal* **11:9**, 1-7.

105. Yang, H. C. and Chao, A. (2005). Modelling Animals' Behavioral Response by Markov Chain Models for Capture-Recapture Experiments. *Biometrics* **61**, 1010-1017.

106. Yang, X. and Pal, N. (2010). Estimation of a population size through capture-mark-recapture method: a comparison of various point and interval estimators. *Journal of Statistical Computation and Simulation* **80**, 335-354.

107. Zaslavsky, A. M. and Wolfgang, G. S. (1993). Triple-System Modelling of Census, Post-Enumeration Survey, and Administrative-List Data. *Journal of Business & Economic Statistics* **11**, 279-288.

108. Zelterman, D. (1988). Robust estimation in truncated discrete distributions with applications to capture–recapture experiments. Journal of Statistical Planning and Inference **18**, 225–237.

Publications

Published

1. Chatterjee, Kiranmoy and Mukherjee, Diganta. (2016a). An Improved Integrated Likelihood Population Size Estimation in Dual-record System. *Statistics and Probability Letters* **110**, 146-154.
2. Chatterjee, Kiranmoy and Mukherjee, Diganta. (2016b). An Improved Estimator of Omission Rate for Census Count: With Particular Reference to India. *Communications in Statistics - Theory and Methods* **45**, 1047-1062.
3. Chatterjee, Kiranmoy and Mukherjee, Diganta. (2016c). On the Estimation of Homogeneous Population Size from a Complex Dual-record System. *Journal of Statistical Computation and Simulation* **86**, 3562–3581.

Unpublished Technical Reports

1. Chatterjee, K. and Mukherjee, D. (2015). *Approximate Bayesian Solution for Estimating Population Size from Dual-record System.* ResearchGate DOI: 10.13140/RG.2.1.3726.7921
2. Chatterjee, K. and Mukherjee, D. (2015). *On the Population Size Estimation from Dual-record System: Profile-Likelihood Approaches.* ResearchGate DOI: 10.13140/RG.2.1.2168.0082
3. Chatterjee, K. and Mukherjee, D. (2015). *An Integrated Likelihood Method for Estimating Population Size in Dependent Dual-record System.* ResearchGate DOI: 10.13140/RG.2.1.4592.8405
4. Chatterjee, K. and Mukherjee, D. (2016). *On the Identification of the Nature of Behavioural Dependence in a Two-sample Capture-Recapture Study.* ResearchGate DOI: 10.13140/RG.2.1.2146.7926

Conference/Workshop Presentations

List of Oral Presentations by the Author related to this Thesis

1. **Ninth International Triennial Calcutta Symposium on Probability and Statistics**, jointly organized by Department of Statistics, University of Calcutta and Calcutta Statistical Association. **Venue:** Department of Statistics, University of Calcutta, Kolkata, India. **Date:** 28th December, 2015. **Title:** *On the Integrated Likelihood Inference for Population Size in Dual-record System.*

2. **Data Dissemination Workshop on Census Data**, jointly organized by Indian Statistical Institute Library, Kolkata and Directorate of Census Operations, West Bengal, India. **Venue:** Indian Statistical Institute, Kolkata, India. **Date:** 10th March, 2015. **Title:** *Estimation of Omission Rate in Census Count.* (**Invited Oral Presentation**)

3. **21st International Conference on Computational Statistics and also hosting the 5th IASC World Conference**, organized by University of Geneva, Switzerland. **Venue:** CICG, Geneva, Switzerland. **Date:** 21 August, 2014. **Title:** *Estimating the homogeneous population size via empirical Bayes method in a complex dual-record system.*

4. **ISBA Regional Meeting and International Workshop/Conference on Bayesian Theory and Applications**, organized by Department of Statistics, Benaras Hindu University, Varanasi, India. **Venue:** Benaras Hindu University, Varanasi, India. **Date:** 8th January, 2013. **Title:** *Census Coverage Error Estimation through Bayesian Approach with Time and Behavioral Response Variation.*

5. **Eighth International Triennial Calcutta Symposium on Probability and Statistics**, jointly organized by Department of Statistics, University of Calcutta and Calcutta Statistical Association. **Venue:** Department of Statistics, University of Calcutta, Kolkata, India. **Date:** 29th December, 2012. **Title:** *On the estimation of omission rate for Indian census count.*

6. **National Conference on Application of Statistics in Industry and Planning**, organized by Department of Statistics, Visva-Bharati in collaboration with Calcutta Statistical Association. **Venue:** Department of Statistics, Visva-Bharati, Shantiniketan, India. **Date:** 26th February, 2012. **Title:** *Evaluation of Coverage Error Estimation in Indian Census through Simulated Hierarchical Population.*