Mixture Model Analysis with Rank-Based Samples

by

Armin Hatefi

A Thesis submitted to
the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Department of Statistics
University of Manitoba
Winnipeg, Manitoba

Abstract

Simple random sampling (SRS) is the most commonly used sampling design in data collection. In many applications (e.g., in fisheries and medical research) quantification of the variable of interest is either time-consuming or expensive but ranking a number of sampling units, without actual measurement on them, can be done relatively easily and at low cost. In these situations, one may use rank-based sampling (RBS) designs to obtain more representative samples from the underlying population and improve the efficiency of the statistical inference. In this thesis, we study the theory and application of the finite mixture models (FMMs) under RBS designs. In Chapter 2, we study the problems of maximum likelihood (ML) estimation and classification in a general class of FMMs under different ranked set sampling (RSS) designs. In Chapter 3, deriving Fisher information (FI) content of different RSS data structures including complete and incomplete RSS data, we show that the FI contained in each variation of the RSS data about different features of FMMs is larger than the FI contained in their SRS counterparts. There are situations where it is difficult to rank all the sampling units in a set with high confidence. Forcing rankers to assign unique ranks to the units (as RSS) can lead to substantial ranking error and consequently to poor statistical inference. We hence focus on the partially rank-ordered set (PROS) sampling design, which is aimed at reducing the ranking error and the burden on rankers by allowing them to declare ties (partially ordered subsets) among the sampled units. Studying the information and uncertainty structures of the PROS data in a general class of distributions,

in Chapter 4, we show the superiority of the PROS design in data analysis over RSS and SRS schemes. In Chapter 5, we also investigate the ML estimation and classification problems of FMMs under the PROS design. Finally, we apply our results to estimate the age structure of a short-lived fish species based on the length frequency data, using SRS, RSS and PROS designs.

Keywords: Finite mixture models; Ranked set sampling; Partial ranking; Latent variables; Expectation-Maximization algorithm; Classification; Fisher information; Entropy; Age structures of Spot fish.

Acknowledgments

First and foremost, I would like to thank my advisor, Dr. Mohammad Jafari Jozani, for all his constant encouragement, guidance and friendship as well as unconditional support throughout these years at the University of Manitoba (UofM) and Winnipeg. His exceptional professional and personal qualities have made these four years of my PhD program a positive and unforgettable experience. To Mohammad, I give me deepest thanks!

Special thanks to Prof. Omer Ozturk (The Ohio State University, OSU) for his invaluable support, productive advise from which I learned many, as well as his kindness and friendliness during my visit at the OSU. Special thanks also go to Prof. Djemel Ziou (Université de Sherbrooke, UdeS) for his unforgettable support and all those stimulating discussions and advice as well as his hospitality and cordiality in my visits at the UdeS and MOIVRE (Modelling Images, Vision and Neural Networks Centre). I gained many valuable skills and knowledge (e.g. in machine learning and image processing) from him. The impact of these potentials in my prospective towards statistical studies is undeniable.

I would like to express my appreciation to Dr. Alexandre Leblanc for his invaluable support, helpfulness as well as all the skills, I gained from him and his courses during my PhD program. I would like to acknowledge him and other members of my PhD committee including Prof. Liqun Wang and Prof. Miroslaw Pawlak (Department of Electrical and Computing engineering, UofM) as well as Prof. Haikady Nagaraja (Division of Biostatistics, OSU), as external referee of the thesis. I am very thankful to Dr. Saumen Mandal for his support and advice regarding my teaching as well as his kindness to me. In addition, I own many thanks to the outstanding Department of Statistics at UofM; from Dr. Brad Johnson (graduate chair and chair of my candidacy examination) to all of my professors, fellow students and department staff.

I have had a good fortune of having a wonderful circle of friends whose companionships and favours helped me throughout these years in Uof M and Winnipeg. Maryam, Elahe, Rojiar, Shabnam, Paymoun, Mohssen and Mohammad, to name a few. To all of my friends mentioned or been in my heart, thanks!

I am eternally grateful to my family for believing in me and for continuously supporting my desire to pursue my education for 22 years. Moreover, I would have never connected with Statistics, if it were not Atefe (my sister and my first teacher in Statistics) to lead me towards the academic education in statistical sciences.

Last but not least, financial support by the University of Manitoba Graduate Fellowship (UMGF), Manitoba Graduate Scholarship (MGS), Gordon P. Osler Scholarship and Faculty of Science Studentship are gratefully acknowledged.

This work is dedicated to

my parents:

Tahere & Aliasghar,

my siblings:

Arefe & Adele & Iman & Ehsan and Atefe,

who have supported me with heart and soul.

Vita

B.Sc. Statistics, September 2004 - August 2008
Ferdowsi University of Mashhad, Iran

M.Sc. Mathematical Statistics, September 2008 - May 2010
Ferdowsi University of Mashhad, Iran

PhD Statistics, September 2010 - August 2014
University of Manitoba, Canada

Publications

- Hatefi, A., Jafari Jozani, M. and Ziou, D. (2014). *Estimation and classification for finite mixture models under ranked set sampling.* Statistica Sinica, **24**, 675–698.
- Hatefi, A. and Jafari Jozani, M. (2013). *Fisher Information in different types of perfect and imperfect ranked set samples from finite mixture models.* Journal of Multivariate Analysis, **119**, 16–31.
- Hatefi, A. and Jafari Jozani, M. (2014). *Information content of partially rank ordered set samples.* Revision Submitted.
- Hatefi, A., Jafari Jozani, M. and Ozturk, O. (2013). *Mixture model analysis of partially rank ordered set samples: estimating the age-groups of fish from length-frequency data.* Under Revision.
- Hatefi, A., Jafari Jozani, M. (2014). *Proportion estimation based on a PROS samples with multiple concomitants in a breast cancer study.* Submitted.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Rank-based sampling designs and finite mixture models are two important statistical tools in data analysis. Rank-based sampling designs, such as ranked set sampling, partially rank-ordered sampling and some of their variations, have attracted the attention of many statisticians and practitioners in the past few years. Finite mixture models as convenient and flexible statistical tools, have been extensively employed in not only mainstream statistical literature but also in various scientific disciplines. Considering these two statistical tools, this thesis studies the problem of finite mixture modelling under rank-based sampling designs. In this chapter, we provide an overview of rank-based sampling designs as well as finite mixture models. To this end, we briefly describe the ranked set sampling design and some of its variations. The partially rank-ordered set sampling design is introduced as a generalization of the ranked set sampling technique. Then, we present finite mixture modelling as a flexible tool in statistical analysis. Finally we give an overview of the thesis.

1.1 Ranked set sampling

Simple random sampling (SRS) is the most commonly-used probability sampling design in data analysis; in many applications, however, the actual measurement of sampling (experimental) units is very difficult to obtain, in the sense that measurements are costly, time-consuming, invasive or even destructive. Despite this difficulty in data collection, a small number of sampling units may be easily ranked through visual inspection, judgment ranking or available auxiliary variables without actual measurements of the variable of interest and this can be done at low cost. Ranked set sampling (RSS), as a powerful and cost-effective sampling design, can be employed as an alternate to SRS in these sampling situations. The RSS design was introduced by McIntyre (1952) to estimate mean pasture yields. He argued that quantification of pasture yield plots is a costly and time-consuming process (e.g., requiring mowing and weighing the hay); however, an experienced person can fairly accurately rank a small number of plots without actual measurement.

To construct a ranked set sample of size nN with replacement, we proceed as follows. First, we take a simple random sample of size n, say $X_{111}, \ldots, X_{11n}$ (called hereafter a *set* of size n), from the population of interest. Using a ranking operator, say $O_r(\cdot)$, we rank the sample as $O_r(X_{111}, \ldots, X_{11n}) = (X_{[1]}, \ldots, X_{[n]})$ from the smallest to the largest. Note that this ranking may be done using any means other than the actual measurement of the variable of interest. RSS selects the item with the smallest rank, denoted $X_{[1]1}$, from this set for full measurement. We then take another independent random sample of size n from the population; perform ranking and choose the unit with the second smallest rank, denoted $X_{[2]1}$, for measurement.

Table 1.1: An example of RSS sample

cycle	set	ranking the units within the sets	Observation
1	1	$\{X_{111}, X_{112}, X_{113}\} \rightarrow \{\boldsymbol{X_{[1]}}, X_{[2]}, X_{[3]}\}$	$X_{[1]1}$
	2	$\{X_{121}, X_{122}, X_{123}\} \rightarrow \{X_{[1]}, \boldsymbol{X_{[2]}}, X_{[3]}\}$	$X_{[2]1}$
	3	$\{X_{131}, X_{132}, X_{133}\} \rightarrow \{X_{[1]}, X_{[3]}, \boldsymbol{X_{[3]}}\}$	$X_{[3]1}$
2	1	$\{X_{211}, X_{212}, X_{213}\} \rightarrow \{\boldsymbol{X_{[1]}}, X_{[2]}, X_{[3]}\}$	$X_{[1]2}$
	2	$\{X_{221}, X_{222}, X_{223}\} \rightarrow \{X_{[1]}, \boldsymbol{X_{[2]}}, X_{[3]}\}$	$X_{[2]2}$
	3	$\{X_{231}, X_{232}, X_{233}\} \rightarrow \{X_{[1]}, X_{[3]}, \boldsymbol{X_{[3]}}\}$	$X_{[3]2}$

Finally, for the n-th independent SRS sample of size n, the item with the largest rank is measured and it is denoted by $X_{[n]1}$. This whole process is referred to as a cycle. We repeat this process for N cycles deriving the total number of nN observations from the population. This is called a balanced RSS of size nN. In this setting $X_{[r]i}$ denote the measured value of the r-th ordered unit in the i-th cycle. We call $X_{[r]i}$ the r-th judgment order statistic in the i-th cycle. Our measured balanced RSS is then given by $\{X_{[r]i}, r = 1, \ldots, n; i = 1, \ldots, N\}$, based on the set size n and N cycles. The construction of a balanced RSS design is shown in Table 1.1 with set size $n = 3$ and $N = 2$ cycles. Once the items are judgement ranked from the smallest to the largest within each set, the boldfaced items are selected for full measurement. The measured balanced RSS, in the example, are denoted by $X_{[r]i}, r = 1, 2, 3$ and $i = 1, 2$. One can similarly construct an unbalanced RSS sample of size $\sum_{r=1}^{n} N_r$, say $\{X_{[r]i}, r = 1, \ldots, n; i = 1, \ldots, N_r\}$, where N_r stands for the number of times the r-th judgment order statistic, namely $X_{[r]}$, has been selected for full measurement from the sets.

The essence of RSS is conceptually similar to the stratified sampling technique. The RSS technique uses inherent heterogeneity among the sampling units through a ranking process to create artificial strata. One can consider RSS as a stratification of the units during the sampling process based on their ranks in the sample, although balanced RSS requires identification of Nn^2 units from the population but only nN of them are actually measured. Note also that in SRS, observations are independent and identically distributed ($i.i.d$) and each of them represents a typical value from the population and there is no additional structure imposed on their relationship to one another. In RSS, however, additional information and structure has been provided through the ranking process. Since all the units selected for the final measurement in our sample are obtained from independent sets, the order statistics are mutually independent, however, they are not identically distributed. On the other hand, samples with judgment ranks have the same distribution, provided that the ranking procedure is consistent within the sets. If the ranking process is perfect (i.e. no ranking error), then the distributions of the measured judgment order statistics agree with those of the usual order statistics. Therefore, $X_{[1]1}, \ldots, X_{[n]N}$ are independent judgment order statistics and each of them provides information about different aspects of the population. Indeed, it is this extra structure that allows RSS data to provide more representative samples from the population than SRS data with the same number of measurements.

RSS has many applications in industrial statistics, environmental and ecological studies as well as medical research. For example, in analysis of environmental risks of hazardous waste sites, measuring toxic chemicals and assessing their environmental impact requires substantial scientific processing of materials and, consequently, high cost. However, Barabesi and El-Sharaawi (2001) show that the hazardous waste sites can be easily ranked according to their contamination levels by utilizing a visual inspection of defoliation or soil discolouration.

Another setting where RSS is found to be useful is medical studies. Biomarkers such as polyphenol DNA adducts, micronuclei and sister chromatid exchanges play important roles in the assessment of lung cancer status. Measurement of these biomarkers involves expensive and time-consuming laboratory investigation; however, ranking the sampling units according to their smoking exposure levels can be easily done, for example, using the records of smoking exposure in pack-years. Using the association between smoking exposure and three carcinogenic biomarkers, Chen and Wang (2004) explored the properties of RSS protocols in the analysis of lung cancer. Some other examples of applications of RSS include estimating phytomass (Muttlak and McDonald, 1992), stream habitat area (Mode et al., 1999), mean and variance in flock management (Ozturk et al., 2005) and the mean stock abundance of fish species using the catch-rate data available from previous years as a concomitant variable (Wang et al., 2009). Moreover, there are many applications of RSS in agriculture (e.g., Halls and Dell, 1966 and Cobby et al., 1985) and in environmental studies (e.g., Johnson et al., 1993 and Patil et al., 1994).

Estimation problem of population mean μ can be considered as an example of statistical inference using RSS data. An intuitive estimator for μ under balanced RSS is the average of the RSS observations given by

$$\hat{\mu}_{RSS} = \frac{1}{nN} \sum_{i=1}^{N} \sum_{r=1}^{n} X_{[r]i},$$

which is an unbiased estimator of μ and is always at least as precise as the SRS estimator based on the same number of measured observations. For more details, see

Chen et al. (2004) and Patil (2006). There are a number of factors influencing the superiority of RSS over SRS in different inferential problems (e.g., estimation of the population mean). A key factor here lies in the ranking step of the sampling process. If there is no error in ranking of the sampling units within each set, which results in a perfect RSS, the efficiency (e.g., precision in the estimation of the population mean) of statistical inference gained in RSS is very high. In many practical applications, however, ranking error is inevitable which results in an imperfect RSS. Minimal ranking error may not cause an excessive decrease in the superiority of RSS over SRS, but an increase in the ranking error reduces the efficiency of the RSS technique. In the worst scenario, where ranking is done by chance, RSS and SRS techniques lead to the same statistical inference about the population. Hence the more accuracy in the ranking within each set, the more efficiency is gained (e.g., Barnett and Moore, 1997; Chen, 2000 and Barabesi and El-Sharaawi, 2001). Throughout the thesis, the square brackets are used to show the possibility of ranking errors; hence imperfect RSS data is denoted by $\{X_{[r]i}, r = 1, \ldots, n; i = 1, \ldots, N\}$. If there is no ranking error, square brackets are replaced with round ones so that the perfect RSS data are represented by $\{X_{(r)i}, r = 1, \ldots, n; i = 1, \ldots, N\}$.

As noted earlier, for balanced RSS we measure the same number of observations N (i.e., cycle size) from each judgment ordered statistic; however, if we measure a different number of observations from each rank, the resulting sample is called an unbalanced RSS. In unbalanced RSS, the units are allocated unequally to the ranks. We still sample n independent sets of size n units each from the population and perform the ranking of the sampling units within each set as before; however, we measure N_r units with the r-th judgment rank so that the total number of measured units is then $\sum_{r=1}^{n} N_r$. There are many variations of RSS. For more details on them, readers are referred to Takahasi and Wakimoto (1968), Dell and Clutter (1972), Muttlak and Mc Donald (1990), Ozturk and Wolfe (2000), Jafari Jozani and Johnson (2011) as well as Jafari Jozani and Perron (2011).

RSS was initially introduced and explored in various aspects of nonparametric inference. For example, see Bohn (1996), Presnell and Bohn (1999), Ozturk (1999), Barabesi (2001) and references therein. On the other hand, parametric inference based on RSS has attracted a lot of attention, for example, Sinha et al. (1996), Tarn et al. (1998) and Barnett (1999). The focus of research on parametric problems is mainly placed on different aspects of the best linear unbiased estimator (BLUE) (e.g., Kim and Arnold, 1999 and Lavin, 1999) and theoretical analysis of the efficiency of RSS with respect to SRS (e.g., Stokes, 1995; Chen, 2000; Barabesi and El-Sharaawi, 2001 and Chen et al., 2004). For an overview of the theory and applications of RSS and its variations, readers are referred to the monographs of Chen et al. (2004) and Wolfe (2012).

1.2 Partially rank ordered set sampling

In RSS, if the within-set ranking process is accurate, there will be a big separation among strata (judgment classes) and the inference based on RSS could be highly efficient. On the other hand, error in ranking reduces the efficiency, and may also produce invalid inference. In RSS, rankers have to declare unique ranks for each unit inside the sets. There are many situations where it is difficult to rank all of the sampling units in a set with high confidence, particularly when subjective information is utilized in the ranking process. Forcing rankers to declare unique ranks can lead to inflated within-set judgment ranking error and consequently to invalid statistical inference including tests with inflated type-I error rates, confidence intervals with incorrect coverage probabilities and biased estimates. For this reason, it is desirable to have a sampling design that is less sensitive to within-set ranking error. The partially rank-ordered set (PROS) sampling design is a generalization of RSS, due to Ozturk (2011), which is aimed at reducing the impact of ranking error and the burden on rankers by not requiring them to provide a full ranking of all the units in each set. Under the PROS sampling technique, rankers have more flexibility by being able to divide the sampling units into subsets. These subsets are partially rank-ordered so that each unit in subset h has a rank smaller than the rank of the units in subset h' for all $h' \geq h$. An observation is then collected from one of these subsets in each set.

Let $D = \{d_1, \ldots, d_n\}$ denote a partition of the integers $\{1, \ldots, S\}$ into n mutually exclusive subsets d_r, each of size m, where $d_r = \{(r-1)m+1, \ldots, rm\}$, $r = 1, \ldots, n$, and $m = S/n$. To construct a PROS sample from the population of interest, we first select a set of S units. Then, a ranker is asked to assign these units to subsets $d_r, r = 1, \ldots, n$. This assignment can be made based on the visual inspection, concomitant variables or any other means that does not require a full measurement of the variable of interest. We note that the subsetting process does not require a full ranking of all units, since there is no need to know the ranks of the units in the subset d_r. From the subset d_1, we select a unit at random for full measurement. The measurement from this unit is denoted by $X_{[d_1]1}$. We select another set of size S and again assign the units into the subsets $d_r, r = 1, \ldots, n$. This time we select a unit at random from the subset d_2 for a full measurement and denote it by $X_{[d_2]1}$. We continue this process until we obtain $X_{[d_n]1}$ as the final measurement from the subset d_n. These fully measured observations, $X_{[d_1]1}, \ldots, X_{[d_n]1}$, constitute the PROS sample obtained from the first cycle of the sampling. In order to increase the sample size, this process can be repeated N times to generate a PROS sample of size nN denoted by $\{X_{[d_r]i}; r = 1, \ldots, n; i = 1, \ldots, N\}$. Throughout this thesis, we only focus on balanced PROS sampling design in which the number of observations in each judgment class is the same; however, with slight modifications the results can be obtained for an unbalanced PROS data. Note that, for the special case with $m = 1$, this design reduces to a balanced RSS design with the set size $n = S$, and when $S = 1$ it results in the usual SRS design. On the other hand, it should be noted that there are different variations of PROS sampling design in which the subset sizes and even the number of subsets involved in each set may not be necessarily the same from set to set and cycle to cycle. For more information about the other variations of the PROS sampling design, readers are referred to Ozturk (2011).

The construction of a PROS design is illustrated in Table 1.2 when $S = 9$, $n = 3$ and the cycle size is $N = 2$. In this example, the design $D = \{d_1, d_2, d_3\} = \{\{1, 2, 3\}, \{4, 5, 6\}, \{7, 8, 9\}\}$ is used to generate the data. Each set contains nine units, which are placed into three subsets with partial ranking information. The partial ranking information indicates that the units in d_1 have the smallest three judgment ranks among the nine units, units in subset d_2 have judgment ranks greater than the judgment ranks of units in d_1 and less than the judgment ranks of units in d_3. The partial ranking process dose not assign any ranks to units within subsets. These units are equally likely to take any rank in that subset. One of the units in

Table 1.2: An example of PROS sample

cycle	set	Subsets	Observation
1	S_1	$D_1 = \{\boldsymbol{d_1}, d_2, d_3\} = \{\mathbf{\{1,2,3\}}, \{4,5,6\}, \{7,8,9\}\}$	$X_{[d_1]1}$
	S_2	$D_2 = \{d_1, \boldsymbol{d_2}, d_3\} = \{\{1,2,3\}, \mathbf{\{4,5,6\}}, \{7,8,9\}\}$	$X_{[d_2]1}$
	S_3	$D_3 = \{d_1, d_2, \boldsymbol{d_3}\} = \{\{1,2,3\}, \{4,5,6\}, \mathbf{\{7,8,9\}}\}$	$X_{[d_3]1}$
2	S_1	$D_1 = \{\boldsymbol{d_1}, d_2, d_3\} = \{\mathbf{\{1,2,3\}}, \{4,5,6\}, \{7,8,9\}\}$	$X_{[d_1]2}$
	S_2	$D_2 = \{d_1, \boldsymbol{d_2}, d_3\} = \{\{1,2,3\}, \mathbf{\{4,5,6\}}, \{7,8,9\}\}$	$X_{[d_2]2}$
	S_3	$D_2 = \{d_1, d_2, \boldsymbol{d_3}\} = \{\{1,2,3\}, \{4,5,6\}, \mathbf{\{7,8,9\}}\}$	$X_{[d_3]2}$

each set is selected at random for full measurement, from the bold faced subsets in Table 1.2. The fully measured units are denoted by $X_{[d_r]i}$, $r = 1, 2, 3$ and $i = 1, 2$.

The PROS design has been used successfully in a wide range of problems. Gao and Ozturk (2012) developed a two-sample distribution-free rank-sum test based on PROS data. Ozturk (2012) used a PROS sampling design to draw inference for population quantiles. Frey (2012) studied nonparametric estimation of the population mean with PROS data while Arslan and Ozturk (2013) developed parametric inference for the location and scale parameters of a location-scale family of distributions based on this design. Ozturk and Jafari Jozani (2014) used the properties of PROS sampling for estimation in the finite population settings. Nazari et al. (2014) developed nonparametric kernel density estimators with PROS data and compared them with their SRS and RSS data counterparts.

1.3 Finite mixture models

Due to the recent advances in simulation and computational techniques, finite mixture models (FMMs) have been extensively employed as flexible and convenient statistical tools in data analysis. They are used not only in mainstream statistical analysis such as modelling unknown distributional shapes, analyzing data with group-structures, model-based classification and clustering analysis, but also in a wide range of applications. For example, Sodium and Lithium Counter-transport (SLC) activity in red blood cell is an important trait in quantitive genetics, since it relates to blood pressure and the prevalence of hypertension. Moreover, SLC activity is easier to study than blood pressure. Suppose SLC trait is determined by the action of a single gene with alleles A and a. Using FMMs for modelling the SLC population, Chen et al. (2012) explored the existence of a major gene and if it exists, whether or not it is dominant. FMMs have a wide range of applications in scientific disciplines including quantitative genetics (e.g., Roeder, 1994, Schork et al., 1996 and Chen and Chen, 2003), medical studies (e.g., Schlattmann, 2009) and different engineering fields, such as speech recognition, medical imaging and pattern recognition (e.g., El Zaart et al., 2002). More applications can be found in McLachlan and Peel (2004), McLachlan et al. (2005), McLachlan and Krishnan (2007) as well as Mengersen et al. (2011).

Suppose X is a random variable associated with the random phenomenon of interest and the distribution of X is a mixture of M component densities in some unknown proportions $\pi = (\pi_1, \ldots, \pi_M)$ with $\pi_j > 0$ and $\sum_{j=1}^{M} \pi_j = 1$. Equivalently, each observed data point $X = x$ is taken to be a realization of a FMM with the following probability density function (pdf)

$$f(x_i; \mathbf{\Psi}) = \pi_1 f_1(x_i; \theta_1) + \cdots + \pi_M f_M(x_i; \theta_M), \tag{1.1}$$

where $f_j(\cdot; \theta_j)$, $j = 1, \ldots, M$, refers to the pdf of the j-th component of the model which is specified up to a vector θ_j of unknown parameters, known a priori to be distinct. In this context, it is assumed that the data have come from M different classes $C_1, \ldots, C_M$ where $f_j(\cdot; \theta_j)$ represents the pdf of the variable of interest in the j-th class. The vector of all unknown parameters is denoted by $\boldsymbol{\Psi} = (\pi_1, \ldots, \pi_{M-1}, \boldsymbol{\xi})^\top$, where $\boldsymbol{\xi} = (\theta_1^\top, \ldots, \theta_M^\top)^\top$ and the superscript $\top$ refers to vector transposition. Throughout the thesis, we assume that the number of components of the underlying FMM, M, is known and given.

Among the various statistical inference procedures considered in the literature for FMMs, the maximum likelihood (ML) estimation of parameters of FMM via the EM algorithm (Dempster et al., 1977 and McLachlan and Krishnan, 2007) has dominated the field mainly, because of its simplicity relative to other methods and its monotonic convergence. For a review of the theory and applications of FMMs, readers are referred to McLachlan and Peel (2004), Mengersen et al. (2011) and Titterington et al. (1985).

1.4 An overview of the thesis

In the standard methods of inference for FMM, samples are typically drawn from the population using SRS. There are many problems in finite mixture modelling in which the measurement is costly, destructive or invasive; but ranking a small number of sampling units can be done easily and at little cost; therefore rank-based sampling designs can be used efficiently to tackle the existing problems and also to obtain better inference about the population parameters. Throughout the thesis, we study FMMs based on rank-based samples. Since our approach is model based, whenever we are sampling from a finite population, we assume that sampling schemes including SRS as well as different rank-based samples are performed with replacement. The study of statistical inference for FMMs under rank-based sampling schemes is motivated by various applications ranging from fishery studies to medical research. For example, estimation of the population age structure or describing the length (or weight) distribution of an age class of fish are of high importance in stock assessment and fishery management for monitoring fish populations, especially in the case of short-lived species. The age of the fish is usually determined by examining an individual's growth structures in either the otoliths (ear bones), scales or other bony parts, as there are annular rings (much like a tree) laid down in these structures over successive years as fish grow. Age structure provides information on age at first maturity, age of recruitment, life span, mortality, reproduction and growth in stock composition (e.g., Summerfelt and Hall, 1987; Beamish, 1987). These kinds of studies are not only time-consuming and costly (e.g., hiring expensive experts and requiring substantial scientific and laboratory investigations) but also destructive (e.g., requiring dissection of fish for age determination). For more details, see Kumar and Adams (1977); MacDonald and Pitcher (1979) and Wang et al. (2009). The age group of fish can be estimated (predicted) indirectly by the use of the less expensive and easily obtained length-frequency data which is often modelled by a FMM. In these settings, rank-based sampling designs are more suitable and efficient in reducing the cost and in providing more representative samples from the population.

In this thesis, we study the problem of finite mixture modelling under rank-based sampling designs including RSS, PROS and some of their variations. The work is aimed at getting more insight into finite mixture modelling based on rank-based sampling techniques by developing new, more suitable and efficient statistical methods. In Chapter 2, we study the problem of maximum likelihood (ML) estimation of FMMs under different variations of RSS design. We propose two variations of perfect RSS from a FMM, called M1-RSS as a prospective sampling scheme and M2-RSS as a separate sampling technique. We also develop a new classification criteria to demonstrate the effect of the extra information associated with RSS data. Chapter 3 is devoted to the calculation of the Fisher information (FI) content of RSS data about the FMM parameters. Deriving FI matrices under different RSS approaches (perfect or imperfect), we show the superiority of RSS estimation of parameters of FMM over their SRS counterparts. In Chapter 4, we explore the concept of information and uncertainty structures of the PROS data as a generalization of RSS data in a general class of distributions. The ML estimation in FMM under PROS sampling design is investigated in Chapter 5. We also explore the model-based classification method under PROS sampling. We show that the proposed criterion results in more precise classification of the observed PROS data than the commonly used classification criteria based on SRS data. Our results are finally applied to a fishery study where we estimate the age structure of a short-lived fish species based on the length frequency data. Lastly, the summary of this thesis and future directions for research in this area are presented in Chapter 6.

Chapter 2

Estimation and Classification for FMM under RSS

In this chapter, we study the problem of ML estimation of unknown parameters of FMMs based on two variations of RSS under the assumption of perfect ranking and compare the obtained results with the corresponding ones with SRS data. We show that, using RSS leads to better inference about unknown features of the underlying model such as the estimation of unknown parameters and identification of unobserved classes and their weights. Also, we explore both theoretically and numerically the problem of classification of RSS data and show how the extra information via the rank of each observation will lead to a more efficient classification of the data compared to the usual one with SRS data. Due to the nature of FMM, RSS from the FMM (1.1) can be carried out using two different approaches to be denoted M1-RSS and M2-RSS. In M1-RSS design, the ranked set samples are obtained from the whole mixture model (1.1) so that, within each set, individuals from different components can possibly be involved in the ranking process. This design, which is more practical, can be considered as a kind of prospective RSS and enables us to make better inference on classification of the observations and estimation of the mixing proportions. The M1-RSS design can be used in a wide range of applied fields such as fisheries research, health related studies, economics and environmental and ecological studies. For example, in the fishery example, to model the age class of a specific type of fish (e.g., Spot) by a mixture of two normal densities, we first note that there are two subpopulations consisting of immature (0 or 1 year-old) and mature (2 year-old or older) Spot. One can easily execute an M1-RSS design to obtain an RSS from this population. To this end, subsamples of Spot are obtained and the r-th shortest Spot ($r = 1, \ldots, n$) is retained and then the mature status of the selected Spot is determined (using some time-consuming methods) in the lab. Here the subsamples consist of fish from both components.

In M2-RSS design, it is assumed that RSS is performed within each component of the FMM separately and individuals in each set are obtained from one and only one component of the model. In addition, we assume that the component to which each measurement belongs is unknown (not be confused with labeled data). One advantage associated with this kind of separate sampling is that it proposes an appropriate method for retrospective studies which are often faced in epidemiological investigations. This in turn enables us to make better inference about all components of the underlying population, even components which are rarely observed (McLachlan and Peel, 2004). The M2-RSS design can also be used in a variety of situations where the RSS is performed within each subpopulation separately, however, the component membership of each data point is missing due to some reason, such as confidentiality or simply because it is not recorded. This could be the case for many research studies involving human subjects as participants. For example, statistical agencies could perform RSS separately within different subpopulations, such as males and females, or different minority groups, but in the end, the membership is unknown (perhaps due to confidentiality or sensitivity of the questions). In other words, the observation $X_{(r)i}$ is known to be obtained from a set consisting of units (subjects) from one of the M subpopulations (e.g., all males or all females), but it is not known which subpopulation.

The outline of this chapter is as follows. Section 2.1 deals with the problem of ML estimation of unknown parameters of the FMM using perfect RSS techniques. We also point out the link between inference based on RSS and SRS techniques. Suitable EM algorithms are developed in Section 2.2. We show that the underlying theory behind the EM algorithm for RSS data is different from its counterpart under the standard situation with SRS data. In Section 2.3, we consider the classification problem for an RSS sample of size n and compare it with the classification of SRS samples. Section 2.4 is devoted to the study of the performance of ML estimators of parameters of finite mixtures of normal distributions via an extensive simulation study.

2.1 Likelihood functions

In this section, we study the problem of ML estimation of the unknown parameters $\mathbf{\Psi}$ of the FMM based on RSS data. Let $f(x; \mathbf{\Psi})$ be defined as in (1.1) and suppose $F(x; \mathbf{\Psi}) = \sum_{j=1}^{M} \pi_j F_j(x; \theta_j)$ is its corresponding cumulative distribution function (cdf), where $F_j(x; \theta_j)$, $j = 1, \ldots, M$, refers to the cdf of the j-th component of the model. The pdf of $X_{(r)}$, the r-th order statistic of a sample of size n from (1.1), is given by

$$f^{(r:n)}(x; \mathbf{\Psi}) = n\binom{n-1}{r-1} f(x; \mathbf{\Psi})\{F(x; \mathbf{\Psi})\}^{r-1}\{\bar{F}(x; \mathbf{\Psi})\}^{n-r}, \tag{2.1}$$

where $\bar{F}(x; \mathbf{\Psi}) = 1 - F(x; \mathbf{\Psi})$, and we have

$$f(x; \mathbf{\Psi}) = \frac{1}{n}\sum_{r=1}^{n} f^{(r:n)}(x; \mathbf{\Psi}). \tag{2.2}$$

Similarly, for each component of the FMM (1.1), we have

$$f_j(x; \theta_j) = \frac{1}{n}\sum_{r=1}^{n} f_j^{(r:n)}(x; \theta_j), \quad j = 1, \ldots, M, \tag{2.3}$$

and so (1.1) can be written in terms of the pdf of the order statistics of each component, $f_j^{(r)}(\cdot; \theta_j)$, as follows

$$f(x; \mathbf{\Psi}) = \frac{1}{n}\sum_{r=1}^{n}\sum_{j=1}^{M} \pi_j f_j^{(r:n)}(x; \theta_j). \tag{2.4}$$

2.1.1 Likelihood function for M1-RSS data

Suppose $\mathbf{X}_{M1,RSS} = \{X_{(r)i}, r = 1, \ldots, n; i = 1, \ldots, N\}$ is an M1-RSS sample of size nN from (1.1) where n is the set size and N is the cycle size. Since for each i, $X_{(r)i}$ corresponds to the r-th order statistic of a sample of size n from (1.1), using (2.1), the likelihood function of $\mathbf{\Psi}$ for M1-RSS sample is

$$L_{M1,RSS}(\mathbf{\Psi}) = \prod_{i=1}^{N}\prod_{r=1}^{n} f^{(r:n)}(x_{(r)i}; \mathbf{\Psi}), \tag{2.5}$$

where $f^{(r:n)}(\cdot;\boldsymbol{\Psi})$ is defined in (2.1). Let $l_{M1,RSS}(\boldsymbol{\Psi}) = \log L_{M1,RSS}(\boldsymbol{\Psi})$. Now, the ML estimate of $\boldsymbol{\Psi}$, denoted by $\hat{\boldsymbol{\Psi}}_{M1,RSS}$, is given as an appropriate root of the likelihood equation,

$$\frac{\partial l_{M1,RSS}(\boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}} = 0. \tag{2.6}$$

Remark 2.1. *Suppose $\mathbf{X}_{SRS} = \{X_1, ..., X_{nN}\}$ is an SRS sample of size nN from (1.1). One can represent $\mathbf{X}_{SRS}$ in the form of a matrix $\mathcal{D} = [X_{(r)i}]$ of size $n \times N$, where n and N refer to the number of columns and rows of $\mathcal{D}$, respectively. Here $X_{(r)i}$ simply denotes the (r,i)-th element of $\mathcal{D}$ and the likelihood function of the SRS data can be written as*

$$L_{SRS}(\boldsymbol{\Psi}) = \prod_{i=1}^{N}\prod_{r=1}^{n} f(x_{(r)i};\boldsymbol{\Psi}) = \prod_{i=1}^{N}\prod_{r=1}^{n}\left\{\sum_{j=1}^{M}\pi_j f_j(x_{(r)i};\theta_j)\right\}. \tag{2.7}$$

Using (2.5) we have

$$L_{M1,RSS}(\boldsymbol{\Psi}) = L_{SRS}(\boldsymbol{\Psi}) \times \left\{\prod_{i=1}^{N}\prod_{r=1}^{n} n\binom{n-1}{r-1}[F(x_{(r)i};\boldsymbol{\Psi})]^{r-1}[\bar{F}(x_{(r)i};\boldsymbol{\Psi})]^{n-r}\right\}. \tag{2.8}$$

Note that the extra term in $L_{M1,RSS}(\boldsymbol{\Psi})$ compared with $L_{SRS}(\boldsymbol{\Psi})$ can be interpreted as the effect of the rank information provided to us using RSS as a more complex sampling design.

2.1.2 Likelihood function for M2-RSS data

Suppose $\mathbf{X}_{M2,RSS} = \{X_{(r)i}, r = 1, \ldots, n; i = 1, \ldots, N\}$ is a sample of size nN from (1.1) obtained through M2-RSS design where n and N are defined as before. For the M2-RSS design, $X_{(r)i}$ corresponds to the r-th order statistic of a sample of size n from one of the components of the FMM (1.1); therefore the likelihood function of $\mathbf{\Psi}$ based on M2-RSS data is

$$L_{M2,RSS}(\mathbf{\Psi}) = \prod_{i=1}^{N} \prod_{r=1}^{n} \left\{ \sum_{j=1}^{M} \pi_j f_j^{(r:n)}(x_{(r)i}; \theta_j) \right\}, \tag{2.9}$$

where $f_j^{(r:n)}(\cdot; \mathbf{\Psi})$ corresponds to the pdf of r-th order statistic of component j. Let $l_{M2,RSS}(\mathbf{\Psi}) = \log L_{M2,RSS}(\mathbf{\Psi})$. Now the ML estimate of $\mathbf{\Psi}$, denoted by $\hat{\mathbf{\Psi}}_{M2,RSS}$, is given as an appropriate root of the likelihood equation

$$\frac{\partial l_{M2,RSS}(\mathbf{\Psi})}{\partial \mathbf{\Psi}} = 0. \tag{2.10}$$

In Section 2.2, we develop new EM algorithms to obtain the solutions of (2.6) and (2.10) corresponding to local maximizers of the likelihood functions (2.5) and (2.9), respectively.

2.2 EM algorithms for RSS techniques

The EM algorithm is a general approach that can be used for ML estimation of the parameters of the FMM (1.1). The EM algorithm alternates between two steps, an E-step in which the conditional expectation of the complete data log-likelihood is computed, and an M-step in which parameters that maximize the expected log-likelihood from the E-step are determined. For more details see Dempster et al. (1977). RSS data has a unique data structure different from the usual SRS data and as a result, the standard EM algorithm developed for FMM is not applicable for RSS data. In this section, we develop new EM-algorithms for data obtained through different variations of RSS.

2.2.1 EM algorithm for M1-RSS data

To use the EM algorithm for estimating the parameters of the FMM (1.1) based on an M1-RSS data, the problem is viewed as being incomplete since the label-component vectors associated with feature variables $x_{(r)i}$ are missing. However, what makes this problem different from the standard EM algorithm is the presence of the terms $[F(\cdot;\mathbf{\Psi})]^{r-1}$ and $[\bar{F}(\cdot;\mathbf{\Psi})]^{n-r}$ in (2.1). To overcome this problem, we propose a different missing data mechanism by introducing three different latent vectors for each $x_{(r)i}$. Let $\mathbf{Z}_i^r$ denote the usual component membership of the observation $x_{(r)i}$, $\mathbf{W}_i^r$ denote the component memberships of the observations less than $x_{(r)i}$ and $\mathbf{V}_i^r$ denote the component memberships of the observations larger than $x_{(r)i}$. More specifically, suppose $\mathbf{Z}_i^{(r)} = \left(Z_{i1}^{(r)}, \ldots, Z_{iM}^{(r)}\right)$ is an M-dimensional vector, where $Z_{ij}^{(r)}$ is one or zero, according to whether or not $x_{(r)i}$ belongs to the j-th component of the mixture model ($j = 1, \ldots, M$). That is

$$Z_{ij}^{(r)} = \begin{cases} 1 & \text{if } x_{(r)i} \text{ belongs to component } j; \\ 0 & \text{otherwise,} \end{cases}$$

with $\sum_{j=1}^{M} Z_{ij}^{(r)} = 1$. Indicator vectors $\mathbf{Z}_1^{(1)}, \ldots, \mathbf{Z}_N^{(n)}$ follow a multinomial distribution consisting of one draw on M classes $(C_1, \ldots, C_M)$ with probabilities $\boldsymbol{\pi} =$

$(\pi_1, \ldots, \pi_M)$, that is, $\mathbf{Z}_1^{(1)}, \ldots, \mathbf{Z}_N^{(n)} \overset{i.i.d.}{\sim} Mult(1, \pi)$, with

$$P(\mathbf{Z}_i^{(r)} = \mathbf{z}_i^{(r)}; \boldsymbol{\pi}) = \prod_{j=1}^{M} \pi_j^{z_{ij}^{(r)}}.$$

In addition, suppose $\mathbf{W}_i^{(r)} = \left(W_{i1}^{(r)}, \ldots, W_{iM}^{(r)}\right)$ is an M-dimensional vector, where $W_{ij}^{(r)}$ denotes the number of observations less than $x_{(r)i}$ which are selected from component j. Note that $\sum_{j=1}^{M} W_{ij}^{(r)} = r - 1$. Accordingly, the latent vectors $\mathbf{W}_1^{(1)}, \ldots, \mathbf{W}_N^{(n)}$ are assumed to be distributed according to a multinomial distribution consisting of $r-1$ draws on M classes $(C_1, \ldots, C_M)$ with probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_M)$; that is, $\mathbf{W}_1^{(1)}, \ldots, \mathbf{W}_N^{(n)} \overset{i.i.d.}{\sim} Mult(r-1, \boldsymbol{\pi})$, with

$$P(\mathbf{W}_i^{(r)} = \mathbf{w}_i^{(r)}; \boldsymbol{\pi}) = \binom{r-1}{w_{i1}^{(r)}, \ldots, w_{iM}^{(r)}} \prod_{j=1}^{M} \pi_j^{w_{ij}^{(r)}}.$$

Likewise, let $\mathbf{V}_i^{(r)} = \left(V_{i1}^{(r)}, \ldots, V_{iM}^{(r)}\right)$ be an M-dimensional vector where $V_{ij}^{(r)}$ denotes the number of observations bigger than $x_{(r)i}$ that are selected from component j, with $\sum_{j=1}^{M} V_{ij}^{(r)} = n - r$. Accordingly, the latent vectors $\mathbf{V}_1^{(1)}, \ldots, \mathbf{V}_N^{(r)}$ are assumed to follow a multinomial distribution consisting of $n-r$ draws on M classes $(C_1, \ldots, C_M)$ with probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_M)$; that is, $\mathbf{V}_1^{(1)}, \ldots, \mathbf{V}_N^{(n)} \overset{i.i.d.}{\sim} Mult(n-r, \boldsymbol{\pi})$, with

$$P(\mathbf{V}_i^{(r)} = \mathbf{v}_i^{(r)}; \boldsymbol{\pi}) = \binom{n-r}{v_{i1}^{(r)}, \ldots, v_{iM}^{(r)}} \prod_{j=1}^{M} \pi_j^{v_{ij}^{(r)}}.$$

Since each set of RSS consists of independent samples from the population and component memberships of those observations are independent of each other, the latent variables $Z_i^{(i)}, W_i^{(r)}$ and $V_i^{(r)}$ are conditionally independent.

Lemma 2.1. *For fixed values i and r, $i = 1, \ldots, N$, $r = 1, \ldots, n$; the joint distribution of $(X_{(r)i}, Z_i^{(r)}, W_i^{(r)}, V_i^{(r)})$ is given by*

$$
\begin{aligned}
f(x_{(r)i}, z_i^{(r)}, w_i^{(r)}, v_i^{(r)}; \mathbf{\Psi}) \quad &= \quad c_1\, c_2\, c_3 \prod_{j=1}^{M} \pi_j^{\{z_{ij}^{(r)} + w_{ij}^{(r)} + v_{ij}^{(r)}\}} \\
&\qquad \times \{f_j(x_{(r)i}; \theta_j)\}^{z_{ij}^{(r)}} \{F_j(x_{(r)i}; \theta_j)\}^{w_{ij}^{(r)}} \{\bar{F}_j(x_{(r)i}; \theta_j)\}^{v_{ij}^{(r)}},
\end{aligned}
$$

where $c_1 = n\binom{n-1}{r-1}$, $c_2 = \binom{r-1}{w_{i1}^{(r)}, \ldots, w_{iM}^{(r)}}$ and $c_3 = \binom{n-r}{v_{i1}^{(r)}, \ldots, v_{iM}^{(r)}}$.

Proof. The conditional pdf of the latent variables $Z_{ij}^{(r)}, W_{ij}^{(r)}, V_{ij}^{(r)}$ given $X_{(r)i}$ are given as follows

$$
\begin{aligned}
f(z_{ij}^{(r)} | x_{(r)i}; \mathbf{\Psi}) &= \frac{f(x_{(r)i} | z_{ij}^{(r)}; \mathbf{\Psi}) f(z_{ij}^{(r)})}{f^{(r)}(x_{(r)i}; \mathbf{\Psi})} \\
&= \frac{c_1 \prod_{j=1}^{M} \{f_j(x_{(r)i}; \theta_j)\}^{z_{ij}^{(r)}} \{F(x_{(r)i}; \mathbf{\Psi})\}^{r-1} \{\bar{F}(x_{(r)i}; \mathbf{\Psi})\}^{n-r} \prod_{j=1}^{M} \pi_j^{z_{ij}^{(r)}}}{c_1 f(x_{(r)i}; \mathbf{\Psi}) \{F(x_{(r)i}; \mathbf{\Psi})\}^{r-1} \{\bar{F}(x_{(r)i}; \mathbf{\Psi})\}^{n-r}} \\
&= \prod_{j=1}^{M} \left(\frac{\pi_j f_j(x_{(r)i}; \theta_j)}{f(x_{(r)i}; \mathbf{\Psi})} \right)^{z_{ij}^{(r)}}, \qquad (2.11)
\end{aligned}
$$

as all $z_{ij}^{(r)}$ are zero except for one.

$$f(w_{ij}^{(r)}|x_{(r)i};\mathbf{\Psi}) = \frac{f(x_{(r)i}|w_{ij}^{(r)};\mathbf{\Psi})f(w_{ij}^{(r)})}{f^{(r)}(x_{(r)i};\mathbf{\Psi})}$$

$$= \frac{c_1 c_2 f(x_{(r)i};\mathbf{\Psi}) \prod_{j=1}^{M}\{F_j(x_{(r)i};\theta_j)\}^{w_{ij}^{(r)}} \{\bar{F}(x_{(r)i};\mathbf{\Psi})\}^{n-r} \prod_{j=1}^{M} \pi_j^{w_{ij}^{(r)}}}{c_1 f(x_{(r)i};\mathbf{\Psi})\{F(x_{(r)i};\mathbf{\Psi})\}^{r-1}\{\bar{F}(x_{(r)i};\mathbf{\Psi})\}^{n-r}}$$

$$= c_2 \prod_{j=1}^{M} \left(\frac{\pi_j F_j(x_{(r)i};\theta_j)}{F(x_{(r)i};\mathbf{\Psi})}\right)^{w_{ij}^{(r)}}, \tag{2.12}$$

as $\sum_{j=1}^{M} w_{ij}^{(r)} = r-1$, and

$$f(v_{ij}^{(r)}|x_{(r)i};\mathbf{\Psi}) = \frac{f(x_{(r)i}|v_{ij}^{(r)};\mathbf{\Psi})f(v_{ij}^{(r)})}{f^{(r)}(x_{(r)i};\mathbf{\Psi})}$$

$$= \frac{c_1 c_3 f(x_{(r)i};\mathbf{\Psi})\{F(x_{(r)i};\mathbf{\Psi})\}^{r-1} \prod_{j=1}^{M}\{\bar{F}_j(x_{(r)i};\theta_j)\}^{v_{ij}^{(r)}} \prod_{j=1}^{M} \pi_j^{v_{ij}^{(r)}}}{c_1 f(x_{(r)i};\mathbf{\Psi})\{F(x_{(r)i};\mathbf{\Psi})\}^{r-1}\{\bar{F}(x_{(r)i};\mathbf{\Psi})\}^{n-r}}$$

$$= c_3 \prod_{j=1}^{M} \left(\frac{\pi_j \bar{F}_j(x_{(r)i};\theta_j)}{\bar{F}(x_{(r)i};\mathbf{\Psi})}\right)^{v_{ij}^{(r)}}, \tag{2.13}$$

as $\sum_{j=1}^{M} v_{ij}^{(r)} = n-r$. From the conditional independence of the latent variables, the proof is completed. $\square$

The following Lemma is useful to show that maximization of the complete data log-likelihood of M1-RSS in the M-step of every iteration of the EM algorithm results in the maximization of the corresponding incomplete data log-likelihood.

Lemma 2.2. *For each $x_{(r)i}$, $i = 1,\ldots,N; r = 1,\ldots,n$, we have*

$$f^{(r:n)}(x_{(r)i};\mathbf{\Psi}) = \sum_{\mathbf{z}}\sum_{\mathbf{w}}\sum_{\mathbf{v}} f(x_{(r)i}, z_i^{(r)}, w_i^{(r)}, v_i^{(r)};\mathbf{\Psi}).$$

Proof. To show the result, we have

$$\sum_{z_i^{(r)}} \sum_{w_i^{(r)}} \sum_{v_i^{(r)}} f(x_{(r)i}, z_{ij}^{(r)}, w_{ij}^{(r)}, v_{ij}^{(r)}; \mathbf{\Psi})$$

$$\begin{aligned}
&= c_1 \left\{ \sum_{z_{i1}^{(r)}+\cdots+z_{iM}^{(r)}=1} \prod_{j=1}^{M} \{\pi_j f_j(x_{(r)i};\theta_j)\}^{z_{ij}^{(r)}} \right\} \left\{ \sum_{w_{i1}^{(r)}+\cdots+w_{iM}^{(r)}=r-1} c_2 \prod_{j=1}^{M} \{\pi_j F_j(x_{(r)i};\theta_j)\}^{w_{ij}^{(r)}} \right\} \\
&\quad \times \left\{ \sum_{v_{i1}^{(r)}+\cdots+v_{iM}^{(r)}=n-r} c_3 \prod_{j=1}^{M} \{\pi_j \bar{F}_j(x_{(r)i};\theta_j)\}^{v_{ij}^{(r)}} \right\} \\
&= c_1 \left\{ \sum_{j=1}^{M} \pi_j f_j(x_{(r)i};\theta_j) \right\} \left\{ \sum_{j=1}^{M} \pi_j F_j(x_{(r)i};\theta_j) \right\}^{r-1} \left\{ \sum_{j=1}^{M} \pi_j \bar{F}_j(x_{(r)i};\theta_j) \right\}^{n-r} \\
&= c_1 f(x_{(r)i};\mathbf{\Psi}) \{F(x_{(r)i};\mathbf{\Psi})\}^{r-1} \{\bar{F}(x_{(r)i};\mathbf{\Psi})\}^{n-r}.
\end{aligned}$$

where c_1, c_2 and c_3 are defined as in Lemma 2.1. □

The complete M1-RSS data $\mathbf{Y}_{M1} = \{(X_{(r)i}, Z_i^{(r)}, W_i^{(r)}, V_i^{(r)}), i = 1, \ldots, N; r = 1, \ldots, n\}$ consist of the feature variables and their associated latent variables. Using Lemma 2.1, the complete data likelihood function is given by

$$\begin{aligned}
L_c(\mathbf{\Psi}|\mathbf{y}_{M1}) &= c_1 \prod_{i=1}^{N} \prod_{r=1}^{n} c_2\, c_3 \prod_{j=1}^{M} \pi_j^{\{z_{ij}^{(r)}+w_{ij}^{(r)}+v_{ij}^{(r)}\}} \\
&\quad \times \{f_j(x_{(r)i};\theta_j)\}^{z_{ij}^{(r)}} \{F_j(x_{(r)i};\theta_j)\}^{w_{ij}^{(r)}} \{\bar{F}_j(x_{(r)i};\theta_j)\}^{v_{ij}^{(r)}}, \quad (2.14)
\end{aligned}$$

where the constants c_1, c_2 and c_3 are as in Lemma 2.1. Using (2.14) and Lemma 2.2, it is easy to see that the incomplete-data likelihood function $L_{M1,RSS}(\mathbf{\Psi})$ can be obtained by summing over $\mathbf{Z}, \mathbf{W}$ and $\mathbf{V}$ of the complete-data likelihood, i.e.,

$$\sum_{\mathbf{Z}} \sum_{\mathbf{W}} \sum_{\mathbf{V}} L_c(\mathbf{\Psi}|\mathbf{y}_{M1}) = \prod_{i=1}^{N} \prod_{r=1}^{n} f^{(r)}(x_{(r)i}; \mathbf{\Psi}).$$

Also, the complete-data log-likelihood function of $\mathbf{\Psi}$ is obtained as

$$\begin{aligned} l_{M1}(\mathbf{\Psi}) \;\; \propto \;\; & \sum_{i=1}^{N} \sum_{r=1}^{n} \sum_{j=1}^{M} \Big[Z_{ij}^{(r)} \left\{ \log \pi_j + \log(f_j(x_{(r)i}; \theta_j)) \right\} \\ & + W_{ij}^{(r)} \left\{ \log \pi_j + \log(F_j(x_{(r)i}; \theta_j)) \right\} \\ & + V_{ij}^{(r)} \left\{ \log \pi_j + \log(\bar{F}_j(x_{(r)i}; \theta_j)) \right\} \Big] . \end{aligned} \tag{2.15}$$

Now, we can formulate the EM algorithm for the M1-RSS data as follows; see McLachlan and Peel (2004) for more details.

E-Step: Let $\mathbf{\Psi}^{(0)}$ be the initial value specified for $\mathbf{\Psi}$ and define the conditional expectation of the complete data log-likelihood function (2.15) given the observed data $\mathbf{Y}_{M1} = \mathbf{y}_{M1}$ by

$$Q_{M1}(\mathbf{\Psi}, \mathbf{\Psi}^{(0)}) = \mathbb{E}_{\mathbf{\Psi}^{(0)}}[l_{M1}(\mathbf{\Psi})|\mathbf{y}_{M1}], \tag{2.16}$$

where the expectation will be computed by using $\mathbf{\Psi}^{(0)}$ instead of $\mathbf{\Psi}$ in the conditional distribution. On the $(p+1)$-th iteration, the E-step needs the calculation of $Q_{M1}(\mathbf{\Psi}, \mathbf{\Psi}^{(p)})$, where $\mathbf{\Psi}^{(p)}$ is the value of $\mathbf{\Psi}$ after the p-th iteration. This involves the calculation of the expectations of $Z_{ij}^{(r)}$, $W_{ij}^{(r)}$ and $V_{ij}^{(r)}$ given the observation $x_{(r)i}$. Now from (2.11), (2.12) and (2.13), it is easy to see that

$$Z_{ij}^{(r)}|X_{(r)i} = x_{(r)i} \sim \text{Bin}\left(1, \frac{\pi_j f_j(x_{(r)i};\theta_j)}{f(x_{(r)i};\boldsymbol{\Psi})}\right),$$

$$W_{ij}^{(r)}|X_{(r)i} = x_{(r)i} \sim \text{Bin}\left(r-1, \frac{\pi_j F_j(x_{(r)i};\theta_j)}{F(x_{(r)i};\boldsymbol{\Psi})}\right),$$

$$V_{ij}^{(r)}|X_{(r)i} = x_{(r)i} \sim \text{Bin}\left(n-r, \frac{\pi_j \bar{F}_j(x_{(r)i};\theta_j)}{\bar{F}(x_{(r)i};\boldsymbol{\Psi})}\right),$$

where $i = 1,\ldots,N$; $r = 1,\ldots,n$ and $j = 1,\ldots,M$. Hence, the conditional expectations of $Z_{ij}^{(r)}$, $W_{ij}^{(r)}$ and $V_{ij}^{(r)}$ given the observation $x_{(r)i}$ are given by follow

$$\tau_{j,M1}(x_{(r)i};\boldsymbol{\Psi}^{(p)}) = \mathbb{E}_{\boldsymbol{\Psi}^{(p)}}[Z_{ij}^{(r)}|x_{(r)i}] = \frac{\pi_j^{(p)} f_j(x_{(r)i};\theta_j^{(p)})}{\sum_{h=1}^{M}\pi_h^{(p)} f_h(x_{(r)i};\theta_h^{(p)})}, \tag{2.17}$$

$$\beta_{j,M1}(x_{(r)i};\boldsymbol{\Psi}^{(p)}) = \mathbb{E}_{\boldsymbol{\Psi}^{(p)}}[W_{ij}^{(r)}|x_{(r)i}] = \frac{(r-1)\pi_j^{(p)} F_j(x_{(r)i};\theta_j^{(p)})}{\sum_{h=1}^{M}\pi_h^{(p)} F_h(x_{(r)i};\theta_h^{(p)})}, \tag{2.18}$$

$$\gamma_{j,M1}(x_{(r)i};\boldsymbol{\Psi}^{(p)}) = \mathbb{E}_{\boldsymbol{\Psi}^{(p)}}[V_{ij}^{(r)}|x_{(r)i}] = \frac{(n-r)\pi_j^{(p)} \bar{F}_j(x_{(r)i};\theta_j^{(p)})}{\sum_{h=1}^{M}\pi_h^{(p)} \bar{F}_h(x_{(r)i};\theta_h^{(p)})}. \tag{2.19}$$

Using (2.17), (2.18) and (2.19), we have

$$Q_{M1}(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(p)}) = cst + \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M} \tau_{j,M1}(x_{(r)i}; \boldsymbol{\Psi}^{(p)}) \left\{\log \pi_j + \log(f_j(x_{(r)i}; \theta_j))\right\}$$

$$+ \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M} \beta_{j,M1}(x_{(r)i}; \boldsymbol{\Psi}^{(p)})\{\log \pi_j + \log(F_j(x_{(r)i}; \theta_j))\}$$

$$+ \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M} \gamma_{j,M1}(x_{(r)i}; \boldsymbol{\Psi}^{(p)})\{\log \pi_j + \log(\bar{F}_j(x_{(r)i}; \theta_j))\}. \tag{2.20}$$

Now, we are ready to implement the M-step.

M-Step: In this step, the maximization of $Q_{M1}(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(p)})$ with respect to $\boldsymbol{\Psi}$ will be done over the parameter space to obtain the updated estimates $\boldsymbol{\Psi}^{(p+1)} = (\pi_1^{(p+1)}, \ldots, \pi_{M-1}^{(p+1)}, \boldsymbol{\xi}^{(p+1)})^\top$. Note that according to (2.15) the updated estimates $\pi_j^{(p+1)}$ of the mixing proportions π_j can be calculated independently of the updated estimates $\boldsymbol{\xi}^{(p+1)}$ of the parameters $\boldsymbol{\xi}$ in $\boldsymbol{\Psi}$. If $Z_{ij}^{(r)}, W_{ij}^{(r)}$ and $V_{ij}^{(r)}$ were observed by $z_{ij}^{(r)}, w_{ij}^{(r)}$ and $v_{ij}^{(r)}$, the (complete data) ML estimate of π_j would be given by

$$\hat{\pi}_j = \frac{1}{Nn^2}\sum_{i=1}^{N}\sum_{r=1}^{n}\left(z_{ij}^{(r)} + w_{ij}^{(r)} + v_{ij}^{(r)}\right), \tag{2.21}$$

by using the Lagrangian multipliers method over the constraint $\sum_{j=1}^{M}\pi_j = 1$; however, since they are not observable, we update the estimates of π_j, $j =$

$1, \ldots, M$, by

$$\pi_{j,M1}^{(p+1)} = \frac{1}{Nn^2} \sum_{i=1}^{N} \sum_{r=1}^{n} \left\{ \tau_{j,M1}(x_{(r)i}; \boldsymbol{\Psi}^{(p)}) + \beta_{j,M1}(x_{(r)i}; \boldsymbol{\Psi}^{(p)}) + \gamma_{j,M1}(x_{(r)i}; \boldsymbol{\Psi}^{(p)}) \right\}. \tag{2.22}$$

Also, the updated value $\boldsymbol{\xi}^{(p+1)}$ is obtained as the solution of

$$\sum_{i=1}^{N} \sum_{r=1}^{n} \sum_{j=1}^{M} \frac{\tau_{j,M1}(x_{(r)i}; \boldsymbol{\Psi}^{(p)})}{f_j(x_{(r)i}; \theta_j)} \frac{\partial}{\partial \boldsymbol{\xi}} f_j(x_{(r)i}; \theta_j)$$
$$+ \sum_{i=1}^{N} \sum_{r=1}^{n} \sum_{j=1}^{M} \frac{\partial}{\partial \boldsymbol{\xi}} F_j(x_{(r)i}; \theta_j) \left(\frac{\beta_{j,M1}(x_{(r)i}; \boldsymbol{\Psi}^{(p)})}{F_j(x_{(r)i}; \theta_j)} - \frac{\gamma_{j,M1}(x_{(r)i}; \boldsymbol{\Psi}^{(p)})}{\bar{F}_j(x_{(r)i}; \theta_j)} \right) = 0, \tag{2.23}$$

with respect to $\boldsymbol{\xi}$.

The E- and M-steps are iterated repeatedly until $|l_{M1}(\boldsymbol{\Psi}^{(p+1)}) - l_{M1}(\boldsymbol{\Psi}^{(p)})|$ becomes negligible.

2.2.2 EM algorithm for M2-RSS data

To formalize the EM algorithm based on M2-RSS, we only need to define one latent variable. Suppose $\mathbf{Z}_i^{(r)} = \left(Z_{i1}^{(r)}, \ldots, Z_{iM}^{(r)}\right)$ is an M-dimensional vector where $Z_{ij}^{(r)}$ is one or zero, according to whether or not $X_{(r)i}$ corresponds to the r-th order statistic of the j-th component of the FMM ($j = 1, \ldots, M$). The conditional pdf of the feature variables $X_{(1)1}, \ldots, X_{(n)N}$ given $\mathbf{Z}_1^{(1)}, \ldots, \mathbf{Z}_N^{(n)}$ is as follows

$$f(x_{(1)1}, \ldots, x_{(n)N} | \mathbf{z}_1^{(1)}, \ldots, \mathbf{z}_N^{(n)}; \boldsymbol{\xi}) = \prod_{i=1}^{N} \prod_{r=1}^{n} f^{(r:n)}(x_{(r)i} | \mathbf{z}_i^{(r)}; \boldsymbol{\xi}),$$

where

$$f^{(r:n)}(x_{(r)i}|\mathbf{z}_i^{(r)};\boldsymbol{\xi}) = \prod_{j=1}^{M}\left\{f_j^{(r:n)}(x_{(r)i};\theta_j)\right\}^{z_{ij}^{(r)}}.$$

Also, the likelihood function of $\boldsymbol{\Psi}$ based on the complete M2-RSS data which is denoted by $\mathbf{Y}_{M2} = \left\{(X_{(r)i}, \mathbf{Z}_i^{(r)}), i = 1, \ldots, N; r = 1, \ldots, n\right\}$ can be expressed as

$$\begin{aligned}
L_{M2}(\boldsymbol{\Psi}) &= \prod_{i=1}^{N}\prod_{r=1}^{n}\prod_{j=1}^{M}\left\{\pi_j f_j^{(r:n)}(x_{(r)i};\theta_j)\right\}^{Z_{ij}^{(r)}} \\
&= \prod_{i=1}^{N}\prod_{r=1}^{n}\prod_{j=1}^{M}\left\{n\binom{n-1}{r-1}\pi_j f_j(x_{(r)i};\theta_j)[F_j(x_{(r)i};\theta_j)]^{r-1}[1-F_j(x_{(r)i};\theta_j)]^{n-r}\right\}^{Z_{ij}^{(r)}}.
\end{aligned} \tag{2.24}$$

Here again the incomplete M2-RSS likelihood function (2.9) can be obtained by summing $\mathbf{Z_i^{(r)}}$ out of the complete M2-RSS likelihood function $L_{M2}(\boldsymbol{\Psi})$. The likelihood function (2.24) leads to the complete data log-likelihood function of $\boldsymbol{\Psi}$ as

$$\begin{aligned}
l_{M2}(\boldsymbol{\Psi}) &= \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M} Z_{ij}^{(r)}\left\{\log \pi_j + \log f_j^{(r:n)}(x_{(r)i};\theta_j)\right\} \\
&= N\sum_{r=1}^{n}\log n\binom{n-1}{r-1} + \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M} Z_{ij}^{(r)}\left\{\log \pi_j + \log f_j(x_{(r)i};\theta_j)\right\} \\
&+ \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M} Z_{ij}^{(r)}\left\{(r-1)\log F_j(x_{(r)i};\theta_j) + (n-r)\log(\bar{F}_j(x_{(r)i};\theta_j))\right\}.
\end{aligned} \tag{2.25}$$

Now, the EM algorithm can be applied to obtain estimates of $\boldsymbol{\Psi}$. To this end, we first calculate

$$\tau_{j,M2}(x_{(r)i}; \mathbf{\Psi}^{(p)}) = \mathbb{E}_{\mathbf{\Psi}^{(p)}}[Z_{ij}^{(r)}|x_{(r)i}]$$

$$= \frac{\pi_j^{(p)} f_j(x_{(r)i}; \theta_j^{(p)})[F_j(x_{(r)i}; \theta_j^{(p)})]^{r-1}[\bar{F}_j(x_{(r)i}; \theta_j^{(p)})]^{n-r}}{\sum_{h=1}^{M} \pi_h^{(p)} f_h(x_{(r)i}; \theta_h^{(p)})[F_h(x_{(r)i}; \theta_h^{(p)})]^{r-1}[\bar{F}_h(x_{(r)i}; \theta_h^{(p)})]^{n-r}}, \tag{2.26}$$

where $j = 1, \ldots, n; i = 1, \ldots, N$ and $\mathbf{\Psi}^{(0)}$ denotes the initial value for $\mathbf{\Psi}$. Using (2.26), the conditional expectation of the complete M2-RSS data log-likelihood function (2.25) given the observed data $\mathbf{X}_{RSS} = \mathbf{x}_{RSS}$ (on the p-th iteration) of the EM algorithm yields

$$\begin{aligned} Q_{M2}(\mathbf{\Psi}, \mathbf{\Psi}^{(p)}) = N \sum_{r=1}^{n} \log n \binom{n-1}{r-1} \\ + \sum_{i=1}^{N} \sum_{r=1}^{n} \sum_{j=1}^{M} \Big[\tau_{j,M2}(x_{(r)i}; \mathbf{\Psi}^{(p)}) \left\{\log \pi_j + \log f_j(x_{(r)i}; \theta_j)\right\} \\ + \tau_{j,M2}(x_{(r)i}; \mathbf{\Psi}^{(p)})\{(r-1)\log F_j(x_{(r)i}; \theta_j) + (n-r)\log \bar{F}_j(x_{(r)i}; \theta_j)\}\Big]. \end{aligned} \tag{2.27}$$

In the M-step, using the Lagrangian multipliers method, we update mixing proportions by (2.26) independently of the other parameters of the model through

$$\pi_j^{(p+1)} = \frac{1}{nN} \sum_{i=1}^{N} \sum_{r=1}^{n} \tau_{j,M2}(x_{(r)i}; \mathbf{\Psi}^{(p)}), \ j = 1, \ldots, M-1. \tag{2.28}$$

Furthermore, $\xi^{(p+1)}$ is updated by an appropriate root of

$$\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\frac{\tau_{j,M2}(x_{(r)i};\mathbf{\Psi}^{(p)})}{f_j(x_{(r)i};\theta_j)}\frac{\partial}{\partial\boldsymbol{\xi}}f_j(x_{(r)i};\theta_j)$$

$$+\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\tau_{j,M2}(x_{(r)i};\mathbf{\Psi}^{(p)})\frac{\partial}{\partial\boldsymbol{\xi}}F_j(x_{(r)i};\theta_j)\left(\frac{r-1}{F_j(x_{(r)i};\theta_j)}-\frac{n-r}{\bar{F}_j(x_{(r)i};\theta_j)}\right)=0. \tag{2.29}$$

The E- and M- steps are iterated repeatedly until $|l_{M2}(\mathbf{\Psi}^{(p+1)})-l_{M2}(\mathbf{\Psi}^{(p)})|$ becomes negligible.

Remark 2.2. *The complete data likelihood function of $\mathbf{\Psi}$ for the SRS sample can be written as*

$$L_{C,SRS}(\mathbf{\Psi})=\prod_{i=1}^{n}\prod_{j=1}^{M}\{\pi_j f_j(x_i;\theta_j)\}^{z_{ij}}, \tag{2.30}$$

where $x_{(r)i}$ refers to the (r,i)-th element of the matrix of observations $\mathcal{D}$ and $Z_{ij}^{(r)}$ is its corresponding indicator function which specifies whether $x_{(r)i}$ is an observation from the j-th component of the FMM (1.1) or not. It can be seen that $\log L_{C,SRS}(\mathbf{\Psi})=l_{C,SRS}(\mathbf{\Psi})=l_{M2}(\mathbf{\Psi})-\Lambda_{M2}(\boldsymbol{\xi})$, *with*

$$\begin{aligned}\Lambda_{M2}(\boldsymbol{\xi}) \quad = \quad & N\sum_{r=1}^{n}\log n\binom{n-1}{r-1}\\ & +\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}Z_{ij}^{(r)}\left\{(r-1)\log F_j(x_{(r)i};\theta_j)+(n-r)\log\bar{F}_j(x_{(r)i};\theta_j)\right\}.\end{aligned}$$

In addition, $l_{C,SRS}(\mathbf{\Psi}) = l_{M1}(\mathbf{\Psi}) - \Lambda_{M1}(\boldsymbol{\xi})$, *and*

$$\begin{aligned}\Lambda_{M1}(\boldsymbol{\xi}) &= \sum_{i=1}^{N}\sum_{r=1}^{n}\log\left\{n\binom{n-1}{r-1}\binom{r-1}{w_{i1}^{(r)},\ldots,w_{iM}^{(r)}}\binom{n-r}{v_{i1}^{(r)},\ldots,v_{iM}^{(r)}}\right\} \\ &\quad+\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}W_{ij}^{(r)}\{\log\pi_j+\log F_j(x_{(r)i};\theta_j)\} \\ &\quad+\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}V_{ij}^{(r)}\{\log\pi_j+\log\bar{F}_j(x_{(r)i};\theta_j)\}.\end{aligned}$$

One can easily rewrite the EM algorithm based on SRS data in terms of their new matrix representations defined in Remark 2.1. To this end, for the E-step, let $\mathbf{\Psi}^{(0)}$ *be the initial value specified for* $\mathbf{\Psi}$ *and* $Q(\mathbf{\Psi},\mathbf{\Psi}^{(0)}) = \mathbb{E}_{\mathbf{\Psi}^{(0)}}[l_{C,SRS}(\mathbf{\Psi})|\mathbf{x}_{SRS}]$*. In the* $(p+1)$*-th iteration of this step we compute*

$$Q(\mathbf{\Psi},\mathbf{\Psi}^{(p)}) = \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\tau_{j,SRS}(x_{(r)i};\mathbf{\Psi}^{(p)})\left\{\log\pi_j+\log f_j(x_{(r)i};\theta_j)\right\}, \tag{2.31}$$

where

$$\tau_{j,SRS}(x_{(r)i};\mathbf{\Psi}^{(p)}) = \frac{\pi_j^{(p)}f_j(x_{(r)i};\theta_j^{(p)})}{\sum_{h=1}^{M}\pi_h^{(p)}f_h(x_{(r)i};\theta_h^{(p)})}. \tag{2.32}$$

Similarly, for the M-step, and at its $(p+1)$*-th iteration, a local maximization of* $Q(\mathbf{\Psi},\mathbf{\Psi}^{(p)})$ *with respect to* $\mathbf{\Psi}$ *will be done over the parameter space to obtain* $\mathbf{\Psi}^{(p+1)}$*. The updated estimate of* π_j *is given by*

$$\pi_j^{(p+1)} = \frac{1}{Nn}\sum_{i=1}^{N}\sum_{r=1}^{n}\tau_{j,SRS}(x_{(r)i};\mathbf{\Psi}^{(p)}),\ j=1,\ldots,M-1, \tag{2.33}$$

while the updated estimate of $\boldsymbol{\xi}$ is obtained as an appropriate root of

$$\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\frac{\tau_{j,SRS}(x_{(r)i};\boldsymbol{\Psi}^{(p)})}{f_j(x_{(r)i};\theta_j)}\frac{\partial}{\partial\boldsymbol{\xi}}f_j(x_{(r)i};\theta_j)=0. \tag{2.34}$$

A comparison of (2.23) *and* (2.29) *with* (2.34) *shows the contribution of the ranks provided by the RSS technique in obtaining the ML estimates of* $\boldsymbol{\Psi}$.

2.2.3 Modified EM algorithm for M2-RSS

ML estimation of $\boldsymbol{\Psi}$ in the FMM (1.1) based on different RSS samples through the EM algorithms consist of updating ξ in the corresponding M-step for the $(p+1)$-th iteration by solving (2.23) and (2.29) in θ_j's. However, it may not be easy to solve those equations and/or find closed form expressions for the solutions. This, of course, is due to the presence of the terms involving forms of the hazard and inverse hazard rate functions, i.e. $\frac{\partial}{\partial\boldsymbol{\xi}}\log F_j(\cdot;\theta_j)$ and $\frac{\partial}{\partial\boldsymbol{\xi}}\log(1-F_j(\cdot;\theta_j))$ in the obtained equations. In this section, we provide an approximate ML method to reduce the computational complexity for iteratively solving the likelihood functions and to obtain simpler and possibly closed forms for the solutions to (2.29). The results, with slight modifications, can be similarly extended to the case of M1-RSS. We use the method of Johnson et al. (1972) as well as Mehrotra and Nanda (1974) which are based on replacing the hazard rate terms by their expectations. Mehrotra and Nanda (1974) applied this method in estimating parameters of normal and gamma distributions based on type II censored data and showed that their approximate ML estimators are still unbiased and are highly efficient relative to the ML estimators based on the complete ML equation. For more details we refer to Bhattacharyya

(1985) which provides a rigorous treatment of this method along with some of its optimal properties.

We first obtain a modified version of (2.29) by replacing the terms involving $\frac{\partial}{\partial \boldsymbol{\xi}} \log F_j(\cdot;\theta_j)$ and $\frac{\partial}{\partial \boldsymbol{\xi}} \log(1 - F_j(\cdot;\theta_j))$ by their expectations. This results in simpler expression for updating ξ in the M-step of the $(p+1)$-th iteration of the EM algorithm. Note that, this method will not affect the estimation procedure of the mixing proportions π_j's, $j = 1, \ldots, M$. We need the following lemma whose proof can be found in Chen (2000) as well as in Chen et al. (2004).

Lemma 2.3. *Let $Y_r = X_{(r)i}$, $r = 1, \ldots, n$, $i = 1, \ldots, N$, with pdf $f^{(r:n)}(\cdot;\boldsymbol{\Psi})$ be the r-th order statistic of a simple random sample of size n from $f(x;\boldsymbol{\Psi})$ with cdf $F(x;\boldsymbol{\Psi})$. Then, for any function $G(\cdot)$,*

$$\mathbb{E}\left[\sum_{r=1}^{n}(r-1)\frac{G(Y_r)}{F(Y_r;\boldsymbol{\Psi})}\right] = \mathbb{E}\left[\sum_{r=1}^{n}(n-r)\frac{G(Y_r)}{1-F(Y_r;\boldsymbol{\Psi})}\right] = n(n-1)\mathbb{E}_{\boldsymbol{\Psi}}[G(X)].$$

We need to compute

$$I_{ijr} = \mathbb{E}\left[\tau_{j,M2}(X_{(r)i};\boldsymbol{\Psi}^{(p)})\left(\frac{\frac{\partial}{\partial\xi}F_j(X_{(r)i};\theta_j)}{F_j(X_{(r)i};\theta_j)}\right)\right],$$

and

$$J_{ijr} = \mathbb{E}\left[\tau_{j,M2}(X_{(r)i};\boldsymbol{\Psi}^{(p)})\left(\frac{\frac{\partial}{\partial\xi}F_j(X_{(r)i};\theta_j)}{1-F_j(X_{(r)i};\theta_j)}\right)\right].$$

Using Lemma 2.3 with $G_j(X_{(r)i}) = \tau_{j,M2}(X_{(r)i};\boldsymbol{\Psi}^{(p)})\frac{\partial}{\partial\boldsymbol{\xi}}F_j(X_{(r)i};\theta_j)$, it is easy to show that

$$\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}(r-1)I_{ijr} = N\sum_{j=1}^{M}\sum_{r=1}^{n}(r-1)\mathbb{E}\left[\frac{G_j(X_{(r)1})}{F_j(X_{(r)1};\theta_j)}\right]$$

$$= Nn(n-1)\sum_{j=1}^{M}\mathbb{E}[G_j(X)]. \tag{2.35}$$

Similarly, it can be shown that

$$\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}(n-r)J_{ijr} = N\sum_{j=1}^{M}\sum_{r=1}^{n}(n-r)\mathbb{E}\left[\frac{G_j(X_{(r)1})}{1-F_j(X_{(r)1};\theta_j)}\right]$$

$$= Nn(n-1)\sum_{j=1}^{M}\mathbb{E}[G_j(X)]. \tag{2.36}$$

Now, upon substituting (2.35) and (2.36) in (2.29) we get a modified expression for updating $\boldsymbol{\xi}$ in the M-step of the EM-M2 algorithm which leads to an approximate ML estimate of $\boldsymbol{\xi}$ as a solution to

$$\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\frac{\tau_{j,M2}(x_{(r)i};\boldsymbol{\Psi}^{(p)})}{f_j(x_{(r)i};\theta_j)}\frac{\partial}{\partial\boldsymbol{\xi}}f_j(x_{(r)i};\theta_j) = 0. \tag{2.37}$$

Note that (2.37) is similar to the updating equation (2.34) obtained in McLachlan and Peel (2004) under the SRS technique. Similarly to the SRS case, one nice feature of this modified version of the EM algorithm is that the solutions of (2.37) often exist in closed form and that they can be obtained by replacing $\tau_{j,M2}(x_{(r)i};\boldsymbol{\Psi}^{(p)})$ by $\tau_{j,SRS}(x_{(r)i};\boldsymbol{\Psi}^{(p)})$ in the available solutions for their SRS counterparts.

2.3 Classification of the RSS sample

Once $\hat{\mathbf{\Psi}}$ is obtained (based on either M1-RSS or M2-RSS data) estimates of the posterior probabilities of the population membership can be formed for each observation to perform a probabilistic classification of the data. Suppose $i \in \{1, \dots, N\}$ is fixed and $y_r = x_{(r)i}$ is observed. For the M2-RSS design, classification of y_r is based on the posterior probability that y_r belongs to the j-th component of the mixture model, that is $\tau_{j,M2}(y_r; \mathbf{\Psi})$ as in (2.26). Also, from (2.17), (2.18) and (2.19), classification of y_r under M1-RSS design is done via the following formula

$$\alpha_{j,M1} = \frac{1}{n} \left\{ \tau_{j,M1}(y_r, \mathbf{\Psi}) + \beta_{j,M1}(y_r, \mathbf{\Psi}) + \gamma_{j,M1}(y_r, \mathbf{\Psi}) \right\}. \tag{2.38}$$

Now, y_r is classified into the component j, i.e., $y_r \in C_j$, if

$$\tau_{j,M2}(y_r; \hat{\mathbf{\Psi}}) > \tau_{t,M2}(y_r; \hat{\mathbf{\Psi}}) \quad \text{or} \quad \alpha_{j,M1}(y_r; \hat{\mathbf{\Psi}}) > \alpha_{t,M1}(y_r; \hat{\mathbf{\Psi}}),$$

for all $t \neq j$, $t = 1, \dots, M$. In this section, we focus on this classification approach and demonstrate the effect of the extra information obtained from the ranks in RSS when designs compared to SRS.

For an SRS sample of size Nn the posterior probability that $y_r = x_{(r)i}$ belongs to the j-th component of the mixture model is given by

$$\tau_{j,SRS}(y_r; \mathbf{\Psi}) = \frac{\pi_j f_j(y_r; \theta_j)}{\sum_{h=1}^{M} \pi_h f_h(y_r; \theta_h)}, \tag{2.39}$$

Now, it can be seen that

$$\alpha_{j,M1}(y_r; \mathbf{\Psi}) = \frac{1}{n} \left\{ \tau_{j,SRS}(y_r; \mathbf{\Psi}) + A_{j,M1}(y_r; \mathbf{\Psi}) \right\}, \tag{2.40}$$

where

$$A_{j,M1}(y_r;\mathbf{\Psi}) = (r-1)\left(\frac{\pi_j F_j(x_{(r)i};\theta_j)}{F(x_{(r)i};\mathbf{\Psi})}\right) + (n-r)\left(\frac{\pi_j \bar{F}_j(x_{(r)i};\theta_j)}{\bar{F}(x_{(r)i};\mathbf{\Psi})}\right). \tag{2.41}$$

Using M2-RSS design instead, we get

$$\tau_{j,M2}(y_r;\mathbf{\Psi}) = \tau_{j,SRS}(y_r;\mathbf{\Psi})A_{j,M2}(y_r;\mathbf{\Psi}), \tag{2.42}$$

where

$$A_{j,M2}(y_r;\mathbf{\Psi}) = \frac{\left(\sum_{v=1}^{M}\pi_v f_v(y_r;\theta_v)\right)[F_j(y_r;\theta_j)]^{r-1}[1-F_j(y_r;\theta_j)]^{n-r}}{\sum_{h=1}^{M}\pi_h f_h(y_r;\theta_h)[F_h(y_r;\theta_h)]^{r-1}[1-F_h(y_r;\theta_h)]^{n-r}}. \tag{2.43}$$

The following example demonstrates the effect of the rank information on the classification decision. Suppose $x=0$ is observed from a population. Assume that the underlying population consists of two components C_1 and C_2 with pdfs

$$0.5\phi(x;-2,1)+0.5\phi(x;1,1) \quad \text{or} \quad 0.6\phi(x;-2,1)+0.4\phi(x;1,3),$$

where $\phi(x;\mu,\sigma)$ refers to the pdf of a normal distribution with mean μ and variance σ^2. Assume that $x=0$ is observed through an RSS with the set size $n=3$. As it is shown in Table 2.1, if x is treated as an observation obtained from an SRS (i.e., ignoring its rank information), since $\tau_{2,SRS}(0;\mathbf{\Psi}) = 0.8176 \geq \tau_{1,SRS}(0;\mathbf{\Psi}) = 0.1824$, then it should be classified into the second component, C_2, of the population. However, using the rank information attached to $x=0$ one could get a different classification result. For example, if $x=0$ is to be the observation of the third order statistic via the M1-RSS technique with $n=3$ then we have $\alpha_{1,M1}(0;\mathbf{\Psi}) = 0.6344 \geq \alpha_{2,M1}(0;\mathbf{\Psi}) = 0.3656$. That is, $x=0$ should be classified into C_1. Similarly, from

Table 2.1: Classification of $x = 0$ under SRS and M1-RSS designs.

Mixture Model	Mixture Component	SRS $x = 0$	M1-RSS with $n = 3$ $y_1 = 0$	$y_2 = 0$	$y_3 = 0$
$0.5\phi(x; -2, 1) + 0.5\phi(x; 1, 1)$	C_1	0.1824	0.0783	0.3563	0.6344
	C_2	0.8176	0.9217	0.6437	0.3656
$0.6\phi(x; -2, 1) + 0.4\phi(x; 1, 3)$	C_1	0.3917	0.1647	0.4138	0.6631
	C_2	0.6083	0.8353	0.5862	0.3369

Table 2.2: Classification of $x = 0$ under SRS and M2-RSS designs.

Mixture Model	Mixture Component	SRS $x = 0$	M2-RSS with $n = 3$ $y_1 = 0$	$y_2 = 0$	$y_3 = 0$
$0.5\phi(x; -2, 1) + 0.5\phi(x; 1, 1)$	C_1	0.1824	0.0002	0.0358	0.8944
	C_2	0.8176	0.9998	0.9642	0.1056
$0.6\phi(x; -2, 1) + 0.4\phi(x; 1, 3)$	C_1	0.3917	0.0008	0.0578	0.8184
	C_2	0.6083	0.9992	0.9422	0.1816

Table 2.2, if $x = 0$ is actually the observation of the third order statistic in an M2-RSS design with $n = 3$, we have $\tau_{1,M2}(0; \mathbf{\Psi}) = 0.8944 \geq \tau_{2,M2}(0; \mathbf{\Psi}) = 0.1056$ which means that $x = 0$ should be classified into C_1. Tables 2.1 and 2.2 provide the values of the posterior probabilities for the classification of $x = 0$ under the SRS, M1-RSS and M2-RSS designs for two different normal mixtures.

Note that $T_j \sim \text{Bin}(n-1, F_j(y_r; \theta_j))$. Equation (2.43) can be written as follows,

$$A_{j,M2}(y_r, \mathbf{\Psi}) = \frac{\left(\sum_{v=1}^{M} \pi_v f_v(y_r; \theta_v)\right) P(T_j = r-1)}{\sum_{h=1}^{M} \pi_h f_h(y_r; \theta_h) P(T_h = r-1)}, \tag{2.44}$$

and if $P(T_j = r-1) \geq P(T_h = r-1)$ for all $h \neq j \in \{1, \ldots, M\}$, then the posterior probability that y_r belongs to the j-th component of the FMM is bigger in RSS than in SRS.

2.4 Simulation studies

In this section, using simulation studies, the performance of ML estimators of the unknown parameters of the finite mixture of normal distributions based on M1-RSS and M2-RSS is investigated. In the first simulation study, the emphasis is placed on the comparison between M1-RSS, M2-RSS and SRS designs for estimating the mixing proportions of the model. In our second simulation study, the performance of the ML estimators of all parameters of the model based on M1-RSS and M2-RSS are compared with the corresponding ones under SRS.

2.4.1 Simulation study 1

Here the goal is to compare the performance of the estimators of the mixing proportion π using M1-RSS, M2-RSS and SRS data for a mixture of two normal distributions

$$f(x; \boldsymbol{\Psi}) = \pi\phi(x; \mu_1, \sigma) + (1 - \pi)\phi(x; \mu_2, \sigma). \tag{2.45}$$

We first compare the performance of $\hat{\pi}$ under M1-RSS and SRS designs. To this end, we generate two data sets each of size $Nn = 120$ form the model (2.45) with $\boldsymbol{\Psi} = (\pi, \mu_1, \sigma) = (0.8, -1, 1)$. To study the effect of the set size as an important design parameter of RSS on the performance of the estimators, we let $n \in \{1, 2, 3, 4, 5\}$. Note that $n = 1$ corresponds to the usual SRS method. To investigate the effect of the distance between the components of the model on the performance of $\hat{\pi}$, we let $d = \mu_2 - \mu_1$ and perform the simulation study for two different values of $d \in \{1, 3\}$. The EM algorithm is used, assuming $\hat{\pi}^{(0)} = 1/2$ and stopping criteria

Table 2.3: $\hat{\pi}_{k,MLE}$ based on SRS (i.e., $n = 1$) and M1-RSS designs, its standard error (in round bracket), MSE [in square bracket] and RE.

	$d = 1$			$d = 3$		
	$\hat{\pi}_{k,MLE}$	RE	iterations	$\hat{\pi}_{k,MLE}$	RE	iterations
$n = 1$	0.8175 (0.0901) [0.0084]	1	50	0.8056 (0.0418) [0.0017]	1	10
$n = 2$	0.8233 (0.0756) [0.0068]	1.235	70	0.8052 (0.0340) (0.0012)	1.416	14
$n = 3$	0.8103 (0.0747) [0.0058]	1.448	72	0.8036 (0.0299) [0.0009]	1.889	17
$n = 4$	0.8026 (0.0754) [0.0055]	1.527	79	0.8083 (0.0287) [0.0008]	2.213	20
$n = 5$	0.8072 (0.0661) [0.0044]	1.909	88	0.8095 (0.0279) [0.0008]	2.213	21

$|\hat{\pi}^{(p+1)} - \hat{\pi}^{(p)}| < 10^{-5}$. Table 2.3 provides ML estimates and their corresponding standard errors (presented in parenthesis) and mean square errors (MSE) (presented in brackets) under SRS and M1-RSS with different set sizes. Using Basford et al. (1997), the estimates of standard errors and biases (used for MSE's) are obtained via a bootstrap with $b = 100$ replications.

We also calculated the observed relative efficiency of the ML estimators $\hat{\pi}_{n,MLE}$ of π under SRS (i.e., $\hat{\pi}_{1,MLE}$) and RSS-based designs (i.e., $\hat{\pi}_{n,MLE}$, $n \geq 2$) using

$$RE(\hat{\pi}_{n,MLE}, \hat{\pi}_{1,MLE}) = \frac{1/MSE(\hat{\pi}_{n,MLE})}{1/MSE(\hat{\pi}_{1,MLE})} = \frac{MSE(\hat{\pi}_{1,MLE})}{MSE(\hat{\pi}_{n,MLE})}.$$

Table 2.3 presents the values of the relative efficiencies for different set sizes when $d \in \{1, 3\}$. The results indicate that ML estimates of π under M1-RSS design are more efficient than their corresponding estimators under SRS and the relative efficiency is an increasing function of n. In addition, when two components are

Table 2.4: $\hat{\pi}_{k,MLE}$ based on SRS (i.e., $n = 1$) and M2-RSS designs, their standard error (in round bracket), MSE [in square bracket] and RE.

	$d = 1$			$d = 3$		
	$\hat{\pi}_{k,MLE}$	RE	iterations	$\hat{\pi}_{k,MLE}$	RE	iterations
$n = 1$	0.8175 (0.0901) [0.0084]	1	50	0.8056 (0.0418) [0.0017]	1	10
$n = 2$	0.8287 (0.0746) [0.0063]	1.333	43	0.8234 (0.0374) [0.0017]	1	7
$n = 3$	0.8016 (0.0754) [0.0059]	1.423	27	0.8044 (0.0375) [0.0015]	1.133	5
$n = 4$	0.8069 (0.0663) [0.0048]	1.75	25	0.8189 (0.0370) [0.0014]	1.214	5
$n = 5$	0.8111 (0.0629) [0.0040]	2.1	23	0.8061 (0.0367) [0.0014]	1.214	4

separated (i.e., $d = 3$) the performance of the M1-RSS design for estimating π is much better than that of SRS. We obtained similar results under M2-RSS design as they are shown in Table 2.4. However, the efficiency of ML estimators under M2-RSS design reduces slightly as the distance between components increases.

2.4.2 Simulation study 2

In this simulation study, the performance of ML estimates of all parameters of the mixture model using M1-RSS, M2-RSS and SRS designs are investigated. To this end, the underlying distribution is chosen to be a homoscedastic mixture of two univariate normal distributions. To investigate the effect of the distance between two components of the model on parameter estimation, we generated samples of size $nN = 300$ from two mixture of normal densities of the form (2.45) with $\mathbf{\Psi}_1 = (\pi, \mu_1, \mu_2, \sigma) = (0.4, -2, 1, 1)$, i.e. Model 1 and $\mathbf{\Psi}_2 = (0.4, -1, 1, 1.5)$, i.e. Model 2 using SRS, M1-RSS and M2-RSS with $n = 5$.

We examined three different methods for setting the initial values of $\boldsymbol{\Psi}$ consisting of (a) fixed initial values method, (b) Finch's method, and (c) the method of moments (see Finch et al., 1989 and Karlis and Xekalaki, 2003). The stopping criteria here is $\|\hat{\boldsymbol{\Psi}}^{(p+1)} - \hat{\boldsymbol{\Psi}}^{(p)}\|_\infty < 10^{-5}$. The bias, standard error and MSE are used as three performance measures for each estimator. These measures are obtained via the bootstrap method of Basford et al. (1997) with $b = 100$ replications and then this procedure was repeated 10 times. The mean and standard errors of these measures under fixed initial values, Finch's method and the method of moments are respectively reported in Tables 2.5, 2.6 and 2.7. We also calculated the observed relative efficiency of the estimators based on RSS design compared with their SRS competitors using the ratio of the average of their MSE's. We used $\boldsymbol{\Psi}_1^{(0)} = (\pi^{(0)}, \mu_1^{(0)}, \mu_2^{(0)}, \sigma^{(0)}) = (0.3, -1.95, 0.95, 0.95)$ for Model 1 and $\boldsymbol{\Psi}_2^{(0)} = (0.3, -0.95, 0.95, 1.40)$ for Model 2 as initial values for the fixed initial values method. The initial values for Finch's method correspond to case (v) in Karlis and Xekalaki (2003). To obtain these initial values, $\pi^{(0)}$ is chosen to be $1/2$ and observations are ordered; then we used the first half of the data to form the first class and the second half to form the second class. Now, the sample means of these two classes (denoted by $\mu_j^{(0)}$ $j = 1, 2$) are considered as initial values of the means of the components. The initial value for the common standard deviation is calculated by $\hat{\sigma}_0 = \sqrt{s^2 - s_0^2}$, where $s_0^2 = \frac{1}{2}\{(\bar{x} - \mu_1^{(0)})^2 + (\bar{x} - \mu_2^{(0)})^2\}$, and s^2 is the sample variance. For the method of moments the initial values are obtained via the method developed by Furman and Lindsay (1994) by simply treating RSS data as SRS.

According to the results of our simulation study, Model 1, for all designs, converged with 100% of times; however, under Model 2, rates of convergence for each method of selecting initial values (a), (b) and (c) are, respectively (SRS: 99%, 97%, 99%), (M1-RSS: 95%, 96%, 95%) and (M2-RSS: 100%, 100%, 99%). The mean number of iterations required for the convergence under initial values (a), (b) and (c) for Model 1 are (SRS: 19, 24, 22), (M1-RSS: 46, 44, 30) and (M2-RSS: 4, 7, 5), respectively. Similarly for Model 2 these values are given by (SRS: 432, 837, 249), (M1-RSS: 776, 501, 738) and (M2-RSS: 50, 47, 40), respectively.

As seen in Tables 2.5, 2.6 and 2.7, for all initial values, RSS-based estimators perform significantly better than their SRS-based competitors in estimating the parameters of the model (in terms of both the bias and the standard error). Comparing the M1-RSS and M2-RSS designs, it is evident that the M2-RSS proposes more efficient estimators for component parameters of the model than M1-RSS. This is due to the fact that the M2-RSS design has a more informative assumption that $x_{(r)i}$ is indeed the r-th order statistic of one of the components of the model. It is also seen that the performance of RSS-based estimators is better than their SRS competitors under both models. In other words, when two components of the model are close together (i.e., distinguishing between two components is difficult) RSS-based estimators performed better for estimating the parameters than their SRS counterparts. For instance, from Table 2.7, one can easily observe that, using the method of moments for initial values, the relative efficiency of M2-RSS to SRS in estimating the mixing proportion is 5.88; in other words, each observation in M1-RSS design has as much information as roughly 6 SRS observations. Moreover, under Model 1, it is seen that the performance of M1-RSS and SRS in estimating the mixing proportion is slightly better than that of M2-RSS; however, M2-RSS still is very good at estimating parameters of the component densities.

Table 2.5: The (average) and [standard error] of the Bias, standard error SE and MSE of ML estimators of $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$, based on fixed's initial values method.

Techniques		Model 1				Model 2			
		π	μ_1	μ_2	σ	π	μ_1	μ_2	σ
SRS	Bias	(0.0058)	(-0.0114)	(-0.0153)	(-0.0348)	(0.0300)	(-0.1301)	(0.1900)	(-0.1527)
		[0.0323]	[0.0489]	[0.0426]	[0.0508]	[0.0757]	[0.2262]	[0.2351]	[0.1116]
	SE	(0.0344)	(0.1148)	(0.0928)	(0.0502)	(0.1168)	(0.5224)	(0.3337)	(0.1265)
		[0.0027]	[0.0195]	[0.0102]	[0.0053]	[0.04371]	[0.3280]	[0.1247]	[0.03087]
	MSE	(0.0021)	(0.0158)	(0.0105)	(0.0060)	(0.0214)	(0.4328)	(0.2112)	(0.0514)
		[0.0008]	[0.0060]	[0.0021]	[0.0034]	[0.0123]	[0.4451]	[0.1092]	[0.0277]
M1-RSS	Bias	(-0.0136)	(-0.0455)	(-0.0051)	(-0.0217)	(0.0035)	(-0.1933)	(0.1024)	(-0.1223)
		[0.0104]	[0.0914]	[0.0402]	[0.0299]	[0.0642]	[0.2788]	[0.1940]	[0.1150]
	SE	(0.0251)	(0.1021)	(0.0715)	(0.0468)	(0.0952)	(0.4770)	(0.3053)	(0.1335)
		[0.0022]	[0.0078]	[0.0067]	[0.0052]	[0.0234]	[0.2249]	[0.1048]	[0.0356]
	MSE	(0.0009)	(0.0200)	(0.0066)	(0.0035)	(0.0132)	(0.3804)	(0.1475)	(0.0458)
		[0.00002]	[0.0106]	[0.0023]	[0.0007]	[0.0086]	[0.4347]	[0.1382]	[0.0214]
M2-RSS	Bias	(0.0058)	(-0.0332)	(0.0011)	(-0.0035)	(-0.0003)	(0.0448)	(-0.0214)	(-0.0051)
		[0.0303]	[0.0391]	[0.0447]	[0.0348]	[0.0224]	[0.1098]	[0.0812]	[0.1036]
	SE	(0.0273)	(0.0534)	(0.0434)	(0.0343)	(0.0578)	(0.1452)	(0.1052)	(0.0690)
		[0.0019]	[0.0047]	[0.0033]	[0.0029]	[0.0097]	[0.0217]	[0.0143]	[0.0076]
	MSE	(0.0016)	(0.0053)	(0.0037)	(0.0022)	(0.0038)	(0.0344)	(0.0176)	(0.0145)
		[0.0013]	[0.0024]	[0.0019]	[0.0017]	[0.0014]	[0.0153]	[0.0070]	[0.0114]

Table 2.6: The (average) and [standard error] of the Bias, standard error (SE) and MSE of ML estimators of $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$, based on Finch's method for initial values.

Techniques		Model 1				Model 2			
		π	μ_1	μ_2	σ	π	μ_1	μ_2	σ
SRS	Bias	(0.0039)	(0.0797)	(-0.0053)	(0.0104)	(-0.0037)	(-0.3310)	(0.1118)	(-0.0659)
		[0.0301]	[0.0823]	[0.0593]	[0.0458]	[0.1352]	[0.5036]	[0.3206]	[0.0833]
	SE	(0.0369)	(0.1242)	(0.1005)	(0.0547)	(0.1483)	(0.5718)	(0.4828)	(0.1316)
		[0.0037]	[0.0160]	[0.0133]	[0.0032]	[0.0482]	[0.1438]	[0.2187]	[0.0165]
	MSE	(0.0022)	(0.0281)	(0.0134)	(0.0050)	(0.0405)	(0.6836)	(0.3812)	(0.0281)
		[0.0015]	[0.0182]	[0.0055]	[0.0024]	[0.0184]	[0.5274]	[0.3646]	[0.0150]
M1-RSS	Bias	(0.0023)	(-0.0250)	(-0.0200)	(-0.0196)	(0.0503)	(-0.0575)	(0.2318)	(-0.1047)
		[0.0314]	[0.1768]	[0.0802]	[0.0634]	[0.0176]	[0.1057]	[0.0795]	[0.0785]
	SE	(0.0266)	(0.1027)	(0.0734)	(0.0480)	(0.1230)	(0.4006)	(0.3557)	(0.1230)
		[0.0039]	[0.0192]	[0.0079]	[0.0051]	[0.0368]	[0.1401]	[0.1404]	[0.0150]
	MSE	(0.0016)	(0.0396)	(0.0116)	(0.0063)	(0.0191)	(0.1916)	(0.2037)	(0.0318)
		[0.0016]	[0.0327]	[0.0062]	[0.0033]	[0.0093]	[0.0995]	[0.1157]	[0.0129]
M2-RSS	Bias	(-0.0097)	(-0.0095)	(-0.0230)	(0.0014)	(0.0185)	(0.0551)	(0.0154)	(-0.0257)
		[0.0222]	[0.0424]	[0.0446]	[0.0284]	[0.0698]	[0.1714]	[0.1477]	[0.0826]
	SE	(0.0277)	(0.0586)	(0.0449)	(0.0343)	(0.0535)	(0.1312)	(0.1087)	(0.0744)
		[0.0018]	[0.0055]	[0.0021]	[0.0021]	[0.0108]	[0.0334]	[0.0130]	[0.0070]
	MSE	(0.0013)	(0.0051)	(0.0043)	(0.0019)	(0.0077)	(0.0477)	(0.0318)	(0.0124)
		[0.0005]	[0.0026]	[0.0023]	[0.0009]	[0.0047]	[0.0240]	[0.0273]	[0.0054]

Table 2.7: The (average) and [standard error] of the Bias, standard error(SE) and MSE of ML estimators of $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$, based on the method of moments for initial values.

Techniques		Model 1				Model 2			
		π	μ_1	μ_2	σ	π	μ_1	μ_2	σ
SRS	Bias	(-0.0089)	(-0.0515)	(-0.0303)	(-0.0202)	(0.0642)	(-0.1611)	(0.3664)	(-0.1502)
		[0.0258]	[0.1249]	[0.1387]	[0.0369]	[0.1157]	[0.3031]	[0.3117]	[0.0699]
	SE	(0.0351)	(0.1238)	(0.0899)	(0.0526)	(0.1140)	(0.4131)	(0.3906)	(0.1190)
		[0.0024]	[0.0113]	[0.0097]	[0.0054]	[0.0315]	[0.1550]	[0.2160]	[0.0204]
	MSE	(0.0019)	(0.0321)	(0.0264)	(0.0044)	(0.0300)	(0.3010)	(0.4163)	(0.0415)
		[0.0006]	[0.0195]	[0.0173]	[0.0016]	[0.0187]	[0.2482]	[0.4602]	[0.0204]
M1-RSS	Bias	(0.0111)	(-0.0272)	(0.0017)	(0.0107)	(-0.0193)	(-0.3288)	(0.1430)	(-0.1248)
		[0.0218]	[0.1009]	[0.0630]	[0.0535]	[0.0725]	[0.2830]	[0.1574]	[0.0847]
	SE	(0.0267)	(0.1040)	(0.0832)	(0.0503)	(0.1110)	(0.4102)	(0.4044)	(0.1171)
		[0.0031]	[0.0125]	[0.0090]	[0.0047]	[0.0408]	[0.1605]	[0.2503]	[0.0191]
	MSE	(0.0012)	(0.0208)	(0.0105)	(0.0052)	(0.0189)	(0.3717)	(0.2627)	(0.0361)
		[0.0003]	[0.0203]	[0.0050]	[0.0032]	[0.0117]	[0.1965]	[0.2704]	[0.0183]
M2-RSS	Bias	(-0.0067)	(-0.0004)	(0.0229)	(-0.0160)	(0.0192)	(0.0090)	(0.0819)	(-0.0591)
		[0.0394]	[0.0548]	[0.0521]	[0.0186]	[0.0415]	[0.1760]	[0.1070]	[0.0425]
	SE	(0.0297)	(0.0535)	(0.0444)	(0.0338)	(0.0547)	(0.1353)	(0.3003)	(0.0719)
		[0.0014]	[0.0055]	[0.0055]	[0.0028]	[0.0163]	[0.0284]	[0.4115]	[0.0125]
	MSE	(0.0023)	(0.0056)	(0.0049)	(0.0017)	(0.0051)	(0.0470)	(0.2596)	(0.0104)
		[0.0017]	[0.0038]	[0.0045]	[0.0008]	[0.0021]	[0.0338]	[0.5060]	[0.0051]

Chapter 3

Fisher Information of RSS Data from FMM

The Fisher information (FI) matrix is a measure to quantify the amount of information that observations from a statistical experiment designed to investigate a parametric model carry about the unknown parameters of interest. The FI is also useful, for example, to obtain a matrix lower bound for the covariance matrix of unbiased estimators of the vector of unknown parameters of the model, or to study the asymptotic properties of the ML estimators of the unknown parameters. In Chapter 2, we considered the problem of ML estimation of the vector of unknown parameters $\mathbf{\Psi}$ for the FMM (1.1) using RSS data. We showed that RSS-based estimators are more efficient than their SRS counterparts and explained this using the structural differences between SRS and RSS. In this chapter, we obtain the FI of RSS data from FMM (1.1) and show that it is bigger than that of SRS data. In SRS, observations are $i.i.d$ and each of them represents a typical value from the underlying model. However, there is no additional structure imposed on their relationship to one another. But in RSS, additional information and structure has been provided through the ranking process. The samples obtained through RSS are independent (judgment) order statistics and each of them provides information about different aspects of the distribution. Indeed, it is this extra information provided by the ranking and the independence of the resulting order statistics that makes procedures based on RSS to be more efficient than their counterparts based on SRS data with the same number of measurements. So, it would be of interest to quantify the amount of information that RSS data carry about the parameters of the underlying FMM and compare it with that of SRS data. It should be noted that a theoretical analysis of the efficiency of RSS with respect to SRS was done by Stokes (1995) for the case of the location-scale family of distributions and by Chen (2000) and Barabesi and El-Sharaawi (2001) for multi-parameter family of distributions.

In this chapter, we focus on three types of ranked set samples from the FMM (1.1). These three types correspond to different variations of RSS data considered in Chapter 2. The first case referred to Type-M0 or incomplete RSS data where the data are obtained from the whole FMM and no more information about the component of origin of each observation is available. For Type-M1 or complete M1-RSS data, the observations and their latent variables are known. In other words, we have access to the observations and the component of origin of each observation $x_{(r)i}$ and the number of observations smaller than and larger than $x_{(r)i}$ that are used from the j-th component of the FMM (1.1) in the ranking process. Finally, we consider Type-M2 or complete M2-RSS data where each $x_{(r)i}$ and its component of origin are known to be obtained from a set consisting of units selected from only one of the components of the mixture model. Examples of these types of data under SRS can be found in Titterington et al. (1985) as well as Hosmer Jr (1973). Since there is no ranking involved in the SRS scheme, the SRS counterparts of both Type-M1 and Type-M2 RSS data will be the same. Adapting the notation of Titterington et al. (1985), we refer to the SRS counterpart of Type-M1 and Type-M2 RSS data as Type-C SRS data. We provide a general expression for the FI matrix for each type of RSS data. The FI matrices are decomposed into the sum of the SRS FI matrix and semi-positive definite matrices. Hence, the superiority of each type of RSS protocol over their associated SRS scheme is generally established in parametric inference for FMMs.

The outline of the chapter is as follows. Section 3.1 develops the FI matrices for three types of RSS data under the perfect ranking assumption. We show that the FI contained in each type of RSS data is larger than the FI contained in its SRS counterpart. We obtain some interesting results concerning the FI about mixing proportions contained in RSS data relative to the corresponding FI contained in SRS data. In particular, we find cases where the ratio of the determinant of the RSS and SRS FI matrices neither depends on the component densities nor the number of components of the underlying mixture model and is always equal to the set size. In Section 3.2, we consider the FI matrices of imperfect ranked set samples and explore the effect of ranking error on the amount of information contained in each type of RSS data. In Section 3.3, we make a comparison among different types of perfect RSS using the missing information criterion. Section 3.4 is devoted to several numerical studies using a mixture of two exponential distributions.

3.1 FI in perfect RSS from FMM

In this section, we study the FI content of three RSS data structures from the FMM (1.1) under the perfect ranking assumption. The impact of imperfect ranking will be studied in Section 3.2. The considered RSS data structures are as follow:

Type-M0: This is incomplete RSS data, where each $x_{(r)i}, i = 1, \ldots, N; r = 1, \ldots, n$ is obtained from a set consisting of observations from the whole FMM. The joint pdf of the resulting sample is given by (2.5) which means that Type-M0 RSS is essentially the incomplete variation of M1-RSS considered in Chapter 2. The FI matrix for Type-M0 perfect RSS data will be denoted by $\mathbb{I}_{M0,RSS}(\boldsymbol{\Psi})$.

Type-M1: This is the complete M1-RSS data, where we observe $(x_{(r)i}, \mathbf{z}_i^{(r)}, \mathbf{w}_i^{(r)}, \mathbf{v}_i^{(r)})$, $i = 1, \ldots, N; r = 1, \ldots, n$ from the FMM (1.1). The joint pdf of the complete M1-RSS data is given by (2.14) with latent variables defined in Section 2.2.1. The FI matrix for Type-M1 perfect RSS data is also denoted by $\mathbb{I}_{M1,RSS}(\boldsymbol{\Psi})$.

Type-M2: This is the complete M2-RSS data, where we observe $(x_{(r)i}, z_i^{(r)})$ as defined in Section 2.2.2. So the joint pdf of the sample is given by (2.24). The FI matrix for Type-M2 perfect RSS data is denoted by $\mathbb{I}_{M2,RSS}(\boldsymbol{\Psi})$.

Remark 3.1. *The SRS counterparts of Types M0, M1 and M2 RSS data can easily be obtained by setting the set size $n = 1$ in (2.5), (2.14) and (2.24), respectively. This, however, results in the same pdf for the SRS counterparts of Types M1 and M2 RSS data. Adapting the notation of Titterington et al. (1985), we refer to this type of SRS data as Type-C SRS data. To be more specific, Type-C SRS data consists of observations $(\mathbf{X}, \mathbf{Z})$ from the FMM (1.1), with the joint pdf (2.30) where $\mathbf{Z_i}$ is the* *usual indicator random vector whose realization identifies the component of origin of x_i. The FI matrix for Type-C SRS data is denoted by $\mathbb{I}_{C,SRS}(\mathbf{\Psi})$.*

Throughout the chapter, we use the following notation.

1. We use $A \geq 0$ and $A \geq B$ to indicate that A and $A - B$ are non-negative definite matrices.
2. We let $\phi_r(\lambda) = (r-1)I(\lambda = 0) + (n-r)I(\lambda = 1)$ with $\lambda \in \{0, 1\}$, $r = 1, \ldots, n$ and where I is the usual indicator function.
3. Under the usual regularity conditions Chen et al. (2004), we use $\mathbb{I}(\mathbf{\Psi}) = -\mathbb{E}[D^2_{\mathbf{\Psi}} \log L(\mathbf{\Psi})]$ to denote the FI matrix, where $D^l_{\mathbf{\Psi}}$ refers to the l-th derivative of the log-likelihood function with respect to $\mathbf{\Psi}$ with $D^1_{\mathbf{\Psi}} = D_{\mathbf{\Psi}}$.

3.1.1 FI in Type-M0 perfect RSS data

Our framework here closely resembles the one introduced by Chen (2000) as well as Chen et al. (2004), and some of the following preliminary results are reproduced here for the sake of completeness. We first obtain $\mathbb{I}_{M0,RSS}(\mathbf{\Psi})$, the FI matrix of Type-M0 RSS data from the FMM (1.1), and compare it with its counterpart under SRS. Using (2.5), the log-likelihood function of $\mathbf{\Psi}$ under Type-M0 RSS design is $\log L_{0,RSS}(\mathbf{\Psi}) \propto \log L_{0,SRS}(\mathbf{\Psi}) + \Gamma_{M0}(\mathbf{\Psi})$, where

$$\Gamma_{M0}(\mathbf{\Psi}) = \sum_{i=1}^{N} \sum_{r=1}^{n} \sum_{\lambda=0}^{1} \phi_r(\lambda) \log[\lambda + (1 - 2\lambda) F(x_{(r)i}; \mathbf{\Psi})].$$

It is easy to see that

$$D^2_{\mathbf{\Psi}}\Gamma_{M0}(\mathbf{\Psi}) = \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{\lambda=0}^{1}(-1)^{\lambda}\phi_r(\lambda)$$

$$\times\left\{\frac{D^2_{\mathbf{\Psi}}F(x_{(r)i};\mathbf{\Psi})}{\lambda+(1-2\lambda)F(x_{(r)i};\mathbf{\Psi})} - \frac{D_{\mathbf{\Psi}}F(x_{(r)i};\mathbf{\Psi})[D_{\mathbf{\Psi}}F(x_{(r)i};\mathbf{\Psi})]^{\top}}{[\lambda+(1-2\lambda)F(x_{(r)i};\mathbf{\Psi})]^2}\right\}, \tag{3.1}$$

and, $\mathbb{I}_{M0,RSS}(\mathbf{\Psi}) = \mathbb{I}_{M0,SRS}(\mathbf{\Psi}) - \mathbb{E}[D^2_{\mathbf{\Psi}}\Gamma_{M0}(\mathbf{\Psi})]$.

We need the following result whose proof can be obtained in Chen (2000) as well as Chen et al. (2004).

Lemma 3.1. *Let $Y_r = X_{(r)i}$, $r = 1,\dots,n$, $i = 1,\dots,N$, with pdf $f^{(r)}(\cdot;\mathbf{\Psi})$ be the r-th order statistic of an SRS sample of size n from $f(x;\mathbf{\Psi})$ with cdf $F(x;\mathbf{\Psi})$. Then, for any function $G(\cdot)$,*

$$\mathbb{E}\left[\sum_{r=1}^{n}\frac{\phi_r(\lambda)\,G(Y_r)}{\lambda+(1-2\lambda)F(Y_r;\mathbf{\Psi})}\right] = n(n-1)\mathbb{E}[G(X)],$$

subject to the existence of the expectations, when $\lambda = 0$ or 1.

Theorem 3.1. *The FI matrix of Type-M0 RSS data about $\mathbf{\Psi}$ in the FMM (1.1) is given by*

$$\mathbb{I}_{M0,RSS}(\mathbf{\Psi}) = \mathbb{I}_{M0,SRS}(\mathbf{\Psi}) + Nn(n-1)\mathbb{E}\left(\frac{D_{\mathbf{\Psi}}F(X;\mathbf{\Psi})[D_{\mathbf{\Psi}}F(X;\mathbf{\Psi})]^{\top}}{F(X;\mathbf{\Psi})\bar{F}(X;\mathbf{\Psi})}\right)$$

$$= \mathbb{I}_{M0,SRS}(\mathbf{\Psi}) + Nn(n-1)\Xi_{M0}(\mathbf{\Psi}), \tag{3.2}$$

where

$$\mathbb{I}_{M0,SRS}(\mathbf{\Psi}) = Nn\,\mathbb{E}\left(\frac{D^2_{\mathbf{\Psi}}f(X;\mathbf{\Psi})}{f(X;\mathbf{\Psi})}\right) + Nn\,\mathbb{E}\left(\frac{D_{\mathbf{\Psi}}f(X;\mathbf{\Psi})[D_{\mathbf{\Psi}}f(X;\mathbf{\Psi})]^{\top}}{[f(X;\mathbf{\Psi})]^2}\right),$$

is the FI matrix of Type-M0 SRS data of the same size and $\Xi_{M0}(\mathbf{\Psi})$ is a non-negative definite matrix.

Proof. Using Lemma 3.1 with $G(u) = D^2_{\mathbf{\Psi}} F(u; \mathbf{\Psi})$, we get

$$\mathbb{E}\left[\sum_{i=1}^{N}\sum_{r=1}^{n}(r-1)\frac{D^2_{\mathbf{\Psi}} F(X_{(r)i}; \mathbf{\Psi})}{F(X_{(r)i}; \mathbf{\Psi})}\right] = \mathbb{E}\left[\sum_{i=1}^{N}\sum_{r=1}^{n}(n-r)\frac{D^2_{\mathbf{\Psi}} F(X_{(r)i}; \mathbf{\Psi})}{\bar{F}(X_{(r)i}; \mathbf{\Psi})}\right].$$

Similarly, in Lemma 3.1, choosing $G(u)$ to be $G_\lambda(u) = \frac{D_{\mathbf{\Psi}} F(u;\mathbf{\Psi})[D_{\mathbf{\Psi}} F(u;\mathbf{\Psi})]^\top}{\lambda+(1-2\lambda)F(u;\mathbf{\Psi})}$, $\lambda \in \{0, 1\}$, we obtain

$$\mathbb{E}\left[\sum_{i=1}^{N}\sum_{r=1}^{n}\frac{\phi_r(\lambda) G_\lambda(X_{(r)i})}{\lambda + (1-2\lambda)F(X_{(r)i}; \mathbf{\Psi})}\right] = Nn(n-1)\mathbb{E}[G_\lambda(X)],$$

where $\phi_r(\lambda) = (r-1)I(\lambda = 0) + (n-r)I(\lambda = 1)$. Now, the result follows upon taking the expectation of (3.1). $\square$

Theorem 3.1 shows that the FI of Type-M0 RSS sample about $\mathbf{\Psi}$ is not smaller than the FI of Type-M0 SRS sample of the same size.

3.1.2 FI in Type-M1 perfect RSS data

To obtain the FI matrix of Type-M1 RSS data about $\mathbf{\Psi}$ we need to work with the joint pdf of $X_{(r)i}, i = 1, \ldots, N; r = 1, \ldots, n$ and their latent variables $(\mathbf{Z}_i^{(r)}, \mathbf{W}_i^{(r)}, \mathbf{V}_i^{(r)})$ as defined in Section 2.2.1. Considering the joint pdf of Type-M1 RSS data (2.14), we have the following log-likelihood function

$$l_{M1}(\mathbf{\Psi}) \propto l_{C,SRS}(\mathbf{\Psi}) + \Gamma_{M1}(\mathbf{\Psi}), \tag{3.3}$$

where

$$l_{C,SRS}(\mathbf{\Psi}) = \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M} Z_{ij}^{(r)}\left\{\log \pi_j + \log f_j(x_{(r)i};\theta_j)\right\}, \tag{3.4}$$

and

$$\Gamma_{M1}(\mathbf{\Psi}) = \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M} \log \pi_j \left\{W_{ij}^{(r)} + V_{ij}^{(r)}\right\}$$
$$+ \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\left\{V_{ij}^{(r)} \log \bar{F}_j(x_{(r)i};\theta_j) + W_{ij}^{(r)} \log F_j(x_{(r)i};\theta_j)\right\}. \tag{3.5}$$

Before pursuing, we need the following additional technical result.

Lemma 3.2. *Let $(X_{(r)i}, \mathbf{Z}_i^{(r)}, \mathbf{W}_i^{(r)}, \mathbf{V}_i^{(r)}), i = 1, \ldots, N$ be Type-M1 RSS data from FMM (1.1), where $\mathbf{Z}_i^{(r)}$, $\mathbf{V}_i^{(r)}$ and $\mathbf{W}_i^{(r)}$ are defined as before. For any $G(\cdot)$, subject to the existence of the expectations, we have*

(i) $\sum_{j=1}^{M}\sum_{r=1}^{n} \mathbb{E}\left(Z_{ij}^{(r)} G(X_{(r)i})\right) = n\sum_{j=1}^{M} \pi_j \mathbb{E}_j\left[G(X)\right],$ *where $\mathbb{E}_j$ indicates the expectation with respect to the j-th component of the mixture.*

(ii) $\sum_{j=1}^{M}\sum_{r=1}^{n} \mathbb{E}\left(\{W_{ij}^{(r)}\}^{1-\lambda}\{V_{ij}^{(r)}\}^{\lambda} G(X_{(r)i})\right) = c_1 \sum_{j=1}^{M} \mathbb{E}\left[G(X)(\lambda + (1-2\lambda)F_j(X;\theta_j))\right]$

where $\lambda \in \{0, 1\}$ and $c_1 = n(n-1)$.

Proof. For part (i), from (2.11) we have

$$\sum_{j=1}^{M}\sum_{r=1}^{n}\mathbb{E}_{\boldsymbol{\Psi}}\left[Z_{ij}^{(r)}G(X_{(r)i})\right]=\sum_{j=1}^{M}\sum_{r=1}^{n}\mathbb{E}_{\boldsymbol{\Psi}}\left[G(X_{(r)i})\mathbb{E}(Z_{ij}^{(r)}|X_{(r)i})\right]$$

$$=\sum_{j=1}^{M}\int\pi_j G(x)f_j(x;\theta_j)\left\{\sum_{r=1}^{n}c_2[F(x;\boldsymbol{\Psi})]^{r-1}[\bar{F}(x;\boldsymbol{\Psi})]^{n-r}\right\}dx$$

$$=n\sum_{j=1}^{M}\pi_j\mathbb{E}_j[G(X)],$$

where $c_2=n\binom{n-1}{r-1}$ and $\mathbb{E}_j$ is the expectation with respect to the pdf of the j-th component of the mixture. For part (ii), when $\lambda=0$, using (2.12), a straightforward calculation shows that

$$\sum_{j=1}^{M}\sum_{r=1}^{n}\mathbb{E}_{\boldsymbol{\Psi}}\left[W_{ij}^{(r)}G(X_{(r)i})\right]=\sum_{j=1}^{M}\sum_{r=1}^{n}\mathbb{E}_{\boldsymbol{\Psi}}\left[G(X_{(r)i})\ \mathbb{E}(W_{ij}^{(r)}|X_{(r)i})\right]$$

$$=n(n-1)\sum_{j=1}^{M}\pi_j\mathbb{E}_{\boldsymbol{\Psi}}\left[G(X)F_j(X;\theta_j)\right],$$

where the last equality is due to Lemma 3.1. A similar proof can be given for the case $\lambda=1$, and hence is omitted. □

Now, we obtain the FI contained in Type-M1 RSS data from the FMM (1.1) and compare it with its SRS counterpart.

Theorem 3.2. *The FI matrix of a Type-M1 RSS sample of size nN about $\boldsymbol{\Psi}$ from the FMM (1.1) is given by*

$$\mathbb{I}_{M1,RSS}(\boldsymbol{\Psi})=\mathbb{I}_{C,SRS}(\boldsymbol{\Psi})-\mathbb{E}[D^2_{\boldsymbol{\Psi}}\Gamma_{M1}(\boldsymbol{\Psi})],\tag{3.6}$$

where

$$\mathbb{E}[D^2_{\mathbf{\Psi}}\Gamma_{M1}(\mathbf{\Psi})] = Nn(n-1)\sum_{j=1}^{M}\pi_j(D^2_{\mathbf{\Psi}}\log\pi_j)$$

$$- Nn(n-1)\sum_{j=1}^{M}\pi_j\mathbb{E}_{\mathbf{\Psi}}\left(\frac{D_{\mathbf{\Psi}}F_j(X;\theta_j)[D_{\mathbf{\Psi}}F_j(X;\theta_j)]^{\top}}{\bar{F}_j(X;\theta_j)F_j(X;\theta_j)}\right).$$

Furthermore, $-E[D^2_{\mathbf{\Psi}}\Gamma_{M1}(\mathbf{\Psi})]$ *is a non-negative definite matrix.*

Proof. Using Lemma 3.2, a straightforward calculation shows that

$$\begin{aligned}
-\mathbb{E}\left[D^2_{\mathbf{\Psi}}l_{C,SRS}(\mathbf{\Psi})\right] = & - Nn\sum_{j=1}^{M}\pi_j[D^2_{\mathbf{\Psi}}\log(\pi_j)] - Nn\sum_{j=1}^{M}\pi_j\mathbb{E}_j\left[\frac{D^2_{\mathbf{\Psi}}f_j(X;\theta_j)}{f_j(X;\theta_j)}\right] \\
& + Nn\sum_{j=1}^{M}\pi_j\mathbb{E}_j\left[\frac{D_{\mathbf{\Psi}}f_j(X;\theta_j)[D_{\mathbf{\Psi}}f_j(X;\theta_j)]^{\top}}{[f_j(X;\theta_j]^2}\right] \\
= & \, \mathbb{I}_{C,SRS}(\mathbf{\Psi}).
\end{aligned} \tag{3.7}$$

Moreover, we observe that $D^2_{\mathbf{\Psi}}\Gamma_{M1}(\mathbf{\Psi})$ can be written as

$$\begin{aligned}
D^2_{\mathbf{\Psi}}\Gamma_{M1}(\mathbf{\Psi}) = & \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\left\{W^{(r)}_{ij}D^2_{\mathbf{\Psi}}\log\pi_j + V^{(r)}_{ij}D^2_{\mathbf{\Psi}}\log\pi_j\right\} \\
& + \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\sum_{\lambda=0}^{1}\Bigg[(-1)^{\lambda}\{W^{(r)}_{ij}\}^{1-\lambda}\{V^{(r)}_{ij}\}^{\lambda} \\
& \times\left\{\frac{D^2_{\mathbf{\Psi}}F_j(x_{(r)i};\theta_j)}{\lambda+(1-2\lambda)F_j(x_{(r)i};\theta_j)} - \frac{D_{\mathbf{\Psi}}F_j(x_{(r)i};\theta_j)[D_{\mathbf{\Psi}}F_j(x_{(r)i};\theta_j)]^{\top}}{[\lambda+(1-2\lambda)F_j(x_{(r)i};\theta_j)]^2}\right\}\Bigg].
\end{aligned}$$

Using the distributions of $\mathbf{W}^{(r)}_{ij}$ and $\mathbf{V}^{(r)}_{ij}$, $r = 1,\ldots,n$, we have $\mathbb{E}\left\{W^{(r)}_{ij}D^2_{\mathbf{\Psi}}\log\pi_j\right\} = \pi_j(r-1)(D^2_{\mathbf{\Psi}}\log\pi_j)$ and $\mathbb{E}\left\{V^{(r)}_{ij}D^2_{\mathbf{\Psi}}\log\pi_j\right\} = \pi_j(n-r)(D^2_{\mathbf{\Psi}}\log\pi_j)$. Using Lemma

3.2, it is easy to show that

$$\sum_{j=1}^{M}\sum_{r=1}^{n}\mathbb{E}\left(\frac{\{W_{ij}^{(r)}\}^{1-\lambda}\{V_{ij}^{(r)}\}^{\lambda}D^2_{\boldsymbol{\Psi}}F_j(X_{(r)i};\theta_j)}{\lambda+(1-2\lambda)F_j(X_{(r)i};\theta_j)}\right)=n(n-1)\sum_{j=1}^{M}\pi_j\mathbb{E}_{\boldsymbol{\Psi}}\left[D^2_{\boldsymbol{\Psi}}F_j(X;\theta_j)\right],$$

and

$$\sum_{j=1}^{M}\sum_{r=1}^{n}\mathbb{E}\left(\frac{\{W_{ij}^{(r)}\}^{1-\lambda}\{V_{ij}^{(r)}\}^{\lambda}D_{\boldsymbol{\Psi}}F_j(X_{(r)i};\theta_j)[D_{\boldsymbol{\Psi}}F_j(X_{(r)i};\theta_j)]^{\top}}{[\lambda+(1-2\lambda)F_j(X_{(r)i};\theta_j)]^2}\right)$$

$$=n(n-1)\sum_{j=1}^{M}\pi_j\mathbb{E}_{\boldsymbol{\Psi}}\left(\frac{D_{\boldsymbol{\Psi}}F_j(X;\theta_j)[D_{\boldsymbol{\Psi}}F_j(X;\theta_j)]^{\top}}{\lambda+(1-2\lambda)F_j(X;\theta_j)}\right).$$

Now, the result follows by taking the second derivative of (3.3) and calculating its expectation using the above equations. □

Theorem 3.2 shows that the FI contained in Type-M1 RSS data about $\boldsymbol{\Psi}$ is not smaller than its counterpart under SRS. In the next result we consider the specific case of a mixture of two known densities and obtain a very interesting result concerning the relative efficiency of Type-M1 RSS data relative to Type-C SRS data in the estimation of the mixing proportion.

Theorem 3.3. *Suppose $f(x;\pi)=\pi f_1(x)+(1-\pi)f_2(x)$ is a finite mixture of two known component densities $f_1(\cdot)$ and $f_2(\cdot)$ where π is unknown. Let $\mathrm{RE}(\pi)=\frac{\mathbb{I}_{M1,RSS}(\pi)}{\mathbb{I}_{C,SRS}(\pi)}$ be the relative efficiency of a Type-M1 perfect RSS of size nN for estimating the mixing proportion π compared with its SRS counterpart with a sample of the same size. Then, under the usual regularity conditions Chen et al. (2004), $\mathrm{RE}(\pi)=n$.*

Proof. When π in the only unknown parameter of the model, the second term on the right-hand side of (3.6) becomes $\mathbb{E}\left[D^2_{\mathbf{\Psi}} l_{C,SRS}(\mathbf{\Psi})\right] = -\frac{Nn(n-1)}{\pi(1-\pi)}$. From (3.7) we have $\mathbb{I}_{C,SRS}(\pi) = \frac{Nn}{\pi(1-\pi)}$. Therefore, using Theorem 3.2 we get

$$\begin{aligned}\mathbb{I}_{M1,RSS}(\pi) &= \mathbb{I}_{C,SRS}(\pi) + \frac{Nn(n-1)}{\pi(1-\pi)} \\ &= \frac{Nn^2}{\pi(1-\pi)} \\ &= n\,\mathbb{I}_{C,srs}(\pi),\end{aligned}$$

which completes the proof. □

Theorem 3.3 shows that the relative efficiency of a Type-M1 perfect RSS of size nN for estimating the mixing proportion π compared with its SRS counterpart with a sample of the same size does not depend on the component densities f_1 and f_2.

Remark 3.2. ***(A modified FI matrix for Type-M1 perfect RSS)*** *As proposed in Section 2.2.3, to reduce the computational burden of calculating the FI matrix for type M1-RSS data and obtaining the ML estimates, we derive a modification of the log-likelihood function $l_{M1}(\mathbf{\Psi})$ by replacing the derivative of the second term on the right side of the expression obtained for $\Gamma_{M1}(\mathbf{\Psi})$ in (3.5) with its expectation. With this modification, it is easy to see that the modified log-likelihood function is*

$$\begin{aligned} l^*_{M1}(\mathbf{\Psi}) \propto \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M} Z_{ij}^{(r)} \left\{\log \pi_j + \log f_j(x_{(r)i};\theta_j)\right\} \\ + \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M} \log \pi_j \left\{W_{ij}^{(r)} + V_{ij}^{(r)}\right\}. \end{aligned} \tag{3.8}$$

Now, the approximate FI matrix for Type-M1 RSS about $\mathbf{\Psi}$ is given by

$$\mathbb{I}^*_{M1,RSS}(\mathbf{\Psi}) = \mathbb{I}_{C,SRS}(\mathbf{\Psi}) - \mathbb{E}[D^2_{\mathbf{\Psi}}\Gamma^*_{M1}(\mathbf{\Psi})], \tag{3.9}$$

where $\mathbb{E}[D^2_{\mathbf{\Psi}}\Gamma^*_{M1}(\mathbf{\Psi})] = Nn(n-1)\sum_{j=1}^{M}\pi_j\, D^2_{\mathbf{\Psi}}\log\pi_j$. Here, again $-\mathbb{E}[D^2_{\mathbf{\Psi}}\Gamma^*_{M1}(\mathbf{\Psi})]$ is a non-negative definite matrix. Hence, the approximate FI contained in Type-M1 RSS data is larger that its SRS counterpart.

The method proposed in Remark 3.2 can be used to obtain a modified relative efficiency of Type-M1 RSS data for estimating the parameters $\mathbf{\Psi}$ of the FMM (1.1) relative to its counterpart under SRS.

Theorem 3.4. *Let $RE^*(\mathbf{\Psi}) = \frac{det\{\mathbb{I}^*_{M1,RSS}(\mathbf{\Psi})\}}{det\{\mathbb{I}_{C,SRS}(\mathbf{\Psi})\}}$ be an approximation to the RE relative efficiency of a Type-M1 RSS data of size Nn relative to its SRS counterpart for estimating the parameters $\mathbf{\Psi}$ of the FMM (1.1). Then, $RE^*(\mathbf{\Psi}) = n$.*

Proof. Let $c_1 = Nn(n-1)$. It is easy to see that

$$\mathbb{E}[D^2_{\mathbf{\Psi}}\Gamma^*_{M1}(\mathbf{\Psi})] = c_1\sum_{j=1}^{M}\pi_j(D^2_{\mathbf{\Psi}}\log\pi_j) = c_1\begin{pmatrix} \mathbb{I}(\boldsymbol{\pi}) & \underline{0} \\ \underline{0}^t & \mathbf{0} \end{pmatrix}, \tag{3.10}$$

where $\mathbf{0}$ is a square matrix of zeros of size equal to the number of component parameters and $\underline{0}$ is a vector of zeros. Also, $\mathbb{I}(\boldsymbol{\pi}) = \sum_{j=1}^{M}\pi_j(D^2_{\boldsymbol{\pi}}\log\pi_j)$. On the other hand, as presented earlier in (3.7), one gets

$$\mathbb{I}_{C,SRS}(\mathbf{\Psi}) = Nn\begin{pmatrix} \mathbb{I}(\boldsymbol{\pi}) & \underline{0} \\ \underline{0}^t & \mathbb{I}(\boldsymbol{\xi}) \end{pmatrix}, \tag{3.11}$$

where

$$\mathbb{I}(\boldsymbol{\xi}) = -Nn\sum_{j=1}^{M}\pi_j\mathbb{E}_j\left[\frac{D^2_{\boldsymbol{\Psi}}f_j(X;\theta_j)}{f_j(X;\theta_j)}\right] + Nn\sum_{j=1}^{M}\pi_j\mathbb{E}_j\left[\frac{D_{\boldsymbol{\Psi}}f_j(X;\theta_j)[D_{\boldsymbol{\Psi}}f_j(X;\theta_j)]^\top}{[f_j(X;\theta_j]^2}\right].$$

Now from (3.9), (3.10) and (3.11), we have

$$\mathbb{I}^*_{M1,RSS}(\boldsymbol{\Psi}) = \begin{pmatrix} Nn\,\mathbb{I}(\boldsymbol{\pi}) & \underline{0} \\ \underline{0}^t & Nn\,\mathbb{I}(\boldsymbol{\xi}) \end{pmatrix} + \begin{pmatrix} c_1\,\mathbb{I}(\boldsymbol{\pi}) & \underline{0} \\ \underline{0}^t & \underline{0} \end{pmatrix} = \begin{pmatrix} Nn^2\,\mathbb{I}(\boldsymbol{\pi}) & \underline{0} \\ \underline{0}^t & Nn\,\mathbb{I}(\boldsymbol{\xi}) \end{pmatrix}.$$

Finally,

$$\mathrm{RE}(\boldsymbol{\Psi}) = \frac{det\{\mathbb{I}^*_{M1,RSS}(\boldsymbol{\Psi})\}}{det\{\mathbb{I}_{C,SRS}(\boldsymbol{\Psi})\}} = \frac{N^2n^3\,det\{\mathbb{I}(\boldsymbol{\pi})\}\times det\{\mathbb{I}(\boldsymbol{\xi})\}}{N^2n^2\,det\{\mathbb{I}(\boldsymbol{\pi})\}\times det\{\mathbb{I}(\boldsymbol{\xi})\}} = n,$$

and this completes the proof. □

Theorem 3.4 shows that the efficiency of the modified Type-M1 RSS data relative to its SRS counterpart in the estimation of the parameters $\boldsymbol{\Psi}$ of the FMM (1.1) is independent of the component densities and the number of components of the model.

3.1.3 FI in Type-M2 perfect RSS data

Suppose $(\mathbf{X}, \mathbf{Z})$ is a Type-M2 complete ranked set sample of size Nn from the FMM (1.1). For each observation $x_{(r)i}, i = 1, \ldots, N; r = 1, \ldots, n$, $\mathbf{Z}^{(r)}_i = (Z^{(r)}_{i1}, \ldots, Z^{(r)}_{iM})$ is an M-dimensional vector defined as in Section 2.2.2 and $\mathbf{Z}^{(1)}_1, \ldots, \mathbf{Z}^{(n)}_N$ are *i.i.d.* samples from a Mult$(1, \boldsymbol{\pi})$ distribution. From (2.25), the log-likelihood function is written as

$$l_{M2,RSS}(\boldsymbol{\Psi}) \propto l_{C,SRS}(\boldsymbol{\Psi}) + \Gamma_{M2}(\boldsymbol{\Psi}), \tag{3.12}$$

where $\Gamma_{M2}(\mathbf{\Psi}) = \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\sum_{\lambda=0}^{1} Z_{ij}^{(r)}\phi_r(\lambda)\log[\lambda + (1-2\lambda)F_j(x_{(r)i};\theta_j)]$. Using Lemma 3.2, since $Z_{ij}^{(r)}|X_{(r)i} = x_{(r)i} \sim \text{Bin}\left(1, \frac{\pi_j f_j^{(r:n)}(x_{(r)i};\theta_j)}{f^{(r:n)}(x_{(r)i};\mathbf{\Psi})}\right)$, we can easily show that $\mathbb{I}_{C,SRS}(\mathbf{\Psi}) = -\mathbb{E}[D^2_{\mathbf{\Psi}} l_{C,SRS}(\mathbf{\Psi})]$. On the other hand, taking the second derivatives of Γ_{M2} with respect to $\mathbf{\Psi}$ yields

$$
\begin{aligned}
D^2_{\mathbf{\Psi}}\Gamma_{M2}(\mathbf{\Psi}) = &\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\sum_{\lambda=0}^{1}(-1)^{\lambda}\phi_r(\lambda)Z_{ij}^{(r)} \\
&\times\left\{\frac{D^2_{\mathbf{\Psi}}F_j(x_{(r)i};\theta_j)}{\lambda+(1-2\lambda)F_j(x_{(r)i};\theta_j)} - \frac{D_{\mathbf{\Psi}}F_j(x_{(r)i};\theta_j)[D_{\mathbf{\Psi}}F_j(x_{(r)i};\theta_j)]^{\top}}{[\lambda+(1-2\lambda)F_j(x_{(r)i};\theta_j)]^2}\right\}.
\end{aligned}
$$

We pursue with a very useful lemma which is necessary to obtain some of our theoretical results.

Lemma 3.3. *Let $(X_{(r)i}, \mathbf{Z}_i^{(r)})$, $r = 1,\ldots,n$, $i = 1,\ldots,N$, be Type-M2 RSS data obtained from the FMM (1.1). Further, let $f_j(\cdot,\theta_j)$ and $F_j(\cdot,\theta_j)$, respectively, represent the pdf and cdf of the j-th component of (1.1), $j = 1,\ldots,M$. Then, for any function $G(\cdot)$ (subject to the existence of the expectations), we have*

$$
\mathbb{E}\left[\sum_{j=1}^{M}\sum_{r=1}^{n}\frac{\phi_r(\lambda)Z_{ij}^{(r)}\,G(X_{(r)i})}{\lambda+(1-2\lambda)F_j(X_{(r)i};\theta_j)}\right] = n(n-1)\sum_{j=1}^{M}\pi_j\mathbb{E}_j[G(X)], \quad \lambda\in\{0,1\},
$$

where $\mathbb{E}_j$ denotes the expectation with respect to $f_j(\cdot,\theta_j)$, $j = 1,\ldots,M$.

Proof. To show the result, first note that $\phi_r(\lambda)\binom{n-1}{r-1} = (n-1)\binom{n-2}{r+\lambda-2}$, $\lambda\in\{0,1\}$. Also,

$$
\frac{\phi_r(\lambda)f_j^{(r)}(x;\theta_j)}{\lambda+(1-2\lambda)F_j(x;\theta_j)} = n(n-1)\binom{n-2}{r+\lambda-2}f(x;\theta_j)[F_j(x;\theta_j)]^{r+\lambda-2}[\bar{F}_j(x;\theta_j)]^{n-r-\lambda}. \tag{3.13}
$$

The result then follows from (3.13), Lemma 3.1, and

$$\mathbb{E}_{\boldsymbol{\Psi}}\left[\sum_{j=1}^{M}\sum_{r=1}^{n}\frac{\phi_r(\lambda)Z_{ij}^{(r)}\,G(X_{(r)i})}{\lambda+(1-2\lambda)F_j(X_{(r)i};\theta_j)}\right]$$

$$=\mathbb{E}_{\boldsymbol{\Psi}}\left[\sum_{j=1}^{M}\sum_{r=1}^{n}\frac{\phi_r(\lambda)\,G(X_{(r)i})}{\lambda+(1-2\lambda)F_j(X_{(r)i};\theta_j)}E[Z_{ij}^{(r)}|X_{(r)i}]\right]$$

$$=\mathbb{E}_{\boldsymbol{\Psi}}\left[\sum_{j=1}^{M}\sum_{r=1}^{n}\frac{\pi_j f_j^{(r:n)}(X_{(r)i};\theta_j)\phi_r(\lambda)\,G(X_{(r)i})}{[\lambda+(1-2\lambda)F_j(X_{(r)i};\theta_j)](f^{(r:n)}(X_{(r)i};\boldsymbol{\Psi}))}\right].$$

□

Now, setting $G(u)=D^2_{\boldsymbol{\Psi}}F_j(u;\theta_j)$, for $\lambda\in\{0,1\}$, Lemma 3.3 leads us to

$$\mathbb{E}\left[\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\frac{\phi_r(\lambda)Z_{ij}^{(r)}D^2_{\boldsymbol{\Psi}}F_j(X_{(r)i};\theta_j)}{\lambda+(1-2\lambda)F_j(X_{(r)i};\theta_j)}\right]=Nn(n-1)\sum_{j=1}^{M}\pi_j\mathbb{E}_j[D^2_{\boldsymbol{\Psi}}F_j(X;\theta_j)].$$

Similarly, by choosing $G_\lambda(u)=\frac{D_{\boldsymbol{\Psi}}F_j(u;\theta_j)[D_{\boldsymbol{\Psi}}F_j(u;\theta_j)]^\top}{[\lambda+(1-2\lambda)F_j(u;\theta_j)]^2}$, $\lambda\in\{0,1\}$, it is easy to show

$$\mathbb{E}\left[\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\frac{\phi_r(\lambda)\,Z_{ij}^{(r)}\,G_\lambda(X_{(r)i};\theta_j)}{[\lambda+(1-2\lambda)F_j(X_{(r)i};\theta_j)]^2}\right]$$

$$=Nn(n-1)\sum_{j=1}^{M}\pi_j\mathbb{E}_j\left(\frac{D_{\boldsymbol{\Psi}}F_j(X;\theta_j)[D_{\boldsymbol{\Psi}}F_j(X;\theta_j)]^\top}{\lambda+(1-2\lambda)F_j(X;\theta_j)}\right).$$

These allow us to derive of the FI matrix of Type-M2 RSS data about $\boldsymbol{\Psi}$ as follows.

Theorem 3.5. *The FI matrix associated with Type-M2 RSS data from the FMM*

(1.1) *is given by*

$$\mathbb{I}_{M2,RSS}(\boldsymbol{\Psi}) = \mathbb{I}_{C,SRS}(\boldsymbol{\Psi}) + Nn(n-1)\sum_{j=1}^{M}\pi_j\mathbb{E}_j\left[\frac{D_{\boldsymbol{\Psi}}F_j(X;\theta_j)[D_{\boldsymbol{\Psi}}F_j(X;\theta_j)]^\top}{F_j(X;\theta_j)\bar{F}_j(X;\theta_j))}\right]$$

$$= \mathbb{I}_{C,SRS}(\boldsymbol{\Psi}) + Nn(n-1)\sum_{j=1}^{M}\pi_j\Xi_{M2,j}(\boldsymbol{\xi}). \tag{3.14}$$

Furthermore, $\sum_{j=1}^{M}\pi_j\Xi_{M2,j}(\boldsymbol{\xi})$ *is a non-negative definite matrix.*

Theorem 3.5 shows that the FI contained in the Type-M2 RSS sample about $\boldsymbol{\Psi}$ is not smaller than the FI contained in Type-C SRS sample of the same size. We now present a result similar to Theorem 3.3, concerning the relative efficiency of Type-M2 RSS data compared with its SRS counterpart for estimating the mixing proportion of a mixture of two known densities. Theorem 3.6 says that Type-M2 RSS data does not provide extra information about the mixing proportion when compared to Type-C SRS data.

Theorem 3.6. *Under the conditions of Theorem 3.3, the relative efficiency of Type-M2 RSS of size nN for estimating the mixing proportion π relative to its SRS counterpart is equal to one, regardless of the component densities f_1 and f_2.*

Proof. When π in the only unknown parameter of the model, $\sum_{j=1}^{M}\pi_j\Xi_{M2,j}(\boldsymbol{\xi}) = 0$. Hence, the result follows from Theorem 3.5. □

3.2 FI in imperfect RSS data from FMM

The FI matrices were obtained for different RSS data structures under the assumption of perfect ranking in the previous section. In practice, however, error in ranking is inevitable and ranks assigned to the sample units could possibly be different from their actual ranks. So, it is natural to study the effects of ranking error on the FI contained in RSS data. In an imperfect RSS, the pdf of $X_{[r]}$, the r-th judgemental order statistic in a random sample of size n is

$$f^{[r]}(x; \mathbf{\Psi}) = \sum_{s=1}^{n} p_{sr}\, f^{(s:n)}(x; \mathbf{\Psi}) = g_r(x, \mathbf{\Psi}) f(x; \mathbf{\Psi}),$$

where $p_{sr} = P(X_{[r]} = X_{(s)})$, $\sum_{r=1}^{n} p_{sr} = \sum_{s=1}^{n} p_{sr} = 1$ and

$$g_r(x; \mathbf{\Psi}) = \sum_{s=1}^{n} p_{sr} f^{(s:n)}(x; \mathbf{\Psi}) / f(x; \mathbf{\Psi}),$$

with

$$\sum_{r=1}^{n} g_r(x; \mathbf{\Psi}) = n. \qquad (3.15)$$

Suppose $\widetilde{\mathbf{X}} = \{X_{[r]i}, r = 1, \ldots, n; i = 1, \ldots, N\}$ is data from an imperfect RSS data of size nN. We can consider three different structures for the imperfect RSS data as in Section 3.1. The description of each structure and the underlying likelihood function is similar to the corresponding ones under perfect ranking. For example, in Type-M0 imperfect RSS, the joint pdf of the sample is given by

$$f(\widetilde{\mathbf{x}}; \mathbf{\Psi}) = \prod_{i=1}^{N} \prod_{r=1}^{n} f^{[r]}(x_{[r]i}; \mathbf{\Psi}). \qquad (3.16)$$

We here consider the FI of Type-M0 and Type-M2 imperfect RSS data about $\mathbf{\Psi}$ and compare them with their counterparts under SRS. The problem of ML estimation of $\mathbf{\Psi}$ and the derivation of the FI matrix based on Type-M1 RSS data is very complicated and computationally intensive. This will be the subject of future investigation.

3.2.1 FI in Type-M0 imperfect RSS data

To obtain the FI matrix of Type-M0 imperfect RSS data, using (3.16) we first obtain the likelihood function as

$$L_{M0,IRSS}(\mathbf{\Psi}) = \prod_{i=1}^{N}\prod_{r=1}^{n} g_r(x_{[r]i};\mathbf{\Psi}) f(x_{[r]i};\mathbf{\Psi}).$$

Therefore, the FI matrix of Type-M0 imperfect RSS data can be written as the sum of two quantities as given by

$$\mathbb{I}_{M0,IRSS}(\mathbf{\Psi}) = -\mathbb{E}\left[D^2_{\mathbf{\Psi}}\sum_{i=1}^{N}\sum_{r=1}^{n}\log g_r(X_{[r]i};\mathbf{\Psi})\right] - \mathbb{E}\left[D^2_{\mathbf{\Psi}}\sum_{i=1}^{N}\sum_{r=1}^{n}\log f(X_{[r]i};\mathbf{\Psi})\right].$$

Using (3.15), an argument similar to the one used in the Type-M0 perfect RSS case, we get

$$-\mathbb{E}\left[D^2_{\mathbf{\Psi}}\sum_{i=1}^{N}\sum_{r=1}^{n}\log f(X_{[r]i};\mathbf{\Psi})\right] = \mathbb{I}_{M0,SRS}(\mathbf{\Psi}).$$

In addition, it is easy to show that

$$(a)\ \mathbb{E}\left[D^2_{\mathbf{\Psi}}\sum_{r=1}^{n}\log g_r(X_{[r]};\mathbf{\Psi})\right] = \sum_{r=1}^{n}\mathbb{E}\left[\frac{D^2_{\mathbf{\Psi}} g_r(X_{[r]};\mathbf{\Psi})}{g_r(X_{[r]};\mathbf{\Psi})}\right] - \sum_{r=1}^{n}\mathbb{E}\left[\frac{D_{\mathbf{\Psi}} g_r(X_{[r]};\mathbf{\Psi})[D_{\mathbf{\Psi}} g_r(X_{[r]};\mathbf{\Psi})]^\top}{[g_r(X_{[r]};\mathbf{\Psi})]^2}\right].$$

(b) $\sum_{r=1}^{n} \mathbb{E}\left[\frac{D_{\mathbf{\Psi}} g_r(X_{[r]};\mathbf{\Psi})[D_{\mathbf{\Psi}} g_r(X_{[r]};\mathbf{\Psi})]^\top}{[g_r(X_{[r]};\mathbf{\Psi})]^2}\right] = \sum_{r=1}^{n} \mathbb{E}\left[\frac{D_{\mathbf{\Psi}} g_r(X;\mathbf{\Psi})[D_{\mathbf{\Psi}} g_r(X;\mathbf{\Psi})]^\top}{g_r(X;\mathbf{\Psi})}\right].$

(c) $\sum_{r=1}^{n} \mathbb{E}\left[\frac{D^2_{\mathbf{\Psi}} g_r(X_{[r]};\mathbf{\Psi})}{g_r(X_{[r]};\mathbf{\Psi})}\right] = 0.$

Now, we are ready to obtain the FI matrix of Type-M0 imperfect RSS about $\mathbf{\Psi}$, i.e. $\mathbb{I}_{M0,IRSS}(\mathbf{\Psi})$.

Theorem 3.7. *The FI matrix of Type-M0 imperfect RSS data about $\mathbf{\Psi}$ in the FMM (1.1) is given by*

$$\begin{aligned}
\mathbb{I}_{M0,IRSS}(\mathbf{\Psi}) &= \mathbb{I}_{M0,SRS}(\mathbf{\Psi}) + N\sum_{r=1}^{n} \mathbb{E}\left[\frac{D_{\mathbf{\Psi}} g_r(X;\mathbf{\Psi})[D_{\mathbf{\Psi}} g_r(X;\mathbf{\Psi})]^\top}{g_r(X;\mathbf{\Psi})}\right] \\
&= \mathbb{I}_{M0,SRS}(\mathbf{\Psi}) + \tilde{\Delta}_0(\mathbf{\Psi}), \qquad (3.17)
\end{aligned}$$

where $\tilde{\Delta}_0(\mathbf{\Psi})$ is a non-negative definite matrix.

Theorem 3.7 shows that the FI of Type-M0 imperfect RSS sample of size nN about $\mathbf{\Psi}$ is not smaller than its counterpart under SRS.

3.2.2 FI in Type-M2 imperfect RSS data

Let $(\widetilde{\mathbf{X}}, \mathbf{Z})$ be a Type-M2 imperfect RSS sample of size Nn. Suppose for each $X_{[r]i}$, $\mathbf{Z}_i^{[r]} = (Z_{i1}^{[r]}, \ldots, Z_{iM}^{[r]})$ is an M-dimensional vector, where

$$Z_{ij}^{[r]} = \begin{cases} 1 & \text{if } X_{[r]i} \text{ is the r-th judgemental order statistic of component } j; \\ 0 & \text{otherwise,} \end{cases}$$

with $\sum_{j=1}^{M} Z_{ij}^{[r]} = 1$. Indicator vectors $\mathbf{Z}_1^{[1]}, \ldots, \mathbf{Z}_N^{[n]}$ are $i.i.d.$ samples from a Mult$(1, \boldsymbol{\pi})$ distribution. Also, the log-likelihood function of $\boldsymbol{\Psi}$ can be written as follow

$$l_{M2,IRSS}(\boldsymbol{\Psi}) = \log \left\{ \prod_{i=1}^{N} \prod_{r=1}^{n} \prod_{j=1}^{M} \left\{ \pi_j f_j(x_{[r]i}; \theta_j) \right\}^{z_{ij}^{[r]}} \prod_{i=1}^{N} \prod_{r=1}^{n} \prod_{j=1}^{M} \left\{ \pi_j g_r(x_{[r]i}; \theta_j) \right\}^{z_{ij}^{[r]}} \right\}$$

$$\propto l_{C,SRS}(\boldsymbol{\Psi}) + \widetilde{\Gamma}_{M2}(\boldsymbol{\Psi}),$$

where $\widetilde{\Gamma}_{M2}(\boldsymbol{\Psi}) = \sum_{i=1}^{N} \sum_{r=1}^{n} \sum_{j=1}^{M} z_{ij}^{[r]} \log g_r(x_{[r]i}; \theta_j)$. Since

$$Z_{ij}^{[r]} \big| X_{[r]i} = x_{[r]i} \sim \text{Bin}\left(1, \frac{\pi_j f_j^{[r]}(x_{[r]i}; \theta_j)}{f^{[r]}(x_{(r)i}; \boldsymbol{\Psi})}\right),$$

it is easy to see that $\mathbb{I}_{C,SRS}(\boldsymbol{\Psi}) = -\mathbb{E}[D_{\boldsymbol{\Psi}}^2 l_{C,SRS}(\boldsymbol{\Psi})]$. Also,

$$D_{\boldsymbol{\Psi}}^2 \widetilde{\Gamma}_{M2}(\boldsymbol{\Psi}) = \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{r=1}^{n} z_{ij}^{[r]} \left\{ \frac{D_{\boldsymbol{\Psi}}^2 g_r(x_{[r]i}; \theta_j)}{g_r(x_{[r]i}; \theta_j)} \right\}$$

$$- \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{r=1}^{n} z_{ij}^{[r]} \left\{ \frac{D_{\boldsymbol{\Psi}} g_r(x_{[r]i}; \theta_j) [D_{\boldsymbol{\Psi}} g_r(x_{[r]i}; \theta_j)]^\top}{[g_r(x_{[r]i}; \theta_j)]^2} \right\}.$$

Using the conditional expectation and $f(\widetilde{\mathbf{X}}, \mathbf{Z}; \theta_j) = f(\widetilde{\mathbf{X}}; \mathbf{Z}, \theta_j) f(\mathbf{Z}; \theta_j)$, a straightforward calculation shows that

$$\mathbb{E}\left\{ \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{r=1}^{n} Z_{ij}^{[r]} \frac{D_{\boldsymbol{\Psi}}^2 g_r(X_{[r]i}; \theta_j)}{g_r(X_{[r]i}; \theta_j)} \right\} = \sum_{i=1}^{N} \sum_{j=1}^{M} \pi_j \mathbb{E}_j \left\{ \sum_{r=1}^{n} \frac{D_{\boldsymbol{\Psi}}^2 g_r(X_{[r]i}; \theta_j)}{g_r(X_{[r]i}; \theta_j)} \right\}$$

$$= N \sum_{j=1}^{M} \pi_j \int D_{\boldsymbol{\Psi}}^2 [\sum_{r=1}^{n} g_r(x; \theta_j)] f_j(x; \theta_j) dx = 0.$$

Similarly,

$$\mathbb{E}\left\{\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{r=1}^{n} Z_{ij}^{[r]} \frac{D_{\mathbf{\Psi}} g_r(X_{[r]i};\theta_j)[D_{\mathbf{\Psi}} g_r(X_{[r]i};\theta_j)]^\top}{[g_r(X_{[r]i};\theta_j)]^2}\right\}$$

$$= N\sum_{j=1}^{M}\pi_j \sum_{r=1}^{n}\mathbb{E}_j\left\{\frac{D_{\mathbf{\Psi}} g_r(X;\theta_j)[D_{\mathbf{\Psi}} g_r(X;\theta_j)]^\top}{g_r(X;\theta_j)}\right\}.$$

Now, we obtain the FI matrix of Type-M2 imperfect RSS data about $\mathbf{\Psi}$ in the following theorem.

Theorem 3.8. *The FI matrix of a Type-M2 imperfect RSS sample of size nN from the FMM (1.1) is given by*

$$\mathbb{I}_{M2,IRSS}(\mathbf{\Psi}) = \mathbb{I}_{C,SRS}(\mathbf{\Psi}) + N\sum_{j=1}^{M}\pi_j \widetilde{\mathbb{E}}_{M2,j}(\boldsymbol{\xi}), \tag{3.18}$$

where

$$\widetilde{\mathbb{E}}_{M2,j}(\boldsymbol{\xi}) = \sum_{r=1}^{n}\mathbb{E}_j\left\{\frac{D_{\mathbf{\Psi}} g_r(X;\theta_j)[D_{\mathbf{\Psi}} g_r(X;\theta_j)]^\top}{g_r(X;\theta_j)}\right\}.$$

Furthermore, $\sum_{j=1}^{M}\pi_j\widetilde{\mathbb{E}}_{M2,j}(\boldsymbol{\xi})$ is a non-negative definite matrix.

Theorem 3.18 shows that the FI of Type-M2 imperfect RSS sample is not smaller than its SRS counterpart.

3.3 Missing information principle

In this section, the estimation problem of unknown parameters of FMM (1.1) using RSS data is considered through a missing information principle (MIP) by noting that in practice usually the latent variables $\mathbf{Z}$, $\mathbf{V}$ and $\mathbf{W}$ are missing. If these were available then one could use either Type-M1 or Type-M2 RSS data to make inference about the unknown parameters $\mathbf{\Psi}$ using their corresponding likelihood functions. The MIP of Orchard and Woodbury (1972) can be used as a useful tool to quantify the amount of information that one may lose by using an incomplete or Type-M0 RSS instead of Type-M1 or Type-M2 RSS to make inference about $\mathbf{\Psi}$. We first give a brief introduction to the MIP and then study the impact of for using Type-M0 RSS instead of Type-M1 or Type-M2 RSS in a mixture of two exponential distributions from the perspective of missing information. We only consider prefect RSS of different types of data. Results for the imperfect cases can be obtained similarly. Suppose $\mathbf{Y} = (\mathbf{X}, \mathbf{M})$ denotes the complete data, where $\mathbf{X}$ is the incomplete data that we usually observe in practice and $\mathbf{M}$ is the missing observations. For example, for comparing Type-M0 RSS to Type-M1 RSS the missing observations are $\mathbf{M} = (\mathbf{Z}, \mathbf{W}, \mathbf{V})$. Let $L_{\mathbf{Y}}(\mathbf{\Psi})$ and $L_{\mathbf{X}}(\mathbf{\Psi})$ be the likelihood functions of the complete and incomplete data, respectively. Also, assume that $g(\mathbf{M}|\mathbf{X}; \mathbf{\Psi})$ is the conditional distribution of the missing data $\mathbf{M}$ given $\mathbf{X}$. Then, the complete data likelihood function may be written as

$$L_{\mathbf{Y}}(\mathbf{\Psi}) = L_{\mathbf{X}}(\mathbf{\Psi}) \times g(\mathbf{M}|\mathbf{X}; \mathbf{\Psi}),$$

resulting in the relationship $l_{\mathbf{Y}}(\mathbf{\Psi}) = l_{\mathbf{X}}(\mathbf{\Psi}) + \log g(\mathbf{M}|\mathbf{X}; \mathbf{\Psi})$ among the log-likelihood functions. Under suitable regularity conditions and after taking the second derivatives with respect to $\mathbf{\Psi}$ and the required expectations, we have

$$\mathbb{I}_{\mathbf{Y}}(\mathbf{\Psi}) = \mathbb{I}_{\mathbf{X}}(\mathbf{\Psi}) + \mathbb{I}_{\mathbf{M}|\mathbf{X}}(\mathbf{\Psi}). \tag{3.19}$$

Now, the loss of information due to the use of $\mathbf{X}$ instead of $\mathbf{Y}$ for making inference about $\mathbf{\Psi}$, which is denoted by $\text{MIP}_{\mathbf{X},\mathbf{Y}}(\mathbf{\Psi})$, is the last quantity on the right side of

(3.19) and it is calculated as follows

$$\mathrm{MIP}_{\mathbf{X},\mathbf{Y}}(\mathbf{\Psi}) = \mathbb{I}_{\mathbf{M}|\mathbf{X}}(\mathbf{\Psi}) = \mathbb{I}_{\mathbf{Y}}(\mathbf{\Psi}) - \mathbb{I}_{\mathbf{X}}(\mathbf{\Psi}). \tag{3.20}$$

Unfortunately, this elegant equality is computationally unattractive and it is indeed very hard to use in practice, especially in the case of FMMs with ranked set samples. In order to reduce the computational burden, we can estimate (3.19) or equivalently (3.20) using their observed versions (i.e., conditional on $\mathbf{x}$). To this end, let

$$Q(\mathbf{\Psi}, \mathbf{\Psi}^*) = \mathbb{E}_{\mathbf{\Psi}^*}\left[D^2_{\mathbf{\Psi}} l_{\mathbf{Y}}(\mathbf{\Psi})|\mathbf{x}\right] \quad \text{and} \quad K(\mathbf{\Psi}, \mathbf{\Psi}^*) = \mathbb{E}_{\mathbf{\Psi}^*}[D^2_{\mathbf{\Psi}} \log g(\mathbf{M}|\mathbf{X}; \mathbf{\Psi})|\mathbf{x}].$$

Suppose $H^{\mathbf{x}}_{Q}(\mathbf{\Psi}, \mathbf{\Psi}) = Q(\mathbf{\Psi}, \mathbf{\Psi}^*)|_{\mathbf{\Psi}^*=\mathbf{\Psi}}$ is the conditional expectation of the Hessian matrix associated with $l_{\mathbf{Y}}(\mathbf{\Psi})$ or equivalently $Q(\mathbf{\Psi}, \mathbf{\Psi}^*)$ evaluated at $\mathbf{\Psi}^* = \mathbf{\Psi}$. Similarly, let $H^{\mathbf{x}}_{K}(\mathbf{\Psi}, \mathbf{\Psi}) = K(\mathbf{\Psi}, \mathbf{\Psi}^*)|_{\mathbf{\Psi}^*=\mathbf{\Psi}}$ and $H^{\mathbf{x}}(\mathbf{\Psi}) = D^2_{\mathbf{\Psi}} l_{\mathbf{x}}(\mathbf{\Psi})$. Then, we obtain the observed MIP as follows

$$\mathrm{MIP}(\mathbf{\Psi}|\mathbf{x}) = H^{\mathbf{x}}_{K}(\mathbf{\Psi}, \mathbf{\Psi}) = H^{\mathbf{x}}_{Q}(\mathbf{\Psi}, \mathbf{\Psi}) - H^{\mathbf{x}}(\mathbf{\Psi}),$$

which can be estimated by $\widehat{\mathrm{MIP}}(\widehat{\mathbf{\Psi}}|\mathbf{x}) = H^{\mathbf{x}}_{Q}(\widehat{\mathbf{\Psi}}, \widehat{\mathbf{\Psi}}) - H^{\mathbf{x}}(\widehat{\mathbf{\Psi}})$, where $\widehat{\mathbf{\Psi}}$ is a suitable estimator of $\mathbf{\Psi}$. We should also mention that when $\mathrm{MIP}(\mathbf{\Psi}|\mathbf{x})$ is calculated, it can be used to derive the asymptotic variance-covariance matrix of the ML estimator of $\mathbf{\Psi}$. Let $\widehat{\mathbf{\Psi}}_{MLE}$ be the ML estimator of $\mathbf{\Psi}$ using Type-M0 RSS data (see Chapter 2). Under suitable regularity conditions, the asymptotic variance of $\widehat{\mathbf{\Psi}}_{MLE}$ is given by $\{\widehat{\mathbb{I}}_{\mathbf{X}}(\widehat{\mathbf{\Psi}}_{MLE})\}^{-1}$, or equivalently,

$$\widehat{Cov}(\widehat{\mathbf{\Psi}}_{MLE}) = \left(H^{\mathbf{x}}_{K}(\widehat{\mathbf{\Psi}}_{MLE}, \widehat{\mathbf{\Psi}}_{MLE}) - H^{\mathbf{x}}_{Q}(\widehat{\mathbf{\Psi}}_{MLE}, \widehat{\mathbf{\Psi}}_{MLE})]\right)^{-1},$$

where $K(\mathbf{\Psi}, \mathbf{\Psi}^*)|_{(\mathbf{\Psi}^*=\mathbf{\Psi})} = H^{\mathbf{x}}_{K}(\mathbf{\Psi}, \mathbf{\Psi}) = -Cov_{\mathbf{\Psi}}[D_{\mathbf{\Psi}} l_{\mathbf{Y}}(\mathbf{\Psi})|\mathbf{x}]$.

In the following examples, we show how to obtain the missing information about $\mathbf{\Psi}$ due to the use of Type-M0 perfect RSS instead of Type-M1 and Type-M2 perfect RSS for a finite mixture of two exponential distributions.

Example 3.1. (Type-M0 RSS vs. Type-M1 RSS) *Suppose $\mathbf{X}$ is a Type-M0 perfect RSS sample of size Nn from a finite mixture of two exponential distribution as*

$$f(x;\mathbf{\Psi}) = \pi\,\alpha e^{-\alpha\,x} + (1-\pi)\beta e^{-\beta\,x}, \quad x>0,$$

where $\mathbf{\Psi} = (\pi,\alpha,\beta)$, $\alpha,\beta > 0$ and $\pi \in (0,1)$. To obtain the missing information due to the use of Type-M0 RSS instead of Type-M1 RSS, using (2.5), (2.14) and the conditional distribution $f_{\mathbf{Z},\mathbf{W},\mathbf{V}|\mathbf{X}}(\mathbf{z},\mathbf{w},\mathbf{v}|\mathbf{x},\mathbf{\Psi})$, one can easily show that $H_Q^{\mathbf{x}}(\mathbf{\Psi},\mathbf{\Psi}) = -diag(L^1_{\mathbf{\Psi}}(\mathbf{x}), L^2_{\mathbf{\Psi}}(\mathbf{x}), L^3_{\mathbf{\Psi}}(\mathbf{x}))$, where

$$\begin{aligned}
L^1_{\mathbf{\Psi}}(\mathbf{x}) =& \frac{(1-2\pi)}{\pi(1-\pi)^2}\sum_{i=1}^{N}\sum_{r=1}^{n}\left\{\frac{\alpha e^{-\alpha x_{(r)i}}}{f(x_{(r)i};\mathbf{\Psi})} + \frac{(r-1)(1-e^{-\alpha x_{(r)i}})}{F(x_{(r)i};\mathbf{\Psi})} + \frac{(n-r)e^{-\alpha x_{(r)i}}}{\bar{F}(x_{(r)i};\mathbf{\Psi})}\right\} \\
&+ \frac{Nn^2}{(1-\pi)^2},
\end{aligned}$$

$$L^2_{\mathbf{\Psi}}(\mathbf{x}) = \pi\sum_{i=1}^{N}\sum_{r=1}^{n}\left\{\frac{e^{-\alpha x_{(r)i}}}{\alpha f(x_{(r)i};\mathbf{\Psi})} + \frac{(r-1)x^2_{(r)i}e^{-\alpha x_{(r)i}}}{(1-e^{-\alpha x_{(r)i}})F(x_{(r)i};\mathbf{\Psi})}\right\},$$

$$L^3_{\mathbf{\Psi}}(\mathbf{x}) = (1-\pi)\sum_{i=1}^{N}\sum_{r=1}^{n}\left\{\frac{e^{-\beta x_{(r)i}}}{\beta f(x_{(r)i};\mathbf{\Psi})} + \frac{(r-1)x^2_{(r)i}e^{-\beta x_{(r)i}}}{(1-e^{-\beta x_{(r)i}})F(x_{(r)i};\mathbf{\Psi})}\right\},$$

and $F(x;\mathbf{\Psi}) = 1-\{\pi e^{-\alpha x}+(1-\pi)e^{-\beta x}\}$. It is worth mentioning that by further taking the expectation with respect to $f_{\mathbf{X}}(\mathbf{x};\mathbf{\Psi})$, one gets $\mathbb{I}_{M0,RSS}(\mathbf{\Psi}) = -\mathbb{E}\left(H_Q^{\mathbf{X}}(\mathbf{\Psi},\mathbf{\Psi})\right)$. Let $\mathbf{Y} = (\mathbf{X},\mathbf{Z},\mathbf{W},\mathbf{V})$ denote the Type-M1 perfect RSS data. Then, $D_{\mathbf{\Psi}} l_{\mathbf{Y}}^{M1,RSS}(\mathbf{\Psi}) =$

$A_1(\mathbf{X})\mathbf{M} + B_1(\mathbf{X})$, *where* $\mathbf{M}$ *is a* $3Nn$*-dimensional vector defined as*

$$\mathbf{M}^t = \left(Z_{11}^{(1)}, \ldots, Z_{N1}^{(n)}, W_{11}^{(1)}, \ldots, W_{N1}^{(n)}, V_{11}^{(1)}, \ldots, V_{N1}^{(n)}\right).$$

Also, $A_1(\mathbf{x})$ *is a* $3 \times (3Nn)$ *matrix and* $B_1(\mathbf{x})$ *is a* 3*-dimensional vector given by*

$$\begin{bmatrix} \frac{1}{\pi(1-\pi)} & \cdots & \frac{1}{\pi(1-\pi)} & \frac{1}{\pi(1-\pi)} & \cdots & \frac{1}{\pi(1-\pi)} & \frac{1}{\pi(1-\pi)} & \cdots & \frac{1}{\pi(1-\pi)} \\ \frac{1}{\alpha} - x_{(1)1} & \cdots & \frac{1}{\alpha} - x_{(n)N} & \frac{x_{(1)1}e^{-\alpha\, x_{(1)1}}}{1-e^{-\alpha\, x_{(1)1}}} & \cdots & \frac{x_{(n)N}e^{-\alpha\, x_{(n)N}}}{1-e^{-\alpha\, x_{(n)N}}} & -x_{(1)1} & \cdots & -x_{(n)N} \\ x_{(1)1} - \frac{1}{\beta} & \cdots & x_{(n)N} - \frac{1}{\beta} & -\frac{x_{(1)1}e^{-\beta\, x_{(1)1}}}{1-e^{-\beta\, x_{(1)1}}} & \cdots & -\frac{x_{(n)N}e^{-\beta\, x_{(n)N}}}{1-e^{-\beta\, x_{(n)N}}} & x_{(1)1} & \cdots & x_{(n)N} \end{bmatrix},$$

and

$$\left(-\frac{Nn^2}{1-\pi}, 0, \frac{nN}{\beta} - \sum_{i=1}^{N}\sum_{r=1}^{n}\left\{(n-r+1)x_{(r)i} - \frac{(r-1)x_{(r)i}e^{-\beta x_{(r)i}}}{1-e^{-\beta x_{(r)i}}}\right\}\right)^{\top},$$

respectively. Hence, $H^{\mathbf{x}}(\mathbf{\Psi}) = A_1(\mathbf{x})Cov_{\mathbf{\Psi}}(\mathbf{M}|\mathbf{x})A_1^{\top}(\mathbf{x})$*, and the observed (or conditional) missing information is* $\mathrm{MIP}(\mathbf{\Psi}|\mathbf{x}) = H_Q^{\mathbf{x}}(\mathbf{\Psi}, \mathbf{\Psi}) - H^{\mathbf{x}}(\mathbf{\Psi})$*, which can be estimated using a suitable estimator* $\widehat{\mathbf{\Psi}}$ *by* $\mathrm{MIP}(\widehat{\mathbf{\Psi}}|\mathbf{x})$*. Using the structure of RSS and the definition of* $\mathbf{Z}$*,* $\mathbf{W}$ *and* $\mathbf{V}$*, it is easy to show that* $Cov_{\mathbf{\Psi}}(\mathbf{M}|\mathbf{x}) = diag(L_1, L_2, L_3)$*, where* $L_1 = \pi(1-\pi)I_{Nn\times Nn}$*, and* L_2, L_3 *are* $(Nn \times Nn)$ *diagonal matrices such that* $L_2 = \pi(1-\pi)diag(0, 1, \ldots, n-1, \ldots, 0, 1, \ldots, n-1)$*, and* $L_3 = \pi(1-\pi)diag(n-1, \ldots, 1, 0, \ldots, n-1, \ldots, 1, 0)$.

Example 3.2. (Type-M0 RSS against Type-M2 RSS) *In Example 3.1, suppose we are interested in calculating the missing information due to the use of Type-M0 RSS instead of Type-M2 RSS. Here, using (2.5) and (2.24) and by working with the conditional distribution* $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x};\mathbf{\Psi})$ *we get*

$$H_Q^{\mathbf{x}}(\mathbf{\Psi}, \mathbf{\Psi}) = -diag(L_{\mathbf{\Psi}}^1(\mathbf{x}), L_{\mathbf{\Psi}}^2(\mathbf{x}), L_{\mathbf{\Psi}}^3(\mathbf{x})),$$

where

$$L^1_{\mathbf{\Psi}}(\mathbf{x}) = \frac{(1-2\pi)}{\pi(1-\pi)^2}\sum_{i=1}^{N}\sum_{r=1}^{n}\left\{\frac{f_1^{(r:n)}(x_{(r)i};\alpha)}{f^{(r:n)}(x_{(r)i};\mathbf{\Psi})}\right\} + \frac{nN}{(1-\pi)^2},$$

$$L^2_{\mathbf{\Psi}}(\mathbf{x}) = \sum_{i=1}^{N}\sum_{r=1}^{n}\frac{\pi f_1^{(r:n)}(x_{(r)i};\alpha)}{f^{(r:n)}(x_{(r)i};\mathbf{\Psi})}\left\{\frac{1}{\alpha^2} + \frac{(r-1)x^2_{(r)i}\bar{F}_1(x_{(r)i};\alpha)}{[F_1(x_{(r)i};\alpha)]^2}\right\},$$

$$L^3_{\mathbf{\Psi}}(\mathbf{x}) = \sum_{i=1}^{N}\sum_{r=1}^{n}\frac{(1-\pi) f_2^{(r:n)}(x_{(r)i};\beta)}{f^{(r:n)}(x_{(r)i};\mathbf{\Psi})}\left\{\frac{1}{\beta^2} + \frac{(r-1)x^2_{(r)i}\bar{F}_2(x_{(r)i};\beta)}{[F_2(x_{(r)i};\beta)]^2}\right\},$$

and $f_j^{(r:n)}(x;\theta_j)$ *and* $F_j^{(r:n)}(x;\theta_j)$ *are respectively the pdf and cdf of the* r*-th order statistic from the* j*-th component of the mixture of two exponential distributions,* $j=1,2$*, with* $\theta_1=\alpha$ *and* $\theta_2=\beta$*. Note that, again, by taking the expectation with respect to* $f_{\mathbf{X}}(\mathbf{x})$ *we have* $\mathbb{I}_{M0,RSS}(\mathbf{\Psi}) = \mathbb{E}\left(H_Q^{\mathbf{X}}(\mathbf{\Psi},\mathbf{\Psi})\right)$*. Let* $\mathbf{Y}=(\mathbf{X},\mathbf{Z})$ *be the Type-M2 RSS sample of size* nN*. Then* $D_{\mathbf{\Psi}} l_{\mathbf{Y}}^{M2,RSS}(\mathbf{\Psi}) = A_2(\mathbf{X})\mathbf{Z} + B_2(\mathbf{X})$*, where* $\mathbf{Z}$ *is an* nN*-dimensional vector defined in (2.24). Also,* $A_2(\mathbf{x})$ *is a* $3\times nN$ *matrix and* $B_2(\mathbf{x})$ *is a 3-dimensional vector defined as*

$$\begin{bmatrix} \frac{1}{\pi(1-\pi)} & \cdots & \frac{1}{\pi(1-\pi)} & \cdots & \frac{1}{\pi(1-\pi)} \\ \frac{1}{\alpha} - nx_{(1)1} & \cdots & \frac{1}{\alpha} - k_1 x_{(r)i} + \frac{(r-1)x_{(r)i}e^{-\alpha x_{(r)i}}}{1-e^{-\alpha x_{(r)i}}} & \cdots & \frac{1}{\alpha} - x_{(n)N} + \frac{(n-1)x_{(n)N}e^{-\alpha x_{(n)N}}}{1-e^{-\alpha x_{(n)N}}} \\ nx_{(1)1} - \frac{1}{\beta} & \cdots & k_1 x_{(r)i} - \frac{1}{\beta} - \frac{(r-1)x_{(r)i}e^{-\beta x_{(r)i}}}{1-e^{-\beta x_{(r)i}}} & \cdots & x_{(n)N} - \frac{1}{\beta} - \frac{(n-1)x_{(n)N}e^{-\beta x_{(n)N}}}{1-e^{-\beta x_{(n)N}}} \end{bmatrix},$$

and

$$B_2(\mathbf{x}) = \left(-\frac{Nn}{1-\pi}, 0, \frac{Nn}{\beta} - \sum_{i=1}^{N}\sum_{r=1}^{n}\left\{k_1 x_{(r)i} - \frac{(r-1)x_{(r)i}e^{-\beta x_{(n)N}}}{1-e^{-\beta x_{(n)N}}}\right\}\right)^{\top},$$

respectively, where $k_1 = n-r+1$*. Hence, using the structure of RSS and* $\mathbf{Z}$ *we have*

$$H^{\mathbf{x}}(\mathbf{\Psi}) = A_2(x)Cov_{\mathbf{\Psi}}(\mathbf{Z}|\mathbf{x})A_2^{\top}(\mathbf{x}) = \pi(1-\pi)A_2(\mathbf{x})A_2^{\top}(\mathbf{x}).$$

Also, the observed (or conditional) missing information is obtained as $\mathrm{MIP}(\boldsymbol{\Psi}|\mathbf{x}) = H_Q^{\mathbf{x}}(\boldsymbol{\Psi}, \boldsymbol{\Psi}) - H^{\mathbf{x}}(\boldsymbol{\Psi})$ *which can be estimated by* $\mathrm{MIP}(\widehat{\boldsymbol{\Psi}}|\mathbf{x})$*, where* $\widehat{\boldsymbol{\Psi}}$ *is any suitable estimate of* $\boldsymbol{\Psi}$*.*

3.4 Numerical results

In this section, we study the FI content of different types of RSS data and compare them with their SRS counterparts for a mixture of two exponential distributions

$$f(x; \boldsymbol{\Psi}) = \pi\alpha e^{-\alpha x} + (1-\pi)\beta e^{-\beta x} \quad x > 0, \tag{3.21}$$

where $\pi \in (0, 1)$, $\alpha, \beta > 0$ and $\boldsymbol{\Psi} = (\pi, \alpha, \beta)$. Model (3.21) has many applications in environmental studies, reliability and life testing problems. For example, Liu et al. (2002) used (3.21) for modelling the diameter distribution of mixed-species forest stands. To calculate the numerical values of the efficiencies, we followed the method presented by Hill (1963) for handling the mixture of exponential distributions. To this end, we introduce the new parameter $h = \frac{\alpha}{\beta}$ and transform the model (3.21) with three unknown parameters (π, α, β) to a mixture density with two parameters (π, h). This enables us to investigate the effect of the parameters (π, h) as well as the set size n on the relative efficiencies of different Types of RSS data compared with their SRS counterparts using the ratio of the determinant of the FI matrices as defined in Theorems 3.3 and 3.4. We first compare the FI of Type-M0 RSS with that of Type-M0 SRS. Then the results are presented for other types of perfect RSS data. Considering a case where the observed sample is of mixed types of (Type-M0 and Type-M1) perfect RSS data, we investigate the contribution of these two types of data in the estimation of the mixing proportion. Finally, the effects of ranking error on the FI matrices and the relative efficiencies are investigated.

3.4.1 RSS versus SRS

We first compare the FI of Type-M0 perfect RSS data and Type-M0 SRS data about the mixing proportion π for the model (3.21) with (partially) known component parameters when the set size is $n \in \{2, 3, 4, 5\}$ and $h = \frac{\alpha}{\beta} \in \{\frac{2}{3}, \frac{1}{3}, \frac{1}{10}, \frac{1}{1000}\}$. We use the ratio of the determinant of the FI matrices to obtain the relative efficiency of RSS data relative to SRS data. A value of the relative efficiency bigger than one shows the superiority of RSS over SRS. The results are presented in Table 3.1. We also show the graphs of the relative efficiency as a function of π for different values of n and h in Figure 3.1. When h is large, (i.e., $h \in \{\frac{1}{3}, \frac{2}{3}\}$), the relative efficiency seems to be increasing in π. This was expected intuitively, since the mixture model (3.21) will be more influenced by the component with higher mixing proportion and we expect to observe more order statistics in the RSS data selected from that component. On the other hand, for other cases (i.e., $h \in \{\frac{1}{10}, \frac{1}{1000}\}$) especially when the first component is associated with a very high weight, the mixture model is still influenced to some extent by the second component and consequently a portion of order statistics (usually with higher ranks) are selected from the second component of the model so that it still performs better (although with a smaller magnitude) than its SRS counterpart.

Table 3.1 shows the relative efficiency of Type-M0 perfect RSS relative to Type-M0 SRS data for $\pi = 0.1(0.1)0.9$ and $h \in \{\frac{2}{3}, \frac{1}{2}, \frac{1}{10}\}$ and $n \in \{2, 3, 4\}$. We also

Table 3.1: Relative efficiencies of Type-M0 perfect RSS data to Type-M0 SRS data.

n	h	π 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	2/3	1.505	1.537	1.562	1.584	1.603	1.619	1.633	1.646	1.658
	1/2	1.508	1.554	1.588	1.616	1.640	1.659	1.675	1.688	1.698
	1/10	1.522	1.568	1.599	1.622	1.638	1.648	1.652	1.647	1.632
3	2/3	2.041	2.118	2.180	2.234	2.281	2.322	2.358	2.392	2.422
	1/2	2.046	2.157	2.242	2.313	2.372	2.423	2.466	2.501	2.530
	1/10	2.057	2.160	2.233	2.289	2.330	2.357	2.368	2.358	2.321
4	2/3	2.608	2.744	2.854	2.949	3.034	3.109	3.175	3.237	3.293
	1/2	2.615	2.809	2.963	3.091	3.20	3.293	3.373	3.44	3.494
	1/10	2.621	2.802	2.936	3.040	3.117	3.168	3.189	3.172	3.105

calculated the relative efficiencies of both Type-M1 and Type-M2 perfect RSS data compared with Type-C SRS data for $\pi = 0.1(0.1)0.9$, $h \in \{\frac{2}{3}, \frac{1}{2}, \frac{1}{10}\}$ and $n \in \{2, 3, 4\}$. These results are presented in Table 3.2.

3.4.2 Mixed sampling

As exemplified by Titterington et al. (1985), sometimes the available sample may be of mixed type, i.e., composed of different types of data. Here, we consider the case where the observed sample from the FMM (3.21) consists of both Type-M0 and Type-M1 RSS data. Suppose that ρ_c denotes the percentage of the mixed sample which is of Type-M0 RSS data. Then, the likelihood function of the mixed sample can be written as

$$L_{mixed}(\mathbf{\Psi}) = L_{M0,RSS}(\mathbf{\Psi}) L_{M1,RSS}(\mathbf{\Psi}).$$

Table 3.3 provides the values of

$$\frac{(1-\rho_c)\mathbb{I}_{M0,RSS}(\pi) + \rho_c \mathbb{I}_{M1,RSS}(\pi)}{\mathbb{I}_{M1,RSS}(\pi)},$$

Table 3.2: Relative efficiencies of Type-M1 and Type-M2 perfect RSS to Type-C SRS.

n	h \ π	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	M2-RSS
2	2/3	3.82	3.84	3.86	3.88	3.89	3.91	3.93	3.95	3.97	1.971536
	1/2	3.70	3.73	3.76	3.79	3.82	3.85	3.88	3.91	3.94	1.971536
	1/10	3.10	3.13	3.17	3.20	3.23	3.25	3.27	3.29	3.31	1.971538
3	2/3	9.34	9.41	9.49	9.56	9.63	9.71	9.78	9.86	9.94	3.269688
	1/2	8.88	8.99	9.11	9.22	9.34	9.45	9.57	9.68	9.80	3.269688
	1/10	6.58	6.74	6.89	7.01	7.03	7.22	7.30	7.36	7.41	3.269692
4	2/3	18.43	18.61	18.79	18.97	19.16	19.34	19.52	19.71	19.89	4.894456
	1/2	17.31	17.59	17.87	18.15	18.43	18.71	18.99	19.28	19.56	4.894455
	1/10	11.72	12.15	12.53	15.58	13.13	13.36	13.54	13.67	13.75	4.894464

that is, the efficiency (in percentage) of the mixed sample relative to a Type-M1 perfect RSS sample when the mixing proportion of the model is unknown and $h \in \{\frac{1}{10}, \frac{1}{2}, \frac{2}{3}\}$. We consider four values of $\rho_c \in \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$.

The case $\rho_c = 0$ corresponds to the FI of Type-M0 RSS data, $\mathbb{I}_{M0,RSS}(\boldsymbol{\Psi})$, and the inverse of the values in that column can be interpreted as the number of Type-M0 observations required to obtain the same efficiency based on a single fully categorized observation from Type-M1 RSS structure. The efficiency is smaller when $\pi = 0.1$ than the case when $\pi = 0.5$. From the Table 3.1, it is seen that the relative efficiency of Type-M0 RSS to Type-M1 RSS decreases as ρ_c decreases. This suggests that Type-M0 RSS data does not contribute very much information about the mixing proportion relative to Type-M1 RSS data. It is interesting to note that, for example, when $\rho_c = 0$, $h = \frac{2}{3}$ and $n = 5$, one single Type-M1 RSS data point gives as much information as $\frac{1}{0.009} = 111$ observations obtained via Type-M0 RSS. So, if the main interest is to make inference about the mixing proportion, it is highly recommended to try to observe some Type-M1 RSS data in the final sample.

3.4.3 Effect of ranking errors

In the previous subsections, we established through simulations that perfect RSS data contain more FI about the unknown parameters of the FMM than SRS data. We now consider the case where ranking is not perfect. Suppose the set size is

Table 3.3: Percentage efficiencies of mixed samples relative to Type-M1 perfect RSS data.

		$\pi = 0.1$				$\pi = 0.5$			
				ρ_c				ρ_c	
n	h	0	1/4	1/2	3/4	0	1/4	1/2	3/4
	2/3	3.1	26	50.6	75.3	2.6	27	51.3	75.6
2	1/2	4.4	28.3	52.2	76.1	7.1	30.3	53.5	76.7
	1/10	35.3	51.5	67.6	83.8	40.6	55.4	70.3	85.1
	2/3	1.1	25.8	50.5	75.2	2.3	26.7	51.1	75.5
3	1/2	3.5	27.6	51.7	75.8	6.2	29.6	53.1	76.5
	1/10	27.9	45.9	63.9	81.9	35.6	51.7	67.8	83.9
	2/3	0.9	25.6	50.4	75.2	2	26.5	51	75.5
5	1/2	2.9	27.1	51.4	75.7	5.4	29.1	52.7	76.5
	1/10	22	41.5	61	80.5	31.5	48.6	65.7	82.8

$n \in \{2, 3\}$ and consider the following doubly stochastic matrices for the probability of ranking error, respectively, for $n = 2$ and $n = 3$,

$$P = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}, \quad P = \begin{bmatrix} p & \frac{(1-p)}{2} & \frac{(1-p)}{2} \\ \frac{(1-p)}{2} & p & \frac{(1-p)}{2} \\ \frac{(1-p)}{2} & \frac{(1-p)}{2} & p \end{bmatrix}.$$

It is straightforward to verify that the relative efficiency of Type-M2 imperfect RSS data relative to Type-C SRS data does not depend on the mixing proportion π. The relative efficiencies of Type-M0 RSS relative to Type-M0 SRS and Type-M2 RSS relative to Type-C SRS are presented in Table 3.4 for different values of p and π. From Table 3.4, it is apparent that errors in ranking will slightly influence the FI content of RSS data. Obviously, the FI content of RSS data is equal to that of the SRS data when $p = 0.5$ for the case $n = 2$, i.e. in the case when ranking is done at random. The relative efficiency shows a symmetric behaviour ,as a function of p, when $n = 2$.

Table 3.4: RE of Type-M0 RSS to Type-M0 SRS. The last column shows the RE of Type-M2 imperfect RSS to Type-C SRS data.

n	p \ π	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	M2-RSS
2	0.9	1.427	1.449	1.463	1.474	1.481	1.486	1.488	1.490	1.490	1.486
	0.75	1.158	1.171	1.180	1.188	1.194	1.199	1.203	1.206	1.208	1.161
	0.50	1	1	1	1	1	1	1	1	1	1
	0.25	1.158	1.171	1.180	1.188	1.194	1.199	1.203	1.206	1.208	1.161
	0.1	1.427	1.449	1.463	1.474	1.481	1.486	1.488	1.490	1.490	1.486
3	0.9	1.748	1.754	1.752	1.745	1.735	1.724	1.710	1.697	1.683	2.280
	0.75	1.404	1.423	1.437	1.448	1.456	1.463	1.467	1.471	1.474	1.582
	0.50	1.066	1.070	1.074	1.077	1.079	1.081	1.083	1.084	1.085	1.083
	0.25	1.018	1.019	1.020	1.020	1.021	1.022	1.022	1.023	1.023	1.022
	0.1	1.157	1.166	1.174	1.179	1.184	1.189	1.192	1.195	1.198	1.205

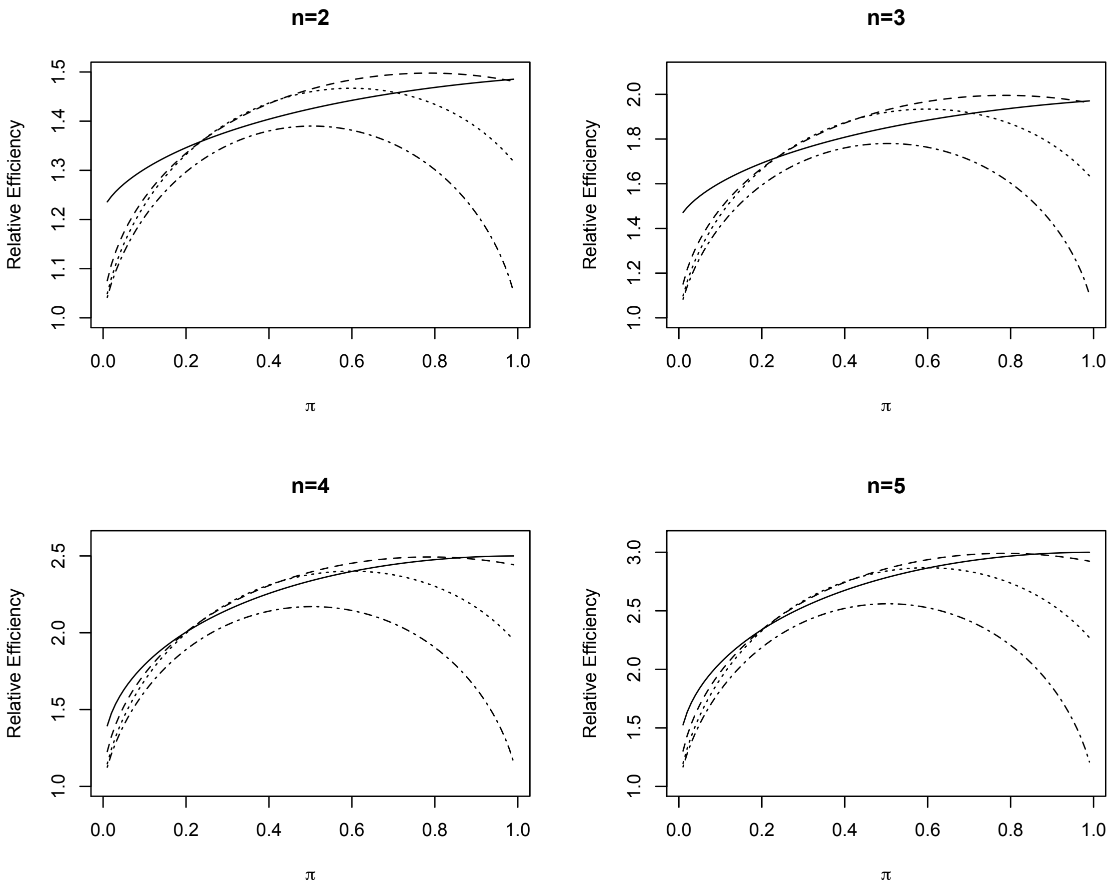

Figure 3.1: RE of estimators of Type-M0 perfect RSS to Type-M0 SRS data for the mixing proportion π. The solid line refers to $h = \frac{2}{3}$, dashed line to $h = \frac{1}{3}$, dotted line to $h = \frac{1}{10}$ and finally dash-dotted line to $h = \frac{1}{1000}$.

Chapter 4

Information Content of PROS Samples

In previous chapters, we focused on RSS designs as a powerful and cost-effective data collection technique that results in more representative samples from the underlying population. There are many situations where it is difficult to rank all of the sampling units in a set with high confidence, particularly when subjective information is utilized in the ranking process. Forcing rankers to declare unique ranks can lead to inflated within-set judgment ranking error and consequently to invalid statistical inference. The partially rank-ordered set (PROS) sampling design is a generalization of RSS due to Ozturk (2011). It is aimed at reducing the impact of ranking error and the burden on rankers by not requiring them to provide a full ranking of all the units in each set. Under the PROS sampling design, rankers have more flexibility by being able to divide the sampling units into subsets of pre-specified sizes. As indicated in Chapter 1 in the construction of PROS samples, these subsets are partially rank-ordered so that each unit in subset h has a rank smaller than the rank of units in subset $h^{'}$ for all $h^{'} \geq h$. An observation is then collected from one of these subsets in each set. PROS sampling has a wide range of applications in different fields ranging from environmental and ecological studies to medical research and it has been shown to be superior to RSS and SRS for estimating the population mean. For more details regarding the theory and applications of PROS sampling, readers are referred to Ozturk (2011).

In this chapter, we study the information content and uncertainty structure (i.e., entropy) of PROS samples. To this end, in Section 4.1, we provide some preliminary results on distributional properties of PROS samples. In Section 4.2, we obtain the FI content of PROS samples and show that it is more than the FI content of their SRS and RSS counterparts of the same size. Several examples including the FI of PROS samples from a location-scale family of distributions as well as a simple linear regression model are also discussed in this section. In addition, we explore the effect of sub-setting errors when using the PROS sampling design on the FI content of samples. Finally in Section 4.3, we study information and uncertainty of PROS samples using the Shannon entropy, Réyni entropy and Kullback-Leibler (KL) information measures and compare them with their SRS and RSS counterparts.

4.1 Distributional properties of PROS samples

To obtain a PROS sample of size n, we choose a set size S and a design parameter $D = \{d_1, \ldots, d_n\}$ that partitions the set $\{1, \ldots, S\}$ into n mutually exclusive subsets. Throughout the chapter, without loss of generality, we assume that $N = 1$ (cycle size is one). We use PROS(n, S, D) to denote a PROS sampling design with set size S, number of subsets n and the design parameter $D = \{d_r, r = 1, \ldots, n\}$ where $d_r = \{(r-1)m+1, \ldots, rm\}$, in which $m = S/n$ is the number of unranked observations in each subset. We note that RSS and SRS can be expressed as special cases of the PROS(n, S, D) design when $S = n$ and $S = 1$, respectively.

Suppose X is a continuous random variable with *pdf* $f(\cdot;\boldsymbol{\theta})$ and *cdf* $F(\cdot;\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector of unknown parameters with $\boldsymbol{\theta} \in \mathbb{R}^p$. Let $\mathbf{X}_{pros} = \{X_{(d_r)}, r = 1, \ldots, n\}$ be a PROS(n, S, D) sample of size n from $f(\cdot, \boldsymbol{\theta})$. The PROS data likelihood function of $\boldsymbol{\theta}$ is given by the joint pdf of $\mathbf{X}_{pros}$ as follows:

$$L(\boldsymbol{\theta}|\mathbf{x}_{pros}) = f(\mathbf{x}_{pros};\boldsymbol{\theta}) = \prod_{r=1}^{n} \left\{ \frac{1}{m} \sum_{u \in d_r} f^{(u:S)}(x_{(d_r)};\boldsymbol{\theta}) \right\},$$

where $f^{(u:S)}(\cdot;\boldsymbol{\theta})$ is the pdf of the u-th order statistic of an SRS of size S from $f(\cdot;\boldsymbol{\theta})$. For each $X_{(d_r)}$ define the latent vector

$$\boldsymbol{\Delta}^{(d_r)} = \left(\Delta^{(d_r)}(u), u \in d_r = \{(r-1)m+1, \ldots m\}\right),$$

where

$$\Delta^{(d_r)}(u) = \begin{cases} 1 & \text{if } X_{(d_r)} \text{ is selected from the } u\text{-th position within the subset } d_r; \\ 0 & \text{otherwise,} \end{cases}$$

with $\sum_{u \in d_r} \Delta^{(d_r)}(u) = 1$. Denote $\mathbf{Y}_{pros} = \{(X_{(d_r)}, \boldsymbol{\Delta}^{(d_r)}), r = 1, \ldots, n\}$ as the complete PROS data consisting of $X_{(d_r)}$ and their corresponding latent vectors $\boldsymbol{\Delta}^{(d_r)}$, $r = 1, \ldots, n$. The complete PROS data likelihood function of $\boldsymbol{\theta}$ using the joint pdf of $\mathbf{Y}_{pros}$ is given by

$$L(\boldsymbol{\theta}|\mathbf{y}_{pros}) = f(\mathbf{y}_{pros};\boldsymbol{\theta}) = \prod_{r=1}^{n} \prod_{u \in d_r} \left\{ \frac{1}{m} f^{(u:S)}(x_{(d_r)};\boldsymbol{\theta}) \right\}^{\boldsymbol{\delta}^{(d_r)}(u)}. \tag{4.1}$$

Furthermore, by summing the joint distribution of $(X_{(d_r)}, \boldsymbol{\Delta}^{(d_r)})$ over $\boldsymbol{\Delta}^{(d_r)} = \boldsymbol{\delta}^{(d_r)}$, the marginal distribution of $X_{(d_r)}$ is obtained as follows

$$f_{(d_r)}(x_{(d_r)};\boldsymbol{\theta}) = \sum_{\boldsymbol{\delta}^{(d_r)}} f(x_{(d_r)}, \boldsymbol{\delta}^{(d_r)};\boldsymbol{\theta}) = \frac{1}{m}\sum_{u\in d_r} f^{(u:S)}(x_{(d_r)};\boldsymbol{\theta}). \tag{4.2}$$

Also, one can easily check that

$$\frac{1}{n}\sum_{r=1}^{n} f_{(d_r)}(x;\boldsymbol{\theta}) = \frac{1}{S}\sum_{v=1}^{S} f^{(v:S)}(x;\boldsymbol{\theta}) = f(x;\boldsymbol{\theta}). \tag{4.3}$$

In addition, the conditional distribution of $\boldsymbol{\Delta}^{(d_r)}$ given $X_{(d_r)}$ is

$$f(\boldsymbol{\delta}^{(d_r)}\big|x_{(d_r)};\boldsymbol{\theta}) = \prod_{u\in d_h}\left\{\frac{f^{(u:S)}(x_{(d_r)};\boldsymbol{\theta})}{\sum_{u\in d_r} f^{(u:S)}(x_{(d_r)};\boldsymbol{\theta})}\right\}^{\boldsymbol{\delta}^{(d_r)}(u)}. \tag{4.4}$$

4.2 FI content of PROS samples

In this section, we first obtain the FI content of $\mathbf{Y}_{pros}$, the complete PROS data, and derive analytical results to compare it with the FI content of SRS and RSS data of the same size. We give examples regarding a location-scale family of distributions as well as a simple linear regression model. Then, we study the FI content of $\mathbf{X}_{pros}$ by modelling an imperfect PROS design involving misplacement errors in the sub-setting process. The FI of PROS samples can play a key role in theory and application to study the asymptotic behaviour of the ML estimators of $\boldsymbol{\theta}$ as well as the derivation of the Cramer-Rao lower bound for unbiased estimators of $\boldsymbol{\theta}$ or some of its functions based on PROS samples.

Similar to Section 3.1, in this chapter, we use the following notation. Under the usual regularity conditions (Chen et al., 2004), the FI matrix is given by $\mathbb{I}(\boldsymbol{\theta}) = -\mathbb{E}[D^2_{\boldsymbol{\theta}} \log f(X; \boldsymbol{\theta})]$, provided the expectation exists, where $D^l_{\boldsymbol{\theta}}$ refers to the l-th derivatives of the log-likelihood function with respect to $\boldsymbol{\theta}$ with $D^1_{\boldsymbol{\theta}} = D_{\boldsymbol{\theta}}$. For any two matrices A and B of the same size, we use $A \geq 0$ and $A \geq B$ to indicate that A and $A - B$ are nonnegative matrices. We also let $\phi_u(\lambda) = (u-1)\, I(\lambda = 0) + (S-u)\, I(\lambda = 1)$ with $\lambda \in \{0,1\}, u = 1, \ldots, S$, where again I is the usual indicator function.

4.2.1 FI matrix of complete PROS data $\mathbf{Y}_{pros}$

To obtain the FI matrix of $\mathbf{Y}_{pros}$, we need the following useful result.

Lemma 4.1. *Suppose $Y_r = X_{(d_r)}$, with pdf $f_{(d_r)}(\cdot; \boldsymbol{\theta})$, is observed from a continuous distribution with pdf $f(\cdot, \boldsymbol{\theta})$ and cdf $F(\cdot; \boldsymbol{\theta})$, respectively, using a* PROS(n, S, D) *design. Let $\boldsymbol{\delta}^{(d_r)}(u)$ be the latent variable associated with $X_{(d_r)}$. For any $\lambda \in \{0,1\}$ and any function $G(\cdot)$,*

$$\mathbb{E}\left\{ \sum_{r=1}^{n} \sum_{u \in d_r} \frac{\phi_u(\lambda)\, \boldsymbol{\delta}^{(d_r)}(u)\, G(Y_r)}{\lambda + (1-2\lambda)\, F(Y_r; \boldsymbol{\theta})} \right\} = n(S-1)\mathbb{E}[G(X)],$$

subject to the existence of all involved the expectations.

Proof. Let $\lambda = 0$. By the total law of expectations and equation (4.4) we get

$$\mathbb{E}\left\{\sum_{r=1}^{n}\sum_{u\in d_r}(u-1)\frac{\boldsymbol{\delta}^{(d_r)}(u)\,G(Y_r)}{F(Y_r;\boldsymbol{\theta})}\right\}$$

$$= \frac{1}{m}\sum_{r=1}^{n}\sum_{u\in d_r}(u-1)\int\frac{G(x)}{F(x;\boldsymbol{\theta})}f^{(u:S)}(x;\boldsymbol{\theta})dx$$

$$= \frac{S}{m}\int G(x)f(x;\boldsymbol{\theta})\left\{\sum_{v=1}^{S}(v-1)\binom{S-1}{v-1}[F(x;\boldsymbol{\theta})]^{v-2}[\bar{F}(x;\boldsymbol{\theta})]^{S-v}\right\}dx$$

$$= n(S-1)\mathbb{E}[G(X)],$$

The proof for $\lambda = 1$ is similar and hence omitted. □

Now, we obtain the FI content of $\mathbf{Y}_{pros}$ and compare it with its SRS counterpart of the same size.

Theorem 4.1. *Under the usual regularity conditions, the FI matrix of a complete* PROS(n, S, D) *sample of size n from $f(\cdot;\boldsymbol{\theta})$ is given by*

$$\mathbb{I}_{pros}(\boldsymbol{\theta}) = \mathbb{I}_{srs}(\boldsymbol{\theta}) + \mathbb{K}(\boldsymbol{\theta}),$$

where $\mathbb{I}_{srs}(\boldsymbol{\theta})$ denotes the FI matrix of an SRS of size n,

$$\mathbb{K}(\boldsymbol{\theta}) = n(S-1)\mathbb{E}\left\{\frac{[D_{\boldsymbol{\theta}}F(X;\boldsymbol{\theta})][D_{\boldsymbol{\theta}}F(X;\boldsymbol{\theta})]^{\top}}{F(X;\boldsymbol{\theta})\bar{F}(X;\boldsymbol{\theta})}\right\},$$

is a non-negative definite matrix and the expectation is taken with respect to X.

Proof. Let $Y_r = X_{(d_r)}, r = 1, \ldots, n$. Using (4.1), the log-likelihood function of $\boldsymbol{\theta}$ can be written as

$$l_{pros}(\boldsymbol{\theta}) \propto l_{srs}(\boldsymbol{\theta}) + \Gamma_p(\boldsymbol{\theta}),$$

where

$$\Gamma_p(\boldsymbol{\theta}) = \sum_{r=1}^{n} \sum_{u \in d_r} \sum_{\lambda=0}^{1} \phi_u(\lambda)\, \boldsymbol{\delta}^{(d_r)}(u) \log[\lambda + (1-\lambda)\, F(y_r; \boldsymbol{\theta})],$$

and $l_{srs}(\boldsymbol{\theta})$ is the log-likelihood function of $\boldsymbol{\theta}$ under an SRS sample of size n so that $-\mathbb{E}[D^2_{\boldsymbol{\theta}} l_{srs}(\boldsymbol{\theta})] = \mathbb{I}_{srs}(\boldsymbol{\theta})$. Taking the second derivative of $\Gamma_p(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, one gets

$$\begin{aligned} D^2_{\boldsymbol{\theta}}\Gamma_p(\boldsymbol{\theta}) &= \sum_{r=1}^{n} \sum_{u \in d_r} \sum_{\lambda=0}^{1} (-1)^{\lambda} \phi_u(\lambda) \boldsymbol{\delta}^{(d_r)}(u) \\ &\quad \times \left\{ \frac{D^2_{\boldsymbol{\theta}} F(y_r; \boldsymbol{\theta})}{\lambda + (1-\lambda)\, F(y_r; \boldsymbol{\theta})} - \frac{[D_{\boldsymbol{\theta}} F(y_r; \boldsymbol{\theta})][D_{\boldsymbol{\theta}} F(y_r; \boldsymbol{\theta})]^\top}{[\lambda + (1-\lambda)\, F(y_r; \boldsymbol{\theta})]^2} \right\}. \end{aligned} \quad (4.5)$$

Using Lemma 4.1 with $G(x) = D^2_{\boldsymbol{\theta}} F(x; \boldsymbol{\theta})$, we have

$$\mathbb{E}\left\{ \sum_{r=1}^{n} \sum_{u \in d_r} (u-1) \frac{\boldsymbol{\delta}^{(d_r)}(u)\, G(Y_r)}{F(Y_r; \boldsymbol{\theta})} \right\} = \mathbb{E}\left\{ \sum_{r=1}^{n} \sum_{u \in d_r} (S-u) \frac{\boldsymbol{\delta}^{(d_r)}(u)\, G(Y_r)}{\bar{F}(Y_r; \boldsymbol{\theta})} \right\}. \quad (4.6)$$

Similarly, by choosing $G(x) = \frac{[D_{\boldsymbol{\theta}} F(x;\boldsymbol{\theta})][D_{\boldsymbol{\theta}} F(x;\boldsymbol{\theta})]^\top}{\lambda + (1-2\lambda) F(x;\boldsymbol{\theta})}$ in Lemma 4.1, we obtain

$$\mathbb{E}\left\{ \sum_{r=1}^{n} \sum_{u \in d_r} \frac{\phi_u(\lambda)\, \boldsymbol{\delta}^{(d_r)}(u)\, G(Y_r)}{\lambda + (1-2\lambda) F(Y_r; \boldsymbol{\theta})} \right\} = n(S-1)\mathbb{E}\left\{G(X)\right\}, \quad \lambda \in \{0, 1\}. \quad (4.7)$$

Finally, taking the expectation of $D^2_{\boldsymbol{\theta}}\Gamma_P(\boldsymbol{\theta})$ and using (4.6) and (4.7), we obtain

$$\mathbb{K}(\boldsymbol{\theta}) = -\mathbb{E}[D^2_{\boldsymbol{\theta}}\Gamma_p(\boldsymbol{\theta})] = n(S-1)\mathbb{E}\left\{ \frac{[D_{\boldsymbol{\theta}} F(X; \boldsymbol{\theta})][D_{\boldsymbol{\theta}} F(X; \boldsymbol{\theta})]^\top}{F(X; \boldsymbol{\theta})\bar{F}(X; \boldsymbol{\theta})} \right\}, \quad (4.8)$$

which completes the proof. □

Theorem 4.1 shows that the FI matrix of the complete PROS(n, S, D) sample can be decomposed into the FI matrix of the SRS data and a nonnegative definite matrix, hence $\mathbb{I}_{pros}(\boldsymbol{\theta}) \geq \mathbb{I}_{srs}(\boldsymbol{\theta})$. In other words, complete PROS samples provide more information about the unknown parameter $\boldsymbol{\theta}$ than SRS samples of the same size. It is worth noting that the result of Barabesi and El-Sharaawi (2001) about FI of RSS data can be obtained as a special case of Theorem 4.1 by setting $S = n$. We now compare the FI content about the unknown parameter $\boldsymbol{\theta}$ of the complete PROS sample with that of a RSS sample of the same size.

Theorem 4.2. *Under the conditions of Theorem 4.1, the FI matrix of a complete* PROS(n, S, D) *sample may be decomposed as*

$$\mathbb{I}_{pros}(\boldsymbol{\theta}) = \mathbb{I}_{rss}(\boldsymbol{\theta}) + \mathbb{H}(\boldsymbol{\theta}),$$

where $\mathbb{I}_{rss}(\boldsymbol{\theta})$ is the FI matrix of an RSS of size n (when the set size is n), and

$$\mathbb{H}(\boldsymbol{\theta}) = n(S-n)\mathbb{E}\left\{\frac{[D_{\boldsymbol{\theta}}F(X;\boldsymbol{\theta})][D_{\boldsymbol{\theta}}F(X;\boldsymbol{\theta})]^\top}{F(X;\boldsymbol{\theta})\bar{F}(X;\boldsymbol{\theta})}\right\},$$

is a non-negative definite matrix.

Proof. Using Theorem 4.1 for $S = n$, we have

$$\mathbb{I}_{rss}(\boldsymbol{\theta}) = \mathbb{I}_{srs}(\boldsymbol{\theta}) + n(n-1)\mathbb{E}\left\{\frac{[D_{\boldsymbol{\theta}}F(X;\boldsymbol{\theta})][D_{\boldsymbol{\theta}}F(X;\boldsymbol{\theta})]^\top}{F(X;\boldsymbol{\theta})\bar{F}(X;\boldsymbol{\theta})}\right\},$$

where $\mathbb{I}_{srs}(\boldsymbol{\theta})$ denotes the FI matrix of a SRS of size n. Now, the result follows from the above equation and the expression for $\mathbb{I}_{pros}(\boldsymbol{\theta})$ in Theorem 4.1. $\square$

Theorem 4.2 shows the superiority of a complete PROS sample over an RSS of the same size in terms of the FI content about the unknown vector of parameters $\boldsymbol{\theta}$. In the following examples we obtain the FI content of a complete PROS sample from a location-scale family of distributions as well as a simple linear regression model and compare them with those based on SRS and RSS data of the same size. To this end, let

$$RE_1(\boldsymbol{\theta}) = \frac{det\{\mathbb{I}_{pros}(\boldsymbol{\theta})\}}{det\{\mathbb{I}_{srs}(\boldsymbol{\theta})\}} \quad \text{and} \quad RE_2(\boldsymbol{\theta}) = \frac{det\{\mathbb{I}_{pros}(\boldsymbol{\theta})\}}{det\{\mathbb{I}_{rss}(\boldsymbol{\theta})\}}.$$

From Theorems 4.1 and 4.2 one can notice that the set size (S) and the number of subsets (n) are two important parameters that influence the FI content of PROS samples. In addition, the matrix $\mathbb{I}_{pros}$ and $\mathbb{I}_{rss}$ are of order nS and n^2 while $\mathbb{I}_{srs}$ is of order n, then RE_1 and RE_2 are obviously of order S^p and $\{S/n\}^p$, respectively, where p indicates the number of unknown parameters of the underlying model. So, both RE_1 and RE_2 increase with the number of the parameters of the model.

Example 4.1. ***(Location-Scale family of distributions)***. *Under the assumptions of Theorem 4.1, if $f(x;\boldsymbol{\theta})$ is a member of the location-scale family of distributions with pdf*

$$f(x;\boldsymbol{\theta}) = \frac{1}{\sigma} g(\frac{x-\mu}{\sigma}), \quad \boldsymbol{\theta} = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+,$$

where $g(\cdot)$ is a pdf with corresponding cdf $G(\cdot)$, then

$$\begin{aligned}
\mathbb{I}_{pros}(\boldsymbol{\theta}) =& \mathbb{I}_{srs}(\boldsymbol{\theta}) + \mathbb{K}(\boldsymbol{\theta}) \\
=& \frac{n}{\sigma^2} \begin{pmatrix} \mathbb{E}\{\frac{g^{'}(Z)^2}{g(Z)^2}\} & \mathbb{E}\{\frac{Zg^{'}(Z)^2}{g(Z)^2}\} \\ \mathbb{E}\{\frac{Zg^{'}(Z)^2}{g(Z)^2}\} & \mathbb{E}\{\frac{Z^2g^{'}(Z)^2}{g(Z)^2} - 1\} \end{pmatrix} \\
&+ \frac{n(S-1)}{\sigma^2} \begin{pmatrix} \mathbb{E}\{\frac{g(Z)^2}{G(Z)[1-G(Z)]}\} & \mathbb{E}\{\frac{Zg(Z)^2}{G(Z)[1-G(Z)]}\} \\ \mathbb{E}\{\frac{Zg(Z)^2}{G(Z)[1-G(Z)]}\} & \mathbb{E}\{\frac{Z^2g(Z)^2}{G(Z)[1-G(Z)]}\} \end{pmatrix}.
\end{aligned}$$

In the specific case where $f(x;\boldsymbol{\theta})$ is symmetric about μ, the FI matrix reduces to

$$\begin{aligned}
\mathbb{I}_{pros}(\boldsymbol{\theta}) =& \frac{n}{\sigma^2} \begin{pmatrix} \mathbb{E}\{\frac{g^{'}(Z)^2}{g(Z)^2}\} & 0 \\ 0 & \mathbb{E}\{\frac{Z^2g^{'}(Z)^2}{g(Z)^2} - 1\} \end{pmatrix} \\
&+ \frac{n(S-1)}{\sigma^2} \begin{pmatrix} \mathbb{E}\{\frac{g(Z)^2}{G(Z)[1-G(Z)]}\} & 0 \\ 0 & \mathbb{E}\{\frac{Z^2g(Z)^2}{G(Z)[1-G(Z)]}\} \end{pmatrix}.
\end{aligned}$$

Tables 4.1 shows the values of RE_1 and RE_2 for some location-scale families of distributions. As expected, the largest values of RE_1 and RE_2 are achieved for the cases where both location and scale parameters are unknown.

Example 4.2. *(**Linear Regression Model**). In this example, the* $\mathrm{PROS}(n,S,D)$ *sampling design is used for inference on the simple regression model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where, for each x_i, $i = 1,\ldots,k$, we have a PROS sample of Y's denoted by $(Y_{i(d_r)},\ldots,Y_{i(d_n)})$. Suppose ϵ_i are i.i.d random variables from a symmetric distribution with pdf $f(\cdot)$ and cdf $F(\cdot)$, respectively. Let $\mathbb{E}(\epsilon_i) = 0$ and $var(\epsilon_i) = \sigma^2$. Without loss of generality, we take $\bar{x} = \frac{1}{k}\sum_{i=1}^{k} x_i = 0$, $s_x^2 = \frac{1}{k}\sum_{i=1}^{k} x_i^2$ and let*

$\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma)$. Using Example 4.1, it is easy to show that

$$\begin{aligned}
\mathbb{I}_{srs}(\boldsymbol{\theta}) &= \sum_{i=1}^{k} \frac{n}{\sigma^2} \begin{pmatrix} \mathbb{E}\{\frac{f^{'}(Z)^2}{f(Z)^2}\} & x_i\mathbb{E}\{\frac{f^{'}(Z)^2}{f(Z)^2}\} & 0 \\ x_i\mathbb{E}\{\frac{f^{'}(Z)^2}{f(Z)^2}\} & x_i^2\mathbb{E}\{\frac{f^{'}(Z)^2}{f(Z)^2}\} & 0 \\ 0 & 0 & \mathbb{E}\{\frac{Z^2 f^{'}(Z)^2}{f(Z)^2}\} - 1 \end{pmatrix} \\
&= \frac{nk}{\sigma^2} diag\left(\mathbb{E}\{\frac{f^{'}(Z)^2}{f(Z)^2}\}, s_x^2\mathbb{E}\{\frac{f^{'}(Z)^2}{f(Z)^2}\}, \mathbb{E}\{\frac{Z^2 f^{'}(Z)^2}{f(Z)^2}\} - 1\right),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{K}(\boldsymbol{\theta}) &= \sum_{i=1}^{k} \frac{2n(S-1)}{\sigma^2} \begin{pmatrix} \mathbb{E}\{\frac{f(Z)^2}{F(Z)}\} & x_i\mathbb{E}\{\frac{f(Z)^2}{F(Z)}\} & 0 \\ x_i\mathbb{E}\{\frac{f(Z)^2}{F(Z)}\} & x_i^2\mathbb{E}\{\frac{f(Z)^2}{F(Z)}\} & 0 \\ 0 & 0 & \mathbb{E}\{\frac{Z^2 f^{'}(Z)^2}{F(Z)}\} \end{pmatrix} \\
&= \frac{2kn(S-1)}{\sigma^2} diag\left(\mathbb{E}\{\frac{f(Z)^2}{F(Z)}\}, s_x^2\mathbb{E}\{\frac{f(Z)^2}{F(Z)}\}, \mathbb{E}\{\frac{Z^2 f^{'}(Z)^2}{F(Z)}\}\right).
\end{aligned}$$

Note that $RE_1(\boldsymbol{\theta})$ is independent of x_i and $\boldsymbol{\theta}$ and it only depends on the pdf $f(\cdot)$ and its corresponding cdf $F(\cdot)$. As a special case, when the ϵ_is are normally distributed, we get

$$RE_1(\boldsymbol{\theta}) = \{1 + 0.4805(S-1)\}^2 \{1 + 0.1350(S-1)\}.$$

When $S = n$, this reduces to the result of Barabesi and El-Sharaawi (2001) for RSS data.

Table 4.1: The values of $RE_i(\boldsymbol{\theta})$, $i = 1, 2$ for comparing the FI content of the complete PROS(n, S, D) sample with its SRS and RSS counterparts using the same size for some distributions.

Distributions	Location	Scale	Shape	RE_1	RE_2
Exponential	0	σ	-	$1{+}0.4041(S-1)$	$1{+}0.4041\{\frac{(S-n)}{1+0.4041(n-1)}\}$
Normal	μ	1	-	$1{+}0.4805(S-1)$	$1{+}0.4805\{\frac{(S-n)}{1+0.4805(n-1)}\}$
	0	σ	-	$1{+}0.1350(S-1)$	$1{+}0.1350\{\frac{(S-n)}{1+0.1350(n-1)}\}$
	μ	σ	-	$1{+}0.6155(S-1)\ {+}0.0649(S-1)^2$	$1{+}(\frac{0.6155(S-n)+0.0649[(S-1)^2-(n-1)^2]}{1+0.6155(n-1)+0.0649(n-1)^2})$
Logistic	μ	1	-	$1{+}0.0050(S-1)$	$1{+}0.1666\{\frac{(S-n)}{0.3332+0.1666(n-1)}\}$
	0	σ	-	$1{+}0.1513(S-1)$	$1{+}0.2149\{\frac{(S-n)}{1.4189+0.2149(n-1)}\}$
	μ	σ	-	$1{+}0.6516(S-1)\ {+}0.0757(S-1)^2$	$1{+}(\frac{0.3081(S-n)+0.0358[(S-1)^2-(n-1)^2]}{0.4728+0.3081(n-1)+0.0358(n-1)^2})$
Extreme-value	μ	1	-	$1{+}0.4041(S-1)$	$1{+}0.4041\{\frac{(S-n)}{1+0.4041(n-1)}\}$
	0	σ	-	$1{+}0.2519(S-1)$	$1{+}0.2518\{\frac{(S-n)}{1+0.2518(n-1)}\}$
	μ	σ	-	$1{+}0.6012(S-1)\ {+}0.0686(S-1)^2$	$1{+}(\frac{0.6560(S-n)+0.1017[(S-1)^2-(n-1)^2]}{1+0.6560(n-1)+0.1017(n-1)^2})$
Gamma	0	σ	2	$1{+}0.4393(S-1)$	$1{+}0.7296\{\frac{(S-n)}{1.6609+0.7296(n-1)}\}$
	0	σ	3	$1{+}0.4523(S-1)$	$1{+}1.1690\{\frac{(S-n)}{2.5846+1.1690(n-1)}\}$
	0	σ	4	$1{+}0.4591(S-1)$	$1{+}1.6161\{\frac{(S-n)}{3.5200+1.6161(n-1)}\}$
	0	σ	10	$1{+}0.4718(S-1)$	$1{+}4.2396\{\frac{(S-n)}{8.9820+4.2396(n-1)}\}$

4.2.2 FI matrix of $\mathbf{X}_{pros}$ and the effect of misplacement errors

In this section we obtain the FI matrix of $\mathbf{X}_{pros}$. We study a general setting where it is assumed that the sub-setting process of the PROS(n, S, D) design is subject to misplacement errors between the groups. For example, when the actual rank of a unit is in the judgment subset d_r, due to judgment ranking error it could have been misplaced into another judgment subset, say d_s, $r \neq s$, which leads to a different kind of ranking error than the one usually happening in RSS. This is a very general setting where the FI matrix of $\mathbf{X}_{pros}$ under the perfect sub-setting assumption can also be obtained as a special case. We use the missing data model proposed by Arslan and Ozturk (2013) to model possible misplacement errors in PROS sampling design. Let $\mathbf{X}_{pros} = \{X_{[d_r]}, r = 1, \ldots, n\}$ denote an imperfect PROS sample where $[\cdot]$ is used to show the presence of misplacement errors in the sub-setting process. When the sub-setting process is perfect we simply use $X_{(d_r)}$ to denote PROS observations. Let $\boldsymbol{\alpha}$ denote the misplacement probability matrix,

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{d_1,d_1} & \alpha_{d_1,d_2} & \cdots & \alpha_{d_1,d_n} \\ \alpha_{d_2,d_1} & \alpha_{d_2,d_2} & \cdots & \alpha_{d_2,d_n} \\ \vdots & \vdots & \cdots & \vdots \\ \alpha_{d_n,d_1} & \alpha_{d_n,d_2} & \cdots & \alpha_{d_n,d_n} \end{bmatrix}_{n\times n},$$

where α_{d_r,d_h} is the probability of misplacing a unit from subset d_h into subset d_r. Since the design parameter D creates a partition over the sets, the matrix $\boldsymbol{\alpha}$ should be a doubly stochastic matrix, that is, a matrix such that $\sum_{r=1}^{n} \alpha_{d_r,d_h} = \sum_{h=1}^{n} \alpha_{d_r,d_h} = 1$. Suppose $f_{[d_r]}(\cdot; \boldsymbol{\theta})$ is the pdf of $X_{[d_r]}$, $r = 1, \ldots, n$. One can easily show that

$$f_{[d_r]}(x_{[d_r]};\boldsymbol{\theta}) = \sum_{h=1}^{n} \alpha_{d_r,d_h} f_{(d_h)}(x_{[d_r]};\boldsymbol{\theta}) = f(x_{[d_r]};\boldsymbol{\theta}) g_r(x_{[d_r]};\boldsymbol{\theta}), \tag{4.9}$$

where

$$g_r(x;\boldsymbol{\theta}) = n \sum_{h=1}^{n} \sum_{u \in d_h} \alpha_{d_r,d_h} \binom{S-1}{u-1} [F(x;\boldsymbol{\theta})]^{u-1} [1 - F(x;\boldsymbol{\theta})]^{S-u}. \tag{4.10}$$

The likelihood function under an imperfect PROS(n, S, D) design is now given by

$$L(\Omega) = \prod_{r=1}^{n} f_{[d_r]}(x_{[d_r]};\boldsymbol{\theta}) = \prod_{r=1}^{n} f(x_{[d_r]};\boldsymbol{\theta}) g_r(x_{[d_r]};\boldsymbol{\theta}),$$

where $\Omega = (\boldsymbol{\theta}, \boldsymbol{\alpha})$. To obtain the FI matrix of an imperfect PROS sample and compare it with its SRS and RSS counterparts, we need the following result, the proof of which is left to the reader.

Lemma 4.2. *Let $Y_r = X_{[d_r]}$, $r = 1, \ldots, n$ be observed from a continuous distribution with pdf $f(\cdot;\boldsymbol{\theta})$ using an imperfect* PROS(n, S, D) *sampling design. Suppose $f_{[d_r]}(\cdot;\boldsymbol{\theta})$ and $g_r(\cdot,\boldsymbol{\theta})$ are defined as in (4.9) and (4.10), respectively. Under the regularity conditions Chen et al. (2004), we have*

(i) $\sum_{r=1}^{n} f_{[d_r]}(x;\boldsymbol{\theta}) = n f(x;\boldsymbol{\theta})$,

(ii) $\sum_{r=1}^{n} g_r(x;\boldsymbol{\theta}) = n$,

(iii) $\sum_{r=1}^{n} \mathbb{E}\left\{\frac{D^2_{\boldsymbol{\theta}} g_r(Y_r;\boldsymbol{\theta})}{g_r(Y_r;\boldsymbol{\theta})}\right\} = 0$,

(iv) $\sum_{r=1}^{n} \mathbb{E}\left\{\frac{[D_{\boldsymbol{\theta}} g_r(Y_r;\boldsymbol{\theta})][D_{\boldsymbol{\theta}} g_r(Y_r;\boldsymbol{\theta})]^\top}{g_r^2(Y_r;\boldsymbol{\theta})}\right\} = \sum_{r=1}^{n} \mathbb{E}\left\{\frac{[D_{\boldsymbol{\theta}} g_r(X;\boldsymbol{\theta})][D_{\boldsymbol{\theta}} g_r(X;\boldsymbol{\theta})]^\top}{g_r(X;\boldsymbol{\theta})}\right\}.$

Now, we show that the FI content of $\mathbf{X}_{pros}$ is more than that of its SRS counterpart. Unfortunately, it is hard to obtain analytical results to compare the FI content of PROS and RSS data, therefore, we should rely on numerical studies for this case (see Tables 4.2 and 4.3).

Theorem 4.3. *Under the conditions of Lemma 4.2, the FI matrix of an imperfect* PROS(n, S, D) *sample about unknown parameters* $\Omega = (\boldsymbol{\alpha}, \boldsymbol{\theta})$ *is given by*

$$\begin{aligned}\mathbb{I}_{ipros}(\Omega) &= \mathbb{I}_{srs}(\boldsymbol{\theta}) + \sum_{r=1}^{n} \mathbb{E}\left\{\frac{[D_{\boldsymbol{\theta}} g_r(X;\boldsymbol{\theta})][D_{\boldsymbol{\theta}} g_r(X;\boldsymbol{\theta})]^\top}{g_r(X;\boldsymbol{\theta})}\right\} \\ &= \mathbb{I}_{srs}(\boldsymbol{\theta}) + \sum_{r=1}^{n} \tilde{\Delta}_r,\end{aligned}$$

where $\sum_{r=1}^{n} \tilde{\Delta}_r$ *is a nonnegative definite matrix.*

Proof. The proof is similar to the proof of Theorem 4.1 and hence is omitted. □

To study the effect of misplacement errors in the sub-setting process of PROS(n, S, D) design, on the information content of data, we consider the following misplacement probability matrices when $n = 2$ and $n = 3$,

$$\boldsymbol{\alpha}_1 = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} \quad \text{and} \quad \boldsymbol{\alpha}_2 = \begin{bmatrix} p & \frac{1-p}{2} & \frac{1-p}{2} \\ \frac{1-p}{2} & p & \frac{1-p}{2} \\ \frac{1-p}{2} & \frac{1-p}{2} & p \end{bmatrix}.$$

Table 4.2: Values of RE_1 and RE_2 for comparing the FI content of imperfect PROS data with its SRS and RSS counterparts with the same sample size for some distributions (Dist.) including Normal (Nor.), Exponential (Exp.) and Logistic (Log.) when $S = 6$.

			p										
Dist.	n	Design	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Nor.	2	RE_1	2.37	1.65	1.33	1.13	1.03	1.00	1.03	1.13	1.33	1.65	2.37
		RE_2	1.44	1.24	1.13	1.06	1.01	1.00	1.01	1.06	1.13	1.24	1.44
	3	RE_1	1.75	1.27	1.07	1.00	1.01	1.10	1.27	1.52	1.87	2.39	2.78
		RE_2	1.24	1.10	1.03	1.00	1.00	1.04	1.10	1.17	1.24	1.33	1.44
Exp.	2	RE_1	1.91	1.46	1.23	1.10	1.02	1.00	1.02	1.10	1.23	1.46	1.91
		RE_2	1.36	1.20	1.11	1.04	1.01	1.00	1.01	1.04	1.11	1.20	1.36
	3	RE_1	1.47	1.18	1.05	1.00	1.01	1.07	1.18	1.35	1.58	1.90	2.45
		RE_2	1.18	1.07	1.02	1.00	1.00	1.03	1.07	1.12	1.19	1.26	1.35
Log.	2	RE_1	2.70	1.77	1.38	1.16	1.03	1.00	1.03	1.16	1.38	1.77	2.70
		RE_2	1.56	1.29	1.16	1.07	1.01	1.00	1.01	1.07	1.16	1.29	1.56
	3	RE_1	1.88	1.30	1.08	1.00	1.01	1.11	1.31	1.61	2.05	2.75	4.16
		RE_2	1.31	1.12	1.03	1.00	1.00	1.05	1.11	1.20	1.30	1.42	1.59

For some members of the location-scale family of distributions, numerical values of $RE_1(\boldsymbol{\theta})$ and $RE_2(\boldsymbol{\theta})$ are calculated to compare the FI content of imperfect PROS data with their SRS and imperfect RSS (with ranking error models $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ for $n = 2$ and $n = 3$, respectively) counterparts of the same size when $S = 6$ and $S = 12$. These values are reported in Tables 4.2 and 4.3, respectively. Both tables show that misplacement errors in the subsetting process of PROS sampling have considerable effect on the information content of PROS data about the unknown parameters of the model. Note that, when the subsetting process is done randomly, i.e., $p = 1/2$ when $n = 2$ and $p = 1/3$ in the case $n = 3$, the FI content of PROS data is the same as the FI content of SRS and RSS samples of the same size.

4.3 Other information criteria

The concept of information and uncertainty of random samples is so rich that several measures have been proposed to study its different aspects. For example, in engineering studies, the Shannon entropy, Rényi entropy and KL information mea-

Table 4.3: Values of RE_1 and RE_2 for comparing the FI content of imperfect PROS data with its SRS and RSS counterparts with the same sample size for some distributions (Dist.) including Normal (Nor.), Exponential (Exp.) and Logistic (Log.) when $S = 12$.

			p										
Dist.	n	Design	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Nor.	2	RE_1	3.01	1.94	1.48	1.20	1.04	1.00	1.04	1.20	1.48	1.94	3.01
		RE_2	1.83	1.46	1.26	1.12	1.03	1.00	1.03	1.12	1.26	1.46	1.83
	3	RE_1	2.34	1.47	1.13	1.00	1.02	1.17	1.45	1.85	2.43	3.27	4.89
		RE_2	1.67	1.27	1.08	1.00	1.01	1.11	1.25	1.42	1.61	1.81	2.10
Exp.	2	RE_1	2.39	1.68	1.35	1.14	1.03	1.00	1.03	1.14	1.35	1.68	2.39
		RE_2	1.70	1.38	1.21	1.09	1.02	1.00	1.02	1.09	1.21	1.38	1.70
	3	RE_1	1.88	1.32	1.09	1.00	1.02	1.12	1.31	1.59	1.97	2.49	3.40
		RE_2	1.50	1.20	1.06	1.00	1.01	1.08	1.19	1.32	1.48	1.65	1.88
Log.	2	RE_1	3.56	2.13	1.56	1.23	1.05	1.00	1.05	1.23	1.56	2.13	3.56
		RE_2	2.06	1.56	1.31	1.14	1.03	1.00	1.03	1.14	1.31	1.56	2.06
	3	RE_1	2.73	1.55	1.15	1.00	1.03	1.20	1.53	2.04	2.83	4.07	6.72
		RE_2	1.90	1.33	1.10	1.00	1.02	1.12	1.30	1.53	1.80	2.10	2.57

sures are used more than FI to quantify the information and uncertainty structures of random samples. These measures quantify the amount of uncertainty inherent in the joint probability distribution of a random sample and have been applied in many areas such as ecological studies, computer science and information technology, in different context including order statistics, spacings, censored data, reliability, life testing, record data and text analysis. For more details see Jafari Jozani and Ahmadi (2014) and Johnson (2004) and references therein.

In this section, we compare the uncertainty measures including Shannon entropy, Rényi entropy and KL information of PROS samples with SRS and RSS samples of the same size. Jafari Jozani and Ahmadi (2014) compared these uncertainty measures of the RSS data with those under SRS data. The results of the following sections extend the results of Jafari Jozani and Ahmadi (2014) to PROS sampling technique. Throughout this section, the sub-setting process of PROS design and the ranking process of RSS are assumed to be perfect.

4.3.1 Shannon entropy of PROS samples

Let X be a continuous random variable with pdf $f(\cdot;\boldsymbol{\theta})$. The Shannon entropy associated with X, is defined as

$$H(X;\boldsymbol{\theta}) = -\int f(x;\boldsymbol{\theta}) \log f(x;\boldsymbol{\theta}) dx,$$

subject to the existence of the integral. The Shannon entropy, as a quantitative measure of information (uncertainty) is extensively used in information technology and computer science and other engineering fields. In practice, smaller values of the Shannon entropy are more desirable (see Johnson, 2004). The Shannon entropy content of an SRS of size n is given by

$$H_n(\mathbf{X}_{srs};\boldsymbol{\theta}) = -\sum_{i=1}^{n} \int f(x;\boldsymbol{\theta}) \log f(x;\boldsymbol{\theta}) dx = n\, H(X_1;\boldsymbol{\theta}).$$

Similarly, for an RSS of size n (with the set size n)

$$H_n(\mathbf{X}_{rss};\boldsymbol{\theta}) = -\sum_{i=1}^{n} \int f^{(i:n)}(x;\boldsymbol{\theta}) \log f^{(i:n)}(x;\boldsymbol{\theta}) dx,$$

where $f^{(i:n)}(\cdot;\boldsymbol{\theta})$ is the pdf of the i-th order statistic in a SRS of size n from $f(\cdot;\boldsymbol{\theta})$. Furthermore, for a PROS(n, S, D) sample, it is easy to see that

$$H_n(\mathbf{X}_{pros};\boldsymbol{\theta}) = -\sum_{r=1}^{n} \int f_{(d_r)}(y;\boldsymbol{\theta}) \log f_{(d_r)}(y;\boldsymbol{\theta}) dy.$$

In the following lemma, we show that the Shannon entropy of a PROS sample is smaller than that of an SRS sample of the same size. Unfortunately, we are not able to obtain an ordering relationship among the Shannon entropy of RSS and PROS data for sample of the same size. Instead, we obtain a lower bound for the Shannon entropy of a PROS(n, S, D) sample in terms of the Shannon entropy of an RSS data of size S when the set size is S.

Lemma 4.3. *Let $\mathbf{X}_{pros}$ be a PROS(n, S, D) sample from a population with pdf $f(\cdot;\boldsymbol{\theta})$ and let $m = S/n$ be the number of observations in each subset. Suppose $\mathbf{X}_{srs}$ is a SRS of size n from $f(\cdot;\boldsymbol{\theta})$ with Shannon entropy $H_n(\mathbf{X}_{srs};\boldsymbol{\theta})$ and let $H_S(\mathbf{X}_{rss};\boldsymbol{\theta})$ represent the Shannon entropy of an RSS of size S when the set size is S. Then,*

$$\frac{1}{m}H_S(\mathbf{X}_{rss};\boldsymbol{\theta}) \leq H_n(\mathbf{X}_{pros};\boldsymbol{\theta}) \leq H_n(\mathbf{X}_{srs};\boldsymbol{\theta}), \qquad \textit{for all } n \in N.$$

Proof. Using (4.3) and let convexity of $h(t) = t\log t, t > 0$, we have

$$\begin{aligned} H_n(\mathbf{X}_{pros};\boldsymbol{\theta}) &\leq -n\int\left(\frac{1}{n}\sum_{r=1}^{n} f_{(d_r)}(x;\boldsymbol{\theta})\right)\left(\log\left[\frac{1}{n}\sum_{r=1}^{n} f_{(d_r)}(x;\boldsymbol{\theta})\right]\right)dx \\ &= H_n(\mathbf{X}_{srs};\boldsymbol{\theta}). \end{aligned}$$

Furthermore, using (4.2) and the convexity of $h(t) = t\log t, t > 0$, we have

$$\begin{aligned} H_n(\mathbf{X}_{pros};\boldsymbol{\theta}) &= -\sum_{r=1}^{n}\int\left(\frac{1}{m}\sum_{u\in d_r} f^{(u:S)}(x;\boldsymbol{\theta})\right)\left(\log\left[\frac{1}{m}\sum_{u\in d_r} f^{(u:S)}(x;\boldsymbol{\theta})\right]\right)dx \\ &\geq -\frac{1}{m}\sum_{r=1}^{n}\sum_{u\in d_r}\int f^{(u:S)}(x;\boldsymbol{\theta})\log f^{(u:S)}(x;\boldsymbol{\theta})dx \\ &= \frac{1}{m}H_S(\mathbf{X}_{rss};\boldsymbol{\theta}), \end{aligned}$$

which completes the proof. □

4.3.2 Rényi entropy of PROS samples

In this section we use Rényi entropy as a quantitative measure to quantify the entropy associated with PROS data $\mathbf{X}_{pros}$. The Réyni entropy of a random variable X with pdf $f(\cdot;\boldsymbol{\theta})$ is defined as follows

$$H_{\alpha,1}(X;\boldsymbol{\theta}) = \frac{1}{1-\alpha}\log \mathbb{E}[f^{\alpha-1}(X;\boldsymbol{\theta})],$$

where $\alpha > 0, \alpha \neq 1$. The Rényi entropy is a very general measure and includes the Shannon entropy as its special case due to the following relationship

$$\lim_{\alpha\to 1} H_{\alpha,1}(X;\boldsymbol{\theta}) = -\int f(x;\boldsymbol{\theta})\log f(x;\boldsymbol{\theta})dx = H(X;\boldsymbol{\theta}).$$

Due to the flexibility of the Rényi entropy, $H_{\alpha,1}(X;\boldsymbol{\theta})$ has been used in many fields such as statistics, ecology, engineering and etc. We derive the Rényi entropy of $\mathbf{X}_{pros}$ and compare it with the Rényi entropy of $\mathbf{X}_{srs}$. We present the results for $0 < \alpha < 1$ and the case with $\alpha > 1$, which requires further investigation, will be presented in later work. To this end, the Rényi entropy of an SRS of size n is given by

$$H_{\alpha,n}(\mathbf{X}_{srs};\boldsymbol{\theta}) = \frac{1}{1-\alpha}\sum_{i=1}^{n}\log\int f^{\alpha}(x_i;\boldsymbol{\theta})\,dx_i = n\,H_{\alpha,1}(X;\boldsymbol{\theta});$$

and for an RSS with set size n,

$$H_{\alpha,n}(\mathbf{X}_{rss};\boldsymbol{\theta}) = \frac{1}{1-\alpha}\sum_{i=1}^{n}\log\int [f^{(i:n)}(x;\boldsymbol{\theta})]^{\alpha}dx.$$

Also, for a PROS(n,S,D) sample, one gets

$$H_{\alpha,n}(\mathbf{X}_{pros};\boldsymbol{\theta}) = \frac{1}{1-\alpha}\sum_{r=1}^{n}\log\int[f_{(d_r)}(x;\boldsymbol{\theta})]^{\alpha}dx.$$

Lemma 4.4. *Let $H_{\alpha,n}(\mathbf{X}_{pros};\boldsymbol{\theta})$ represent the Rényi entropy of a* PROS(n,S,D) *sample of size n from a population with pdf $f(\cdot;\boldsymbol{\theta})$. Suppose $\mathbf{X}_{srs}$ and $\mathbf{X}^{*}_{rss}$ are, respective an SRS of size n and a RSS of size S (with the set size S) from $f(\cdot;\boldsymbol{\theta})$. For any $0<\alpha<1$ and all $n\in N$, we have*

$$\frac{1}{m}H_{\alpha,S}(\mathbf{X}^{*}_{rss};\boldsymbol{\theta}) \leq H_{\alpha,n}(\mathbf{X}_{pros};\boldsymbol{\theta}) \leq H_{\alpha,n}(\mathbf{X}_{srs};\boldsymbol{\theta}).$$

Proof. By using (4.2) and the concavity of the functions $h_1(t)=\log t$ and $h_2(t)=t^{\alpha}$, we have

$$\begin{aligned}
H_{\alpha,n}(\mathbf{X}_{pros};\boldsymbol{\theta}) &\leq \frac{n}{1-\alpha}\left[\log\int\frac{1}{n}\sum_{r=1}^{n}\left(\frac{1}{m}\sum_{u\in d_r}f^{(u:S)}(x;\boldsymbol{\theta})\right)^{\alpha}dx\right] \\
&\leq \frac{n}{1-\alpha}\log\int\left(\frac{1}{S}\sum_{r=1}^{n}\sum_{u\in d_r}f^{(u:S)}(x;\boldsymbol{\theta})\right)^{\alpha}dx \\
&= H_{\alpha,n}(\mathbf{X}_{srs};\boldsymbol{\theta}).
\end{aligned}$$

Similarly, one can show the following inequalities

$$\begin{aligned}
H_{\alpha,n}(\mathbf{X}_{pros};\boldsymbol{\theta}) &\geq \frac{1}{1-\alpha}\sum_{r=1}^{n}\log\left(\frac{1}{m}\sum_{u\in d_r}\int[f^{(u:S)}(x;\boldsymbol{\theta})]^{\alpha}dx\right) \\
&\geq \frac{1}{m(1-\alpha)}\sum_{r=1}^{n}\sum_{u\in d_r}\log\left(\int[f^{(u:S)}(x;\boldsymbol{\theta})]^{\alpha}dx\right) \\
&= \frac{1}{m}H_{\alpha,S}(\mathbf{X}^{*}_{rss};\boldsymbol{\theta}),
\end{aligned}$$

which complete the proof. $\square$

4.3.3 KL information in PROS samples

The Kullback-Leibler (KL) distance is another measure to quantify the information of a random phenomenon by comparing two probability density functions of the random experiment. Consider two pdfs $f(\cdot;\boldsymbol{\theta})$ and $g(\cdot;\boldsymbol{\theta})$. The KL information measure based on $f(\cdot;\boldsymbol{\theta})$ and $g(\cdot;\boldsymbol{\theta})$ is defined by

$$K(f,g) = \int f(t;\boldsymbol{\theta}) \log\left(\frac{f(t;\boldsymbol{\theta})}{g(t;\boldsymbol{\theta})}\right) dt,$$

which quantifies the information lost by using $g(\cdot;\boldsymbol{\theta})$ for the density of the random variable X instead of $f(\cdot;\boldsymbol{\theta})$. In this section, using the KL measure we make a comparison between PROS, SRS and RSS designs to determine which design provides more informative samples from the underlying population. To this end, we use

$$K\left(L_{pros}(\boldsymbol{\theta}|\mathbf{y}), L_{srs}(\boldsymbol{\theta}|\mathbf{y})\right) = \int \cdots \int L_{pros}(\boldsymbol{\theta}|\mathbf{y}) \log\left(\frac{L_{pros}(\boldsymbol{\theta}|\mathbf{y})}{L_{srs}(\boldsymbol{\theta}|\mathbf{y})}\right) d\mathbf{y}, \tag{4.11}$$

to compare the PROS(n,S,D) and SRS designs, where $L_{pros}(\boldsymbol{\theta}|\mathbf{y})$ and $L_{srs}\boldsymbol{\theta}|\mathbf{y})$ denote the likelihood functions of PROS and SRS samples of the same size, respectively. The KL information measure for comparing RSS and SRS is defined similarly by using (4.11) and setting $S = n$ in the PROS sampling design. One can interpret (4.11) in terms of a hypothesis testing problem within the Neyman-Pearson log-likelihood ratio testing framework (see Johnson, 2004).

Lemma 4.5. *Let $L_{pros}(\boldsymbol{\theta}|\mathbf{y})$ and $L_{srs}(\boldsymbol{\theta}|\mathbf{y})$ denote, respectively, the likelihood functions of a* PROS(n,S,D) *sample and a SRS of size n from a population with pdf*

$f(\cdot;\boldsymbol{\theta})$. *Then we have*

$$K\left(L_{pros}(\boldsymbol{\theta}|\mathbf{y}), L_{srs}(\boldsymbol{\theta}|\mathbf{y})\right) = \sum_{r=1}^{n} \int f_{(d_r)}(y;\boldsymbol{\theta}) \log\left(\frac{f_{(d_r)}(y;\boldsymbol{\theta})}{f(y;\boldsymbol{\theta})}\right) dy.$$

Proof. To show the result, using (4.11) we have

$$\begin{aligned} K\left(L_{pros}(\boldsymbol{\theta}|\mathbf{y}), L_{srs}(\boldsymbol{\theta}|\mathbf{y})\right) &= \sum_{r=1}^{n} \int \cdots \int \left\{\prod_{h=1}^{n} f_{(d_h)}(y_h;\boldsymbol{\theta})\right\} \log\left(\frac{f_{(d_r)}(y_r;\boldsymbol{\theta})}{f(y_r;\boldsymbol{\theta})}\right) \left\{\prod_{j=1}^{n} dy_j\right\} \\ &= \sum_{r=1}^{n} \int f_{(d_r)}(y;\boldsymbol{\theta}) \log\left(\frac{f_{(d_r)}(y;\boldsymbol{\theta})}{f(y;\boldsymbol{\theta})}\right) dy, \end{aligned}$$

where the last equality follows from the independence of observations and the fact that $n-1$ of the integrals are 1. □

In the following lemma, we show that the KL distance between the likelihoods of PROS and SRS samples is greater than the KL information distance between the likelihoods of two SRS samples. Hence, PROS sampling scheme provides more informative sample from the underlying population compared with SRS sampling scheme. We also obtain a lower bound for the KL information between the likelihoods of PROS and SRS data of the same size.

Lemma 4.6. *Let $L_{pros}(\boldsymbol{\theta}|\mathbf{y})$ denote the likelihood function of a* PROS(n, S, D) *sample from a population with pdf $f(\cdot,\boldsymbol{\theta})$. Suppose $L_{srs,1}(\boldsymbol{\theta}|\mathbf{y})$ and $L_{srs,2}(\boldsymbol{\theta}|\mathbf{y})$ denote the likelihood functions of simple random samples of size n from $f(\cdot;\boldsymbol{\theta})$ and $g(\cdot;\boldsymbol{\theta})$, respectively. In addition, let $L_{rss^*}(\boldsymbol{\theta}|\mathbf{y})$ represent the likelihood function of a RSS of size S with set size is S. Then,*

$$K\left(L_{srs,1}(\boldsymbol{\theta}|\mathbf{y}), L_{srs,2}(\boldsymbol{\theta}|\mathbf{y})\right) \le K\left(L_{pros}(\boldsymbol{\theta}|\mathbf{y}), L_{srs,2}(\boldsymbol{\theta}|\mathbf{y})\right) \le \frac{1}{m} K\left(\tilde{L}_{rss^*}(\boldsymbol{\theta}|\mathbf{y}), L_{srs,2}(\boldsymbol{\theta}|\mathbf{y})\right).$$

Proof. Applying Lemma 4.5 and using the convexity of $h(t) = t \log t$, $t > 0$, we derive

$$\begin{aligned}
& K\left(L_{pros}(\boldsymbol{\theta}|\mathbf{y}), L_{srs,2}(\boldsymbol{\theta}|\mathbf{y})\right) \\
= & \sum_{r=1}^{n} \int g(y;\boldsymbol{\theta}) \left(\frac{f_{(d_r)}(y;\boldsymbol{\theta})}{g(y;\boldsymbol{\theta})}\right) \log\left(\frac{f_{(d_r)}(x;\boldsymbol{\theta})}{g(y;\boldsymbol{\theta})}\right) dy \\
\geq & \; n \int g(y;\boldsymbol{\theta}) \left[\frac{1}{n}\sum_{r=1}^{n} \frac{f_{(d_r)}(y;\boldsymbol{\theta})}{g(y;\boldsymbol{\theta})}\right] \log\left[\frac{\frac{1}{n}\sum_{r=1}^{n} f_{(d_r)}(y;\boldsymbol{\theta})}{g(y;\boldsymbol{\theta})}\right] dy \\
= & \; n \int f(y;\boldsymbol{\theta}) \log\left(\frac{f(y;\boldsymbol{\theta})}{g(y;\boldsymbol{\theta})}\right) dy \\
= & \; K\left(L_{srs,1}(\boldsymbol{\theta}), L_{srs,2}(\boldsymbol{\theta})\right),
\end{aligned}$$

which shows the first inequality. Similarly,

$$\begin{aligned}
& K\left(L_{pros}(\boldsymbol{\theta}|\mathbf{y}), L_{srs,2}(\boldsymbol{\theta}|\mathbf{y})\right) \\
= & \sum_{r=1}^{n} \int g(y;\boldsymbol{\theta}) \left(\frac{1}{m}\sum_{u\in d_r} \frac{f^{(u:S)}(y;\boldsymbol{\theta})}{g(y;\boldsymbol{\theta})}\right) \log\left(\frac{1}{m}\sum_{u\in d_r} \frac{f^{(u:S)}(y;\boldsymbol{\theta})}{g(y;\boldsymbol{\theta})}\right) dy \\
\leq & \; \frac{1}{m}\sum_{v=1}^{S} \int f^{(v:S)}(y;\boldsymbol{\theta}) \log\left(\frac{f^{(v:S)}(x;\boldsymbol{\theta})}{g(y;\boldsymbol{\theta})}\right) dy \\
= & \; \frac{1}{m} K\left(L_{rss^*}(\boldsymbol{\theta}|\mathbf{y}), L_{srs,2}(\boldsymbol{\theta}|\mathbf{y})\right),
\end{aligned}$$

which completes the proof. □

Chapter 5

Mixture Model Analysis of PROS Samples

In Chapter 2, we developed parametric inference for the FMMs based on an RSS design under the perfect ranking assumption and showed that it improved the efficiency for the estimation of the mixing parameters and parameters of the component distributions compared with the usual case under the SRS design. In this chapter, we investigate the FMMs under PROS design with an imperfect ranking error model. This extends the results of Chapter 2 to an imperfect setting. In order to minimize the magnitude of the ranking error, we propose to use a PROS sampling design in which subsets are partially ordered in each set and units within subsets are not ordered. The partial ordering does not eliminate the possibility of ranking error completely, but reduces it to the misplacement errors of the units into subsets. Implementing the ML method to estimate the parameters of the model is much more challenging when working with PROS samples compared to the traditional SRS and even RSS samples. One significant difference is that the full likelihood based on PROS data involves powers of convex combinations of the survival and cumulative distribution functions of the underlying FMM. We propose a new EM algorithm for the ML estimation of the unknown parameters of the FMM as well as the ranking error probabilities under the PROS sampling design.

To this end, in Section 5.1, we introduce the PROS sampling design for FMMs. In Section 5.2, we present the likelihood function of the PROS sample from a FMM and introduce suitable indicator variables to obtain the complete-data likelihood function. In Section 5.3, we introduce an EM algorithm for the estimation of the parameters of the model given the PROS sample. One major challenge to ML estimation, when it involves PROS data from FMMs, is computational intractability. In Section 5.4, a modified version of the proposed EM algorithm is suggested to reduce the computational burden of the estimation process. In Section 5.5, we study the problem of classification of the PROS sample into the components of the FMM. We describe our simulation studies in Section 5.6. An application of our method to a fishery data example is presented in Section 5.7. We apply the proposed method to estimate the age groups of a fish species in the Chesapeake Bay area using length-frequency data which can be considered to be a mixture of two normal densities.

5.1 PROS sample from FMM

Suppose X is a random variable associated with a random phenomenon of interest following the FMM (1.1). Let $D = \{d_1, \ldots, d_n\}$ denote a partition of the integers $\{1, \ldots, S\}$ into n mutually exclusive subsets d_r, each of size m, where $d_r = \{(r-1)m+1, \ldots, rm\}$, $r = 1, \ldots, n$, and $m = S/n$. Let $\{X_{[d_r]i}; r = 1, \ldots, n; i = 1, \ldots, N\}$ denote a balanced PROS sample of size Nn constructed through the PROS sampling design described in Chapter 1, with set size S, number of subsets n of equal size $m = S/n$ and number of cycles N.

To provide a model for misplacement errors in PROS design, we consider a missing data model proposed in Arslan and Ozturk (2013). Let

$$\mathbf{X}_i = \left\{X_{(1)i} < X_{(2)i} < \cdots < X_{(S)i}\right\}$$

be the order statistics in set i, $i = 1, \ldots, N$. The PROS design selects one unit, $X_{[d_r]i}$, from the subset d_r in this set. The construction of d_r and the selection of $X_{[d_r]i}$ is modelled by a missing data model which accommodates the possibility of ranking error between judgment subsets d_r, $r = 1, \ldots, n$. Let $\boldsymbol{\alpha}$ denote the misplacement probability matrix,

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_{d_1,d_1} & \boldsymbol{\alpha}_{d_1,d_2} & \ldots & \boldsymbol{\alpha}_{d_1,d_n} \\ \boldsymbol{\alpha}_{d_2,d_1} & \boldsymbol{\alpha}_{d_2,d_2} & \ldots & \boldsymbol{\alpha}_{d_2,d_n} \\ \vdots & \vdots & \ldots & \vdots \\ \boldsymbol{\alpha}_{d_n,d_1} & \boldsymbol{\alpha}_{d_n,d_2} & \ldots & \boldsymbol{\alpha}_{d_n,d_n} \end{bmatrix}_{S\times S} ; \; \boldsymbol{\alpha}_{d_r,d_h} = \alpha_{d_r,d_h} \begin{bmatrix} \frac{1}{m^2} & \frac{1}{m^2} & \ldots & \frac{1}{m^2} \\ \frac{1}{m^2} & \frac{1}{m^2} & \ldots & \frac{1}{m^2} \\ \vdots & \vdots & \ldots & \vdots \\ \frac{1}{m^2} & \frac{1}{m^2} & \ldots & \frac{1}{m^2} \end{bmatrix}_{m\times m},$$

where α_{d_r,d_h} is the probability that any order statistic from the subset d_h is misplaced into the subset d_r. Since the design D is a partition, we must have the constraint $\sum_{h=1}^{n} \alpha_{d_r,d_h} = 1$. In order to have a valid probability model, we also need to have the constraint $\sum_{r=1}^{n} \alpha_{d_r,d_h} = 1$. These two constraints define a doubly stochastic matrix $\boldsymbol{\alpha}$. Thus, the matrix $\boldsymbol{\alpha}$ contains the misplacement probabilities of an order statistic from the subset d_h to a judgment order statistic in another subset d_r, for $h, r \in \{1, \ldots, n\}$. Entries $1/m^2$ indicate that random selection from d_h and random replacement to d_r are all equally likely.

Now, for each $X_{[d_r]i}$, define an $m \times S$ dimensional multinomial random matrix,

$\mathbf{\Delta}_i^{[d_r]}$, with parameters 1 and $\boldsymbol{\alpha}_{[d_r]}$, where $\boldsymbol{\alpha}_{[d_r]} = (\boldsymbol{\alpha}_{d_r,d_1}, \ldots, \boldsymbol{\alpha}_{d_r,d_n})$, and $\boldsymbol{\alpha}_{d_r,d_h}$ are as defined above. It is clear that $\boldsymbol{\alpha}_{[d_r]}$ is an $m \times S$ dimensional probability matrix whose entries sum to 1. With this structure, $\mathbf{\Delta}_i^{[d_r]}$ can be written as an $m \times S$ random matrix $\mathbf{\Delta}_i^{[d_r]} = \left(\Delta_i^{[d_r,d_h]}(k,u) : k \in d_r, u \in d_h\right)$ which has only one nonzero (exactly equal to 1) entry and all the other entries are zero. The position of the nonzero entry determines the identity of the unit and the subset from which the unit is selected at random; that is,

$$X_{[d_r]i} = \mathbf{1}_m^\top \mathbf{\Delta}_i^{[d_r]} \mathbf{X}_{(i)},$$

where $\mathbf{1}_m$ is a column vector of ones. Since $\mathbf{\Delta}_i^{[d_r]}$ is a multinomial random matrix, its pdf is given by

$$\mathbb{P}(\mathbf{\Delta}_i^{[d_r]} = \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\alpha}) = \prod_{k \in d_r} \prod_{h=1}^{n} \prod_{u \in d_h} \left(\frac{\alpha_{d_r,d_h}}{m^2}\right)^{\delta_i^{[d_r,d_h]}(k,u)}. \tag{5.1}$$

Also, the conditional distribution of $X_{[d_r]i}$ given $\mathbf{\Delta}_i^{[d_r]}$ can be written as

$$f(x_{[d_r]i} | \boldsymbol{\delta}_i^{[d_r]}; \mathbf{\Psi}) = \prod_{k \in d_r} \prod_{h=1}^{n} \prod_{u \in d_h} \left\{f^{(u:S)}(x_{[d_r]i}; \mathbf{\Psi})\right\}^{\delta_i^{[d_r,d_h]}(k,u)}, \tag{5.2}$$

where $f^{(u:S)}(x; \mathbf{\Psi})$ is the pdf of the u-th order statistic in a set of size S from the FMM (1.1) as follows

$$f^{(u:S)}(x; \mathbf{\Psi}) = S \binom{S-1}{u-1} f(x; \mathbf{\Psi}) \{F(x; \mathbf{\Psi})\}^{u-1} \{\bar{F}(x; \mathbf{\Psi})\}^{S-u}, \tag{5.3}$$

and $\bar{F}(x; \mathbf{\Psi}) = 1 - F(x; \mathbf{\Psi})$. Finally, the joint distribution of $(X_{[d_r]i}, \mathbf{\Delta}_i^{[d_r]})$ follows from (5.1) and (5.2) and is given by

$$f(x_{[d_r]i}, \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\Omega}) = \prod_{k\in d_r}\prod_{h=1}^{n}\prod_{u\in d_h}\left\{\frac{\alpha_{d_r,d_h}}{m^2} f^{(u:S)}(x_{[d_r]i}; \boldsymbol{\Psi})\right\}^{\delta_i^{[d_r,d_h]}(k,u)}, \tag{5.4}$$

where $\boldsymbol{\Omega} = (\boldsymbol{\Psi}, \boldsymbol{\alpha})$. The marginal distribution of $X_{[d_r]i}$ is then obtained by summing $f(x_{[d_r]i}, \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\Omega})$ over $\boldsymbol{\delta}_i^{[d_r]}$, that is

$$f(x_{[d_r]i}; \boldsymbol{\Omega}) = \sum_{\boldsymbol{\delta}_i^{[d_r]}} f(x_{[d_r]i}, \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\Omega}) = \frac{1}{m}\sum_{h=1}^{n}\sum_{u\in d_r}\alpha_{d_r,d_h} f^{(u:S)}(x_{[d_r]i}; \boldsymbol{\Psi}). \tag{5.5}$$

Form (5.4) and (5.5), it is easy to see that

$$f(\boldsymbol{\delta}_i^{[d_r]}|x_{[d_r]i}; \boldsymbol{\Omega}) = \prod_{k\in d_r}\prod_{h=1}^{n}\prod_{u\in d_h}\left\{\frac{\alpha_{d_r,d_h} B_{u,S+1-u}(F(x_{[d_r]}; \boldsymbol{\Psi}))}{m\sum_{h=1}^{n}\sum_{u'\in d_h}\alpha_{d_r,d_h} B_{u',S+1-u'}(F(x_{[d_r]}; \boldsymbol{\Psi}))}\right\}^{\delta_i^{[d_r,d_h]}(k,u)}. \tag{5.6}$$

where $B_{a,b}(\cdot)$ is a beta density function with parameters a and b.

5.2 Likelihood functions for PROS sample

In this section, we derive the ML estimator of the unknown parameter $\boldsymbol{\Psi}$ of the FMM (1.1) based on a PROS sample. Let $F(x; \boldsymbol{\Psi}) = \sum_{j=1}^{M}\pi_j F_j(x; \theta_j)$ be the cumulative distribution function (cdf) of the FMM (1.1), where $F_j(x; \theta_j)$, $j = 1, \ldots, M$, refers to the cdf of the j-th component of the model. Using (5.5), the incomplete likelihood function under the imperfect PROS sampling design is given by

$$L_1(\boldsymbol{\Omega}) = \prod_{i=1}^{N}\prod_{r=1}^{n} f(x_{[d_r]i}; \boldsymbol{\Omega}). \tag{5.7}$$

Let $l_1(\boldsymbol{\Omega}) = \log L_1(\boldsymbol{\Omega})$. Now, the ML estimator of $\boldsymbol{\Psi}$, denoted by $\widehat{\boldsymbol{\Psi}}$, is given as an appropriate root of the likelihood equation, $\frac{\partial l_1(\boldsymbol{\Omega})}{\partial \boldsymbol{\Psi}} = 0$. Note that finding $\widehat{\boldsymbol{\Psi}}$ from (5.7) is not tractable since the pdf involves summations over h and u in the subset d_r. To overcome this problem we rewrite the likelihood function $L_1(\boldsymbol{\Omega})$ from a missing data model perspective. Considering the unmeasured (latent) multinomial random matrices $\boldsymbol{\Delta}_i^{[d_r]}$ defined in (5.4), the complete likelihood function of $\boldsymbol{\Omega}$ becomes

$$L_2(\boldsymbol{\Omega}) = \prod_{i=1}^{N} \prod_{r=1}^{n} f(x_{[d_r]i}, \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\Omega}). \tag{5.8}$$

Using the marginalization principle in (5.5), it is easy to see that the likelihood function (5.7) can be obtained by summing the complete-data likelihood function (5.8) over $\boldsymbol{\Delta}_i^{[d_r]}$.

Note that the complete-data likelihood function L_2 is more tractable than L_1 when interest lies in the estimation of the parameter $\boldsymbol{\alpha}$. Using the EM algorithm, the ML estimator of $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\alpha}}_{ML}$, can easily be obtained. However, from equations (5.3) and (5.8), we note that, due to the presence of the terms $[F(\cdot; \boldsymbol{\Psi})]^{u-1}$ and $[\bar{F}(\cdot; \boldsymbol{\Psi})]^{S-u}$ in L_2, the complete likelihood function L_2 is still intractable for estimating $\boldsymbol{\Psi}$. To reduce the computational complexity in the estimation of the parameter $\boldsymbol{\Psi}$, we provide an alternative representation for the likelihood function L_2. We first insert (5.3) into equation (5.2) and rewrite the conditional density of $X_{[d_r]i}$ given the latent vector $\boldsymbol{\Delta}_i^{[d_r]} = \boldsymbol{\delta}_i^{[d_r]}$ and $\boldsymbol{\Psi}$ as follows

$$f(x_{[d_r]i} | \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\Psi}) \tag{5.9}$$

$$= \prod_{k \in d_r} \prod_{h=1}^{n} \prod_{u \in d_h} \left\{ S \binom{S-1}{u-1} f(x_{[d_r]}; \boldsymbol{\Psi}) [F(x_{[d_r]}; \boldsymbol{\Psi})]^{u-1} [\bar{F}(x_{[d_r]}; \boldsymbol{\Psi})]^{S-u} \right\}^{\delta_i^{[d_r, d_h]}(k,u)}.$$

In equation (5.9), to keep track of the component membership of the judgment order statistics $X_{[d_r]i}$ and to simplify the density to a manageable form, we introduce three additional latent variables $\mathbf{Z}_i^{[d_r]}, \mathbf{W}_i^{[d_r]}, \mathbf{V}_i^{[d_r]}$ for each $x_{[d_r]i}$, given $\boldsymbol{\Delta}_i^{[d_r]} = \boldsymbol{\delta}_i^{[d_r]}$. Let $\mathbf{Z}_i^{[d_r]}|\boldsymbol{\delta}_i^{[d_r]}$ denote the conditional component membership of the observation $x_{[d_r]i}$ and $\mathbf{W}_i^{[d_r]}|\boldsymbol{\delta}_i^{[d_r]}$ denote the conditional component membership of the observations less than $x_{[d_r]i}$ in a set of size S from which $x_{[d_r]i}$ is obtained. Similarly, let $\mathbf{V}_i^{[d_r]}|\boldsymbol{\delta}_i^{[d_r]}$ denote the conditional component membership of the observations larger than $x_{[d_r]i}$ in its corresponding set of size S. We note that these second-level latent variables are handled conditionally, given the first-level latent matrix $\boldsymbol{\Delta}_i^{[d_r]}$, since the rank and the exact position of $x_{[d_r]i}$ (i.e., the subset and the position within the subset) are determined by the latent matrix $\boldsymbol{\Delta}_i^{[d_r]}$.

For a given judgment order statistic $x_{[d_r]i}$, let $\{\Delta_i^{[d_r,d_h]}(k,u) = 1\}$ denote the event that the entry of the matrix $\boldsymbol{\Delta}_i^{[d_r]}$ at the k-th row of the subset d_r and u-th column of the subset d_h is one. The latent vector $\mathbf{Z}_i^{[d_r]}|\{\Delta_i^{[d_r,d_h]}(k,u) = 1\}$ is then an M-dimensional vector, where $Z_{ij}^{[d_r]}|\{\Delta_i^{[d_r,d_h]}(k,u) = 1\}$ is one or zero, according to whether or not $x_{[d_r]i}$ belongs to the j-th component of the mixture model ($j = 1, \ldots, M$), that is,

$$Z_{ij}^{[d_r]}\big|\{\Delta_i^{[d_r,d_h]}(k,u) = 1\} = \begin{cases} 1 & \text{if } x_{[d_r]i} \text{ belongs to component } j; \\ 0 & \text{otherwise,} \end{cases}$$

with $\sum_{j=1}^{M} \left(Z_{ij}^{[d_r]}\big|\{\Delta_i^{[d_r,d_h]}(k,u) = 1\} \right) = 1$. For $i = 1, \ldots, N; r = 1, \ldots, n$, we have that each $\mathbf{Z}_i^{[d_r]}|\{\Delta_i^{[d_r,d_h]}(k,u) = 1\}$ follows a multinomial distribution consisting of one draw on M classes $(C_1, \ldots, C_M)$ with probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_M)$. Hence, it is easy to show that

$$f(\mathbf{z}_i^{[d_r]} | \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\pi}) = \prod_{k \in d_r} \prod_{h=1}^{n} \prod_{u \in d_h} \left\{ \binom{1}{z_{i1}^{[d_r]}, \ldots, z_{i1}^{[d_r]}} \prod_{j=1}^{M} \pi_j^{z_{ij}^{[d_r]}} \right\}^{\delta_i^{[d_r,d_h]}(k,u)} . \quad (5.10)$$

In addition, suppose M-dimensional latent vector $\mathbf{W}_i^{[d_r]} | \{\Delta_i^{[d_r,d_h]}(k,u) = 1\}$ denote the number of observations less than $x_{[d_r]i}$ which are selected from component j of the FMM (1.1). It should be noted that $\sum_{j=1}^{M} \left(W_{ij}^{[d_r]} | \{\Delta_i^{[d_r,d_h]}(k,u) = 1\} \right) = u - 1$, and the latent vectors $\mathbf{W}_i^{[d_r]} | \{\Delta_i^{[d_r,d_h]}(k,u) = 1\}, i = 1, \ldots, N; r = 1, \ldots, n$, each have a multinomial distribution consisting of $u-1$ draws on M classes $(C_1, \ldots, C_M)$ with probabilities $\boldsymbol{\pi}$. Therefore, we can show that

$$f(\mathbf{w}_i^{[d_r]} | \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\pi}) = \prod_{k \in d_r} \prod_{h=1}^{n} \prod_{u \in d_h} \left\{ \binom{u-1}{w_{i1}^{[d_r]}, \ldots, w_{i1}^{[d_r]}} \prod_{j=1}^{M} \pi_j^{w_{ij}^{[d_r]}} \right\}^{\delta_i^{[d_r,d_h]}(k,u)} . \quad (5.11)$$

Similarly, let $\mathbf{V}_i^{[d_r]} | \{\Delta_i^{[d_r,d_h]}(k,u) = 1\}$ be an M-dimensional vector, such that variable $V_{ij}^{[d_r]} | \{\Delta_i^{[d_r,d_h]}(k,u) = 1\}$ denotes the number of observations bigger than $x_{[d_r]i}$ that are selected from component j of the FMM (1.1), with the fact that $\sum_{j=1}^{M} \left(V_{ij}^{[d_r]} | \{\Delta_i^{[d_r,d_h]}(k,u) = 1\} \right) = S - u$. Accordingly, $\mathbf{V}_i^{[d_r]} | \{\Delta_i^{[d_r,d_h]}(k,u) = 1\}, i = 1, \ldots, N; r = 1, \ldots, n$, follows a multinomial distribution consisting of $S - u$ draws on M classes $(C_1, \ldots, C_M)$ with probabilities $\boldsymbol{\pi}$, and

$$f(\mathbf{v}_i^{[d_r]} | \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\pi}) = \prod_{k \in d_r} \prod_{h=1}^{n} \prod_{u \in d_h} \left\{ \binom{S-u}{v_{i1}^{[d_r]}, \ldots, v_{i1}^{[d_r]}} \prod_{j=1}^{M} \pi_j^{v_{ij}^{[d_r]}} \right\}^{\delta_i^{[d_r,d_h]}(k,u)} . \quad (5.12)$$

Since each set in the PROS design consists of independent samples from the population, and since the component memberships of those observations are independent of each other, the latent variables $\mathbf{Z}_i^{[d_r]}$, $\mathbf{W}_i^{[d_r]}$ and $\mathbf{V}_i^{[d_r]}$ are conditionally independent, given $\boldsymbol{\Delta}_i^{[d_r]}$. The joint distribution of the judgment order statistic $X_{[d_r]i}$ and the latent variables $\mathbf{Z}_i^{[d_r]}$, $\mathbf{W}_i^{[d_r]}$ and $\mathbf{V}_i^{[d_r]}$ is given in the following lemma.

Lemma 5.1. *For fixed values i and r, $(i = 1, \ldots, N,\ r = 1, \ldots, n)$, we have*

$$
\begin{aligned}
& f(x_{[d_r]i}, \boldsymbol{\delta}_i^{[d_r]}, \mathbf{z}_i^{[d_r]}, \mathbf{w}_i^{[d_r]}, \mathbf{v}_i^{[d_r]}; \boldsymbol{\Omega}) \\
& \propto \prod_{k\in d_r} \prod_{h=1}^{n} \prod_{u\in d_h} \prod_{j=1}^{M} \left\{ \frac{\alpha_{d_r,d_h}}{m^2}\, \pi_j^{\{z_{ij}^{[d_r]} + w_{ij}^{[d_r]} + v_{ij}^{[d_r]}\}} \right\}^{\delta_i^{[d_r,d_h]}(k,u)} \\
& \quad \times \left\{ [f_j(x_{[d_r]i}, \theta_j)]^{z_{ij}^{[d_r]}} [F_j(x_{[d_r]i}, \theta_j)]^{w_{ij}^{[d_r]}} [\bar{F}_j(x_{[d_r]i}, \theta_j)]^{v_{ij}^{[d_r]}} \right\}^{\delta_i^{[d_r,d_h]}(k,u)}.
\end{aligned}
$$

Proof. Let $c_1 = S\binom{S-1}{u-1}$, $c_2 = \binom{1}{z_{i1}^{[d_r]},\ldots,z_{iM}^{[d_r]}}$, $c_3 = \binom{u-1}{w_{i1}^{[d_r]},\ldots,w_{iM}^{[d_r]}}$ and $c_4 = \binom{S-u}{v_{i1}^{[d_r]},\ldots,v_{iM}^{[d_r]}}$.

$$
\begin{aligned}
& f(x_{[d_r]i} | \{\mathbf{z}_i^{[d_r]}, \boldsymbol{\delta}_i^{[d_r]}\}; \boldsymbol{\Psi}) \\
& = \prod_{k\in d_r} \prod_{h=1}^{n} \prod_{u\in d_h} \left\{ c_1 \prod_{j=1}^{M} [f_j(x_{[d_r]i}; \theta_j)]^{z_{ij}^{[d_r]}} [F(x_{[d_r]i}; \boldsymbol{\Psi})]^{u-1} [\bar{F}(x_{[d_r]i}; \boldsymbol{\Psi})]^{S-u} \right\}^{\delta_i^{[d_r,d_h]}(k,u)}.
\end{aligned}
$$

Using (5.1) and (5.10), we write

$$
\begin{aligned}
f(x_{[d_r]i}, \mathbf{z}_i^{[d_r]}, \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\Omega}) = \prod_{k\in d_r} \prod_{h=1}^{n} \prod_{u\in d_h} & \left\{ \frac{\alpha_{d_r,d_h}}{m^2}\, c_1\, c_2 \prod_{j=1}^{M} [\pi_j f_j(x_{[d_r]i}; \theta_j)]^{z_{ij}^{[d_r]}} \right. \\
& \left. \times [F(x_{[d_r]i}; \boldsymbol{\Psi})]^{u-1} [\bar{F}(x_{[d_r]i}; \boldsymbol{\Psi})]^{S-u} \right\}^{\delta_i^{[d_r,d_h]}(k,u)}. \qquad (5.13)
\end{aligned}
$$

Furthermore, using (5.4), one gets

$$f(\mathbf{z}_i^{[d_r]}|\{x_{[d_r]i},\boldsymbol{\delta}_i^{[d_r]}\};\boldsymbol{\Psi})=\prod_{k\in d_r}\prod_{h=1}^{n}\prod_{u\in d_h}\left\{c_2\prod_{j=1}^{M}\left(\frac{\pi_j f_j(x_{[d_r]i};\theta_j)}{f(x_{[d_r]i};\boldsymbol{\Psi})}\right)^{z_{ij}^{[d_r]}}\right\}^{\delta_i^{[d_r,d_h]}(k,u)}. \tag{5.14}$$

On the other hand, the conditional distribution of $X_{[d_r]i}$ given $\mathbf{W}_i^{[d_r]}=\mathbf{w}_i^{[d_r]}$ and $\boldsymbol{\Delta}_i^{[d_r]}=\boldsymbol{\delta}_i^{[d_r]}$ can be written as

$$\begin{aligned}&f(x_{[d_r]i}|\{\mathbf{w}_i^{[d_r]},\boldsymbol{\delta}_i^{[d_r]}\};\boldsymbol{\Psi})\\&=\prod_{k\in d_r}\prod_{h=1}^{n}\prod_{u\in d_h}\left\{c_1\prod_{j=1}^{M}f(x_{[d_r]i};\boldsymbol{\Psi})[F_j(x_{[d_r]i};\theta_j)]^{w_{ij}^{[d_r]}}[\bar{F}(x_{[d_r]i};\boldsymbol{\Psi})]^{S-u}\right\}^{\delta_i^{[d_r,d_h]}(k,u)}.\end{aligned}$$

From (5.1) and (5.11), it is easy to write

$$\begin{aligned}f(x_{[d_r]i},\mathbf{w}_i^{[d_r]},\boldsymbol{\delta}_i^{[d_r]};\boldsymbol{\Omega})=\prod_{k\in d_r}\prod_{h=1}^{n}\prod_{u\in d_h}&\left\{\frac{\alpha_{d_r,d_h}}{m^2}\,c_1\,c_3\prod_{j=1}^{M}f(x_{[d_r]i};\boldsymbol{\Psi})\right.\\&\left.\times[\pi_jF_j(x_{[d_r]i};\theta_j)]^{w_{ij}^{[d_r]}}[\bar{F}(x_{[d_r]i};\boldsymbol{\Psi})]^{S-u}\right\}^{\delta_i^{[d_r,d_h]}(k,u)}.\end{aligned}$$

Now, using equation (5.4) shows

$$\begin{aligned}&f(\mathbf{w}_i^{[d_r]}|\{x_{[d_r]i},\boldsymbol{\delta}_i^{[d_r]}\};\boldsymbol{\Psi})\\&=\prod_{k\in d_r}\prod_{h=1}^{n}\prod_{u\in d_h}\left\{c_3\prod_{j=1}^{M}\left(\frac{\pi_jF_j(x_{[d_r]i};\theta_j)}{F(x_{[d_r]i};\boldsymbol{\Psi})}\right)^{w_{ij}^{[d_r]}}\right\}^{\delta_i^{[d_r,d_h]}(k,u)}.\end{aligned} \tag{5.15}$$

Similarly, we have

$$\begin{aligned}&f(x_{[d_r]i}|\{\mathbf{v}_i^{[d_r]},\boldsymbol{\delta}_i^{[d_r]}\};\boldsymbol{\Psi})\\&=\prod_{k\in d_r}\prod_{h=1}^{n}\prod_{u\in d_h}\left\{c_1\prod_{j=1}^{M}f(x_{[d_r]i};\boldsymbol{\Psi})[F(x_{[d_r]i};\boldsymbol{\Psi})]^{u-1}[\bar{F}_j(x_{[d_r]i};\theta_j)]^{v_{ij}^{[d_r]}}\right\}^{\delta_i^{[d_r,d_h]}(k,u)}.\end{aligned}$$

Again, making use of equations (5.1) and (5.12) yields the following joint distribution

$$f(x_{[d_r]i}, \mathbf{v}_i^{[d_r]}, \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\Omega}) = \prod_{k\in d_r}\prod_{h=1}^{n}\prod_{u\in d_h}\left\{\frac{\alpha_{d_r,d_h}}{m^2}\, c_1\, c_4 \prod_{j=1}^{M} f(x_{[d_r]i}; \boldsymbol{\Psi})\right.$$

$$\left.\times\, [F(x_{[d_r]i}; \boldsymbol{\Psi})]^{u-1}[\pi_j \bar{F}_j(x_{[d_r]i}; \theta_j)]^{v_{ij}^{[d_r]}}\right\}^{\delta_i^{[d_r,d_h]}(k,u)}.$$

Once again, using (5.4), we write

$$f(\mathbf{v}_i^{[d_r]}|\{x_{[d_r]i}, \boldsymbol{\delta}_i^{[d_r]}\}; \boldsymbol{\Psi}) = \prod_{k\in d_r}\prod_{h=1}^{n}\prod_{u\in d_h}\left\{c_4\prod_{j=1}^{M}\left(\frac{\pi_j \bar{F}_j(x_{[d_r]i}; \theta_j)}{\bar{F}(x_{[d_r]i}; \boldsymbol{\Psi})}\right)^{v_{ij}^{[d_r]}}\right\}^{\delta_i^{[d_r,d_h]}(k,u)}. \tag{5.16}$$

Finally, based on the conditional independence of the latent variables and from (5.14), (5.15), (5.16) and (5.4), the lemma is easily proved. □

The key difference between the joint pdf in Lemma 5.1 and the pdf in equation (5.7) is that the summation in equation (5.7) is replaced by products with the help of the latent variables. The following lemma shows that the conditional distribution of $X_{[d_r]i}$ given $\boldsymbol{\delta}_i^{[d_r]}$ can be obtained from Lemma 5.1. Therefore, we can use the EM algorithm to obtain the ML estimate of $\boldsymbol{\Omega}$.

Lemma 5.2. *For each $x_{[d_r]i}$, $i = 1, \ldots, N; r = 1, \ldots, n$, we have*

$$f(x_{[d_r]i}, \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\Omega}) = \sum_{\mathbf{Z}|\boldsymbol{\delta}}\sum_{\mathbf{W}|\boldsymbol{\delta}}\sum_{\mathbf{V}|\boldsymbol{\delta}} f(x_{[d_r]i}, \boldsymbol{\delta}_i^{[d_r]}, \mathbf{z}_i^{[d_r]}, \mathbf{w}_i^{[d_r]}, \mathbf{v}_i^{[d_r]}; \boldsymbol{\Omega}).$$

where $\mathbf{Z}|\boldsymbol{\delta}$, $\mathbf{W}|\boldsymbol{\delta}$ and $\mathbf{V}|\boldsymbol{\delta}$ in the summations indicate all the possible values of $\mathbf{Z}_i^{[d_r]}$, $\mathbf{W}_i^{[d_r]}$ and $\mathbf{V}_i^{[d_r]}$ given $\boldsymbol{\Delta}_i^{[d_r]} = \boldsymbol{\delta}_i^{[d_r]}$, respectively.

Proof. To show the result, we have

$$
\begin{aligned}
&\sum_{\mathbf{Z}|\boldsymbol{\delta}} \sum_{\mathbf{W}|\boldsymbol{\delta}} \sum_{\mathbf{V}|\boldsymbol{\delta}} f(x_{[d_r]i}, \boldsymbol{\delta}_i^{[d_r]}, \mathbf{z}_i^{[d_r]}, \mathbf{w}_i^{[d_r]}, \mathbf{v}_i^{[d_r]}; \boldsymbol{\Omega}) \\
=& \prod_{k\in d_r} \prod_{h=1}^{n} \prod_{u\in d_h} \Bigg\{ c_1 \frac{\alpha_{d_r,d_h}}{m^2} \left[\sum_{\mathbf{Z}|\boldsymbol{\delta}} c_2 \prod_{j=1}^{M} \{\pi_j f_j(x_{[d_r]i}; \theta_j)\}^{z_{ij}^{[d_r]}} \right] \\
&\times \left[\sum_{\mathbf{W}|\boldsymbol{\delta}} c_3 \prod_{j=1}^{M} \{\pi_j F_j(x_{[d_r]i}; \theta_j)\}^{w_{ij}^{[d_r]}} \right] \left[\sum_{\mathbf{V}|\boldsymbol{\delta}} c_4 \prod_{j=1}^{M} \{\pi_j \bar{F}_j(x_{[d_r]i}; \theta_j)\}^{v_{ij}^{[d_r]}} \right] \Bigg\}^{\delta_i^{[d_r,d_h]}(k,u)} \\
=& \prod_{k\in d_r} \prod_{h=1}^{n} \prod_{u\in d_h} \Bigg\{ c_1 \frac{\alpha_{d_r,d_h}}{m^2} \left[\sum_{j=1}^{M} \pi_j f_j(x_{[d_r]i}; \theta_j) \right] \left[\sum_{j=1}^{M} \pi_j F_j(x_{[d_r]i}; \theta_j) \right]^{u-1} \\
&\times \left[\sum_{j=1}^{M} \pi_j \bar{F}_j(x_{[d_r]i}; \theta_j) \right]^{S-u} \Bigg\}^{\delta_i^{[d_r,d_h]}(k,u)} \\
=& \prod_{k\in d_r} \prod_{h=1}^{n} \prod_{u\in d_h} \left\{ c_1 \frac{\alpha_{d_r,d_h}}{m^2} f(x_{[d_r]i}; \boldsymbol{\Psi}) [F(x_{[d_r]i}; \boldsymbol{\Psi})]^{u-1} [\bar{F}(x_{[d_r]i}; \boldsymbol{\Psi})]^{S-u} \right\}^{\delta_i^{[d_r,d_h]}(k,u)} \\
=& f(x_{[d_r]i}, \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\Omega}).
\end{aligned}
$$

Since the random matrix $\boldsymbol{\Delta}_i^{[d_r]}$ follows a multinomial distribution with one draw from mS cells (only one entity is 1 and other entries are 0), we could switch the order of the triple summations with the triple products in the first equality. $\square$

Let $\mathbf{Y}$ denote the collection of all latent variables and $\mathbf{X}$ denote the observed PROS data, that is, $\mathbf{Y} = \{(\boldsymbol{\Delta}_i^{[d_r]}, \mathbf{Z}_i^{[d_r]}, \mathbf{W}_i^{[d_r]}, \mathbf{V}_i^{[d_r]}), i = 1, \dots, N; r = 1, \dots, n\}$ and $\mathbf{X} = \{X_{[d_r]i}, i = 1, \dots, N; r = 1, \dots, n\}$, respectively. Using Lemma 5.1, the likelihood function can be written as follows

$$L(\boldsymbol{\Omega}|\mathbf{y},\mathbf{x}) = \prod_{i=1}^{N}\prod_{r=1}^{n} f(x_{[d_r]i}, \boldsymbol{\delta}_i^{[d_r]}, \mathbf{z}_i^{[d_r]}, \mathbf{w}_i^{[d_r]}, \mathbf{v}_i^{[d_r]}; \boldsymbol{\Omega}).$$

It is easy to see that the likelihood function $L_2(\boldsymbol{\Psi}, \alpha)$ can be obtained from $L(\boldsymbol{\Omega}|\mathbf{y},\mathbf{x})$ by summing over $\mathbf{Z}|\boldsymbol{\Delta}, \mathbf{W}|\boldsymbol{\delta}$ and $\mathbf{V}|\boldsymbol{\delta}$, i.e., $\sum_{\mathbf{Z}|\boldsymbol{\delta}}\sum_{\mathbf{W}|\boldsymbol{\delta}}\sum_{\mathbf{V}|\boldsymbol{\delta}} L(\boldsymbol{\Omega}|\mathbf{y},\mathbf{x}) = L_2(\boldsymbol{\Omega})$. The log-likelihood function of $\boldsymbol{\Omega}$ based on the full data is now given by

$$\begin{aligned}
&l(\boldsymbol{\Omega}|\mathbf{y},\mathbf{x}) \\
&\propto \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{k\in d_r}\sum_{h=1}^{n}\sum_{u\in d_h} \delta_i^{[d_r,d_h]}(k,u)\log\alpha_{d_r,d_h} \\
&+ \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{k\in d_r}\sum_{h=1}^{n}\sum_{u\in d_h} \delta_i^{[d_r,d_h]}(k,u)\sum_{j=1}^{M}\left[\left\{z_{ij}^{[d_r]} + w_{ij}^{[d_r]} + v_{ij}^{[d_r]}\right\}\log\pi_j\right] \\
&+ \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{k\in d_r}\sum_{h=1}^{n}\sum_{u\in d_h} \delta_i^{[d_r,d_h]}(k,u) \\
&\quad\times\sum_{j=1}^{M}\left[z_{ij}^{[d_r]}\log f_j(x_{[d_r]i};\theta_j) + w_{ij}^{[d_r]}\log F_j(x_{[d_r]i};\theta_j) + v_{ij}^{[d_r]}\log\bar{F}_j(x_{[d_r]i};\theta_j)\right] \\
&= l_1(\boldsymbol{\alpha}|\mathbf{y},\mathbf{x}) + l_2(\boldsymbol{\pi}|\mathbf{y},\mathbf{x}) + l_3(\boldsymbol{\xi}|\mathbf{y},\mathbf{x}).
\end{aligned} \tag{5.17}$$

Note that the log-likelihood function is partitioned into three parts, $l_1(\boldsymbol{\alpha}|\mathbf{y},\mathbf{x})$, $l_2(\boldsymbol{\pi}|\mathbf{y},\mathbf{x})$, and $l_3(\boldsymbol{\xi}|\mathbf{y},\mathbf{x})$, where $l_1(\boldsymbol{\alpha}|\mathbf{y},\mathbf{x})$ depends only on $\boldsymbol{\alpha}$, $l_2(\boldsymbol{\pi}|\mathbf{y},\mathbf{x})$ depends only on $\boldsymbol{\pi}$ and $l_3(\boldsymbol{\xi}|\mathbf{y},\mathbf{x})$ depends only on $\boldsymbol{\xi}$. Thus, maximization of the full likelihood is done by maximizing each piece separately.

5.3 EM algorithm based on PROS sample

In this section, we introduce an EM algorithm to maximize the log-likelihood function (5.17). The EM algorithm starts with an initial value of the population parameter $\mathbf{\Omega}^{(0)}$ and involves two steps.

5.3.1 E-Step

This step computes the conditional expectation of the full data log-likelihood function (5.17) given the observed data $\mathbf{X} = \mathbf{x}$ by

$$Q(\mathbf{\Omega}, \mathbf{\Omega}^{(0)}) = \mathbb{E}_{\mathbf{\Omega}^{(0)}}[l(\mathbf{\Omega}, \mathbf{Y})|\mathbf{x}], \tag{5.18}$$

where the expectation will be computed by using $\mathbf{\Omega}^{(0)}$ instead of $\mathbf{\Omega}$ in the conditional distribution. On the $(p+1)$-th iteration, the E-step requires one to compute $Q(\mathbf{\Omega}, \mathbf{\Omega}^{(p)})$, where $\mathbf{\Omega}^{(p)}$ is the value of $\mathbf{\Omega}$ after the p-th iteration. This involves the calculation of the conditional expectation of the appropriate latent variables. For the expected value of $\Delta_i^{[d_r,d_h]}(k,u)$, from (5.6), we have

$$\Delta_i^{[d_r,d_h]}(k,u)|\{X_{[d_r]i} = x_{[d_r]i}\} \sim \text{Bin}\left(1, \phi_{i,k,u}^{[d_r,d_h]}(\mathbf{\Omega})\right),$$

where $\text{Bin}(a,b)$ is a binomial distribution with parameters a and b,

$$\phi_{i,k,u}^{[d_r,d_h]}(\mathbf{\Omega}) = \frac{\alpha_{d_r,d_h} B_{u,S-u+1}(F(x_{[d_r]i}; \mathbf{\Psi}))}{m \sum_{h=1}^{n} \sum_{u' \in d_h} \alpha_{d_r,d_h} B_{u',S-u'+1}(F(x_{[d_r]i}; \mathbf{\Psi}))},$$

and $B_{a,b}(\cdot)$ is a beta density function with parameters a and b. By summing $\phi^{[d_r,d_h]}_{i,k,u}(\mathbf{\Omega}^{(p)})$ over k and u, we obtain

$$\phi^{[d_r,d_h]}_{i}(\mathbf{\Omega}^{(p)}) = \sum_{k\in d_r}\sum_{u\in d_h}\phi^{[d_r,d_h]}_{i,k,u}(\mathbf{\Omega}^{(p)}),$$

which is the conditional probability that a randomly selected judgment unit in the subset d_r came from a randomly selected order statistic in the subset d_h given that $X_{[d_r]i}$ has been measured from the subset d_r. For notational convenience, we also write

$$\phi^{[d_r,d_h]}_{i,\cdot,u}(\mathbf{\Omega}^{(p)}) = \sum_{k\in d_r}\phi^{[d_r,d_h]}_{i,k,u}(\mathbf{\Omega}^{(p)}), \quad \text{and} \quad \phi^{(p)}_{r,h} = \sum_{i=1}^{N}\phi^{[d_r,d_h]}_{i}(\mathbf{\Omega}^{(p)}).$$

To obtain the conditional expectation of the latent variables $\mathbf{Z}_i$, $\mathbf{W}_i$ and $\mathbf{V}_i$, we first establish their conditional distributions given $X_{[d_r]i}$ and the latent variable $\Delta^{[d_r,d_h]}_i(k,u)$. Note that from equations (5.14), (5.15) and (5.16), we observe that

$$Z^{[d_r]}_{ij}|\{x_{[d_r]i}, \Delta^{[d_r,d_h]}_i(k,u)=1\} \sim \text{Bin}\left(1, \frac{\pi_j f_j(x_{[d_r]i};\theta_j)}{f(x_{[d_r]i};\mathbf{\Psi})}\right),$$

$$W^{[d_r]}_{ij}|\{x_{[d_r]i}, \Delta^{[d_r,d_h]}_i(k,u)=1\} \sim \text{Bin}\left(u-1, \frac{\pi_j F_j(x_{[d_r]i};\theta_j)}{F(x_{[d_r]i};\mathbf{\Psi})}\right),$$

$$V^{[d_r]}_{ij}|\{x_{[d_r]i}, \Delta^{[d_r,d_h]}_i(k,u)=1\} \sim \text{Bin}\left(S-u, \frac{\pi_j \bar{F}_j(x_{[d_r]i};\theta_j)}{\bar{F}(x_{[d_r]i};\mathbf{\Psi})}\right),$$

where $i=1,\ldots,N$; $r=1,\ldots,n$ and $j=1,\ldots,M$. Let $\tau^{[d_r]}_{u,i,j}(\mathbf{\Omega})$, $\beta^{[d_r]}_{u,i,j}(\mathbf{\Omega})$ and $\gamma^{[d_r]}_{u,i,j}(\mathbf{\Omega})$ denote the conditional expectations of $Z^{[d_r]}_{ij}$, $W^{[d_r]}_{ij}$ and $V^{[d_r]}_{ij}$, given the observed measurement $x_{[d_r]i}$ and the latent variable $\Delta_i^{[d_r,d_h]}(k,u) = 1$, respectively. Using the conditional expectation we have

$$\mathbb{E}_{\mathbf{\Omega}^{(p)}}\left[\Delta_i^{[d_r,d_h]}(k,u) Z_{ij}^{[d_r]} \middle| x_{[d_r]i}\right] = \phi_{i,k,u}^{[d_r,d_h]}(\mathbf{\Omega}^{(p)})\, \tau_{u,i,j}^{[d_r]}(\mathbf{\Omega}^{(p)}),$$

$$\mathbb{E}_{\mathbf{\Omega}^{(p)}}\left[\Delta_i^{[d_r,d_h]}(k,u) W_{ij}^{[d_r]} \middle| x_{[d_r]i}\right] = \phi_{i,k,u}^{[d_r,d_h]}(\mathbf{\Omega}^{(p)})\, \beta_{u,i,j}^{[d_r]}(\mathbf{\Omega}^{(p)}),$$

$$\mathbb{E}_{\mathbf{\Omega}^{(p)}}\left[\Delta_i^{[d_r,d_h]}(k,u) V_{ij}^{[d_r]} \middle| x_{[d_r]i}\right] = \phi_{i,k,u}^{[d_r,d_h]}(\mathbf{\Omega}^{(p)})\, \gamma_{u,i,j}^{[d_r]}(\mathbf{\Omega}^{(p)}).$$

The above expressions will help us to calculate the conditional expectation of the log-likelihood function in (5.18) given the observed data as follows

$$Q(\mathbf{\Omega}, \mathbf{\Omega}^{(p)}) = Q_1(\boldsymbol{\alpha}, \mathbf{\Omega}^{(p)}) + Q_2(\boldsymbol{\pi}, \mathbf{\Omega}^{(p)}) + Q_3(\boldsymbol{\xi}, \mathbf{\Omega}^{(p)}),$$

where

$$\begin{aligned} Q_1(\boldsymbol{\alpha}, \mathbf{\Omega}^{(p)}) &= \sum_{i=1}^{N} \sum_{r=1}^{n} \sum_{k \in d_r} \sum_{h=1}^{n} \sum_{u \in d_h} \phi_{i,k,u}^{[d_r,d_h]}(\mathbf{\Omega}^{(p)}) \log \alpha_{d_r,d_h} \\ &= \sum_{r=1}^{n} \sum_{h=1}^{n} \phi_{r,h}^{(p)} \log \alpha_{d_r,d_h}, \\ Q_2(\boldsymbol{\pi}, \mathbf{\Omega}^{(p)}) &= \sum_{i=1}^{N} \sum_{r=1}^{n} \sum_{k \in d_r} \sum_{h=1}^{n} \sum_{u \in d_h} \sum_{j=1}^{M} \phi_{i,k,u}^{[d_r,d_h]}(\mathbf{\Omega}^{(p)}) \log \pi_j \\ &\quad \times \left\{ \tau_{u,i,j}^{[d_r]}(\mathbf{\Omega}^{(p)}) + \beta_{u,i,j}^{[d_r]}(\mathbf{\Omega}^{(p)}) + \gamma_{u,i,j}^{[d_r]}(\mathbf{\Omega}^{(p)}) \right\}, \end{aligned}$$

and

$$Q_3(\boldsymbol{\xi}, \boldsymbol{\Omega}^{(p)}) = \sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\Bigg\{ \log f_j(x_{[d_r]i};\theta_j)\sum_{h=1}^{n}\sum_{u\in d_h}\phi_{i,\cdot,u}^{[d_r,d_h]}(\boldsymbol{\Omega}^{(p)})\tau_{u,i,j}^{[d_r]}(\boldsymbol{\Omega}^{(p)})$$

$$+\log F_j(x_{[d_r]i};\theta_j)\sum_{h=1}^{n}\sum_{u\in d_h}\phi_{i,\cdot,u}^{[d_r,d_h]}(\boldsymbol{\Omega}^{(p)})\beta_{u,i,j}^{[d_r]}(\boldsymbol{\Omega}^{(p)})$$

$$+\log \bar{F}_j(x_{[d_r]i};\theta_j)\sum_{h=1}^{n}\sum_{u\in d_h}\phi_{i,\cdot,u}^{[d_r,d_h]}(\boldsymbol{\Omega}^{(p)})\beta_{u,i,j}^{[d_r]}(\boldsymbol{\Omega}^{(p)})\Bigg\}.$$

5.3.2 M-Step

The M-step considers the maximization of the conditional log-likelihood function $Q(\boldsymbol{\Omega}, \boldsymbol{\Omega}^{(p)})$ over the parameter space $\boldsymbol{\Omega}$ to obtain the updated estimates $\boldsymbol{\Omega}^{(p+1)} = (\boldsymbol{\Psi}^{(p+1)}, \boldsymbol{\alpha}^{(p+1)})$. The maximization of $Q_1(\boldsymbol{\alpha}, \boldsymbol{\Omega}^{(p)})$ should be done under the constraint that $\boldsymbol{\alpha}$ is a doubly stochastic matrix. In order to force the constraint, we use the Lagrangian multipliers $\lambda = (\lambda_1, \ldots, \lambda_n)$ and rewrite the expected log-likelihood function as

$$Q_1(\boldsymbol{\alpha}, \boldsymbol{\Omega}^{(p)};\lambda) = \sum_{h=1}^{n}\left\{\sum_{h'=1}^{h-1}\phi_{h,h'}(\boldsymbol{\Omega}^{(p)})\log\alpha_{d_h,d_{h'}} + \sum_{h'=h}^{n}\phi_{h,h'}(\boldsymbol{\Omega}^{(p)})\log\alpha_{d_{h'},d_h}\right\}$$

$$+\sum_{h=1}^{n}\lambda_h\left\{\sum_{h'=1}^{h-1}\alpha_{d_h,d_{h'}} + \sum_{h'=h}^{n}\alpha_{d_{h'},d_h} - 1\right\}.$$

In this equation, we used the fact that $\alpha_{d_h,d_{h'}} = \alpha_{d_{h'},d_h}$. The details of the maximization algorithm are given in Arslan and Ozturk (2013).

The maximization of $Q_2(\boldsymbol{\pi}, \boldsymbol{\Omega}^{(p)})$ with respect to π_j, $j = 1, \ldots, M$, does not depend on the updated estimates $\boldsymbol{\xi}^{(p+1)}$ of the parameters $\boldsymbol{\xi}$ in $\boldsymbol{\Psi}$ and it is accomplished by using the Lagrangian multipliers to enforce the constraint $\sum_{j=1}^{M} \pi_j = 1$. After a little algebra, the estimator is given by

$$\hat{\pi}_j^{(p+1)} = c_4 \sum_{i=1}^{N} \sum_{r=1}^{n} \sum_{h=1}^{n} \sum_{u \in d_h} \phi_{i,\cdot,u}^{[d_r,d_h]}(\boldsymbol{\Omega}^{(p)}) \left\{ \tau_{u,i,j}^{[d_r]}(\boldsymbol{\Omega}^{(p)}) + \beta_{u,i,j}^{[d_r]}(\boldsymbol{\Omega}^{(p)}) + \gamma_{u,i,j}^{[d_r]}(\boldsymbol{\Omega}^{(p)}) \right\}.$$

where $c_4 = \frac{1}{N(nm)^3}$. Finally, the maximization of $Q_3(\boldsymbol{\xi}, \boldsymbol{\Omega}^{(p)})$ with respect to $\boldsymbol{\xi}$ (to obtain the updated estimate $\boldsymbol{\xi}^{(p+1)}$) is achieved by solving the following estimating equation in $\boldsymbol{\xi}$

$$\begin{aligned} &\sum_{i=1}^{N} \sum_{r=1}^{n} \sum_{j=1}^{M} \frac{\frac{\boldsymbol{\partial}}{\boldsymbol{\partial \xi}} f_j(x_{[d_r]i}; \theta_j)}{f_j(x_{[d_r]i}; \theta_j)} \sum_{h=1}^{n} \sum_{u \in d_h} \phi_{i,\cdot,u}^{[d_r,d_h]}(\boldsymbol{\Omega}^{(p)}) \tau_{u,i,j}^{[d_r]}(\boldsymbol{\Omega}^{(p)}) \\ &+ \sum_{i=1}^{m} \sum_{r=1}^{k} \sum_{j=1}^{M} \frac{\boldsymbol{\partial}}{\boldsymbol{\partial \xi}} F_j(x_{[d_r]i}; \theta_j) \left(\sum_{h=1}^{n} \sum_{u \in d_h} \phi_{i,\cdot,u}^{[d_r,d_h]}(\boldsymbol{\Omega}^{(p)}) \left\{ \frac{\beta_{u,i,j}^{[d_r]}(\boldsymbol{\Omega}^{(p)})}{F_j(x_{[d_r]i}; \theta_j)} - \frac{\gamma_{u,i,j}^{[d_r]}(\boldsymbol{\Omega}^{(p)})}{\bar{F}_j(x_{[d_r]i}; \theta_j)} \right\} \right) \\ &= 0. \end{aligned} \tag{5.19}$$

In order to compute the MLE of $\boldsymbol{\Omega}$, the E- and M-steps are alternated repeatedly until $\|\boldsymbol{\Omega}^{(p+1)} - \boldsymbol{\Omega}^{(p)}\|_\infty$ becomes negligible, where $|| \cdot ||_\infty$ is the sup-norm.

5.4 Modified EM algorithm

Similarly to Subsection 2.2.3, in this section we propose a modified EM algorithm by replacing the hazard rate function in the log-likelihood function (5.19) with its expectation. This enables us to reduce the computational complexity of ML estimation based on PROS sampling to the level of computational complexity of ML estimation under SRS. We also note that, this approach only affects the maximization of the conditional log-likelihood function $Q_3(\boldsymbol{\xi}, \boldsymbol{\Omega}^{(p)})$.

Lemma 5.3. *Let $X_{[d_r]i}$ be a fully measured observation from the subset d_r of the PROS design with pdf (5.5). Suppose $W_{ij}^{[d_r]}$ and $V_{ij}^{[d_r]}$ are the j-th elements of the latent variables $W_i^{[d_r]}$ and $V_i^{[d_r]}$ associated with $X_{[d_r]i}$. Then, for any function $G(\cdot)$ (subject to the finiteness of the expectations) we have*

$$(a) \quad \sum_{r=1}^{n}\sum_{k\in d_r}\sum_{h=1}^{n}\sum_{u\in d_h}\sum_{j=1}^{M}\mathbb{E}\left(\Delta_i^{[d_r,d_h]}(k,u)W_{ij}^{[d_r]}G(X_{[d_r]i})\right)$$

$$= S(S-1)\sum_{j=1}^{M}\pi_j\mathbb{E}_{\boldsymbol{\Psi}}\left[G(X)F_j(X;\theta_j)\right],$$

$$(b) \quad \sum_{r=1}^{n}\sum_{k\in d_r}\sum_{h=1}^{n}\sum_{u\in d_h}\sum_{j=1}^{M}\mathbb{E}\left(\Delta_i^{[d_r,d_h]}(k,u)V_{ij}^{[d_r]}G(X_{[d_r]i})\right)$$

$$= S(S-1)\sum_{j=1}^{M}\pi_j\mathbb{E}_{\boldsymbol{\Psi}}\left[G(X)\bar{F}_j(X;\theta_j)\right],$$

and the expectations on the right sides are with respect to the FMM (1.1).

Proof. Despite the lack of independence between the vectors $\Delta_i^{[d_r,d_h]}(k,u)|x_{[d_r]i}$ and $W_{ij}^{[d_r]}|\{x_{[d_r]i}, \delta_i^{[d_r,d_h]}\}$, using the distribution of $\Delta_i^{[d_r,d_h]}(k,u)|\{X_{[d_r]i} = x_{[d_r]i}\}$ and the total law of expectation, we have

$$\sum_{r=1}^{n}\sum_{k\in d_r}\sum_{h=1}^{n}\sum_{u\in d_h}\sum_{j=1}^{M}\mathbb{E}\left(\Delta_i^{[d_r,d_h]}(k,u)W_{ij}^{[d_r]}G(X_{[d_r]i})\right)$$

$$=\sum_{r=1}^{n}\sum_{k\in d_r}\sum_{h=1}^{n}\sum_{u\in d_h}\sum_{j=1}^{M}\mathbb{E}\left(G(X_{[d_r]i})\frac{(u-1)\pi_j F_j(X_{[d_r]i};\theta_j)}{F(X_{[d_r]i};\boldsymbol{\Psi})}\frac{\alpha_{d_r,d_h}f^{(u:S)}(X_{[d_r]i};\boldsymbol{\Psi})}{f_{[d_r]}(X_{[d_r]i};\boldsymbol{\Psi})}\right)$$

$$=\sum_{r=1}^{n}\sum_{k\in d_r}\sum_{h=1}^{n}\sum_{u\in d_h}\sum_{j=1}^{M}\int G(x)\frac{(u-1)\pi_j F_j(x;\theta_j)}{F(x;\boldsymbol{\Psi})}\alpha_{d_r,d_h}f^{(u:S)}(x;\boldsymbol{\Psi})dx$$

$$=\sum_{r^*=1}^{S}\sum_{u^*=1}^{S}\sum_{j=1}^{M}\int G(x)\frac{(u^*-1)\pi_j F_j(x;\theta_j)}{F(x;\boldsymbol{\Psi})}\alpha_{r^*,u^*}f^{(u^*:S)}(x;\boldsymbol{\Psi})dx$$

$$=\sum_{u^*=1}^{S}\sum_{j=1}^{M}\int G(x)\frac{(u^*-1)\pi_j F_j(x;\theta_j)}{F(x;\boldsymbol{\Psi})}f^{(u^*:S)}(x;\boldsymbol{\Psi})dx$$

$$=S\sum_{j=1}^{M}\pi_j\int G(x)F_j(x;\theta_j)f(x;\boldsymbol{\Psi})\left(\sum_{u^*=1}^{S}(u^*-1)\binom{S-1}{u^*-1}[F(x;\boldsymbol{\Psi})]^{u^*-2}[1-F(x;\boldsymbol{\Psi})]^{S-u^*}\right)dx$$

$$=S(S-1)\sum_{j=1}^{M}\pi_j\mathbb{E}_{\boldsymbol{\Psi}}\left[G(X)F_j(X;\theta_j)\right],$$

where the last equality holds as

$$\sum_{u^*=1}^{S}(u^*-1)\binom{S-1}{u^*-1}[F(x;\boldsymbol{\Psi})]^{u^*-2}[1-F(x;\boldsymbol{\Psi})]^{S-u^*}=S-1.$$

$\square$

Using Lemma 5.3 and considering $G_1(x_{[d_r]i})=\frac{\partial}{\partial\boldsymbol{\xi}}F_j(x_{[d_r]i};\theta_j)/F_j(x_{[d_r]i};\theta_j)$, one can easily show that

$$\sum_{r=1}^{n}\sum_{k\in d_r}\sum_{h=1}^{n}\sum_{u\in d_h}\sum_{j=1}^{M}\mathbb{E}\left(\Delta_i^{[d_r,d_h]}(k,u)W_{ij}^{[d_r]}G_1(X_{[d_r]i})\right)=c_5\sum_{j=1}^{M}\pi_j\mathbb{E}_{\boldsymbol{\Psi}}\left[\frac{\partial}{\partial\boldsymbol{\xi}}F_j(X;\theta_j)\right],$$

and similarly, considering $G_2(x_{(r)i}) = \frac{\partial}{\partial \xi} F_j(x_{(r)i};\theta_j)/\bar{F}_j(x_{(r)i};\theta_j)$ in Lemma 5.3, we get

$$\sum_{r=1}^{n}\sum_{k\in d_r}\sum_{h=1}^{n}\sum_{u\in d_h}\sum_{j=1}^{M}\mathbb{E}\left(\Delta_i^{[d_r,d_h]}(k,u)V_{ij}^{[d_r]}G_2(X_{[d_r]i})\right) = c_5\sum_{j=1}^{M}\pi_j\mathbb{E}_{\boldsymbol{\Psi}}\left[\frac{\partial}{\partial \boldsymbol{\xi}}F_j(X;\theta_j)\right].$$

where $c_5 = S(S-1)$. Now, using (5.19) and the above equalities, we get the following modified estimating equation to update $\boldsymbol{\xi}$ in the M-step of the EM algorithm. This approach leads to approximate ML estimate of $\boldsymbol{\xi}$:

$$\sum_{i=1}^{N}\sum_{r=1}^{n}\sum_{j=1}^{M}\frac{\frac{\partial}{\partial \boldsymbol{\xi}}f_j(x_{[d_r]i};\theta_j)}{f_j(x_{[d_r]i};\theta_j)}\left\{\sum_{h=1}^{n}\sum_{u\in d_h}\phi_{i,\cdot,u}^{[d_r,d_h]}(\boldsymbol{\Omega}^{(p)})\tau_{u,i,j}^{[d_r]}(\boldsymbol{\Omega}^{(p)})\right\} = 0. \qquad (5.20)$$

Note that (5.20) is similar to the updating equation for parameters of the component densities under the SRS design. Therefore, the modified version of the proposed EM algorithm for PROS design requires the same computational efforts as the EM algorithm based on SRS to update $\boldsymbol{\xi}$. However, in the current setup, we still take advantage of updating the mixing proportions by using the information contained in all the latent variables. Similar to the SRS case, one nice feature of this modified version of the EM algorithm is that the solutions to (5.20) often exist in closed form.

5.5 Classifications of PROS samples

In this section we first consider the problem of classifying an observed PROS sample into the components of the FMM (1.1) by inferring the component membership of each observation. To achieve this, we use the model-based classification technique by assigning each observation to different components according to their posterior probabilities. Suppose $i \in \{1,\ldots,N\}$ is fixed and $x_{[d_r]i}$ is observed. To classify $x_{[d_r]i}$, we estimate the component membership vector $\mathbf{Z}_i^{[d_r]}$ by $\hat{\mathbf{Z}}_i^{[d_r]}$, where $\hat{\mathbf{Z}}_i^{[d_r]} = (\hat{Z}_{ij}^{[d_r]})$ and $\hat{Z}_{ij}^{[d_r]}$ is defined by

$$\hat{Z}_{ij}^{[d_r]} = \begin{cases} 1, & \text{if } j = \text{argmax}_h \eta_h(x_{[d_r]i}; \Omega), \\ 0, & \text{otherwise,} \end{cases}$$

for $j = 1, \ldots, M$, where

$$\eta_h(x_{[d_r]i}; \boldsymbol{\Omega}) = \mathbb{P}(Z_{ih}^{[d_r]} = 1 | x_{[d_r]i}).$$

Also, by using (5.13) and the definitions of $\mathbf{z}_i^{[d_r]}$ and $\boldsymbol{\delta}_i^{[d_r]}$, we have

$$\begin{aligned} \mathbb{P}(\mathbf{Z}_i^{[d_r]} = \mathbf{z}_i^{[d_r]} | x_{[d_r]i}) &= \frac{\sum_{\boldsymbol{\delta}_i^{[d_r]}} f(x_{[d_r]i}, \mathbf{z}_i^{[d_r]}, \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\Omega})}{\sum_{\mathbf{z}_i^{[d_r]}} \sum_{\boldsymbol{\delta}_i^{[d_r]}} f(x_{[d_r]i}, \mathbf{z}_i^{[d_r]}, \boldsymbol{\delta}_i^{[d_r]}; \boldsymbol{\Omega})} \\ &= \frac{\prod_{j=1}^{M} \left\{ \pi_j f_j(x_{[d_r]i}; \theta_j) \right\}^{z_{ij}^{[d_r]}}}{\sum_{j=1}^{M} \pi_j f_j(x_{[d_r]i}; \theta_j)}, \end{aligned}$$

and so

$$\eta_h(x_{[d_r]i}; \boldsymbol{\Omega}) = \eta_h(x_{[d_r]i}; \boldsymbol{\Psi}) = \frac{\pi_h f_h(x_{[d_r]i}; \theta_h)}{f(x_{[d_r]i}; \boldsymbol{\Psi})}. \tag{5.21}$$

In practice, the so-called plug-in rule $\eta_h(x_{[d_r]i}; \hat{\boldsymbol{\Psi}})$ can be used to estimate the posterior probabilities $\eta_h(x_{[d_r]i}; \boldsymbol{\Psi})$, where $\hat{\boldsymbol{\Psi}}$ denotes the PROS estimates of the unknown parameter vector $\boldsymbol{\Psi}$.Every observation is then assigned to the component having the highest estimated posterior probability that the observation originates from this component. It is interesting to note that the expression we obtained in (5.21) as the posterior probability of component membership of each PROS observation is equal to the commonly used expression for the SRS design. However, as we will see in Section 5.7, the PROS estimates of the FMM tend to provide a better fit to the model and the mixing proportions are estimated with better precision than their SRS counterparts. Thus we observe a better component-wise classification for PROS samples when compared to their SRS counterparts.

One can also perform a model-based subset-wise classification of the PROS sample which helps to estimate the true subset-membership of a new observation obtained through the PROS design. This could also be useful to estimate and monitor the misplacement error probability matrix associated with the PROS design. To this end, once $\hat{\mathbf{\Omega}}$ is obtained, estimates of the posterior probabilities of the subset misplacement error of each observed data can be formed to perform a probabilistic subset-wise classification of the data. Under the PROS design, the subset misplacement error of an observation $x_{[d_r]i}$, obtained from the subset d_r, can be defined based on the posterior probability that $x_{[d_r]i}$ is judged to belong to the subset d_r when its true subset is d_h using the following formula

$$\phi^{[d_r,d_h]}(x_{[d_r]i};\hat{\mathbf{\Omega}}) = \frac{\hat{\alpha}_{d_r,d_h} \sum_{u\in d_h} B_{u,S-u+1}(F(x_{[d_r]i};\hat{\mathbf{\Psi}}))}{\sum_{h=1}^{n} \sum_{u'\in d_h} \hat{\alpha}_{d_r,d_h} B_{u',S-u'+1}(F(x_{[d_r]i};\hat{\mathbf{\Psi}}))}.$$

Now, $x_{[d_r]i}$ will be classified into the subset d_h if

$$\phi^{[d_r,d_h]}(x_{[d_r]i};\hat{\mathbf{\Omega}}) > \phi^{[d_r,d_{h'}]}(x_{[d_r]i};\hat{\mathbf{\Omega}}),$$

for all $h^{'} \neq h$, $h^{'} = 1, \ldots, n$.

5.6 Simulation study

In this section, we investigate the finite sample properties of the ML estimates of the parameters of a FMM based on PROS data. The ML estimates are computed using the EM- and modified EM-algorithms. We generated data sets from a homoscedastic mixture of two univariate normal distributions based on the PROS, SRS and RSS sampling designs. The parameter of the mixture model and the number of judgment subsets in the PROS design are selected as $\boldsymbol{\Psi} = (\pi, \mu_1, \mu_2, \sigma) = (0.8, -2, 1, 1)$ and $n = 3$, respectively. Each study comprised 3000 replications with a cycle size of $N = 30$. In each replication of the simulation, the initial values of $\boldsymbol{\Psi} = (\pi, \mu_1, \mu_2, \sigma)$ in the EM-algorithm are computed following the method of Furman and Lindsay (1994) by treating the PROS sample as a simple random sample.

We aim to assess the small sample performance of the ML estimates under the PROS sampling design and compare it with their competitors under SRS and RSS designs. In particular, the simulation study investigates two main features of the estimators: robustness against possible ranking error and the efficiency of the estimators. For robustness, we use four different misplacement error structures $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\boldsymbol{\alpha}_3$ and $\boldsymbol{\alpha}_4$. In these models $\boldsymbol{\alpha}_1$ leads to an error-free ranking model, while $\boldsymbol{\alpha}_2$, $\boldsymbol{\alpha}_3$, and $\boldsymbol{\alpha}_4$ induce misplacement errors among judgment subsets. Defining

$$\boldsymbol{\alpha}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \boldsymbol{\alpha}_2 = \begin{bmatrix} 0.9 & 0.10 & 0 \\ 0.10 & 0.80 & 0.10 \\ 0 & 0.10 & 0.90 \end{bmatrix},$$

$$\boldsymbol{\alpha}_3 = \begin{bmatrix} 0.75 & 0.15 & 0.10 \\ 0.15 & 0.70 & 0.15 \\ 0.10 & 0.15 & 0.75 \end{bmatrix}, \boldsymbol{\alpha}_4 = \begin{bmatrix} 0.5 & 0.35 & 0.15 \\ 0.35 & 0.3 & 0.35 \\ 0.15 & 0.35 & 0.50 \end{bmatrix},$$

the misplacement error is gradually more severe. Since $\boldsymbol{\alpha}_i,\ i = 1, \ldots, 4,$ is a doubly stochastic and symmetric matrix, we only need to estimate three independent parameters (entries). The other parameters of $\boldsymbol{\alpha}_i$ can be computed by using the constraints, $\sum_{h=1}^{n} \alpha_{h',h} = \sum_{h'=1}^{n} \alpha_{h',h} = 1$. For the initial values of $\boldsymbol{\alpha}$, we use the random ranking assumption, namely, $\{\alpha_{i,j}^{(0)} = \frac{1}{3}; i, j = 1, 2, 3\}$. The estimators are computed based on the stopping criteria $\|\hat{\boldsymbol{\Psi}}^{(k+1)} - \hat{\boldsymbol{\Psi}}^{(k)}\|_\infty < 10^{-5}$.

The ML estimates of the FMM based on RSS data under perfect ranking assumption is developed in Chapter 2. For each one of the ranking models $\boldsymbol{\alpha}_i$, $i = 1, \ldots, 4$, the ML estimates of the parameters of the model are computed for PROS and RSS sampling designs. For the RSS design, samples are generated based on the PROS design with $m = 1$ (number of units in each subset) and ranking probability models $\boldsymbol{\alpha}_i$ so that both PROS and RSS sampling designs have the same judgment ranking error. In all three sampling designs (PROS, RSS, SRS), the number of measured units is matched so that a meaningful efficiency comparison can be made. The ML estimates based on the RSS design are computed under perfect ranking by using the estimators developed in Chapter 2.

The first part of the simulation study considers the robustness of the estimators against imperfect ranking. Tables 5.1 and 5.2 present the amount of biases for the parameter estimates for ranking models $\boldsymbol{\alpha}_i$, $i = 1, \ldots, 4$. It is clear that PROS estimators of $\boldsymbol{\alpha}_i$ have very little bias compared with their RSS counterparts and one may consider them as unbiased estimators of $\boldsymbol{\alpha}_i$. Also, we observe that the PROS estimators of the FMM parameters $\boldsymbol{\Psi}$ have the lowest bias compared with their SRS and RSS counterparts, regardless of the quality of ranking errors. It is interesting to note that the results under the modified EM algorithm show similar

Table 5.1: Bias estimates of the ML estimates under complete (Comp.) and modified (Mod.) EM-algorithm techniques with between-judgment class ranking error model $\boldsymbol{\alpha}_i$, $i = 1, 2, 3, 4$, when sample size is 30 such that $N = 10$, $n = 3$, and $S = 12$.

EM	$\boldsymbol{\alpha}$	Design	α_{11}	α_{12}	α_{22}	π	μ_1	μ_2	σ
		SRS				-0.0549	-0.1204	-0.1113	-0.1071
Comp.	$\boldsymbol{\alpha}_1$	PROS	-0.0143	0.0140	-0.0288	-0.0148	-0.0255	-0.0251	-0.0466
		RSS	-0.0615	0.0540	-0.1070	-0.0204	-0.0357	-0.0377	-0.0728
	$\boldsymbol{\alpha}_2$	PROS	-0.0015	-0.0044	0.0142	-0.0161	-0.0378	-0.0036	-0.0463
		RSS	-0.0314	0.0151	-0.0275	-0.0255	-0.0400	0.0240	-0.0764
	$\boldsymbol{\alpha}_3$	PROS	0.0018	-0.0045	0.0140	-0.0166	-0.0340	0.0174	-0.0312
		RSS	-0.0332	0.0393	-0.0538	-0.0220	-0.0611	-0.0274	-0.0693
	$\boldsymbol{\alpha}_4$	PROS	0.0010	-0.0022	0.0085	-0.0199	-0.0395	0.0303	-0.0431
		RSS	0.0243	-0.0326	0.0488	-0.0311	-0.0882	0.0335	-0.0698
Mod.	$\boldsymbol{\alpha}_1$	PROS	-0.0171	0.0169	-0.0330	-0.0190	-0.0416	0.0117	-0.0482
		RSS	-0.0634	0.0554	-0.1160	-0.0289	-0.0661	-0.0121	-0.0689
	$\boldsymbol{\alpha}_2$	PROS	-0.0019	-0.0050	0.0154	-0.0179	-0.0374	0.0110	-0.0579
		RSS	-0.0400	0.0238	-0.0425	-0.0262	-0.0773	-0.0220	-0.0754
	$\boldsymbol{\alpha}_3$	PROS	-0.0011	0.0046	-0.0183	-0.0326	-0.0248	-0.0714	-0.0398
		RSS	-0.0224	0.0265	-0.0414	-0.0297	-0.0679	-0.0237	-0.0701
	$\boldsymbol{\alpha}_4$	PROS	0.0047	-0.0071	0.0117	-0.0246	-0.0535	-0.1617	-0.0492
		RSS	0.0129	-0.0332	0.0480	-0.0394	-0.0789	-0.0465	-0.0750

behaviour, which is promising especially for practical purposes.

Table 5.2: Bias estimates of the ML estimates under complete (Comp.) and modified (Mod.) EM-algorithm techniques with between-judgment class ranking error model $\boldsymbol{\alpha}_i$, $i = 1, 2, 3, 4$, when sample size is 30 such that $N = 10$, $n = 3$, and $S = 18$.

EM	$\boldsymbol{\alpha}$	Design	α_{11}	α_{12}	α_{22}	π	μ_1	μ_2	σ
		SRS				-0.0549	-0.1204	-0.1113	-0.1071
Comp.	$\boldsymbol{\alpha}_1$	PROS	-0.0090	0.0090	-0.0179	-0.0113	-0.0219	0.0219	-0.0337
		RSS	-0.0605	0.0532	-0.1044	-0.0257	-0.0520	-0.0406	-0.0692
	$\boldsymbol{\alpha}_2$	PROS	0.0014	-0.0073	0.0168	-0.0143	-0.0338	0.0305	-0.0541
		RSS	-0.0283	0.0110	-0.0232	-0.0263	-0.0428	-0.0250	-0.0727
	$\boldsymbol{\alpha}_3$	PROS	0.0032	-0.0022	0.0135	-0.0205	-0.0484	0.0091	-0.0480
		RSS	-0.0213	0.0237	-0.0389	-0.0290	-0.0693	0.0135	-0.0706
	$\boldsymbol{\alpha}_4$	PROS	0.0075	-0.0021	0.0047	-0.0269	-0.0411	0.0150	-0.0558
		RSS	0.0218	-0.0302	0.0635	-0.0303	-0.0686	-0.0206	-0.0814
Mod.	$\boldsymbol{\alpha}_1$	PROS	-0.0125	0.0125	-0.0244	-0.0215	-0.0483	-0.0132	-0.0472
		RSS	-0.0645	0.0575	-0.1170	-0.0205	-0.0629	0.0281	-0.0623
	$\boldsymbol{\alpha}_2$	PROS	-0.0011	-0.0043	0.0112	-0.0258	-0.0367	-0.0406	-0.0508
		RSS	-0.0396	0.0247	-0.0335	-0.0244	-0.0524	0.0327	-0.0780
	$\boldsymbol{\alpha}_3$	PROS	-0.0053	0.0094	-0.0073	-0.0361	-0.0303	-0.0851	-0.0262
		RSS	-0.0240	0.0267	-0.0485	-0.0307	-0.0549	-0.0322	-0.0735
	$\boldsymbol{\alpha}_4$	PROS	0.0007	-0.0031	0.0054	-0.0702	-0.0231	-0.2559	-0.0168
		RSS	0.0202	-0.0314	0.0668	-0.0377	-0.0657	-0.0411	-0.0713

The second part of the simulation study considers the relative efficiencies of

Table 5.3: MSEs of PROS and RSS estimators for between-judgment class ranking error model $\boldsymbol{\alpha}_i$, $i = 2, 3, 4$, when same size is 30 such that $N = 10$, $n = 3$, and $S \in \{12, 18\}$.

			$S = 12$			$S = 18$		
EM	$\boldsymbol{\alpha}$	Design	α_{11}	α_{12}	α_{22}	α_{11}	α_{12}	α_{22}
Comp.	$\boldsymbol{\alpha}_2$	PROS	0.0089	0.0089	0.0176	0.0076	0.0076	0.0157
		RSS	0.0240	0.0228	0.0488	0.0234	0.0218	0.0477
	$\boldsymbol{\alpha}_3$	PROS	0.0156	0.0136	0.0284	0.0147	0.0121	0.0251
		RSS	0.0387	0.0374	0.0768	0.0380	0.0361	0.0767
	$\boldsymbol{\alpha}_4$	PROS	0.0215	0.0192	0.0393	0.0192	0.0166	0.0318
		RSS	0.0431	0.0436	0.0948	0.0442	0.0434	0.0959
Mod.	$\boldsymbol{\alpha}_2$	PROS	0.0090	0.0091	0.0184	0.0086	0.0085	0.0165
		RSS	0.0269	0.0251	0.0520	0.0256	0.0243	0.0501
	$\boldsymbol{\alpha}_3$	PROS	0.0169	0.0136	0.0290	0.0145	0.0119	0.0239
		RSS	0.0394	0.0353	0.0807	0.0386	0.0368	0.0801
	$\boldsymbol{\alpha}_4$	PROS	0.0252	0.0212	0.0377	0.0211	0.0170	0.0318
		RSS	0.0433	0.0423	0.0890	0.0432	0.0441	0.0859

the PROS sample estimators with respect to RSS and SRS estimators. Relative efficiencies are given in terms of the ratio of mean square errors of the estimators based on RSS, SRS and PROS sampling designs as

$$RE_1 = \frac{\text{MSE(RSS)}}{\text{MSE(PROS)}} \text{ and } RE_2 = \frac{\text{MSE(SRS)}}{\text{MSE(PROS)}}.$$

The values of RE_1 and RE_2 greater then one indicate that the PROS sample estimators have higher efficiency than their competitor estimators based on RSS and SRS designs. These relative efficiencies are given in Table 5.4. We observe that the PROS estimators of the $\boldsymbol{\alpha}_i$s and the FMM parameters $\boldsymbol{\Psi}$ are significantly better than their SRS and RSS counterparts, regardless of the quality of ranking errors. The efficiency of PROS estimators is increased with the set size S. Also, results under the modified EM algorithm show similar behaviour, which is very important from the computational cost perspective.

Table 5.4: REs of PROS estimators with respect to RSS and SRS estimator for between-judgment class ranking error model $\boldsymbol{\alpha}_i$, $i = 1, 2, 3, 4$, when same size is 30 such that $N = 10$, $n = 3$, and $S \in \{12, 18\}$.

			$S = 12$				$S = 18$			
EM	$\boldsymbol{\alpha}$	RE	π	μ_1	μ_2	σ	π	μ_1	μ_2	σ
Comp.	$\boldsymbol{\alpha}_1$	RE_1	1.4081	1.7439	1.3527	1.2700	2.0129	2.3710	1.2618	1.3723
		RE_2	2.7894	2.9834	1.7912	1.9892	3.2669	3.5963	1.4324	2.3886
	$\boldsymbol{\alpha}_2$	RE_1	1.4014	1.5492	1.6203	1.2840	1.4144	1.2910	1.1445	1.2447
		RE_2	2.5806	2.8769	1.9890	1.9800	2.5893	2.7652	1.7126	2.0365
	$\boldsymbol{\alpha}_3$	RE_1	1.4776	1.5843	1.7131	1.2665	1.4404	1.5050	1.4255	1.3199
		RE_2	2.5433	2.8185	1.1307	1.9481	2.4740	2.8612	1.5921	1.9051
	$\boldsymbol{\alpha}_4$	RE_1	1.5387	1.8249	1.4806	1.3433	1.0700	1.5838	1.4469	1.2167
		RE_2	2.3743	2.4780	1.0839	1.8424	1.8875	1.9660	1.0892	1.8055
Mod.	$\boldsymbol{\alpha}_1$	RE_1	1.6518	1.3953	1.3487	1.2491	1.9302	1.9012	1.2646	1.3388
		RE_2	2.5581	2.4729	1.6801	1.7701	3.1324	3.4532	1.6440	2.1511
	$\boldsymbol{\alpha}_2$	RE_1	1.3183	1.4449	1.4498	1.1365	1.3893	1.3179	1.1288	1.2318
		RE_2	2.3436	2.7380	1.8935	1.7525	2.4236	2.6897	1.9505	1.7445
	$\boldsymbol{\alpha}_3$	RE_1	1.3944	1.4111	1.6176	1.1845	1.4540	1.5203	1.5516	1.2262
		RE_2	2.4141	2.6083	1.6633	1.4266	2.3146	2.1266	1.2486	1.5064
	$\boldsymbol{\alpha}_4$	RE_1	1.2641	1.7560	1.6420	1.1103	1.2007	1.8976	1.7582	1.2415
		RE_2	2.0461	2.2601	1.0577	1.8156	1.6762	1.9619	1.0927	1.7476

5.7 Application

Spot (Leiostomus xanthurus) is an important fish species in the Chesapeake Bay area. They usually appear in all areas of the Chesapeake Bay in the late spring and will remain until fall when water temperature starts to go down. Spots are a substantial food source for other fish species including striped bass, bluefish, weakfish, shark and flounder, as well as many bird species. As one of the most frequently caught species, they are important for both commercial and recreational fisheries in the area. The commercial catch of Spot has averaged 8.2 million lbs per year since 1950. Recreational catches throughout the region averaged 3.9 million lbs per year since 1981. Because of an increase in commercial and recreational fishing pressure, the Chesapeake Bay Program (CBP) adopted a fishery management plan for Spot in 1991. The goal of the plan is to protect the Spot as a resource in the Chesapeake Bay, its tributaries, and coastal waters, while providing the greatest long term ecological, economic, and social benefits from their usage over time. In spite of these efforts, there continues to be a downward trend in the number of Spots in the Chesapeake Bay. Recent research suggests that chemical pollution may have a serious impact on maturation of Spot adults, which may account for the continued decline in population, see Rickabaugh and Capossela (2011). Analyzing the age structure of the population of Spots (and any other species under threat) is vital and provides an understanding of multiple effects on the species population dynamics. Age data are especially important when managing fish species, like Spot, which migrate in and out of Chesapeake Bay waters and that are targeted by both recreational and commercial fishermen.

In this section, we consider the age (from otoliths) and length of 403 Virginia-Chesapeake Bay Spot as our population. The data set is available online in R under the FSAdata package (Ogle, 2013) and it is extracted from Table 1 in Chapter 8 of the Virginia Marine Resources Commissions's Final Report on fish aging, 2002 (see, http://ww2.odu.edu/sci/cqfe/). The maximum life span of a Spot is about five years, although Spots over three years of age are uncommon. Spot mature between ages one and two and have lengths of seven to eight inches. We are especially interested in two classes of Spot. Age zero and one year old, which are sexually immature and usually smaller, and two years and older fish, which are sexually mature and usually longer. Analyzing these two groups is critical. As is often the case in exploited species, the larger size class are harvested first, which clearly influences the abundance of reproductively active individuals and therefore would have serious consequences for the population size of subsequent generations. Conversely, the presence of young Spot is important because it indicates Spots are

Table 5.5: Summary statistics for Spot classes.

Spot Class	Group Size	Min	Q_1	Median	Q_3	Max	μ	σ
0 and 1 year	273	6.3	8.2	8.9	9.7	12	9.01	1.15
2 years and older	130	8.5	11.1	11.6	12.5	13.9	11.7	1.11

naturally reproducing and if removal of this size class can be postponed for one to two years the population productivity may increase greatly. Table 5.5 provides the summary statistics for these two classes for the underlying population.

Figures 5.1 and 5.2 indicate that a two-component mixture model appears to be appropriate for the length measurement of Spots in this data set since both components are reasonably close to a normal distribution. We then treat these 403 fish as a population and perform a simulation study with 3000 repetitions by generating PROS, SRS and RSS samples to illustrate the use of our proposed estimators. We generated PROS samples with set size $S = 12$, subset size $n = 3$, cycle size $N = 5$ and design

$$D = \{d_1, d_2, d_3\} = \{\{1, 2, 3, 4\}, \{5, 6, 7, 8\}, \{9, 10, 11, 12\}.\}$$

Since Spots are ranked according to their length, there is no misplacement error in PROS and RSS sample. But we still estimated the misplacement probabilities $\alpha_{h,h'}$ in PROS data to illustrate the use of the estimators. The misplacement probabilities are not estimated in RSS data. The PROS data is constructed by selecting a set of 12 Spots out of 403 for each fully measured observation $X_{(d_r)i}$, $r = 1, 2, 3$ and $i = 1, \ldots, 5$. These 12 Spots are ranked from smallest to largest and divided into three subsets d_1, d_2 and d_3. One of the fish from subset d_r in set i is selected at random for full measurement to create a balanced PROS sample. We also constructed a

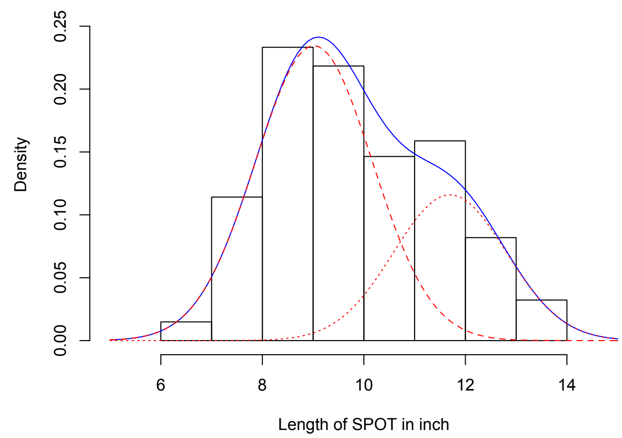

Figure 5.1: The length distribution of Spots in a population of size 403 as a mixture of two components. Solid, dashed and dotted lines are for the mixture, zero and one group and two years and older group densities, respectively.

ranked set sample from the same population. The ranked set sample is constructed in the same way as in the PROS design, but with parameters $N = 5$, $S = 3$ and $D = \{d_1, d_2, d_3\} = \{\{1\}, \{2\}, \{3\}\}$.

Table 5.6 presents the bias and efficiency of the estimates of the two-component mixture model parameters based on PROS and RSS samples. The column labeled precision provides the percentage of correct classification of Spots into one of the components of the mixture model based on the posterior probabilities developed in Section 5.5.

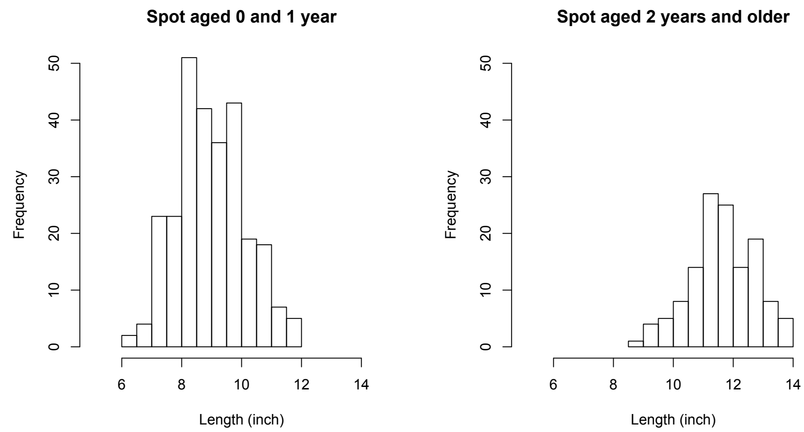

Figure 5.2: The length frequency distributions of Spots of (a) age zero and one and (b) two years and older. The total population size is 403.

Table 5.6: Biases and REs of estimators based on PROS and RSS designs with respect to SRS estimators for $\mathbf{\Psi} = (\pi, \mu_1, \mu_2, \sigma)$ when $N = 5$, $n = 3$, and $S = 12$. The last column shows the classification precision associated with each design.

EM-Tech.	Design		π	μ_1	μ_2	σ	Precision
	SRS	Bias	-0.0751	-0.2566	-0.0577	-0.2164	83.49%
Complete	PROS	Bias	-0.0572	-0.2006	-0.0897	-0.1514	94.38%
		RE	2.4444	2.3546	2.2693	1.8366	
	RSS	Bias	-0.0614	-0.2164	-0.0611	-0.1744	91.57%
		RE	1.6046	1.6688	1.6991	1.4925	
Modified	PROS	Bias	-0.0572	-0.2046	-0.0593	-0.1570	93.46%
		RE	2.3162	2.2725	2.1875	1.7415	
	RSS	Bias	-0.0544	-0.2131	-0.0234	-0.1701	90.58%
		RE	1.5175	1.5281	1.5921	1.4622	

Chapter 6

Summary and Future Work

In this thesis, we developed statistical inference for finite mixture models (FMMs) using rank-based sampling designs. The work is aimed at getting more insight into finite mixture modelling using rank-based sampling techniques by developing new, more suitable and efficient statistical methods. Due to advances in computational techniques, FMMs provide flexible and convenient statistical tools in the data analysis and play important roles in not only mainstream statistical analysis (e.g., modelling unknown distributional shapes, analyzing data with group-structures), but also in various scientific disciplines such as quantitative genetics, medical studies, engineering and so on. Ranked set sampling (RSS) is a powerful and cost-effective sampling design which is used to obtain more representative samples from the underlying population when the exact measurement of the variable of interest is costly or difficult to obtain, but a small number of sampling units can be ranked via auxiliary variables or judgment ranking, without actual measurement and at little cost. There are many situations where it is even difficult to rank all of the sampling units in a set with high confidence. Forcing rankers to declare unique ranks (as in RSS) can lead to substantial ranking error and, consequently, to poor statistical inference. The partially rank-ordered set (PROS) sampling design, a generalization of RSS, aiming to reduce ranking errors as well as the burden of ranking was studied. In standard methods of inference for FMMs, samples are typically drawn from the population using simple random sampling (SRS). It is important to develop statistical inference for FMMs using non-standard sampling designs used often in practice. To this end, we use rank-based sampling designs to take samples from population whose distribution can be represented by a FMM to tackle the existing problems and also to draw better inference about the underlying population. In fishery studies, for example, interest lies in estimating (or predicting) the population age structure or describing the length distribution of an age class of fish, especially in the case of short-lived species. These kinds of studies are usually time consuming and costly. Rank-based sampling designs are more suitable and efficient sampling designs for these situations. To reduce the cost, the length of the fish caught, which can be modelled a FMM, can be measured in the field. The age of a retained subsample can later be determined by more time consuming methods in the lab. This example can be treated as a problem of inference based on rank-based samples from FMMs.

In Chapter 2, we studied the analysis of FMMs based on different variations of RSS designs under perfect ranking assumption. RSS from an FMM results in two different approaches including prospective RSS, called M1-RSS, and separate RSS, called M2-RSS. The M1-RSS samples are obtained from the whole FMM so that, within each set, individuals from different components could possibly be involved in the ranking process. Under the M2-RSS design, it is assumed that RSS is performed within each component of the FMM separately and individuals in each set are obtained from one and only one component of the model. Proposing new EM-algorithm for each RSS approach, we studied the estimation and classification problems of the FMM under RSS designs. The developed methods are valid under balanced RSS designs. We conjecture that using a suitable unbalanced RSS can lead to better estimation of the parameters of the FMM over the balanced RSS. To this end, we considered the problem of estimating the mixing proportion of a mixture of two normal densities studied in Section 2.4.1 using unbalanced RSS samples consisting of (1) only minimums, (2) only maximums, and (3) both minimums and maximums. The results for a small simulation study are presented in Table 6.1. The performance of these unbalanced RSS designs compared with their SRS counterparts are calculated using the ratio of the MSEs of the maximum likelihood (ML) estimators of π based on unbalanced RSS and SRS samples of size 120. We also investigated the relative efficiency as a function of the set size $n \in \{2, 3, 4, 5\}$. A comparison of the results in Table 6.1 with Tables 2.3 and 2.4 shows that the performance of the proposed unbalanced RSS design could be better than that of the balanced RSS. In this example, selecting only maximums seems to be better than using an unbalanced RSS consisting of only minimums, or even both minimums and maximums. Note that when $d = \mu_2 - \mu_1$ is large, the SRS design performs better than an unbalanced M1-RSS consisting of only minimums. This simulation study shows that finding an optimal unbalanced RSS design for making inference about FMMs requires more investigation. A comprehensive study in this area is essential to study the statistical inference about the unknown parameters of FMMs based on unbalanced RSS as well as the effect of imperfect ranking.

In Chapter 3, we considered three different RSS data structures corresponding

Table 6.1: The REs of the ML estimates of π using unbalanced M1-RSS and M2-RSS samples consisting of the minimums, the maximums and both minimums and maximums for model (2.45) when $\pi = 0.8$, $\mu_1 = -1$, $\mu_2 = \mu_1 + d$ and $\sigma = 1$.

d	design	RSS	$n = 2$	$n = 3$	$n = 4$	$n = 5$
1	Min	M1	0.980	1.079	1.219	1.282
		M2	1.250	1.704	1.704	1.704
	Max	M1	1.562	2.027	2.586	2.830
		M2	1.162	1.339	1.648	1.829
	Both	M1	1.327	1.851	1.578	1.764
		M2	1.200	1.648	1.875	1.898
3	Min	M1	0.666	0.615	0.524	0.500
		M2	1.032	1.333	1.230	1.142
	Max	M1	1.777	1.777	2.666	4.000
		M2	1.032	1.066	1.333	1.103
	Both	M1	1.142	1.523	1.882	2.461
		M2	1.142	1.103	1.454	1.230

to different RSS approaches from FMMs under both perfect and imperfect ranking. These data structures include two complete M1-RSS, M2-RSS data (as explored in Chapter 2) as well as an incomplete M0-RSS data containing only the observed values of the underlying FMM. Deriving Fisher information (FI) matrices under these data structures, we showed that the FI contained in each variation of the RSS data is larger than the FI contained in their SRS counterparts. The missing information principle was also utilized as a useful tool to quantify how much information one may lose through using these data structures for statistical inference. We here derived the FI of Type-M0 and Type-M2 imperfect RSS data about the parameters of FMMs and compared them with their counterparts under SRS. Due to the complexity of the imperfect M1-RSS design, the problem of deriving the FI matrix based on imperfect Type-M1 RSS data has not considered and will require further investigation in the future. In addition, missing information principle under the imperfect ranking assumption (similar to Section 3.3) would be of interest.

We focused on the information content and uncertainty associated with PROS samples for a general class of distributions in Chapter 4. We compared the FI content of PROS samples with the FI content of SRS and RSS data of the same size under both perfect and imperfect sub-setting. We showed that PROS sampling designs generally result in more informative observations from the underlying population than SRS and RSS. Some examples were presented to show the amount of extra information provided by the PROS sampling design. We then studied other information and uncertainty measures such as Shannon entropy, Rényi entropy and the Kulback-Leibler (KL) distance. Similar results were obtained an assumption of perfect sub-setting. It would naturally be of interest to extend these results to imperfect sub-setting situations. The results of Chapter 4 suggest that one might be able to obtain more powerful tests for hypothesis testing or model selection problems based on PROS data. For example, it seems promising to develop goodness-of-fit tests based on a PROS sampling design and KL distance. We believe that further investigation of the PROS sampling design using the missing information criterion as in Chapter 3 is of interest and appealing as well.

In Chapter 5, we examined the problem of ML estimation of the parameters of FMMs based on PROS samples. We explored the use of the PROS sampling design for FMMs and a model was used to incorporate the subset misplacement error associated with the sampling into the estimation process. A suitable EM-algorithm was developed to obtain the ML estimates of the parameters. One major challenge to ML estimation based on PROS data from FMMs is its computational intractability. To overcome this difficulty, we presented a modified version of the commonly used EM-algorithm which reduces the computational complexity to the level of the usual EM-algorithm based on SRS data. Our numerical studies showed that the modified

EM-algorithm performs very well in estimating the parameters. We further studied the problem of classification of a PROS sample into the components of the underlying FMM. Simulation studies indicated a very good performance of the PROS data for estimating the parameters of the FMM in terms of bias and precision of the estimates. The success rate of classification under the PROS sampling design was substantially better than those of SRS and RSS designs. The proposed methods were applied to estimate the age group of fish in the Chesapeake Bay area using length-frequency data which can be considered as a mixture of two normal densities. The developed methodologies are valid under balanced PROS sampling (i.e., pre-specified number of subsets of the same size); however, future investigation for the analysis of FMMs under unbalanced PROS protocols is required.

In many settings, in order to reduce the impact of ranking errors and deficiencies of ranking operators, one can use rank-based sampling designs based on multi-observers. This not only overcomes the problems associated with the ranking error involved in the data, but also implies significant improvements in the inference on the underlying population. We would like to investigate the effect of using multi-observer approach on methodologies developed for the analysis of FMM under rank-based sampling designs.

As a long-term research project, we shall study the number of components involved in (or the order of) the mixture model. Testing for the order of a mixture model is one of the most important problems in mixture modelling and related fields (e.g., classification and clustering analysis) which has not been completely resolved. Many methods have been proposed to tackle the problem based on non-parametric and parametric approaches including penalized likelihood functions, likelihood ratio testing (LRT), bootstrapping LRTs and EM-algorithm tests; however there are still many drawbacks associated with the proposed methods. For instance, the components have to be well-separated in order to be detected, finding the distribution of test statistics is difficult, etc. As rank-based sampling designs provide extra information about different aspects of the population and lead to better estimation of the parameters of the model, it is expected that this extra information can be used to obtain better results in assessing the number of components of FMMs.

Bibliography

ARSLAN, G. AND OZTURK, O. 2013. Parametric inference based on partially rank ordered set samples. *Journal of The Indian Statistical Association* 51:1–24.

BARABESI, L. 2001. The unbalanced ranked-set sample sign test. *Journal of Non-parametric Statistics* 13:279–289.

BARABESI, L. AND EL-SHARAAWI, A. 2001. The efficiency of ranked set sampling for parameter estimation. *Statistics and Probability Letters* 53:189–199.

BARNETT, V. 1999. Ranked set sample design for environmental investigations. *Environmental and Ecological Statistics* 6:59–74.

BARNETT, V. AND MOORE, K. 1997. Best linear unbiased estimates in ranked-set sampling with particular reference to imperfect ordering. *Journal of Applied Statistics* 24:697–710.

BASFORD, K., GREENWAY, D., MCLACHLAN, G., AND PEEL, D. 1997. Standard errors of fitted component means of normal mixtures. *Computational Statistics* 12:1–18.

BEAMISH, R. 1987. Current trends in age determination methodology. *Age and growth of fish* pp. 15–42.

Bhattacharyya, G. K. 1985. The asymptotics of maximum likelihood and related estimators based on type ii censored data. *Journal of the American Statistical Association* 80:398–404.

Bohn, L. L. 1996. A review of nonparametric ranked-set sampling methodology. *Communications in Statistics–Theory and Methods* 25:2675–2685.

Chen, H. and Chen, J. 2003. Tests for homogeneity in normal mixtures in the presence of a structural parameter. *Statistica Sinica* 13:351–366.

Chen, J., Li, P., and Fu, Y. 2012. Inference on the order of a normal mixture. *Journal of the American Statistical Association* 107:1096–1105.

Chen, Z. 2000. The efficiency of ranked-set sampling relative to simple random sampling under multi-parameter families. *Statistica Sinica* 10:247–264.

Chen, Z., Bai, Z., and Sinha, B. 2004. Ranked set sampling: theory and applications, volume 176. Springer.

Chen, Z. and Wang, Y.-G. 2004. Efficient regression analysis with ranked-set sampling. *Biometrics* 60:997–1004.

Cobby, J., Ridout, M., Bassett, P., and Large, R. 1985. An investigation into the use of ranked set sampling on grass and grass-clover swards. *Grass and Forage Science* 40:257–263.

Dell, T. and Clutter, J. 1972. Ranked set sampling theory with order statistics background. *Biometrics* pp. 545–555.

Dempster, A. P., Laird, N. M., and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 1–38.

El Zaart, A., Ziou, D., Wang, S., and Jiang, Q. 2002. Segmentation of sar images. *Pattern Recognition* 35:713–724.

Finch, S. J., Mendell, N. R., and Thode Jr, H. C. 1989. Probabilistic measures of adequacy of a numerical search for a global maximum. *Journal of the American Statistical Association* 84:1020–1023.

Frey, J. 2012. Nonparametric mean estimation using partially ordered sets. *Environmental and Ecological Statistics* 19:309–326.

Furman, W. D. and Lindsay, B. G. 1994. Measuring the relative effectiveness of moment estimators as starting values in maximizing likelihoods. *Computational Statistics and Data Analysis* 17:493–507.

Gao, J. and Ozturk, O. 2012. Two sample distribution-free inference based on partially rank-ordered set samples. *Statistics and Probability Letters* 82:876–884.

Halls, L. K. and Dell, T. R. 1966. Trial of ranked-set sampling for forage yields. *Forest Science* 12:22–26.

Hatefi, A. and Jafari Jozani, M. 2013a. Fisher information in different types of perfect and imperfect ranked set samples from finite mixture models. *Journal of Multivariate Analysis* 119:16–31.

Hatefi, A. and Jafari Jozani, M. 2013b. Information content of partially rank ordered set samples. *Submitted* .

HATEFI, A. AND JAFARI JOZANI, M. 2014. Proportion estimation based on a pros samples with multiple concomitants in a breast cancer study. *Submitted* .

HATEFI, A., JAFARI JOZANI, M., AND OZTUK, O. 2013. Mixture model analysis of partially rank ordered set samples: Estimating the age-groups of fish from length-frequency data. *Submitted* .

HATEFI, A., JAFARI JOZANI, M., AND ZIOU, D. 2014. Estimation and classification for finite mixture models under ranked set sampling. *Statistica Sinica* 24:675–698.

HILL, B. M. 1963. Information for estimating the proportions in mixtures of exponential and normal distributions. *Journal of the American Statistical Association* 58:918–932.

HOSMER JR, D. W. 1973. A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. *Biometrics* pp. 761–770.

JAFARI JOZANI, M. AND AHMADI, J. 2014. On uncertainty and information properties of ranked set samples. *Information Sciences* 264:291–301.

JAFARI JOZANI, M. AND JOHNSON, B. C. 2011. Design based estimation for ranked set sampling in finite populations. *Environmental and Ecological Statistics* 18:663–685.

JAFARI JOZANI, M. AND PERRON, F. 2011. On the use of antithetic variables to improve over the ranked set sampling estimator of the population mean. *Sankhya A* 73:142–161.

Johnson, G., Patil, G., and Sinha, A. 1993. Ranked set sampling for vegetation research. *Abstracta Botanica* 17:87–102.

Johnson, O. 2004. Information Theory And The Central Limit Theorem. Imperial College Press.

Johnson, R. A., Mehrotra, K. G., et al. 1972. Locally most powerful rank tests for the two-sample problem with censored data. *The Annals of Mathematical Statistics* 43:823–831.

Karlis, D. and Xekalaki, E. 2003. Choosing initial values for the em-algorithm for finite mixtures. *Computational Statistics and Data Analysis* 41:577–590.

Kim, Y. and Arnold, B. C. 1999. Parameter estimation under generalized ranked set sampling. *Statistics and Probability Letters* 42:353–360.

Kumar, K. D. and Adams, S. M. 1977. Estimation of age structure of fish populations from length-frequency data. Technical report, Oak Ridge National Lab., TN (USA).

Lavin, M. 1999. The 'bayesics' of ranked set sampling. *Environmental and Ecological Statistics* 6:47–57.

Liu, C., Zhang, L., Davis, C. J., Solomon, D. S., and Gove, J. H. 2002. A finite mixture model for characterizing the diameter distributions of mixed-species forest stands. *Forest Science* 48:653–661.

MacDonald, P. and Pitcher, T. 1979. Age-groups from size-frequency data: a versatile and efficient method of analyzing distribution mixtures. *Journal of the Fisheries Board of Canada* 36:987–1001.

McIntyre, G. 1952. A method for unbiased selective sampling, using ranked sets. *Crop and Pasture Science* 3:385–390.

McLachlan, G., Do, K.-A., and Ambroise, C. 2005. Analyzing microarray gene expression data, volume 422. John Wiley and Sons.

McLachlan, G. and Krishnan, T. 2007. The EM algorithm and extensions, volume 382. John Wiley and Sons.

McLachlan, G. and Peel, D. 2004. Finite mixture models. Wiley. com.

Mehrotra, K. and Nanda, P. 1974. Unbiased estimation of parameters by order statistics in the case of censored samples. *Biometrika* 61:601–606.

Mengersen, K., Robert, C., and Titterington, M. 2011. Mixtures: estimation and applications, volume 896. John Wiley and Sons.

Mode, N. A., Conquest, L. L., and Marker, D. A. 1999. Ranked set sampling for ecological research: accounting for the total costs of sampling. *Environmetrics* 10:179–194.

Muttlak, H. and McDonald, L. 1992. Ranked set sampling and the line intercept method: A more efficient procedure. *Biometrical Journal* 34:329–346.

Muttlak, H. A. and Mc Donald, L. L. 1990. Ranked set sampling with respect to concomitant variables and with size biased probability of selection. *Communications in Statistics-Theory and Methods* 19:205–219.

Nazari, S., Jafari Jozani, M., and Kharrati-Kopaei, M. 2014. Nonparametric density estimation using partially rank-ordered set samples with application in estimating the distribution of wheat yield. *Electronic Journal of Statistics* 8:738–761.

Ogle, D. 2013. Fisheries stock assessment (fsa) methods package for r. *R package version 0.4.13* .

Orchard, T. and Woodbury, M. A. 1972. A missing information principle: Theory and applications. *In* Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability, volume 1, pp. 697–715. McGraw-Hill.

Ozturk, O. 1999. Two-sample inference based on one-sample ranked set sample sign statistics. *Journal of Nonparametric Statistics* 10:197–212.

Ozturk, O. 2011. Sampling from partially rank-ordered sets. *Environmental and Ecological Statistics* 18:757–779.

Ozturk, O. 2012. Quantile inference based on partially rank-ordered set samples. *Journal of Statistical Planning and Inference* 142:2116–2127.

Ozturk, O., Bilgin, O. C., and Wolfe, D. A. 2005. Estimation of population mean and variance in flock management: a ranked set sampling approach in a finite population setting. *Journal of Statistical Computation and Simulation* 75:905–919.

Ozturk, O. and Jafari Jozani, M. 2014. Inclusion probabilities in partially rank ordered set sampling. *Computational Statistics and Data Analysis* 69:122–132.

Ozturk, O. and Wolfe, D. A. 2000. Alternative ranked set sampling protocols for the sign test. *Statistics and probability letters* 47:15–23.

Patil, G. 2006. Ranked set sampling. *Encyclopedia of environmetrics* .

Patil, G., Gore, S., and Sinha, A. 1994. Environmental chemistry, statistical modeling, and observational economy. Lewis Publishers, Boca Raton.

Presnell, B. and Bohn, L. L. 1999. U-statistics and imperfect ranking in ranked set sampling. *Journal of Nonparametric Statistics* 10:111–126.

Rickabaugh, H. and Capossela, K. 2011. Evaluation of the Status of Spot in Maryland. Maryland DNR Fisheries Service doc.

Roeder, K. 1994. A graphical technique for determining the number of components in a mixture of normals. *Journal of the American Statistical Association* 89:487–495.

Schlattmann, P. 2009. Medical applications of finite mixture models. Springer.

Schork, N. J., Allison, D. B., and Thiel, B. 1996. Mixture distributions in human genetics research. *Statistical Methods in Medical Research* 5:155–178.

Sinha, B. K., Sinha, B. K., and Purkayastha, S. 1996. On some aspects of ranked set sampling for estimation of normal and exponential parameters. *Statistics and Decisions-International Journal for Stochastic Methods and Models* 14:223–240.

Stokes, L. 1995. Parametric ranked set sampling. *Annals of the Institute of Statistical Mathematics* 47:465–482.

Summerfelt, R. C. and Hall, G. E. 1987. Age and Growth of Fish. Ames: Iowa State University Press.

Takahasi, K. and Wakimoto, K. 1968. On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics* 20:1–31.

Tarn, C. Y., Yu, P. L., and Fung, T. W. 1998. Sensitivity analysis of blue for the population mean based on a ranked set sample. *Communications in Statistics-Simulation and Computation* 27:1075–1091.

Titterington, D. M., Smith, A. F., Makov, U. E., et al. 1985. Statistical analysis of finite mixture distributions, volume 7. Wiley New York.

Wang, Y.-G., Ye, Y., and Milton, D. A. 2009. Efficient designs for sampling and subsampling in fisheries research based on ranked sets. *ICES Journal of Marine Science: Journal du Conseil* 66:928–934.

Wolfe, D. A. 2012. Ranked set sampling: Its relevance and impact on statistical inference. *ISRN Probability and Statistics* 2012:1–32.