Statistica Sinica **24** (2014), 675-698
doi:http://dx.doi.org/10.5705/ss.2012.178

ESTIMATION AND CLASSIFICATION FOR FINITE MIXTURE MODELS UNDER RANKED SET SAMPLING

Armin Hatefi[1], Mohammad Jafari Jozani[1] and Djemel Ziou[2]

[1] *University of Manitoba and* [2] *Université de Sherbrooke*

Abstract: We consider maximum likelihood estimation of the parameters of a finite mixture model for independent order statistics data arising from ranked set sampling, as well as classification of the observed data. We propose two ranked-based sampling designs from a finite mixture density and explain how to estimate the unknown parameters of the model for each design. To exploit the special structure of the ranked set sampling, we develop a new expectation-maximization algorithm that turns out to be different from its counterpart with simple random sample data. Our findings are that estimators based on ranked set sampling are more efficient than their counterparts based on the commonly used simple random sampling technique. Theoretical results are augmented with simulation studies.

key words and phrases: Classification, complete-data likelihood, expectation maximization algorithm, finite mixture models, order statistics, ranked set samples.

1. Introduction

In many sampling situations, a small number of sampling units can be fairly accurately ordered with respect to a variable of interest without actual measurements on them and at little cost. On the other hand, exact measurements of these units may be very expensive. For example, for environmental risks such as radiation (soil contamination and disease clusters) or pollution (water contamination and root disease of crops), exact measurements involve substantial scientific processing of materials and consequently high cost, while the variable of interest from a small number of experimental (sampling) units may easily be ranked. Ranked set sampling (RSS), as proposed by McIntyre (1952) in estimating the mean of pasture yields, provides an interesting alternative to simple random sampling (SRS) in these situations. Compared to SRS, RSS has been proven theoretically by Takahasi and Wakimoto (1986) and shown empirically (see Kaur et al. (1995) and Chen, Bai, and Sinha (2004) for more details) to yield a more precise estimator of the population mean. Wang, Ye and Milton (2009) explained how RSS can be used as an efficient sampling design in fisheries surveys where interest lies in either estimating the population age structure of fish or describing the length distribution of an age or sex class of fish, especially in the case of short-lived species. These kind of studies are usually time consuming and costly. RSS could be then considered as a more efficient sampling design; for example, samples of the catch could be measured and weighed in the field and only a subsample retained and later aged, with the sex of the fish determined by more time consuming methods, to reduce the cost. The goal is to estimate the proportion of males and females and the mean and variance of length of both males and females for each age class. We can state this problem in terms of inference based on ranked set sample data from a finite mixture model.

In the standard methods of modelling and inference for finite mixture models, it is assumed that the samples are drawn from the underlying population using SRS technique, McLachlan and Peel (2004), and there are only a few results available when the sampling design is different, Wedel, Ter Hofstede and Steenkamp (1998). However in some applications, such as the one explained in Wang, Ye and Milton (2009), using RSS may be cheaper and result in better and more informative samples from the underlying population. In this paper, we study the problem of maximum likelihood estimation of unknown parameters for finite mixture models based on two variations of RSS techniques and compare the results with the corresponding ones with simple random sample data. We show that using RSS leads to a better inference about the unknown features of the underlying model, such as the estimation of the unknown parameters and identification of unobserved classes and their weights. We also explore, both theoretically and numerically, the problem of classification of the RSS data and show how the extra information via the rank of each observation in RSS leads to a more efficient classification of the data compared with the usual SRS data.

To this end, in Section 2 we explain RSS techniques and propose two ranked-based sampling designs from a finite mixture model. We also discuss the differences between RSS and SRS techniques. Section 3 deals with the problem of maximum likelihood estimation of unknown parameters of the finite mixture model (FMM) using RSS techniques. We also point out the link between inference based on the RSS and SRS techniques. Suitable EM algorithms are developed in Sections 4. We show that the underlying theory behind the EM algorithm for RSS data is different from its counterpart under SRS data. In Section 5, we consider the problem of classification of an RSS sample of size n and compare it with the usual one based on SRS techniques. Section 6 is devoted to the study of the performance of ML estimators of parameters of finite mixture of normal distributions via extensive simulation studies.

2. RSS from a Finite Mixture Model

Suppose X is a random variable with distribution a mixture of M densities in some unknown proportions $\pi = (\pi_1, \ldots, \pi_M)$, with $\pi_j > 0$, and $\sum_{j=1}^{M} \pi_j = 1$.

The probability density function (pdf) of X is written as

$$f(x_i; \mathbf{\Psi}) = \pi_1 f_1(x_i; \theta_1) + \cdots + \pi_M f_M(x_i; \theta_M), \tag{2.1}$$

where $f_j(\cdot; \theta_j)$, $j = 1, \ldots, M$, refers to the pdf of the jth component of the model that is specified up to a vector θ_i of unknown parameters, known a priori to be distinct. Thus the data come from M different classes, $C_1, \ldots, C_M$, and $f_j(\cdot; \theta_j)$ is the pdf of the variable of interest in the jth class. The vector of all unknown parameters is denoted by $\mathbf{\Psi} = (\pi_1, \ldots, \pi_{M-1}, \xi)^\top$, where $\xi = (\theta_1^\top, \ldots, \theta_M^\top)^\top$ and the superscript $\top$ refers to vector transpose.

To obtain a ranked set sample of size $n = mk$, we proceed as follows. If $X_{11}, \ldots, X_{k1}$ is a simple random sample of size k taken from the population then, using a ranking operator $O_r(\cdot)$, it can be ranked as $O_r(X_{11}, \ldots, X_{k1}) = (X_{(1)1}, \ldots, X_{(k)1})$; RSS uses only the lowest observation, $X_{(1)1}$, from this sample (called hereafter a *set* of size k). We take another independent SRS sample of size k from the population and the second lowest observation, namely $X_{(2)1}$, is chosen from this set. Finally, for the kth independent SRS sample of size k, the largest observation, $X_{(k)1}$, is chosen from this set. This whole process is referred to as a *cycle*; it is repeated m times resulting in a ranked set sample of size mk, $\mathbf{X}_{RSS} = \{X_{(r)i}, r = 1, \ldots, k; i = 1, \ldots, m\}$. The RSS technique uses inherent heterogeneity among the sampling units through a ranking process to create artificial strata, it can be considered as an stratification of the units during the sampling process based on their ranks in a sample. Although RSS requires identification of mk^2 units from the population only mk of them are actually measured. In RSS, information and structure is provided through the ranking process: $X_{(1)i}, \ldots, X_{(k)i}$ are independent order statistics and each provides information about some aspect of the distribution. It is this extra structure that makes procedures based on RSS more efficient than their counterparts based on SRS data with the same number of measurements.

In this paper, we assume that ranking is done perfectly; RSS with imperfect ranking and its comparison with perfect RSS for mixture models will be discussed in future work. We consider two variants of RSS designs, denoted by M1-RSS and M2-RSS. In the M1-RSS design, the ranked set samples are obtained from the whole model so that, within each set, individuals from different components can possibly be involved in the ranking process. This design is more practical and enables one to make better inference on classification of observations and estimation of mixing proportions. It is useful in such applied fields as fisheries research, health related studies, economics, and environmental and ecological studies. For example, in the fisheries example in Section 1, to model the age distribution of a specific type of fish (e.g., halibut) by a mixture of two normal densities, we first note that there are two subpopulations consisting of male and female halibut. One can easily execute an M1-RSS design to obtain a ranked set sample from this population. To this end, subsamples of halibut are obtained and the rth shortest halibut ($r = 1, \ldots, k$) is retained and the age and the sex of the selected halibut is determined (using some time consuming methods) in the lab. Here the subsamples consist of both male and female halibut.

In 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. In addition, we assume that the component to which each measurement belongs is unknown. One advantage associated with this kind of separate sampling is that it is appropriate for the retrospective studies often needed in epidemiological investigations. They also allow better inference about all components of the underlying population, even components that are rarely observed (McLachlan and Peel (2004)). The M2-RSS design can also be used when the component membership of each data point is missing for some reason such as confidentiality, or simply because it is not recorded. For example, statistical agencies could perform RSS separately within different subpopulations, such as males and females or minority groups, but where membership is unknown (perhaps due to confidentiality or sensitivity of the questions). Here the observation $X_{(r)i}$ is known to be obtained from a set consisting of units from one of the M subpopulations, but it is not known which subpopulation.

3. Likelihood Functions for RSS Data

In this section, we study the problem of maximum likelihood (ML) estimation of unknown parameters $\mathbf{\Psi}$ at (2.1) based on RSS data. We limit ourselves to the case where the representation (2.1) is identifiable.

If $F(x; \mathbf{\Psi}) = \sum_{j=1}^{M} \pi_j F_j(x; \theta_j)$ is the cumulative distribution function (cdf) at (2.1), the pdf of $X_{(r)}$, the rth order statistic of a sample of size k from (2.1), is

$$f^{(r)}(x; \mathbf{\Psi}) = k \binom{k-1}{r-1} f(x; \mathbf{\Psi}) \{F(x; \mathbf{\Psi})\}^{r-1} \{\bar{F}(x; \mathbf{\Psi})\}^{k-r}, \tag{3.1}$$

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

$$f(x; \mathbf{\Psi}) = \frac{1}{k} \sum_{r=1}^{k} f^{(r)}(x; \mathbf{\Psi}). \tag{3.2}$$

Similarly, for each component of (2.1),

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

and so (2.1) can be written in terms of the pdf of the order statistics of each component, $f_j^{(r)}(\cdot;\theta_j)$, as

$$f(x;\mathbf{\Psi}) = \frac{1}{k}\sum_{r=1}^{k}\sum_{j=1}^{M}\pi_j f_j^{(r)}(x;\theta_j). \tag{3.3}$$

3.1. Likelihood function for M1-RSS data

Suppose $\mathbf{X}_{M1,RSS} = \{x_{(r)i}, r = 1,\ldots,k; i = 1,\ldots,m\}$ is an M1-RSS data of size mk from (2.1), where k is the set size and m is the cycle number. Since for each i, $x_{(r)i}$ corresponds to the rth order statistic of a sample of size k from (2.1), using (3.1), the likelihood function of $\mathbf{\Psi}$ for M1-RSS data is

$$L_{M1,RSS}(\mathbf{\Psi}) = \prod_{i=1}^{m}\prod_{r=1}^{k} f^{(r)}(x_{(r)i};\mathbf{\Psi}). \tag{3.4}$$

If $l_{M1,RSS}(\mathbf{\Psi}) = \log L_{M1,RSS}(\mathbf{\Psi})$, the ML estimate of $\mathbf{\Psi}$, denoted by $\hat{\mathbf{\Psi}}_{M1,RSS}$, is an appropriate root of the likelihood equation,

$$\frac{\partial l_{M1,RSS}(\mathbf{\Psi})}{\partial \mathbf{\Psi}} = 0. \tag{3.5}$$

Remark 1. Suppose $\mathbf{X}_{SRS} = \{x_1,\ldots,x_{mk}\}$ is a simple random sample of size mk from (2.1). One can represent $\mathbf{X}_{SRS}$ in the form of the matrix $\mathcal{D} = [X_{(r)i}]$ of size $k \times m$. Here $X_{(r)i}$ is the (r,i)th element of $\mathcal{D}$ and the likelihood function of the SRS data is

$$L_{SRS}(\mathbf{\Psi}) = \prod_{i=1}^{m}\prod_{r=1}^{k} f(x_{(r)i};\mathbf{\Psi}) = \prod_{i=1}^{m}\prod_{r=1}^{k}\left\{\sum_{j=1}^{M}\pi_j f_j(x_{(r)i};\theta_j)\right\}.$$

Using (3.4) we have

$$L_{M1,RSS}(\mathbf{\Psi}) = L_{SRS}(\mathbf{\Psi}) \times \left\{\prod_{i=1}^{m}\prod_{r=1}^{k} k\binom{k-1}{r-1}[F(x_{(r)i};\mathbf{\Psi})]^{r-1}[\bar{F}(x_{(r)i};\mathbf{\Psi})]^{k-r}\right\}. \tag{3.6}$$

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

3.2. Likelihood function for M2-RSS data

Suppose $\mathbf{X}_{M2,RSS} = \{x_{(r)i}, r = 1,\ldots,k; i = 1,\ldots,m\}$ is a sample of size mk from (2.1) obtained through M2-RSS design, k and m defined as before. For

M2-RSS design, $x_{(r)i}$ corresponds to the rth order statistic of a sample of size k from one of the components of (2.1), and the likelihood function of $\mathbf{\Psi}$ based on M2-RSS data is

$$L_{M2,RSS}(\mathbf{\Psi}) = \prod_{i=1}^{m} \prod_{r=1}^{k} \Big\{ \sum_{j=1}^{M} \pi_j f_j^{(r)}(x_{(r)i}; \theta_j) \Big\}, \tag{3.7}$$

where $f_j^{(r)}(\cdot; \mathbf{\Psi})$ corresponds to the pdf of rth 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 an appropriate root of the likelihood equation

$$\frac{\partial l_{M2,RSS}(\mathbf{\Psi})}{\partial \mathbf{\Psi}} = 0. \tag{3.8}$$

In Section 4, we develop new expectation-maximization (EM) algorithms to obtain the solutions of (3.5) and (3.8) corresponding to local maximizers of the likelihood functions (3.4) and (3.7), respectively.

4. EM Algorithms Based on Ranked Set Samples

The EM algorithm is a general approach that can be used for the ML estimation of the parameters at (2.1), see Dempster, Laird, and Rubin (1977). RSS data have a unique structure and, as a result, the standard EM algorithm is not applicable in its standard form. In this section, we develop new EM-algorithms for RSS data.

4.1. EM algorithm for M1-RSS data

To use the EM algorithm for estimating the parameters of the mixture density $f(x_{(r)i}; \mathbf{\Psi})$ based on the M1-RSS data $X_{M1,RSS} = \{x_{(r)i}, r = 1, \ldots, k; i = 1, \ldots, m\}$, the problem is viewed as incomplete since the label-component vectors associated with the feature variables $x_{(r)}$ are missing. What makes this problem non-standard is the presence of the terms $\{F(x_{(r)i}; \mathbf{\Psi})\}^{r-1}$ and $\{\bar{F}(x_{(r)i}, \mathbf{\Psi})\}^{k-r}$ in (3.1).

We propose a missing mechanism that introduces three latent vectors for each $x_{(r)i}$. Let $\mathbf{Z}_i^{(r)} = \big(Z_{i1}^{(r)}, \ldots, Z_{iM}^{(r)}\big)$ denote the component membership of the observation $x_{(r)i}$. So, for $j = 1, \ldots, M$,

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

with $\sum_{j=1}^{M} Z_{ij}^{(r)} = 1$. Thus $\mathbf{Z}_1^{(1)}, \ldots, \mathbf{Z}_m^{(k)} \overset{i.i.d.}{\sim} Mult(1, \pi)$, with

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

Let $\mathbf{W}_i^{(r)} = \big(W_{i1}^{(r)}, \ldots, W_{iM}^{(r)}\big)$, where $W_{ij}^{(r)}$ is the number of observations less than $x_{(r)i}$ that are selected from the component j, $\sum_{j=1}^M W_{ij}^{(r)} = r-1$. Thus $\mathbf{W}_1^{(1)}, \ldots, \mathbf{W}_m^{(k)} \overset{i.i.d.}{\sim} Mult(r-1, \pi)$, with

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

Likewise, let $\mathbf{V}_i^{(r)} = \big(V_{i1}^{(r)}, \ldots, V_{iM}^{(r)}\big)$, where $V_{ij}^{(r)}$ is the number of observations larger than $x_{(r)i}$ that are selected from the component j, $\sum_{j=1}^M V_{ij}^{(r)} = k-r$. Here $\mathbf{V}_1^{(1)}, \ldots, \mathbf{V}_m^{(k)} \overset{i.i.d.}{\sim} Mult(k-r, \pi)$, with

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

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 independent.

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

$$\begin{aligned} f(x_{(r)i}, z_i^{(r)}, w_i^{(r)}, v_i^{(r)}; \mathbf{\Psi}) &= k\binom{k-1}{r-1}\binom{r-1}{w_{i1}^{(r)}, \ldots, w_{iM}^{(r)}}\binom{k-r}{v_{i1}^{(r)}, \ldots, v_{iM}^{(r)}} \\ &\times \prod_{j=1}^M \pi_j^{\{z_{ij}^{(r)}+w_{ij}^{(r)}+v_{ij}^{(r)}\}} \{f_j(x_{(r)i}, \theta_j)\}^{z_{ij}^{(r)}} \{F_j(x_{(r)i}, \theta_j)\}^{w_{ij}^{(r)}} \{\bar{F}_j(x_{(r)i}, \theta_j)\}^{v_{ij}^{(r)}}. \end{aligned}$$

Proof. The proof is in Section A.1.

Lemma 2. *For each $x_{(r)i}$, $i = 1, \ldots, m; r = 1, \ldots, k$,*

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

Proof. The proof is in Section A.2.

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

$$L_c(\mathbf{\Psi}|\mathbf{y}_{M1}) = \prod_{i=1}^m \prod_{r=1}^k f(x_{(r)i}, z_i^{(r)}, w_i^{(r)}, v_i^{(r)}; \mathbf{\Psi})$$

$$= k\binom{k-1}{r-1}\prod_{i=1}^{m}\prod_{r=1}^{k}\binom{r-1}{w_{i1}^{(r)},\ldots,w_{iM}^{(r)}}\binom{k-r}{v_{i1}^{(r)},\ldots,v_{iM}^{(r)}}$$
$$\times\prod_{j=1}^{M}\pi_j^{\{z_{ij}^{(r)}+w_{ij}^{(r)}+v_{ij}^{(r)}\}}\{f_j(x_{(r)i},\theta_j)\}^{z_{ij}^{(r)}}\{F_j(x_{(r)i},\theta_j)\}^{w_{ij}^{(r)}}\{\bar{F}_j(x_{(r)i},\theta_j)\}^{v_{ij}^{(r)}}. \quad (4.1)$$

Using (4.1) and Lemma 2, the incomplete-data likelihood function $L_{M1,RSS}(\mathbf{\Psi})$ can be obtained by summing over $\mathbf{Z},\mathbf{V}$ and $\mathbf{W}$ of the complete-data likelihood,

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

The complete-data log-likelihood function of $\mathbf{\Psi}$ is

$$\begin{aligned} l_{C,M1,RSS}(\mathbf{\Psi}) = & \sum_{i=1}^{m}\sum_{r=1}^{k}\log\left\{k\binom{k-1}{r-1}\binom{r-1}{w_{i1}^{(r)},\ldots,w_{iM}^{(r)}}\binom{k-r}{v_{i1}^{(r)},\ldots,v_{iM}^{(r)}}\right\} \\ & +\sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M}Z_{ij}^{(r)}\{\log\pi_j+\log(f_j(x_{(r)i};\theta_j))\} \\ & +W_{ij}^{(r)}\{\log\pi_j+\log(F_j(x_{(r)i};\theta_j))\} \\ & +\sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M}V_{ij}^{(r)}\{\log\pi_j+\log(\bar{F}_j(x_{(r)i};\theta_j))\}. \end{aligned} \quad (4.2)$$

We 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 take

$$Q_{M1}(\mathbf{\Psi},\mathbf{\Psi}^{(0)}) = E_{\mathbf{\Psi}^{(0)}}[l_{C,M1,RSS}(\mathbf{\Psi})|\mathbf{y}_{M1}], \quad (4.3)$$

with the expectation computed using $\mathbf{\Psi}^{(0)}$ instead of $\mathbf{\Psi}$ in the conditional distribution. At the $(p+1)$th iteration, the calculation of $Q_{M1}(\mathbf{\Psi},\mathbf{\Psi}^{(p)})$, where $\mathbf{\Psi}^{(p)}$ is the value of $\mathbf{\Psi}$ after the pth iteration, involves the expectations of $Z_{ij}^{(r)}$, $W_{ij}^{(r)}$ and $V_{ij}^{(r)}$ given the observation $x_{(r)i}$. Since

$$Z_{ij}^{(r)}|X_{(r)i}=x_{(r)i}\sim Bin\left(1,\frac{\pi_j f_j(x_{(r)i};\theta_j)}{f(x_{(r)i};\mathbf{\Psi})}\right),$$
$$W_{ij}^{(r)}|X_{(r)i}=x_{(r)i}\sim Bin\left(r-1,\frac{\pi_j F_j(x_{(r)i};\theta_j)}{F(x_{(r)i};\mathbf{\Psi})}\right),$$
$$V_{ij}^{(r)}|X_{(r)i}=x_{(r)i}\sim Bin\left(k-r,\frac{\pi_j \bar{F}_j(x_{(r)i};\theta_j)}{\bar{F}(x_{(r)i};\mathbf{\Psi})}\right),$$

where $i = 1, \ldots, m$; $r = 1, \ldots, k$ and $j = 1, \ldots, M$, one has

$$\tau_{j,M1,RSS}(x_{(r)i}; \mathbf{\Psi}^{(p)}) = E_{\mathbf{\Psi}^{(p)}}[Z_{ij}^{(r)}|x_{(r)i}] = \frac{\pi_j^{(p)} f_j(x_{(r)i}; \theta_j^{(p)})}{\sum_{h=1}^{M} \pi_h^{(p)} f_h(x_{(r)i}; \theta_h^{(p)})}, \tag{4.4}$$

$$\beta_{j,M1,RSS}(x_{(r)i}; \mathbf{\Psi}^{(p)}) = E_{\mathbf{\Psi}^{(p)}}[W_{ij}^{(r)}|x_{(r)i}] = \frac{(r-1)\pi_j^{(p)} F_j(x_{(r)i}; \theta_j^{(p)})}{\sum_{h=1}^{M} \pi_h^{(p)} F_h(x_{(r)i}; \theta_h^{(p)})}, \tag{4.5}$$

$$\gamma_{j,M1,RSS}(x_{(r)i}; \mathbf{\Psi}^{(p)}) = E_{\mathbf{\Psi}^{(p)}}[V_{ij}^{(r)}|x_{(r)i}] = \frac{(k-r)\pi_j^{(p)} \bar{F}_j(x_{(r)i}; \theta_j^{(p)})}{\sum_{h=1}^{M} \pi_h^{(p)} \bar{F}_h(x_{(r)i}; \theta_h^{(p)})}. \tag{4.6}$$

Using (4.4), (4.5) and (4.6), we have

$$\begin{aligned} Q_{M1}&(\mathbf{\Psi}, \mathbf{\Psi}^{(p)}) \\ &= cst + \sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M} \tau_{j,M1,RSS}(x_{(r)i}; \mathbf{\Psi}^{(p)})\{\log \pi_j + \log(f_j(x_{(r)i}; \theta_j))\} \\ &\quad + \sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M} \beta_{j,M1,RSS}(x_{(r)i}; \mathbf{\Psi}^{(p)})\{\log \pi_j + \log(F_j(x_{(r)i}; \theta_j))\} \\ &\quad + \sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M} \gamma_{j,M1,RSS}(x_{(r)i}; \mathbf{\Psi}^{(p)})\{\log \pi_j + \log(\bar{F}_j(x_{(r)i}; \theta_j))\}. \end{aligned} \tag{4.7}$$

M-Step: In this step, the maximization of $Q_{M1}(\mathbf{\Psi}, \mathbf{\Psi}^{(p)})$ with respect to $\mathbf{\Psi}$ is done over the parameter space to obtain the updated estimates $\mathbf{\Psi}^{(p+1)} = (\pi_1^{(p+1)}, \ldots, \pi_{M-1}^{(p+1)}, \xi^{(p+1)})^\top$. According to (4.2) the updated estimates $\pi_j^{(p+1)}$ of the mixing proportions π_j can be calculated independently of the updated estimates $\xi^{(p+1)}$ of the parameters ξ in $\mathbf{\Psi}$. If $Z_{ij}^{(r)}, W_{ij}^{(r)}$ and $V_{ij}^{(r)}$ were observed as $z_{ij}^{(r)}, w_{ij}^{(r)}$ and $v_{ij}^{(r)}$, the (complete data) ML estimate of π_j would be

$$\hat{\pi}_j = \frac{1}{mk^2}\sum_{i=1}^{m}\sum_{r=1}^{k}\left(z_{ij}^{(r)} + w_{ij}^{(r)} + v_{ij}^{(r)}\right). \tag{4.8}$$

They are not observable and, as showed in the Appendix, the updated estimate of π_j, $j = 1, \ldots, M$, is

$$\begin{aligned} \pi_{j,M1}^{(p+1)} = \frac{1}{mk^2}\sum_{i=1}^{m}\sum_{r=1}^{k}\Big\{&\tau_{j,M1,RSS}(x_{(r)i}; \mathbf{\Psi}^{(p)}) + \beta_{j,M1,RSS}(x_{(r)i}; \mathbf{\Psi}^{(p)}) \\ &+\gamma_{j,M1,RSS}(x_{(r)i}; \mathbf{\Psi}^{(p)})\Big\}. \end{aligned} \tag{4.9}$$

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

$$\sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M}\frac{\tau_{j,M1,RSS}(x_{(r)i};\boldsymbol{\Psi}^{(p)})}{f_j(x_{(r)i};\theta_j)}\frac{\partial}{\partial\xi}f_j(x_{(r)i};\theta_j)$$
$$+\sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M}\frac{\partial}{\partial\xi}F_j(x_{(r)i};\theta_j)\left(\frac{\beta_{j,M1,RSS}(x_{(r)i};\boldsymbol{\Psi}^{(p)})}{F_j(x_{(r)i};\theta_j)}-\frac{\gamma_{j,M1,RSS}(x_{(r)i};\boldsymbol{\Psi}^{(p)})}{\bar{F}_j(x_{(r)i};\theta_j)}\right)$$
$$=0, \tag{4.10}$$

with respect to ξ.

The E- and M-steps are alternated until $|l_{C,M1,RSS}(\boldsymbol{\Psi}^{(p+1)})-l_{C,M1,RSS}(\boldsymbol{\Psi}^{(p)})|$ is negligible.

4.2. EM algorithm for M2-RSS data

To formalize the EM algorithm based on M2-RSS, we only need to define one latent variable. Let $\mathbf{Z}_i^{(r)}=\left(Z_{i1}^{(r)},\ldots,Z_{iM}^{(r)}\right)$ be such that $Z_{ij}^{(r)}$ is one or zero, according to whether or not $X_{(r)i}$ corresponds to the rth order statistic of the jth component of the finite mixture model, $j=1,\ldots,M$. The conditional pdf of $X_{(1)1},\ldots,X_{(k)m}$ given $\mathbf{Z}_1^{(1)},\ldots,\mathbf{Z}_m^{(k)}$ is

$$f(x_{(1)1},\ldots,x_{(k)m}|\mathbf{z}_1^{(1)},\ldots,\mathbf{z}_m^{(k)};\xi)=\prod_{i=1}^{m}\prod_{r=1}^{k}f^{(r)}(x_{(r)i}|\mathbf{z}_i^{(r)};\xi),$$

with

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

The likelihood function of $\boldsymbol{\Psi}$ based on the complete M2-RSS data, denoted by $\mathbf{Y}_{M2}=\{(X_{(r)i},\mathbf{Z}_i^{(r)}),\ i=1,\ldots,m\ ;\ r=1,\ldots,k\}$, can be expressed as

$$L_{C,M2,RSS}(\boldsymbol{\Psi})$$
$$=\prod_{i=1}^{m}\prod_{r=1}^{k}\prod_{j=1}^{M}\left\{\pi_j f_j^{(r)}(x_{(r)i};\theta_j)\right\}^{Z_{ij}^{(r)}} \tag{4.11}$$
$$=\prod_{i=1}^{m}\prod_{r=1}^{k}\prod_{j=1}^{M}\left\{k\binom{k-1}{r-1}\pi_j f_j(x_{(r)i};\theta_j)[F_j(x_{(r)i};\theta_j)]^{r-1}[1-F_j(x_{(r)i};\theta_j)]^{k-r}\right\}^{Z_{ij}^{(r)}}.$$

The incomplete M2-RSS likelihood function (3.7) can be obtained by summing $\mathbf{Z}_\mathbf{i}^{(\mathbf{r})}$ out of the complete M2-RSS likelihood function $L_{C,M2,RSS}(\boldsymbol{\Psi})$, and then the complete data log-likelihood function of $\boldsymbol{\Psi}$ is

$l_{C,M2,RSS}(\boldsymbol{\Psi})$

$$= \sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M} Z_{ij}^{(r)} \left\{\log \pi_j + \log f_j^{(r)}(x_{(r)i};\theta_j)\right\}$$

$$= m\sum_{r=1}^{k}\log k\binom{k-1}{r-1} + \sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M} Z_{ij}^{(r)} \left\{\log \pi_j + \log f_j(x_{(r)i};\theta_j)\right\}$$

$$+ \sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M} Z_{ij}^{(r)} \left\{(r-1)\log F_j(x_{(r)i};\theta_j) + (k-r)\log(\bar{F}_j(x_{(r)i};\theta_j))\right\}. \quad (4.12)$$

The EM algorithm can be applied to obtain estimates of $\mathbf{\Psi}$. First,

$$\tau_{j,M2,RSS}(x_{(r)i};\mathbf{\Psi}^{(p)}) = E_{\mathbf{\Psi}^{(p)}}[Z_{ij}^{(r)}|x_{(r)i}]$$
$$= \frac{\pi_j^{(p)} f_j(x_{(r)i};\theta_j^{(p)})[F_j(x_{(r)i};\theta_j^{(p)})]^{r-1}[\bar{F}_j(x_{(r)i};\theta_j^{(p)})]^{k-r}}{\sum_{h=1}^{M}\pi_h^{(p)} f_h(x_{(r)i};\theta_h^{(p)})[F_h(x_{(r)i};\theta_h^{(p)})]^{r-1}[\bar{F}_h(x_{(r)i};\theta_h^{(p)})]^{k-r}}, \quad (4.13)$$

where $j = 1,\ldots,k$ and $i = 1,\ldots,m$ and $\mathbf{\Psi}^{(0)}$ denotes an initial value for $\mathbf{\Psi}$. Using (4.13), the conditional expectation of the complete M2-RSS data log-likelihood function (4.12) given the observed data $\mathbf{X}_{RSS} = \mathbf{x}_{RSS}$ (at the pth iteration of the EM algorithm) is

$$Q_{M2}(\mathbf{\Psi},\mathbf{\Psi}^{(p)})$$
$$= m\sum_{r=1}^{k}\log k\binom{k-1}{r-1} + \sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M}\tau_{j,M2,RSS}(x_{(r)i};\mathbf{\Psi}^{(p)})\left\{\log\pi_j + \log f_j(x_{(r)i};\theta_j)\right\}$$
$$+ \sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M}\tau_{j,M2,RSS}(x_{(r)i};\mathbf{\Psi}^{(p)})\{(r-1)\log F_j(x_{(r)i};\theta_j)$$
$$+(k-r)\log\bar{F}_j(x_{(r)i};\theta_j)\}. \quad (4.14)$$

In the M-step, mixing proportions are updated by (4.13) independently of the other parameters of model through

$$\pi_j^{(p+1)} = \frac{1}{mk}\sum_{i=1}^{m}\sum_{r=1}^{k}\tau_{j,M2,RSS}(x_{(r)i};\mathbf{\Psi}^{(p)}),\ j = 1,\ldots,M-1. \quad (4.15)$$

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

$$\sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M}\frac{\tau_{j,M2,RSS}(x_{(r)i};\mathbf{\Psi}^{(p)})}{f_j(x_{(r)i};\theta_j)}\frac{\partial}{\partial\xi}f_j(x_{(r)i};\theta_j)$$
$$+\sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M}\tau_{j,M2,RSS}(x_{(r)i};\mathbf{\Psi}^{(p)})\frac{\partial}{\partial\xi}F_j(x_{(r)i};\theta_j)\left(\frac{r-1}{F_j(x_{(r)i};\theta_j)} - \frac{k-r}{\bar{F}_j(x_{(r)i};\theta_j)}\right)$$

$$= 0. \tag{4.16}$$

The E- and M- steps are alternated repeatedly until $|l_{C,M2,RSS}(\mathbf{\Psi}^{(p+1)}) - l_{C,M2,RSS}(\mathbf{\Psi}^{(p)})|$ is negligible.

Remark 2. The complete data log-likelihood function of $\mathbf{\Psi}$ for the SRS sample can be written as

$$l_{C,SRS}(\mathbf{\Psi}) = \sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M} Z_{ij}^{(r)} \left\{\log \pi_j + \log f_j(x_{(r)i};\theta_j)\right\}, \tag{4.17}$$

where $x_{(r)i}$ refers to the (r,i)th element of the matrix of observations $\mathcal{D}$ and $Z_{ij}^{(r)}$ is its component indicator function. Then $l_{C,SRS}(\mathbf{\Psi}) = l_{C,M2,RSS}(\mathbf{\Psi}) - \Lambda_{M2}(\xi)$, with

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

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

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

For the EM algorithm's E-step, let $\mathbf{\Psi}^{(0)}$ be an initial value and $Q(\mathbf{\Psi},\mathbf{\Psi}^{(0)}) = E_{\mathbf{\Psi}^{(0)}}[l_{C,SRS}(\mathbf{\Psi})|\mathbf{x}_{SRS}]$. In the $(p+1)$th iteration we compute

$$Q(\mathbf{\Psi},\mathbf{\Psi}^{(p)}) = \sum_{i=1}^{m}\sum_{r=1}^{k}\sum_{j=1}^{M} \tau_{j,SRS}(x_{(r)i};\mathbf{\Psi}^{(p)}) \left\{\log \pi_j + \log f_j(x_{(r)i};\theta_j)\right\}, \tag{4.18}$$

where

$$\tau_{j,SRS}(x_{(r)i};\mathbf{\Psi}^{(p)}) = \frac{\pi_j^{(p)} f_j(x_{(r)i};\theta_j^{(p)})}{\sum_{h=1}^{M}\pi_h^{(p)} f_h(x_{(r)i};\theta_h^{(p)})}. \tag{4.19}$$

For the M-step, at its $(p+1)$th iteration, a local maximization of $Q(\mathbf{\Psi},\mathbf{\Psi}^{(p)})$ with respect to $\mathbf{\Psi}$ is done to obtain the updated estimate $\mathbf{\Psi}^{(p+1)}$. The updated estimate of π_j is

$$\pi_j^{(p+1)} = \frac{1}{mk} \sum_{i=1}^{m} \sum_{r=1}^{k} \tau_{j,SRS}(x_{(r)i}; \mathbf{\Psi}^{(p)}), \; j = 1, \ldots, M-1, \tag{4.20}$$

while the updated estimate of ξ is obtained as an appropriate root of

$$\sum_{i=1}^{m} \sum_{r=1}^{k} \sum_{j=1}^{M} \frac{\tau_{j,SRS}(x_{(r)i}; \mathbf{\Psi}^{(p)})}{f_j(x_{(r)i}; \theta_j)} \frac{\partial}{\partial \xi} f_j(x_{(r)i}; \theta_j) = 0. \tag{4.21}$$

Comparing (4.10) and (4.16) with (4.21) shows the contribution of the ranks provided via the RSS technique in obtaining the ML estimates of $\mathbf{\Psi}$.

5. Classification of an 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 give a probabilistic classification of the data. Suppose $i \in \{1, \ldots, m\}$ is fixed and $y_r = x_{(r)i}$ is observed. For M2-RSS design, classification of y_r is based on the posterior probability that y_r belongs to the jth component of the mixture model, that is $\tau_{j,M2,RSS}(y_r; \mathbf{\Psi})$, as in (4.13). For M1-RSS design classification of y_r is done using

$$\alpha_{j,M1,RSS} = \frac{1}{k} \left\{ \tau_{j,M1,RSS}(y_r, \mathbf{\Psi}) + \beta_{j,M1,RSS}(y_r, \mathbf{\Psi}) + \gamma_{j,M1,RSS}(y_r, \mathbf{\Psi}) \right\}. \tag{5.1}$$

We classify y_r into the component j if

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

for all $t \neq j$, $t = 1, \ldots, M$. Here we focus on this approach to classification to demonstrate the effect of the extra information through ranks in RSS designs compared with SRS. However, as explained by Celeux and Govaert (1993) or McLachlan and Peel (2004), one can do better by using other approaches to classification.

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

$$\tau_{j,SRS}(y_r; \mathbf{\Psi}) = \frac{\pi_j f_j(y_r; \theta_j)}{\sum_{h=1}^{M} \pi_h f_h(y_r; \theta_h)}. \tag{5.2}$$

It can be seen that

$$\alpha_{j,M1,RSS}(y_r; \mathbf{\Psi}) = \frac{1}{k} \left\{ \tau_{j,SRS}(y_r; \mathbf{\Psi}) + A_{j,M1,RSS}(y_r; \mathbf{\Psi}) \right\}, \tag{5.3}$$

where

$$A_{j,M1,RSS}(y_r;\mathbf{\Psi}) = (r-1)\left(\frac{\pi_j F_j(x_{(r)i};\theta_j)}{F(x_{(r)i};\mathbf{\Psi})}\right) + (k-r)\left(\frac{\pi_j \bar{F}_j(x_{(r)i};\theta_j)}{\bar{F}(x_{(r)i};\mathbf{\Psi})}\right). \quad (5.4)$$

For M2-RSS, we get

$$\tau_{j,M2,RSS}(y_r;\mathbf{\Psi}) = \tau_{j,SRS}(y_r;\mathbf{\Psi})A_{j,M2,RSS}(y_r;\mathbf{\Psi}), \quad (5.5)$$

where

$$A_{j,M2,RSS}(y_r;\mathbf{\Psi}) = \frac{\left(\sum_{h=1}^{M}\pi_h f_h(y_r;\theta_h)\right)[F_j(y_r;\theta_j)]^{r-1}[1-F_j(y_r;\theta_j)]^{k-r}}{\sum_{h=1}^{M}\pi_h f_h(y_r;\theta_h)[F_h(y_r;\theta_h)]^{r-1}[1-F_h(y_r;\theta_h)]^{k-r}}. \quad (5.6)$$

An example demonstrates the effect of the rank information on classification. Suppose $x = 0$ is observed from a population. Assume that the underlying population has components C_1 and C_2 with a pdf

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

where $\phi(x;\mu,\sigma)$ is the pdf of a normal distribution with mean μ and variance σ^2. Assume that $x = 0$ is observed through an RSS with set size $k = 3$. As shown in Table 1, by treating $x = 0$ as an observation obtained from a SRS and since $\tau_{2,SRS}(0;\mathbf{\Psi}) = 0.8176 \geq \tau_{1,SRS}(0;\mathbf{\Psi}) = 0.1824$, then $x = 0$ should be classified into the second component, C_2, of the population. Now, if $x = 0$ is the observation of the third order statistic via the M1-RSS technique with $k = 3$, then $\alpha_{1,M1,RSS}(0;\mathbf{\Psi}) = 0.6344 \geq \alpha_{2,M1,RSS}(0;\mathbf{\Psi}) = 0.3656$ and $x = 0$ should be classified into C_1. Similarly, from Table 2, if $x = 0$ is the observation of the third order statistic in an M2-RSS design with $k = 3$, we have $\tau_{1,M2,RSS}(0;\mathbf{\Psi}) = 0.8944 \geq \tau_{2,M2,RSS}(0;\mathbf{\Psi}) = 0.1056$, and $x = 0$ should be classified into C_1. Tables 1 and 2 provide the values of the posterior probabilities for classification of $x = 0$ under SRS, M1-RSS, and M2-RSS designs for two different mixtures of normal distributions.

Numerical evaluations through different simulation studies show that with increasing set size k, the RSS technique results in better classification of the observations than the SRS method based on its misclassification error rate. Observe also that if $T_j \sim Bin(k-1, F_j(y_r;\theta_j))$, (5.6) can be written as

$$A_{j,M2,RSS}(y_r,\mathbf{\Psi}) = \frac{\left(\sum_{h=1}^{M}\pi_h f_h(y_r;\theta_h)\right)P(T_j = r-1)}{\sum_{h=1}^{M}\pi_h f_h(y_r;\theta_h)P(T_h = r-1)}, \quad (5.7)$$

Table 1. Posterior probabilities for classification of $x = 0$ under SRS and M1-RSS techniques.

Mixture Model	Mixture Component	SRS	M1-RSS with $k=3$		
		$x=0$	$y_1=0$	$y_2=0$	$y_3=0$
$0.5\phi(x;-2,1)+0.5\phi(x;1,1)$	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. Posterior probabilities for classification of $x = 0$ under SRS and M2-RSS techniques.

Mixture Model	Mixture Component	SRS	M2-RSS with $k=3$		
		$x=0$	$y_1=0$	$y_2=0$	$y_3=0$
$0.5\phi(x;-2,1)+0.5\phi(x;1,1)$	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

and if $P(T_j = r-1) \geq P(T_h = r-1)$ for all $h \neq j \in \{1,\ldots,M\}$ then RSS results in a bigger value than its SRS counterpart for the posterior probability that y_r belongs to the jth component of the mixture model.

6. Simulation Studies for Mixture of Normal Densities

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 their corresponding one under SRS.

6.1. Simulation study 1

In the first simulation, our goal was to compare the performance of the estimators of the mixing proportion π using M1-RSS, M2-RSS, and SRS data for the mixture

$$f(x,\boldsymbol{\Psi}) = \pi\phi(x;\mu_1,\sigma) + (1-\pi)\phi(x;\mu_2,\sigma). \tag{6.1}$$

We first compared the performance of $\hat{\pi}$ under M1-RSS and SRS designs. We generated two data sets each of size $mk = 120$ from the model (6.1) with

Table 3. $\hat{\pi}_{k,MLE}$ based on SRS ($k = 1$) and M1-RSS designs ($k \geq 2$), their (standard error), [MSE] and RE for model (6.1).

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

$(\pi, \mu_1, \sigma) = (0.8, -1, 1)$. To study the effect of the set size as design parameter of RSS on performance, we let $k \in \{1, 2, 3, 4, 5\}$. Note that $k = 1$ 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 $d = \mu_2 - \mu_1$ with $d \in \{1, 3\}$. The EM algorithm was performed, assuming equal initial values for mixing proportions of two components and stopping criteria $|\hat{\mathbf{\Psi}}^{(k+1)} - \hat{\mathbf{\Psi}}^{(k)}| < 10^{-5}$. Table 3 provides ML estimates with their standard errors and mean square errors (MSE) under SRS and M1-RSS with different set sizes. The estimates of standard errors and biases (used for MSE's) were obtained via a bootstrap with $l = 100$ repeats. We followed Basford et al. (1997) to find the approximations.

We also calculated the observed relative efficiency of ML estimators $\hat{\pi}_{k,MLE}$ of π under SRS and RSS-based designs using

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

Table 3 presents the values of the relative efficiencies for different set sizes, $d \in \{1, 3\}$. The results indicate that ML estimates of π under the M1-RSS design are more efficient than their corresponding estimators under SRS and that relative efficiency is an increasing function of k. In addition, when $d = 3$, the performance of M1-RSS design for estimating π was much better than that of SRS. We obtained similar results under the M2-RSS design. However, the

Table 4. $\hat{\pi}_{k,MLE}$ based on SRS ($k = 1$) and M2-RSS designs ($k \geq 2$), their (standard error), [MSE] and RE for model (6.1).

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

efficiency of ML estimators under the M2-RSS design, when components were separated, was slightly reduced compared to the non-separated case.

6.2. Simulation study 2

In the second simulation study, the performance of ML estimates of all parameters of the mixture model using M1-RSS, M2-RSS and SRS designs was investigated. The underlying distribution was chosen to be a homosedastic 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 $mk = 300$ from two mixture of normal densities of the form (6.1) with $\mathbf{\Psi}_1 = (\pi, \mu_1, \mu_2, \sigma) = (0.4, -2, 1, 1)$, Model 1, and $\mathbf{\Psi}_2 = (0.4, -1, 1, 1.5)$, Model 2, using SRS, M1-RSS, and M2-RSS techniques with $k = 5$.

We examined three methods for setting the initial values of $\mathbf{\Psi}$: fixed initial values; Finch's method; and the method of moments (see Finch, Mendell, and Thode (1989) and Karlis and Xekalaki (2003)). The stopping criteria was $\|\hat{\mathbf{\Psi}}^{(k+1)} - \hat{\mathbf{\Psi}}^{(k)}\|_\infty < 10^{-5}$. The bias, standard error, and MSE were used as performance measures for each estimator. These were obtained via the bootstrap, Basford et al. (1997), with $b = 100$ and 10 repetition. The mean and standard errors of these measures under fixed initial values, Finch's method, and the method of moments are reported in Tables 5, 6, and 7, respectively. 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.

Table 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]

We used $\boldsymbol{\Psi}_1^{(0)} = (\pi^{(0)}, \mu_1^{(0)}, \mu_2^{(0)}, \sigma^{(0)}) = (0.3, -1.95, 0.95, 0.95)$, and $\boldsymbol{\Psi}_2^{(0)} = (0.3, -0.95, 0.95, 1.40)$ as initial values for the fixed initial values method; the initial values for the Finch's method corresponded to the case (v) in Karlis and Xekalaki (2003); for the method of moments the initial values were obtained via the formula developed by Furman and Lindsay (1994).

In Tables 5, 6 and 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). It is evident that the M2-RSS technique is more efficient than M1-RSS; an M2-RSS design has the more informative assumption that $x_{(r)i}$ is indeed the rth order statistic of one of the components of the model.

7. Conclusion

In this paper, we studied maximum likelihood estimation of the unknown parameters of a finite mixture model using RSS. We proposed two ranked-based sampling designs and developed new EM algorithms to calculate the ML estimates of the parameters of the model for each design. According to our simulation studies, RSS-based designs resulted in more efficient estimates than did SRS-based design. The validity of the developed methods under RSS designs rely

Table 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]

on the assumption of perfect ranking; when RSS is imperfect, it is understood that the efficiency of ML estimators of the parameters of the model decreases. However, a numerical study, not presented here, found the basic behaviour presented in this article did not change under imperfect ranking. We think that using a suitably unbalanced RSS can lead to a better inference about the parameters of the FMM over the balanced RSS. To this end, we considered the problem of estimating the mixing proportion of the densities, studied in Subsection 6.1, using unbalanced ranked set samples consisting of only minimums, only maximums, and both minimums and maximums. The results for a small simulation study are presented in Table 8. The performance of the unbalanced RSS designs compared with their SRS counterpart was calculated using the ratio of the MSEs of the maximum likelihood estimators of π based on unbalanced RSS and SRS samples of size 120. We also investigated relative efficiency as a function of the set size $k \in \{2, 3, 4, 5\}$. Comparing the results in Table 8 with Tables 3 and 4 shows that the performance of the unbalanced RSS design could be better than the balanced RSS. For large $d = \mu_2 - \mu_1$, the SRS performs better than an unbalanced M1-RSS consisting of only the minimums. A comprehensive study of the problem of finite mixture modelling based on unbalanced RSS, as well as the effect of imperfect ranking, is under investigation and results will appear in another paper.

Table 7. 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 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]

Table 8. The relative efficiency of the maximum likelihood estimators of π using unbalanced M1-RSS and M2-RSS samples consisting of the minimums, the maximums and both minimums and maximums for model (6.1), when $\pi = 0.8$, $\mu_1 = -1$, $\mu_2 = \mu_1 + d$, and $\sigma = 1$.

d	design	RSS	$k=2$	$k=3$	$k=4$	$k=5$
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

Acknowledgements

Armin Hatefi's research is supported by the University of Manitoba Graduate Fellowship (UMGF) and by Manitoba Graduate Scholarship (MGS). Authors gratefully acknowledges the research support of the Natural Sciences and Engineering Research Council of Canada. The authors are grateful to two anonymous referees, a consulting editor and an associate editor for their constructive comments and suggestions.

Appendices

Throughout this section, we let $c_1 = k\binom{k-1}{r-1}$, $c_2 = \binom{r-1}{w_{i1}^{(r)},\ldots,w_{iM}^{(r)}}$ and $c_3 = \binom{k-r}{v_{i1}^{(r)},\ldots,v_{iM}^{(r)}}$.

A.1. Proof of Lemma 1

The conditional *pdf* of latent variables $Z_{ij}^{(r)}, W_{ij}^{(r)}, V_{ij}^{(r)}$ given $X_{(r)i}$ are

$$
\begin{aligned}
f(z_{ij}^{(r)}|x_{(r)i},\boldsymbol{\Psi}) &= \frac{f(x_{(r)i}|z_{ij}^{(r)},\boldsymbol{\Psi})f(z_{ij}^{(r)})}{f^{(r)}(x_{(r)i};\boldsymbol{\Psi})}\\
&= \frac{c_1\prod_{j=1}^{M}\{f_j(x_{(r)i};\theta_j)\}^{z_{ij}^{(r)}}\{F(x_{(r)i};\boldsymbol{\Psi})\}^{r-1}\{\bar{F}(x_{(r)i};\boldsymbol{\Psi})\}^{k-r}\prod_{j=1}^{M}\pi_j^{z_{ij}^{(r)}}}{c_1 f(x_{(r)i};\boldsymbol{\Psi})\{F(x_{(r)i};\boldsymbol{\Psi})\}^{r-1}\{\bar{F}(x_{(r)i};\boldsymbol{\Psi})\}^{k-r}}\\
&= \prod_{j=1}^{M}\left(\frac{\pi_j f_j(x_{(r)i};\theta_j)}{f(x_{(r)i};\boldsymbol{\Psi})}\right)^{z_{ij}^{(r)}},\\
f(w_{ij}^{(r)}|x_{(r)i},\boldsymbol{\Psi}) &= \frac{f(x_{(r)i}|w_{ij}^{(r)},\boldsymbol{\Psi})f(w_{ij}^{(r)})}{f^{(r)}(x_{(r)i};\boldsymbol{\Psi})}\\
&= \frac{c_1c_2 f(x_{(r)i};\boldsymbol{\Psi})\prod_{j=1}^{M}\{F_j(x_{(r)i};\theta_j)\}^{w_{ij}^{(r)}}\{\bar{F}(x_{(r)i};\boldsymbol{\Psi})\}^{k-r}\prod_{j=1}^{M}\pi_j^{w_{ij}^{(r)}}}{c_1 f(x_{(r)i};\boldsymbol{\Psi})\{F(x_{(r)i};\boldsymbol{\Psi})\}^{r-1}\{\bar{F}(x_{(r)i};\boldsymbol{\Psi})\}^{k-r}}\\
&= c_2\prod_{j=1}^{M}\left(\frac{\pi_j F_j(x_{(r)i};\theta_j)}{F(x_{(r)i};\boldsymbol{\Psi})}\right)^{w_{ij}^{(r)}},\\
f(v_{ij}^{(r)}|x_{(r)i},\boldsymbol{\Psi}) &= \frac{f(x_{(r)i}|v_{ij}^{(r)},\boldsymbol{\Psi})f(v_{ij}^{(r)})}{f^{(r)}(x_{(r)i};\boldsymbol{\Psi})}\\
&= \frac{c_1c_3 f(x_{(r)i};\boldsymbol{\Psi})\{F(x_{(r)i};\boldsymbol{\Psi})\}^{r-1}\prod_{j=1}^{M}\{\bar{F}_j(x_{(r)i};\theta_j)\}^{v_{ij}^{(r)}}\prod_{j=1}^{M}\pi_j^{v_{ij}^{(r)}}}{c_1 f(x_{(r)i};\boldsymbol{\Psi})\{F(x_{(r)i};\boldsymbol{\Psi})\}^{r-1}\{\bar{F}(x_{(r)i};\boldsymbol{\Psi})\}^{k-r}}\\
&= c_3\prod_{j=1}^{M}\left(\frac{\pi_j \bar{F}_j(x_{(r)i};\theta_j))}{\bar{F}(x_{(r)i};\boldsymbol{\Psi})}\right)^{v_{ij}^{(r)}}.
\end{aligned}
$$

From the independence of latent variables, we have

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

A.2. Proof of Lemma 2

We have

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

A.3. Proof of (4.9)

The maximization of $Q_{M1}(\mathbf{\Psi}, \mathbf{\Psi}^{(p)})$ over π_j is done under the constraint $\sum_{j=1}^{M} \pi_j = 1$. We use the Lagrangian multiplier method, taking

$$
\mathcal{L}(\mathbf{\Psi}, \lambda) = Q_{M1}(\mathbf{\Psi}, \mathbf{\Psi}^{(p)}) - \lambda \Big(\sum_{j=1}^{M} \pi_j - 1 \Big).
$$

Differentiating $\mathcal{L}(\boldsymbol{\Psi}, \lambda)$ with respect to π_j leads to

$$\pi_{j,M1}^{(p+1)} = \frac{1}{\lambda}\sum_{i=1}^{m}\sum_{r=1}^{k}\Big\{\tau_{j,M1,RSS}(x_{(r)i};\boldsymbol{\Psi}^{(p)}) + \beta_{j,M1,RSS}(x_{(r)i};\boldsymbol{\Psi}^{(p)}) + \gamma_{j,M1,RSS}(x_{(r)i};\boldsymbol{\Psi}^{(p)})\Big\},$$

where

$$\lambda = \sum_{j=1}^{M}\sum_{i=1}^{m}\sum_{r=1}^{k}\Big\{\tau_{j,M1,RSS}(x_{(r)i};\boldsymbol{\Psi}^{(p)}) + \beta_{j,M1,RSS}(x_{(r)i};\boldsymbol{\Psi}^{(p)}) + \gamma_{j,M1,RSS}(x_{(r)i};\boldsymbol{\Psi}^{(p)})\Big\}.$$

Using (4.4), (4.5) and (4.6),

$$\sum_{j=1}^{M}\Big\{\tau_{j,M1,RSS}(x_{(r)i};\boldsymbol{\Psi}^{(p)}) + \beta_{j,M1,RSS}(x_{(r)i};\boldsymbol{\Psi}^{(p)}) + \gamma_{j,M1,RSS}(x_{(r)i};\boldsymbol{\Psi}^{(p)})\Big\} = k,$$

and so $\lambda = \sum_{i=1}^{m}\sum_{r=1}^{k} k = mk^2$. This completes the proof.

References

Basford, K. E., Greenway, D. R., McLachlan, G. J. and Peel, D. (1997). Standard errors of fitted component means of normal mixtures. *Comput. Statist.* **12**, 1-18.

Celeux, G. and Govaert, G. (1993). Comparison of the mixture and the classification maximum likelihood in cluster analysis. *J. Statist. Comput. Simulation* **47**, 127-146.

Chen, Z., Bai, Z. and Sinha, B. K. (2004). *Ranked Set Sampling: Theory and Applications.* Volume 176 of Lecture Notes in Statistics. Springer-Verlag, New York.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* (Methodological), 1-38.

Finch, S. J., Mendell, N. R. and Thode, J. H. C. (1989). Probabilistic measures of adequacy of a numerical search for a global maximum. *J. Amer. Statist. Assoc.* **84**, 1020-1023.

Furman, W. D. and Lindsay, B. G. (1994). Measuring the relative effectiveness of moment estimators as starting values in maximizing likelihoods. *Comput. Statist. Data Anal.* **17**, 493-507.

Karlis, D. and Xekalaki, E. (2003). Choosing initial values for the EM algorithm for finite mixtures. *Comput. Statist. Data Anal.* **41**, 577-590.

Kaur, A., Patil, G. P., Sinha, A. K. and Taillie, C. (1995). Ranked set sampling: an annotated bibliography. *Environmental and Ecological Statistics* **2**, 25-54.

McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets. *Crop and Pasture Sci.* **3**, 385-390.

McLachlan, G. and Peel, D. (2004). *Finite Mixture Models.* Wiley-Interscience.

Takahasi, K. and Wakimoto, K. (1986). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Statist. Math.* **20**, 1-31.

Wang, Y. G., Ye, Y. and Milton, D. A. (2009). Efficient designs for sampling and subsampling in fisheries research based on ranked sets. **66**, 928-934.

Wedel, M., Ter Hofstede, F. and Steenkamp, J. B. E. M. (1998). Mixture model analysis of complex samples. *J. Classification* **15**, 225-244.

Department of Statistics, University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2.

E-mail: umhatefi@cc.umanitoba.ca

Department of Statistics, University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2.

E-mail: m_jafari_jozani@umanitoba.ca

Université de Sherbrooke, Département d'Informatique, Sherbrooke, Quebec, Canada, J1K 2R1.

E-mail: djemel.ziou@usherbrooke.ca

(Received June 2012; accepted March 2013)