Statistica Sinica Preprint No: SS-2018-0266	
Title	Finite Mixture Modeling, Classification and Statistical Learning with Order Statistics
Manuscript ID	SS-2018-0266
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0266
Complete List of Authors	Armin Hatefi Nancy Reid Mohammad Jafari Jozani and Omer Ozturk
Corresponding Author	Armin Hatefi
E-mail	ahatefi@mun.ca; hatefi.ar@gmail.com
Notice: Accepted version subject to English editing.	

Statistica Sinica

Finite Mixture Modeling, Classification and Statistical Learning with Order Statistics

Armin Hatefi[1], Nancy Reid[2], Mohammad Jafari Jozani[3] and Omer Ozturk[4]

[1]*Memorial University,* [2]*University of Toronto,* [3]*University of Manitoba and* [4]*The Ohio State University*

Abstract: In this paper, we provide a unified approach to maximum likelihood estimation, classification and statistical learning in the context of finite mixture models based on observations that can be considered as a collection of order statistics. We consider these problems under both supervised and unsupervised learning approaches. New missing data mechanisms and EM algorithms are developed to exploit the structure of the observed data in the estimation process under each learning strategy. Also, we present some model-based classification criteria and show how they can be used to make better inference about rarely observed components in finite mixture models. Through simulation studies, we evaluate the performance of the estimation and classification methodologies. Finally the proposed methods are applied to a fishery study to estimate the age structure of Spot as a short-lived fish species.

Key words and phrases: Finite mixture models; order statistics; ranked Set Sam-

pling; classification, latent variables; EM algorithm.

1. Introduction

Consider a population consisting of M subpopulations and suppose that we are interested in a random phenomenon X with a probability density function (pdf) that can be written as a finite mixture model (FMM). Suppose we randomly select n sampling units from the population. There are many situations where some of the observations in the sample may be missing, either at random or not, although, we might easily be able to assign ranks to the observed values and thus retain order statistics. A typical situation occurs in life-testing when an experiment is terminated after the first r out of n items under the test have failed, where each item is composed of M different components with different lifetime distributions. Observations of this kind are called censored samples and can lead to the selection of different types of order statistics out of samples of size n. A collection of order statistics can be found in another setting, where finding the final measurements on all the sampling units is expensive due to the budget cut and/or some other constraint. Experiment may be scaled back to select a subset of the sampled units for final study. For example, in studies involving the age determination of fish populations, it is a common practice to first catch

a large number of fish and then subsample for age determination. In this case, the subsampling within the larger sample is often done after ordering the larger sample using the length of the fish and then implementing systematic sampling. This approach, for example selecting every 3rd fish in the ordered sample, is easy to explain and easy for field workers to follow. We use the term *selected order statistics* when observations are obtained from specific designs that lead to specific choices of order statistics. Examples of designs leading to selected order statistics are

- single censored samples from FMMs, where either the r_1 smallest (left censored) X values or the r_2 largest (right censored) X values are not observed, when r_1 and r_2 are fixed by design (Miyata, 2011; Mendenhall and Hader, 1958).

- doubly censored samples from FMMs, where the r_1 smallest and r_2 largest X values are not observed, with fixed values of r_1 and r_2 (Sindhu et al., 2016; Saleem et al., 2010).

- compressed data from FMMs, where a large number of data points are replaced by a small number of selected order statistics (Bishop, 2006).

- systematic subsamples with auxiliary information enabling the order-

ing of sampled units, as described in fish example above.

We also use the term *induced order statistics* when, after observing a simple random sample with missing observations, auxiliary information is used to assign a rank to each observation. In all these examples, observations can be considered as collections of order statistics of a sample of size n from a FMM, whether or not they are labelled or unlabelled. In other words, we might or might not know the subpopulation from which the data is observed. The research interest is now to estimate the unknown parameters of the underlying FMM using this data.

There are several variations of rank-based sampling (RBS) designs that lead to independent order statistics. Inference for FMMs in these settings is discussed in Hatefi et al. (2014, 2015). In this paper, the order statistics are correlated and finite mixture modelling is a more challenging problem. We provide a unified approach to statistical inference about FMMs based on various collection of order statistics. We consider the problem under both supervised and unsupervised learning methods. To obtain ML estimates of the parameters, we introduce new missing data mechanisms and EM-algorithms, which accommodate the dependence structure among the order statistics. This imposes various kinds of difficulties in the estimation process, as the log-likelihood function contains terms that are convex

combinations of survival functions which typically do not have closed form for many statistical distributions. Moreover, new model-based classification criteria are developed for the FMM with rarely observed components.

Section 2 discusses the likelihood functions based on unlabelled order statistics of FMMs. The needed EM algorithm and its modified version are explained in Section 3. Section 4 presents different model-based classification criteria. In Section 5, we study estimators of the parameters of FMMs under supervised learning method. Section 6 compares the performance of different estimation procedures through numerical studies. The proposed estimation procedures are employed in a fishery study to determine the age structure of fish in Section 7. Finally, some concluding remarks are presented in Section 8. All proofs, some further remarks as well as a simulation study are provided in the Appendix as supplementary materials.

2. Order Statistics of the FMM

Suppose that the pdf of a random variable of interest X follows a mixture of M component densities

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

where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_M)$ is the vector of unknown mixing proportions with $\pi_j > 0$, $\sum_{j=1}^{M} \pi_j = 1$, and $f_j(\cdot;\theta_j)$; $j = 1, \ldots, M$, refers to the pdf of the j-th

component of the FMM, specified up to a vector θ_j of unknown parameters, known priori to be distinct. Let $\boldsymbol{\Psi} = (\pi_1, \ldots, \pi_{M-1}, \boldsymbol{\xi})^\top$, denote the vector of all unknown parameters, where $\boldsymbol{\xi} = (\theta_1^\top, \ldots, \theta_M^\top)^\top$ and superscript $\top$ refers to vector transpose. The cumulative distribution function (cdf) of X is $F(x; \boldsymbol{\Psi}) = \sum_{j=1}^{M} \pi_j F_j(x; \theta_j)$, where $F_j(\cdot; \theta_j)$ represents the cdf of the j-th component. For more details regarding the theory and applications of FMMs, see McLachlan and Peel (2004).

Suppose $\tilde{\mathbf{X}} = \{X_{(i_1)}, X_{(i_2)}, \ldots, X_{(i_k)}\}$ with $1 \leq i_1 \leq i_2 \leq \ldots \leq i_k \leq n$ is a collection of k, $k = 2, \ldots, n-1$ order statistics in a random sample of size n from (2.1), where $X_{(i_l)}$ is the i_lth smallest observation in the sample.

According to the theory of order statistics, the log-likelihood function of $\boldsymbol{\Psi}$ based on $\tilde{\mathbf{X}} = \tilde{\mathbf{x}}$ is

$$l(\boldsymbol{\Psi}|\tilde{\mathbf{x}}) \propto \sum_{r=1}^{k} \log f(x_{i_r}; \boldsymbol{\Psi}) + (i_1 - 1) \log F(x_{i_1}; \boldsymbol{\Psi}) + (n - i_k) \log \bar{F}(x_{i_k}; \boldsymbol{\Psi}) + \sum_{s=2}^{k} (i_s - i_{s-1} - 1) \log \left[F(x_{i_s}; \boldsymbol{\Psi}) - F(x_{i_{s-1}}; \boldsymbol{\Psi}) \right] \tag{2.2}$$

and the MLE of $\boldsymbol{\Psi}$, $\widehat{\boldsymbol{\Psi}}_{MLE}$ is obtained as the solution of $\frac{\partial}{\partial \boldsymbol{\Psi}} l(\boldsymbol{\Psi}|\tilde{\mathbf{x}}) = 0$ in $\boldsymbol{\Psi}$. Due to the complexity of (2.2), this is typically intractable due to the presence of convex combinations of components of the form $\log f(x_{i_r}; \boldsymbol{\Psi})$, $\log F(x_{i_1}; \boldsymbol{\Psi})$, $\log \left[F(x_{i_s}; \boldsymbol{\Psi}) - F(x_{i_{s-1}}; \boldsymbol{\Psi}) \right]$ and $\log \bar{F}(x_{i_k}; \boldsymbol{\Psi})$. To tackle the problem, we model $\tilde{\mathbf{X}} = \tilde{\mathbf{x}}$ as an incomplete data. The likelihood and log-

likelihood functions based on $\tilde{\mathbf{X}} = \tilde{\mathbf{x}}$ are then called incomplete likelihood and log-likelihood functions, respectively.

To obtain $\widehat{\mathbf{\Psi}}_{MLE}$, we construct a new EM algorithm following the idea of Dempster et al. (1977). Let $\mathbf{\Delta} = \{\mathbf{Z}_1, \ldots, \mathbf{Z}_k, \mathbf{W}_1, \ldots, \mathbf{W}_{k+1}\}$ be a collection of $2k+1$ latent vectors, each of length M. For each order statistic $X_{(i_r)}$, $r = 1, \ldots, k$, we define $\mathbf{Z}_r = (Z_{r1}, \ldots, Z_{rM})$, with $\mathbf{Z}_r \overset{i.i.d}{\sim} Mult(1, \boldsymbol{\pi})$. We also introduce

- $\mathbf{W}_1 = (W_{11}, \ldots, W_{1M})$, with $\mathbf{W}_1 \sim Mult(i_1 - 1, \boldsymbol{\pi})$,
- $\mathbf{W}_s = (W_{s1}, \ldots, W_{sM})$, with $\mathbf{W}_s \sim Mult(i_s - i_{s-1} - 1, \boldsymbol{\pi})$, for $s = 2, \ldots, k$, and
- $\mathbf{W}_{k+1} = (W_{k+1\,1}, \ldots, W_{k+1\,M})$, with $\mathbf{W}_{k+1} \sim Mult(n - i_k, \boldsymbol{\pi})$.

The complete likelihood function is given by the following Lemma, where the proof is provided in the Appendix.

Lemma 1. *Let* $\tilde{\mathbf{X}} = \{X_{(i_1)}, X_{(i_2)}, \ldots, X_{(i_k)}\}$ *be a collection of* $k = 2, \ldots, n-1$ *order statistics from a random sample of size* n *from* (2.1) *and let* $\mathbf{\Delta} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_k, \mathbf{W}_1, \ldots, \mathbf{W}_{k+1})$ *be the collection of latent vectors as defined*

above. The complete-data likelihood function based on $(\tilde{\mathbf{X}}, \boldsymbol{\Delta})$ *is given by*

$$f(\tilde{\mathbf{x}}, \boldsymbol{\delta}; \boldsymbol{\Psi}) \propto \prod_{j=1}^{M} \{\pi_j F_j(x_{i_1}; \theta_j)\}^{w_{1j}} \left\{\pi_j \bar{F}_j(x_{i_k}; \theta_j)\right\}^{w_{k+1\,j}} \prod_{r=1}^{k} \{\pi_j f_j(x_{i_r}; \theta_j)\}^{z_{rj}} \times \left(\prod_{s=2}^{k} \left[\pi_j \{F_j(x_{i_s}; \theta_j) - F_j(x_{i_{s-1}}; \theta_j)\}\right]^{w_{sj}} \right).$$

Using Lemma 1, the joint distribution of $(\tilde{\mathbf{X}}, \mathbf{Z}_r); r = 1, \ldots, k$, is

$$f(\tilde{\mathbf{x}}, \mathbf{z}_r) \propto \{F(x_{i_1}; \boldsymbol{\Psi})\}^{i_1 - 1} \prod_{j=1}^{M} \{\pi_j f_j(x_{i_r}; \theta_j)\}^{z_{rj}} \prod_{\substack{s=1 \\ s \neq r}}^{k} f(x_{i_s}; \boldsymbol{\Psi}) \times \prod_{s=2}^{k} \left\{F(x_{i_s}; \boldsymbol{\Psi}) - F(x_{i_{s-1}}; \boldsymbol{\Psi})\right\}^{i_s - i_{s-1} - 1} \left\{\bar{F}(x_{i_k}; \boldsymbol{\Psi})\right\}^{n - i_k} \quad (2.3)$$

In the Appendix, we provide further remarks to discuss the joint pdf of order statistics and their latent variables.

From (2.3) and the pdf of order statistics, one can easily show that

$$f_{\mathbf{Z}_r|\tilde{\mathbf{X}}}(\mathbf{z}_r|\tilde{\mathbf{x}}) = \prod_{j=1}^{M} \left\{ \frac{\pi_j f_j(x_{i_r}; \theta_j)}{f(x_{i_r}; \boldsymbol{\Psi})} \right\}^{z_{rj}}, \quad (2.4)$$

and conclude that $\mathbf{Z}_r|\tilde{\mathbf{X}} = \tilde{\mathbf{x}} \sim Mult\left(1, \frac{\pi_1 f_1(x_{i_r};\theta_1)}{f(x_{i_r};\boldsymbol{\Psi})}, \ldots, \frac{\pi_M f_M(x_{i_r};\theta_M)}{f(x_{i_r};\boldsymbol{\Psi})}\right)$ for each $r = 1, \ldots, k$.

Lemma 2. *Let* $\mathbf{Z}_r$ *be the latent vector associated with* $X_{(r)}$, $r = 1, \ldots, k$. *Given order statistics,* $\mathbf{Z}_r$ *are independent and identically distributed.*

The proof, due to Yang (1977), is given in Appendix.

Based on Remark 5 in the Appendix and pdf of order statistics, we have

$$f_{\mathbf{W}_1|\tilde{\mathbf{X}}}(\mathbf{w}_1|\tilde{\mathbf{x}}) = \binom{i_1 - 1}{w_{11}, \ldots, w_{1M}} \prod_{j=1}^{M} \left(\frac{\pi_j F_j(x_{i_1}; \theta_j)}{F(x_{i_1}; \boldsymbol{\Psi})} \right)^{w_{1j}}, \quad (2.5)$$

that is $\mathbf{W}_1|\tilde{\mathbf{X}} = \tilde{\mathbf{x}} \sim Mult\left(i_1 - 1, \frac{\pi_1 F_1(x_{i_1};\theta_1)}{F(x_{i_1};\boldsymbol{\Psi})}, \ldots, \frac{\pi_M F_M(x_{i_1};\theta_M)}{F(x_{i_1};\boldsymbol{\Psi})}\right)$. Similarly, due to Remark 6 in the Appendix, we have

$$f(\mathbf{w}_r|\tilde{\mathbf{x}}) = \prod_{j=1}^{M} \binom{i_r - i_{r-1} - 1}{w_{r1}, \ldots, w_{rM}} \left(\frac{\pi_j[F_j(x_{i_r};\theta_j) - F_j(x_{i_{r-1}};\theta_j)]}{F(x_{i_r};\boldsymbol{\Psi}) - F(x_{i_{r-1}};\boldsymbol{\Psi})}\right)^{w_{rj}}, \quad (2.6)$$

for each $r = 2, \ldots, k$. Finally, from Remark 7 in Appendix, we have

$$f(\mathbf{w}_{k+1}|\tilde{\mathbf{x}}) = \binom{n - i_k}{w_{k+1\,1}, \ldots, w_{k+1\,M}} \prod_{j=1}^{M} \left(\frac{\pi_j \bar{F}_j(x_{i_k};\theta_j)}{\bar{F}(x_{i_k};\boldsymbol{\Psi})}\right)^{w_{k+1\,j}}. \quad (2.7)$$

From Lemma 1, the complete data log-likelihood function is

$$\begin{aligned} l(\boldsymbol{\Psi}|\tilde{\mathbf{x}}, \boldsymbol{\delta}) \propto \sum_{j=1}^{M} \Big\{ & w_{1j} \log\left[\pi_j F_j(x_{i_1};\theta_j)\right] + w_{k+1\,j} \log\left[\pi_j \bar{F}_j(x_{i_k};\theta_j)\right] \\ & + \sum_{r=1}^{k} z_{rj} \log\left[\pi_j f_j(x_{i_r};\theta_j)\right] \\ & + \sum_{s=2}^{k} w_{sj} \log\{\pi_j[F_j(x_{i_s};\theta_j) - F_j(x_{i_{s-1}};\theta_j)]\} \Big\}. \end{aligned} \quad (2.8)$$

3. EM Algorithm

Here, we use the EM algorithm of Dempster et al. (1977) to obtain $\widehat{\boldsymbol{\Psi}}_{MLE}$ using (2.8). To this end, let $\boldsymbol{\Psi}^{(0)}$ be an initial value for $\boldsymbol{\Psi}$.

E-Step: Given $\tilde{\mathbf{X}} = \tilde{\mathbf{x}}$, the conditional expectation of the complete-data log-likelihood function is $Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(0)}) = E_{\boldsymbol{\Psi}^{(0)}}[l(\boldsymbol{\Psi}|\tilde{\mathbf{x}})]$, where the expectation is taken under $\boldsymbol{\Psi}^{(0)}$. On the $(p+1)$-th iteration, $Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(p)})$ is computed in the E-step, where $\boldsymbol{\Psi}^{(p)}$ is the estimate of $\boldsymbol{\Psi}$ obtained from the p-th iteration.

From (2.4), (2.5), (2.6) and (2.7), we have

$$\tau_{r,j}(\mathbf{\Psi}) = \mathbb{E}\left(Z_{rj}|\tilde{\mathbf{x}}\right) = \frac{\pi_j f_j(x_{i_r};\theta_j)}{f(x_{i_r};\mathbf{\Psi})},\ r = 1,\ldots,k;\ j = 1,\ldots,M. \quad (3.1)$$

$$\beta_{1,j}(\mathbf{\Psi}) = \mathbb{E}\left(W_{1j}|\tilde{\mathbf{x}}\right) = (i_1 - 1)\frac{\pi_j F_j(x_{i_1};\theta_j)}{F(x_{i_1};\mathbf{\Psi})},\ j = 1,\ldots,M. \quad (3.2)$$

$$\beta_{s,j}(\mathbf{\Psi}) = \mathbb{E}\left(W_{sj}|\tilde{\mathbf{x}}\right) = (i_s - i_{s-1} - 1)\frac{\pi_j[F_j(x_{i_s};\theta_j) - F_j(x_{i_{s-1}};\theta_j)]}{[F(x_{i_s};\mathbf{\Psi}) - F(x_{i_{s-1}};\mathbf{\Psi})]},$$

$$s = 2,\ldots,k;\ j = 1,\ldots,M. \quad (3.3)$$

$$\beta_{k+1,j}(\mathbf{\Psi}) = \mathbb{E}\left(W_{k+1\,j}|\tilde{\mathbf{x}}\right) = (n - i_k)\frac{\pi_j \bar{F}_j(x_{i_k};\theta_j)}{\bar{F}(x_{i_k};\mathbf{\Psi})},\ j = 1,\ldots,M. \quad (3.4)$$

Combining these with (2.8), the expectation at the $(p+1)$-th iteration is

$$Q(\mathbf{\Psi}, \mathbf{\Psi}^{(p)}) = Q_1(\boldsymbol{\pi}, \mathbf{\Psi}^{(p)}) + Q_2(\boldsymbol{\xi}, \mathbf{\Psi}^{(p)}), \quad (3.5)$$

$$Q_1(\boldsymbol{\pi}, \mathbf{\Psi}^{(p)}) = \sum_{j=1}^{M} \log \pi_j \left\{ \sum_{r=1}^{k} \tau_{r,j}(\mathbf{\Psi}^{(p)}) + \sum_{s=1}^{k+1} \beta_{s,j}(\mathbf{\Psi}^{(p)}) \right\},$$

$$\begin{aligned} Q_2(\boldsymbol{\xi}, \mathbf{\Psi}^{(p)}) = \sum_{j=1}^{M} \Bigg[& \beta_{1,j}(\mathbf{\Psi}^{(p)}) \log F_j(x_{i_1};\theta_j) + \beta_{k+1,j}(\mathbf{\Psi}^{(p)}) \log \bar{F}_j(x_{i_k};\theta_j) \\ & + \sum_{r=1}^{k} \tau_{r,j}(\mathbf{\Psi}^{(p)}) \log f_j(x_{i_r};\theta_j) \\ & + \sum_{s=2}^{k} \beta_{s,j}(\mathbf{\Psi}^{(p)}) \log\{F_j(x_{i_s};\theta_j) - F_j(x_{i_{s-1}};\theta_j)\} \Bigg]. \end{aligned}$$

M-Step: On the $(p+1)$-th iteration of the M-step, $Q(\mathbf{\Psi}, \mathbf{\Psi}^{(p)})$ is maximized with respect to $\mathbf{\Psi}$ to obtain $\mathbf{\Psi}^{(p+1)}$. From (3.5), the estimate $\hat{\boldsymbol{\pi}}^{(p+1)}$

is updated by maximizing $Q_1(\boldsymbol{\pi}, \boldsymbol{\Psi}^{(p)})$ with respect to $\boldsymbol{\pi}$. Due to the constraint $\sum_{j=1}^{M} \pi_j = 1$, using the Lagrangian multiplier, the mixing proportions π_j, $j = 1, \ldots, M-1$ are updated as

$$\hat{\pi}_j^{(p+1)} = \frac{1}{n} \left\{ \sum_{s=1}^{k} \tau_{s,j}(\boldsymbol{\Psi}^{(p)}) + \sum_{s=1}^{k+1} \beta_{s,j}(\boldsymbol{\Psi}^{(p)}) \right\} \tag{3.6}$$

Using $Q_2(\boldsymbol{\xi}, \boldsymbol{\Psi}^{(p)})$ in (3.5), we obtain $\boldsymbol{\xi}^{(p+1)}$ as the solution of

$$\boldsymbol{\xi}^{(p+1)} = \arg\max_{\boldsymbol{\xi}} Q_2(\boldsymbol{\xi}, \boldsymbol{\Psi}^{(p)}). \tag{3.7}$$

Finally, $\widehat{\boldsymbol{\Psi}}_{MLE}$ of FMM (2.1) is computed iteratively through the E- and M-step until the algorithm converges.

3.1 Modified EM algorithm

In the algorithm proposed above, each M-step requires finding a solution to (3.7). Thus updating $\boldsymbol{\xi}$ is cumbersome, computationally expensive and affects the convergence rate of the algorithm. This intractability is due to the terms of $\frac{\partial}{\partial \boldsymbol{\xi}} \log F_j(x_{(i_1)}; \theta_j)$, $\frac{\partial}{\partial \boldsymbol{\xi}} \log(1 - F_j(x_{(i_k)}; \theta_j))$ and $\frac{\partial}{\partial \boldsymbol{\xi}} \log\{F_j(x_{(i_s)}; \theta_j) - F_j(x_{(i_{s-1})}; \theta_j)\}$ in the log-likelihood function. When the cdf of component densities does not have a closed form, which is the case for most commonly used distributions, the dependence structures among order statistics makes the computations very extensive and time consuming. To tackle the problem, Johnson and Mehrotra (1972) and Mehrotra

and Nanda (1974) proposed a modification technique in which the expectation of the likelihood function is maximized to obtain MLE. Recently Hatefi et al. (2015) took advantage of this modified approach for FMM analysis under different RBS designs. Using the properties of the RBS, where order statistics are independent, they showed that M-step for $\boldsymbol{\xi}$ in EM algorithm reduces to the M-step in the usual SRS EM-algorithm. Unfortunately, due to the dependence structure among the order statistics, this is not the case in the EM-algorithm under correlated order statistics. Based on their work, we propose computing the M-step of the EM-algorithm for estimation of $\boldsymbol{\xi}$ using M-step of $\boldsymbol{\xi}$ in an EM-algorithm for SRS data. Despite the similarity in updating $\boldsymbol{\xi}$, note that the observations are order statistics of the FMMs. Accordingly, instead of equation (3.7), the following modified estimating equation is used to update $\boldsymbol{\xi}$

$$\hat{\boldsymbol{\xi}}^{(p+1)} = \arg\max_{\boldsymbol{\xi}} \sum_{s=1}^{k} \sum_{j=1}^{M} \left\{ \tau_{s,j}(\boldsymbol{\Psi}^{(p)}) \log f_j(x_{i_s}; \theta_j) \right\}. \tag{3.8}$$

where $\tau_{s,j}(\boldsymbol{\Psi}^{(p)})$ is defined at (3.1). This updating step for $\boldsymbol{\xi}$ is the same as that under SRS data, but we still take advantage of the information in the order statistics and their latent variables when updating the mixing proportions in each step. This indirectly affects the estimation of $\boldsymbol{\xi}$.

4. Classification

Once the parameters of FMM are estimated, we can determine the component membership of each observation. According to the characteristics of the order statistics from FMM, we propose different model-based classification criteria. These criteria not only enable us to determine the component membership of the observations, but also assist us through a probabilistic view-point to make inference about rarely observed component(s) in FMMs. We first focus on the classification of a sample of order statistics from FMM.

Suppose that we have observed $X_{(r)} = x_{(r)}$. To classify $x_{(r)}$, we estimate its component membership vector $\mathbf{Z}_r = (Z_{r1}, \ldots, Z_{rM})$ by $\widehat{\mathbf{Z}}_r$ where

$$\hat{Z}_{rj} = \begin{cases} 1, & \text{if } j = \text{argmax}_h \eta_h(x_{(r)}; \mathbf{\Psi}), \\ 0, & \text{otherwise,} \end{cases}$$

for $j = 1, \ldots, M$, and $\eta_h(x_{(r)}; \mathbf{\Psi}) = \mathbb{P}(Z_{rh} = 1 | x_{(r)}; \mathbf{\Psi})$. From (2.4), the posterior distribution of $\mathbf{Z}_r$ given $X_{(r)} = x_{(r)}$ is given by

$$\mathbb{P}(\mathbf{Z}_r = \mathbf{z}_r | x_{(r)}) = \binom{1}{z_{r1}, \ldots, z_{rM}} \prod_{h=1}^{M} \left\{ \frac{\pi_h f_h(x_{(r)}; \theta_h)}{f(x_{(r)}; \mathbf{\Psi})} \right\}^{z_{rh}},$$

so

$$\eta_h(x_{(r)}; \mathbf{\Psi}) = \frac{\pi_h \, f_h(x_{(r)}; \theta_h)}{f(x_{(r)}; \mathbf{\Psi})}. \tag{4.1}$$

The posterior probabilities $\eta_h(x_{(r)}; \mathbf{\Psi})$ are then estimated by $\eta_h(x_{(r)}; \hat{\mathbf{\Psi}}_{MLE})$. Using the classifier (4.1), we assign the observations into a component with

the highest estimated posterior probability. It is interesting to note that the expression we obtained in (4.1) as the posterior probability of component membership of each order statistic is equal to the commonly used expression for the SRS design. However, the parameters are estimated using the order statistics of FMM (2.1).

The following remark describes the classification of unobserved X_l given observed order statistics X_r where $l \leq r$, with other classification scenarios summarized as Remarks 8 and 9 in the Appendix.

Remark 1. Given $X_{(r)} = x_{(r)}$ and its label $\mathbf{Z}_{(r)} = \mathbf{z}_{(r)}$, suppose we are now interested in classifying an unobserved order statistic $X_{(l)}$ for $l \leq r$. To this end, the component membership vector $\mathbf{Z}_l = (Z_{l1}, \ldots, Z_{lM})$ can be estimated by $\widehat{\mathbf{Z}}_l$ where

$$\hat{Z}_{lj} = \begin{cases} 1, & \text{if } j = \text{argmax}_h \alpha_h(x_{(l)}; \mathbf{\Psi}), \\ 0, & \text{otherwise,} \end{cases}$$

and $\alpha_h(x_{(l)}; \mathbf{\Psi}) = \mathbb{P}(Z_{lj} = 1 | x_{(r)}, \mathbf{z}_{(r)}; \mathbf{\Psi})$. From Remark 2 in Appendix, the posterior distribution of $\mathbf{Z}_l$ is derived by

$$\mathbb{P}(\mathbf{Z}_l = \mathbf{z}_l | \mathbf{Z}_r = \mathbf{z}_r, x_{(r)}) = \binom{1}{z_{l1}, \ldots, z_{lM}} \prod_{h=1}^{M} \left\{ \frac{\pi_h F_h(x_{(r)}; \theta_h)}{F(x_{(r)}; \mathbf{\Psi})} \right\}^{z_{lh}},$$

and consequently $\alpha_h(x_{(r)}; \mathbf{\Psi}) = \pi_h \, F_h(x_{(r)}; \theta_h) / F(x_{(r)}; \mathbf{\Psi})$. In other words, given the observed value y for the the r-th order statistic $X_{(r)}$ that was

selected out of a sample of size n from the FMM, the missing (unselected) order statistics that are smaller than y are classified into the j-th component of FMM, if $\alpha_j(y;\hat{\mathbf{\Psi}}) > \alpha_h(y;\hat{\mathbf{\Psi}})$ for all $h = 1,\ldots,M; j \neq h$. Now we would like to investigate how we can use the properties of order statistics for FMMs with rarely observed component(s). In other words, how likely we expect to observe at least m observations from these rare components. These probabilities are studied in Lemmas 4, 5, 6 whose proofs are provided in the Appendix. We first state the following result from David and Nagaraja (1981).

Lemma 3. *Let X be a random variable with cdf $F(\cdot;\mathbf{\Psi})$. Then*

$$\sum_{i=r}^{n} \binom{i}{n} [F(x;\mathbf{\Psi})]^i [\bar{F}(x;\mathbf{\Psi})]^{n-i} = \mathbb{I}_{F(x;\mathbf{\Psi})}(r, n-r+1), \tag{4.2}$$

where $\mathbb{I}_{F(x;\mathbf{\Psi})}(r,n-r+1) = \frac{1}{B(r,n-r+1)} \int_0^{F(x;\mathbf{\Psi})} t^{r-1}(1-t)^{n-r}dt$, and $B(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$.

Lemma 4. *Let $X_{(r)} = x_r$ be the observed r-th order statistic from FMM (2.1) based on a random sample of size n. For $m = 1,\ldots,r-1$, let $T^1_{m,j}$ denote the event of observing at least m sample points less than $X_{(r)}$ from component j, then we have $\mathbb{P}(T^1_{m,j}|x_r) = \mathbb{I}_{G_1(x_r)}(m, r-m)$, where $G_1(x_r) = \pi_j F_j(x_r;\theta_j)/F(x_r;\mathbf{\Psi})$ and $j = 1,\ldots,M$. In addition, let S^1_j denote the event of observing no sample points less than $X_{(r)}$ from component j, then*

we have $\mathbb{P}(S_j^1|x_r) = 1 - \mathbb{I}_{G_1(x_r)}(1, r-1)$.

Lemma 5. *Let* $X_{(r)} = x_r$ *and* $X_{(l)} = x_l$ *be the observed* r*-th and* l*-th order statistics* $r < l$ *from FMM* (2.1) *from a sample of size* n*. Let* $T_{m,j}^2$ *denote the event of observing at least* m *sample points between* $X_{(r)}$ *and* $X_{(l)}$ *from component* j*, then we have* $\mathbb{P}(T_{m,j}^2|x_r, x_l) = \mathbb{I}_{G_2(x_r,x_l)}(m, l-r-m)$*, for* $m = 1, \ldots, l-r-1$*, where* $G_2(x_r, x_l) = \pi_j[F_j(x_l;\theta_j) - F_j(x_r;\theta_j)]/[F(x_l;\mathbf{\Psi}) - F(x_r;\mathbf{\Psi})]$ *and* $j = 1, \ldots, M$*. Therefore, let* S_j^2 *denote the event of observing no sample points between* $X_{(r)}$ *and* $X_{(l)}$ *from component* j*, then we have* $\mathbb{P}(S_j^2|x_r, x_l) = 1 - \mathbb{I}_{G_2(x_r,x_l)}(1, l-r-1)$.

Lemma 6. *Let* $X_{(l)} = x_l$ *be the observed* l*-th order statistic from FMM* (2.1) *based on a random sample of size* n*. For* m*;* $m = 1, \ldots, n-l-1$*, let* $T_{m,j}^3$ *denote the event of observing at least* m *sample points bigger than* $X_{(l)}$ *from component* j*, then we have* $\mathbb{P}(T_{m,j}^3|x_l) = \mathbb{I}_{G_3(x_l)}(m, n-l-m+1)$ *where* $G_3(x_l) = \pi_j \bar{F}_j(x_l;\theta_j)/\bar{F}(x_l;\mathbf{\Psi})$ *and* $j = 1, \ldots, M$*. Further, let* S_j^3 *denote the event of observing no sample points bigger than* $X_{(l)}$ *from component* j*, then we have* $\mathbb{P}(S_j^3|x_l) = 1 - \mathbb{I}_{G_3(x_l)}(1, n-l)$.

As mentioned in Section 1, in many environmental, ecological and medical studies, measuring the variable of interest is difficult and time-consuming; however, rank information can be obtained easily; for example, the age determination of fish based on the length information as described

in introduction. Hatefi et al. (2015) exploited properties of ranked set sampling (RSS) design under perfect ranking to analyze the age of fish based on length frequency data. To obtain a sample of k fish, a simple random sample of k^2 fish are first selected and they are randomly divided into k sets of size k. Then in each set, fish are ranked based on their length and finally the i-th smallest fish from set i is selected for age determination. In the following example, we use Lemma 6, for a perfect RSS (there is no ranking error in sampling process) as an example of the order statistics of FMMs.

Example 1. Consider a perfect RSS with set size $H = 10$ from mixture of two normal distributions with $\mathbf{\Psi} = \{\pi, \mu_1, \mu_2, \sigma_1, \sigma_2\} = \{0.8, 4.87, 8, 1, 2\}$. Figure 1 demonstrates the probability of observing at least one observation from the second component. For example, Given $x_{(5)} = 4$, the probability of observing at least $m = 3$ units out of $H = 10$ sampling units, from a rare population (second component with $\pi = 0.2$) is 0.0856. It is observed from Figure 1, while rank is considered fixed, as the value of x increases, the probability of observing sample(s) from the rare component increases. It also illustrates that once x is considered fixed, as the rank of x increases, the probability of observing sample(s) from the rare event decreases.

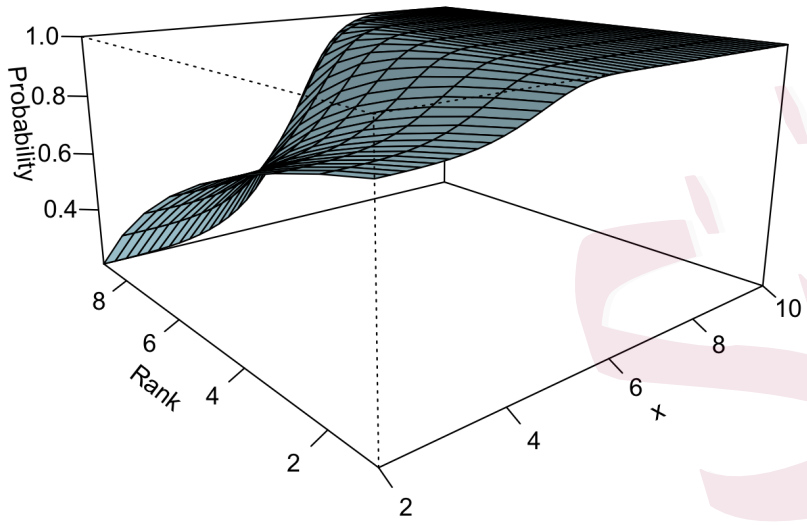

Figure 1: The probability of observing at least one observation from the second component when the set size is 10.

5. Statistical Learning with Order Statistics

In this section, we study how the notion of order statistics can be incorporated into supervised and unsupervised learning in the context of FMMs. As in the previous section, we use the properties of order statistics to make inference about FMMs in the context of unsupervised learning where the information about the component membership of order statistics is not available. As the cost of obtaining k order statistics is the same as that of ordering the entire sample, order statistics under unsupervised learning are

studied here for the sake of completeness in the context of estimation and classification and consistency of the results. This enables us to better compare the performance of proposed methods with their counterparts under the supervised learning, particularly in the settings where measuring the labeled data is difficult. Unlike unsupervised learning, in this section, we study the problem of order statistics of FMMs in the context of supervised learning. In this case, both the measured values of order statistics and their component memberships are available.

In Subsection 5.1, we revisit the results of Section 2 for order statistics of the FMMs in an unsupervised learning setting. We then study the problem of order statistics of the FMM in the context of supervised learning. Suppose $\mathbf{X} = (X_1, \ldots, X_k)$ represents a collection of unlabeled SRS data of size k from FMM (2.1). In the case of labeled SRS data, for each observation X_i, $i = 1, \ldots, k$, let $\mathbf{Z}_i^* = \{z_{i1}^*, \ldots, z_{iM}^*\}$ be the observed label such that $z_{ij}^* = 1$ if the X_i is from component j otherwise z_{ij}^* is zero.

5.1 Unsupervised Learning with OS from FMMs

Suppose we only have access to the unlabeled SRS data $\mathbf{x} = (x_1, \ldots, x_k)$, hence, the likelihood function becomes $L_{un}(\mathbf{\Psi}|\mathbf{x}) = \prod_{i=1}^{k} \sum_{j=1}^{M} \pi_j f_j(x_i; \theta_j)$. As in Section 2, we introduce latent variables $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iM}); i =$

$1, \ldots, k$ for each x_i so that $Z_{ij} = 1$ if x_i comes from component j of the FMM and $Z_{ij} = 0$ otherwise. Now let $\mathcal{Y}_{un} = (\mathbf{X}, \mathbf{Z})$ denote the complete data with likelihood function

$$L_{un}(\boldsymbol{\Psi}|\mathcal{Y}_{un}) = \prod_{i=1}^{k}\prod_{j=1}^{M} \{\pi_j f_j(y_i;\theta_j)\}^{z_{ij}}\,. \tag{5.1}$$

As in Section 3, we obtain ML estimates of the parameters using the EM algorithm. The conditional expectation of $Z_{ij}|\mathbf{y}$, computed in the E-step, is used in the $(p+1)$-th step to update $\boldsymbol{\Psi}_{un}^{(p+1)} = (\pi_{un}^{(p+1)}, \boldsymbol{\xi}_{un}^{(p+1)})$ by

$$\hat{\pi}_{un,j}^{(p+1)} = \frac{1}{k}\sum_{i=1}^{k} \tau_{r,j}(\boldsymbol{\Psi}^{(p)}), \qquad j = 1, \ldots, M, \tag{5.2}$$

$$\hat{\boldsymbol{\xi}}_{un}^{(p+1)} = \arg\max_{\boldsymbol{\xi}} \sum_{i=1}^{k}\sum_{j=1}^{M} \left\{\tau_{r,j}(\boldsymbol{\Psi}^{(p)}) \log f_j(y_i;\theta_j)\right\}, \tag{5.3}$$

where $\tau_{r,j}(\boldsymbol{\Psi}^{(p)}) = \mathbb{E}(Z_{rj}|\mathbf{y}); r = 1, \ldots, k$.

Let $\tilde{\mathbf{X}}_{ou} = \{X_{(i_1)}, \ldots, X_{(i_k)}\}$ be the collection of order statistics of unlabeled data $\mathbf{X}$ in the sample of size n. Let $\mathcal{Y}_{ou} = (\tilde{\mathbf{X}}_{ou}, \boldsymbol{\Delta})$ denote the complete order statistics consisting of unlabeled order statistics and their latent variables. According to Lemma 1, the likelihood function based on $\mathcal{Y}_{ou}$ can be written as

$$L(\boldsymbol{\Psi}|\mathcal{Y}_{ou}) \propto L(\boldsymbol{\Psi}|\mathcal{Y}_{un})\, \kappa(\boldsymbol{\Psi}|\mathcal{Y}_{ou}), \tag{5.4}$$

where

$$\kappa(\boldsymbol{\Psi}|\mathcal{Y}_{ou}) = \prod_{j=1}^{M} \Big\{ \{\pi_j F_j(y_{(i_1)};\theta_j)\}^{w_{1j}} \{\pi_j \bar{F}_j(y_{(i_k)};\theta_j)\}^{w_{m+1,j}} \times \prod_{s=2}^{k} \{\pi_j [F_j(y_{(i_s)};\theta_j) - F_j(y_{(i_{s-1})};\theta_j)]\}^{w_{sj}} \Big\}. \tag{5.5}$$

From (5.1), it is apparent that the $\kappa(\boldsymbol{\Psi}|\mathcal{Y}_{ou})$ is the contribution of k order statistics into unsupervised learning of FMMs.

5.2 Supervised Learning with OS of FMMs

In this subsection, we focus on FMM analysis with labeled data. For SRS supervised learning, we estimate the parameters based on the labeled data. The likelihood function based on these observations is

$$L_{us}(\boldsymbol{\Psi}|\mathbf{x},\mathbf{z}^*) = \prod_{i=1}^{k}\prod_{j=1}^{M} \{\pi_j f_j(x_i;\theta_j)\}^{z^*_{ij}}. \tag{5.6}$$

Using (5.6), the ML estimate $\hat{\boldsymbol{\Psi}}_{us}$ is

$$\hat{\pi}_{us,j} = \frac{1}{k}\sum_{i=1}^{k} z^*_{ij}, \tag{5.7}$$

$$\hat{\theta}_j = \arg\max_{\theta_j} \sum_{i=1}^{k} \log f_j(x_i;\theta_j), \quad j = 1,\ldots,M. \tag{5.8}$$

Here, we show how one can exploit the properties of order statistics to make inference for FMMs based on labeled data. Let $\tilde{\mathbf{X}}_{os} = \{X_{(i_1)},\ldots,X_{(i_k)}\}$ be the collection of k order statistics of labeled data $\mathbf{X}$ of the sample of size

n with labels $\mathbf{Z}^* = \{Z_1^*, \ldots, Z_k^*\}$. Using the pdf of order statistics, the likelihood function based on $(\tilde{\mathbf{X}}_{os}, \mathbf{Z}^*)$ is

$$\begin{aligned} L_{os}(\boldsymbol{\Psi}|\tilde{\mathbf{x}}_{os}, \mathbf{z}^*) \propto & \left\{F(x_{(i_1)}; \boldsymbol{\Psi})\right\}^{i_1-1} \left\{\bar{F}(x_{(i_k)}; \boldsymbol{\Psi})\right\}^{n-i_k} \\ & \times \prod_{s=2}^{k} \left\{F(x_{(i_s)}; \boldsymbol{\Psi}) - F(x_{(i_{s-1})}; \boldsymbol{\Psi})\right\}^{i_s - i_{s-1}-1} \\ & \times \prod_{r=1}^{k} \prod_{j=1}^{M} \left\{\pi_j f_j(x_{(i_r)}; \theta_j)\right\}^{z_{rj}^*}. \end{aligned} \tag{5.9}$$

In order to obtain the ML estimate of $\boldsymbol{\Psi}$, we introduce the latent vectors $\mathbf{W_s} = (W_{s1}, \ldots, W_{sM}); s = 1, \ldots, k+1$. Let $\mathcal{Y}_{os} = (\tilde{\mathbf{X}}_{os}, \mathbf{Z}^*, \mathbf{W})$ denote the complete labelled order statistics. Similar to Lemma 1, the complete likelihood function version of (5.9) is given by

$$L(\boldsymbol{\Psi}|\mathcal{Y}_{os}) \propto L(\boldsymbol{\Psi}|\mathcal{Y}_{us})\, \kappa(\boldsymbol{\Psi}|\mathcal{Y}_{os}), \tag{5.10}$$

where $\kappa(\boldsymbol{\Psi}|\mathcal{Y}_{os})$ is defined in (5.5) by replacing y_{i_j}s with $x_{(i_j)}$s. From (5.10), it is apparent that $\kappa(\boldsymbol{\Psi}|\mathcal{Y}_{os})$ shows the contribution of k order statistics of the sample of size n into the supervised FMM. Now we estimate the parameters of the FMM using the EM-algorithm presented in Section 3. The E-step only requires the conditional expectation of the latent variables $\mathbf{W}_s; s = 1, \ldots, n$ given $\tilde{\mathbf{x}}_{os}, \mathbf{z}^*$. As in Section 3, using (3.2), (3.3) and (3.4), the parameters are updated on the $(p+1)$-th step by

$$\hat{\pi}_{os,j}^{(p+1)} = \frac{1}{n} \left\{ \sum_{s=1}^{k} z_{sj}^* + \sum_{s=1}^{k+1} \beta_{s,j}(\boldsymbol{\Psi}^{(p)}) \right\}, \tag{5.11}$$

where $j = 1, \ldots, M-1$ and on the $(p+1)$-th iteration of M-step, the estimate of component parameters $\boldsymbol{\xi}_{os}^{(p+1)}$ are updated by

$$\boldsymbol{\xi}_{os}^{(p+1)} = \arg\max_{\boldsymbol{\xi}} Q_{os}(\boldsymbol{\xi}, \boldsymbol{\Psi}^{(p)}). \tag{5.12}$$

where

$$\begin{aligned} Q_{os}(\boldsymbol{\xi}, \boldsymbol{\Psi}^{(p)}) = \sum_{j=1}^{M} \Big\{ & \beta_{1,j}(\boldsymbol{\Psi}^{(p)}) \log F_j(x_{i_1}; \theta_j) + \beta_{k+1,j}(\boldsymbol{\Psi}^{(p)}) \log \bar{F}_j(x_{i_k}; \theta_j) \\ & + \sum_{r=1}^{k} z_{rj}^{*} \log f_j(x_{i_r}; \theta_j) \\ & + \sum_{s=2}^{k} \beta_{s,j}(\boldsymbol{\Psi}^{(p)}) \log[F_j(x_{i_s}; \theta_j) - F_j(x_{i_{s-1}}; \theta_j)] \Big\}. \end{aligned}$$

The E- and M-steps are alternated repeatedly until the algorithm converges.

6. Numerical Studies

In this section, we empirically study the performance of the MLEs of FMM parameters under various order statistics designs $D_i, i = 1, \ldots, 6$, as in Table 1. In all these designs, the original simple random sample size is assumed to be $n = 30$, where we only observe k order statistics, $k \in \{6, 8, 10\}$. We select D_i's such that the performance of $\boldsymbol{\Psi}_{MLE}$ can be evaluated under different scenarios including right and left censoring schemes (D_1, D_2), modified version of maxima-minima nominated sampling (D_3, D_4) as well as systematic sampling (D_5). The MLEs of parameters of FMMs are com-

puted assuming we have labeled order statistics, unlabeled order statistics, labeled SRS data and unlabeled SRS data. We used the modified EM algorithm to compute $\mathbf{\Psi}_{MLE}$. The underlying FMM is assumed to be a mixture of two univariate normal distributions,

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

with parameters $\mathbf{\Psi} = \{\pi, \mu_1, \mu_2, \sigma\}$. Due to the key role of mixing proportion parameters in mixture modelling, we investigate two simulation studies. The first simulation study, described in Subsection 6.1, investigates estimation of mixing proportion, while the component parameters are assumed to be known. The second study, provided in the Appendix, is devoted to the estimation of all the parameters of the model. We investigate the performance of estimation and classification procedures based on designs D_i and compare this with the case where observations are simple random samples. It is worth emphasizing that we do not necessarily motivate to use order statistics for finite mixture modeling as a sampling scheme to replace SRS but as a natural setting that happens in many real world applications. The goal is to show how the rank information provided by different collections of order statistics can affect the estimation and classification processes. To generate observations using D_i's, for each simulation, we take a sample of size $n = 30$ from (6.1). Ranking observations, we then select order statistics

according to the designs of Table 1. Under unsupervised approach, we only consider the value of the selected order statistics, while in supervised approach, we observe both the selected order statistics and their component memberships.

Table 1: Various collections of order statistics.

Design	Collection of Order Statistics	Experiment (k=size)
D_1	{1, 2, 3, 4, 5, 6}	Right censored data (6)
D_2	{23, 24, 25, 26, 27, 28, 29, 30}	Left censored data (8)
D_3	{1, 2, 3 , 28, 29, 30}	Modified MMN sample (6)
D_4	{1, 2, 3, 4, 5, 26, 27, 28, 29, 30}	Modified MMN sample (10)
D_5	{1, 5, 10, 20, 25, 30}	Systematic selection (6)

6.1 Simulation Study 1

We first consider estimating π and evaluating the classification performance when the component parameters of the FMM are assumed to be known. Using Table 1, we generate samples from model (6.1). We consider $(\mu_1, \mu_2, \sigma) = (9.01, 11.70, 1.15)$ and $\pi \in \{0.35, 0.50, 0.60, 0.67, 0.80\}$ such that the component parameters are chosen exactly the same as component parameters of Spot data analyzed in Section 7. The modified EM algorithm, as described at Subsection 3.1, is carried out 5000 times with initial

value of 0.5 for π with stopping criteria $|\pi^{(k+1)} - \pi^{(k)}| < 10^{-6}$.

Tables 2 and 3 provide the biases, square root of mean squared errors ($\sqrt{MSE}$), classification precisions (CLP%) and convergence rates (CVR%) for all estimation procedures. The classification precision rate (CLP%) is the average proportion of correct classification rates over 5000 simulations. The simulation studies are devised so that we have access to the true component membership of sampling units under all estimation procedures. Comparing true and predicted memberships of the test data, we compute the correct classification rate of the classifiers for each estimator in each simulation. The rate of convergence (CVR%) is calculated as the average number of times that the estimation procedure converged over 5000 replications. Comparing ML estimates of π under each design D_i, we clearly observe the impact of various collections of order statistics on estimation and classification procedures. For instance, from Table 2 when $\pi = 0.8$, design D_1 practically fails to capture the rare event (i.e., second component) so that the convergence rate of estimation procedure is about 1%. On the other hand, using the collection of upper order statistics (Design D_2) guarantees observing data from the rare component and consequently revives the convergence rate of estimation procedures by 93%.

The relative efficiency (RE) of the proposed estimator depends on the

sampling design D_i. The estimator based on the design D_5 provides substantial amount of improvement over the MLE of SRS sampling design. For example, the relative efficiencies RE = MSE(SRS)/MSE(D_5) from Table 2 are $(0.18^2/0.11^2 =)$2.7, 2.98, 2.7, 2.7, 3.16 for $\pi = 0.35, 0.50, 0.60, 0.67, 0.8$, respectively. These empirical results show that the MLE based on design D_5 is at least 2.7 times more efficient than the corresponding SRS estimator. The same efficiencies under unsupervised learning in Table 3 are $4.76, 4.69, 4.34, 4.76, 4.41$. These RE values indicate that design D_5 works much better with unsupervised learnings.

7. Data Analysis

Age structure of fish is very important in various fishery studies, as it provides valuable information about age of recruitment, maturity, etc. Estimation of the age structure of fish plays a key role in stock assessments and dynamics of fish population. In this section, we study the problem of age determination of Spot as a short-lived fish species, using the length frequency data. Due to the both commercial and recreational purposes and food source for other fish, Spot can be categorized as one of the most important and frequently caught fish in the Chesapeake Bay area. The existence of several environmental threads towards such a short-lived fish

Table 2: Bias, $\sqrt{MSE}$, (CLP%) and (CVR%) under supervised learning based on designs of Table 1, against those of SRS data of the same size, when π is the only unknown parameter of model (6.1).

		OS					SRS				
	π	0.35	0.50	0.60	0.67	0.80	0.35	0.50	0.60	0.67	0.80
D_1	Bias	-0.09	-0.15	-0.19	-0.24	-0.30	0.02	-0.00	-0.01	-0.03	-0.07
	$\sqrt{MSE}$	0.15	0.24	0.28	0.35	0.44	0.18	0.19	0.18	0.17	0.17
	CLP%	87.7	86.6	87.1	84.7	87.7	85.2	84.0	84.6	85.5	87.7
	CVR%	31.1	9.5	4.6	2.8	1.2	92.1	97.0	94.8	91.1	72.6
D_2	Bias	0.17	0.11	0.07	0.04	0.00	0.01	0.00	-0.00	-0.02	-0.04
	$\sqrt{MSE}$	0.26	0.18	0.13	0.10	0.08	0.16	0.17	0.17	0.16	0.13
	CLP%	87.4	87.4	87.5	88.4	90.8	86.0	85.2	85.6	86.2	88.5
	CVR%	7.3	21.1	40.8	61.8	93.8	97.0	99.3	98.2	96.1	83.9
D_3	Bias	0.04	-0.00	-0.03	-0.04	-0.05	0.03	-0.00	-0.02	-0.03	-0.07
	$\sqrt{MSE}$	0.16	0.15	0.16	0.16	0.15	0.18	0.19	0.18	0.18	0.16
	CLP%	87.7	86.9	87.2	88.1	90.4	85.2	84.1	84.6	85.5	87.9
	CVR%	100	100	100	99.9	99.5	92.4	96.7	94.9	90.8	74.0
D_4	Bias	0.02	-0.00	-0.02	-0.03	-0.02	0.01	-0.00	-0.00	-0.01	-0.02
	$\sqrt{MSE}$	0.13	0.13	0.13	0.13	0.11	0.15	0.16	0.15	0.15	0.12
	CLP%	88.2	87.2	87.4	88.2	90.5	86.4	85.6	85.8	86.6	88.9
	CVR%	99.9	100	100	100	99.7	98.7	99.8	99.4	98.3	88.4
D_5	Bias	0.00	-0.00	-0.00	-0.00	-0.00	0.03	0.00	-0.02	-0.04	-0.07
	$\sqrt{MSE}$	0.11	0.11	0.11	0.11	0.09	0.18	0.19	0.18	0.18	0.16
	CLP%	88.0	87.3	87.8	88.4	90.7	85.1	84.1	84.7	85.3	87.8
	CVR%	99.8	100	99.9	99.8	97.7	92.6	97.1	94.5	90.8	73.3

Table 3: Bias, $\sqrt{MSE}$, (CLP%) and (CVR%) under unsupervised learning based on designs of Table 1, against those of SRS data of the same size, when π is the only unknown parameter of model (6.1).

		OS					SRS				
	π	0.35	0.50	0.60	0.67	0.80	0.35	0.50	0.60	0.67	0.80
D_1	Bias	0.04	0.04	0.01	-0.01	-0.04	0.00	0.00	-0.00	0.00	-0.01
	$\sqrt{MSE}$	0.17	0.19	0.18	0.18	0.19	0.24	0.26	0.25	0.24	0.21
	CLP%	87.3	85.2	85.0	85.6	87.0	82.6	81.5	82.2	83.1	86.0
	CVR%	96.9	91.0	83.1	76.8	68.7	99.2	99.2	99.2	99.3	98.9
D_2	Bias	-0.02	-0.04	-0.04	-0.02	-0.01	0.00	-0.00	-0.00	-0.00	-0.01
	$\sqrt{MSE}$	0.17	0.18	0.16	0.15	0.11	0.21	0.22	0.22	0.21	0.18
	CLP%	85.6	86.0	87.0	87.8	90.3	84.8	84.3	84.5	84.9	87.3
	CVR%	86.8	95.8	98.3	99.5	99.8	99.5	99.6	99.4	99.5	99.2
D_3	Bias	0.03	-0.00	-0.03	-0.04	-0.04	0.01	-0.01	-0.00	-0.00	-0.01
	$\sqrt{MSE}$	0.16	0.15	0.16	0.16	0.15	0.25	0.26	0.25	0.24	0.21
	CLP%	87.6	86.9	87.2	87.9	90.1	82.9	82.0	82.3	83.5	86.1
	CVR%	99.9	100	100	99.9	99.7	99.2	99.5	99.5	99.2	99.1
D_4	Bias	0.02	0.00	-0.02	-0.02	-0.02	0.00	-0.00	-0.00	-0.00	-0.00
	$\sqrt{MSE}$	0.14	0.13	0.14	0.14	0.12	0.19	0.20	0.20	0.19	0.16
	CLP%	88.0	87.2	87.3	88.1	90.2	86.1	85.3	85.4	86.2	88.2
	CVR%	99.9	100	100	100	99.8	99.7	99.8	99.6	99.4	99.2
D_5	Bias	0.01	-0.00	-0.00	0.00	0.00	0.00	0.00	-0.00	0.00	-0.01
	$\sqrt{MSE}$	0.11	0.12	0.12	0.11	0.10	0.24	0.26	0.25	0.24	0.21
	CLP%	88.0	87.3	87.7	88.2	90.3	82.8	81.7	82.5	83.0	86.0
	CVR%	99.9	100	99.9	100	99.7	99.4	99.4	99.2	99.3	99.1

species (Thomas, 1990; Rickabaugh and Capossela, 2011) has increased the importance of analyzing the age structures of Spot.

Recently, there have been several attempts to use different sampling designs based on ranks and order statistics in fishery surveys. These fishery studies include the determination of the mercury level of fish (Nourmohammadi et al., 2015) stock abundance of fish (Wang et al., 2009), as well as RBS designs for age structure determination (Hatefi et al., 2015).

Here, we employ ML estimation for parameters of FMM in a fishery study to determine the age structure of Spot fish. Due to the cost associated with age determination, sometimes a large sample of caught fish is first examined and a subsample of fish is submitted for age determination. As the length of the fish is correlated with the fish age, length is often used as a concomitant to select the final sample. In this section, we consider the length and age determined by otoliths of 403 Virginia-Chesapeake Bay Spot as our population of interest. The data set is available online in the FSAdata package (Ogle, 2013). In this study, the interest is centred in two classes of Spot. The classes include ages 0 and 1 year, which are sexually immature and usually smaller, and 2 years and older fish, which are sexually mature and usually longer. Statistical analysis of the two groups is important as the longer fish group play a vital role in current reproductivity of

population, and the smaller fish group influence dynamics and reproduction of the population in future. Hatefi et al. (2015) showed that the length distribution of Spot is well-modeled by a mixture of two normal distributions with parameters $\mathbf{\Psi} = (\pi, \mu_1, \mu_2, \sigma) = (0.67, 9.01, 11.70, 1.15)$.

We perform a simulation study with 5000 repetitions by generating samples using two common approaches for selecting the final sample to send to the lab. We generate samples of size $n = 30$ and then select the following ordered elements (rank collections) for each sample for age determination: The 30 fish in the original sample are modelled according to their length, which is readily obtained. These collections include $D_1^* = \{1, 4, 7, 10, 13, 16, 19, 22, 25, 28\}$, $D_2^* = \{2, 5, 8, 11, 14, 17, 20, 23, 26, 29\}$, $D_3^* = \{3, 6, 9, 12, 15, 18, 21, 24, 27, 30\}$, $D_4^* = \{5, 10, 15, 20, 25\}$, and finally $D_5^* = \{1, 5, 10, 15, 20, 25, 30\}$. Now, we employ the proposed methods for estimation and classification of the observation to determine the age structure of Spot. We study the effect of different collections of order statistics in observed samples using D_i^*, $i = 1, 2, 3, 4, 5$.

Table 4 and 6 present the bias and square root of the MSE for the estimates of $\mathbf{\Psi}$ based on D_i^*'s designs under supervised and unsupervised learning approaches, respectively. In addition, Tables 5 and 7 present the different computational aspects of the estimation procedures in the analysis

of the Spot dataset. The estimate $\hat{\pi}_{MLE}$ using either labeled or unlabeled order statistics almost outperforms their SRS-based estimate. This is because $\hat{\pi}_{MLE}$ under order statistics approaches take full and direct advantages of rank information of order statistics. Although estimation of the component parameters of the FMM based on order statistics, using the modified EM algorithm, can't take full advantage of rank information, indirectly through $\hat{\pi}$ updates, it uses rank information. It is seen in Table 5 and Table 7, in which estimation procedures under both supervised and unsupervised approaches outperform their SRS-counterparts in classification precision as well as convergence rate.

8. Summary and Concluding Remarks

In this manuscript, we propose estimation and classification methods based on order statistics of FMMs. Our paper is uniquely different both in focus and structure from the two recent papers deal with order statistics in FMMs, Hatefi et al. (2014, 2015). The key aim of this paper is to develop statical inference for classifying the labeled and/or unlabeled current or future observations based on correlated order statistics. In contrast, in Hatefi et al. (2014, 2015), the goal is to estimate the parameters of FMM and to classify the the observations into subpopulations using independent

Table 4: Bias, $\sqrt{MSE}$ of Spot Data under supervised learning approach based on designs $D_i^*; i = 1, \ldots, 5$ against those of SRS data of the same size.

		OS				SRS			
		π	μ_1	μ_2	σ	π	μ_1	μ_2	σ
D_1^*	Bias	-0.03	-0.13	-0.15	-0.08	-0.00	-0.01	-0.01	-0.16
	$\sqrt{MSE}$	0.11	0.36	0.56	0.23	0.14	0.45	0.69	0.32
D_2^*	Bias	-0.01	-0.01	0.04	-0.12	-0.01	0.01	-0.01	-0.16
	$\sqrt{MSE}$	0.10	0.32	0.52	0.26	0.14	0.45	0.67	0.32
D_3^*	Bias	0.01	0.09	0.26	-0.11	-0.01	-0.01	-0.00	-0.16
	$\sqrt{MSE}$	0.11	0.34	0.64	0.25	0.14	0.45	0.69	0.32
D_4^*	Bias	-0.06	0.12	-0.50	-0.37	-0.05	0.00	-0.01	-0.33
	$\sqrt{MSE}$	0.13	0.42	0.93	0.57	0.19	0.69	0.89	0.57
D_5^*	Bias	0.00	-0.29	0.52	0.01	-0.02	0.01	-0.01	-0.23
	$\sqrt{MSE}$	0.13	0.55	0.97	0.21	0.16	0.56	0.79	0.43

order statistics in ranked set sampling designs. In this paper the order statistics are correlated, so requires different latent structures, missing data mechanisms, and EM algorithms than those in Hatefi et al. (2014, 2015).

We used the properties of the correlatet order statistics in estimation and classification of FMMs under both supervised and unsupervised learning methods. Using the correlation structure of the order statistics, we obtained various model-based classification criteria. These criteria not only assist us to determine the group membership of data, but also enable infer-

Table 5: Computational aspects of the estimators of Spot Data under supervised learning based on designs $D_i^*; i = 1, \ldots, 5$, against those of SRS data of the same size.

	OS				SRS			
	interation	CLP%	time	Conv.	iteration	CLP%	time	Conv.
D_1^*	4.26	86.40	0.0049	98.86	1.00	86.60	0.0004	98.06
D_2^*	3.00	86.83	0.0036	99.84	1.00	86.51	0.0004	98.10
D_3^*	3.69	86.89	0.0042	99.98	1.00	86.54	0.0004	97.96
D_4^*	4.60	85.37	0.0035	87.80	1.00	84.84	0.0003	85.72
D_5^*	3.58	85.86	0.0030	99.92	1.00	85.66	0.0003	93.52

ence about rarely-observed components. Our framework is general enough to apply to several sampling designs from FMMs, including left censoring, right censoring, double censoring, minimal-maximal nomination sampling and systematic sampling. Empirical evidence illustrates that the selection of appropriate collection of order statistics provides substantial improvement over their SRS counterparts in both supervised and unsupervised learning. For example, systematic sampling can be two or three times more efficient than their SRS counterparts in the estimation of mixing proportion in supervised and unsupervised learning, respectively. The proposed methodologies finally were employed in determining the age structure of Spot fish using length frequency data. Numerical results illustrate that es-

Table 6: Bias, $\sqrt{MSE}$ of Spot Data under unsupervised learning based on designs $D_i^*; i = 1, \ldots, 5$, against those of SRS data of the same size.

		OS				SRS			
		π	μ_1	μ_2	σ	π	μ_1	μ_2	σ
D_1^*	Bias	-0.21	-0.62	-0.64	-0.19	-0.10	-0.35	-0.05	-0.34
	$\sqrt{MSE}$	0.35	0.97	1.12	0.36	0.23	0.77	0.85	0.53
D_2^*	Bias	-0.07	-0.24	-0.07	-0.23	-0.09	-0.33	-0.03	-0.34
	$\sqrt{MSE}$	0.16	0.48	0.54	0.36	0.23	0.77	0.86	0.53
D_3^*	Bias	0.06	0.19	0.56	-0.12	-0.10	-0.35	-0.06	-0.34
	$\sqrt{MSE}$	0.18	0.56	1.01	0.32	0.24	0.78	0.86	0.53
D_4^*	Bias	-0.15	-0.25	-0.72	-0.54	-0.13	-0.42	-0.20	-0.56
	$\sqrt{MSE}$	0.23	0.48	1.13	0.78	0.27	1.00	1.13	0.83
D_5^*	Bias	-0.01	-0.33	0.48	0.02	-0.11	-0.39	-0.12	-0.45
	$\sqrt{MSE}$	0.19	0.68	1.02	0.27	0.25	0.88	1.00	0.67

timation procedures under both supervised and unsupervised approaches almost outperform their SRS-counterparts in estimation and classification precision.

Supplementary Materials

All the proofs, eight remarks as well as one simulation study are provided in Appendix as Supplementary Materials.

Table 7: Computational aspects of the estimators of Spot Data under unsupervised learning based on designs $D_i^*; i = 1, \ldots, 5$, against those of SRS data of the same size.

	OS				SRS			
	interation	CLP%	time	Conv.	iteration	CLP%	time	Conv.
D_1^*	18.34	75.12	0.0220	92.30	12.43	82.17	0.0043	99.06
D_2^*	14.78	85.02	0.0186	99.52	12.64	81.91	0.0044	98.90
D_3^*	21.04	83.75	0.0251	94.88	12.46	81.91	0.0043	99.00
D_4^*	9.95	81.85	0.0080	99.88	8.20	79.91	0.0028	98.82
D_5^*	21.68	82.59	0.0189	93.76	10.34	80.83	0.0035	99.30

References

Biradar, B. and Santosha, C. (2015). Estimation of the population mean based on extremes ranked set sampling. *American Journal of Mathematics and Statistics* **5**, 32–36.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Singapore: Springer.

Chen, Z., Bai, Z., and Sinha, B. (2004). Ranked Set Sampling: Theory and Applications. Springer.

David, H. A. and Nagaraja, H. N. (1981). Order Statistics. New Jersey: Wiley Online Library.

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)* **39**, 1–38.

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.

Hatefi, A., Jafari Jozani, M., and Ozturk, O. (2015). Mixture model analysis of partially rank-ordered set samples: Age groups of fish from length-frequency data. *Scandinavian Journal of Statistics* **42**, 848–871.

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.

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

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., Tenn.(USA).

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. and Peel, D. (2004). Finite Mixture Models. New York: Wiley.

Mehrotra, K. and Nanda, P. (1974). Unbiased estimation of parameters by order statistics in the case of censored samples. *Biometrika* **61**, 601–606.

Mendenhall, W. and Hader, R. (1958). Estimation of parameters of mixed exponentially dis-

tributed failure time distributions from censored life test data. *Biometrika* **45**, 504–520.

Miyata, Y. (2011). Maximum likelihood estimators in finite mixture models with censored data. *Journal of Statistical Planning and Inference* **141**, 56–64.

Nourmohammadi, M., Jafari Jozani, M., and Johnson, B. C. (2015). Distribution-free tolerance intervals with nomination samples: Applications to mercury contamination in fish. *Statistical Methodology* **26**, 16–33.

Ogle, D. 2013. Fisheries stock assessment (fsa) methods package for r. R package version 0.4.13.

Ozturk, O. (2011). Sampling from partially rank-ordered sets. *Environmental and Ecological Statistics* **18**, 757–779.

Rickabaugh, H. and Capossela, K. (2011). Evaluation of the status of spot in maryland 2010. *Maryland DNR Fisheries Services Doc 6-23-2011* .

Saleem, M., Aslam, M., and Economou, P. (2010). On the bayesian analysis of the mixture of power function distribution using the complete and the censored sample. *Journal of Applied Statistics* **37**, 25–40.

Sindhu, T., Feroze, N., and Aslam, M. (2016). Doubly censored data from two-component mixture of inverse weibull distributions: Theory and applications. *Journal of Modern Applied Statistical Methods* **15**, 1–21.

Summerfelt, R. and Hall, G. (1987). Age and Growth of Fish. Ames: Iowa State University

Press.

Thomas, P. (1990). Teleost model for studying the effects of chemicals on female reproductive endocrine function. *The journal of experimental zoology* **256**, 126–128.

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. (2004). Ranked set sampling: an approach to more efficient data collection. *Statistical Science,* 636–643.

Yang, S.-S. (1977). General distribution theory of the concomitants of order statistics. *The Annals of Statistics* **5**, 996–1002.

Zheng, G. and Al-Saleh, M. F. (2002). Modified maximum likelihood estimators based on ranked set samples. *Annals of the Institute of Statistical Mathematics* **54**, 641–658.

Department of Mathematics and Statistics, Memorial University, St. John's, NL, Canada.

E-mail: ahatefi@mun.ca, hatefi.ar@gmail.com

Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada.

E-mail: reid@utstat.utoronto.ca

Department of Statistics, University of Manitoba, Winnipeg, MB, Canada.

E-mail: m_jafari_jozani@umanitoba.ca

Department of Statistics, The Ohio State University, Columbus, OH 43210, USA.

E-mail: omer@stat.osu.edu

Appendix (Supplementary Materials)

Remark 2. Let $X_{(r)} = x_r$ be the r-th order statistic of a random sample of size n from FMM (2.1) and $\mathbf{Z}_r = \mathbf{z}_r$ be associated latent vector with x_r. Let $\mathbf{Z}_l = \mathbf{z}_l$ be the latent vector associated with unobserved l-th order statistic where $l \leq r$. From Lemma 1, the joint distribution of $(X_{(r)}, \mathbf{Z}_r, \mathbf{Z}_l)$ is

$$f(x_r, \mathbf{z}_r, \mathbf{z}_l) \propto \{F(x_r; \mathbf{\Psi})\}^{r-2} \prod_{j=1}^{M} \{\pi_j F_j(x_r; \theta_j)\}^{z_{lj}} \{\pi_j f_j(x_r; \theta_j)\}^{z_{rj}} \times \left\{\bar{F}(x_r; \mathbf{\Psi})\right\}^{n-r}.$$

Remark 3. Let $X_{(r)} = x_r$ and $X_{(s)} = x_s$ be the r-th and s-th order statistics of a random sample of size n from FMM (2.1) where $r < s$. Let $\mathbf{Z}_r = \mathbf{z}_r$ and $\mathbf{Z}_s = \mathbf{z}_s$ be the latent vector associated with x_r and x_s, respectively. Let $\mathbf{Z}_l = \mathbf{z}_l$ be the latent vector associated with unobserved l-th order statistic, where $r \leq l \leq s$. From Lemma 1, the joint distribution of $(X_{(r)}, X_{(s)}, \mathbf{Z}_r, \mathbf{Z}_l, \mathbf{Z}_s)$ is given by

$$\begin{aligned} f(x_r, x_s, \mathbf{z}_r, \mathbf{z}_l, \mathbf{z}_s) \propto & \{F(x_r; \mathbf{\Psi})\}^{r-1} \prod_{j=1}^{M} \{\pi_j f_j(x_r; \theta_j)\}^{z_{rj}} \{\pi_j f_j(x_s; \theta_j)\}^{z_{sj}} \\ & \times \left[\pi_j \left\{F_j(x_s; \theta_j) - F_j(x_r; \theta_j)\right\}\right]^{z_{lj}} \\ & \times \left\{F(x_s; \mathbf{\Psi}) - F(x_r; \mathbf{\Psi})\right\}^{s-r-2} \left\{\bar{F}(x_s; \mathbf{\Psi})\right\}^{n-s}. \end{aligned}$$

Remark 4. Let $X_{(r)} = x_r$ be the r-th order statistic of a random sample of size n from FMM (2.1) and $\mathbf{Z}_r = \mathbf{z}_r$ be the latent vector associated with x_r. Let $\mathbf{Z}_l = \mathbf{z}_l$ be the latent vector associated with unobserved l-th order statistic, where $r \leq l$. From Lemma 1, the joint distribution of $(\mathbf{X}, \mathbf{Z}_r, \mathbf{Z}_l)$ is given by

$$f(x_r, \mathbf{z}_r, \mathbf{z}_l) \propto \{F(x_r; \mathbf{\Psi})\}^{r-1} \prod_{j=1}^{M} \{\pi_j f_j(x_r; \theta_j)\}^{z_{rj}} \left\{\pi_j \bar{F}_j(x_r; \theta_j)\right\}^{z_{lj}} \times \left\{\bar{F}(x_r; \mathbf{\Psi})\right\}^{n-r-1}.$$

Remark 5. Let $\tilde{\mathbf{X}} = \{X_{(i_1)}, X_{(i_2)}, \ldots, X_{(i_k)}\}$ be a collection of $k = 2, \ldots, n-1$ order statistics from a random sample of size n from (2.1) and let $\mathbf{\Delta} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_k, \mathbf{W}_1, \ldots, \mathbf{W}_{k+1})$ be the collection of latent vectors defined above. Using Lemma 1, the pdf of $(\tilde{\mathbf{X}}, \mathbf{W}_1)$ can be derived as

$$f(\tilde{\mathbf{x}}, \mathbf{w}_1) \propto \prod_{j=1}^{M} \{\pi_j F_j(x_{i_1}; \theta_j)\}^{w_{1j}} \prod_{s=1}^{k} f(x_{i_s}; \mathbf{\Psi}) \left\{\bar{F}(x_{i_k}; \mathbf{\Psi})\right\}^{n-i_k} \times \prod_{s=2}^{k} \left\{F(x_{i_s}; \mathbf{\Psi}) - F(x_{i_{s-1}}; \mathbf{\Psi})\right\}^{i_s - i_{s-1} - 1} \tag{8.1}$$

Remark 6. In a similar vein to Remark 5, the joint distribution of $(\tilde{\mathbf{X}}, \mathbf{W}_s)$;

$s = 2, \ldots, k$, is given by

$$f(\tilde{\mathbf{x}}, \mathbf{w}_s) \propto \{F(x_{i_1}; \boldsymbol{\Psi})\}^{i_1 - 1} \prod_{r=1}^{k} f(x_{i_r}; \boldsymbol{\Psi}) \prod_{j=1}^{M} \left[\pi_j \left\{F_j(x_{i_s}; \theta_j) - F_j(x_{i_{s-1}}; \theta_j)\right\}\right]^{w_{sj}} \times \left[\prod_{\substack{l=2 \\ l \neq s}}^{k} \left\{F(x_{i_l}; \boldsymbol{\Psi}) - F(x_{i_{l-1}}; \boldsymbol{\Psi})\right\}^{i_l - i_{l-1} - 1}\right] \left\{\bar{F}(x_{i_k}; \boldsymbol{\Psi})\right\}^{n - i_k} \tag{8.2}$$

Remark 7. In a similar vein to Remark 5, the joint pdf of $(\tilde{\mathbf{X}}, \mathbf{W}_{k+1})$ is given by

$$f(\tilde{\mathbf{x}}, \mathbf{w}_{k+1}) \propto \{F(x_{i_1}; \boldsymbol{\Psi})\}^{i_1 - 1} \prod_{s=2}^{k} \left\{F(x_{i_s}; \boldsymbol{\Psi}) - F(x_{i_{s-1}}; \boldsymbol{\Psi})\right\}^{i_s - i_{s-1} - 1} \times \prod_{j=1}^{M} \left\{\pi_j \bar{F}_j(x_{(i_k)}; \theta_j)\right\}^{w_{k+1\,j}} \prod_{s=1}^{k} f(x_{i_s}; \boldsymbol{\Psi}). \tag{8.3}$$

Remark 8. Given $X_{(r)} = x_{(r)}$ and $\mathbf{Z}_{(r)} = \mathbf{z}_{(r)}$, suppose we are interested in classifying an unobserved order statistic $X_{(l)}$ for $l \geq r$. The component membership vector $\mathbf{Z}_l = (Z_{l1}, \ldots, Z_{lM})$ can be estimated similarly as explained above. From Remark 4, the posterior distribution of $\mathbf{Z}_l$ given $x_{(r)}$ and $\mathbf{z}_{(r)}$ is

$$\mathbb{P}(\mathbf{Z}_l = \mathbf{z}_l | \mathbf{Z}_r = \mathbf{z}_r, x_{(r)}) = \binom{1}{z_{l1}, \ldots, z_{lM}} \prod_{h=1}^{M} \left\{\frac{\pi_h \bar{F}_h(x_{(r)}; \theta_h)}{\bar{F}(x_{(r)}; \boldsymbol{\Psi})}\right\}^{z_{lh}},$$

where $\gamma_h(x_{(r)}; \boldsymbol{\Psi}) = \pi_h \, \bar{F}_h(x_{(r)}; \theta_h) / \bar{F}(x_{(r)}; \boldsymbol{\Psi})$. Hence, given the observation y from the FMM, an unobserved data but bigger than y will be classified

into the j-th component of the FMM, if $\gamma_j(y;\hat{\mathbf{\Psi}}) > \gamma_h(y;\hat{\mathbf{\Psi}})$ for all $h = 1,\ldots,M; j \neq h$.

Remark 9. Given $X_{(r)} = x_{(r)}, \mathbf{Z}_{(r)} = \mathbf{z}_{(r)}, X_{(s)} = x_{(s)}$ and $\mathbf{Z}_{(s)} = \mathbf{z}_{(s)}$, if the interest is to classify an unobserved order statistic $X_{(l)}$ for $s \leq l \leq r$, we can estimate the component membership vector $\mathbf{Z}_l = (Z_{l1},\ldots,Z_{lM})$. From Remark 3, the posterior distribution of $\mathbf{Z}_l$ given $x_{(r)}, \mathbf{z}_{(r)}$ and $x_{(s)}, \mathbf{z}_{(s)}$ becomes

$$\mathbb{P}(\mathbf{Z}_l = \mathbf{z}_l | \mathbf{z}_r, x_{(r)}, \mathbf{z}_s, x_{(s)}) \propto \prod_{h=1}^{M} \left\{ \frac{\pi_h[F_h(x_{(r)};\theta_h) - F_h(x_{(s)};\theta_h)]}{F(x_{(r)};\mathbf{\Psi}) - F(x_{(s)};\mathbf{\Psi})} \right\}^{z_{lh}},$$

so $\gamma_h(x_{(s)}, x_{(r)};\mathbf{\Psi}) = \pi_h[F_h(x_{(r)};\theta_h) - F_h(x_{(s)};\theta_h)]/[F(x_{(r)};\mathbf{\Psi}) - F(x_{(s)};\mathbf{\Psi})]$. Hence, given the observations y_1, y_2 such that $y_1 < y_2$ from the underlying FMM, an unobserved data between y_1 and y_2 will be classified into the j-th component of the FMM, if $\gamma_j(y_1, y_2;\hat{\mathbf{\Psi}}) > \gamma_h(y_1, y_2;\hat{\mathbf{\Psi}})$ for all $h = 1,\ldots,M; j \neq h$.

Proof of Lemma 4

Proof. From (2.5), we have $W_{1j}|X_{(r)} = x_r \sim B\left(r-1, \frac{\pi_j F_j(x_r;\theta_j)}{F(x_r;\mathbf{\Psi})}\right)$, where W_{1j} represents the number of the order statistics smaller than x_r from component j; $j = 1,\ldots,M$. Lemma 3 for the variable $W_{1j}|x_r$ completes the proof. □

Proof of Lemma 5

Proof. From (2.6), $W_{rl,j}|\{x_r, x_l\} \sim B\left(l-r-1, \frac{\pi_j[F_j(x_l;\theta_j)-F_j(x_r;\theta_j)]}{F(x_l;\mathbf{\Psi})-F(x_r;\mathbf{\Psi})}\right)$, where $W_{rl,j}$ represents the number of the order statistics between $X_{(r)}$ and $X_{(l)}$ from component $j = 1, \ldots, M$. One completes the proof by applying Lemma 3 for $W_{rl,j}|\{x_r, x_l\}$. □

Proof of Lemma 6

Proof. From (2.7), $W_{lj}|x_r \sim B\left(n-l, \frac{\pi_j \bar{F}_j(x_l;\theta_j)}{\bar{F}(x_l;\mathbf{\Psi})}\right)$, where W_{lj} represents the number of the order statistics bigger than x_l from component j, $j = 1, \ldots, M$. Applying Lemma 3 to $\mathbf{W}_{lj}|x_r$ completes the proof. □

8.1 Simulation Study 2

In the second simulation study, we investigate the performance of the ML estimates of all parameters of FMM (6.1) with $\mathbf{\Psi} = \{\pi, \mu_1, \mu_2, \sigma\} = \{0.80, 9.01, 11.70, 1.15\}$ using the designs in Table 1. We evaluate the estimation procedures for both supervised and unsupervised learning approaches over 5000 simulations. In each simulation, we use the stopping criteria $||\mathbf{\Psi}^{(k+1)} - \mathbf{\Psi}^{(k)}||_\infty < 10^{-6}$ and the initial values in the EM-algorithms are computed using the method of moments by treating the order statistics as a simple random sample data (Furman and Lindsay, 1994). In addition,

for each simulation, we generate a fixed test data of size $n = 30$ for estimation procedures under both supervised and unsupervised order statistics and SRS counterparts. Tables 8 and 10 show the bias and square root of MSE as performance measures for each estimation procedure. Tables 9 and 11 present different computational aspects associated with each estimation procedure. For each simulation, CVR% and CLP% are obtained as explained in simulation study 1. In addition to these criteria, we also compare the performance of MLEs based on the average of the number of iterations required for convergence (ITR) and the average time (in seconds) required for convergence (TIME). Tables 9 and 11 illustrate the significant impact of various collection order statistics on the estimation and classification procedures of FMMs suffering from rarely observed components. From Table 9, using collection of lower order statistics (design D_1), we are not capable of observing the rare event (second component) and the estimation procedures are practically not convergent; however, appropriate collection of order statistics (e.g., design D_2) guarantees observation of rare component and convergence of the procedures under labelled data. Table 11 shows that the convergence rate is almost stable and high under various collection of unlabeled order statistics (except for D_1) similar to that of SRS. Unlike the convergence rate, the impact of various collection of unlabeled order

statistics is evident on classification precision compared to unlabeled SRS counterparts.

Table 8: Bias, $\sqrt{MSE}$ under supervised learning approach based on designs $D_i; i = 1, \ldots, 5$ in Table 1, against those of SRS data of the same size.

		OS				SRS			
		π	μ_1	μ_2	σ	π	μ_1	μ_2	σ
D_1	Bias	-0.43	-1.27	-3.55	-0.69	-0.08	-0.01	0.02	-0.26
	$\sqrt{MSE}$	0.67	1.82	5.04	1.00	0.17	0.56	0.99	0.49
D_2	Bias	-0.01	1.85	0.26	-0.47	-0.04	0.00	0.01	-0.19
	$\sqrt{MSE}$	0.14	2.66	0.62	0.71	0.13	0.47	0.92	0.3900
D_3	Bias	-0.11	-1.45	0.86	-0.34	-0.07	-0.02	0.00	-0.27
	$\sqrt{MSE}$	0.33	2.16	1.35	0.77	0.16	0.56	1.00	0.50
D_4	Bias	-0.04	-0.94	0.54	-0.03	-0.02	0.01	0.01	-0.15
	$\sqrt{MSE}$	0.17	1.44	0.93	0.46	0.11	0.42	0.87	0.33
D_5	Bias	-0.01	-0.44	0.76	0.10	-0.07	-0.02	-0.02	-0.26
	$\sqrt{MSE}$	0.12	0.74	1.38	0.29	0.16	0.56	0.99	0.49

Table 9: Computational aspects of the estimators under supervised learning approach based on designs $D_i; i = 1, \ldots, 5$ in Table 1, against those of SRS data of the same size.

	OS				SRS			
	iteration	CLP%	time	Conv.	iteration	CLP%	time	Conv.
D_1	7.60	38.76	0.0032	0.92	1.00	87.72	0.0003	73.94
D_2	4.78	84.52	0.0022	94.14	1.00	88.54	0.0003	83.62
D_3	7.90	83.23	0.0034	99.44	1.00	87.75	0.0003	73.14
D_4	6.14	86.49	0.0030	99.72	1.00	89.10	0.0003	89.38
D_5	5.03	88.01	0.0036	97.42	1.00	87.74	0.0003	73.54

Table 10: Bias, $\sqrt{MSE}$ under unsupervised learning approach based on designs $D_i; i = 1, \ldots, 5$ in Table 1, against those of SRS data of the same size.

		OS				SRS			
		π	μ_1	μ_2	σ	π	μ_1	μ_2	σ
D_1	Bias	-0.77	-2.12	-3.90	-0.84	-0.23	-0.58	-0.55	-0.52
	$\sqrt{MSE}$	1.08	3.07	5.53	1.20	0.39	1.17	1.48	0.79
D_2	Bias	0.15	2.20	1.00	-0.61	-0.20	-0.53	-0.39	-0.42
	$\sqrt{MSE}$	0.22	3.16	1.66	0.91	0.36	1.09	1.35	0.65
D_3	Bias	-0.18	-1.78	0.79	-0.66	-0.23	-0.59	-0.56	-0.53
	$\sqrt{MSE}$	0.40	2.56	1.27	0.98	0.39	1.18	1.48	0.79
D_4	Bias	-0.15	-1.45	0.37	-0.50	-0.18	-0.49	-0.31	-0.35
	$\sqrt{MSE}$	0.30	2.12	0.78	0.79	0.34	1.02	1.28	0.56
D_5	Bias	-0.05	-0.44	0.51	0.18	-0.23	-0.59	-0.54	-0.52
	$\sqrt{MSE}$	0.25	0.9400	1.46	0.53	0.39	1.18	1.46	0.79

Table 11: Computational aspects of the estimators under unsupervised learning approach based on designs $D_i; i = 1, \ldots, 5$ in Table 1, against those of SRS data of the same size.

	OS				SRS			
	iteration	CLP%	time	Conv.	iteration	CLP%	time	Conv.
D_1	15.98	23.14	0.0073	79.50	9.77	73.63	0.0033	99.16
D_2	18.34	84.22	0.0088	97.34	11.71	75.54	0.0040	98.96
D_3	7.55	79.70	0.0034	99.72	9.73	73.59	0.0031	99.16
D_4	7.12	78.96	0.0036	99.80	13.68	76.67	0.0045	98.42
D_5	25.16	80.51	0.0190	92.10	9.73	73.78	0.0031	99.08