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*

Supplementary Material

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

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

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

$$f(\tilde{\mathbf{x}}, \mathbf{w}_1) \propto \prod_{j=1}^{M} \{\pi_j F_j(x_{i_1}; \theta_j)\}^{w_{1j}} \prod_{s=1}^{k} f(x_{i_s}; \mathbf{\Psi}) \left\{\bar{F}(x_{i_k}; \mathbf{\Psi})\right\}^{n-i_k} \times \prod_{s=2}^{k} \left\{F(x_{i_s}; \mathbf{\Psi}) - F(x_{i_{s-1}}; \mathbf{\Psi})\right\}^{i_s - i_{s-1} - 1} \tag{S0.1}$$

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

$$f(\tilde{\mathbf{x}}, \mathbf{w}_s) \propto \{F(x_{i_1}; \mathbf{\Psi})\}^{i_1 - 1} \prod_{r=1}^{k} f(x_{i_r}; \mathbf{\Psi}) \prod_{j=1}^{M} \left[\pi_j \left\{F_j(x_{i_s}; \theta_j) - F_j(x_{i_{s-1}}; \theta_j)\right\}\right]^{w_{sj}} \times \left[\prod_{\substack{l=2 \\ l \neq s}}^{k} \left\{F(x_{i_l}; \mathbf{\Psi}) - F(x_{i_{l-1}}; \mathbf{\Psi})\right\}^{i_l - i_{l-1} - 1}\right] \left\{\bar{F}(x_{i_k}; \mathbf{\Psi})\right\}^{n-i_k} \tag{S0.2}$$

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

$$f(\tilde{\mathbf{x}}, \mathbf{w}_{k+1}) \propto \{F(x_{i_1}; \boldsymbol{\Psi})\}^{i_1-1} \prod_{s=2}^{k} \left\{F(x_{i_s}; \boldsymbol{\Psi}) - F(x_{i_{s-1}}; \boldsymbol{\Psi})\right\}^{i_s - i_{s-1} - 1}$$
$$\times \prod_{j=1}^{M} \left\{\pi_j \bar{F}_j(x_{(i_k)}; \theta_j)\right\}^{w_{k+1\,j}} \prod_{s=1}^{k} f(x_{i_s}; \boldsymbol{\Psi}). \tag{S0.3}$$

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 $\mathbf{Z}_l = (Z_{l1}, \dots, Z_{lM})$ 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

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

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 $\gamma_j(y; \hat{\boldsymbol{\Psi}}) > \gamma_h(y; \hat{\boldsymbol{\Psi}})$ for all $h = 1, \dots, M; j \neq h$.

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 $\mathbf{Z}_l = (Z_{l1}, \dots, Z_{lM})$. 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. $\square$

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 $W_{rl,j}|\{x_r, x_l\}$. $\square$

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. □

S0.1 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