Inference on the Order of a Normal Mixture: Supplementary Material

Jiahua Chen, Pengfei Li, and Yuejiao Fu*

April 26, 2012

This is a supplementary document to the corresponding paper submitted to the *Journal of the American Statistical Association*. It contains the proof of Theorem 1 and a brief description of the R-command used to test the null hypotheses of order two or three.

---

*Jiahua Chen is Professor, Department of Statistics, University of British Columbia, Vancouver, BC, Canada, V6T 1Z2 (Email: *jhchen@stat.ubc.ca*). Pengfei Li is Assistant Professor, Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, ON, Canada, N2L 3G1 (Email: *pengfei.li@uwaterloo.ca*). Yuejiao Fu is Associate Professor, Department of Mathematics and Statistics, York University, Toronto, ON, Canada, M3J 1P3 (Email: *yuejiao@mathstat.yorku.ca*). The authors thank the editor, the associate editor, and three referees for constructive comments and suggestions that lead to significant improvements in the article. The research is supported by the Natural Sciences and Engineering Research Council of Canada and by a startup grant from the University of Waterloo.

1 Proof

Throughout the proofs, we assume that the true value of Ψ under the null hypothesis of order m_0 is Ψ_0, where

$$\Psi_0(\theta, \sigma) = \sum_{h=1}^{m_0} \alpha_{0h} I(\theta_{0h} \leq \theta, \sigma_{0h} \leq \sigma),$$

the means $\theta_{01} < \theta_{02} < \cdots < \theta_{0m_0}$, and the mixing proportions α_{0h} are nonzero. For convenience of presentation, let $\boldsymbol{\alpha}_0 = (\alpha_{01}, \ldots, \alpha_{0m_0})^\tau$, $\boldsymbol{\theta}_0 = (\theta_{01}, \ldots, \theta_{0m_0})^\tau$, and $\boldsymbol{\sigma}_0 = (\sigma_{01}, \ldots, \sigma_{0m_0})^\tau$. Let $\hat{\Psi}_0$ be the penalized MLE of Ψ under the null hypothesis such that

$$\hat{\Psi}_0(\theta, \sigma) = \sum_{h=1}^{m_0} \hat{\alpha}_{0h} I(\hat{\theta}_{0h} \leq \theta, \hat{\sigma}_{0h} \leq \sigma)$$

with $\hat{\theta}_{01} \leq \cdots \leq \hat{\theta}_{0m_0}$.

Preliminary results

In the following, for any given $\boldsymbol{\beta}_0 \in \mathrm{B}$, suppose $\Psi^{(k)}(\boldsymbol{\beta}_0)$ is obtained iteratively as specified in Section 2.1 of the paper. Let $\boldsymbol{\alpha}^{(k)}$, $\boldsymbol{\theta}_1^{(k)}$, $\boldsymbol{\theta}_2^{(k)}$, $\boldsymbol{\sigma}_1^{(k)}$, $\boldsymbol{\sigma}_2^{(k)}$, and $\boldsymbol{\beta}^{(k)}$ be the constituent entries of $\Psi^{(k)}(\boldsymbol{\beta}_0)$. The proof of Theorem 1 depends on the consistency of $\boldsymbol{\alpha}^{(k)}$, $\boldsymbol{\theta}_1^{(k)}$, $\boldsymbol{\theta}_2^{(k)}$, $\boldsymbol{\sigma}_1^{(k)}$, $\boldsymbol{\sigma}_2^{(k)}$, and $\boldsymbol{\beta}^{(k)}$, which is shown by the following proposition.

Proposition 1. *Suppose the conditions in Theorem 1 are satisfied. Under the null distribution $f(x; \Psi_0)$, and for each given $\boldsymbol{\beta}_0 \in B^{m_0}$, after any finite k iterations, we have*

$$\boldsymbol{\alpha}^{(k)} - \boldsymbol{\alpha}_0 = o_p(1),\ \boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}_0 = o_p(1),\ \boldsymbol{\theta}_1^{(k)} - \boldsymbol{\theta}_0 = o_p(1),$$

$$\boldsymbol{\theta}_2^{(k)} - \boldsymbol{\theta}_0 = o_p(1),\ \boldsymbol{\sigma}_1^{(k)} - \boldsymbol{\sigma}_0 = o_p(1),\ \boldsymbol{\sigma}_2^{(k)} - \boldsymbol{\sigma}_0 = o_p(1).$$

When the data are generated from a null model, the EM-iteration only changes the value of $\boldsymbol{\beta}$ by a $o_p(1)$ quantity. Therefore the updated value of $\boldsymbol{\beta}$, $\boldsymbol{\beta}^{(k)}$, stays in a small neighbourhood of the initial value $\boldsymbol{\beta}_0$. This is a technical result useful for future derivations.

We prove Proportion 1 in several steps. The consistency result for $k = 1$ is proved in Lemma 1. We then show in Lemmas 2 and 3 that consistency extends from $k = k_0$ to $k = k_0 + 1$. Lemma 2 proves that the EM-iterations retain $\boldsymbol{\beta}^{(k_0+1)}$ in a small neighborhood of $\boldsymbol{\beta}_0$. Lemma 3 further establishes the consistency of $\boldsymbol{\alpha}^{(k_0+1)}$, $\boldsymbol{\theta}_1^{(k_0+1)}$, $\boldsymbol{\theta}_2^{(k_0+1)}$, $\boldsymbol{\sigma}_1^{(k_0+1)}$, and $\boldsymbol{\sigma}_2^{(k_0+1)}$. Hence, Proposition 1 follows by mathematical induction.

Lemma 1. *Suppose the conditions in Theorem 1 are satisfied. Under the null distribution $f(x;\Psi_0)$, and for each given $\boldsymbol{\beta}_0 \in B^{m_0}$, we have*

$$\boldsymbol{\alpha}^{(1)} - \boldsymbol{\alpha}_0 = o_p(1),\ \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}_0 = o_p(1),\ \boldsymbol{\theta}_1^{(1)} - \boldsymbol{\theta}_0 = o_p(1),$$

$$\boldsymbol{\theta}_2^{(1)} - \boldsymbol{\theta}_0 = o_p(1),\ \boldsymbol{\sigma}_1^{(1)} - \boldsymbol{\sigma}_0 = o_p(1),\ \boldsymbol{\sigma}_2^{(1)} - \boldsymbol{\sigma}_0 = o_p(1).$$

Proof.

For given $\boldsymbol{\beta}_0 \in \mathrm{B}^{m_0}$, we can write

$$\hat{\Psi}_0 = \sum_{h=1}^{m_0} \{\hat{\alpha}_{0h}\beta_{0h} I(\hat{\theta}_{0h} \le \theta, \hat{\sigma}_{0h} \le \sigma) + \hat{\alpha}_{0h}(1-\beta_{0h}) I(\hat{\theta}_{0h} \le \theta, \hat{\sigma}_{0h} \le \sigma)\}.$$

Hence, $\hat{\Psi}_0 \in \Omega_{2m_0}(\boldsymbol{\beta}_0)$ as defined in Section 2.1 of the paper.

By definition, $\Psi^{(1)}(\boldsymbol{\beta}_0)$ maximizes $pl_n(\Psi)$ over $\Omega_{2m_0}(\boldsymbol{\beta}_0)$. Hence,

$$pl_n(\Psi^{(1)}(\boldsymbol{\beta}_0)) \ge pl_n(\hat{\Psi}_0) \ge l_n(\hat{\Psi}_0) + \sum_{h=1}^{m_0} p(\beta_{0h}). \tag{1}$$

The second inequality is true because $p_n(\hat{\sigma}_{0h}^2; \hat{\sigma}_{0h}^2) = 0$ by Condition C2 and the penalties $\sum_{h=1}^{m_0} p(\beta_{0h})$ are balanced. By the definition of $\hat{\Psi}_0$, we see that

$$l_n(\hat{\Psi}_0) + \sum_{h=1}^{m_0} p_n(\hat{\sigma}_{0h}^2; s_n^2) \geq l_n(\Psi_0) + \sum_{h=1}^{m_0} p_n(\sigma_{0h}^2; s_n^2).$$

By removing the nonpositive $p_n(\hat{\sigma}_{0h}^2; s_n^2)$ from the left-hand side, we get

$$l_n(\hat{\Psi}_0) \geq l_n(\Psi_0) + \sum_{h=1}^{m_0} p_n(\sigma_{0h}^2; s_n^2) \geq l_n(\Psi_0) + o(n), \tag{2}$$

where the second inequality comes from Condition C3 on $p_n(\cdot;\cdot)$.

Combining (1) and (2), we get

$$pl_n(\Psi^{(1)}(\boldsymbol{\beta}_0)) \geq l_n(\Psi_0) + o(n). \tag{3}$$

Theorem 3 of Chen et al. (2008) showed that any estimator $\hat{\Psi}$ satisfying inequality (3) is consistent. Therefore, $\Psi^{(1)}(\boldsymbol{\beta}_0)$ is consistent.

Because $\boldsymbol{\beta}_0$ is fixed and $\Psi^{(1)}(\boldsymbol{\beta}_0)$ is confined in $\Omega_{2m_0}(\boldsymbol{\beta}_0)$, the consistency of $\Psi^{(1)}$ implies $\boldsymbol{\alpha}^{(1)} - \boldsymbol{\alpha}_0 = o_p(1)$ and the other conclusions in this lemma. $\square$

Next, we continue the process of mathematical induction. We show that under the null distribution $f(x; \Psi_0)$, the EM-iteration changes the fitted values of $\boldsymbol{\beta}$ by only $o_p(1)$. For a given $k_0 \geq 1$, let

$$H_{nh}(\beta) = \sum_{i=1}^{n} w_{i1h}^{(k_0)} \log(\beta) + \sum_{i=1}^{n} w_{i2h}^{(k_0)} \log(1-\beta),$$

and $Q_{nh}(\beta) = H_{nh}(\beta) + p(\beta)$. The updated value of β_h is $\beta_h^{(k_0+1)} = \arg\max_\beta Q_{nh}(\beta)$.

Lemma 2. *Suppose that Conditions C1–C5 are satisfied. If under the null distribution $f(x; \Psi_0)$ and for each given $\boldsymbol{\beta}_0 \in B^{m_0}$,*

$$\boldsymbol{\alpha}^{(k_0)} - \boldsymbol{\alpha}_0 = o_p(1),\ \ \boldsymbol{\beta}^{(k_0)} - \boldsymbol{\beta}_0 = o_p(1),\ \ \boldsymbol{\theta}_1^{(k_0)} - \boldsymbol{\theta}_0 = o_p(1),$$

$$\boldsymbol{\theta}_2^{(k_0)} - \boldsymbol{\theta}_0 = o_p(1),\ \boldsymbol{\sigma}_1^{(k_0)} - \boldsymbol{\sigma}_0 = o_p(1),\ \boldsymbol{\sigma}_2^{(k_0)} - \boldsymbol{\sigma}_0 = o_p(1),$$

then we have $\beta_h^{(k_0+1)} - \beta_{0h} = o_p(1)$ *for* $h = 1, \ldots, m_0$.

PROOF. Note that $H_{nh}(\beta)$ is maximized at $\hat{\beta}_h = \sum_{i=1}^n w_{i1h}^{(k_0)}/(n\alpha_n^{(k_0)})$. We show that $\beta_h^{(k_0+1)} - \beta_{0h} = o_p(1)$ through the proofs of $\hat{\beta}_h - \beta_{0h} = o_p(1)$ and $\beta_h^{(k_0+1)} - \hat{\beta}_h = o_p(1)$.

By the consistency of $\Psi^{(k_0)}$ assumed in this lemma, we have

$$\begin{aligned}
\sum_{i=1}^n \frac{f(X_i; \theta_{1h}^{(k_0)}, \sigma_{1h}^{(k_0)})}{f(X_i; \Psi^{(k_0)}(\boldsymbol{\beta}_0))} &= \sum_{i=1}^n \frac{f(X_i; \theta_{0h}, \sigma_{0h})}{f(X_i; \Psi_0)} + O_p(n) \sum_{h=1}^{m_0} |\alpha_h^{(k_0)} - \alpha_{0h}| \\
&\quad + O_p(n) \sum_{h=1}^{m_0} \sum_{j=1}^{2} \{|\theta_{jh}^{(k_0)} - \theta_{0h}| + |(\sigma_{jh}^{(k_0)})^2 - \sigma_{0h}^2|\} \\
&= \sum_{i=1}^n \frac{f(X_i; \theta_{0h}, \sigma_{0h})}{f(X_i; \Psi_0)} + o_p(n) \\
&= n(1 + o_p(1)).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{i=1}^n w_{i1h}^{(k_0)} &= \alpha_h^{(k_0)} \beta_h^{(k_0)} \sum_{i=1}^n \frac{f(X_i; \theta_{1h}^{(k_0)}, \sigma_{1h}^{(k_0)})}{f(X_i; \Psi^{(k_0)}(\boldsymbol{\beta}_0))} \\
&= n\alpha_h^{(k_0)} \beta_h^{(k_0)} (1 + o_p(1)) \qquad (4)
\end{aligned}$$

and hence

$$\hat{\beta}_h = \frac{\sum_{i=1}^n w_{i1h}^{(k_0)}}{n\alpha_n^{(k_0)}} = \beta_h^{(k_0)} + o_p(1) = \beta_{0h} + o_p(1). \qquad (5)$$

The next step is to show $\beta_h^{(k_0+1)} - \hat{\beta}_h = o_p(1)$. Note that $H_{nh}(\beta)$ attains its maximum at $\hat{\beta}_h$ and it is monotone on both sides. Hence, for any $\epsilon > 0$ and $\beta > \hat{\beta}_h + \epsilon$, we have

$$H_{nh}(\beta) - H_{nh}(\hat{\beta}_h) \leq H_{nh}(\hat{\beta}_h + \epsilon) - H_{nh}(\hat{\beta}_h).$$

Thus, by (4) and (5), we have

$$\begin{aligned} n^{-1}\{H_{nh}(\hat\beta_h+\epsilon)-H_{nh}(\hat\beta_h)\} &= \left[n^{-1}\sum_{i=1}^{n}\omega_{i1h}^{(k_0)}\right]\{\log(\hat\beta_h+\epsilon)-\log(\hat\beta_h)\} \\ &\quad+\left[n^{-1}\sum_{i=1}^{n}\omega_{i2h}^{(k_0)}\right]\{\log(1-\hat\beta_h-\epsilon)-\log(1-\hat\beta_h)\} \\ &\to c \end{aligned}$$

in probability for some $c<0$. Therefore, $H_{nh}(\beta)-H_{nh}(\hat\beta_h)\to-\infty$ in probability uniformly for any $\beta>\hat\beta_h+\epsilon$. Hence, by Condition C1, with probability approaching one,

$$Q_{nh}(\beta)-Q_{nh}(\hat\beta_h)\le p(0.5)-p(\hat\beta_h)+H_{nh}(\beta)-H_{nh}(\hat\beta_h)\to-\infty$$

uniformly for any $\beta>\hat\beta_h+\epsilon$. Then we must have $\beta_h^{(k_0+1)}\le\hat\beta_h+\epsilon$ in probability. Similarly, we can show that $\beta_h^{(k_0+1)}\ge\hat\beta_h-\epsilon$ in probability. Therefore, we have $\beta_h^{(k_0+1)}-\hat\beta_h=o_p(1)$ as claimed. □

The next lemma proves that the EM-iterations keep the iterated values of $\boldsymbol\alpha$, $\boldsymbol\theta_1$, $\boldsymbol\theta_2$, $\boldsymbol\sigma_1$, and $\boldsymbol\sigma_2$ in small neighborhoods of $\boldsymbol\alpha_0$, $\boldsymbol\theta_0$, $\boldsymbol\theta_0$, $\boldsymbol\sigma_0$, and $\boldsymbol\sigma_0$, respectively.

Lemma 3. *Assume the conditions of Lemma 2. Under the null distribution $f(x;\Psi_0)$, we have*

$$\boldsymbol\alpha^{(k_0+1)}-\boldsymbol\alpha_0=o_p(1),\ \boldsymbol\beta^{(k_0+1)}-\boldsymbol\beta_0=o_p(1),\ \boldsymbol\theta_1^{(k_0+1)}-\boldsymbol\theta_0=o_p(1),$$

$$\boldsymbol\theta_2^{(k_0+1)}-\boldsymbol\theta_0=o_p(1),\ \boldsymbol\sigma_1^{(k_0+1)}-\boldsymbol\sigma_0=o_p(1),\ \boldsymbol\sigma_2^{(k_0+1)}-\boldsymbol\sigma_0=o_p(1).$$

Proof. We first define a mixing distribution obtained after a partial EM-iteration. For each $l=1,2,\ldots,m_0$, let

$$\begin{aligned} \Psi_l^{(k_0+1)}(\theta,\sigma) &= \sum_{h\ne l}\alpha_h^{(k_0)}\{\beta_h^{(k_0)}I(\theta_{1h}^{(k_0)}\le\theta,\sigma_{1h}^{(k_0)}\le\sigma)+(1-\beta_h^{(k_0)})I(\theta_{2h}^{(k_0)}\le\theta,\sigma_{2h}^{(k_0)}\le\sigma)\} \\ &\quad+\alpha_l^{(k_0)}\{\beta_l^{(k_0)}I(\theta_{1l}^{(k_0+1)}\le\theta,\sigma_{1l}^{(k_0+1)}\le\sigma)+(1-\beta_l^{(k_0)})I(\theta_{2l}^{(k_0+1)}\le\theta,\sigma_{2l}^{(k_0+1)}\le\sigma)\}. \end{aligned}$$

That is, the two mixing distributions $\Psi_l^{(k_0+1)}(\theta,\sigma)$ and $\Psi^{(k_0)}(\theta,\sigma)$ are identical except for the lth pair of support points, which have been updated by the EM-iteration. For convenience, we have omitted their dependence on $\boldsymbol{\beta}_0$ in the above definition. Because the EM-iteration always increases the likelihood (Dempster, Laird, and Rubin, 1977; Wu, 1981), we have

$$pl_n(\Psi_l^{(k_0+1)}) \geq pl_n(\Psi^{(k_0)}) \geq pl_n(\Psi^{(1)}).$$

Using the inequality in (3), this further implies that

$$pl_n(\Psi_l^{(k_0+1)}) \geq l_n(\Psi_0) + o(n).$$

For the same reason as in the proof of Lemma 1, the above result implies the consistency of $\Psi_l^{(k+1)}$ for Ψ_0. Since $\boldsymbol{\alpha}^{(k_0)}$, $\boldsymbol{\beta}^{(k_0)}$, $\boldsymbol{\theta}_1^{(k_0)}$, $\boldsymbol{\theta}_2^{(k_0)}$, $\boldsymbol{\sigma}_1^{(k_0)}$, and $\boldsymbol{\sigma}_2^{(k_0)}$ are consistent as assumed, the consistency of $\Psi_l^{(k_0+1)}$ is possible only if

$$\theta_{1l}^{(k_0+1)} = \theta_{0l} + o_p(1),\ \theta_{2l}^{(k_0+1)} = \theta_{0l} + o_p(1),\ \sigma_{1l}^{(k_0+1)} = \sigma_{0l} + o_p(1),\ \sigma_{2l}^{(k_0+1)} = \sigma_{0l} + o_p(1).$$

Because this result is applicable to all $l = 1, 2, \ldots, m_0$, we have the consistency properties for $\boldsymbol{\theta}_1^{(k_0+1)}$, $\boldsymbol{\theta}_2^{(k_0+1)}$, $\boldsymbol{\sigma}_1^{(k_0+1)}$, and $\boldsymbol{\sigma}_2^{(k_0+1)}$.

Next, we apply the same idea to the completely updated $\Psi^{(k_0+1)}$ and conclude that it too must be a consistent estimator of Ψ_0. Because

$$\boldsymbol{\theta}_1^{(k_0+1)} = \boldsymbol{\theta}_0 + o_p(1),\ \boldsymbol{\theta}_2^{(k_0+1)} = \boldsymbol{\theta}_0 + o_p(1),\ \boldsymbol{\sigma}_1^{(k_0+1)} = \boldsymbol{\sigma}_0 + o_p(1),\ \boldsymbol{\sigma}_2^{(k_0+1)} = \boldsymbol{\sigma}_0 + o_p(1),$$

and because $\boldsymbol{\beta}^{(k_0+1)} = \boldsymbol{\beta}_0 + o_p(1)$ as shown in the previous Lemma, the overall consistency of $\Psi^{(k_0+1)}$ implies that $\boldsymbol{\alpha}^{(k_0+1)} = \boldsymbol{\alpha}_0 + o_p(1)$. This completes the proof. □

With the proof of Lemma 3, we have completed the mathematical induction. Thus, we have shown Proposition 1.

Proof of Theorem 1

We need some notation. For $h = 1, 2, \ldots, m_0$, let

$$
\begin{aligned}
\Delta_{ih} &= \frac{f(X_i;\theta_{0h},\sigma_{0h}) - f(X_i;\theta_{0m_0})}{f(X_i;\Psi_0)},\\
Y_{ih} &= \frac{\partial f(X_i;\theta_{0h},\sigma_{0h})/\partial\theta}{f(X_i;\Psi_0)},\\
Z_{ih} &= \frac{\partial^2 f(X_i;\theta_{0h},\sigma_{0h})/\partial\theta^2}{f(X_i;\Psi_0)},\\
U_{ih} &= \frac{\partial^3 f(X_i;\theta_{0h},\sigma_{0h})/\partial\theta^3}{f(X_i;\Psi_0)},\\
V_{ih} &= \frac{\partial^4 f(X_i;\theta_{0h},\sigma_{0h})/\partial\theta^4}{f(X_i;\Psi_0)}.
\end{aligned}
$$

The above quantities represent derivatives of $f(X_i;\theta,\sigma)$ with respect to σ^2. It is easy to verify that

$$
\begin{aligned}
Z_{ih} &= \frac{2\partial f(X_i;\theta_{0h},\sigma_{0h})/\partial(\sigma^2)}{f(X_i;\Psi_0)},\\
U_{ih} &= \frac{2\partial^2 f(X_i;\theta_{0h},\sigma_{0h})/\partial\theta\partial(\sigma^2)}{f(X_i;\Psi_0)},\\
V_{ih} &= \frac{2\partial^3 f(X_i;\theta_{0h},\sigma_{0h})/\partial\theta^2\partial(\sigma^2)}{f(X_i;\Psi_0)} = \frac{4\partial^2 f(X_i;\theta_{0h},\sigma_{0h})/\partial(\sigma^2)^2}{f(X_i;\Psi_0)}.
\end{aligned}
$$

The above relationships are the reasons behind the loss of strong identifiability (Chen, 1995) of finite normal mixture models. Many asymptotic results on finite mixture models, such as those in Dacunha-Castelle and Gassiat (1999) and Liu and Shao (2003), are not applicable to finite normal mixture models. Denote

$$
\begin{aligned}
R_{1n}(\Psi^{(K)}(\boldsymbol{\beta}_0)) &= 2\{l_n(\Psi^{(K)}(\boldsymbol{\beta}_0)) - l_n(\Psi_0)\},\\
R_{2n} &= 2\{l_n(\Psi_0) - l_n(\hat{\Psi}_0)\},\\
R_{3n}(\Psi^{(K)}(\boldsymbol{\beta}_0)) &= 2\left[\sum_{h=1}^{m_0}\{p_n((\sigma_{1h}^{(K)})^2;\hat{\sigma}_{0h}^2) + p_n((\sigma_{2h}^{(K)})^2;\hat{\sigma}_{0h}^2)\} + \sum_{h=1}^{m_0} p(\beta_h^{(K)})\right].
\end{aligned}
$$

Then

$$M_n^{(K)}(\boldsymbol{\beta}_0) = R_{1n}(\Psi^{(K)}(\boldsymbol{\beta}_0)) + R_{2n} + R_{3n}(\Psi^{(K)}(\boldsymbol{\beta}_0)).$$

We now derive a quadratic approximation for $R_{1n}(\Psi^{(K)}(\boldsymbol{\beta}_0))$. We proceed in two steps: first finding an upper bound, and then showing that this upper bound is attained.

Set

$$R_{1n}(\Psi^{(K)}(\boldsymbol{\beta}_0)) = 2\sum_{i=1}^{n} \log(1+\delta_i)$$

with

$$\begin{aligned}
\delta_i &= \frac{f(X_i;\Psi^{(K)}(\boldsymbol{\beta}_0)) - f(X_i;\Psi_0)}{f(X_i;\Psi_0)} \\
&= \sum_{h=1}^{m_0}(\alpha_h^{(K)} - \alpha_{0h})\frac{f(X_i;\theta_{0h},\sigma_{0h})}{f(X_i;\Psi_0)} + \sum_{h=1}^{m_0}\alpha_h^{(K)}\beta_h^{(K)}\frac{f(X_i;\theta_{1h}^{(K)},\sigma_{1h}^{(K)}) - f(X_i;\theta_{0h},\sigma_{0h})}{f(X_i;\Psi_0)} \\
&\quad + \sum_{h=1}^{m_0}\alpha_h^{(K)}(1-\beta_h^{(K)})\frac{f(X_i;\theta_{2h}^{(K)},\sigma_{2h}^{(K)}) - f(X_i;\theta_{0h},\sigma_{0h})}{f(X_i;\Psi_0)}. \qquad (6)
\end{aligned}$$

Since $\sum_{h=1}^{m_0}\alpha_h^{(K)} = \sum_{h=1}^{m_0}\alpha_{0h} = 1$, we have

$$\sum_{h=1}^{m_0}(\alpha_h^{(K)} - \alpha_{0h})\frac{f(X_i;\theta_{0h},\sigma_{0h})}{f(X_i;\Psi_0)} = \sum_{h=1}^{m_0-1}(\alpha_h^{(K)} - \alpha_{0h})\frac{f(X_i;\theta_{0h},\sigma_{0h}) - f(X_i;\theta_{0m_0},\sigma_{0m_0})}{f(X_i;\Psi_0)}. \qquad (7)$$

For $h = 1,\ldots,m_0$, $l = 0,1,2,3,4$, and $s = 0,1,2,3,4$, we define

$$m_{h,ls} = \alpha_h^{(K)}\beta_h^{(K)}(\theta_{1h}^{(K)} - \theta_{0h})^l\{(\sigma_{1h}^{(K)})^2 - \sigma_{0h}^2\}^s + \alpha_h^{(K)}(1-\beta_h^{(K)})(\theta_{2h}^{(K)} - \theta_{0h})^l\{(\sigma_{2h}^{(K)})^2 - \sigma_{0h}^2\}^s.$$

Denoting

$$f^{(l,s)}(x;\theta,\sigma) = \frac{\partial^{l+s} f(x;\theta,\sigma)}{\partial\theta^l \partial(\sigma^2)^s},$$

and expanding $f(X_i;\theta_{jh}^{(K)},\sigma_{jh}^{(K)})$ for $j=1,2$ to order 4, we find

$$\begin{aligned}
&\sum_{h=1}^{m_0}\alpha_h^{(K)}\beta_h^{(K)}\frac{f(X_i;\theta_{1h}^{(K)},\sigma_{1h}^{(K)}) - f(X_i;\theta_{0h},\sigma_{0h})}{f(X_i;\Psi_0)} + \sum_{h=1}^{m_0}\alpha_h^{(K)}(1-\beta_h^{(K)})\frac{f(X_i;\theta_{2h}^{(K)},\sigma_{2h}^{(K)}) - f(X_i;\theta_{0h},\sigma_{0h})}{f(X_i;\Psi_0)} \\
&= \sum_{h=1}^{m_0}\sum_{l+s=1}^{4}\binom{l+s}{s}\frac{m_{h,ls}}{(l+s)!}\frac{f^{(l,s)}(X_i;\theta_{0h},\sigma_{0h})}{f(X_i;\Psi_0)} + \epsilon_{in}^{(1)},
\end{aligned}$$

where $\epsilon_{in}^{(1)}$ is the remainder term. Absorbing the terms $m_{h,ls}$ with $l+2s \geq 5$ into the remainder term, we further have

$$\begin{aligned}
&\sum_{h=1}^{m_0} \alpha_h^{(K)} \beta_h^{(K)} \frac{f(X_i; \theta_{1h}^{(K)}, \sigma_{1h}^{(K)}) - f(X_i; \theta_{0h}, \sigma_{0h})}{f(X_i; \Psi_0)} + \sum_{h=1}^{m_0} \alpha_h^{(K)} (1 - \beta_h^{(K)}) \frac{f(X_i; \theta_{2h}^{(K)}, \sigma_{2h}^{(K)}) - f(X_i; \theta_{0h}, \sigma_{0h})}{f(X_i; \Psi_0)} \\
&= \sum_{h=1}^{m_0} \sum_{l+2s=1}^{4} \binom{l+s}{s} \frac{m_{h,ls}}{(l+s)!} \frac{f^{(l,s)}(X_i; \theta_{0h}, \sigma_{0h})}{f(X_i; \Psi_0)} + \epsilon_{in}. \qquad (8)
\end{aligned}$$

Combining (6), (7), and (8) plus the notational substitution of Δ_{ih}, Y_{ih}, Z_{ih}, U_{ih}, and V_{ih}, we get

$$\delta_i = \sum_{h=1}^{m_0-1} (\alpha_h^{(K)} - \alpha_{0h}) \Delta_{ih} + \sum_{h=1}^{m_0} \{t_{1h} Y_{ih} + t_{2h} Z_{ih} + t_{3h} U_{ih} + t_{4h} V_{ih}\} + \epsilon_{in}$$

with $t_{1h} = m_{h,10}$, $t_{2h} = m_{h,20} + m_{h,01}$, $t_{3h} = m_{h,30} + 3m_{h,11}$, and $t_{4h} = m_{h,40} + 6m_{h,21} + 3m_{h,02}$. It follows that

$$\sum_{i=1}^{n} \delta_i = \sum_{i=1}^{n} \left[\sum_{h=1}^{m_0-1} (\alpha_h^{(K)} - \alpha_{0h}) \Delta_{ih} + \sum_{h=1}^{m_0} \{t_{1h} Y_{ih} + t_{2h} Z_{ih} + t_{3h} U_{ih} + t_{4h} V_{ih}\} \right] + \epsilon_n \qquad (9)$$

with, using techniques similar to those in Chen and Li (2008),

$$|\epsilon_n| = \left| \sum_{i=1}^{n} \epsilon_{in} \right| = O_p(n^{1/2}) \sum_{h=1}^{m_0} \sum_{j=1}^{2} \{|\theta_{jh}^{(K)}|^5 + |\theta_{jh}^{(K)}|((\sigma_{jh}^{(K)})^2 - \sigma_{0h}^2)^2 + |(\sigma_{jh}^{(K)})^2 - \sigma_{0h}^2|^3\}.$$

With Lemma 8 of Chen and Li (2008), we further have

$$\sum_{h=1}^{m_0} \sum_{j=1}^{2} \{|\theta_{jh}^{(K)}|^5 + |\theta_{jh}^{(K)}|((\sigma_{jh}^{(K)})^2 - \sigma_{0h}^2)^2 + |(\sigma_{jh}^{(K)})^2 - \sigma_{0h}^2|^3\} = o_p\left(\sum_{h=1}^{m_0} \sum_{l=1}^{4} |t_{lh}| \right).$$

Therefore

$$|\epsilon_n| = o_p(n^{1/2}) \left(\sum_{h=1}^{m_0} \sum_{l=1}^{4} |t_{lh}| \right). \qquad (10)$$

Let

$$
\begin{aligned}
\mathbf{t}_1 &= \left(\alpha_1^{(K)} - \alpha_{01}, \ldots, \alpha_{m_0-1}^{(K)} - \alpha_{0m_0-1}, t_{11}, \ldots, t_{1m_0}, t_{21}, \ldots, t_{2m_0}\right)^\tau, \\
\mathbf{t}_2 &= (t_{31}, \ldots, t_{3m_0}, t_{41}, \ldots, t_{4m_0})^\tau, \\
\mathbf{b}_{1i} &= (\Delta_{i1}, \ldots, \Delta_{im_0-1}, Y_{i1}, \ldots, Y_{im_0}, Z_{i1}, \ldots, Y_{im_0})^\tau, \\
\mathbf{b}_{2i} &= (U_{i1}, \ldots, U_{im_0}, V_{i1}, \ldots, V_{im_0})^\tau.
\end{aligned}
$$

Further, set $\mathbf{t} = (\mathbf{t}_1^\tau, \mathbf{t}_2^\tau)^\tau$ and $\mathbf{b}_i = (\mathbf{b}_{1i}^\tau, \mathbf{b}_{2i}^\tau)^\tau$. Then (9) becomes

$$
\sum_{i=1}^n \delta_i = \sum_{i=1}^n \mathbf{t}^\tau \mathbf{b}_i + \epsilon_n \tag{11}
$$

with the order of ϵ_n assessed by (10).

By $\log(1+x) \leq x - x^2/2 + x^3/3$, we have

$$
\begin{aligned}
R_{1n}(\Psi^{(K)}(\boldsymbol{\beta}_0)) &= 2\sum_{i=1}^n \log(1+\delta_i) \\
&\leq 2\sum_{i=1}^n \delta_i - \sum_{i=1}^n \delta_i^2 + 2/3\sum_{i=1}^n \delta_i^3 \\
&= 2\sum_{i=1}^n \mathbf{t}^\tau \mathbf{b}_i - \sum_{i=1}^n (\mathbf{t}^\tau \mathbf{b}_i)^2 + 2/3\sum_{i=1}^n (\mathbf{t}^\tau \mathbf{b}_i)^3 + O_p(\epsilon_n)
\end{aligned}
$$

in which the expansions of the quadratic and cubic terms can be done similarly to (11).

Set $\mathbf{B} = \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i^\tau$. Then the quadratic term

$$
\sum_{i=1}^n (\mathbf{t}^\tau \mathbf{b}_i)^2 = \mathbf{t}^\tau \mathbf{B} \mathbf{t},
$$

the cubic term

$$
\sum_{i=1}^n (\mathbf{t}^\tau \mathbf{b}_i)^3 = o_p(n) \mathbf{t}^\tau \mathbf{t}.
$$

Applying $|2x| \le 1 + x^2$, the order assessment (10) implies that

$$|\epsilon_n| \le o_p(1) + o_p(n) \sum_{h=1}^{m_0} \sum_{l=1}^{4} t_{lh}^2 = o_p(1) + o_p(n)\mathbf{t}^\tau \mathbf{t}.$$

Combining the above order assessments, the upper bound becomes

$$\begin{aligned} R_{1n}(\Psi^{(K)}(\boldsymbol{\beta}_0)) &\le 2\mathbf{t}^\tau \sum_{i=1}^{n} \mathbf{b}_i - \mathbf{t}^\tau \mathbf{B}\mathbf{t}\{1 + o_p(1)\} + o_p(1) \\ &\le \left(\sum_{i=1}^{n} \mathbf{b}_i\right)^\tau \mathbf{B}^{-1} \left(\sum_{i=1}^{n} \mathbf{b}_i\right) + o_p(1). \end{aligned}$$

Note that R_{2n} is the minus likelihood ratio statistic defined under the assumption that the order of the finite normal mixture model is correctly specified. Hence, it can easily be expanded as

$$R_{2n} = -\left(\sum_{i=1}^{n} \mathbf{b}_{1i}\right)^\tau \mathbf{B}_{11}^{-1} \left(\sum_{i=1}^{n} \mathbf{b}_{1i}\right) + o_p(1)$$

with $\mathbf{B}_{11} = \sum_{i=1}^{n} \mathbf{b}_{1i}\mathbf{b}_{1i}^\tau$.

Since two penalty functions are non-positive, we have

$$R_{3n}(\Psi^{(K)}(\boldsymbol{\beta}_0)) \le 0.$$

Therefore,

$$\begin{aligned} M_n^{(K)}(\boldsymbol{\beta}_0) &= R_{1n}(\Psi^{(K)}(\boldsymbol{\beta}_0)) + R_{2n} + R_{3n}(\Psi^{(K)}(\boldsymbol{\beta}_0)) \\ &\le \left(\sum_{i=1}^{n} \mathbf{b}_i\right)^\tau \mathbf{B}^{-1} \left(\sum_{i=1}^{n} \mathbf{b}_i\right) - \left(\sum_{i=1}^{n} \mathbf{b}_{1i}\right)^\tau \mathbf{B}_{11}^{-1} \left(\sum_{i=1}^{n} \mathbf{b}_{1i}\right) + o_p(1). \end{aligned}$$

Set

$$\tilde{\mathbf{b}}_{2i} = \mathbf{b}_{2i} - \left(\sum_{i=1}^{n} \mathbf{b}_{1i}\mathbf{b}_{2i}^\tau\right)^\tau \mathbf{B}_{11}^{-1}\mathbf{b}_{1i},$$

and $\tilde{\mathbf{B}}_{22} = \sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}\tilde{\mathbf{b}}_{2i}^{\tau}$. After some algebra work, we find that the upper bound for $M_n^{(K)}(\boldsymbol{\beta}_0)$ reduces to

$$M_n^{(K)}(\boldsymbol{\beta}_0) \le \left(\sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}\right)^{\tau} \tilde{\mathbf{B}}_{22}^{-1} \left(\sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}\right) + o_p(1).$$

We note that the above upper bound also serves as the upper bound for $EM_n^{(K)}$. Therefore,

$$EM_n^{(K)}(\boldsymbol{\beta}_0) \le \left(\sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}\right)^{\tau} \tilde{\mathbf{B}}_{22}^{-1} \left(\sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}\right) + o_p(1).$$

Now we show that $EM_n^{(K)}$ asymptotically attains this upper bound. Since the EM-iteration increases $pl_n(\Psi)$, we need only show that this is the case when $k = 1$. Consider the special choice of $\boldsymbol{\beta}_0$ with $\beta_{0h} = 0.5$ for $h = 1, \ldots, m_0$. It suffices to show that we can find a $\tilde{\Psi} \in \Omega_{2m_0}(\boldsymbol{\beta}_0)$ such that

$$2\{pl_n(\tilde{\Psi}) - l_n(\hat{\Psi}_0)\} = \left(\sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}\right)^{\tau} \tilde{\mathbf{B}}_{22}^{-1} \left(\sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}\right) + o_p(1).$$

This is equivalent to finding a set of values such that

$$\mathbf{t} = \mathbf{B}^{-1} \left(\sum_{i=1}^{n} \mathbf{b}_i\right) + o_p(n^{-1/2}).$$

Recall that $\mathbf{t}$ is defined through parameters in $\tilde{\Psi}$. With $2m_0$ support points built in, there are $5m_0 - 1$ parameters in $\tilde{\Psi}$ involved in the definition of $\mathbf{t}$, not including $\boldsymbol{\beta}_0$. Hence, such a solution can be shown to exist; see Chen and Li (2008, page 38) for a detailed proof. It is easy to verify that such a choice satisfies

$$\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = O_p(n^{-1/2}),\ \tilde{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0 = O_p(n^{-1/8}),\ \tilde{\boldsymbol{\sigma}}_j - \boldsymbol{\sigma}_0 = O_p(n^{-1/4}),$$

for $j = 1, 2$. Here $\tilde{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{\theta}}_1$, $\tilde{\boldsymbol{\theta}}_2$, $\tilde{\boldsymbol{\sigma}}_1$, and $\tilde{\boldsymbol{\sigma}}_2$ are constituent entries of $\tilde{\Psi}$. The above order assessment information leads to the expansion

$$R_{1n}(\tilde{\Psi}) = \left(\sum_{i=1}^{n} \mathbf{b}_i\right)^{\tau} \mathbf{B}^{-1} \left(\sum_{i=1}^{n} \mathbf{b}_i\right) + o_p(1).$$

Further, by Condition C5 on the penalty function $p_n(\cdot;\cdot)$, we have

$$R_{3n}(\tilde{\Psi}) = o_p(1).$$

With the quadratic approximation of R_{2n}, we get

$$2\{pl_n(\tilde{\Psi}) - l_n(\hat{\Psi}_0)\} = \left(\sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}\right)^{\tau} \tilde{\mathbf{B}}_{22}^{-1} \left(\sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}\right) + o_p(1).$$

Therefore,

$$EM_n^{(K)} \geq EM_n^{(1)} \geq 2\{pl_n(\tilde{\Psi}) - l_n(\hat{\Psi}_0)\} = \left(\sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}\right)^{\tau} \tilde{\mathbf{B}}_{22}^{-1} \left(\sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}\right) + o_p(1).$$

Combining the lower and upper bounds, we arrive at

$$EM_n^{(K)} = \left(\sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}\right)^{\tau} \tilde{\mathbf{B}}_{22}^{-1} \left(\sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}\right) + o_p(1).$$

The result of Theorem 1 follows from the asymptotic normality of $\tilde{\mathbf{B}}_{22}^{-1/2} \sum_{i=1}^{n} \tilde{\mathbf{b}}_{2i}$.

2 Use of R-function *emtest.norm()*

The input of *emtest.norm()* has two parts: (1) a vector containing assumed iid observations from a finite normal mixture model; (2) the null order. It can be used to test the null hypothesis of order one, two, or three. The first component of its output is a vector of values for $EM_n^{(1)}$, $EM_n^{(2)}$, and $EM_n^{(3)}$. The second component is a vector of corresponding P-values.

Suppose that the SLC data is contained in file *slc.txt* in the same working directory as R. Testing the null order of $m = 2$ can be accomplished by a single command:

```
> emtest.norm(read.table("slc.txt")[,1], 2)
```

```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Testing H_0: order= 2 under normal mixture model without equal variance assumption
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
EM-test statistics: 4.595385 4.637419 4.656803


p-values: 0.3313866 0.3265626 0.3243574
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

Similarly, testing the null order of $m = 3$ can be accomplished by a single command:

```
> emtest.norm(read.table("slc.txt")[,1], 3)
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Testing H_0: order= 3 under normal mixture model without equal variance assumption
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
EM-test statistics: 4.627216 4.71304 4.74831


p-values: 0.5924334 0.5811087 0.5764755
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

The data sets used in the examples and the source R code are in the following files provided separately:

1. *acidity.txt*: contains the lake chemistry data;

2. *analyze.R*: contains the R code for analyzing all four application examples in Section 5 of the paper;

3. *emnorm.R*: contains all the source R code including the R function *emtest.norm()*;

4. *prostz.txt*: contains the 6033 z-scores for the prostate cancer data;

5. *slc.txt*: contains the SLC data;

6. *winesugar.txt*: contains the adulteration data.

References

Chen, J. (1995) Optimal rate of convergence in finite mixture models. *The Annals of Statistics*, **23**, 221-234.

Chen, J. and Li, P. (2008) Homogeneity test in normal mixture models: The EM approach. *Techinical Report.* University of British Columbia.

Chen, J., Tan, X., and Zhang, R. (2008) Inference for normal mixtures in mean and variance. *Statistica Sinica*, **18**, 443-465.

Dacunha-Castelle, D. and Gassiat, E. (1999) Testing the order of a model using locally conic parametrization: Population mixtures and stationary ARMA processes. *The Annals of Statistics*, **27**, 1178-1209.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) Maximum likelihood from incomplete data via EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.

Liu, X. and Shao, Y. (2003) Asymptotics for likelihood ratio tests under loss of identifiability. *The Annals of Statistics*, **31**, 807-832.

Wu, C. F. J. (1981) On the convergence properties of the EM algorithm. *The Annals of Statistics*, **11**, 95-103.