PENALIZED LIKELIHOOD FOR LOGISTIC-NORMAL MIXTURE MODELS WITH UNEQUAL VARIANCES

Juan Shen[1], Yingchuan Wang[2], and Xuming He[3]

[1]*Fudan University and* [2,3]*University of Michigan*

Supplementary Material

In this supplement, we provide the proofs for **Theorem 1** and **Theorem 2**, and give more details for the analysis of the data in Section 4 in the main paper.

S1 Proofs

We first provide a useful lemma which is proved in the end of this section.

Lemma 1. *Suppose* $\{(P_k, Q_k)\}_{k=1}^{\infty}$ *are i.i.d. continuous random variables with finite means. Also suppose that the density of* Q_k *and the conditional density of* $P_k|Q_k$ *are bounded by C, then uniformly in* σ_n *between* n^{-1} *and* e^{-1}*, there exists a constant* C^* *such that for sufficiently large n,*

$$P\Big(\sup_{a,b\in R} n^{-1}\sum_{k=1}^{n} 1\Big(|P_k - aQ_k - b| \le |\sigma_n \log \sigma_n|\Big) > C^*|\sigma_n \log \sigma_n|\Big) \le Cn^{-2}.$$

S1.1 Proof of Theorem 1

We note that **S1** and **S2** play the role of **Lemma 1** in Chen et al. (2008). With these two properties, it then follows from the Borel-Cantelli Lemma that as $n \to \infty$ and almost surely,

1. *for each given σ between n^{-1} and e^{-1},*

$$\sup_{\boldsymbol{\beta} \in R^{q_1}} n^{-1} \sum_{i=1}^{n} 1\Big(|Y_i - \boldsymbol{Z}_i^T \boldsymbol{\beta}| \leq |\sigma \log \sigma|\Big) \leq C|\sigma \log \sigma|,$$

2. *uniformly for σ between 0 and n^{-1},*

$$\sup_{\boldsymbol{\beta} \in R^{q_1}} n^{-1} \sum_{i=1}^{n} 1\Big(|Y_i - \boldsymbol{Z}_i^T \boldsymbol{\beta}| \leq |\sigma \log \sigma|\Big) \leq 4(\log n)^2/n.$$

These almost sure results are stated for a given σ. However, following the arguments in **Lemma 2** of Chen et al. (2008), we have a stronger result as follows.

Except for a zero-probability event not depending on σ, we have for all large enough n:

1. *for σ between n^{-1} and e^{-1},* $\sup_{\boldsymbol{\beta} \in R^{q_1}} n^{-1} \sum_{i=1}^{n} 1\Big(|Y_i - \boldsymbol{Z}_i^T \boldsymbol{\beta}| \leq |\sigma \log \sigma|\Big) \leq C|\sigma \log \sigma|$,

2. *for σ between 0 and n^{-1},* $\sup_{\boldsymbol{\beta} \in R^{q_1}} n^{-1} \sum_{i=1}^{n} 1\Big(|Y_i - \boldsymbol{Z}_i^T \boldsymbol{\beta}| \leq |\sigma \log \sigma|\Big) \leq 4(\log n)^2/n.$

We partition the parameter space with respect to σ as in Chen et al. (2008).

Let $\Gamma_1 = \{\mathbf{\Theta} : \sigma_1 \leq \sigma_2 \leq \epsilon_0\}$, $\Gamma_2 = \{\mathbf{\Theta} : \sigma_1 \leq \tau_0, \sigma_2 \geq \epsilon_0\}$, $\Gamma_3 = \Gamma - (\Gamma_1 \bigcup \Gamma_2)$, where ϵ_0, τ_0 and Γ are specified in Chen et al. (2008). Note that $\boldsymbol{Z}_i^T\boldsymbol{\beta}$ in our setting plays the same role as θ in Chen et al. (2008), where the model has no covariates. Hence, with the above almost surely results and **Theorem 1** and **Theorem 2** of Chen et al. (2008), we have as $n \to \infty$ and almost surely, the penalized maximum likelihood estimators of our model will be attained in Γ_3. Note that σ is bounded away from *zero* in Γ_3, standard techniques of proving the consistency of the maximum likelihood estimators lead to the consistency of our proposed penalized maximum likelihood estimators.

Next, we show **S1** and **S2**. Since the proof of **S2** is essentially the same as that for **S1**, we only provide the details of the proof of **S1**. For convenience, we allow the constants used in the proofs vary line by line.

Recall that $\boldsymbol{Z} = (1, \boldsymbol{U}, \boldsymbol{V})$, where 1 represents the intercept in the model and $\boldsymbol{U}$ consists of only discrete variables with a finite sample space and $\boldsymbol{V}$ consists of only continuous variables. We prove **S1** for the following three cases.

Case 1: If $\boldsymbol{Z}$ only has three dimensions, that is, $\boldsymbol{Z} = (1, U, V)$. Further, we assume $U \sim Ber(1/2)$.

Case 2: If $\boldsymbol{Z} = (1, \boldsymbol{U}, V)$, where $\boldsymbol{U}$ is a random vector taking any finite values and V is one dimensional continuous variable.

Case 3: If $\boldsymbol{Z} = (1, \boldsymbol{U}, \boldsymbol{V})$, where $\boldsymbol{U}$ is a random vector taking finite values and $\boldsymbol{V}$

ia a vector of continuous random variables.

From **Case 1** to **Case 3**, we will prove **S1** from the simplest case to the most general situation. Then we complete the proof of **Theorem 1**.

Next we provide the detailed proof under Cases 1-3.

Proof for Case 1: We prove **S1** when $\boldsymbol{Z}$ only has three dimensions, that is, $\boldsymbol{Z} = (1, U, V)$. Further, we assume $U \sim Ber(1/2)$.

Let $\bar{U}_n = n^{-1} \sum_{i=1}^{n} U_i$ and let $f_X(x)$ *and* $f_{X|Y}(x|y)$ denote the density of X and the conditional density of $X|Y$, respectively. Then, for any given $\sigma_n \in (n^{-1}, e^{-1})$, let

$$\epsilon_n = \{n^{-1}(8 \log n)\}^{1/2},$$

$$I = P\Big(\sup_{\boldsymbol{\beta} \in R^3} W_n(\boldsymbol{\beta}) > C|\sigma_n \log \sigma_n| \Big| |\bar{U}_n - 1/2| \le \epsilon_n \Big),$$

$$II = P(|\bar{U}_n - 1/2| > \epsilon_n).$$

We have,

$$\begin{aligned} P(A_n(C)) &= P\Big(\sup_{\boldsymbol{\beta} \in R^3} W_n(\boldsymbol{\beta}) > C|\sigma_n \log \sigma_n| \Big) \\ &\le P\Big(\sup_{\boldsymbol{\beta} \in R^3} W_n(\boldsymbol{\beta}) > C|\sigma_n \log \sigma_n| \Big| |\bar{U}_n - 1/2| \le \epsilon_n \Big) + \\ &\quad P(|\bar{U}_n - 1/2| > \epsilon_n) \\ &= I + II. \end{aligned} \tag{S1.1}$$

We verify the following two claims:

CL1 $II \le Cn^{-2}$;

CL2 $I \le Cn^{-2}$.

Proof of CL1: By Bernstein's inequality, for sufficient large n,

$$II = 2P\Big(\sum_{i=1}^{n} U_i/n - 1/2 > \epsilon_n\Big) \le \exp\Big\{-\frac{n^2\epsilon_n^2/2}{n + n\epsilon_n/3}\Big\} \le Cn^{-2}.$$

Proof of CL2: Note that

$$\begin{aligned} I &= P\Big(\sup\nolimits_{\beta\in R^3} W_n(\boldsymbol{\beta}) > C|\sigma_n \log\sigma_n|\Big||\bar{U}_n - 1/2| \le \epsilon_n\Big) \\ &= \sum\nolimits_{u_1,\cdots,u_n} P\Big(\sup\nolimits_{\beta\in R^3} W_n(\boldsymbol{\beta}) > C|\sigma_n \log\sigma_n|\Big|U_1 = u_1,\cdots,U_n = u_n, |\bar{U}_n - 1/2| \le \epsilon_n\Big) \\ &\quad \times f_{(u_1,\cdots,u_n|\bar{U}_n)}(u_1,\cdots,u_n). \end{aligned}$$

For any $u_1,\cdots,u_n$ such that $\bar{U}_n = n^{-1}\sum_{i=1}^n u_i \in [2^{-1}-\epsilon_n, 2^{-1}+\epsilon_n]$, let $\boldsymbol{U} = (U_1,\cdots,U_n)$, $\boldsymbol{u} = (u_1,\cdots,u_n)$, and $\{i_1,\cdots,i_{n\bar{U}_n}\}$ are indices for $\boldsymbol{u} = 1$, and $\{j_1,\cdots,j_{n-n\bar{U}_n}\}$ are indices for $\boldsymbol{u} = 0$. Also let $\boldsymbol{U}_{i_k} = (U_{i_1},\cdots,U_{i_{n\bar{U}_n}})$, $\boldsymbol{U}_{j_k} = (U_{j_1},\cdots,U_{j_{n-n\bar{U}_n}})$ and let the variables (P_k, Q_k) and (P'_k, Q'_k) be specified with the following distributions:

$$\{(P'_k, Q'_k)\}_{k=1}^{n\bar{U}_n} \overset{D}{=} \{(Y_{i_k}, V_{i_k})\}_{k=1}^{n\bar{U}_n} | \boldsymbol{U}_{i_k} = \mathbf{1}$$

and

$$\{(P_k, Q_k)\}_{k=1}^{n-n\bar{U}_n} \stackrel{D}{=} \{(Y_{j_k}, V_{j_k})\}_{k=1}^{n-n\bar{U}_n} | \boldsymbol{U}_{j_k} = \mathbf{0}.$$

By the independence of $\{\boldsymbol{Z}_k\}_{k=1}^n = \{(1, U_k, V_k)\}_{k=1}^n$, we have

$$\begin{aligned}
& P\Big(\sup_{\beta \in R^3} W_n(\boldsymbol{\beta}) > C|\sigma_n \log \sigma_n| \Big| U_1 = u_1, \cdots, U_n = u_n, |\bar{U}_n - 1/2| \leq \epsilon_n \Big) \\
= \; & P\Big(\sup_{\beta \in R^3} W_n(\boldsymbol{\beta}) > C|\sigma_n \log \sigma_n| \Big| U_1 = u_1, \cdots, U_n = u_n \Big) \\
= \; & P\Big(\sup_{\beta \in R^3} \{ n^{-1} \textstyle\sum_{k=1}^{n\bar{U}_n} 1(|Y_{i_k} - \beta_1 - \beta_2 - \beta_3 V_{i_k}| \leq |\sigma_n \log \sigma_n|) \\
& + n^{-1} \textstyle\sum_{k=1}^{n-n\bar{U}_n} 1(|Y_{j_k} - \beta_1 - \beta_3 V_{j_k}| \leq |\sigma_n \log \sigma_n|) \} > C|\sigma_n \log \sigma_n| \Big| \boldsymbol{U} = \boldsymbol{u} \Big) \\
\leq \; & P\Big(\sup_{\beta \in R^3} n^{-1} \textstyle\sum_{k=1}^{n\bar{U}_n} 1(|Y_{i_k} - \beta_1 - \beta_2 - \beta_3 V_{i_k}| \leq |\sigma_n \log \sigma_n|) > (C/2)|\sigma_n \log \sigma_n| \Big| \boldsymbol{U}_{i_k} = \mathbf{1} \Big) \\
& + P\Big(\sup_{\beta \in R^3} n^{-1} \textstyle\sum_{k=1}^{n-n\bar{U}_n} 1(|Y_{j_k} - \beta_1 - \beta_3 V_{j_k}| \leq |\sigma_n \log \sigma_n|) > (C/2)|\sigma_n \log \sigma_n| \Big| \boldsymbol{U}_{j_k} = \mathbf{0} \Big) \\
\leq \; & P\Big(\sup_{a,b \in R} (n\bar{U}_n)^{-1} \textstyle\sum_{k=1}^{n\bar{U}_n} 1(|(P_k' - aQ_k' - b| \leq |\sigma_n \log \sigma_n|) > (C/2)|\sigma_n \log \sigma_n| \Big) \\
& + P\Big(\sup_{a,b \in R} (n - n\bar{U}_n)^{-1} \textstyle\sum_{k=1}^{n-n\bar{U}_n} 1(|P_k - aQ_k - b| \leq |\sigma_n \log \sigma_n|) > (C/2)|\sigma_n \log \sigma_n| \Big).
\end{aligned}$$

Since $\{Y_i, \boldsymbol{Z}_i, \boldsymbol{X}_i\}_{i=1}^n$ are *i.i.d.*, $\{(P_k, Q_k)\}_{k=1}^{n-n\bar{U}_n}$ are *i.i.d.* and so are $\{(P_k', Q_k')\}_{k=1}^{n\bar{U}_n}$. We now prove the following two properties under both the null hypothesis and the alternative hypothesis.

CL3 (P_k, Q_k) and (P_k', Q_k') have finite means;

CL4 The densities of P_k, P_k' and the conditional densities of $P_k|Q_k$, $P_k'|Q_k'$ are bounded.

Then, by the choice of $\bar{U}_n$, $n\bar{U}_n = O(n/2)$ and $n - n\bar{U}_n = O(n/2)$ almost surely.

Taken **CL3** and **CL4** and **Lemma 1** together, we conclude that there exist constant C' such that $I \le C'n^{-2}$ for sufficiently large n. The proof for **CL1** and **CL2** is then complete.

Proof of CL3 and CL4: Recall

$$Y|\{(U,V),\boldsymbol{X}\} \sim \pi(\boldsymbol{X}^T\boldsymbol{\gamma})N(\boldsymbol{Z}^T(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2),\sigma_1^2) + (1-\pi(\boldsymbol{X}^T\boldsymbol{\gamma}))N(\boldsymbol{Z}^T\boldsymbol{\beta}_1,\sigma_2^2).$$

Note that the null model is just a special case of the above in that $\boldsymbol{\beta}_2 = 0$ and $\sigma_1 = \sigma_2$. By the definitions of (P_k, Q_k) and $(P_k^{'}, Q_k^{'})$, for any $\boldsymbol{\beta}^T = (\boldsymbol{\gamma}^T, \boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \sigma_1, \sigma_2)$, if suffices to show the following two statements:

S(i) $E\Big(|Y|\Big|U\Big) < \infty,\ E\Big(|V|\Big|U\Big) < \infty$;

S(ii) the conditional densities of $V \mid U$ and $Y \mid U, V$ are bounded.

The statement **S(ii)** is obvious, since $Y \mid V, U, \boldsymbol{X}$ follows the logistic mixture of normals, its density is uniformly bounded by $\{(2\pi)^{1/2}\min\{\sigma_1,\sigma_2\}\}^{-1}$, where σ_1, σ_2 are the true parameters in the model. Therefore, the conditional density of $Y|V, U$ is bounded, and by Condition C4, the conditional density of $V \mid U$ is bounded. For **S(i)**, by Condition C5, $E\Big(|V|\Big|U\Big) < \infty$, again because $Y \mid V, U, \boldsymbol{X}$

follows logistic mixture of normals,

$$\begin{aligned} E\Big(|Y|\Big|U\Big) = E\Big\{E\Big(|Y|\Big|U,V,\boldsymbol{X}\Big)\Big|U\Big\} &= E\Big\{(E\Big(\pi(\boldsymbol{X}^T\boldsymbol{\gamma})|Y_1| + [1-\pi(\boldsymbol{X}^T\boldsymbol{\gamma})]|Y_2|\Big|U,V,\boldsymbol{X}))\Big|U\Big\} \\ &= E\Big\{[\pi(\boldsymbol{X}^T\boldsymbol{\gamma})E\Big(|Y_1|\Big|U,V,\boldsymbol{X}\Big) \\ &\quad +[1-\pi(\boldsymbol{X}^T\boldsymbol{\gamma})]E\Big(|Y_2|\Big|U,V,\boldsymbol{X}\Big)]\Big|U\Big\} \end{aligned}$$

where $Y_1 \mid (U,V,\boldsymbol{X}) \sim N(\boldsymbol{Z}^T(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2),\sigma_1^2)$, $Y_2 \mid (U,V,\boldsymbol{X}) \sim N(\boldsymbol{Z}^T\boldsymbol{\beta}_1,\sigma_2^2)$, and $\boldsymbol{Z} = (1,U,V)$. Note that

$$\begin{aligned} E\Big(|Y_1|\Big|U,V,\boldsymbol{X}\Big) &\le \sigma_1 E\Big(|\tfrac{Y_1-\boldsymbol{Z}^T\boldsymbol{\beta}_1-\boldsymbol{Z}^T\boldsymbol{\beta}_2}{\sigma_1}|\Big|U,V,\boldsymbol{X}\Big) + (1+|V|+|U|)\|\boldsymbol{\beta}_1+\boldsymbol{\beta}_2\|_\infty \\ &= \tfrac{2}{\sqrt{2\pi}}\sigma_1 + (1+|V|+|U|)\|\boldsymbol{\beta}_1+\boldsymbol{\beta}_2\|_\infty, \end{aligned}$$

where $\|\cdot\|_\infty$ is the supreme norm and the last equation is due to the fact that $E|Z| = 2(2\pi)^{-1/2}$ if $Z \sim N(0,1)$. Similarly,

$$E\Big(|Y_2|\Big|U,V,\boldsymbol{X}\Big) \le \frac{2}{\sqrt{2\pi}}\sigma_2 + (1+|V|+|U|)\|\boldsymbol{\beta}_1\|_\infty.$$

Therefore,

$$E\Big(|Y|\Big|U\Big) \le \frac{2}{\sqrt{2\pi}}\max\{\sigma_1,\sigma_2\} + \max\{\|\boldsymbol{\beta}_1\|_\infty, \|\boldsymbol{\beta}_1+\boldsymbol{\beta}_2\|_\infty\}E\Big(1+|V|+|U|\Big|U\Big) < \infty,$$

where the last inequality is due to Condition C5.

We have now verified properties **CL3** and **CL4** for any $\boldsymbol{\beta}^T = (\boldsymbol{\gamma}^T, \boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \sigma_1, \sigma_2)$, under both the null and the alternative hypotheses. By the results from **CL1**, **CL2** and Equation (S1.1), we finished the proof of **Case 1**.

Proof for Case 2: We prove **S1** with $\boldsymbol{Z} = (1, \boldsymbol{U}, V)$, where $\boldsymbol{U}$ is a random vector taking any finite values and V is one dimensional continuous variable.

Let $P(\boldsymbol{U} = \boldsymbol{u}^t) = p_t > 0$, $t = 1, 2, \cdots, r$, and $\sum_{t=1}^{r} p_t = 1$. Also let $\bar{U}_n^t = n^{-1} \sum_{i=1}^{n} 1(\boldsymbol{U}_i = \boldsymbol{u}^t)$, $t = 1, 2, \cdots, r$. As in the earlier proof, we set $\epsilon_{nt} = \{n^{-1}(8 \log n)\}^{1/2}$, and we bound $P(A_n(C))$ by

$$P\Big(A_n(C)\Big|\bar{U}_n^t \in [p_t - \epsilon_{nt}, p_t + \epsilon_{nt}], t = 1, \cdots, r\Big) + \sum_{t=1}^{r} P\Big(|\bar{U}_n^t - p_t| > \epsilon_{nt}\Big) \triangleq I + II.$$

By Bernstein's inequality, we know $II < Cn^{-2}$. For part I, we use arguments conditional on $\boldsymbol{U}_i = \boldsymbol{u}_i$, $i = 1, 2, \cdots, n$, such that the values of $\boldsymbol{u}_i$ satisfy $\bar{U}_n^t \in [p_t - \epsilon_{nt}, p_t + \epsilon_{nt}]$, $t = 1, \cdots, r$. We then group the $\boldsymbol{U}_i = \boldsymbol{u}_i$ which have the same value of $\boldsymbol{u}^t$. Note that the number of the items in each group is of the order of $O(p_t n)$, and by the independence of the vectors of $\mathbf{Z}_i$, we can directly apply **Lemma 1** and get the desired results.

Proof for Case 3: We prove **S1** for general $\boldsymbol{Z} = (1, \boldsymbol{U}, \boldsymbol{V})$, where $\boldsymbol{U}$ is a random vector taking finite values and $\boldsymbol{V}$ ia a vector of continuous random variables.

We bound $P(A_n(C))$ by conditioning on the possible values of $\boldsymbol{U}$ as we did previously, then it suffices to show

$$P\Big(\sup_{b\in R,\rho\in R^+,\|\alpha\|=1} n^{-1}\sum_{k=1}^{n} 1\Big(|P_k-\rho\alpha^T\boldsymbol{Q}_k-b|\le|\sigma_n\log\sigma_n|\Big)>C^*|\sigma_n\log\sigma_n|\Big)\le Cn^{-2},$$

for some C^* and C and sufficiently large n. However, the set of α with $\|\alpha\|=1$ is a compact set, we can prove it by using standard empirical process argument and the same techniques as those used to prove **Lemma 1** in the next subsection.

S1.2 Proof of Lemma 1

In this subsection, we prove Lemma 1 which is needed for the proof of Theorem 1. We allow the constants below to vary line by line. Let

$$G_n(a,b,\sigma_n)=n^{-1}\textstyle\sum_{k=1}^{n}1\Big(|P_k-aQ_k-b|\le|\sigma_n\log\sigma_n|\Big),$$

$$L_{n1}=P(\sup_{|a|\le n^2,b\in R}G_n(a,b,\sigma_n)>C^*|\sigma_n\log\sigma_n|),$$

$$L_{n2}=P(\sup_{|a|>n^2,b\in R}G_n(a,b,\sigma_n)>C^*|\sigma_n\log\sigma_n|).$$

Note that

$$\sup_{a,b\in R}G_n(a,b,\sigma_n)=\max\Big\{\sup_{|a|\le n^2,b\in R}G_n(a,b,\sigma_n),\sup_{|a|>n^2,b\in R}G_n(a,b,\sigma_n)\Big\}.$$

Thus we have,

$$P\Big(\sup_{a,b\in R} G_n(a,b,\sigma_n) > C^*|\sigma_n \log\sigma_n|\Big) \le L_{n1} + L_{n2}. \tag{S1.2}$$

Step 1: We show $L_{n2} \le Cn^{-2}$.

Note that for any given $\sigma_n \in (n^{-1},\ e^{-1})$, $|\sigma_n \log\sigma_n| \ge n^{-1}\log n$. We have

$$\begin{aligned} G_n(a,b,\sigma_n) \;\; \le \;\; & n^{-1}\textstyle\sum_{k=1}^n 1\left(\frac{P_k}{|a|} - \frac{|\sigma_n\log\sigma_n|}{|a|} \le Q_k + \frac{b}{|a|} \le \frac{P_k}{|a|} + \frac{|\sigma_n\log\sigma_n|}{|a|}\right) \\ & \times 1\Big(|P_k| \le (|a|-1)|\sigma_n\log\sigma_n|\Big) \\ & + n^{-1}\textstyle\sum_{k=1}^n 1\Big(|P_k| > (|a|-1)|\sigma_n\log\sigma_n|\Big). \end{aligned}$$

Therefore,

$$\begin{aligned} & \sup_{|a|>n^2, b\in R} G_n(a,b,\sigma_n) \\ \le \;\; & \sup_{|a|>n^2, b\in R}\Big\{n^{-1}\textstyle\sum_{k=1}^n 1\left(\frac{P_k}{|a|} - \frac{|\sigma_n\log\sigma_n|}{|a|} \le Q_k + \frac{b}{|a|} \le \frac{P_k}{|a|} + \frac{|\sigma_n\log\sigma_n|}{|a|}\right) \\ & \times 1(|P_k| \le (|a|-1)|\sigma_n\log\sigma_n|)\Big\} \\ & + \sup_{|a|>n^2}\Big\{n^{-1}\textstyle\sum_{k=1}^n 1\Big(|P_k| > (|a|-1)|\sigma_n\log\sigma_n|\Big)\Big\} \\ \le \;\; & \sup_{\theta\in R} n^{-1}\Big\{\textstyle\sum_{k=1}^n 1(-|\sigma_n\log\sigma_n| \le Q_k - \theta \le |\sigma_n\log\sigma_n|)\Big\} \\ & + n^{-1}\textstyle\sum_{k=1}^n 1(|P_k| > n). \end{aligned}$$

Let

$$L_{n21} = P\Big(\sup_{\theta\in R}\Big\{n^{-1}\sum_{k=1}^n 1(-|\sigma_n\log\sigma_n| \le Q_k - \theta \le |\sigma_n\log\sigma_n|)\Big\} > (C^*/2)|\sigma_n\log\sigma_n|\Big)$$

and

$$L_{n22} = P\Big(n^{-1}\sum_{k=1}^{n} 1(|P_k| > n) > (C^*/2)|\sigma_n \log \sigma_n|\Big).$$

Then,

$$L_{n2} = P\Big(\sup_{|a|>n^2, b\in R} G_n(a,b,\sigma_n) > C^*|\sigma_n \log \sigma_n|\Big) \le L_{n21} + L_{n22}. \tag{S1.3}$$

Step 1-1: We show $L_{n21} \le Cn^{-2}$.

Observe that in L_{n21},

$$\begin{aligned} & n^{-1}\textstyle\sum_{k=1}^{n} 1\Big(-|\sigma_n \log \sigma_n| \le Q_k - \theta \le |\sigma_n \log \sigma_n|\Big) \\ = \;& F_n(\theta + |\sigma_n \log \sigma_n|) - F_n(\theta - |\sigma_n \log \sigma_n|), \end{aligned}$$

where F_n is the empirical distribution for Q. Since the density of Q is bounded, a direct application of **Lemma 1** of Chen et al. (2008) yields $L_{n21} \le Cn^{-2}$.

Step 1-2: We show $L_{n22} \le Cn^{-2}$.

Note that $E\{1(|P_k| > n)\} \le n^{-1}E(|P_k|) \le n^{-1}\log n \le |\sigma_n \log \sigma_n|$, for sufficiently large n. Then, by Bernstein's inequality, we have

$$\begin{aligned} L_{n22} &\le P\Big(\textstyle\sum_{k=1}^{n}\Big[1\{|P_k| > n\} - E\Big(1\{|P_k| > n\}\Big)\Big] > \tilde{C}n|\sigma_n \log \sigma_n|\Big) \\ &\le \exp\Big\{-\frac{(\tilde{C}n)^2|\sigma_n \log \sigma_n|^2}{2n|\sigma_n \log \sigma_n| + 2\tilde{C}n|\sigma_n \log \sigma_n|}\Big\} \le Cn^{-2}, \end{aligned}$$

where $\tilde{C} = C^*/2 - 1$.

By **Step 1-1**, **Step 1-2** and Equation (S1.3), we have

$$L_{n2} \le L_{n21} + L_{n22} \le Cn^{-2}, \tag{S1.4}$$

which completes the proof of **Step 1**.

Step 2: We show $L_{n1} \le Cn^{-2}$.

Let $\delta_n = n^{-1}|\sigma_n \log \sigma_n| \ge n^{-2}(\log n)$. Divide $|a| \le n^2$ into the union of k_n subsets $\{\Omega_{nj}\}_{j=1}^{k_n}$, such that, the distance between any two points in each subset is no greater than δ_n. It is clear that we can achieve this with $k_n \le (\log n)^{-1} 2n^4 \le O(n^4)$. Let $U_k(a, b, \sigma_n) = 1(|P_k - aQ_k - b| \le |\sigma_n \log \sigma_n|)$, then

$$\begin{aligned}
& \sup_{|a|\le n^2, b\in R} G_n(a, b, \sigma_n) \\
= \; & \max_{1\le j\le k_n} \Big[\sup_{a\in\Omega_{nj}, b\in R} \{G_n(a, b, \sigma_n)\} \Big] \\
\le \; & \max_{1\le j\le k_n} \Big[\sup_{b\in R} G_n(a_j, b, \sigma_n) + \sup_{|a-a_j|\le\delta_n, b\in R} \Big\{ |G_n(a, b, \sigma_n) - G_n(a_j, b, \sigma_n)| \Big\} \Big] \\
\le \; & \max_{1\le j\le k_n} \Big[\sup_{b\in R} G_n(a_j, b, \sigma_n) + \sup_{|a-a_j|\le\delta_n, b\in R} \Big\{ n^{-1} \textstyle\sum_{k=1}^n |U_k(a, b, \sigma_n) \\
& -U_k(a_j, b, \sigma_n)| \Big\} \Big],
\end{aligned}$$

where a_j is any fixed point in Ω_{nj}. Let

$$L_{n11} = k_n \sup_{a\in R} P\Big\{ \sup_{b\in R} G_n(a, b, \sigma_n) > (C^*/2)|\sigma_n \log \sigma_n| \Big\},$$

and

$$L_{n12} = k_n \sup_{a' \in R} P\Big\{ \sup_{|a-a'|\le\delta_n, b\in R} n^{-1} \sum_{k=1}^{n} |U_k(a,b,\sigma_n) - U_k(a',b,\sigma_n)| > (C^*/2)|\sigma_n \log\sigma_n| \Big\}.$$

Then we have

$$L_{n1} \le L_{n11} + L_{n12}. \tag{S1.5}$$

Step 2-1: We show $L_{n11} \le Cn^{-2}$.

In L_{n11}, for any $a \in R$, let $R_k^a = P_k - aQ_k$. Since P_k, Q_k are continuous, and R_k^a is continuous and its density $f_{R_k^a}(r) = \int f_{R_k^a|Q_k}(r|q_k) f_{Q_k}(q_k)dq_k = \int f_{P_k|Q_k}(r + aq_k|q_k) f_{Q_k}(q_k)dq_k \le C$. Therefore,

$$\begin{aligned} G_n(a,b,\sigma_n) &= n^{-1} \textstyle\sum_{k=1}^{n} 1\Big(|R_k^a - b| \le |\sigma_n \log\sigma_n|\Big) \\ &= F_n(b + |\sigma_n \log\sigma_n|) - F_n(b - |\sigma_n \log\sigma_n|), \end{aligned}$$

where F_n is the empirical distribution for R_k^a, $k = 1, \cdots, n$. Since the density of R_k^a is uniformly bounded over a, a direct application of **Lemma 1** of Chen et al. (2008) yields

$$P\Big(\sup_{b\in R} n^{-1} \sum_{k=1}^{n} 1\Big(|R_k^a - b| \le |\sigma_n \log\sigma_n|\Big) > (C^*/2)|\sigma_n \log\sigma_n|\Big) < Cn^{-6},$$

for any $a \in R$ and for some fixed constant C^*. By using the order of k_n, we have for some C^*,

$$L_{n11} \leq C^* n^{-2} \tag{S1.6}$$

for sufficiently large n.

Step 2-2: We show $L_{n12} \leq Cn^{-2}$.

For any $a^{'} \in R$, let

$$\begin{aligned}
M_{n1}(a,b,a^{'},\sigma_n) &= n^{-1}\textstyle\sum_{k=1}^{n} 1\Big(P_k - a^{'}Q_k - b \geq -|\sigma_n \log\sigma_n|\Big)1\Big(P_k - aQ_k - b \leq -|\sigma_n \log\sigma_n|\Big),\\
M_{n2}(a,b,a^{'},\sigma_n) &= n^{-1}\textstyle\sum_{k=1}^{n} 1\Big(P_k - a^{'}Q_k - b \leq |\sigma_n \log\sigma_n|\Big)1\Big(P_k - aQ_k - b \geq |\sigma_n \log\sigma_n|\Big),\\
M_{n3}(a,b,a^{'},\sigma_n) &= n^{-1}\textstyle\sum_{k=1}^{n} 1\Big(P_k - a^{'}Q_k - b \leq -|\sigma_n \log\sigma_n|\Big)1\Big(P_k - aQ_k - b \geq -|\sigma_n \log\sigma_n|\Big),\\
M_{n4}(a,b,a^{'},\sigma_n) &= n^{-1}\textstyle\sum_{k=1}^{n} 1\Big(P_k - a^{'}Q_k - b \geq |\sigma_n \log\sigma_n|\Big)1\Big(P_k - aQ_k - b \leq |\sigma_n \log\sigma_n|\Big);
\end{aligned}$$

$$\begin{aligned}
N_{n1}(b,a^{'},\sigma_n) &= n^{-1}\textstyle\sum_{k=1}^{n} 1\Big(-|\sigma_n \log\sigma_n| + \delta_n|Q_k| \geq P_k - a^{'}Q_k - b \geq -|\sigma_n \log\sigma_n|\Big),\\
N_{n2}(b,a^{'},\sigma_n) &= n^{-1}\textstyle\sum_{k=1}^{n} 1\Big(-|\sigma_n \log\sigma_n| - \delta_n|Q_k| \leq P_k - a^{'}Q_k - b \leq -|\sigma_n \log\sigma_n|\Big),\\
N_{n3}(b,a^{'},\sigma_n) &= n^{-1}\textstyle\sum_{k=1}^{n} 1\Big(|\sigma_n \log\sigma_n| - \delta_n|Q_k| \leq P_k - a^{'}Q_k - b \leq |\sigma_n \log\sigma_n|\Big),\\
N_{n4}(b,a^{'},\sigma_n) &= n^{-1}\textstyle\sum_{k=1}^{n} 1\Big(|\sigma_n \log\sigma_n| + \delta_n|Q_k| \geq P_k - a^{'}Q_k - b \geq |\sigma_n \log\sigma_n|\Big).
\end{aligned}$$

Then,

$$
\begin{aligned}
& n^{-1}\sum_{k=1}^{n}|U_k(a,b,\sigma_n)-U_k(a',b,\sigma_n)| \\
= \;& n^{-1}\sum_{k=1}^{n}|U_k(a,b,\sigma_n)-U_k(a',b,\sigma_n)|U_k(a',b,\sigma_n) \\
& +n^{-1}\sum_{k=1}^{n}|U_k(a,b,\sigma_n)-U_k(a',b,\sigma_n)|(1-U_k(a',b,\sigma_n)) \\
\le \;& M_{n1}(a,b,a',\sigma_n)+M_{n2}(a,b,a',\sigma_n) \\
& +M_{n3}(a,b,a',\sigma_n)+M_{n4}(a,b,a',\sigma_n).
\end{aligned}
$$

Note that for any a, such that $|a-a'|\le\delta_n$,

$$P_k-a'Q_k-b = P_k-aQ_k-b-(a'-a)Q_k \in [P_k-aQ_k-b-\delta_n|Q_k|, P_k-aQ_k-b+\delta_n|Q_k|].$$

Thus, for any a, such that $|a-a'|\le\delta_n$,

$$
\begin{aligned}
& M_{n1}(a,b,a',\sigma_n)+M_{n2}(a,b,a',\sigma_n)+M_{n3}(a,b,a',\sigma_n)+M_{n4}(a,b,a',\sigma_n) \\
\le \;& N_{n1}(b,a',\sigma_n)+N_{n2}(b,a',\sigma_n)+N_{n3}(b,a',\sigma_n)+N_{n4}(b,a',\sigma_n).
\end{aligned}
$$

Therefore, for any $a'\in R$,

$$
\begin{aligned}
& \sup_{|a-a'|\le\delta_n,b\in R} n^{-1}\sum_{k=1}^{n}|U_k(a,b,\sigma_n)-U_k(a',b,\sigma_n)| \\
\le \;& \sup_{b\in R} N_{n1}(b,a',\sigma_n)+\sup_{b\in R} N_{n2}(b,a',\sigma_n)+\sup_{b\in R} N_{n3}(b,a',\sigma_n) \\
& +\sup_{b\in R} N_{n4}(b,a',\sigma_n).
\end{aligned}
$$

Let $L_{n12i} = k_n \sup_{a' \in R} P(N_{ni}(b, a', \sigma_n) > (C^*/8)|\sigma_n \log \sigma_n|)$, $i = 1, 2, 3, 4$. Then

$$L_{n12} \leq \sum_{i=1}^{4} L_{n12i}. \tag{S1.7}$$

By the choice of δ_n,

$$\begin{aligned} N_{n1}(b, c, \sigma_n) \quad &\leq \quad \sup_{b \in R} n^{-1} \textstyle\sum_{k=1}^{n} 1\Big(-|\sigma_n \log \sigma_n| + \delta_n |Q_k| \geq P_k - a' Q_k - b \\ &\qquad \geq -|\sigma_n \log \sigma_n|\Big) \times 1(|Q_k| \leq n) + n^{-1} \textstyle\sum_{k=1}^{n} 1(|Q_k| > n) \\ &\leq \quad \sup_{b \in R} n^{-1} \textstyle\sum_{k=1}^{n} 1\Big(0 \geq P_k - a' Q_k - b \geq -|\sigma_n \log \sigma_n|\Big) \\ &\qquad + n^{-1} \textstyle\sum_{k=1}^{n} 1(|Q_k| > n). \end{aligned}$$

Therefore,

$$\begin{aligned} L_{n121} \quad &\leq k_n \sup_{a' \in R} P\Big(\sup_{b \in R} n^{-1} \textstyle\sum_{k=1}^{n} 1(0 \geq P_k - a' Q_k - b \geq -|\sigma_n \log \sigma_n|) \\ &> (C^*/16)|\sigma_n \log \sigma_n|\Big) + k_n P\Big(n^{-1} \textstyle\sum_{k=1}^{n} 1(|Q_k| > n) > (C^*/16)|\sigma_n \log \sigma_n|\Big). \end{aligned}$$

Analogous to the proof for (S1.4), we have $L_{n121} \leq Cn^{-2}$. Similarly, the results hold for $L_{n12i}, i = 2, 3, 4$. Therefore, by (S1.7), $L_{n12} \leq \sum_{i=1}^{4} L_{n12i} \leq Cn^{-2}$.

By **Step 2-1**, **Step 2-2** and Equation (S1.5), we have

$$L_{n1} \leq \sum_{i=1}^{2} L_{n1i} \leq Cn^{-2}, \tag{S1.8}$$

which completes the proof of **Step 2**.

By **Step1**, **Step 2** and Equation (S1.2), we complete the proof of **Lemma 1**.

S1.3 Proof of Theorem 2

For Theorem 2, we note that the Taylor expansion of $pEM_j^{(K)}$ together with Condition C6 which implies that the penalty vanishes almost surely. Then, the results follow from similar arguments to those for **Theorem 2** in Shen and He (2015).

S2 Additional Results for Empirical Studies

In this section, we provide additional tables and figures for the simulation and real data examples in Section 4 of the paper.

Firstly, we show the type-1 errors of the pEM test with $\lambda = 50$, which gives similar results to Table 1 in the main paper.

Table 1: Type I errors of the pEM test with bootstrap approximations in 1000 data sets with standard errors in the parenthesis, with $\lambda = 50$.

n	Nominal level α	$pEM^{(0)}$	$pEM^{(3)}$	$pEM^{(9)}$
n=60	0.01	0.013(0.004)	0.013 (0.004)	0.014(0.004)
	0.05	0.044(0.006)	0.050(0.007)	0.051(0.007)
	0.10	0.089 (0.009)	0.088(0.009)	0.094(0.009)
n=100	0.01	0.010(0.003)	0.010(0.003)	0.008(0.003)
	0.05	0.049(0.007)	0.049(0.007)	0.048(0.007)
	0.10	0.103 (0.010)	0.116(0.010)	0.113(0.010)

For the NSW data, we have the descriptions of the variables in Table 2.

For the AIDS data, we have the estimates from the unequal variance model

Table 2: Summary statistics for the NSW study. In the first six rows, we give the mean and quantiles for the continuous variables, and in the last two rows we give the frequencies of the four binary variables. In the table, Y: log(RE78+1)-log(RE75+1) in which RE78 and RE75 is the earning of the individual in 1975 and in 1978, respectively; *trt*: treatment indicator, that is, if the subject joins the training program; X_1: education years; X_2: whether the subject is Black; X_3: whether the baseline income is zero; and X_4: whether the baseline income is above the median of the positive part.

	Y	X_1		
Min.	-9.81	3		
1st Qu.	-0.38	9		
Median	0.41	10		
Mean	1.43	10		
3rd Qu.	6.94	11		
Max.	10.44	16		
	trt	X_2	X_3	X_4
0	425	144	433	506
1	297	578	289	216

in Table 3, and the plot of estimated *membership scores* from (4.3) and (4.4) in Figure 1.

Bibliography

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

Shen, J. and He, X. (2015). Inference for subgroup analysis with a structured logistic-normal mixture model. *Journal of the American Statistical Association* **110,** 303–312.

Table 3: Parameter estimates and their standard errors when the unequal variance structured logistic-normal mixture model was used to fit the data in the ACTG study with $\lambda = 400$.

	$\boldsymbol{\beta}_1(1)$	$\boldsymbol{\beta}_1(\text{trt})$	$\boldsymbol{\beta}_1\big(\log(cd4.0)\big)$	$\boldsymbol{\beta}_1\big(\log_{10}(rna.0)\big)$	$\boldsymbol{\beta}_1(\text{Age})$
est	-46.00	41.76	-0.68	7.41	0.72
se	44.46	6.34	3.59	6.73	0.38
	$\boldsymbol{\beta}_2(1)$	$\boldsymbol{\beta}_2(\text{trt})$	$\boldsymbol{\beta}_2\big(\log(cd4.0)\big)$	$\boldsymbol{\beta}_2\big(\log_{10}(rna.0)\big)$	$\boldsymbol{\beta}_2(\text{Age})$
est	3.23	51.74	8.75	-1.23	-0.71
se	63.63	9.73	5.79	10.23	1.06
	$\boldsymbol{\gamma}(1)$		$\boldsymbol{\gamma}\big(\log(cd4.0)\big)$	$\boldsymbol{\gamma}\big(\log_{10}(rna.0)\big)$	$\boldsymbol{\gamma}(\text{ Age})$
est	-9.16		0.67	1.40	-0.02
se	1.02		0.08	0.16	0.01
	σ_1	σ_2			
est	57.65	48.29			
se	1.24	0.97			

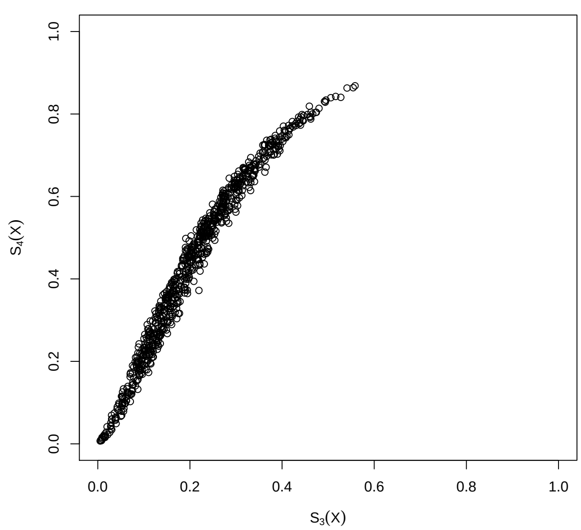

Figure 1: AIDS data. *Membership scores* for all the subjects estimates from the equal variance structured mixture model (4.3) and the unequal variance structured mixture model (4.4), respectively.