SUPPLEMENTARY MATERIAL

BY JIMING JIANG

University of California, Davis

Throughout this supplementary material, the paper, "The Subset Argument and Consistency of MLE in GLMM: Answer to An Open Problem and Beyond", is referred to as Jiang2012. Equation numbers without referring to Jiang2012 conrespond to those in this supplementary material.

1. Proof of (7) in Jiang2012.

Consider the bivariate function

$$g(p,x) \quad = \quad \left(\frac{p}{p_0}\right)^{p_0+x}\left(\frac{1-p}{1-p_0}\right)^{1-p_0-x}.$$

Let $\delta_\epsilon = p_0(\mu+\epsilon) - p_0 > 0$. It can be shown that, provided that $|x| \leq \delta_\epsilon$, then $g(p,x)$ is decreasing with p for $p \geq p_0 + \delta_\epsilon$; hence, we have $g(p,x) \leq g(p_0+\delta_\epsilon, x)$ if $p \geq p_0+\delta_\epsilon$ and $|x| \leq \delta_\epsilon$. On the other hand, it is easy to show that $g(p_0+\delta_\epsilon, 0) < 1$; thus, by continuity, there is $0 < \delta \leq \delta_\epsilon$ and $0 < \gamma < 1$ such that $g(p_0+\delta_\epsilon, x) \leq \gamma$, if $|x| \leq \delta$.

Next, we divide the interval $(\mu+\epsilon, K]$ by $\lambda_j = \mu+\epsilon+(j/J)(K-\mu-\epsilon)$, $j = 1,\dots,J$, where $J = [\rho mn/m\wedge n]+1$ and $\rho = -2(K-\mu-\epsilon)/\log\gamma > 0$. Then, we have $p_j \equiv p_0(\lambda_j) > p_0(\mu+\epsilon) = p_0+\delta_\epsilon$, implying $g(p_j,x) \leq g(p_0+\delta_\epsilon,x) \leq \gamma$, $1 \leq j \leq J$, if $|x| \leq \delta$. It follows that, on $\mathcal{A}_\delta$, we have

$$\frac{p_{\lambda_j}(y_{[1]})}{p_\mu(y_{[1]})} = \{g(p_j,\Delta)\}^{m\wedge n} \leq \gamma^{m\wedge n}, \;\; 1 \leq j \leq J. \tag{1}$$

For any $\lambda \in (\mu+\epsilon, K]$, there is $1 \leq j \leq J$ such that $|\lambda-\lambda_j| \leq (K-\mu-\epsilon)/J$. Then, by the Taylor expansion, and proof of Theorem 2 in Jiang2012, we have $\log p_\lambda(y_{[1]}, y_{[2]}) - \log p_{\lambda_j}(y_{[1]}, y_{[2]}) \leq mn(K-\mu-\epsilon)/J$. It follows that

$$\sup_{\lambda\in(\mu+\epsilon,K]} \frac{p_\lambda(y_{[1]},y_{[2]})}{p_\mu(y_{[1]},y_{[2]})} \quad \leq \quad \exp\left\{mn\left(\frac{K-\mu-\epsilon}{J}\right)\right\} \max_{1\leq j\leq J} \frac{p_{\lambda_j}(y_{[1]},y_{[2]})}{p_\mu(y_{[1]},y_{[2]})}.$$

Thus, by the subsect argument of Jiang2012 and (1), we have, on $\mathcal{A}_\delta$,

$$\mathrm{P}_\mu\left\{\sup_{\lambda\in(\mu+\epsilon,K]} \frac{p_\lambda(y_{[1]},y_{[2]})}{p_\mu(y_{[1]},y_{[2]})} > 1 \middle| y_{[1]}\right\}$$

$$\begin{aligned}
&\leq \mathrm{P}_\mu\left[\max_{1\leq j\leq J}\frac{p_{\lambda_j}(y_{[1]},y_{[2]})}{p_\mu(y_{[1]},y_{[2]})} > \exp\left\{-mn\left(\frac{K-\mu-\epsilon}{J}\right)\right\}\middle| y_{[1]}\right]\\
&\leq \sum_{j=1}^J \mathrm{P}_\mu\left[\frac{p_{\lambda_j}(y_{[1]},y_{[2]})}{p_\mu(y_{[1]},y_{[2]})} > \exp\left\{-mn\left(\frac{K-\mu-\epsilon}{J}\right)\right\}\middle| y_{[1]}\right]\\
&\leq \exp\left\{mn\left(\frac{K-\mu-\epsilon}{J}\right)\right\}\sum_{j=1}^J \mathrm{E}_\mu\left\{\frac{p_{\lambda_j}(y_{[1]},y_{[2]})}{p_\mu(y_{[1]},y_{[2]})}\middle| y_{[1]}\right\}\\
&= \exp\left\{mn\left(\frac{K-\mu-\epsilon}{J}\right)\right\}\sum_{i=1}^J \frac{p_{\lambda_j}(y_{[1]})}{p_\mu(y_{[1]})} \quad \text{[see (2) of Jiang2012]}\\
&\leq \exp\left\{mn\left(\frac{K-\mu-\epsilon}{J}\right)\right\} J\gamma^{m\wedge n}\\
&\leq \exp\left[(m\wedge n)\left\{\frac{\log\gamma}{2}+\frac{\log(m\vee n)+\log(2\rho)}{m\wedge n}\right\}\right],
\end{aligned}$$

if $m\wedge n \geq N_0$ for some $N_0 \geq 1$, using the definition of J. Therefore, there is $N_1 \geq N_0$ such that, when $m\wedge n\geq N_1$, we have

$$\mathrm{P}_\mu\left\{\sup_{\lambda\in(\mu+\epsilon,K]}\frac{p_\lambda(y_{[1]},y_{[2]})}{p_\mu(y_{[1]},y_{[2]})} > 1\middle| y_{[1]}\right\} \quad\leq\quad \exp\left\{\frac{\log\gamma}{4}(m\wedge n)\right\}$$

on $\mathcal{A}_\delta$ (note that $\log\gamma$ is negative), or, equivalently,

$$\begin{aligned}
&\mathrm{P}_\mu\left\{\sup_{\lambda\in(\mu+\epsilon,K]}\frac{p_\lambda(y_{[1]},y_{[2]})}{p_\mu(y_{[1]},y_{[2]})} > 1, |\Delta|\leq\delta\middle| y_{[1]}\right\}\\
&\leq \exp\left\{\frac{\log\gamma}{4}(m\wedge n)\right\}1_{\mathcal{A}_\delta} \qquad (2)
\end{aligned}$$

everywhere. Now, for any $0<\eta<1$, there is $N_2\geq N_1$ such that, when $m\wedge n\geq N_2$, we have $\mathrm{P}_\mu(|\Delta|>\delta)<\eta/2$. Also, let $N_3=[4(\log\eta-\log 2)/\log\gamma]+1$. By taking expectation on both sides of (2), we have, when $m\wedge n\geq N_3$,

$$\mathrm{P}_\mu\left\{\sup_{\lambda\in(\mu+\epsilon,K]}\frac{p_\lambda(y_{[1]},y_{[2]})}{p_\mu(y_{[1]},y_{[2]})} > 1, |\Delta|\leq\delta\right\} \leq \exp\left\{\frac{\log\gamma}{4}(m\wedge n)\right\} < \eta/2.$$

Thus, when $m\wedge n\geq N_2\vee N_3$, we have

$$\mathrm{P}_\mu\left\{\sup_{\lambda\in(\mu+\epsilon,K]}\frac{p_\lambda(y_{[1]},y_{[2]})}{p_\mu(y_{[1]},y_{[2]})} > 1\right\} \quad<\quad \eta.$$

A similar result, with $(\mu+\epsilon,K]$ replaced by $[-K,\mu-\epsilon)$, can be proved.

2. Proof of Theorem 3 of Jiang2012. For any $\theta \in \Theta$, $\theta \neq \theta_0$, by *A2* of Jiang2012, there is $1 \leq a \leq b$ such that

$$\limsup_{N\to\infty} \frac{1}{m_a}\sum_{j=1}^{m_a} \mathrm{E}_{\theta_0}\left[\log\left\{\frac{p_\theta(y_{a,j})}{p_{\theta_0}(y_{a,j})}\right\}\right] \quad < \quad 0. \tag{3}$$

Let $y_{[1]}$ denote the combined vector of $y_{a,j}, 1 \leq j \leq m_a$, and $y_{[2]}$ the vector of the rest of the y's. By the subset argument [see (2) of Jiang2012], we have

$$\mathrm{P}_{\theta_0}\{p_{\theta_0}(y_{[1]}, y_{[2]}) \leq p_\theta(y_{[1]}, y_{[2]})|y_{[1]}\} \quad \leq \quad \frac{p_\theta(y_{[1]})}{p_{\theta_0}(y_{[1]})}. \tag{4}$$

On the other hand, we have

$$\begin{aligned}\frac{1}{m_a}\log\left\{\frac{p_\theta(y_{[1]})}{p_{\theta_0}(y_{[1]})}\right\} &= \frac{1}{m_a}\sum_{j=1}^{m_a}\log\left\{\frac{p_\theta(y_{a,j})}{p_{\theta_0}(y_{a,j})}\right\}\\ &= \frac{1}{m_a}\sum_{j=1}^{m_a}\mathrm{E}_{\theta_0}\left[\log\left\{\frac{p_\theta(y_{a,j})}{p_{\theta_0}(y_{a,j})}\right\}\right] + \frac{1}{m_a}\sum_{j=1}^{m_a}\Delta_j, \end{aligned}\tag{5}$$

where $\Delta_j = \log\{p_\theta(y_{a,j})/p_{\theta_0}(y_{a,j})\} - \mathrm{E}_{\theta_0}[\log\{p_\theta(y_{a,j})/p_{\theta_0}(y_{a,j})\}]$. By *A3* of Jiang2012, the second term on the right side of (5) is $o_{\mathrm{P}}(1)$. Thus, combined with (3), there is a constant $\lambda > 0$ such that, with probability tending to one, we have $m_a^{-1}\log\{p_\theta(y_{[1]})/p_{\theta_0}(y_{[1]})\} \leq -\lambda$; hence, by (4), we have

$$\mathrm{P}_{\theta_0}\{p_{\theta_0}(y_{[1]}, y_{[2]}) \leq p_\theta(y_{[1]}, y_{[2]})|y_{[1]}\} \quad \leq \quad e^{-\lambda m_a}. \tag{6}$$

The arguments have shown that the left side of (6) is $O_{\mathrm{P}}(e^{-\lambda m_a})$. Thus, by *A1* of Jiang2012 and the dominated convergence theorem, we have

$$\mathrm{P}_{\theta_0}\{p_{\theta_0}(y_{[1]}, y_{[2]}) \leq p_\theta(y_{[1]}, y_{[2]})\} \to 0 \quad\Rightarrow\quad \mathrm{P}_{\theta_0}(\hat\theta \neq \theta) \to 1, \tag{7}$$

as $N \to \infty$. Because (7) holds for every $\theta \in \Theta \setminus \{\theta_0\}$, and Θ is finite, the proof is complete.

3. Some details of the proof of Theorem 4 of Jiang2012. First we establish (14) of Jiang2012. It is easy to show that $|D| \leq 2dK^{d-1}$. Furthermore, for any $\theta \in \partial C_\epsilon \cap \Theta_{N,a}$, there is a point $\theta_l \in D$ such that $|\theta_c - \theta_{l,c}| \leq 2\epsilon/K, 1 \leq c \leq d$. Thus, by the Taylor expansion, there is a point $\tilde\theta$ that lies between θ and θ_l such that

$$|\log\{p_\theta(y)\} - \log\{p_{\theta_l}(y)\}| = \left|\sum_{c=1}^{d}\left\{\frac{\partial}{\partial\theta_c}\log p_\theta(y)\bigg|_{\theta=\tilde\theta}\right\}(\theta_c - \theta_{l,c})\right| \leq \frac{2d\epsilon B}{K},$$

implying $p_\theta(y) \le \exp(2d\epsilon B/K) p_{\theta_l}(y)$ [B is the left side of (9) in Jiang2012]. It follows that $\sup_{\theta\in\partial C_\epsilon \cap \Theta_{N,a}} p_\theta(y) \le \exp(2d\epsilon B/K)\max_{\theta\in D} p_\theta(y)$; hence

$$\begin{aligned} & \mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y) \le \sup_{\theta\in\partial C_\epsilon\cap\Theta_{N,a}} p_\theta(y)\right\} \\ \le\ & \mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y) \le \exp\left(\frac{2d\epsilon B}{K}\right)\max_{\theta\in D} p_\theta(y)\right\} \\ \le\ & \mathrm{P}_{\theta_0}\left\{\exp\left(\frac{2d\epsilon B}{K}\right) > 2\right\} + \mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y) \le 2\max_{\theta\in D} p_\theta(y)\right\}. \end{aligned}$$

Next, we show that $\mathrm{P}_{\theta_0}\{p_{\theta_0}(y) \le 2\max_{\theta\in D} p_\theta(y)|y_{[1]}\} = o_{\mathrm{P}}(1)$. By the subset inequality, that is, (15) of Jiang2012, we have

$$\begin{aligned} \mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y) \le 2\max_{\theta\in D} p_\theta(y)\middle| y_{[1]}\right\} &\le \sum_{\theta\in D}\mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y) \le 2p_\theta(y)|\, y_{[1]}\right\} \\ &\le 2\sum_{\theta\in D}\frac{p_\theta(y_{[1]})}{p_{\theta_0}(y_{[1]})}. \end{aligned} \tag{8}$$

We now define a new collection of points. Let $L = [s_{a,N}^{-1}K] + 1$. Let G be the largest integer such that $GL \le K$. Then, we have $G \le s_{a,N}$. For any $g = (g_1, \dots, g_d)$, where $g_1, \dots, g_d$ are integers such that $0 \le g_c \le G-1, 1 \le c \le d$, select a point $\theta_{(g)}$ from the subset $\{\theta : \theta_{0c} - \epsilon + 2\epsilon g_c L/K \le \theta_c \le \theta_{0c} - \epsilon + 2\epsilon(g_c+1)L/K, 1 \le c \le d\} \cap \partial C_\epsilon \cap \Theta_{N,a}$, if the latter is not empty; otherwise, do not select. Let D_1 be the collection of all such points selected. Similarly, we have $|D_1| \le 2dG^{d-1} \le 2ds_{a,N}^{d-1}$. Furthermore, for any $\theta \in D$, there is a $\theta_{(g)} \in D_1$ such that $|\theta_c - \theta_{(g),c}| \le 2\epsilon L/K, 1 \le c \le d$. Thus, by the Taylor expansion, there is a $\tilde\theta$ that lies between θ and $\theta_{(g)}$ such that

$$\begin{aligned} \log\{p_\theta(y_{[1]})\} - \log\{p_{\theta_{(g)}}(y_{[1]})\} &= \sum_{j=1}^{m_a}\log\{p_\theta(y_{a,j})\} - \sum_{j=1}^{m_a}\log\{p_{\theta_{(g)}}(y_{a,j})\} \\ &= \sum_{c=1}^{d}\sum_{j=1}^{m_a}\frac{\partial}{\partial\theta_c}\log\{p_{\tilde\theta}(y_{a,j})\}\{\theta_c - \theta_{(g),c}\} \\ &= \sum_{j=1}^{m_a}\sum_{c=1}^{d}\frac{\partial}{\partial\theta_c}\log\{p_{\tilde\theta}(y_{a,j})\}\{\theta_c - \theta_{(g),c}\}. \end{aligned}$$

It is then easy to derive that

$$\max_{\theta\in D}\frac{p_\theta(y_{[1]})}{p_{\theta_0}(y_{[1]})} \le \exp\left(\frac{2d\epsilon L}{K}B_a m_a\right)\max_{\theta\in D_1}\frac{p_\theta(y_{[1]})}{p_{\theta_0}(y_{[1]})}, \tag{9}$$

where B_a is the left side of (10) in Jiang2012.

Next, by *B2* of Jiang2012, there are constant $\lambda \in (0,\infty)$ and positive integer N_1 such that

$$\sup_{\theta\in\Theta,\epsilon\le|\theta-\theta_0|\le M}\min_{1\le a'\le b}\frac{1}{m_{a'}}\sum_{j=1}^{m_{a'}}\mathrm{E}_{\theta_0}\left[\log\left\{\frac{p_\theta(y_{a',j})}{p_{\theta_0}(y_{a',j})}\right\}\right]\le-\lambda, \tag{10}$$

if $N\ge N_1$. For any $\theta\in D_1\subset\partial C_\epsilon\cap\Theta_{N,a}$, we have

$$\frac{1}{m_a}\sum_{j=1}^{m_a}\mathrm{E}_{\theta_0}\left[\log\left\{\frac{p_\theta(y_{a,j})}{p_{\theta_0}(y_{a,j})}\right\}\right]=S_{N,a}(\theta)=\min_{1\le a'\le b}S_{N,a'}(\theta)\le-\lambda,$$

by the definition of $\partial C_\epsilon,\Theta_{N,a}$ and (10). Thus, by (5), we have

$$\frac{1}{m_a}\log\left\{\frac{p_\theta(y_{[1]})}{p_{\theta_0}(y_{[1]})}\right\}\quad\le\quad-\lambda+\max_{\theta\in D_1}\frac{1}{m_a}\sum_{j=1}^{m_a}\Delta_j(\theta), \tag{11}$$

if $N\ge N_1$, where $\Delta_j(\theta)$ is the Δ_j below (5). Because (11) holds for any $\theta\in D_1$, we have

$$\max_{\theta\in D_1}\frac{p_\theta(y_{[1]})}{p_{\theta_0}(y_{[1]})}\quad\le\quad\exp\left[-m_a\left\{\lambda-\max_{\theta\in D_1}\frac{1}{m_a}\sum_{j=1}^{m_a}\Delta_j(\theta)\right\}\right], \tag{12}$$

if $N\ge N_1$. Combining (8), (9) and (12), we have, for $N\ge N_1$,

$$\begin{aligned}&\mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y)\le2\max_{\theta\in D}p_\theta(y)\middle|\,y_{[1]}\right\}\\ \le\;& 4dK^{d-1}\exp\left[-m_a\left\{\lambda-\frac{2d\epsilon L}{K}B_a-\max_{\theta\in D_1}\frac{1}{m_a}\sum_{j=1}^{m_a}\Delta_j(\theta)\right\}\right].\end{aligned} \tag{13}$$

Let $\xi_N=2d\epsilon LB_a/K$, $\eta_N=\max_{\theta\in D_1}m_a^{-1}\sum_{j=1}^{m_a}\Delta_j(\theta)$. By (13), we have

$$\begin{aligned}&\mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y)\le2\max_{\theta\in D}p_\theta(y)\middle|\,y_{[1]}\right\}\\ =\;&\mathrm{P}_{\theta_0}\{\cdots|y_{[1]}\}1_{(\xi_N>\lambda/4\text{ or }\eta_N>\lambda/4)}+\mathrm{P}_{\theta_0}\{\cdots|y_{[1]}\}1_{(\xi_N\le\lambda/4,\eta_N\le\lambda/4)}\\ \le\;&1_{(\xi_N>\lambda/4)}+1_{(\eta_N>\lambda/4)}+4dK^{d-1}\exp\left(-\frac{\lambda}{2}m_a\right),\end{aligned} \tag{14}$$

if $N\ge N_1$. It remains to evaluate the three terms on the right side of (14). The last term is bounded by

$$4d(2e^{\delta m_a})^{d-1}\exp\left\{-\frac{\lambda}{2}m_a\right\}=2^{d+1}d\exp\left[-m_a\left\{\frac{\lambda}{2}-(d-1)\delta\right\}\right],$$

which goes to zero if $(d-1)\delta < \lambda/2$. Next, it is easy to show that $s_{a,N}/K \le \exp[-m_a\{\delta - \log(s_{a,N})/m_a\}] \to 0$, by *B3* of Jiang2012. Thus, there is $N_2 \ge 1$ such that $s_{a,N}^{-1}K \ge 1$, implying $L/K \le 2s_{a,N}^{-1}$, if $N \ge N_2$. It follows that $\xi_N \le 2d\epsilon(B_a/s_{a,N}) = o_{\rm P}(1)$, by *B3* of Jiang2012. It follows that the first term on the right side of (14) $\stackrel{L^1}{\longrightarrow} 0$, hence is $o_{\rm P}(1)$. Also, we have

$$\begin{aligned} \mathrm{P}_{\theta_0}\left(\eta_N \ge \frac{\lambda}{4}\right) &\le \sum_{\theta \in D_1} \mathrm{P}_{\theta_0}\left\{\frac{1}{m_a}\sum_{j=1}^{m_a}\Delta_j(\theta) \ge \frac{\lambda}{4}\right\} \\ &\le \frac{16}{\lambda^2}\sum_{\theta \in D_1} \mathrm{E}_{\theta_0}\left\{\frac{1}{m_a}\sum_{j=1}^{m_a}\Delta_j(\theta)\right\}^2 \\ &= \frac{16}{\lambda^2}\sum_{\theta \in D_1}\frac{1}{m_a^2}\sum_{j=1}^{m_a}\mathrm{var}_{\theta_0}\left[\log\left\{\frac{p_\theta(y_{a,j})}{p_{\theta_0}(y_{a,j})}\right\}\right] \\ &\le \frac{32d}{\lambda^2} \times \text{the left side of (11) of Jiang2012,} \end{aligned}$$

because $|D_1| \le 2ds_{a,N}^{d-1}$ [see the note below (8)]. It follows, again, that the second term on the right side of (14) $\stackrel{L^1}{\longrightarrow} 0$, hence is $o_{\rm P}(1)$. Therefore, we have $\mathrm{P}_{\theta_0}\{p_{\theta_0}(y) \le 2\max_{\theta \in D} p_\theta(y)|y_{[1]}\} = o_{\rm P}(1)$.

Next, we show (12) of Jiang2012. It then follows by the dominated convergence theorem that

$$\mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y) \le 2\max_{\theta \in D} p_\theta(y)\right\} \longrightarrow 0. \tag{15}$$

In addition, we have

$$\frac{2d\epsilon B}{K} = 2d\epsilon\left(\frac{B}{s_N}\right)\left(\frac{s_N}{K}\right) \le 2d\epsilon\left(\frac{B}{s_N}\right)\exp\left[-m_a\left\{\delta - \frac{\log(s_N)}{m_a}\right\}\right],$$

which is $o_{\rm P}(1)$ by *B3* of Jiang2012. It follows that

$$\mathrm{P}_{\theta_0}\left\{\exp\left(\frac{2d\epsilon B}{K}\right) > 2\right\} = \mathrm{P}_{\theta_0}\left\{\frac{2d\epsilon B}{K} > \log(2)\right\} \longrightarrow 0. \tag{16}$$

The result follows by combining (13), (14) of Jiang2012, and (15), (16).

4. Proof of Theorem 5 of Jiang2012. Let $\epsilon = (\rho - \delta)/2$. For any $M > 0$, define $J_N = [\sqrt{d} \vee (dMc_N b_k)] + 1$. For any $l = (l_1, \dots, l_d)$ and $j = (j_1, \dots, j_d)$, where $l_c = 1, 2, \dots$ and $j_c \in \{0, \dots, J_N - 1\}$, $1 \le c \le d$, define $r(l, j) = \{w \in R^d : l_c + j_c/J_N \le w_c \le l_c + (j_c + 1)/J_N, 1 \le c \le$

$d\}$. Then, we can find a collection $r(l,j),(l,j)\in C_k$ such that $S_d(k)\subset \cup_{(l,j)\in C_k} r(l,j)$ and $|C_k|\le (2J_N)^d(k+1)^{d_1}$, where $d_1=d1_{(d>1)}$. To see this, note that this is clearly true when $d=1$. If $d>1$, then for any $w\in S_d(k)$, we must have $|w_c|\le|w|<k+1, 1\le c\le d$. Therefore, $S_d(k)\subset \cup_{-(k+1)\le l_c\le k, 0\le j_c\le J_N-1, 1\le c\le d} c(l,j)$, and there are $\{2(k+1)J_N\}^d$ such $c(l,j)$'s in the union. [Note: It can be shown that, if $d>1$, the number of $c(l,j)$ that are entirely inside $S_d(k)$ is in the order of $(kJ_N)^d$; thus, although the upper bound $\{2(k+1)J_N\}^d$ may not be very accurate, at least it gets the order right, which is all that matters.] For each $(l,j)\in C_k$, select a point $\theta(l,j)\in c(l,j)\cap\Theta\cap S_d(k)$, if the latter is not empty; otherwise, do not select. Let D_k be the collection of all such points. Then, we have $|D_k|\le (2J_N)^d(k+1)^{d_1}$. For any $\theta\in\Theta\cap S_d(k)$, there is $\theta(l,j)\in D_k$ such that $|\theta_c-\theta(l,j)_c|\le 1/J_N, 1\le c\le d$. Thus, by the Taylor expansion and *C1* of Jiang2012, it is easy to show that $\log p_\theta(y)-\log p_{\theta(l,j)}(y)\le 1$, if $\zeta_N\le M$. Note that the convexity of Θ implies that $(1-t)\theta_1+t\theta_2\in\Theta\cap S_d[k-1,k+2)$, if $\theta_j\in\Theta\cap S_d(k), j=1,2$ and $|\theta_{1c}-\theta_{2c}|\le 1/J_N, 1\le c\le d$. It follows that $\sup_{\theta\in\Theta\cap S_d(k)} p_\theta(y)\le e\max_{\theta\in D_k} p_\theta(y)$, if $\zeta_N\le M$. Therefore, we have

$$
\begin{aligned}
& \mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y)\le \sup_{\theta\in\Theta\cap S_d(k)} p_\theta(y), \zeta_N\le M, \eta_N\le M, \Delta_N\le\epsilon\right\} \\
(17)\qquad \le\ & \mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y)\le e\max_{\theta\in D_k} p_\theta(y), \eta_N\le M, \Delta_N\le\epsilon\right\},
\end{aligned}
$$

where η_N,Δ_N are defined in the sequel.

Next, define $y_{[1]}$ as the combined vector of $y_{(1)},\dots,y_{(m_N)}$, and $y_{[2]}$ as the rest of the y data. By the subset argument of Jiang2012, we have

$$
\begin{aligned}
\mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y)\le e\max_{\theta\in D_k} p_\theta(y)\middle| y_{[1]}\right\} &\le \sum_{\theta\in D_k}\mathrm{P}_{\theta_0}\{p_{\theta_0}(y)\le e p_\theta(y)|y_{[1]}\} \\
(18)\qquad &\le e\sum_{\theta\in D_k}\frac{p_\theta(y_{[1]})}{p_{\theta_0}(y_{[1]})}.
\end{aligned}
$$

For every $\theta\in D_k$, we have $\log\{p_\theta(y_{[1]})/p_{\theta_0}(y_{[1]})\}=\sum_{j=1}^{m_N}\log\{p_{j,\theta}(y_{(j)})\}-\sum_{j=1}^{m_N}\log\{p_{j,\theta_0}(y_{(j)})\}=I_1-I_2$. Let $t\in\mathcal{T}_N$ be the one that satisfies (ii) of *C2* in Jiang2012. Define $\mathcal{J}_{N,s}=\{1\le j\le m_N: y_{(j)}=s\}, s\in\mathcal{T}_N$. Then, we have $I_1\le\sum_{j\in\mathcal{J}_{N,t}}\log\{p_{j,\theta}(t)\}\le-\gamma_k|\mathcal{J}_{N,t}|\le-\gamma_k\min_{s\in\mathcal{T}_N}|\mathcal{J}_{N,s}|$. Also, for any $s\in\mathcal{T}_N$, we have $|\mathcal{J}_{N,s}|=m_N[m_N^{-1}\sum_{j=1}^{m_N}p_{j,\theta_0}(s)+m_N^{-1}\sum_{j=1}^{m_N}\Delta_{j,s}]$, where $\Delta_{j,s}=1_{\{y_{(j)}=s\}}-p_{j,\theta_0}(s)$. Let $\Delta_N=m_N^{-1}\max_{t\in\mathcal{T}_N}|\sum_{j=1}^{m_N}\Delta_{j,t}|$. Then, by (iii) of *C2* in Jiang2012, we have $|\mathcal{J}_{N,s}|\ge m_N(\rho-\epsilon)$, for every $s\in\mathcal{T}_N$, hence $I_1\le-(\rho-\epsilon)m_N\gamma_k$, if $\Delta_N\le\epsilon$. On the other hand, we have $-I_2\le m_N\eta_N$, where $\eta_N = m_N^{-1}\sum_{j=1}^{m_N}|\log\{p_{j,\theta_0}(y_{(j)})\}|$. Thus, if $\eta_N \le M$ and $\Delta_N \le \epsilon$, we have $\log\{p_\theta(y_{[1]})/p_{\theta_0}(y_{[1]})\} \le -(\rho-\epsilon)m_N\gamma_k + Mm_N = -m_N\{(\rho-\epsilon)\gamma_k - M\} \le -m_N(\rho-2\epsilon)\gamma_k = -\delta m_N\gamma_k$ if k is large, say, $k \ge K_1$ for some K_1. Note that both η_N and Δ_N are $\mathcal{F}(y_{[1]})$ measurable. Thus, we have, by (18),

$$\begin{aligned}
& \mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y) \le e\max_{\theta\in D_k} p_\theta(y), \eta_N \le M, \Delta_N \le \epsilon \middle| y_{[1]}\right\} \\
= \;& \mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y) \le e\max_{\theta\in D_k} p_\theta(y) \middle| y_{[1]}\right\} 1_{(\eta_N\le M, \Delta_N\le\epsilon)} \\
\le \;& e\sum_{\theta\in D_k}\frac{p_\theta(y_{[1]})}{p_{\theta_0}(y_{[1]})}1_{(\eta_N\le M,\Delta_N\le\epsilon)} \\
\le \;& e|D_k|e^{-\delta m_N\gamma_k} \\
\le \;& c(M)c_N^d k^{d_1} b_k^d e^{-\delta m_N\gamma_k}, \qquad (19)
\end{aligned}$$

where $c(M) = 2^{d_1}e\{4(dM+\sqrt{d})\}^d$. Combining (17) and (19), we have

$$\begin{aligned}
& \mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y) \le \sup_{\theta\in\Theta\cap S_d(k)} p_\theta(y), \zeta_N\vee\eta_N \le M, \Delta_N \le \epsilon\right\} \\
\le \;& c(M)c_N^d k^{d_1} b_k^d e^{-\delta m_N\gamma_k}. \qquad (20)
\end{aligned}$$

As (20) holds for every $k \ge K_1$, we have

$$\begin{aligned}
& \mathrm{P}_{\theta_0}\{p_{\theta_0}(y) \le p_\theta(y) \text{ for some } \theta\in\Theta \text{ with } |\theta| \ge K\vee K_1\} \\
\le \;& \mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y) \le \sup_{\theta\in\Theta\cap S_d(k)} p_\theta(y) \text{ for some } k \ge K\vee K_1\right\} \\
\le \;& \mathrm{P}_{\theta_0}(\zeta_N\vee\eta_N > M) + \mathrm{P}_{\theta_0}(\Delta_N > \epsilon) \\
& +\mathrm{P}_{\theta_0}\left[\cup_{k=K\vee K_1}^\infty\left\{p_{\theta_0}(y) \le \sup_{\theta\in\Theta\cap S_d(k)} p_\theta(y), \zeta_N\vee\eta_N \le M, \Delta_N \le \epsilon\right\}\right] \\
\le \;& \mathrm{P}_{\theta_0}(\zeta_N\vee\eta_N > M) + \mathrm{P}_{\theta_0}(\Delta_N > \epsilon) \\
& + \sum_{k=K\vee K_1}^\infty \mathrm{P}_{\theta_0}\left\{p_{\theta_0}(y) \le \sup_{\theta\in\Theta\cap S_d(k)} p_\theta(y), \zeta_N\vee\eta_N \le M, \Delta_N \le \epsilon\right\} \\
\le \;& \mathrm{P}_{\theta_0}(\zeta_N\vee\eta_N > M) + \mathrm{P}_{\theta_0}(\Delta_N > \epsilon) + c(M)c_N^d\sum_{k=K\vee K_1}^\infty k^{d_1}b_k^d e^{-\delta m_N\gamma_k} \\
\le \;& \mathrm{P}_{\theta_0}(\zeta_N\vee\eta_N > M) + \mathrm{P}_{\theta_0}(\Delta_N > \epsilon) + c(M)c_N^d\sum_{k=K}^\infty k^{d_1}b_k^d e^{-\delta m_N\gamma_k}.
\end{aligned}$$

It is then straightforward to argue that, as $N\to\infty$,

$$\mathrm{P}_{\theta_0}\{p_{\theta_0}(y) \le p_\theta(y) \text{ for some } \theta\in\Theta \text{ with } |\theta| \ge K\vee K_1\} \longrightarrow 0. \qquad (21)$$

On the other hand, by almost the same arguments as in the proof of Theorem 4 of Jiang2012, it can be shown that, for any $0 < \epsilon < M$, we have

$$\mathrm{P}_{\theta_0}\{p_{\theta_0}(y) \le p_\theta(y) \text{ for some } \theta \in \Theta \text{ with } \epsilon < |\theta - \theta_0| \le M\} \longrightarrow 0, \tag{22}$$

as $N \to \infty$. The result thus follows.

5. Some detailed derivations in Section 4 of Jiang2012. Regarding identity (20) of Jiang2012, we have

$$\begin{aligned} p_\gamma(1,1) &= \mathrm{E}_\gamma\{h(\mu_0 + X)h(\mu_0 + Y)\} \\ &= \mathrm{E}_\gamma\left\{\int_0^\infty 1_{(s \le h(\mu_0+X))} ds \int_0^\infty 1_{(t \le h(\mu_0+Y))} dt\right\} \\ &= \mathrm{E}_\gamma\left\{\int_0^\infty \int_0^\infty 1_{(s \le h(\mu_0+X), t \le h(\mu_0+Y))} ds dt\right\} \\ &= \int_0^\infty \int_0^\infty P_\gamma\{s \le h(\mu_0 + X), t \le h(\mu_0 + Y)\} ds dt \\ &= \int_0^\infty \int_0^\infty P_\gamma\{X \ge \mathrm{logit}(s) - \mu_0, Y \ge \mathrm{logit}(t) - \mu_0\} ds dt. \end{aligned}$$

Regarding the bounds for the partial derivatives, note that, because $\sigma_0^2 > 0$ and $\tau_0^2 > 0$, there is a neighborhood of θ_0, $\mathcal{N}(\theta_0)$, and constants $A, B, C > 0$ such that the following hold uniformly for $\theta \in \mathcal{N}(\theta_0)$:

$$\left(\frac{m}{\sigma^2}\right) \vee \left(\frac{n}{\tau^2}\right) \le C \cdot N, \tag{23}$$

$$\{|\mu| + 2 + \log(1 + e^{|\mu|+2}) - B\}N + \frac{m}{2\sigma^2} + \frac{n}{2\tau^2} \le 0, \tag{24}$$

$$\begin{aligned} &\log(4\sqrt{\pi}\sigma^3) + \frac{m-1}{2}\log(4\pi\sigma^2) + \frac{n}{2}\log(2\pi\tau^2) + \log(m) \\ \le\ &\left(\frac{A}{4\sigma^2} - B\right) N + (m + n)\log(2), \end{aligned} \tag{25}$$

$$\begin{aligned} &\log(4\sqrt{\pi}\tau^3) + \frac{n-1}{2}\log(4\pi\tau^2) + \frac{m}{2}\log(2\pi\sigma^2) + \log(n) \\ \le\ &\left(\frac{A}{4\tau^2} - B\right) N + (m + n)\log(2), \end{aligned} \tag{26}$$

where $N = \sum_{(i,j)\in S} c_{ij}$, the total sample size. Note that the irreducibility of S implies $N \geq m \vee n$. It can be shown that, in this case, we have

$$
\begin{aligned}
\log\{p_\theta(y)\} &= c - \frac{m}{2}\log\sigma^2 - \frac{n}{2}\log\tau^2 \\
&+ \log\int\cdots\int \exp\left(s_0 + s_1 + s_2 - s_3 - \frac{1}{2\sigma^2}\sum_{i=1}^m u_i^2 - \frac{1}{2\tau^2}\sum_{j=1}^n v_j^2\right) du dv, \qquad (27)
\end{aligned}
$$

where $s_0 = \mu\sum_{(i,j)\in S} y_{i,j,\cdot}$ with $y_{i,j,\cdot} = \sum_{k=1}^{c_{ij}} y_{i,j,k}$, $s_1 = \sum_{(i,j)\in S} y_{i,j,\cdot}u_i$, $s_2 = \sum_{(i,j)\in S} y_{i,j,\cdot}v_j$, $s_3 = \sum_{(i,j)\in S} c_{ij}\log(1 + e^{\mu+u_i+v_j})$, $du = \prod_{i=1}^m du_i$, and $dv = \prod_{j=1}^n dv_j$. Thus, we have the expression $(\partial/\partial\mu)\log\{p_\theta(y)\} = I_\mu/I$, where $I = \int\cdots\int e^\eta du dv$, η being the expression inside the exponential on the right side of (27), and

$$
I_\mu = \int\cdots\int\left[\sum_{(i,j)\in S}\sum_{k=1}^{c_{ij}}\{y_{i,j,k} - h(\mu + u_i + v_j)\}\right] e^\eta du dv
$$

$[h(x) = e^x/(1 + e^x)]$. It follows that $|(\partial/\partial\mu)\log\{p_\theta(y)\}| \leq N$. Similarly, we have $(\partial/\partial\sigma^2)\log\{p_\theta(y)\} = -m/2\sigma^2 + I_{\sigma^2}/I$, where

$$
\begin{aligned}
I_{\sigma^2} &= \frac{1}{2\sigma^4}\int\cdots\int\left(\sum_{i=1}^m u_i^2\right) e^\eta du dv \\
&= \int\cdots\int \zeta 1_{(\sum_{i=1}^m u_i^2 \leq A\cdot N)} du dv \\
&\quad + \int\cdots\int \zeta 1_{(\sum_{i=1}^m u_i^2 > A\cdot N)} du dv \\
&= I_{\sigma^2,1} + I_{\sigma^2,2}
\end{aligned}
$$

with $\zeta = (\sum_{i=1}^m u_i^2)e^\eta$. We have $0 \leq I_{\sigma^2,1} \leq (A\cdot N)I$. Also, we have $\sum_{i=1}^m u_i^2 \geq (A/2)N + (1/2)\sum_{i=1}^m u_i^2$, if $\sum_{i=1}^m u_i^2 > A\cdot N$; and $y_{i,j,k}(\mu + u_i + v_j) - \log(1 + e^{\mu+u_i+v_j}) \leq 0$ for any $y_{i,j,k} = 0$ or 1, u_i, and v_j. It follows that

$$
\begin{aligned}
& s_0 + s_1 + s_2 - s_3 \\
&= \sum_{(i,j)\in S}\sum_{k=1}^{c_{ij}}\{y_{i,j,k}(\mu + u_i + v_j) - \log(1 + e^{\mu+u_i+v_j})\} \leq 0 \qquad (28)
\end{aligned}
$$

always holds. Thus, we have, by (25),

$$
\begin{aligned}
0 &\leq I_{\sigma^2,2} \\
&\leq \exp\left(-\frac{A}{4\sigma^2}N\right)\int\cdots\int\left(\sum_{i=1}^m u_i^2\right)
\end{aligned}
$$

$$\times \exp\left(-\frac{1}{4\sigma^2}\sum_{i=1}^m u_i^2 - \frac{1}{2\tau^2}\sum_{j=1}^n v_j^2\right) dudv$$

$$= 4\sqrt{\pi}\sigma^3 m(4\pi\sigma^2)^{(m-1)/2}(2\pi\tau^2)^{n/2}\exp\left(-\frac{A}{4\sigma^4}N\right)$$

$$= \exp\{\text{LS of (25)} - \text{RS of (25)} + (m+n)\log(2) - BN\}$$

$$\leq 2^{m+n}e^{-BN} \tag{29}$$

for $\theta \in \mathcal{N}(\theta_0)$, where LS (RS) stands for left (right) side. On the other hand, if $|u_i| \leq 1, |v_j| \leq 1$ for all i, j, we have, by (24) and (28),

$$|\eta| \leq \{|\mu| + 2 + \log(1 + e^{|\mu|+2})\}N + \frac{m}{2\sigma^2} + \frac{n}{2\tau^2} \leq BN,$$

hence $I \geq e^{-BN}\int_{-1}^1 \cdots \int_{-1}^1 dudv = 2^{m+n}e^{-BN}$. Therefore, by (29), we have $0 \leq I_{\sigma^2,2} \leq I$ for all $\theta \in \mathcal{N}(\theta_0)$. It follows that $|(\partial/\partial\sigma^2)\log\{p_\theta(y)\}| \leq (A + C + 1)N, \theta \in \mathcal{N}(\theta_0)$. By a similar argument, it can be shown that $|(\partial/\partial\tau^2)\log\{p_\theta(y)\}| \leq (A + C + 1)N, \theta \in \mathcal{N}(\theta_0)$.

6. Checking assumptions *C1*, *C2* of Jiang2012. First consider the open problem (Sections 1 and 2 of Jiang2012). Assumption *C1* has, in fact, already been verified in the previous section, with $c_N = mn$, $\zeta_N = 1$ and $b_k = 1$. Note that this bound holds uniformly for all k.

As for assumption *C2*, let μ_0 denote the true μ. Consider $y_{(j)} = y_{jj}, 1 \leq j \leq m \wedge n$. Then, $\log\{p_{j,\mu}(y_{jj})\} = y_{jj}\log p + (1 - y_{jj})\log(1 - p)$, where $p = \mathrm{E}h(\mu + \xi)$, $h(x) = e^x/(1 + e^x)$, and $\xi \sim N(0, 2)$. It follows that $\mathrm{E}_{\mu_0}|\log\{p_{j,\mu_0}(y_{jj})\}| \leq |\log p_0| + |\log(1 - p_0)|$, where p_0 is p with $\mu = \mu_0$. Thus, (i) is satisfied.

Next, for any μ such that $k \leq |\mu| < k + 1$, if $k \leq \mu < k + 1$, we have $\log\{p_{j,\mu}(0)\} = \log(1 - p) \leq 1 - k$; if $-(k + 1) < \mu \leq -k$, we have $\log\{p_{j,\mu}(1)\} = \log p \leq 1 - k$ (see the proof of Theorem 2 in Jiang2012). Thus, (ii) is satisfied with $\gamma_k = k - 1$.

(iii) holds because $(m \wedge n)^{-1}\sum_{j=1}^{m\wedge n} p_{j,\mu_0}(t) = \mathrm{P}_{\mu_0}(y_{11} = t) > 0, t = 0, 1$. (iv) holds because $\mathcal{T}_N = \{0, 1\}$, $m_N = m \wedge n$; and $c_N\sum_{k=2}^\infty e^{-\delta(m\wedge n)(k-1)} = \{1 - e^{-\delta(m\wedge n)}\}^{-1}\exp\{-(m \wedge n)(\delta - (m \wedge n)^{-1}(\log m + \log n)\} \to 0$ for any $\delta > 0$, provided that $(m \wedge n)^{-1}\log(m \vee n) \to 0$.

As another example, we consider the example of Section 4 in Jiang2012. It is more convenient to consider $\theta = (\mu, \sigma, \tau)'$ as the parameter vector. The likelihood function can be expressed as $p_\theta(y) = \mathrm{E}(e^\zeta)$, where $\zeta = \sum_{(i,j)\in S}\sum_{k=1}^{c_{ij}}\{(\mu + \sigma\xi_i + \tau\eta_j)y_{i,j,k} - \log(1 + e^{\mu+\sigma\xi_i+\tau\eta_j})\}$, and $\xi_i, 1 \leq i \leq$

m, $\eta_j, 1 \le j \le n$ are independent $N(0,1)$ random variables. Write $l = \log p_\theta(y)$. It is shown in Jiang2012 that $|\partial l/\partial\mu| \le N$. Next, we have $\partial l/\partial\sigma = \sum_{(i,j)\in S}\sum_{k=1}^{c_{ij}} \mathrm{E}\{e^\zeta(y_{i,j,k} - h_{ij})\xi_i\}/\mathrm{E}(e^\zeta)$, where $h_{ij} = h(\mu + \sigma\xi_i + \tau\eta_j)$ and $h(\cdot)$ is defined above. Define

$$\alpha_k \;=\; k + 4 + \log 2 + 2\{\log(k+2) - \log c_0\}, \tag{30}$$

where $c_0 = \sqrt{2/\pi}e^{-1/18}$. For any $\theta \in \Theta \cap S_3[k-1, k+2)$, we have

$$\begin{aligned}\mathrm{E}(e^\zeta|\xi_i|) &= \mathrm{E}\{e^\zeta|\xi_i|1_{(|\xi_i|\le 2\sqrt{N\alpha_k})}\} + \mathrm{E}\{e^\zeta|\xi_i|1_{(|\xi_i|>2\sqrt{N\alpha_k})}\}\\ &\le 2\sqrt{N\alpha_k}\mathrm{E}(e^\zeta) + 2\sqrt{2/\pi}e^{-N\alpha_k},\end{aligned}$$

because $e^\zeta \le 1$ (the conditional pmf of y given ξ, η), and $\mathrm{E}\{|\xi_i|1_{(|\xi_i|>a)}\} = \int_{|x|>a}|x|(2\pi)^{-1/2}e^{-x^2/2}dx \le (2\pi)^{-1/2}e^{-a^2/4}\int|x|e^{-x^2/4}dx = 2\sqrt{2/\pi}e^{-a^2/4}$ for any $a \ge 0$. On the other hand, it is easy to show that $x - \log(1 + e^x) \ge x \wedge 0 - \log 2$ and $-\log(1+e^x) \ge -x \vee 0 - \log 2$. It follows that $(\mu + \sigma\xi_i + \tau\eta_j)y_{i,j,k} - \log(1 + e^{\mu+\sigma\xi_i+\tau\eta_j}) \ge -|\mu + \sigma\xi_i + \tau\eta_j| - \log 2$. Thus, if $|\xi_{i'}| \le (k+2)^{-1}$ and $|\eta_{j'}| \le (k+2)^{-1}$ for all i', j', we have $|\mu + \sigma\xi_{i'} + \tau\eta_{j'}| \le k + 4$ for all i', j', hence $\zeta \ge -N(k + 4 + \log 2)$. It follows that

$$\begin{aligned}\mathrm{E}(e^\zeta) &\ge \mathrm{E}\{e^\zeta 1_{(|\xi_{i'}|\le(k+2)^{-1},1\le i'\le m,|\eta_{j'}|\le(k+2)^{-1},1\le j'\le n)}\}\\ &\ge e^{-N(k+4+\log 2)}[\mathrm{P}\{|\xi_1| \le (k+2)^{-1}\}]^{m+n}\\ &\ge e^{-N(k+4+\log 2)}\{c_0(k+2)^{-1}\}^{m+n},\end{aligned}$$

where c_0 is defined below (30). It is then easy to show that

$$\begin{aligned}&\frac{|\mathrm{E}\{e^\zeta(y_{ij} - h_{ij})\xi_i\}|}{\mathrm{E}(e^\zeta)}\\ \le\;& 2\sqrt{N\alpha_k} + 2\sqrt{\frac{2}{\pi}}\\ &\times\exp\left(N\left[k + 4 + \log 2 + 2\{\log(k+2) - \log c_0\} - \alpha_k\right]\right)\\ =\;& 2\sqrt{N\alpha_k} + 2\sqrt{\frac{2}{\pi}}\\ \le\;& c\sqrt{Nk}\end{aligned}$$

for some constant $c > 2$, by (30). Therefore, we have $|\partial l/\partial\sigma| \le cN^{3/2}\sqrt{k}$. The same upper bound can be obtained for $|\partial l/\partial\tau|$, if $\theta \in \Theta \cap S_d[k-1, k+2)$. Thus, assumption *C1* holds with $c_N = N^{3/2}$, $b_k = \sqrt{k}$, and $\zeta_N = c$.

As for assumption *C2*, consider the first subset considered in Jiang2012, Section 4, that is, $y_{i,i} = (y_{i,i,k})_{k=1,2}$ for $(i,i) \in S_2$. It is shown that

$$p_\theta(y_{i,i}) = \mathrm{E}\left[\frac{\exp\{y_{i,i,\cdot}(\mu + \psi\xi)\}}{\{1 + \exp(\mu + \psi\xi)\}^2}\right],$$

where $y_{i,i,\cdot} = y_{i,i,1} + y_{i,i,2}$, $\psi = \sqrt{\sigma^2 + \tau^2}$, and $\xi \sim N(0,1)$. Define $g_\theta(s) = \mathrm{E}\{e^{s(\mu+\psi\xi)}(1+e^{\mu+\psi\xi})^{-2}\}$, $s = 0,1,2$. Then,

$$\mathrm{E}_{\theta_0}|\log\{p_{\theta_0}(y_{i,i})\}| = \sum_{s=0}^{2} |\log\{g_{\theta_0}(s)\}|g_{\theta_0}(s),$$

which is a finite constant, hence (i) is satisfied.

For any $\theta \in \Theta \cap S_3(k)$, consider two cases. I: $\psi \geq \sqrt{k}/2$. Then, we have

$$\begin{aligned} g_\theta(1) &= \frac{1}{\sqrt{2\pi}} \int \frac{e^{\mu+\psi x}}{(1+e^{\mu+\psi x})^2} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}\psi} \int \frac{e^{\mu+u}}{(1+e^{\mu+u})^2} e^{-u^2/2\psi^2} du \\ &\leq \frac{1}{\sqrt{2\pi}\psi} \int \frac{e^{\mu+u}}{(1+e^{\mu+u})^2} du \\ &= \frac{1}{\sqrt{2\pi}\psi} \\ &\leq \sqrt{\frac{2}{\pi k}}, \end{aligned}$$

hence $\log\{p_\theta(1)\} \leq -(1/2)\{\log k + \log(\pi/2)\}$. II: $\psi < \sqrt{k}/2$. Then, we must have $k^2 \leq \mu^2 + \psi^2 < \mu^2 + k/4$, implying $|\mu| > k\sqrt{1-1/4k}$. Therefore, there are two subcases. II.1: $\mu > k\sqrt{1-1/4k}$. Then, using the moment-generating function of the standard normal distribution, we have

$$\begin{aligned} g_\theta(0) &\leq e^{-2\mu}\mathrm{E}(e^{-2\psi\xi}) \\ &= e^{-2\mu+2\psi^2} \\ &\leq \exp\left(-2k\sqrt{1-\frac{1}{4k}} + \frac{k}{2}\right) \\ &= \exp\left\{-2k\left(\sqrt{1-\frac{1}{4k}} - \frac{1}{4}\right)\right\} \\ &\leq \exp\left\{-\left(\frac{2\sqrt{3}-1}{2}\right)k\right\}, \end{aligned}$$

hence $\log\{p_\theta(0)\} \leq -\{(2\sqrt{3}-1)/2\}k$. II.2: $\mu < -k\sqrt{1-1/4k}$. By a similar argument, it can be shown that $\log\{p_\theta(2)\}$ has the same upper bound. It is easy to show that the upper bound under case I is larger than the upper bound under case II for all $k \geq 1$. Therefore, (ii) is satisfied with $\gamma_k = (1/2)\{\log k + \log(\pi/2)\}$.

Furthermore, (iii) holds with $\rho = \min_{s=0,1,2} g_{\theta_0}(s) > 0$. (iv) holds because $\mathcal{T}_N = \{0, 1, 2\}$, $m_N = m_1 \to \infty$ (assumed in Section 4 of Jiang2012); and

$$\begin{aligned} & c_N^3 \sum_{k=K}^{\infty} k^3 b_k^3 e^{-\delta m_1 \gamma_k} \\ = & (mn)^{9/2} \exp\left\{-\frac{\delta}{2}\log\left(\frac{\pi}{2}\right) m_1\right\} \sum_{k=K}^{\infty} k^{-(\delta m_1 - 9)/2}. \end{aligned} \tag{31}$$

It is easy to show that, for $K = 4$, the right side of (31) is bounded by

$$2\exp\left[-\frac{\delta}{2}\left\{1 + \log\left(\frac{\pi}{2}\right)\right\} m_1 - \log(\delta m_1 - 11) + \frac{9}{2}(\log m + \log n) + \frac{11}{2}\right],$$

which goes to zero for any $\delta > 0$, provided that $m_1^{-1}\log(m \vee n) \to 0$.

7. Derivation of (21), (22) of Jiang2012. For any $x \in R^d$, we have, using the expression above (21) of Jiang2012,

$$\begin{aligned} x' I_{\mathrm{f}}(\theta) x &= x' I_{\mathrm{f},1}(\theta) x - x' I_{\mathrm{f},2}(\theta) x \\ &= \mathrm{E}_\theta\left(\mathrm{E}_\theta\left[\left\{x'\frac{\partial}{\partial\theta}\log p_\theta(y)\right\}^2 \middle| y_{[1]}\right]\right) \\ &\quad -\mathrm{E}_\theta\left[x'\mathrm{E}_\theta\left\{\frac{1}{p_\theta(y)}\frac{\partial^2}{\partial\theta\partial\theta'}p_\theta(y)\middle| y_{[1]}\right\} x\right] \end{aligned} \tag{32}$$

The conditional expectation inside the second term on the right side of (32)

$$\begin{aligned} &= \int \frac{1}{p_\theta(y_{[1]}, y_{[2]})}\left\{\frac{\partial^2}{\partial\theta\partial\theta'}p_\theta(y_{[1]}, y_{[2]})\right\}\frac{p_\theta(y_{[1]}, y_{[2]})}{p_\theta(y_{[1]})}\nu(dy_{[2]}) \\ &= \frac{1}{p_\theta(y_{[1]})}\frac{\partial^2}{\partial\theta\partial\theta'}\int p_\theta(y_{[1]}, y_{[2]})\nu(dy_{[2]}) \\ &= \frac{1}{p_\theta(y_{[1]})}\frac{\partial^2}{\partial\theta\partial\theta'}p_\theta(y_{[1]}). \end{aligned}$$

Thus, continuing with (32), we have

$$\begin{aligned} x' I_{\mathrm{f}}(\theta) x &= \mathrm{E}_\theta\left[\mathrm{var}_\theta\left\{x'\frac{\partial}{\partial\theta}\log p_\theta(y)\middle| y_{[1]}\right\}\right] \\ &\quad +\mathrm{E}\left(\left[x'\mathrm{E}_\theta\left\{\frac{\partial}{\partial\theta}\log p_\theta(y)\middle| y_{[1]}\right\}\right]^2\right) - x' I_{\mathrm{s},2}(\theta) x. \end{aligned} \tag{33}$$

Once again, the conditional expectation inside the second term on the right side of (33) is equal to

$$\begin{aligned}
& \int \frac{1}{p_\theta(y_{[1]}, y_{[2]})} \left\{ \frac{\partial}{\partial\theta} p_\theta(y_{[1]}, y_{[2]}) \right\} \frac{p_\theta(y_{[1]}, y_{[2]})}{p_\theta(y_{[1]})} \nu(dy_2) \\
= & \frac{1}{p_\theta(y_{[1]})} \frac{\partial}{\partial\theta} \int p_\theta(y_{[1]}, y_{[2]}) \nu(dy_2) \\
= & \frac{1}{p_\theta(y_{[1]})} \frac{\partial}{\partial\theta} p_\theta(y_{[1]}).
\end{aligned}$$

Therefore, going back to (33), we have

$$\begin{aligned}
x' I_{\mathrm{f}}(\theta) x &= x' \mathrm{E}_\theta \left[\mathrm{Var}_\theta \left\{ \frac{\partial}{\partial\theta} \log p_\theta(y) \middle| y_{[1]} \right\} \right] x \\
& \quad + x' I_{\mathrm{s},1}(\theta) x - x' I_{\mathrm{s},2}(\theta) x \\
&= x' \mathrm{E}_\theta \left[\mathrm{Var}_\theta \left\{ \frac{\partial}{\partial\theta} \log p_\theta(y) \middle| y_{[1]} \right\} \right] x + x' I_{\mathrm{s}}(\theta) x.
\end{aligned}$$

Because x is arbitrary, (22) of Jiang2012 must hold, which is a nonnegative definite matrix.

8. Consistency of the DC MLE. Let Y_N denote the data vector under the sample size N. Lele *et al.* (2010, Corollary of Lemma A.2 in the Appendix) shows that, under regularity conditions, we have $\theta^{(1)} \xrightarrow{\mathrm{d}} \delta_{\hat{\theta}}$, as $K \to \infty$, conditional on Y_N, where $\theta^{(1)}$ has the posterior distribution (24) of Jiang2012, and $\delta_{\hat{\theta}}$ is the degenerate distribution at $\hat{\theta}$, the MLE. Then (e.g., Jiang 2010, p. 45), we have $\limsup_{K\to\infty} \mathrm{P}\{\theta^{(1)} \in C | Y_N\} \le \mathrm{P}(\xi \in C)$ for every closed set C, where $\xi \sim \delta_{\hat{\theta}}$. Thus, for any $\epsilon > 0$, by considering $C = \{\theta \in R^d : |\theta - \hat{\theta}| \ge \epsilon\}$, we have $\mathrm{P}\{\theta^{(1)} \in C | Y_N\} = \mathrm{P}\{|\theta^{(1)} - \hat{\theta}| \ge \epsilon | Y_N\}$ and $\mathrm{P}(\xi \in C) = \mathrm{P}(|\xi - \hat{\theta}| \ge \epsilon) = 0$, implying $\mathrm{P}\{|\theta^{(1)} - \hat{\theta}| \ge \epsilon | Y_N\} \to 0$, as $K \to \infty$. It then follows, by the dominated convergence theorem, that $\mathrm{P}\{|\theta^{(1)} - \hat{\theta}| \ge \epsilon\} \to 0$, as $K \to \infty$, for any fixed ϵ, n. In particular, there is $K(B, n)$ such that $\mathrm{P}\{|\theta^{(1)} - \hat{\theta}| \ge n^{-1}\} \le (Bn)^{-1}$, if $K \ge K(B, n)$.

On the other hand, note that $|\bar{\theta}^{(\cdot)} - \hat{\theta}| \ge n^{-1}$ implies that $|\theta^{(b)} - \hat{\theta}| \ge n^{-1}$ for some $1 \le b \le B$. Therefore, we have $\mathrm{P}\{|\bar{\theta}^{(\cdot)} - \hat{\theta}| \ge n^{-1}\} \le \sum_{b=1}^B \mathrm{P}\{|\theta^{(b)} - \hat{\theta}| \ge n^{-1}\} = B\mathrm{P}\{|\theta^{(1)} - \hat{\theta}| \ge n^{-1}\} \le n^{-1}$, if $K \ge K(B, n)$. Thus, for any ϵ, we have, for any $n \ge 2/\epsilon$, $\mathrm{P}\{|\bar{\theta}^{(\cdot)} - \theta_0| \ge \epsilon\} \le \mathrm{P}\{|\bar{\theta}^{(\cdot)} - \hat{\theta}| \ge n^{-1}\} + \mathrm{P}(|\hat{\theta} - \theta_0| \ge \epsilon/2) \le n^{-1} + \mathrm{P}(|\hat{\theta} - \theta_0| \ge \epsilon/2)$, if $n \ge 2/\epsilon$ and $K \ge K(n, B)$. The result then follows by the consistency (as $n \to \infty$) of $\hat{\theta}$.