Identifying the Most Appropriate Order
for Categorical Responses

Tianmeng Wang and Jie Yang

University of Illinois at Chicago

Supplementary Materials

Proof. of Theorem 2.1: The log-likelihood of the model at $\theta_1$ with permuted responses $Y_i^{\sigma_1}$'s is

$$l_1(\theta_1) = \sum_{i=1}^{m} \log(n_i!) - \sum_{i=1}^{m} \sum_{j=1}^{J} \log(Y_{i\sigma_1(j)}!) + \sum_{i=1}^{m} \sum_{j=1}^{J} Y_{i\sigma_1(j)} \log \pi_{ij}(\theta_1)$$

while the log-likelihood at $\theta_2$ with $Y_i^{\sigma_2}$'s is

$$l_2(\theta_2) = \sum_{i=1}^{m} \log(n_i!) - \sum_{i=1}^{m} \sum_{j=1}^{J} \log(Y_{i\sigma_2(j)}!) + \sum_{i=1}^{m} \sum_{j=1}^{J} Y_{i\sigma_2(j)} \log \pi_{ij}(\theta_2)$$

Since $Y_i^{\sigma_1}$ and $Y_i^{\sigma_2}$ are different only at the order of individual terms,

$$\sum_{j=1}^{J} \log(Y_{i\sigma_1(j)}!) = \sum_{j=1}^{J} \log(Y_{i\sigma_2(j)}!)$$
On the other hand, \( \pi_{ij}(\theta_2) = \pi_{i\sigma_2(\sigma_1^{-1}(j))}(\theta_1) \) implies that
\[
\sum_{j=1}^{J} Y_{i\sigma_2(j)} \log \pi_{ij}(\theta_2) = \sum_{j=1}^{J} Y_{i\sigma_2(j)} \log \pi_{i\sigma_2(\sigma_1^{-1}(j))}(\theta_1)
\]
\[
= \sum_{j=1}^{J} Y_{ij} \log \pi_{i\sigma_1^{-1}(j)}(\theta_1)
\]
\[
= \sum_{j=1}^{J} Y_{i\sigma_1(j)} \log \pi_{ij}(\theta_1)
\]

Therefore, \( l_1(\theta_1) = l_2(\theta_2) \), which implies \( \max_{\theta_1} l_1(\theta_1) \leq \max_{\theta_2} l_2(\theta_2) \). Similarly, \( \max_{\theta_2} l_2(\theta_2) \leq \max_{\theta_1} l_1(\theta_1) \). Thus \( \max_{\theta_1} l_1(\theta_1) = \max_{\theta_2} l_2(\theta_2) \).

Given that (2.8) is true for \( \sigma_1 \) and \( \sigma_2 \), for any permutation \( \sigma \in \mathcal{P} \),
\[
\pi_{i\sigma_1^{-1}(\sigma^{-1}(j))}(\theta_1) = \pi_{i\sigma_2^{-1}(\sigma^{-1}(j))}(\theta_2)
\]
for all \( i \) and \( j \). That is, \( \pi_{i(\sigma_1^{-1}(\sigma^{-1}(j)))}(\theta_1) = \pi_{i(\sigma_2^{-1}(\sigma^{-1}(j)))}(\theta_2) \) for all \( i \) and \( j \).

Following the same proof above, we have \( \sigma_1 \sim \sigma_2 \).

**Proof.** of Theorem 2.2: We first show that \( \sigma_1 = \text{id} \), the identity permutation, and any permutation \( \sigma_2 \) satisfying \( \sigma_2(J) = J \) are equivalent. Actually, given any \( \theta_1 = (\beta_1^T, \ldots, \beta_{J-1}^T, \zeta^T)^T \) for \( \sigma_1 = \text{id} \), we let \( \theta_2 = (\beta_{\sigma_2(1)}^T, \ldots, \beta_{\sigma_2(J-1)}^T, \zeta^T)^T \) for \( \sigma_2 \). Then \( \eta_{ij}(\theta_2) = h_j^T(x_i) \beta_{\sigma_2(j)} + h_c^T(x_i) \zeta = h_{\sigma_2(j)}(x_i) \beta_{\sigma_2(j)} + h_c^T(x_i) \zeta = \eta_{i\sigma_2(j)}(\theta_1) \) for all \( i = 1, \ldots, m \) and \( j = 1, \ldots, J-1 \). According to (2.6) and (2.7), \( \pi_{ij}(\theta_2) = \pi_{i\sigma_2(j)}(\theta_1) \) for all \( i = 1, \ldots, m \) and \( j = 1, \ldots, J \). Then \( \text{id} \sim \sigma_2 \) is obtained by Theorem 2.1.
SUPPLEMENTARY MATERIALS

For general $\sigma_1$ and $\sigma_2$ satisfying $\sigma_1(J) = \sigma_2(J) = J$, id $\sim \sigma_1^{-1}\sigma_2$ implies $\sigma_1 \sim \sigma_2$ by Theorem 2.1.

Proof. of Theorem 2.3: Given $\theta_1 = (\beta_1^T, \beta_2^T, \ldots, \beta_{J-1}^T, \zeta^T)^T$ with $\sigma_1$, we let $\theta_2 = (-\beta_{J-1}^T, -\beta_{J-2}^T, \ldots, -\beta_1^T, -\zeta^T)^T$ for $\sigma_2$. Then $\eta_{ij}(\theta_2) = -h_j^T(x_i)\beta_{J-j} - h_c^T(x_i)\zeta = -\eta_{i,J-j}(\theta_1)$ and thus $\rho_{ij}(\theta_2) = 1 - \rho_{i,J-j}(\theta_1)$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, J - 1$. It can be verified that $\pi_{ij}(\theta_2) = \pi_{i,J+1-j}(\theta_1)$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, J$ according to (2.6) and (2.7). Then for all $i = 1, \ldots, m$ and $j = 1, \ldots, J$,

$$
\pi_{i\sigma_2^{-1}(j)}(\theta_2) = \pi_{i,J+1-\sigma_2^{-1}(j)}(\theta_1) = \pi_{i\sigma_2^{-1}(\sigma_2^{-1}(J+1-\sigma_2^{-1}(j))))(\theta_1) = \pi_{i\sigma_2^{-1}(\sigma_2^{-1}(j))))(\theta_1) = \pi_{i\sigma_2^{-1}(j)}(\theta_1)
$$

That is, (2.8) holds given $\theta_1$. Since it is one-to-one from $\theta_1$ to $\theta_2$, (2.8) holds given $\theta_2$ as well. According to Theorem 2.1, $\sigma_1 \sim \sigma_2$.

Proof. of Theorem 2.4: Similar as the proof of Theorem 2.3, for ppo models satisfying $h_1(x_i) = \cdots = h_{J-1}(x_i)$, given $\theta_1 = (\beta_1^T, \beta_2^T, \ldots, \beta_{J-1}^T, \zeta^T)^T$ with $\sigma_1$, we let $\theta_2 = (-\beta_{J-1}^T, -\beta_{J-2}^T, \ldots, -\beta_1^T, -\zeta^T)^T$ for $\sigma_2$. Then $\eta_{ij}(\theta_2) = -h_j^T(x_i)\beta_{J-j} - h_c^T(x_i)\zeta = -\eta_{i,J-j}(\theta_1)$ and thus $\rho_{ij}(\theta_2) = 1 - \rho_{i,J-j}(\theta_1)$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, J - 1$. It
can be verified that for \( j = 1, \ldots, J - 1, \)
\[
\prod_{l=j}^{J-1} \frac{\rho_{il}(\theta_2)}{1 - \rho_{il}(\theta_2)} = \prod_{l=j}^{J-1} \frac{1 - \rho_{i,j-l}(\theta_1)}{\rho_{i,j-l}(\theta_1)} = \prod_{l=1}^{J-j} \frac{1 - \rho_{il}(\theta_1)}{\rho_{il}(\theta_1)}
\]
implies \( \pi_{ij}(\theta_2) = \pi_{i,J+1-j}(\theta_1) \) for all \( i = 1, \ldots, m \) and \( j = 1, \ldots, J \) due to (2.6) and (2.7). Then \( \pi_{i\sigma^{-1}(j)}(\theta_2) = \pi_{i\sigma^{-1}(j)}(\theta_1) \) similarly as in the proof of Theorem 2.3, which leads to \( \sigma_1 \sim \sigma_2 \) based on Theorem 2.1.

**Proof.** of Theorem 2.5: According to Theorem 2.1, we only need to show that (2.8) holds for \( \sigma_1 = \text{id} \) and an arbitrary permutation \( \sigma_2 \in \mathcal{P} \).

**Case one:** \( \sigma_2(J) = J \). In this case, for any \( \theta_1 = (\beta_1^T, \ldots, \beta_{J-1}^T)^T \), we let \( \theta_2 = (\beta_{\sigma_2(1)}^T, \ldots, \beta_{\sigma_2(J-1)}^T)^T \). Then \( \eta_{ij}(\theta_2) = h_j^T(x_i)\beta_{\sigma_2(j)} = h_{\sigma_2(j)}(x_i)\beta_{\sigma_2(j)} = \eta_{i\sigma_2(j)}(\theta_1) \), which leads to \( \rho_{ij}(\theta_2) = \rho_{i\sigma_2(j)}(\theta_1) \) for all \( i = 1, \ldots, m \) and \( j = 1, \ldots, J - 1 \). According to (2.6) in Lemma ??, \( \pi_{ij}(\theta_1) = \pi_{i\sigma_2^{-1}(j)}(\theta_2) \), which is (2.8) in this case. Since the correspondence between \( \theta_1 \) and \( \theta_2 \) is one-to-one, then (2.8) holds for given \( \theta_2 \) as well.

**Case two:** \( \sigma_2(J) \neq J \). Given \( \theta_1 = (\beta_1^T, \ldots, \beta_{J-1}^T)^T \), we let \( \theta_2 = (\beta_{21}^T, \ldots, \beta_{2,J-1}^T)^T \) such that for \( j = 1, \ldots, J - 1, \)
\[
\beta_{2j} = \begin{cases} 
\beta_{\sigma_2(j)} - \beta_{\sigma_2(J)} & \text{if } j \neq \sigma_2^{-1}(J) \\
-\beta_{\sigma_2(J)} & \text{if } j = \sigma_2^{-1}(J)
\end{cases}
\]
Then for $i = 1, \ldots, m$ and $j = 1, \ldots, J - 1$,

$$
\eta_{ij}(\theta_2) = h_j^T(x_i) \beta_{2j} = \begin{cases} 
\eta_{\sigma_2(j)}(\theta_1) - \eta_{\sigma_2(j)}(\theta_1) & \text{if } j \neq \sigma_2^{-1}(J) \\
-\eta_{\sigma_2(j)}(\theta_1) & \text{if } j = \sigma_2^{-1}(J)
\end{cases}
$$

It can be verified that: (i) If $\sigma_2^{-1}(j) \neq J$ and $j \neq J$, then $\pi_{i\sigma_2^{-1}(j)}(\theta_2) = \pi_{ij}(\theta_1)$ according to (2.6); (ii) $\pi_{i\sigma_2^{-1}(J)}(\theta_2) = \pi_{iJ}(\theta_1)$ according to (2.6); and (iii) if $\sigma_2^{-1}(j) = J$, then $\pi_{i\sigma_2^{-1}(j)}(\theta_2) = \pi_{ij}(\theta_1)$ according to (2.7). Thus (2.8) holds given $\theta_1$. Given $\sigma_2$, the correspondence between $\theta_1$ and $\theta_2$ is one-to-one, then (2.8) holds given $\theta_2$ as well. Thus $\text{id} \sim \sigma_2$ according to Theorem 2.1.

For general $\sigma_1$ and $\sigma_2$, $\text{id} \sim \sigma_1^{-1}\sigma_2$ implies $\sigma_1 \sim \sigma_2$ according to Theorem 2.1.

Proof. of Theorem 2.6: Similar as the proof of Theorem 2.5, we first show that (2.8) holds for $\sigma_1 = \text{id}$ and an arbitrary permutation $\sigma_2$.

For any $\theta_1 = (\beta_{11}^T, \ldots, \beta_{J-1}^T)^T$ with $\sigma_1 = \text{id}$, we let $\theta_2 = (\beta_{21}^T, \ldots, \beta_{2,J-1}^T)^T$ for $\sigma_2$, where

$$
\beta_{2j} = \begin{cases} 
\sum_{l=\sigma_2(j)}^{\sigma_2(j+1)-1} \beta_l & \text{if } \sigma_2(j) < \sigma_2(j+1) \\
-\sum_{l=\sigma_2(j+1)}^{\sigma_2(j)-1} \beta_l & \text{if } \sigma_2(j) > \sigma_2(j+1)
\end{cases} \tag{S.1}
$$

For $j = 1, \ldots, J-1$, $\eta_{ij}(\theta_2) = h_j^T(x_i) \beta_{2j} = \sum_{l=\sigma_2(j)}^{\sigma_2(j+1)-1} h_l^T(x_i) \beta_l = \sum_{l=\sigma_2(j)}^{\sigma_2(j+1)-1} \eta_{il}(\theta_1)$ if $\sigma_2(j) < \sigma_2(j+1)$; and $\eta_{ij}(\theta_2) = -\sum_{l=\sigma_2(j+1)}^{\sigma_2(j)-1} \eta_{il}(\theta_1)$ if $\sigma_2(j) > \sigma_2(j+1)$.
\[ \sigma_2(j + 1). \] It can be verified that (S.1) implies for any \( 1 \leq j < k \leq J - 1, \]
\[ \sum_{l=j}^{k} \beta_{2j} = \begin{cases} 
\sum_{l=\sigma_2(j)}^{\sigma_2(k+1)-1} \beta_l & \text{if } \sigma_2(j) < \sigma_2(k + 1) \\
- \sum_{l=\sigma_2(j+1)}^{\sigma_2(k) - 1} \beta_l & \text{if } \sigma_2(j) > \sigma_2(k + 1) 
\end{cases} \tag{S.2} \]

Then for \( j = 1, \ldots, J - 1, \)
\[ \prod_{l=j}^{J-1} \frac{\rho_{il}(\theta_2)}{1 - \rho_{il}(\theta_2)} = \exp \left\{ \sum_{l=j}^{J-1} \eta_{il}(\theta_2) \right\} = \begin{cases} 
\prod_{l=\sigma_2(j)}^{\sigma_2(J)-1} \frac{\rho_{il}(\theta_1)}{1 - \rho_{il}(\theta_1)} & \text{if } \sigma_2(j) < \sigma_2(J) \\
\prod_{l=\sigma_2(J)}^{\sigma_2(j)-1} \frac{\rho_{il}(\theta_1)}{1 - \rho_{il}(\theta_1)} & \text{if } \sigma_2(j) > \sigma_2(J) 
\end{cases} \]

According to (2.6) and (2.7), it can be verified that \( \pi_{ij}(\theta_2) = \pi_{i\sigma_2(j)}(\theta_1) \) for all \( i = 1, \ldots, m \) and \( j = 1, \ldots, J. \) That is, (2.8) holds given \( \theta_1. \)

On the other hand, (S.2) implies an inverse transformation from \( \theta_2 \) to \( \theta_1 \)
\[ \beta_j = \sum_{l=\sigma_2^{-1}(j)}^{\sigma_2^{-1}(j+1)} \beta_{2l} \tag{S.3} \]
with \( j = 1, \ldots, J - 1. \) That is, it is one-to-one from \( \theta_1 \) to \( \theta_2. \) According to Theorem 2.1, \( \text{id} \sim \sigma_2. \)

Similar as in the proof of Theorem 2.5, we have \( \sigma_1 \sim \sigma_2 \) for any two permutations \( \sigma_1 \) and \( \sigma_2. \)

**Proof.** of Theorem 2.7: We first verify condition (2.8) for \( \sigma_1 = \text{id} \) and \( \sigma_2 = (J - 1, J). \) In this case, given \( \theta_1 = (\beta_{1}^T, \ldots, \beta_{J-2}^T, \beta_{J-1}^T)^T \) for \( \sigma_1, \) we let \( \theta_2 = (\beta_1^T, \ldots, \beta_{J-2}^T, -\beta_{J-1}^T)^T \) for \( \sigma_2. \) Then \( \eta_{ij}(\theta_2) = h_j^T(x_i) \beta_j = \eta_{ij}(\theta_1) \)
for \( j = 1, \ldots, J - 2, \) and \( \eta_{i,J-1}(\theta_2) = -h_{J-1}^T(x_i) \beta_{J-1} = -\eta_{i,J-1}(\theta_1). \) We
further obtain $\rho_{ij}(\theta_2) = \rho_{ij}(\theta_1)$ for $j = 1, \ldots, J - 2$ and $\rho_{i,J-1}(\theta_2) = 1 - \rho_{i,J-1}(\theta_1)$. According to (2.6) and (2.7), we obtain $\pi_{ij}(\theta_2) = \pi_{ij}(\theta_1)$ for $j = 1, \ldots, J - 2$; $\pi_{i,J-1}(\theta_2) = \pi_{i,j}(\theta_1)$; and $\pi_{i,J}(\theta_2) = \pi_{i,J-1}(\theta_1)$. That is, (2.8) holds given $\theta_1$, which also holds given $\theta_2$ since it is one-to-one from $\theta_1$ to $\theta_2$. According to Theorem 2.1, $\text{id} \sim (J - 1, J)$ and thus $\sigma_1 \sim \sigma_1(J - 1, J)$ for any permutation $\sigma_1$.

Proof. of Lemma 2: It is well known that for each $i = 1, \ldots, m$, $\left(\frac{Y_{i1}}{N_i}, \ldots, \frac{Y_{iJ}}{N_i}\right)$ maximizes $\sum_{j=1}^J Y_{ij} \log \pi_{ij}$ as a function of $(\pi_{i1}, \ldots, \pi_{iJ})$ under the constraints $\sum_{j=1}^J \pi_{ij} = 1$ and $\pi_{ij} \geq 0$, $j = 1, \ldots, J$ (see, for example, Section 3.5.6 of Johnson et al. [1997]). If $\hat{\theta} \in \Theta$ and $\hat{\sigma} \in \mathcal{P}$ satisfy $\pi_{i\hat{\sigma}^{-1}(j)}(\hat{\theta}) = \frac{Y_{ij}}{N_i}$ for all $i$ and $j$, then $\{\pi_{i\hat{\sigma}^{-1}(j)}(\hat{\theta})\}_{ij}$ maximizes $\sum_{i=1}^m \sum_{j=1}^J Y_{ij} \log \pi_{i\hat{\sigma}^{-1}(j)}(\hat{\theta})$, which implies $(\hat{\theta}, \hat{\sigma})$ maximizes $l_N(\theta, \sigma)$ and thus $l(\theta, \sigma)$.

Proof. of Lemma 3: According to the strong law of large numbers (see, for example, Chapter 4 in Ferguson [1996]), $\frac{N_i}{N} = N^{-1} \sum_{l=1}^N 1_{\{X_l = x_i\}} \rightarrow E(1_{\{X_l = x_i\}}) = \frac{m_i}{n}$ almost surely, as $N \rightarrow \infty$, for each $i = 1, \ldots, m$. Since $n_0 \geq 1$, it can be verified that $\min\{N_1, \ldots, N_m\} \rightarrow \infty$ almost surely, as $N \rightarrow \infty$. Similarly, we have $\frac{Y_{ij}}{N_i} \rightarrow \pi_{i\hat{\sigma}^{-1}(j)}(\theta_0)$ almost surely, as $N_i \rightarrow \infty$. 

SUPPLEMENTARY MATERIALS
for each \(i = 1, \ldots, m\) and \(j = 1, \ldots, J\). Then as \(N\) goes to infinity,

\[
N^{-1} l_N(\theta_0, \sigma_0) = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{J} Y_{ij} \log \pi_{i \sigma_0^{-1}(j)}(\theta_0)
\]

\[
= \sum_{i=1}^{m} \sum_{j=1}^{J} \frac{N_i}{N} Y_{ij} \log \pi_{i \sigma_0^{-1}(j)}(\theta_0)
\]

\[
\xrightarrow{a.s.} \sum_{i=1}^{m} \sum_{j=1}^{J} \frac{n_i}{N} \pi_{i \sigma_0^{-1}(j)}(\theta_0) \log \pi_{i \sigma_0^{-1}(j)}(\theta_0)
\]

\[
= \sum_{i=1}^{m} \sum_{j=1}^{J} \frac{n_i}{N} \pi_{ij}(\theta_0) \log \pi_{ij}(\theta_0) = l_0 < 0
\]

\[\square\]

Proof. of Theorem 3.1: First we claim that for large enough \(N\), all \(i = 1, \ldots, m\) and \(j = 1, \ldots, J\), 

\[0 > \log \pi_{i \hat{\sigma}_N^{-1}(j)}(\hat{\theta}_N) \geq \frac{2n(l_0 - 1)}{m_0 \pi_0}
\]

which is a finite constant. Actually, since \((\hat{\theta}_N, \hat{\sigma}_N)\) is an MLE, we have

\[
N^{-1} l_N(\hat{\theta}_N, \hat{\sigma}_N) \geq N^{-1} l_N(\theta_0, \sigma_0)
\]

for each \(N\). According to Lemma 3, \(N^{-1} l_N(\theta_0, \sigma_0) \to l_0\) almost surely, then \(N^{-1} l_N(\hat{\theta}_N, \hat{\sigma}_N) > l_0 - 1\) for large enough \(N\) almost surely. On the other hand, since \(\log \pi_{i \hat{\sigma}_N^{-1}(j)}(\hat{\theta}_N) < 0\) for all \(i\) and \(j\), then

\[
N^{-1} l_N(\hat{\theta}_N, \hat{\sigma}_N) = \sum_{i=1}^{m} \sum_{j=1}^{J} \frac{N_i}{N} \cdot \frac{Y_{ij}}{N} \log \pi_{i \hat{\sigma}_N^{-1}(j)}(\hat{\theta}_N) < \sum_{i=1}^{m} \sum_{j=1}^{J} \frac{N_i}{N} \cdot \frac{Y_{ij}}{N} \log \pi_{i \hat{\sigma}_N^{-1}(j)}(\hat{\theta}_N)
\]

for each \(i\) and \(j\). Since \(\frac{N_i}{N} \to \frac{m_i}{n}\) almost surely and \(\frac{Y_{ij}}{N} \to \pi_{i \sigma_0^{-1}(j)}(\theta_0)\) almost surely, then \(\sum_{i=1}^{m} \sum_{j=1}^{J} \frac{n_i}{N} \log \pi_{i \hat{\sigma}_N^{-1}(j)}(\hat{\theta}_N) < \frac{1}{2} \cdot \frac{n_i}{n} \pi_{i \sigma_0^{-1}(j)}(\theta_0) \log \pi_{i \hat{\sigma}_N^{-1}(j)}(\hat{\theta}_N)\) for large enough \(N\) almost surely. Then we have

\[
0 > \frac{1}{2} \cdot \frac{n_i}{n} \pi_{i \sigma_0^{-1}(j)}(\theta_0) \log \pi_{i \hat{\sigma}_N^{-1}(j)}(\hat{\theta}_N) > l_0 - 1
\]
almost surely for large enough $N$ and each $i$ and $j$. Since $l_0 - 1 < 0$, we further have almost surely for large enough $N$,

$$0 > \log \pi_{\hat{\sigma}_N^{-1}(j)}(\hat{\theta}_N) > \frac{l_0 - 1}{2} \cdot \frac{n_i}{n} \pi_{\hat{\sigma}_N^{-1}(j)}(\theta_0) \geq \frac{l_0 - 1}{2} \cdot \frac{n_i}{n} \pi_0 = \frac{2n(l_0 - 1)}{n_0 \pi_0}$$

Now we are ready to check the asymptotic difference between $N^{-1}l_N(\hat{\theta}_N, \hat{\sigma}_N)$ and $N^{-1}l_N(\theta_0, \sigma_0)$. According to Lemma 2 and its proof, $(\theta_0, \sigma_0)$ maximizes

$$\sum_{i=1}^m \sum_{j=1}^J \frac{N_i}{N} \pi_{\hat{\sigma}_N^{-1}(j)}(\theta_0) \log \pi_{\hat{\sigma}_N^{-1}(j)}(\theta).$$

Then

$$0 \leq N^{-1}[l_N(\hat{\theta}_N, \hat{\sigma}_N) - l_N(\theta_0, \sigma_0)]$$

$$= \sum_{i=1}^m \sum_{j=1}^J \frac{N_i}{N} \pi_{\hat{\sigma}_N^{-1}(j)}(\theta_0) \log \pi_{\hat{\sigma}_N^{-1}(j)}(\hat{\theta}_N)$$

$$- \sum_{i=1}^m \sum_{j=1}^J \frac{N_i}{N} \pi_{\hat{\sigma}_N^{-1}(j)}(\theta_0) \log \pi_{\hat{\sigma}_N^{-1}(j)}(\theta_0)$$

$$+ \sum_{i=1}^m \sum_{j=1}^J \frac{N_i}{N} \left[ \frac{Y_{ij}}{N_i} - \pi_{\hat{\sigma}_N^{-1}(j)}(\theta_0) \right] \cdot \left[ \log \pi_{\hat{\sigma}_N^{-1}(j)}(\hat{\theta}_N) - \log \pi_{\hat{\sigma}_N^{-1}(j)}(\theta_0) \right]$$

$$\leq \sum_{i=1}^m \sum_{j=1}^J \frac{N_i}{N} \left[ \frac{Y_{ij}}{N_i} - \pi_{\hat{\sigma}_N^{-1}(j)}(\theta_0) \right] \cdot \left[ \log \pi_{\hat{\sigma}_N^{-1}(j)}(\hat{\theta}_N) - \log \pi_{\hat{\sigma}_N^{-1}(j)}(\theta_0) \right]$$

Then for large enough $N$, we have almost surely

$$\frac{1}{N} |l_N(\hat{\theta}_N, \hat{\sigma}_N) - l_N(\theta_0, \sigma_0)| \leq \sum_{i=1}^m \sum_{j=1}^J \left| \frac{Y_{ij}}{N_i} - \pi_{\hat{\sigma}_N^{-1}(j)}(\theta_0) \right| \cdot \left[ \frac{-2n(l_0 - 1)}{n_0 \pi_0} - \log \pi_0 \right]$$

Since $\frac{Y_{ij}}{N_i} \to \pi_{\hat{\sigma}_N^{-1}(j)}(\theta_0)$ almost surely for each $i$ and $j$, then $N^{-1}|l_N(\hat{\theta}_N, \hat{\sigma}_N) - l_N(\theta_0, \sigma_0)| \to 0$ almost surely as $N$ goes to infinity. The rest parts of the theorem are straightforward. $\square$
Proof. of Corollary 1: Since $\text{AIC} - \text{AIC}(\theta_0, \sigma_0) = \text{BIC} - \text{BIC}(\theta_0, \sigma_0) = -2l(\hat{\theta}_N, \hat{\sigma}_N) + 2l(\theta_0, \sigma_0) = -2l_N(\hat{\theta}_N, \hat{\sigma}_N) + 2l_N(\theta_0, \sigma_0)$, the conclusion follows directly by Theorem 3.1.

References
