SUPPLEMENTARY TO "ON THE IDENTIFIABILITY OF DIAGNOSTIC CLASSIFICATION MODELS"

In this supplementary, we provide the technical proofs for six theorems which are stated in the main paper. We also give the computation scheme for the proposed non-parametric Bayes method.

Technical Proofs

Lemma 1. (Kruskal (1977)) Suppose $A, B, C, \bar{A}, \bar{B}, \bar{C}$ are six matrices with L columns. There exist integers I_0, J_0, and K_0 such that $I_0 + J_0 + K_0 \geq 2L + 2$. In addition, every I_0 columns of A are linearly independent, every J_0 columns of B are linearly independent, and every K_0 columns of C are linearly independent. Define a triple product to be a three-way array $[A, B, C] = (d_{ijk})$ where $d_{ijk} = \sum_{r=1}^{L} a_{ir} b_{ir} c_{kr}$. Suppose that the following two triple products are equal $[A, B, C] = [\bar{A}, \bar{B}, \bar{C}]$. Then, there exists a column permutation matrix P, we have $\bar{A} = AP\Lambda, \bar{B} = BPM, \bar{C} = CPN$, where Λ, M, N are diagonal matrices such that $\Lambda M N =$ identity. Column permutation matrix is a matrix acts on the righthand side of another matrix and permutes the columns of that matrix.

In the following notation, we use n to denote the number of respondents instead of R to make proof easier to read.

Proof of Theorem 1. For each item i, let $\pi_{i\boldsymbol{\alpha}} = P(X^i = 1|\boldsymbol{\alpha})$. that takes two possible values. Let p_{i-} or p_{i+} be these two values. According to Lemma 1, it is sufficient to show that the T-matrices corresponding to the three subsets of items T_{I_1}, T_{I_2}, and T_{I_3} are all of full column rank.

Suppose that there are n_l items in $I_l (l = 1, 2, 3)$. For each item $l \in I_l$, define

$$\mathbf{P}_i = \begin{pmatrix} p_{i-} & p_{i+} \\ 1 - p_{i-} & 1 - p_{i+} \end{pmatrix}.$$

We further define

$$\mathcal{P}_l = \bigotimes_{i \in I_l} \mathbf{P}_i$$

which is a 2^{n_l} by 2^{n_l} matrix. Because $p_{i-} \neq p_{i+}$, each $\mathbf{P}_i$ is a full-rank matrix and is of rank 2. Thus, $\mathcal{P}_l$ is rank 2^{n_i} matrix and is a full-rank matrix. Each column of T_{I_l} is precisely one of the column vector in $\mathcal{P}_l$. In addition, there is no identical columns in T_{I_l}, thus its columns vectors are linearly independent. Thus, T_{I_l} is of full column rank.

We construct three groups of items $\tilde{I}_1 = I_1, \tilde{I}_2 = I_2$ and $\tilde{I}_3 = \{1, \ldots, I\} \backslash (I_1 \bigcup I_2)$. These three groups are non-overlapping and $I_3 \subset \tilde{I}_3$. Notice that $T_{\tilde{I}_1} = T_{I_1}, T_{\tilde{I}_2} = T_{I_2}$ and T_{I_3} is a submatrix of $T_{\tilde{I}_3}$. Therefore, $T_{\tilde{I}_1}, T_{\tilde{I}_2}$ and $T_{\tilde{I}_3}$ are all full column rank. We define

$$W = [T_{\tilde{I}_1}\Lambda, T_{\tilde{I}_2}, T_{\tilde{I}_3}],$$

where Λ is a C by C diagonal matrix with $v_{\boldsymbol{\alpha}}$ being its $\boldsymbol{\alpha}$-th element. It is not hard to see that every entry of array W corresponds to a probability $P(X^1 = x^1, \ldots, X^I = x^I)$.

Suppose that there is another decomposition of W say $W = [T'_{\tilde{I}_1}\Lambda', T'_{\tilde{I}_2}, T'_{\tilde{I}_3}]$. Notice that each $T_{\tilde{I}_i}$ has rank C and $C + C + C \geq 2C + 2$ provided $C \geq 2$. Then we apply Lemma 1 and obtain that $T_{\tilde{I}_1}\Lambda = T'_{\tilde{I}_1}\Lambda' PF$, $T_{\tilde{I}_2} = T'_{\tilde{I}_2}PG$, and $T_{\tilde{I}_3} = T'_{\tilde{I}_3}PH$. Here, F, G, and H are all diagonal matrix with $FGH = I$ and I is an identity and P is a column permutation matrix. Each column of $T_{\tilde{I}_i}$ and $T'_{\tilde{I}_i}$ corresponds to a probability distribution and thus sums up to one. It means F, G and H must be identity matrix. Hence, we conclude that $T_{\tilde{I}_1}\Lambda = T'_{\tilde{I}_1}\Lambda' P$ which implies $\Lambda = P'\Lambda' P$. Then, we have $T_{\tilde{I}_1} = T'_{\tilde{I}_1}P$, $T_{\tilde{I}_2} = T'_{\tilde{I}_2}P$ and $T_{\tilde{I}_3} = T'_{\tilde{I}_3}P$. This is equivalent that the item parameters $\pi_{j\boldsymbol{\alpha}}$ and the latent class population $v_{\boldsymbol{\alpha}}$ are identifiable up to a permutation of the class label. $\square$

Proof of Corollary 1. Without loss of generality, we assume that the first, second, and third A rows of Q each form an identity matrix. The attributes are binary and each of the first $3A$ items only depends on one attribute. Thus, their item response function $\pi_{i\boldsymbol{\alpha}}$ can only take two possible values. Furthermore, we divide these $3A$ items into 3 groups $I_1 = \{1, \ldots, A\}$, $I_2 = \{A + 1, \cdots, 2A\}$, and $I_3 = \{2A + 1, \cdots, 3A\}$. It is straightforward to check that these three subsets of items satisfy condition A1 in Theorem 1. The corollary is an application of Theorem 1. $\square$

Proof of Theorem 2. Under condition $B1$, we define

$$\mathbf{P}_i = \begin{pmatrix} p_{i-}^1 & p_{i+}^1 \\ p_{i-}^2 & p_{i+}^2 \\ \vdots & \vdots \\ p_{i-}^{k_i} & p_{i+}^{k_i} \end{pmatrix},$$

whose column vectors are the two positive $P_{i\boldsymbol{\alpha}}$. For each I_l, we define

$$\mathcal{P}_l = \bigotimes_{i \in I_l} \mathbf{P}_i$$

which is a $\prod_{i \in I_l} k_i$ by 2^{n_l} matrix. n_l is the number of items in I_l. Each $\mathbf{P}_i$ is a full column rank matrix of rank 2. Thus, $\mathcal{P}_l$ is rank of 2^{n_l} matrix and is a full column rank matrix.

Each column vector of T_{I_l} is a column vector of $\mathcal{P}_l$. We can show that for two classes $\boldsymbol{\alpha}_1 \neq \boldsymbol{\alpha}_2$, $\boldsymbol{\alpha}_1$-th and $\boldsymbol{\alpha}_2$-th columns of T_{I_l} are not identical. We prove this by contradiction. Suppose that they are the same. It is easy to see that the $\boldsymbol{\alpha}_l$-th column in T_{I_l} has the form

$$\bigotimes_{i \in I_l} \begin{pmatrix} \pi_{i\boldsymbol{\alpha}_j}^1 \\ \pi_{i\boldsymbol{\alpha}_j}^2 \\ \vdots \\ \pi_{i\boldsymbol{\alpha}_j}^{k_i} \end{pmatrix},$$

$j = 1, 2$. So

$$\bigotimes_{i \in I_l} \begin{pmatrix} \pi^1_{i\boldsymbol{\alpha}_1} \\ \pi^2_{i\boldsymbol{\alpha}_1} \\ \vdots \\ \pi^{k_i}_{i\boldsymbol{\alpha}_1} \end{pmatrix} = \bigotimes_{i \in I_l} \begin{pmatrix} \pi^1_{i\boldsymbol{\alpha}_2} \\ \pi^2_{i\boldsymbol{\alpha}_2} \\ \vdots \\ \pi^{k_i}_{i\boldsymbol{\alpha}_2} \end{pmatrix}. \tag{A1}$$

However, we can find item $i^* \in I_l$ such that

$$\pi^1_{i^*\boldsymbol{\alpha}_1} + ... + \pi^k_{i^*\boldsymbol{\alpha}_1} \neq \pi^1_{i\boldsymbol{\alpha}_2} + ... + \pi^k_{i^*\boldsymbol{\alpha}_2}$$

for some $k = 1, ..., k^{i^*} - 1$ which means

$$\begin{pmatrix} \pi^1_{i^*\boldsymbol{\alpha}_1} \\ \pi^2_{i^*\boldsymbol{\alpha}_1} \\ \vdots \\ \pi^{k_{i^*}}_{i^*\boldsymbol{\alpha}_1} \end{pmatrix} \neq \begin{pmatrix} \pi^1_{i^*\boldsymbol{\alpha}_2} \\ \pi^2_{i^*\boldsymbol{\alpha}_2} \\ \vdots \\ \pi^{k_{i^*}}_{i^*\boldsymbol{\alpha}_2} \end{pmatrix}.$$

It contradicts with equation (A1) due to the fact that two different marginal distributions of item i^* leads to the two different joint distributions. Hence, each column of T_{I_l} is precisely one column of $\mathcal{P}_l$. T_{I_l} is of full column rank with rank M as a result. Then $C + C + C \geq 2C + 2$ whenever $C \geq 2$. We apply Lemma 1 and use the same argument as in the proof of Theorem 1. □

Proof of Theorem 3. There exist three non-overlapp subsets of items I_1, I_2, and I_3 such that $I_1 \bigcup I_2 \bigcup I_3 = \{1, \ldots, I\}$. We write the three-way array $W = [T_{I_1}\Lambda, T_{I_2}, T_{I_3}]$, where T_{I_1}, T_{I_2}, and T_{I_3} are the T-matrices of subsets I_1, I_2, and I_3 respectively and Λ is a $\prod_{i=1}^{I} k_i$ by $\prod_{i=1}^{I} k_i$ diagonal matrix with $\boldsymbol{\alpha}$-th diagonal element being $v_{\boldsymbol{\alpha}}$. Thus W is a $\prod_{i\in I_1} k_i$ by $\prod_{i\in I_2} k_i$ by $\prod_{i\in I_3} k_i$ array. It is not hard to see that $W(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \sum_{\boldsymbol{\alpha}} t^1_{\mathbf{x}_1\boldsymbol{\alpha}} t^2_{\mathbf{x}_2\boldsymbol{\alpha}} t^3_{\mathbf{x}_3\boldsymbol{\alpha}}$, where $t^l_{\mathbf{x}_l\boldsymbol{\alpha}}$ is the $(\mathbf{x}_l, \boldsymbol{\alpha})$-th element of matrix T_{I_l}. In other words, $W(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = P(\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3))$.

Suppose that there exists another set of parameters of the model giving the same distribution; that is, another decomposition of $W = [T'_{I_1}\Lambda', T'_{I_2}, T'_{I_3}]$. Because T_{I_l} are all full column rank. By applying the Lemma 1, we have that $T_{I_1}\Lambda = T'_{I_1}\Lambda' PF$, $T_{I_2} = T'_{I_2}PG$, and $T_{I_3} = T'_{I_3}PH$. Here, F, G, and H are all diagonal matrix with $FGH = \mathcal{I}$ and $\mathcal{I}$ is an identity and P is a column permutation matrix.

The sum of each column of $T_{I_2}, T_{I_3}, T'_{I_2}$, and T'_{I_3} equals 1. Then, G and H must be both identity matrices. As a result, F is an identity too. Due to the same reason that the sum of each column of T_{I_1} and T'_{I_1} is 1, we have $\Lambda = P'\Lambda' P$ and $T_{I_1} = T'_{I_1}P$. Besides, $T_{I_2} = T'_{I_2}P$ and $T_{I_3} = T'_{I_3}P$. We conclude that all parameters are identifiable up to a permutation of the columns. $\square$

Proof of Theorem 4. According to Theorem 3, it is sufficient to find three non-overlap subsets of items I_1, I_2, and I_3 such that $I_1 \bigcup I_2 \bigcup I_3 = \{1, \ldots, I\}$ and their corresponding T-matrices $T_{I_1}, T_{I_2}, T_{I_3}$ are all full column rank.

We construct the three subsets as follows: $I_1 = \bigcup_{a=1}^{A} I_{1,a}, I_2 = \bigcup_{a=1}^{A} I_{2,a}, I_3 = \{1, \ldots, I\} \backslash (I_1 \bigcup I_2)$. Then we need to show that I_1, I_2, I_3 are non-overlap and their T-matrix $T_{I_1}, T_{I_2}, T_{I_3}$ are of full column rank.

We know that $I_{l,a} \bigcap I_{j,b} = \emptyset$ for all $l \neq j$, $a \neq b, a, b \in \{1, \ldots, A\}$ and therefore $(\bigcup_{a=1}^{A} I_{l,a}) \bigcap (\bigcup_{b=1}^{A} I_{j,b}) = \emptyset$. This also implies that $\bigcup_{a=1}^{A} I_{3,a} \subset I_3$ and $I_1 \bigcap I_2 = \emptyset = I_3 \bigcap I_1 = I_3 \bigcap I_2 = \emptyset$. Hence, I_1, I_2 and I_3 are non-overlap.

Next, we need to prove $T_{I_l}, l = 1, 2, 3$ are of full column rank. Notice that $\bigcup_{a=1}^{A} I_{l,a} \subset I_l$. Thus the rank of $T_{\bigcup_{a=1}^{A} I_{l,a}}$ is less than or equals the rank of T_{I_3}. Thus if we can prove $T_{\bigcup_{a=1}^{A} I_{l,a}}$ is of full column rank then T_{I_3} is also of full column rank. As a result, we only need to show $T_{\bigcup_{a=1}^{A} I_{l,a}}$ are of full column rank.

Recall that the class label $\boldsymbol{\alpha} = (\alpha^1, \ldots, \alpha^A)$ and $\alpha_a \in \{1, \ldots, d_a\}$. It is s-traightforward to see that $T_{\bigcup_{a=1}^{A} I_{l,a}} = \bigotimes_{a=1}^{A} T_{l,a}$ since each column of $\bigotimes_{a=1}^{A} T_{l,a}$ is indexed by $(\alpha^1, \ldots, \alpha^A)$ and each row in $\bigotimes_{a=1}^{A} T_{l,a}$ is indexed by all the possible values $(x^1, \ldots, x^I)$. By the property of tensor product, the rank of $\bigotimes_{a=1}^{A} T_{l,a}$ equals the product of the rank of $T_{l,a}$. That is $\mathrm{rank}(\bigotimes_{a=1}^{A} T_{l,a}) = \prod_{a=1}^{A} d_a$. The number of columns in $\bigotimes_{a=1}^{A} T_{l,a}$ is also $\prod_{a=1}^{A} d_a$. Thus $T_{\bigcup_{a=1}^{A} I_{l,a}}$ is of full column rank. □

Proof of Theorem 5. It is sufficient to construct a consistent estimator of the partial information. Notice that the estimator does not have to be practically implementable. The strategy is to first consider the maximum likelihood estimator and merge the estimated item response probabilities based on their asymptotic properties.

Recall that $\pi_{i\boldsymbol{\alpha}}^{x} = P(X^i = x|\boldsymbol{\alpha})$ is the response probability to item i for latent class $\boldsymbol{\alpha}$. Let

$$L(\mathbf{x}; \boldsymbol{\pi}, v) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} v_{\boldsymbol{\alpha}} \Big\{ \prod_{i=1}^{I} \pi_{i\boldsymbol{\alpha}}^{x^i} \Big\}. \tag{A2}$$

be the likelihood of a single observation, where

$$\boldsymbol{\pi} = (\pi_{i\boldsymbol{\alpha}}^{x^i} : 1 \leq i \leq I, \boldsymbol{\alpha} \in \mathcal{M}, x^i \in \{1, ..., k_i\})$$

and $v = (v_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{M})$. Then, the maximum likelihood estimator is defined as

$$(\hat{\boldsymbol{\pi}}, \hat{v}) = \arg\max_{\boldsymbol{\pi}, v} \prod_{r=1}^{R} L(\mathbf{x}_r; \boldsymbol{\pi}, v).$$

According to the identifiability results in theorems and the asymptotic property of the M-estimator (Chapter 5.1 of Van der Vaart (1998)), $(\hat{\boldsymbol{\pi}}, \hat{v})$ converges weakly to the true parameter. Furthermore, according to chapter 5.3 Van der Vaart (1998), the MLE is asymptotically normally distributed. Thus, for each i, $\boldsymbol{\alpha}$, and x, we have

$$\hat{\pi}^x_{i\boldsymbol{\alpha}} - \pi^x_{i\boldsymbol{\alpha}} = O_p(R^{-1/2}). \tag{A3}$$

In the following, we use n to denote R for convenience. We say a random sequence $a_n = O_p(n^{-1/2})$ if $\sqrt{n}a_n$ is tight. Notice that the identifiability is subject to a permutation of the latent class labels. To simplify notation, we assume that the class labels of $\hat{\pi}^x_{i\boldsymbol{\alpha}}$ have been arranged in an appropriate order. Otherwise, we need to write $\hat{\pi}^x_{i\boldsymbol{\alpha}} - \pi^x_{i\boldsymbol{\alpha}} = O_p(n^{-1/2})$. Thus, we proceed assuming that the permutation λ is identity.

We now construct an estimator of the partial information for each item. The basic idea is that if $[\boldsymbol{\alpha}_1]_i = [\boldsymbol{\alpha}_2]_i$, then $\pi^x_{i\boldsymbol{\alpha}_1} = \pi^x_{i\boldsymbol{\alpha}_2}$ for all x. Together with (A3), we have that

$$d_i(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2) = \sum_{x=1}^{k_i} (\hat{\pi}^x_{i\lambda(\boldsymbol{\alpha}_1)} - \hat{\pi}^x_{i\lambda(\boldsymbol{\alpha}_2)})^2 = O_p(n^{-1}). \tag{A4}$$

Based on this fact, we define an equivalent class such that

$$\boldsymbol{\alpha}_1 \stackrel{i}{=} \boldsymbol{\alpha}_2 \quad \text{if} \quad d_i(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2) \leq n^{-1/2}. \tag{A5}$$

Based on (A4), we have that

$$P(\boldsymbol{\alpha}_1 \overset{i}{=} \boldsymbol{\alpha}_2) \to 1$$

as $n \to \infty$. If $[\boldsymbol{\alpha}_1]_i \neq [\boldsymbol{\alpha}_2]_i$, then there exists an ε and x such that $(\hat{\pi}^x_{i\lambda(\boldsymbol{\alpha}_1)} - \hat{\pi}^x_{i\lambda(\boldsymbol{\alpha}_2)})^2 > \varepsilon$ and thus

$$P(\boldsymbol{\alpha}_1 \overset{i}{\neq} \boldsymbol{\alpha}_2) \to 1$$

as $n \to \infty$. Let "$\langle\ \ \rangle_i$" be the canonical map of the estimated equivalence class as in (A5). Based on the above argument, we have that for each j,

$$P(\langle \boldsymbol{\alpha} \rangle_i = [\boldsymbol{\alpha}]_i) \to 1$$

as the sample size $n \to \infty$. Hence, the estimation of equivalence classes is the same as the true one up to a permutation. □

Proof of Theorem 6. It suffices to show that for any $\epsilon^0 > 0$ there exists N such that for any $n > N$,

$$P(\theta \in \Theta_c \setminus \mathcal{N}_\epsilon(\theta^*)|X) < \epsilon^0 \qquad P^* - a.s..$$

Let $\delta = \inf_{\theta \in \Theta^c \setminus \mathcal{N}_\epsilon(\theta^*)} \|P_\theta - P^*\|_1$, where P_θ and P^* are response distribution under parameters θ and θ^* respectively. According to the definition of Θ_c and identifiability results, we know that $P_\theta \neq P_{\theta^*}$ for any $\theta \in \Theta^c \setminus \mathcal{N}_\epsilon(\theta^*)$. Since Θ^c is a compact set and $\mathcal{N}_\epsilon(\theta^*)$ is an open set, then $\Theta^c \setminus \mathcal{N}_\epsilon(\theta^*)$ is also an compact set. With these facts, we know that $\delta > 0$.

Therefore, we have

$$P(\theta \in \Theta_c \setminus \mathcal{N}_\epsilon(\theta^*)|X) \leq P(\|P - P_0\|_1 > \delta/2|X) < \epsilon^0 \qquad P^* - a.s. \qquad \text{(A6)}$$

for sufficiently large n. The last inequality of equation (A6) follows by Theorem 2 in Dunson and Xing (2009). Hence, we conclude the proof.

Sliced Gibbs Sampler

We now present the sliced sampler for the simulation from the posterior distribution of model. The likelihood function is

$$\prod_{r=1}^{R}\Big\{\sum_{\boldsymbol{\alpha}=1}^{\infty} v_{\boldsymbol{\alpha}} \prod_{i=1}^{I}\prod_{x=1}^{k_i}(\pi_{i\boldsymbol{\alpha}}^{x})^{I(x_r^i=x)}\Big\}.$$

For each observation, we augment an independent index u_i following the uniform distribution in $[0,1]$. Thus, the complete data likelihood is

$$L(\boldsymbol{\pi}, v; \mathbf{x}_1, ..., \mathbf{x}_n, \boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_R, \mathbf{u}) = \prod_{r=1}^{R}\Big\{I(u_r < v_{\boldsymbol{\alpha}_r})\prod_{i=1}^{I}\prod_{x=1}^{k_i}(\pi_{i\boldsymbol{\alpha}}^{x})^{I(x_r^i=x)}\Big\}.$$

With this augmentation scheme, a Gibbs sampler iterates according the following conditional distributions.

1. Update u_r, for $r = 1, ..., R$, by sampling from the conditional posterior, $U(0, v_{\boldsymbol{\alpha}_r})$.

2. For $\boldsymbol{\alpha} = 1, ..., C^*$, where $C^* = \max\{\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_R\}$, update $\boldsymbol{\pi}_{i\boldsymbol{\alpha}}$ from the full conditional posterior distribution

$$\text{Dirichlet}\Big(1 + \sum_{r:\boldsymbol{\alpha}_r=\boldsymbol{\alpha}} I(x_r^i = 1), ..., 1 + \sum_{r:\boldsymbol{\alpha}_r=\boldsymbol{\alpha}} I(x_r^i = k_i)\Big)$$

3. For $\boldsymbol{\alpha} = 1, ..., C^*$, update $V_{\boldsymbol{\alpha}}$ from the conditional distribution that is Beta$(1, \beta)$ truncated to the interval

$$\left[\max_{r:\boldsymbol{\alpha}_r=\boldsymbol{\alpha}}\left\{\frac{u_r}{\prod_{l<\boldsymbol{\alpha}}(1-V_l)}\right\}, 1 - \max_{r:\boldsymbol{\alpha}_r>\boldsymbol{\alpha}}\left\{\frac{u_r}{V_{\boldsymbol{\alpha}}\prod_{l<\boldsymbol{\alpha}, l\neq\boldsymbol{\alpha}}(1-V_l)}\right\}\right].$$

4. Update each $\boldsymbol{\alpha}_r$ from the multinomial conditional distribution

$$P(\boldsymbol{\alpha}_r = \boldsymbol{\alpha}|...) = \frac{I(\boldsymbol{\alpha} \in A_r) \prod_{i=1}^{I} \pi_{i\boldsymbol{\alpha}}^{x_r^i}}{\sum_{l \in A_r} \prod_{i=1}^{I} \pi_{i\boldsymbol{\alpha}}^{x_r^i}}$$

   where $A_r = \{\boldsymbol{\alpha} : v_{\boldsymbol{\alpha}} > u_r\}$.

5. Assuming a gamma$(1, 1)$ hyperprior for β, update β by its conditional posterior

$$\text{gamma}(1 + C^*, 1 - \sum_{\boldsymbol{\alpha}=1}^{C^*} \log(1 - V_{\boldsymbol{\alpha}})).$$

At last, we point out that the computation complexity of the full Bayesian algorithm is $O(JRC^*)$ for each iteration and class labels practically will not switch after Markov chain becomes stable. In the simulation study, we order class labels based on their response possibility ($\pi_{i\boldsymbol{\alpha}}, i = 1, \ldots, I$) from low to high and then compare the results with the underlying true parameters.

References

Dunson, D. and Xing, C. (2009) Nonparametric bayes modeling of multivariate categorical data. *Journal of the American Statistical Association*, **104**, 1042–1051.

Kruskal, J. (1977) Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Application*, **18**, 95–138.

Van der Vaart, A. W. (1998) *Asymptotic statistics.* Cambridge University Press.