

ON THE IDENTIFIABILITY OF DIAGNOSTIC CLASSIFICATION MODELS

GUANHUA FANG[✉], JINGCHEN LIU AND ZHILIANG YING

COLUMBIA UNIVERSITY

This paper establishes fundamental results for statistical analysis based on diagnostic classification models (DCMs). The results are developed at a high level of generality and are applicable to essentially all diagnostic classification models. In particular, we establish identifiability results for various modeling parameters, notably item response probabilities, attribute distribution, and Q -matrix-induced partial information structure. These results are stated under a general setting of latent class models. Through a nonparametric Bayes approach, we construct an estimator that can be shown to be consistent when the identifiability conditions are satisfied. Simulation results show that these estimators perform well under various model settings. We also apply the proposed method to a dataset from the National Epidemiological Survey on Alcohol and Related Conditions (NESARC).

Key words: identifiability, diagnostic classification models, Dirichlet allocation.

1. Introduction

Diagnostic classification models (DCMs) are an important tool for cognitive diagnosis that has become increasingly recognized in educational assessment, psychiatric evaluation, and many other disciplines. A cognitive diagnostic test, consisting of a set of items, aims to examine each test taker's ability detailing his/her mastery of skills (often called attributes) based on item responses. For instance, a teacher may wish to find out a student's skill mastery based on his/her examination results; in psychiatry, patients are diagnosed based on their answers to diagnostic questions. There have been a number of DCMs developed: the rule space method (Tatsuoka 1985, 2009), the reparameterized unified/fusion model (RUM) (DiBello et al. 1995; Hartz 2002; Templin et al. 2003), the conjunctive (noncompensatory) models including deterministic-input, noisy-and-gate (DINA) model and noisy-input, deterministic-and-gate (NIDA) model (Junker and Sijtsma 2001, De La Torre and Douglas 2004), the compensatory models including deterministic-input, noisy-or-gate (DINO) and noisy-input, deterministic-or-gate (NIDO) model (Templin and Henson 2006), the attribute hierarchy method (Leighton et al. 2004). These core DCMs are followed by a development of more complicated models: the general diagnostic model (GDM) (von Davier 2005), the log-linear cognitive diagnosis model (LCDM) (Rupp and Templin 2008b), and the generalized DINA (G-DINA) model (De La Torre 2011). Many of the core DCMs are special cases of GDM, LCDM, and G-DINA, which are more flexible models with fewer parameter constraints. Generally speaking, a DCM is a restricted latent class model sharing the statistical nature of relating the latent unobserved characteristics to observed responses. We refer readers to Rupp et al. (2010) for a comprehensive review of DCMs, additional approaches to cognitive diagnosis, and detailed discussions of diagnostic evaluation.

In the application of DCMs to real data analysis, the related inference problems are also being investigated including model parameter estimation (Roussos et al. 2007; Stout 2007), Q -matrix estimation (Chen et al. 2015a, 2015b; Liu et al. 2012, 2013), hypothesis testing, and model

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11336-018-09658-x>) contains supplementary material, which is available to authorized users.

Correspondence should be made to Jingchen Liu, Columbia University, New York, USA.
Email: jlcliu@stat.columbia.edu; URL: <http://stat.columbia.edu/~jlcliu/>

diagnosis (Gu et al. 2018). A key issue common to DCMs and more generally to latent variable models is the model identifiability, which, generally speaking, is a property whether the unknown model or its parameters can be consistently estimated under a suitably defined asymptotic regime. It is worthwhile to point out that identifiability is always developed at certain level. For example, in exploratory factor analysis, the factor loading matrix is identifiable only up to a rotation and a reflection; in latent class models, the class labels are determined only up to a class permutation. The specific definition of identifiability will be provided in Sect. 2.2.

The Q -matrix is a key quantity in the specification of DCMs, and it characterizes the item–attribute relationship and specifies which items measure which attributes. A Q -matrix under different DCMs may lead to different model structures and distinct functional forms of correct answering probability, adding a great challenge to the development of unified theory. We consider a slightly different estimand, that is, the *partial information structure* implied by the Q -matrix. Specifically, a single item usually fails to provide information to differentiate all attribute profiles. People with different profiles may have same ability to answer this item. Therefore, only part of attribute profiles could be measured by an item. We develop identifiability results for this partial information structure for each item. Under specific model parameterizations, the Q -matrix can also be reconstructed based on the estimated partial information structure.

The primary focus of this paper is to ascertain sufficient conditions that ensure the identification for DCMs. We aim at developing theories under a general latent class model framework, of which diagnostic classification models are special cases with special parameterizations and constraints. Our results consist of two parts: 1. the identification of parameters for the item response functions and the attribute population that live on a continuous space and are considered as “regular” parameters and 2. the identification of item partial information structure.

The results of this paper make use of the fundamental results of Kruskal (1977), which provide sufficient conditions for the uniqueness decomposition of three-way arrays up to a column permutations and scalar scaling. We formulate DCMs into the form of three-way arrays and adapt Kruskal’s sufficient conditions to obtain the unique parameterization and, hence, develop identifiability results. The main contribution of the present paper is to provide the sufficient conditions for the identifiability of DCMs with multi-categorical responses and attributes. We discuss a few pieces of works most relevant to ours. A recent work by Allman et al. (2009), which is technically close to our work, discusses generic identifiability of general latent class models, hidden Markov models, and several other models with discrete latent variables. The identifiability results are *generic* in the sense that they hold when the parameters do not lie on some measure zero set. As we will discuss in the sequel, the parameters of DCMs always live on a low-dimensional manifold that is indeed a measure zero set. Therefore, the results in Allman et al. (2009) cannot be applied directly for DCMs. In the psychometrics literature, Xu and Zhang (2016, 2018) and Xu (2017) develop several identifiability results for Q -matrix and other model parameters under DINA model and more generally restricted latent class models. These analyses focus on binary response and binary attribute models. Our work takes a different point of view by starting with general latent class models and considers Q -matrix estimation by using item partial information structure. We extend the identifiability results for more general DCMs with polytomous responses and without parameter restrictions. The results are presented under the settings of general latent class models and DCMs, including many well-known cognitive diagnosis models as special cases.

It is customary to work with a pre-specified Q -matrix; for example, an examination maker specifies the set of skills associated with each examination question and a psychologist designs a questionnaire to measure different personalities. However, such a specification is usually very subjective and may not always be precise. Therefore, we develop an exploratory estimation approach starting from a general latent class model. We adopt a nonparametric Bayes approach by making use of a Dirichlet allocation method (Dunson and Xing 2009). Specifically, it adopts a prior distribution on the latent class probabilities via the stick-breaking representation that is originated from

the derivation of the Dirichlet processes. The major advantage of this approach is that the number of latent classes need not be specified and the infinitely many latent classes are allowed. In the estimation, we adopt a full Bayesian setting. The posterior distribution is obtained by means of a Gibbs sampling scheme. An estimator of the partial information structure is constructed based on a clustering algorithm of the Bayesian estimator. It can be used for recovery of Q -matrix when additional model information is provided.

The rest is organized as follows. In Sect. 2, we present basic concepts including latent class models, diagnostic classification models, identifiability, and estimation consistency. In Sect. 3, the main results are presented including the identifiability of item parameters, attribute distribution, and the partial information structure. A latent class model with Dirichlet allocation along with its inference is presented in Sect. 4. Simulation studies and real data analysis are presented in Sects. 5 and 6, respectively. A concluding remark is given in Sect. 7.

2. Preliminaries

2.1. Latent Class Models and Diagnostic Classification Models

We start with formally introducing latent class models and DCMs as well as their connections. We consider an I -dimensional multivariate categorical response random vector $\mathbf{X} = (X^1, \dots, X^I)$. We use subscripts to index independent replications, that is, \mathbf{X}_r and X_r^i . Here, r indicates respondent and i indicates item. Let X^i be a discrete random variable taking k_i possible values, $X^i \in \{1, \dots, k_i\}$. Their dependence is incorporated into a discrete latent variable/class α taking values in a discrete set \mathcal{A} . In other words, $P(X^i = x^i, i = 1, \dots, I | \alpha) = \prod_{i=1}^I P(X^i = x^i | \alpha)$ which is known as the local independence. Let v_α ($\alpha \in \mathcal{A}$) be the probability mass function of the latent class membership. We parameterize $P(X^i = x^i | \alpha)$ as $f_i(x^i | \theta_i, \alpha)$ where θ_i is the item-specific parameter. Thus, the joint marginal distribution of \mathbf{X} can be expressed as

$$P(X^1 = x^1, \dots, X^I = x^I) = \sum_{\alpha \in \mathcal{A}} \left\{ v_\alpha \prod_{i=1}^I f_i(x^i | \theta_i, \alpha) \right\}. \quad (1)$$

Here, we adopt the following parameterization, $\mathcal{A} = \{(\alpha_1, \dots, \alpha_A) | \alpha_a \in \{1, \dots, d_a\}\}$. Under this representation, each α_a is known as an attribute or trait indicating the presence/absence or level of a -th latent characteristic.

Formulation (1) contains DCMs as special cases where the item response function f_i also admits some low-dimensional structures that distinguish themselves from the general latent class models. We introduce the Q -matrix, a $I \times A$ matrix with 0/1 entries, indicating item–attribute association. Each row of Q corresponds to an item, and each column corresponds to an attribute. We write

$$Q = \begin{pmatrix} q_{11} \\ \vdots \\ q_{1I} \end{pmatrix},$$

where $q_i = (q_{i1}, \dots, q_{iA})$ and $q_{ia} = 1$ or 0 , indicating whether or not f_i depends on α_a . Thus, $f_i(x | \theta, \alpha) = f_i(x | \theta, \alpha_1, \alpha_2)$ if $q_i = (1, 1)$, and $f_i(x | \theta, \alpha) = f_i(x | \theta, \alpha_1)$ if $q_i = (1, 0)$. The Q -matrix qualitatively describes the relationship between items and attributes. The specific forms of f_i as functions of α are determined by the parametric model subject to the constraints implied by the Q -matrix and additional item parameters. We provide two examples of parametric DCMs for readers to easily understand these points.

Example 1. (DINA model, Junker and Sijtsma 2001) The DINA model assumes a conjunctive (noncompensatory) relationship among attributes with the i th item response function being:

$$f_i(x|\theta, \boldsymbol{\alpha}) = (1 - s_i)^{\xi_{i\boldsymbol{\alpha}}} g_i^{1-\xi_{i\boldsymbol{\alpha}}}. \quad (2)$$

where $\xi_{i\boldsymbol{\alpha}} = \prod_{a=1}^A (\alpha_a)^{q_{ia}} = I(\alpha_a \geq q_{ia} \text{ for all } a)$ is the ideal response, that is, whether $\boldsymbol{\alpha}$ has all the attributes required by item i . Item-specific parameters s_i and g_i are known as the slipping and guessing parameters, respectively. Parameter s_i represents the probability that the respondents who have mastered all required attributes incorrectly answer the i th item and g_i represents the probability that the respondents who do not possess at least one of the required attributes correctly answer the i -th item. If $\xi_{i\boldsymbol{\alpha}} = 1$ (the subject is capable of solving a problem), then the positive response probability is $1 - s_i$; otherwise, the probability is g_i . The DINA model is popular in the educational testing (DeCarlo 2011; De La Torre 2009).

Example 2. (LCDM, Henson et al. 2009) The LCDM allows for complicated relationships between categorical variables. Many classical DCMs can be parameterized as special cases of LCDM including DINA, DINO, etc. The item response function takes the following form:

$$f_i(x|\theta, \boldsymbol{\alpha}) = \frac{\exp(\boldsymbol{\lambda}_i^T h(\boldsymbol{\alpha}, \mathbf{q}_i) + \lambda_{i,0})}{1 + \exp(\boldsymbol{\lambda}_i^T h(\boldsymbol{\alpha}, \mathbf{q}_i) + \lambda_{i,0})},$$

where $\lambda_{i,0}$ is the intercept parameter and $\boldsymbol{\lambda}_i^T h(\boldsymbol{\alpha}, \mathbf{q}_i) = \sum_{a=1}^A \lambda_{i,1,(a)} (\alpha_a q_{ia}) + \sum_{a=1}^A \sum_{\tilde{a}>a} \lambda_{i,2,(a,\tilde{a})} (\alpha_a \alpha_{\tilde{a}} q_{ia} q_{i\tilde{a}}) + \dots$. More specifically, $\lambda_{i,1,(a)}$ represents the main effect for Attribute a and $\lambda_{i,2,(a,\tilde{a})}$ is pairwise interaction effect for Attribute a and Attribute \tilde{a} . Because of its generality, the LCDM may be used for identifying a suitable DCM by imposing parameter restriction. It can be seen that the response probability is linked to linear combination of attributes by a logit function. The linear additive form and hierarchical decomposition of $\boldsymbol{\lambda}_i$ allow for a systematic way to interpret the parameters.

2.2. Identifiability of Model Parameters

We discuss the identifiability for two types of parameters separately: (1) item parameters and the attribute distribution and (2) item partial information structure and the Q -matrix. Let θ denote the vector of all item parameters and the attribute distribution. Its identifiability is defined as follows.

Definition 1. The parameter θ is said to be identifiable if, for any $\theta' \neq \theta$, the resulting marginal distributions of the responses \mathbf{X} in (1) are distinct.

If θ is identifiable, then, thanks to the entropy inequality and under very mild conditions, the maximum likelihood estimator is consistent (Van der Vaart 1998). In the context of latent class models and DCMs, the list of parameters includes the item response probabilities and attribute/latent class distribution.

The Q -matrix is different from the regular item parameters and is usually assumed to be known. However, the misspecification of the Q -matrix could possibly lead to biased estimations of model parameters and inaccurate classification of the latent attribute profiles (Rupp and Templin 2008a). Therefore, an objective construction of the Q -matrix estimated based on the data is useful and important. In what follows, we provide some discussions on the estimation of the Q -matrix for a generic DCM. The problem of Q -matrix estimation can be viewed from different perspectives.

The most straightforward approach is to treat Q as part of the model parameters and to consider it as a usual estimation of parametric models. This is often difficult from the computational aspect in that Q is a discrete matrix living on a high-dimensional space. Even with a reasonably small number of items and a few attributes, this space is often too large to explore thoroughly by any existing numerical method as the dimension grows exponentially fast with both I and A . Generally speaking, maximizing likelihood, which is usually a discrete and nonlinear function over $Q \in \{0, 1\}^{I \times A}$, is computationally intensive and sometimes infeasible. Estimators developed based on this idea, even though theoretically sound, often suffer from substantial computational overhead. A different approach is to cast the Q -matrix estimation in the context of variable selection. If both the response X^i and the latent variable α were observed, then the estimation of Q is a regular variable selection problem. In most situations, f_i takes the form of a generalized linear model, in which the responses to items are the dependent variables, the attributes play the role of covariates, and the item parameters θ are the regression coefficients. Thus, the Q -matrix estimation is equivalent to a variable selection problem. However, in the context of latent class models, the covariates α 's are all missing and therefore the task is, rigorously speaking, to select latent variables. Chen et al. (2015a) took this viewpoint and developed estimation methods for the Q -matrix via regularized likelihood.

The Q -matrix suggests that a single item usually does not provide information to differentiate all dimensions of the attribute profile. In particular, $q_{ia} = 0$ means that item i is irrelevant to attribute a . Under the setting of latent class models (not necessarily possessing a specific parameterization), this corresponds to an item-specific *partial information structure*. If an item does not differentiate all dimensions of α , then some distinct attribute profiles may admit the same response distribution. In other words, there exist $\alpha_1 \neq \alpha_2$ such that $f_i(x|\theta, \alpha_1) = f_i(x|\theta, \alpha_2)$ for all x . In this case, responses to this item of subjects in latent classes α_1 and α_2 admit the same probability law. Thus, each item usually provides partial information of the entire attribute profile. This information structure will be formulated mathematically in the sequel. Each Q -matrix along with a specific model parameterization (such as the DINA and DINO models) maps to a item-specific partial information structure. We develop identifiability results and a computation approach for the partial information structure.

3. On the Identifiability of Diagnostic Classification Models

3.1. Identifiability of Item Parameters and the Attribute Distribution

Our main identifiability results consist of two parts: 1. item parameters and attribute population and 2. item partial information structure. These results are presented in the form of four theorems which are applicable to different situations. We start with the simplest case that the responses are binary and each item only has two possible response distributions. We provide all the technical statements of the theorems in Sect. 3.1.1, and the discussions on the conditions and their implications are given in Sect. 3.1.2.

3.1.1. Main Results In the following, we use the simplified notation $\pi_{i\alpha} = P(X^i = 1|\alpha)$, $\pi_{i\alpha}^k = P(X^i = k|\alpha)$ and $\boldsymbol{\pi}_{i\alpha} = (\pi_{i\alpha}^1, \dots, \pi_{i\alpha}^{k_i})$ ($i = 1, \dots, I$) to represent item response probability (vector) and always assume (v_α) are strictly positive for all α 's. Theorems 1 and 2 are for general latent class models, and Theorems 3 and 4 are for DCMs with polytomous attributes and responses.

Theorem 1. *We consider a latent class model with $C > 2$ latent classes. The responses are binary and take values in $\{0, 1\}$. Suppose that there exist three nonoverlap subsets of items denoted by I_1, I_2 , and I_3 such that*

- A1 for each $l = 1, 2$ or 3 , the conditional distributions of $(X^l : i \in I_l)$ on classes α_1 and α_2 are distinct if $\alpha_1 \neq \alpha_2$;
- A2 for each $i \in I_1 \cup I_2 \cup I_3$, the response probabilities $(\pi_{i\alpha} : \alpha = 1, \dots, C)$ take only two possible values. That is, the cardinality of set $\{\pi_{i\alpha} | \alpha = 1, \dots, C\}$ is two for each i .

Then, the item parameters $\pi_{i\alpha}$ and the latent class population v_α are identifiable up to a permutation of the class label.

The identifiability up to a permutation of class label refers to the following fact. Suppose that there exist a set of item parameters and attribute prior distribution, denoted by $\tilde{\pi}_{i\alpha}^k$ and \tilde{v}_α , yielding the same marginal distribution as that in (1) with parameters $\pi_{i\alpha}^k$ and v_α . Then, there exists a one-to-one map on the latent class space, λ , such that $\tilde{\pi}_{i\lambda(\alpha)}^k = \pi_{i\alpha}^k$ and $\tilde{v}_{\lambda(\alpha)} = v_\alpha$.

Corollary 1. Consider a DCM for I binary responses with A binary attributes. Suppose that we can rearrange the columns and rows of Q such that it contains three distinct identity submatrices. Then, the item response function and the attribute population are identifiable up to a relabeling of the attributes.

In DCM, there is no relabeling issue once we parameterize each latent class into 0-1 attribute profile such that each attribute has specific meaning.

Theorem 2. We consider a latent class model with $C > 2$ classes. The response to item i takes k_i possible values $\{1, \dots, k_i\}$. Suppose that there exist three nonoverlap subsets of items denoted by I_1, I_2 , and I_3 such that

- B1 for each $i \in I_1 \cup I_2 \cup I_3$, the response vectors $(\pi_{i\alpha} : \alpha = 1, \dots, C)$ take only two distinct values, that is to say, the cardinality of vector set $\{\pi_{i\alpha} | \alpha = 1, \dots, C\}$ is two.
- B2 for each $l = 1, 2$ or 3 , $\pi_{i\alpha_1}^1 + \dots + \pi_{i\alpha_1}^k \neq \pi_{i\alpha_2}^1 + \dots + \pi_{i\alpha_2}^k$ for some $k = 1, \dots, k_i - 1$ if $\alpha_1 \neq \alpha_2$.

Then, the item parameters $\pi_{i\alpha}$ and the latent class population v_α are identifiable up to a permutation of the class labels.

In order to present results for general DCMs with polytomous attributes and responses, we first introduce a T -matrix following Liu et al. (2013). Consider a test with A attributes and I items. The response to item i takes k_i different values $\{1, \dots, k_i\}$, and attribute α_a takes d_a possible values $\{1, \dots, d_a\}$. There are $\kappa = \prod_{i=1}^I k_i$ response patterns and $C = \prod_{a=1}^A d_a$ latent classes. We defined the T -matrix as a $\kappa \times C$ matrix. Each column of the matrix corresponds to one attribute profile or one latent class, and each row corresponds to one response pattern. The particular order of the columns and rows of the T -matrix does not affect the results. For technical convenience, we use α and $\mathbf{x} = (x^1, \dots, x^I)$ to label its columns and rows, that is, $t_{\mathbf{x}\alpha}$ is the element in row \mathbf{y} and column α and $t_{\mathbf{x}\alpha} = P(\mathbf{X} = \mathbf{x} | \alpha) = \prod_{i=1}^I \pi_{i\alpha}^{x^i}$. Very often, we construct a T -matrix for a subset of items $I_0 \subset \{1, \dots, I\}$. We use T_{I_0} to denote the corresponding matrix of the items in I_0 and T to denote the matrix to all items.

For illustration, we construct the T -matrix for a LCDM with two attributes and two items:

$$\pi_{1\alpha} = \frac{e^{\alpha^1}}{1 + e}, \quad \pi_{2\alpha} = \frac{e^{2\alpha^2}}{1 + e^2}$$

that corresponds to an identity Q -matrix. In this case, the latent attribute α belongs to $\{(0, 0), (1, 0), (0, 1), (1, 1)\}$ and the response \mathbf{X} belongs to $\{(1, 1), (1, 0), (0, 1), (0, 0)\}$. According to the above item response function, we obtain $P(X^1 = 1 | \alpha^1 = 1) = 0.731$, $P(X^1 = 1 | \alpha^1 =$

$0) = 0.269$, $P(X^2 = 1|\alpha^2 = 1) = 0.881$ and $P(X^2 = 1|\alpha^2 = 0) = 0.119$. Then, the T -matrix is

$$T = \begin{pmatrix} (0, 0) & (0, 1) & (1, 0) & (1, 1) \\ 0.03 & 0.24 & 0.09 & 0.64 \\ 0.24 & 0.03 & 0.64 & 0.09 \\ 0.09 & 0.64 & 0.03 & 0.24 \\ 0.64 & 0.09 & 0.24 & 0.03 \end{pmatrix} \begin{matrix} (1, 1) \\ (1, 0) \\ (0, 1) \\ (0, 0) \end{matrix}.$$

We refer to Liu et al. (2012, 2013) for more discussions on the T -matrix.

Theorem 3. Consider a DCM of I items and A attributes. Suppose that items can be partitioned into three nonoverlap subsets I_1 , I_2 , and I_3 , that is, $I_1 \cup I_2 \cup I_3 = \{1, \dots, I\}$ so that T_{I_1} , T_{I_2} , and T_{I_3} are all of full column rank. Then, the item parameters $\pi_{i\alpha}$ and the latent class population v_α are identifiable up to a permutation of the class labels.

Evaluation of column ranks could be computationally intensive. The following theorem provides easy-to-check conditions.

Theorem 4. Consider a DCM of I items and A multi-category attributes. For each attribute $a = 1, \dots, A$, there exist three nonoverlap subsets of items $I_{1,a}$, $I_{2,a}$, and $I_{3,a}$ satisfying the following conditions.

- C1 The items in $I_{1,a}$ are only associated with attribute a ; that is, their corresponding row vector in Q is e_a .
- C2 Let $T_{I_{1,a}}$ be the corresponding T -matrix of this reduced simple attribute model. The matrix $T_{I_{1,a}}$ is of full column rank.

Then, the item parameter $\pi_{i\alpha}^k$ and the latent class population parameter v_α are identifiable up to a permutation of the class labels.

3.1.2. Discussion of the Theorems We provide further discussions on the above theorems. Assumption A2 is applicable to simple models, such as the DINA and the DINO models. Regarding Assumption A1, it is necessary for a test to have at least one set of items I that is able to differentiate among all latent classes; otherwise, it is always possible to merge some latent classes and reduce the model to satisfy the condition. Following the idea of repeated measurements (Vonesh and Chinchilli 1997), Assumption A1 requires three such sets of items. It is often satisfied for many tests in practice. Corollary 1 of Theorem 1 presents easy-to-check conditions for DCMs. Theorem 2 extends 1 to the polytomous response case.

Theorem 1 is close to Theorem 1 in Xu (2017) where Xu develops results for restricted latent class model (RLCM) that has binary attributes and binary responses. Xu’s results require similar but weaker conditions as Condition A1. The difference lies in that Xu’s model requires that the item response probabilities admit the same partial order as that of the attribute profiles. Theorems 1 and 3 in our paper are for general latent class models (DCMs) and do not require partial order structure of the response probabilities. When additional constraints on parameters are assumed, the sufficient conditions could be reduced accordingly. There are several results developed for specific models and under various conditions. For instance, Xu and Zhang (2016) develop results for the DINA model and a follow-up work by Gu and Xu (2018) develops sufficient and necessary conditions for the parameter identifiability of the DINA model. Their results are developed by assuming that the Q -matrix is correctly specified in the model estimation. Other

core DCM model like DINO is mathematically equivalent to DINA model, and the same reduced sufficient condition will guarantee its identifiability.

In the statements of all four theorems, there is a similar condition that requires three distinct subsets of items to identify each attribute. This is one of the key sufficient conditions. It can be relaxed with additional parametric assumptions, for instance, the sufficient and necessary conditions in Gu and Xu (2018) for the DINA model with a correctly specified Q -matrix. To gain intuition, we provide identifiable and nonidentifiable examples for the binary attribute case. A DCM with binary attribute is identifiable if it admits a Q -matrix that contains at least three identity submatrices, that is,

$$Q^1 = \begin{pmatrix} \mathcal{I}_A \\ \mathcal{I}_A \\ \mathcal{I}_A \\ * \end{pmatrix},$$

where \mathcal{I}_A is the $A \times A$ identity matrix and “*” denotes any entry. Such a Q -matrix ensures the existence of three subsets of items identifying each attributes. If we reduce the number of identity submatrices to two, for instance,

$$Q = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix},$$

we are able to construct two sets of parameters leading to the same response distribution. To simplify notation, we write $\alpha_1 = (1, 1)$, $\alpha_2 = (1, 0)$, $\alpha_3 = (0, 1)$, and $\alpha_4 = (0, 0)$ and $\pi_{\alpha_c} = (\pi_{1\alpha_c}, \pi_{2\alpha_c}, \pi_{3\alpha_c}, \pi_{4\alpha_c})$. The first set of parameters is $\pi_{\alpha_1}^{(1)} = (0.7, 0.7, 1/2, 1/2)$, $\pi_{\alpha_2}^{(1)} = (0.7, 0.7, 1/6, 1/6)$, $\pi_{\alpha_3}^{(1)} = (0.3, 0.3, 1/2, 1/2)$, $\pi_{\alpha_4}^{(1)} = (0.3, 0.3, 1/6, 1/6)$, and $v^{(1)} = (1/12, 5/12, 1/12, 5/12)$. The second set of parameters is $\pi_{\alpha_1}^{(2)} = (0.7, 0.7, 1/3, 1/3)$, $\pi_{\alpha_2}^{(2)} = (0.7, 0.7, 1/12, 1/12)$, $\pi_{\alpha_3}^{(2)} = (0.3, 0.3, 1/3, 1/3)$, $\pi_{\alpha_4}^{(2)} = (0.3, 0.3, 1/12, 1/12)$, and $v^{(2)} = (5/18, 4/18, 5/18, 4/18)$. It is easy to check that these two sets of parameters give the same response distribution.

The conditions in Theorem 3 are mild, but they are sometimes difficult to verify. This is due to the fact that the T -matrix is computationally costly to construct. For instance, consider a subset of 20 items having binary responses and its T -matrix has $2^{20} = 1,048,576$ rows. Thus, construction of T -matrix for reasonably large-scale studies is impossible. Theorem 4 requires stronger conditions, but they are much easier to check since we only need to consider constructions of much smaller T -matrices. In particular, the construction of $T_{I_{l,a}}$ often contains very few items and the matrix only contains d_a columns. Generally speaking, we need to include a sufficient number of items in each $I_{l,a}$ so that their responses contain information to differentiate different latent classes defined by Attribute a . Theorem 4 seemingly requires many single-attribute items, that is, three distinct sets of items for each attribute. However, it is reasonable. Notice that matrix $T_{I_{l,a}}$ for the reduced single-attribute model contains d_a columns. In the case of binary attribute, it is sufficient to include one single-attribute item in each $I_{l,a}$; see the proof of Theorem 2. In other words, we only need three single items totally for each attribute in binary attribute model, which is not that many. Besides, it still remains possibly sufficient to include a single-attribute items in each $I_{l,a}$ if the response to the item also takes more than two possible values.

The identifiability results in the paper are all subject to a latent class permutation; that is, the class label cannot be identified based on the data. In practice, the model estimation is usually

subject to additional constraints. For instance, Xu (2017), Rupp et al. (2010) impose monotonicity constraints on the item response functions. Under such constraints, the class labels are no longer exchangeable and furthermore stronger identifiability results can be developed.

3.2. Identifiability of Partial Information Structure

The previous section provides identifiability results of item parameters and the attribute distribution. We now proceed to the discussion of Q -matrix and partial information structure. The Q -matrix specifies the relationship between the items and attributes. The specific form of the item response function depends on the model parameterization. As the aim of this study is to provide results applicable to general DCMs, we take a slightly different viewpoint and consider the partial information structure that is a mathematically more general concept than the Q -matrix. Distinct Q -matrices could lead to the same partial information structure under different DCMs. Studying the partial information structure allows to perform analysis without specifying the particular DCM.

To proceed, we start with a description of the item partial information in the context of a general latent class model. Let $\alpha \in \mathcal{M} = \{1, \dots, C\}$ denote the latent class membership. The partial information of item i characterizes the latent classes it is capable of differentiating. Mathematically, we define an item-specific equivalence relation on \mathcal{M} , denoted by “ $\stackrel{i}{=}$.” For $\alpha_1, \alpha_2 \in \mathcal{M}$,

$$\alpha_1 \stackrel{i}{=} \alpha_2 \quad \text{if} \quad \pi_{i\alpha_1}^x = \pi_{i\alpha_2}^x \quad \text{for all } x = 1, \dots, k_i.$$

It is not hard to verify that “ $\stackrel{i}{=}$ ” is an equivalence relation. We define the quotient set $\mathcal{M}/\stackrel{i}{=}$ as the partial information of item i and use $[\alpha]_i$ to denote the corresponding equivalence class that latent class α belongs to. The map $[\cdot]_i$ is known as the canonical projection which leads to a partition of \mathcal{M} . Two latent classes are mapped to the same equivalence class, $[\alpha_1]_i = [\alpha_2]_i$ if $\alpha_1 \stackrel{i}{=} \alpha_2$, and in this case, item i does not provide information to differentiate α_1 and α_2 .

From the modeling point of view, the Q -matrix along with a particular loading parameterization determines the partial information of each item. Consider a particular item i whose corresponding row vector in Q has the first l entries being one and others being zero. Then, the conditional response distribution is reduced to

$$P(X^i = x^i | \alpha) = P(X^i = x^i | \alpha^1, \dots, \alpha^l).$$

Consider two attribute profiles α_1 and α_2 . If their first l components are identical, then $[\alpha_1]_i = [\alpha_2]_i$. The following theorem presents identifiability of the partial information structure, whose proof is in Appendix.

Theorem 5. *Under each set of conditions of Theorems 1, 2, 3, or 4, the partial information of each item can be consistently estimated up to a permutation of the latent class label. That is, letting “ $\langle \cdot \rangle_i$ ” be the estimated canonical projection of item i , there exists a permutation of latent class labels λ such that*

$$P(\langle \lambda(\alpha) \rangle_i = [\alpha]_i) \rightarrow 1 \quad \text{for all } i,$$

as the sample size $R \rightarrow \infty$, $\alpha \in \mathcal{M}$.

4. Estimation via a Latent Class Model with Dirichlet Allocation

DCM with a Q -matrix structure can be utilized to facilitate computation; see Chen et al. (2015a) and Chen et al. (2018). In the absence of such a structure (i.e., general latent class model), we need to explore alternative approaches to estimating model parameters. In this section, we develop an estimation method for the partial information structure by making use of the Dirichlet allocation (Dunson and Xing 2009). We adopt the general setup of the latent class models in Sect. 2.1. The Dirichlet allocation approach does not require the upper bound on the number of latent classes which is typically the case of exploratory analysis. Without loss of generality, we assume that the latent classes are labeled by natural numbers $\mathcal{M} = \{1, 2, \dots\}$. The marginal distribution of the responses in (1) becomes

$$P(X^1 = x^1, \dots, X^I = x^I) = \sum_{\alpha=1}^{\infty} v_{\alpha} \prod_{i=1}^I \pi_{i\alpha}^{x_i^i}, \quad (3)$$

where $\boldsymbol{\pi}$ and v are unknown parameters. This is an infinite mixture model, which is overly parameterized. We adopt a Bayesian model and a proper prior distribution to regularize this overparameterization. Specifically, we assume the item response probabilities follow a Dirichlet prior distribution

$$\boldsymbol{\pi}_{i\alpha} = (\pi_{i\alpha}^x : x = 1, \dots, k_i) \sim \text{Dirichlet}(1, \dots, 1). \quad (4)$$

We assume the class probability functions (v_{α} , $\alpha = 1, 2, \dots$) follow a stick-breaking prior (Sethuraman 1994). Specifically,

$$v_{\alpha} = V_{\alpha} \prod_{l < \alpha} (1 - V_l), \quad (5)$$

where $\{V_i : i = 1, 2, \dots\}$ is a sequence of i.i.d. random variables following the Beta distribution $\text{Beta}(1, \beta)$. It is easy to verify that v_{α} under the above construction is a well-defined probability mass function. This is known as the stick-breaking representation originated from the Dirichlet process. We borrow this representation mostly due to its technical convenience for modeling a discrete distribution. The likelihood function (3) and the prior distributions (4) and (5) completely specify a Bayesian model.

We adopt this model for several reasons. First, it does not require to specify the number of latent classes. The stick-breaking representation penalizes the ‘‘tail’’ latent classes. In addition, the posterior distribution of this model can be obtained through a sliced sampler that is a Gibbs sampler via a data augmentation scheme without truncating the model to a finite mixture (Walker 2007). The posterior simulation scheme is presented in the supplementary.

Once posterior distribution is obtained, we can calculate the posterior mean as the point estimator, that is,

$$(\hat{\boldsymbol{\pi}}, \hat{v}) = E\{(\boldsymbol{\pi}, v) | \mathbf{X}_1, \dots, \mathbf{X}_R\}.$$

The item-wise partial information structure is estimated by clustering the item response probabilities. For each item i and latent attribute profile α , let $\hat{\boldsymbol{\pi}}_{i\alpha} = (\hat{\pi}_{i\alpha}^1, \dots, \hat{\pi}_{i\alpha}^{k_i})$ be the response probability. We treat $\hat{\boldsymbol{\pi}}_{i1}, \hat{\boldsymbol{\pi}}_{i2}, \hat{\boldsymbol{\pi}}_{i3}, \dots$ as (infinitely many) samples, each of which is a k_i dimensional vector, and apply the K -means clustering algorithm to group these samples. Although there are seemingly infinitely many samples, $\hat{\boldsymbol{\pi}}_{i\alpha}$'s are almost same as the prior mean for large

α . This is because there will be no respondents assigned to those large α labels in computation which implies the posterior mean is equal to the prior mean. Finally, we select the largest C such that

$$\hat{\pi}_\alpha \ll R^{-1/2} \quad \alpha = C + 1, C + 2, \dots$$

That is, we treat latent classes of very small proportion ($O(\frac{1}{\sqrt{R}})$) as practically nonexistent. In applying the K -means algorithm, we truncate $\hat{\pi}_{i\alpha}, \alpha = 1, \dots$ into finite samples $\hat{\pi}_{i1}, \hat{\pi}_{i2}, \dots, \hat{\pi}_{iC}$. Then, the estimated partial information structure is given by $\alpha_1 \stackrel{i}{=} \alpha_2$ if $\hat{\pi}_{i\alpha_1}$ and $\hat{\pi}_{i\alpha_2}$ are in the same cluster.

Notice that different items could have different partial information structures; that is, the number of clusters may vary from item to item. To determine the number of clusters for each item, we use the following criterion. We apply K -means method to data points $\hat{\pi}_{i1}, \dots, \hat{\pi}_{iC}$ with number of clusters starting from 2 up to C . We pooled the estimates for each cluster, that is, $\hat{\pi}_{ih} = \sum_{\alpha \in \mathcal{C}_h} v_\alpha \hat{\pi}_{i\alpha} / \sum_{\alpha \in \mathcal{C}_h} v_\alpha$ where \mathcal{C}_h is the set of classes belonging to cluster h . We then compute the information criterion $IC_i(n_i) = -2l_i(n_i) + \text{const} \times \log(R) \times n_i$. Here, $l_i(n_i)$ is the log-likelihood computed under $(\tilde{\pi}, \hat{v})$ in which $\tilde{\pi}$ is same as $\hat{\pi}$ except that $\hat{\pi}_i$ is replaced by pooled estimates with the number of clusters for i -th item being n_i . Then, the optimal n_i^* is $\text{argmin}_{n_i} IC_i(n_i)$.

In the following paragraph, we provide two examples showing how to reconstruct a Q -matrix based on partial information structure when we have additional information.

Example 3. Suppose we have the following partial information structure results under the setting of DINA model with $1 - s_i > g_i$.

$$\text{Partial information} = \begin{array}{c} \text{Item} \\ 1-2 \\ 3-4 \\ 5-6 \\ \dots \end{array} \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ \circ & \circ & \bullet & \bullet \\ \circ & \bullet & \circ & \bullet \\ \circ & \circ & \circ & \bullet \\ \dots & \dots & \dots & \dots \end{pmatrix}.$$

Here, we use symbol \circ and \bullet to denote level 1 (low probability) and level 2 (high probability), respectively. Therefore, we could parameterize α_1 as $(0, 0)$, α_2 as $(0, 1)$, α_3 as $(1, 0)$, α_4 as $(1, 1)$ after ordering of attributes. Then, the Q -matrix can be constructed as follows:

$$Q = \begin{array}{c} \text{Item} \\ 1-2 \\ 3-4 \\ 5-6 \\ \dots \end{array} \begin{pmatrix} \alpha^1 & \alpha^2 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ \dots & \dots \end{pmatrix}.$$

Example 4. Under the setting of the reduced noncompensatory reparameterized unified model (reduced NC-RUM, Hartz 2002), the response probability takes form

$$\pi_{i\alpha} = \pi_i^* \prod_{a=1}^A r_{ia}^{*(1-\alpha_a)q_{ia}}.$$

π_i^* is the probability of a correct response when all measured attributes have been mastered. r_{ia}^* is the item- and attribute-specific penalty for not having attribute a . Parameters π_i^* and r_{ia}^* are between 0 and 1.

Suppose we have the following estimated partial information structure:

$$\text{Partial information} = \begin{array}{c} \text{Item} \\ 1-3 \\ 4-6 \\ 7 \\ \dots \end{array} \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ \circ & \circ & \bullet & \bullet \\ \circ & \bullet & \circ & \bullet \\ \circ & \bullet & \bullet & \dagger \\ \dots & \dots & \dots & \dots \end{pmatrix},$$

where symbols \circ , \bullet and \dagger are used to denote level 1 (lowest probability) to level 3 (highest probability), respectively. Then, we can also parameterize each latent class and reconstruct the Q -matrix. That is, after an attribute ordering,

$$Q = \begin{array}{c} \text{Item} \\ 1-3 \\ 4-6 \\ 7 \\ \dots \end{array} \begin{pmatrix} \alpha^1 & \alpha^2 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ \dots & \dots \end{pmatrix}.$$

To end this section, we point out that the posterior consistency was proved in Theorem 2 of Dunson and Xing (2009), where they showed that the posterior distribution will fall into ϵ -neighborhood of underlying true distribution in terms of L_1 distance almost surely. Thanks to the identifiability results, we have the following consistency result:

Theorem 6. Let $\mathcal{N}_\epsilon(P^0) = \{P \mid \|P - P^0\|_1 < \epsilon\}$ denote a L_1 neighborhood around an arbitrary distribution P^0 and let $\mathcal{N}_\epsilon(\theta^0) = \{\theta \mid \|\theta - \theta^0\|_1 < \epsilon\}$ denote a L_1 neighborhood around an arbitrary parameter θ . Denote $S_{k_i} = \{\pi_i \mid \sum_{k=1}^{k_i} \pi_i^k = 1, \pi_i^k \geq 0\}$ and $S_\infty = \{v \mid \sum v_h = 1, v_h \geq 0\}$. Let θ^* denote true parameter and P^* is true response distribution. Assume the following conditions:

- $Q_i(\mathcal{N}_\epsilon(\pi_i^0)) > 0$ for any $\epsilon > 0$ and $\pi_i^0 \in S_{k_i}$, $i = 1, \dots, I$. Here, Q_i is the prior for π_i^0 .
- $Q_v(\mathcal{N}_\epsilon(v^0)) > 0$ for any $\epsilon > 0$ and $v^0 \in S_\infty$ such that $v_h^0 = 0$ for $h > C$. Here, Q_v is the prior for v^0 .

Then, under each set of conditions of Theorems 1, 2, 3, or 4,

$$P(\theta \in \Theta_c \setminus \mathcal{N}_\epsilon(\theta^*) \mid X) \rightarrow 0 \quad P^* - a.s.$$

for any compact set Θ_c such that $v_h = 0$ for $h > C$.

5. Simulation Study

In this section, simulation studies are conducted to assess the performance of the proposed approach. We consider three models for the data generation: LCDM model, polytomous response model, and a third model containing nonidentifiable parameters. The results are presented assuming all the model parameters, including Q -matrix, attribute distribution, and item response probabilities, are unknown.

TABLE 1.
 Q -matrix and item parameters of the first simulation setting. For notation simplicity, we omit subscript i in table.

Item	Q -matrix					Parameters
						λ
1	1	0	0	0	0	$\lambda_{1,(1)} = 4, \lambda_0 = -2$
2	1	0	0	0	0	$\lambda_{1,(1)} = 4, \lambda_0 = -2$
3	1	0	0	0	0	$\lambda_{1,(1)} = 4, \lambda_0 = -2$
4	0	1	0	0	0	$\lambda_{1,(2)} = 4, \lambda_0 = -2$
5	0	1	0	0	0	$\lambda_{1,(2)} = 4, \lambda_0 = -2$
6	0	1	0	0	0	$\lambda_{1,(2)} = 4, \lambda_0 = -2$
7	0	0	1	0	0	$\lambda_{1,(3)} = 4, \lambda_0 = -2$
8	0	0	1	0	0	$\lambda_{1,(3)} = 4, \lambda_0 = -2$
9	0	0	1	0	0	$\lambda_{1,(3)} = 4, \lambda_0 = -2$
10	0	0	0	1	0	$\lambda_{1,(4)} = 4, \lambda_0 = -2$
11	0	0	0	1	0	$\lambda_{1,(4)} = 4, \lambda_0 = -2$
12	0	0	0	1	0	$\lambda_{1,(4)} = 4, \lambda_0 = -2$
13	0	0	0	0	1	$\lambda_{1,(5)} = 4, \lambda_0 = -2$
14	0	0	0	0	1	$\lambda_{1,(5)} = 4, \lambda_0 = -2$
15	0	0	0	0	1	$\lambda_{1,(5)} = 4, \lambda_0 = -2$
16	1	1	0	0	0	$\lambda_{1,(1)} = 2, \lambda_{1,(2)} = 2, \lambda_{2,(12)} = 0, \lambda_0 = -2$
17	0	1	1	0	0	$\lambda_{1,(2)} = 2, \lambda_{1,(3)} = 2, \lambda_{2,(23)} = 0, \lambda_0 = -2$
18	0	0	1	1	0	$\lambda_{1,(3)} = 2, \lambda_{1,(4)} = 2, \lambda_{2,(34)} = 0, \lambda_0 = -2$
19	0	0	0	1	1	$\lambda_{1,(4)} = 2, \lambda_{1,(5)} = 2, \lambda_{2,(45)} = 0, \lambda_0 = -2$
20	1	0	0	0	1	$\lambda_{1,(1)} = 2, \lambda_{1,(5)} = 2, \lambda_{2,(15)} = 0, \lambda_0 = -2$

5.1. LCDM

We consider a five-binary-attribute LCDM model in the first simulation setting. The Q -matrix and parameters are listed in Table 1. The following sample sizes are considered, $R = 500, 1000, 2000,$ and 4000 . The latent attribute profile is generated as the following mechanisms under two situations.

1. For each respondent, we generate $\theta = (\theta_1, \dots, \theta_5)$ that follows a multivariate normal distribution $N(\mathbf{0}, \Sigma_1)$, where the covariance matrix

$$\Sigma_1 = \begin{pmatrix} 1 & 0.999 & 0 & 0 & 0 \\ 0.999 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0.999 & 0 \\ 0 & 0 & 0.999 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then, the latent attribute profile α is given as $(\alpha^1, \dots, \alpha^5)$ with $\alpha^k = \mathbf{1}(\theta_k > 0)$ for $k = 1, \dots, 5$. In this situation, Attributes 1 and 2, Attributes 3 and 4 are strongly correlated. It means respondents tend to have these attributes pairwise.

2. For each respondent, we generate $\theta = (\theta_1, \dots, \theta_5)$ which follows another multivariate normal distribution $N(\mathbf{0}, \Sigma_2)$, where the covariance matrix

TABLE 2.
RMSE for item parameters under different sample sizes for the LCDM simulation setting.

n	RMSE for λ							
	Case 1				Case 2			
	500	1000	2000	4000	500	1000	2000	4000
Intercept	0.40	0.38	0.12	0.09	0.23	0.17	0.12	0.08
Main	0.60	0.38	0.17	0.13	0.33	0.23	0.17	0.12
Interaction	0.64	0.48	0.25	0.18	0.47	0.33	0.25	0.17

$$\Sigma_2 = \begin{pmatrix} 1 & -, 0.999 & 0 & 0 & 0 \\ -0.999 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -0.999 & 0 \\ 0 & 0 & -0.999 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then, the latent attribute profile α is given as $(\alpha^1, \dots, \alpha^5)$ with $\alpha^k = \mathbf{1}(\theta_k > 0)$ for $k = 1, \dots, 5$. Attributes 1 and 2, Attributes 3 and 4 have strong negative correlations. It means the attributes in each pair are not easily obtained by respondents at the same time.

Under both cases, we can see that there are eight main classes that have large positive class probabilities among all $2^5 = 32$ possible latent attribute profiles and the others are relatively close to zero. That is to say, the true underlying model can be viewed as a eight-class model and it satisfies the identifiability condition of Theorem 1.

We fit the model in Sect. 4. The eight latent classes are successfully identified—the estimated probabilities of the remaining classes are very small (less than $5e-3$). The root-mean-squared errors (RMSEs) of item parameters are listed in Table 2. The RMSEs decrease to zero as sample size increases, indicating the consistency of the Bayesian estimator. Moreover, we provide the attribute classification accuracy in Table 3.

Further, we define CP to be the proportion of items whose partial information structures are estimated correctly,

$$CP = \frac{\#\{i | (\alpha)_i = [\alpha]_i, \alpha \in \mathcal{M}\}}{I}. \quad (6)$$

The results of recovering item partial information structures under various sample sizes are listed in Table 3. In addition, Q -matrix recovery rate is also provided in Table 3 if we assume the LCDM is a restricted latent class model (Xu 2017). These show that the proposed method can recover latent structure very well.

5.2. Polytomous Response Model

Now we consider simulation for a polytomous response DCM model with $I = 15$ items and $A = 3$ attributes. There are $C = 8$ latent classes, $\alpha_1 = (1, 0, 0)$, $\alpha_2 = (0, 1, 0)$, $\alpha_3 = (0, 0, 1)$, $\alpha_4 = (1, 1, 0)$, $\alpha_5 = (1, 0, 1)$, $\alpha_6 = (0, 1, 1)$, $\alpha_7 = (0, 0, 0)$, and $\alpha_8 = (1, 1, 1)$. The response for each item has three categories. The attributes are generated independently from Bernoulli(0.5). The Q -matrix and item response probabilities for each class are specified in

TABLE 3.

Table contains the proportion of correct attribute classification and the proportion of items with correctly estimated item partial information structures and Q -matrix recovery rate under different sample sizes for the first simulation setting.

		Attribute classification accuracy			
		$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$
Case 1	α^1	95.8%	95.7%	97.5%	97.6%
	α^2	95.9%	95.7%	97.5%	97.6%
	α^3	96.5%	96.5%	97.5%	97.7%
	α^4	96.5%	96.5%	97.5%	97.7%
	α^5	93.0%	93.4%	93.7%	93.9%
Case 2	α^1	98.2%	98.3%	98.3%	98.4%
	α^2	98.3%	98.3%	98.3%	98.4%
	α^3	98.4%	98.4%	98.3%	98.4%
	α^4	98.3%	98.4%	98.4%	98.4%
	α^5	94.0%	93.7%	93.9%	93.9%

		Partial information recovery			
		$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$
Case 1	CP	98.5%	99.1%	99.8%	100%
	Q recovery	82%	90%	98%	100%
Case 2	CP	99.2%	99.8%	100%	100%
	Q recovery	94%	98%	100%	100%

TABLE 4.

Q -matrix and item parameters of the second simulation setting.

Item	Q -matrix		Parameters Probability vectors
1	1	0	$\mathbf{p}_1, \mathbf{p}_2$
2	1	0	$\mathbf{p}_1, \mathbf{p}_2$
3	1	0	$\mathbf{p}_1, \mathbf{p}_2$
4	0	1	$\mathbf{p}_1, \mathbf{p}_2$
5	0	1	$\mathbf{p}_1, \mathbf{p}_2$
6	0	1	$\mathbf{p}_1, \mathbf{p}_2$
7	0	0	$\mathbf{p}_1, \mathbf{p}_2$
8	0	0	$\mathbf{p}_1, \mathbf{p}_2$
9	0	0	$\mathbf{p}_1, \mathbf{p}_2$
10	1	1	$\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$
11	1	1	$\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$
12	1	0	$\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$
13	1	0	$\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$
14	0	1	$\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$
15	0	1	$\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$

Here, $\mathbf{p}_1 = (0.8, 0.1, 0.1)$, $\mathbf{p}_2 = (0.1, 0.1, 0.8)$, and $\mathbf{p}_3 = (0.4, 0.3, 0.3)$. If the subject has all required attributes of that item, then the response follows \mathbf{p}_1 . If the subject has none of the required attributes of that item, then the response follows \mathbf{p}_2 . If the subject possesses partial required attribute of that item, then the response follows \mathbf{p}_3 .

Table 4. This model setup satisfies the identifiability condition of Theorem 4. We consider sample sizes $R = 500, 1000, 2000$ and 4000 .

TABLE 5.

RMSE of item response probability, correct proportion of attribute classification and proportions of items with correctly estimated item partial information structures under different sample sizes for the second simulation setting.

RMSE of \mathbf{p}_1 , \mathbf{p}_2 and \mathbf{p}_3				
Item	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$
\mathbf{p}_1	$5e-2$	$3e-2$	$2e-2$	$2e-2$
\mathbf{p}_2	$6e-2$	$4e-2$	$2e-2$	$2e-2$
\mathbf{p}_3	$7e-2$	$5e-2$	$3e-2$	$2e-2$
Attribute classification accuracy				
	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$
α^1	94.9%	97.0%	97.3%	97.4%
α^2	94.7%	97.2%	97.3%	97.4%
α^3	94.8%	97.0%	97.3%	97.4%
Partial information structure recovery				
	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$
CP	96.9%	98.3%	99.8%	100%

We fit the model in Sect. 4 which turns out to identify all latent classes well, and the estimated probabilities beyond the eight largest classes are very small (below $5e-3$). The RMSEs of response probabilities are listed in Table 5. The RMSEs decrease as sample size goes large. It indicates that the proposed estimator is consistent in this setting. Table 5 also provides attribute classification accuracy and item partial information structure recovery results.

5.3. A Nonidentifiable Example

In the third simulation setting, we consider a binary attribute DCM which does not satisfy the identifiability condition. Specifically, we let $I = 6$ and $A = 2$. The first three items measure the first attribute. The fourth item measures the second. The last two items measure both attributes. The specific item parameters and parameterization of latent classes are given in Table 6, and latent class probabilities for four classes are set as $\nu = (1/12, 5/12, 1/12, 5/12)$. Then, we cannot find three sets of items to meet the conditions of Theorem 3. Moreover, it can be checked that there exists the infinite number of sets of parameters leading to the same marginal distribution. For example, we can replace 0.5 and $1/6$ by p_a and p_b , respectively, in Table 6 and set ν to be $(w/2, (1-w)/2, w/2, (1-w)/2)$ as long as $p_a = \frac{bw - \sqrt{w(1-w)(d-b^2)}}{w}$, $p_b = \frac{b(1-w) + \sqrt{w(1-w)(d-b^2)}}{1-w}$ and $(d-b^2)/d < w < (1-b)^2/(d-b^2 + (1-b)^2)$ with $b = 2/9$, $d = 7/108$.

We still let sample size vary from 500 to 4000 and generate datasets for 50 replications from the above setting. We observe that none of these estimates is close to the true parameters. The RMSEs of parameters are provided in Table 7. We can see the errors do not decrease even if sample size increases. This suggests the current setting is nonidentifiable.

6. Real Data Analysis

We apply the proposed method to a subset of the National Epidemiological Survey on Alcohol and Related Conditions (NESARC) (Grant et al. 2003). We extract the subset of items concerning social phobia. There are in total 13 diagnostic questions of binary responses presented in Table 8.

This subset contains 728 white male respondents aged from 25 through 50. These 13 questions are designed according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (American Psychiatric Association, 1994). We fit the latent Dirichlet allocation model and estimate the partial information structure via the procedure in Sect. 4. The results are summarized as follows.

To obtain meaningful and stable estimates, we consider large latent classes whose probabilities are over $\frac{1}{\sqrt{728}} \approx 3.7\%$. According to the fitted model, there are five such latent classes. The estimated posterior probability of each class is $v_1 = 0.37$, $v_2 = 0.18$, $v_3 = 0.14$, $v_4 = 0.12$, and

TABLE 6.

Q -matrix and item parameters of the third simulation setting. The parameterization for latent class is set as: $\alpha_1 = (1, 1)$, $\alpha_2 = (1, 0)$, $\alpha_3 = (0, 1)$ and $\alpha_4 = (0, 0)$.

Item	Q -matrix		Probability for each latent class	
1-3	1	0	$p(Y = 1 \alpha^1 = 1) = 0.88$	$p(Y = 1 \alpha^1 = 0) = 0.12$
4	0	1	$p(Y = 1 \alpha^2 = 1) = 0.5$	$p(Y = 1 \alpha^2 = 0) = 1/6$
5	1	1	$p(Y = 1 \alpha^1 = 0, \alpha^2 = 0) = 1/6$	$p(Y = 1 \alpha^1 = 1, \alpha^2 = 0) = 0.88$ $p(Y = 1 \alpha^1 = 0, \alpha^2 = 1) = 0.5$
6	1	1	$p(Y = 1 \alpha^1 = 0, \alpha^2 = 0) = 0.12$	$p(Y = 1 \alpha^1 = 1, \alpha^2 = 0) = 1/6$ $p(Y = 1 \alpha^1 = 0, \alpha^2 = 1) = 0.12$
			$p(Y = 1 \alpha^1 = 1, \alpha^2 = 1) = 0.5$	

TABLE 7.

Table provides RMSEs of p_1 and p_2 under various sample sizes in Simulation 3.

	RMSEs in simulation 3			
	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$
p_1	0.07	0.06	0.06	0.07
p_2	0.17	0.25	0.27	0.28

Here, parameters p_1 and p_2 denote $p(Y_5 = 1|\alpha^1 = 0, \alpha^2 = 0)$, $p(Y_5 = 1|\alpha^1 = 0, \alpha^2 = 1)$, respectively

TABLE 8.

Content of 13 items for the social anxiety disorder data.

ID	Have you ever had a strong fear or avoidance of
1	Speaking in front of other people?
2	Taking part or speaking in class?
3	Taking part or speaking at a meeting?
4	Performing in front of other people?
5	Being interviewed?
6	Writing when someone watches?
7	Taking an important exam?
8	Speaking to an authority figure?
9	Eating or drinking in front of other people?
10	Having conversations with people you don't know well?
11	Going to parties and social gatherings?
12	Dating?
13	Being in a small group situation?

TABLE 9.

Estimated probability matrix based on latent Dirichlet allocation model for the social anxiety disorder data. Each row corresponding to the item response probability for each class.

Item	α_1	α_2	α_3	α_4	α_5
1	0.92	0.23	0.98	0.90	0.95
2	0.76	0.07	0.97	0.92	0.73
3	0.54	0.02	0.96	0.79	0.62
4	0.56	0.11	0.95	0.82	0.81
5	0.14	0.02	0.71	0.71	0.37
6	0.06	0.06	0.41	0.22	0.10
7	0.18	0.12	0.76	0.69	0.43
8	0.18	0.05	0.77	0.58	0.47
9	0.01	0.02	0.41	0.03	0.08
10	0.17	0.11	0.98	0.42	0.85
11	0.07	0.13	0.90	0.26	0.85
12	0.07	0.03	0.62	0.34	0.39
13	0.01	0.01	0.45	0.08	0.13

TABLE 10.

Estimated cluster matrix based on estimated posterior probability matrix by using k-means method.

Item	(1,0,0)	(0,0,0)	(1,1,1)	(1,1,0)	(1,0,1)	Q -matrix		
1	●	○	●	●	●	1	0	0
2	●	○	†	†	●	1	1	0
3	●	○	†	†	●	1	1	0
4	●	○	†	†	†	1	1	1
5	○	○	●	●	○	0	1	0
6	○	○	●	●	○	0	1	0
7	○	○	●	●	○	0	1	0
8	○	○	●	●	●	0	1	1
9	○	○	●	○	○	1	1	1
10	○	○	●	○	●	0	0	1
11	○	○	●	○	●	0	0	1
12	○	○	●	●	●	0	1	1
13	○	○	†	●	●	0	1	1

Here, we use symbols ○, ●, and † to represent levels 1, 2, and 3, respectively. Level 1 represents the lowest probability level, and level 3 represents the highest probability level. The estimated Q -matrix based on the three-dimensional LCDM model is in the last three columns.

$v_5 = 0.11$. They add up to 92% of the population. The estimated item response probabilities are presented in Table 9.

We apply the K -means method to the item response probabilities of each item to select the number of clusters. The partial information is then obtained via this cluster analysis. The results are summarized in Table 10. We can see that 13 items may be divided into three groups according to their functioning. Items 1–4 can differentiate between Classes 1, 3, 4, 5 and Class 2. Items 5–8 differentiate Classes 1, 2 and Classes 3,4. Items 9, 10, 11, 13 differentiate Class 3 and Classes 1, 2, 4. Furthermore, we can see that Items 2, 3, 4, 13 differentiate multiple groups, indicating that these items are more informative.

TABLE 11.
Table of parameters for the three-dimensional LCDM model for the social anxiety disorder data.

Item	Estimated parameters (SD)				
1	$\lambda_0 : -1.18 (0.28)$	$\lambda_{1,(1)} : 3.80 (0.35)$			
2	$\lambda_0 : -2.64 (0.54)$	$\lambda_{1,(1)} : 3.76 (0.57)$	$\lambda_{1,(2)} : 1.72 (0.43)$		
3	$\lambda_0 : -3.95 (1.38)$	$\lambda_{1,(1)} : 4.17 (1.40)$	$\lambda_{1,(2)} : 1.80 (0.35)$		
4	$\lambda_0 : -2.10 (0.31)$	$\lambda_{1,(1)} : 2.33 (0.35)$	$\lambda_{1,(2)} : 1.65 (0.56)$	$\lambda_{1,(3)} : 1.65 (0.48)$	$\lambda_{2,(2,3)} : -1.65 (0.88)$
5	$\lambda_0 : -1.76 (0.17)$	$\lambda_{1,(2)} : 2.65 (0.28)$			
6	$\lambda_0 : -2.57 (0.17)$	$\lambda_{1,(2)} : 1.82 (0.28)$			
7	$\lambda_0 : -1.35 (0.13)$	$\lambda_{1,(2)} : 2.32 (0.26)$			
8	$\lambda_0 : -1.87 (0.17)$	$\lambda_{1,(2)} : 2.36 (0.38)$	$\lambda_{1,(3)} : 2.36 (0.38)$	$\lambda_{2,(2,3)} : -2.36 (0.60)$	
9	$\lambda_0 : -3.70 (0.36)$	$\lambda_{3,(1,2,3)} : 3.30 (0.43)$			
10	$\lambda_0 : -1.41 (0.14)$	$\lambda_{1,(3)} : 3.87 (0.57)$			
11	$\lambda_0 : -2.00 (0.16)$	$\lambda_{1,(3)} : 3.97 (0.44)$			
12	$\lambda_0 : -2.77 (0.24)$	$\lambda_{1,(2)} : 2.61 (0.43)$	$\lambda_{1,(3)} : 2.61 (0.41)$	$\lambda_{2,(2,3)} : -2.61 (0.62)$	
13	$\lambda_0 : -4.87 (0.54)$	$\lambda_{1,(2)} : 2.70 (0.80)$	$\lambda_{1,(3)} : 2.70 (0.75)$	$\lambda_{2,(2,3)} : -0.74 (1.00)$	

We further construct a parameterization of the item response function based on the estimated partial information structure through using LCDM with $K = 3$ attributes. The estimated partial information structure indicates that we could parameterize each class when the number of attributes equals three, specifically, $\alpha_1 = (1, 0, 0)$, $\alpha_2 = (0, 0, 0)$, $\alpha_3 = (1, 1, 1)$, $\alpha_4 = (1, 1, 0)$, and $\alpha_5 = (1, 0, 1)$. Then, an estimated Q -matrix and item parameters under this parameterization are provided in Tables 10 and 11, respectively. The identifiability conditions of Theorem 3 hold. In particular, for $C = 5$, we identify $I_1 = (1, 5, 9)$, $I_2 = (2, 6, 10)$, $I_3 = (3, 7, 11)$ and their T -matrices are of full rank. We also perform a model goodness-of-fit test and consider the maximized log-likelihood as the test statistic that has a p value equal to 0.432. It suggests that the model fits the data reasonably well.

By comparing the results from Iza et al. (2014), Attribute 1 appears to be associated with “public performance,” Attribute 2 with “close scrutiny,” and Attribute 3 with “interaction”. Further, we calculate the attribute mastery proportions. There are around 79% of people who have the first attribute. Around 28% of people who have the second attribute. About 27% of people who possess the third attribute. Based on latent class parameterization, we can see that most people are afraid of public performances. People who are fear of close scrutiny or interaction also suffer from public performance, which indicates the evidence of hierarchical structure of these three attributes.

Table 9 presents posterior mean of the response probability for each item under each class. Based on the loadings in this table, we may interpret these latent classes as follows. Class 1 has high response probabilities of items 1–4. This shows that it possesses Attribute 1 only, i.e., those people in this class are afraid of public performance, but not close scrutiny and interaction. Class 2 is loosely associated with all items without any strong signals, indicating it may be connected to none of attributes. Class 3 is strongly associated with all items. Hence, people from Class 3 are likely to possess 3 attributes. In other words, Class 3 corresponds to, using a technical term, the generalized social anxiety disorder subtype (“fears most social situations”). Class 4 is strongly associated with items 1–4 which are related to public performance and items 5–8 which are related to close scrutiny. Thus, this group is characterized by strong fear of public performance and close scrutiny. Finally, Class 5 has relatively high response probabilities of items 1–4 and 9–13 and relatively low response probability of items 5–8, which means that this class is more likely related to Attributes 1 and 3. In other words, people from Class 5 may have “fears” of public performance and interaction with other people.

7. Discussion

This paper concerns identifiability issues of DCMs. Model identifiability is a fundamental question that ensures the feasibility of parameter estimation and subsequent statistical inference. It is also a long-standing question for latent variable models mostly due to the fact that some variables are not directly observed. In this paper, we establish identifiability results for DCM via four theorems that cover various model settings including binary response, multi-categorical response, and multi-category attribute. We provide both general results and easy-to-check conditions that are applicable to a large class of DCMs including well-known core models, DINA, DINO, LCDM, etc. Following the model parameter identifiability, we also discuss the partial information structure implied by the Q -matrix that is another key quantity in the model specification. We show that item partial information structure can be consistently estimated if the model parameters are identifiable. We also provide some examples showing the reconstruction of Q -matrix based on partial information structure when additional information about model is available.

To apply the theories under a specific model framework, we consider a latent class model with infinitely many latent classes regularized by a Dirichlet process prior and use slice Gibbs sampler to get nonparametric Bayesian estimates for item parameters. Further, we propose an estimator of the partial information structure via the K -means method applied to the estimated item response probabilities.

Simulation results show the proposed method performs well under a variety of settings and consistently recovers model parameters with reasonable sample sizes. We also show via simulations that the model may not be consistently identified if some of the conditions are violated. Real data analysis is also provided as an illustration. In this analysis, we start with an exploratory analysis via the infinite mixture model and reconstructed the partial information structure largely matching the understanding of the items.

Many works (De La Torre 2011; Xu and Zhang 2016; Chiu et al. 2009) deployed special DCMs like DINA or reduced NC-RUM for fitting data under a known Q -matrix as a priori information. However, it suffers the problem that mis-specification of Q -matrix may lead to errors in parameter estimation and attribute classification. In addition, directly fitting a simple DCM may lead to bad interpretations. In contrast, our work is featured by exploratory analysis starting from general latent class model and does not require a pre-specification of the number of latent classes, which avoids the number of classes strictly being 2^A . Therefore, the results could be interpreted at a very general level. Moreover, the induced partial information structure is data driven, which provides a benchmark for examining the differentiability of each item. This helps test developers to re-check the expert-specified Q -matrix. Hence, a systematic tool based on the current method for checking Q -matrix could be developed in the future.

Supplementary Material

Supplementary Material is available online, including the technical proofs for Theorem 1–6 and details of computation procedure.

Acknowledgments

This research is supported in part by NSF IIS-1633360 and SES-1826540.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Allman, E. S., Matias, C., & Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, *37*, 3099–3132.
- Chen, Y., Culppepper, S. A., Chen, Y., & Douglas, J. (2018). Bayesian estimation of the DINA Q matrix. *Psychometrika*, *83*, 89–108.
- Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015a). Statistical analysis of Q -matrix based diagnostic classification models. *Journal of the American Statistical Association*, *110*, 850–866.
- Chen, Y., Liu, J., & Ying, Z. (2015b). Online item calibration for Q -matrix in CD-CAT. *Applied Psychological Measurement*, *39*, 5–15.
- Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, *74*, 633.
- De La Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of educational and behavioral statistics*, *34*, 115–130.
- De La Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*, 179–199.
- De La Torre, J., & Douglas, J. A. (2004). Higher order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333–353.
- DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q -matrix. *Applied Psychological Measurement*, *35*, 8–26.
- DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In S. F. Chipman, P. D. Nichols, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361–389). Hillsdale, NJ: Erlbaum.
- Dunson, D., & Xing, C. (2009). Nonparametric Bayes modeling of multivariate categorical data. *Journal of the American Statistical Association*, *104*, 1042–1051.
- Grant, B. F., Kaplan, K., Shepard, J., & Moore, T. (2003). *Source and accuracy statement for wave 1 of the 2001–2002 national epidemiologic survey on alcohol and related conditions*. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.
- Gu, Y., & Xu, G. (2018). The sufficient and necessary condition for the identifiability and estimability of the DINA model. *Psychometrika*. <https://doi.org/10.1007/s11336-018-9619-8>.
- Gu, Y., Liu, J., Xu, G., & Ying, Z. (2018). Hypothesis testing of the Q -matrix. *Psychometrika*, *83*, 515–537.
- Hartz, S. M. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality*. Doctoral Dissertation, University of Illinois, Urbana-Champaign.
- Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*, 191–210.
- Iza, M., Wall, M., Heimberg, R., Rodebaugh, T., Schneier, F., Liu, S.-M., et al. (2014). Latent structure of social fears and social anxiety disorders. *Psychological medicine*, *44*, 361–370.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*, 258–272.
- 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.
- Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy model for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement*, *41*, 205–237.
- Liu, J., Xu, G., & Ying, Z. (2012). Data-driven learning of Q -matrix. *Applied Psychological Measurement*, *36*, 609–618.
- Liu, J., Xu, G., & Ying, Z. (2013). Theory of the self-learning Q -matrix. *Bernoulli: Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, *19*, 1790.
- Roussos, L. A., Templin, J. L., & Henson, R. A. (2007). Skills diagnosis using IRT-based latent class models. *Journal of Educational Measurement*, *44*, 293–311.
- Rupp, A. A., & Templin, J. (2008a). The effects of Q -matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, *68*, 78–96.
- Rupp, A. A., & Templin, J. L. (2008b). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspective*, *6*, 219–262.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York: Guilford Press.
- Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, *4*, 639–650.
- Stout, W. (2007). Skills diagnosis using IRT-based continuous latent trait models. *Journal of Educational Measurement*, *44*, 313–324.
- Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics*, *12*, 55–73.
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. Boca Raton: CRC Press.
- Templin, J., He, X., Roussos, L., & Stout, W. (2003). *The pseudo-item method: a simple technique for analysis of polytomous data with the fusion model*. External diagnostic research group technical report.
- Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*, 287–305.
- Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge: Cambridge University Press.
- von Davier, M. (2005). *A general diagnosis model applied to language testing data*. Research report: Educational testing service.

- Vonesh, E. F., & Chinchilli, V. G. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. London: Chapman and Hall.
- Walker, S. G. (2007). Sampling the dirichlet mixture model with slices. *Communications in Statistics-Simulation and Computation*, 36, 45–54.
- Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *The Annals of Statistics*, 45, 675–707.
- Xu, G., & Shang, Z. (2018). Identifying latent structures in restricted latent class models. *Journal of the American Statistical Association*, 113(523), 1284–1295.
- Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika*, 81, 625–649.

Manuscript Received: 22 FEB 2017

Published Online Date: 23 JAN 2019