

Statistical Inference for Diagnostic Classification Models

Gongjun Xu

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2013

©2013

Gongjun Xu

All Rights Reserved

ABSTRACT

Statistical Inference for Diagnostic Classification Models

Gongjun Xu

Diagnostic classification models (DCM) are an important recent development in educational and psychological testing. Instead of an overall test score, a diagnostic test provides each subject with a profile detailing the concepts and skills (often called attributes) that he/she has mastered. Central to many DCMs is the so-called Q -matrix, an incidence matrix specifying the item-attribute relationship. It is common practice for the Q -matrix to be specified by experts when items are written, rather than through data-driven calibration. Such a non-empirical approach may lead to misspecification of the Q -matrix and substantial lack of model fit, resulting in erroneous interpretation of testing results. This motivates our study and we consider the identifiability, estimation, and hypothesis testing of the Q -matrix. In addition, we study the identifiability of diagnostic model parameters under a known Q -matrix.

The first part of this thesis is concerned with estimation of the Q -matrix. In particular, we present definitive answers to the learnability of the Q -matrix for one

of the most commonly used models, the DINA model, by specifying a set of sufficient conditions under which the Q -matrix is identifiable up to an explicitly defined equivalence class. We also present the corresponding data-driven construction of the Q -matrix. The results and analysis strategies are general in the sense that they can be further extended to other diagnostic models.

The second part of the thesis focuses on statistical validation of the Q -matrix. The purpose of this study is to provide a statistical procedure to help decide whether to accept the Q -matrix provided by the experts. Statistically, this problem can be formulated as a pure significance testing problem with null hypothesis $H_0 : Q = Q_0$, where Q_0 is the candidate Q -matrix. We propose a test statistic that measures the consistency of observed data with the proposed Q -matrix. Theoretical properties of the test statistic are studied. In addition, we conduct simulation studies to show the performance of the proposed procedure.

The third part of this thesis is concerned with the identifiability of the diagnostic model parameters when the Q -matrix is correctly specified. Identifiability is a prerequisite for statistical inferences, such as parameter estimation and hypothesis testing. We present sufficient and necessary conditions under which the model parameters are identifiable from the response data.

Contents

List of Tables	iv
List of Figures	v
Acknowledgments	vi
Chapter 1 Introduction	1
1.1 Background	1
1.1.1 Classical test theory	2
1.1.2 Item response theory	3
1.2 Diagnostic classification models	9
1.2.1 Notation	11
1.2.2 Conjunctive diagnostic models	13
1.2.3 Compensatory diagnostic models	16
1.3 Main results of the thesis	16
Chapter 2 Estimation of Q-matrix	19
2.1 Introduction	19
2.2 Estimation of Q -matrix	21
2.2.1 An illustration example	24

2.2.2	Main results for ideal responses	27
2.3	DINA model with known slipping and guessing parameters	31
2.3.1	A general T -matrix	31
2.3.2	Estimation of the Q -matrix and consistency results	33
2.4	DINA with unknown slipping probabilities	35
2.5	DINA with unknown slipping and guessing	40
2.6	Simulation	42
2.6.1	Estimation of the Q -matrix with no special structure	42
2.6.2	Estimation of the Q -matrix with partial information	49
2.7	Discussion	50
2.8	Proofs of theorems	55
2.8.1	Proof of theorems	55
2.8.2	Several propositions and lemmas	59
2.9	Technical proofs	65
Chapter 3 Hypothesis Testing of Q-matrix		72
3.1	Introduction	72
3.2	Q -matrix validation procedure	74
3.2.1	Notation	74
3.2.2	Validation procedure	77
3.2.3	Computations	84
3.3	Distribution of test statistic $S(Q_0)$	86
3.3.1	Case of known item parameters	87
3.3.2	Case of unknown item parameters	90
3.3.3	An example: DINA Model	95
3.4	Simulation study	96

3.4.1	DINA model with known item parameters	97
3.4.2	DINA model with unknown model parameters	98
3.5	Appendix	102
Chapter 4 Identifiability of DCM model parameters		105
4.1	Introduction	105
4.2	Main results	107
4.2.1	Notation	107
4.2.2	Identifiability conditions	109
4.3	Proof of theorems	113
Bibliography		130

List of Tables

2.1	Estimation of Q -matrix with uniformly distributed attribute profiles.	44
2.2	Estimation of Q -matrix with uniformly distributed attribute profiles.	48
2.3	Estimation of Q -matrix with non-uniformly distributed attribute profiles.	49
2.4	Estimation of Q -matrix with partial information.	50
3.1	Validation of Q -matrix with known slipping and guessing parameters in the DINA model	98
3.2	Validation of Q -matrix with unknown slipping and guessing parameters	100
3.3	Validation of Q -matrix with unknown slipping and guessing parameters	100
3.4	Test power of the Q -matrix validation procedure	101

List of Figures

1.1	Item characteristic curves for the Rasch model.	4
1.2	Item characteristic curves for the two parameter logistic model.	6
1.3	Item characteristic curves for the three parameter logistic model.	8
1.4	Item characteristic curves for the DINA model with $K = 1$	15
2.1	Results of a simulated data set, for which the estimated Q-matrix does not pass the true one.	47
2.2	Results of a simulated data set, for which the estimated Q-matrix passes the true one but does not converge to it.	47

Acknowledgments

I would like to express my deeply-felt thanks to my Ph.D. advisors Prof. Zhiliang Ying and Prof. Jingchen Liu. This thesis would not have been possible without their constant guidance, support, and encouragement during my graduate study. Their advice and help is always invaluable on both an academic and a personal level, for which I am truly indebted and thankful.

I am grateful to Prof. Lawrence T. DeCarlo, Dr. Xiaolong Luo, Prof. Bodhisattva Sen, and Prof. Tian Zheng for their helpful suggestions and for serving on my doctoral oral/thesis committee.

I am very thankful to all the faculty and staff in the Department of Statistics at Columbia University for their support and assistance throughout the last five years. I am obliged to many of my friends for their enthusiasm, encouragement, and inspiration.

I especially thank my parents and grandparents for their constant care and support. Finally, I owe my special thanks to my wife, Liuyan Zhao, for her unwavering support and love.

To my family

Chapter 1

Introduction

1.1 Background

Measurement theory deals with assigning numbers to subjects in a systematic way to represent their properties (Allen and Yen 1979). Our daily experience is to use certain physical device to assign such numbers. For instance, we use thermometers to measure temperature and directly obtain measurement values from them. *Educational and psychological measurement* focuses on people's latent traits, for instance, a high school student's mathematical ability or a person's intelligence level, and its primary goal is to numerically assess these unobservable latent traits. Unlike measurement based on physical devices, such as thermometers and speedometers, from which we can obtain direct measurement values, educational and psychological measurement uses *tests* as the main measurement tool and uses people's responses to infer their latent traits indirectly.

Just as in any type of measurement, a key issue in educational and psychological measurement is to control and adjust measurement error and make valid inference of

the latent traits from the response data. To adjust the measurement error, different statistical procedures have been developed in the literature. Below we give a brief introduction to two widely used test theories: classical test theory (Spearman 1904; Novick 1966) and item response theory (Lord and Novick 1968; Lord 1980).

1.1.1 Classical test theory

Classical test theory (CTT) is one of the earliest formalizations of measurement theory for educational and psychological tests. It was put forward in the pioneering work of Spearman (1904); see also Novick (1966); Lord and Novick (1968); Allen and Yen (1979) for more details. CTT borrows concepts, such as measurement error, from measurement in the physical sciences (Mislevy 1996). CTT decomposes the observed test score into a true score term and an error term, and assumes that response variations are due to variation in subjects' ability (the true score), which is of our interest, and variation from other external sources (the measurement error), such as rater or instrument error. Specifically, let X be a subject's observed score for a test. The classical test model assumes that

$$X = T + e,$$

where T is the subject's true score defined as the expected value of X and e is the measurement error term, which is assumed to be independent of T . The true score T is the latent variable that we are interested in.

Note that CTT combines all sources of error into one term e and that the model equation $X = T + e$ is essentially the linear equation of the one-way analysis of variance (ANOVA) model with a random factor. In this line of research, people further decompose the variance of the error term into variances from different sources (factors); this decomposition corresponds to the multi-way ANOVA. This approach

is known as generalizability theory in the literature (Cronbach, Nageswari and Gleser 1963); see also Shavelson (1991); Brennan (2001) for more details.

1.1.2 Item response theory

Item response theory (IRT) models were developed in the 1950s and 1960s for the purpose of analyzing test items that were dichotomously scored. Unlike the classical test theory, which focuses on subjects' total scores calculated using the whole test form, IRT deals with subjects' responses to each test item. Consider a binary response test. Let $R_{i,j}$ be the j th subject response to the i th item, where $R_{i,j} = 1$ if the answer is correct and 0 otherwise. IRT uses a latent parameter $\theta \in (-\infty, +\infty)$ to denote a subject's ability, and represents the correct response probability to the i th item by a mathematical function $p_i(\theta) := P(R_{i,j} = 1|\theta)$. This function is known as the item response function, or item characteristic curve (Tucker 1946). The item response function $p_i(\theta)$, $\theta \in (-\infty, +\infty)$, is a non-linear function in θ due to the restriction that $0 \leq p_i(\theta) \leq 1$. In addition, it is usually a monotonically increasing function in θ , meaning that subjects with higher ability have larger probabilities of obtaining correct responses to item i .

Different item response functions (models) have been proposed in the literature to model subjects' response data (Rasch 1960; Lord and Novick 1968; Lord 1980). Below we give a brief introduction to several of the commonly used models. More details about the corresponding statistical issues, such as item and ability parameters estimation and goodness of fit, can be found in van der Linden and Hambleton (1996); Embretson and Reise (2000); Baker (2001); Baker and Kim (2004); Reckase (2009); Fox (2010).

Rasch Model. The Rasch model (see Rasch 1960; 1961) assumes that the item

response function takes the following form:

$$p_i(\theta) = \frac{1}{1 + \exp\{-(\theta - b_i)\}},$$

where b_i is the difficulty parameter for item i . Since the Rasch model uses the logistic link function and has only a single item parameter for each item, it is also known as the one-parameter logistic (1-PL) model. The parameter b_i is also called the location parameter and shows where the item response function achieves its central value between its lower and upper asymptotes. Figure 1.1 gives three item characteristic curves for the Rasch model with different item difficulties $b_i = -1, 0, 1$. We can see that when $\theta = b_i$, the correct response probability takes value $1/2$. In addition, larger values of b_i indicate more difficult items, with smaller success probabilities given the same ability level θ , while smaller values of b_i indicate the reverse.

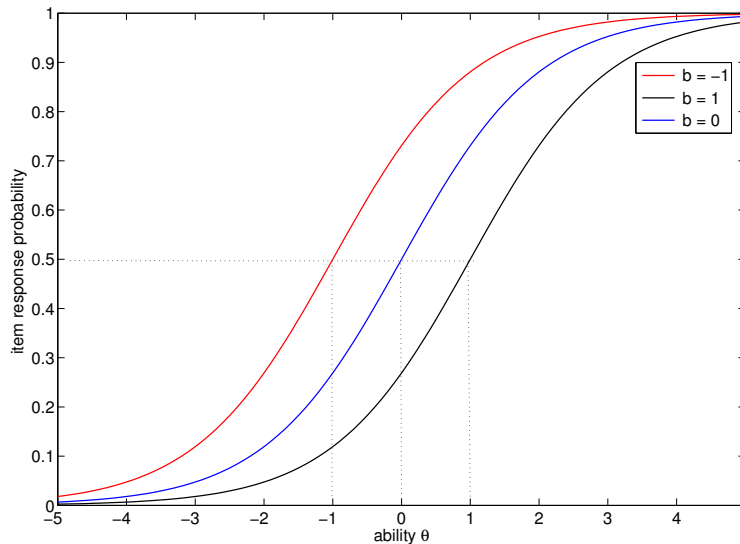


Figure 1.1: Item characteristic curves for the Rasch model.

For the Rasch model, we need to estimate both item parameters b and the ability

parameters θ . Consider the likelihood function of N subjects' response data for n items. It takes the form of

$$\begin{aligned}
 L(\theta, b; \mathbf{R}) &= \prod_{j=1}^N \prod_{i=1}^n p_i(\theta_j)^{R_{i,j}} (1 - p_i(\theta_j))^{1-R_{i,j}} \\
 &= \prod_{j=1}^N \prod_{i=1}^n \left(\frac{1}{1 + \exp\{-(\theta_j - b_i)\}} \right)^{R_{i,j}} \left(\frac{\exp\{-(\theta_j - b_i)\}}{1 + \exp\{-(\theta_j - b_i)\}} \right)^{1-R_{i,j}} \\
 &= \frac{\prod_{j=1}^N \exp\{\sum_{i=1}^n R_{i,j} \theta_j\} \prod_{i=1}^n \exp\{-\sum_{j=1}^N R_{i,j} b_i\}}{\prod_{j=1}^N \prod_{i=1}^n (1 + \exp\{(\theta_j - b_i)\})}.
 \end{aligned}$$

We can see that under the Rasch model, the subject's total score $\sum_{i=1}^n R_{i,j}$ is a sufficient statistic for the ability parameter θ_j and the item score $\sum_{j=1}^N R_{i,j}$ is a sufficient statistic for the difficulty parameter b_i . Therefore, when estimating item parameters, ability parameters can be eliminated through conditioning on the subjects' raw sum scores. This estimation procedure is called conditional maximum likelihood (CML) estimation (see e.g. Liou 1994). As an alternative to the CML approach, estimates of the item parameters can be found through maximization of the marginal likelihood function (MML) (see e.g. Thissen 1982) by integrating the likelihood function over a density function for θ , i.e.,

$$L(b) = \int_{-\infty}^{\infty} L(\theta, b; \mathbf{R}) f(\theta; \mu, \sigma) d\theta, \quad (1.1)$$

where $f(\theta; \mu, \sigma)$ is usually taken as a normal density function with mean $\mu = 0$ and standard deviation σ . See Baker and Kim (2004) for more details about parameter estimation.

Two Parameter Logistic Model. A limitation of the Rasch model is that all items only differ in item difficulty, which indicates that all items discriminate subjects in a similar way, i.e., all item characteristic curves have the same shape (see Figure 1.1).

From a practical point of view, it is desirable to parameterize both item difficulty and item discrimination. Birnbaum proposed the two parameter logistic model (2-PL model) in the chapters he contributed to Lord and Novick (1968). The item response function for the 2-PL model is given by

$$p_i(\theta) = \frac{1}{1 + \exp\{-a_i(\theta - b_i)\}},$$

where a_i is the discrimination parameter and b_i is the difficulty parameter. The parameter a_i is related to the maximum slope of the item characteristic curve and indicates how well an item discriminates among subjects. When $a_i = 0$, the 2-PL model is equivalent to the Rasch model. See Figure 1.2 for an example of two item characteristic curves following 2-PL model corresponding to two different discrimination levels ($a_i = 1$ and 2) and the same difficulty level ($b_i = 0$).

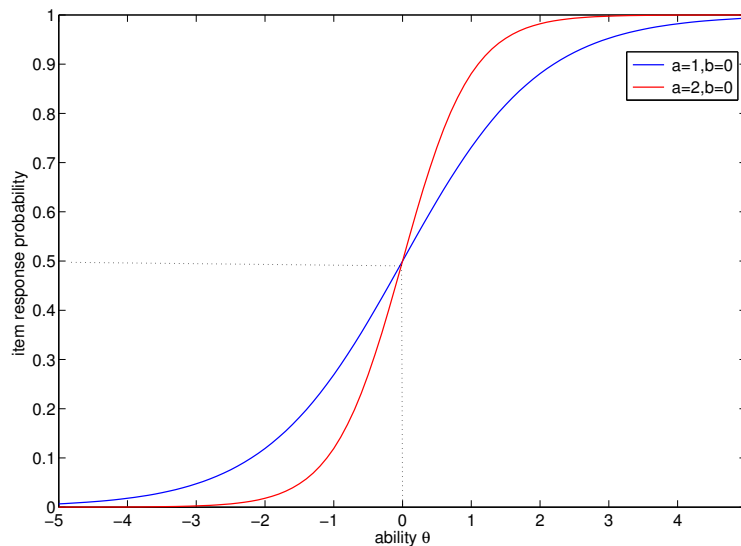


Figure 1.2: Item characteristic curves for the two parameter logistic model.

The likelihood function of the two parameter logistic model takes the following

form:

$$\begin{aligned} L(\theta, b) &= \prod_{j=1}^N \prod_{i=1}^n p_i(\theta_j)^{R_{i,j}} (1 - p_i(\theta_j))^{1-R_{i,j}} \\ &= \frac{\exp\{\sum_{j=1}^N \sum_{i=1}^n R_{i,j} a_i(\theta_j - b_i)\}}{\prod_{j=1}^N \prod_{i=1}^n (1 + \exp\{a_i(\theta_j - b_i)\})}. \end{aligned}$$

Unlike in the Rasch model, the total scores here are no longer sufficient statistics for the ability parameters. As a result, the conditional maximum likelihood estimation approach is no longer possible. Bock and Lieberman (1970) and Bock and Aitken (1981) developed an estimation procedure based on the marginal likelihood function for the two-parameter model. The item parameters are estimated from the marginal distribution by integrating the likelihood function over the ability distribution as in equation (1.1).

Three Parameter Logistic Model. The three parameter logistic model (3-PL) introduces an additional item parameter c to capture the “guessing” probability in the multiple-choice items; see Lord and Novick (1968); Lord (1980). This model has the following item response function:

$$p_i(\theta) = c_i + (1 - c_i) \frac{1}{1 + \exp\{-a_i(\theta - b_i)\}},$$

where a_i is the discrimination parameter, b_i is the difficulty parameter, and c_i is the guessing parameter. When $c_i = 0$, the 3-PL model is equivalent to the 2-PL model. Figure 1.3 gives an example of two item response functions following the 3-PL model with $(a_i, b_i, c_i) = (1, 0, 0.2)$ and $(a_i, b_i, c_i) = (1, 0, 0)$. We can see that the parameter c_i equals the smallest probability of correctly answering the i th item, and when $\theta = b_i$, the item response function takes its central value between its lower asymptote (c_i)

and upper asymptote (1), i.e., $p_i(b_i) = (1 - c_i)/2$. For the 3-PL model, maximization of the marginal likelihood function approach is usually used to estimate the model parameters a , b and c . See Baker and Kim (2004) for more details.

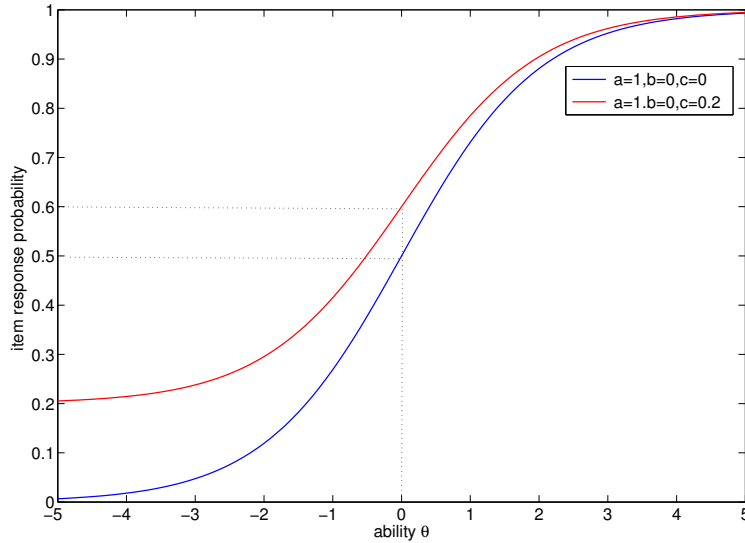


Figure 1.3: Item characteristic curves for a three parameter logistic model.

Besides the logistic models introduced above, other item response functions have also been proposed in the literature. A widely used alternative is normal-ogive models (Lord 1952; Lord and Novick 1968). A two parameter normal-ogive model takes the following form:

$$p_i(\theta) = \int_{-\infty}^{a_i(\theta-b_i)} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx,$$

where a_i is the slope parameter and b_i is the difficulty parameter as in the logistic IRT models. Moreover, researchers have also proposed nonparametric modeling approaches (Guttman 1947; van der Linden and Hambleton 1996).

In addition to the above models which focus on a unidimensional latent variable and dichotomous response data, other developments in item response theory include polytomous response models (Samejima 1969; Bock 1972; van der Linden and Hambleton 1996; Embretson and Reise 2000; Ostini and Nering 2006), multidimensional IRT models (Lord and Novick 1968; McDonald 1967; Samejima 1974; van der Linden and Hambleton 1996; Reckase 2009), and so on.

1.2 Diagnostic classification models

Cognitive diagnosis has recently gained prominence in educational assessment, psychiatric evaluation, and many other disciplines (Rupp and Templin 2008b; Rupp, Templin and Henson 2010). Instead of an overall test score, a cognitive diagnostic test provides each subject with a profile detailing the concepts and skills (often called attributes) that he/she has mastered. Take the PSAT/NMSQTTM (Preliminary SAT/National Merit Scholarship Qualifying Test) as an example. The College Board currently provides a *Score Report PlusTM* to each student, which not only tells his/her total scores for the test subjects including math, writing, and reading, but also provides more detailed information about various skills in each subject such as probability, statistics, algebra, etc. This is the first nationally standardized test to give diagnostic skills-based feedback (Roussos, Templin and Henson 2007b). Such feedback could have a significant impact on learning process by providing students and teachers detailed information on students' strengths and weaknesses.

Traditional test theory, such as the classical test theory and the unidimensional item response theory models introduced in Section 1.1, mainly focuses on scaling and ranking students on a latent proficiency continuum. However, the overall information (such as test scores or proficiency estimators) obtained based on such models is one

dimensional and it does not contain enough information for the multiple targeted skills or attributes in terms of designing effective instruction and providing intervention. As an alternative, diagnostic classification models, which are also known as cognitive diagnostic models (Rupp *et al.* 2010), have been developed for the purpose of identifying the presence or absence of multiple fine-grained skills or attributes.

Statistically speaking, diagnostic classification models belong to the family of latent structure models. In particular, they are restricted latent class models within the broader family of generalized linear and nonlinear mixed models (von Davier, 2009). Such restrictions are usually based on the so-called Q -matrix (Tatsuoka 1983), which specifies the relationship between test items and latent attributes. A short list of diagnostic classification models based on the Q -matrix includes the conjunctive (noncompensatory) DINA and NIDA models (Junker and Sijtsma 2001; Tatsuoka 2002; de la Torre and Douglas 2004; Templin 2006; de la Torre 2008b; DeCarlo 2011), the reparameterized unified/fusion model (RUM) (DiBello, Stout and Roussos 1995; Hartz 2002; Templin, He, Roussos and Stout 2003), the compensatory DINO and NIDO models (Templin and Henson 2006; Templin 2006), the rule space method (Tatsuoka 1985; 2009), the attribute hierarchy method (Leighton, Gierl and Hunka 2004), and nonparametric clustering method (Chiu, Douglas and Li 2009); see also (Junker 1999; von Davier 2005; Rupp *et al.* 2010) for more developments and approaches to cognitive diagnosis.

In the following, we give a detailed description of the diagnostic classification models. See also Rupp *et al.* (2010) for a recent review.

1.2.1 Notation

This thesis is concerned with N subjects taking a test consisting of J items. The responses are binary, so that the data will be an $N \times J$ matrix with entries being 0 or 1. The diagnostic classification model to be considered for such data envisions K attributes that are related to both the subjects and the items. Throughout this thesis, we assume that the number of attributes K is known and that the number of items J is fixed.

The following notation and specifications are needed to describe the diagnostic classification models.

Q -matrix The Q -matrix is the key component of diagnostic models. It specifies the link between the test items and the latent attributes. In particular, $Q = (q_{jk})_{J \times K}$ is a $J \times K$ matrix with binary entries. For each j and k , $q_{jk} = 1$ indicates that item j requires attribute k and $q_{jk} = 0$ otherwise.

We take the following 3×2 Q -matrix as an example.

$$\begin{array}{rcc}
 & \hline \hline
 & \text{addition} & \text{multiplication} \\
 \hline
 Q = & \begin{array}{cc}
 2 + 3 & 1 & 0 \\
 5 \times 2 & 0 & 1 \\
 (2 + 3) \times 2 & 1 & 1
 \end{array} & (1.2)
 \end{array}$$

There are two attributes and three items. The first item requires addition, the second item requires multiplication, and the third one requires both addition and multiplication skills.

Responses to items We use

$$\mathbf{R} = (R^1, \dots, R^J)^\top$$

to denote the vector of responses to the J test items. In this thesis, we focus on binary responses. For each j , R^j is a binary variable taking 0 or 1, and superscript “ \top ” denotes transpose. For the example in (1.2), $J = 2$ and there are $2^2 = 4$ possible response vectors.

Attribute profile There are K attributes and we use

$$\boldsymbol{\alpha} = (\alpha^1, \dots, \alpha^K)^\top$$

to denote the vector of attributes, where $\alpha^k = 1$ or 0, indicating the presence or absence of the k th attribute, $k = 1, \dots, K$. For the example in (1.2), $K = 2$ and all the possible attributes profiles are $(0, 0)$, $(0, 1)$, $(1, 0)$, and $(1, 1)$, where the first element represents addition and the second one represents multiplication.

Note that both $\boldsymbol{\alpha}$ and \mathbf{R} are subject-specific. We use subscripts to indicate different subjects. For instance, $\mathbf{R}_i = (R_i^1, \dots, R_i^J)^\top$ is the response vector for subject i . Similarly, $\boldsymbol{\alpha}_i = (\alpha_i^1, \dots, \alpha_i^K)^\top$ is the attribute vector for subject i .

With N subjects, we observe $\mathbf{R}_1, \dots, \mathbf{R}_N$ but not $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N$. The primary purpose of cognitive diagnosis is to accurately estimate attribute profile $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N$ from the response data $\mathbf{R}_1, \dots, \mathbf{R}_N$. Sometimes, we are also interested in the proportion of subjects with certain attribute profiles, which is specified as follows.

Population proportion We assume that the attribute profiles are i.i.d. and denote the proportion of subjects with attribute profile $\boldsymbol{\alpha}$ by

$$p_{\boldsymbol{\alpha}} = P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}).$$

We write $\mathbf{p} = (p_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \{0, 1\}^K)^\top$. For statement simplicity, we let the first element of \mathbf{p} be

$$p_{\mathbf{0}} = P(\boldsymbol{\alpha}_i = (0, \dots, 0)_{K \times 1}^\top),$$

and define the other part of \mathbf{p} as \mathbf{p}^* , i.e.,

$$\mathbf{p} = \begin{pmatrix} p_0 \\ \mathbf{p}^* \end{pmatrix}. \quad (1.3)$$

We use vector $\mathbf{0}_K$ and $\mathbf{1}_K$ to denote the K dimensional zero and one column vectors respectively. When there is no ambiguity, we omit the index of length and only write $\mathbf{0}$ and $\mathbf{1}$.

Ideal response Let $\xi^j(\boldsymbol{\alpha}, Q)$ denote the ideal response, which indicates whether a subject possessing attribute profile $\boldsymbol{\alpha}$ is capable of providing a positive response to item j if the item-attribute relationship is specified by matrix Q .

Different ideal response structures give rise to different diagnostic classification models. In general, there are two categories: conjunctive models and compensatory models. In the following, we give a detailed description and introduce several popular diagnostic models in the literature. A more thorough review of those models can be found in Rupp *et al.* (2010).

1.2.2 Conjunctive diagnostic models

For conjunctive models, the ideal response is specified by

$$\xi^j(\boldsymbol{\alpha}, Q) = I(\alpha^k \geq q_{jk} \text{ for all } k = 1, \dots, K), \quad (1.4)$$

where $I(\cdot)$ is the usual indicator function. Equation (1.4) shows that all the specified skills by the Q -matrix are required for successful performance on the corresponding item ($\xi = 1$) and lack of competency on any one required attribute leads to a negative ideal response ($\xi = 0$). In other words, having additional unnecessary attributes does not compensate for the lack of the necessary attributes.

A basic but popular conjunctive diagnostic model is the DINA (Deterministic Input, Noisy output “AND” gate) model (Junker and Sijtsma 2001). The ideal response under the DINA model take the form of (1.4). In addition, to count for the randomness of the responses, the DINA model introduces the so-called slipping and guessing parameters (Junker and Sijtsma 2001). The concept is due to Macready and Dayton (1977) for mastery testing; see also van der Linden (1978). The slipping parameter is the probability that a subject (with attribute profile $\boldsymbol{\alpha}$) responds negatively to an item if the ideal response to that item $\xi(\boldsymbol{\alpha}, Q) = 1$; similarly, the guessing parameter refers to the probability that a subject responds positively if his or her ideal response $\xi(\boldsymbol{\alpha}, Q) = 0$.

We use $s = (s_1, \dots, s_J)^\top$ to denote the slipping probability and $g = (g_1, \dots, g_J)^\top$ to denote the guessing probability (with corresponding subscript indicating different J items). In the discussion, it is more convenient to work with the complement of the slipping parameter. Therefore, we define $c = \mathbf{1} - s$ to be the probability of answering correctly, with c_j being the corresponding item-specific notation. Given a specific subject’s profile $\boldsymbol{\alpha}$, the response to item j under the DINA model follows a Bernoulli distribution

$$P(R^j = 1 | Q, \boldsymbol{\alpha}, c_j, g_j) = c_j^{\xi^j(\boldsymbol{\alpha}, Q)} g_j^{1 - \xi^j(\boldsymbol{\alpha}, Q)}. \quad (1.5)$$

In addition, conditional on $\boldsymbol{\alpha}$, (R^1, \dots, R^J) are jointly independent.

Note that under the DINA model, if $K = 1$, the item characteristic curve can be taken as a discretized version of the item characteristic function of a IRT model. See Figure 1.4 for an illustration, where subjects with low ability are taken as nonmastering the necessary ability and have correct response probability equal to g while those with high ability are considered as mastering the ability and have correct response probability $1 - s$. In the literature, this is also known as the mastery testing, see

Macready and Dayton (1977); van der Linden (1978) for more details.

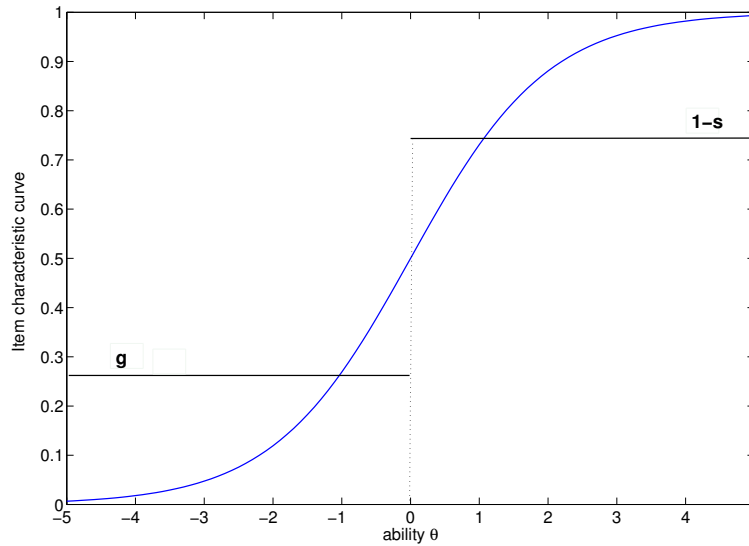


Figure 1.4: Item characteristic curves for the DINA model with $K = 1$.

A generalization of the DINA model is the reduced version of the Reparameterized Unified Model (RUM); see Hartz (2002). Under the reduced-RUM model, we have that for the j th item,

$$P(R^j = 1 | \boldsymbol{\alpha}, \pi_j, r_{jk}) = \pi_j \prod_{k=1}^K r_{jk}^{q_{jk}(1-\alpha^k)}, \quad (1.6)$$

where π_j is the correct response probability for subjects who possess all required attributes and $r_{j,k}$, $0 < r_{j,k} < 1$, is the penalty parameter for not possessing the k th attribute. The reduced RUM model is also a conjunctive model, and it generalizes the DINA model by allowing slipping and guessing parameters to vary across different attribute profiles.

1.2.3 Compensatory diagnostic models

Compensatory models only need one of the specified attributes for a successful performance. The presence of one required attribute can compensate for the lack of others. For example, consider the case in which different attributes represent different strategies for solving an item. Then a positive response to the item only requires the successful performance of one strategy (attribute). The ideal response for a compensatory model is then

$$\xi^j(\boldsymbol{\alpha}, Q) = I(\alpha^k \geq q_{jk} \text{ for some } k = 1, \dots, K) = 1 - \prod_{k=1}^K (1 - \alpha^k)^{q_{jk}}. \quad (1.7)$$

The DINO (Deterministic Input, Noisy output “OR” gate) model is a compensatory model (Templin and Henson 2006; Templin 2006). The ideal response $\xi_{DINO}^j(\boldsymbol{\alpha}, Q)$ takes the above form. Therefore, it only needs to possess one of the attributes required by the Q -matrix to be capable of providing a positive response to an item. With the same definition of c_j and g_j as in the DINA model, the response under the DINO model follows

$$P(R^j = 1 | \boldsymbol{\alpha}, c_j, g_j) = c_j^{\xi_{DINO}^j(\boldsymbol{\alpha}, Q)} g_j^{1 - \xi_{DINO}^j(\boldsymbol{\alpha}, Q)}. \quad (1.8)$$

1.3 Main results of the thesis

This thesis focuses on statistical inference for diagnostic classification models. Central to many diagnostic models is the Q -matrix, which specifies the item-attribute relationship. It is common practice for the Q -matrix to be specified by experts when items are written, rather than through data-driven calibration. Such a non-empirical approach may lead to misspecification of the Q -matrix and substantial lack of model fit, resulting in erroneous interpretation of testing results. This motivates our study

and we consider the identifiability, estimation, and hypothesis testing of the Q -matrix. In addition, we study the identifiability issue of the model parameters under a known Q -matrix.

In Chapter 2¹, we develop identifiability conditions for the Q -matrix to be learnable from the response data. Despite the importance of the Q -matrix in cognitive diagnosis, its estimation problem is largely an unexplored area. Unlike typical inference problems, inference for the Q -matrix is particularly challenging for the following reasons. First, in many cases, the Q -matrix is simply nonidentifiable, i.e., multiple Q -matrices lead to an identical response distribution. Therefore, we only expect to identify the Q -matrix up to some equivalence relation. In other words, two Q -matrices in the same equivalence class are not distinguishable based on data. Our first task is to define a meaningful and identifiable equivalence class. Second, the Q -matrix lives on a discrete space – the set of $J \times K$ matrices with binary entries. This discrete nature makes analysis particularly difficult because calculus tools are not applicable. In fact, most theoretical analyses in this thesis are combinatorics based. In this chapter, we present definitive answers to the learnability of the Q -matrix for one of the most commonly used models, the DINA model, by specifying a set of sufficient conditions under which the Q -matrix is identifiable up to an explicitly defined equivalence class.

Chapter 3 focuses on hypothesis testing of the Q -matrix. The purpose of this study is to provide a statistical procedure to help decide whether to accept the Q -matrix provided by the experts. Let Q_0 be the prespecified Q -matrix. Then this problem is equivalent to testing the null hypothesis $H_0 : Q = Q_0$. Based on the theoretical developments in Chapter 2, we propose a test statistic that measures the consistency of observed data with the proposed Q -matrix Q_0 . Asymptotic distributions of the

¹Part of Chapter 2 has been published/accepted in Bernoulli (Liu, Xu and Ying 2012b) and Applied Psychological Measurement (Liu, Xu and Ying 2012a)

test statistic are derived and the corresponding test procedures are established. In addition, we conduct simulation studies to assess the performance of the proposed method.

Chapter 4 is concerned with the identifiability of the diagnostic model parameters with a specified Q -matrix. Identifiability is a prerequisite for statistical inferences, such as parameter estimation and hypothesis testing. In this chapter, we focus on the DINA model and present sufficient and necessary conditions under which the model parameters are identifiable. The analysis method developed in this chapter is generic in the sense that it can be employed for the analysis of other diagnostic classification or latent class models.

Chapter 2

Estimation of Q -matrix

2.1 Introduction

Diagnostic classification models are important statistical tools in cognitive diagnosis and can be employed in a number of disciplines, including educational assessment and clinical psychology (Rupp and Templin 2008b). A key issue of cognitive diagnosis is to correctly specify the item-attribute relationships, which is specified by the Q -matrix Tatsuoka (1983). Different diagnostic classification models have been proposed in literature based on the Q -matrix. One simple and widely studied model among them is the DINA model (Deterministic Input, Noisy output “AND” gate; see Junker and Sijtsma 2001). Other important models and developments can be found in Tatsuoka (1985); DiBello, Stout and Roussos (1995); Hartz (2002); Tatsuoka (2002); Leighton, Gierl and Hunka (2004); von Davier (2005); Templin and Henson (2006); Chiu, Douglas and Li (2009). A more thorough review of cognitive diagnostic models can be found in Rupp, Templin and Henson (2010).

Statistical analysis with diagnostic models typically assumes a known Q -matrix

provided by experts such as those who developed the questions (Rupp 2002; Henson and Templin 2005; Roussos, Templin and Henson 2007b; Stout 2007). Such *a priori* knowledge when correct is certainly very helpful for both model estimation and eventually identification of subjects' latent attributes. On the other hand, model fitting is usually sensitive to the choice of Q -matrix and its misspecification could seriously affect the goodness of fit. This is one of the main sources for lack of fit. Various diagnostic tools and testing procedures have been developed (Rupp and Templin 2008a; de la Torre 2008a; Henson and Douglas 2005; Liu, Douglas and Henson 2007; Henson, Roussos, Douglas and He 2008; DeCarlo 2012). A comprehensive review of diagnostic classification models can be found in Rupp and Templin (2008b).

Despite the importance of the Q -matrix in cognitive diagnosis, its estimation problem is largely an unexplored area. Unlike typical inference problems, the inference for the Q -matrix is particularly challenging for the following reasons. First, in many cases, the Q -matrix is simply nonidentifiable. One typical situation is that multiple Q -matrices lead to an identical response distribution. Therefore, we only expect to identify the Q -matrix up to some equivalence relation (Definition 2). In other words, two Q -matrices in the same equivalence class are not distinguishable based on data. Our first task is to define a meaningful and identifiable equivalence class. Second, the Q -matrix lives on a discrete space – the set of $J \times K$ matrices with binary entries. This discrete nature makes analysis particularly difficult because calculus tools are not applicable. In fact, most analyses are combinatorics based. Third, the model makes explicit distributional assumptions on the (unobserved) attributes, dictating the law of observed responses. The dependence of responses on attributes via Q -matrix is a highly nonlinear discrete function. The nonlinearity also adds to the difficulty of the analysis.

The primary purpose of this chapter is to provide theoretical analyses on the learn-

ability of the underlying Q -matrix. In particular, we obtain definitive answers to the identifiability of Q -matrix for one of the most commonly used models – the DINA model – by specifying a set of sufficient conditions under which the Q -matrix is identifiable up to an explicitly defined equivalence class. We also present the corresponding consistent estimators. We believe that the results (especially the intermediate results) and analysis strategies can be extended to other conjunctive models (Maris 1999; Junker and Sijtsma 2001; Templin 2006; Templin and Henson 2006; Roussos, DiBello, Stout, Hartz, Henson and Templin 2007a).

The rest of this chapter is organized as follows. In Section 2.2, we present the basic inference result for Q -matrices in a conjunctive model with no slipping or guessing. In addition, we introduce all the necessary terminologies and technical conditions. In Section 2.3, we extend the results in Section 2.2 to the DINA model with known slipping and guessing parameters. In Section 2.4, we further generalize the results to the DINA model with unknown slipping parameters. In Section 2.5, we discuss the estimation of Q -matrix when both slipping and guessing parameters are unknown. Simulation results are given in Section 2.6 and further discussion is provided in Section 2.7. Proofs are given in Section 2.8. Lastly, the proofs of two key propositions are given in Section 2.9.

2.2 Estimation of Q -matrix

In this section, we focus on the conjunctive model and consider the simplest situation that there is no uncertainty in the response, that is, for a subject with attribute profile $\boldsymbol{\alpha}$, the response to item j is

$$R^j = \xi^j = \prod_{k=1}^K (\alpha^k)^{q_{jk}}, \quad (2.1)$$

where $\{\xi^j, j = 1, \dots, J\}$ are the ideal responses defined in Chapter 1. Therefore, the responses are completely determined by the Q -matrix and the attributes.

We assume that all items require at least one attribute. Equivalently, the Q -matrix does not have zero row vectors. Subjects who do not possess any attribute are not capable of responding positively to any item.

In order to provide an estimator of the Q -matrix, we first introduce one central quantity, the T -matrix, which connects the Q -matrix with the response and attribute distributions.

T -matrix. The T -matrix $T(Q)$ has 2^K columns each of which corresponds to one attribute vector, $\alpha \in \{0, 1\}^K$, with the same order of α in the population proportion vector $\mathbf{p} = (p_\alpha, \alpha \in \{0, 1\}^K)^\top$. So the columns of $T(Q)$ can be labeled by α instead of ordinal numbers. For instance, the α -th column of $T(Q)$ is the column that corresponds to attribute α .

Let I_i be a generic notation for positive responses to item i . Let “ \wedge ” stand for “and” combination. For instance, $I_{i_1} \wedge I_{i_2}$ denotes positive responses to both items i_1 and i_2 . Each row of $T(Q)$ corresponds to one item or one “and” combination of items, for instance, I_{i_1} , $I_{i_1} \wedge I_{i_2}$, or $I_{i_1} \wedge I_{i_2} \wedge I_{i_3}$, ... If $T(Q)$ contains all the single items and all “and” combinations, $T(Q)$ contains $2^J - 1$ rows. We will later say that such a $T(Q)$ is *saturated* (Definition 1 in Section 2.2.2).

We now describe each row vector of $T(Q)$. We define that $B_Q(j)$ is a 2^K dimensional row vector. Using the same labeling system as that of the columns of $T(Q)$, the α -th element of $B_Q(j)$ is defined as $\prod_{k=1}^K (\alpha^k)^{q_{jk}}$, which indicates if a subject with attribute α is able to solve item j .

Using a similar notation, we define that

$$B_Q(i_1, \dots, i_l) = \Upsilon_{h=1}^l B_Q(i_h), \quad (2.2)$$

where the operator “ $\Upsilon_{h=1}^l$ ” is element-by-element multiplication from $B_Q(i_1)$ to $B_Q(i_l)$. For instance, for vectors $W = (W^1, \dots, W^{2^k})$ and $V_h = (V_h^1, \dots, V_h^{2^k})$,

$$W = \Upsilon_{h=1}^l V_h$$

means that $W^j = \prod_{h=1}^l V_h^j$. Therefore, $B_Q(i_1, \dots, i_l)$ is the vector indicating the attributes that are capable of responding positively to items i_1, \dots, i_l . The row in $T(Q)$ corresponding to $I_{i_1} \wedge \dots \wedge I_{i_l}$ is $B_Q(i_1, \dots, i_l)$.

γ -vector. Let γ be a column vector the length of which equals to the number of rows of $T(Q)$. Each element of γ corresponds to one row vector of $T(Q)$. The element in γ corresponding to $I_{i_1} \wedge \dots \wedge I_{i_l}$ is defined as $N_{I_{i_1} \wedge \dots \wedge I_{i_l}}/N$, where $N_{I_{i_1} \wedge \dots \wedge I_{i_l}}$ denotes the number of people who have positive responses to items i_1, \dots, i_l , that is

$$N_{I_{i_1} \wedge \dots \wedge I_{i_l}} = \sum_{r=1}^N I(R_r^{i_j} = 1 : \text{for all } j = 1, \dots, l).$$

For each $\alpha \in \{0, 1\}^k$, let

$$\tilde{p}_\alpha = \frac{1}{N} \sum_{r=1}^N I(\alpha_r = \alpha). \quad (2.3)$$

If (2.1) is strictly respected and Q matrix is true, then

$$T(Q)\tilde{\mathbf{p}} = \gamma, \quad (2.4)$$

where $\tilde{\mathbf{p}} = (\tilde{p}_\alpha : \alpha \in \{0, 1\}^K)$ is arranged in the same order as the columns of $T(Q)$. This is because each row of $T(Q)$ indicates the attribute profiles corresponding to subjects capable of responding positively to that set of item(s). Vector $\tilde{\mathbf{p}}$ contains the proportions of subjects with each attribute profile. For each set of items, matrix multiplication sums up the proportions corresponding to each attribute profile capable

of responding positively to that set of items, giving us the total proportion of subjects who respond positively to the corresponding items.

Estimator of the Q -matrix. For each $J \times K$ binary matrix Q , we define

$$S(Q) = \inf_{\mathbf{p} \in [0,1]^{2^K}} |T(Q)\mathbf{p} - \boldsymbol{\gamma}|, \quad (2.5)$$

where $\mathbf{p} = (p_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \{0,1\}^K)$. The above minimization is subject to the constraint that

$$\sum_{\boldsymbol{\alpha} \in \{0,1\}^K} p_{\boldsymbol{\alpha}} = 1,$$

where $|\cdot|$ is the usual Euclidean norm.

An estimator of Q is then obtained by minimizing $S(Q)$, that is,

$$\hat{Q} = \arg \inf_Q S(Q), \quad (2.6)$$

where “arg inf” is the minimizer of the minimization problem over all $J \times K$ binary matrices. Note that the minimizers are not unique. We will later prove that the minimizers are in the same meaningful equivalence class. Because of (2.4), we can see that the true Q -matrix is always among the minimizers.

2.2.1 An illustration example

We illustrate the above construction by one simple example. We emphasize that this example is discussed to explain the estimation procedure for a concrete and simple example. The proposed estimator is certainly able to handle much larger Q -matrices. We consider the following 3×2 Q -matrix,

	addition	multiplication	
$Q_0 =$	$2 + 3$	1	0
	5×2	0	1
	$(2 + 3) \times 2$	1	1

(2.7)

Assume Q_0 is true. We consider the contingency table of attributes (addition and multiplication),

	multiplication	
addition	\tilde{p}_{00}	\tilde{p}_{01}
	\tilde{p}_{10}	\tilde{p}_{11}

In the above table, \tilde{p}_{00} is the proportional of people who do not master either addition or multiplication. Similarly, we define \tilde{p}_{01} , \tilde{p}_{10} , and \tilde{p}_{11} . Note that $\{\tilde{p}_{ij}; i, j = 0, 1\}$ is not observed.

Just for illustration, we construct a simple non-saturated T -matrix. Suppose the relationship in (2.1) is strictly respected. Then, we should be able to establish the following identities:

$$N(\tilde{p}_{10} + \tilde{p}_{11}) = N_{I_1}, \quad N(\tilde{p}_{01} + \tilde{p}_{11}) = N_{I_2}, \quad N\tilde{p}_{11} = N_{I_3}. \quad (2.8)$$

Therefore, if we let $\tilde{\mathbf{p}} = (\tilde{p}_{00}, \tilde{p}_{10}, \tilde{p}_{01}, \tilde{p}_{11})^\top$, the above display imposes three linear constraints on the vector $\tilde{\mathbf{p}}$. Together with the natural constraint that $\sum_{ij} \tilde{p}_{ij} = 1$, $\tilde{\mathbf{p}}$ solves the linear equation

$$T(Q_0)\tilde{\mathbf{p}} = \boldsymbol{\gamma}, \quad (2.9)$$

subject to the constraints that $\tilde{\mathbf{p}} \in [0, 1]^4$ and $\tilde{p}_{00} + \tilde{p}_{10} + \tilde{p}_{01} + \tilde{p}_{11} \in [0, 1]$, where

$$T(Q_0) = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \gamma = \begin{pmatrix} N_{I_1}/N \\ N_{I_2}/N \\ N_{I_3}/N \end{pmatrix}. \quad (2.10)$$

Each column of $T(Q_0)$ corresponds to one attribute profile. The first column corresponds to $\alpha = (0, 0)$, the second column to $\alpha = (1, 0)$, the third column to $\alpha = (0, 1)$, and the last column to $\alpha = (1, 1)$. The first row corresponds to item $2 + 3$, the second row to 5×2 , and the last row to $(2 + 3) \times 2$. For this particular situation, $T(Q)$ has rank equal to 3 and there exists one unique solution to (2.9) subject to the constraints of $\tilde{\mathbf{p}}$. In fact, we would not expect the constrained solution to the linear equation in (2.9) to always exist unless (2.1) is strictly followed. This is the topic of the next section.

The identities in (2.8) only consider the marginal rate of each question. There are additional constraints if one considers “combinations” among items. For instance,

$$N\tilde{p}_{11} = N_{I_1 \wedge I_2}.$$

People who are able to solve problem 3 must have both attributes and therefore are able to solve both problems 1 and 2. Again, if (2.1) is not strictly followed, this is not necessarily respected in the real data, though it is a logical conclusion. The DINA in the next section handles such a case. Upon considering I_1 , I_2 , I_3 , and $I_1 \wedge I_2$, the new T -matrix is

$$T(Q_0) = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \gamma = \begin{pmatrix} N_{I_1}/N \\ N_{I_2}/N \\ N_{I_3}/N \\ N_{I_1 \wedge I_2}/N \end{pmatrix}. \quad (2.11)$$

The last row is added corresponding to $I_1 \wedge I_2$. With (2.1) in force, we have

$$S(Q_0) = \inf_{\mathbf{p} \in [0,1]^4} |T(Q_0)\mathbf{p} - \boldsymbol{\gamma}| = |T(Q)\check{\mathbf{p}} - \boldsymbol{\gamma}| = 0. \quad (2.12)$$

This shows that the true Q -matrix Q_0 is always among the minimizers of the objective function $S(Q)$.

2.2.2 Main results for ideal responses

Before stating the main result, we provide a list of notations, which will be used in the discussions. We use Q_0 and $\mathbf{p}_0 = (p_{0,\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \{0,1\}^K)$ to denote the true Q -matrix and the true population parameter. We use \mathbf{e}_i to denote a row vector such that the i th element is one and the rest are zeros. When there is no ambiguity, we omit the length index of \mathbf{e}_i . In addition, we write the $k \times k$ identity matrix as \mathcal{I}_k .

The following definitions will be used in subsequent discussions.

Definition 1. *We say that a T -matrix is saturated if all combinations of form $I_{i_1} \wedge \dots \wedge I_{i_l}$, for $l = 1, \dots, J$, are included in $T(\cdot)$.*

For the example in the last section, the saturated T -matrix and the corresponding $\boldsymbol{\gamma}$ -vector are

$$T(Q_0) = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} N_{I_1}/N \\ N_{I_2}/N \\ N_{I_3}/N \\ N_{I_1 \wedge I_2}/N \\ N_{I_1 \wedge I_3}/N \\ N_{I_2 \wedge I_3}/N \\ N_{I_1 \wedge I_2 \wedge I_3}/N \end{pmatrix}. \quad (2.13)$$

It is not hard to see that not all the Q -matrices can be identified based on the response data \mathbf{R} . Since each column of Q represents an attribute, permuting the columns of Q is equivalent to relabeling the attributes, and therefore we can not distinguish the difference between Q and Q' if they have the same column vectors. We need to specify an equivalent class up to which our data can distinguish the Q -matrices. Such an *equivalence relation* is defined as follows.

Definition 2. We write $Q \sim Q'$ if and only if Q and Q' have identical column vectors, which could be arranged in different orders; otherwise, we write $Q \not\sim Q'$.

The primary objective of cognitive diagnosis is to identify the subjects' attributes (Rupp and Templin 2008b). It has been established (see more details in Chapter 4 and Chiu *et al.* (2009)) that in the ideal responses case, the sufficient and necessary condition for a set of items to consistently identify attributes is the Q -matrix is *complete*. Thus, it is usually recommended to use a complete Q -matrix.

Definition 3. A Q -matrix is said to be complete if $\{\mathbf{e}_i : i = 1, \dots, K\} \subset R_Q$ (R_Q is the set of row vectors of Q); otherwise, we say that Q is incomplete.

From the above definition, a Q -matrix is complete if and only if for any attribute there exists at least one item only requiring that attribute. Completeness implies that $J \geq K$. We will show that completeness is among the sufficient conditions to identify Q .

Listed below are assumptions which will be used in subsequent development.

C1 The true Q -matrix Q_0 is complete.

C2 T -matrix is saturated.

C3 The true population proportion vector $\mathbf{p}_0 \succ \mathbf{0}$, i.e., $p_{0,\alpha} > 0$ for all $\alpha \in \{0, 1\}^K$.

C4 Each attribute has been required by at least two items.

With these preparations, we are ready to introduce the first theorem, the proof of which is given in Section 2.8.

Theorem 1. *Assume that conditions C1-C4 hold. Suppose that for the i th subject with attribute profile α_i , the response to item j follows equation (2.1). Let \hat{Q} , defined in (2.6), be a minimizer of $S(Q)$ among all $J \times K$ binary matrices, where $S(Q)$ is defined in (2.5). Then,*

$$\lim_{N \rightarrow \infty} P(\hat{Q} \sim Q_0) = 1. \quad (2.14)$$

Further, let

$$\hat{\mathbf{p}} = \arg \inf_{\mathbf{p}} |T(\hat{Q})\mathbf{p} - \gamma|^2. \quad (2.15)$$

With an appropriate rearrangement of the columns of \hat{Q} , for any $\varepsilon > 0$

$$\lim_{N \rightarrow \infty} P(|\hat{\mathbf{p}} - \mathbf{p}_0| \leq \varepsilon) = 1.$$

Remark 1. *Conditions C1 and C2 are imposed to guarantee consistency. They may not be always necessary. Furthermore, constructing a saturated T -matrix is sometimes computationally not feasible, especially when the number of items is large. In practice, one may include combinations of one item, two items, ..., h items. The choice of h depends on the sample size and the computational resources. Condition C3 makes sure that \mathbf{p}_0 is not living on some sub-manifold. To see a counter example, suppose that $P(\alpha_i = (1, \dots, 1)^\top) = p_{0,(1,\dots,1)} = 1$. Then, for all Q -matrices, $P(\mathbf{R}_i = (1, \dots, 1)^\top) = 1$, that is, all subjects are able to solve all problems. Therefore, the distribution of \mathbf{R} is independent of Q . In other words, the Q -matrix is not identifiable. More generally, if there exist α_i^k and α_i^h such that $P(\alpha_i^k = \alpha_i^h) = 1$, then the Q -matrix is not identifiable*

based on the data. This is because one cannot tell if an item requires attribute k alone, attribute h alone, or both; see Reckase (1990; 2009) for similar cases for the multidimensional IRT models. The condition C_4 is required for technical purposes. Nonetheless, one can in fact construct counterexamples, i.e., the Q -matrix is not identifiable up to the relationship “ \sim ”, if C_4 is violated.

Remark 2. If $Q_1 \sim Q_2$, the two matrices only differ by a column permutation and will be considered to be the “same”. Therefore, we expect to identify the equivalence class that Q belongs to. Also, note that $S(Q_1) = S(Q_2)$ if $Q_1 \sim Q_2$.

Remark 3. Note that the estimator of the attribute distribution, $\hat{\mathbf{p}}$, in (2.15) depends on the order of columns of \hat{Q} . In order to achieve consistency, we will need to arrange the columns of \hat{Q} such that $\hat{Q} = Q_0$ whenever $\hat{Q} \sim Q_0$.

One practical issue associated with the proposed procedure is the computation. For a specific Q , the computation of $S(Q)$ only involves a constraint minimization of a quadratic function. However, if J or K is large, the computation overhead of searching the minimizer of $S(Q)$ over the space of $J \times K$ matrices could be substantial. One practical solution is to break the Q -matrix into smaller sub-matrices. For instance, one may divide the J items into m groups (possibly with nonempty overlap across different groups) and then apply the proposed estimator to each of the m group of items. This is equivalent to breaking a big J by K Q -matrix into several smaller matrices and estimating each of them separately. Lastly, combine the m estimated sub-matrices together to form a single estimate. The consistency results can be applied to each of the m sub-matrices and therefore the combined matrix is also a consistent estimator. A similar technique has been discussed in Chapter 8.6 of Tatsuoka (2009). See also Section 2.5 and 2.7 for more details.

2.3 DINA model with known slipping and guessing parameters

In this section, we extend the inference results in the previous section to the situation under which the responses do not deterministically depend on the attributes. In particular, we consider the DINA model.

Recall that for the DINA model, the j th item has two parameters: the slipping parameter (s_j) and the guessing parameter (g_j). And $c_j = 1 - s_j$ is the probability of a subject's responding positively to item j given that s/he is capable of solving that problem. Under the DINA model assumption, the response probabilities are specified as

$$P(R^j = 1|\xi^j) = c_j^{\xi^j} g_j^{1-\xi^j}, \quad (2.16)$$

where ξ^j is the capability indicator defined in (1.4). In addition, conditional on $\{\xi^1, \dots, \xi^J\}$, $\{R^1, \dots, R^J\}$ are jointly independent.

We write $c \succ g$ if $c_j > g_j$ for all $1 \leq j \leq J$, and write $c \not\asymp g$ if $c_j \neq g_j$ for all $j = 1, \dots, J$. In the following, we use $c_0 = (c_{0,1}, \dots, c_{0,J})^\top$ and $g_0 = (g_{0,1}, \dots, g_{0,J})^\top$ to denote the true slipping and guessing parameters. In this section we assume c_0 and g_0 are known. We discuss the more general cases in the next sections.

2.3.1 A general T -matrix

Consider a general Q -matrix Q and a general set of parameters (c, g, \mathbf{p}) . We modify the T -matrix defined in last section accordingly in this context. We first consider the case that $g_j = 0$ for all $j = 1, \dots, J$. We introduce a diagonal matrix D_c . If the h th row of matrix $T_c(Q)$ corresponds to $I_{i_1} \wedge \dots \wedge I_{i_l}$, then the h th diagonal element of D_c

is $c_{i_1} \times \dots \times c_{i_l}$. Then, we let

$$T_c(Q) = D_c T(Q), \quad (2.17)$$

where $T(Q)$ is the binary matrix defined previously. In other words, we multiply each row of $T(Q)$ by a common factor and obtain $T_c(Q)$. Note that in absence of slipping ($c_j = 1$ for each j) we have that $T_c(Q) = T(Q)$.

There is another equivalent way of constructing $T_c(Q)$. We define

$$B_{c,Q}(j) = c_j B_Q(j),$$

and

$$B_{c,Q}(i_1, \dots, i_l) = \Upsilon_{h=1}^l B_{c,Q}(i_h), \quad (2.18)$$

where “ Υ ” refers to element by element multiplication. Let the row vector in $T_c(Q)$ corresponding to $I_{i_1} \wedge \dots \wedge I_{i_l}$ be $B_{c,Q}(i_1, \dots, i_l)$.

For instance, for $c_0 = (c_{0,1}, c_{0,2}, c_{0,3})$, $T_{c_0}(Q_0)$ corresponding to the T -matrix in (2.11) would be

$$T_{c_0}(Q_0) = \begin{pmatrix} 0 & c_{0,1} & 0 & c_{0,1} \\ 0 & 0 & c_{0,2} & c_{0,2} \\ 0 & 0 & 0 & c_{0,3} \\ 0 & 0 & 0 & c_{0,1}c_{0,2} \end{pmatrix}. \quad (2.19)$$

Lastly, we consider the situation that both the probability of making a mistake and the probability of guessing correctly could be strictly positive. By this, we mean that the probability that a subject responds positively to item j is c_j if s/he is capable of doing so; otherwise the probability is g_j . We create a corresponding $T_{c,g}(Q)$ by slightly modifying $T_c(Q)$. We define the row vector

$$\mathbb{E} = (1, \dots, 1).$$

When there is no ambiguity, we omit the length index of \mathbb{E} . Now, let

$$B_{c,g,Q}(j) = g_j \mathbb{E} + (c_j - g_j) B_Q(j)$$

and

$$B_{c,g,Q}(i_1, \dots, i_l) = \Upsilon_{h=1}^l B_{c,g,Q}(i_h). \quad (2.20)$$

Let the row vector in $T_{c,g}(Q)$ corresponding to $I_{i_1} \wedge \dots \wedge I_{i_l}$ be $B_{c,g,Q}(i_1, \dots, i_l)$. For instance, the matrix $T_{c_0,g_0}(Q_0)$ corresponding to the $T_{c_0}(Q_0)$ in (2.19) is

$$T_{c_0,g_0}(Q_0) = \begin{pmatrix} g_{0,1} & c_{0,1} & g_{0,1} & c_{0,1} \\ g_{0,2} & g_{0,2} & c_{0,2} & c_{0,2} \\ g_{0,3} & g_{0,3} & g_{0,3} & c_{0,3} \\ g_{0,1}g_{0,2} & c_{0,1}g_{0,2} & g_1c_{0,2} & c_{0,1}c_{0,2} \end{pmatrix}. \quad (2.21)$$

Note that the first column of T -matrix contains the probabilities of providing positive responses to items simply by guessing.

By the law of large number, we know that

$$\gamma - T_{c_0,g_0}(Q_0) \mathbf{p}_0 \rightarrow \mathbf{0} \quad (2.22)$$

almost surely as $N \rightarrow \infty$.

2.3.2 Estimation of the Q -matrix and consistency results

Having concluded our preparations, we are now ready to introduce our estimators for Q . We introduce the objective function

$$S_{c,g,\mathbf{p}}(Q) = |T_{c,g}(Q)\mathbf{p} - \gamma|. \quad (2.23)$$

If we know the true parameters (c_0, g_0, \mathbf{p}_0) , then a natural estimator of the Q -matrix is

$$\hat{Q} = \arg \inf_Q S_{c_0, g_0, \mathbf{p}_0}(Q).$$

Since we don't know \mathbf{p}_0 , we will use the profiled objective function instead. Given c_0 and g_0 , we define our objective function for Q as

$$S_{c_0, g_0}(Q) = \inf_{\mathbf{p} \in [0, 1]^{2^K}} |T_{c_0, g_0}(Q)\mathbf{p} - \boldsymbol{\gamma}|, \quad (2.24)$$

where $\mathbf{p} = (p_\alpha : \alpha \in \{0, 1\}^K)$. The minimization in (2.24) is subject to constraints that

$$p_\alpha \in [0, 1], \quad \text{and} \quad \sum_{\alpha} p_\alpha = 1.$$

We propose an estimator of the Q -matrix through a minimization problem, that is,

$$\hat{Q}(c_0, g_0) = \arg \inf_Q S_{c_0, g_0}(Q). \quad (2.25)$$

We write c_0 and g_0 in the argument to emphasize that the estimator depends on c_0 and g_0 . The computation of the minimization in (2.24) consists of minimizing a quadratic function subject to finitely many linear constraints. Therefore, it can be done efficiently.

Theorem 2. *Suppose that c_0 and g_0 are known and that conditions C1-C4 are in force. For subject i , the responses are generated independently such that*

$$P(R_i^j = 1 | \xi_i^j) = c_{0,j}^{\xi_i^j} g_{0,j}^{1-\xi_i^j}, \quad (2.26)$$

where ξ_i^j is defined as in Theorem 1. Let $\hat{Q}(c_0, g_0)$ be defined as in (2.25). If $c_{0,j} \neq g_{0,j}$ for all j , then

$$\lim_{N \rightarrow \infty} P(\hat{Q}(c_0, g_0) \sim Q_0) = 1.$$

Furthermore, let

$$\hat{\mathbf{p}}(c_0, g_0) = \arg \inf_{\mathbf{p}} \left| T_{c_0, g_0}(\hat{Q}(c_0, g_0))\mathbf{p} - \boldsymbol{\gamma} \right|^2,$$

subject to constraint that $\sum_{\alpha} p_{\alpha} = 1$. Then, with an appropriate rearrangement of the columns of \hat{Q} , for any $\varepsilon > 0$,

$$\lim_{N \rightarrow \infty} P(|\hat{\mathbf{p}}(c_0, g_0) - \mathbf{p}_0| \leq \varepsilon) = 1.$$

Remark 4. There are various metrics one can employ to measure the distance between the vectors $T_{c,g}(\hat{Q}(c_0, g_0))\mathbf{p}$ and $\boldsymbol{\gamma}$. In fact, any metric that generates the same topology as the Euclidian metric is sufficient to obtain the consistency results in the theorem. For instance, a principled choice of objective function would be the likelihood with \mathbf{p} profiled out. The reason we prefer the Euclidian metric (versus, for instance, the full likelihood) is that the evaluation of $S(Q)$ is easier than the evaluation based on other metrics. More specifically, the computation of current $S(Q)$ consists of quadratic programming types of well oiled optimization techniques.

2.4 DINA with unknown slipping probabilities

In this section, we further extend our results to the situation where the slipping probabilities are unknown and the guessing probabilities are known. In the context of standard exams, the guessing probabilities can typically be set to zero for open problems. For instance, the chance of guessing the correct answer to “ $(3 + 2) \times 2 = ?$ ” is very small if the student does not know addition. On the other hand, for multiple choice problems, the guessing probabilities cannot be ignored. In that case, g_j can be considered as $1/n$ when there are n choices.

We first provide several estimators of c_0 given a Q -matrix Q and g_0 . The first method is applicable to all Q -matrices, but computationally intensive. The second

one is computationally easy, but requires certain structures of Q . The third one is the usual maximum likelihood estimator and is easy to compute using the EM algorithm. We introduce the first two methods here due to theoretical interest.

A estimator based on S function We first provide an estimator of c_0 that is applicable to all Q -matrices. For a given Q -matrix Q and true guessing parameters g_0 , we propose the following estimator of c :

$$\hat{c}(Q, g_0) = \arg \inf_{c \in [0,1]^J} S_{c, g_0}(Q). \quad (2.27)$$

A moment estimator When the Q -matrix has a certain structure, we are able to estimate c consistently based on estimating equations.

We need a result which will be given in the proof of Proposition 6 (Section 2.8.2). For general c and g , let $c - g = (c_1 - g_1, \dots, c_J - g_J)^\top$ and

$$\tilde{T}_{c,g}(Q) = \begin{pmatrix} T_{c,g}(Q) \\ \mathbb{E} \end{pmatrix}.$$

Then there exists a matrix D_g , which only depending on g , such that

$$D_g \tilde{T}_{c,g}(Q) = T_{c-g}(Q).$$

Under the specification of Q_0 , for a particular item i , suppose that there exist items i_1, \dots, i_l (different from i) such that

$$B_{Q_0}(i, i_1, \dots, i_l) = B_{Q_0}(i_1, \dots, i_l), \quad (2.28)$$

that is, the attributes required by item i are a subset of the attributes required by i_1, \dots, i_l .

Let a_{g_0} and a_{*g_0} be the row vectors in D_{g_0} corresponding to $I_{i_1} \wedge \dots \wedge I_{i_l}$ and $I_i \wedge I_{i_1} \wedge \dots \wedge I_{i_l}$ in matrix $T_{c_0-g_0}(Q_0)$. Then, for the true Q -matrix Q_0 and true parameters (\mathbf{p}_0, c_0, g_0) , the law of large number implies that

$$\begin{aligned} \frac{a_{*g_0}^\top \begin{pmatrix} \gamma \\ 1 \end{pmatrix}}{a_{g_0}^\top \begin{pmatrix} \gamma \\ 1 \end{pmatrix}} &= \frac{a_{*g_0}^\top \tilde{T}_{c_0, g_0}(Q_0) \mathbf{p}_0}{a_{g_0}^\top \tilde{T}_{c_0, g_0}(Q_0) \mathbf{p}_0} + o_p(1) \\ &= \frac{B_{c_0-g_0, Q_0}(i, i_1, \dots, i_l) \mathbf{p}_0}{B_{c_0-g_0, Q_0}(i_1, \dots, i_l) \mathbf{p}_0} + o_p(1) \\ &\xrightarrow{p} (c_{0,i} - g_{0,i}), \end{aligned} \tag{2.29}$$

where the vectors a_{g_0} and a_{*g_0} only depend on guessing parameters g_0 .

Therefore, the corresponding estimator of $c_{0,i}$ given Q_0 and g_0 would be

$$\hat{c}_i(Q_0, g_0) = g_{0,i} + \frac{a_{*g_0}^\top \begin{pmatrix} \gamma \\ 1 \end{pmatrix}}{a_{g_0}^\top \begin{pmatrix} \gamma \\ 1 \end{pmatrix}}. \tag{2.30}$$

Note that the computation of $\bar{c}_i(Q, g)$ only consists of affine transformations and therefore is very fast.

Proposition 1. *Suppose (2.26) and (2.28) are true. Then $\hat{c}_i(Q_0, g_0) \rightarrow c_{0,i}$, for $i = 1, \dots, J$, in probability as $N \rightarrow \infty$.*

Proof of Proposition 1. By the law of large numbers,

$$\begin{aligned} a_{*g_0}^\top \begin{pmatrix} \gamma \\ 1 \end{pmatrix} - a_{*g}^\top \tilde{T}_{c,g}(Q_0) \mathbf{p}_0 &\rightarrow 0, \\ a_{g_0}^\top \begin{pmatrix} \gamma \\ 1 \end{pmatrix} - a_{g_0}^\top \tilde{T}_{c_0,g_0}(Q_0) \mathbf{p}_0 &\rightarrow 0, \end{aligned}$$

in probability as $N \rightarrow \infty$. By the construction of a_{*g_0} and a_{g_0} , we have

$$\begin{aligned} a_{*g_0}^\top \tilde{T}_{c_0,g_0}(Q_0) \mathbf{p}_0 &= B_{c_0-g_0, Q_0}(i, i_1, \dots, i_l) \mathbf{p}_0, \\ a_{g_0}^\top \tilde{T}_{c_0,g_0}(Q_0) \mathbf{p}_0 &= B_{c_0-g_0, Q_0}(i_1, \dots, i_l) \mathbf{p}_0. \end{aligned}$$

Thanks to (2.28), we have

$$\frac{a_{*g_0}^\top \begin{pmatrix} \gamma \\ 1 \end{pmatrix}}{a_{g_0}^\top \begin{pmatrix} \gamma \\ 1 \end{pmatrix}} \rightarrow c_{0,i} - g_{0,i}.$$

□

Maximum likelihood estimator A natural estimator of c_0 is the maximum likelihood estimator (MLE). For a given Q -matrix Q and true guessing parameters g_0 , the MLE of c_0 is defined as:

$$\hat{c}(Q, g_0) = \arg \sup_{c \in [0,1]^J} L_{c,g_0}(Q), \quad (2.31)$$

where $L_{c,g}(Q)$ is the likelihood function taking the form of

$$L_{c,g}(Q) = \prod_{i=1}^N \left\{ \sum_{\alpha \in \{0,1\}^K} \left(p_\alpha \prod_{j=1}^J P(R_i^j = 1 | c, g, \alpha, Q)^{R_i^j} (1 - P(R_i^j = 1 | c, g, \alpha, Q))^{1-R_i^j} \right) \right\}.$$

The computation of the MLE can be done efficiently by the EM algorithm (Dempster, Laird and Rubin (1977); de la Torre (2009)).

Under certain regular conditions, (\mathbf{p}_0, c_0) is identifiable based on the response data with a correctly specified Q -matrix (see Chapter 4). Nonetheless, non-identifiability issue does exist and we can easily construct counter examples. For instance, consider a complete matrix $Q = \mathcal{I}_K$. There are in total 2^K constraints based on the T -matrix and restriction on \mathbf{p} . On the other hand, the parameter dimension of (\mathbf{p}_0, c_0) is $2^K + K$. Therefore, without additional information \mathbf{p}_0 and c_0 cannot be consistently identified. A typically approach to tackle this problem is to introduce addition parametric assumptions such as \mathbf{p}_0 satisfying certain restrictions or in the Bayesian setting (weakly) informative prior distributions (Gelman, Jakulin, Pittau and Su 2008). This example shows that it is not always possible to consistently estimate c_0 and \mathbf{p}_0 given the guessing parameter g and Q -matrix. The identifiability of (\mathbf{p}_0, c_0) will be further pursued in Chapter 4, where sufficient and necessary conditions for parameter identifiability are proposed.

To estimate the Q -matrix, we replace the unknown slipping parameter in S function with the estimator \hat{c} described in the above methods, and have the following objective function

$$S_{\hat{c}(Q, g_0), g_0}(Q) = \inf_{\mathbf{p}} S_{\hat{c}(Q, g_0), g_0, \mathbf{p}}(Q) = \inf_{\mathbf{p}} |T_{\hat{c}(Q, g_0), g_0}(Q)\mathbf{p} - \boldsymbol{\gamma}|, \quad (2.32)$$

where the minimization is subject to the natural constraints that $p_{\alpha} \in [0, 1]$ and $\sum_{\alpha} p_{\alpha} = 1$. Then, the corresponding estimator is

$$\hat{Q}_{\hat{c}}(g_0) = \arg \inf_Q S_{\hat{c}(Q, g_0), g_0}(Q). \quad (2.33)$$

The consistency of the Q -matrix estimator is given in the following theorem.

Theorem 3. *Suppose that g_0 is known and the conditions in Theorem 2 hold. Let $\hat{Q}_{\hat{c}}(g_0)$ is as defined in (2.33). Then,*

$$\lim_{N \rightarrow \infty} P \left(\hat{Q}_{\hat{c}}(g_0) \sim Q_0 \right) = 1.$$

Note that under the conditions of Theorem 3, the moment estimator is consistent. In addition, we have the consistency of \hat{c} when it is the likelihood estimator or the estimator minimizing S function; see Proposition 11 and Theorem 6 in Chapter 4 for more details.

2.5 DINA with unknown slipping and guessing

In this subsection, we consider the computation of the estimator of the Q -matrix when neither the slipping nor the guessing parameters are known.

Based on the construction and the discussions in the previous sections, we consider the objective function for any c , g , \mathbf{p} , and Q ,

$$S_{c,g,\mathbf{p}}(Q) = |T_{c,g}(Q)\mathbf{p} - \boldsymbol{\gamma}|. \quad (2.34)$$

If we know the true parameters (c_0, g_0, \mathbf{p}_0) , then a natural estimator of the Q -matrix is

$$\hat{Q} = \arg \inf_Q S_{c_0, g_0, \mathbf{p}_0}(Q).$$

Most of the time, the parameters (c_0, g_0, \mathbf{p}_0) are unknown. Under these situations, we consider the profiled objective functions

$$S(Q) = \inf_{c,g,\mathbf{p}} S_{c,g,\mathbf{p}}(Q), \quad (2.35)$$

where the minimization is subject to the natural constraints that $c_i, g_i, p_\alpha \in [0, 1]$ and $\sum_\alpha p_\alpha = 1$. Then, the corresponding estimator is

$$\hat{Q} = \arg \inf_Q S(Q). \quad (2.36)$$

The minimization of \mathbf{p} in (2.35) consists of a quadratic optimization with linear constraints, and therefore can be done efficiently. The minimization with respect to c and g is usually not straightforward. One may alternatively replace the minimization by other estimators (such as the maximum likelihood estimator) $(\hat{c}(Q), \hat{g}(Q), \hat{\mathbf{p}}(Q))$. Thus, the objective function becomes

$$\hat{S}(Q) = S_{\hat{c}(Q), \hat{g}(Q), \hat{\mathbf{p}}(Q)}(Q). \quad (2.37)$$

The corresponding estimator is

$$\hat{Q} = \arg \inf_Q \hat{S}(Q). \quad (2.38)$$

This alternative allows certain flexibility in the estimation procedure. The S -function in (2.37) is usually easier to compute. Therefore, we often work with the estimator (2.38) and a hill-climbing algorithm to compute \hat{Q} is given in the next subsection.

We consider the estimator in (2.38) and the objective function (2.37). Let $(\hat{c}, \hat{g}, \hat{\mathbf{p}})$ be the maximum likelihood estimator (MLE). The computation of the MLE can be done efficiently by the EM algorithm (Dempster, Laird and Rubin (1977); de la Torre (2009)). Furthermore, we consider the optimization of (2.36) and (2.38).

The optimization of a general nonlinear discrete function is a very challenging problem. A simple-minded search of the entire space consists of evaluating the function S up to $2^{J \times K}$ times. In our current setting, an *a priori* Q -matrix, denoted by Q^* , is usually available. We expect that Q^* is reasonably close to the true matrix Q .

For each Q , let $U_j(Q)$ be the set of $J \times K$ matrices that are identical to Q except for the j th row (item). Then our algorithm is described as follows.

Algorithm 1. *Choose a starting point $Q(0) = Q^*$. For each iteration m , given the matrix from the previous iteration $Q(m-1)$, we perform the following steps.*

1. Let

$$Q_j = \arg \inf_{Q \in U_j(Q^{(m-1)})} S(Q). \quad (2.39)$$

2. Let $j_* = \arg \inf_j S(Q_j)$.

3. Let $Q(m) = Q_{j_*}$.

Repeat steps 1-3 until $Q(m) = Q(m-1)$.

At each step m , the algorithm considers updating one of the J items. In particular, if the j th item is updated, the Q -matrix for the next iteration would be Q_j . Then, $Q(m)$ is set to be the Q_{j_*} that admits the smallest objective function among all the Q_j 's. The optimization (2.39) consists of evaluating the function S 2^K times. Thus, the total computation complexity of each iteration is $J \times 2^K$ evaluations of S .

Remark 5. *The simulation study in Section 2.6 shows that if Q^* is different from Q by 3 items (out of 20 items) Algorithm 1 has a very high chance of recovering the true matrix with reasonably large samples.*

2.6 Simulation

In this section, we conduct simulation studies to illustrate the performance of the proposed method. We generate the data from the DINA model under different settings and compare the estimated Q -matrix and the true Q -matrix.

2.6.1 Estimation of the Q -matrix with no special structure

The simulation setting. We start with a test of $J = 20$ items that requires $K = 3$ attributes. The true Q -matrix Q_0 is given by

$$Q_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \quad (2.40)$$

We further generate the attributes from a uniform distribution, i.e.,

$$p_{0,\alpha} = 2^{-K}, \quad \forall \alpha \in \{0, 1\}^K.$$

The slipping parameters and the guessing parameters are set to be $s_{0,j} = g_{0,j} = 0.2$ for $j = 1, \dots, 20$. In addition, for each sample size $N = 500, 1000, 2000$, and 4000 ,

	N=500		N=1000		N=2000		N=4000	
	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$
$Q_0 = Q_1$	94	6	100	0	100	0	100	0
$Q_0 = Q_2$	82	18	100	0	100	0	100	0
$Q_0 = Q_3$	38	62	98	2	100	0	100	0

Table 2.1: Numbers of correctly estimated Q -matrices out of 100 simulations with $N = 500, 1000, 2000,$ and 4000 for $Q_1, Q_2,$ and Q_3 .

100 data sets were generated under the DINA model assumption.

To reduce computational complexity, we choose the T -matrix containing combinations of up to four items. More generally, the simulation study shows that a T -matrix containing all the $(K + 1)$ (and lower) combinations delivers good estimates. We implement Algorithm 1 with a starting Q -matrix Q^* specified as follows. The Q^* is constructed based on the true Q -matrix by misspecifying three items. In particular, we randomly selected 3 items out of the total 20 items without replacement. For each of the selected items, the corresponding row of Q^* is sampled uniformly from all the possible K dimensional binary vectors excluding the true vector (of Q_1) and the zero vector. That is, each of these rows is a uniform sample of $2^K - 2$ vectors. Thus, it is guaranteed that Q^* does not have zero-vectors and is different from the true Q -matrix $Q_0 = Q_1$ by precisely three items. The simulation results are given by the first row of Table 2.1. The columns “ $\hat{Q} = Q_0$ ” and “ $\hat{Q} \neq Q_0$ ” contains the frequencies of the events “ $\hat{Q} = Q_0$ ” and “ $\hat{Q} \neq Q_0$ ” respectively. Based on 100 independent simulations, \hat{Q} recovers the true Q -matrix 98 times when the sample size is 500. For larger samples $N = 1000, 2000,$ and $4000,$ the estimate \hat{Q} never misses the true Q -matrix.

We further simulate the data from Q -matrices with 4 and 5 attributes

$$Q_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad Q_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}. \quad (2.41)$$

With exactly the same settings, the results are given by the corresponding rows in the Table 2.1. The estimator performs well except for the cases where $N = 500$. This is mainly because the sample size is small relative to the dimension K .

An improved estimation procedure for small samples. We further investigate

the data sets generated according to Q_2 and Q_3 with $N = 500$ when the estimator \hat{Q} did not perform as well as other situations. In particular, we look into the cases when $\hat{Q} \neq Q_0$, with $Q_0 = Q_1, Q_2$, or Q_3 . We observe that Q -matrices with more misspecified entries do not necessarily admit larger S values. In some cases, the true Q -matrix Q_0 does not minimize the objective function S ; nonetheless, $S(Q_0)$ is not much larger than the global minimum $\inf_Q S(Q)$. Figures 2.1 and 2.2 show two typical cases. For each of the two figures, two plots are provided. The x -axis shows the number of iterations of the optimization algorithm. The y -axis of the left plot shows the number of misspecified entries of $Q(m)$ at iteration m ; the plot on the right shows the objective function $S(Q(m))$. For the case shown in Figure 2.1, the algorithm just misses the true Q -matrix by one entry; for the case in Figure 2.2, the algorithm in fact passes the true Q -matrix and move to another one. Both cases show that the true Q -matrix does not minimize the objective function S . In fact, the values of the S function have basically dropped to a very low level after three iterations. The algorithm tends to correct one misspecified item at each of the first 2 iterations. After iteration 3, the reduction of the S function is marginal, and there are several Q -matrices that fits the data approximately equally well. For such situations where there are several matrices whose S values are close to the global minimum, we recommend careful investigation of all those matrices and selection of the most sensible one from a practical point of view.

Motivated by this, we consider a modified algorithm with an early stopping rule, i.e., we stop the algorithm when the reduction of the S -function value is below some threshold. In particular, we choose a threshold value of 4.5% of $S(Q_{m-1})$ at the m -th iteration. With this early stopping rule, the estimator for Q_2 and Q_3 can be improved substantially. The results based on the *same* samples as in Table 2.1 are shown in Table 2.2 which is show much high frequency of recovering the true Q -matrix.

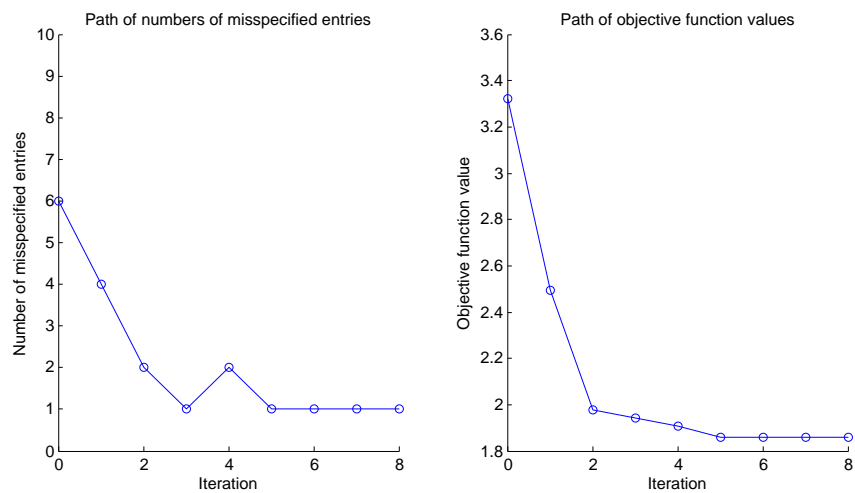


Figure 2.1: Results of a simulated data set with $N = 500$ and $K = 5$, for which the estimated \hat{Q} -matrix does not pass the true one.

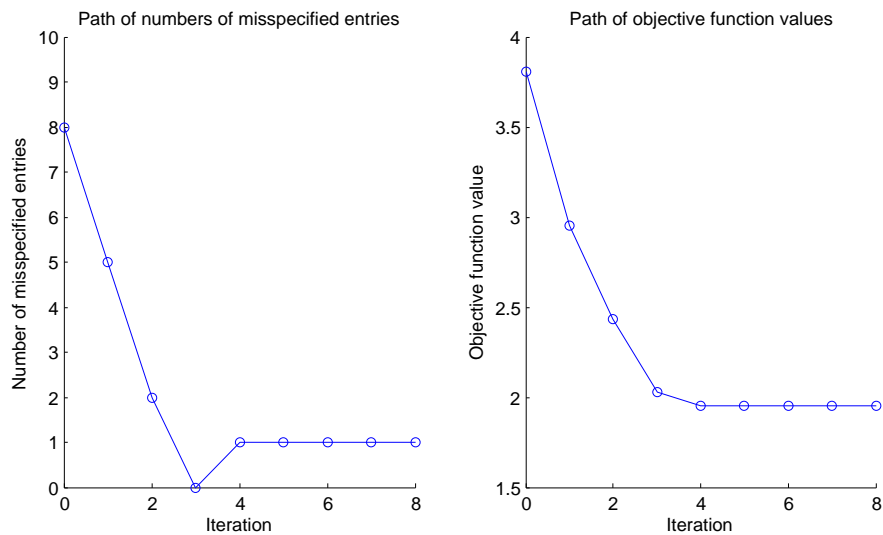


Figure 2.2: Results of a simulated data set with $N = 500$ and $K = 5$, for which the estimated \hat{Q} -matrix passes the true one but does not converge to it.

N=500		
	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$
$Q_0 = Q_2$	94	6
$Q_0 = Q_3$	70	30

Table 2.2: The results of algorithm with an early stopping rule for Q_2 and Q_3 based on the same samples as in Table 2.1.

When attribute profile α follows a non-uniform distribution. We consider the situation where the attribute profile α follows a non-uniform distribution. We adopt a similar setting as in Chiu *et al.* (2009), where attributes are correlated and unequal prevalence. We assume a multivariate probit model. In particular, for each subject, let $\theta = (\theta_1, \dots, \theta_k)$ be the underlying ability following a multivariate normal distribution $MVN(0, \Sigma)$, where the covariance matrix Σ has unit variance and common correlation ρ taking values of 0.05, 0.15 and 0.25. Then the attribute profile $\alpha = (\alpha^1, \dots, \alpha^K)$ is determined by

$$\alpha^k = \begin{cases} 1 & \text{if } \theta_k \geq \Phi^{-1}\left(\frac{k}{K+1}\right) \\ 0 & \text{otherwise.} \end{cases}$$

The other settings are similar as the previous simulations. We consider the true Q -matrix given as in (2.40) and $K = 3$. The slipping and guessing parameters are set to be 0.2. 100 independent datasets are generated. Table 2.3 shows the frequency of Q_1 being recovered by the estimator (after applying the early stopping method introduced in the above subsection). We can find that the more correlated the attributes are, the more difficult it is to estimate a Q -matrix. This is mostly because the samples are unevenly distributed over the 2^K possible attribute profiles and thus the “effective sample size” becomes smaller.

	N=1000		N=2000		N=4000	
	$\hat{Q} = Q_1$	$\hat{Q} \neq Q_1$	$\hat{Q} = Q_1$	$\hat{Q} \neq Q_1$	$\hat{Q} = Q_1$	$\hat{Q} \neq Q_1$
$\rho = 0.05$	78	22	98	2	100	0
$\rho = 0.15$	71	29	94	6	99	1
$\rho = 0.25$	41	59	76	24	95	5

Table 2.3: Numbers of correctly estimated Q_1 out of 100 simulations with $N = 500$, 1000, 2000, and 4000 for different ρ values.

2.6.2 Estimation of the Q -matrix with partial information

In this subsection, we consider the situation where partial knowledge of the true Q -matrix is available. We consider one of the situations discussed in Section 2.7. Consider a $J \times K$ Q -matrix where the attribute requirements of $J - 1$ items are known among the total J items. We are interested in learning the unknown J th item's Q -matrix structure. In this simulation we let $J = 2K + 1$. The first $2K$ rows of Q are known to form two complete matrices, i.e.,

$$Q_0 = \begin{pmatrix} \mathcal{I}_K \\ \mathcal{I}_K \\ V_J \end{pmatrix},$$

where \mathcal{I}_K is the identity matrix of dimension K and V_J is the row corresponding to the J th item to be learnt. The the corresponding estimator becomes

$$\hat{Q} = \arg \inf_{Q \in U_J(Q)} S(Q),$$

where $U_J(Q)$ is defined in Algorithm 1, as the set of Q -matrices identical to Q_0 for the first $J - 1$ rows.

With a similar setting to the previous simulations, the slipping and guessing parameters are set to be 0.2 and the population is set to be uniform, i.e., $p_\alpha = 2^{-K}$.

For each combination of $K = 3, 4,$ and $5,$ we consider different V_J 's. 100 independent datasets are generated. Table 2.4 shows the frequency of V_J being recovered by the estimator. One empirical finding is that the more "1"'s V_J contains, the more difficult it is to estimate V_J .

V_J	N=250		N=500		N=1000		N=2000	
	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$
(1 0 0)	91	9	98	2	100	0	100	0
(1 1 0)	82	18	97	3	99	1	100	0
(1 1 1)	70	30	83	17	100	0	100	0

V_J	N=500		N=1000		N=2000		N=4000	
	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$
(1 0 0 0)	91	9	98	2	100	0	100	0
(1 1 0 0)	84	16	94	6	100	0	100	0
(1 1 1 0)	71	29	87	13	99	1	100	0
(1 1 1 1)	39	61	62	38	94	6	100	0

V_J	N=1000		N=2000		N=4000		N=8000	
	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$	$\hat{Q} = Q_0$	$\hat{Q} \neq Q_0$
(1 0 0 0 0)	95	5	100	0	100	0	100	0
(1 1 0 0 0)	88	12	99	1	100	0	100	0
(1 1 1 0 0)	77	23	98	2	100	0	100	0
(1 1 1 1 0)	47	53	76	24	92	8	100	0
(1 1 1 1 1)*	29	71	37	63	56	44	88	12

Table 2.4: Numbers of correctly estimated Q -matrices out of 100 simulations with $K = 3, 4, 5.$ * In the case of (1 1 1 1 1), \hat{Q} recovers Q_0 100 times when $N = 12000.$

2.7 Discussion

Estimation of the Q -matrix for other DCMs. The differences among DCMs

lie mostly in their ideal response structures and the distribution of the response vectors implied by the Q -matrices. The distribution of response vector \mathbf{R} takes an additive form if responses to different items are conditionally independent given the attribute profile $\boldsymbol{\alpha}$, which is usually assumed for DCMs. With such a structure, one can construct the corresponding B -vectors that contain the corresponding conditional probabilities of the response vectors given each attribute profile $\boldsymbol{\alpha}$. Furthermore, a T -matrix is constructed by stacking all the B -vectors and an S -function is defined as the L^2 distance between the observed frequencies and those implied by the Q matrix. An estimator is then obtained by minimizing the S -function. Thus, this estimation procedure can be applied to other DCM's. For instance, one immediate extension of the current estimation procedure is to the DINO model.

Incorporating available information in the estimation procedure. Sometimes partial information is available for the parameters (Q, c, g, \mathbf{p}) . For instance, it is often reasonable to assume that some entries of the Q -matrix are known. Suppose we can separate the attributes into “hard” and “soft” ones. By “hard”, we mean those that are concrete and easily recognizable in a given problem and, by “soft”, we mean those that are subtle and not obvious. We can then assume that entries in columns which correspond to “hard” attributes are known. Alternatively, there may be a subset of items whose attribute requirements are known, while the item-attribute relationships of all other items need to be learnt, for example, when new items need to be calibrated according to the existing ones. Furthermore, even if an estimated Q -matrix may not be an appropriate replacement of the *a priori* Q -matrix provided by the “expert” (such as exam makers), it can serve as validation as well as a method of calibration using existing knowledge about the Q -matrix. When such information is available and correct, computation can be substantially reduced. This is because the optimization,

for instance that in (2.38), can be performed subject to existing knowledge of the Q -matrix. In particular, once the attribute requirements of a subset of $J - 1$ items are known, one can calibrate other items, one at a time, using those known items. More specifically, consider a $J \times K$ matrix Q^* , the first $J - 1$ items of which are known. We estimate the last item by $\hat{Q} = \arg \sup_{Q \in U_J(Q^*)} S(Q)$, i.e., we minimize the S -function subject to our knowledge about the first $J - 1$ items. Note that this optimization requires 2^K evaluations of the S -function and is therefore efficient. Thus, to calibrate M items, the total computation complexity is $O(M \times 2^K)$, which is typically of a manageable order.

Information about other parameters such as c , g , and \mathbf{p} can also be included in the estimation procedure. For instance, the attribute population is typically modeled to admit certain parametric form such as a log-linear model with certain interactions (von Davier and Yamamoto 2004; Henson and Templin 2005; Xu and von Davier 2008). This type of information can be incorporated in to the definition of (2.35) and (2.37), where the minimization and estimation of (c, g, \mathbf{p}) can be subject to additional parametric form or constraints. Such addition information is helpful enhancing the identifiability of the Q -matrix.

Theoretical properties of the estimator. Under weaker conditions, such as absence of completeness in the Q -matrix or the presence of unknown guessing parameter, the identifiability of the Q -matrix may be weaker, which corresponds to a coarser quotient set. One empirical finding is that Q -matrices with more diversified items tend to be easier to identify. For instance, one simple yet surprising example of a

non-identifiable Q -matrix is that

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

with slipping and guessing probabilities being 0.2 for all items and $p_{\alpha} = 1/4$ for all α . This Q -matrix cannot be distinguished from

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

that is, one can find another set of slipping, guessing probabilities and p'_{α} that implies the same distribution of the response vector. See Chapter 4 for more details about the identifiability of the model parameters.

Model Validation The proposed framework is applicable to not only the estimation of the Q -matrix but also the validation of an existing Q -matrix. If the Q -matrix is correctly specified and the assumptions of the DINA model are in place, then one may expect

$$|\gamma - T_{\hat{c}, \hat{g}}(Q)\hat{\mathbf{p}}| \rightarrow 0$$

in probability as $N \rightarrow \infty$. The above convergence requires no additional conditions to establish the consistency of \hat{Q} (such as completeness or diversified attribute distribution). In fact, it suffices that the responses are conditionally independent given the attributes and (\hat{c}, \hat{g}) are consistent estimators of (c, g) . Then, one may expect that when Q is true,

$$\hat{S}(Q) \rightarrow 0.$$

If the convergence rate of the estimators (\hat{c}, \hat{g}) is known, for instance, $(\hat{c} - c, \hat{g} - g) = O_p(N^{-1/2})$, then a necessary condition for a correctly specified Q -matrix is that $S_{\hat{c}, \hat{g}}(Q) = O_p(N^{-1/2})$. The asymptotic distribution of S depends on the specific form of (\hat{c}, \hat{g}) . Consequently, checking the closeness of S to zero forms a procedure for validation of the existing knowledge of the Q -matrix and the DINA model assumption. See Chapter 3 for more details about the Q -matrix validation.

Sample size. As the simulation results show that the estimator misses the true Q -matrix with non-ignorable probability (over 50%). This probability is substantially reduced (to 2%) when the sample size is increased to $N = 1000$. This suggests that a practically large sample N should be at least 30×2^K . Note that the K binary attributes partition the population into 2^K groups. In order to have the estimator yield reasonably accurate estimate there should be on average at least 30 samples in each group. In addition, performance of the estimator maybe further affected by the underlying attribute distribution. For instance, if the attributes are very correlated, the probabilities of certain attributes will be substantially smaller than others. For such cases, estimation for some rows in the Q -matrix (those corresponding to the small probability attributes) will be less accurate. For such situations, the “effective sample size” is even smaller.

Computation. The optimization of $S(Q)$ over the space of $J \times K$ binary matrices is a nontrivial problem. This is a substantial computational load if J and K are reasonably large. This computation might be reduced by splitting the Q -matrix into small sub-matrices. For typical statistical models, dividing the parameter space is usually not possible. The Q -matrix adopts a particular structure with which there is certain independence among items so that splitting the Q -matrix is valid. Similar techniques have been employed in the literature, such as Chapter 8.6 in the Tatsuoka

(2009) with large scale empirical studies in that chapter. In particular, for instance if there are 100 items, one can handle such a situation as follows. First, split the 100 items into 10 groups (possibly with overlapping items between groups if necessary); then apply the estimator to each of the 20 groups of items respectively. This is equivalent to breaking a big $100 \times K$ Q -matrix into 20 smaller matrices and estimating each of them separately. Lastly, combine the 20 estimated sub-matrices together to form a single estimate. Given that the computation for smaller scale matrices is much easier than those big ones, the splitting approach reduces the computation overhead. Nonetheless, developing a fast computation algorithm is an important line of future research.

Summary. As a concluding remark, we emphasize that learning the Q -matrix based on the data is an important problem even if *a priori* knowledge is sometimes available. In this chapter, we propose an estimation procedure of the Q -matrix under the setting of the DINA model. This method can also be adapted to the DINO model that is considered as the dual model of the DINA model. Simulation study shows that the estimator performs well when the sample size is reasonably large.

2.8 Proofs of theorems

2.8.1 Proof of theorems

To prove the main theorems, we need several propositions and lemmas, which are postponed to the next subsection. We now proceed to prove our theorems.

Proof of Theorem 1. Consider a general Q -matrix Q such that $Q \approx Q_0$. Suppose

$T(Q)$ and $T(Q_0)$ are saturated. Let $\tilde{\mathbf{p}}_0$ to be the vector containing $\tilde{p}_{0,\alpha}$'s with

$$\tilde{p}_{0,\alpha} = \frac{1}{N} \sum_{i=1}^N I(\alpha_i = \alpha).$$

For $\mathbf{p}_0 \succ \mathbf{0}$, since $\tilde{\mathbf{p}}_0 \rightarrow \mathbf{p}_0$ almost surely by the law of large number, then Corollary 1 and the fact that $\boldsymbol{\gamma} = T(Q_0)\tilde{\mathbf{p}}_0$ by (2.4) imply that $T(Q_0)\mathbf{p}_0 \notin C(T(Q))$, where $C(T)$ is the linear space generated by the column vectors of T , and there exists $\delta > 0$ such that,

$$\lim_{N \rightarrow \infty} P \left(\inf_{\mathbf{p} \in [0,1]^{2K}} |T(Q)\mathbf{p} - \boldsymbol{\gamma}| > \delta \right) = 1$$

and

$$P \left(\inf_{\mathbf{p} \in [0,1]^{2K}} |T(Q_0)\mathbf{p} - \boldsymbol{\gamma}| = 0 \right) = 1.$$

Given the fact that there are finitely many candidate $J \times K$ binary matrices, we have

$$P(\hat{Q} \sim Q_0) \rightarrow 1$$

as $N \rightarrow \infty$. In fact, we can arrange the columns of \hat{Q} such that $P(\hat{Q} = Q_0) \rightarrow 1$ as $N \rightarrow \infty$.

Note that $\tilde{\mathbf{p}}_0$ satisfies the identity

$$T(Q_0)\tilde{\mathbf{p}}_0 = \boldsymbol{\gamma}.$$

In addition, since $T(Q_0)$ is of full rank (Proposition 2), the solution to the above linear equation is unique. Therefore, the solution to the optimization problem $\inf_{\mathbf{p}} |T(Q_0)\mathbf{p} - \boldsymbol{\gamma}|$ is unique and is $\tilde{\mathbf{p}}$. Notice that when $\hat{Q} = Q_0$, $\hat{\mathbf{p}} = \arg \inf_{\mathbf{p}} |T(\hat{Q})\mathbf{p} - \boldsymbol{\gamma}| = \tilde{\mathbf{p}}$. Therefore,

$$\lim_{N \rightarrow \infty} P(\tilde{\mathbf{p}} = \hat{\mathbf{p}}) = 1.$$

Together with the consistency of $\tilde{\mathbf{p}}$, the conclusion of the theorem follows immediately. \square

Proof of Theorem 2. By the law of large number,

$$|T_{c_0, g_0}(Q_0)\mathbf{p}_0 - \boldsymbol{\gamma}| \rightarrow 0$$

almost surely as $N \rightarrow \infty$. Therefore,

$$S_{c_0, g_0}(Q_0) \rightarrow 0$$

almost surely as $N \rightarrow \infty$.

As previously defined, let

$$\tilde{T}_{c, g}(Q) = \begin{pmatrix} T_{c, g}(Q) \\ \mathbb{E} \end{pmatrix}. \quad (2.42)$$

Note that the last row of $\tilde{T}_{c, g}(Q)$ consists entirely of ones. Then we have that

$$\begin{pmatrix} \boldsymbol{\gamma} \\ 1 \end{pmatrix} \rightarrow \tilde{T}_{c_0, g_0}(Q_0)\mathbf{p}_0.$$

For any $Q \approx Q_0$, according to Proposition 6 and the fact that $\mathbf{p}_0 \succ \mathbf{0}$, there exists $\delta(c) > 0$ such that $\delta(c)$ is continuous in c and

$$\inf_{\mathbf{p}} \left| \tilde{T}_{c, g_0}(Q)\mathbf{p} - \tilde{T}_{c_0, g_0}(Q_0)\mathbf{p}_0 \right| > \delta(c).$$

By elementary calculus,

$$\delta \triangleq \inf_{c \in [0, 1]^J} \delta(c) > 0$$

and

$$\inf_{c, \mathbf{p}} \left| \tilde{T}_{c, g_0}(Q)\mathbf{p} - \tilde{T}_{c_0, g_0}(Q_0)\mathbf{p}_0 \right| > \delta.$$

Therefore,

$$P \left(\inf_{c, \mathbf{p}} \left| \tilde{T}_{c, g_0}(Q)\mathbf{p} - \begin{pmatrix} \boldsymbol{\gamma} \\ 1 \end{pmatrix} \right| > \delta/2 \right) \rightarrow 1,$$

as $N \rightarrow \infty$. For the same δ , we have

$$P\left(\inf_{c, \mathbf{p}} |T_{c, g_0}(Q))\mathbf{p} - \gamma| > \delta/2\right) = P(\inf_c S_{c, g_0}(Q) > \delta/2) \rightarrow 1.$$

The above minimization on the left of the equation is subject to the constraint that

$$\sum_{\alpha \in \{0,1\}^K} p_\alpha = 1.$$

Together with the fact that there are only finitely many $J \times K$ binary matrices, we have

$$P(\hat{Q}(c_0, g_0) \sim Q_0) = 1.$$

We arrange the columns of $\hat{Q}(c_0, g_0)$ so that $P(\hat{Q}(c_0, g_0) = Q_0) \rightarrow 1$ as $N \rightarrow \infty$.

Now we proceed to the proof of consistency for $\hat{\mathbf{p}}(c_0, g_0)$. Note that

$$\begin{aligned} \left| \tilde{T}_{c_0, g_0}(\hat{Q}(c_0, g_0))\hat{\mathbf{p}}(c_0, g_0) - \begin{pmatrix} \gamma \\ 1 \end{pmatrix} \right| &\xrightarrow{p} 0, \\ \left| \tilde{T}_{c_0, g_0}(Q_0)\mathbf{p}_0 - \begin{pmatrix} \gamma \\ 1 \end{pmatrix} \right| &\xrightarrow{p} 0. \end{aligned}$$

Since $\tilde{T}_{c_0, g_0}(Q_0)$ is a full column rank matrix and $P(\hat{Q}(c_0, g_0) = Q_0) \rightarrow 1$, $\hat{\mathbf{p}}(c_0, g_0) \rightarrow \mathbf{p}_0$ in probability. \square

Proof of Theorem 3. Assuming g_0 is known, note that for a given Q -matrix Q

$$\inf_{\mathbf{p}} \left| \tilde{T}_{c, g_0}(Q)\mathbf{p} - \begin{pmatrix} \gamma \\ 1 \end{pmatrix} \right|$$

is a continuous function of c . According to the results of Proposition 1, the definition in (2.27), and the definition of \hat{c} in Section 2.4, we obtain that

$$\inf_{\mathbf{p}} \left| \tilde{T}_{\hat{c}(Q_0, g_0), g_0}(Q_0) \mathbf{p} - \begin{pmatrix} \gamma \\ 1 \end{pmatrix} \right| \rightarrow 0,$$

in probability as $N \rightarrow \infty$ (Note that under the conditions C1-C4, we have the consistency of \hat{c} when it is the likelihood estimator; see Theorem 6 in Chapter 4). In addition, thanks to Proposition 6 and with a similar argument as in the proof of Theorem 2, $\hat{Q}_{\hat{c}}(g_0)$ is a consistent estimator. \square

2.8.2 Several propositions and lemmas

To make the discussion smooth, we postpone several long proofs to Appendix 2.9. For statement convenience, we introduce the following notations.

- Linear space spanned by vectors V_1, \dots, V_l :

$$\mathcal{L}(V_1, \dots, V_l) = \left\{ \sum_{j=1}^l a_j V_j : a_j \in \mathbf{R} \right\}.$$

- For a matrix M , $M_{1:l}$ denotes the submatrix containing the first l rows and all columns of M .
- For a matrix M , $C(M)$ is the linear space generated by the column vectors of M . It is usually called the *column space* of M .
- C_M denotes the set of column vectors of M .
- R_M denotes the set of row vectors of M .
- $T^*(Q)$ denotes the matrix that contains the second to the last columns of $T(Q)$. Note that the first column of $T(Q)$ is zero vector.

Proposition 2. *Suppose that Q is complete and matrix $T(Q)$ is saturated. Then, we are able to arrange the columns and rows of Q and $T^*(Q)$ such that $T^*(Q)_{1:(2^K-1)}$ has full rank and $T^*(Q)$ has full column rank.*

Proof of Proposition 2. Provided that Q is complete, without loss of generality we assume that the i th row vector of Q is \mathbf{e}_i^\top for $i = 1, \dots, K$, that is, item i only requires attribute i for each $i = 1, \dots, K$. Let the first $2^K - 1$ rows of $T^*(Q)$ be associated with $\{I_1, \dots, I_K\}$. In particular, we let the first K rows correspond to I_1, \dots, I_K and the first K columns of $T^*(Q)$ correspond to $\boldsymbol{\alpha}$'s that only have one attribute. We further arrange the next C_2^K rows of $T^*(Q)$ to correspond to combinations of two items, $I_i \wedge I_j$, $i \neq j$. The next C_2^K columns of $T^*(Q)$ correspond to $\boldsymbol{\alpha}$'s that only have two positive attributes. Similarly, we arrange $T^*(Q)$ for combinations of three, four, and up to K items. Therefore, the first $2^K - 1$ rows of $T^*(Q)$ admit a block upper triangle form. In addition, we are able to further arrange the columns within each block such that the diagonal matrices are identities, so that $T^*(Q)$ has form

$$\begin{array}{l} I_1, I_2, \dots \\ I_1 \wedge I_2, I_1 \wedge I_3, \dots \\ I_1 \wedge I_2 \wedge I_3, \dots \\ \vdots \end{array} \begin{pmatrix} \mathcal{I}_K & * & * & * & \dots \\ 0 & \mathcal{I}_{C_2^K} & * & * & \\ 0 & 0 & \mathcal{I}_{C_3^K} & * & \\ \vdots & \vdots & \vdots & & \end{pmatrix}. \quad (2.43)$$

Note that $T^*(Q)$ has $2^K - 1$ columns and $T^*(Q)_{1:(2^K-1)}$ obviously has full rank, therefore $T^*(Q)$ has full column rank. \square

When the guessing parameter is zero, we have the following result.

Proposition 3. *Suppose that Q is complete, $T(Q)$ is saturated, and $c \not\equiv \mathbf{0}$. Then, $T_c^*(Q)$ and $T_c^*(Q)_{1:(2^K-1)}$ have full column rank.*

Proof of Proposition 3. By Proposition 2, (2.17) and the fact that D_c is a diagonal matrix of full rank as long as $c \not\cong \mathbf{0}$, we have that

$$T_c^*(Q) = D_c T^*(Q),$$

is of full column rank. □

Consider two Q -matrices Q and Q' , the following two propositions compare the column spaces of $T_c(Q)$ and $T_c(Q')$, which are central to the proof of all the theorems. Their proofs are delayed to the appendix.

We assume that Q is a complete matrix and $T(Q)$ is saturated. We further assume that $Q_{1:K} = \mathcal{I}_K$ and the first $2^K - 1$ rows of $T(Q)$ are arranged in the order as in (2.43).

The first proposition discusses the case where $Q'_{1:K}$ is complete. We can always rearrange the columns of Q' so that $Q_{1:K} = Q'_{1:K}$. In addition, according to the proof of Proposition 2, the last column vector of $T_c(Q)$ corresponds to attribute $\boldsymbol{\alpha} = (1, \dots, 1)^\top$. Therefore, this column vector is all of nonzero entries.

Proposition 4. *Assume that Q satisfies conditions C1, C2, and C4. For $Q' \neq Q$, suppose $Q_{1:K} = Q'_{1:K} = \mathcal{I}_K$. Then $T_c(Q)\mathbf{p}$ is not in the column space $C(T_{c'}(Q'))$ for all $c' \in \mathbb{R}^J$ if $c \not\cong \mathbf{0}$ and $\mathbf{p} \succ 0$.*

The next proposition discusses the case where $Q'_{1:K}$ is incomplete.

Proposition 5. *Assume that Q is a complete matrix, $T(Q)$ is saturated, and $\mathbf{p} \succ 0$. Without loss of generality, let $Q_{1:K} = \mathcal{I}_K$. If $c \not\cong \mathbf{0}$ and $Q'_{1:K}$ is incomplete, $T_c(Q)\mathbf{p}$ is not in the column space $C(T_{c'}(Q'))$ for all $c' \in \mathbb{R}^J$.*

Consider the Q -matrix Q_0 in Theorem 3. We have the next result, which is a direct corollary of the above two propositions.

Corollary 1. *Under the conditions of Theorem 3, $T_{c_0}(Q_0)\mathbf{p}_0$ is not in the column space $C(T_c(Q))$ for all $c \in [0, 1]^J$ and all $Q \approx Q_0$.*

To obtain a similar proposition for the cases where the g_i 's are non-zero, we will need to expand the $T_{c,g}(Q)$ as follows. As previously defined, let

$$\tilde{T}_{c,g}(Q) = \begin{pmatrix} T_{c,g}(Q) \\ \mathbb{E} \end{pmatrix}. \quad (2.44)$$

Note that the last row of $\tilde{T}_{c,g}(Q)$ consists entirely of ones.

Proposition 6. *Suppose that Q satisfies conditions C1, C2, and C4, $Q' \approx Q$, $\mathbf{p} \succ 0$, and $c \not\approx g$. Then $\tilde{T}_{c,g}(Q)\mathbf{p}$ is not in the column space $C(\tilde{T}_{c',g}(Q'))$ for all $c' \in [0, 1]^J$. In addition, $\tilde{T}_{c,g}(Q)$ is of full column rank.*

To prove Proposition 6, we will need the following lemma.

Lemma 1. *Consider two matrices T_1 and T_2 of the same dimension. If $T_1\mathbf{p} \in C(T_2)$, then for any matrix D of appropriate dimension for multiplication, we have*

$$DT_1\mathbf{p} \in C(DT_2).$$

Conversely, if for some D , $DT_1\mathbf{p}$ does not belong to $C(DT_2)$, then $T_1\mathbf{p}$ does not belong to $C(T_2)$.

Proof of Lemma 1. Note that DT_i is just a linear row transform of T_i for $i = 1, 2$. The conclusion is immediate by basic linear algebra. \square

Proof of Proposition 6. Thanks to Lemma 1, we only need to find a matrix D such that $D\tilde{T}_{c,g}(Q)\mathbf{p}$ does not belong to the column space of $D\tilde{T}_{c',g}(Q')$ for all $c' \in [0, 1]^m$.

For c, g, c', g' , write

$$c - g = (c_1 - g_1, \dots, c_m - g_m), \quad c' - g' = (c'_1 - g_1, \dots, c'_m - g_m).$$

We claim that there exists a matrix D such that

$$D\tilde{T}_{c,g}(Q) = T_{c-g}(Q) = \left(0, T_{c-g}^*(Q) \right)$$

and

$$D\tilde{T}_{c',g}(Q') = T_{c'-g}(Q') = \left(0, T_{c'-g}^*(Q') \right),$$

where the choice of D does not depend on c or c' .

In the rest of the proof, we will construct such a D -matrix for $\tilde{T}_{c,g}(Q)$ satisfying the above conditions. The verification for $\tilde{T}_{c',g}(Q')$ is completely analogous. Note that each row in $D\tilde{T}_{c,g}(Q)$ is just a linear combination of rows of $\tilde{T}_{c,g}(Q)$. Therefore, it suffices to show that every row vector of the form

$$B_{c-g,Q}(i_1, \dots, i_l)$$

can be written as a linear combination of the row vectors of $\tilde{T}_{c,g}(Q)$. We prove this by induction. First note that for each $1 \leq i \leq J$,

$$B_{c-g,Q}(i) = (c_i - g_i)B_Q(i) = B_{c,g,Q}(i) - g_i\mathbb{E}. \quad (2.45)$$

Suppose that all rows of the form

$$B_{c-g,Q}(i_1, \dots, i_l)$$

for all $1 \leq l \leq j$ can be written as linear combinations of the row vectors of $\tilde{T}_{c,g}(Q)$ with coefficients only depending on g_1, \dots, g_J . Thanks to (2.45), the case of $j = 1$ holds. Suppose the statement holds for some general j . We consider the case of $j + 1$. By definition,

$$\begin{aligned} B_{c,g,Q}(i_1, \dots, i_{j+1}) &= \Upsilon_{h=1}^{j+1} B_{c,g,Q}(i_h) \\ &= \Upsilon_{h=1}^{j+1} (g_{i_h}\mathbb{E} + B_{c-g,Q}(i_h)). \end{aligned} \quad (2.46)$$

Let “*” denote element-by-element multiplication. For every generic vector V' of appropriate length,

$$\mathbb{E} * V' = V'.$$

We expand the right hand side of (2.46). The last term would be

$$B_{c-g,Q}(i_1, \dots, i_{j+1}) = \Upsilon_{h=1}^{j+1} B_{c-g,Q}(i_h),$$

From the induction assumption and definition (2.18), the other terms on both sides of (2.46) belong to the row space of $\tilde{T}_{c,g}(Q)$. Therefore, $(0, B_{c-g,Q}(i_1, \dots, i_{j+1}))$ is also in the row space of $\tilde{T}_{c,g}(Q)$. In addition, all the corresponding coefficients only consist of g_i . Therefore, one can construct a $(2^J - 1) \times 2^J$ matrix D such that

$$D\tilde{T}_{c,g}(Q) = T_{c-g}(Q) = \left(0, T_{c-g}^*(Q) \right).$$

Because D is free of c and Q , we have

$$D\tilde{T}_{c',g}(Q') = T_{c'-g}(Q') = \left(0, T_{c'-g}^*(Q') \right).$$

In addition, thanks to Propositions 4 and 5, $D\tilde{T}_{c,g}(Q)\mathbf{p} = T_{c-g}(Q)\mathbf{p}$ is not in the column space $C(T_{c'-g}(Q')) = C(D\tilde{T}_{c',g}(Q'))$ for all $c' \in [0, 1]^J$. Therefore, by Lemma 1, $\tilde{T}_{c,g}(Q)\mathbf{p}$ is not in the column space $C(\tilde{T}_{c',g}(Q'))$ for all $c' \in [0, 1]^J$.

In addition,

$$\begin{pmatrix} D \\ \mathbf{e}_{2^J}^\top \end{pmatrix} \tilde{T}_{c,g}(Q)$$

is of full column rank, where $\mathbf{e}_{2^J}^\top$ is a 2^J dimension row vector with last element being one and rest being zero. Therefore, $\tilde{T}_{c,g}(Q)$ is also of full column rank. \square

2.9 Technical proofs

Proof of Proposition 4. Note that $Q_{1:k} = Q'_{1:k} = \mathcal{I}_k$. Let $T(\cdot)$ be arranged as in (2.43). Then, $T(Q)_{1:(2^k-1)} = T(Q')_{1:(2^k-1)}$. Given that $Q \neq Q'$, we have $T(Q) \neq T(Q')$. We assume that $T(Q)_{li} \neq T(Q')_{li}$, where $T(Q)_{li}$ is the entry in the l th row and i th column. Since $T(Q)_{1:(2^k-1)} = T(Q')_{1:(2^k-1)}$, it is necessary that $l \geq 2^k$.

Suppose that the l th row of the $T(Q')$ corresponds to an item that requires attributes i_1, \dots, i_ν . Then, we consider $1 \leq h \leq 2^k - 1$, such that the h th row of $T(Q')$ is $B_{Q'}(i_1, \dots, i_\nu)$. Then, the h -th row vector and the l th row vector of $T(Q')$ are identical.

Since $T(Q)_{1:(2^k-1)} = T(Q')_{1:(2^k-1)}$, we have $T(Q)_{hj} = T(Q')_{hj} = T(Q')_{lj}$ for $j = 1, \dots, 2^k - 1$. If $T(Q)_{li} = 0$ and $T(Q')_{li} = 1$, the matrices $T(Q)$ and $T(Q')$ look like

$$T(Q') = \begin{array}{c} \text{column } i \\ \downarrow \\ \text{row } h \rightarrow \\ \text{row } l \rightarrow \end{array} \left(\begin{array}{cccccc} 0 & \mathcal{I} & * & \dots & * & \dots \\ 0 & \vdots & \vdots & & \dots & \dots \\ 0 & \vdots & \vdots & \mathcal{I} & \dots & \dots \\ 0 & \vdots & \vdots & \vdots & & \\ 0 & * & 1 & * & & \\ 0 & * & * & * & & \end{array} \right),$$

and

$$T(Q) = \begin{array}{c} \text{column } i \\ \downarrow \\ \text{row } h \rightarrow \\ \text{row } l \rightarrow \end{array} \begin{pmatrix} 0 & \mathcal{I} & * & \dots & * & \dots \\ 0 & \vdots & \vdots & & \dots & \dots \\ 0 & \vdots & \vdots & \mathcal{I} & \dots & \dots \\ 0 & \vdots & \vdots & \vdots & & \\ 0 & * & 0 & * & & \\ 0 & * & * & * & & \end{pmatrix}.$$

Case 1 Either the h th or l th row vector of $T_{c'}(Q')$ is a zero vector. The conclusion is immediate because all the entries of $T_c(Q)\mathbf{p}$ are non-zero.

Case 2 The h th and l th row vectors of $T_{c'}(Q')$ are nonzero vectors. Suppose that the l th row corresponds to an item. There are three different situations: according to the true Q -matrix (a) the item in row l requires strictly more attributes than row h , (b) the item in row l requires strictly fewer attributes than row h , (c) otherwise. We consider these three situations respectively.

(a) Under the true Q -matrix, there are two types of sub-populations in consideration: people who are able to answer item(s) in row h (p_1) only and people who are able to answer items in both row h and row l (p_2). Then, the sub-matrix of $T_c(Q)$ and $T_{c'}(Q)$ are like

		$T_c(Q)$		$T_{c'}(Q')$		
		p_1	p_2	p_1	p_2	
row h	c_h	c_h	c_h	row h	c'_h	c'_h
row l	0	c_l	c_l	row l	c'_l	c'_l

We now claim that c_l and c'_l must be equal (otherwise the conclusion hold) for the following reason.

Consider the following two rows of $T(Q)$: row A corresponding to the combination that contains all the items; row B corresponding to the row that contains all the items except for the one in row l .

Rows A and B are in fact identical in $T(Q)$. This is because all the attributes are used at least twice (condition C4). Then, the attributes in row l are also required by some other item(s) and rows A and B require the same combination of items. Thus, the corresponding entries of all the column vectors of $T_c(Q)$ are different by a factor of c_l .

For $T(Q')$, rows A and B are also identical. This is because row h and row l have identical attribute requirements. Then, Thus, the corresponding entries of all the column vectors of $T_{c'}(Q)$ are different by a factor of c'_l . Thus, c'_l and c_l must be identical otherwise $T_c(Q)\mathbf{p}$ is not in the column space of $T_{c'}(Q)$.

Similarly, we obtain that $c_h = c'_h$. Given that $c_h = c'_h$ and $c_l = c'_l$, we now consider row h and row l . Notice that all the column vectors in $T_{c'}(Q')$ have their entries in row h and row l different by a factor of c_h/c_l . On the other hand, the h and l th entries of $T_c(Q)\mathbf{p}$ are NOT different by a factor of c_h/c_l as long as the proportion of p_1 is positive. Thereby, we conclude this case.

- (b) Consider the following two types of sub-populations: people who are able to answer item(s) in row l (p_1) only and people who are able to answer items in both row h and row l (p_2). Similar to the analysis of (a), the sub-matrices look like:

$$\begin{array}{cc}
& T_c(Q) & & T_{c'}(Q') \\
& p_1 & p_2 & p_1 & p_2 \\
\text{row } h & 0 & c_h & \text{row } h & 0 & c'_h \\
\text{row } l & c_l & c_l & \text{row } l & 0 & c'_l
\end{array}$$

With exactly the same argument as in (a), we conclude that $c_j = c'_j$, $c_h = c'_h$, and further $T_c(Q)\mathbf{p}$ is not in the column space of $T_{c'}(Q')$.

- (c) Consider the following three types of sub-populations: people who are able to answer item(s) in row l only (p_1), people who are able to answer item(s) in row h only (p_2), and people who are able to answer items in both row h and row l (p_3). The sub-matrices look like:

$$\begin{array}{cc}
& T_c(Q) & & T_{c'}(Q') \\
& p_1 & p_2 & p_3 & p_1 & p_2 & p_3 \\
\text{row } h & 0 & c_h & c_h & \text{row } h & 0 & c'_h & c'_h \\
\text{row } l & c_l & 0 & c_l & \text{row } l & 0 & c'_l & c'_l \\
\text{row } l \wedge h & 0 & 0 & c_h c_l & \text{row } l \wedge h & 0 & c'_h c'_l & c'_h c'_l
\end{array}$$

With the same argument, we have that $c_l = c'_l$ and $c_h = c'_h$. On considering row h and row $l \wedge h$, we conclude that $T_c(Q)\mathbf{p}$ is not in the column space of $T_{c'}(Q')$.

□

Proof of Proposition 5. $T(\cdot)$ is arranged as in (2.43). Consider Q' such that $Q'_{1:k}$ is incomplete. We discuss the following situations.

1. There are two row vectors, say the h th and l th row vectors ($1 \leq i, j \leq k$), in $Q'_{1:k}$ that are identical. Equivalently, two items require exactly the same

attributes according to Q' . With exactly the same argument as in the previous proof, under condition C4, we have that $c_h = c'_h$ and $c_l = c'_l$. We now consider the rows corresponding to l and $l \wedge h$. Note that the elements corresponding to row l and row $l \wedge h$ for all the vectors in the column space of $T_{c'}(Q')$ are different by a factor of c_h . However, the corresponding elements in the vector $T_c(Q)\mathbf{p}$ are NOT different by a factor of c_h as long as the population is fully diversified.

2. No two row vectors in $Q'_{1:k}$ are identical. Then, among the first k rows of Q' there is at least one row vector containing two or more non-zero entries. That is, there exists $1 \leq i \leq k$ such that

$$\sum_{j=1}^k q'_{ij} > 1.$$

This is because if each of the first k items requires only one attribute and $Q'_{1:k}$ is not complete, there are at least two items that require the same attribute. Then, there are two identical row vectors in $Q'_{1:k}$ and it belongs to the first situation. We define

$$a_i = \sum_{j=1}^k q'_{ij},$$

the number of attributes required by item i according to Q' .

Without loss of generality, assume $a_i > 1$ for $i = 1, \dots, n$ and $a_i = 1$ for $i = n+1, \dots, k$. Equivalently, among the first k items, only the first n items require more than one attribute while the $(n+1)$ -th through the k th items require only one attribute each, all of which are distinct. Without loss of generality, we assume $q'_{ii} = 1$ for $i = n+1, \dots, k$ and $q_{ij} = 0$ for $i = n+1, \dots, k$ and $i \neq j$.

- (a) $n = 1$. Since $a_1 > 1$, there exists an $l > 1$ such that $q'_{1l} = 1$. We now consider rows 1 and l . With the same argument as before (i.e., the attribute required by row l is also required by item 1 in Q'), we have that $c_l = c'_l$ (be careful that we cannot claim that $c_1 = c'_1$). We now consider the rows 1 and $1 \wedge l$. Note that in $T_{c'}(Q')$ these two rows are different by a factor of c_l ; while the corresponding entries in $T_c(Q)\mathbf{p}$ are NOT different by a factor of c_l . Thereby, we conclude the result in this situation.
- (b) $n > 1$ and there exists $j > n$ and $i \leq n$ such that $q'_{ij} = 1$. The argument is identical to that in (2a).
- (c) $n > 1$ and for each $j > n$ and $i \leq n$, $q'_{ij} = 0$. Let the i^* -th row in $T(Q')$ correspond to $I_1 \wedge, \dots, \wedge I_n$. Let the i_h^* -th row in $T(Q')$ correspond to $I_1 \wedge, \dots, \wedge I_{h-1} \wedge I_{h+1} \wedge, \dots, \wedge I_n$ for $h = 1, \dots, n$.

We claim that there exists an h such that the i^* -th row and the i_h^* -th row are identical in $T(Q')$, that is

$$B_{Q'}(1, \dots, h-1, h+1, \dots, n) = B_{Q'}(1, \dots, n). \quad (2.47)$$

If the above claim is true, then the attributes required by item h have been required by some other items. Then, we conclude that c_h and c'_h must be identical. In addition, rows in $T_{c'}(Q')$ corresponding to $I_1 \wedge, \dots, \wedge I_{h-1} \wedge I_{h+1} \wedge, \dots, \wedge I_n$ and $I_1 \wedge, \dots, \wedge I_n$ are different by a factor of c_h . On the other hand, the corresponding entries in $T_c(Q)\mathbf{p}$ are NOT different by a factor of c_h . Then, we are able to conclude the results for all the cases.

In what follows, we prove the claim in (2.47) by contradiction. Suppose that there does not exist such an h . This is equivalent to saying that for each $j \leq n$ there exists an α_j such that $q'_{j\alpha_j} = 1$ and $q'_{i\alpha_j} = 0$ for all

$1 \leq i \leq n$ and $i \neq j$. Equivalently, for each $j \leq n$, item j requires at least one attribute that is not required by other first n items. Consider

$$\mathcal{C}_i = \{j : \text{there exists } i \leq i' \leq n \text{ such that } q'_{i'j} = 1\}.$$

Let $\#(\cdot)$ denote the cardinality of a set. Since for each $i \leq n$ and $j > n$, $q'_{ij} = 0$, we have that $\#(\mathcal{C}_1) \leq n$. Note that $q'_{1\alpha_1} = 1$ and $q'_{i\alpha_1} = 0$ for all $2 \leq i \leq n$. Therefore, $\alpha_1 \in \mathcal{C}_1$ and $\alpha_1 \notin \mathcal{C}_2$. Therefore, $\#(\mathcal{C}_2) \leq n - 1$. By a similar argument and induction, we have that $a_n = \#(\mathcal{C}_n) \leq 1$. This contradicts the fact that $a_n > 1$. Therefore, there exists an h such that (2.47) is true. As for $T(Q)$, we have that

$$B_Q(1, \dots, h-1, h+1, \dots, n) \neq B_Q(1, \dots, n).$$

Summarizing the cases in 1, 2(a), 2(b), and 2(c), we conclude the proof.

□

Chapter 3

Hypothesis Testing of Q -matrix

3.1 Introduction

There is a growing interest in statistical inference of Q -matrix-based diagnostic classification models (Rupp 2002; Henson and Templin 2005; Roussos, Templin and Henson 2007b; Stout 2007). One simple and widely studied model among them is the DINA model (Deterministic Input, Noisy output “AND” gate; see Junker and Sijtsma 2001), under which it is required to master all attributes specified in Q for an item to get a correct answer. Other important models and developments can be found in Tatsuoka (1985); Hartz (2002); Leighton, Gierl and Hunka (2004); von Davier (2005); Templin and Henson (2006); DeCarlo (2011). See Rupp, Templin and Henson (2010) for a more thorough review of diagnostic models.

A correctly specified Q -matrix is crucial both for parameter estimation (such as the slipping and the guessing parameters in the DINA model) and for the specification of subjects’ latent attributes (Rupp and Templin 2008a; de la Torre and Douglas 2004). A misspecified Q -matrix may lead to substantial lack of fit and, consequently,

erroneous classification of subjects. In the literature, however, the Q -matrix is usually assumed as correct after its construction. Thus, it is desirable to be able to detect misspecification of the Q -matrix.

Chapter 2 gives definitive answers to the learnability of the Q -matrix from the response data and provides consistent estimators of the Q -matrix (see also Liu, Xu and Ying 2012b;a). In application, however, the estimated Q -matrix may not be the true one, especially when the sample size (the number of examinees) is limited. In this case, it is of more interest is to check whether a Q -matrix provided by experts fits the response data. This imposes the need of efficient Q -matrix validation procedures. While empirically based methods of validating the Q -matrix in the DINA model analysis have been proposed in the literature (de la Torre 2008a; DeCarlo 2012), the corresponding theoretical justification is still lacking.

This chapter focuses on the problem of validating a prespecified Q -matrix, either from experts or estimated based upon the response data, under a general cognitive diagnosis model. Statistically, this problem can be formulated as a pure significance testing problem with null hypothesis $H_0 : Q = Q_0$, where Q_0 is the candidate Q -matrix. In this chapter we construct a test statistic that measures the consistency of observed data with the proposed Q -matrix Q_0 . Asymptotic distributions of the test statistic are derived under different diagnostic models. In addition, we provide computational algorithms, and conduct simulation studies to assess the performance of the proposed testing procedure.

It is worth pointing out that our testing procedure is generic in the sense that it covers a large class of diagnostic models, including the DINA, DINO (Deterministic Input, Noisy output “OR” gate) model, the NIDA (Noisy Inputs, Deterministic “And” Gate) model, the NIDO (Noisy Inputs, Deterministic “Or” Gate) model, and the RUM (Reparameterized Unified Model) among others.

The remainder of this chapter is organized as follows. We introduce in Section 3.2 the validation (testing) procedure. Section 3.3 derives the asymptotic distributions of the testing statistic under different assumptions. Section 3.4 includes simulation results to help assess the performance of the proposed testing procedure. Finally, technical proofs are provided in the appendix section.

3.2 Q -matrix validation procedure

As in Chapter 2, we are concerned with the situation where N subjects take a test consisting of J items. We assume that the responses are binary, so that the data will be an $N \times J$ matrix with entries being 0 or 1. The diagnostic classification model to be considered for such data envisions K attributes that are related to both the subjects and the items. We assume that the number of attributes K is known and that the number of items J is observed.

3.2.1 Notation

We first recall some notation introduced in Chapter 1. The Q -matrix provides a link between the items and the attributes. In particular, $Q = (q_{jk})_{J \times K}$ is a $J \times K$ matrix with binary entries. For each j and k , $q_{jk} = 1$ indicates that item j requires attribute k and $q_{jk} = 0$ otherwise. Moreover, we use \mathbf{q}_j to denote the j th row of Q .

We say a Q -matrix is *complete* if for any attribute, there exists an item only requiring that attribute, i.e., for any integer k ($1 \leq k \leq K$), we have \mathbf{e}_k lying in the vector set $\{\mathbf{q}_j, j = 1, \dots, J\}$, where \mathbf{e}_k is a K dimensional row vector such that the k th element is one and the rest are zeroes; see Definition 3 in Chapter 2 for more details. An example of a complete Q -matrix is the K by K identity matrix.

We use $\boldsymbol{\alpha} = (\alpha^1, \dots, \alpha^K)^\top$ to denote the vector of attributes, where $\alpha^k = 1$ or 0 , indicating the presence or absence of the k th attribute, $k = 1, \dots, K$, and superscript \top denotes transpose. In addition, let $\mathbf{R} = (R^1, \dots, R^J)^\top$ denote the vector of responses to the J test items. Note that both $\boldsymbol{\alpha}$ and \mathbf{R} are subject-specific.

Consider a general diagnostic model. For notational convenience, we use $\boldsymbol{\theta}$ to denote the vector of unknown item parameters. Given a specific subject's profile $\boldsymbol{\alpha}$, the response R^j to item j under the corresponding model follows a Bernoulli distribution

$$P(R^j|Q, \boldsymbol{\alpha}, \boldsymbol{\theta}) = (c_{j,\boldsymbol{\alpha}})^{R^j} (1 - c_{j,\boldsymbol{\alpha}})^{1-R^j}, \quad (3.1)$$

where $c_{j,\boldsymbol{\alpha}}$ is the probability of providing correct response to item j for subjects with $\boldsymbol{\alpha}$, i.e.,

$$c_{j,\boldsymbol{\alpha}} = P(R^j = 1|Q, \boldsymbol{\alpha}, \boldsymbol{\theta}).$$

In addition, conditional on $\boldsymbol{\alpha}$, we assume (R^1, \dots, R^J) are jointly independent.

Note that the specific form of $c_{j,\boldsymbol{\alpha}}$ depends on the Q -matrix, the item parameter vector $\boldsymbol{\theta}$, and diagnostic model assumptions. We use the following examples for an illustration.

Example 1 (DINA model). *Under the DINA model, the item parameters are specified by*

$$\boldsymbol{\theta} = \{s_j, g_j; j = 1, \dots, J\},$$

where s_j and g_j represent the slipping and guessing parameters for the j th item (Macready and Dayton 1977; Junker and Sijtsma 2001). For a Q -matrix Q , $c_{j,\boldsymbol{\alpha}}$ takes the form

$$c_{j,\boldsymbol{\alpha}} = (1 - s_j)^{\xi_{DINA}^j(\boldsymbol{\alpha}, Q)} g_j^{1 - \xi_{DINA}^j(\boldsymbol{\alpha}, Q)}, \quad (3.2)$$

where

$$\xi_{DINA}^j(\boldsymbol{\alpha}, Q) = \prod_{k=1}^K (\alpha^k)^{q_{jk}}. \quad (3.3)$$

The DINA model assumes conjunctive relationship among attributes, that is, it is necessary to possess all the attributes indicated by the Q -matrix to be capable of providing a positive response to an item. In addition, having additional unnecessary attributes does not compensate for a lack of the necessary attributes.

Example 2 (DINO model). Under the DINO model, the item parameters are specified by

$$\boldsymbol{\theta} = \{s_j, g_j; j = 1, \dots, J\}$$

with s_j and g_j the slipping and guessing parameters for the j th item. Then $c_{j,\boldsymbol{\alpha}}$ takes the form

$$c_{j,\boldsymbol{\alpha}} = (1 - s_j)^{\xi_{DINO}^j(\boldsymbol{\alpha}, Q)} g_j^{1 - \xi_{DINO}^j(\boldsymbol{\alpha}, Q)}, \quad (3.4)$$

where

$$\xi_{DINO}^j(\boldsymbol{\alpha}, Q) = 1 - \prod_{k=1}^K (1 - \alpha_k)^{q_{jk}}. \quad (3.5)$$

In contrast to the DINA model, the DINO model assumes non-conjunctive relationship among attributes, that is, it only needs to possess one of the attributes required by the Q -matrix to be capable of providing a positive response to an item.

Example 3 (Reduced RUM model). Under the reduced version of the Reparameterized Unified Model (RUM), we have

$$c_{j,\boldsymbol{\alpha}} = \pi_j \prod_{k=1}^K r_{jk}^{q_{jk}(1 - \alpha_k)}, \quad (3.6)$$

where π_j is the correct response probability for subjects who possess all required attributes and $r_{j,k}$, $0 < r_{j,k} < 1$, is the penalty parameter for not possessing the k th

attribute. Then the corresponding item parameters are

$$\boldsymbol{\theta} = \{\pi_j, r_{j,k}; j = 1, \dots, J, k = 1, \dots, K\}.$$

The reduced RUM model is also a conjunctive model, and it generalizes the DINA model by allowing the slipping and guessing parameters to vary across different attribute profiles.

Lastly, we use subscripts to indicate different subjects. For instance, $\mathbf{R}_i = (R_i^1, \dots, R_i^J)^\top$ is the response vector of subject i . Similarly, $\boldsymbol{\alpha}_i = (\alpha_i^1, \dots, \alpha_i^K)$ is the attribute vector of subject i . With N subjects, we observe $\mathbf{R}_1, \dots, \mathbf{R}_N$ but not $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N$. We further assume that the attribute profiles are i.i.d. and let

$$p_{\boldsymbol{\alpha}} = P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}) \quad \text{and} \quad \mathbf{p} = (p_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \{0, 1\}^K)^\top.$$

For statement simplicity, we let the first element of \mathbf{p} be

$$p_{\mathbf{0}} = P(\boldsymbol{\alpha}_i = (0, \dots, 0)^\top),$$

and define the other part of \mathbf{p} as \mathbf{p}^* , i.e.,

$$\mathbf{p} = \begin{pmatrix} p_{\mathbf{0}} \\ \mathbf{p}^* \end{pmatrix}. \quad (3.7)$$

In this section, we assume that $\mathbf{p} \succ \mathbf{0}$, i.e., $p_{\boldsymbol{\alpha}} > 0$ for all $\boldsymbol{\alpha} \in \{0, 1\}^K$. We use vector $\mathbf{0}$ and $\mathbf{1}$ to denote the zero vector and one vector, i.e., $(0, \dots, 0)$ and $(1, \dots, 1)$ respectively.

3.2.2 Validation procedure

In application, the Q -matrix is usually assumed to be correct (denoted by Q_0) after it has been constructed, without further validation. Consequently, one raising issue is

whether such Q_0 is appropriate for the cognitive diagnosis analysis. This problem can be formulated as a pure significance testing problem with null hypothesis $H_0 : Q = Q_0$ (see Chapter 3 of Cox and Hinkley 1974), with the goal of checking whether the observed data is consistent with the current model formulation while there is not an alternative model particularly specified. In the following, we propose a general statistical procedure to carry out this pure significance test under a general diagnostic model.

Following the hypothesis framework, our first step is to construct a test statistic that measures how well a given matrix Q fits the data. This is based on the development of Chapter 2, where a Q -matrix estimation method was proposed. To formalize, we need the following notation. First we define a T -matrix for a general diagnostic classification model.

T -matrix. Recall that the T -matrix serves as a connection between the observed response distribution and the model structure. We first specify each row vector of the T -matrix for a general diagnostic model. For each item j , we have

$$P(R^j = 1|Q, \mathbf{p}, \boldsymbol{\theta}) = \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} P(R^j = 1|Q, \boldsymbol{\alpha}, \boldsymbol{\theta}) = \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} c_{j,\boldsymbol{\alpha}}, \quad (3.8)$$

If we create a row vector $B_{\boldsymbol{\theta},Q}(j)$ of length 2^K containing the probabilities $c_{j,\boldsymbol{\alpha}}$ for all $\boldsymbol{\alpha}$'s and arrange those elements in an appropriate order, then for all j we can write (3.8) in the form of a matrix product

$$\sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} c_{j,\boldsymbol{\alpha}} = B_{\boldsymbol{\theta},Q}(j) \mathbf{p},$$

where \mathbf{p} is the column vector containing the probabilities $p_{\boldsymbol{\alpha}}$. Similarly, for each pair of items, we may establish that the probability of responding positively to both items

j_1 and j_2 is

$$P(R^{j_1} = 1, R^{j_2} = 1 | Q, \mathbf{p}, \boldsymbol{\theta}) = \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} c_{j_1, \boldsymbol{\alpha}} c_{j_2, \boldsymbol{\alpha}} = B_{\boldsymbol{\theta}, Q}(j_1, j_2) \mathbf{p},$$

where $B_{\boldsymbol{\theta}, Q}(j_1, j_2)$ is a row vector containing the probabilities $c_{j_1, \boldsymbol{\alpha}} \cdot c_{j_2, \boldsymbol{\alpha}}$ for each $\boldsymbol{\alpha}$. Note that each element of $B_{\boldsymbol{\theta}, Q}(j_1, j_2)$ is the product of the corresponding elements of $B_{\boldsymbol{\theta}, Q}(j_1)$ and $B_{\boldsymbol{\theta}, Q}(j_2)$. With a completely analogous construction, we have that

$$P(R^{j_1} = 1, \dots, R^{j_l} = 1 | Q, \mathbf{p}, \boldsymbol{\theta}) = B_{\boldsymbol{\theta}, Q}(j_1, \dots, j_l) \mathbf{p},$$

for each combination of distinct (j_1, \dots, j_l) . Similarly, $B_{\boldsymbol{\theta}, Q}(j_1, \dots, j_l)$ is the element-by-element product of $B_{\boldsymbol{\theta}, Q}(j_1), \dots, B_{\boldsymbol{\theta}, Q}(j_l)$. From a computational point of view, one only needs to construct the $B_{\boldsymbol{\theta}, Q}(j)$'s for each individual item j and then take products to obtain the corresponding combinations.

The T -matrix has 2^K columns. Each row vector of the T -matrix is one of the vectors $B_{\boldsymbol{\theta}, Q}(j_1, \dots, j_l)$, i.e., the T -matrix is a stack of B -vectors

$$T_{\boldsymbol{\theta}}(Q) = \begin{pmatrix} B_{\boldsymbol{\theta}, Q}(1) \\ \vdots \\ B_{\boldsymbol{\theta}, Q}(J) \\ B_{\boldsymbol{\theta}, Q}(1, 2) \\ \vdots \end{pmatrix}. \quad (3.9)$$

By the definition of the B -vectors, we have that

$$T_{\boldsymbol{\theta}}(Q)\mathbf{p} = \begin{pmatrix} P(R_i^1 = 1|Q, \boldsymbol{\theta}, \mathbf{p}) \\ \vdots \\ P(R_i^J = 1|Q, \boldsymbol{\theta}, \mathbf{p}) \\ P(R_i^1 = 1, R_i^2 = 1|Q, \boldsymbol{\theta}, \mathbf{p}) \\ \vdots \end{pmatrix} = \begin{pmatrix} \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} c_{1,\boldsymbol{\alpha}} \\ \vdots \\ \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} c_{J,\boldsymbol{\alpha}} \\ \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} c_{1,\boldsymbol{\alpha}} c_{2,\boldsymbol{\alpha}} \\ \vdots \end{pmatrix} \quad (3.10)$$

is a vector containing the corresponding probabilities associated with a particular set of parameters $(Q, \boldsymbol{\theta}, \mathbf{p})$.

In the following, we use n to denote the number of rows of T . We say a T -matrix is *saturated* if it contains all the possible combinations of items, i.e., $n = 2^J - 1$. For each Q and the corresponding parameters $\boldsymbol{\theta}$, we denote the saturated T -matrix by $T_{\boldsymbol{\theta}}^{all}(Q)$.

$\boldsymbol{\gamma}$ -vector. We further define $\boldsymbol{\gamma}$ to be the $n \times 1$ vector containing the probabilities (corresponding to those in (3.10)) of the empirical distribution, e.g., the first element of $\boldsymbol{\gamma}$ is $\frac{1}{N} \sum_{i=1}^N I(R_i^1 = 1)$ and the $(J + 1)$ -th element is $\frac{1}{N} \sum_{i=1}^N I(R_i^1 = 1, R_i^2 = 1)$, i.e.,

$$\boldsymbol{\gamma} = \begin{pmatrix} \frac{1}{N} \sum_{i=1}^N I(R_i^1 = 1) \\ \vdots \\ \frac{1}{N} \sum_{i=1}^N I(R_i^J = 1) \\ \frac{1}{N} \sum_{i=1}^N I(R_i^1 = 1, R_i^2 = 1) \\ \vdots \end{pmatrix}. \quad (3.11)$$

For the saturated matrix $T_{\boldsymbol{\theta}}^{all}(Q)$, we denote the corresponding $\boldsymbol{\gamma}$ vector by $\boldsymbol{\gamma}^{all}$.

We use an example to illustrate the T -matrix and $\boldsymbol{\gamma}$ -vector under different diagnostic models.

Example 4. Suppose that we are interested in testing two attributes. The population is naturally divided into four strata with the corresponding distribution vector $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})^\top$. Consider a test containing three problems and admitting the following Q -matrix,

$$Q_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}. \quad (3.12)$$

Choose $n = 4$, that is, T is a 4×4 matrix. Let the first to fourth columns of $T_\theta(Q_0)$ be indexed by attribute profiles $(0, 0)$, $(1, 0)$, $(0, 1)$, and $(1, 1)$ respectively. In addition, the first three rows of $T(Q_0)$ correspond to items one, two and three, and the fourth row corresponds to the combination of items one and two. Then under the above Q_0 , the T -matrix and the corresponding γ are

$$T_\theta(Q_0) = \begin{pmatrix} c_{1,(0,0)} & c_{1,(1,0)} & c_{1,(0,1)} & c_{1,(1,1)} \\ c_{2,(0,0)} & c_{2,(1,0)} & c_{2,(0,1)} & c_{2,(1,1)} \\ c_{3,(0,0)} & c_{3,(1,0)} & c_{3,(0,1)} & c_{3,(1,1)} \\ c_{1,(0,0)}c_{2,(0,0)} & c_{1,(1,0)}c_{2,(1,0)} & c_{1,(0,1)}c_{2,(0,1)} & c_{1,(1,1)}c_{2,(1,1)} \end{pmatrix}, \quad (3.13)$$

and

$$\gamma = \begin{pmatrix} N_1/N \\ N_2/N \\ N_3/N \\ N_{1\wedge 2}/N \end{pmatrix}, \quad (3.14)$$

where N_j is the total number correct responses to item j and $N_{1\wedge 2}$ is the number of correct responses to both item 1 and 2.

Under the DINA model in Example 1, the T -matrix in (3.13) becomes

$$T_{DINA,\boldsymbol{\theta}}(Q_0) = \begin{pmatrix} g_1 & 1 - s_1 & g_1 & 1 - s_1 \\ g_2 & g_2 & 1 - s_2 & 1 - s_2 \\ g_3 & g_3 & g_3 & 1 - s_3 \\ g_1 g_2 & (1 - s_1)g_2 & g_1(1 - s_2) & (1 - s_1)(1 - s_2) \end{pmatrix},$$

where s and g are the slipping and guessing parameters in the DINA model. Similarly under the DINO model in Example 2 and the reduced RUM in Example 3, we have

$$T_{DINO,\boldsymbol{\theta}}(Q_0) = \begin{pmatrix} g_1 & 1 - s_1 & g_1 & 1 - s_1 \\ g_2 & g_2 & 1 - s_2 & 1 - s_2 \\ g_3 & 1 - s_3 & 1 - s_3 & 1 - s_3 \\ g_1 g_2 & (1 - s_1)g_2 & g_1(1 - s_2) & (1 - s_1)(1 - s_2) \end{pmatrix}$$

and

$$T_{RUM,\boldsymbol{\theta}}(Q_0) = \begin{pmatrix} \pi_1 r_{1,1} & \pi_1 & \pi_1 r_{1,1} & \pi_1 \\ \pi_2 r_{2,2} & \pi_2 r_{2,2} & \pi_2 & \pi_2 \\ \pi_3 r_{3,1} r_{3,2} & \pi_3 r_{3,2} & \pi_3 r_{3,1} & \pi_3 \\ \pi_1 r_{1,1} \pi_2 r_{2,2} & \pi_1 \pi_2 r_{2,2} & \pi_1 r_{1,1} \pi_2 & \pi_1 \pi_2 \end{pmatrix}.$$

Testing Statistic. Under the null hypothesis that Q_0 is true, let $(\boldsymbol{\theta}_0, \mathbf{p}_0)$ be the corresponding true model parameters.

By the the law of large number, we have that under the null hypothesis,

$$\boldsymbol{\gamma} = \begin{pmatrix} \frac{1}{N} \sum_{i=1}^N I(R_i^1 = 1) \\ \vdots \\ \frac{1}{N} \sum_{i=1}^N I(R_i^J = 1) \\ \frac{1}{N} \sum_{i=1}^N I(R_i^1 = 1, R_i^2 = 1) \\ \vdots \end{pmatrix} \rightarrow \begin{pmatrix} P(R_i^1 = 1 | Q_0, \boldsymbol{\theta}_0, \mathbf{p}_0) \\ \vdots \\ P(R_i^J = 1 | Q_0, \boldsymbol{\theta}_0, \mathbf{p}_0) \\ P(R_i^1 = 1, R_i^2 = 1 | Q_0, \boldsymbol{\theta}_0, \mathbf{p}_0) \\ \vdots \end{pmatrix}$$

almost surely as $N \rightarrow \infty$. Then by the relationship in (3.10), the above result is equivalent to

$$\boldsymbol{\gamma} - T_{\boldsymbol{\theta}_0}(Q_0) \mathbf{p}_0 \rightarrow \mathbf{0} \quad (3.15)$$

almost surely as $N \rightarrow \infty$.

Our test statistic $S(Q_0)$ is constructed based on the above observation. In particular, if $\boldsymbol{\theta}_0$ and \mathbf{p}_0 are known, $S(Q_0)$ is defined as

$$S(Q_0) = |T_{\boldsymbol{\theta}_0}(Q_0)\mathbf{p}_0 - \boldsymbol{\gamma}|^2 \quad (3.16)$$

where $|\cdot|$ is the Euclidean norm. If all the parameters are correctly specified, we expect that $S(Q_0) \rightarrow 0$ as $N \rightarrow \infty$.

Most of the time, the true parameters $(\boldsymbol{\theta}_0, \mathbf{p}_0)$ are unknown. Under these situations, we consider the profiled objective function and define

$$S(Q_0) = \inf_{\boldsymbol{\theta}, \mathbf{p}} |T_{\boldsymbol{\theta}}(Q_0)\mathbf{p} - \boldsymbol{\gamma}|^2, \quad (3.17)$$

where the minimization is subject to the natural constraints that $\boldsymbol{\theta}, p_\alpha \in (0, 1)$ and $\sum_\alpha p_\alpha = 1$.

The minimization of \mathbf{p} in (3.17) consists of a quadratic optimization with linear constraints, and therefore can be done efficiently. The minimization with respect to $\boldsymbol{\theta}$ is usually not straightforward. One may alternatively replace the minimization by other estimators (such as the maximum likelihood estimators $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{p}})$ derived under Q_0). Thus, the test statistic $S(Q_0)$ becomes

$$\hat{S}(Q_0) = |T_{\hat{\boldsymbol{\theta}}}(Q_0)\hat{\mathbf{p}} - \boldsymbol{\gamma}|^2. \quad (3.18)$$

This alternative allows certain flexibility in the estimation procedure. The S -function in (3.18) is usually easier to compute.

Validation Procedure

The statistic $\hat{S}(Q_0)$ constructed above is similar to the deviance statistics. It shows the L_2 distance between the observed data frequencies of the responses and those suggested by the matrix Q_0 . Therefore, under the null hypothesis $H_0 : Q = Q_0$, one may expect that the testing statistic

$$\hat{S}(Q_0) \rightarrow 0$$

almost surely as $N \rightarrow \infty$.

Then we construct the test procedure based on $\hat{S}(Q_0)$ and reject the null hypothesis H_0 if $s(Q_0)$, the realization of $\hat{S}(Q_0)$ calculated based on response data, is above some threshold number close to zero. Specifically, we reject $H_0 : Q = Q_0$ if $s(Q_0) > v$ for $v > 0$ satisfying

$$P\left(\hat{S}(Q_0) > v_\alpha \mid Q_0\right) = \alpha,$$

where $\alpha \in (0, 1)$ is a prespecified test significance level. The value of v_α can be calculated according the asymptotic distribution of $\hat{S}(Q_0)$ under the null hypothesis, which is shown in Section 3.3.

3.2.3 Computations

In this subsection, we consider the computation of testing procedure when $(\boldsymbol{\theta}_0, \mathbf{p}_0)$ are all unknown. We focus on the case that estimators $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{p}})$ in (2.37) are maximum likelihood estimators (MLE). The computation of $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{p}})$ can be done efficiently by the EM algorithm (Dempster, Laird and Rubin 1977) for most diagnostic classification models, including the DINA model (de la Torre 2009). Then for each candidate Q_0 , the test procedure proposed in the last subsection can be summarized in the following algorithm:

Algorithm 2. *We perform the following steps.*

1. Evaluate MLE $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{p}})$ under the null hypothesis $H_0 : Q = Q_0$.
2. Calculate test statistic

$$s(Q_0) = |T_{\hat{\boldsymbol{\theta}}}(Q_0)\hat{\mathbf{p}} - \boldsymbol{\gamma}|^2$$

3. Following the asymptotic distribution derived in Section 3.3.2, find value v_α such that

$$P\left(\hat{S}(Q_0) > v_\alpha \mid Q_0\right) = \alpha,$$

where $\alpha \in (0, 1)$ is the prespecified testing significance level.

4. Reject the null hypothesis $H_0 : Q = Q_0$ if $s(Q_0) > v_\alpha$, otherwise accept it.

One potential computational concern of the above algorithm is with the case where the item number J is large. The evaluation of the asymptotic distribution of $\hat{S}(Q_0)$ needs to specify a $(2^J - 1) \times (2^J - 1)$ covariance matrix Σ of $\boldsymbol{\gamma}^{all}$ (details will be provided in Section 3.3). This is a substantial computational load if J are reasonably large, for instance $J > 20$. The computation may be reduced by splitting the Q -matrix into small sub-matrices. In particular, if there are 100 items, one can handle such a situation as follows. First, split the 100 items into 10 groups (without overlapping items between groups); then apply the procedure to each of the 10 groups of items respectively. This is equivalent to breaking a big $100 \times K$ Q -matrix into 10 smaller matrices and estimating each of them separately. Lastly, combine the 10 calculated U values together to form a single test statistic.

The idea is described in the following algorithm:

Algorithm 3. *We perform the following steps.*

1. Split the Q_0 into M smaller Q -matrices, $\{Q_{0,1}, \dots, Q_{0,M}\}$, by row (without overlapping items among sub-matrices) such that for each sub-matrix $Q_{0,m}$, there are no more than 15 items.
2. For each sub-matrix $Q_{0,m}$, evaluate MLE $(\hat{\theta}_m, \hat{\mathbf{p}}_m)$ and response vector γ_m only using the response data related to $Q_{0,m}$.
3. Calculate test statistic for each $Q_{0,m}$,

$$s_m(Q_0) = |T_{\hat{\theta}_m}(Q_{0,m})\hat{\mathbf{p}}_m - \gamma_m|^2, \quad m = 1, \dots, M.$$

4. Following the derivation in Section 3.3.2, find value v_α such that

$$P\left(\sum_{m=1}^M \hat{S}_m(Q_0) > v_\alpha \mid Q_0\right) = \alpha,$$

where $\alpha \in (0, 1)$ is the prespecified test significance level.

5. Reject the null hypothesis $H_0 : Q = Q_0$ if $\sum_{m=1}^M s_m(Q_0) > v_\alpha$, otherwise accept it.

Remark 6. For typical statistical models, dividing the parameter space is usually impossible. The Q -matrix adopts a particular structure with which there is certain independence among items so that splitting the Q -matrix is valid. Similar techniques have been employed in the literature, such as Chapter 8.6 in the Tatsuoka (2009) with large scale empirical studies in that chapter.

3.3 Distribution of test statistic $S(Q_0)$

In this section, we derive the asymptotic distribution of the testing statistic $S(Q_0)$ under the null hypothesis $H_0 : Q = Q_0$. We start with the ideal case where the item parameters θ_0 are known, and then move on to the more general case.

3.3.1 Case of known item parameters

When \mathbf{p}_0 is known. To illustrate the idea, we start with the simplest case where \mathbf{p}_0 is also known. In this case,

$$S(Q_0) = |T_{\boldsymbol{\theta}_0}(Q_0)\mathbf{p}_0 - \boldsymbol{\gamma}|^2$$

and we only need to derive the distribution of $T_{\boldsymbol{\theta}_0}(Q_0)\mathbf{p}_0 - \boldsymbol{\gamma}$. A direct application of the central limit theorem gives the following result:

$$\sqrt{N}(\boldsymbol{\gamma} - T_{\boldsymbol{\theta}_0}(Q_0)\mathbf{p}_0) \xrightarrow{d} \mathcal{N}(0, \Sigma), \text{ as } N \rightarrow \infty, \quad (3.19)$$

where Σ is an $n \times n$ covariance matrix specified as follows.

Suppose the i th row of Σ corresponds to the combination of items (i_1, \dots, i_k) , same as the i th row of $T_{\boldsymbol{\theta}_0}(Q_0)$, and suppose that the j th column corresponds to items (j_1, \dots, j_l) . Denote the set $\{i_1, \dots, i_k\}$ and $\{j_1, \dots, j_l\}$ by S_i and S_j respectively. Then $\Sigma_{i,j}$ is given by

$$\begin{aligned} \Sigma_{i,j} &= P_0(R^h = 1, h \in S_i \cup S_j) - P_0(R^h = 1, h \in S_i)P_0(R^h = 1, h \in S_j) \\ &= B_{Q_0, \boldsymbol{\theta}_0}(S_i \cup S_j)\mathbf{p}_0 - (B_{Q_0, \boldsymbol{\theta}_0}(S_i)\mathbf{p}_0) \cdot (B_{Q_0, \boldsymbol{\theta}_0}(S_j)\mathbf{p}_0), \end{aligned} \quad (3.20)$$

where

$$P_0(R^h = 1, h \in S_i) = P(R^h = 1, \text{ for all } h \in S_i | Q_0, \boldsymbol{\theta}_0, \mathbf{p}_0)$$

is the probability of providing correct answer to all the items in the set $S_i = \{i_1, \dots, i_k\}$ under the null hypothesis and the corresponding parameters.

Therefore, we have the asymptotic distribution

$$N S(Q_0) = N|\boldsymbol{\gamma} - T_{\boldsymbol{\theta}_0}(Q_0)\mathbf{p}_0|^2 \xrightarrow{d} \sum_{l=1}^n \lambda_l Z_l^2,$$

where $\{\lambda_l, l = 1, \dots, n\}$ are eigenvalues of Σ , and $(Z_1, \dots, Z_n)^\top \sim^d \mathcal{N}(0, \mathcal{I}_n)$.

When \mathbf{p}_0 is unknown. In this case, we have the test statistic as given in (3.17):

$$S(Q_0) = \inf_{\mathbf{p}} |T_{\theta_0}(Q_0)\mathbf{p} - \gamma|^2. \quad (3.21)$$

The minimization of \mathbf{p} in (3.21) consists of a quadratic optimization with linear constraints, and therefore can be done efficiently. Let $\hat{\mathbf{p}}$ be the estimator obtained in the above minimization equation, i.e.,

$$\hat{\mathbf{p}} = \arg \inf_{\mathbf{p}} |T_{\theta_0}(Q_0)\mathbf{p} - \gamma|^2$$

subject to the constraint that $\sum_{\alpha} p_{\alpha} = 1$. Then the objective function becomes

$$S(Q_0) = |T_{\theta_0}(Q_0)\hat{\mathbf{p}} - \gamma|^2.$$

Note that we have the following transformation:

$$|\gamma - T_{\theta_0}(Q_0)\mathbf{p}|^2 = |\gamma - \mathbf{t}_1 - (T_{\theta_0}^*(Q_0) - \mathbf{t}_1\mathbf{1})\mathbf{p}^*|^2,$$

where \mathbf{t}_1 is the first column of matrix $T_{\theta_0}(Q_0)$ and $T_{\theta_0}^*(Q_0)$ is the other part, i.e.,

$$T_{\theta_0}(Q_0) = (\mathbf{t}_1, T_{\theta_0}^*(Q_0)),$$

and \mathbf{p}^* is defined as in Section 3.2.1 such that $\mathbf{p} = (p_0, (\mathbf{p}^*)^\top)^\top$. Let

$$\tilde{T} = T_{\theta_0}^*(Q_0) - \mathbf{t}_1\mathbf{1}.$$

Then, to find $\hat{\mathbf{p}}$, it is equivalent to obtain $\hat{\mathbf{p}}^*$ minimizing the equation

$$|\gamma - \mathbf{t}_1 - \tilde{T}\mathbf{p}^*|^2.$$

This can be taken as a linear regression problem, and we have the solution that (suppose matrix $\tilde{T}^\top\tilde{T}$ is invertible)

$$\hat{\mathbf{p}}^* = \left(\tilde{T}^\top\tilde{T}\right)^{-1} \tilde{T}^\top(\gamma - \mathbf{t}_1).$$

Thanks to the law of large numbers,

$$\begin{aligned}\hat{\mathbf{p}}^* &= \left(\tilde{T}^\top \tilde{T}\right)^{-1} \tilde{T}^\top (\boldsymbol{\gamma} - \mathbf{t}_1) \rightarrow \left(\tilde{T}^\top \tilde{T}\right)^{-1} \tilde{T}^\top (T_{\boldsymbol{\theta}_0}(Q_0)\mathbf{p}_0 - \mathbf{t}_1) \\ &= \left(\tilde{T}^\top \tilde{T}\right)^{-1} \tilde{T}^\top \tilde{T}\mathbf{p}_0^* = \mathbf{p}_0^*,\end{aligned}$$

which gives the consistency property of the estimator $\hat{\mathbf{p}}^*$. Therefore under our assumption that $\mathbf{p}_0 \succ 0$, the constraint that $\hat{\mathbf{p}} \succ \mathbf{0}$ is satisfied when the sample size is large enough.

Plug $\hat{\mathbf{p}}$ into $S(Q_0)$, and we have that

$$\begin{aligned}S(Q_0) &= \left| \boldsymbol{\gamma} - \mathbf{t}_1 - \tilde{T} \left(\tilde{T}^\top \tilde{T}\right)^{-1} \tilde{T}^\top (\boldsymbol{\gamma} - \mathbf{t}_1) \right|^2 \\ &= (\boldsymbol{\gamma} - \mathbf{t}_1)^\top \cdot \left(\mathcal{I}_n - \tilde{T} \left(\tilde{T}^\top \tilde{T}\right)^{-1} \tilde{T}^\top \right) \cdot (\boldsymbol{\gamma} - \mathbf{t}_1) \\ &= \left(\boldsymbol{\gamma} - \mathbf{t}_1 - \tilde{T}\mathbf{p}_0^* \right)^\top \cdot \left(\mathcal{I}_n - \tilde{T} \left(\tilde{T}^\top \tilde{T}\right)^{-1} \tilde{T}^\top \right) \cdot \left(\boldsymbol{\gamma} - \mathbf{t}_1 - \tilde{T}\mathbf{p}_0^* \right) \\ &= \left(\boldsymbol{\gamma} - T_{\boldsymbol{\theta}_0}(Q_0)\mathbf{p}_0 \right)^\top \cdot \left(\mathcal{I}_n - \tilde{T} \left(\tilde{T}^\top \tilde{T}\right)^{-1} \tilde{T}^\top \right) \cdot \left(\boldsymbol{\gamma} - T_{\boldsymbol{\theta}_0}(Q_0)\mathbf{p}_0 \right) \\ &= \left(\boldsymbol{\gamma} - T_{\boldsymbol{\theta}_0}(Q_0)\mathbf{p}_0 \right)^\top \Sigma^{-1/2} \cdot \Gamma \cdot \Sigma^{-1/2} \left(\boldsymbol{\gamma} - T_{\boldsymbol{\theta}_0}(Q_0)\mathbf{p}_0 \right),\end{aligned}\tag{3.22}$$

where

$$\Gamma = \Sigma^{1/2} \left(\mathcal{I}_n - \tilde{T} \left(\tilde{T}^\top \tilde{T}\right)^{-1} \tilde{T}^\top \right) \Sigma^{1/2}.\tag{3.23}$$

By (3.19),

$$\Sigma^{-1/2} \left(\boldsymbol{\gamma} - T_{\boldsymbol{\theta}_0}(Q_0)\mathbf{p}_0 \right) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}_n),$$

which implies the following result:

Proposition 7. *If matrix $\tilde{T}^\top \tilde{T}$ is invertible and the item parameters are known, then under the null hypothesis $H_0 : Q = Q_0$,*

$$N S(Q_0) \xrightarrow{d} \sum_{l=1}^n \lambda_l Z_l^2,$$

where λ_l 's are the eigenvalues of the Γ matrix in (3.23) and Z 's are independent standard Gaussian random variables.

Remark 7. Under the DINA model, matrix $\tilde{T}^\top \tilde{T}$ is invertible if Q_0 is complete and T is saturated. This follows from the result in Chapter 2 (see also Liu, Xu and Ying 2012b) that if Q_0 is complete then \tilde{T} has full column rank.

To simulate the asymptotic distribution in Proposition 7, the difficulty lies in the construction of the Γ matrix. This can be done following the definition formula (3.23), where Σ and T can be computed by (3.20) and (3.9) respectively .

3.3.2 Case of unknown item parameters

When the true parameters $\boldsymbol{\theta}_0$ are unknown, we use the plug-in method and replace $\boldsymbol{\theta}_0$ with the corresponding estimators. In this subsection, we focus on the MLE $\hat{\boldsymbol{\theta}}$. Following (3.18), the test statistic $S(Q_0)$ takes the form of

$$\hat{S}(Q_0) = |T_{\hat{\boldsymbol{\theta}}}(Q_0)\hat{\mathbf{p}} - \gamma|^2.$$

Then, if $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{p}})$ are consistent and $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0, \hat{\mathbf{p}} - \mathbf{p}_0)$ is normally distributed, we have the following approximation

$$\begin{aligned} \hat{S}(Q_0) &= |\gamma - T_{\hat{\boldsymbol{\theta}}}(Q_0)\hat{\mathbf{p}}|^2 \\ &= \left| \gamma - T_{\boldsymbol{\theta}_0}(Q_0)\mathbf{p}_0 + (T_{\boldsymbol{\theta}_0}(Q_0) - T_{\hat{\boldsymbol{\theta}}}(Q_0))\mathbf{p}_0 - T_{\boldsymbol{\theta}_0}(Q_0)(\hat{\mathbf{p}} - \mathbf{p}_0) \right. \\ &\quad \left. - (T_{\hat{\boldsymbol{\theta}}}(Q_0) - T_{\boldsymbol{\theta}_0}(Q_0))(\hat{\mathbf{p}} - \mathbf{p}_0) \right|^2 \\ &= (1 + o(1)) |\gamma - T_{\boldsymbol{\theta}_0}(Q_0)\mathbf{p}_0 + (T_{\boldsymbol{\theta}_0}(Q_0) - T_{\hat{\boldsymbol{\theta}}}(Q_0))\mathbf{p}_0 - T_{\boldsymbol{\theta}_0}(Q_0)(\hat{\mathbf{p}} - \mathbf{p}_0)|^2. \end{aligned} \tag{3.24}$$

From the above equation, we can see that the distribution of $\hat{S}(Q_0)$ depends on the joint distribution of $(\gamma, T_{\hat{\boldsymbol{\theta}}}(Q_0), \hat{\mathbf{p}})$, which relies on the joint distribution of $(\gamma, \hat{\boldsymbol{\theta}}, \hat{\mathbf{p}})$.

In the following, we derive the joint distribution of $(\boldsymbol{\gamma}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{p}})$ for a general diagnostic model. Our first result (Lemma 2) connects the MLE $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{p}})$ with the saturated response vector $\boldsymbol{\gamma}^{all}$, which is defined in Section 3.2.2. Before stating the lemma, we introduce some notation.

Let $\boldsymbol{\gamma}^*$ be a $(2^J - 1) \times 1$ vector defined as

$$\boldsymbol{\gamma}^* = \frac{1}{N} \left(\sum_{i=1}^N I(\mathbf{R}_i = \mathbf{R}), \mathbf{R} \in \{0, 1\}^J \setminus \mathbf{0} \right)^\top.$$

There is a one-to-one mapping between vectors $\boldsymbol{\gamma}^{all}$ and $\boldsymbol{\gamma}^*$. Then without loss of generality, let

$$\boldsymbol{\gamma}^* = L^* \boldsymbol{\gamma}^{all}. \quad (3.25)$$

Note that L^* is a $(2^J - 1) \times (2^J - 1)$ invertible matrix. Further, let $\boldsymbol{\gamma}$ be a vector defined by

$$\boldsymbol{\gamma} = \left(\frac{1}{N} \sum_{i=1}^N I(\mathbf{R}_i = \mathbf{0}), (\boldsymbol{\gamma}^*)^\top \right)^\top.$$

Let the true attribute profile probabilities be

$$\mathbf{p}_0 = (p_{0,\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \{0, 1\}^K)^\top$$

with the first element defined as $p_{0,\mathbf{0}}$ and the other part as \mathbf{p}_0^* (see the definition of \mathbf{p} in Section 3.2.1).

Under Q_0 , we have the likelihood function taking the form of

$$L_N(\boldsymbol{\theta}, \mathbf{p}^*) = \prod_{i=1}^N \left\{ \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} \left(p_{\boldsymbol{\alpha}} \prod_{j=1}^J c_{j,\boldsymbol{\alpha}}^{R_i^j} (1 - c_{j,\boldsymbol{\alpha}})^{1-R_i^j} \right) \right\}. \quad (3.26)$$

Here recall that $c_{j,\boldsymbol{\alpha}} = P(R_i^j = 1 \mid Q_0, \boldsymbol{\alpha}, \boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$ and Q_0 as described in Section 3.2.1. In addition, we write L_N as a function of \mathbf{p}^* instead of \mathbf{p} due to the constraint that $\sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} = 1$.

The following result shows that the MLE derived from (3.26) can be expressed as a linear function of the saturated response vector $\boldsymbol{\gamma}^{all}$:

Lemma 2. *Under the null hypothesis that $Q = Q_0$, suppose MLE $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{p}}^*)$ are consistent. Then we have that as $N \rightarrow \infty$,*

$$\sqrt{N} \begin{pmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \hat{\mathbf{p}}^* - \mathbf{p}_0^* \end{pmatrix} = (1 + o(1)) \cdot \mathbf{I}_0^{-1} \boldsymbol{\eta}^\top L \cdot \sqrt{N} (\boldsymbol{\gamma}^{all} - T_{\boldsymbol{\theta}_0}^{all}(Q_0) \mathbf{p}_0)$$

and further

$$\sqrt{N} \begin{pmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \hat{\mathbf{p}}^* - \mathbf{p}_0^* \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \mathbf{I}_0^{-1}),$$

where \mathbf{I}_0 is the Fisher information of the likelihood function (3.26) evaluated at $(\boldsymbol{\theta}_0, \mathbf{p}_0^*)$, L is a $2^J \times (2^J - 1)$ matrix defined by

$$L = \begin{pmatrix} -\mathbf{1}L^* \\ L^* \end{pmatrix} \quad (3.27)$$

with L^* as in (3.25), and $\boldsymbol{\eta}$ is a $2^J \times \dim(\boldsymbol{\theta})$ matrix.

With the help of Lemma 2, we can replace the last two terms in equation (3.24) with a linear transformation of $\boldsymbol{\gamma}^{all}$. We consider them one by one. Note that generally, for parameters $(\theta_1, \dots, \theta_h)$ and their estimators $(\hat{\theta}_1, \dots, \hat{\theta}_h)$ such that $\sqrt{N}((\hat{\theta}_1, \dots, \hat{\theta}_h) - (\theta_1, \dots, \theta_h))$ follows a multivariate normal distribution, we have that

$$\hat{\theta}_1 \cdots \hat{\theta}_h - \theta_1 \cdots \theta_h = (1 + o(1)) \sum_{l=1}^h \theta_1 \cdots \theta_{l-1} \theta_{l+1} \cdots \theta_h (\hat{\theta}_l - \theta_l).$$

This implies that there exists an $n \times 2J$ matrix $W_{\boldsymbol{\theta}_0, \mathbf{p}_0}$ such that

$$(T_{\hat{\boldsymbol{\theta}}}(Q_0) - T_{\boldsymbol{\theta}_0}(Q_0)) \mathbf{p}_0 = (1 + o(1)) W_{\boldsymbol{\theta}_0, \mathbf{p}_0} \begin{pmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix}. \quad (3.28)$$

In addition, we have that

$$T_{\theta_0}(Q_0)(\hat{\mathbf{p}} - \mathbf{p}_0) = T_{\theta_0}(Q_0) \begin{pmatrix} -\mathbf{1} \\ \mathcal{I}_{2^{K-1}} \end{pmatrix} (\hat{\mathbf{p}}^* - \hat{\mathbf{p}}_0^*).$$

The above results imply that for the main quantity in (3.24)

$$\begin{aligned} & \sqrt{N} \{ \boldsymbol{\gamma} - T_{\theta_0}(Q_0)\mathbf{p}_0 + (T_{\theta_0}(Q_0) - T_{\hat{\theta}}(Q_0))\mathbf{p}_0 - T_{\theta_0}(Q_0)(\hat{\mathbf{p}} - \mathbf{p}_0) \} \\ = & \sqrt{N} \left\{ \boldsymbol{\gamma} - T_{\theta_0}(Q_0)\mathbf{p}_0 - \left(W_{\theta_0, \mathbf{p}_0}, T_{\theta_0}(Q_0) \begin{pmatrix} -\mathbf{1} \\ \mathcal{I}_{2^{K-1}} \end{pmatrix} \right) \begin{pmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \hat{\mathbf{p}}^* - \hat{\mathbf{p}}_0^* \end{pmatrix} \right\} \\ = & \sqrt{N} \{ A (\boldsymbol{\gamma}^{all} - T_{\theta_0}^{all}(Q_0)\mathbf{p}_0) \} \\ \xrightarrow{d} & \mathcal{N}(0, \Sigma_u), \end{aligned}$$

where

$$\Sigma_u = A \Sigma_{all} A^\top \quad (3.29)$$

with Σ_{all} the covariance matrix of $\boldsymbol{\gamma}^{all}$ as defined in (3.20) and

$$A = \begin{pmatrix} \mathcal{I}_n & \mathbf{0} \end{pmatrix}_{n \times (2^J - 1)} - \left(W_{\theta_0, \mathbf{p}_0}, T_{\theta_0}(Q_0) \begin{pmatrix} -\mathbf{1} \\ \mathcal{I}_{2^{K-1}} \end{pmatrix} \right) \mathbf{I}_0^{-1} \boldsymbol{\eta}^\top L. \quad (3.30)$$

Then the distribution of equation (3.24) is given as follows:

Theorem 4. *Under the conditions of Lemma 2, we have that as $N \rightarrow \infty$,*

$$N\hat{S}(Q_0) \xrightarrow{d} \sum_{l=1}^{2^J - 1} \lambda_l^* Z_l^2,$$

where $\lambda_1^* \geq \dots \geq \lambda_{2^J - 1}^*$ are Σ_u 's eigenvalues, and

$$(Z_1, \dots, Z_{2^J - 1})^\top \sim^d \mathcal{N}(0, \mathcal{I}_{2^J - 1}).$$

Further, in the case of Algorithm 3, for each sub-matrix $Q_{0,m}$, Theorem 4 implies that there exist eigenvalues $\lambda_{m,1}, \dots, \lambda_{m,2^{J_m-1}}$ such that

$$N\hat{S}_m(Q_0) \xrightarrow{d} \sum_{l=1}^{2^{J_m-1}} \lambda_{m,l}^* Z_{m,l}^2,$$

where J_m is the number of rows in $Q_{0,m}$ and $(Z_{m,1}, \dots, Z_{m,2^{J_m-1}})^\top \sim^d \mathcal{N}(0, \mathcal{I}_{2^{J_m-1}})$. Therefore, due to the independence structure of the data, we have the following result.

Corollary 2. *Under the conditions of Lemma 2 and the setup of Algorithm 3,*

$$N \sum_{m=1}^M \hat{S}_m(Q_0) \xrightarrow{d} \sum_{m=1}^M \sum_{l=1}^{2^{J_m-1}} \lambda_{m,l}^* Z_{m,l}^2,$$

with $\{Z_{m,l}\}$ independent and following standard normal distribution.

Another issue concerns with the computation of the covariance matrix Σ_u in application, which depends on the unknown true parameters $(\boldsymbol{\theta}_0, \mathbf{p}_0)$. A natural approach is to replace $(\boldsymbol{\theta}_0, \mathbf{p}_0)$ in Σ_u with their MLE constructed under Q_0 . Simulation studies in the next section show that this approach works reasonably well.

The covariance matrix Σ_u follows from the definition equation (3.29). The covariance matrix Σ_{all} takes the form as in (3.20). The A matrix is given in (2.43), where W matrix can be computed from (3.28), and \mathbf{I}_0 can be calculated by taking the second derivatives of $\log L_N$, or equivalently, thanks to Lemma 2,

$$\mathbf{I}_0 = N \cdot Cov(\eta^\top L \cdot (\boldsymbol{\gamma}^{all} - T_{\boldsymbol{\theta}_0}^{all}(Q_0)\mathbf{p}_0)) = N \cdot \eta^\top L \Sigma L^\top \eta.$$

Then the only unspecified quantity is the matrix η .

The form of matrix η depends on the model setup of the diagnostic model. It can be derived from the corresponding likelihood function. To illustrate the idea, we use the DINA model as an example.

3.3.3 An example: DINA Model

Under the DINA model, the item parameters are

$$\boldsymbol{\theta}_{DINA} = (s_1, \dots, s_J, g_1, \dots, g_J)^\top,$$

where s_j and g_j are the slipping and guessing parameters.

Under the null hypothesis $H_0 : Q = Q_0$, let $(\boldsymbol{\theta}_{DINA,0}, \mathbf{p}_0)$ be the true model parameters. For a response vector \mathbf{R} , denote the corresponding probability mass function by

$$P_0(\mathbf{R}) = P(\mathbf{R}|Q_0, \mathbf{p}_0, \boldsymbol{\theta}_{DINA,0}).$$

Moreover, let $\mathbf{R}^{-j} = (R^1, \dots, R^{j-1}, R^j, \dots, R^J)^\top$ and write

$$P_0(\mathbf{R}^{-j}) = P(\mathbf{R}^{-j}|Q_0, \mathbf{p}_0, \boldsymbol{\theta}_{DINA,0}).$$

Lemma 3. *Under the DINA model and the null hypothesis $H_0 : Q = Q_0$, suppose MLE $(\hat{\boldsymbol{\theta}}_{DINA}, \hat{\mathbf{p}}^*)$ are consistent. Then as $N \rightarrow \infty$,*

$$\begin{aligned} & \sqrt{N} \begin{pmatrix} \hat{\boldsymbol{\theta}}_{DINA} - \boldsymbol{\theta}_{DINA,0} \\ \hat{\mathbf{p}}^* - \mathbf{p}_0^* \end{pmatrix} \\ &= (1 + o(1)) \mathbf{I}_{DINA,0}^{-1} \eta_{DINA}^\top L \cdot \sqrt{N} \left(\boldsymbol{\gamma}^{all} - T_{\boldsymbol{\theta}_{DINA,0}}^{all}(Q_0) \mathbf{p}_0 \right) \end{aligned}$$

and

$$\sqrt{N} \begin{pmatrix} \hat{\boldsymbol{\theta}}_{DINA} - \boldsymbol{\theta}_{DINA,0} \\ \hat{\mathbf{p}}^* - \mathbf{p}_0^* \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \mathbf{I}_{DINA,0}^{-1}),$$

where $\mathbf{I}_{DINA,0}$ is the Fisher information of the likelihood function (3.26) evaluated at $(\boldsymbol{\theta}_{DINA,0}, \mathbf{p}_0^*)$ under the DINA model, L is a $2^J \times (2^J - 1)$ matrix defined in (3.27), and η_{DINA} is a $2^J \times (2J + 2^K - 1)$ matrix defined as

$$\eta_{DINA} = \left(\eta_{s_1} \cdots \eta_{s_J}, \eta_{g_1}, \dots, \eta_{g_J}, \eta_{p_{\alpha_1}}, \dots, \eta_{p_{\alpha_{2^K-1}}} \right). \quad (3.31)$$

Here with \mathbf{R} arranged in the same order as in the response vector $\boldsymbol{\gamma}$ and $\xi_{DINA}^j(\boldsymbol{\alpha}, Q_0)$ as defined in (3.3), we have

$$\eta_{s_j} = \left((I(R^j = 0) - I(R^j = 1)) \cdot \frac{\sum_{\xi_{DINA}^j(\boldsymbol{\alpha}, Q_0)=1} p_{0,\boldsymbol{\alpha}} P_0(\mathbf{R}^{-j}|\boldsymbol{\alpha})}{P_0(\mathbf{R})}, \mathbf{R} \in \{0, 1\}^J \right)^\top, \quad (3.32)$$

$$\eta_{g_j} = \left((I(R^j = 1) - I(R^j = 0)) \cdot \frac{\sum_{\xi_{DINA}^j(\boldsymbol{\alpha}, Q_0)=0} p_{0,\boldsymbol{\alpha}} P_0(\mathbf{R}^{-j}|\boldsymbol{\alpha})}{P_0(\mathbf{R})}, \mathbf{R} \in \{0, 1\}^J \right)^\top, \quad (3.33)$$

and

$$\eta_{p_{\alpha_h}} = \left(\frac{P_0(\mathbf{R}|\boldsymbol{\alpha}_h) - P_0(\mathbf{R}|\boldsymbol{\alpha} = \mathbf{0})}{P_0(\mathbf{R})}, \mathbf{R} \in \{0, 1\}^J \right)^\top. \quad (3.34)$$

Remark 8. Consistency of MLE is assumed in Lemma 3 to make the corresponding local expansion of the log likelihood function in driving the final result. Note that not all the MLE are consistent under the model assumption. For instance, in the ideal case that $\boldsymbol{\theta}_{DINA,0} = \mathbf{0}$, completeness of the Q -matrix is sufficient and necessary to identify the \mathbf{p} vector; see Chapter 4 for more details about the identifiability of (s, g, \mathbf{p}) .

3.4 Simulation study

We conduct simulation studies to illustrate the theoretical results developed in the last section. We focus on the DINA model and study the performance of the Q -matrix validation procedure under different settings. We start with the ideal case where both the slipping and the guessing parameters are known and then move on to more general cases with unknown item parameters.

3.4.1 DINA model with known item parameters

In this subsection we study through simulation the performance of the testing procedure under the ideal assumption that both s and g are known.

$$Q_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad Q_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}. \quad (3.35)$$

We consider $J \times K$ Q -matrices, $J = 20$ items and $K = 3, 4, 5$ attributes, given by

(3.35). We further generate the attributes from a uniform distribution, i.e.,

$$p_{\alpha} = 2^{-K}, \quad \forall \alpha \in \{0, 1\}^K.$$

The slipping parameters and the guessing parameters are set to be $s_i = g_i = 0.2$ for $j = 1, \dots, 20$.

	Q_1		Q_2		Q_3	
	$D = 2$	$D = 3$	$D = 2$	$D = 3$	$D = 2$	$D = 3$
$N = 250$	45	62	57	68	70	75
$N = 500$	52	57	48	65	60	53
$N = 1000$	62	58	56	57	52	59
$N = 2000$	52	49	56	47	51	52

Table 3.1: Numbers of rejections out of 1000 simulations with $N = 250, 500, 1000$ and 2000 for Q_1, Q_2 , and Q_3 .

We consider various sample sizes $N = 250, 500, 1000$, and 2000 in the simulation. To reduce the computational complexity, the T -matrix contains combinations of up to D items, with $D = 2$ and 3. The testing significance level α is taken as 0.05. We simulate the data 1000 times independently under the above settings. The numbers of rejections of the true Q -matrices are given in Table 3.1. Table 3.1 shows that the proposed testing procedure performs quite well for all the considered situations.

3.4.2 DINA model with unknown model parameters

When J is small. We consider $J \times K$ Q -matrices, $J = 10$ items and $K = 3, 4, 5$

attributes, given by

$$Q_{11} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad Q_{21} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \quad Q_{31} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \quad (3.36)$$

where Q_{11} , Q_{21} and Q_{31} correspond to the first 10 items of Q_1 , Q_2 and Q_3 respectively. With a similar setting to the previous simulations, the slipping and guessing parameters are set to be 0.2 and the population is set to be uniform, i.e., $p_{\alpha} = 2^{-K}, \forall \alpha \in \{0, 1\}^K$.

For each sample size $N = 250, 500, 1000$, and 2000 , the T -matrix contains combinations of up to D items, with $D = 2, 3, 4$. Table 3.2 shows the numbers of rejections of the true Q -matrices out of 1000 independent simulations with respect to the significance level $\alpha = 0.05$.

When J is large. Consider matrices Q_1 , Q_2 and Q_3 , we split each of them into two parts, i.e.,

$$Q_1 = \begin{pmatrix} Q_{11} \\ Q_{12} \end{pmatrix}, \quad Q_2 = \begin{pmatrix} Q_{21} \\ Q_{22} \end{pmatrix}, \quad Q_3 = \begin{pmatrix} Q_{31} \\ Q_{32} \end{pmatrix}.$$

Here Q_{12} , Q_{22} and Q_{32} correspond to the last 10 items of Q_1 , Q_2 and Q_3 respectively.

	Q_{11}			Q_{21}			Q_{31}		
	$D = 2$	$D = 3$	$D = 4$	$D = 2$	$D = 3$	$D = 4$	$D = 2$	$D = 3$	$D = 4$
$N = 250$	46	44	55	45	46	55	63	55	62
$N = 500$	47	54	48	56	55	54	56	53	50
$N = 1000$	50	44	56	42	40	47	53	47	42
$N = 2000$	55	37	53	44	45	54	44	37	56

Table 3.2: Numbers of rejections out of 1000 simulations with $N = 250, 500, 1000$ and 2000 for $Q_{11}, Q_{21},$ and Q_{31} .

With a similar setting to the previous simulations, the slipping and guessing parameters are set to be 0.2 and the population is set to be uniform. We apply Algorithm 3 and obtain the results in Table 3.3. Under each simulation setting, the rejections rate (rejection number/1000) is close to the prespecified significance level $\alpha = 0.05$.

	Q_1			Q_2			Q_3		
	$D = 2$	$D = 3$	$D = 4$	$D = 2$	$D = 3$	$D = 4$	$D = 2$	$D = 3$	$D = 4$
$N = 250$	40	51	50	47	55	58	60	56	59
$N = 500$	45	56	46	54	46	56	56	58	48
$N = 1000$	56	48	46	48	51	47	38	52	43
$N = 2000$	51	43	43	46	46	50	48	44	55

Table 3.3: Numbers of rejections out of 1000 simulations with $N = 250, 500, 1000$ and 2000 for $Q_1, Q_2,$ and Q_3 with unknown parameters.

Testing power. We check the power of the testing procedure through simulation study. We construct misspecified Q -matrices Q_{10}, Q_{20}, Q_{30} based on the true ones Q_1, Q_2, Q_3 , as defined in (3.35), by misspecifying one item respectively.

For Q_{10} , we set the fourth item as $(1 \ 1 \ 0)$ and all the other 19 items having the same row vectors as those in Q_1 ; for Q_{20} , we set the fifth item as $(1 \ 0 \ 0 \ 0)$ and all the others having the same vectors as in Q_2 ; for Q_{30} , we set the sixth item as $(1 \ 0 \ 0 \ 0 \ 0)$

and all the others the same as in Q_3 . Specifically,

$$Q_{10} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \mathbf{1} & \mathbf{1} & \mathbf{0} \\ \vdots & \vdots & \vdots \end{pmatrix}, \quad Q_{20} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \quad Q_{30} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \quad (3.37)$$

Data sets are generated under the matrices Q_{10}, Q_{20}, Q_{30} and under the same model settings as in the previous simulations. We apply Algorithm 3 to the simulated data under the Q -matrices Q_1, Q_2 , and Q_3 . 1000 independent simulations are conducted, and numbers of rejections are given in Table 3.4 with respect to different sample sizes. We can see that the test power decreases as the attributes numbers K increases. For all situations considered in the simulation, Algorithm 3 works reasonably well, especially when the sample size is larger than 500.

	Q_{10}			Q_{20}			Q_{30}		
	$D = 2$	$D = 3$	$D = 4$	$D = 2$	$D = 3$	$D = 4$	$D = 2$	$D = 3$	$D = 4$
$N = 250$	683	475	301	305	411	267	183	191	194
$N = 500$	983	892	661	718	797	776	549	519	427
$N = 1000$	1000	1000	979	988	994	996	908	899	842
$N = 2000$	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table 3.4: Numbers of rejections out of 1000 simulations with $N = 250, 500, 1000$ and 2000 for Q_{10}, Q_{20} , and Q_{30} .

3.5 Appendix

In the Appendix, we provide the proof of Lemma 3. Similar derivation leads to Lemma 2 and we omit the details here.

Proof of Lemma 3. Under the DINA model and null hypothesis H_0 , the likelihood takes the form of

$$L_N(\boldsymbol{\theta}_{DINA}, \mathbf{p}^*) = \prod_{i=1}^N \left\{ \sum_{\boldsymbol{\alpha}} \left(p_{\boldsymbol{\alpha}} \prod_{j=1}^J P(R_i^j | Q_0, \boldsymbol{\alpha}, \boldsymbol{\theta}_{DINA}) \right) \right\}.$$

The log-likelihood function is

$$l_N(\boldsymbol{\theta}_{DINA}, \mathbf{p}^*) = \sum_{i=1}^N \log \left\{ \sum_{\boldsymbol{\alpha}} \left(p_{\boldsymbol{\alpha}} \prod_{j=1}^J P(R_i^j | Q_0, \boldsymbol{\alpha}, \boldsymbol{\theta}_{DINA}) \right) \right\}.$$

We start with the derivative of l_N with respect to s_1 .

$$\begin{aligned} & \left. \frac{\partial l_N(\boldsymbol{\theta}_{DINA}, \mathbf{p}^*)}{\partial s_1} \right|_{\boldsymbol{\theta}_{DINA,0}, \mathbf{p}_0^*} \\ &= \sum_{i=1}^N \left. \frac{\sum_{\boldsymbol{\alpha}} \left(p_{\boldsymbol{\alpha}} \cdot \frac{\partial P(R_i^1 | Q_0, \boldsymbol{\alpha}, \boldsymbol{\theta}_{DINA})}{\partial s_1} \cdot \prod_{j=2}^J P(R_i^j | Q_0, \boldsymbol{\alpha}, \boldsymbol{\theta}_{DINA}) \right)}{\sum_{\boldsymbol{\alpha}} \left(p_{\boldsymbol{\alpha}} \prod_{j=1}^J P(R_i^j | Q_0, \boldsymbol{\alpha}, \boldsymbol{\theta}_{DINA}) \right)} \right|_{\boldsymbol{\theta}_{DINA,0}, \mathbf{p}_0^*} \\ &= \sum_{i=1}^N \left\{ \frac{(I(R_i^1 = 0) - I(R_i^1 = 1)) \cdot \sum_{\xi_{DINA}^1(\boldsymbol{\alpha}, Q_0)=1} \left(p_{0,\boldsymbol{\alpha}} \cdot \prod_{j=2}^J P_0(R_i^j | \boldsymbol{\alpha}) \right)}{\sum_{\boldsymbol{\alpha}} \left(p_{0,\boldsymbol{\alpha}} \prod_{j=1}^J P_0(R_i^j | \boldsymbol{\alpha}) \right)} \right\} \\ &= \sum_{\mathbf{R} \in \{0,1\}^J} \left(\sum_{i=1}^N I(\mathbf{R}_i = \mathbf{R}) \right) \\ & \quad \times \left\{ \frac{(I(R^1 = 0) - I(R^1 = 1)) \cdot \sum_{\xi_{DINA}^1(\boldsymbol{\alpha}, Q_0)=1} \left(p_{0,\boldsymbol{\alpha}} P_0(\mathbf{R}^{-1} | \boldsymbol{\alpha}) \right)}{P_0(\mathbf{R})} \right\}, \end{aligned}$$

where $\xi_{DINA}^1(\boldsymbol{\alpha}, Q_0)$ is defined as in equation (3.3) and $\mathbf{R}^{-1} := (R^2, \dots, R^J)^\top$.

Since

$$\boldsymbol{\gamma}^* = L^* \boldsymbol{\gamma}^{all}, \quad (3.38)$$

we have that

$$\begin{aligned} \boldsymbol{\gamma} &= \begin{pmatrix} 1 & -\mathbf{1} \\ 0 & \mathcal{I}_{2^J-1} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{\gamma}^* \end{pmatrix} \\ &= \begin{pmatrix} 1 & -\mathbf{1} \\ 0 & \mathcal{I}_{2^J-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & L^* \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{\gamma}^{all} \end{pmatrix} \\ &= \begin{pmatrix} 1 & -\mathbf{1}L^* \\ 0 & L^* \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{\gamma}^{all} \end{pmatrix}. \end{aligned}$$

Then by the definition of η_{s_1} that

$$\eta_{s_1} = \left((I(R^1 = 0) - I(R^1 = 1)) \cdot \frac{\sum_{\xi_{DINA}^1(\boldsymbol{\alpha}, Q_0)=1} p_{0,\boldsymbol{\alpha}} P_0(\mathbf{R}^{2:J} | \boldsymbol{\alpha})}{P_0(\mathbf{R})}, \mathbf{R} \in \{0, 1\}^J \right)^\top, \quad (3.39)$$

we have

$$\frac{1}{\sqrt{N}} \frac{\partial l_N(\boldsymbol{\theta}_{DINA}, \mathbf{P}^*)}{\partial s_1} \Big|_{\boldsymbol{\theta}_{DINA,0}, \mathbf{P}_0^*} = \sqrt{N} \eta_{s_1}^\top \boldsymbol{\gamma} = \sqrt{N} \eta_{s_1}^\top \begin{pmatrix} 1 & -\mathbf{1}L^* \\ 0 & L^* \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{\gamma}^{all} \end{pmatrix}.$$

Recall that $T_{\boldsymbol{\theta}_{DINA,0}}^{all}(Q_0)$ denotes the saturated T -matrix with respect to parameters $(\boldsymbol{\theta}_{DINA,0}, Q_0)$ and $\boldsymbol{\gamma}^{all}$ is the $(2^J - 1) \times 1$ saturated response vector. Some calculation implies that

$$\sqrt{N} \eta_{s_1}^\top \begin{pmatrix} 1 & -\mathbf{1}L^* \\ 0 & L^* \end{pmatrix} \begin{pmatrix} 1 \\ T_{\boldsymbol{\theta}_{DINA,0}}^{all}(Q_0) \mathbf{P}_0 \end{pmatrix} = 0.$$

Therefore, we have that

$$\frac{1}{\sqrt{N}} \frac{\partial l_N(\boldsymbol{\theta}_{DINA}, \mathbf{P}^*)}{\partial s_1} \Big|_{\boldsymbol{\theta}_{DINA,0}, \mathbf{P}_0^*} = \sqrt{N} \eta_{s_1}^\top L \cdot \left(\boldsymbol{\gamma}^{all} - T_{\boldsymbol{\theta}_{DINA,0}}^{all}(Q_0) \mathbf{P}_0 \right).$$

Similarly, we have that for general derivatives of l_N with respect to $\{s_i, g_i, i = 1, \dots, J\}$, and $\{p_{\alpha_h}, \alpha_h \in \{0, 1\}^K \setminus \mathbf{0}\}$,

$$\begin{aligned} \frac{1}{\sqrt{N}} \frac{\partial l_N(\boldsymbol{\theta}_{DINA}, \mathbf{p}^*)}{\partial s_i} \Big|_{\boldsymbol{\theta}_{DINA,0}, \mathbf{p}_0^*} &= \sqrt{N} \eta_{s_i}^\top L \cdot \left(\boldsymbol{\gamma}^{all} - T_{\boldsymbol{\theta}_{DINA,0}}^{all}(Q_0) \mathbf{p}_0 \right), \\ \frac{1}{\sqrt{N}} \frac{\partial l_N(\boldsymbol{\theta}_{DINA}, \mathbf{p}^*)}{\partial g_i} \Big|_{\boldsymbol{\theta}_{DINA,0}, \mathbf{p}_0^*} &= \sqrt{N} \eta_{g_i}^\top L \cdot \left(\boldsymbol{\gamma}^{all} - T_{\boldsymbol{\theta}_{DINA,0}}^{all}(Q_0) \mathbf{p}_0 \right), \\ \frac{1}{\sqrt{N}} \frac{\partial l_N(\boldsymbol{\theta}_{DINA}, \mathbf{p}^*)}{\partial p_{\alpha_h}} \Big|_{\boldsymbol{\theta}_{DINA,0}, \mathbf{p}_0^*} &= \sqrt{N} \eta_{p_{\alpha_h}}^\top L \cdot \left(\boldsymbol{\gamma}^{all} - T_{\boldsymbol{\theta}_{DINA,0}}^{all}(Q_0) \mathbf{p}_0 \right), \end{aligned}$$

where η_{s_i} , η_{g_i} and $\eta_{p_{\alpha_h}}$ are defined as in the statement of Lemma 2.

Since $\mathbf{I}_{DINA,0}$ is the Hessian matrix of l_N with respect to $(\boldsymbol{\theta}_{DINA}, \mathbf{p}^*)$ evaluated at $(\boldsymbol{\theta}_{DINA,0}, \mathbf{p}_0^*)$. Then by Taylor's expansion, we have that

$$\begin{aligned} & \sqrt{N} \begin{pmatrix} \hat{\boldsymbol{\theta}}_{DINA} - \boldsymbol{\theta}_{DINA,0} \\ \hat{\mathbf{p}}^* - \mathbf{p}^* \end{pmatrix} \\ &= (1 + o(1)) \mathbf{I}_{DINA,0}^{-1} \frac{1}{\sqrt{N}} \frac{\partial l_N(\boldsymbol{\theta}_{DINA}, \mathbf{p})}{\partial (\boldsymbol{\theta}_{DINA}^\top, \mathbf{p}^\top)^\top} \Big|_{\boldsymbol{\theta}_{DINA,0}, \mathbf{p}_0^*} \\ &= (1 + o(1)) \mathbf{I}_{DINA,0}^{-1} \eta_{DINA}^\top L \cdot \sqrt{N} \left(\boldsymbol{\gamma}^{all} - T_{\boldsymbol{\theta}_{DINA,0}}^{all}(Q_0) \mathbf{p}_0 \right), \end{aligned}$$

where η_{DINA} is a $2^J \times (2J + 2^K - 1)$ matrix defined as

$$\eta_{DINA} = \left(\eta_{s_1} \cdots \eta_{s_J}, \eta_{g_1}, \dots, \eta_{g_J}, \eta_{p_{\alpha_1}}, \dots, \eta_{p_{\alpha_{2^K-1}}} \right).$$

Further, by the central limit theorem, we have the second result in Lemma 2. \square

Chapter 4

Identifiability of DCM model parameters

4.1 Introduction

The main purpose of cognitive diagnosis is to accurately evaluate subjects' strengths and weaknesses. Diagnostic classification models are important statistical tools in cognitive diagnosis and have gained increasing interest in recent years. These models provide specific attribute profiles for each subject, which allows for effective intervention for personal improvement. One simple and widely studied model among them is the DINA model (Junker and Sijtsma 2001). Other important models and developments can be found in Tatsuoka (1985); Hartz (2002); Leighton, Gierl and Hunka (2004); von Davier (2005); Templin and Henson (2006); DeCarlo (2011). A more thorough review of diagnostic models can be found in Rupp *et al.* (2010).

In order to specify attribute profiles, we need correct estimation of the diagnostic model parameters. Estimation of the diagnostic model parameters has been studied in

the literature and different estimation procedures have been proposed. For instance, de la Torre (2009) uses the EM algorithm and the MCMC method to estimate the slipping and guessing parameters in the DINA model. However, the fundamental question of the identifiability of the diagnostic model parameters (such as the slipping and guessing parameters in the DINA model) can be difficult to address. Under the DINA model, when both the slipping and the guessing parameters are zero, Chiu, Douglas and Li (2009) proposes the completeness assumption of the Q -matrix, which turns out to be sufficient and necessary for the model to be identifiable in this ideal case. See DeCarlo (2011) for more discussion. In the case that neither the slipping nor the guessing parameters are known, the corresponding theoretical justification of their consistency is still lacking.

The study of identifiability dates back to Koopmans (1950); Koopmans and Reiersl (1950). The key issue is to know whether the model parameters can be recovered based on the observed data. Identifiability is a prerequisite for statistical inferences, such as parameter estimation and hypothesis testing. In Chapter 2 we addressed the identifiability of the Q -matrix and proposed sufficient identifiability conditions. This chapter focuses on the identifiability of the diagnostic model parameters. In particular, we propose sufficient and necessary conditions under which the slipping and guessing parameters are estimable from the data under the DINA model assumption. The analysis method developed in this chapter is based on the theoretical framework of Chapter 2. This method is generic in the sense that it can be employed for the analysis of other diagnostic classification or latent class models, which is an interesting future research topic.

The remainder of this chapter is organized as follows. We introduce in Section 4.2 sufficient and necessary conditions for the DINA model parameters to be identifiable. The corresponding proofs are given in Section 4.3.

4.2 Main results

This chapter is concerned with the general statistical concept of identifiability, as applied to cognitive diagnosis, and the DINA model in particular. This is the issue of primary concern when examining the consistency of estimates in diagnostic classification models.

We say that a set of parameters θ for a family of distributions $\{f(x|\theta) : \theta \in \Theta\}$ is identifiable if distinct values of θ correspond to distinct probability density(mass) functions, i.e., for any θ there is no $\tilde{\theta} \in \Theta \setminus \{\theta\}$ such that $f(x|\theta) \equiv f(x|\tilde{\theta})$; see Definition 11.2.2 in Casella and Berger (2001). In addition, we say that a set of parameters θ is locally identifiable if there exists a neighborhood of θ , $\mathcal{N}_\theta \in \Theta$, such that there is no $\tilde{\theta} \in \mathcal{N}_\theta \setminus \{\theta\}$ such that $f(x|\theta) \equiv f(x|\tilde{\theta})$.

Both identifiability and local identifiability of latent class models are well-established concepts in latent class analysis (e.g. McHugh 1956; Goodman 1974). Identifiability is an important prerequisite for many types of statistical inference, such as parameter estimation and hypothesis testing. Local identifiability is a weaker form of identifiability, which ensures that the model parameters are identifiable in a neighborhood of the true parameter values. This is necessary for the model parameters to be estimable.

4.2.1 Notation

We first recall some notation introduced in Chapter 1. We consider a test of J items which requires K latent attributes. The Q -matrix $Q = (q_{jk})_{J \times K}$ is a $J \times K$ matrix with binary entries. For each j and k , $q_{jk} = 1$ indicates that item j requires attribute k and $q_{jk} = 0$ otherwise. Moreover, we use \mathbf{q}_j to denote the j th row of Q .

We say a Q -matrix is *complete* if for any attribute, there exists an item only requiring that attribute. In other words, Q is complete if there exist K rows of Q

such that they consist an identity matrix; see Definition 3 in Chapter 2 for more details. An simple example of a complete Q -matrix is the $K \times K$ identity matrix.

We use $\boldsymbol{\alpha} = (\alpha^1, \dots, \alpha^K)^\top$ to denote the vector of attributes, where $\alpha^k = 1$ or 0 , indicating the presence or absence of the k th attribute, $k = 1, \dots, K$, and superscript \top denotes transpose. In addition, let $\mathbf{R} = (R^1, \dots, R^J)^\top$ be the vector of responses to the J test items. Note that both $\boldsymbol{\alpha}$ and \mathbf{R} are subject-specific.

We assume that attribute profiles $\boldsymbol{\alpha}_i$, $i = 1, \dots, N$, are i.i.d. random variables, with the following distribution

$$P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}) = p_{\boldsymbol{\alpha}}, \quad (4.1)$$

where, for each $\boldsymbol{\alpha} \in \{0, 1\}^K$, $p_{\boldsymbol{\alpha}} \in [0, 1]$ and $\sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} = 1$. We use $\mathbf{p} = (p_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \{0, 1\}^K)$ to denote the distribution of the attribute profiles.

In this chapter we focus on the DINA model. We use $s = (s_1, \dots, s_J)^\top$ and $g = (g_1, \dots, g_J)^\top$ to denote slipping and guessing parameters. For notational convenience, let $c = \mathbf{1} - s$. Given a subject's profile $\boldsymbol{\alpha}$, the response to item j under the DINA model follows a Bernoulli distribution

$$P(R^j = 1 | Q, \boldsymbol{\alpha}, c_j, g_j) = c_j^{\xi^j(\boldsymbol{\alpha}, Q)} g_j^{1 - \xi^j(\boldsymbol{\alpha}, Q)}, \quad (4.2)$$

where

$$\xi^j(\boldsymbol{\alpha}, Q) = I(\alpha^k \geq q_{jk} \text{ for all } k = 1, \dots, K) \quad (4.3)$$

is as defined in (1.4). In addition, conditional on $\boldsymbol{\alpha}$, (R^1, \dots, R^J) are assumed to be jointly independent.

We write $c \succ g$ if $c_j > g_j$ for all $1 \leq j \leq J$. Through out this chapter, we assume $c \succ g$, $\mathbf{p} \succ 0$ and the Q -matrix is prespecified and correct.

4.2.2 Identifiability conditions

We list below conditions that will be used in the upcoming identifiability theorems. It will be shown under various model assumptions that certain specific combinations of these conditions are either necessary and/or sufficient for the identifiability of the unknown parameters.

C1 Q is complete. Without loss of generality, we assume that the Q -matrix takes the following form:

$$Q = \begin{pmatrix} \mathcal{I}_K \\ Q_1 \end{pmatrix}, \quad (4.4)$$

where matrix \mathcal{I}_K denotes the $K \times K$ identity matrix.

C2 Each attribute has been required by at least two items.

C3 Each attribute has been required by at least three items.

C4 There are at least $2^K + K - 1$ different rows in $T_{\tilde{c}, \tilde{g}}(Q)$, where $T_{\tilde{c}, \tilde{g}}(Q)$ is the T -matrix of Q as defined in Chapter 2.3 with $\tilde{c} = \underbrace{(0, \dots, 0)}_K, \underbrace{(1, \dots, 1)}_{J-K}$ and $\tilde{g} = \underbrace{(g_1, \dots, g_K)}_{J-K}, \underbrace{(0, \dots, 0)}_{J-K}$.

C5 For any $k \in \{1, \dots, K\}$, there exists an item or item combination in Q_1 requiring all attributes except the k th one.

Note that condition C5 is equivalent to $T(\mathbf{1} - \mathbf{e}_k) \in R(T(Q_1))$ for $k = 1, \dots, K$, where \mathbf{e}_k is a K dimensional row vector such that the j th element is one and the rest are zeroes, $T(\mathbf{1} - \mathbf{e}_k)$ is the T -matrix of item $\mathbf{1} - \mathbf{e}_k$ only, and $R(T(Q_1))$ is the set of row vectors of $T(Q_1)$.

We start with the ideal case, in which both the slipping and guessing parameters are known.

Theorem 5. *Population proportion parameters \mathbf{p} are identifiable only if condition C1 is satisfied. Moreover, C1 is also the sufficient condition if both slipping and guessing parameters are known.*

Theorem 5 shows that when s and g are known, the completeness of the Q -matrix is a sufficient and necessary condition for \mathbf{p} to be identifiable. Completeness ensures that we have enough information in the response data for each attribute profile to have its own distinct ideal response vector. When a Q -matrix is incomplete, we can easily construct a non-identifiable example. For instance, consider incomplete Q -matrix

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

Under the condition that $s = g = \mathbf{0}$, parameters $p_{(1,0)}$ and $p_{(0,0)}$ are nonidentifiable based on the response data since subjects with attribute profiles $(1, 0)$ and $(0, 0)$ have the same response vector $\mathbf{R} = (0, 0)$.

When the guessing parameters are known but the slipping parameters are unknown, we need stronger conditions for the DINA model to be identifiable. The corresponding results are given in the following theorem.

Theorem 6. *Under the DINA model with known guessing parameters g , the slipping parameters s and the population proportion parameters \mathbf{p} are identifiable if and only if conditions C1 and C2 hold. In addition, if C1 and C2 do not hold, then for any (s, \mathbf{p}) , there exist infinitely many $(\hat{s}, \hat{\mathbf{p}})$ having the same likelihood value.*

From the above theorem, we can see that when the guessing parameters g are known and Q -matrix takes the form of

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

the DINA model is non-identifiable. On the other hand,

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

satisfies the conditions of Theorem 6 and ensures the identifiability of the model parameters s and \mathbf{p} .

Note that $C1$ and $C2$ are also sufficient conditions for a Q -matrix to be identifiable; see Theorem 3 in Chapter 2. These two conditions ensure that we have enough information to estimate the unknown slipping and population proportion parameters, which helps separate the response vector space generated by the true Q -matrix with those generated by wrong Q -matrices; see Section 2.6 for more details.

In the most difficult setting, neither the slipping nor the guessing parameters are known. Then we have the following result.

Theorem 7. *Under the DINA model, s , g and \mathbf{p} are locally identifiable if and only if conditions $C1$, $C3$, and $C4$ hold.*

Theorem 7 follows from the next two propositions.

Proposition 8. (Necessary Conditions) *Under the DINA model, s , g and \mathbf{p} are locally identifiable only if conditions $C1$, $C3$, and $C4$ hold. Otherwise, there exist two parameter sets $(\hat{c}, \hat{g}, \hat{\mathbf{p}}) \neq (\bar{c}, \bar{g}, \bar{\mathbf{p}})$ such that they have the same likelihood value.*

Proposition 9. (Sufficient Conditions) *Suppose conditions C1 and C3 hold. Then $s = (s_1, \dots, s_J), g^* = (g_{K+1}, \dots, g_J)$ are identifiable.*

In addition, if condition C4 is true, then s and g are locally identifiable.

Moreover, if condition C5 is true, then s and g are identifiable.

Proposition 9 implies the following corollary. This is because the completeness of Q_1 implies condition C5.

Corollary 3. *If C1 and C3 hold and Q_1 is complete, s and g are identifiable.*

Based on Propositions 8 and 9, we have the following identifiability result, whose proof is similar to that of Theorem 7 and therefore is omitted. Let

$$\tilde{T}_{c,g}(Q) = \begin{pmatrix} T_{c,g}(Q) \\ \mathbb{E} \end{pmatrix},$$

where \mathbb{E} is the row vector with all elements being 1.

Theorem 8. *Under the DINA model, s , g and \mathbf{p} are identifiable if and only if C1, C3, and the following condition hold:*

For parameter sets $(g_1, \dots, g_K, \mathbf{p})$ and $(\hat{g}_1, \dots, \hat{g}_K, \hat{\mathbf{p}})$, equation $\tilde{T}_{\tilde{c},\tilde{g}}(Q)\mathbf{p} = \tilde{T}_{\tilde{c},\hat{\tilde{g}}}(Q)\hat{\mathbf{p}}$ holds if and only if $(g_1, \dots, g_K, \mathbf{p}) = (\hat{g}_1, \dots, \hat{g}_K, \hat{\mathbf{p}})$, where $\tilde{c} = \underbrace{(0, \dots, 0)}_K, \underbrace{(1, \dots, 1)}_{J-K}^\top$, $\tilde{g} = (g_1, \dots, g_K, 0, \dots, 0)^\top$, and $\hat{\tilde{g}} = (\hat{g}_1, \dots, \hat{g}_K, 0, \dots, 0)^\top$.

Based on the above results, we can see that when the Q -matrix takes the form of

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

the DINA model is non-identifiable, i.e., there exist two sets $(\hat{c}, \hat{g}, \hat{\mathbf{p}}) \neq (\bar{c}, \bar{g}, \bar{\mathbf{p}})$ such that they have the same likelihood value. This is because for the above Q -matrix the Q_1 part (the last three rows of Q) only have one kind of row vector, i.e., $(1, 1)$; this results in the unsatisfactory of condition C4. Therefore Proposition 8 gives the non-identifiability result. In particular, Proposition 9 indicates that the guessing parameters for the first two items (g_1, g_2) are non-identifiable. On the other hand, the following Q -matrix

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

satisfies conditions C1, C3, and C5 and therefore ensures the identifiability of the model parameters.

4.3 Proof of theorems

We begin with two important propositions necessary to prove the main results; their own proofs are postponed to the end of this section.

Recall that identifiability and local identifiability depend on the probability density function $f(x; \theta)$, which, when written as a function of the parameters θ becomes the likelihood $L(\theta)$. For N subjects' response data, the likelihood function under the DINA model is

$$L(c, g, \mathbf{p}) = \prod_{i=1}^N \left\{ \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} \left(p_{\boldsymbol{\alpha}} \prod_{j=1}^J P(R_i^j = 1 | c, g, \boldsymbol{\alpha}, Q)^{R_i^j} (1 - P(R_i^j = 1 | c, g, \boldsymbol{\alpha}, Q))^{1-R_i^j} \right) \right\}.$$

Note that the likelihood function L depends on the Q -matrix. Since in this chapter we assume that the Q -matrix is given and correctly specified, we omit the Q -matrix argument in L and write $L(c, g, \mathbf{p})$.

Proposition 10. *For two sets of parameters $(\hat{s}, \hat{g}, \hat{\mathbf{p}})$ and $(\bar{s}, \bar{g}, \bar{\mathbf{p}})$,*

$$L(\hat{s}, \hat{g}, \hat{\mathbf{p}}) = L(\bar{s}, \bar{g}, \bar{\mathbf{p}})$$

if and only if the following equation holds:

$$\tilde{T}_{\bar{c}, \bar{g}}(Q)\bar{\mathbf{p}} = \tilde{T}_{\hat{c}, \hat{g}}(Q)\hat{\mathbf{p}}. \quad (4.5)$$

Proposition 11. *There exists an invertible matrix D depending only on $g^* = (g_1^*, \dots, g_J^*)$, such that*

$$D\tilde{T}_{c, g}(Q) = \tilde{T}_{c-g^*, g-g^*}(Q).$$

We now prove our main theorems and propositions.

Proof of Theorem 5. Under the DINA model, if s and g are both equal to zero, the completeness of the Q -matrix is a sufficient and necessary condition for \mathbf{p} to be identifiable (Chiu *et al.* 2009; Liu *et al.* 2012b).

Consider the case where s and g are not zeroes and the Q -matrix is incomplete. Without loss of generality, we assume $\mathbf{e}_1 = (1, 0, \dots, 0)$ is not in the set of row vectors of Q . Then in the T -matrix $T_{c, g}(Q)$, the columns corresponding to attribute profiles

$\mathbf{0}$ and \mathbf{e}_1 are the same. In particular, these two columns take the following form:

$$\begin{pmatrix} g_1 & g_1 \\ \vdots & \vdots \\ g_J & g_J \\ g_1 g_2 & g_1 g_2 \\ \vdots & \vdots \\ g_1 g_2 \cdots g_J & g_1 g_2 \cdots g_J \end{pmatrix}. \quad (4.6)$$

Therefore, by Proposition 10, $p_{\mathbf{0}}$ and $p_{\mathbf{e}_1}$ are nonidentifiable.

When Q is complete, condition C1 assumes that first K rows of Q are $Q_{1:K} = \mathcal{I}_K$. The matrix $\tilde{T}_{c,g}(Q)$ has full column rank if $\tilde{T}_{c-g,0}(Q)$ has full column rank, since, by Proposition 11, $\tilde{T}_{c-g,0}(Q) = D\tilde{T}_{c,g}(Q)$ and D is invertible. Consider the rows of $\tilde{T}_{c-g,0}(Q)$ corresponding to combinations of the first K items and row vector \mathbb{E} . This constitutes an upper-triangular submatrix of size $2^K \times 2^K$ with nonzero diagonal entries. Thus, $\tilde{T}_{c,g}(Q)$ is of full column rank, and \mathbf{p} is identifiable. \square

Proof of Theorem 6. When g is known, Propositions 10 and 11 indicate that two sets of parameters $(\hat{c}, g, \hat{\mathbf{p}})$ and $(\bar{c}, g, \bar{\mathbf{p}})$ produce equal likelihoods if and only if

$$\begin{aligned} T_{\hat{c}-g,0}(Q)\hat{\mathbf{p}} &= DT_{\hat{c}-g,0}(Q)\hat{\mathbf{p}} \\ &= T_{\bar{c}-g,0}(Q)\bar{\mathbf{p}} = DT_{\bar{c}-g,0}(Q)\bar{\mathbf{p}} \end{aligned}$$

Then under the assumption that $c \succ g$, we only need to consider the situation in which $g = 0$.

(a) Sufficient Condition.

For each item $i \in 1, \dots, J$, condition $C2$ implies that there exist items i_1, \dots, i_l (different from i) such that

$$B_Q(i, i_1, \dots, i_l) = B_Q(i_1, \dots, i_l), \quad (4.7)$$

that is, the attributes required by item i are a subset of the attributes required by items i_1, \dots, i_l . Recall that B_Q 's are the row vectors of the T -matrix; see Chapter 2.2 for more details.

Let a and a_* be the row vectors in the T -matrix corresponding to item combinations $I_{i_1} \wedge \dots \wedge I_{i_l}$ and $I_i \wedge I_{i_1} \wedge \dots \wedge I_{i_l}$. Let $(\hat{c}, \hat{\mathbf{p}})$ and $(\bar{c}, \bar{\mathbf{p}})$ be two sets of parameters such that $L(\hat{c}, 0, \hat{\mathbf{p}}) = L(\bar{c}, 0, \bar{\mathbf{p}})$. Then by Proposition 10, they must satisfy the following equation:

$$\frac{a_*^\top T_{\hat{c},0}(Q) \hat{\mathbf{p}}}{a^\top T_{\hat{c},0}(Q) \hat{\mathbf{p}}} = \frac{a_*^\top T_{\bar{c},0}(Q) \bar{\mathbf{p}}}{a^\top T_{\bar{c},0}(Q) \bar{\mathbf{p}}}.$$

On the other hand, we have that

$$\begin{aligned} \frac{a_*^\top T_{\hat{c},0}(Q) \hat{\mathbf{p}}}{a^\top T_{\hat{c},0}(Q) \hat{\mathbf{p}}} &= \frac{B_{\hat{c},0;Q}(i, i_1, \dots, i_l) \hat{\mathbf{p}}}{B_{\hat{c},0;Q}(i_1, \dots, i_l) \hat{\mathbf{p}}} = \hat{c}_i, \\ \frac{a_*^\top T_{\bar{c},0}(Q) \bar{\mathbf{p}}}{a^\top T_{\bar{c},0}(Q) \bar{\mathbf{p}}} &= \frac{B_{\bar{c},0;Q}(i, i_1, \dots, i_l) \bar{\mathbf{p}}}{B_{\bar{c},0;Q}(i_1, \dots, i_l) \bar{\mathbf{p}}} = \bar{c}_i. \end{aligned} \quad (4.8)$$

Therefore, $\hat{c}_i = \bar{c}_i$ for all $i = 1, \dots, J$, which gives the identifiability of the slipping parameters.

In addition, the completeness of the Q -matrix ensures that we have enough equations to estimate \mathbf{p} uniquely, therefore s and \mathbf{p} are identifiable under conditions $C1$ and $C2$.

(b) Necessary Condition.

Thanks to Theorem 5, we only need to show that condition $C2$ is necessary for the identifiability of s and \mathbf{p} . Without loss of generality, suppose that the first attribute

only appears once in the first column of the Q -matrix, i.e., the Q -matrix takes the following form:

$$Q = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathcal{I}_{K-1} \\ \mathbf{0} & Q' \end{pmatrix}. \quad (4.9)$$

In the following we show that s and \mathbf{p} are non-identifiable. We only need to show that there are two different parameter sets $(\hat{c}, \hat{\mathbf{p}})$ and $(\bar{c}, \bar{\mathbf{p}})$ such that equation (4.5) holds.

Consider items $\{2, \dots, J\}$ and the corresponding Q -matrix

$$Q_{2:J} = \begin{pmatrix} \mathbf{0} & \mathcal{I}_{K-1} \\ \mathbf{0} & Q' \end{pmatrix}. \quad (4.10)$$

Based on the responses to items $\{2, \dots, J\}$, let $(\hat{c}_2, \dots, \hat{c}_J)$ and $\{\hat{p}_a, a \in \{0, 1\}^{K-1}\}$ be the corresponding parameter estimators. Note that in this case we only have $K-1$ attributes and $\{a : a \in \{0, 1\}^{K-1}\}$ include all possible attribute profiles for attributes $\{2, \dots, K\}$.

Now consider all J items. Following the above notation, let the population proportion estimators take the form of $\{\hat{p}_{(b,a)}, b \in \{0, 1\}, a \in \{0, 1\}^{K-1}\}$. Note that $\hat{p}_a = \hat{p}_{(1,a)} + \hat{p}_{(0,a)}$ for all $a \in \{0, 1\}^{K-1}$. Let $1 - \hat{c}_1$ be the estimator for the first item's slipping parameter. Then

$$(\bar{c}_1, \bar{c}_2, \dots, \bar{c}_J) := (x\hat{c}_1, \hat{c}_2, \dots, \hat{c}_J),$$

with x close to 1, satisfies equation (4.5) with the corresponding population proportion estimators $\bar{\mathbf{p}}$ defined as follows:

$$\bar{p}_{(1,a)} = \hat{p}_{(1,a)}/x \quad \text{and} \quad \bar{p}_{(0,a)} = \hat{p}_a - \hat{p}_{(1,a)}/x, \quad \text{for all } a \in \{0, 1\}^{K-1}.$$

The reason is given as follows. Consider the row in the T -matrix related to the first item. The corresponding row in equation (4.5) is

$$\hat{c}_1 \sum_{a \in \{0,1\}^{K-1}} \hat{p}_{(1,a)} + \hat{g}_1 \sum_{a \in \{0,1\}^{K-1}} \hat{p}_{(0,a)} = \bar{c}_1 \sum_{a \in \{0,1\}^{K-1}} \bar{p}_{(1,a)} + \bar{g}_1 \sum_{a \in \{0,1\}^{K-1}} \bar{p}_{(0,a)}, \quad (4.11)$$

which is satisfied under the above construction of \bar{c} and $\bar{\mathbf{p}}$. For other items and items combinations, the corresponding equations in equation (4.5) also hold. Therefore equation (4.5) is satisfied and we complete the proof. \square

Proof of Proposition 8. Thanks to Theorem 5, condition C1 is necessary for the model parameters to be identifiable. In the following we first show the necessity of condition C3. We consider the next two cases where C3 does not hold:

Case 1 There exists an attribute such that it only appears in one item.

Case 2 All attributes are required by at least two items and there exists an attribute such that it only appears in two items.

Case 1 We use a similar argument as in the proof of Theorem 6. Without loss of generality, suppose that the first attribute only appear once in the first column of the Q -matrix and the Q -matrix takes the following form:

$$Q = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathcal{I}_{K-1} \\ \mathbf{0} & Q' \end{pmatrix} \quad (4.12)$$

Consider items $\{2, \dots, J\}$, and let $(\hat{c}_2, \hat{g}_2, \dots, \hat{c}_J, \hat{g}_J)$ and $\{\hat{p}_a, a \in \{0, 1\}^{K-1}\}$ be the corresponding model parameter estimators. Here $\{a : a \in \{0, 1\}^{K-1}\}$ are all possible attribute profiles for attributes $\{2, \dots, K\}$.

Now consider all J items. Following the above notation, let $\hat{\mathbf{p}}$ take the form of $\{\hat{p}_{(b,a)}, b \in \{0, 1\}, a \in \{0, 1\}^{K-1}\}$. Note that $\hat{p}_a = \hat{p}_{(1,a)} + \hat{p}_{(0,a)}$ for all $a \in \{0, 1\}^{K-1}$. Let $1 - \hat{c}_1$ and \hat{g}_1 be the slipping and guessing estimators for the first item. Then

$$\begin{aligned} (\bar{c}_1, \bar{c}_2, \dots, \bar{c}_J) &:= (x\hat{c}_1, \hat{c}_2, \dots, \hat{c}_J), \\ (\bar{g}_1, \bar{g}_2, \dots, \bar{g}_J) &:= \left(\frac{\hat{P}_{(0,a)}}{\hat{p}_a - \hat{p}_{(0,a)}/x} \hat{g}_1, \hat{g}_2, \dots, \hat{g}_J \right), \end{aligned}$$

with x close to 1, satisfy equation (4.5) with the corresponding attribute profiles estimators $\bar{\mathbf{p}}$ defined as follows:

$$\bar{p}_{(1,a)} = \hat{p}_{(1,a)}/x \quad \text{and} \quad \bar{p}_{(1,a)} = \hat{p}_a - \hat{p}_{(0,a)}/x, \quad \text{for all } a \in \{0, 1\}^{K-1}.$$

Therefore, in Case 1, (s, g, \mathbf{p}) are not identifiable.

Case 2 We consider the case that all attributes are required by at least two items and there exists an attribute such that it only appears in two items. In the following two cases, we show that there exist two different parameter sets $(\hat{s}, \hat{g}, \hat{\mathbf{p}})$ and $(\bar{c}, \bar{g}, \bar{\mathbf{p}})$ satisfying equation (4.5).

Under the Q -matrices considered below in Case 2.1 and 2.2, we assume that $(\hat{s}_3, \dots, \hat{s}_J) = (\bar{s}_3, \dots, \bar{s}_J)$, $(\hat{g}_3, \dots, \hat{g}_J) = (\bar{g}_3, \dots, \bar{g}_J)$, and $(\hat{p}_a, a \in \{0, 1\}^{K-1}) = (\bar{p}_a, a \in \{0, 1\}^{K-1})$, where a corresponds to attribute profiles of the last $K - 1$ attributes.

Case 2.1 Without loss of generality, we write the Q -matrix as

$$Q = \begin{pmatrix} 1 & \mathbf{0}^\top \\ 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathcal{I}_{K-1} \\ \mathbf{0} & Q' \end{pmatrix}. \quad (4.13)$$

In this case, equation (4.5) holds if we have the following equations:

$$\left\{ \begin{array}{l} \hat{c}_1 \hat{p}_{(1,a)} + \hat{g}_1 \hat{p}_{(0,a)} = \bar{c}_1 \bar{p}_{(1,a)} + \bar{g}_1 \bar{p}_{(0,a)} \\ \hat{c}_2 \hat{p}_{(1,a)} + \hat{g}_2 \hat{p}_{(0,a)} = \bar{c}_2 \bar{p}_{(1,a)} + \bar{g}_2 \bar{p}_{(0,a)} \\ \hat{c}_1 \hat{c}_2 \hat{p}_{(1,a)} + \hat{g}_1 \hat{g}_2 \hat{p}_{(0,a)} = \bar{c}_1 \bar{c}_2 \bar{p}_{(1,a)} + \bar{g}_1 \bar{g}_2 \bar{p}_{(0,a)} \\ \hat{p}_{(0,a)} + \hat{p}_{(1,a)} = \bar{p}_{(0,a)} + \bar{p}_{(1,a)} \quad \forall a \in \{0, 1\}^{K-1} \end{array} \right. . \quad (4.14)$$

Choose $\hat{p}_{(1,a)}/\hat{p}_{(0,a)}$ to be certain constant for all $a \in \{0, 1\}^{K-1}$. Then for the pre-chosen $(\hat{c}, \hat{g}, \hat{\mathbf{p}})$, we have infinity many $(\bar{c}, \bar{g}, \bar{\mathbf{p}})$ satisfying (4.14) under restrictions that $c \succ g$ and $\mathbf{p} \succ 0$, which leads to the non-identifiability of (c, g, \mathbf{p}) .

Case 2.2 Without loss of generality, we write the Q -matrix as

$$Q = \begin{pmatrix} 1 & \mathbf{0}^\top \\ 1 & \mathbf{v}^\top \\ \mathbf{0} & \mathcal{I}_{K-1} \\ \mathbf{0} & Q' \end{pmatrix}, \quad (4.15)$$

where \mathbf{v} is a $K - 1$ dimensional nonzero vector. We borrow a result in Case 2 of the Proof of Proposition 9. It implies that for the sub-matrix related to the first three items

$$Q_{1:3} = \begin{pmatrix} 1 & \mathbf{0}^\top \\ 1 & \mathbf{v}^\top \\ \mathbf{0} & \mathbf{e}_1 \end{pmatrix}, \quad (4.16)$$

we have $\hat{c}_1 = \bar{c}_1$ and $\hat{g}_2 = \bar{g}_2$. Then equation (4.5) holds if we have the following equations:

$$\left\{ \begin{array}{l} \hat{c}_1 \hat{p}_{(1,a)} + \hat{g}_1 \hat{p}_{(0,a)} = \hat{c}_1 \bar{p}_{(1,a)} + \bar{g}_1 \bar{p}_{(0,a)} \\ (\hat{c}_2 - \hat{g}_2) \hat{p}_{(1,a)} = (\bar{c}_2 - \hat{g}_2) \bar{p}_{(1,a)}, \mathbf{v} \subseteq a \\ \hat{p}_{(0,a)} + \hat{p}_{(1,a)} = \bar{p}_{(0,a)} + \bar{p}_{(1,a)} \quad \forall a \in \{0, 1\}^{K-1} \end{array} \right. . \quad (4.17)$$

Then for certain pre-chosen $(\hat{c}, \hat{g}, \hat{\mathbf{p}})$, we can have infinity many estimators $(\bar{c}, \bar{g}, \bar{\mathbf{p}})$ satisfying equation (4.5) under the restrictions that $c \succ g$ and $\mathbf{p} \succ 0$.

Therefore, we obtain the necessity of condition C3.

To show the necessity of C4, without loss of generality we assume C1 and C3 hold. Then Proposition 9 gives that for (\hat{c}, \hat{g}) and (\bar{c}, \bar{g}) satisfying $L(\hat{c}, \hat{g}, \hat{\mathbf{p}}) = L(\bar{c}, \bar{g}, \bar{\mathbf{p}})$, the following holds:

$$\begin{cases} \hat{c}_j = \bar{c}_j & j = 1, \dots, J \\ \hat{g}_j = \bar{g}_j & j = (K+1), \dots, J \end{cases} . \quad (4.18)$$

Therefore equation (4.5) holds if the following is true:

$$\tilde{T}_{\bar{c}-g^*, \bar{g}-g^*}(Q)\bar{\mathbf{p}} = \tilde{T}_{\hat{c}-g^*, \hat{g}-g^*}(Q)\hat{\mathbf{p}}, \quad (4.19)$$

where

$$g^* = (\hat{c}_1, \dots, \hat{c}_K, \hat{g}_{K+1}, \dots, \hat{g}_J).$$

Then if C4 is not satisfied, there are less than $2^K + K$ equations but $2^K + K$ unknown parameters. Therefore for fixed $(\hat{c}, \hat{g}, \hat{\mathbf{p}})$, we have infinity many $(\bar{c}, \bar{g}, \bar{\mathbf{p}})$ satisfying equation (4.5) under the restrictions that $c \succ g$ and $\mathbf{p} \succ 0$. \square

Proof of Proposition 9. Under the completeness assumption, we only need to show the identifiability of s and g . We focus on the identifiability of s_1 and g_1 below. We consider the following three cases:

Case 1 There exist at least three items with Q -matrix row vector \mathbf{e}_1 . Without loss of generality, we write the Q -matrix as

$$Q = \begin{pmatrix} 1 & \mathbf{0}^\top \\ 1 & \mathbf{0}^\top \\ 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathcal{I}_{K-1} \\ \mathbf{0} & Q' \end{pmatrix}. \quad (4.20)$$

Suppose there are two sets of parameters (\hat{c}, \hat{g}) and (\bar{c}, \bar{g}) such that $L(\hat{c}, \hat{g}, \hat{\mathbf{p}}) = L(\bar{c}, \bar{g}, \bar{\mathbf{p}})$. In the following, we show that

$$\begin{cases} \hat{c}_j = \bar{c}_j & j = 1, 2, 3 \\ \hat{g}_j = \bar{g}_j & j = 1, 2, 3 \end{cases}. \quad (4.21)$$

Let $c - g = (c_1 - g_1, \dots, c_J - g_J)$. By Proposition 11 there exists a matrix D_g (only depending on g) such that

$$D_g^* \tilde{T}_{c,g}(Q) = \tilde{T}_{c-g^*, g-g^*}(Q).$$

Recall that

$$\tilde{T}_{c,g}(Q) = \begin{pmatrix} T_{c,g}(Q) \\ \mathbb{E} \end{pmatrix}.$$

Let $a_g(i_1, \dots, i_h)$ be the row vectors in D_g corresponding to $I_{i_1} \wedge \dots \wedge I_{i_h}$ (in $T_{c-g}(Q)$). For $T_{\hat{c}, \hat{g}}$ we have

$$\frac{a_{\hat{g}}(1, 3)^\top T_{\hat{c}, \hat{g}}(Q) \hat{\mathbf{p}}}{a_{\hat{g}}(1)^\top T_{\hat{c}, \hat{g}}(Q) \hat{\mathbf{p}}} = \frac{B_{\hat{c}-\hat{g}, 0; Q}(1, 3) \hat{\mathbf{p}}}{B_{\hat{c}-\hat{g}, 0; Q}(1) \hat{\mathbf{p}}} = \hat{c}_3 - \hat{g}_3, \quad (4.22)$$

$$\frac{a_{\hat{g}}(1, 2, 3)^\top T_{\hat{c}, \hat{g}}(Q) \hat{\mathbf{p}}}{a_{\hat{g}}(1, 2)^\top T_{\hat{c}, \hat{g}}(Q) \hat{\mathbf{p}}} = \frac{B_{\hat{c}-\hat{g}, 0; Q}(1, 2, 3) \hat{\mathbf{p}}}{B_{\hat{c}-\hat{g}, 0; Q}(1, 2) \hat{\mathbf{p}}} = \hat{c}_3 - \hat{g}_3. \quad (4.23)$$

Therefore, by Proposition 10, we have

$$\frac{B_{\bar{c}-\bar{g}, \bar{g}-\bar{g}; Q}(1 \wedge 3) \bar{\mathbf{p}}}{B_{\bar{c}-\bar{g}, \bar{g}-\bar{g}; Q}(1) \bar{\mathbf{p}}} = \frac{B_{\bar{c}-\bar{g}, \bar{g}-\bar{g}; Q}(1, 2, 3) \bar{\mathbf{p}}}{B_{\bar{c}-\bar{g}, \bar{g}-\bar{g}; Q}(1, 2) \bar{\mathbf{p}}}, \quad (4.24)$$

which implies

$$\begin{aligned}
& \left\{ (\bar{g}_1 - \hat{g}_1)(\bar{g}_3 - \hat{g}_3) \sum_{a \in \{0,1\}^{K-1}} \bar{p}_{(0,a)} + (\bar{c}_1 - \hat{g}_1)(\bar{c}_3 - \hat{g}_3) \sum_{a \in \{0,1\}^{K-1}} \bar{p}_{(1,a)} \right\} \\
& / \left\{ (\bar{g}_1 - \hat{g}_1) \sum_{a \in \{0,1\}^{K-1}} \bar{p}_{(0,a)} + (\bar{c}_1 - \hat{g}_1) \sum_{a \in \{0,1\}^{K-1}} \bar{p}_{(1,a)} \right\} \\
= & \left\{ (\bar{g}_1 - \hat{g}_1)(\bar{g}_2 - \hat{g}_2)(\bar{g}_3 - \hat{g}_3) \sum_{a \in \{0,1\}^{K-1}} \bar{p}_{(0,a)} \right. \\
& \left. + (\bar{c}_1 - \hat{g}_1)(\bar{c}_2 - \hat{g}_2)(\bar{c}_3 - \hat{g}_3) \sum_{a \in \{0,1\}^{K-1}} \bar{p}_{(1,a)} \right\} \\
& / \left\{ (\bar{g}_1 - \hat{g}_1)(\bar{g}_2 - \hat{g}_2) \sum_{a \in \{0,1\}^{K-1}} \bar{p}_{(0,a)} + (\bar{c}_1 - \hat{g}_1)(\bar{c}_2 - \hat{g}_2) \sum_{a \in \{0,1\}^{K-1}} \bar{p}_{(1,a)} \right\}.
\end{aligned} \tag{4.25}$$

From equation (4.25), we have

$$(\bar{g}_1 - \hat{g}_1)(\bar{c}_1 - \hat{g}_1)(\bar{c}_2 - \bar{g}_2)(\bar{c}_3 - \bar{g}_3) = 0.$$

Then under the constraint that $\bar{c} \succ \bar{g}$, we have

$$\bar{g}_1 = \hat{g}_1 \text{ or } \bar{c}_1 = \hat{g}_1.$$

Similarly consider different item combinations and we obtain that

$$\begin{cases} \bar{g}_2 = \hat{g}_2 \text{ or } \bar{c}_2 = \hat{g}_2 \\ \bar{g}_3 = \hat{g}_3 \text{ or } \bar{c}_3 = \hat{g}_3 \end{cases}. \tag{4.26}$$

Moreover,

$$\begin{cases} \hat{g}_1 = \bar{g}_1 \text{ or } \hat{c}_1 = \bar{g}_1 \\ \hat{g}_2 = \bar{g}_2 \text{ or } \hat{c}_2 = \bar{g}_2 \\ \hat{g}_3 = \bar{g}_3 \text{ or } \hat{c}_3 = \bar{g}_3 \end{cases}. \tag{4.27}$$

Therefore, for $j = 1, 2$, and 3 , if $\hat{g}_j \neq \bar{g}_j$ we have $\hat{c}_j = \bar{g}_j$ and $\bar{c}_j = \hat{g}_j$, which is impossible under the assumption that $c \succ g$. Thus we have $\hat{g}_j = \bar{g}_j$ for $j = 1, 2, 3$. Thanks to the proof of Theorem 6, we also have $\hat{c}_j = \bar{c}_j$ for $i = 1, 2, 3$.

Case 2 There exit two items with row vector \mathbf{e}_1 . Without loss of generality, we write the Q -matrix as

$$Q = \begin{pmatrix} 1 & \mathbf{0}^\top \\ 1 & \mathbf{0}^\top \\ 1 & \mathbf{v}^\top \\ \mathbf{0} & \mathcal{I}_{K-1} \\ \mathbf{0} & Q' \end{pmatrix}, \quad (4.28)$$

where \mathbf{v} is a non-zero vector. Without loss of generality we assume $\mathbf{v}^\top = (1, \mathbf{v}_*^\top)$. Consider the sub-matrix containing the first four items:

$$Q_{1:4} = \begin{pmatrix} 1 & 0 & \mathbf{0}^\top \\ 1 & 0 & \mathbf{0}^\top \\ 1 & 1 & \mathbf{v}_*^\top \\ 0 & 1 & \mathbf{0}^\top \end{pmatrix}, \quad (4.29)$$

Similarly as in the proof of Case 1, suppose that there are two sets of parameters (\hat{c}, \hat{g}) and (\bar{c}, \bar{g}) such that $L(\hat{s}, \hat{g}, \hat{\mathbf{p}}) = L(\bar{c}, \bar{g}, \bar{\mathbf{p}})$. We show that for the first four items, the following holds:

$$\begin{cases} \hat{c}_j = \bar{c}_j & j = 1, 2, 4 \\ \hat{g}_j = \bar{g}_j & j = 1, 2, 3 \end{cases} \quad (4.30)$$

Proposition 11 implies that there exists a matrix D_g (only depending on g) such that

$$D_g \tilde{T}_{c,g}(Q) = \tilde{T}_{c-g,0}(Q).$$

Let $a_g(i_1, \dots, i_h)$ be the row vectors in D_g corresponding to $I_{i_1} \wedge \dots \wedge I_{i_h}$ (in $T_{c-g}(Q)$). For $T_{\hat{c}, \hat{g}}$ we have

$$\begin{aligned} \frac{a_{\hat{g}}(1, 3)^\top T_{\hat{c}, \hat{g}}(Q) \hat{\mathbf{p}}}{a_{\hat{g}}(3)^\top T_{\hat{c}, \hat{g}}(Q) \hat{\mathbf{p}}} &= \frac{B_{\hat{c}-\hat{g}, 0; Q}(1, 3) \hat{\mathbf{p}}}{B_{\hat{c}-\hat{g}, 0; Q}(3) \hat{\mathbf{p}}} = \hat{c}_1 - \hat{g}_1, \\ \frac{a_{\hat{g}}(1, 4, 3)^\top T_{\hat{c}, \hat{g}}(Q) \hat{\mathbf{p}}}{a_{\hat{g}}(4, 3)^\top T_{\hat{c}, \hat{g}}(Q) \hat{\mathbf{p}}} &= \frac{B_{\hat{c}-\hat{g}, 0; Q}(1, 4, 3) \hat{\mathbf{p}}}{B_{\hat{c}-\hat{g}, 0; Q}(4, 3) \hat{\mathbf{p}}} = \hat{c}_1 - \hat{g}_1. \end{aligned}$$

Therefore, by Proposition 10, we have

$$\frac{B_{\bar{c}-\hat{g}, \bar{g}-\hat{g}; Q}(1, 3) \bar{\mathbf{p}}}{B_{\bar{c}-\hat{g}, \bar{g}-\hat{g}; Q}(3) \bar{\mathbf{p}}} = \frac{B_{\bar{c}-\hat{g}, \bar{g}-\hat{g}; Q}(1, 4, 3) \bar{\mathbf{p}}}{B_{\bar{c}-\hat{g}, \bar{g}-\hat{g}; Q}(4, 3) \bar{\mathbf{p}}},$$

which implies

$$\begin{aligned} &\frac{\tilde{g}_1 \tilde{g}_4 \tilde{g}_3 \bar{\mathbf{p}}_{0,0} + \tilde{c}_1 \tilde{g}_4 \tilde{g}_3 \bar{\mathbf{p}}_{1,0} + \tilde{g}_1 \tilde{c}_4 \tilde{g}_3 \bar{\mathbf{p}}_{0,1} + \tilde{c}_1 \tilde{c}_4 \tilde{c}_3 \bar{\mathbf{p}}_{1,1}}{\tilde{g}_4 \tilde{g}_3 \bar{\mathbf{p}}_{0,0} + \tilde{g}_4 \tilde{g}_3 \bar{\mathbf{p}}_{1,0} + \tilde{c}_4 \tilde{g}_3 \bar{\mathbf{p}}_{0,1} + \tilde{c}_4 \tilde{c}_3 \bar{\mathbf{p}}_{1,1}} \\ &= \frac{\tilde{g}_1 \tilde{g}_3 \bar{\mathbf{p}}_{0,0} + \tilde{c}_1 \tilde{g}_3 \bar{\mathbf{p}}_{1,0} + \tilde{g}_1 \tilde{g}_3 \bar{\mathbf{p}}_{0,1} + \tilde{c}_1 \tilde{c}_3 \bar{\mathbf{p}}_{1,1}}{\tilde{g}_3 \bar{\mathbf{p}}_{0,0} + \tilde{g}_3 \bar{\mathbf{p}}_{1,0} + \tilde{g}_3 \bar{\mathbf{p}}_{0,1} + \tilde{c}_3 \bar{\mathbf{p}}_{1,1}}, \end{aligned} \quad (4.31)$$

where $\tilde{g}_j = \bar{g}_j - \hat{g}_j$ for $j = 1, 3, 4$, $\tilde{c}_j = \bar{c}_j - \hat{c}_j$ for $j = 1, 4$,

$$\tilde{c}_3 = \frac{(\bar{c}_3 - \hat{c}_3) \sum_{v_* \subseteq a} \bar{p}_{(1,1,a)} + (\bar{g}_3 - \hat{g}_3) \sum_{v_* \not\subseteq a} \bar{p}_{(1,1,a)}}{\sum_{a \in \{0,1\}^{K-2}} \bar{p}_{(1,1,a)}},$$

and $\bar{\mathbf{p}}_{i,j} = \sum_{a \in \{0,1\}^{K-2}} \bar{p}_{(i,j,a)}$ for $i, j \in \{0, 1\}$.

From equation (4.31), we obtain

$$\bar{\mathbf{p}}_{0,0} \bar{\mathbf{p}}_{1,1} \tilde{g}_3 \tilde{c}_3 (\tilde{g}_1 - \tilde{c}_1) = \bar{\mathbf{p}}_{1,0} \bar{\mathbf{p}}_{0,1} \tilde{g}_3 \tilde{g}_3 (\tilde{g}_1 - \tilde{c}_1). \quad (4.32)$$

Since $\tilde{g}_1 - \tilde{c}_1 \neq 0$, we have

$$\tilde{g}_3 = 0 \text{ or } \bar{\mathbf{p}}_{0,0} \bar{\mathbf{p}}_{1,1} \tilde{c}_3 = \bar{\mathbf{p}}_{1,0} \bar{\mathbf{p}}_{0,1} \tilde{g}_3. \quad (4.33)$$

We show the second equation in (4.33) can not be true. Otherwise, we have

$$\bar{\mathbf{p}}_{0,0} \bar{\mathbf{p}}_{1,1} (\bar{c}_3 - \hat{g}_3) = \bar{\mathbf{p}}_{1,0} \bar{\mathbf{p}}_{0,1} (\bar{g}_3 - \hat{g}_3). \quad (4.34)$$

Similarly we have

$$\hat{\mathbf{P}}_{0,0}\hat{\mathbf{P}}_{1,1}(\hat{c}_3 - \bar{g}_3) = \hat{\mathbf{P}}_{1,0}\hat{\mathbf{P}}_{0,1}(\hat{g}_3 - \bar{g}_3). \quad (4.35)$$

Equations (4.34) and (4.35) imply that

$$\hat{c}_3 > \hat{g}_3 > \bar{c}_3 > \bar{g}_3 \text{ or } \bar{c}_3 > \bar{g}_3 > \hat{c}_3 > \hat{g}_3,$$

which conflicts the equation that

$$\hat{g}_3(\hat{\mathbf{P}}_{0,0} + \hat{\mathbf{P}}_{1,0} + \hat{\mathbf{P}}_{0,1}) + \hat{c}_3\mathbf{P}_{1,1} = \bar{g}_3(\bar{\mathbf{P}}_{0,0} + \bar{\mathbf{P}}_{1,0} + \bar{\mathbf{P}}_{0,1}) + \bar{c}_3\mathbf{P}_{1,1}.$$

Therefore, we have $\tilde{g}_3 = \bar{g}_3 - \hat{g}_3 = 0$. Let $\underline{g} = (0, 0, \hat{g}_3, 0, \dots, 0)$. Proposition 10 indicates the following equations:

$$\begin{aligned} \bar{c}_1 &= \frac{B_{\bar{c}-\underline{g},\bar{g}-\underline{g};Q}(1,4,3)\bar{\mathbf{P}}}{B_{\bar{c}-\underline{g},\bar{g}-\underline{g};Q}(4,3)\bar{\mathbf{P}}} = \frac{B_{\bar{c}-\underline{g},\hat{g}-\underline{g};Q}(1,4,3)\hat{\mathbf{P}}}{B_{\bar{c}-\underline{g},\hat{g}-\underline{g};Q}(4,3)\hat{\mathbf{P}}} = \hat{c}_1, \\ \bar{c}_2 &= \frac{B_{\bar{c}-\underline{g},\bar{g}-\underline{g};Q}(2,4,3)\bar{\mathbf{P}}}{B_{\bar{c}-\underline{g},\bar{g}-\underline{g};Q}(4,3)\bar{\mathbf{P}}} = \frac{B_{\bar{c}-\underline{g},\hat{g}-\underline{g};Q}(2,4,3)\hat{\mathbf{P}}}{B_{\bar{c}-\underline{g},\hat{g}-\underline{g};Q}(4,3)\hat{\mathbf{P}}} = \hat{c}_2, \\ \bar{c}_4 &= \frac{B_{\bar{c}-\underline{g},\bar{g}-\underline{g};Q}(1,4,3)\bar{\mathbf{P}}}{B_{\bar{c}-\underline{g},\bar{g}-\underline{g};Q}(1,3)\bar{\mathbf{P}}} = \frac{B_{\bar{c}-\underline{g},\hat{g}-\underline{g};Q}(1,4,3)\hat{\mathbf{P}}}{B_{\bar{c}-\underline{g},\hat{g}-\underline{g};Q}(1,3)\hat{\mathbf{P}}} = \hat{c}_4. \end{aligned}$$

Consider items 1 and 2. Let $\underline{c} = (\hat{c}_1, \hat{c}_2, 0, \dots, 0)$. Proposition 10 indicates that

$$\begin{aligned} \bar{g}_1 &= \frac{B_{\bar{c}-\underline{c},\bar{g}-\underline{c};Q}(1,2)\bar{\mathbf{P}}}{B_{\bar{c}-\underline{c},\bar{g}-\underline{c};Q}(2)\bar{\mathbf{P}}} = \frac{B_{\bar{c}-\underline{c},\hat{g}-\underline{c};Q}(1,2)\hat{\mathbf{P}}}{B_{\bar{c}-\underline{c},\hat{g}-\underline{c};Q}(2)\hat{\mathbf{P}}} = \hat{g}_1, \\ \bar{g}_2 &= \frac{B_{\bar{c}-\underline{c},\bar{g}-\underline{c};Q}(1,2)\bar{\mathbf{P}}}{B_{\bar{c}-\underline{c},\bar{g}-\underline{c};Q}(1)\bar{\mathbf{P}}} = \frac{B_{\bar{c}-\underline{c},\hat{g}-\underline{c};Q}(1,2)\hat{\mathbf{P}}}{B_{\bar{c}-\underline{c},\hat{g}-\underline{c};Q}(1)\hat{\mathbf{P}}} = \hat{g}_2. \end{aligned}$$

Therefore, we complete the proof in Case 2.

Case 3 There exists only one item with row vector \mathbf{e}_1 , i.e., the Q -matrix can be written as

$$Q = \begin{pmatrix} 1 & \mathbf{0}^\top \\ 1 & \mathbf{v}_2^\top \\ 1 & \mathbf{v}_3^\top \\ \mathbf{0} & \mathcal{I}_{K-1} \\ \mathbf{0} & Q' \end{pmatrix}. \quad (4.36)$$

Consider the sub-matrices:

$$Q_a = \begin{pmatrix} 1 & \mathbf{0}^\top \\ 1 & \mathbf{v}_2^\top \\ \mathbf{0} & \mathbf{e}_{h_2} \end{pmatrix} \text{ and } Q_b = \begin{pmatrix} 1 & \mathbf{0}^\top \\ 1 & \mathbf{v}_3^\top \\ \mathbf{0} & \mathbf{e}_{h_3} \end{pmatrix}, \quad (4.37)$$

where $\mathbf{e}_{h_2} \subseteq \mathbf{v}_2$ and $\mathbf{e}_{h_3} \subseteq \mathbf{v}_3$. By the proof in Case 2, we have

$$\begin{cases} \hat{c}_1 = \bar{c}_1 \\ \hat{g}_2 = \bar{g}_2 \\ \hat{g}_3 = \bar{g}_3 \end{cases}. \quad (4.38)$$

Now combining the results in Cases 1-3, we have that for the Q -matrix

$$Q = \begin{pmatrix} \mathcal{I}_K \\ Q_1 \end{pmatrix}, \quad (4.39)$$

the following holds:

$$\begin{cases} \hat{c}_j = \bar{c}_j & j = 1, \dots, K \\ \hat{g}_j = \bar{g}_j & j = (K+1), \dots, J \end{cases}. \quad (4.40)$$

By the proof of Theorem 6, we have that for the sub-matrix Q_1 ,

$$\hat{c}_j = \bar{c}_j, \quad j = (K+1), \dots, J. \quad (4.41)$$

Equations (4.40) and (4.41) implies that equation (4.5) is equivalent to

$$\tilde{T}_{\bar{c}-g^*, \bar{g}-g^*}(Q)\bar{\mathbf{p}} = \tilde{T}_{\hat{c}-g^*, \hat{g}-g^*}(Q)\hat{\mathbf{p}}, \quad (4.42)$$

where

$$g^* = (\hat{c}_1, \dots, \hat{c}_K, \hat{g}_{K+1}, \dots, \hat{g}_J).$$

Then if condition *C4* is true, by a similar argument in the proof of Proposition 8, we have the local identifiability.

In addition, if $T(\mathbf{1} - \mathbf{e}_k) \in R(T(Q_1))$ for $k = 1, \dots, K$, we have $T(\mathbf{1}) \in R(T(Q_1))$. Then we obtain $\hat{p}_{\mathbf{1}-\mathbf{e}_k} = \bar{p}_{\mathbf{1}-\mathbf{e}_k}$ from equation

$$\tilde{T}_{\bar{c}-g^*, \bar{g}-g^*}(Q_1)\bar{\mathbf{p}} = \tilde{T}_{\hat{c}-g^*, \hat{g}-g^*}(Q_1)\hat{\mathbf{p}}.$$

Given $\hat{p}_{\mathbf{1}-\mathbf{e}_k} = \bar{p}_{\mathbf{1}-\mathbf{e}_k}$, equation $\tilde{T}_{\bar{c}-g^*, \bar{g}-g^*}(Q)\bar{\mathbf{p}} = \tilde{T}_{\hat{c}-g^*, \hat{g}-g^*}(Q)\hat{\mathbf{p}}$ implies $\hat{g}_j = \bar{g}_j$, $j = 1, \dots, K$, and $\hat{\mathbf{p}} = \bar{\mathbf{p}}$. Therefore, (s, g, \mathbf{p}) are identifiable.

Therefore, we conclude the proof of Proposition 9. \square

Proof of Proposition 10. A subject's responses follow a multinomial distribution and the likelihood function has the form of

$$L(c, g, \mathbf{p}) = \prod_{h=1}^{2^J} \tilde{\pi}_h^{I(\mathbf{R}=\tilde{\mathbf{R}}_h)},$$

where \mathbf{R} is the observed response vector, $\{\tilde{\mathbf{R}}_h = (\tilde{R}_h^1, \dots, \tilde{R}_h^J) : h = 1, \dots, 2^J\}$ are all the 2^J possible response vectors for J items indexed by h , and $\tilde{\pi}_h$ is the probability in the multinomial distribution of observing a response vector \mathbf{R}_h , i.e.,

$$\tilde{\pi}_h = \sum_{\alpha \in \{0,1\}^K} p_\alpha \prod_{j=1}^J \pi_{j,h,\alpha}^{\tilde{R}_h^j} (1 - \pi_{j,h,\alpha})^{1-\tilde{R}_h^j}.$$

In the above equation, $\pi_{j,h,\alpha}$ is the probability of subjects with ability profile α having the response \tilde{R}_h^j , i.e.,

$$\begin{aligned}\pi_{j,h,\alpha} &= P(R^j = \tilde{R}_h^j | Q, \alpha, c, g) \\ &= I(\tilde{R}_h^j = 1) \cdot c_j^{\xi^j(\alpha, Q)} g_j^{1-\xi^j(\alpha, Q)} + I(\tilde{R}_h^j = 0) \cdot (1 - c_j)^{\xi^j(\alpha, Q)} (1 - g_j)^{1-\xi^j(\alpha, Q)}.\end{aligned}$$

Therefore for two sets of parameters $(\hat{c}, \hat{g}, \hat{\mathbf{p}})$ and $(\bar{c}, \bar{g}, \bar{\mathbf{p}})$,

$$L(\hat{s}, \hat{g}, \hat{\mathbf{p}}) = L(\bar{s}, \bar{g}, \bar{\mathbf{p}})$$

holds if and only if

$$\bar{\pi}_h = \hat{\pi}_h \text{ for } h = 1, \dots, 2^J, \quad (4.43)$$

where

$$\begin{aligned}\bar{\pi}_h &= \sum_{\alpha \in \{0,1\}^K} \bar{p}_\alpha \prod_{j=1}^J \bar{\pi}_{j,h,\alpha}^{\tilde{R}_h^j} (1 - \bar{\pi}_{j,h,\alpha})^{1-\tilde{R}_h^j}, \\ \hat{\pi}_h &= \sum_{\alpha \in \{0,1\}^K} \hat{p}_\alpha \prod_{j=1}^J \hat{\pi}_{j,h,\alpha}^{\tilde{R}_h^j} (1 - \hat{\pi}_{j,h,\alpha})^{1-\tilde{R}_h^j}, \\ \hat{\pi}_{j,h,\alpha} &= P(R^j = \tilde{R}_h^j | Q, \alpha, \hat{c}, \hat{g}), \text{ and } \bar{\pi}_{j,h,\alpha} = P(R^j = \tilde{R}_h^j | Q, \alpha, \bar{c}, \bar{g}).\end{aligned}$$

Next we show that equation (4.5) is equivalent to equation (4.43). For any row in $T_{c,g}$, we have

$$B_{c,g;Q}(j_1, \dots, j_l) \mathbf{p} = \sum_{\tilde{R}_h^{j_1}=1, \dots, \tilde{R}_h^{j_l}=1} \sum_{\alpha \in \{0,1\}^K} p_\alpha \prod_{j=1}^J \pi_{j,h,\alpha}^{\tilde{R}_h^j} (1 - \pi_{j,h,\alpha})^{1-\tilde{R}_h^j}.$$

This gives a one to one linear transformation between $\tilde{T}_{c,g} \mathbf{p}$ and $\{\tilde{\pi}_h : h = 1, \dots, 2^J\}$.

Therefore, we finish the proof. \square

Proof of the Proposition 11. In what follows, we construct a D matrix satisfying the condition in the proposition. We show that there exists a matrix D_{g^*} only depending on g^* so that $D_{g^*}\tilde{T}_{c,g}(Q) = \tilde{T}_{c-g^*,g-g^*}(Q)$. Note that each row of $D_{g^*}\tilde{T}_{c,g}(Q)$ is just a row linear transform of $\tilde{T}_{c,g}(Q)$. Then, it is sufficient to show that each row vector of $T_{c-g^*,g-g^*}(Q)$ is a linear transform of rows of $\tilde{T}_{c,g}(Q)$ with coefficients only depending on g^* . We prove this by induction.

First, note that

$$B_{c-g^*,g-g^*;Q}(j) = B_{c,g;Q}(j) - g_j^*\mathbb{E}.$$

Then all row vectors of $T_{c-g^*,g-g^*}(Q)$ of the form $B_{c-g^*,g-g^*;Q}(j)$ are inside the row space of $\tilde{T}_{c,g}(Q)$ with coefficients only depending on g^* . Suppose that all the vectors of the form

$$B_{c-g^*,g-g^*;Q}(j_1, \dots, j_l)$$

for all $1 \leq l \leq \iota$ can be written linear combinations of the row vectors of $\tilde{T}_{c,g}(Q)$ with coefficients only depending on g^* . Then, we consider

$$B_{c,g;Q}(j_1, \dots, j_{\iota+1}) = \Upsilon_{h=1}^{\iota+1} (B_{c-g^*,g-g^*;Q}(j_h) + g_{j_h}^*\mathbb{E}).$$

The left hand side is just a row vector of $\tilde{T}_{c,g}(Q)$. We expand the right hand side of the above display. Note that the last term is precisely

$$B_{c-g^*,g-g^*;Q}(j_1, \dots, j_{\iota+1}) = \Upsilon_{h=1}^{\iota+1} B_{c-g^*,g-g^*;Q}(j_h).$$

The rest terms are all of the form $B_{c-g^*,g-g^*;Q}(j_1, \dots, j_l)$ for $1 \leq l \leq \iota$ multiplied by coefficients only depending on g^* . Therefore, according to the induction assumption, we have that $B_{c-g^*,g-g^*;Q}(j_1, \dots, j_{\iota+1})$ can be written as linear combinations of rows of $\tilde{T}_{c,g}(Q)$ with coefficients only depending on g^* .

From the above construction, we conclude the proof. \square

Bibliography

- Allen, M. and Yen, W. M. (1979) *Introduction to Measurement Theory*. Long Grove, IL: Waveland Press.
- Baker, F. (2001) *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.
- Baker, F. B. and Kim, S.-H. (2004) *Item Response Theory: Parameter Estimation Techniques (2nd ed.)*. Marcel Dekker.
- Bock, R. (1972) Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, **37**, 29–51.
- Bock, R. and Aitken, M. (1981) Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika*, **46**, 443–459.
- Bock, R. and Lieberman, M. (1970) Fitting a response model for n dichotomously scored items. *Psychometrika*, **35**, 179–197.
- Brennan, R. L. (2001) *Generalizability Theory*. New York: Springer-Verlag.
- Casella, G. and Berger, R. L. (2001) *Statistical Inference*. Duxbury Press.
- Chiu, C., Douglas, J. and Li, X. (2009) Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, **74**, 633–665.
- Cox, D. and Hinkley, D. (1974) *Theoretical Statistics*. CRC press: Florida, U.S.
- Cronbach, L., Nageswari, R. and Gleser, G. (1963) Theory of generalizability: A liberation of reliability theory. *The British Journal of Statistical Psychology*, **16**, 137–163.

- de la Torre, J. (2008a) An empirically-based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, **45**, 343–362.
- de la Torre, J. (2008b) The generalized DINA model. *Paper presented at the annual International Meeting of the Psychometric Society, Durham, NH.*
- 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. and Douglas, J. (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.
- DeCarlo, L. T. (2012) Recognizing uncertainty in the Q-matrix via a Bayesian extension of the DINA model. *Applied Psychological Measurement*, **36**, 447–468.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society Series B-Methodological*, **39**, 1–38.
- DiBello, L., Stout, W. and Roussos, L. (1995) Unified cognitive psychometric assessment likelihood-based classification techniques. *Cognitively diagnostic assessment. Hillsdale, NJ: Erlbaum*, 361–390.
- Embretson, S. and Reise, S. P. (2000) *Item response theory for psychologists*. Psychology Press.
- Fox, J.-P. (2010) *Bayesian Item Response Modeling: Theory and Applications*. Springer.
- Gelman, A., Jakulin, A., Pittau, M. G. and Su, Y.-S. (2008) A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, **2**, 1360–1383.
- Goodman, L. A. (1974) Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, **61**, 215–231.
- Guttman, L. (1947) The cornell technique for scale and intensity analysis. *Educational and Psychological Measurement*, **7**, 247–280.

- Hartz, S. (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. and Douglas, J. (2005) Test construction for cognitive diagnosis. *Applied Psychological Measurement*, **29**, 262–277.
- Henson, R., Roussos, L., Douglas, J. and He, X. (2008) Cognitive diagnostic attribute level discrimination indices. *Applied Psychological Measurement*, **32**, 275–288.
- Henson, R. and Templin, J. (2005) Hierarchical log-linear modeling of the skill joint distribution. *External Diagnostic Research Group Technical Report*.
- Junker, B. (1999) Some statistical models and computational methods that may be useful for cognitively-relevant assessment. *Technical Report, available from <http://www.stat.cmu.edu/~brian/nrc/cfa/documents/final.pdf>*.
- Junker, B. and Sijtsma, K. (2001) Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, **25**, 258–272.
- Koopmans, T. C., ed. (1950) *Statistical Inference in Dynamic Economic Models*. Wiley, New York.
- Koopmans, T. C. and Reiersl, O. (1950) The identification of structural characteristics. *Ann. Math. Statist.*, **21**, 165–181.
- Leighton, J. P., Gierl, M. J. and Hunka, S. M. (2004) The attribute hierarchy model for cognitive assessment: A variation on tatsuoaka’s rule-space approach. *Journal of Educational Measurement*, **41**, 205–237.
- Liou, M. (1994) More on the computation of higher-order derivatives of the elementary symmetric functions in the Rasch model. *Applied Psychological Measurement*, **18**, 53–62.
- Liu, J., Xu, G. and Ying, Z. (2012a) Data-driven learning of Q -matrix. *Applied Psychological Measurement*, **36**, 548–564.
- Liu, J., Xu, G. and Ying, Z. (2012b) Theory of self-learning Q -matrix. *Bernoulli*, to appear.

- Liu, Y., Douglas, J. and Henson, R. (2007) Testing person fit in cognitive diagnosis. *Paper presented at the annual meeting of the National Council on Measurement in Education (NCME), Chicago, IL, April.*
- Lord, F. M. (1952) A theory of test scores. *Psychometric Monographs*, **7**.
- Lord, F. M. (1980) *Applications of item response theory to practical testing problems*. Mahwah, NJ: Erlbaum.
- Lord, F. M. and Novick, M. R. (1968) *Statistical Theories of Mental Test Scores*. Addison-Wesley: Reading, MA.
- Macready, G. and Dayton, C. (1977) The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, **2**, 99–120.
- Maris, E. (1999) Estimating multiple classification latent class models. *Psychometrika*, **64**, 187–212.
- McDonald, R. (1967) Non-linear Factor Analysis. *Psychometric Monographs*, **15**.
- McHugh, R. B. (1956) Efficient estimation and local identification in latent class analysis. *Psychometrika*, **21**, 331–347.
- Mislevy, R. (1996) Test theory reconceived. *Journal of Educational Measurement*, **33**, 379–416.
- Novick, M. R. (1966) The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, **3**, 1–18.
- Ostini, R. and Nering, M. L. (2006) *Polytomous Item Response Theory Models*. Sara Miller McCune, Sage Publication, Inc.
- Rasch, G. (1960) *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen, Danish Institute for Educational Research.
- Rasch, G. (1961) On general laws and the meaning of measurement in psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability IV*. Berkeley, California: University of California Press., 321–334.
- Reckase, M. D. (1990) Unidimensional data from multidimensional tests and multidimensional data from unidimensional tests. *Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).*

- Reckase, M. D. (2009) *Multidimensional Item Response Theory*. Springer.
- Roussos, L., DiBello, L., Stout, W., Hartz, S., Henson, R. and Templin, J. (2007a) The fusion model skills diagnosis system. in J. P. Leighton, & Gierl, M. J. (Ed.). *Cognitively diagnostic assessment for education: Theory and practice*, 275–318.
- Roussos, L. A., Templin, J. L. and Henson, R. A. (2007b) Skills diagnosis using IRT-based latent class models. *Journal of Educational Measurement*, **44**, 293–311.
- Rupp, A. (2002) Feature selection for choosing and assembling measurement models: A building-block-based organization. *Psychometrika*, **2**, 311–360.
- Rupp, A. and Templin, J. (2008a) Effects of Q-matrix misspecification on parameter estimates and misclassification rates in the DINA model. *Educational and Psychological Measurement*, **68**, 78–98.
- Rupp, A. and Templin, J. (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., Templin, J. and Henson, R. A. (2010) *Diagnostic Measurement: Theory, Methods, and Applications*. Guilford Press.
- Samejima, F. (1969) Estimation of Latent Ability Using a Response Pattern of Graded Scores. *Psychometric Monographs*, **17**.
- Samejima, F. (1974) Normal ogive model on the continuous response level in the multidimensional latent space. *Psychometrika*, **39**, 111–121.
- Shavelson, R.J., . W. N. (1991) *Generalizability Theory: A Primer*. Thousand Oaks, CA: Sage.
- Spearman, C. (1904) The proof and measurement of association between two things. *American Journal of Psychology*, **15**, 72–101.
- Stout, W. (2007) Skills diagnosis using IRT-based continuous latent trait models. *Journal of Educational Measurement*, **44**, 313–324.
- Tatsuoka, C. (2002) Data-analytic methods for latent partially ordered classification models. *Applied Statistics (JRSS-C)*, **51**, 337–350.
- Tatsuoka, K. (1983) Rule space: an approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, **20**, 345–354.

- Tatsuoka, K. (1985) A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics*, **12**, 55–73.
- Tatsuoka, K. (2009) *Cognitive assessment: an introduction to the rule space method*. CRC Press.
- Templin, J. (2006) CDM: cognitive diagnosis modeling with Mplus [computer software]. (available from <http://jtemplin.myweb.uga.edu/cdm/cdm.html>).
- Templin, J., He, X., Roussos, L. and 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. and Henson, R. (2006) Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, **11**, 287–305.
- Thissen, D. (1982) Marginal maximum-likelihood estimation for the one-parameter logistic model. *Psychometrika*, **47**, 175–186.
- Tucker, L. (1946) Maximum validity of a test with equivalent items. *Psychometrika*, **11**, 1–13.
- van der Linden, W. (1978) Forgetting, guessing, and mastery: The macready and dayton models revisited and compared with a latent trait approach. *Journal of Educational Statistics*, **3**, 305–317.
- van der Linden, W. J. and Hambleton, R. K., eds. (1996) *Handbook of Modern Item Response Theory*. Springer.
- von Davier, M. (2005) *A general diagnosis model applied to language testing data*. Educational Testing Service, Research Report.
- von Davier, M. and Yamamoto, K. (2004) A class of models for cognitive diagnosis. (slides available from www.von-davier.com). *The Spearman Conference*.
- Xu, X. and von Davier, M. (2008) Fitting the structured general diagnostic model to NAEP data (rp-08-27). *Princeton, NJ: Educational Testing Service*.