

CONSISTENCY THEORY FOR THE GENERAL NONPARAMETRIC CLASSIFICATION METHOD

CHIA-YI CHIU

RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY

HANS-FRIEDRICH KÖHN

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Parametric likelihood estimation is the prevailing method for fitting cognitive diagnosis models—also called diagnostic classification models (DCMs). Nonparametric concepts and methods that do not rely on a parametric statistical model have been proposed for cognitive diagnosis. These methods are particularly useful when sample sizes are small. The general nonparametric classification (GNPC) method for assigning examinees to proficiency classes can accommodate assessment data conforming to any diagnostic classification model that describes the probability of a correct item response as an increasing function of the number of required attributes mastered by an examinee (known as the “monotonicity assumption”). Hence, the GNPC method can be used with any model that can be represented as a general DCM. However, the statistical properties of the estimator of examinees’ proficiency class are currently unknown. In this article, the consistency theory of the GNPC proficiency-class estimator is developed and its statistical consistency is proven.

Key words: cognitive diagnosis, Q-matrix, DINA model, DINO model, general DCM, G-DINA model, nonparametric classification, general nonparametric classification method.

1. Introduction

Cognitive diagnosis (CD) in educational measurement (DiBello, Roussos, & Stout, 2007; Haberman & von Davier, 2007; Leighton & Gierl, 2007; Nichols, Chipman, & Brennan, 1995; Rupp, Templin, & Henson, 2010; Tatsuoka, 2009) explicitly targets mastery of the instructional content and seeks to provide immediate feedback to students on their strengths and weaknesses in terms of attributes mastered and attributes needing study. Within the CD framework, skills, specific knowledge, talent, ability—any aptitude required to perform cognitive tasks—are collectively referred to as “attributes” that an examinee may or may not possess. CD models—or “Diagnostic Classification Models” (DCMs), as they are called here—describe an examinee’s ability as a composite of these attributes. Mastery of attributes is recorded as a binary string. Different zero-one combinations define attribute profiles of distinct proficiency classes to which examinees are assigned in estimating their individual attribute profiles from their test performance. Parametric methods for fitting DCMs prevail. They either use marginal maximum likelihood estimation relying on the expectation–maximization algorithm (MMLE-EM) or Markov chain Monte Carlo (MCMC) techniques (de la Torre, 2009, 2011; DiBello et al., 2007; von Davier, 2008).

A number of researchers (Ayers, Nugent, & Dean, 2008; Chiu, 2008; Chiu & Douglas, 2013; Chiu, Douglas, & Li, 2009; Park & Lee, 2011; Willse, Henson, & Templin, 2007) have explored the potential of methods that do not rely on a parametric statistical model—nonparametric methods for short—as alternatives to MMLE-EM and MCMC for assigning examinees to proficiency

Correspondence should be made to Chia-Yi Chiu, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA. Email: chia-yi.chiu@gse.rutgers.edu

classes. The need for nonparametric methods might arise in situations where sample sizes are insufficient to provide reliable maximum likelihood estimates; for example, when assessment data have been collected in educational micro-environments, say, for monitoring the instruction and learning process at the classroom level. The NonParametric Classification (NPC) method by Chiu and Douglas (2013) and its generalization, and the General NPC (GNPC) method (Chiu, Sun, & Bian, 2018) are two recent examples of efficient and effective nonparametric methodological developments for assigning examinees to proficiency classes. The algorithms of the NPC and GNPC method can handle small sample sizes (Chiu et al. 2018). In addition, they are easy to implement and computationally inexpensive using minimal CPU times. These features support the use of the NPC and GNPC methods as computational engines for CD-based computerized adaptive tests (CAT) tailored to small and very small teaching units like individual classrooms.

Both methods, have certain drawbacks. The NPC estimator of examinees' proficiency class was proven by Wang and Douglas (2015) to be statistically consistent for any DCM under certain regularity conditions. However, these consistency conditions are often difficult to meet for more complex DCMs that model the probability of a correct item response as an increasing function of the number of required attributes mastered by an examinee (known as the “monotonicity assumption”). The NPC method does not provide the flexibility to account for such complex relations between required and mastered attributes. The GNPC method, on the other hand, provides this flexibility, but the statistical properties of the resulting proficiency-class estimator are currently unknown.

In this article, the statistical consistency of the GNPC estimator of examinees' proficiency class is proven. The next sections provide brief summaries of some key technical concepts of the NPC and the GNPC method that are prerequisites of the proof of consistency. The paper concludes with a brief discussion of possible applications of this important result.

2. Review of Technical Key Concepts

2.1. Cognitive Diagnosis and Diagnostic Classification Models

Assume ability in a given domain is conceptualized as a composite of K latent binary attributes $\alpha_1, \alpha_2, \dots, \alpha_K$. The K -dimensional binary vector $\alpha_m = (\alpha_{m1}, \alpha_{m2}, \dots, \alpha_{mK})^T$ denotes the attribute profile of proficiency class C_m , $m = 1, 2, \dots, M$, where the k th entry, $\alpha_{mk} \in \{0, 1\}$, indicates (non-)mastery of the corresponding attribute. (The transpose of vectors or matrices is denoted by a superscripted T ; the conventional “prime” notation is reserved here for distinguishing between vectors or their scalar entries.) If the attributes do not have a hierarchical structure, then there are $2^K = M$ distinct proficiency classes. The attribute profile $\alpha_{i \in C_m}$ of examinee $i \in C_m$ is usually written as $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iK})^T$. Throughout the text, the terms “profile” and “vector” are used interchangeably; for brevity, the examinee index i , $i = 1, 2, \dots, N$, is omitted if the context permits; for example, α_i is simply written as $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_K)^T$.

The individual items of a test are also characterized by K -dimensional attribute profiles \mathbf{q}_j that specify for each individual item j of a test, $j = 1, 2, \dots, J$, which attributes are required for a correct response ($q_{jk} = 1$, if a correct answer requires mastery of the k th attribute, and 0 otherwise). If a domain is characterized by K attributes, then there are at most $2^K - 1$ distinct item-attribute profiles because item-attribute profiles that consist entirely of zeroes are inadmissible. The J item-attribute profiles of a test form its Q-matrix, $\mathbf{Q} = \{q_{jk}\}_{(J \times K)}$ (Tatsuoka, 1985). The Q-matrix must be identified and complete. A Q-matrix is said to be complete if it guarantees the identifiability of all realizable proficiency classes among examinees (Chiu et al., 2009; Köhn & Chiu, 2016a, 2017). Q-completeness is formally defined as $S(\alpha) = S(\alpha') \Rightarrow \alpha = \alpha'$, where

$S(\boldsymbol{\alpha}) = E(\mathbf{Y} \mid \boldsymbol{\alpha})$ denotes the conditional expectation of the item response vector \mathbf{Y} , given attribute vector $\boldsymbol{\alpha}$. Verbally stated, a Q-matrix is complete if the equality of two expected item response vectors, $S(\boldsymbol{\alpha})$ and $S(\boldsymbol{\alpha}')$, implies that the underlying attribute profiles, $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$, are also identical. Completeness of the Q-matrix is a general requirement for any diagnostic classification regardless of whether MMLE-EM, MCMC, or nonparametric methods are used to assign examinees to proficiency classes. An incomplete Q-matrix causes examinees to be assigned to proficiency classes to which they do not belong.

DCMs are constrained latent class models such that the latent variable proficiency-class membership—associated with mastery of a particular attribute set—determines the probability of a correct item response. The specific functional relation between mastery of attributes and the probability of a correct item response distinguishes DCMs (de la Torre & Douglas, 2004; Henson et al., 2009; Maris, 1999). Criteria for describing these differences are, for example, compensatory versus non-compensatory models, and disjunctive versus conjunctive models. DCMs that allow for compensating the lack of certain attributes by the mastery of other attributes are called compensatory models, in contrast to non-compensatory models that do not provide this possibility. The second criterion distinguishes between disjunctive DCMs, where mastery of a subset of the required attributes is a sufficient condition for maximizing the probability of a correct item response, and conjunctive DCMs, where mastery of only a subset of the required attributes results in a success probability equal to that of an examinee mastering none of the attributes. The **D**eterministic **I**nput **N**oisy “**A**ND” **G**ate (**DINA**) **M**odel (Haertel, 1989; Junker & Sijtsma, 2001; Macready & Dayton, 1977) is the standard example of a conjunctive DCM; its item response function (IRF) is

$$P(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i) = (1 - s_j)^{\eta_{ij}} g_j^{(1-\eta_{ij})}$$

where monotonicity is typically imposed through the restriction $0 < g_j < 1 - s_j < 1 \forall j$. The response of examinee i to binary item j , $j = 1, 2, \dots, J$ is denoted as Y_{ij} . The conjunctive ideal response η_{ij} , defined as $\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}$, indicates whether examinee i has mastered all the attributes needed to answer item j correctly. The item-related parameters $s_j = P(Y_{ij} = 0 \mid \eta_{ij} = 1)$ and $g_j = P(Y_{ij} = 1 \mid \eta_{ij} = 0)$ formalize the probabilities of slipping (failing to answer item j correctly despite having the skills required to do so) and guessing (answering item j correctly despite lacking the skills required to do so), respectively. Thus, η_{ij} can be interpreted as the ideal item response when neither slipping nor guessing occurs. The **D**eterministic **I**nput **N**oisy “**O**R” gate (**DINO**) model (Templin & Henson, 2006) is the prototypic disjunctive DCM (i.e., mastery of a subset of the required attributes is a sufficient condition for maximizing the probability of a correct item response). The IRF of the DINO model is

$$P(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i) = (1 - s_j)^{\omega_{ij}} g_j^{(1-\omega_{ij})}$$

where the disjunctive ideal response $\omega_{ij} = 1 - \prod_{k=1}^K (1 - \alpha_{ik})^{q_{jk}}$ indicates whether at least one of the attributes required for item j has been mastered. (Like η_{ij} in the DINA model, ω_{ij} is the ideal item response.) The DINA model and the DINO model are rather limited in their flexibility to model the relation between response probabilities and attribute mastery. For example, the DINA model cannot distinguish between examinees who master none and those who master a subset of the attributes required for an item. Only if all required attributes are mastered can an examinee realize a high probability of answering the item correctly. In contrast, more complex models like the general DCMs (de la Torre, 2011; Henson et al., 2009; Rupp et al., 2010; von Davier, 2005, 2008) offer far greater flexibility in modeling the probability of correct item responses for different attribute profiles.

2.2. *The Nonparametric Classification Method*

The NPC method for estimating an examinee’s proficiency class developed by Chiu and Douglas (2013) does not—as the name suggests—rely on parametric estimation, but uses a distance-based algorithm on the observed item responses for classifying examinees. Specifically, proficiency-class membership is determined by comparing an examinee’s observed item response profile \mathbf{y} with each of the ideal item response profiles of the realizable $2^K = M$ proficiency classes. Let $\boldsymbol{\eta}_i$ denote the J -dimensional ideal item response vector of examinee i . (All examinees in proficiency class C_m share the same attribute profile; hence, $\boldsymbol{\eta}_i = \boldsymbol{\eta}_{i \in C_m} = \boldsymbol{\eta}_m$.) As the Q-matrix and the M realizable proficiency classes are known, the construction of their ideal item response profiles $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots, \boldsymbol{\eta}_M$ is straightforward. The NPC estimator $\hat{\boldsymbol{\alpha}}$ of an examinee’s attribute profile $\boldsymbol{\alpha}_i$ is defined as the attribute profile underlying the ideal item response profile that among all ideal item response profiles minimizes the distance to an examinee’s observed item response vector—formally,

$$\hat{\boldsymbol{\alpha}}_i = \arg \min_{m \in \{1, 2, \dots, M\}} d(\mathbf{Y}_i, \boldsymbol{\eta}_m) \tag{1}$$

with \mathbf{Y}_i denoting the J -dimensional item response vector of examinee i . Hence, the choice of the specific distance measure $d(\cdot)$ for the loss function of Eq. 1 is of critical importance in determining $\hat{\boldsymbol{\alpha}}_i$.

A distance measure often used with binary data is the Hamming distance defined as the number of disagreements between two vectors—for the NPC method:

$$d_H(\mathbf{Y}, \boldsymbol{\eta}) = \sum_{j=1}^J |Y_j - \eta_j|$$

Weighted Hamming distances allow to adjust for different levels of variability in the item responses; for example, in using the inverse of the item sample variance to increase the impact on the distance function of items with smaller variance:

$$d_{wH}(\mathbf{Y}, \boldsymbol{\eta}) = \sum_{j=1}^J \frac{1}{\bar{p}_j(1 - \bar{p}_j)} |Y_j - \eta_j|$$

(\bar{p}_j is the proportion of correct responses to the j th item). A purported advantage of the weighted Hamming distance is the substantial reduction in the number of ties, which can be an issue especially with short tests.

Wang and Douglas (2015) proved that under certain regularity conditions statistical consistency of the NPC estimator $\hat{\boldsymbol{\alpha}}$ is guaranteed for any DCM as long as the probability of a correct response is greater than 0.5 for examinees who master the required attributes and less than 0.5 for examinees who do not master all the required attributes. An implementation of the NPC method is available in the R package NPCD (Zheng & Chiu, 2014).

2.3. *The General Nonparametric Classification Method*

The consistency conditions of the NPC estimator $\hat{\boldsymbol{\alpha}}$ identified by Wang and Douglas (2015) are sometimes difficult to meet for complex DCMs that model the probability of a correct response to an item as an increasing function of the number of attributes that are required for an item and mastered by an examinee. As an example, consider a domain characterized by two attributes; the realizable proficiency classes are $\boldsymbol{\alpha}_1 = (00)$, $\boldsymbol{\alpha}_2 = (10)$, $\boldsymbol{\alpha}_3 = (01)$, and $\boldsymbol{\alpha}_4 = (11)$. For an item having attribute vector $\mathbf{q} = (11)$, the corresponding ideal item responses are $\eta_1 = 0$, $\eta_2 = 0$, $\eta_3 = 0$, and $\eta_4 = 1$. Assume this item conforms to the DINA model, with $g = 0.1$ and $1 - s =$

0.9. (The equivalent parameterization using the G-DINA model is $\beta = (\beta_0, \beta_1, \beta_2, \beta_{12})' = (0.1, 0, 0, 0.8)'$.) The probabilities of answering the item correctly for the four proficiency classes are 0.1, 0.1, 0.1, and 0.9, respectively. Thus, the ideal responses 0, 0, 0, and 1 are, indeed, the most likely responses of the four proficiency classes. However, this may no longer be true if the data conform to a more complex model, say, the saturated G-DINA model with parameter vector $\beta = (\beta_0, \beta_1, \beta_2, \beta_{12})^T = (0.1, 0.4, 0.6, -0.2)^T$. Then, for the four proficiency classes, the probabilities of a correct item response are 0.1, 0.5, 0.7, and 0.9. Thus, the ideal item responses 0, 0, 0, and 1 are no longer the most likely responses. Said differently, using the conjunctive η in this instance to model the relation between \mathbf{q} and α may result in examinee misclassifications.

Recall that the DINA model and the DINO model are the prototypic conjunctive and disjunctive DCM, respectively. They define two extremes of a continuum describing the relation between \mathbf{q} and α . Based on this observation, Chiu et al. (2018) proposed a weighted ideal item response defined as the convex combination of the conjunction parameter η and the disjunction parameter ω . The weights are estimated from the data; hence, the relative contribution of η and ω to the weighted ideal item response is tailored to the complexity of the DCM underlying the data, which elegantly resolves the limitations of the NPC method. Hence, the GNPC method can be used with any DCM that can be represented as a general DCM. Also, no a priori knowledge of the model underlying the data is required, which is presumably the most distinctive feature of the GNPC method.

The conjunctive ideal item response vector of examinee i is henceforth written as $\eta_i^{(c)}$, with elements $\eta_{ij}^{(c)} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}$ defined earlier as the conjunction parameter of item j for the DINA model; $\eta_i^{(d)}$ denotes the disjunctive ideal item response vector, with elements $\eta_{ij}^{(d)} = 1 - \prod_{k=1}^K (1 - \alpha_{ik})^{q_{jk}}$ defined previously for the DINO model. For each item j and proficiency class \mathcal{C}_m , the weighted ideal response $\eta_{mj}^{(w)}$ is defined as the convex combination

$$\eta_{mj}^{(w)} = w_{mj} \eta_{mj}^{(c)} + (1 - w_{mj}) \eta_{mj}^{(d)} \quad (2)$$

subject to $0 \leq w_{mj} \leq 1$. The distance between the responses to item j and the weighted ideal responses $\eta_{mj}^{(w)}$ of examinees in \mathcal{C}_m is defined as the sum of squared deviations:

$$d_{mj} = \sum_{i \in \mathcal{C}_m} (Y_{ij} - \eta_{mj}^{(w)})^2 = \sum_{i \in \mathcal{C}_m} (Y_{ij} - w_{mj} \eta_{mj}^{(c)} - (1 - w_{mj}) \eta_{mj}^{(d)})^2 \quad (3)$$

Thus, w_{mj} can be estimated by minimizing d_{mj} :

$$\tilde{w}_{mj} = \frac{\sum_{i \in \mathcal{C}_m} (Y_{ij} - \eta_{mj}^{(d)})}{\|\mathcal{C}_m\| (\eta_{mj}^{(c)} - \eta_{mj}^{(d)})} \quad (4)$$

subject to $\eta_{mj}^{(c)} - \eta_{mj}^{(d)} \neq 0$; $\|\mathcal{C}_m\|$ indicates the number of examinees in proficiency class \mathcal{C}_m . Some algebra shows

$$\tilde{w}_{mj} = 1 - \bar{Y}_{j\mathcal{C}_m} \quad (5)$$

(As Y_{ij} is a binary random variable, $0 \leq \bar{Y}_{j\mathcal{C}_m} \leq 1$; hence, $0 \leq \tilde{w}_{mj} \leq 1$, which is in accord with the constraint $0 \leq w_{mj} \leq 1$.) After \tilde{w}_{mj} has been determined, $\tilde{\eta}_{mj}^{(w)}$ is computed as

$$\tilde{\eta}_{mj}^{(w)} = \tilde{w}_{mj} \eta_{mj}^{(c)} + (1 - \tilde{w}_{mj}) \eta_{mj}^{(d)} \quad (6)$$

In this manner, the estimators of all possible weighted ideal response vectors $\tilde{\eta}_1^{(w)}, \tilde{\eta}_2^{(w)}, \dots, \tilde{\eta}_M^{(w)}$ can be constructed for the M proficiency classes (each of which identified by its attribute profile α_m). Finally, note that the relation

$$\tilde{\eta}_{mj}^{(w)} = \bar{Y}_j C_m \tag{7}$$

follows directly from Eqs. 5 and 6.

Comment: Eq. 4 relies on $\eta_{mj}^{(c)}$ and $\eta_{mj}^{(d)}$, which implies that an initial classification of examinees is required as input to the estimation of w_{mj} (and, of course, $\eta_{mj}^{(w)}$). Said differently, in the previous exposition, examinees' true attribute profiles α were assumed to be known and used to construct $\tilde{\eta}_1^{(w)}, \tilde{\eta}_2^{(w)}, \dots, \tilde{\eta}_M^{(w)}$; hence, they are global minimizers of Eq. 3. The “tilde” notation is used to distinguish the case where the α are known from a situation where the attribute profiles α are unknown and must be estimated from the data to determine the ideal response vectors; this case is discussed next.

Examinees' true attribute profiles α are never known, so they must be estimated from the data to provide the input to initialize the GNPC algorithm. Denote this initial estimator of α by $\hat{\alpha}^{(0)}$, which allows to derive the estimators $\hat{\eta}^{(c)}$ and $\hat{\eta}^{(d)}$ that are then used to obtain the estimators

$$\hat{w}_{mj} = \frac{\sum_{i \in \hat{C}_m} (Y_{ij} - \hat{\eta}_{mj}^{(d)})}{\|\hat{C}_m\| (\hat{\eta}_{mj}^{(c)} - \hat{\eta}_{mj}^{(d)})} \tag{8}$$

(\hat{C}_m indicates that membership in proficiency class m is determined based on $\hat{\alpha}^{(0)}$) and

$$\hat{\eta}_{mj}^{(w)} = \hat{w}_{mj} \hat{\eta}_{mj}^{(c)} + (1 - \hat{w}_{mj}) \hat{\eta}_{mj}^{(d)} \tag{9}$$

From Eq. 8, the estimator \hat{w}_{mj} is derived:

$$\hat{w}_{mj} = 1 - \bar{Y}_j \hat{C}_m \tag{10}$$

(notice that $0 \leq \hat{w}_{mj} \leq 1$), and from Eqs. 9 and 10

$$\hat{\eta}_{mj}^{(w)} = \bar{Y}_j \hat{C}_m \tag{11}$$

The GNPC estimator $\hat{\alpha}$ of an examinee's attribute profile is defined as the attribute profile underlying the estimated weighted ideal item response profile that among all estimated weighted ideal item response profiles minimizes the loss function defined in terms of the distance to an examinee's observed item response vector:

$$\hat{\alpha}_i = \arg \min_{m \in \{1, 2, \dots, M\}} d(Y_i, \hat{\eta}_m^{(w)})$$

where

$$d(Y_i, \hat{\eta}_m^{(w)}) = \sum_{j=1}^J d(Y_{ij}, \hat{\eta}_{mj}^{(w)}) = \sum_{j=1}^J (Y_{ij} - \hat{\eta}_{mj}^{(w)})^2$$

3. Consistency Theory of the Proficiency-Class Estimator $\hat{\alpha}$ of the General Nonparametric Classification Method

The consistency theory of the proficiency-class estimator $\hat{\alpha}$ of the GNPC method consists of four lemmas and one proposition that specify the theoretical requirements for two consistency theorems. A few technical preliminaries are warranted. First, note that Eq. 2 implies that if $w_{mj} = 1$, then $\eta_{mj}^{(w)}$ reduces to $\eta_{mj}^{(c)}$, and if $w_{mj} = 0$, then $\eta_{mj}^{(w)}$ reduces to $\eta_{mj}^{(d)}$. In these two cases, the underlying DCM corresponds to the DINA model and the DINO model, respectively. For the DINA model, $\hat{\alpha}$ was proven to be consistent by Wang and Douglas (2015); the proof can be readily extended to the DINO model based on the “dual” relation of the two models (see Köhn & Chiu, 2016b). Hence, these two cases are of no further concern here. Second, Eq. 4 shows that w_{mj} can only be estimated if $\eta_{mj}^{(c)} - \eta_{mj}^{(d)} \neq 0$ is satisfied. If $\eta_{mj}^{(c)} = \eta_{mj}^{(d)}$, then either $\eta_{mj}^{(c)} = \eta_{mj}^{(d)} = 0$, implying $\eta_{mj}^{(w)} = \eta_{mj}^{(c)} = \eta_{mj}^{(d)} = 0$ (i.e., all examinees in \mathcal{C}_m have failed all attributes required for item j), or $\eta_{mj}^{(c)} = \eta_{mj}^{(d)} = 1$ implying $\eta_{mj}^{(w)} = \eta_{mj}^{(c)} = \eta_{mj}^{(d)} = 1$ (i.e., all examinees in \mathcal{C}_m have mastered all attributes required for item j). Hence, the DCM underlying these two instances is the DINA model or the DINO model, which means that the proof of the consistency of $\hat{\alpha}$ presented in Wang and Douglas (2015) also applies to these two cases that are, therefore, of no further concern here either. Third, the case $\eta_{mj}^{(c)} = 1$ and $\eta_{mj}^{(d)} = 0$ is logically impossible because $\eta_{mj}^{(c)} \leq \eta_{mj}^{(d)}$ is always true. ($\eta_{mj}^{(c)} = 1$ is true only if all item attributes required for item j are mastered—of course, then $\eta_{mj}^{(d)} = 1$ too; if only a subset of the required item attributes is mastered, then $\eta_{mj}^{(c)} = 0$, whereas $\eta_{mj}^{(d)} = 1$; finally, if none of the attributes are mastered, then $\eta_{mj}^{(c)} = \eta_{mj}^{(d)} = 0$ —a scenario just discussed.) Fourth, consequently, w_{mj} is estimable only if $\eta_{mj}^{(c)} = 0$ and $\eta_{mj}^{(d)} = 1$. Therefore, proving the consistency of $\hat{\alpha}$ needs to concern only the case $\hat{\eta}_{mj}^{(w)} = \bar{Y}_j \hat{c}_m$ (see Eq. 11). In summary, the complex relations between \hat{w}_{mj} and $\hat{\eta}_{mj}^{(w)}$ can be described for the different values of $\eta_{mj}^{(c)}$ and $\eta_{mj}^{(d)}$ as

$$\hat{w}_{mj} = \begin{cases} \text{not defined} & \text{if } \eta_{mj}^{(c)} = \eta_{mj}^{(d)} = \eta_{mj}^{(w)} = 0 \\ 1 - \bar{Y}_j \hat{c}_m & \text{if } \eta_{mj}^{(c)} = 0 \wedge \eta_{mj}^{(d)} = 1 \\ \text{not defined} & \text{if } \eta_{mj}^{(c)} = \eta_{mj}^{(d)} = \eta_{mj}^{(w)} = 1 \end{cases}$$

and

$$\hat{\eta}_{mj}^{(w)} = \begin{cases} \eta_{mj}^{(c)} & \text{if } w_{mj} = 1 \\ \eta_{mj}^{(d)} & \text{if } w_{mj} = 0 \\ \bar{Y}_j \hat{c}_m & \text{otherwise} \end{cases}$$

The following assumptions are made:

- A1** The item response vectors Y_1, Y_2, \dots, Y_N for examinees 1, 2, \dots , N are statistically independent.
- A2** For examinee i , the item responses $Y_{i1}, Y_{i2}, \dots, Y_{iJ}$ are locally independent.
- A3** The Q-matrix is complete.
- A4** For the number of examinees and the length of the test, the relation $N e^{-2J\varepsilon^2} \rightarrow \infty$ as $J \rightarrow \infty$ holds $\forall \varepsilon > 0$.

Recall the distinction between a scenario where examinees’ attribute profiles α are assumed to be known and a scenario where the attribute profiles α are unknown and must be estimated

from the data to initialize the GNPC algorithm. First, consider the scenario assuming that attribute profiles α are known.

Lemma 1. *Suppose the data conform to any DCM that can be expressed in terms of the G-DINA model. For each item j and each proficiency class C_m , $\tilde{\eta}_{mj}^{(w)}$ is defined as in Eq. 6. N_m denotes the number of examinees in proficiency class C_m . Let $S_j(\alpha_m) = E(Y_j \mid \alpha_m)$. Then, $\sqrt{N_m} (\tilde{\eta}_{mj}^{(w)} - S_j(\alpha_m)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, S_j(\alpha_m)(1 - S_j(\alpha_m)))$ for all j .*

Proof. Suppose examinee i belongs to proficiency class C_m —that is, $\alpha_i = \alpha_m$. The response Y_{ij} is a binary random variable, with (conditional) expectation $S_j(\alpha_m)$. Hence, the variance of Y_{ij} is

$$\text{Var}(Y_{ij} \mid \alpha_m) = S_j(\alpha_m)(1 - S_j(\alpha_m))$$

with $0 < S_j(\alpha_m)(1 - S_j(\alpha_m)) < \infty$. Due to the central limit theorem

$$\sqrt{N_m} (\bar{Y}_{jC_m} - S_j(\alpha_m)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, S_j(\alpha_m)(1 - S_j(\alpha_m))) \quad \forall j \quad (12)$$

where \bar{Y}_{jC_m} is the mean of the responses Y_{ij} in proficiency class C_m . Because $\tilde{\eta}_{mj}^{(w)} = \bar{Y}_{jC_m}$ (see Eq. 7), Eq. 12 implies

$$\sqrt{N_m} (\tilde{\eta}_{mj}^{(w)} - S_j(\alpha_m)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, S_j(\alpha_m)(1 - S_j(\alpha_m))) \quad \forall j \quad (13)$$

□

Lemma 1 specifies the asymptotic distribution of $\tilde{\eta}_{mj}^{(w)}$.

Second, consider the scenario where examinees' proficiency classes α are unknown and must be estimated. The asymptotic distribution of the estimator $\hat{\eta}_{mj}^{(w)}$ is presented in Theorem 1; convergence of $\hat{\eta}_{mj}^{(w)}$ to $\tilde{\eta}_{mj}^{(w)}$ —a requirement for Theorem 1—is established in Lemma 2.

Lemma 2. *Suppose $\hat{\alpha}^{(0)}$ is a consistent estimator of α . $\tilde{\eta}_{mj}^{(w)}$ and $\hat{\eta}_{mj}^{(w)}$ are estimators of $\eta_{mj}^{(w)}$ as defined in Eqs. 6 and 9, respectively. Assume $J \rightarrow \infty$ and $J < N$. Then $\hat{\eta}_{mj}^{(w)} \xrightarrow{\mathcal{P}} \tilde{\eta}_{mj}^{(w)}$ for all j .*

Proof. Because $\hat{\alpha}^{(0)}$ is a consistent estimator of α

$$P\left(\bigcup_{i=1}^N \{|\hat{\alpha}_i^{(0)} - \alpha_i| > \varepsilon\}\right) \rightarrow 0 \quad (14)$$

as $J \rightarrow \infty$. Equations 6 and 9 show that $\hat{\eta}_{mj}^{(w)} = \tilde{\eta}_{mj}^{(w)}$ if $\hat{\alpha}_i^{(0)} = \alpha_i$ for all i , which can be expressed as

$$P(\{\hat{\alpha}_i^{(0)} = \alpha_i\}) \leq P(\{\hat{\eta}_{mj}^{(w)} = \tilde{\eta}_{mj}^{(w)}\})$$

or equivalently as

$$P(\{\hat{\eta}_{mj}^{(w)} \neq \tilde{\eta}_{mj}^{(w)}\}) \leq P(\{\hat{\alpha}_i^{(0)} \neq \alpha_i\}) \quad \forall i$$

Hence, for every $\varepsilon > 0$,

$$P(|\hat{\eta}_{mj}^{(w)} - \tilde{\eta}_{mj}^{(w)}| > \varepsilon) \leq P\left(\bigcup_{i=1}^N \{|\hat{\alpha}_i^{(0)} - \alpha_i| > \varepsilon\}\right) \quad (15)$$

Due to Eqs. 14, 15 also implies

$$P(|\hat{\eta}_{mj}^{(w)} - \tilde{\eta}_{mj}^{(w)}| > \varepsilon) \rightarrow 0$$

as $J \rightarrow \infty$. Note that $J \rightarrow \infty$ implies $N \rightarrow \infty$ because $N > J$. Therefore,

$$\hat{\eta}_{mj}^{(w)} \xrightarrow{\mathcal{P}} \tilde{\eta}_{mj}^{(w)}$$

for all j . □

Theorem 1. Let $\tilde{\eta}_{mj}^{(w)}$ and $\hat{\eta}_{mj}^{(w)}$ be the estimators of $\eta_{mj}^{(w)}$ with properties as specified in Lemmas 1 and 2. Assume $N > J$ and $J \rightarrow \infty$. Also, assume $\hat{\alpha}_i^{(0)}$ is a consistent estimator of α_i . Then $(\hat{\eta}_{mj}^{(w)} - S_j(\alpha_m)) \xrightarrow{\mathcal{D}} \frac{1}{\sqrt{N_m}} \mathcal{N}(0, S_j(\alpha_m)(1 - S_j(\alpha_m)))$ for all j .

Proof. Lemma 1 states that

$$\sqrt{N_m} (\tilde{\eta}_{mj}^{(w)} - S_j(\alpha_m)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, S_j(\alpha_m)(1 - S_j(\alpha_m)))$$

which can be written as

$$(\tilde{\eta}_{mj}^{(w)} - S_j(\alpha_m)) \xrightarrow{\mathcal{D}} \frac{1}{\sqrt{N_m}} \mathcal{N}(0, S_j(\alpha_m)(1 - S_j(\alpha_m))) \quad (16)$$

Lemma 2 states that

$$\hat{\eta}_{mj}^{(w)} - \tilde{\eta}_{mj}^{(w)} \xrightarrow{\mathcal{P}} 0 \quad \forall j \quad (17)$$

Applying Slutsky's theorem to Eqs. 16 and 17 results in

$$\hat{\eta}_{mj}^{(w)} - \tilde{\eta}_{mj}^{(w)} + \tilde{\eta}_{mj}^{(w)} - S_j(\alpha_m) = \hat{\eta}_{mj}^{(w)} - S_j(\alpha_m) \xrightarrow{\mathcal{D}} \frac{1}{\sqrt{N_m}} \mathcal{N}(0, S_j(\alpha_m)(1 - S_j(\alpha_m)))$$

for all j . □

Theorem 1 asserts that the distribution of $\hat{\eta}_{mj}^{(w)}$ is asymptotically normal.

Define the distance $d_i(\alpha_m) = d(Y_i, \hat{\eta}_m^{(w)})$, with $d_{ij}(\alpha_m) = d(Y_{ij}, \hat{\eta}_{mj}^{(w)})$; the notation $d_i(\alpha_m)$ is chosen for emphasis that the distance between the item response vector Y of examinee i and the weighted ideal item response vector $\hat{\eta}_m^{(w)}$ of proficiency class \mathcal{C}_m is a function of its attribute profile.

Lemma 3. Suppose the item response vectors Y_1, Y_2, \dots, Y_N are statistically independent and the Q -matrix is complete. Then, for each examinee i , the true attribute profile minimizes $E(d_i(\alpha_m))$ for all m .

Proof. Due to completeness of the Q-matrix, the identifiability of all realizable attribute profiles is guaranteed. Suppose α_m is the true attribute profile of examinee i ; $\alpha_{m'}$ is some other attribute profile. In case of the true attribute profile α_m

$$\begin{aligned} E(d_{ij}(\alpha_m)) &= E((Y_{ij} - \hat{\eta}_{mj}^{(w)})^2) \\ &= E(Y_{ij}^2 - 2\hat{\eta}_{mj}^{(w)}Y_{ij} + (\hat{\eta}_{mj}^{(w)})^2) \\ &= E(Y_{ij}^2) - 2E(\hat{\eta}_{mj}^{(w)}Y_{ij}) + E((\hat{\eta}_{mj}^{(w)})^2) \end{aligned} \tag{18}$$

Y_{ij} is a binary random variable; hence, $E(Y_{ij}^2) = E(Y_{ij})$, and Eq. 18 can be reduced to

$$\begin{aligned} &E(Y_{ij}) - 2E(\bar{Y}_j c_m Y_{ij}) + E((\hat{\eta}_{mj}^{(w)})^2) \\ &= S_j(\alpha_m) - \frac{2}{N_m} E\left(Y_{ij}Y_{ij} + \sum_{i' \neq i, i'=1}^{N_m} Y_{ij}Y_{i'j}\right) + \text{Var}(\hat{\eta}_{mj}^{(w)}) + E^2(\hat{\eta}_{mj}^{(w)}) \end{aligned} \tag{19}$$

Independence of Y_{ij} and $Y_{i'j}$ implies $E(Y_{ij}Y_{i'j}) = E(Y_{ij})E(Y_{i'j})$ so that the RHS of Eq. 19 becomes

$$\begin{aligned} &S_j(\alpha_m) - \frac{2}{N_m}(E(Y_{ij}^2) + (N_m - 1)E(Y_{ij})E(Y_{i'j})) + \frac{S_j(\alpha_m)(1 - S_j(\alpha_m))}{N_m} + S_j^2(\alpha_m) \\ &= S_j(\alpha_m) - \frac{2}{N_m}(S_j(\alpha_m)(1 - S_j(\alpha_m)) + S_j^2(\alpha_m) + (N_m - 1)S_j^2(\alpha_m)) \\ &\quad + \frac{S_j(\alpha_m)(1 - S_j(\alpha_m))}{N_m} + S_j^2(\alpha_m) \\ &= S_j(\alpha_m) - \frac{2}{N_m}S_j(\alpha_m)(1 - S_j(\alpha_m)) - 2S_j^2(\alpha_m) + \frac{S_j(\alpha_m)(1 - S_j(\alpha_m))}{N_m} + S_j^2(\alpha_m) \\ &= S_j(\alpha_m) - S_j^2(\alpha_m) - \frac{1}{N_m}S_j(\alpha_m)(1 - S_j(\alpha_m)) \\ &= \frac{N_m - 1}{N_m}S_j(\alpha_m)(1 - S_j(\alpha_m)) \end{aligned}$$

Conversely, if the incorrect attribute profile $\alpha_{m'}$ is assumed to be the true attribute profile of examinee i , then

$$\begin{aligned} E(d_{ij}(\alpha_{m'})) &= E((Y_{ij} - \hat{\eta}_{m'j}^{(w)})^2) \\ &= E(Y_{ij}) - 2E(\bar{Y}_j c_{m'} Y_{ij}) + E((\hat{\eta}_{m'j}^{(w)})^2) \\ &= S_j(\alpha_m) - \frac{2}{N_{m'}} E\left(Y_{ij} \sum_{i'=1}^{N_{m'}} Y_{i'j}\right) + \text{Var}(\hat{\eta}_{m'j}^{(w)}) + E^2(\hat{\eta}_{m'j}^{(w)}) \\ &= S_j(\alpha_m) - \frac{2}{N_{m'}} E(Y_{ij}) \sum_{i'=1}^{N_{m'}} E(Y_{i'j}) + \frac{S_j(\alpha_{m'})(1 - S_j(\alpha_{m'}))}{N_{m'}} + S_j^2(\alpha_{m'}) \\ &= S_j(\alpha_m) - \frac{2}{N_{m'}} N_{m'} S_j(\alpha_m) S_j(\alpha_{m'}) + \frac{S_j(\alpha_{m'})(1 - S_j(\alpha_{m'}))}{N_{m'}} + S_j^2(\alpha_{m'}) \end{aligned}$$

$$= S_j(\boldsymbol{\alpha}_m) - 2S_j(\boldsymbol{\alpha}_m)S_j(\boldsymbol{\alpha}_{m'}) + \frac{S_j(\boldsymbol{\alpha}_{m'})(1 - S_j(\boldsymbol{\alpha}_{m'}))}{N_{m'}} + S_j^2(\boldsymbol{\alpha}_{m'})$$

Thus, for each item j ,

$$\begin{aligned} & E(d_{ij}(\boldsymbol{\alpha}_m)) - E(d_{ij}(\boldsymbol{\alpha}_{m'})) \\ &= E((Y_{ij} - \hat{\eta}_{mj}^{(w)})^2) - E((Y_{ij} - \hat{\eta}_{m'j}^{(w)})^2) \\ &= \frac{N_m - 1}{N_m} S_j(\boldsymbol{\alpha}_m)(1 - S_j(\boldsymbol{\alpha}_m)) - S_j(\boldsymbol{\alpha}_m) + 2S_j(\boldsymbol{\alpha}_m)S_j(\boldsymbol{\alpha}_{m'}) \\ &\quad - \frac{S_j(\boldsymbol{\alpha}_{m'})(1 - S_j(\boldsymbol{\alpha}_{m'}))}{N_{m'}} - S_j^2(\boldsymbol{\alpha}_{m'}) \\ &= -S_j^2(\boldsymbol{\alpha}_m) - \frac{1}{N_m} S_j(\boldsymbol{\alpha}_m)(1 - S_j(\boldsymbol{\alpha}_m)) + 2S_j(\boldsymbol{\alpha}_m)S_j(\boldsymbol{\alpha}_{m'}) \\ &\quad - \frac{1}{N_{m'}} S_j(\boldsymbol{\alpha}_{m'})(1 - S_j(\boldsymbol{\alpha}_{m'})) - S_j^2(\boldsymbol{\alpha}_{m'}) \\ &= -(S_j(\boldsymbol{\alpha}_m) - S_j(\boldsymbol{\alpha}_{m'}))^2 - \frac{1}{N_m} S_j(\boldsymbol{\alpha}_m)(1 - S_j(\boldsymbol{\alpha}_m)) - \frac{1}{N_{m'}} S_j(\boldsymbol{\alpha}_{m'})(1 - S_j(\boldsymbol{\alpha}_{m'})) \\ &= -(S_j(\boldsymbol{\alpha}_m) - S_j(\boldsymbol{\alpha}_{m'}))^2 - \text{Var}(\hat{\eta}_{mj}^{(w)}) - \text{Var}(\hat{\eta}_{m'j}^{(w)}) \\ &< 0 \end{aligned}$$

Therefore, $E(d_{ij}(\boldsymbol{\alpha}_m)) < E(d_{ij}(\boldsymbol{\alpha}_{m'}))$ for all j , which suggests

$$\sum_{j=1}^J E(d_{ij}(\boldsymbol{\alpha}_m)) < \sum_{j=1}^J E(d_{ij}(\boldsymbol{\alpha}_{m'}))$$

Then,

$$\begin{aligned} E\left(\sum_{j=1}^J d_{ij}(\boldsymbol{\alpha}_m)\right) &< E\left(\sum_{j=1}^J d_{ij}(\boldsymbol{\alpha}_{m'})\right) \\ E(d_i(\boldsymbol{\alpha}_m)) &< E(d_i(\boldsymbol{\alpha}_{m'})) \end{aligned}$$

due to $d_i(\boldsymbol{\alpha}_m) = \sum_{j=1}^J d_{ij}(\boldsymbol{\alpha}_m)$. These inequalities imply that the expected distance between the ideal and manifest item response vectors is minimized by the true attribute profile. \square

Lemma 3 maintains that for examinee i in \mathcal{C}_m , the distance between the item response vector \mathbf{Y}_i and the ideal response vector $\hat{\boldsymbol{\eta}}_m^{(w)}$ is minimized if the true attribute profile is used to construct $\hat{\boldsymbol{\eta}}_m^{(w)}$.

Proposition. *Suppose the item response vectors $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N$ are statistically independent and the Q -matrix is complete. If $\boldsymbol{\alpha}_m$ is the true attribute profile of examinee i and $\boldsymbol{\alpha}_{m'}$ a different attribute profile, then $\lim_{J \rightarrow \infty} (E(d_i(\boldsymbol{\alpha}_{m'})) - E(d_i(\boldsymbol{\alpha}_m))) = \infty$ is true.*

Proof. Based on Lemma 3,

$$\begin{aligned}
 & \lim_{J \rightarrow \infty} \left(E(d_i(\boldsymbol{\alpha}_{m'})) - E(d_i(\boldsymbol{\alpha}_m)) \right) \\
 &= \lim_{J \rightarrow \infty} \sum_{j=1}^J \left((S_j(\boldsymbol{\alpha}_m) - S_j(\boldsymbol{\alpha}_{m'}))^2 + \text{Var}(\hat{\eta}_{mj}^{(w)}) + \text{Var}(\hat{\eta}_{m'j}^{(w)}) \right) \\
 &> \lim_{J \rightarrow \infty} \sum_{j=1}^J (\text{Var}(\hat{\eta}_{mj}^{(w)}) + \text{Var}(\hat{\eta}_{m'j}^{(w)})) \\
 &> \lim_{J \rightarrow \infty} J \left(\min_j (\text{Var}(\hat{\eta}_{mj}^{(w)})) + \min_j (\text{Var}(\hat{\eta}_{m'j}^{(w)})) \right) \\
 &= \infty
 \end{aligned} \tag{20}$$

□

Let $\bar{d}_i(\boldsymbol{\alpha}_m) = \frac{1}{J} \sum_{j=1}^J d_{ij}(\boldsymbol{\alpha}_m)$ denote the average distance for examinee i across all j .

Lemma 4. *Assume local independence; then, for examinee i , $\bar{d}_i(\boldsymbol{\alpha}_m) \xrightarrow{\mathcal{P}} E(d_{ij}(\boldsymbol{\alpha}_m))$ uniformly as $J \rightarrow \infty$.*

Proof. Hoeffding’s (1963) inequality states $P\left(\left|\frac{1}{N} \sum_{i=1}^N X_i - E(X_i)\right| \geq \epsilon\right) < 2 \exp(-2N\epsilon^2)$, where X_1, X_2, \dots, X_N are iid random variables and $0 \leq X_i \leq 1$ for all i . Hoeffding’s inequality is used to show that $\lim_{J \rightarrow \infty} P\left(\max_m \left| \bar{d}_i(\boldsymbol{\alpha}_m) - E(d_{ij}(\boldsymbol{\alpha}_m)) \right| \geq \epsilon\right) = 0$. Specifically, due to local independence, $Y_{i1}, Y_{i2}, \dots, Y_{iJ}$ are independent conditional on $\boldsymbol{\alpha}$. Hence, $d_{i1}(\boldsymbol{\alpha}_m), d_{i2}(\boldsymbol{\alpha}_m), \dots, d_{iJ}(\boldsymbol{\alpha}_m)$ are independent as well. Also, because $0 \leq d_{ij}(\boldsymbol{\alpha}_m) \leq 1$, the conditions for using Hoeffding’s inequality are satisfied, and thus, for every $\epsilon > 0$,

$$P\left(\left|\bar{d}_i(\boldsymbol{\alpha}_m) - E(d_{ij}(\boldsymbol{\alpha}_m))\right| \geq \epsilon\right) < 2 \exp(-2J\epsilon^2)$$

is true. Proposition 3 in Wang and Douglas (2015) is then used to show $\lim_{J \rightarrow \infty} P\left(\max_m \left| \bar{d}_i(\boldsymbol{\alpha}_m) - E(d_{ij}(\boldsymbol{\alpha}_m)) \right| \geq \epsilon\right) = 0$. Hence, $\bar{d}_i(\boldsymbol{\alpha}_m) \xrightarrow{\mathcal{P}} E(d_{ij}(\boldsymbol{\alpha}_m))$ uniformly as $J \rightarrow \infty$. □

Lemma 4 asserts that $\bar{d}_i(\boldsymbol{\alpha}_m)$ converges to $E(d_{ij}(\boldsymbol{\alpha}_m))$, which is required for the proof of Theorem 2.

Theorem 2. *Suppose all assumptions of the preceding lemmas and the proposition hold, including a) statistical independence of the item response vectors $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N$, b) local independence, and c) completeness of the Q -matrix. In addition, assume $N \exp(-J\epsilon^2) \rightarrow 0$ as $J \rightarrow \infty$. Then, $\hat{\boldsymbol{\alpha}}$ obtained from the GNPC method is a uniformly consistent estimator of $\boldsymbol{\alpha}$ for each examinee i .*

Proof. Theorem 2 is proven using Theorems 1 and 2 in Wang and Douglas (2015). □

Theorem 2 establishes the consistency of the estimator $\hat{\boldsymbol{\alpha}}$ obtained by the GNPC method.

Comment: Wang and Douglas (2015) presented two consistency theorems; Theorem 1 asserts point-wise convergence in probability; Theorem 2 uniform convergence, a stronger form of statistical consistency, because all parameter estimators are guaranteed to converge. As uniform convergence implies point-wise convergence, here, only uniform convergence was considered (as

it is stated in Theorem 2). Instead of an elaborate proof, reference is just made to Theorems 1 and 2 in Wang and Douglas (2015). The connection, however, between these two theorems, and how they actually allow for constructing a proof of Theorem 2 presented above, might not be obvious. Thus, as a courtesy to the reader, a more detailed summary of the argument developed in Wang and Douglas (2015), and how it applies to the proof of Theorem 2 are provided in the subsequent paragraphs.

Key to Wang and Douglas' (2015) work are two assumptions about the independence of examinees' item response vectors and the item responses of individual examinees:

Assumption (1) The item response vectors Y_1, Y_2, \dots, Y_N of examinees 1, 2, \dots , N are statistically independent.

Assumption (2) For examinee i , the item responses $Y_{i1}, Y_{i2}, \dots, Y_{iJ}$ are statistically independent conditional on attribute vector α_i .

Let q_{jk} be the entry in the j th row and k th column of the $J \times K$ Q-matrix; define the set $\mathcal{B}_j = \{k \mid q_{jk} = 1\}$. For some number $\delta \in (0, 0.5)$, the following conditions on the parameters of the true model underlying the data must be fulfilled:

Condition (a.1) If the data conform to the DINA model, then the parameters g_j and s_j must satisfy $g_j < 0.5 - \delta$ and $s_j < 0.5 - \delta$.

Condition (a.2) If the data conform to the NIDA model, then $g_k < 0.5 - \delta$, for $k = 1, 2, \dots, K$, and $\prod_{k \in \mathcal{B}_j} (1 - s_k) > 0.5 + \delta$, for $j = 1, 2, \dots, J$ must be true.

Condition (a.3) If the data conform to the Reduced RUM, then $\pi_j^* > 0.5 + \delta$ for every j , and $r_{jk}^* < 0.5 - \delta$ for some $k \in \mathcal{B}_j$ must be true.

Condition (b) Define the set $\mathcal{A}_{m,m'} = \{j \mid \eta_{mj} \neq \eta_{m'j}\}$, where m and m' index attribute profiles of different proficiency classes among all the $M = 2^K$ realizable proficiency classes; $\text{Card}(\mathcal{A}_{m,m'}) \rightarrow \infty$ as $J \rightarrow \infty$.

Condition (c) The total number of examinees N and of items J satisfy the relation $Ne^{-2J\epsilon^2} \rightarrow 0$ as $J \rightarrow \infty \quad \forall \epsilon > 0$.

How do Wang and Douglas' (2015) assumptions and conditions relate to those made in this article? Assumption (1) in Wang and Douglas (2015) is equivalent to A_1 in this article (p. 12). Assumption (2) essentially concerns local independence and corresponds to assumption A_2 (p. 12). Conditions (a.1), (a.2), and (a.3) concern the DINA model, the NIDA model, and the Reduced RUM, respectively; they are irrelevant for the GNPC method. Condition (b) is equivalent to the completeness condition of the Q-matrix in A_3 (p. 12). Condition (c) corresponds to A_4 (p. 12). Hence, all assumptions and conditions in Wang and Douglas (2015)—with the exception of Conditions (a.1), (a.2), and (a.3)—were also used in this article.

For the proof of Theorem 1 in Wang and Douglas (2015), Assumptions (1) and (2) as well as Conditions (a.1), (a.2), (a.3), and (b) are needed; for the proof of Theorem 2, Assumptions (1) and (2) as well as Conditions (a.1), (a.2), (a.3), (b), and (c) are needed. In addition to these assumptions and conditions, Wang and Douglas (2015) presented three propositions that are required for proving Theorems 1 and 2.

The proofs of Theorems 1 and 2 of the GNPC method have a different structure than those of Wang and Douglas (2015). However, similar to their work, the proofs presented in this article required three additional antecedents: Lemma 3 (p. 15), the proposition presented on p. 18, and Lemma 4 (p. 19). The proof of Theorem 1 for the GNPC method is self-explanatory; for the proof of Theorem 2, assumptions A_1, A_2, A_3, A_4 , the proposition presented on p. 18, and Lemmas 3 and 4 were needed.

4. Discussion and Conclusion

The GNPC method does not rely on a parametric statistical model for identifying an examinee’s proficiency class. As a key advantage in comparison with other nonparametric classification procedures, the GNPC method can also account for item response probabilities that are modeled as an increasing function of the number of attributes required by that item and those mastered by an examinee. This includes all models that can be represented as a general DCM. In the past, a major obstacle to the use of the GNPC method was that the statistical properties of the proficiency-class estimator were unknown. In this article, the GNPC estimator of an examinee’s proficiency class, $\hat{\alpha}$, was proven to be statistically consistent. The consistency theory of $\hat{\alpha}$ consists of four lemmas and one proposition that specify the theoretical requirements for two theorems that establish uniform convergence of $\hat{\alpha}$ to the true α if $J \rightarrow \infty$. In conclusion, three topics remain to be briefly addressed here.

First, recall that the GNPC algorithm must be initialized by the estimator $\hat{\alpha}^{(0)}$ of an examinee’s proficiency class. The NPC method is typically used to obtain $\hat{\alpha}^{(0)}$ —this is equivalent to initializing the GNPC method by setting \hat{w}_{mj} to 1; hence,

$$\hat{\eta}_{mj}^{(w)} = \hat{w}_{mj} \hat{\eta}_{mj}^{(c)} + (1 - \hat{w}_{mj}) \hat{\eta}_{mj}^{(d)} = \hat{\eta}_{mj}^{(c)}$$

(see Eq. 9). Then, based on $\hat{\eta}_{mj}^{(w)}$, examinees’ estimated proficiency classes are updated, resulting in updated estimates of $\hat{\eta}^{(c)}$ and $\hat{\eta}^{(d)}$ that are subsequently used to obtain an update of \hat{w}_{mj} (see Eq. 8), and so on.

Second, the performance of the GNPC method was evaluated in a series of simulation studies using various conditions. These included also increasing the number of items, J , for studying the effect on the consistency of $\hat{\alpha}$. The findings corroborated the theoretical derivations and proofs presented in this article. The description of the design of these simulation studies and their results are available in Chiu et al. (2018) that the reader may wish to consult for further details. (They are not reported here to avoid redundancy.) Chiu et al. (2018) also present the results of the analysis of a real-world data set with the GNPC method.

Third, one could raise the general question whether nonparametric approaches to analyzing CD-data might not be obsolete in light of the availability of specialized software offering efficient implementations of parametric, maximum-likelihood-based methods for fitting DCMs to (educational) assessment data—for example, the R packages CDM (Robitzsch, Kiefer, George, & Uenlue, 2018) and GDINA (Ma & de la Torre, 2019). However, as was mentioned earlier, in some situations, parametric methods may be difficult to implement, or they fail entirely. Algorithms like MMLE-EM work best for large-scale assessments, where the data of at least several hundred examinees are available. But when sample sizes are small, as is typically the case with assessment data from educational micro-environments, say, for monitoring the instruction and learning process at the classroom level, then sample sizes may be simply insufficient to guarantee reliable maximum likelihood estimates of examinees’ proficiency class. (Within an applied context, the focus is typically on the evaluation of instruction and the assessment of students’ learning; hence, estimation of the item parameters is not necessarily a primary goal.) Under these circumstances, nonparametric methods may be the only viable tools for monitoring and assisting “the teaching and learning process while it is occurring” (Stout, 2002, p. 506)—that is, at the classroom level, where CD-based methods are most useful and needed. Similar restrictions apply to the availability of CD-based computerized adaptive testing (CD-CAT) in small-scale educational settings, where it would be most beneficial. But the lack of an efficient and effective computational device for the reliable assessment of examinees’ proficiency class has been a serious obstacle that, so far, has prevented the use of CD-CAT in classrooms. The algorithms of the NPC and GNPC methods,

however, can easily handle even smallest sample sizes. They are immune to the difficulties arising from unstable and unreliable estimates that parametric methods typically encounter under such conditions (Chiu et al. 2018). In addition, the NPC and GNPC methods are easy to implement and computationally inexpensive. Their minimal CPU times predestine them to be used with CD-based computerized adaptive tests (CAT) to be administered for monitoring teaching and learning at the level of individual classrooms.

As a final comment, parametric and nonparametric methods for cognitive diagnosis require that the Q-matrix of a test be known and correctly specified; this applies also to the theoretical results concerning the nonparametric GNPC method, as they have been presented in this article. However, as a remarkable aside, informal simulations (not reported here due to space restrictions) showed that the GNPC method was surprisingly robust to the misspecification of individual entries of the Q-matrix. The Q-matrices involved $K = 5$ attributes and were composed of 30 items conforming to the DINA model and 30 items to the saturated GDINA model. Even if 20% of the entries of the Q-matrix were misspecified, the GNPC method outperformed the MMLE-EM algorithm in the correct classification of examinees in all instances where samples consisted of 100 or fewer examinees.

Acknowledgments

Funding was provided by National Science Foundation (Grant No. 1552563).

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

References

- Ayers, E., Nugent, R., & Dean, N. (2008). Skill set profile clustering based on student capability vectors computed from online tutoring data. In R. S. J. de Baker, T. Barnes, & J. E. Beck (Eds.), *Educational data mining 2008: Proceedings of the 1st international conference on educational data mining, Montreal, Quebec, Canada* (pp. 210–217). Retrieved from <http://www.educationaldatamining.org/EDM2008/uploads/proc/full%20proceedings.pdf>.
- Chiu, C.-Y. (2008). *Cluster analysis for cognitive diagnosis: Theory and applications* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3337778).
- Chiu, C.-Y., & Douglas, J. A. (2013). A nonparametric approach to cognitive diagnosis by proximity to ideal response profiles. *Journal of Classification*, *30*, 225–250.
- Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, *74*, 633–665.
- Chiu, C.-Y., Sun, Y., & Bian, Y. (2018). Cognitive diagnosis for small educational programs: The general nonparametric classification method. *Psychometrika*, *83*, 355–375.
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, *34*, 115–130.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*, 179–199.
- de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333–353.
- DiBello, L. V., Roussos, L. A., & Stout, W. F. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Volume 26. Psychometrics* (pp. 979–1030). Amsterdam: Elsevier.
- Haberman, S. J., & von Davier, M. (2007). Some notes on models for cognitively based skill diagnosis. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics: Volume 26. Psychometrics* (pp. 1031–1038). Amsterdam: Elsevier.
- Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*, 333–352.
- Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*, 191–210.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, *58*(301), 13–30.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*, 258–272.

- Köhn, H.-F., & Chiu, C.-Y. (2016a). Conditions of completeness of the Q-matrix of tests for cognitive diagnosis. In A. L. van der Ark, D. M. Bolt, W.-C. Wang, J. A. Douglas, & M. Wiberg (Eds.), *Quantitative psychology research: The 80th annual meeting of the psychometric society, Beijing, 2015* (pp. 255–264). New York: Springer.
- Köhn, H.-F., & Chiu, C.-Y. (2016b). A proof of the duality of the DINA model and the DINO model. *Journal of Classification*, *33*, 171–184.
- Köhn, H.-F., & Chiu, C.-Y. (2017). A procedure for assessing completeness of the Q-matrix of cognitive diagnosis models. *Psychometrika*, *82*, 112–132.
- Leighton, J., & Gierl, M. (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge: Cambridge University Press.
- Ma, W., & de la Torre, J. (2019). *GDINA: The generalized DINA model framework. R package version 2.3.2*. Retrieved from the Comprehensive R Archive Network [CRAN] website <https://cran.r-project.org/web/packages/GDINA/>.
- Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, *33*, 379–416.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*, 187–212.
- Nichols, P. D., Chipman, S. E., & Brennan, R. L. (1995). *Cognitively diagnostic assessment*. Mahwah, NJ: Erlbaum.
- Park, Y. S., & Lee, Y.-S. (2011). Diagnostic cluster analysis of mathematics skills. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, *4*, 75–107.
- Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2018). *CDM: Cognitive diagnosis modeling. R package version 7.1-20*. Retrieved from the Comprehensive R Archive Network [CRAN] website <https://cran.r-project.org/web/packages/CDM/>.
- Rupp, A. A., Templin, J. L., & Henson, R. A. (2010). *Diagnostic Measurement. Theory, Methods, and Applications*. New York: Guilford.
- Stout, W. (2002). Psychometrics: From practice to theory and back. *Psychometrika*, *67*, 485–518.
- Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconception in the pattern classification approach. *Journal of Educational and Behavioral Statistics*, *12*, 55–73.
- Tatsuoka, K. K. (2009). *Cognitive assessment. An introduction to the rule space method*. New York: Routledge/Taylor & Francis.
- Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*, 287–305.
- von Davier, M. (2005). *A general diagnostic model applied to language testing data* (Research report No. RR-05-16). Princeton, NJ: Educational Testing Service.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*, 287–301.
- Wang, S., & Douglas, J. (2015). Consistency of nonparametric classification in cognitive diagnosis. *Psychometrika*, *80*, 85–100.
- Willse, J., Henson, R., & Templin, J. (2007). *Using sum scores or IRT in place of cognitive diagnosis models: Can existing or more familiar models do the job?* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.
- Zheng, Y., & Chiu, C.-Y. (2014). *NPCD: Nonparametric methods for cognitive diagnosis. R package version 1.0-5*. Retrieved from the Comprehensive R Archive Network [CRAN] website <http://CRAN.R-project.org/package=NPCD>.

Manuscript Received: 6 MAR 2018

Published Online Date: 6 FEB 2019