# On the Estimation of Standard Errors in Cognitive Diagnosis Models

**Michel Philipp**
Universität Zürich

**Carolin Strobl**
Universität Zürich

**Jimmy de la Torre**
The Hong Kong University

**Achim Zeileis**
Universität Innsbruck

**Abstract**

Cognitive diagnosis models (CDMs) are an increasingly popular method to assess mastery or nonmastery of a set of fine-grained abilities in educational or psychological assessments. Several inference techniques are available to quantify the uncertainty of model parameter estimates, to compare different versions of CDMs or to check model assumptions. However, they require a precise estimation of the standard errors (or the entire covariance matrix) of the model parameter estimates. In this article, it is shown analytically that the currently widely used form of calculation leads to underestimated standard errors because it only includes the item parameters, but omits the parameters for the ability distribution. In a simulation study, we demonstrate that including those parameters in the computation of the covariance matrix consistently improves the quality of the standard errors. The practical importance of this finding is discussed and illustrated using a real data example.

*Keywords*: cognitive diagnosis model, G-DINA, standard errors, information matrix.

## 1. Introduction

Cognitive diagnosis models (CDMs) are restricted latent class models that can be used to analyze response data from educational or psychological tests. In the educational context, they are becoming a popular method for measuring mastery or nonmastery of a set of fine-grained abilities (called attributes) that can be used, for example, to support teachers to recognize strengths and weaknesses of students. Lee, Park, and Taylan (2011) and Li (2011) are examples of cognitive diagnostic analyses of mathematics and language skills in large-scale assessments. However, the method has also been suggested to identify the presence or absence of psychological disorders (de la Torre, van der Ark, and Rossi 2015; Templin and a Henson 2006), or can be used for a detailed measurement of fluid intelligence using abstract reasoning tasks (Yang and Embretson 2007; Rupp, Templin, and Henson 2010).

The field of cognitive diagnostic assessments has also become a popular area for methodological research over the past 20 years. Many different versions of CDMs have been proposed to analyze responses from tests with various characteristics (e.g., models for dichotomous and polytomous responses, compensatory and noncompensatory processes). See Rupp *et al.*

This is a preprint of an article published in *Journal of Educational and Behavioral Statistics*, **43**(1), 88–115. doi:10.3102/1076998617719728

(2010) for a taxonomy of CDMs. Many of these models can be subsumed within a more general framework, such as the generalized deterministic input, noisy "and" gate (G-DINA; de la Torre 2011) model, the log-linear CDM (LCDM; Henson, Templin, and Willse 2009), or the general diagnostic model (GDM; von Davier 2008). Aside from Bayesian approaches, which are presented in the literature for different versions of CDMs (see e.g., Culpepper 2015), the model parameters are usually estimated via marginal maximum likelihood estimation (MMLE) using, for example, the EM algorithm (Dempster, Laird, and Rubin 1977; McLachlan and Krishnan 2007). In the marginal formulation of the model, a probability distribution that models the attribute space is imposed in conjunction with the traditional item response function, that models the conditional probability of a correct response given the attributes.

An important step of any practical analysis is to assess the uncertainty of the estimated model parameters using confidence intervals or significance tests. Furthermore, several techniques are available to investigate the model fit or to check the model assumptions of a CDM, including tests for (item-level) model comparisons (de la Torre and Lee 2013) and to detect differential item functioning (Hou, de la Torre, and Nandakumar 2014). These methods require a precise estimation of the model parameters and their standard errors (or the entire covariance matrix).

However, according to the CDM literature (see e.g., Chen and de la Torre 2013; George 2013; Rojas 2013; Song, Wang, Dai, and Ding 2012; de la Torre 2009, 2011) and open source software implementations (e.g., in the R package **cdm**, version 4.991-1), it is common to compute the standard errors only for the parameters which are used to specify the item response function while ignoring the parameters used to specify the joint distribution of the attributes. Consequently, this approach is frequently applied in substantive as well as in many methodological research applications.

Unfortunately, this widely used approach can lead to underestimated standard errors, as we will demonstrate in this paper. The aim of this article is to provide detailed guidance on how standard errors for cognitive diagnosis models should be computed correctly. In addition to analytic arguments, we will investigate the quality of the standard errors using simulations.

The severity of the underestimation varies considerably depending on some known factors (e.g., test length and number of attributes in the assessment), as well as unknown factors (e.g., parameters of the item response function and distribution of the attributes). In some situations, the incremental improvement with the correct approach may become negligibly small (e.g., for high test lengths). However, because the factors potentially causing underestimation are manifold, practioners cannot know upfront whether the data being analyzed is subject to underestimation of standard errors, and how severe the underestimation might be. Given that the necessary computations are straightforward, using the correct approach presented in this article is recommended to be on the "safe side". The additional computations only involve components that are already provided by the results of the estimation routine, and we provide free and open-source software for obtaining the results in practice.

In many situations the underestimation can seriously deteriorate the quality of confidence intervals and statistical tests. Hou *et al.* (2014), for example, proposed the Wald test to detect differential item functioning in CDMs, and encountered serious Type I error inflation (up to 18%). Li and Wang (2015) later found that this was caused by a substantial underestimation of the standard errors with the marginal maximum likelihood estimation (MMLE) approach.

Although it is not clear whether the underestimation they observed in their study was caused by the incorrect computation of the standard errors or otherwise, it demonstrates how the performance of the Wald test can be negatively affected by underestimated standard errors (or the entire covariance matrix). Several studies in the field of item response theory (IRT) have also demonstrated the influence of the estimation approach on the quality of procedures that require a covariance matrix. Woods, Cai, and Wang (2012), for example, found better controlled Type I error in the Wald test to detect differential item functioning in the Rasch model if the covariance matrix was computed using the supplemented EM algorithm (Cai 2008).

Other statistical issues might also cause biases in standard errors for CDMs when using MMLE. Similar to traditional latent class analysis, for example, parameter estimates sometimes converge towards the boundary of the parameter space for small data sets. This causes numerical problems in the calculation of the information matrix, which is inverted to get the covariance matrix. Posterior mode (PM) estimation has been suggested to overcome these problems (DeCarlo 2011; Garre and Vermunt 2006). However, in the CDM literature and in some frequently used software packages, the traditional maximum likelihood (ML) estimation is prevalent. Therefore, we will focus on the estimation of standard errors in this framework for this article.

The rest of the article is organized as follows. The next section contains a short formal introduction of CDMs before the correct estimation of the standard errors is discussed in detail. Later in that section, the G-DINA model will be introduced for the remaining aspects discussed in the article. In the section after next, the quality of the standard errors is investigated using simulation studies and a real data example. The last section concludes with a discussion. To simplify notation and language, we will focus on CDMs for dichotomous responses in the context of educational assessments for the rest of the article. Please note, however, that the calculation of the standard errors described here holds for all types of CDMs estimated via MMLE.

## 2. Cognitive diagnosis models

The primary goal in cognitive diagnosis modeling is to infer mastery or nonmastery of $K$ attributes from the responses of each individual to $J$ items in an assessment. For this task a $J \times K$ $Q$-matrix (Tatsuoka 1983) must be specified to identify the cognitive specification of the items, where $Q = \{q_{jk}\}$ and $q_{jk} = 1$ if attribute $k$ ($k = 1, \ldots, K$) is required to solve item $j$ ($j = 1, \ldots, J$), and 0 otherwise. The $Q$-matrix requires domain-specific knowledge, and should ideally be specified together with experts from the field for which the assessment will be needed.

Let $\boldsymbol{X}_i = \{X_{ij}\}$ be the binary response pattern of individual $i$ ($i = 1, \ldots, N$). The conditional probability of a correct response to item $j$ given the unobserved attribute profile $\boldsymbol{\alpha}_i = \{\alpha_{ik}\}$ is parametrized using a specific item response function, denoted by $P_j(\boldsymbol{\alpha}_i) = \Pr(X_{ij} = 1|\boldsymbol{\alpha}_i)$. Furthermore, let $\boldsymbol{\delta}_j$ denote the vector of all parameters used to specify $P_j(\boldsymbol{\alpha}_i)$ and, let $\boldsymbol{\delta} = (\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_J)^\top$ denote the vector of parameters that contains all item parameters. For reasons of consistency, it is usually suggested to estimate $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}_i$ using a marginal maximum likelihood approach (de la Torre 2009; Neyman and Scott 1948). The marginal probability is given by the sum over all $L = 2^K$ possible attribute patterns, called latent classes:

$$\Pr(\boldsymbol{X}_i = \boldsymbol{x}_i) = \sum_{l=1}^{L} p(\boldsymbol{\alpha}_l) \cdot \Pr(\boldsymbol{X}_i = \boldsymbol{x}_i | \boldsymbol{\alpha}_l),$$

where $\Pr(\boldsymbol{X}_i = \boldsymbol{x}_i | \boldsymbol{\alpha}_l) = \prod_{j=1}^{J} P_j(\boldsymbol{\alpha}_l)^{x_{ij}} \left[1 - P_j(\boldsymbol{\alpha}_l)\right]^{1-x_{ij}}$.
A distribution $p(\boldsymbol{\alpha}_l)$ is imposed to specify a prior probability for each latent class. Let $\boldsymbol{\pi}$ be the vector of all parameters used in the model that specifies $p(\boldsymbol{\alpha}_l)$. For this article, we choose a saturated model by estimating a probability $\pi_l = p(\boldsymbol{\alpha}_l)$ for each latent class, where $\pi_L = 1 - \sum_{l=1}^{L-1} \pi_l$. Different models can be assumed to reduce the number of parameters (de la Torre and Douglas 2004; Rupp *et al.* 2010).
Thus, let $\boldsymbol{\vartheta} = (\boldsymbol{\delta}, \boldsymbol{\pi})^\top$ be the complete vector of all model parameters of a CDM, and further $p = \dim(\boldsymbol{\delta})$ and $q = \dim(\boldsymbol{\pi})$. The marginal log-likelihood that is maximized to estimate $\boldsymbol{\vartheta}$ given the data sample $\boldsymbol{X} = \{\boldsymbol{x}_i\}$ for individuals $i = 1, \ldots, N$, is given by

$$\ell(\boldsymbol{\vartheta}; \boldsymbol{X}) = \log\left[L(\boldsymbol{\vartheta}; \boldsymbol{X})\right] = \log \prod_{i=1}^{N} \sum_{l=1}^{L} \pi_l \cdot \Pr(\boldsymbol{X}_i = \boldsymbol{x}_i | \boldsymbol{\alpha}_l),$$

and can be maximized using the EM algorithm as described in de la Torre (2009). The estimation procedure provides the posterior probability for each latent class, $\widehat{\Pr}(\boldsymbol{\alpha}_l | \boldsymbol{x}_i)$, that can be used to find $\hat{\boldsymbol{\pi}}$ and the attribute profiles $\widehat{\boldsymbol{\alpha}}_i$. However, the aim of this article is to discuss the estimation of standard errors for the estimated model parameters $\widehat{\boldsymbol{\vartheta}}$, which will be the focus of the next section.

## 2.1. Theory and estimation of standard errors

The standard errors of the estimated model parameters $\widehat{\boldsymbol{\vartheta}} = \left(\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\pi}}\right)^\top$ can be computed as the square root of the diagonal elements of the covariance matrix of $\widehat{\boldsymbol{\vartheta}}$. Regarding the two types of parameters, $\boldsymbol{\delta}$ and $\boldsymbol{\pi}$, the covariance matrix of $\widehat{\boldsymbol{\vartheta}}$ can be divided into four blocks:

$$\mathrm{Cov}(\widehat{\boldsymbol{\vartheta}}) = V_{\boldsymbol{\vartheta}} = \begin{pmatrix} V_{\boldsymbol{\delta}} & V_{\boldsymbol{\delta},\boldsymbol{\pi}} \\ V_{\boldsymbol{\pi},\boldsymbol{\delta}} & V_{\boldsymbol{\pi}} \end{pmatrix},$$

where $V_{\boldsymbol{\delta}} = \mathrm{Cov}(\widehat{\boldsymbol{\delta}})$ is the covariance matrix of the parameters used to specify the item response function, $V_{\boldsymbol{\pi}} = \mathrm{Cov}(\widehat{\boldsymbol{\pi}})$ is the covariance matrix of the parameters used to specify the distribution of the latent classes and $V_{\boldsymbol{\delta},\boldsymbol{\pi}} = V_{\boldsymbol{\pi},\boldsymbol{\delta}}^\top = \mathrm{Cov}(\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\pi}})$ is the covariance matrix between the two types of parameters.

*Complete and incomplete information matrix*

The (asymptotic) covariance matrix of $\widehat{\boldsymbol{\vartheta}}$ is equal to the inverse of the information matrix, $V_{\boldsymbol{\vartheta}} = \mathcal{I}_{\boldsymbol{\vartheta}}^{-1}$, which is defined as

$$\mathcal{I}_{\boldsymbol{\vartheta}} = E\left[\psi(\boldsymbol{\vartheta})\psi(\boldsymbol{\vartheta})^\top\right], \tag{1}$$

where

$$\psi(\boldsymbol{\vartheta}) = \left(\psi(\boldsymbol{\delta}), \psi(\boldsymbol{\pi})\right)^\top = \left(\frac{\partial \ell(\boldsymbol{\vartheta}; \boldsymbol{x})}{\partial \delta_1}, \ldots, \frac{\partial \ell(\boldsymbol{\vartheta}; \boldsymbol{x})}{\partial \delta_p}, \frac{\partial \ell(\boldsymbol{\vartheta}; \boldsymbol{x})}{\partial \pi_1}, \ldots, \frac{\partial \ell(\boldsymbol{\vartheta}; \boldsymbol{x})}{\partial \pi_q}\right)^\top$$

is the score function (i.e., the partial derivatives of the log-likelihood with respect to all model parameters).

Similar to the covariance matrix, the information matrix can be divided into four blocks:

$$\mathcal{I}_{\boldsymbol{\vartheta}} = \begin{pmatrix} \mathcal{I}_{\boldsymbol{\delta}} & \mathcal{I}_{\boldsymbol{\delta},\boldsymbol{\pi}} \\ \mathcal{I}_{\boldsymbol{\pi},\boldsymbol{\delta}} & \mathcal{I}_{\boldsymbol{\pi}} \end{pmatrix} = E\left[\begin{pmatrix} \psi(\boldsymbol{\delta})\psi(\boldsymbol{\delta})^\top & \psi(\boldsymbol{\delta})\psi(\boldsymbol{\pi})^\top \\ \psi(\boldsymbol{\pi})\psi(\boldsymbol{\delta})^\top & \psi(\boldsymbol{\pi})\psi(\boldsymbol{\pi})^\top \end{pmatrix}\right],$$

where $\mathcal{I}_{\boldsymbol{\delta}}$ is the information matrix for the parameters used to specify the item response function, $\mathcal{I}_{\boldsymbol{\pi}}$ is the information matrix for the parameters used to specify the distribution of the latent classes and $\mathcal{I}_{\boldsymbol{\delta},\boldsymbol{\pi}} = \mathcal{I}_{\boldsymbol{\pi},\boldsymbol{\delta}}^\top$ is the information matrix for the two types of parameters.

In many practical applications (e.g., tests for differential item functioning) researchers are primarily interested in the parameters $\boldsymbol{\delta}$, and thus they incorrectly compute the covariance matrix for $\widehat{\boldsymbol{\delta}}$ via the inverse of the *incomplete* information matrix $\mathcal{I}_{\boldsymbol{\delta}}$. This approach, however, considers only a submatrix of the *complete* information matrix including all model parameters $\mathcal{I}_{\boldsymbol{\vartheta}}$. It is important to note that, since $\boldsymbol{\delta}$ and $\boldsymbol{\pi}$ are generally not mutually independent in CDMs (i.e., $\mathcal{I}_{\boldsymbol{\delta},\boldsymbol{\pi}} = \mathcal{I}_{\boldsymbol{\pi},\boldsymbol{\delta}}^\top \neq \mathbf{0}$), inverting the incomplete information matrix $\mathcal{I}_{\boldsymbol{\delta}}$ systematically underestimates the standard errors for $\widehat{\boldsymbol{\delta}}$. In some cases, only the *item-wise* information matrix $\mathcal{I}_{\boldsymbol{\delta}_j}$ (a submatrix of $\mathcal{I}_{\boldsymbol{\delta}}$) is computed and inverted to get the covariance matrix of the parameter vector $\boldsymbol{\delta}_j$. However, similar to traditional IRT models (Yuan, Cheng, and Patton 2014), $\mathcal{I}_{\boldsymbol{\delta}}$ is not block-diagonal. And thus, inverting the item-wise information matrix underestimates the standard errors even stronger compared to the incomplete information matrix approach.

The above statement can be derived in a formal way using matrix algebra. Let $(\mathcal{I}_{\boldsymbol{\delta}})^{-1}$ be the covariance of $\widehat{\boldsymbol{\delta}}$, based on the incomplete information matrix and let $V_{\boldsymbol{\delta}}$ be the covariance of $\widehat{\boldsymbol{\delta}}$, based on the complete information matrix. From blockwise matrix inversion (see e.g., Banerjee and Roy 2014), it follows, that

$$V_{\boldsymbol{\delta}} \;=\; (\mathcal{I}_{\boldsymbol{\delta}})^{-1} + \Delta, \tag{2}$$

with $\Delta = (\mathcal{I}_{\boldsymbol{\delta}})^{-1}\mathcal{I}_{\boldsymbol{\delta},\boldsymbol{\pi}}V_{\boldsymbol{\pi}}\mathcal{I}_{\boldsymbol{\pi},\boldsymbol{\delta}}(\mathcal{I}_{\boldsymbol{\delta}})^{-1}$. If the inverse of $\mathcal{I}_{\boldsymbol{\vartheta}}$ exists[1] and $\mathcal{I}_{\boldsymbol{\delta},\boldsymbol{\pi}} = \mathcal{I}_{\boldsymbol{\pi},\boldsymbol{\delta}}^\top \neq \mathbf{0}$, then the diagonal elements of all terms in (2) are positive (see Appendix A), which implies,

$$\sqrt{\left[V_{\boldsymbol{\delta}}\right]_{r,r}} > \sqrt{\left[(\mathcal{I}_{\boldsymbol{\delta}})^{-1}\right]_{r,r}} \qquad r = 1, \ldots, p.$$

This means that the standard errors of the estimated parameters $\widehat{\boldsymbol{\delta}}$ are consistently underestimated if the incomplete or the item-wise – instead of the complete – information matrix is used. Later, in Section 3, we will demonstrate by means of simulations that standard errors computed using the complete information matrix are of better quality. But first, we will discuss two important techniques to estimate the information matrix.

### *Estimating the information matrix and standard errors*

Computing the (expected) information matrix by evaluating the expected value at the maximum likelihood estimate is infeasible for large assessments. The expectation must be taken

[1]The inverse exists in many practical cases. However, it does not exist, e.g., when the parameters lie at the boundary of the parameter space (but estimating standard errors for such parameters is not meaningful anyway), or when the latent classes are not completely identified by the items in the test.

over the support of the random response vector $\boldsymbol{x}_i$, which becomes very large even if $J$ (the number of items) is only moderately large (e.g., $J = 25$) and computation becomes very slow due to memory limitation.

Thus, the information matrix is often estimated by the empirical counterpart of Equation 1, given by

$$\mathcal{J}_{\boldsymbol{\vartheta},OPG} = \frac{1}{N}\left[\sum_{i=1}^{N}\psi(\boldsymbol{\vartheta};\boldsymbol{x}_i)\psi(\boldsymbol{\vartheta};\boldsymbol{x}_i)^\top\right]\Bigg|_{\vartheta=\widehat{\vartheta}}, \tag{3}$$

also known as the "outer product of gradients" (OPG) estimator, where $\psi(\boldsymbol{\vartheta};\boldsymbol{x}_i)$ is the contribution of individual $i$ to the score function.

Another estimator follows from the fact that under the true parameter values and standard regularity conditions the information matrix (as defined in Equation 1) is equivalent to the expected value of the negative Hessian matrix of the log-likelihood. Thus, the information matrix may also be estimated via

$$\mathcal{J}_{\boldsymbol{\vartheta},Hess} = -\frac{1}{N}\left[\sum_{i=1}^{N}\frac{\partial^2\ell(\boldsymbol{\vartheta};\boldsymbol{x}_i)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^\top}\right]\Bigg|_{\vartheta=\widehat{\vartheta}}. \tag{4}$$

In practice, however, (3) and (4) are evaluated at the estimated parameter values and, thus, the two estimators differ by

$$\mathcal{J}_{\boldsymbol{\vartheta},Hess} - \mathcal{J}_{\boldsymbol{\vartheta},OPG} = \frac{1}{N}\left[\sum_{i=1}^{N}\frac{1}{L(\boldsymbol{\vartheta};\boldsymbol{x}_i)}\frac{\partial^2 L(\boldsymbol{\vartheta};\boldsymbol{x}_i)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^\top}\right]\Bigg|_{\vartheta=\widehat{\vartheta}}.$$

Often (3) is easier to compute, but (4) promises a better finite sample approximation of the information matrix (McLachlan and Krishnan 2007).

From the above definitions, the standard error for the parameter $\vartheta_r$ ($r = 1, \ldots, p+q$), can be computed via the inverse of the complete information matrix, using

$$\widehat{\mathrm{se}}(\widehat{\vartheta}_r) = \sqrt{\left[(\mathcal{J}_{\boldsymbol{\vartheta},OPG})^{-1}\right]_{r,r}} \qquad \text{or} \qquad \widehat{\mathrm{se}}(\widehat{\vartheta}_r) = \sqrt{\left[(\mathcal{J}_{\boldsymbol{\vartheta},Hess})^{-1}\right]_{r,r}},$$

estimated via the outer-products of gradients or the Hessian matrix, respectively. Since the differences between the OPG and the Hessian approach turned out to be negligibly small for simple CDMs (i.e., for the DINA model introduced below, but results are not shown), we will only consider the OPG estimator for the rest of the article.

In Section 3, the improvement of the quality of the standard errors by using the inverse of the complete information matrix will be illustrated using three specific versions of CDMs. Therefore, we will briefly introduce the generalized DINA model framework proposed by de la Torre (2011), which covers other CDMs as special cases. For a comprehensive description of the framework, its relation to other general CDMs and parameter estimation, we refer the reader to the original article.

## 2.2. The G-DINA model

A comprehensive and very flexible version of a CDM is the generalized deterministic input, noisy "and" gate (G-DINA) model (de la Torre 2011). Due to its general formulation, it includes many other (more restrictive) CDMs as special cases.

For each item in the assessment, the individuals are separated into $2^{K_j^*}$ latent groups, where $K_j^*$ is the number of attributes required by item $j$ (i.e., the sum of the $j$th row in the $Q$-matrix). Presence or absence of all the other attributes does not affect the group membership of an individual. Consequently, the attribute vector $\boldsymbol{\alpha}_i$ can be reduced to the attributes required by the particular item.

Let $\boldsymbol{\alpha}_{ij}^* = (\alpha_{i1}, \ldots, \alpha_{iK_j^*})$ denote the reduced attribute vector of individual $i$ for item $j$. The conditional probability to answer item $j$ correctly is then defined by the item response function

$$P_j(\boldsymbol{\alpha}_{ij}^*) = h^{-1}\left(\delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{ik} + \sum_{k=1}^{K_j^*-1} \sum_{k'=k+1}^{K_j^*} \delta_{jkk'}\alpha_{ik}\alpha_{ik'} + \ldots + \delta_{j12\ldots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{ik}\right),$$

where $h(\cdot)$ is a link function, such as *identity*, *log* or *logit*.

The $\boldsymbol{\delta}_j$ are the model parameters of item $j$. In case of the identity link, $\delta_{j0}$ represents the baseline probability for correctly answering item $j$ when none of the required attribute has been mastered (i.e., a lucky guess); $\delta_{jk}$ is the main effect that increases (or in rare cases decreases) the probability for correctly answering item $j$ when attribute $k$ has been mastered; and the rest of the parameters represent interaction terms that can increase or decrease the response probability when two or more of the required attributes have been mastered.

Other CDMs can be obtained by restricting the parameters in the G-DINA model. An intuitive, simple and parsimonious CDM is the deterministic input, noisy "and" gate (DINA; Haertel 1989; Junker and Sijtsma 2001) model. In the DINA model the individuals are separated into two latent groups, depending on whether they have mastered all the attributes required to solve the item or not. Thus, the DINA model is a completely noncompensatory (or conjunctive) model, which means that having mastered only part of the required attributes does not increase the probability of answering the item correctly. It can be obtained from the G-DINA model by restricting all parameters except $\delta_{j0}$ and $\delta_{j12\ldots K_j^*}$ to zero. Thus, $= g_j$ is called the *guessing* probability, since individuals that have not mastered all attributes required by the item can only guess the correct response. On the other hand, $1 - (\delta_{j0} + \delta_{j12\ldots K_j^*}) = s_j$ is called the *slip* probability, since in this probabilistic model individuals that have mastered all attributes required by the item may still slip and give the wrong response.

Another CDM that can be obtained from the G-DINA model is the *additive* CDM (A-CDM). It is slightly more flexible than the DINA model because the conditional response probability can increase (or in some cases decrease) for each attribute that has been mastered. It can be obtained from the G-DINA model by restricting all interaction parameters to zero.

### *Score contributions for parameters in the G-DINA model*

To estimate the information matrix of the model parameters of the G-DINA model via OPG, the contributions of individual $i$ to the score function, $\psi(\boldsymbol{\vartheta}; \boldsymbol{x}_i)$, are required. They are given by the first-order derivative of the casewise log-likelihood contribution with respect to the model parameters:

$$\begin{aligned} \psi(\boldsymbol{\vartheta}; \boldsymbol{x}_i) &= \frac{\partial \ell(\boldsymbol{\vartheta}; \boldsymbol{x}_i)}{\partial \boldsymbol{\vartheta}} = \frac{\partial \log L(\boldsymbol{\vartheta}; \boldsymbol{x}_i)}{\partial \boldsymbol{\vartheta}} \\ &= \frac{1}{L(\boldsymbol{\vartheta}; \boldsymbol{x}_i)} \cdot \frac{\partial L(\boldsymbol{\vartheta}; \boldsymbol{x}_i)}{\partial \boldsymbol{\vartheta}} = \frac{1}{L(\boldsymbol{\vartheta}; \boldsymbol{x}_i)} \cdot \frac{\partial}{\partial \boldsymbol{\vartheta}} \left( \sum_{l=1}^{L} \pi_l \cdot \Pr(\boldsymbol{x}_i | \boldsymbol{\alpha}_l) \right). \end{aligned}$$

Using formula (A6) from the Appendix in de la Torre (2009) for the partial derivative of the conditional likelihood, the score contributions of the parameters of item $j$ can be computed via

$$\frac{\partial \ell(\boldsymbol{\vartheta}; \boldsymbol{x}_i)}{\partial \boldsymbol{\delta}_j} = \sum_{l=1}^{L} \Pr(\boldsymbol{\alpha}_l|\boldsymbol{x}_i) \cdot \left[ \frac{x_{ij} - P_j(\boldsymbol{\alpha}^*_{lj})}{P_j(\boldsymbol{\alpha}^*_{lj})(1 - P_j(\boldsymbol{\alpha}^*_{lj}))} \right] \cdot \frac{\partial P_j(\boldsymbol{\alpha}^*_{lj})}{\partial \boldsymbol{\delta}_j}. \tag{5}$$

To estimate the score contributions, we plug-in the estimated parameters $\widehat{\boldsymbol{\delta}}_j$ to get $P_j(\boldsymbol{\alpha}^*_{lj})$ and use $\Pr(\boldsymbol{\alpha}_l|\boldsymbol{x}_i)$ that is also available from the estimation procedure. The last term in Equation (5) depends on the type of CDM that is used. It is also possible to compute the score contributions directly for the conditional response probabilities. In this case, the last term in Equation (5) needs to be derived with respect to the conditional response probability of interest.

For the score contributions of the latent class probabilities, the constraint $\pi_L = 1 - \sum_{l=1}^{L-1} \pi_l$ must be taken into account, and thus,

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\vartheta}; \boldsymbol{x}_i)}{\partial \pi_l} &= \frac{1}{L(\boldsymbol{\vartheta}; \boldsymbol{x}_i)} \frac{\partial}{\partial \pi_l} \left( \sum_{l=1}^{L-1} \pi_l \cdot \Pr(\boldsymbol{x}_i|\boldsymbol{\alpha}_l) + \pi_L \cdot \Pr(\boldsymbol{x}_i|\boldsymbol{\alpha}_L) \right) \\
&= \frac{1}{L(\boldsymbol{\vartheta}; \boldsymbol{x}_i)} \frac{\partial}{\partial \pi_l} \left( \sum_{l=1}^{L-1} \pi_l \cdot \Pr(\boldsymbol{x}_i|\boldsymbol{\alpha}_l) + \left( 1 - \sum_{l=1}^{L-1} \pi_l \right) \cdot \Pr(\boldsymbol{x}_i|\boldsymbol{\alpha}_L) \right) \\
&= \frac{1}{L(\boldsymbol{\vartheta}; \boldsymbol{x}_i)} \frac{\partial}{\partial \pi_l} \left( \sum_{l=1}^{L-1} \pi_l \cdot \Big( \Pr(\boldsymbol{x}_i|\boldsymbol{\alpha}_l) - \Pr(\boldsymbol{x}_i|\boldsymbol{\alpha}_L) \Big) + \Pr(\boldsymbol{x}_i|\boldsymbol{\alpha}_L) \right) \\
&= \frac{1}{L(\boldsymbol{\vartheta}; \boldsymbol{x}_i)} \Big( \Pr(\boldsymbol{x}_i|\boldsymbol{\alpha}_l) - \Pr(\boldsymbol{x}_i|\boldsymbol{\alpha}_L) \Big).
\end{aligned}$$

Since the parameters in the last iteration of the EM algorithm are computed from the posterior values $\Pr(\boldsymbol{\alpha}_l|\boldsymbol{x}_i)$, it is more precise to also compute the score function for the latent class probabilities using the posterior values, via

$$\frac{\partial \ell(\boldsymbol{\vartheta}; \boldsymbol{x}_i)}{\partial \pi_l} = \frac{1}{\pi_l} \Big( \Pr(\boldsymbol{\alpha}_l|\boldsymbol{x}_i) - \Pr(\boldsymbol{\alpha}_L|\boldsymbol{x}_i) \Big).$$

### *Nonidentifiability of latent classes*

In the theory about standard errors of parameters that is presented above, it is assumed that the inverse of the complete information matrix $\mathcal{I}_{\boldsymbol{\vartheta}}$ exists. This, however, is not always the case in practical applications due to different causes. The most common cause has previously been discussed in Haertel (1989) as the nonidentifiability of latent classes. The problem arises whenever a test does not involve a single-attribute item for each of the $K$ attributes (see Chiu, Douglas, and Li 2009, for a theoretical discussion of the completeness of a $Q$-matrix in the DINA model, and Chiu and Köhn 2015, for CDMs in general). The G-DINA model can still be estimated, but some of the latent classes are not identified and the estimates of the corresponding latent class probabilities are equivalent. Moreover, when computing the covariance matrix using the complete information matrix, the corresponding columns and rows in the information matrix are alike (i.e., they are linearly dependent). Thus, the information matrix is nonsingular and cannot be inverted.

To avoid problems of identification in practice, it is therefore recommended that, whenever possible a single-attribute item is included for each of the $K$ attributes when developing new tests for cognitive diagnostic assessment. For researchers who perform a cognitive diagnostic analysis of data from an existing assessment (so-called retrofitting), the inversion problem can be circumvented by pooling latent classes that cannot be separated from each other.

# 3. Illustrations

Following the theoretical derivation of the underestimation of the standard errors – resulting from the inversion of the incomplete or the item-wise information matrix – the goal of this section is to illustrate the extent of this underestimation, and its effect on confidence intervals for the parameter estimates. In addition, we show for an exemplary real data set how much the standard errors may be underestimated in practice when the wrong methods are used. For both illustrations, the OPG estimator was used to estimate the covariance matrix of the model parameter estimates.

## 3.1. Coverage study

In the first study, we compare the quality of the standard error estimates based on the complete, the incomplete, and the item-wise information matrix, by estimating the coverage probability of the true parameter in a Wald-type confidence interval that uses a normal approximation given by $\left[\widehat{\vartheta} \pm z_{\frac{\alpha}{2}} \cdot \widehat{\text{se}}(\widehat{\vartheta})\right]$, and by computing the empirical bias of the standard errors.

Four different sample sizes ($N = 500, 1000, 2000, 5000$) were investigated using the $Q$-matrix given in Table 1. The $Q$-matrix included five attributes and was constructed such that each attribute was measured equally often (equal row sums in the table) and that the number of items that required the same number of attributes was equally distributed (i.e., five single-attribute items, five two-attribute items, and five three-attribute items). Thus, the $Q$-matrix represented a test with $J = 15$ items.

The DINA model and the A-CDM were used to generate response data. For each item, the true value of the baseline parameter ($\delta_{j0}$) was set to 0.2. In case of the DINA model, the true value of the interaction parameter between all attributes required by the item ($\delta_{j12\ldots K_j^*}$) was set to 0.6. Therefore, the guessing and the slip probabilities were both equal to 0.2. In case of the A-CDM, the main effect parameters were set to $\delta_{jk} = 0.6/K_j^*$. Thus, with each additionally mastered attribute, the conditional response probability increased by the same amount. The $K$ attributes for each individual were sampled independently from a Bernoulli

Table 1: Transposed $Q$-matrix used in the simulation study.

| | Items | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Attributes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | $\sum_k$ |
| $\alpha_1$ | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 6 |
| $\alpha_2$ | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 6 |
| $\alpha_3$ | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 6 |
| $\alpha_4$ | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 6 |
| $\alpha_5$ | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 6 |
| $\sum_j$ | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | |

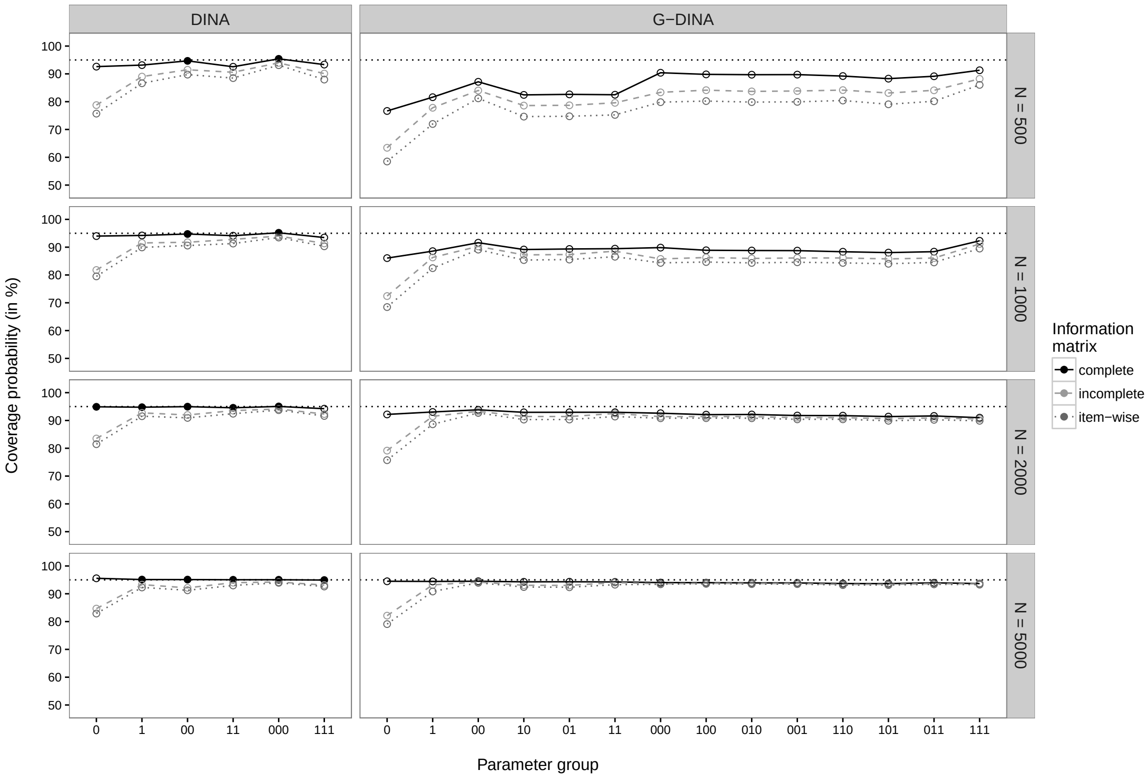


Figure 1: Coverage probabilities of 95% Wald-type confidence intervals for data generated under the DINA model are illustrated (on the $y$-axis) separately for parameters of items that require the same number of attributes (= parameter groups on the $x$-axis) using three different calculation methods for the standard errors. For ease of readability, values sufficiently close to the nominal coverage probability are depicted as solid circles, all others as empty circles.

distribution with probability $\Pr(\alpha_k = 1) = 0.5$, for all $k = 1 \dots K$. The joint distribution of the attributes (i.e., the latent class distribution) is then given by a categorical distribution with equal probabilities $\pi_l = \Pr(\alpha_l) = 1/(2^K)$. Responses that were simulated under the DINA model were analyzed using the DINA and the G-DINA model using the identity link. Note, that the G-DINA is also correct for data that were generated under the DINA model. It was fitted in addition to the DINA model because in practice the true model is usually unknown. However, in this situation the G-DINA model is overspecified, due to the many additional parameters estimated, for which the true values are zero according to the data generating model. Responses that were simulated under the A-CDM were accordingly analyzed using the A-CDM and the G-DINA model using the identity link. Again, the G-DINA is also correct – yet overspecified – for data generated under the A-CDM. To estimate the models and the standard errors, the EM algorithm was implemented in R (R Core Team 2016) based on the description in de la Torre (2009), but including our new suggestions on how the standard errors should be estimated.

Figures 1 and 2 illustrate the coverage probabilities for the data generated under the DINA model and the A-CDM, respectively. For all sample sizes and models, the coverage probabilities were computed for the $\boldsymbol{\delta}$ parameters using standard errors based on the complete

Table 2: Coverage probabilities of 95% Wald-type confidence intervals and average estimated bias of the standard errors for data generated under the DINA model and fitted to the DINA model.

| | | Coverage probabilities | | | Average estimated bias | | |
|---|---|---|---|---|---|---|---|
| $N$ | $\eta$ | Complete | Incomplete | Item-wise | Complete | Incomplete | Item-wise |
| 500 | 0 | 0.9261 | 0.7877 | 0.7571 | 0.0061 | 0.0248 | 0.0275 |
| | 1 | 0.9315 | 0.8902 | 0.8662 | 0.0056 | 0.0147 | 0.0191 |
| | 00 | 0.9468 | 0.9148 | 0.8974 | 0.0004 | 0.0037 | 0.0050 |
| | 11 | 0.9256 | 0.9057 | 0.8849 | 0.0044 | 0.0094 | 0.0136 |
| | 000 | 0.9541 | 0.9397 | 0.9312 | −0.0006 | 0.0007 | 0.0014 |
| | 111 | 0.9334 | 0.9010 | 0.8796 | 0.0036 | 0.0127 | 0.0173 |
| 5000 | 0 | 0.9556 | 0.8467 | 0.8289 | −0.0005 | 0.0051 | 0.0057 |
| | 1 | 0.9511 | 0.9328 | 0.9228 | −0.0002 | 0.0016 | 0.0024 |
| | 00 | 0.9511 | 0.9213 | 0.9126 | 0.0000 | 0.0009 | 0.0011 |
| | 11 | 0.9504 | 0.9399 | 0.9299 | 0.0000 | 0.0008 | 0.0016 |
| | 000 | 0.9504 | 0.9423 | 0.9394 | 0.0000 | 0.0002 | 0.0003 |
| | 111 | 0.9494 | 0.9313 | 0.9262 | 0.0000 | 0.0019 | 0.0024 |

information matrix $\mathcal{J}_{\boldsymbol{\vartheta}}$ (correct approach), and the incomplete information matrix $\mathcal{J}_{\boldsymbol{\delta}}$ and the item-wise information matrix $\mathcal{J}_{\boldsymbol{\delta}_j}$ (incorrect approaches). It turned out that the asymptotically expected standard errors of the item parameters are identical across items that require the same number of attributes. In the DINA model, for example, the baseline (guessing) probabilities of all single-attribute items share the same asymptotic standard error, no matter which of the attributes is required. This also holds for other item parameters, items that require more attributes and different models. Therefore, the coverage probabilities were averaged over the parameters within those groups, which are illustrated on the $x$-axis of the graph. The parameter group "0", for example, represents the baseline probability of all single-attribute items. The parameter group "111" represents the parameter of the three-way interaction of all three-attribute items.

By definition, the coverage probability of a 95% confidence interval has an expected nominal coverage rate of 95%. However, due to sampling error, the estimated coverage probabilities may randomly deviate from this nominal value. To achieve a high precision of the estimated coverage probabilities, each configuration was repeated 10,000 times. Assuming an exact binomial distribution for the coverage probabilities, the sampling error was equal to $\sqrt{\frac{0.95 \cdot 0.05}{10{,}000}} \approx 0.002$. Thus, based on a Wald-type confidence interval, we would consider coverage probabilities within $\left[94.6\%, 95.4\%\right]$ as sufficiently close to the nominal rate. Numbers within this interval are depicted with solid circles (otherwise empty circles) in Figures 1 and 2.

Additionally, Tables 2 to 5 list the exact values of the coverage probabilities and the empirical bias of the standard errors for the smallest ($N = 500$) and the largest ($N = 5000$) sample sizes (the intemediate sample sizes were omitted for brevity, but can be requested from the corresponding author) and each parameter group (labeled by $\eta$). The average empirical bias corresponds to the average of the empirical biases over all replications, that were computed by subtracting the estimated standard errors $\widehat{\text{se}}(\widehat{\vartheta}_r)$ from the empirical standard error of the estimated parameter values over all replications.

Figure 1 shows the coverage probabilities for the data generated under the DINA model. When the DINA model was used to analyze the data (see left column in Figure 1 and exact values reported in Table 2), the coverage probabilities for the standard errors based on the complete information matrix (solid line) were reasonably close to the expected coverage rate for small

Table 3: Coverage probabilities of 95% Wald-type confidence intervals and average estimated bias of the standard errors for data generated under the DINA model and fitted to the overspecified G-DINA model.

| | | Coverage probabilities | | | Average estimated bias | | |
|---|---|---|---|---|---|---|---|
| $N$ | $\eta$ | Complete | Incomplete | Item-wise | Complete | Incomplete | Item-wise |
| 500 | 0 | 0.7666 | 0.6343 | 0.5850 | 0.0304 | 0.0440 | 0.0482 |
| | 1 | 0.8162 | 0.7782 | 0.7198 | 0.0346 | 0.0403 | 0.0483 |
| | 00 | 0.8715 | 0.8403 | 0.8121 | 0.0194 | 0.0231 | 0.0269 |
| | 10 | 0.8244 | 0.7861 | 0.7465 | 0.0441 | 0.0512 | 0.0583 |
| | 01 | 0.8265 | 0.7871 | 0.7477 | 0.0444 | 0.0515 | 0.0587 |
| | 11 | 0.8250 | 0.7960 | 0.7527 | 0.0622 | 0.0702 | 0.0824 |
| | 000 | 0.9040 | 0.8335 | 0.7982 | 0.0443 | 0.0643 | 0.0712 |
| | 100 | 0.8981 | 0.8413 | 0.8022 | 0.0505 | 0.0790 | 0.0903 |
| | 010 | 0.8969 | 0.8368 | 0.7984 | 0.0534 | 0.0854 | 0.0973 |
| | 001 | 0.8973 | 0.8383 | 0.7993 | 0.0508 | 0.0801 | 0.0917 |
| | 110 | 0.8920 | 0.8415 | 0.8041 | 0.0548 | 0.0980 | 0.1159 |
| | 101 | 0.8829 | 0.8312 | 0.7906 | 0.0643 | 0.1069 | 0.1252 |
| | 011 | 0.8913 | 0.8407 | 0.8014 | 0.0549 | 0.0994 | 0.1175 |
| | 111 | 0.9130 | 0.8820 | 0.8603 | 0.0411 | 0.0999 | 0.1270 |
| 5000 | 0 | 0.9449 | 0.8213 | 0.7910 | 0.0004 | 0.0063 | 0.0073 |
| | 1 | 0.9444 | 0.9319 | 0.9082 | 0.0006 | 0.0018 | 0.0036 |
| | 00 | 0.9451 | 0.9423 | 0.9394 | 0.0004 | 0.0005 | 0.0007 |
| | 10 | 0.9429 | 0.9304 | 0.9243 | 0.0009 | 0.0024 | 0.0030 |
| | 01 | 0.9432 | 0.9301 | 0.9235 | 0.0010 | 0.0025 | 0.0031 |
| | 11 | 0.9426 | 0.9402 | 0.9327 | 0.0014 | 0.0019 | 0.0030 |
| | 000 | 0.9402 | 0.9361 | 0.9342 | 0.0017 | 0.0019 | 0.0021 |
| | 100 | 0.9399 | 0.9377 | 0.9357 | 0.0027 | 0.0029 | 0.0033 |
| | 010 | 0.9391 | 0.9369 | 0.9348 | 0.0028 | 0.0030 | 0.0034 |
| | 001 | 0.9390 | 0.9370 | 0.9348 | 0.0028 | 0.0030 | 0.0034 |
| | 110 | 0.9368 | 0.9337 | 0.9313 | 0.0044 | 0.0051 | 0.0057 |
| | 101 | 0.9365 | 0.9340 | 0.9314 | 0.0046 | 0.0053 | 0.0059 |
| | 011 | 0.9394 | 0.9363 | 0.9340 | 0.0039 | 0.0047 | 0.0053 |
| | 111 | 0.9366 | 0.9355 | 0.9329 | 0.0061 | 0.0066 | 0.0077 |

data samples, and converged quickly toward the nominal rate with increasing sample size $N$. The coverage rates for the standard errors based on the incomplete (dashed line) or the item-wise (dotted line) information matrix, however, were systematically smaller than the nominal coverage probability, particularly for the first parameter groups. Even for the largest sample size considered, their coverage probability does not converge towards the nominal rate. This is caused by the structural underestimation of the standard errors discussed earlier. We observed the largest underestimation for the baseline probabilities of single-attribute items (parameter group "0"). For the other parameters, the difference to the correct approach is smaller, but still lower than for the correct approach and notably below the nominal rate. A similar pattern can be observed when the G-DINA model was used to analyze the data generated under the DINA model (see right column in Figure 1 and exact values reported in Table 3). However, for smaller sample sizes the coverage probabilities were generally estimated considerably below the nominal coverage rate of 95%. This artifact may be explained by several circumstances. First, the normal approximation underlying the Wald-type confidence intervals might fail, particularly for the baseline probabilities that are restricted between zero and one. Second, for smaller data sets and more complex models, the conditional response probabilities and the parameters used to specify the attribute distribution are often estimated on the boundary of the parameter space. As mentioned earlier, this causes numerical problems in the calculation of the information matrix. Finally, the ratio between the number of estimated parameters per observation is larger for more general models. Thus, inferior asymptotic convergence

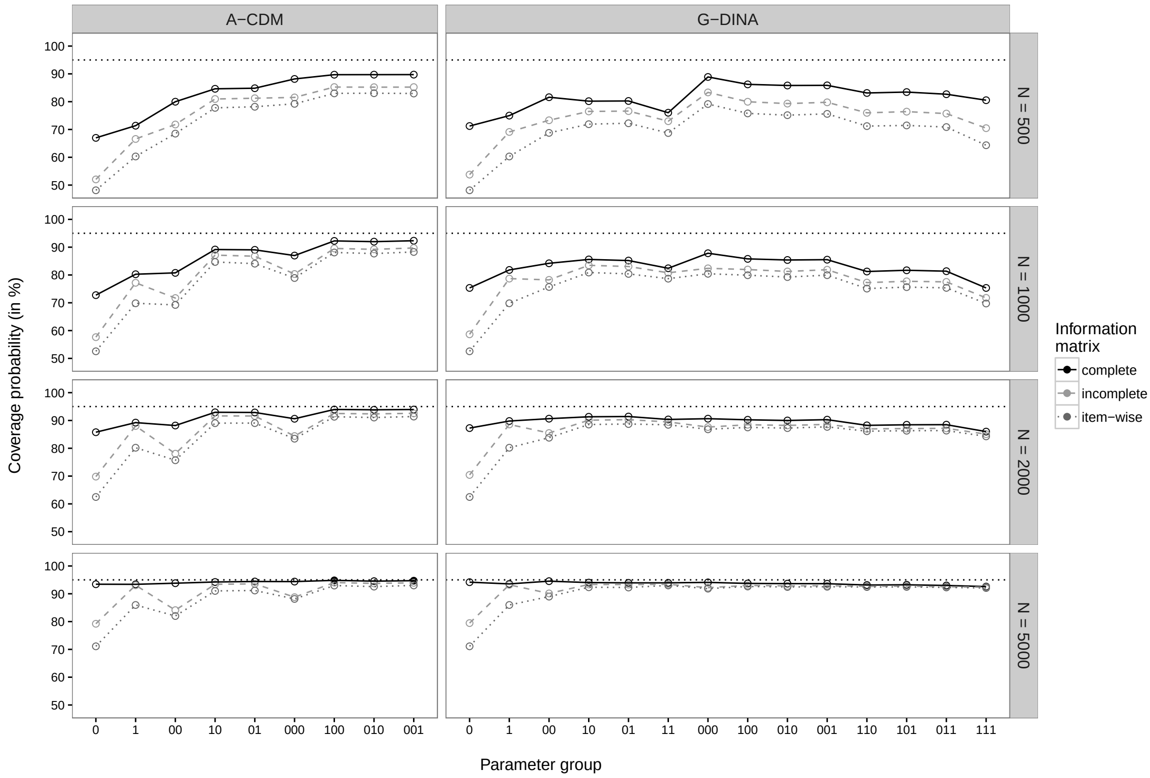


Figure 2: Coverage probabilities of 95% Wald-type confidence intervals for data generated under the A-CDM are illustrated (on the $y$-axis) separately for parameters of items that require the same number of attributes (= parameter groups on the $x$-axis) using three different calculation methods for the standard errors. For ease of readability, values sufficiently close to the nominal coverage probability are depicted as solid circles, all others as empty circles.

has to be reckoned with the G-DINA when compared to the DINA model. Nevertheless, the complete information matrix approach clearly provided more accurate results in all conditions considered.

Similar and related conclusions can be drawn from the average empirical biases reported in Tables 2 and 3. They were (in absolute terms) always smaller when the complete information matrix instead of the incomplete or the item-wise information matrix approaches were used. Please note, that when the correct DINA model was fitted to the simulated data (see values reported in Table 2), and when the standard errors were estimated with the complete information matrix approach, the bias almost completely vanished for the larger sample size, whereas with the two incorrect approaches still had significant biases at the larger sample size. This, however, was not the case when the overspecified G-DINA model was fitted to the data simulated under the DINA model (see values reported in Table 3). The average estimated biases reported for the complete information matrix approach did not coverage zero, although it was always smaller than for the incomplete and the item-wise approaches. Despite this finding, the complete information matrix approach provided the most accurate standard error estimates of all estimation approaches considered in this study.

Figure 2 shows the coverage probabilities for the data generated under the A-CDM (for exact

Table 4: Coverage probabilities of 95% Wald-type confidence intervals and average estimated bias of the standard errors for data generated under the A-CDM and fitted to the A-CDM.

| | | Coverage probabilities | | | Average estimated bias | | |
|---|---|---|---|---|---|---|---|
| $N$ | $\eta$ | Complete | Incomplete | Item-wise | Complete | Incomplete | Item-wise |
| 500 | 0 | 0.6698 | 0.5205 | 0.4813 | 0.0496 | 0.0633 | 0.0678 |
| | 1 | 0.7137 | 0.6659 | 0.6030 | 0.0687 | 0.0757 | 0.0843 |
| | 00 | 0.7999 | 0.7179 | 0.6852 | 0.0268 | 0.0352 | 0.0382 |
| | 10 | 0.8463 | 0.8096 | 0.7778 | 0.0252 | 0.0308 | 0.0350 |
| | 01 | 0.8484 | 0.8124 | 0.7817 | 0.0247 | 0.0302 | 0.0343 |
| | 000 | 0.8817 | 0.8153 | 0.7923 | 0.0139 | 0.0243 | 0.0270 |
| | 100 | 0.8971 | 0.8526 | 0.8299 | 0.0134 | 0.0217 | 0.0250 |
| | 010 | 0.8972 | 0.8521 | 0.8304 | 0.0142 | 0.0225 | 0.0258 |
| | 001 | 0.8973 | 0.8525 | 0.8294 | 0.0137 | 0.0220 | 0.0252 |
| 5000 | 0 | 0.9343 | 0.7922 | 0.7111 | 0.0016 | 0.0099 | 0.0127 |
| | 1 | 0.9340 | 0.9307 | 0.8597 | 0.0025 | 0.0029 | 0.0087 |
| | 00 | 0.9380 | 0.8411 | 0.8200 | 0.0009 | 0.0061 | 0.0069 |
| | 10 | 0.9424 | 0.9341 | 0.9102 | 0.0007 | 0.0014 | 0.0031 |
| | 01 | 0.9442 | 0.9357 | 0.9119 | 0.0006 | 0.0013 | 0.0029 |
| | 000 | 0.9439 | 0.8878 | 0.8815 | 0.0005 | 0.0041 | 0.0044 |
| | 100 | 0.9483 | 0.9403 | 0.9296 | 0.0003 | 0.0010 | 0.0017 |
| | 010 | 0.9453 | 0.9373 | 0.9258 | 0.0004 | 0.0011 | 0.0019 |
| | 001 | 0.9471 | 0.9394 | 0.9300 | 0.0002 | 0.0008 | 0.0016 |

Table 5: Coverage probabilities of 95% Wald-type confidence intervals and average estimated bias of the standard errors for data generated under the A-CDM and fitted to the overspecified G-DINA model.

| | | Coverage probabilities | | | Average estimated bias | | |
|---|---|---|---|---|---|---|---|
| $N$ | $\eta$ | Complete | Incomplete | Item-wise | Complete | Incomplete | Item-wise |
| 500 | 0 | 0.7126 | 0.5377 | 0.4815 | 0.0453 | 0.0614 | 0.0678 |
| | 1 | 0.7497 | 0.6909 | 0.6029 | 0.0639 | 0.0720 | 0.0843 |
| | 00 | 0.8160 | 0.7332 | 0.6877 | 0.0322 | 0.0412 | 0.0464 |
| | 10 | 0.8017 | 0.7649 | 0.7190 | 0.0679 | 0.0762 | 0.0853 |
| | 01 | 0.8027 | 0.7663 | 0.7225 | 0.0680 | 0.0762 | 0.0854 |
| | 11 | 0.7604 | 0.7300 | 0.6874 | 0.1240 | 0.1344 | 0.1476 |
| | 000 | 0.8890 | 0.8336 | 0.7918 | 0.0228 | 0.0405 | 0.0479 |
| | 100 | 0.8622 | 0.7997 | 0.7579 | 0.0612 | 0.0913 | 0.1048 |
| | 010 | 0.8583 | 0.7933 | 0.7516 | 0.0642 | 0.0957 | 0.1095 |
| | 001 | 0.8587 | 0.7978 | 0.7562 | 0.0628 | 0.0922 | 0.1056 |
| | 110 | 0.8312 | 0.7598 | 0.7120 | 0.1125 | 0.1642 | 0.1867 |
| | 101 | 0.8344 | 0.7640 | 0.7145 | 0.1136 | 0.1649 | 0.1873 |
| | 011 | 0.8269 | 0.7573 | 0.7087 | 0.1168 | 0.1676 | 0.1899 |
| | 111 | 0.8049 | 0.7044 | 0.6430 | 0.1607 | 0.2422 | 0.2772 |
| 5000 | 0 | 0.9417 | 0.7945 | 0.7111 | 0.0008 | 0.0098 | 0.0127 |
| | 1 | 0.9355 | 0.9322 | 0.8597 | 0.0023 | 0.0027 | 0.0087 |
| | 00 | 0.9453 | 0.9013 | 0.8898 | 0.0005 | 0.0040 | 0.0047 |
| | 10 | 0.9402 | 0.9335 | 0.9234 | 0.0015 | 0.0024 | 0.0037 |
| | 01 | 0.9393 | 0.9332 | 0.9226 | 0.0016 | 0.0025 | 0.0038 |
| | 11 | 0.9393 | 0.9340 | 0.9300 | 0.0029 | 0.0041 | 0.0048 |
| | 000 | 0.9410 | 0.9219 | 0.9184 | 0.0017 | 0.0039 | 0.0042 |
| | 100 | 0.9372 | 0.9299 | 0.9262 | 0.0038 | 0.0051 | 0.0058 |
| | 010 | 0.9365 | 0.9283 | 0.9244 | 0.0042 | 0.0056 | 0.0063 |
| | 001 | 0.9362 | 0.9290 | 0.9253 | 0.0040 | 0.0053 | 0.0060 |
| | 110 | 0.9315 | 0.9267 | 0.9243 | 0.0081 | 0.0095 | 0.0103 |
| | 101 | 0.9323 | 0.9275 | 0.9247 | 0.0081 | 0.0094 | 0.0102 |
| | 011 | 0.9299 | 0.9254 | 0.9228 | 0.0087 | 0.0101 | 0.0109 |
| | 111 | 0.9259 | 0.9234 | 0.9207 | 0.0166 | 0.0178 | 0.0189 |

values, see Tables 4 and 5). For the same reasons as discussed above, the coverage probabilities were estimated below the nominal rate for smaller samples. As the sample size increased, the coverage probabilities computed with the standard errors based on the complete information matrix again approached the nominal rate for the A-CDM and the G-DINA model. The coverage probabilities computed with the standard errors based on the incomplete or the item-wise information matrix, however, were again systematically underestimated. Overall, the complete information matrix approach again provided more accurate results across all conditions considered.

The average empirical biases reported in Tables 4 and 5 were again always smaller with the complete information approach for all parameter groups and sample sizes. However, as discussed above for the data simulated under the DINA model, the bias did not converge toward zero for the larger sample size, when the overspecified G-DINA model was used to estimate the data simulated under the A-CDM (see values reported in Table 5).

### 3.2. Empirical example

To illustrate the practical importance of estimating standard errors via the complete information matrix, data from a real assessment was analyzed using CDMs. The data stem from a learning experiment at the University of Tuebingen in Germany and is available in the R package **pks** (Heller and Wickelmaier 2013). The participants were required to answer 12 items about elementary probability theory. For example, "A box contains 30 marbles in the following colors: 8 red, 10 black, 12 yellow. What is the probability that a randomly drawn marble is yellow?". Four different attributes (concepts) were tested: How to calculate

- the classic probability of an event (pb),
- the probability of the complement of an event (cp),
- the probability of the union of two disjoint events (un),
- the probability of two independent events (id).

These concepts were combined to form the 12 items. Therefore, the $Q$-matrix (see Table 6) was defined by the design of the items. The first four items required only one attribute, the items 5 to 10 required two attributes and the items 11 and 12 required three attributes. For this illustration, the responses of 504 participants from the first part of the experiment were analyzed.

Table 6: Transposed $Q$-matrix used for analyzing the elementary probability theory data.

| | Items | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Attributes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | $\sum_k$ |
| pb | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 8 |
| cp | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 5 |
| un | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 4 |
| id | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 5 |
| $\sum_j$ | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | |

Table 7: Estimates and standard errors of parameters for the elementary probability theory data. Numbers in brackets correspond to the relative change to the standard errors based on the complete information matrix.

| | | | Standard errors | | | | |
|---|---|---|---|---|---|---|---|
| Item | Attribute | Estimate | Complete | Incomplete | | Item-wise | |
| 1 | baseline | 0.224 | 0.065 | 0.061 | (−0.071) | 0.052 | (**−0.203**) |
| | pb | 0.710 | 0.067 | 0.063 | (−0.061) | 0.055 | (**−0.186**) |
| 2 | baseline | 0.275 | 0.105 | 0.080 | (**−0.241**) | 0.068 | (**−0.356**) |
| | cp | 0.699 | 0.105 | 0.081 | (**−0.232**) | 0.069 | (**−0.346**) |
| 3 | baseline | 0.097 | 0.060 | 0.055 | (−0.082) | 0.048 | (**−0.194**) |
| | un | 0.864 | 0.061 | 0.056 | (−0.082) | 0.050 | (**−0.188**) |
| 4 | baseline | 0.125 | 0.038 | 0.035 | (−0.072) | 0.032 | (**−0.159**) |
| | id | 0.837 | 0.039 | 0.037 | (−0.064) | 0.034 | (**−0.140**) |
| 5 | baseline | 0.201 | 0.067 | 0.055 | (**−0.187**) | 0.048 | (**−0.288**) |
| | pb | 0.364 | 0.116 | 0.101 | (**−0.130**) | 0.094 | (**−0.191**) |
| | cp | 0.293 | 0.125 | 0.111 | (**−0.116**) | 0.103 | (**−0.181**) |
| 6 | baseline | 0.194 | 0.062 | 0.058 | (−0.058) | 0.051 | (**−0.185**) |
| | pb | 0.462 | 0.085 | 0.080 | (−0.053) | 0.074 | (**−0.125**) |
| | cp | 0.308 | 0.083 | 0.081 | (−0.021) | 0.077 | (−0.071) |
| 7 | baseline | 0.278 | 0.071 | 0.068 | (−0.049) | 0.062 | (**−0.126**) |
| | pb | 0.292 | 0.095 | 0.088 | (−0.078) | 0.083 | (**−0.127**) |
| | un | 0.372 | 0.116 | 0.105 | (−0.094) | 0.097 | (**−0.164**) |
| 8 | baseline | 0.430 | 0.087 | 0.076 | (**−0.132**) | 0.063 | (**−0.277**) |
| | pb | 0.065 | 0.095 | 0.066 | (**−0.297**) | 0.059 | (**−0.371**) |
| | un | 0.462 | 0.111 | 0.088 | (**−0.212**) | 0.079 | (**−0.293**) |
| 9 | baseline | 0.116 | 0.045 | 0.043 | (−0.042) | 0.038 | (**−0.145**) |
| | pb | 0.510 | 0.084 | 0.079 | (−0.060) | 0.074 | (**−0.113**) |
| | id | 0.154 | 0.075 | 0.070 | (−0.065) | 0.065 | (**−0.124**) |
| 10 | baseline | 0.083 | 0.050 | 0.044 | (**−0.115**) | 0.037 | (**−0.248**) |
| | cp | −0.056 | 0.060 | 0.055 | (−0.086) | 0.048 | (**−0.190**) |
| | id | 0.781 | 0.036 | 0.035 | (−0.027) | 0.034 | (−0.062) |
| 11 | baseline | 0.053 | 0.049 | 0.045 | (−0.086) | 0.038 | (**−0.229**) |
| | pb | 0.010 | 0.106 | 0.086 | (**−0.184**) | 0.080 | (**−0.244**) |
| | cp | −0.037 | 0.094 | 0.084 | (**−0.109**) | 0.078 | (**−0.173**) |
| | id | 0.672 | 0.034 | 0.033 | (−0.030) | 0.032 | (−0.060) |
| 12 | baseline | 0.000 | 0.039 | 0.036 | (−0.090) | 0.029 | (**−0.269**) |
| | pb | 0.140 | 0.469 | 0.191 | (**−0.592**) | 0.169 | (**−0.640**) |
| | un | 0.000 | 0.452 | 0.181 | (**−0.600**) | 0.162 | (**−0.643**) |
| | id | 0.660 | 0.046 | 0.042 | (−0.067) | 0.042 | (−0.084) |

*Note.* Strongest relative changes are printed in bold letters for better readability.

The data was fitted using the DINA, the A-CDM and the G-DINA model with the resulting BIC values of 5200.46 ($df = 39$), 5154.58 ($df = 49$) and 5241.70 ($df = 63$), respectively. The results of the A-CDM – which had the lowest BIC value – are illustrated in Table 7. The table summarizes the estimated parameters, the corresponding standard errors based on the complete, the incomplete and the item-wise information matrix, and the relative change in the standard errors between the correct and the two incorrect approaches (in parentheses).

For each item, the first parameter estimate represents the baseline probability (i.e., the probability of correctly answering the item when the attributes required by the item have not been mastered). Thus, large values for this guessing probability are unusual. For item 8, however, a value of over 0.4 is reported. A possible explanation is that the item – "What is the probability of obtaining an odd number when throwing a dice?" – was not very difficult, even for individuals without knowledge in basic probability theory. Further parameter estimates represent the amount of increase (or seldom decrease) in probability of answering an item correctly when the corresponding attribute had been mastered. For example, the probability of answering item 1 increased by 0.71 when attribute “pb” had been mastered.

The relative change between the standard errors based on the complete and the incomplete information matrix showed substantial differences (highlighted by bold letters in Table 7) for both parameters of the single-attribute item 2, for some of the parameters of the two-attribute items 5, 8 and 10, and for some of the parameters of the three-attribute items 11 and 12. The underestimation of the standard errors based on the item-wise information matrix was even worse. For 30 out of 34 item parameters the standard error was underestimated.

It should be noted that ten out of 48 conditional response probabilities and four out of 16 parameters of the latent class probabilities were estimated at the boundary of the parameter space (not displayed in Table 7). As mentioned earlier, this can cause numerical problems in computing the information matrix. According to the previous simulation study, where a similar scenario was investigated (see top-left panel in Figure 2 for the same model and a nearly equal sample size), it must be assumed that some of the standard errors reported for this data are generally underestimated. Nevertheless, just like in the simulation study – and as expected from our theoretical considerations – the additional severe underestimation caused by the wrong computation of the information matrix can easily be avoided by using the complete information matrix.

# 4. Discussion

Standard errors are an important measure to quantify the uncertainty of an estimate. They are required for many different statistical techniques to evaluate model fit or to check model assumptions. It is therefore crucial in practical research to estimate standard errors as precisely as possible. In the commonly used approach for computing standard errors in CDMs, however, the information matrix is based only on those parameters which are used to specify the item response function. The parameters used to specify the joint distribution of the attributes (i.e., latent class distribution) are not incorporated in the computation.

In this article, we have shown that with this approach, the standard errors for the parameters of the item response function are systematically underestimated. We therefore strongly recommend to compute the standard errors based on the complete information matrix, which also includes the parameters used to specify the latent class distribution. In addition to the clear theoretical result, we have also illustrated by means of simulations that our approach leads to a higher quality of Wald-type confidence intervals and lower empirical bias. An additional benefit of using the complete instead of the incomplete information matrix is that it also provides the information required to compute standard errors for the parameters used to specify the latent class distribution.

We assume that the incomplete information matrix approaches have only become widely used in the CDM literature because previous authors might have assumed that the off-diagonal elements of the information matrix would have negligible impact under certain conditions. With respect to the item-wise computation of the standard errors, the CDM literature may be partially influenced by the traditional IRT literature, where approaches exist that lead to block diagonal information matrices (e.g., in Thissen and Wainer 1982), in which case an item-wise computation of the standard errors is possible. However for CDMs, as we showed analytically and illustrated with examples, the complete information matrix approach clearly generates better standard errors than the incomplete and the item-wise approaches and is computationally well feasible. Similar to our results, Yuan *et al.* (2014) showed that the item-wise computation of the standard errors in IRT models also leads to undersized standard errors.

In the simulation study, we did not specifically vary design factors, such as the $Q$-matrix, the true values of the item parameters, or the latent class distribution. Varying these factors might positively or negatively affect the severity of underestimation. In a preliminary study with the DINA model, we found that longer tests and highly discriminating items can alleviate the underestimation. It should be highlighted, however, that the proposed method for estimating the standard errors cannot make the quality of the standard errors worse. In practical situations, however, it is difficult (or even impossible) to control the factors that have a large impact on the underestimation. As such it is always preferable to compute standard errors using the complete information matrix.

We note that differences between the approaches are not only expected for the standard errors, but for the entire covariance matrix of the model parameters (although not generally in the same direction). Thus, many techniques used to investigate a fitted model may be affected. The impact of under- or overestimation of the entire covariance matrix will be multiplied for multivariate methods. It is therefore worth in any circumstances to estimate standard errors (and also the entire covariance matrix) from the complete information matrix. As we did not specifically investigate the impact of misestimating the entire covariance matrix on multivariate techniques, it will be interesting for future research to investigate how much the quality of the covariance matrix can be improved by using the complete information matrix in computing it.

The results of the simulation study revealed problems of asymptotic convergence when more complex models were fitted to smaller data sets. This might partially be caused by boundary problems that often occur for smaller data sets. DeCarlo (2011) suggested posterior mode (PM) estimation to overcome these problems. Whether PM estimation leads to more accurate parameter and standard error estimates than the traditional ML approach in CDMs was not the scope of this work, but something that can be investigated in future research. Moreover, the normal approximation of the ML estimates might be more accurate on the real line under the logit link rather than on the (bounded) probability scale under the identity link. However, this not only concerns the estimation of standard errors but of the model parameters in general. Therefore, this is beyond the scope of this manuscript and is not pursued here. It might be of interest for future research, though, to explore the potential benefits of different link functions. In general, the results from our simulation study suggest that it is recommended to use simpler models whenever possible and appropriate because it may avoid boundary problems or problems with asymptotic convergence.

Finally, in the present article, we assumed that the $Q$-matrix is known or well specified for an assessment. However, in practice (especially when retrofitting CDMs to existing data), the $Q$-matrix may be unknown or misspecified, which can affect parameter estimation and classification accuracy (de la Torre 2008; Rupp and Templin 2007). To minimize the impact of a misspecified $Q$-matrix, several methods have been proposed. De la Torre (2008) proposed an iterative procedure to evaluate the correctness of the $Q$-matrix specification in the context of the DINA model. The approach was extended by de la Torre and Chiu (2016) to apply generally to other CDMs. Other recent approaches include that of Chen, Liu, Xu, and Ying (2015), which estimates the $Q$-matrix of the DINA model using regularization, whereas Chiu

(2013) proposed a nonparametric approach to $Q$-matrix validation that does not require specifying the exact form of the CDM, only that the underlying process is conjunctive in nature. Future research should examine the extent of the impact of $Q$-matrix mispecifications on standard error estimation, and whether specific steps can be taken to minimize such an impact.

## Computational details

The estimation routines used in this study were written in the free and open-source software R (R Core Team 2016) for statistical computing. Functions to estimate the parameters and the standard errors in the G-DINA model are provided in the form of the add-on package **Rcdm**, available online at `https://github.com/mphili/cdm` under the terms of the GNU General Public License (Version 2 or 3).

## References

Banerjee S, Roy A (2014). *Linear Algebra and Matrix Analysis for Statistics*. CRC Press, Boca Raton.

Cai L (2008). "SEM of Another Flavour: Two New Applications of the Supplemented EM Algorithm." *British Journal of Mathematical and Statistical Psychology*, **61**(2), 309–329. doi:10.1348/000711007x249603.

Chen J, de la Torre J (2013). "A General Cognitive Diagnosis Model for Expert-Defined Polytomous Attributes." *Applied Psychological Measurement*, **37**(6), 419–437. doi:10.1177/0146621613479818.

Chen Y, Liu J, Xu G, Ying Z (2015). "Statistical Analysis of Q-Matrix Based Diagnostic Classification Models." *Journal of the American Statistical Association*, **110**(510), 850–866. doi:10.1080/01621459.2014.934827.

Chiu CY (2013). "Statistical Refinement of the Q-Matrix in Cognitive Diagnosis." *Applied Psychological Measurement*, **37**(8), 598–618. doi:10.1177/0146621613488436.

Chiu CY, Douglas JA, Li X (2009). "Cluster Analysis for Cognitive Diagnosis: Theory and Applications." *Psychometrika*, **74**(4), 633–665. doi:10.1007/s11336-009-9125-0.

Chiu CY, Köhn HF (2015). "A General Proof of Consistency of Heuristic Classification for Cognitive Diagnosis Models." *British Journal of Mathematical and Statistical Psychology*, **68**(3), 387–409. doi:10.1111/bmsp.12055.

Culpepper SA (2015). "Bayesian Estimation of the DINA Model with Gibbs Sampling." *Journal of Educational and Behavioral Statistics*, **40**(5), 454–476. doi:10.3102/1076998615595403.

de la Torre J (2008). "An Empirically Based Method of Q-Matrix Validation for the DINA Model: Development and Applications." *Journal of Educational Measurement*, **45**(4), 343–362. doi:10.1111/j.1745-3984.2008.00069.x.

de la Torre J (2009). "DINA Model and Parameter Estimation: A Didactic." *Journal of Educational and Behavioral Statistics*, **34**(1), 115–130. doi:10.3102/1076998607309474.

de la Torre J (2011). "The Generalized DINA Model Framework." *Psychometrika*, **76**(2), 179–199. doi:10.1007/s11336-011-9207-7.

de la Torre J, Chiu CY (2016). "A General Method of Empirical Q-Matrix Validation." *Psychometrika*, **81**, 253–273. doi:10.1007/s11336-015-9467-8.

de la Torre J, Douglas JA (2004). "Higher-Order Latent Trait Models for Cognitive Diagnosis." *Psychometrika*, **69**(3), 333–353. doi:10.1016/j.pse.2014.11.001.

de la Torre J, Lee YS (2013). "Evaluating the Wald Test for Item-Level Comparison of Saturated and Reduced Models in Cognitive Diagnosis." *Journal of Educational Measurement*, **50**(4), 355–373. doi:10.1111/jedm.12022.

de la Torre J, van der Ark LA, Rossi G (2015). "Analysis of Clinical Data from Cognitive Diagnosis Modeling Framework." *Measurement and Evaluation in Counseling and Development*. doi:10.1177/0748175615569110. Advance online publication.

DeCarlo LT (2011). "On the Analysis of Fraction Subtraction Data: The DINA Model, Classification, Latent Class Sizes, and the Q-Matrix." *Applied Psychological Measurement*, **35**(1), 8–26. doi:10.1177/0146621610377081.

Dempster AP, Laird NM, Rubin DB (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society B*, **39**(1), 1–38.

Garre FG, Vermunt JK (2006). "Avoiding Boundary Estimates in Latent Class Analysis by Bayesian Posterior Mode Estimation." *Behaviormetrika*, **33**(1), 43–59. doi:10.2333/bhmk.33.43.

George AC (2013). *Investigating CDMs: Blending Theory with Practicality*. Ph.D. thesis, TU Dortmund University, Dortmund, Germany. URL https://eldorado.tu-dortmund.de/.

Haertel EH (1989). "Using Restricted Latent Class Models to Map the Skill of Achievement Structure Items." *Journal of Educational Measurement*, **26**(4), 301–321. doi:10.1111/j.1745-3984.1989.tb00336.x.

Harville DA (2008). *Matrix Algebra from a Atatistician's Perspective*. Springer-Verlag, New York.

Heller J, Wickelmaier F (2013). "Minimum Discrepancy Estimation in Probabilistic Knowledge Structures." *Electronic Notes in Discrete Mathematics*, **42**, 49–56. doi:10.1016/j.endm.2013.05.145.

Henson RA, Templin JL, Willse JT (2009). "Defining a Family of Cognitive Diagnosis Models Using Log-Linear Models with Latent Variables." *Psychometrika*, **74**(2), 191–210. doi:10.1007/s11336-008-9089-5.

Hou L, de la Torre J, Nandakumar R (2014). "Differential Item Functioning Assessment in Cognitive Diagnostic Modeling: Application of the Wald Test to Investigate DIF in the DINA Model." *Journal of Educational Measurement*, **51**(1), 98–125. doi:10.1111/jedm.12036.

Junker BW, Sijtsma K (2001). "Cognitive Assessment Models with Few Assumptions, and Connections with Nonparametric Item Response Theory." *Applied Psychological Measurement*, **25**(3), 258–272. doi:10.1177/01466210122032064.

Lee YS, Park YS, Taylan D (2011). "A Cognitive Diagnostic Modeling of Attribute Mastery in Massachusetts, Minnesota, and the U.S. National Sample Using the TIMSS 2007." *International Journal of Testing*, **11**(2), 144–177. doi:10.1080/15305058.2010.534571.

Li H (2011). "A Cognitive Diagnostic Analysis of the MELAB Reading Test." *Spaan Fellow Working Papers in Second or Foreign Language Assessment*, **9**, 17–46.

Li X, Wang WC (2015). "Assessment of Differential Item Functioning under Cognitive Diagnosis Models: The DINA Model Example." *Journal of Educational Measurement*, **52**(1), 28–54. doi:10.1111/jedm.12061.

McLachlan G, Krishnan T (2007). *The EM Algorithm and Extensions*. 2nd edition. John Wiley & Sons, New York.

Neyman J, Scott EL (1948). "Consistent Estimates Based on Partially Consistent Observations." *Econometrica*, **16**(1), 1–32. doi:10.2307/1914288.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rojas G (2013). *Cognitive Diagnosis Models: Attribute Classification, Differential Item Functioning and Applications*. Ph.D. thesis, Universidad Autónomia de Madrid, Madrid, Spain. URL https://repositorio.uam.es/.

Rupp AA, Templin J (2007). "The Effects of Q-Matrix Misspecification on Parameter Estimates and Classification Accuracy in the DINA Model." *Educational and Psychological Measurement*, **68**(1), 78–96. doi:10.1177/0013164407301545.

Rupp AA, Templin J, Henson RA (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. Methodology in the Social Sciences. Guilford Press, New York.

Song L, Wang W, Dai H, Ding S (2012). "The Revised DINA Model Parameter Estimation with EM Algorithm." *International Journal of Digital Content Technology and Its Applications*, **6**(9), 85–92. doi:10.4156/jdcta.vol6.issue9.11.

Tatsuoka KK (1983). "Rule Space: An Approach for Dealing with Misconceptions Based in Item Response Theory." *Journal of Educational Measurement*, **20**(4), 345–354. doi:10.1111/j.1745-3984.1983.tb00212.x.

Templin JL, a Henson R (2006). "Measurement of Psychological Disorders Using Cognitive Diagnosis Models." *Psychological Methods*, **11**(3), 287–305. doi:10.1037/1082-989x.11.3.287.

Thissen D, Wainer H (1982). "Some Standard Errors in Item Response Theory." *Psychometrika*, **47**(4), 397–412. doi:10.1007/bf02293705.

von Davier M (2008). "A General Diagnostic Model Applied to Language Testing Data." *British Journal of Mathematical and Statistical Psychology*, **61**(2), 287–307. doi:10.1348/000711007x193957.

Woods CM, Cai L, Wang M (2012). "The Langer-Improved Wald Test for DIF Testing with Multiple Groups: Evaluation and Comparison to Two-Group IRT." *Educational and Psychological Measurement*, **73**(3), 532–547. doi:10.1177/0013164412464875.

Yang X, Embretson SE (2007). "Construct Validity and Cognitive Diagnostic Assessment." In JP Leighton, MJ Gierl (eds.), *Cognitive Diagnostic Assessment for Education: Theory and Applications*, pp. 119–145. Cambridge University Press, New York.

Yuan KH, Cheng Y, Patton J (2014). "Information Matrices and Standard Errors for MLEs of Item Parameters in IRT." *Psychometrika*, **79**(2), 232–254. doi:10.1007/s11336-013-9334-4.

# A. Blockwise matrix inversion

The following statements about blockwise matrix inversion of a symmetric matrix can be used to establish the inequality between standard errors based on the complete and the incomplete information matrix discussed in the section on the estimation of the standard errors. The corresponding theorems (and proofs) can be found in Chapter 13 of Banerjee and Roy (2014), if not stated otherwise.

Let $\boldsymbol{A}$ be a positive definite (*p.d.*) symmetric matrix, i.e. the inverse $\boldsymbol{A}^{-1}$ exists and is also *p.d.*. Suppose $\boldsymbol{A}$ is partitioned as

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{12}^\top & \boldsymbol{A}_{22} \end{pmatrix},$$

where $\boldsymbol{A}_{11}$ is $p \times p$, $\boldsymbol{A}_{12}$ is $p \times q$ and $\boldsymbol{A}_{22}$ is $q \times q$. Then its principal submatrices $\boldsymbol{A}_{11}$ and $\boldsymbol{A}_{22}$ are also invertible and *p.d.*. Let $\boldsymbol{B} = \boldsymbol{A}^{-1}$ be partitioned (similar to $\boldsymbol{A}$) as

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{12}^\top & \boldsymbol{B}_{22} \end{pmatrix},$$

where $\boldsymbol{B}_{11} = \left(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12}^\top\right)^{-1}$ and $\boldsymbol{B}_{22} = \left(\boldsymbol{A}_{22} - \boldsymbol{A}_{12}^\top\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\right)^{-1}$ are given by the inverse of the *Schur* complements of $\boldsymbol{A}_{22}$ and $\boldsymbol{A}_{11}$, respectively, which are also *p.d.*. By the *Sherman-Woodbury-Morrison* formula (see e.g., Banerjee and Roy 2014, p. 82),

$$\begin{aligned} \left(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12}^\top\right)^{-1} &= \boldsymbol{A}_{11}^{-1} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\left(\boldsymbol{A}_{22} - \boldsymbol{A}_{12}^\top\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\right)^{-1}\boldsymbol{A}_{12}^\top\boldsymbol{A}_{11}^{-1} \\ \boldsymbol{B}_{11} &= \boldsymbol{A}_{11}^{-1} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{B}_{22}\boldsymbol{A}_{12}^\top\boldsymbol{A}_{11}^{-1} \\ \boldsymbol{B}_{11} &= \boldsymbol{A}_{11}^{-1} + \boldsymbol{C}^\top\boldsymbol{B}_{22}\boldsymbol{C}. \end{aligned}$$

where $\boldsymbol{C} = \boldsymbol{A}_{12}^\top\boldsymbol{A}_{11}^{-1} = (\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^\top$. For the diagonal elements, we have

$$\mathrm{diag}(\mathrm{B}_{11}) = \mathrm{diag}(\mathrm{A}_{11}^{-1}) + \mathrm{diag}(\mathrm{C}^\top\mathrm{B}_{22}\mathrm{C}),$$

where $\boldsymbol{B}_{11}$ and $\boldsymbol{A}_{11}^{-1}$ are both positive definite, i.e., their diagonal elements are positive.

**Lemma 1.** *If $\boldsymbol{B}_{22}$ and $\boldsymbol{A}_{11}^{-1}$ are positive definite and $\boldsymbol{A}_{12} \neq \boldsymbol{0}$, then each diagonal element of $\boldsymbol{C}^\top\boldsymbol{B}_{22}\boldsymbol{C}$ is positive.*

*Proof.* Since $\boldsymbol{B}_{22}$ is positive definite, $\boldsymbol{x}^\top\boldsymbol{B}_{22}\boldsymbol{x} > 0$ whenever $\boldsymbol{x} \neq 0$. Choosing $\boldsymbol{x} = \boldsymbol{C}\boldsymbol{e}_i$ reveals that

$$\boldsymbol{x}^\top\boldsymbol{B}_{22}\boldsymbol{x} = \boldsymbol{e}_i^\top\boldsymbol{C}^\top\boldsymbol{B}_{22}\boldsymbol{C}\boldsymbol{e}_i > 0,$$

where $\boldsymbol{e}_i$ is the $i$th unit vector that is used to extract the $i$th diagonal element from $\boldsymbol{C}^\top\boldsymbol{B}_{22}\boldsymbol{C}$. Hence, the diagonal elements in $\boldsymbol{C}^\top\boldsymbol{B}_{22}\boldsymbol{C}$ are also positive. □

So, if $\boldsymbol{A}_{12} \neq \boldsymbol{0}$, all diagonal elements in $\boldsymbol{C}^\top\boldsymbol{B}_{22}\boldsymbol{C}$ are positive and therefore,

$$\mathrm{diag}(\boldsymbol{B}_{11})_r > \mathrm{diag}(\boldsymbol{A}_{11}^{-1})_r \qquad \forall\, r \in \{1, \ldots, p\}.$$

To obtain the inequality of the standard errors as stated in the section on the estimation of the standard errors, use $\boldsymbol{A} = \mathcal{I}_{\boldsymbol{\vartheta}}$ and $\boldsymbol{B} = V_{\boldsymbol{\vartheta}}$ and let $\mathcal{I}_{\boldsymbol{\beta},\boldsymbol{\pi}} \neq \boldsymbol{0}$.

Please note, that the symmetric information matrix $\mathcal{I}_{\boldsymbol{\vartheta}}$ is only positive semidefinite. A positive semidefinite symmetric matrix is, however, positive definite if and only if it is nonsingular (see e.g., Harville 2008, Corollary 14.3.12). Thus, the inequality holds if $\mathcal{I}_{\boldsymbol{\vartheta}}$ is invertible, which is required anyway to compute the standard errors.

**Affiliation:**

Michel Philipp, Carolin Strobl
Department of Psychology
Universität Zürich
Binzmühlestr. 14
8050 Zürich, Switzerland
E-mail: Michel.Philipp.mp@gmail.com, Carolin.Strobl@uzh.ch
URL: http://www.psychologie.uzh.ch/methoden.html

Jimmy de la Torre
Faculty of Education
The University of Hong Kong
E-mail: j.delatorre@hku.hk
URL: http://web.edu.hku.hk/staff/academic/j.delatorre

Achim Zeileis
Department of Statistics
Faculty of Economics and Statistics
Universität Innsbruck
Universitätsstr. 15
6020 Innsbruck, Austria
E-mail: Achim.Zeileis@R-project.org
URL: https://eeecon.uibk.ac.at/~zeileis/