ASYMPTOTIC POSTERIOR NORMALITY OF MULTIVARIATE LATENT TRAITS IN AN IRT MODEL

MIA J. K. KORNELY AND MARIA KATERI

RWTH AACHEN UNIVERSITY

The asymptotic posterior normality (APN) of the latent variable vector in an item response theory (IRT) model is a crucial argument in IRT modeling approaches. In case of a single latent trait and under general assumptions, Chang and Stout (Psychometrika, 58(1):37–52, 1993) proved the APN for a broad class of latent trait models for binary items. Under the same setup, they also showed the consistency of the latent trait's maximum likelihood estimator (MLE). Since then, several modeling approaches have been developed that consider multivariate latent traits and assume their APN, a conjecture which has not been proved so far. We fill this theoretical gap by extending the results of Chang and Stout for multivariate latent traits. Further, we discuss the existence and consistency of MLEs, maximum a-posteriori and expected a-posteriori estimators for the latent traits under the same broad class of latent trait models.

Key words: multidimensional item response theory, empirical Bayes, posterior distribution, ability estimation, consistency, normal approximation, Bernstein–von Mises theorem .

1. Introduction

In the context of item response theory (IRT) methodology, statistical inference for the examinee's ability relies often on the assumption that its posterior distribution given the test response is a normal distribution. As this is usually hard to justify and in contradiction to common models of the examinees abilities distribution in the population, it can assumed to be, for a long test, well approximated by a normal distribution. This assumption of asymptotic posterior normality (APN) is part of the famous Dutch identity conjecture of Holland (1990), who mentioned then that he was not aware of a thorough discussion of APN of latent variables and this would be an interesting area for future research. Shortly after, Chang and Stout (1993) proved the APN for univariate latent traits (LTs), mentioning that APN for multivariate LTs can be proved, but without providing further details or discussing the associated regularity conditions required.

As far as we know, APN of multivariate LTs has not been proved so far for IRT models of a general context, although posterior normality or APN is assumed quite often under various IRT setups (e.g., Anderson & Vermunt, 2000; Anderson & Yu, 2017; Anderson et al., 2007; Hessen, 2012; Li, 2010; Paek, 2016). For example, Pelle et al. (2016) assume posterior multivariate normality for the latent variable vector of a log-linear multidimensional Rasch model for capture–recapture analysis of registration data.

Sometimes the APN-assumption is justified by the APN in a Bayesian framework (pointing to the "Bernstein–von Mises Theorem") without however proceeding to further details (e.g., the computationally efficient adaptive quadrature methods for high-dimensional item factor analysis

(Schilling & Bock, 2005) and for generalized linear mixed models (Rabe-Hesketh et al., 2002) are based on the APN assumption).

In this work, we study the APN for multivariate latent trait models, focusing on models for dichotomous items and targeting at conditions that are tailored to IRT models and thus simpler to verify. APN of LTs, univariate or multivariate, is related to Bayesian asymptotics. In the light of this connection, we deepen in the approach of Ghosal et al. (1995), who discussed asymptotic posterior distributions in a very general setup that includes the regular cases and some non-regular cases as well. They also proved a general result on the asymptotic equivalence of the Bayes and maximum likelihood estimators, a well-known result for the regular cases. In particular, we generalize the approach and results of Chang and Stout (1993), CS hereafter, linking them to the semiproper centering concept of Ghosal et al. (1995), GGS hereafter, and embedding them in their approach. We provide conditions for multivariate APN that correspond one to one to the conditions of CS for univariate LTs, which is the standard approach for IRT models, as alternatives to the conditions imposed in Ghosal et al. (1995). Even for the case of univariate LTs, the proposed approach could be an interesting alternative to that of CS, since it has the advantage of applying also to models with non-monotone item response functions, which is not the case in the CS setup. Furthermore, we discuss conditions under which the existence of the maximum likelihood estimators (MLEs) for latent variable vectors is ensured. The consistency of MLEs under mild conditions, which was indicated as an open issue by Sinharay (2015), follows as a natural consequence of the proof of the APN. Finally, we prove the consistency of maximum a-posteriori (MAP) and expected a-posteriori (EAP) estimators for multivariate LTs.

The paper is organized as follows. Basic notation and the adopted IRT framework is set in Sect. 2, while the CS-theory for a univariate LT is briefly reviewed in Sect. 3. The approach of Ghosal et al. (1995) is discussed and linked to the APN of LTs and the CS-results in Sect. 4. The CS-conditions are generalized for the multivariate case and commented in Sect. 5 while they are verified for characteristic examples in Sect. 6. The main result on APN for multivariate LTs and properties of the MLEs, MAPs and EAPs of LTs are provided in Sect. 7 and supported by a simulation study in Sect. 8. Finally, the results are summarized in Sect. 9. A brief version of the proofs of the results of Sect. 7 is given in "Appendix" while their extended version can be found in the web-appendix. For a preliminary version of these results, see also Chapter 3 in Kornely (2021).

2. Preliminaries

Consider a test consisting of d binary response variables Y_i, $i \in [d] := \{1, \ldots, d\}$, with $Y_i \in \{0, 1\}$ for the i-th item, where 1 (0) denotes a correct (incorrect) response, and defined over a probability space $(\Omega, \mathcal{A}, \mathsf{P})$. Consider further the response vector $\mathbf{Y}^{(d)} = (Y_1, \ldots, Y_d)^\top$, with superscript $^\top$ denoting the transpose of a vector. Thus, the manifest probability for a specific response pattern $\mathbf{y}^{(d)}$ is given by $P(\mathbf{y}^{(d)}) = \mathsf{P}(\mathbf{Y}^{(d)} = \mathbf{y}^{(d)})$. In an multidimensional IRT (MIRT) modeling framework, manifest probabilities are derived via conditioning on an absolutely continuous latent variable vector $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_q)^\top \in \Theta \subseteq \mathbb{R}^q$, defined over the same probability space as the binary items with probability density function (pdf) and cumulative distribution function (cdf) $\mathfrak{h}$ and $\mathcal{H}$, respectively. In particular, the conditional probability mass function (pmf) of $Y_i \mid \boldsymbol{\eta}$ is thus given by

$$\mathsf{P}(Y_i = y \mid \boldsymbol{\eta}) = P_i(\boldsymbol{\eta})^y (1 - P_i(\boldsymbol{\eta}))^{1-y}, \, y \in \{0, 1\}, \,\, i \in [d], \tag{1}$$

with $P_i(\boldsymbol{\eta})$ being known as the i-th item response function. In MIRT modeling, specific assumptions are usually imposed on the conditional distribution $\mathsf{P}(\mathbf{Y}^{(d)} = \mathbf{y}^{(d)} \mid \boldsymbol{\eta})$; namely the assump-

tion of *local independence*

$$P^{(d)}(\mathbf{y}^{(d)} \mid \boldsymbol{\eta}) := \mathsf{P}(\mathbf{Y}^{(d)} = \mathbf{y}^{(d)} \mid \boldsymbol{\eta}) = \prod_{i=1}^{d} \mathsf{P}(Y_i = y_i \mid \boldsymbol{\eta}), \quad \mathbf{y}^{(d)} \in \{0, 1\}^d, \tag{2}$$

and that of *monotonicity* for the item response functions $P_i(\boldsymbol{\eta})$, i.e., for $i \in [d]$

$$P_i(\boldsymbol{\eta}) \text{ is strictly monotonic in every dimension of } \boldsymbol{\eta} \text{ being measured.} \tag{3}$$

Note that assumption (3), which is required in the CS-approach, is relaxed in our setup. In the sequel, we denote by $\{Y_i\}_{i \in \mathbb{N}} \sim \mathcal{P}(\boldsymbol{\eta})$ a sequence of Bernoulli random variables that fulfill (1) and (2) for all $d \in \mathbb{N}$.

Due to assumption (2) and using (1), the manifest probabilities are derived through the following integral

$$P(\mathbf{y}^{(d)}) = \int \ldots \int \left(\prod_{i=1}^{d} P_i(\boldsymbol{\eta})^{y_i} (1 - P_i(\boldsymbol{\eta}))^{1-y_i} \right) \mathfrak{h}(\boldsymbol{\eta}) \mathrm{d}\boldsymbol{\eta}. \tag{4}$$

Remark 1. For simplicity of notation, we use $\boldsymbol{\eta}$ to denote the random latent variable vector as well as a realization of it. If not clear from the context, we write explicitly $\boldsymbol{\eta} \in \Theta$ for a realization or $\boldsymbol{\eta} \sim \mathcal{H}$ for the random vector with values in Θ. In the sequel, we abbreviate the term latent variable vector to latent vector.

The posterior density of $\boldsymbol{\eta}$, given an observed response $\mathbf{y}^{(d)} \in \{0, 1\}^d$, is then given by

$$h(\boldsymbol{\eta} \mid \mathbf{y}^{(d)}) := \mathfrak{h}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)} = \mathbf{y}^{(d)}) = \frac{\mathsf{P}(\mathbf{Y}^{(d)} = \mathbf{y}^{(d)} \mid \boldsymbol{\eta}) \mathfrak{h}(\boldsymbol{\eta})}{\mathsf{P}(\mathbf{Y}^{(d)} = \mathbf{y}^{(d)})} = \frac{\exp(\ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{y}^{(d)})) \mathfrak{h}(\boldsymbol{\eta})}{P(\mathbf{y}^{(d)})}, \tag{5}$$

where $\ell^{(d)}(\cdot \mid \mathbf{y}^{(d)}))$ is the log-likelihood corresponding to (1), given by

$$\ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{y}^{(d)})) = \sum_{i=1}^{d} (y_i \lambda_i(\boldsymbol{\eta}) - \psi(\lambda_i(\boldsymbol{\eta}))), \tag{6}$$

with λ_i denoting the item logit, i.e.,

$$\lambda_i(\boldsymbol{\eta}) := \log \left(\frac{P_i(\boldsymbol{\eta})}{1 - P_i(\boldsymbol{\eta})} \right), \tag{7}$$

and the function $\psi(\cdot)$ being defined as $\psi(x) = \log (1 + \exp(x))$, $x \in \mathbb{R}$.

Let $\hat{\boldsymbol{\eta}}_d = \hat{\boldsymbol{\eta}}(\mathbf{y}^{(d)})$ denote the MLE of the true value of the latent vector $\boldsymbol{\eta}_0$, based on a test realization $\mathbf{y}^{(d)}$. Furthermore, the Fisher information matrix of the test at point $\boldsymbol{\eta}$ is given by

$$\mathcal{I}^{(d)}(\boldsymbol{\eta}) := \mathsf{E}_{\boldsymbol{\eta}} \left(\nabla \ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) \, \nabla^{\mathsf{T}} \ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) \right) = \sum_{i=1}^{d} \mathcal{I}_i(\boldsymbol{\eta}), \tag{8}$$

where $\mathcal{I}_i(\cdot)$ is the i-th item information matrix

$$
\begin{aligned}
\mathcal{I}_i(\boldsymbol{\eta}) &:= \mathsf{E}_{\boldsymbol{\eta}} \left(\nabla \log \left(P_i(\boldsymbol{\eta})^{Y_i} (1 - P_i(\boldsymbol{\eta}))^{1-Y_i} \right) \nabla^{\mathsf{T}} \log \left(P_i(\boldsymbol{\eta})^{Y_i} (1 - P_i(\boldsymbol{\eta}))^{1-Y_i} \right) \right) \\
&= P_i(\boldsymbol{\eta})(1 - P_i(\boldsymbol{\eta})) \nabla \lambda_i(\boldsymbol{\eta}) \nabla \lambda_i(\boldsymbol{\eta})^{\mathsf{T}}, \quad \boldsymbol{\eta} \in \Theta.
\end{aligned}
$$

This work studies the APN of $\boldsymbol{\eta}$ for $d \to \infty$, based on a sequence of random variables $\{Y_i\}_{i \in \mathbb{N}} \sim \mathcal{P}(\boldsymbol{\eta})$, as defined above. Particularly, we shall prove that, under certain conditions, (8) is invertible at $\hat{\boldsymbol{\eta}}_d$ and $\boldsymbol{\eta} \mid \mathbf{Y}^{(d)} = \mathbf{y}^{(d)}$ is approximately normal distributed, $\mathcal{N}(\hat{\boldsymbol{\eta}}_d, [\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)]^{-1})$, for a realization $\mathbf{y}^{(d)}$ of $\mathbf{Y}^{(d)}$. This enables the approximation of probabilities of the type

$$
\mathsf{P} \left(\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{1/2} \left(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d \right) \in B \,\middle|\, \mathbf{Y}^{(d)} \right), \quad B \in \mathcal{B}^q, \tag{9}
$$

where $\mathcal{B}^q$ denotes the Borel-σ-algebra of $\mathbb{R}^q$. Practically speaking, a set B can be any countable union or intersection of q-dimensional real cubes.

Next, we define some functions that are useful for the sequel derivations. For all $d \in \mathbb{N}$, set $Z^{(d)} \colon \Theta \times \Theta \to \mathbb{R}$ with

$$
Z^{(d)}(\boldsymbol{\eta}, \boldsymbol{\eta}') := \prod_{i=1}^{d} Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}'), \tag{10}
$$

where $Z_i \colon \Theta \times \Theta \to \mathbb{R}$, $i \in \mathbb{N}$, are defined as

$$
Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}') := \frac{P_i(\boldsymbol{\eta})^{Y_i} (1 - P_i(\boldsymbol{\eta}))^{1-Y_i}}{P_i(\boldsymbol{\eta}')^{Y_i} (1 - P_i(\boldsymbol{\eta}'))^{1-Y_i}}. \tag{11}
$$

Note that for given d and $\boldsymbol{\eta}$, $\boldsymbol{\eta}' \in \Theta$, (10) is the likelihood ratio of the likelihoods for $\boldsymbol{\eta}$ and $\boldsymbol{\eta}'$. Furthermore,

$$
\log \left(Z^{(d)}(\boldsymbol{\eta}, \boldsymbol{\eta}') \right) = \sum_{i=1}^{d} \log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}') = \ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) - \ell^{(d)}(\boldsymbol{\eta}' \mid \mathbf{Y}^{(d)}), \tag{12}
$$

while $-\mathsf{E}_{\boldsymbol{\eta}_0}(\log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}_0))$ is the Kullback–Leibler divergence between the conditional distributions of Y_i given $\boldsymbol{\eta}$ and $\boldsymbol{\eta}_0$, respectively. A basic approach for deriving APN results relies on a quadratic approximation of (12).

3. Review of APN for Univariate Latent Traits

In case of a single latent variable ($q = 1$, $\boldsymbol{\eta} = \eta$), Chang and Stout (1993) proved the APN of the univariate latent trait, adopting the approach of Walker (1969) for binary Y_i, $i \in [d]$, that are independent but not identically distributed (inid). We briefly review their results, so that we can extend in the sequel their approach to the multivariate case ($q > 1$).

Additional to the general assumptions (2) and (3), they also introduced the following regularity conditions.

(CS1)   [i] Let $\eta \in \Theta$, where $\Theta \subseteq (-\infty, \infty)$ is a bounded or unbounded interval.

[ii] Let the prior density $\mathfrak{h}$ be continuous and positive at the true value η_0.

(CS2) $P_i(\eta)$ is twice continuously differentiable with the first two derivatives being uniformly bounded in absolute value with respect to both η and i in some closed interval $\Theta_0 \subset \Theta$ around η_0.

(CS3) For every fixed $\eta \neq \eta_0$, $\eta \in \Theta$, there is a $c(\eta) > 0$ such that

$$\limsup_{d \longrightarrow \infty} \frac{1}{d} \sum_{i=1}^{d} \mathsf{E}_{\eta_0} \log Z_i(\eta, \eta_0) \leq -c(\eta), \tag{13}$$

and $\sup_{i \in \mathbb{N}} |\lambda_i(\eta)| < \infty$.

(CS4) If restricted to Θ_0, the following sets of functions are uniformly bounded:

$$\left\{ \left| \frac{\mathrm{d}\mathcal{I}_i}{\mathrm{d}\eta} \right| \mid i \in \mathbb{N} \right\}, \quad \left\{ \left| \frac{\mathrm{d}^2 \lambda_i}{\mathrm{d}\eta^2} \right| \mid i \in \mathbb{N} \right\}, \quad \left\{ \left| \frac{\mathrm{d}^3 \lambda_i}{\mathrm{d}\eta^3} \right| \mid i \in \mathbb{N} \right\}.$$

(CS5) Asymptotically, the average information at η_0 is bounded away from 0, i.e.,

$$\liminf_{d \longrightarrow \infty} \frac{\mathcal{I}^{(d)}(\eta_0)}{d} > 0.$$

Remark 2. With respect to the prior of η, additional to (CS1[ii]), Chang and Stout (1993) implicitly assumed its properness, which was stated explicitly in the earlier associated technical report (Chang & Stout, 1991, p. 15).

Remark 3. Reasonable models for applications do not depend on a specific compact interval in Θ since usually η_0 is unknown. For this, also the conditions depending on η_0 should be satisfied for almost all $\eta_0 \in \Theta$ and for almost each η_0 there should be some (arbitrary small) interval Θ_0. In the usual models these conditions are satisfied.

Chang and Stout (1993) argued convincingly that conditions (CS1)–(CS5) are realistic and non-restrictive in practice for commonly used IRT models of well-designed tests. They particularly commented condition (CS3) and (13), which plays an important role in the proof of their main theorem. (CS3) is required when the item responses $\{Y_i\}_{i \in \mathbb{N}}$ are independent but not identically distributed. If they are iid, (CS3) is automatically satisfied, which however is not necessarily the case in IRT models. Their main results are expressed in the three theorems given below.

Theorem 1. (Chang & Stout, 1993, Theorem 1) *Suppose that conditions (CS1) through (CS5) hold for a fixed η_0. Let $\hat{\eta}_d$ be the MLE of η_0 and $\hat{\sigma}_d = (\mathcal{I}^{(d)}(\hat{\eta}_d))^{-1/2}$. Then, for $-\infty \leq a < b \leq \infty$, the posterior probability of $\hat{\eta}_d + a\hat{\sigma}_d < \eta < \hat{\eta}_d + b\hat{\sigma}_d$ approaches the probability of $Z \in (a, b)$ in P_{η_0} for $Z \sim \mathcal{N}(0, 1)$, that means*

$$A_d \equiv \int_{\hat{\eta}_d + a\hat{\sigma}_d}^{\hat{\eta}_d + b\hat{\sigma}_d} h(\eta \mid \mathbf{Y}^{(d)}) \, \mathrm{d}\eta \xrightarrow{\mathsf{P}_{\eta_0}} \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{1}{2}\eta^2 \right) \mathrm{d}\eta \equiv A, \quad d \to \infty.$$

Theorem 2. (Chang & Stout, 1993, Theorem 2) *Suppose that conditions (CS1) through (CS5) hold for fixed η_0 and let $\hat{\eta}_d$ and $\hat{\sigma}_d$ be defined as in* Theorem 1. *Then, for $-\infty \leq a < b \leq \infty$, the posterior probability A_d approaches A P_{η_0}-almost surely, as $d \to \infty$.*

Theorem 3. (Chang & Stout, 1993, Theorem 3) *Assume* $\Theta = \Theta_0$, *a finite interval. Suppose that conditions (CS1) through (CS5) hold for all* $\eta_0 \in \Theta_0$ *and let* $\hat{\eta}_d$ *and* $\hat{\sigma}_d$ *be defined as in* Theorem 1. *Then, for* $-\infty \leq a < b \leq \infty$, *the posterior probability* A_d *approaches* A *in manifest probability* P, *as* $d \to \infty$.

The result of Theorem 3 does not depend on the true value η_0 and is thus of special practical interest for estimation and prediction purposes. As Chang and Stout (1993) comment, Theorems 1 and 3 treat sampling from a fixed ability sub-population and from the whole population, respectively. An important by-product of the proof of the APN of latent variables distributions was the establishment of the weak and strong consistency of the MLE of η under milder conditions than Lord (1983).

Due to the theorems above, the following approximation for a large d and any observed response pattern $\mathbf{y} \in \{0, 1\}^d$, i.e., the construction of asymptotic credible intervals, is justified

$$\mathsf{P}(a \leq \eta \leq b \mid \mathbf{Y}^{(d)} = \mathbf{y}) \approx \Phi_1(b \,;\, \hat{\eta}_d, \hat{\sigma}^2) - \Phi_1(a \,;\, \hat{\eta}_d, \hat{\sigma}^2), \tag{14}$$

for $-\infty \leq a \leq b \leq \infty \in \mathbb{R}$, where $\hat{\eta}_d$ is the MLE of η_0 based on the sample $\mathbf{y}$, $\hat{\sigma}^2 = (\mathcal{I}^{(d)}(\hat{\eta}_d))^{-1}$ and $\Phi_1(\cdot \,;\, \hat{\eta}_d, \hat{\sigma}^2)$ denotes the cdf of $\mathcal{N}(\hat{\eta}_d, \hat{\sigma}^2)$. Approximation (14) is of special practical importance in the context of long tests where the exact computation of posterior probabilities for latent variables is commonly intractable. Furthermore, (14) allows the approximation of the posterior if the exact distribution $\mathcal{H}$ of η is unavailable or uncertain.

Finally, Chang and Stout (1993) noted that their theory, under suitable regularity conditions, can be extended to prove the APN for latent vectors of general multidimensional IRT models, without however commenting further the proving procedure or the regularity conditions required. Next, we discuss the asymptotic posterior distribution of multivariate latent traits in the context of MIRT.

4. APN for Multivariate Latent Traits

The theory of APN of the latent variables is naturally linked to Bayesian procedures and results on the convergence of posterior distributions. In particular, interesting and inspiring is the fundamental contribution by GGS (Ghosal et al., 1995), who consider asymptotic multivariate posterior distributions (not necessarily normal) in a very general and flexible framework discussing different types of convergence, relying on earlier works by Ghosh et al. (1994) and Ibragimov and Has'minskii (1981), denoted as IH hereafter. In particular, they studied posterior convergence of suitably centered and normalized posteriors. Their results provide a very general framework, which can be adopted for the APN in the IRT setup. Next, we adjust the GGS approach for MIRT models and discuss their conditions, embedding the CS approach in the GGS framework.

Following Ghosal et al. (1995, Definition 2), we distinguish two types of APN and link them to the statistic used for the centering of the posterior distribution of the latent vector.

Definition 1. Let $\mathbf{Z} \sim \mathcal{N}_q(\mathbf{0}, \mathsf{I}_q)$ be a q-variate standard normal distributed random vector. A $\mathbb{R}^q$-valued statistic $\tilde{\boldsymbol{\eta}}_d$ is called a *proper centering (with limiting normal distribution)* if

$$\sup_{A \in \mathcal{B}^q} \left| \mathsf{P}\left(\mathcal{I}^{(d)}(\tilde{\boldsymbol{\eta}}_d)^{1/2}(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}_d) \in A \mid \mathbf{Y}^{(d)} \right) - \mathsf{P}(\mathbf{Z} \in A) \right| \xrightarrow{\mathsf{P}_{\eta_0}} 0, \text{ as } d \to \infty. \tag{15}$$

A statistic $\tilde{\boldsymbol{\eta}}_d$ is called *semiproper centering (with limiting normal distribution)* if, for all $A \in \mathcal{B}^q$,

$$\mathsf{P}\left(\mathcal{I}^{(d)}(\tilde{\boldsymbol{\eta}}_d)^{1/2}(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}_d) \in A \mid \mathbf{Y}^{(d)} \right) \xrightarrow{\mathsf{P}_{\eta_0}} \mathsf{P}(\mathbf{Z} \in A), \text{ as } d \to \infty. \tag{16}$$

A statistic $\tilde{\boldsymbol{\eta}}_d$ is called *compatible* (with the posterior), if

$$\left(\mathcal{I}^{(d)}(\tilde{\boldsymbol{\eta}}_d)^{1/2}(\tilde{\boldsymbol{\eta}}_d - \boldsymbol{\eta}_0), h^* \left(\cdot \left| \mathbf{Y}^{(d)} \right) \right),$$

as a random element in $\mathbb{R}^q \times L^1(\mathbb{R}^q)$, converges in distribution for $d \to \infty$, where $h^*(\cdot \mid \mathbf{Y}^{(d)})$ denotes the density of the posterior distribution of $\mathcal{I}^{(d)}(\boldsymbol{\eta}_0)^{1/2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)$ and $L^1(\mathbb{R}^q)$ stands for the space of all q-variate Lebesgue-integrable real functions on $\mathbb{R}^q$.

Proper and semiproper centering correspond to uniform and pointwise convergence of the posterior of the standardized latent vector $\mathcal{I}^{(d)}(\tilde{\boldsymbol{\eta}}_d)^{1/2}(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}_d)$, respectively. Hence, proper centering is a stronger property than semiproper centering, and is consequently expected to require stronger assumptions.

Under this view, one can easily recognize that Theorem 1 of Chang and Stout (1993) is the semiproper centering of the MLE, since it can be formulated as

$$\mathsf{P}\left(\mathcal{I}^{(d)}(\hat{\eta}_d)^{1/2} \left(\eta - \hat{\eta}_d \right) \in [a, b] \left| \mathbf{Y}^{(d)} \right. \right) \xrightarrow{\mathsf{P}_{\eta_0}} \mathsf{P}\left(Z \in [a, b] \right), \quad d \longrightarrow \infty,$$

for $Z \sim \mathcal{N}(0, 1)$, $A = [a, b] \subset \mathbb{R}$ and $\tilde{\eta}_d = \hat{\eta}_d$ being the MLE of η_0 based on $\mathbf{Y}^{(d)}$. Thus, for the extension of the CS-theory for multivariate LTs, we focus on semiproper centering.

The asymptotic results of GGS adjusted in our setup, primarily focus on the convergence of the posterior distribution of the standardized latent vector

$$\boldsymbol{\eta}^* = \mathcal{I}^{(d)}(\tilde{\boldsymbol{\eta}}_d)^{1/2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0), \tag{17}$$

with $\boldsymbol{\eta}^* \in \Theta_d := \mathcal{I}^{(d)}(\tilde{\boldsymbol{\eta}}_d)^{1/2}(\Theta - \boldsymbol{\eta}_0)$. We need the likelihood ratio (10) expressed in terms of $\boldsymbol{\eta}^*$, which is denoted by

$$Z^{*(d)}(\boldsymbol{\eta}^*) := Z^{(d)}(\boldsymbol{\eta}_0 + \mathcal{I}^{(d)}(\tilde{\boldsymbol{\eta}}_d)^{-1/2}\boldsymbol{\eta}^*, \boldsymbol{\eta}_0). \tag{18}$$

In our setup, for binary response variables Y_i, $i \in [d]$, and log-likelihoods given by (6), the likelihood ratio $Z^{*(d)}$ takes the form

$$Z^{*(d)}(\boldsymbol{\eta}^*) = \exp \left(\sum_{i=1}^d \left(Y_i \left(\lambda_i \left(\boldsymbol{\eta}_0 + \mathcal{I}^{(d)}(\tilde{\boldsymbol{\eta}}_d)^{-1/2}\boldsymbol{\eta}^* \right) - \lambda_i \left(\boldsymbol{\eta}_0 \right) \right) \right. \right.$$
$$\left. \left. - \left(\psi \left(\lambda_i \left(\boldsymbol{\eta}_0 + \mathcal{I}^{(d)}(\tilde{\boldsymbol{\eta}}_d)^{-1/2}\boldsymbol{\eta}^* \right) \right) - \psi \left(\lambda_i \left(\boldsymbol{\eta}_0 \right) \right) \right) \right) \right),$$

with the item logits λ_i provided in (7).

The primary conditions of GGS for APN are given as follows:

(GGS1) For some $M > 0$, $m_1 \geq 0$ and $\alpha > 0$ holds

$$\mathsf{E}_{\boldsymbol{\eta}_0} \left(\left| Z^{*(d)}(\boldsymbol{\eta}_1^*)^{1/2} - Z^{*(d)}(\boldsymbol{\eta}_2^*)^{1/2} \right|^2 \right) \leq M(1 + R^{m_1})\|\boldsymbol{\eta}_1^* - \boldsymbol{\eta}_2^*\|^\alpha,$$

for all $\boldsymbol{\eta}_j^* \in \Theta_d$, satisfying $\|\boldsymbol{\eta}_j^*\| \leq R$, $j = 1, 2$, where $\|\cdot\|$ is the Euclidean norm.

(GGS2) For all $\boldsymbol{\eta}^* \in \Theta_d$ holds

$$\mathsf{E}_{\boldsymbol{\eta}_0} \left(Z^{*(d)}(\boldsymbol{\eta}^*)^{1/2} \right) \leq \exp \left(- g_d(\|\boldsymbol{\eta}^*\|) \right),$$

where $\{g_d\}_{d \in \mathbb{N}}$ is a sequence of real-valued functions on $[0, \infty)$ satisfying the following: (a) for a fixed $d \geq 1$, $\lim_{x \to \infty} g_d(x) = \infty$; (b) for any $N > 0$,

$$\lim_{x \to \infty} \lim_{d \to \infty} x^N \exp(-g_d(x)) = \lim_{d \to \infty} \lim_{x \to \infty} x^N \exp(-g_d(x)) = 0.$$

(GGS3) For all $n \in \mathbb{N}$ and $\boldsymbol{\eta}_1^*, \ldots, \boldsymbol{\eta}_n^* \in \mathbb{R}^q$, the vector of the likelihood-ratios, defined in (18), satisfies

$$(Z^{*(d)}(\boldsymbol{\eta}_1^*), \ldots, Z^{*(d)}(\boldsymbol{\eta}_n^*)) \xrightarrow{\mathcal{D}} (Z(\boldsymbol{\eta}_1^*), \ldots, Z(\boldsymbol{\eta}_n^*)),$$

for $d \longrightarrow \infty$, where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and $Z(\boldsymbol{\eta}^*) = \exp(\boldsymbol{\xi}^T \boldsymbol{\eta}^* - \frac{1}{2}\|\boldsymbol{\eta}^*\|^2)$, $\boldsymbol{\eta}^* \in \mathbb{R}^q$, where $\boldsymbol{\xi} \sim \mathcal{N}_q(\mathbf{0}, \mathrm{I}_q)$.

Under these conditions, Ghosal et al. (1995) provided the following general result. Notice that they discussed a far more general framework, allowing further distributions for the response variable and considering cases for which the posterior may converge to another distribution than a normal. We refer to GGS for further details regarding these cases.

Theorem 4. (Ghosal et al., 1995, Theorem 1) *Assume that conditions (GGS1) through (GGS3) hold. If either a proper centering or a semiproper compatible centering sequence $\{\tilde{\boldsymbol{\eta}}_d\}_{d \in \mathbb{N}}$ exists, then it exists a random vector $\boldsymbol{W}$, such that (a) $\mathcal{I}^{(d)}(\tilde{\boldsymbol{\eta}}_d)^{1/2}(\tilde{\boldsymbol{\eta}}_d - \boldsymbol{\eta}_0) \xrightarrow{\mathcal{D}} \boldsymbol{W}$ for $d \longrightarrow \infty$ and (b) for almost all $\boldsymbol{x} \in \mathbb{R}^q$, $\frac{Z(\boldsymbol{x}-\boldsymbol{W})}{\int_{\mathbb{R}^q} Z(\boldsymbol{x}^*-\boldsymbol{W}) \, d\boldsymbol{x}^*}$ is nonrandom, where Z is as defined in condition (GGS3).*

Conversely, if (b) holds for a random vector $\boldsymbol{W}$, then any Bayes estimator (with respect to a prior and loss considered by Ghosal et al. (1995)) *is a compatible proper centering.*

Applying Theorem 4 for an appropriate Bayes estimator for $\boldsymbol{\eta}_0$, the APN of an MIRT model under conditions (GGS1) to (GGS3) is derived. The extension of Theorem 4 for an MLE, i.e., for $\tilde{\boldsymbol{\eta}}_d = \hat{\boldsymbol{\eta}}_d$, is based on its asymptotic equivalence to an arbitrary Bayes estimator, which has been proved by Ghosal et al. (1995, cf. Corollary 1) under (GGS2)–(GGS3) and the following strengthened form of (GGS1):

(GGS1') For some $M > 0$, $m_1 \geq 0$ and $m \geq \alpha > q$ holds

$$\mathsf{E}_{\boldsymbol{\eta}_0} \left(\left| Z^{*(d)}(\boldsymbol{\eta}_1^*)^{1/m} - Z^{*(d)}(\boldsymbol{\eta}_2^*)^{1/m} \right|^m \right) \leq M(1 + R^{m_1})\|\boldsymbol{\eta}_1^* - \boldsymbol{\eta}_2^*\|^{\alpha},$$

for all $\boldsymbol{\eta}_j^* \in \Theta_d$, satisfying $\|\boldsymbol{\eta}_j^*\| \leq R$, $j = 1, 2$.

Remark 4. Alternatively to the GGS conditions discussed above, one could consider the conditions of Ibragimov & Has'minskii (1981, Section III.4) for general regular models for independent non-necessarily identical distributed (inid) random variables. They proved that these conditions are sufficient for the set of conditions N1–N4 of IH, Section III.1, where N1 is the *uniform asymptotic normality* and corresponds to (GGS3), while N3 and N4 correspond to (GGS1) and (GGS2), respectively.

5. Regularity Conditions for Asymptotic Properties of Latent Vectors

Aiming to generalize the CS approach, we provide conditions for APN of (multivariate) LTs that correspond one to one to the conditions of CS for univariate LTs, which is the standard approach for IRT models, as alternatives to the conditions imposed in Ghosal et al. (1995). Throughout, we assume that $\{Y_i\}_{i \in \mathbb{N}} \sim \mathcal{P}(\boldsymbol{\eta})$, i.e., $\{Y_i\}_{i \in \mathbb{N}}$ are Bernoulli random variables fulfilling (1) and (2), and that the true latent vector $\boldsymbol{\eta}_0$ lies in the interior of the parameter space, i.e., $\boldsymbol{\eta}_0 \in \Theta \setminus \partial\Theta$, where $\partial\Theta$ denotes the boundary of Θ. The asymptotic results of Sect. 7 rely on the following regularity conditions.

(CS1')   [i] The set Θ is closed, convex and has non-empty interior.
   [ii] The prior density $\mathfrak{h}$ of $\boldsymbol{\eta}$ is proper and continuous at $\boldsymbol{\eta}_0$ with $\mathfrak{h}(\boldsymbol{\eta}_0) > 0$.

(CS2') P_i is thrice continuously differentiable, $i \in \mathbb{N}$. If restricted to a compact subset $K \subseteq \Theta$, all $\left|\frac{\partial P_i}{\partial \eta_k}\right|$ and $\left|\frac{\partial^2 P_i}{\partial \eta_k \partial \eta_j}\right|$ are uniformly bounded for all $i \in \mathbb{N}$, $1 \leq j, k \leq q$. Moreover, there exist constants $0 < \zeta_0(K) < \zeta_1(K) < 1$, which are independent of $i \in \mathbb{N}$, such that

$$\zeta_0(K) \leq \inf_{(i,\boldsymbol{\eta}) \in \mathbb{N} \times K} P_i(\boldsymbol{\eta}) \leq \sup_{(i,\boldsymbol{\eta}) \in \mathbb{N} \times K} P_i(\boldsymbol{\eta}) \leq \zeta_1(K). \tag{19}$$

(CS3') For each $\boldsymbol{\eta} \in \Theta$, $\boldsymbol{\eta} \neq \boldsymbol{\eta}_0$, there is a $c(\boldsymbol{\eta}) < 0$ such that

$$\limsup_{d \to \infty} \frac{1}{d} \sum_{i=1}^{d} \mathsf{E}_{\boldsymbol{\eta}_0}(\log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}_0)) = \limsup_{d \to \infty} \frac{1}{d} \mathsf{E}_{\boldsymbol{\eta}_0}(\ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) - \ell^{(d)}(\boldsymbol{\eta}_0 \mid \mathbf{Y}^{(d)})) \leq c(\boldsymbol{\eta}),$$

and if Θ is unbounded holds additionally

$$\sup_{\boldsymbol{\eta} \in \Theta \setminus B_\delta(\boldsymbol{\eta}_0)} c(\boldsymbol{\eta}) < 0, \qquad \text{for all } \delta > 0, \tag{20}$$

where $B_\delta(\boldsymbol{\eta}_0) := \{\boldsymbol{\eta} \in \mathbb{R}^q : \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| < \delta\}$ is the open ball of radius δ and center $\boldsymbol{\eta}_0$.

(CS4') If restricted to any compact set $K \subseteq \Theta$, the following set of functions is uniformly bounded

$$\left\{ \left| \frac{\partial^3 P_i}{\partial \eta_k \partial \eta_g \partial \eta_u} \right| : i \in \mathbb{N}, 1 \leq k, g, u \leq q \right\}.$$

(CS5') For all $\boldsymbol{\eta} \in \Theta$ holds

$$\liminf_{d \to \infty} \nu_{\min} \left(\frac{1}{d} \sum_{i=1}^{d} \nabla\lambda_i(\boldsymbol{\eta}) \nabla\lambda_i(\boldsymbol{\eta})^\mathsf{T} \right) > 0, \tag{21}$$

where $\nu_{\min}$ denotes the smallest eigenvalue.

These regularity conditions correspond one to one to conditions (CS1)–(CS5), given in Sect. 3. For the comparison of these conditions, have in mind that convexity and connectivity are equivalent properties in $\mathbb{R}$. The convexity condition in (CS1') is at first place stronger but it does not impose a real practical restriction, since non-convex Θ are only rarely needed in MIRT. Analogue to the CS-theory (s. Remark 3), conditions involving η_0, like $\mathfrak{h}(\eta_0) > 0$, should be interpreted as $\mathfrak{h} > 0$ almost surely. Condition (CS1'[ii]) on $\mathfrak{h}$ seems more strict than (CS1[ii]). However, Chang and Stout (1993) require additional a proper prior (s. Remark 2). Thus, under the consideration that η_0 is still unknown and we consider $\mathbb{R}^q$ instead of $\mathbb{R}$, the requirements on proper priors in (CS1'[ii]) are analogue to (CS1[ii]). Finally note that in the generalization of conditions (CS3) and (CS4), some requirements have been removed as the remaining requirements on λ_i, $i \in \mathbb{N}$, and its derivatives are implied by conditions (CS2') and (CS4').

A common assumption in one-dimensional IRT models ($q = 1$) is the strict monotonicity assumption (3) of P_i in η, for all $i \in \mathbb{N}$. Conceptually, this represents the notion that a more able subject has a higher probability of responding correct in any item of an educational test. Thus, models fulfilling this strict monotonicity assumption are easier to interpret. However, models with non-generalized-linear latent variable effects can be more adequate in practice. For example, Rizopoulos and Moustaki (2008) considered IRT models with possibly non-monotonic latent variable dependencies (like polynomial effects). Due to this reason, in order to allow for more flexible modeling options, in our semiproper centering theory, we abandon the requirement on strict monotonicity of $\eta \mapsto P_i(\eta)$, for all $i \in \mathbb{N}$, in each component. Since the results of Chang and Stout (1993) rely on this monotonicity assumption, the merit of the current contribution is not only the extension of the CS-results for latent vectors ($q > 1$) but also for univariate latent variables in case of a non-monotonic latent variable effect.

If all P_i, $i \in \mathbb{N}$, are strictly monotonic in each component, then requirement (19) of condition (CS2') is satisfied as in the univariate case. Otherwise, the requirement in (19) is generally not really restrictive; it is the technical formulation of the notion that the response probabilities can (but not necessarily have to) approach zero or one only if $\|\eta\| \longrightarrow \infty$. Assumption (20) in (CS3') serves for ensuring the identifiability of the latent vector in case of $\|\eta\| \longrightarrow \infty$. Hence, this condition is quite natural for a statistical model. In the univariate case, (20) of (CS3') is implied by the strict monotonicity, too. But in contrast to (19), (20) cannot be concluded directly from the strict monotonicity of all P_i in each component if $q > 1$. Moreover, while a single item can suffice in the univariate case for identifiability, there are always at least q needed in the q-dimensional one. Similarly, the average test information $\frac{1}{d}\mathcal{I}^{(d)}(\eta)$ is always singular for $d < q$, since it is a sum of d rank-one matrices. Condition (CS5') ensures that $\frac{1}{d}\mathcal{I}^{(d)}(\eta)$ becomes regular for $d \to \infty$ and can be interpreted as a condition to ensure that the asymptotic posterior of η is regular q-dimensional distributed and does not have a lower dimensional support (cf. Lemma W.6 in the web-appendix).

To get a better impression of the conditions, we exemplary discuss them next for model (22), the multidimensional version of a model of Lee and Bolt (2018) and a logit model with an interaction of the latent variables (Rizopoulos, 2006).

6. Verification of the CS Regularity Conditions for Multidimensional IRT Models

We shall verify the proposed conditions (CS1') to (CS5') for a multidimensional version of a model by Lee and Bolt (2018) and discuss them also for the models of Pelle et al. (2016) and a logit model with interaction of the latent variables (Rizopoulos, 2006).

Consider first the IRT model by Lee and Bolt (2018) or its multidimensional version

$$P_i(\boldsymbol{\eta}) = \Phi_1\left(\frac{\boldsymbol{\alpha}_i^\mathsf{T}\boldsymbol{\eta} + \beta_i}{\sqrt{2}(1 + \exp(-\boldsymbol{\delta}_i^\mathsf{T}\boldsymbol{\eta}))^{-1/2}}\right), \qquad \boldsymbol{\eta} \in \Theta, \; i \in \mathbb{N}. \tag{22}$$

If $\mathcal{H}$ is one of the usual structural models or any other regular distribution, for example $\mathcal{N}_q(\mathbf{0}, \mathbf{I}_q)$, a mixture of normals or a uniform distribution on some compact set, then (CS1') is directly satisfied. Further, considering the model parameters as random variables, we assume that the parameter sequences $\{\boldsymbol{\alpha}_i\}_{i\in\mathbb{N}}$ and $\{\boldsymbol{\delta}_i\}_{i\in\mathbb{N}}$ behave as they were two independent iid sequences drawn from absolutely continuous regular distributions in some bounded region in $\mathbb{R}^q$ and the sequence $\{\beta_i\}_{i\in\mathbb{N}}$ is in an arbitrary bounded subset of $\mathbb{R}$, then conditions (CS2') and (CS4') are directly satisfied. The assumption of regular distributions with a bounded support for the model parameters is reasonable, since in IRT practice items with arbitrarily large discrimination are not realistic and items of arbitrarily high or low difficulty are avoided. Furthermore, in almost all cases, the latent vector is identifiable if q arbitrary items are given. Hence, (CS3') is satisfied, too. The gradient of the response probabilities is given by

$$\begin{aligned}
\nabla P_i(\boldsymbol{\eta}) = {}& \phi_1\left(\frac{\boldsymbol{\alpha}_i^\mathsf{T}\boldsymbol{\eta} + \beta_i}{\sqrt{2}(1 + \exp(-\boldsymbol{\delta}_i^\mathsf{T}\boldsymbol{\eta}))^{-1/2}}\right) \\
& \times \left(\sqrt{\frac{1 + \exp(-\boldsymbol{\delta}_i^\mathsf{T}\boldsymbol{\eta})}{2}}\boldsymbol{\alpha}_i - \frac{\exp(-\boldsymbol{\delta}_i^\mathsf{T}\boldsymbol{\eta})}{2\sqrt{2}}\frac{\boldsymbol{\alpha}_i^\mathsf{T}\boldsymbol{\eta} + \beta_i}{\sqrt{1 + \exp(-\boldsymbol{\delta}_i^\mathsf{T}\boldsymbol{\eta})}}\boldsymbol{\delta}_i\right),
\end{aligned} \tag{23}$$

for all $i \in \mathbb{N}$ and $\boldsymbol{\eta} \in \Theta$, where ϕ_q denotes the pdf of $\mathcal{N}_q(\mathbf{0}, \mathrm{I}_q)$. In particular, we see from (23), that $\{\nabla P_i(\boldsymbol{\eta})\}_{i\in\mathbb{N}}$ behaves in almost all cases for all $\boldsymbol{\eta} \in \Theta$ as an iid sequence drawn from a regular distribution with bounded support in $\mathbb{R}^q$, since the parameters are iid distributed for all items and every $\boldsymbol{\eta} \in \Theta$ is considered separately, i.e., $\boldsymbol{\eta}$ is held fixed. Exceptions are pathological cases like the one in which zero belongs to the support of the distributions of all model parameters and all model parameters equal zero, i.e., $P_i(\boldsymbol{\eta}) = 0.5$ for all $\boldsymbol{\eta} \in \Theta$ and $i \in \mathbb{N}$. However, the subset of such cases is of zero probability for regular continuous distributions, i.e., is a null-set. Thus,

$$\frac{1}{d}\sum_{i=1}^{d} \nabla P_i(\boldsymbol{\eta})\nabla P_i(\boldsymbol{\eta})^\mathsf{T}$$

converges to the second moment of the distribution of $\nabla P_1(\boldsymbol{\eta})$, as a random vector formed by the multivariate transformation of the randomly selected parameter values described directly after equation (23), and is thus positive definite.

With respect to (CS5'), note that (19) in (CS2') implies that (21) is equivalent to

$$\liminf_{d\to\infty} \nu_{\min}\left(\frac{1}{d}\sum_{i=1}^{d}\mathcal{I}_i(\boldsymbol{\eta})\right) > 0 \quad \text{and to} \quad \liminf_{d\to\infty} \nu_{\min}\left(\frac{1}{d}\sum_{i=1}^{d}\nabla P_i(\boldsymbol{\eta})\nabla P_i(\boldsymbol{\eta})^\mathsf{T}\right) > 0, \tag{24}$$

which in our case ensures that (CS5') is satisfied (cf. Lemma W.6 in the web-appendix).

For illustrative purposes, consider an example with $d = 30$ and $q = 2$ and model parameter values, as given in Table 1, which are independently drawn from a uniform distribution on $(-2, 2)$ for β_i, and on $(-0.5, 1)$ for all other parameters.

TABLE 1.
Hypothetical parameter values for model (22) and the first 30 items.

i	β_i	α_{i1}	α_{i2}	δ_{i1}	δ_{i2}	i	β_i	α_{i1}	α_{i2}	δ_{i1}	δ_{i2}
1	-0.906	0.946	-0.398	0.230	0.327	16	-0.104	-0.498	0.298	0.359	0.622
2	-1.337	0.893	-0.085	0.696	0.622	17	0.458	0.496	0.267	0.677	0.551
3	-0.913	-0.363	0.703	0.052	0.605	18	-1.882	-0.259	-0.484	0.101	-0.022
4	0.387	0.166	0.313	-0.352	-0.447	19	0.801	0.066	-0.488	0.282	0.697
5	0.676	0.677	0.199	0.701	0.283	20	1.953	-0.169	0.963	-0.467	0.443
6	-0.084	0.865	0.833	0.957	0.724	21	-1.712	0.326	0.979	-0.146	-0.347
7	0.138	0.688	-0.444	0.942	-0.043	22	1.725	0.146	0.875	0.843	0.092
8	1.608	0.296	0.472	-0.096	-0.101	23	1.980	0.826	0.517	0.195	0.825
9	-0.882	0.016	-0.378	0.088	0.458	24	0.171	0.197	0.722	0.069	-0.442
10	1.240	0.320	0.924	-0.250	0.386	25	0.499	-0.101	-0.319	0.943	0.071
11	-0.919	0.867	-0.391	0.051	0.202	26	-1.775	0.530	-0.188	0.819	-0.336
12	0.216	0.093	0.028	0.152	0.339	27	1.523	-0.302	0.646	0.581	0.718
13	1.900	0.726	0.007	0.842	-0.072	28	1.548	0.272	0.469	-0.207	-0.241
14	0.510	-0.068	-0.105	0.132	-0.256	29	-0.823	-0.113	-0.147	0.027	-0.282
15	-1.109	0.072	-0.017	0.370	0.623	30	0.822	0.661	-0.485	-0.277	-0.017

In Fig. 1 (top) visualizations of $\mathsf{E}_{\boldsymbol{\eta}_0}(\log(Z_5(\boldsymbol{\eta}, \boldsymbol{\eta}_0)))$ are provided for two exemplary values of $\boldsymbol{\eta}_0 \in \Theta$ and the parameters in Table 1. In particular, we can recognize lines in Θ, for which $\mathsf{E}_{\boldsymbol{\eta}_0}(\log(Z_5(\boldsymbol{\eta}, \boldsymbol{\eta}_0))) = 0$ holds. In Fig. 1 (bottom), surfaces of $\frac{1}{30} \sum_{i=1}^{30} \mathsf{E}_{\boldsymbol{\eta}_0}(\log(Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}_0)))$ are illustrated for further two exemplary $\boldsymbol{\eta}_0$ values. The surfaces are drawn over $[-1, 1]^2$. There is a nearly parabolic surface, which illustrates that there is no reason to doubt for (20) in (CS3').

Figure 2 provides the minimal smallest eigenvalue of $\frac{1}{d} \sum_{i=1}^{d} \nabla\lambda_i(\boldsymbol{\eta}) \nabla\lambda_i(\boldsymbol{\eta})^{\mathsf{T}}$ on $[-1, 1]^2$ for $d \in \{1, \ldots, 30\}$, cf. (CS5'). Overall, in this case, the regularity conditions can be considered as justified to apply the APN for arbitrary response patterns on the illustrated 30 items.

Conditions (CS1') to (CS5') for other models can be verified similarly. For example, the multidimensional Rasch model implemented in Pelle et al. (2016) is the logit model

$$\text{logit}(\mathsf{P}(Y_i = 1 \mid \boldsymbol{\eta})) := \log\left(\frac{\mathsf{P}(Y_i = 1 \mid \boldsymbol{\eta})}{1 - \mathsf{P}(Y_i = 1 \mid \boldsymbol{\eta})}\right) = \alpha_{i0} + \boldsymbol{\alpha}_i^{\mathsf{T}}\boldsymbol{\eta}, \qquad \boldsymbol{\eta} \in \mathbb{R}^q, \quad i = 1, \ldots, d,$$

(25)

where $\mathsf{P}(Y_i = 1 \mid \boldsymbol{\eta})$ is the probability of inclusion in registration i, given the vector of latent variables. In this case,

$$\nabla\lambda_i(\boldsymbol{\eta}) = \boldsymbol{\alpha}_i, \qquad i = 1, \ldots, d,$$

would replace (23) for $\boldsymbol{\eta} \in \Theta$, while the subsequent arguments are the same as above.

Another example is the two-dimensional model

$$\lambda_i(\boldsymbol{\eta}) = \alpha_{i0} + \alpha_{i1}\eta_1 + \alpha_{i2}\eta_2 + \alpha_{i3}\eta_1\eta_2, \qquad i = 1, \ldots, d,$$

which is a logit model that contains an interaction term between the two latent variables and was considered by Rizopoulos (2006) (see Section 4). While it is still logit-linear in the model
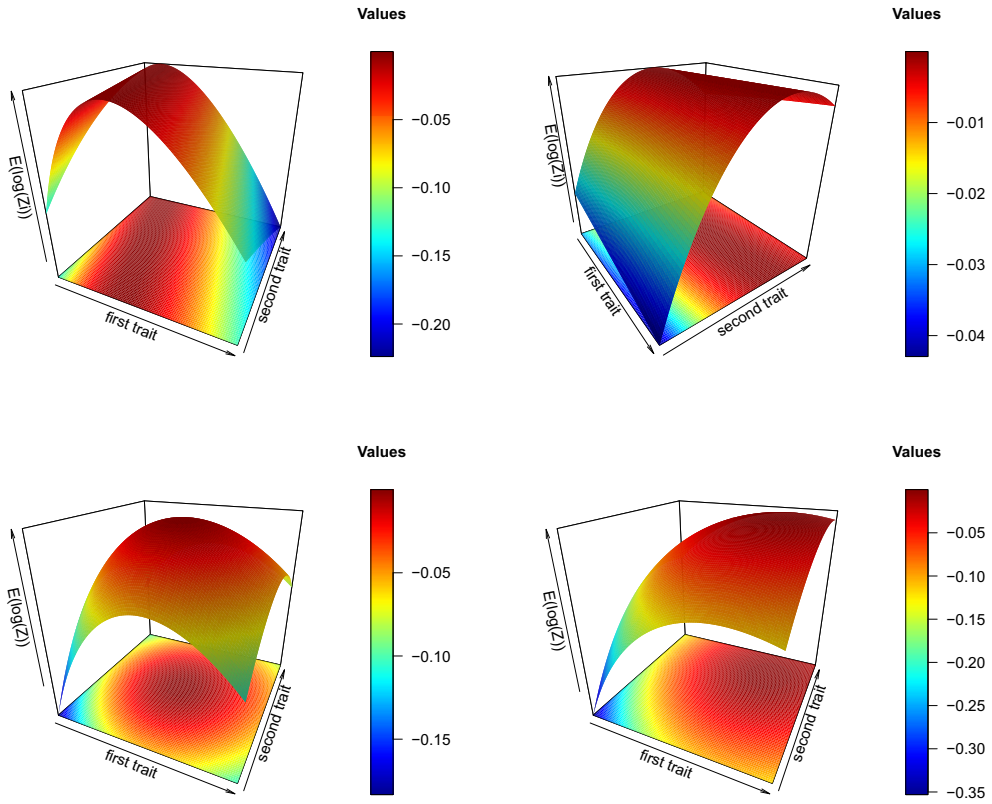
FIGURE 1.
Top left: $\mathsf{E}_{(-0.5,0.1)}(\log(Z_5(\boldsymbol{\eta}, (-0.5, 0.1))))$, top right: $\mathsf{E}_{(-0.1,0.3)}(\log(Z_9(\boldsymbol{\eta}, (-0.1, 0.3))))$,
bottom left: $\frac{1}{30} \sum_{i=1}^{30} \mathsf{E}_{(0,0)}(\log(Z_i(\boldsymbol{\eta}, (0, 0))))$, bottom right: $\frac{1}{30} \sum_{i=1}^{30} \mathsf{E}_{(0.5,0.5)}(\log(Z_i(\boldsymbol{\eta}, (0.5, 0.5))))$.
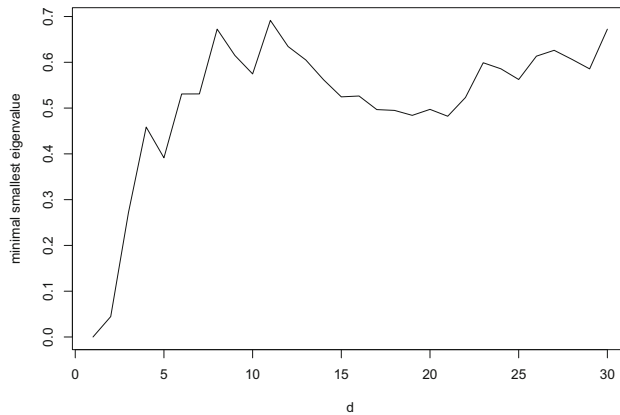
FIGURE 2.
Minimal eigenvalue of $\frac{1}{d} \sum_{i=1}^{d} \nabla\lambda_i(\boldsymbol{\eta})\nabla\lambda_i(\boldsymbol{\eta})^{\mathsf{T}}$ for $\boldsymbol{\eta} \in [-1, 1]^2$.

parameters α_{ij}, $i \in [d]$, $j \in [4]$, it is no longer linear in the latent variables. However, with

$$\nabla \lambda_i(\boldsymbol{\eta}) = \begin{pmatrix} \alpha_{i1} + \alpha_{i3}\eta_2 \\ \alpha_{i2} + \alpha_{i3}\eta_1 \end{pmatrix}, \quad \boldsymbol{\eta} \in \mathbb{R}^2, \quad i = 1, \ldots, d,$$

the same arguments still apply (compare also to Rizopoulos and Moustaki (2008), who discuss MIRT models within a more general form of the generalized latent variable model, allowing nonlinear effects of latent variables).

7. Main Results

Our main contribution is the generalization of Theorems 1 and 3 of Chang and Stout (1993) for $q > 1$, under the assumptions (CS1') to (CS5'). Furthermore, we embed the CS-approach in the GGS framework (see Theorem 5 (iii)). Similarly to Chang and Stout (1993), the consistency of the MLE is received as a by-product, along with an assertion on its existence. Additionally, the consistency of a penalized MLE is derived. The results are provided in the next theorem, while their proofs along with some preliminary required lemmas are given in appendix.

Theorem 5. *Let $\mathbf{Z} \sim \mathcal{N}_q(\mathbf{0}, \mathsf{I}_q)$ be a q-variate standard normal distributed random vector and $\{Y_i\}_{i \in \mathbb{N}} \sim \mathcal{P}(\boldsymbol{\eta}_0)$ is a sequence of binary response variables for a sequence of item response functions $\{Y_i\}_{i \in \mathbb{N}}$ satisfying (CS1'[i]), (CS2') and (CS3') for $\boldsymbol{\eta}_0 \in \Theta \setminus \partial\Theta$. Then, the following statements holds:*

(i) *There is a sequence $\{\hat{\boldsymbol{\eta}}_d\}_{d \in \mathbb{N}}$ of measurable mappings so that*

$$\lim_{d \to \infty} \mathsf{P}_{\boldsymbol{\eta}_0} \left(\nabla \ell^{(d)}(\hat{\boldsymbol{\eta}}_d \mid \mathbf{Y}^{(d)}) = \mathbf{0} \right) = 1,$$

$$\lim_{d \to \infty} \mathsf{P}_{\boldsymbol{\eta}_0} \left(\ell^{(d)}(\hat{\boldsymbol{\eta}}_d \mid \mathbf{Y}^{(d)}) = \max_{\boldsymbol{\eta} \in \Theta} \ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) \right) = 1$$

and $\hat{\boldsymbol{\eta}}_d \xrightarrow{\mathsf{P}_{\boldsymbol{\eta}_0}} \boldsymbol{\eta}_0$ for $d \to \infty$.

(ii) *Statement (i) remains valid if $\ell^{(d)}$ is replaced by the penalized log-likelihood*

$$\tilde{\ell}^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)})) = \ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) + \log(\mathcal{W}(\boldsymbol{\eta})), \quad \boldsymbol{\eta} \in \Theta, \, d \in \mathbb{N},$$

for some continuously differentiable, positive and bounded function $\mathcal{W}$.

(iii) *If additional (CS1'[ii]), (CS4') and (CS5') are satisfied, then the following statement holds: If $\boldsymbol{\eta}_0$ is held fix, then, for all $B \in \mathcal{B}^q$,*

$$\mathsf{P} \left(\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{1/2} \left(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d \right) \in B \mid \mathbf{Y}^{(d)} \right) \xrightarrow{\mathsf{P}_{\boldsymbol{\eta}_0}} \mathsf{P}(\mathbf{Z} \in B). \tag{26}$$

That is, the MLE $\hat{\boldsymbol{\eta}}_d$ is a semiproper centering (cf. Definition 1). If $\boldsymbol{\eta}_0 \sim \mathcal{G}$, where $\mathcal{G}$ is an absolutely continuous proper distribution with $\mathrm{supp}(\mathcal{G}) \subseteq \Theta$, then furthermore

$$\mathsf{P} \left(\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{1/2} \left(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d \right) \in B \mid \mathbf{Y}^{(d)} \right) \xrightarrow{\mathsf{P}} \mathsf{P}(\mathbf{Z} \in B), \tag{27}$$

for all $B \in \mathcal{B}^q$.

Remark 5. For $\mathcal{W} = \mathfrak{h}$, the penalized MLE in part Theorem 5 (ii) becomes the maximum a-posteriori estimator (MAP), which is an important estimator for $\boldsymbol{\eta}$ in IRT, also because it ensures the existence of estimates in cases the MLE becomes infinite (for example when $\sum_{i=1}^{d} y_i = 0$ or d). The restriction on $\mathfrak{h}$ in part (ii) is stronger than (CS1'[ii]), but still mild.

As already noted in Sect. 3, Theorem 5 (iii) can be used for the construction of credible regions for $\boldsymbol{\eta}$. Additionally, it allows the interpretation of the MLE as a Bayesian estimator of $\boldsymbol{\eta}_0$ and thus enables the use of $\hat{\boldsymbol{\eta}}_d$ to derive some kind of *objective posterior*, in the sense that it is prior-free constructed.

An important concept in the asymptotic analysis of Bayesian procedures is the consistency of the posterior distribution, which forms a basis for the asymptotic validity of inferential methods, and is proved in Theorem 6 (i). The consistency of the EAP is stated in Theorem 6 (ii).

Theorem 6. *Consider the setup and the assumptions of* Theorem 5*(iii), the following statements hold:*

(i) *If $\boldsymbol{\eta}_0$ is held fix, then*

$$\mathsf{P}(\boldsymbol{\eta} \in B \mid \mathbf{Y}^{(d)}) \xrightarrow{\mathsf{P}_{\boldsymbol{\eta}_0}} \delta_{\boldsymbol{\eta}_0}(B) := \begin{cases} 1, & \boldsymbol{\eta}_0 \in B \\ 0, & \boldsymbol{\eta}_0 \notin B \end{cases}, \qquad d \to \infty,$$

for all Borel-sets $B \in \mathcal{B}^q$ with $\boldsymbol{\eta}_0 \notin \partial B$.

(ii) *Suppose that $\boldsymbol{\eta}_0$ is held fix and that there is a continuous mapping $f : \Theta \to \mathbb{R}$ so that $\int_{\Theta} f(\boldsymbol{\eta}) \mathfrak{h}(\boldsymbol{\eta}) \, \mathrm{d}(\boldsymbol{\eta})$ exists. Then, the posterior expected value $\mathsf{E}(f(\boldsymbol{\eta}) \mid \mathbf{Y}^{(d)})$ exists for all $d \in \mathbb{N}$ and is weakly consistent for $f(\boldsymbol{\eta}_0)$, i.e.,*

$$\mathsf{E}(f(\boldsymbol{\eta}) \mid \mathbf{Y}^{(d)}) \xrightarrow{\mathsf{P}_{\boldsymbol{\eta}_0}} f(\boldsymbol{\eta}_0), \qquad \text{for } d \to \infty.$$

If in particular $\mathsf{E}(\boldsymbol{\eta})$ exists, then the posterior expected value $\mathsf{E}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)})$ exists for all $d \in \mathbb{N}$ and is weakly consistent for $\boldsymbol{\eta}_0$.

8. Simulation Study

The simulation study that follows examines the convergence to zero of the error for the approximation of the MLE-centered normalized posterior by a standard normal distribution and its relation to the convergence of the MLE, for the case of a bivariate latent variable vector ($q = 2$). Convergences are evaluated based on the following measures. For the MLE, we use the root-mean-square error

$$\mathrm{RMSE}(\hat{\boldsymbol{\eta}}_d, \boldsymbol{\eta}_0) = \sqrt{\frac{1}{2} \left((\hat{\eta}_{d1} - \eta_{01})^2 + (\hat{\eta}_{d2} - \eta_{02})^2 \right)}.$$

For the approximation of the normalized posterior density h^* by a bivariate normal pdf ϕ_2, we compute the *density approximation error* (also known as L^1-distance)

$$\mathrm{DAE}(h^*, \phi_2) = \int_{\mathbb{R}^2} \left| h^*(\boldsymbol{\eta}) - \phi_2(\boldsymbol{\eta}) \right| \, \mathrm{d}\boldsymbol{\eta},$$

the Hellinger-distance

$$\mathrm{HD}(h^*, \phi_2) = \sqrt{\frac{1}{2} \int_{\mathbb{R}^2} \left(\sqrt{h^*(\boldsymbol{\eta})} - \sqrt{\phi_2(\boldsymbol{\eta})} \right)^2 \, \mathrm{d}\boldsymbol{\eta}},$$

and the Kullback–Leibler divergence

$$\mathrm{KLD}(h^*, \phi_2) = \int_{\mathbb{R}^2} \log \left(\frac{\phi_2(\boldsymbol{\eta})}{h^*(\boldsymbol{\eta})} \right) \phi_2(\boldsymbol{\eta}) \, \mathrm{d}\boldsymbol{\eta}.$$

The simulation study is based on model (22) with the same item parameters across all replications, to mimic the situation that different persons respond on the same test. These are generated as in Sect. 6. For the structural model we assume $\mathcal{H} = \mathcal{N}_2(\mathbf{0}, I_2)$, resulting in $\Theta = \mathbb{R}^2$. The number of items d varies from 10 to 70 in steps of ten items, to mimic the asymptotic behavior with *test lengthening*. All involved integrals are approximated using an importance sampling Monte Carlo (MC) approximation with $\mathcal{N}_2(\mathbf{0}, I_2)$ being the importance distribution.

We replicate 1000 times ($\ell = 1, \ldots, 1000$) the following procedure.

1. Draw $\boldsymbol{\eta}_0^{(\ell)} \sim \mathcal{H}$.
2. Draw $\mathbf{y}^{(70,\ell)} = (y_1^{(\ell)}, \ldots, y_{70}^{(\ell)})$ from model (22) with underlying true latent variable vector $\boldsymbol{\eta}_0^{(\ell)}$ and item parameter values as described above (setting $y_i^{(\ell)} = 1$ if $y_i^{*(\ell)} < P_i(\boldsymbol{\eta}_0^{(\ell)})$ and $y_i = 0$ otherwise, where $y_i^{*(\ell)}$ is drawn from iid $\mathcal{U}(0, 1)$, $i = 1, \ldots, 70$). Then set $\mathbf{y}^{(d,\ell)} = (y_1^{(\ell)}, \ldots, y_d^{(\ell)})$, for $d = 10, 20, \ldots, 70$.
3. Compute the MLE $\hat{\boldsymbol{\eta}}_d^{(\ell)}$ and the test information matrix $\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d^{(\ell)})$, based on $\mathbf{y}^{(d,\ell)}$, for $d = 10, 20, \ldots, 70$.
4. Derive the posterior pdf $h^{*(\ell)}$ of the normalized latent vector $\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d)$, estimating its normalization constant by a MC quadrature.
5. Compute $\mathrm{RMSE}_\ell^{(d)} = \mathrm{RMSE}(\hat{\boldsymbol{\eta}}_d^{(\ell)}, \boldsymbol{\eta}_0^{(\ell)})$, $\mathrm{DAE}_\ell^{(d)} = \mathrm{DAE}(h^{*(\ell)}, \phi_2)$, $\mathrm{HD}_\ell^{(d)} = \mathrm{HD}(h^{*(\ell)}, \phi_2)$ and $\mathrm{KLD}_\ell^{(d)} = \mathrm{KLD}(h^{*(\ell)}, \phi_2)$, for $d = 10, 20, \ldots, 70$.

Our results are visualized in Fig. 3, where the box-plots of the RMS, DAE, HD and KLD values computed above are pictured, for all d values considered. As expected, all evaluation measures and their range are decreasing in d.

Table 2 provides the average values of the evaluation measures, i.e., $\overline{\mathrm{RMSE}}^{(d)} = \frac{\sum_\ell \mathrm{RMSE}_\ell^{(d)}}{1000}$, and $\overline{\mathrm{DAE}}^{(d)}, \overline{\mathrm{HD}}^{(d)}$, and $\overline{\mathrm{KLD}}^{(d)}$, defined analogously. Notice that in our simulation study in case of relatively small number of items ($d \leq 30$), we observed simulation cycles for which the Kullback–Leibler divergence was numerically infinite (due to floating point arithmetics), indicating that the divergence between the two compared distributions for these cases was extremely large. In particular, this occurred in 112 cases for $d = 10$, 20 cases for $d = 20$ and one case for $d = 30$ (out of 1000). These cases were excluded from the calculation of the corresponding average KLD-values reported in Table 2.

Figure 4 visualizes the relation of the divergence measures of the normalized posterior from the standardized normal distribution to the RMSE of the MLE for the values of Table 2, pictured for $d \geq 30$. Observe that as d increases, $\overline{\mathrm{DAE}}^{(d)}$ and $\overline{\mathrm{HD}}^{(d)}$ are linear in $\overline{\mathrm{RMSE}}^{(d)}$, while $\overline{\mathrm{KLD}}^{(d)}$ is linear in $\overline{\mathrm{RMSE}}^{(d)}/\sqrt{d}$. This is an indication that DAE and HD have the same rate of convergence to zero as RMSE while that of KLD is scaled by $d^{-1/2}$.
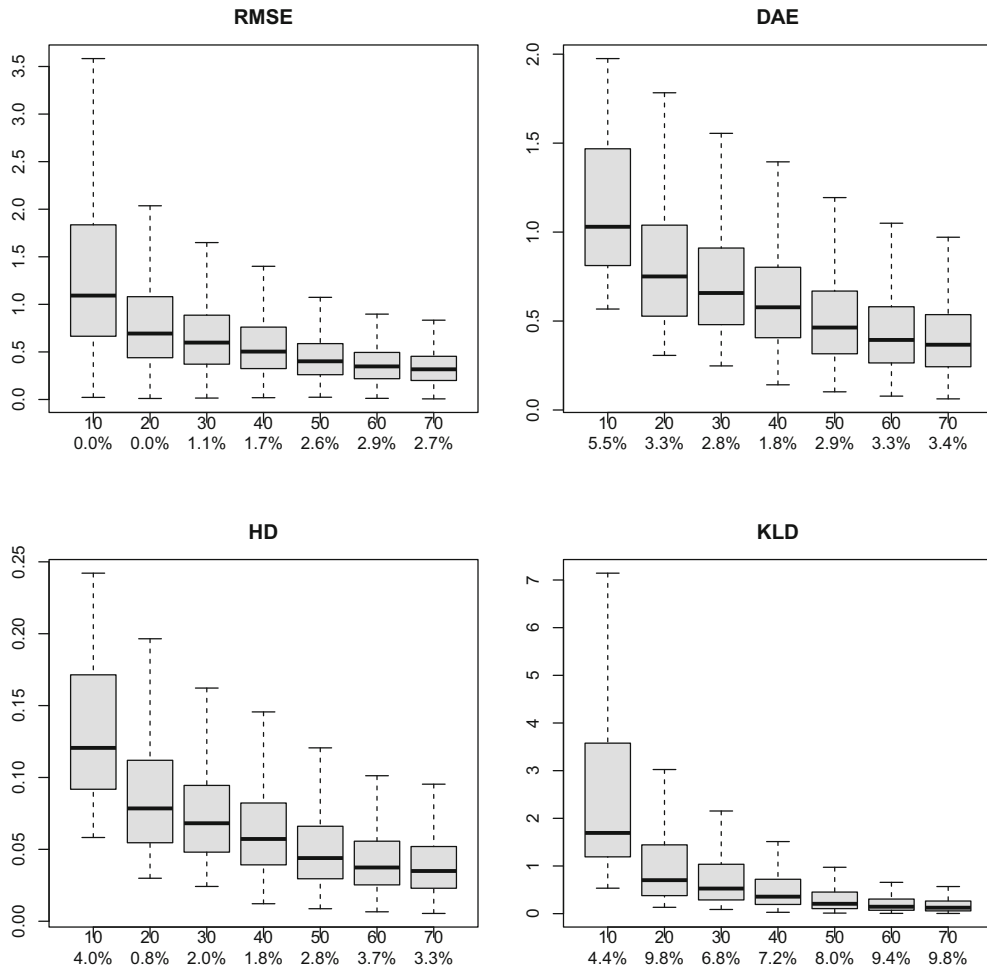
FIGURE 3.
Box-Plots of the RMSE, DAE, HD and KLD simulated values for $d = 10, 20, \ldots, 70$. Under every d value on the horizontal axis, the percentage of points located outside the corresponding whiskers is given.

TABLE 2.
Average values of the root-mean-square error (RMSE) for the MLE along with the density approximation error (DAE), the Hellinger distance (HD) and the Kullback–Leibler divergence (KLD) between the density of the MLE-centered normalized posterior distribution and a bivariate standard normal density, based on 1000 simulations of model (19) with $q = 2$, for different numbers of items d.

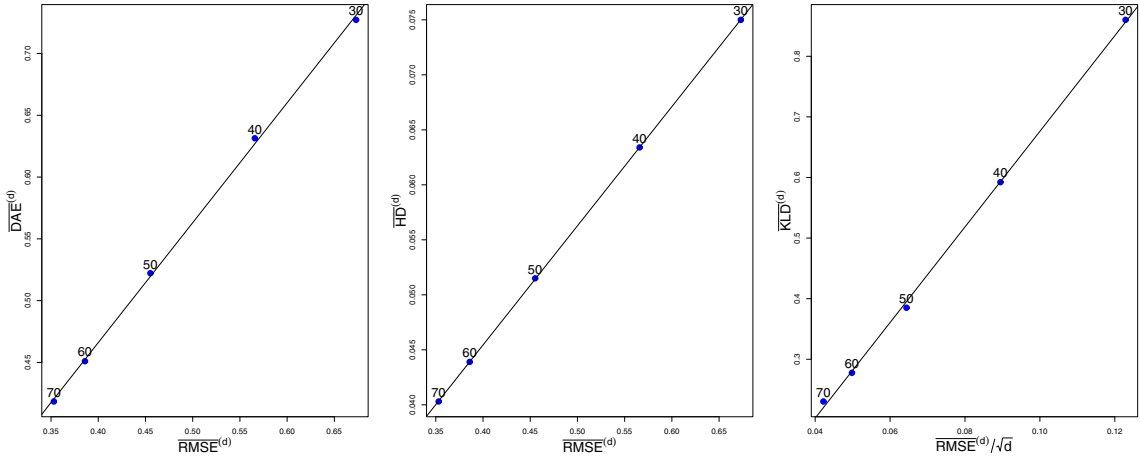| d	$\overline{\mathrm{RMSE}}^{(d)}$	$\overline{\mathrm{DAE}}^{(d)}$	$\overline{\mathrm{HD}}^{(d)}$	$\overline{\mathrm{KLD}}^{(d)}$
10	1.4105	1.1416	0.1330	3.0803
20	0.8206	0.8267	0.0885	1.2698
30	0.6730	0.7272	0.0750	0.8601
40	0.5658	0.6314	0.0634	0.5922
50	0.4552	0.5222	0.0515	0.3850
60	0.3859	0.4510	0.0439	0.2776
70	0.3530	0.4184	0.0403	0.2303

FIGURE 4.
Linear regression of $\overline{\text{DAE}}^{(d)}$ and $\overline{\text{HD}}^{(d)}$ on $\overline{\text{RMSE}}^{(d)}$ and of $\overline{\text{KLD}}^{(d)}$ on $\overline{\text{RMSE}}^{(d)}/\sqrt{d}$ for $d \geq 30$ (s. Table 2).

Notice that the convergence of the DAE to zero in probability is equivalent to proper centering. Thus, our simulation results suggest that the MLE is a proper centering for the multivariate version of the model of Lee and Bolt (2018) with the parameters we consider.

Simulation studies for other IRT models can be conducted similarly, expecting analogous results.

9. Discussion

In this work, we proved the APN of LTs under mild conditions that are fulfilled by a broad class of MIRT models for binary items. Furthermore, we obtained as by-products the existence and consistency of the MLE and the MAP estimator. Note that though the MLE is commonly known as consistent in IRT and MIRT settings, Sinharay (2015) indicated the lack of asymptotic results under milder conditions than some of the usual ones (such as test lengthening by *strictly parallel forms*). Thus, Theorem 5 (i) is a contribution toward this direction.

The distribution $\mathcal{G}$ in Theorem 5 (iii) can be different from $\mathcal{H}$ used in the model. Hence, the asymptotic result above is robust to misspecifications of $\mathcal{H}$ as long as the support is sufficiently large. An interesting task for further investigation, pointed out by one of the reviewers, is the study of the effect of misspecified item response functions.

Under similar mild conditions we provided results on the weak consistency of posterior distributions. In Theorem 6 (ii), we get the existence and consistency of the expected a-posteriori estimator (EAP) for estimating $\boldsymbol{\eta}_0$ as well as for estimating $f(\boldsymbol{\eta}_0)$. To the best of our knowledge, a proof of these properties in such a general setup and under comparably mild or milder conditions on the MIRT model does not exist in the related literature.

Our results are under the assumption of a proper prior $\mathfrak{h}$. This is appropriate in IRT settings, where the prior is a model of the population distribution of the latent traits. However, in a Bayesian framework, improper priors can also be considered. If this is the case, the proper prior assumption in (CS1'[ii]) can be replaced by the following condition if the posterior is still proper

$$\frac{1}{P^{(d)}(\mathbf{Y}^{(d)} \mid \boldsymbol{\eta}_0)} \int_{\Theta \setminus B_\delta(\boldsymbol{\eta}_0)} P^{(d)}(\mathbf{Y}^{(d)} \mid \boldsymbol{\eta})\mathfrak{h}(\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\eta} = o_{\mathsf{P}_{\boldsymbol{\eta}_0}}\left(d^{-q/2}\right), \quad \text{for all} \quad \delta > 0. \quad (28)$$

Condition (28) is sufficient for the derivation of the results stated here and is satisfied by a proper prior.

The APN for a univariate LT for polytomous items was discussed by Chang (1996). The extension of the results for MIRT models with polytomous items is the subject of our current research.

Here, we derived conditions for APN of LTs by generalizing the contribution of Chang and Stout (1993) to $q > 1$. The methodology of GGS/IH, discussed in Sect. 4, provides a general framework for APN in various contexts, including MIRT. The results of Ghosal (1997, 1999) are helpful in deriving alternative conditions for APN tailored for MIRT models.

<div align="center">Acknowledgments</div>

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

<div align="center">Appendix: Proofs of Theorems in Section 7</div>

Here, we prove the main results Theorem 5 and 6 and provide required lemmas for their proof. More detailed versions of all proofs including preliminary results are provided in the web-appendix. For the proof of Theorem 5 (i) the following lemma is required. This lemma ensures that for $d \to \infty$, a global maximum of the log-likelihood has to be in an arbitrarily small area around $\boldsymbol{\eta}_0$, thus being the main step to prove the consistency of the MLE.

Lemma 1. *Consider* $\{Y_i\}_{i \in \mathbb{N}} \sim \mathcal{P}(\boldsymbol{\eta})$ *and assume that conditions (CS1'[i]), (CS2') and (CS3') are satisfied for a fixed* $\boldsymbol{\eta}_0 \in \Theta$. *Then, for any* $\delta > 0$ *there is a* $k(\delta) < 0$ *so that*

$$\lim_{d \to \infty} \mathsf{P}_{\boldsymbol{\eta}_0} \left(\sup_{\boldsymbol{\eta} \in \Theta \setminus B_\delta(\boldsymbol{\eta}_0)} \frac{1}{d} \left(\ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) - \ell^{(d)}(\boldsymbol{\eta}_0 \mid \mathbf{Y}^{(d)}) \right) < k(\delta) \right) = 1.$$

Proof. A more detailed version of the proof can be found in the web-appendix (p. 3). Consider an arbitrary $\delta > 0$. One can show that the associated sequence of item response functions $\{P_i\}_{i \in \mathbb{N}}$ is equicontinuous on each compact set (compare Lemma W.1 in the web-appendix). This implies, applying the strong law of large numbers, that

$$\lim_{d \to \infty} \mathsf{P}_{\boldsymbol{\eta}_0} \left(\sup_{\boldsymbol{\eta} \in K} \frac{1}{d} \left(\ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) - \ell^{(d)}(\boldsymbol{\eta}_0 \mid \mathbf{Y}^{(d)}) \right) < c_K \right) = 1,$$

for each compact $K \subset \Theta$ for which a $\delta > 0$ exists such that $B_\delta(\boldsymbol{\eta}_0) \not\subset K$ with a constant $c_K \leq \sup_{\boldsymbol{\eta} \in K} c(\boldsymbol{\eta})/2 < 0$ (compare the more detailed version of the proof in the web-appendix). This is in particular true for

$$K = \Theta_j := \left\{ \boldsymbol{\eta} \in \Theta \colon \delta + j \leq \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| \leq \delta + j + 1 \right\}$$

for each $j \in \mathbb{N}_0$ and $\delta > 0$. Finally, we get with probability tending to one for $d \to \infty$ that

$$
\begin{aligned}
\sup_{\boldsymbol{\eta} \in \Theta \setminus B_\delta(\boldsymbol{\eta}_0)} &\frac{1}{d} \left(\ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) - \ell^{(d)}(\boldsymbol{\eta}_0 \mid \mathbf{Y}^{(d)}) \right) \\
&= \sup_{j \in \mathbb{N}_0} \left(\sup_{\boldsymbol{\eta} \in \Theta_j} \frac{1}{d} \left(\ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) - \ell^{(d)}(\boldsymbol{\eta}_0 \mid \mathbf{Y}^{(d)}) \right) \right) \\
&\leq \sup_{j \in \mathbb{N}_0} c_j \leq \sup_{\boldsymbol{\eta} \in \Theta \setminus B_\delta(\boldsymbol{\eta}_0)} c(\boldsymbol{\eta})/2 =: 2 \cdot k(\delta).
\end{aligned}
$$

$\square$

Proof of Theorem 5(i)–(ii). A more detailed version of the proof can be found in the web-appendix (p. 7).
Analogously to the proof of Corollary 3.1 of Chang and Stout (1991), notice that

$$\ell^{(d)}(\hat{\boldsymbol{\eta}}_d \mid \mathbf{Y}^{(d)}) - \ell^{(d)}(\boldsymbol{\eta}_0 \mid \mathbf{Y}^{(d)}) = \log \left(\frac{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\boldsymbol{\eta}}_d)}{P^{(d)}(\mathbf{Y}^{(d)} \mid \boldsymbol{\eta}_0)} \right) \geq 0, \qquad \text{(A1)}$$

if the MLE exists due to its definition as global maximum. From Lemma 1 follows, that every global maximum of the log-likelihood has to be in every arbitrary small region around $\boldsymbol{\eta}_0$ with probability tending to one for $d \to \infty$, which implies consistency. The existence of the MLE and further its derivation as solution of the likelihood equations can be shown completely analogous to classical iid cases (e.g., Lehmann & Casella, 1998, Chapter 6, Theorem 5.1, p. 463).
Considering the modified log-likelihood function

$$\tilde{\ell}^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)})) = \ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) + \log(\mathcal{W}(\boldsymbol{\eta})), \qquad \boldsymbol{\eta} \in \Theta, \ d \in \mathbb{N},$$

of part (ii), the consistency is obtained by replacing ℓ by $\tilde{\ell}$ in Lemma 1 and part (i) of this theorem. $\square$

The following lemma ensures the log-likelihood-ratio can be well approximated by a quadratic form of the test information matrix, which is an essential part for the proof of Theorem 5(iii). Lemma 3 and Corollary 1 provided in the sequel, are additionally required for the proof of Theorem 5(iii) and Theorem 6.

Lemma 2. *Suppose that conditions (CS1') through (CS5') hold. Denote by $H_d(\cdot)$ the Hessian matrix of $\ell^{(d)}(\cdot \mid \mathbf{Y}^{(d)})$. Set $\boldsymbol{\Sigma}_d := \mathcal{I}^{(d)}(\boldsymbol{\eta}_0)^{-1}$, which is estimated by $\widehat{\boldsymbol{\Sigma}}_d = \mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{-1}$, if $\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{-1}$ exists, and by $\widehat{\boldsymbol{\Sigma}}_d = \mathrm{I}_q$ otherwise, where I_q is the $q \times q$ identity matrix, $d \in \mathbb{N}$. Then, we have the following.*

1. *There is a sequence* $\{a_d\}_{d \in \mathbb{N}}$, $a_i \in [0, 1]$, *such that for*

$$R_d(\boldsymbol{\eta}) := \widehat{\boldsymbol{\Sigma}}_d \left(\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d) + H_d(\boldsymbol{\eta}_d^*) \right) = \mathrm{I}_q + \mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{-1} H_d(\boldsymbol{\eta}_d^*)$$

where $\boldsymbol{\eta}_d^* := a_d \hat{\boldsymbol{\eta}}_d + (1 - a_d)\boldsymbol{\eta}$, *it holds with probability tending to 1 for* $d \to \infty$, *that*

$$\ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) - \ell^{(d)}(\hat{\boldsymbol{\eta}}_d \mid \mathbf{Y}^{(d)}) = \frac{1}{2}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d)^{\mathsf{T}} H_d(\boldsymbol{\eta}_d^*)(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d) \tag{A2}$$

$$= -\frac{1}{2}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d)^{\mathsf{T}} \mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)(\mathrm{I}_q - R_d(\boldsymbol{\eta}))(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d), \tag{A3}$$

$\boldsymbol{\eta} \in \Theta$.
2. *For any* $\varepsilon > 0$, *there is a* $\delta > 0$ *such that*

$$\lim_{d \to \infty} \mathsf{P}_{\boldsymbol{\eta}_0} \left(\sup_{\boldsymbol{\eta} \in B_\delta(\boldsymbol{\eta}_0)} \|R_d(\boldsymbol{\eta})\| < \varepsilon \right) = 1, \tag{A4}$$

where $\|\boldsymbol{A}\|$ *denotes the spectral norm for a matrix* $\boldsymbol{A}$.
3. *For any* $\varepsilon > 0$, *there is a* $\delta > 0$ *so that for all* $\boldsymbol{\eta} \in B_\delta(\boldsymbol{\eta}_0)$

$$\lim_{d \to \infty} \mathsf{P}_{\boldsymbol{\eta}_0} \bigg((1 + \varepsilon) Q_d(\boldsymbol{\eta}) \le -\frac{1}{2}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d)^{\mathsf{T}} \mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)(\mathrm{I}_q - R_d(\boldsymbol{\eta}))(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d)$$

$$\le (1 - \varepsilon) Q_d(\boldsymbol{\eta}) \bigg) = 1,$$

where

$$Q_d(\boldsymbol{\eta}) := -\frac{1}{2}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d)^{\mathsf{T}} \mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d), \quad \boldsymbol{\eta} \in B_\delta(\boldsymbol{\eta}_0), \ d \in \mathbb{N}.$$

Recall that if a matrix $\boldsymbol{A}$ is symmetric and positive definite, then $\|\boldsymbol{A}\| = \nu_{\max}(\boldsymbol{A})$ and $\|\boldsymbol{A}^{-1}\| = \nu_{\min}(\boldsymbol{A})^{-1}$ hold, where $\nu_{\max}$ and $\nu_{\min}$ denote the largest and smallest eigenvalues of a matrix.

Proof. An extended proof is provided in the web-appendix (p. 14).
Equation (A2) follows directly from a second-order Taylor expansion of $\ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)})$ at $\hat{\boldsymbol{\eta}}_d$. Theorem 5(i) and conditions (CS2') and (CS5') imply the existence of $\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{-1}$ with probability tending to one for $d \to \infty$ (compare Lemma W.6 in the web-appendix) and, therefore, (A3). Condition (CS5') further implies for some constant $C_0 > 0$ and $d \to \infty$ that

$$\|R_d(\boldsymbol{\eta})\| = \left\| \left(\frac{1}{d} \mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d) \right)^{-1} \frac{1}{d} \left(\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d) + H_d(\boldsymbol{\eta}_d^*) \right) \right\|$$

$$\le \left\| \left(\frac{1}{d} \mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d) \right)^{-1} \right\| \cdot \left\| \frac{1}{d} \left(\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d) + H_d(\boldsymbol{\eta}_d^*) \right) \right\|$$

$$\le \frac{1}{C_0} \left\| \frac{1}{d} \left(\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d) + H_d(\boldsymbol{\eta}_d^*) \right) \right\|.$$

One can show that the conditions imposed imply that

$$\left\{\frac{1}{d}\mathcal{I}^{(d)}(\cdot)\right\}_{d\in\mathbb{N}}, \qquad \left\{\frac{1}{d}H_d(\cdot)\right\}_{d\in\mathbb{N}}$$

are equicontinuous in every compact and convex region in Θ (compare Lemma W.4 in the web-appendix on p. 11). Kolmogorov's strong law of large numbers leads then to

$$\|R_d(\boldsymbol{\eta})\| \leq C_1\|\boldsymbol{\eta}_0 - \boldsymbol{\eta}_d^*\| + C_2\|\boldsymbol{\eta}_0 - \hat{\boldsymbol{\eta}}_d\| + o_{\mathsf{P}_{\boldsymbol{\eta}}}(1),$$

for $d \to \infty$ and some appropriate constants $C_1, C_2 > 0$. Since the MLE is consistent and for every $\boldsymbol{\eta} \in B_\delta(\boldsymbol{\eta}_0)$ it holds

$$\|\boldsymbol{\eta}_d^* - \boldsymbol{\eta}_0\| \leq \|\boldsymbol{\eta}_0 - \hat{\boldsymbol{\eta}}_d\| + \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\|,$$

(recall $\boldsymbol{\eta}_d^*$ lies between $\boldsymbol{\eta}$ and $\hat{\boldsymbol{\eta}}_d$), we get for $\varepsilon := 2C_1\delta$

$$\sup_{\boldsymbol{\eta}\in B_\delta(\boldsymbol{\eta}_0)} \|R_d(\boldsymbol{\eta})\| < \varepsilon + o_{\mathsf{P}_{\boldsymbol{\eta}}}(1).$$

Notice that

$$\left|\boldsymbol{x}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{B}\boldsymbol{x}\right| \leq \sqrt{\frac{\nu_{\max}(\boldsymbol{A})}{\nu_{\min}(\boldsymbol{A})}}\|\boldsymbol{B}\|\boldsymbol{x}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{x},$$

for every $\boldsymbol{x} \in \mathbb{R}^q$, symmetric and positive definite $\boldsymbol{A} \in \mathbb{R}^{q\times q}$ and any further matrix $\boldsymbol{B} \in \mathbb{R}^{q\times q}$. The final part follows by selecting $\boldsymbol{A} = \mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)$ and $\boldsymbol{B} = R_d(\boldsymbol{\eta}_d^*)$ for $d \to \infty$. $\square$

Lemma 3. *Let* $\tilde{\Phi}(B) := \mathsf{P}(\boldsymbol{Z} \in B)$ *for all* $B \in \mathcal{B}^q$ *with* $\boldsymbol{Z} \sim \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{\mathsf{I}}_q)$. *Consider a sequence* $\{Y_i\}_{i\in\mathbb{N}}$ *for a fixed* $\boldsymbol{\eta}_0 \in \Theta$, *for which conditions (CS1' [i]), (CS2') through (CS5'), and either (CS1' [ii]) or* (28) *are satisfied. Then, the following holds.*

1. *For every function* f *that is either absolutely bounded by a constant* $c > 0$ *or for which the integral* $\int_{\mathbb{R}^q} f(\boldsymbol{\eta})\mathfrak{h}(\boldsymbol{\eta})\mathrm{d}\boldsymbol{\eta}$ *exists, and for every* $\delta > 0$, *it holds*

$$\frac{\int_{\mathbb{R}^q \setminus B_\delta(\boldsymbol{\eta}_0)} f(\boldsymbol{\eta})P^{(d)}(\boldsymbol{Y}^{(d)} \mid \boldsymbol{\eta})\mathfrak{h}(\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\eta}}{P^{(d)}(\boldsymbol{Y}^{(d)} \mid \hat{\boldsymbol{\eta}}_d)\det(\widehat{\boldsymbol{\Sigma}}_d^{1/2})} \xrightarrow{\mathsf{P}_{\boldsymbol{\eta}_0}} 0, \qquad d \longrightarrow \infty. \tag{A5}$$

2. *Consider a sequence* $\{G_d\}_{d\in\mathbb{N}}$ *with* $G_d\colon (\Theta, \mathcal{B}(\Theta)) \to (\Theta, \mathcal{B}(\Theta))$ *satisfying either*

$$\lim_{d\to\infty} \mathsf{P}_{\boldsymbol{\eta}_0}\left(G_d(B) \subset B_\delta(\boldsymbol{\eta}_0)\right) = 1, \quad \text{for all} \ \ \delta > 0, \tag{A6}$$

or

$$\lim_{d\to\infty} \mathsf{P}_{\boldsymbol{\eta}_0}\left(G_d(B) \supset B_\delta(\boldsymbol{\eta}_0)\right) = 1, \quad \text{for one} \ \ \delta > 0, \tag{A7}$$

for all bounded $B \in \mathcal{B}^q$. Then, for $d \longrightarrow \infty$, it holds

$$\frac{\int_{G_d(B)} P^{(d)}(\mathbf{Y}^{(d)} \mid \boldsymbol{\eta})\mathfrak{h}(\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\eta}}{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\boldsymbol{\eta}}_d)\det(\hat{\boldsymbol{\Sigma}}_d^{1/2})}$$
$$- \tilde{\Phi}\left(\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{1/2}(G_d(B) - \hat{\boldsymbol{\eta}}_d)\right)\mathfrak{h}(\boldsymbol{\eta}_0)(2\pi)^{q/2} = o_{\mathsf{P}_{\boldsymbol{\eta}_0}}(1).$$

In particular, in case of (A7), it holds

$$\tilde{\Phi}\left(\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{1/2}(G_d(B) - \hat{\boldsymbol{\eta}}_d)\right) = \mathsf{P}\left(\mathbf{Z} \in \mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{1/2}(G_d(B) - \hat{\boldsymbol{\eta}}_d)\right) \xrightarrow{\mathsf{P}_{\boldsymbol{\eta}_0}} 1.$$

Proof. A more detailed version of the proof can be found in the web-appendix (p. 17). 1. With regard to the left-hand side of (A5), note that, in terms of the log-likelihood function, it can be written as

$$\frac{\int_{\mathbb{R}^q \setminus B_\delta(\boldsymbol{\eta}_0)} f(\boldsymbol{\eta}) P^{(d)}(\mathbf{Y}^{(d)} \mid \boldsymbol{\eta})\mathfrak{h}(\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\eta}}{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\boldsymbol{\eta}}_d)\det(\widehat{\boldsymbol{\Sigma}}_d^{1/2})} = \exp\left(\ell^{(d)}(\boldsymbol{\eta}_0 \mid \mathbf{Y}^{(d)}) - \ell^{(d)}(\hat{\boldsymbol{\eta}}_d \mid \mathbf{Y}^{(d)})\right)\frac{T_d}{\det\left(\widehat{\boldsymbol{\Sigma}}_d^{1/2}\right)},\tag{A8}$$

where

$$T_d := \int_{\mathbb{R}^q \setminus B_\delta(\boldsymbol{\eta}_0)} f(\boldsymbol{\eta}) \exp\left(\ell^{(d)}(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) - \ell^{(d)}(\boldsymbol{\eta}_0 \mid \mathbf{Y}^{(d)})\right)\mathfrak{h}(\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\eta},$$

while it always fulfills

$$\left|\frac{\int_{\mathbb{R}^q \setminus B_\delta(\boldsymbol{\eta}_0)} f(\boldsymbol{\eta}) P^{(d)}(\mathbf{Y}^{(d)} \mid \boldsymbol{\eta})\mathfrak{h}(\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{\eta}}{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\boldsymbol{\eta}}_d)\det(\widehat{\boldsymbol{\Sigma}}_d^{1/2})}\right| \leq \left|\frac{T_d}{\det\left(\widehat{\boldsymbol{\Sigma}}_d^{1/2}\right)}\right|.$$

If $\mathcal{H}$ is improper and f is bounded by a constant, (28) directly implies

$$\left|\frac{T_d}{\det\left(\widehat{\boldsymbol{\Sigma}}_d^{1/2}\right)}\right| = o_{\mathsf{P}_{\boldsymbol{\eta}_0}}(1).$$

In any other case, Lemma 1 leads to

$$\lim_{d \to \infty} \mathsf{P}_{\boldsymbol{\eta}_0}\left(|T_d| \leq \mathsf{E}(|f(\boldsymbol{\eta})|)\exp(dc(\delta))\right) = 1.$$

Finally, from the polynomial grows of $\det\left(\widehat{\boldsymbol{\Sigma}}_d^{1/2}\right)^{-1}$, it follows

$$\frac{\mathsf{E}(|f(\boldsymbol{\eta})|)\exp(dc(\delta))}{\det\left(\widehat{\boldsymbol{\Sigma}}_d^{1/2}\right)} \xrightarrow{\mathsf{P}_{\boldsymbol{\eta}_0}} 0.$$

2. Let $B \in \mathcal{B}^q$ be an arbitrary bounded Borel set and define for $\delta > 0$ and $d \in \mathbb{N}$ the set

$$M_{\delta,d} := B_\delta(\eta_0) \cap G_d(B)$$

and the integral

$$V_d := \int_{M_{\delta,d}} P^{(d)}(\mathbf{Y}^{(d)} \mid \eta) \mathfrak{h}(\eta) \, \mathrm{d}\eta.$$

Using the definition of $R_d(\eta)$ in Lemma 2, it holds

$$
\begin{aligned}
&\frac{V_d}{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\eta}_d) \det\left(\widehat{\mathbf{\Sigma}}_d^{1/2}\right)} \\
&= \frac{\mathfrak{h}(\eta_0)}{\det\left(\widehat{\mathbf{\Sigma}}_d^{1/2}\right)} \int_{M_{\delta,d}} \frac{\mathfrak{h}(\eta)}{\mathfrak{h}(\eta_0)} \exp\left(-\frac{1}{2}(\eta - \hat{\eta}_d)^\mathsf{T} \mathcal{I}^{(d)}(\hat{\eta}_d)(I - R_d(\eta))(\eta - \hat{\eta}_d) \right) \mathrm{d}\eta.
\end{aligned}
\tag{A9}
$$

By (CS1'), i.e., the continuity of $\mathfrak{h}$ and $\mathfrak{h}(\eta_0) > 0$, it follows that for every $\varepsilon_1 > 0$, it exists a $\delta_1 > 0$, such that

$$1 - \varepsilon_1 \leq \inf_{\eta \in M_{\delta_1,d}} \frac{\mathfrak{h}(\eta)}{\mathfrak{h}(\eta_0)} \leq \sup_{\eta \in M_{\delta_1,d}} \frac{\mathfrak{h}(\eta)}{\mathfrak{h}(\eta_0)} \leq 1 + \varepsilon_1. \tag{A10}$$

Furthermore, by Lemma 2 we get for any $\varepsilon_2 > 0$ and appropriate $\delta_2 = \delta_2(\varepsilon_2) > 0$:

$$
\begin{aligned}
&(1 - o_{\mathsf{P}_{\eta_0}}(1)) \int_{M_{\delta_2,d}} \exp\left(-\frac{1 + \varepsilon_2}{2}(\eta - \hat{\eta}_d)^\mathsf{T} \mathcal{I}^{(d)}(\hat{\eta}_d)(\eta - \hat{\eta}_d) \right) \mathrm{d}\eta \\
&\leq \int_{M_{\delta_2,d}} \exp\left(-\frac{1}{2}(\eta - \hat{\eta}_d)^\mathsf{T} \mathcal{I}^{(d)}(\hat{\eta}_d)(I - R_d(\eta_d^*))(\eta - \hat{\eta}_d) \right) \mathrm{d}\eta \\
&\leq (1 + o_{\mathsf{P}_{\eta_0}}(1)) \int_{M_{\delta_2,d}} \exp\left(-\frac{1 - \varepsilon_2}{2}(\eta - \hat{\eta}_d)^\mathsf{T} \mathcal{I}^{(d)}(\hat{\eta}_d)(\eta - \hat{\eta}_d) \right) \mathrm{d}\eta.
\end{aligned}
\tag{A11}
$$

Next, (A10) and (A11) imply

$$
\begin{aligned}
&\tilde{\Phi}\left(\sqrt{1 + \varepsilon_2} \mathcal{I}^{(d)}(\hat{\eta}_d)^{1/2} (M_{\delta_2,d} - \hat{\eta}_d) \right) \mathfrak{h}(\eta_0)(2\pi)^{q/2} \frac{1 - \varepsilon_1}{(1 + \varepsilon_2)^{q/2}} \left(1 - o_{\mathsf{P}_{\eta_0}}(1) \right) \\
&\leq \frac{V_d}{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\eta}_d) \det\left(\widehat{\mathbf{\Sigma}}_d^{1/2}\right)} \\
&\leq \tilde{\Phi}\left(\sqrt{1 - \varepsilon_2} \mathcal{I}^{(d)}(\hat{\eta}_d)^{1/2} (M_{\delta_2,d} - \hat{\eta}_d) \right) \mathfrak{h}(\eta_0)(2\pi)^{q/2} \frac{1 + \varepsilon_1}{(1 - \varepsilon_2)^{q/2}} \left(1 + o_{\mathsf{P}_{\eta_0}}(1) \right).
\end{aligned}
\tag{A12}
$$

In the case of (A6), it holds $\lim_{d \to \infty} \mathsf{P}_{\eta_0}\left(G_d(B) = M_{\delta_2,d} \right) = 1$. Selecting ε_1 and ε_2 arbitrarily small leads to

$$\frac{\int_{G_d(B)} P^{(d)}(\mathbf{Y}^{(d)} \mid \eta) \mathfrak{h}(\eta) \, \mathrm{d}\eta}{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\eta}_d) \det(\hat{\mathbf{\Sigma}}_d^{1/2})} = \tilde{\Phi}(\mathcal{I}^{(d)}(\hat{\eta}_d)^{1/2}(G_d(B) - \hat{\eta}_d)) \mathfrak{h}(\eta_0)(2\pi)^{q/2} + o_{\mathsf{P}_{\eta_0}}(1).$$

In the case of equation (A7), we get for each $\delta_2 < \delta$: $\lim_{d \to \infty} \mathsf{P}_{\boldsymbol{\eta}_0} \big(B_{\delta_2}(\boldsymbol{\eta}_0) = M_{\delta_2, d} \big) = 1$. Condition (CS5') implies

$$\tilde{\Phi} \left(\sqrt{1 + \varepsilon_2} \mathcal{I}^{(d)} (\hat{\boldsymbol{\eta}}_d)^{1/2} \big(B_{\delta_2}(\boldsymbol{\eta}_0) - \hat{\boldsymbol{\eta}}_d \big) \right) \xrightarrow{\mathsf{P}_{\boldsymbol{\eta}_0}} 1.$$

Finally, the further valid selection of arbitrary small $\varepsilon_1, \varepsilon_2 > 0$ in (A12) and the application of Lemma 3(1.) on $f = \mathbb{1}_{G_d(B) \setminus B_{\delta_2}(\boldsymbol{\eta}_0)}$ completes the proof. $\square$

For $G_d(B) := \mathbb{R}^q$, Lemma 3(2.) leads directly to the following Corollary.

Corollary 1. *Suppose a sequence* $\{Y_i\}_{i \in \mathbb{N}}$ *for a fixed* $\boldsymbol{\eta}_0 \in \Theta$, *for which conditions (CS1') through (CS5') hold. Then holds for* $d \to \infty$

$$\left(\frac{P(\mathbf{Y}^{(d)})}{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\boldsymbol{\eta}}_d) \det \big(\widehat{\boldsymbol{\Sigma}}_d^{1/2} \big)} \right)^{-1} \xrightarrow{\mathsf{P}_{\boldsymbol{\eta}_0}} \frac{1}{\mathfrak{h}(\boldsymbol{\eta}_0)(2\pi)^{q/2}}.$$

Proof of Theorem 5(iii). An extended proof can be found in the web-appendix (p. 21). Analogously to the proof of Lemma 2, we can assume without loss of generality that $\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{-1/2}$ exists. Set

$$G_d(B) := \left\{ \widehat{\boldsymbol{\Sigma}}_d^{1/2} \boldsymbol{x} + \hat{\boldsymbol{\eta}}_d \; : \; \boldsymbol{x} \in B \right\} \equiv \mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{-1/2} B + \hat{\boldsymbol{\eta}}_d, B \in \mathcal{B}^q, \; d \in \mathbb{N}.$$

Then, (26) for bounded B follows directly from the reformulation

$$\mathsf{P} \left(\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{1/2} \big(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d \big) \in B \; \Big| \; \mathbf{Y}^{(d)} \right)$$

$$= \frac{\int_{G_d(B)} P^{(d)}(\mathbf{Y}^{(d)} \mid \boldsymbol{\eta}) \mathfrak{h}(\boldsymbol{\eta}) \, d\boldsymbol{\eta}}{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\boldsymbol{\eta}}_d) \det \big(\widehat{\boldsymbol{\Sigma}}_d^{1/2} \big)} \left(\frac{P(\mathbf{Y}^{(d)})}{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\boldsymbol{\eta}}_d) \det \big(\widehat{\boldsymbol{\Sigma}}_d^{1/2} \big)} \right)^{-1}$$

due to Lemma 3 part 2 and Corollary 1. The case of unbounded B can be shown by considering a decomposition $B = \bigcup_{m \in \mathbb{N}} B_m$ for bounded and pairwise disjoint $\{B_m\}_{m \in \mathbb{N}}$ (compare to the web-appendix, p. 21–22). Next, set for all $d \in \mathbb{N}, \varepsilon > 0$ and $\boldsymbol{\eta}' \in \Theta$

$$H_{d,\varepsilon}(\boldsymbol{\eta}') := \mathsf{P} \left(\Big| \int_{G_d(B)} h(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) \, d\boldsymbol{\eta} - \int_B \phi(\boldsymbol{x}) \, d(\boldsymbol{x}) \Big| > \varepsilon \; \Big| \; \boldsymbol{\eta}_0 = \boldsymbol{\eta}' \right),$$

where ϕ is the pdf of $\mathcal{N}_q(\mathbf{0}, \mathbf{I}_q)$. Then, (27) follows from

$$\lim_{d \to \infty} \mathsf{P} \left(\Big| \int_{G_d(B)} h(\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) \, d\boldsymbol{\eta} - \int_B \phi(\boldsymbol{x}) \, d(\boldsymbol{x}) \Big| > \varepsilon \right) = \lim_{d \to \infty} \int_\Theta H_{d,\varepsilon} \, d\mathcal{G}$$

$$= \int_\Theta \lim_{d \to \infty} H_{d,\varepsilon} \, d\mathcal{G} = 0,$$

for each $\varepsilon > 0$, which is valid due to (26) and Lebesgue's theorem of dominated convergence. $\square$

Proof of Theorem 6. A more detailed proof is provided in the web-appendix (p. 23).
Part (i) follows directly from Lemma 3 and Corollary 1 by using the reformulation

$$\mathcal{H}(B \mid \mathbf{Y}^{(d)}) = \frac{\int_B P^{(d)}(\mathbf{Y}^{(d)} \mid \boldsymbol{\eta})\mathfrak{h}(\boldsymbol{\eta})\, d\boldsymbol{\eta}}{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\boldsymbol{\eta}}_d) \det\left(\widehat{\boldsymbol{\Sigma}}_d^{1/2}\right)} \left(\frac{P(\mathbf{Y}^{(d)})}{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\boldsymbol{\eta}}_d) \det\left(\widehat{\boldsymbol{\Sigma}}_d^{1/2}\right)}\right)^{-1}, \quad \text{(A13)}$$

for an arbitrary $B \in \mathcal{B}^q$ with $\boldsymbol{\eta}_0 \notin \partial B$.
Next, we prove part (ii). In a first step, the existence of $\mathsf{E}(f(\boldsymbol{\eta}) \mid \mathbf{Y}^{(d)})$ for all functions $f : \Theta \to \mathbb{R}$, which are continuous and for which the integral $\int_\Theta f(\boldsymbol{\eta})\mathfrak{h}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ exists, will be proved. In a second step its consistency for $f(\boldsymbol{\eta}_0)$ will be discussed.
For every $d \in \mathbb{N}$, it holds

$$\int_\Theta |f(\boldsymbol{\eta})|\mathcal{H}(d\boldsymbol{\eta} \mid \mathbf{y}^{(d)}) = \frac{\int_\Theta |f(\boldsymbol{\eta})| P^{(d)}(\mathbf{y}^{(d)} \mid \boldsymbol{\eta})\mathfrak{h}(\boldsymbol{\eta})\, d\boldsymbol{\eta}}{P^{(d)}(\mathbf{y}^{(d)})} \leq \frac{\int_\Theta |f(\boldsymbol{\eta})|\mathfrak{h}(\boldsymbol{\eta})\, d\boldsymbol{\eta}}{P^{(d)}(\mathbf{y}^{(d)})} < \infty,$$

for all $\mathbf{y}^{(d)} \in \{0, 1\}^d$, since $P^{(d)}(\mathbf{y}^{(d)} \mid \boldsymbol{\eta}) \in (0, 1)$, $P^{(d)}(\mathbf{y}^{(d)})$ is positive and independent of $\boldsymbol{\eta} \in \Theta$, and $\int_\Theta f(\boldsymbol{\eta})\mathfrak{h}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ exists if and only if $\int_\Theta f(\boldsymbol{\eta})\mathfrak{h}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ exists. Hence, $\mathsf{E}(f(\boldsymbol{\eta}) \mid \mathbf{Y}^{(d)})$ exists. Furthermore, it remains integrable for $d \to \infty$, as shown next. Notice that the last statement does not follow directly, since $P^{(d)}(\mathbf{y}^{(d)}) \longrightarrow 0$ for any sequence $\{y_i\}_{i \in \mathbb{N}}$ and $d \longrightarrow \infty$.
Adjusting representation (A13), we have

$$\int_\Theta |f(\boldsymbol{\eta})|\mathcal{H}(d\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) = \frac{\int_\Theta |f(\boldsymbol{\eta})| P^{(d)}(\mathbf{Y}^{(d)} \mid \boldsymbol{\eta})\mathfrak{h}(\boldsymbol{\eta})\, d\boldsymbol{\eta}}{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\boldsymbol{\eta}}_d) \det(\widehat{\boldsymbol{\Sigma}}_d)^{1/2}} \left(\frac{P^{(d)}(\mathbf{Y}^{(d)})}{P^{(d)}(\mathbf{Y}^{(d)} \mid \hat{\boldsymbol{\eta}}_d) \det\left(\widehat{\boldsymbol{\Sigma}}_d\right)^{1/2}}\right)^{-1}.$$

Lemma 3 and Corollary 1 imply

$$\lim_{d \to \infty} \mathsf{P}_{\boldsymbol{\eta}_0}\left(\int_\Theta |f(\boldsymbol{\eta})|\mathcal{H}(d\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) < C_1\right) = 1,$$

with $\sup_{\boldsymbol{\eta} \in B_\delta(\boldsymbol{\eta}_0)} |f(\boldsymbol{\eta})| =: C_1 < \infty$ for an arbitrary $\delta > 0$.
Further, for every $\delta > 0$ it holds

$$\left|\int_\Theta f(\boldsymbol{\eta})\mathcal{H}(d\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) - f(\boldsymbol{\eta}_0)\right| = \left|\int_\Theta f(\boldsymbol{\eta})\mathcal{H}(d\boldsymbol{\eta} \mid \mathbf{Y}^{(d)}) - f(\boldsymbol{\eta}_0)\mathcal{H}(\Theta \mid \mathbf{Y}^{(d)})\right|$$

$$\leq \left|\int_{B_\delta(\boldsymbol{\eta}_0)} \left(f(\boldsymbol{\eta}) - f(\boldsymbol{\eta}_0)\right)\mathcal{H}(d\boldsymbol{\eta} \mid \mathbf{Y}^{(d)})\right| + \left|\int_{\Theta \setminus B_\delta(\boldsymbol{\eta}_0)} f(\boldsymbol{\eta})\mathcal{H}(d\boldsymbol{\eta} \mid \mathbf{Y}^{(d)})\right|$$

$$+ |f(\boldsymbol{\eta}_0)| \cdot \mathcal{H}(\Theta \setminus B_\delta(\boldsymbol{\eta}_0) \mid \mathbf{Y}^{(d)}),$$

where the last two terms converge to zero in probability by Lemma 3, Corollary 1 and part (i). Last, the continuity of f implies for each $\varepsilon > 0$ and appropriate $\delta' = \delta'(\varepsilon)$:

$$\left|\int_{B_{\delta'}(\boldsymbol{\eta}_0)} \left(f(\boldsymbol{\eta}) - f(\boldsymbol{\eta}_0)\right)\mathcal{H}(d\boldsymbol{\eta} \mid \mathbf{Y}^{(d)})\right| \leq \varepsilon \cdot \mathcal{H}(B_{\delta'}(\boldsymbol{\eta}_0) \mid \mathbf{Y}^{(d)}) \leq \varepsilon.$$

The second part follows directly by considering the mappings $\boldsymbol{\eta} \mapsto \eta_j$, $j \in \{1, \ldots, q\}$, in the first part, which are continuous and by assumption integrable. □

References

Anderson, C. J., Li, Z., & Vermunt, J. K. (2007). Estimation of models in a Rasch family for polytomous items and multiple latent variables. *Journal of Statistical Software, 20*(6), 1–36.

Anderson, C. J., & Vermunt, J. K. (2000). Log-multiplicative association models as latent variable models for nominal and/or ordinal data. *Sociological Methodology, 30,* 81–121.

Anderson, C. J., & Yu, H.-T. (2017). Properties of Second-Order Exponential Models as Multidimensional Response Models. In L. A. van der Ark, M. Wiberg, S. Culpepper, J. A. Douglas, & W. C. Wang (Eds.), *Quantitative Psychology. IMPS 2016. Springer Proceedings in Mathematics & Statistics* (Vol. 196). Springer.

Chang, H.-H. (1996). The asymptotic posterior normality of the latent trait for polytomous IRT models. *Psychometrika, 61*(3), 445–463.

Chang, H.-H., & Stout, W. (1991). The asymptotic posterior normality of the latent trait in an IRT model. Technical Report ONR Research Report 91-4, Department of Statistics, University of Illinois at Urbana-Champaign.

Chang, H.-H., & Stout, W. (1993). The asymptotic posterior normality of the latent trait in an IRT model. *Psychometrika, 58*(1), 37–52.

Ghosal, S. (1997). Normal approximation to the posterior distribution for generalized linear models with many covariates. *Mathematical Methods of Statistics, 6,* 332–348.

Ghosal, S. (1999). Asymptotic normality of posterior distributions in high dimensional linear models. *Bernoulli, 5,* 315–331.

Ghosal, S., Ghosh, J. K., & Samanta, T. (1995). On convergence of posterior distributions. *The Annals of Statistics, 23*(6), 2145–2152.

Ghosh, J. K., Ghosal, S., & Samanta, T. (1994). Stability and convergence of the posterior in non-regular problems. *Statistical Decision Theory and Related Topics V* (pp. 183–199). Springer.

Hessen, D. J. (2012). Fitting and testing conditional multinormal partial credit models. *Psychometrika, 77*(4), 693–709.

Holland, P. W. (1990). The Dutch identity: A new tool for the study of item response models. *Psychometrika, 55*(1), 5–18.

Ibragimov, I. A., & Has'minskii, R. Z. (1981). *Statistical estimation: Asymptotic theory*. Springer.

Kornely, M. J. K. (2021). *Multidimensional Modeling and Inference of Dichotomous Item Response Data*. PhD thesis, RWTH Aachen University, Germany.

Lee, S., & Bolt, D. M. (2018). An alternative to the 3pl: Using asymmetric item characteristic curves to address guessing effects. *Journal of Educational Measurement, 55*(1), 90–111.

Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation* (2nd ed.). Springer.

Li, Z. (2010). *Loglinear models as item response models*. PhD thesis, University of Illinois at Urbana-Champaign.

Lord, F. M. (1983). Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika, 48*(2), 233–245.

Paek, Y. (2016). *Pseudo-Likelihood Estimation of Multidimensional Polytomous Item Response Theory Models*. PhD thesis, University of Illinois at Urbana-Champaign.

Pelle, E., Hesse, D., & van der Heijden, P. G. M. (2016). A log-linear multidimensional rasch model for capture-recapture. *Statistics in Medicine, 35,* 622–634.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal, 2*(1), 1–21.

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software, 17*(5), 1–25.

Rizopoulos, D., & Moustaki, I. (2008). Generalized latent variable models with non-linear effects. *British Journal of Mathematical and Statistical Psychology, 61*(2), 415–438.

Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika, 70*(3), 533–555.

Sinharay, S. (2015). The asymptotic distribution of ability estimats: Beyond dichotomous items and unidimensional IRT models. *Journal of Educational and Behavioral Statistics, 40*(5), 511–528.

Walker, A. M. (1969). On the asymptotic behaviour of posterior distributions. *Journal of the Royal Statistical Society. Series B (Methodological), 31*(1), 80–88.