



# A Note on the Asymptotic Posterior Normality of Multivariate Latent Traits in an IRT Model for Polytomous Items of Mixed Format

Mia Johanna Katharina Kornely

Maria Kateri 

RWTH Aachen University

*The asymptotic posterior normality (APN) of the latent vector in an item response theory model is an important argument in modeling and inference. For a single latent trait, Chang and Stout proved its APN for binary items under general conditions, which generalized Chang for polytomous data and Kornely and Kateri for multivariate latent traits (MLT) and binary items. As MLT and polytomous items are nowadays common in psychometry, an APN-theory covering both simultaneously remains an open and ongoing problem. We generalize the APN-theory accordingly, providing thus a broader foundation for developments relying on APN. We also prove consistency of common estimators for the MLT, extending the according results of Chang and Kornely and Kateri.*

**Keywords:** multidimensional item response theory; polytomous data; posterior distribution; ability estimation; normal approximation

## 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 or exhibits certain properties, resulting from posterior normality. Fortunately, in several cases the abilities distribution can be proven to be asymptotic posterior normal, allowing the approximation of the posterior by a normal distribution, as shown by Chang and Stout (1993) for univariate latent traits (LTs) and binary items under general assumptions. Chang (1996) extended their results to polytomous items, remaining in the framework of a univariate LT. In the meantime, various IRT and relevant statistical modeling approaches have been developed, improving the fit and interpretability by considering multivariate LTs, assuming normality or asymptotic posterior normality (APN) (cf. e.g.,

Anderson and Vermunt, 2000; Hessen, 2012; Li, 2010; Pelle et al., 2016; Rabe-Hesketh et al., 2002; Schilling and Bock, 2005).

Kornely and Kateri (2022) proved APN for the case of multivariate LTs and binary items, under general assumptions, analog to those of Chang and Stout (1993). The recent work of Wang et al. (2022) regarding adaptive testing with polytomous items and multivariate latent traits (MLT) argues under the conjecture of posterior normality, indicating thus the timeliness need for developing APN theory for the case of polytomous items and multivariate LTs. About the same period, Sinharay (2022) revisits and extends approaches for the practical problem of estimating the passing probability of unfinished dichotomous and mixed format tests, some relying on the APN of the LT's distribution. However, extensions of the analyzed *modified Lord–Wingersky approach* to multidimensional IRT models would require the APN for multivariate LTs in mixed format tests, which has not been proved yet. This need is further underlined in a recent revisit of log-multiplicative association models for IRT setups of mixed format items for multivariate LTs and their connection to IRT models under posterior normality by Anderson et al. (2023). In this note, we show that, similarly to the work of Chang (1996), the results of Kornely and Kateri (2022) can be extended naturally to polytomous items with possibly different numbers of response categories per item. It is important to note that the results for binary items cannot be directly extended to models for polytomous items, due to the fact that, though polytomous data can be reformulated to respective sets of binary items, these binary items contain local dependencies that prevent the application of the theorems of Kornely and Kateri (2022). Hence, it is required to formulate a new proof. The general approach along with some advanced technical steps based on proved properties of the log-likelihood function can be transferred. However, the preliminary results establishing these properties cannot directly be carried over and have to be shown. For example, this involves proving certain bounds of the log-likelihood-ratio outside some neighborhood and quadratic approximability inside of it. With this work the results of Chang (1996) are extended for cases (a) with LTs of higher dimension of LTs and (b) with possibly different numbers of response categories. Furthermore, we provide proofs for the consistency of penalized MLE (maximum likelihood estimator)/MAP (maximum a-posteriori estimator) and EAP (expected a-posteriori estimator), which were not considered. The work of Kornely and Kateri (2022), on the other side, is extended (a) with regard to the response variables, from binary to polytomous ones, and (b) by weakening the requirements on the penalization function for the penalized MLE.

The article is organized as follows. After setting the assumed IRT framework and the notation in Section 2, the required regularity conditions are formulated and briefly explained in Section 3. The main result regarding the APN for multivariate LTs, along with results on the existence and consistency of the MLE, penalized MLE/MAP, the posterior and the EAP as well as the APN with

convergence in manifest probabilities, is given and commented in Section 4. The way the proof for multivariate APN extends from the case of binary items to that of polytomous items is discussed in Appendix 1. Finally, the results are summarized in Section 5.

## 2. Preliminaries

This work extends the results of Kornely and Kateri (2022) to polytomous items, adapting their setup and notation accordingly. The set of positive integer numbers is denoted by  $\mathbb{N}$ , that is,  $\mathbb{N} = \{1, 2, 3 \dots\}$ , and the set of real numbers by  $\mathbb{R}$ . Consider a test consisting of  $d \in \mathbb{N}$  categorical response variables  $Y_i$ ,  $i \in [d] := \{1, \dots, d\}$ , where  $Y_i$  may take a value in  $[m_i]_0 := \{0, \dots, m_i\}$  for  $m_i \in \mathbb{N}$ , each  $k \in [m_i]_0$  representing a response category. Consider further the response vector  $\mathbf{Y}^{(d)} = (Y_1, \dots, Y_d)^\top$ , with superscript  $\top$  denoting the transpose of a vector. The manifest probability for a specific response pattern  $\mathbf{y}^{(d)}$  is given by  $P(\mathbf{y}^{(d)}) = \mathbb{P}(\mathbf{Y}^{(d)} = \mathbf{y}^{(d)})$ . In a multidimensional IRT (MIRT) modeling framework, manifest probabilities are derived via conditioning on a  $q$ -dimensional absolutely continuous latent variable vector  $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_q)^\top \in \Theta \subseteq \mathbb{R}^q$ , with  $q \in \mathbb{N}$ , defined over the same probability space as the binary items with probability density function (pdf)  $h$ . For the moment, we simply consider that the latent space  $\Theta$  is a subset of  $\mathbb{R}^q$ . Later on, further conditions will be imposed on  $\Theta$ , along with some required regularity conditions. In the sequel, to simplify notation,  $\boldsymbol{\eta}$  stands for both, the random vector and its realization; they are distinguished by the context. In particular, the conditional probability of the event  $Y_i = k$ , for some  $k \in \{0, \dots, m_i\}$  and conditional on  $\boldsymbol{\eta}$ , is given by

$$\mathbb{P}(Y_i = k | \boldsymbol{\eta}) = P_{ik}(\boldsymbol{\eta}), \quad \boldsymbol{\eta} \in \Theta, \quad (1)$$

where  $P_i = (P_{i0}, \dots, P_{im_i})^\top$  is defined as  $P_{ik}: \Theta \rightarrow (0, 1)$ ,  $k \in [m_i]_0$ , with  $\sum_{k=0}^{m_i} P_{ik} = 1$ . In MIRT modeling, the assumption of *local independence* is usually imposed on the conditional distribution  $\mathbb{P}(\mathbf{Y}^{(d)} = \mathbf{y}^{(d)} | \boldsymbol{\eta})$ , that is,

$$\mathbb{P}(\mathbf{Y}^{(d)} = \mathbf{y}^{(d)} | \boldsymbol{\eta}) = \prod_{i=1}^d \mathbb{P}(Y_i = y_i | \boldsymbol{\eta}) = \prod_{i=1}^d P_{iy_i}(\boldsymbol{\eta}), \quad \mathbf{y}^{(d)} \in \mathbb{Y}_d, \quad (2)$$

where  $\mathbb{Y}_d := [m_1]_0 \times \dots \times [m_d]_0$  is the space of all possible response patterns to the  $d$  items. To prove APN, we need to state an associate assumption on the sequence of response variables, where we also have to impose the practical and natural restriction that the number of response categories  $m_i + 1$  for each item  $i$  is bounded. If a sequence of categorical response variables  $\{Y_i\}_{i \in \mathbb{N}}$  satisfies Equations 1 and 2 for each  $i \in [d]$  and  $d \in \mathbb{N}$  for some  $\boldsymbol{\eta}_0 \in \Theta$ , and if there is some arbitrary  $m_{\max} \in \mathbb{N}$  such that  $m_i \leq m_{\max}$ ,  $i \in \mathbb{N}$ , for some given sequences  $\{m_i\}_{i \in \mathbb{N}}$  and  $\{\mathbf{P}_i\}_{i \in \mathbb{N}}$ , we write  $\{Y_i\}_{i \in \mathbb{N}} \sim \mathcal{M}(\boldsymbol{\eta}_0)$ . Thus,  $\mathcal{M}$  symbolized the

conditional distribution of the response variables given  $\boldsymbol{\eta}_0$ . This  $\boldsymbol{\eta}_0$  can be thought of as a true underlying but unknown  $q$ -dimensional latent vector that might be drawn from some distribution  $\mathcal{G}$  with support  $\text{supp}(\mathcal{G}) \subseteq \Theta$  prior to the experiment. Note that the latent vector does not have to be drawn from the supposed prior with pdf  $h$  as long as the supposed latent space  $\Theta$  covers the actual latent space  $\text{supp}(\mathcal{G})$ .

Assuming Equation 2, the marginal probability mass function of  $\mathbf{Y}^{(d)}$  is derived as

$$P(\mathbf{y}^{(d)}) = \int \dots \int \left( \prod_{i=1}^d \prod_{k=0}^{m_i} P_{ik}(\boldsymbol{\eta}) \mathbb{1}^{(y_i = k)} \right) h(\boldsymbol{\eta}) d\boldsymbol{\eta}, \quad \mathbf{y}^{(d)} \in \mathbb{Y}_d, \quad (3)$$

where  $\mathbb{1}$  is the indicator function. Analogously to Chang (1996), we use the symbol  $\lambda_{ik}$  to denote logarithm of probabilities for responses to item-response-categories, that is,

$$\lambda_{ik}(\boldsymbol{\eta}) := \log(P_{ik}(\boldsymbol{\eta})), \quad i \in \mathbb{N}, k = 1, \dots, m_i, \boldsymbol{\eta} \in \Theta. \quad (4)$$

Given a realization  $\mathbf{y}^{(d)}$  of  $\mathbf{Y}^{(d)}$ , we estimate the true underlying value of the latent vector  $\boldsymbol{\eta}_0$  by its MLE, denoted by  $\hat{\boldsymbol{\eta}}_d = \hat{\boldsymbol{\eta}}(\mathbf{y}^{(d)})$ . The MLE is obtained by maximizing the log-likelihood function corresponding to Equation 2, given by

$$\ell^{(d)}(\boldsymbol{\eta} | \mathbf{y}^{(d)}) = \log(P(\mathbf{Y}^{(d)} = \mathbf{y}^{(d)} | \boldsymbol{\eta})), \quad \boldsymbol{\eta} \in \Theta. \quad (5)$$

The test (or Fisher) information matrix for a test on  $d$  polytomous items is then given by

$$\mathcal{I}^{(d)}(\boldsymbol{\eta}) = \sum_{i=1}^d \sum_{k=0}^{m_i} \frac{1}{P_{ik}(\boldsymbol{\eta})} \nabla P_{ik}(\boldsymbol{\eta}) \nabla^\top P_{ik}(\boldsymbol{\eta}), \quad \boldsymbol{\eta} \in \Theta, \quad (6)$$

where  $\nabla$  denotes the gradient of a function, that is,  $\nabla^\top P_{ik}(\boldsymbol{\eta}) = \left( \frac{\partial P_{ik}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}_1}, \dots, \frac{\partial P_{ik}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}_q} \right)$ .

Kornely and Kateri (2022) have studied the APN of  $\boldsymbol{\eta}$  for  $d \rightarrow \infty$  for tests with binary response variables, that is, for  $m_i = 1$  for all  $i \in [d]$ . This work extends their results to  $m_i \geq 1$  under the condition that there exists a finite  $m_{\max} \in \mathbb{N}$  such that  $\max_{i \in \mathbb{N}} m_i = m_{\max}$ . Similarly, Chang (1996) has studied the case of polytomous items under the restrictions  $m_i = m$  for all  $i \in \mathbb{N}$  and one-dimensional LT, that is,  $q = 1$ , which will be extended to multivariate LTs ( $q > 1$ ) in this work. Hence, we target the approximation of probabilities of the type

$$P\left(\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{1/2}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d) \in B \mid \mathbf{Y}^{(d)}\right), \quad B \in \mathcal{B}^q, \quad (7)$$

for a sequence  $\{Y_i\}_{i \in \mathbb{N}} \sim \mathcal{M}(\boldsymbol{\eta}_0)$ , where  $\mathcal{B}^q$  denotes the Borel- $\sigma$ -algebra of  $\mathbb{R}^q$ , under conditions that are similar to those of Kornely and Kateri (2022) and Chang (1996), for  $q \geq 1$  and possibly different  $m_i \geq 1$ .

Next, we define the functions  $Z_i : \Theta^2 \rightarrow \mathbb{R}$ ,  $i \in \mathbb{N}$ , needed for the proof of APN in our setup, as follows:

$$Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}') := \prod_{k=0}^{m_i} \left( \frac{P_{ik}(\boldsymbol{\eta})}{P_{ik}(\boldsymbol{\eta}_0)} \right)^{\mathbf{1}_{\{y_i=k\}}}, \quad (\boldsymbol{\eta}, \boldsymbol{\eta}') \in \Theta^2, i \in \mathbb{N}.$$

Note that

$$\mathbb{E}_{\boldsymbol{\eta}_0} \left( \sum_{i=1}^d \log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}') \right) = \mathbb{E}_{\boldsymbol{\eta}_0} (\ell^{(d)}(\boldsymbol{\eta} | \mathbf{Y}^{(d)}) - \ell^{(d)}(\boldsymbol{\eta}' | \mathbf{Y}^{(d)})) \quad (8)$$

is the negative Kullback–Leibler divergence between the conditional distributions of  $\mathbf{Y}^{(d)}$  given  $\boldsymbol{\eta}$  and  $\boldsymbol{\eta}_0$ , respectively, providing the key to ensuring identifiability of  $\boldsymbol{\eta}_0$  in the MIRT model.

### 3. Regularity Conditions for Asymptotic Properties of Latent Vectors

Throughout, we assume that  $\{Y_i\}_{i \in \mathbb{N}} \sim \mathcal{M}(\boldsymbol{\eta}_0)$ , that is,  $\{Y_i\}_{i \in \mathbb{N}}$  are categorical random variables fulfilling Equations 1 and 2, that there is some maximum number  $m_{\max}$  of response categories for all items, and that the true latent vector  $\boldsymbol{\eta}_0$  lies in the interior of the parameter space, that is,  $\boldsymbol{\eta}_0 \in \Theta \setminus \partial\Theta$ . The asymptotic results of Section 4 rely on the following regularity conditions:

- (C1) [i] The set  $\Theta$  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  $\mathfrak{h}(\boldsymbol{\eta}_0) > 0$ .
- (C2)  $P_{ik}$  is thrice continuously differentiable,  $i \in \mathbb{N}$ ,  $k = 0, \dots, m_i$ . If restricted to a compact subset  $K \subseteq \Theta$ , all  $\left| \frac{\partial P_{ik}}{\partial \boldsymbol{\eta}_u} \right|$  and  $\left| \frac{\partial^2 P_{ik}}{\partial \boldsymbol{\eta}_u \partial \boldsymbol{\eta}_j} \right|$  are uniformly bounded for all  $k = 0, \dots, m_i$ ,  $i \in \mathbb{N}$ ,  $1 \leq u, j \leq q$ . Moreover, there is a constant  $0 < \zeta(K) < 1$ , which is independent of  $i \in \mathbb{N}$ , such that

$$\zeta(K) \leq \inf_{(i, \boldsymbol{\eta}) \in \mathbb{N} \times K} \inf_{k \in \{0, \dots, m_i\}} P_{ik}(\boldsymbol{\eta}). \quad (9)$$

- (C3) For each  $\boldsymbol{\eta} \in \Theta$ ,  $\boldsymbol{\eta} \neq \boldsymbol{\eta}_0$ , there is a  $c(\boldsymbol{\eta}) < 0$  such that

$$\limsup_{d \rightarrow \infty} \frac{1}{d} \sum_{i=1}^d E_{\boldsymbol{\eta}_0}(\log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}_0)) = \limsup_{d \rightarrow \infty} \frac{1}{d} E_{\boldsymbol{\eta}_0}(\ell^{(d)}(\boldsymbol{\eta} | \mathbf{Y}^{(d)}) - \ell^{(d)}(\boldsymbol{\eta}_0 | \mathbf{Y}^{(d)})) \leq c(\boldsymbol{\eta}), \tag{10}$$

and if  $\Theta$  is unbounded holds additionally

$$\sup_{\boldsymbol{\eta} \in \Theta \setminus B_\delta(\boldsymbol{\eta}_0)} c(\boldsymbol{\eta}) < 0, \quad \text{for all } \delta > 0, \tag{11}$$

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

(C4) If restricted to any compact set  $K \subseteq \Theta$ , the set of functions

$$\left\{ \left| \frac{\partial^3 P_{ik}}{\partial \boldsymbol{\eta}_b \partial \boldsymbol{\eta}_g \partial \boldsymbol{\eta}_u} \right| : k = 0, \dots, m_i, i \in \mathbb{N}, 1 \leq b, g, u \leq q \right\}$$

is uniformly bounded.

(C5) For all  $\boldsymbol{\eta} \in \Theta$  holds

$$0 < \liminf_{n \rightarrow \infty} \nu_{\min} \left( \frac{1}{n} \sum_{i=1}^n \sum_{k=0}^{m_i} \nabla \lambda_{ik}(\boldsymbol{\eta}) \nabla \lambda_{ik}(\boldsymbol{\eta})^\top \right), \tag{12}$$

where  $\nu_{\min}(A)$  denotes the smallest eigenvalue of some matrix  $A$ .

The regularity conditions are essentially the same as that of Chang and Stout (1993), Chang (1996), and Kornely and Kateri (2022), adjusted for MLT and polytomous items. Thus, we just briefly commend them. Comparing these conditions to respective versions for univariate LTs, notice that under the current multivariate setup, we are required to enforce properties on more abstract subsets of the latent space, whereas intervals were appropriate in the univariate case. When considering binary items, all important information is contained in each item’s log-odds as a function of the LTs (or a single function for the probability of responding one, respectively). For polytomous items, in contrast, this is not possible and a function for each item response category is assumed (plus the sum to one condition for each  $\boldsymbol{\eta} \in \Theta$ ). Hence, the conditions (C2) and (C4) are formulated for these response category probability functions instead of item response probability functions. In condition (C3) this aspect of polytomous items is incorporated in the  $Z_i$ s (or log-likelihood-functions, respectively). Similarly, condition (C5) contains a sum over all categories for each item incorporating all  $\lambda_{ik}$  instead of all item logits. Conditions (C1) to (C3) are required to obtain consistency of the MLE and MAP. Imposing further conditions (C4) and (C5) leads to consistency of the posterior and the EAP as well as the desired APN property. Condition (C1) is a condition on the prior distribution of  $\boldsymbol{\eta}$  and is satisfied by all

usually assumed models like  $\boldsymbol{\eta} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}_q)$ . Condition (9) in (C2) ensures that for each item the probability of responding to some arbitrary category is positive and may only approach zero if  $\|\boldsymbol{\eta}\| \rightarrow \infty$ . If  $\boldsymbol{\eta}$  is held fixed and responses to the items are considered sequentially, the response probabilities for any category can never approach zero. The remaining parts of (C2) and (C4) ensure a sufficient smoothness of the model, are easily verified for usual models, and are thus usually not critical. Condition (C3) is a formulation of the asymptotic identifiability of  $\boldsymbol{\eta}_0$ , cases for which identifiability is not given. For example, regardless of the model,  $d \geq q$  is mandatory for being able to identify  $\boldsymbol{\eta}_0$ . Finally, condition (C5) ensures that the test information is asymptotically regular and that  $\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{-1}$  becomes a valid estimate for  $\text{Cov}(\boldsymbol{\eta} | \mathbf{Y}^{(d)} = \mathbf{y}^{(d)})$  for sufficiently large  $d$ .

#### 4. Main Results

Our main result is the direct extension of the APN for multivariate LTs given binary items of Kornely and Kateri (2022, Theorem 5(c)) to polytomous items. At the same time, we extend Theorem 1 of Chang (1996) to MLT. We formulate this result in the following theorem:

**Theorem 1.** Let  $\mathbf{Z} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}_q)$  be a  $q$ -variate standard normal distributed random vector and  $\{Y_i\}_{i \in \mathbb{N}} \sim \mathcal{M}(\boldsymbol{\eta}_0)$  is a sequence of polytomous response variables satisfying conditions (C1) to (C5) for  $\boldsymbol{\eta}_0 \in \Theta \setminus \partial\Theta$ . Then, for all  $B \in \mathcal{B}^q$ ,

$$P\left(\mathcal{I}^{(d)}(\hat{\boldsymbol{\eta}}_d)^{1/2}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_d) \in B \mid \mathbf{Y}^{(d)}\right) \xrightarrow{P_{\boldsymbol{\eta}_0}} P(\mathbf{Z} \in B). \quad (13)$$

The first key challenge for proving Theorem 4 is to prove that  $\ell^{(d)}(\boldsymbol{\eta} | \mathbf{Y}^{(d)}) - \ell^{(d)}(\boldsymbol{\eta}_0 | \mathbf{Y}^{(d)})$  is asymptotically bounded away from zero for all  $\boldsymbol{\eta}$  outside of any neighborhood of  $\boldsymbol{\eta}_0$  with probability approaching one. We tackle this challenge by first proving a negative limit value for each single  $\boldsymbol{\eta} \neq \boldsymbol{\eta}_0$  using the strong law of large numbers. Afterward, this is extended to regions utilizing local equicontinuity of all  $P_{ik}$ ,  $i \in \mathbb{N}$ ,  $k \in [m_i]_0$ , a property that can be shown using their smoothness (requested in the regularity conditions).

The second key challenge is to prove that  $\ell^{(d)}(\boldsymbol{\eta} | \mathbf{Y}^{(d)}) - \ell^{(d)}(\boldsymbol{\eta}_0 | \mathbf{Y}^{(d)})$  can asymptotically be well approximated by a quadratic form of the information matrix at the MLE  $\hat{\boldsymbol{\eta}}_d$  (i.e., the inverse of the covariance estimate of the posterior distribution) in small neighborhoods of  $\boldsymbol{\eta}_0$ . We address it by using a Taylor expansion and, again, arguments of equicontinuity and laws of large numbers.

The subsequent steps can then follow from proofs for binary items and MLTs. Our general approach is adopted from Chang (1996) and Kornely and Kateri (2022) and adjusted for the combined setup. The proof of Theorem 1 is further discussed in Appendix 1. The APN in Theorem 1 is the semiproper centering of

the MLE, that is, the set-wise convergence of the probabilities of the normalized posterior, centered at the MLE (cf. Definition 2 of Ghosal et al., 1995). If  $\boldsymbol{\eta}_0 \sim \mathcal{G}$  for an absolutely continuous proper distribution  $\mathcal{G}$  with  $\text{supp}(\mathcal{G}) \subseteq \Theta$ , then the convergence with respect to  $P_{\boldsymbol{\eta}_0}$  can be replaced by the convergence with respect to  $P$  in Equation 13, that is, convergence in manifest probabilities.

Several preliminary results and by-products of the proof of Theorem 1 are of own interest due to their usage independently of the APN of LTs. We formulate them in the following theorem, noting the interesting relation of these results to APN:

**Theorem 2.** Let  $\mathbf{Z} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}_q)$  be a  $q$ -variate standard normal distributed random vector. Further, let  $\{Y_i\}_{i \in \mathbb{N}} \sim \mathcal{M}(\boldsymbol{\eta}_0)$  be a sequence of polytomous response variables satisfying conditions (C1[i]), (C2) and (C3) for  $\boldsymbol{\eta}_0 \in \Theta \setminus \partial\Theta$ . Then the following statements hold:

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

$$\lim_{d \rightarrow \infty} P_{\boldsymbol{\eta}_0}(\nabla \ell^{(d)}(\hat{\boldsymbol{\eta}}_d | \mathbf{Y}^{(d)}) = \mathbf{0}) = 1,$$

$$\lim_{d \rightarrow \infty} P_{\boldsymbol{\eta}_0} \left( \ell^{(d)}(\hat{\boldsymbol{\eta}}_d | \mathbf{Y}^{(d)}) = \max_{\boldsymbol{\eta} \in \Theta} \ell^{(d)}(\boldsymbol{\eta} | \mathbf{Y}^{(d)}) \right) = 1$$

and  $\hat{\boldsymbol{\eta}}_d \xrightarrow{P_{\boldsymbol{\eta}_0}} \boldsymbol{\eta}_0$  for  $d \rightarrow \infty$ .

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

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

for some continuously differentiable and positive function  $\mathcal{W}: \Theta \rightarrow (0, \infty)$ .

- (iii) If (C1[ii]), (C4) and (C5) are additionally satisfied and if there is a mapping  $f: \Theta \rightarrow \mathbb{R}$  that is continuous in  $\boldsymbol{\eta}_0$ , bounded in each bounded subset of  $\Theta$  and  $E(f(\boldsymbol{\eta}))$  exists, then the posterior expected value  $E(f(\boldsymbol{\eta}) | \mathbf{Y}^{(d)})$  exists for all  $d \in \mathbb{N}$  and is weakly consistent for  $f(\boldsymbol{\eta}_0)$ , that is,

$$E(f(\boldsymbol{\eta}) | \mathbf{Y}^{(d)}) \xrightarrow{P_{\boldsymbol{\eta}_0}} f(\boldsymbol{\eta}_0), \quad \text{for } d \rightarrow \infty.$$

The proof of Theorem 2 is discussed in the Supplemental Appendix (available in the online version of this article). This second theorem has a few implications. Part (i) ensures the asymptotic existence and consistency of the MLE. Part (ii) is a criterion for consistency of the MAP as we can set  $\mathcal{W} = \mathfrak{h}$  to obtain the MAP.

Part (iii) delivers the consistency of the posterior distribution by considering  $f = \mathbb{1}_B$  for Borel-sets  $B \in \mathcal{B}^q$  with  $\boldsymbol{\eta}_0 \notin \partial B$ , as  $E(\mathbb{1}_B(\boldsymbol{\eta})|\mathbf{Y}^{(d)}) = P(\boldsymbol{\eta} \in B|\mathbf{Y}^{(d)})$ . Moreover, part (iii) implies the consistency of the EAP if  $E(\boldsymbol{\eta})$  exists by setting  $f(\boldsymbol{\eta}) = \boldsymbol{\eta}_j, j \in [q]$ .

### 5. Conclusion

In this work, we proved that, for MIRT models for polytomous items with possibly different numbers of response categories per item, under certain assumptions, latent vectors are asymptotic posterior normal distributed. The assumptions are not restrictive and thus the result applies to a large class of MIRT models.

Furthermore, we also extended the results of Kornely and Kateri (2022) from binary to polytomous items regarding (a) the existence and consistency of the MLE and penalized MLE/MAP (with a further weakening of the requirements), (b) the consistency of the posterior and the EAP, and (c) APN with convergence in  $P_{\boldsymbol{\eta}_0}$  and in manifest probabilities  $P$  instead of  $P_{\boldsymbol{\eta}_0}$ . The results of Chang (1996) for (a) (weak) consistency of the MLE and (b) APN with convergence in  $P_{\boldsymbol{\eta}_0}$  and in manifest probabilities  $P$  are extended here to MLTs, while the consistency of the posterior, MAP and EAP is additionally proved.

The consistency of the MLE is often considered as commonly known. However, we are only aware of asymptotic theoretic results for LTs that either restrict to stricter conditions on the model or to special cases, like binary items or univariate LTs (e.g., Sinharay [2015]; Kornely and Kateri [2022]). Thus, this work contributes to the certainty of the maximum likelihood approach for ability estimation in more general setups.

## Appendix 1

### *Proof of Theorem 1*

Chang (1996) showed how the proof of APN for binary items of Chang and Stout (1993) extends naturally to polytomous items. The same is true in the more general setup of latent vectors. In this section, we show how the proofs for binary items of Kornely and Kateri (2022) extend to polytomous items by focusing on the first step. Details regarding the next steps are provided in the Supplemental Appendix (available in the online version of this article).

An important step in the proof is to show that

$$\limsup_{d \rightarrow \infty} \frac{1}{d} (\ell^{(d)}(\boldsymbol{\eta}|\mathbf{Y}^{(d)}) - \ell^{(d)}(\boldsymbol{\eta}_0|\mathbf{Y}^{(d)})) \leq c(\boldsymbol{\eta}) < 0, \quad P_{\boldsymbol{\eta}_0} - \text{almost surely}, \quad (\text{A1})$$

holds for every  $\boldsymbol{\eta} \in \Theta \setminus \{\boldsymbol{\eta}_0\}$ , where  $c(\boldsymbol{\eta})$  is the constant of condition of (C3) (or (CS3')) of Kornely and Kateri (2022) in the binary case, respectively). Due to Equation 8, the proof of (A1) can follow by applying Kolmogorov's strong law of large numbers (SLLN) for independent but not identically distributed (inid) random variables to the sequence  $\{\log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}_0)\}_{i \in \mathbb{N}}$  of inid random variables. The essential step is to show that

$$\sum_{i=1}^{\infty} \frac{\text{Var}_{\boldsymbol{\eta}_0}(\log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}_0))}{i^2} < \infty, \tag{A2}$$

which is the requirement to apply for the SLLN. The SLLN then implies

$$\frac{1}{d} \sum_{i=1}^d \log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}_0) - \frac{1}{d} \sum_{i=1}^d \text{E}_{\boldsymbol{\eta}_0}(\log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}_0)) \xrightarrow{\text{P}_{\boldsymbol{\eta}_0}\text{-a.s.}} 0, \quad d \rightarrow \infty,$$

which further implies Equation 1 due to Equation 10 or the binary version in condition (CS3') of Kornely and Kateri (2022). Condition (A2) is directly obtained if  $\text{Var}_{\boldsymbol{\eta}_0}(\log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}_0))$  has an upper bound for all  $i$  making it sufficient to derive an appropriate estimate.

For binary variables (Kornely and Kateri, 2022), this can be achieved by noting that

$$\text{Var}_{\boldsymbol{\eta}_0}(\log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}_0)) = \text{Var}_{\boldsymbol{\eta}_0}(Y_i)(\lambda_i(\boldsymbol{\eta}) - \lambda_i(\boldsymbol{\eta}_0))^2 = P_i(\boldsymbol{\eta}_0)(1 - P_i(\boldsymbol{\eta}_0))(\lambda_i(\boldsymbol{\eta}) - \lambda_i(\boldsymbol{\eta}_0))^2$$

holds for each  $i \in \mathbb{N}$ . The binary version of condition (C2) ensures that there is a constant  $M(\boldsymbol{\eta}, \boldsymbol{\eta}_0) < \infty$  for each  $\boldsymbol{\eta} \in \Theta \setminus \{\boldsymbol{\eta}_0\}$  such that

$$\sup_{i \in \mathbb{N}} \text{Var}_{\boldsymbol{\eta}_0}(\log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}_0)) \leq 2 \sup_{i \in \mathbb{N}} |\lambda_i(\boldsymbol{\eta})|^2 + 2 \sup_{i \in \mathbb{N}} |\lambda_i(\boldsymbol{\eta}_0)|^2 =: M(\boldsymbol{\eta}, \boldsymbol{\eta}_0) < \infty.$$

For polytomous items, we first set  $X_{ik} := \mathbb{1}(Y_i = k)$  and obtain

$$\begin{aligned} \text{Var}_{\boldsymbol{\eta}_0}(\log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}_0)) &= \text{Var}_{\boldsymbol{\eta}_0} \left( \sum_{k=0}^{m_i} X_{ik} \log \left( \frac{P_{ik}(\boldsymbol{\eta})}{P_{ik}(\boldsymbol{\eta}_0)} \right) \right) \\ &= \sum_{k=0}^{m_i} \log \left( \frac{P_{ik}(\boldsymbol{\eta})}{P_{ik}(\boldsymbol{\eta}_0)} \right)^2 \text{Var}_{\boldsymbol{\eta}_0}(X_{ik}) \\ &\quad + 2 \sum_{a=1}^{m_i} \sum_{b=0}^{a-1} \log \left( \frac{P_{ia}(\boldsymbol{\eta})}{P_{ia}(\boldsymbol{\eta}_0)} \right) \log \left( \frac{P_{ib}(\boldsymbol{\eta})}{P_{ib}(\boldsymbol{\eta}_0)} \right) \text{Cov}_{\boldsymbol{\eta}_0}(X_{ia}, X_{ib}) \end{aligned}$$

for each  $i \in \mathbb{N}$  and  $\boldsymbol{\eta} \in \Theta \setminus \{\boldsymbol{\eta}_0\}$ . By the definition of  $\lambda_{ik}$  and condition (C2), for each  $\boldsymbol{\eta} \in \Theta$  we have  $M(\boldsymbol{\eta}) := |\log(\zeta(\{\boldsymbol{\eta}\}))| < \infty$  with  $\zeta(\{\boldsymbol{\eta}\})$  of condition (C2) and

$$\left| \log \left( \frac{P_{ia}(\boldsymbol{\eta})}{P_{ia}(\boldsymbol{\eta}_0)} \right) \log \left( \frac{P_{ib}(\boldsymbol{\eta})}{P_{ib}(\boldsymbol{\eta}_0)} \right) \right| \leq M(\boldsymbol{\eta})^2 + 2M(\boldsymbol{\eta})M(\boldsymbol{\eta}_0) + M(\boldsymbol{\eta}_0)^2 =: M'(\boldsymbol{\eta}, \boldsymbol{\eta}_0) \quad (\text{A3})$$

for all  $0 \leq a, b \leq m_i$  and all  $i \in \mathbb{N}$ . As  $m_i < m_{\max}$  for all  $i \in \mathbb{N}$  by assumption, we obtain the required estimate for each  $\boldsymbol{\eta} \in \Theta \setminus \{\boldsymbol{\eta}_0\}$ :

$$\sup_{i \in \mathbb{N}} \text{Var}_{\boldsymbol{\eta}_0}(\log Z_i(\boldsymbol{\eta}, \boldsymbol{\eta}_0)) \leq (m_{\max} + m_{\max}(m_{\max} + 1))M'(\boldsymbol{\eta}, \boldsymbol{\eta}_0) < \infty.$$

### Acknowledgments

The authors sincerely thank the reviewers for their constructive and useful comments on an earlier version of the manuscript.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

Maria Kateri  <https://orcid.org/0000-0002-6746-7858>

### References

- Anderson, C. J., Kateri, M., & Moustaki, I. (2023). *Log-linear and log-multiplicative association models for categorical data* (pp. 1–42). Springer International Publishing.
- 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.
- 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. (1993). The asymptotic posterior normality of the latent trait in an IRT model. *Psychometrika*, 58(1), 37–52.
- Ghosal, S., Ghosh, J. K., & Samanta, T. (1995). On convergence of posterior distributions. *The Annals of Statistics*, 23(6), 2145–2152.
- Hessen, D. J. (2012). Fitting and testing conditional multinormal partial credit models. *Psychometrika*, 77(4), 693–709.
- Kornely, M. J. K., & Kateri, M. (2022). Asymptotic posterior normality of multivariate latent traits in an IRT model. *Psychometrika*, 87(3), 1146–1172.

- Li, Z. (2010). *Loglinear models as item response models* [PhD thesis, University of Illinois at Urbana-Champaign].
- Pelle, E., Hessen, D. J., & van der Heijden, P. G. M. (2016). A log-linear multidimensional rasch model for capture–recapture. *Statistics in Medicine*, *35*(4), 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.
- 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.
- Sinharay, S. (2022). Estimating probabilities of passing for examinees with incomplete data in mastery tests. *Educational and Psychological Measurement*, *82*(3), 580–609.
- Wang, Z., Wang, C., & Weiss, D. J. (2022). Termination criteria for grid multiclassification adaptive testing with multidimensional polytomous items. *Applied Psychological Measurement*, *46*(7), 551–570.

### Authors

MIA JOHANNA KATHARINA KORNELY was a research associate at the Manufacturing Technology Institute, RWTH Aachen University, Templergraben 55, 52062 Aachen, Germany; e-mail: mia.kornely@rwth-aachen.de. Her research interests include latent variable modeling, model selection and high-dimensional statistics.

MARIA KATERI is a professor (Chair of Statistics and Data Science) at the Institute of Statistics, RWTH Aachen University, Templergraben 55, 52062 Aachen, Germany; e-mail: maria.kateri@rwth-aachen.de. Her research interests include categorical and ordinal data analysis, stochastic modeling, high-dimensional statistics and model selection, reliability theory and statistical information theory, within both frequentist and Bayesian frameworks.

Manuscript received February 18, 2024

Revision received July 13, 2024

Accepted August 21, 2024