Estimation of Latent Regression Item Response Theory Models Using a Second-Order Laplace Approximation

Björn Andersson
University of Oslo

Tao Xin
Beijing Normal University

The estimation of high-dimensional latent regression item response theory (IRT) models is difficult because of the need to approximate integrals in the likelihood function. Proposed solutions in the literature include using stochastic approximations, adaptive quadrature, and Laplace approximations. We propose using a second-order Laplace approximation of the likelihood to estimate IRT latent regression models with categorical observed variables and fixed covariates where all parameters are estimated simultaneously. The method applies when the IRT model has a simple structure, meaning that each observed variable loads on only one latent variable. Through simulations using a latent regression model with binary and ordinal observed variables, we show that the proposed method is a substantial improvement over the first-order Laplace approximation with respect to the bias. In addition, the approach is equally or more precise to alternative methods for estimation of multidimensional IRT models when the number of items per dimension is moderately high. Simultaneously, the method is highly computationally efficient in the high-dimensional settings investigated. The results imply that estimation of simple-structure IRT models with very high dimensions is feasible in practice and that the direct estimation of high-dimensional latent regression IRT models is tractable even with large sample sizes and large numbers of items.

Keywords: assessment; educational policy; international education/studies; item response theory; measurements; NAEP; performance assessment; psychometrics; regression analyses; research methodology; statistics; student knowledge

Introduction

Item response theory (IRT) is a class of models for categorical observed variables where an underlying latent variable is assumed to generate the observed
variables. Through different formulations of the probability, conditional on the latent variable, for the response in each category of each observed variable, different IRT models are obtained. The primary areas of application of IRT models are in education and psychology, where the models are used to estimate the individual latent variable, to assess the properties of scales and tests, and to infer population characteristics.

Maximum likelihood estimation of IRT models involves the calculation of integrals which, for most IRT models, have no tractable analytical solution and have to be approximated. With three latent dimensions or fewer, computationally efficient implementations using fixed numerical quadrature rules are the most commonly used approaches (Bock & Aitkin, 1981). However, as the dimension increases, the computational expense grows exponentially which makes the fixed quadrature methods excessively computationally and memory intensive when the dimension is higher than three. A second approach is to replace the fixed quadrature rules with adaptive quadrature, which means that each unique integral is approximated with a unique set of quadrature points (Cagnone & Monari, 2013; Rabe-Hesketh et al., 2002; Schilling & Bock, 2005). The adaptive quadrature methods reduce the required number of quadrature points per dimension, but the computational expense still increases exponentially with higher dimensions, making the approach prohibitively computationally demanding in very high dimensions. A third approach uses stochastic approximations such as the Metropolis–Hastings Robbins–Monro (MH-RM) method (Cai, 2010a, 2010b; Chalmers, 2015). The stochastic methods are slower than the alternatives with few dimensions, but the computational expense only grows linearly with an increased number of dimensions. A fourth option is the Laplace approximation (Huber et al., 2004; Joe, 2008; Pinheiro & Bates, 1995; Shun, 1997), which uses asymptotic expansions to approximate the required integral. The first-order Laplace approximation is formally equivalent to adaptive quadrature with only one quadrature point per dimension. Hence, the computational demand of the first-order Laplace approximation grows only linearly with increasing dimensionality and as a result, in high-dimensional models, the first-order Laplace approximation is by far the most computationally efficient method out of the four referenced. The downside is however the inaccuracy of the approximation, especially with few observed variables per dimension and with dichotomous observed variables (Joe, 2008). To improve the computational accuracy, higher order Laplace approximations can be pursued, which has been done with generalized linear models (Raudenbush et al., 2000), generalized linear latent variable models with ordinal data (Bianconcini & Cagnone, 2012), and confirmatory factor analysis with ordinal data and a probit link (Jin et al., 2017). However, a higher order Laplace approximation requires a substantial amount of higher order derivatives and greatly increases the computational expense, especially for high-dimensional models (Bianconcini, 2014).
An extension to the regular IRT model that incorporates a latent regression component is used in many areas, for example, in large-scale educational assessment programs such as the Programme for International Student Assessment, the National Assessment of Education Progress, and the National Assessment of basic Education Quality (NAEQ; Jiang et al., 2019). With a latent regression IRT model, the usual assumption of a fixed mean vector is replaced with an assumption that each individual has a mean vector defined by observed covariates and regression parameters. These models have also been called explanatory IRT models (De Boeck & Wilson, 2013) in the literature and can, for many IRT models, be formulated as nonlinear mixed models (Rijmen et al., 2003). The estimation of latent regression models again requires the approximation of integrals. In current operational procedures in large-scale educational assessment programs, estimation is done in two steps (von Davier & Sinharay, 2013). First, the item parameters are estimated using a unidimensional model without considering the covariates. Second, assuming fixed item parameters, the latent regression and covariance parameters are estimated. The estimation of the latent regression and covariance parameters can be done with an expectation-maximization (EM) algorithm using a second-order Laplace approximation (Thomas, 1993) or using stochastic approximation (von Davier & Sinharay, 2010). It is also possible to utilize adaptive quadrature, stochastic approximations, or a Laplace approximation and estimate the item and regression parameters simultaneously (Chalmers, 2015; Harrell, 2015; Raudenbush et al., 2000), but such methods have typically not been applied to large-scale assessment programs partly because of the large computational requirements and partly because a second-order Laplace approximation has not been available for many IRT models.

The purpose of this article is to introduce an estimation method for latent regression IRT models using second-order Laplace approximations of the integrals in the likelihood function. As a special case of the approach, a second-order Laplace estimation method for regular IRT models without covariates is obtained. The method is introduced for multidimensional simple structure IRT models (also called between-item IRT models) where each item can be individually modeled with either of the nominal response, graded response, generalized partial credit, and three-parameter logistic (3-PL) models. Previous implementations of the higher order Laplace approximation with item response models have focused on estimation of only the latent regression and covariance parameters (Thomas, 1993), Rasch type models (Raudenbush et al., 2000), or particular models for ordinal data without a latent regression (Jin et al., 2017). The approach in this article extends previous research using the second-order Laplace approximation by providing a general method that applies to any item response model and by applying the method to several common models in IRT, for which the second-order Laplace method was not previously available.
Latent Regression IRT Models

Let $Y$ denote the $J \times N$ matrix of item responses and let $X$ denote the $K \times N$ matrix of fixed covariates including a constant equal to 1 for each individual. Also, let $y_i$ and $x_i$ denote the $J$ item response vector and $K$ covariate vector for individual $i$, respectively. Define $\theta$ as the $p \times 1$ latent variable vector and let $P(y_i|\theta; \alpha)$ be the likelihood of the response vector $y_i$ conditional on $\theta$, where $\alpha$ denotes the $p \times 1$ vector of all item parameters. In a latent regression model, the latent variable vector is related to the covariates by the regression model

$$\theta_j = \beta x_i + \varepsilon,$$

where $\varepsilon \sim N(0, \Sigma)$ and $\beta$ is a $p \times K$ matrix of regression coefficients. The marginal likelihood of a response vector $y_i$ is then

$$L_i(\alpha, \beta, \Sigma|x_i, y_i) = \int P(y_i|\theta; \alpha) \phi(\theta; \beta x_i, \Sigma) d\theta,$$

and the marginal log-likelihood for all $n$ individuals is

$$l = \sum_{i=1}^n \log \int P(y_i|\theta; \alpha) \phi(\theta; \beta x_i, \Sigma) d\theta.$$  

IRT Models

The IRT model is defined by the likelihood $P(y_i|\theta; \alpha)$ which is equal to $P(y_i|\theta; \alpha) = \prod_{j=1}^J P(y_{ij}|\theta, \alpha_j)$ due to the local independence assumption, where $y_{ij}$ denotes the item response to the $j$th item, with $y_{ij}$ taking values $k \in \{1, \ldots, m_j\}$, and $\alpha_j$ is a $p \times 1$ vector of item parameters for item $j$. There are numerous models for the item responses $y_i$. We will consider one model for nominal data, namely, the nominal response model with probabilities defined by (Bock, 1972)

$$P(y_{ij} = k|\theta, \alpha_j) = \frac{\exp(a_{jk}/\theta + b_jk)}{\sum_{c=1}^{m_j} \exp(a_{jc}/\theta + b_jc)},$$

where $b_{j1} = 0$ and $a_{jc}$ and $a_{jk}$ are $p \times 1$ vectors such that $a_{j1} = 0$. Let $a_j$ be a $p \times 1$ vector of discrimination parameters. For ordinal data, we consider the graded response model with probabilities defined by (Samejima, 1969)

$$P(y_{ij} = k|\theta, \alpha_j) = P^*(k|\theta, \alpha_j) - P^*(k + 1|\theta, \alpha_j),$$

where

$$P^*(k|\theta, \alpha_j) = \frac{1}{1 + \exp(-a_j/\theta - b_jk)}.$$
with \( P^*(1|\theta; \alpha_j) = 1 \) and \( P^*(m_j + 1|\theta; \alpha_j) = 0 \). We also consider the generalized partial credit model, a special case of the nominal response model, with probabilities defined by (Muraki, 1992)

\[
P(y_{ij} = k|\theta; \alpha_j) = \frac{\exp \left[ \sum_{v=1}^{k} (a_j^v \theta + b_{j,v}) \right]}{\sum_{c=1}^{m_i} \exp \left[ \sum_{v=1}^{c} (a_j^v \theta + b_{j,v}) \right]}.
\] (6)

For dichotomous items, with \( y_{ij} = 2 \) indicating a correct response, these three models coincide, resulting in the two-parameter logistic (2-PL) model

\[
P(y_{ij} = 2|\theta; \alpha_j) = \frac{1}{1 + \exp(-a_j^0 \theta - b_j)}.
\] (7)

With dichotomous data, we will also consider the 3-PL model, with probabilities defined by (Birnbaum, 1968)

\[
P(y_{ij} = 2|\theta; \alpha_j) = c_j + \frac{1 - c_j}{1 + \exp(-a_j^0 \theta - b_j)}.
\] (8)

Note that, for numerical stability in estimation, the parametrization

\[
c_j = \frac{\exp(c_j^*)}{1 + \exp(c_j^*)},
\] (9)

will be used when estimating the unknown parameters with the 3-PL model.

Depending on the exact model formulation, IRT models may be part of other modeling frameworks. For example, the Rasch model can be formulated as a generalized linear mixed model (Clayton, 1996), the graded response model with a probit link as a confirmatory factor analysis model with categorical observed variables (Bartholomew, 1980), and the generalized partial credit model and nominal response model as generalized linear latent variable models (Bartholomew et al., 2011) or as nonlinear mixed models (Rijmen et al., 2003). Meanwhile, the 3-PL model can be viewed as a type of latent class model (Vermunt, 2001).

**Parameter Estimation Using a Second-Order Laplace Approximation**

The Laplace approximation has been proposed to approximate integrals of the form

\[
I(J) = \int e^{-Jh(x)} \, dx,
\] (10)
with $0 < J < \infty$ and $h(x)$ being a smooth function with a unique minimum at $x_0 \in \mathbb{R}^p$. Define the $p \times p$ matrix $H_0 = \frac{\partial^2 h(x)}{\partial x \partial x} \bigg|_{x=x_0}$. We can then write the integral as (Shun, 1997)

$$I(J) = \left( \frac{2\pi}{J} \right)^{p/2} |H_0|^{-1/2} e^{-Jh(x_0)} (1 + R_J + ...)$$

with

$$R_J = - \frac{1}{2J} \left[ \frac{1}{4} \sum_{ijkl} \hat{h}^{ijkl} b_{ik} b_{jl} - \sum_{ijkl} \hat{h}^{ijkl} \hat{h}^{ijkl} \times \left( \frac{1}{4} b_{ij} b_{jk} b_{kl} + \frac{1}{6} b_{ij} b_{jk} b_{kl} \right) \right],$$

and $\{b_{jk}\}$ correspond to the entries in $H_0^{-1}$.

Consider now the latent regression model defined by $\theta_i | x_i \sim N(\beta x_i, \Sigma)$ and a simple-structure item response model. Let $h_i(\theta) = -\log[P(y_i | \theta; \alpha) \phi(\theta; \beta x_i, \Sigma)]$ and let $\hat{\theta}_i$ denote the minimizer of $h_i(\theta)$. For parameter identification, we impose the restrictions $\text{diag}(\Sigma) = (1, \ldots, 1)$ and $\sum_{i=1}^n \beta x_i = 0$. Let the $p_{\beta} \times 1$ vector $\beta^*$ and the $p_{\Sigma} \times 1$ vector $\Sigma^*$ denote the vectors of free parameters in $\beta$ and $\Sigma$, respectively. The proposed estimation method will directly maximize an approximation to the marginal likelihood function with a quasi-Newton method. The second-order Laplace approximation of the likelihood for an individual $i$ is

$$L_{\text{Lap}}^2(\alpha, \beta^*, \Sigma^* | x_i, y_i) = (2\pi)^{p/2} \left| \frac{\partial^2 h_i(\theta)}{\partial \theta \partial \theta} \right|^{\theta=\hat{\theta}_i} e^{-h_i(\hat{\theta}_i)} (1 + \varepsilon_0),$$

and hence the Laplace log-likelihood is

$$l_{\text{Lap}}^2(\alpha, \beta^*, \Sigma^* | x_i, y_i) = \frac{p}{2} \log(2\pi) - \frac{1}{2} \log \left| \frac{\partial^2 h_i(\theta)}{\partial \theta \partial \theta} \right|^{\theta=\hat{\theta}_i} - h_i(\hat{\theta}_i) + \log(1 + \varepsilon_0).$$

Define the $p \times p$ matrix $H_i = \frac{\partial^2 h_i(\theta)}{\partial \theta \partial \theta}$, let $\hat{h}_i = h_i(\hat{\theta}_i)$ and let $b_{i,jj}$ be the $j$th column and $j$th row entry of the inverse of $H_i$. The term $\varepsilon_0$ is for the simple structure assumed here defined by

$$\varepsilon_0 = - \frac{1}{2} \left[ \frac{1}{4} \sum_{j=1}^p \frac{\partial^4 \hat{h}_i}{\partial \theta_j^4} b_{i,jj}^2 - \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \frac{\partial^3 \hat{h}_i}{\partial \theta_j^3} \frac{\partial^3 \hat{h}_i}{\partial \theta_k^3} (b_{i,jk} b_{i,jk} + b_{i,jk}^3) \right],$$

since all third-order and higher cross-derivatives are 0. We then maximize the expression.
Latent Regression Item Response Models

\[ l_{\text{Lap}^2}(\alpha, \beta^*, \Sigma^*|X, Y) = \sum_{i=1}^{n} l_{\text{Lap}^2}(\alpha, \beta^*, \Sigma^*|x_i, y_i), \]  

(17)

with respect to the unknown parameters to obtain the second-order Laplace approximation maximum likelihood estimator of \( \alpha, \beta^* \) and \( \Sigma^* \). To implement the estimation method, the gradient of the Laplace-approximated log-likelihood is needed. Let \( \mathbf{v} \) be a function such that, for column vectors \( \mathbf{x}, \mathbf{y} \) and \( \mathbf{z} \), \( \mathbf{v}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = (x', y', z')' \). The gradient is a \((p_\alpha + p_\beta + p_\Sigma) \times 1\) vector given by

\[
\nabla = \sum_{i=1}^{n} \nabla_i = \sum_{i=1}^{n} \left[ \frac{\partial l^\text{Lap}_i(\alpha, \beta^*, \Sigma^*|x_i, y_i)}{\partial \mathbf{v}(\alpha, \beta^*, \Sigma^*)} \right] + \left( \frac{\partial \mathbf{v}}{\partial \mathbf{v}(\alpha, \beta^*, \Sigma^*)} \right) \frac{\partial l^\text{Lap}_i(\alpha, \beta^*, \Sigma^*|x_i, y_i)}{\partial \mathbf{v}(\alpha, \beta^*, \Sigma^*)},
\]

(18)

where, for \( r \in \{\alpha, \beta^*, \Sigma^*, \theta\}, \)

\[
\frac{\partial l^\text{Lap}_i(\alpha, \beta^*, \Sigma^*|x_i, y_i)}{\partial r} = -\frac{\partial \hat{h}_i}{\partial r} - \frac{1}{2} \text{tr} \left( H^{-1}_i \frac{\partial H_i}{\partial r} \right) - \frac{1}{2(1+e_0)} \sum_{j=1}^{p} \sum_{k=1}^{p} \left[ \frac{\partial^3 \hat{h}_i}{\partial \theta^3_j \partial \theta^3_k \partial r} \left( \frac{1}{4} b_{j,k} b_{j,k} b_{i,k} + \frac{1}{6} b_{i,k}^3 \right) + \frac{\partial^3 \hat{h}_i}{\partial \theta^4_j \partial \theta^4_k \partial r} \left( \frac{1}{4} b_{j,k} b_{j,k} b_{i,k} + \frac{1}{6} b_{i,k}^3 \right) + \frac{\partial^3 \hat{h}_i}{\partial \theta^3_j \partial \theta^4_k \partial r} \left( \frac{1}{4} b_{j,k} b_{j,k} b_{i,k} + \frac{1}{6} b_{i,k}^3 \right) \frac{1}{2} b_{j,k} \frac{\partial b_{j,k}}{\partial r} \right],
\]

(19)

and

\[
\frac{\partial \mathbf{v}}{\partial r} = - \left( \frac{\partial^2 \hat{h}_i}{\partial \theta^2 \partial r} \right)^{-1} \frac{\partial^2 \hat{h}_i}{\partial \theta^2 \partial r}.
\]

(20)

Maximization of the approximated marginal likelihood is done by employing a modified Newton–Raphson method. In iteration \( m \), we update the parameter estimates by

\[
v(\alpha^{(m)}, \beta^{*^{(m)}}, \Sigma^{*^{(m)}}) = v(\alpha^{(m-1)}, \beta^{*^{(m-1)}}, \Sigma^{*^{(m-1)}}) - \lambda_m B_{m-1}^{-1} \nabla^{(m-1)},
\]

(21)

where the \((p_\alpha + p_\beta + p_\Sigma) \times (p_\alpha + p_\beta + p_\Sigma)\) matrix \( B_{m-1} \) is the approximation to the Hessian matrix with the parameter estimates from iteration \( m - 1 \) and \( \lambda_m \) is the step length for iteration \( m \). The matrix \( B_{m-1} \) can be set to \( \sum_{i=1}^{n} \nabla^{(m-1)} \nabla^{(m-1)'} \), the approximation used in the Berndt, Hall, Hall and Hausman (Berndt et al., 1974; BHHH) or to the approximation from the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. In our experience, using the BFGS approximation results in better convergence rates for small sample sizes but requires
considerably more iterations to converge compared to using the BHHH approximation. Hence, BFGS is recommended for small sample sizes, while BHHH is recommended for large sample sizes. The step length $\lambda_m$ is by default set to 0.5 for the first 25 iterations and to 1 for the remaining iterations in order to stabilize the algorithm during the initial steps. A line search (Nocedal & Wright, 2006) may also be employed to find an optimal value of $\lambda_m$ but since the evaluation of the gradient and the likelihood requires the minimization of $h_i(\theta)$ for each individual, it is typically more efficient to set the step length to a fixed value, even though this will require slightly more iterations in the algorithm. The required derivatives for the nominal response model, graded response model, generalized partial credit model, and 3-PL model are provided as Supplementary Material in the online version of the journal.

To obtain an estimator of the asymptotic covariance matrix of the parameter estimator, it is proposed to use one of three different estimators. The first is equal to the $\left(\frac{p}{n} H + \frac{p}{n} H + \frac{p}{n} \Sigma\right) / \left(\frac{p}{n} H + \frac{p}{n} H + \frac{p}{n} \Sigma\right)$ inverse of the empirical cross-product matrix

$$\hat{\Sigma}_{XPD} = \left(\sum_{i=1}^{n} \nabla_i \nabla_i'\right)^{-1},$$

the second is equal to the $\left(\frac{p}{n} H + \frac{p}{n} H + \frac{p}{n} \Sigma\right) / \left(\frac{p}{n} H + \frac{p}{n} H + \frac{p}{n} \Sigma\right)$ inverse of the observed information matrix

$$\hat{\Sigma}_{Obs} = \left(- \sum_{i=1}^{n} \frac{\partial \nabla_i}{\partial \theta^*} \right)^{-1},$$

and the third is the $\left(\frac{p}{n} H + \frac{p}{n} H + \frac{p}{n} \Sigma\right) / \left(\frac{p}{n} H + \frac{p}{n} H + \frac{p}{n} \Sigma\right)$ sandwich estimator (White, 1982)

$$\hat{\Sigma}_{Sand} = \hat{\Sigma}_{Obs} \left(\sum_{i=1}^{n} \nabla_i \nabla_i'\right) \hat{\Sigma}_{Obs}.$$

In practice, the calculation of the observed information matrix is done by numerically differentiating the observed gradient from Equation 18 with respect to the unknown parameters. This calculation can be time-consuming since the gradient evaluation involves the minimization of $h_i(\theta)$ for each individual. For correctly specified models and with large sample sizes, the matrix $\hat{\Sigma}_{XPD}$ is appropriate, while $\hat{\Sigma}_{Sand}$ is needed when the model is misspecified. It can be noted that the matrix $\hat{\Sigma}_{XPD}$ is a by-product of the estimation method and is thus easily obtained within the proposed approach. In this article, the finite-sample properties of the estimator based on the observed information matrix or the empirical cross-product matrix are evaluated through simulations.
Numerical Results

To explore the properties of the proposed estimation method, three simulation studies were performed. In the simulations, the R package mirt (v. 1.30; Chalmers, 2012) was used for the fixed quadrature EM and MH-RM estimation methods, while an alpha version of the R package LaplaceIRT (Andersson & Jin, 2020) was used for the Laplace estimation. Both of the R packages utilize C++ code for the heavy computations needed. In addition, the MH-RM implementation in the commercial software flexMIRT (Cai & Wirth, 2013) was used. In the remainder of the article, we refer to the MH-RM implementation in mirt as MH-RM1 and the MH-RM implementation in flexMIRT as MH-RM2. The first simulation study considered simple-structure dichotomous item response models up to a dimension of six and the second simulation study considered simple-structure polytomous item response models up to a dimension of 12. The covariance matrices used for these two studies were based on the 12 x 12 covariance matrix given in Table 1, which consists of combinations of the covariance terms 0.4, 0.5, and 0.6 and is similar to the setting used in Cai (2010b). The third simulation study used a five-dimensional latent regression model with 60 binary and ordinal items and 100 continuous covariates, designed to mimic a large-scale educational assessment.

The estimation methods EM, first-order Laplace (Lap1), second-order Laplace (Lap2), MH-RM1, and MH-RM2 were considered in the study, but the EM method was only used for models with a dimension of three since estimation in reasonable time was not possible for the higher dimensional models. All estimation methods used the absolute difference in the parameter estimates between successive iterations as the stopping criterion with a tolerance of 0.0001. The EM algorithm with fixed quadrature used 20 quadrature points, the

| Covariance Matrix Used in the 12-Dimensional Simulation |
|--------------------------------------------------------|
| 1.00 .50 .60 .40 .50 .40 .60 .50 .40 .60 .50 .40 |
| .50 1.00 .40 .50 .60 .50 .40 .50 .50 .60 .50 .50 |
| .60 .40 1.00 .50 .50 .50 .40 .60 .60 .40 .40 .40 |
| .40 .50 .50 1.00 .40 .60 .60 .40 .60 .40 .50 .50 |
| .50 .60 .50 .40 1.00 .50 .40 .40 .40 .60 .50 .50 |
| .40 .50 .50 .60 .50 1.00 .40 .40 .50 .60 .50 .60 |
| .60 .40 .40 .40 .40 .40 1.00 .40 .40 .60 .40 .60 |
| .60 .50 .60 .60 .40 .40 .50 .40 .40 .40 .40 .50 |
| .50 .40 .60 .40 .40 .60 .60 .50 .50 .50 1.00 .40 |
| .40 .50 .40 .50 .60 .50 .40 .60 .60 .40 .40 1.00 |
| .60 .50 .40 .40 .50 .60 .60 .40 .60 .40 .40 1.00 |

TABLE 1.

Numerical Results

| Latent Regression Item Response Models |
|----------------------------------------|
| 252 |
MH-RM1 method used five draws per iteration, the gain function \((0.1/t)^{0.75}\), with \(t\) denoting the iteration number, and maximum iterations in the third stage equal to 4,000; the MH-RM2 method used five draws per iteration, the default gain function, and maximum iterations in the third stage equal to 4,000. The settings with the MH-RM1 method were similar to those used in Yang and Cai (2014) and Chalmers and Flora (2014) and worked better than the settings used in Cai (2010b) based on initial testing. The standard errors were estimated based on the Oakes method (Oakes, 1999) for the EM algorithm, with the postconvergence approximation (Yang & Cai, 2014, denoted FMHRM in \textit{mirt}) using 10,000 draws for the MH-RM1 method, and with the recursive approximation using the default tolerance level in \textit{flexMIRT} for the MH-RM2 method. Numerical differentiation of the observed gradient with finite differences was used to calculate the observed information matrix with the Laplace estimation methods. The starting values for the EM, MH-RM1, and MH-RM2 estimation methods were set to the default values used in the respective software implementations. The starting values for the Laplace estimation methods were for the covariance parameters set to 0.5, for the discrimination parameters set to a random number uniformly selected between 0.8 and 1.2, and for the intercept parameters set to 0, (1, –1), (2, 0, –2), (3, 1, –1, –3), and (4, 2, 0, –2, –4), depending on the number of categories (2, 3, 4, 5, and 6). In this study, nonconvergence was identified by the failure to converge within the maximum number of iterations or obtaining a nonpositive definite observed information matrix after convergence. The number of replications was 200 for each of the simulation settings used. The evaluation criteria were, for a parameter \(\xi\) with estimate \(\hat{\xi}_i\) in replication \(i\), the absolute bias \(|\sum_{i=1}^{n}(\hat{\xi}_i - \xi)|/n\) and the root mean square error (RMSE) \(\sqrt{\sum_{i=1}^{n}(\hat{\xi}_i - \xi)^2/n}\). The evaluation criteria were averaged for each type of parameter in each setting. For each simulation setting, we only report the bias and RMSE from the estimation methods that attained a convergence rate of 50% or higher.

**Simulation Study 1: Multidimensional 2-PL Models**

To evaluate the properties of the first- and second-order Laplace approximation estimation methods with dichotomous items, three- and six-dimensional 2-PL models were used where 3 and 6 dichotomous items per dimension were considered. The covariance matrices for the models were set to the corresponding upper submatrices of the matrix in Table 1. Sample sizes 500, 1,000, and 2,000 were used for each condition. The nonconvergence rates for each simulation setting are given in Tables 2 and 3. Due to high nonconvergence rates, the results for Lap1 were excluded for all models except the six-dimensional model with 6 items per dimension. For the six-dimensional model with only 3 items per dimension, MH-RM1 also had excessively high nonconvergence rates, whereas Lap2 had moderate nonconvergence rates (Table 3). MH-RM2 had
no issues with nonconvergence. Note that only the replications for which all compared estimators converged are included in the presented results. The full results can be obtained from the first author upon request. The results for the multidimensional 2-PL models can be found in Tables 4 through 7. Concerning the parameter estimates, the three-dimensional model with 3 items per dimension (Table 4) has higher bias overall with Lap2 compared to EM, MH-RM1, and MH-RM2, while the Lap2 method has lower RMSE, excepting the covariance parameters. The bias is not much reduced for Lap2 with an increased sample size. With 6 items per dimension (Table 5), the differences between EM, MH-RM1, MH-RM2, and Lap2 decrease, and there are no substantial differences between the estimation methods. With six dimensions and 3 items per dimension (Table 6), Lap2 is uniformly better than MH-RM2 with regard to bias and RMSE. Furthermore, the bias is reduced with a larger sample size. For the case of six dimensions and 6 items per dimension (Table 7), the performance of Lap2 is overall somewhat better compared to MH-RM1 and MH-RM2. Meanwhile, both the bias and RMSE are substantially improved with Lap2 compared to Lap1 for all parameters with sample size
TABLE 4.
Average Absolute Bias and RMSE of the Parameter Estimates and Standard Errors for EM, Second-Order Laplace (Lap2), MH-RM1, and MH-RM2 With the Three-Dimensional 3-Item 2-PL Model

| Estimator       | Bias | RMSE |       |       |       |
|-----------------|------|------|-------|-------|-------|
|                 | \(\hat{a}\) | \(\hat{b}\) | \(\hat{\Sigma}\) | \(\hat{SE}_a\) | \(\hat{SE}_b\) | \(\hat{SE}_\Sigma\) | \(\hat{SE}_a\) | \(\hat{SE}_b\) | \(\hat{SE}_\Sigma\) |
| Sample size 500 | .113 | .070 | .010 | .068 | .047 | .003 | .505 | .350 | .111 | .370 | .306 | .015 |
|     EM          | .080 | .045 | .022 | .107 | .081 | .055 | .423 | .289 | .110 | .558 | .417 | .310 |
|     Lap2        | .144 | .088 | .110 | .209 | .107 | .039 | .471 | .322 | .139 | .219 | .122 | .039 |
| Sample size 1,000 | .059 | .044 | .017 | .014 | .009 | .004 | .294 | .224 | .079 | .112 | .103 | .008 |
|     EM          | .088 | .039 | .008 | .025 | .013 | .003 | .246 | .168 | .082 | .055 | .042 | .008 |
|     Lap2        | .061 | .045 | .008 | .030 | .018 | .012 | .302 | .229 | .079 | .166 | .149 | .028 |
|     MH-RM1      | .094 | .062 | .101 | .127 | .075 | .028 | .308 | .224 | .118 | .130 | .079 | .028 |
|     MH-RM2      | .029 | .026 | .004 | .017 | .014 | .003 | .204 | .153 | .056 | .054 | .049 | .005 |
| Sample size 2,000 | .082 | .045 | .016 | .012 | .010 | .003 | .184 | .122 | .061 | .027 | .020 | .005 |
|     EM          | .035 | .029 | .004 | .015 | .009 | .014 | .207 | .153 | .054 | .109 | .094 | .044 |
|     Lap2        | .070 | .055 | .072 | .098 | .059 | .023 | .225 | .167 | .086 | .099 | .061 | .023 |

2,000. The bias and RMSE of the standard errors are the lowest for Lap2 across all sample sizes.

Simulation Study 2: High-Dimensional Graded Response Model

To evaluate the proposed estimation method with a combination of dichotomous and polytomous observed variables, the three-dimensional simple-structure graded response model in Cai (2010b) was used as the basis for a simulation study where, in addition to the three-dimensional case, six- and 12-dimensional models were also considered. As in Simulation Study 1, the covariance matrices for the three- and six-dimensional models were set to the corresponding upper submatrices of the matrix in Table 1. The 12-dimensional model used the full covariance matrix. For the three-dimensional model with a sample size of 500, the second-order Laplace estimation took an average of approximately 3 seconds using a 3.3GHz CPU (Core i5-4590) with 16GB RAM, which can be compared to the time with MH-RM reported in Cai (2010b), which was 20 seconds with a 2.0GHz CPU and 2GB RAM. The estimation methods EM, Lap1, Lap2, MH-RM1, and MH-RM2 were considered in the study, but EM was only used with the three-dimensional models since estimation in reasonable time was not possible for the higher dimensional models. In addition, it was not
### TABLE 5.
**Average Absolute Bias and RMSE of the Parameter Estimates and Standard Errors for EM, Second-Order Laplace (Lap2), MH-RM1, and MH-RM2 With the Three-Dimensional 6-Item 2-PL Model**

| Estimator | Bias | RMSE |
|-----------|------|------|
|           | $\hat{a}$ | $\hat{b}$ | $\hat{\Sigma}$ | $\hat{SE}_a$ | $\hat{SE}_b$ | $\hat{SE}_\Sigma$ | $\hat{a}$ | $\hat{b}$ | $\hat{\Sigma}$ | $\hat{SE}_a$ | $\hat{SE}_b$ | $\hat{SE}_\Sigma$ |
| Sample size 500 |       |       |       |       |       |       |       |       |       |       |       |       |
| EM        | .034   | .025   | .005   | .016   | .010   | .001   | .266   | .209   | .074   | .067   | .060   | .005   |
| Lap2      | .029   | .017   | .004   | .014   | .014   | .001   | .259   | .208   | .074   | .059   | .058   | .005   |
| MH-RM1    | .033   | .023   | .012   | .015   | .009   | .005   | .266   | .208   | .080   | .071   | .063   | .009   |
| MH-RM2    | .043   | .034   | .052   | .079   | .043   | .032   | .277   | .218   | .083   | .087   | .055   | .033   |
| Sample size 1,000 |       |       |       |       |       |       |       |       |       |       |       |       |
| EM        | .017   | .012   | .005   | .006   | .006   | .003   | .177   | .141   | .131   | .027   | .024   | .004   |
| Lap2      | .017   | .007   | .004   | .007   | .005   | .003   | .172   | .137   | .055   | .025   | .022   | .004   |
| MH-RM1    | .017   | .012   | .004   | .005   | .006   | .005   | .178   | .141   | .057   | .028   | .025   | .006   |
| MH-RM2    | .021   | .017   | .040   | .046   | .025   | .021   | .182   | .145   | .063   | .048   | .029   | .021   |
| Sample size 2,000 |       |       |       |       |       |       |       |       |       |       |       |       |
| EM        | .008   | .009   | .001   | .007   | .006   | .002   | .125   | .100   | .034   | .015   | .013   | .002   |
| Lap2      | .013   | .008   | .002   | .007   | .006   | .002   | .123   | .097   | .035   | .014   | .013   | .002   |
| MH-RM1    | .009   | .008   | .005   | .007   | .006   | .002   | .125   | .099   | .036   | .015   | .013   | .003   |
| MH-RM2    | .011   | .013   | .027   | .032   | .018   | .017   | .127   | .102   | .043   | .033   | .019   | .017   |

### TABLE 6.
**Average Absolute Bias and RMSE of the Parameter Estimates and Standard Errors for Second-Order Laplace (Lap2) and MH-RM2 With the Six-Dimensional 3-Item 2-PL Model**

| Estimator | Bias | RMSE |
|-----------|------|------|
|           | $\hat{a}$ | $\hat{b}$ | $\hat{\Sigma}$ | $\hat{SE}_a$ | $\hat{SE}_b$ | $\hat{SE}_\Sigma$ | $\hat{a}$ | $\hat{b}$ | $\hat{\Sigma}$ | $\hat{SE}_a$ | $\hat{SE}_b$ | $\hat{SE}_\Sigma$ |
| Sample size 500 |       |       |       |       |       |       |       |       |       |       |       |       |
| Lap2      | .065   | .031   | .019   | .034   | .026   | .005   | .307   | .203   | .134   | .100   | .088   | .028   |
| MH-RM2    | .167   | .114   | .116   | .182   | .094   | .033   | .457   | .320   | .137   | .193   | .111   | .033   |
| Sample size 1,000 |       |       |       |       |       |       |       |       |       |       |       |       |
| Lap2      | .049   | .026   | .016   | .018   | .013   | .004   | .212   | .145   | .093   | .051   | .047   | .011   |
| MH-RM2    | .115   | .080   | .092   | .104   | .054   | .028   | .296   | .210   | .109   | .108   | .060   | .028   |
| Sample size 2,000 |       |       |       |       |       |       |       |       |       |       |       |       |
| Lap2      | .043   | .025   | .015   | .013   | .010   | .002   | .151   | .105   | .067   | .026   | .023   | .005   |
| MH-RM2    | .086   | .056   | .076   | .065   | .031   | .022   | .201   | .140   | .088   | .066   | .033   | .022   |
possible to use MH-RM1 in the simulation with the 12-dimensional model due to the excessively long time required in estimation and for calculation of the observed information matrix (with sample size 500 one replication took more than 150 minutes to finish). For all conditions considered, the nonconvergence rates were below 1\%.

The results for the three-dimensional model can be seen in Table 8. Overall, Lap1 has higher bias than the other estimation methods. However, Lap1 has lower RMSE for the item parameters compared to the other estimation methods with the smaller sample sizes. For the parameter estimates, EM, Lap2, MH-RM1, and MH-RM2 are virtually identical across all settings and evaluation criteria, with the exception of higher bias and RMSE for the covariance parameters with MH-RM2. The six-dimensional results can be found in Table 9. The results mirror those for the three-dimensional case, with Lap1 having higher bias than the alternatives but maintaining a lower RMSE for the small sample case. Lap2 has the overall lowest bias across all estimation methods for each setting although the differences are small with the largest sample size. The standard errors with the three- and six-dimensional models are very close regarding the accuracy and precision between the different estimation methods, with the exception that MH-RM1 has slightly higher RMSE overall and that MH-RM2 has higher bias and RMSE overall. With the 12-dimensional model, the Laplace and MH-RM2 estimation methods were the only ones possible to use and the results can be seen in Table 10. The results are similar to the lower dimensional
cases in that Lap1 has higher bias but lower variance, resulting in a lower RMSE for the small sample sizes, and that MH-RM2 has higher bias and RMSE with respect to the covariance parameters. Lap2 has bias which is slightly lower on average than with lower dimensional models and has the lowest overall bias among the considered estimation methods. The standard errors are highly similar with respect to the bias and RMSE between Lap1 and Lap2, while the bias and RMSE are slightly higher with MH-RM2.

**Simulation Study 3: Latent Regression in Large-Scale Assessment**

A five-dimensional model based on the NAEQ 2015 mathematics assessment was also used, where the correlations between the latent variables were all set to .8. The model had 60 generalized partial credit model items (12 in each dimension, with 6 dichotomous and 6 polytomous three-category items in each dimension) and 100 uncorrelated covariates, yielding a total of 690 parameters. A sample size of 60,000 was used, similar to the real Grade 8 NAEQ assessment, where each individual answered on average 20 items in total, meaning that on
The simulation results are given in Table 11. The bias is substantially lower with Lap2 for the item and covariance parameters, while the improvement is smaller for the regression parameters. The same applies with respect to the RMSE, except that the differences between Lap1 and Lap2 are somewhat smaller. The standard errors were estimated with the cross-product approximation due to the large computational resources required to calculate the observed information matrix. The standard errors are approximately equally accurate and precise for either of the two estimation methods.
Conclusions

In this article, a second-order Laplace approximation was introduced for the estimation of multidimensional simple structure item response models with a latent regression component. Through numerical illustrations using realistic settings in education and psychology, it was shown that the estimation method gave a substantial improvement over the first-order Laplace approximation for the estimation of latent regression models and multidimensional IRT models with both binary and ordinal data. Furthermore, the estimation method was typically equally or more precise in estimating multidimensional IRT models when compared to alternatives such as the EM algorithm with fixed quadrature and two

| TABLE 10. Average Absolute Bias and RMSE of the Parameter Estimates and Standard Errors for First-Order Laplace (Lap1), Second-Order Laplace (Lap2), and MH-RM2 With the 12-Dimensional Graded Response Model |
| Estimator | $\hat{a}$ | $\hat{b}$ | $\hat{\Sigma}$ | $\hat{SE}_a$ | $\hat{SE}_b$ | $\hat{SE}_\Sigma$ | $\hat{a}$ | $\hat{b}$ | $\hat{\Sigma}$ | $\hat{SE}_a$ | $\hat{SE}_b$ | $\hat{SE}_\Sigma$ |
| Sample size 500 |
| Lap1 | .016 | .009 | .012 | .008 | .006 | .002 | .171 | .141 | .056 | .018 | .011 | .004 |
| Lap2 | .014 | .010 | .003 | .008 | .006 | .002 | .178 | .142 | .054 | .020 | .011 | .004 |
| MH-RM2 | .025 | .012 | .042 | .018 | .017 | .015 | .176 | .142 | .065 | .024 | .018 | .015 |
| Sample size 1,000 |
| Lap1 | .022 | .005 | .008 | .005 | .004 | .002 | .121 | .098 | .040 | .010 | .006 | .002 |
| Lap2 | .006 | .006 | .003 | .005 | .004 | .001 | .124 | .099 | .038 | .010 | .006 | .002 |
| MH-RM2 | .027 | .028 | .040 | .012 | .011 | .012 | .126 | .103 | .054 | .014 | .011 | .012 |
| Sample size 2,000 |
| Lap1 | .023 | .005 | .012 | .003 | .003 | .001 | .088 | .069 | .030 | .005 | .004 | .002 |
| Lap2 | .005 | .004 | .002 | .003 | .003 | .001 | .089 | .070 | .027 | .006 | .004 | .001 |
| MH-RM2 | .029 | .015 | .034 | .009 | .007 | .009 | .092 | .071 | .043 | .010 | .008 | .009 |

| TABLE 11. Average Absolute Bias and RMSE of the Parameter Estimates and Average RMSE of the Standard Errors for First-Order Laplace (Lap1) and Second-Order Laplace (Lap2) With the Five-Dimensional Latent Regression Generalized Partial Credit Model |
| Estimator | $v(\hat{a}, \hat{b})$ | $\beta$ | $\Sigma$ | $\hat{SE}_v(\hat{a}, \hat{b})$ | $\hat{SE}_\beta$ | $\hat{SE}_\Sigma$ | $v(\hat{a}, \hat{b})$ | $\beta$ | $\Sigma$ | $\hat{SE}_v(\hat{a}, \hat{b})$ | $\hat{SE}_\beta$ | $\hat{SE}_\Sigma$ |
| Lap1 | .022 | .002 | .013 | .002 | .001 | .000 | .043 | .020 | .013 | .002 | .001 | .000 |
| Lap2 | .002 | .001 | .001 | .001 | .001 | .000 | .035 | .020 | .004 | .002 | .001 | .000 |

Conclusions

In this article, a second-order Laplace approximation was introduced for the estimation of multidimensional simple structure item response models with a latent regression component. Through numerical illustrations using realistic settings in education and psychology, it was shown that the estimation method gave a substantial improvement over the first-order Laplace approximation for the estimation of latent regression models and multidimensional IRT models with both binary and ordinal data. Furthermore, the estimation method was typically equally or more precise in estimating multidimensional IRT models when compared to alternatives such as the EM algorithm with fixed quadrature and two
implementations of the MH-RM, while considerably reducing the time required for convergence with high-dimensional models. The second-order Laplace approximation provides a fast yet accurate method for estimation of high-dimensional simple structure IRT models and enables the direct estimation of high-dimensional simple structure latent regression IRT models to situations with a large amount of items, covariates, and individuals such as in large-scale educational assessment programs.

In previous research, the first-order Laplace approximation has been shown to function poorly with a low number of observed dichotomous variables (Joe, 2008). The results of this study echo this but suggest that the second-order Laplace approximation can substantially improve over the first-order approximation with dichotomous observed variables. Note that with as few as three observed variables per dimension, the second-order Laplace approximation estimation method can still exhibit bias that is larger than the alternatives. However, when increasing the number of observed dichotomous variables per dimension to six, the second-order Laplace approximation was equally good or better than any of the alternatives considered with respect to the RMSE. The practical performance of the Laplace approximation-based estimation methods is related to the theoretical error rates of the approximations used (Shun, 1997). With an increased number of items, the error of the approximations decrease, and with a higher order approximation, the error decreases at a faster rate compared to a lower order approximation.

Besides the reduced time needed in estimation, the Laplace approximation has several other advantages compared to the stochastic approximation methods. The convergence diagnostics are more straightforward since the observed-likelihood gradient provides an inexpensive convergence check to supplement the stopping criterion. Although the second-order test via the observed information matrix can be expensive to obtain, the second-order test is also required with the stochastic approximation methods, and for the stochastic approximation methods, this matrix has to be approximated via simulations which often requires a substantial amount of replicates in order for the accuracy to be sufficient. Another advantage is the fast calculation of the observed log-likelihood, used for model comparisons and hypothesis tests. With stochastic approximation methods, obtaining an accurate approximation of the log-likelihood requires heavy computations, which for large sample sizes can take even longer to finish than the actual estimation process. Another advantage to the Laplace estimation method is that the method is not heavily dependent on the estimation settings used. With the MH-RM method, choices have to be made regarding how many samples to draw at each iteration, how to select the gain constants, and how accurate the approximations to the observed information matrix and observed log-likelihood should be. The optimal choices for these settings can vary greatly for different problems, which makes the application of the MH-RM somewhat difficult from the user perspective. It can be noted that the mirt implementation of the MH-RM (denoted MH-
RM1 in the article) exhibited convergence problems with small sample sizes and with few dichotomous items. We have used MH-RM estimation settings which are highly similar to previous studies (Chalmers & Flora, 2014; Yang & Cai, 2014) and found that other settings like those in, for example, Cai (2010b) did not improve the performance. However, it is possible that tuning the settings further could have improved the convergence rates, and we also note that the *flexMIRT* implementation (denoted MH-RM2 in the article) did not exhibit the same non-convergence problems. The performance of MH-RM1 and MH-RM2 are dependent on the specific settings used in this study. By changing settings such as the maximum number of iterations, the stopping criterion, or the type of standard error estimation, different performance could be obtained. The settings in the Laplace approximation require less finetuning, but it may be necessary to try multiple optimizers or step lengths during estimation, particularly for small sample sizes. To exemplify the impact of the specific settings used, we can note that the standard errors from the recursive approximation (used with the MH-RM2 method) have a tendency to underestimate the standard errors, while the cross-product matrix used in Simulation Study 3 has a tendency to overestimate the standard errors with small sample sizes.

In the current study, we have only investigated the computational efficiency of the different methods using a single CPU core. With the Laplace approximation methods, multiple cores are easily utilized by parallel computations of each individual contribution to the log-likelihood and gradient. We also note that estimation methods using MH-RM easily utilize multiple cores which will improve their computational efficiency with many cores. In addition, using one draw instead of five draws as done here will increase the computational efficiency with MH-RM. To compare the computational efficiency of estimation methods incorporating parallel computing is a useful future area of research.

A deficiency of estimation with a higher order Laplace approximation is that the computational requirement for models with a more complicated structure increases greatly in higher dimensions. For example, with 12 dimensions and items that load on all latent variables, the summations in Equation 12 contain up to $12^6$ nonzero elements, instead of only up to $12^2$ as in the simple structure case given in Equation 16. Hence, it is unlikely that the second-order Laplace approximation will be more efficient than the stochastic approximation methods for the most general models although it may provide a benefit for models that are slightly more complex than the simple structure models (e.g., when each item loads on only two out of many latent variables) provided that the computer implementation takes the specific structure into account. It is straightforward to implement a general estimation method, and general model structures are supported in the software package *LaplaceIRT*. Additional research is needed to evaluate the properties of the higher order Laplace estimation method for models with a more complex structure.
The proposed method is general but for each new item response model up to the fifth-order derivatives with respect to the latent variables need to be derived and additionally the derivatives with respect to the item parameters have to be derived for the derivatives up to the fourth order with respect to the latent variables. In this article, support for the nominal response, graded response, generalized partial credit, and 3-PL models has been provided, and any combination of these models can be used.

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
Björn Andersson https://orcid.org/0000-0002-9007-2440

References
Andersson, B., & Jin, S. (2020). LaplaceIRT: Estimation of multidimensional item response theory models using a second-order Laplace approximation (R package version 0.0.7) [Computer software manual].
Bartholomew, D. J. (1980). Factor analysis for categorical data. Journal of the Royal Statistical Society. Series B (Methodological), 52, 293–321.
Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). Latent variable models and factor analysis: A unified approach (Vol. 904). John Wiley & Sons.
Berndt, E. R., Hall, B. H., Hall, R. E., & Hausman, J. A. (1974). Estimation and inference in nonlinear structural models. Annals of Economic and Social Measurement, 3, 653–665.
Bianconcini, S. (2014). Asymptotic properties of adaptive maximum likelihood estimators in latent variable models. Bernoulli, 20, 1507–1531.
Bianconcini, S., & Cagnone, S. (2012). Estimation of generalized linear latent variable models via fully exponential Laplace approximation. Journal of Multivariate Analysis, 112, 183–193.
Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee’s ability. In F. Lord & M. Novick (Eds.), Statistical theories of mental test scores (pp. 395–479). Addison-Wesley.
Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 37, 29–51.
Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 46, 443–459.
Cagnone, S., & Monari, P. (2013). Latent variable models for ordinal data by using the adaptive quadrature approximation. Computational Statistics, 28, 597–619.
Cai, L. (2010a). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika*, 75, 33–57.

Cai, L. (2010b). Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, 35, 307–335.

Cai, L., & Wirth, R. (2013). *flexmirt*: Flexible multilevel multidimensional item analysis and test scoring (Version 2) [Computer software]. Vector Psychometric Group.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29.

Chalmers, R. P. (2015). Extended mixed-effects item response models with the MH-RM algorithm. *Journal of Educational Measurement*, 52, 200–222.

Chalmers, R. P., & Flora, D. B. (2014). Maximum-likelihood estimation of noncompensatory IRT models with the MH-RM algorithm. *Applied Psychological Measurement*, 38, 339–358.

Clayton, D. G. (1996). Generalized linear mixed models. In W. Gilks, S. Richardson, & D. J. Spigelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 275–301). Chapman & Hall/CRC.

De Boeck, P., & Wilson, M. (2013). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer-Verlag.

Harrell, L. A. (2015). *Analysis strategies for planned missing data in health sciences and education research* [Unpublished doctoral dissertation]. UCLA.

Huber, P., Ronchetti, E., & Victoria-Feser, M.-P. (2004). Estimation of generalized linear latent variable models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66, 893–908.

Jiang, Y., Zhang, J., & Xin, T. (2019). Toward education quality improvement in china: A brief overview of the national assessment of education quality. *Journal of Educational and Behavioral Statistics*, 44, 733–751.

Jin, S., Noh, M., & Lee, Y. (2017). H-likelihood approach to factor analysis for ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 530–540.

Joe, H. (2008). Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics & Data Analysis*, 52, 5066–5074.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159–176.

Nocedal, J., & Wright, S. (2006). *Numerical optimization*. Springer-Verlag.

Oakes, D. (1999). Direct calculation of the information matrix via the EM. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 479–482.

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4, 12–35.

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

Raudenbush, S. W., Yang, M.-L., & Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics*, 9, 141–157.

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods*, 8, 185–205.
Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometrika, Monograph Supplement No. 17). Psychometrika Society.

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

Shun, Z. (1997). Another look at the salamander mating data: A modified Laplace approximation approach. *Journal of the American Statistical Association*, 92(437), 341–349.

Thomas, N. (1993). Asymptotic corrections for multivariate posterior moments with factored likelihood functions. *Journal of Computational and Graphical Statistics*, 2, 309–322.

Vermunt, J. K. (2001). The use of restricted latent class models for defining and testing nonparametric and parametric item response theory models. *Applied Psychological Measurement*, 25, 283–294.

von Davier, M., & Sinharay, S. (2010). Stochastic approximation methods for latent regression item response models. *Journal of Educational and Behavioral Statistics*, 35, 174–193.

von Davier, M., & Sinharay, S. (2013). Analytics in international large-scale assessments: Item response theory and population models. In L. Rutkowski (Ed.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 155–174). Chapman & Hall/CRC Press.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1–25.

Yang, J. S., & Cai, L. (2014). Estimation of contextual effects through nonlinear multilevel latent variable modeling with a Metropolis–Hastings Robbins–Monro algorithm. *Journal of Educational and Behavioral Statistics*, 39, 550–582.

**Authors**

BJÖRN ANDERSSON is an associate professor at the Centre for Educational Measurement, University of Oslo, Postboks 1161 Blindern, Oslo, 0316 Norway; email: bjur-n.andersson@cemo.uio.no. His research interests include test equating, latent variable model estimation methods, and applications of item response theory.

TAO XIN is a professor at the Collaborative Innovation Center of Assessment toward Basic Education Quality at Beijing Normal University, No. 19, Xinjiekouwai St., Beijing, 100875, China; email: xintao@bnu.edu.cn. His research interests are cognitive diagnostic modeling, latent trait models, and assessment theory.

Manuscript received December 12, 2017
First revision received May 7, 2018
Second revision received May 10, 2019
Third revision received December 8, 2019
Fourth revision received June 30, 2020
Accepted July 3, 2020

265