Computational Statistics and Data Analysis 182 (2023) 107710

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Fast estimation of multiple group generalized linear latent variable models for categorical observed variables

Björn Andersson [a,*], Shaobo Jin [b,c], Maoxin Zhang [a]

[a] Centre for Educational Measurement at the University of Oslo (CEMO), P.O. Box 1161 Forskningsparken, Oslo, 0318, Norway
[b] Department of Statistics, Uppsala University, Box 513, Uppsala, 751 20, Sweden
[c] Department of Mathematics, Uppsala University, Box 480, Uppsala, 751 06, Sweden

ARTICLE INFO

Article history:
Received 28 April 2022
Received in revised form 20 October 2022
Accepted 28 January 2023
Available online 2 February 2023

Keywords:
Latent variable models
Item response theory
Integral approximation
Gauss-Hermite quadrature
Laplace approximation

ABSTRACT

A computationally efficient method for marginal maximum likelihood estimation of multiple group generalized linear latent variable models for categorical data is introduced. The approach utilizes second-order Laplace approximations of the integrals in the likelihood function. It is demonstrated how second-order Laplace approximations can be utilized highly efficiently for generalized linear latent variable models by considering symmetries that exist for many types of model structures. In a simulation with binary observed variables and four correlated latent variables in four groups, the method has similar bias and mean squared error compared to adaptive Gauss-Hermite quadrature with five quadrature points while substantially improving computational efficiency. An empirical example from a large-scale educational assessment illustrates the accuracy and computational efficiency of the method when compared against adaptive Gauss-Hermite quadrature with three, five, and 13 quadrature points.

1. Introduction

When estimating latent variable models for categorical observed variables, such as generalized linear latent variable models or item response theory models, marginal maximum likelihood estimation is typically used. With marginal maximum likelihood, integrals without an explicit solution must be calculated. Standard estimation methods are based on Gauss-Hermite quadrature approximations (Bock and Aitkin, 1981), which are highly efficient for models with one or two latent variables but quickly decrease in efficiency with higher-dimensional problems. Adaptive quadrature (Schilling and Bock, 2005; Cagnone and Monari, 2013), which concentrates the region of integration to the most relevant part for each integral, is a partial solution but with more than four dimensions the computational expense means that adaptive quadrature becomes impractical if high accuracy is desired. Another approach is to use a simulation-based method such as the Metropolis-Hastings Robbins-Monro method (Cai, 2010a) or a Monte Carlo method (Zhu et al., 2005). However, the simulation-based methods can be slow to converge to a local maximum with small sample sizes and ensuring that proper convergence has been attained is often challenging and time-consuming. Efficient approximation methods such as the variational approximation have also been proposed (Hui et al., 2017; Niku et al., 2019; Cho et al., 2021), but these methods perform relatively poorly with few observed variables and have not been implemented for many types of models.

* Corresponding author.
E-mail address: bjorn.andersson@cemo.uio.no (B. Andersson).

https://doi.org/10.1016/j.csda.2023.107710

Another approach is to use a first-order Laplace approximation to approximate the required integrals (Huber et al., 2004). However, estimation methods based on first-order Laplace approximations often have convergence problems and high bias (Joe, 2008), especially with binary data or complex models (Andersson and Xin, 2021). To remedy this, second-order Laplace approximations of the log-likelihood (Shun, 1997; Thomas, 1993; Bianconcini, 2014; Raudenbush et al., 2000) or second-order Laplace approximations of the gradient of the log-likelihood (Bianconcini and Cagnone, 2012) have been used. Such higher-order Laplace approximations have shown promise in providing a computationally efficient yet accurate estimation method for models with many latent variables, such as in Andersson and Xin (2021) where up to 12 correlated latent variables were used with an independent-clusters structure and ordinal observed variables. In contrast to this, some previous studies found that the second-order Laplace approximation was computationally much less efficient compared to adaptive quadrature with 5 quadrature points in each dimension for exploratory factor analysis with ordinal data with up to four dimensions (Bianconcini, 2014). As we will show in the present study, the efficiency of higher-order Laplace approximations is highly dependent on the structure of the model used and implementations that disregard the structure will be inefficient for most types of models. However, if the structure is exploited in the implementation of the second-order Laplace approximation, substantial computational gains can be obtained which makes the method highly computationally efficient for many types of models when compared to adaptive quadrature approximations which have the same theoretical approximation accuracy.

A hindrance to methods which use second-order Laplace approximations is that they require the derivation of higher-order derivatives which depend on the type of measurement model specified. This makes higher-order Laplace approximations difficult to efficiently implement and generalize across different types of models. Hence, so far, the available implementations of second-order Laplace approximations of the log-likelihood which are relevant to latent variable models for categorical data have been limited to generalized linear models (Raudenbush et al., 2000), generalized linear mixed models (Noh and Lee, 2007), confirmatory factor analysis models for ordinal data (Bianconcini, 2014; Jin et al., 2018), independent-clusters item response theory models (Thomas, 1993; Andersson and Xin, 2021) and nonlinear structural equation models (Jin et al., 2020). Thus, additional research is needed to implement estimation of multiple group generalized linear latent variable models with second-order Laplace approximations and to investigate its estimation properties.

The objective of the current work is then to develop a computationally efficient estimation method based on a second-order Laplace approximation to estimate multidimensional generalized linear latent variable models with categorical observed variables, with support for an arbitrary model structure and with multiple groups. There are three main contributions of the present study in relation to the existing literature. First, we derive an estimation algorithm that uses a second-order Laplace approximation to the marginal log-likelihood function for generalized linear latent variable models for categorical observed variables which supports a general model structure. Here, we also detail how the second-order Laplace approximation can be highly efficiently implemented by accounting for the structure of the particular model used. Second, we implement the second-order Laplace approximation estimation method for multiple group models where parameter invariance between groups can be established and where the mean vectors and covariance matrices of the latent variable in multiple groups can be estimated. Third, we compare the second-order Laplace approximation method to an implementation of adaptive Gauss-Hermite quadrature, that uses the same underlying code base as the second-order Laplace method, in terms of the estimation accuracy and precision and in terms of the computational efficiency.

The paper is structured as follows. We first introduce the modeling framework used and then present the second-order Laplace approximation estimation method along with a discussion of some of its properties with models commonly used in applied measurement. Then, based on a simulation study, we contrast and compare the proposed approach to adaptive Gauss-Hermite quadrature and discuss the advantages and disadvantages of the method based on theoretical and practical considerations. Subsequently, an empirical example from an international large-scale assessment is used to illustrate the application of the Laplace approximations and adaptive quadrature methods. Lastly, we discuss our findings and provide recommendations for applied work.

2. Methods

2.1. Models

With latent variable models for categorical data, we model the response probabilities for each category of a set of discrete observed variables $i \in \{1, \ldots, I\}$ conditional on a latent variable. Define $P_{ic}(\boldsymbol{z})$ as the probability, conditional on the $p \times 1$ latent variable vector $\boldsymbol{z}$, to observe category c of observed variable Y_i which has m_i possible outcomes. We assume conditional independence such that the joint probability for multiple random variables $Y_1 = y_1, \ldots, Y_I = y_I$, conditional on $\boldsymbol{z}$, can be factorized as

$$P(Y_1 = y_1, \ldots, Y_I = y_I|\boldsymbol{z}) = \prod_{i=1}^{I} P_{iy_i}(\boldsymbol{z}), \tag{1}$$

where the individual P_{iy_i} can be based on, for example, confirmatory factor analysis with categorical data, the generalized partial credit model (Muraki, 1992), the graded response model (Samejima, 1969), or the nominal response model

(Bock, 1972). These three specific models are all types of generalized linear latent variable models (Huber et al., 2004) and also fall within the framework of generalized linear latent and mixed models (Rabe-Hesketh et al., 2004). Let $\boldsymbol{b}_i$ be a $m_i \times 1$ vector of intercept parameters, with entries b_{ic} such that $b_{i1} = 0$. For the graded response model, with a $p \times 1$ vector $\boldsymbol{a}_i$ of slope parameters, we have

$$P_{ic}(\boldsymbol{z}) = P^*_{ic}(\boldsymbol{z}) - P^*_{i(c+1)}(\boldsymbol{z}), \tag{2}$$

where

$$P^*_{ic}(\boldsymbol{z}) = \frac{1}{1 + \exp(-\boldsymbol{a}_i'\boldsymbol{z} - b_{ic})}, \tag{3}$$

with $P^*_{i1}(\boldsymbol{z}) = 1$ and $P^*_{i(m_i+1)}(\boldsymbol{z}) = 0$. For the nominal response model, with a $p \times 1$ vector $\boldsymbol{a}_{ic}$ of slope parameters for each category c such that $\boldsymbol{a}_{i1} = \boldsymbol{0}$, we have

$$P_{ic}(\boldsymbol{z}) = \frac{\exp\left(\boldsymbol{a}_{ic}'\boldsymbol{z} + b_{ic}\right)}{\sum_{c'=1}^{m_i} \exp\left(\boldsymbol{a}_{ic'}'\boldsymbol{z} + b_{ic'}\right)}, \tag{4}$$

and for the generalized partial credit model, a special case of the nominal response model, we have

$$P_{ic}(\boldsymbol{z}) = \frac{\exp\left[\sum_{v=1}^{c}(\boldsymbol{a}_i'\boldsymbol{z} + b_{iv})\right]}{\sum_{c'=1}^{m_i} \exp\left[\sum_{v=1}^{c'}(\boldsymbol{a}_i'\boldsymbol{z} + b_{iv})\right]}. \tag{5}$$

In principle any probability model that satisfies the conditional independence assumption can be used. Let N_g denote the sample size in group g and define $\boldsymbol{y}_{fg}$ as the $I \times 1$ vector of observed variables for an individual $f \in \{1, \ldots, N_g\}$ in group $g \in \{1, \ldots, G\}$ and let $N = \sum_{g=1}^{G} N_g$. The marginal log-likelihood for an individual f in group g is equal to

$$l_{fg}(\boldsymbol{\theta}_g|\boldsymbol{y}_{fg}) = \log \int P(\boldsymbol{y}_{fg}|\boldsymbol{z})\phi(\boldsymbol{z}; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)d\boldsymbol{z}, \tag{6}$$

where $\boldsymbol{\theta}_g$ are the unknown parameters in group g and ϕ is the multivariate normal density function with mean $\boldsymbol{\mu}_g$ and covariance matrix $\boldsymbol{\Sigma}_g$. Define $\boldsymbol{\theta}$ as the vector of all free parameters of the model across all groups. With multiple group models it is possible to evaluate measurement invariance across groups and estimate the mean vectors and covariance matrices in the groups, provided there exist some observed indicators which exhibit invariance (Muthen and Lehman, 1985). Typically, we cannot solve the integral in Equation (6) analytically and it must be approximated.

2.2. Likelihood approximation

We consider approximating the marginal log-likelihood with either a second-order Laplace approximation or adaptive Gauss-Hermite quadrature. Here, we present these two methods and outline their properties in terms of accuracy and computational efficiency.

2.2.1. A second-order Laplace approximation

We propose to approximate the integrals in the likelihood function with a second-order Laplace approximation (Shun, 1997) and implement an estimation method based on such approximations. Define the function $h_{fg}(\boldsymbol{z}) = -\log P(\boldsymbol{y}_{fg}|\boldsymbol{z})\phi(\boldsymbol{z}; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$. The second-order Laplace approximation is based on 1) the estimation of the posterior mode of the latent variable vector for each individual, and 2) derivatives of h_{fg} with respect to $\boldsymbol{z}$ up to the fourth order. Let $\hat{\boldsymbol{z}}_{fg}$ be the posterior mode, equal to the minimizer of $h_{fg}(\boldsymbol{z})$. Define $\hat{h} = h_{fg}(\hat{\boldsymbol{z}}_{fg})$, $\boldsymbol{H}_{fg} = \frac{\partial^2 \hat{h}}{\partial \boldsymbol{z} \partial \boldsymbol{z}'}$ and let $\hat{\boldsymbol{Z}}$ denote the $p \times N$ matrix of posterior modes. We then obtain the second-order Laplace approximation to the log-likelihood as Shun (1997)

$$l^{\text{Lap2}}_{fg}(\boldsymbol{\theta}_g|\boldsymbol{y}_{fg}) = \frac{p}{2}\log(2\pi) - \frac{1}{2}\log\left|\boldsymbol{H}_{fg}\right| - \hat{h} + \log(1 + \epsilon_{fg}), \tag{7}$$

where, with b_{jk} denoting the j-th row entry of the k-th column in $\boldsymbol{H}^{-1}_{fg}$,

$$\epsilon_{fg} = -\frac{1}{2}\left[\frac{1}{4}\sum_{jklm}^{p} \frac{\partial^4 \hat{h}}{\partial z_j \partial z_k \partial z_l \partial z_m} b_{jl} b_{km} - \frac{1}{4}\sum_{jklrst}^{p} \frac{\partial^3 \hat{h}}{\partial z_j \partial z_k \partial z_l} \frac{\partial^3 \hat{h}}{\partial z_r \partial z_s \partial z_t} b_{jr} b_{kl} b_{st} \right. \\ \left. - \frac{1}{6}\sum_{jklrst}^{p} \frac{\partial^3 \hat{h}}{\partial z_j \partial z_k \partial z_l} \frac{\partial^3 \hat{h}}{\partial z_r \partial z_s \partial z_t} \frac{1}{6} b_{jr} b_{ks} b_{lt}\right]. \tag{8}$$

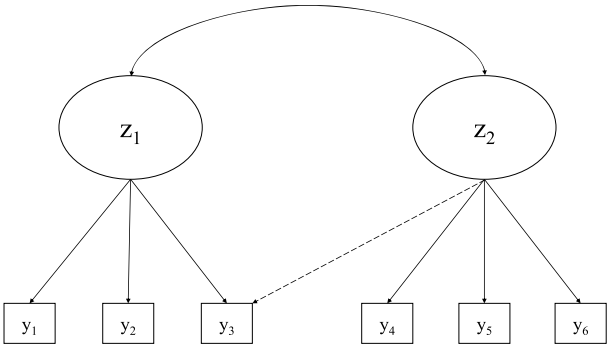

Fig. 1. Illustration of a model with six observed variables and two latent variables.

Denote the sums of Equation (8) as sum A, B and C, for the first, second and third entry, respectively. Without considering the model structure, Equation (8) requires the computation of $p^4 + p^6 + p^6$ entries which quickly becomes computationally demanding as the number of latent variables p increases. However, with models structured in a particular way, many of the terms needed for computing the entries in Equation (8) will be zero or repeated. For example, if all observed variables are each related to just one out of many latent variables, the expression reduces to a simple sum and a two-fold sum instead of the four-fold and six-fold sums in Equation (8) (Andersson and Xin, 2021; Noh and Lee, 2007). In our implementation of the second-order Laplace approximation, we avoid computing the same entries multiple times and identify the unique entries, which are products of multiple terms, in each of the sums in Equation (8). Such a procedure was also suggested in the supplementary material of Jin and Andersson (2020). We accomplish this with a computer algorithm at the first step of the estimation process which identifies the entries that must be computed and thus filters out repeated and zero entries. We subsequently weight the unique entries in accordance with the frequency of each entry in the sum.

In addition to avoiding zero and repeated entries in Equation (8), we exploit the expression of the function $h_{fg}(\boldsymbol{z})$ to gain further computational advantages. To illustrate this, first observe that

$$h_{fg}(\boldsymbol{z}) = -\sum_{i=1}^{I} \log P(y_{ifg}|\boldsymbol{z}; \boldsymbol{\alpha}_{ig}) - \log \phi(\boldsymbol{z}; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g), \tag{9}$$

where y_{ifg} denotes the ith observed variable for individual f in group g and where $\boldsymbol{\alpha}_{ig}$ is the parameter vector for the ith observed variable in group g. Define $h_{ifg}(\boldsymbol{z}) = -\log P(y_{ifg}|\boldsymbol{z}; \boldsymbol{\alpha}_{ig})$. Since derivatives of $\log \phi(\boldsymbol{z}; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ with respect to $\boldsymbol{z}$ of order three and higher are all zero, we have

$$\frac{\partial^u h_{fg}(\boldsymbol{z})}{\partial z_j \partial z_k \dots \partial z_v} = \sum_{i=1}^{I} \frac{\partial^u h_{ifg}(\boldsymbol{z})}{\partial z_j \partial z_k \dots \partial z_v}, \tag{10}$$

for $u > 2$. The implication of expressing the higher-order derivatives in this manner is that, while the higher-order derivative terms in the approximation of the likelihood in Equation (8) may not be zero or equal for a combination of $j, k, \dots, v$, in many cases a term $\frac{\partial^u h_{ifg}(\boldsymbol{z})}{\partial z_j \partial z_k \dots \partial z_v}$ for a single observed variable is indeed zero or equal for this combination of $j, k, \dots, v$. For example, consider the model with six observed variables and two latent variables displayed in Fig. 1 where the ellipses represent the latent variables and the rectangles represent the observed variables. For this model, the unique third derivatives of h_{fg} are $\frac{\partial^3 h_{fg}(\boldsymbol{z})}{\partial z_1^3}$, $\frac{\partial^3 h_{fg}(\boldsymbol{z})}{\partial z_1^2 \partial z_2}$, $\frac{\partial^3 h_{fg}(\boldsymbol{z})}{\partial z_1 \partial z_2^2}$, and $\frac{\partial^3 h_{fg}(\boldsymbol{z})}{\partial z_2^3}$. However, for all observed variables except y_3, the only non-zero third-derivatives are $\frac{\partial^3 h_{ifg}(\boldsymbol{z})}{\partial z_1^3}$, for $i \in \{1, 2\}$, and $\frac{\partial^3 h_{ifg}(\boldsymbol{z})}{\partial z_2^3}$, for $i \in \{4, 5, 6\}$. We thus also account for these patterns when computing the entries in the main approximation, beyond accounting for the symmetries that exist for the entries in the two four-fold and six-fold sums in Equation (8).

The filtering process described above means that the second-order Laplace approximation can be implemented with substantial efficiency gains compared to using, for example, adaptive Gauss-Hermite quadrature with the same order of accuracy. Note that, unlike in Shun (1997), we do not remove terms from the approximation in Equation (8) to improve the computational efficiency. Rather, we account for zero entries and symmetries that exist for the models we use.

2.2.2. Adaptive Gauss-Hermite quadrature

Let $\boldsymbol{\Gamma}_{fg}$ be the Cholesky decomposition of the matrix $\boldsymbol{H}_{fg}^{-1}$. The adaptive quadrature approximation to the log-likelihood is then (Jin and Andersson, 2020)

$$l_{fg}^{\text{AGHQ}}(\boldsymbol{\theta}_g|\boldsymbol{y}_{fg}) = \frac{p}{2}\log(2) - \frac{1}{2}\log\left|\boldsymbol{H}_{fg}\right| + \log \sum_{j_1,\dots,j_p}^{Q} \left[\prod_{k=1}^{p} w_{j_k} \exp\left(q_{j_k}^2\right)\right] \exp\left(h_{fg}(\boldsymbol{z})\big|^{\boldsymbol{z}=\sqrt{2}\boldsymbol{\Gamma}_{fg}\boldsymbol{q}_{j_1,\dots,j_p}+\hat{\boldsymbol{z}}_{fg}}\right), \tag{11}$$

where Q denotes the number of quadrature points per dimension, q_{j_k} is the j_k-th Gauss-Hermite quadrature point with weight w_{j_k} and $\boldsymbol{q}_{j_1,\ldots,j_p} = (q_{j_1}, \ldots, q_{j_p})'$. The theoretical approximation accuracy of adaptive Gauss-Hermite quadrature depends on the number of quadrature points and the error rate is given by $O\left(I^{-\lfloor (Q+2)/3 \rfloor}\right)$ (Jin and Andersson, 2020), implying that using four to six quadrature points has the same theoretical accuracy as the second-order Laplace approximation.

Adaptive Gauss-Hermite quadrature requires Q^p number of quadrature points, meaning that higher-dimensional models quickly become very computationally demanding to estimate. For example, a four-dimensional model requires a total of 81, 625, and 2401 quadrature points for $Q = 3, 5,$ and 7, respectively. Unlike for the number of entries required with the second-order Laplace approximation, the total number of quadrature points needed for a given level of accuracy is unaffected by the model structure.

2.3. Parameter estimation with the approximated likelihood

To estimate the unknown parameters, the gradient of the approximated log-likelihood is needed. We calculate the gradient $\nabla_{\boldsymbol{\theta}}$ of Equations (7) and (11) to obtain, for each $\theta \in \boldsymbol{\theta}$, and for each Method $\in$ {Lap2, AGHQ},

$$\nabla_\theta = \sum_{g=1}^{G} \sum_{f=1}^{N_g} \left(\frac{\partial l_{fg}^{\text{Method}}(\boldsymbol{\theta}_g | \boldsymbol{y}_{fg})}{\partial \theta} + \frac{\partial \hat{\boldsymbol{z}}_{fg}}{\partial \theta} \frac{\partial l_{fg}^{\text{Method}}(\boldsymbol{\theta}_g | \boldsymbol{y}_{fg})}{\partial \boldsymbol{z}} \Bigg|^{\boldsymbol{z}=\hat{\boldsymbol{z}}_{fg}^{\text{Method}}} \right), \tag{12}$$

where the second term in the expression is needed since the mode is dependent on the parameter vector $\boldsymbol{\theta}$ (Huber et al., 2004; Jin et al., 2018), and where $\hat{\boldsymbol{z}}_{fg}^{\text{Lap2}} = \hat{\boldsymbol{z}}_{fg}$ and $\hat{\boldsymbol{z}}_{fg}^{\text{AGHQ}} = \sqrt{2}\boldsymbol{\Gamma}_{fg}\boldsymbol{q}_{j_1,\ldots,j_p} + \hat{\boldsymbol{z}}_{fg}$. Note that AGHQ requires derivatives up to the third-order and the second-order Laplace requires derivatives up to the fifth-order. We analytically derived the derivatives for the generalized partial credit model and nominal response model and give the expressions of the constituents of Equation (12) for each of these in the appendix. The derivatives for the graded response model can be found in the supplementary material of Jin and Andersson (2020). Note that, with the second-order Laplace approximation, the model structure and symmetry of the derivatives also impact the computation of the gradient and just like for the computation of the entries in Equation (8), we compute only the unique entries in Equation (12) and weight them by their frequency.

With the gradient, we implement a quasi-Newton method for parameter estimation where the Hessian matrix is approximated with either the empirical cross-product matrix (Berndt et al., 1974) or the Broyden–Fletcher–Goldfarb–Shanno (Nocedal and Wright, 2006, BFGS) method. Let iter denote the iteration number and define α_{iter} as the step size in the quasi-Newton method. The algorithm proceeds as follows.

1. Let iter $= 0$ and define starting values $\hat{\boldsymbol{\theta}}_{\text{iter}}$.
2. With values $\hat{\boldsymbol{\theta}}_{\text{iter}}$, compute the posterior modes $\hat{\boldsymbol{Z}}$ and the gradient $\nabla_{\hat{\boldsymbol{\theta}}_{\text{iter}}}$.
3. Compute the approximated Hessian matrix $\boldsymbol{H}_{\text{iter}}$ from the gradient $\nabla_{\hat{\boldsymbol{\theta}}_{\text{iter}}}$.
4. Update the parameter estimates with $\hat{\boldsymbol{\theta}}_{\text{iter}+1} = \hat{\boldsymbol{\theta}}_{\text{iter}} + \alpha_{\text{iter}} \times \boldsymbol{H}_{\text{iter}}^{-1} \nabla_{\hat{\boldsymbol{\theta}}_{\text{iter}}}$ and let iter $=$ iter $+ 1$.
5. Repeat steps 2-4 until $\max|\hat{\boldsymbol{\theta}}_{\text{iter}} - \hat{\boldsymbol{\theta}}_{\text{iter}-1}| < \text{TOL}$.

We propose using starting values $a_{ijc} = 1.2$ for the slope parameters, an even sequence from $m_i - 2$ to $-(m_i - 2)$ for the intercept parameters (hence, starting value 0 if $m_i = 2$ and starting values 1 and -1 if $m_i = 3$) and $\sigma_{jk} = 0.5$, with TOL $=$ 0.0001 and $\alpha_{\text{iter}} = 1.0$ as default settings. If $\max|\boldsymbol{H}_{\text{iter}}^{-1} \nabla_{\hat{\boldsymbol{\theta}}_{\text{iter}}}| > 0.25$, we suggest to instead set $\alpha_{\text{iter}} = 0.25/\max|\hat{\boldsymbol{\theta}}_{\text{iter}} - \hat{\boldsymbol{\theta}}_{\text{iter}-1}|$ to avoid changing the parameter estimates too much in each iteration. Note that we directly maximize the approximated marginal log-likelihood function instead of using the EM (Dempster et al., 1977) algorithm. As a result, $\hat{\boldsymbol{z}}_{fg}$ is treated as a function of θ as implied in Equation (12).

2.4. Inference with the approximated likelihood

To draw inference we suggest using the inverse of the observed information matrix. The observed information matrix can be approximated by a numerical approximation to the Jacobian of the observed gradient in Equation (12) or the approximation from the BFGS algorithm. The results in Andersson and Xin (2021) indicated that using the numerical approximation to the Jacobian was accurate with correctly specified independent-clusters models and we therefore use this method in the current study. The numerical approximation to the Jacobian is obtained by defining an objective function with the unknown parameters as a vector-valued input argument, which computes and returns the exact observed gradient of the approximated log-likelihood, given in Equation (12). Before computing the gradient, the objective function updates the mode for each response pattern based on the parameters of the input argument. The Jacobian of this function is then approximated with a finite difference approach as implemented in the R package numDeriv (Gilbert and Varadhan, 2019). This method thus provides an approximation of the second derivatives of the approximated log-likelihood, taking the mode estimation into account when doing so. It is also possible to use the sandwich estimator, based on an approximation of the observed information matrix and the empirical cross-product matrix, to obtain robust standard errors.

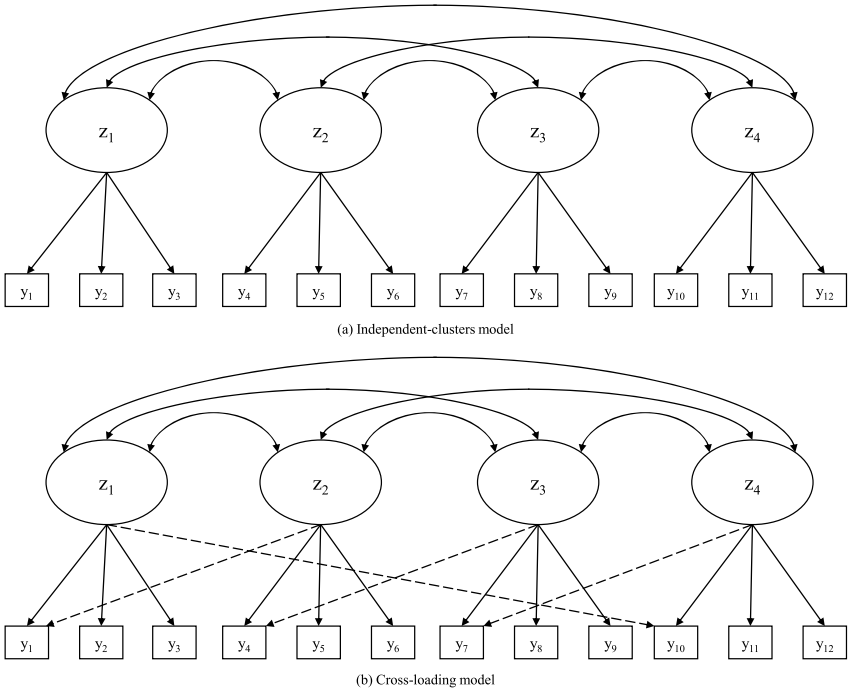

Fig. 2. Illustration of the two models with 12 observed variables used in the simulation study.

3. Simulation study

We performed a simulation study to investigate the parameter recovery and the computational efficiency of the second-order Laplace approximation (Lap2) method for multiple group models by comparing with the first-order Laplace approximation (Lap1) and adaptive Gauss-Hermite quadrature with three (AGHQ3) and five (AGHQ5) quadrature points. The R package lamle (Andersson and Jin, 2022) was used for parameter estimation with all methods.

3.1. Simulation design

In the simulations, we considered models with four latent variables and four groups, with 500 participants per group. Two settings were manipulated in the simulation design: (1) the type of multidimensional model - an independent-clusters model or a cross-loading model, and (2) the number of observed variables - three or four per dimension, for a total of either 12 or 16 observed variables. We considered only binary observed variables because the Laplace approximation has been shown in previous research to perform the worst in this setting (Joe, 2008; Andersson and Xin, 2021). We considered only the case of a low number of observed variables for the same reason. The conditions of experimental setting (1) represent between-item and within-item multidimensional models, respectively (Wang et al., 2004). In the between-item case, each observed variable is assumed to measure a single latent variable while in the within-item case some observed variables measure more than one latent variable via cross-loadings. Specifically, in the scenario with 12 observed variables, we added cross-loadings to one observed variable for each latent variable, resulting in four cross-loadings. Similarly, two observed variables in each dimension load on another dimension in the scenario with 16 observed variables, resulting in eight cross-loadings. The two models in the setting with 12 observed variables are presented in Fig. 2, where the ellipses represent the latent variables, the rectangles represent the observed variables, the solid lines with two arrows represent covariances between the latent variables, and the solid lines with one arrow represent the main loadings, with cross-loadings represented by dotted lines with one arrow. For both types of models, the covariances for the latent variables are freely estimated in each group and the mean vector and variances for the latent variables are freely estimated in each group except the first one, where the means and variances are all fixed to 0 and 1, respectively. The manipulation of the simulation settings led to $2 \times 2 = 4$ conditions. 1000 replications were conducted under each condition.

Data were generated in R version 4.1.1 (R Core Team, 2021). We simulated binary data using the graded response model with slope and intercept parameters given in Tables A.4-A.6 in the Appendix. We selected these parameters to have a setting which closely resembles real-life examples in educational and psychological measurement (Ayala, 2009). The latent variables were generated from a multivariate normal distribution with the mean vectors $(-1, -1, -1, -1)$, $(-.5, -.5, -.5, -.5)$, $(0, 0, 0, 0)$, and $(.25, .25, .25, .25)$ in the respective group, chosen to represent common differences in proficiency between age groups in practice. The covariances, which were identical for each group, were set to values between 0.4 and 0.6 and

Table 1
Convergence rates (in percent), average absolute bias, average coverage rate of 95% confidence intervals (in percent), average root mean squared error, average estimation time (in seconds), and average number of iterations for the four-dimensional multiple group models with sample size 2000 and different numbers of variables.

Outcome measure	Model	J	Lap1	Lap2	AGHQ3	AGHQ5
Convergence rate	Independent-clusters model	12	100	100	100	100
		16	100	100	100	100
	Cross-loading model	12	67.6	100	98.9	99.3
		16	99.7	100	100	100
Average absolute bias	Independent-clusters model	12	0.076	0.013	0.024	0.011
		16	0.050	0.010	0.017	0.009
	Cross-loading model	12	0.080	0.015	0.022	0.012
		16	0.048	0.010	0.016	0.009
Average coverage rate	Independent-clusters model	12	90.0	94.6	94.6	94.8
		16	92.1	95.0	94.9	95.0
	Cross-loading model	12	91.3	94.7	94.8	94.9
		16	92.4	94.8	94.7	94.8
Average root mean squared error	Independent-clusters model	12	0.040	0.027	0.025	0.027
		16	0.025	0.020	0.019	0.020
	Cross-loading model	12	0.047	0.030	0.029	0.031
		16	0.029	0.024	0.023	0.024
Average estimation time	Independent-clusters model	12	23.71	31.64	271.15	2182.70
		16	29.32	40.17	363.10	2987.91
	Cross-loading model	12	78.00	88.22	372.17	2437.80
		16	40.91	164.52	475.47	3365.19
Average number of iterations	Independent-clusters model	12	30.77	29.93	29.75	29.83
		16	30.67	30.32	30.20	30.31
	Cross-loading model	12	113.63	35.82	42.14	34.14
		16	38.09	31.63	31.89	31.56

Notes. J = number of observed variables, Lap1 = first-order Laplace, Lap2 = second-order Laplace, AGHQ3/AGHQ5 = adaptive Gauss-Hermite quadrature with 3 or 5 quadrature points.

are given in Table A.7 in the Appendix. The variances for the latent variables were fixed to one. Hence, the four groups varied in the means of the latent variables but shared the same correlations among the four dimensions. Note that we freely estimated the covariances in each group.

With the generated data, we employed the Lap1, Lap2, AGHQ3, and AGHQ5 estimation methods. To assess the performance of the four methods, we examined their statistical properties in terms of convergence rate, parameter recovery, and computational speed. Successful convergence was determined by fulfilling all of the following three criteria: 1) the algorithm stopped within 500 iterations, 2) the empirical cross-product matrix was positive definite, and 3) the approximated observed information matrix was positive definite. After concluding the simulation, we also inspected the parameter estimates and standard errors with each method to detect outlying replications. We computed the convergence rate in percent for each method and setting. Regarding parameter recovery, for a parameter θ with the estimate $\hat{\theta}^r$ in replication r, define the absolute bias as $|\text{bias}|_\theta = |\sum_{r=1}^{R}(\hat{\theta}^r - \theta)/R|$ and the root mean squared error (RMSE) as $\text{RMSE}_\theta = \sqrt{\sum_{r=1}^{R}(\hat{\theta}^r - \theta)^2/R}$. We computed overall measures of the recovery of the parameters (slopes, intercepts, covariances, variances, and means) in terms of average absolute bias, average RMSE, and average coverage rate of 95% confidence intervals estimated with the standard errors from the observed information matrix, across all parameters in a given setting. To evaluate the computational efficiency of the four estimation methods, we recorded the time information and the number of iterations required in the estimation. The computational efficiency is comparable between the methods since all methods are based on the same code base written in C++. The total simulation time exceeded 3600 core hours.

3.2. Results

In this subsection, we present the results of the simulation study. First in Table 1 are the convergence rates of Lap1, Lap2, AGHQ3, and AGHQ5 estimation methods for the four-dimensional multiple group models. It suggests that all the estimation methods attained 100% convergence for the independent-clusters models. However, with cross-loading models, the Lap2 method outperformed the other methods and was the only method to reach a 100% convergence rate. The Lap1 method was particularly problematic with respect to convergence (convergence rate = 67.6%) in the cross-loading scenario with 12 observed variables. We excluded the non-converged replications in subsequent comparisons of the methods.

Next, we evaluate the recovery of parameters and summarize the results of Table 1 concerning this. With respect to average absolute bias, Lap2 and AGHQ5 estimation methods produced less bias compared with Lap1 and AGHQ3. With an

Table 2
Number of unique nonzero 3rd and 4th derivatives and the number of unique nonzero entries in the sums of the second-order Laplace approximation for the four different models considered in the simulation.

Model	J	Unique 3rd	Unique 4th	Unique sum A	Unique sum B	Unique sum C
Independent-	12	4	4	4	10	10
clusters model	16	4	4	4	10	10
Cross-loading	12	12	16	20	170	114
model	16	16	22	28	322	214

Note. J = number of observed variables.

increasing number of observed variables, the average absolute bias decreased, especially for the Lap1 estimation method. The types of model, independent-clusters or cross-loading, did not impact estimation accuracy much. Regarding average coverage of 95% confidence intervals, Lap2, AGHQ3, and AGHQ5 showed similar results and performed better than Lap1. In addition, the manipulation settings barely had influence on this evaluation criterion. Besides the accuracy of parameter estimates, we also considered the estimation precision. The results suggest that compared with Lap1, other estimation methods produced more precise estimates, especially when the number of observed variables was 12. Increasing the number of observed variables or using a simpler model improved the average RMSE for all estimation methods. In general, the Lap2 estimation method exhibited similar estimation accuracy and precision with the AGHQ5 method, outperforming the AGHQ3 method and especially the Lap1 method.

Regarding computational efficiency, overall, the Lap1 method cost the least time (23.71 to 78.00 seconds per replication on average), followed by the Lap2 method (31.64 to 164.52 seconds), while the adaptive Gauss-Hermite quadrature methods required much longer time - over 270 seconds for AGHQ3 and more than 2100 seconds for AGHQ5 for all settings. Regarding the influence of the experimental factors, all the methods needed more time under the cross-loading conditions compared with the independent-clusters conditions. The computational time increased as the number of observed variables rose, except for the Lap1 method under the cross-loading conditions. The reason for this decrease is the need for additional iterations in the algorithm when using the Lap1 method in the cross-loading setting with 12 observed variables compared to 16 observed variables. It is worth mentioning that the simulation study in Bianconcini (2014) showed that Lap2 needs substantially more steps than AGHQ5, whereas our results show that they tend to converge within similar number of iterations.

To explain these computational results, it is useful to consider their relationship to the expression of the second-order Laplace approximation in terms of the unique quantities that are needed and the size of the sums included in the approximation. With an arbitrary four-dimensional model, the number of unique third derivatives of $h_{fg}(\boldsymbol{z})$ is at most 20 and the number of unique fourth derivatives of $h_{fg}(\boldsymbol{z})$ is at most 35. The sums in Equation (8) consist of entries that are products of these derivatives and the entries of the inverse of $\boldsymbol{H}_{fg}$. When ignoring symmetries and zero entries, sum A with fourth-order derivatives has $4^4 = 256$ entries, and sums B and C with third-order derivatives each have $4^6 = 4096$ entries.

In Table 2, we present the number of unique derivatives that must be computed for each model we considered and the number of unique entries required in the sums of Equation (8). The number of unique entries does not change when increasing the number of observed variables for the independent-clusters model because the model structure remains the same. This means that the computational time is essentially a linear function of the number of observed variables. However, since the model structure changes for the cross-loading model when increasing the number of observed variables from 12 to 16, there is an increased number of unique entries both for the derivatives and for the resulting sums. Nevertheless, the reduction in the number of entries is substantial compared to the unfiltered number since at most 28 out of the total 256 of sum A, 322 of the total 4096 of sum B, and 214 of the total 4096 of sum C have to be computed. This illustrates that the number of derivatives that must be computed for these models is still quite small compared the total which explains the high computational efficiency of the second-order Laplace approximation. Note that the function h_{fg} must be evaluated only once for each unique response pattern in the data for either the first- or second-order Laplace approximation whereas with adaptive quadrature it needs to be evaluated either 81 (AGHQ3) or 625 (AGHQ5) times for each unique response pattern in the data. As seen in the simulation study, this results in drastically increased computational time for adaptive quadrature relative to the Laplace approximations for the four-dimensional models considered.

In sum, the simulation study suggests that the Lap2 estimation method led to desirable outcomes in three aspects. First, unlike the Lap1 method, which suffered from non-convergence problems under some conditions, the Lap2 method achieved convergence in all the simulated data sets. Second, the Lap2 method yielded both accurate and precise parameter estimates, which was comparable to AGHQ5. Both Lap2 and AGHQ5 methods outperformed Lap1 and AGHQ3 methods in terms of parameter recovery. Third, the Laplace approximation methods greatly improved the computational speed compared to the adaptive Gauss-Hermite quadrature methods. Thus, we conclude that the Lap2 estimation method can produce satisfactory parameter estimates with a substantial improvement of computational efficiency compared to adaptive Gauss-Hermite quadrature for estimation of multidimensional multiple group models.

Table 3
Estimated latent means (se) of 2009 PISA mathematics literacy, reading literacy, and science literacy in Hong Kong (the reference group), Macao, Shanghai, and Chinese Taipei.

	Mathematics	Reading	Science
Hong Kong	0	0	0
Macao	-0.31 (0.03)	-0.58 (0.02)	-0.45 (0.02)
Shanghai	0.47 (0.03)	0.20 (0.03)	0.25 (0.03)
Chinese Taipei	-0.07 (0.03)	-0.50 (0.02)	-0.32 (0.03)

4. Empirical illustration

4.1. Data and models

To illustrate the proposed estimation method and compare against alternatives, we utilized data from Hong Kong, Macao, Shanghai and Chinese Taipei in the 2009 Programme for International Student Assesment (Schleicher et al., 2009, PISA). PISA is a large-scale educational assessment run by the Organisation for Economic Co-operation and Development (OECD) which aims to measure student achievement in mathematics, reading, and science, and monitor the outcomes of education systems internationally. We estimated three-dimensional multiple group (i.e., four regions) independent-clusters graded response models, where the dimensions corresponded to mathematics literacy, reading literacy and science literacy. The total number of respondents were 21690 but we removed 18 of these due to excessive numbers of missing values. The sample sizes in each region were 4792 in Hong Kong, 5948 in Macao, 5113 in Shanghai and 5819 in Chinese Taipei. There were 35 mathematics items, 100 reading items and 53 science items for a total of 188 items, out of which 176 items were binary scored and 12 were scored in three categories. Due to the PISA 2009 sampling design, each respondent only answered a subset of the total items included in the study and missing values are assumed missing at random. The average number of responses to the 188 total items was 54.55 across all regions. We simultaneously estimated the item parameters and the mean vectors and covariance matrices in each region using the item response data. Hong Kong was set as the reference group and the means and variances of the latent variable were not estimated in this group. The item parameters were considered invariant between regions. The model had nine free mean parameters, nine free variance parameters, 12 free covariance parameters, and 388 free item parameters, for a total of 418 parameters that were uniquely estimated.

4.2. Estimation settings

Estimation was done by maximizing the approximated log-likelihood, where the first-order Laplace (Lap1), second-order Laplace (Lap2) and adaptive Gauss-Hermite quadrature with 3, 5 and 13 quadrature points (AGHQ3, AGHQ5, and AGHQ13) were used for the integral approximations. The most accurate method out of these according to underlying theory is the AGHQ13 method, which has a fifth-order accuracy (Jin and Andersson, 2020). We thus used this as the reference method to compare the other methods against. The methods used the same starting values and estimation settings: tolerance of 0.0001, step size 0.5 for the first 25 iterations, maximum update direction 0.25, and maximum number of iterations 500. The R package lamle was used for all methods. All methods converged successfully after 36 iterations.

4.3. Results

4.3.1. Estimated distribution parameters

The estimated mean vectors and associated standard errors from the Lap2 method are shown in Table 3, where Hong Kong is set as the reference group and thus has a mean vector of 0 and variances equal to 1. The estimated means indicate that Shanghai is the highest performing region overall. The performance in reading and science are the highest in Hong Kong and Shanghai. Macao and Chinese Taipei have similar profiles, with slightly lower mean estimates compared to Hong Kong in all domains. The estimated covariance matrix for the latent mathematics, reading, and science literacy indicate that the domains are highly correlated, with estimated correlations between 0.86 and 0.92 in the reference group Hong Kong. The results for the three other regions were similar to Hong Kong and the full results are provided as covariance matrices in the Appendix.

4.3.2. Computational efficiency and accuracy

The log-likelihood values, estimation times and the parameter estimates were obtained after convergence. The log-likelihood values and estimation times are given in Table 4. These statistics reflect the results from the simulation study, in that the Lap1 method is the fastest but which approximates the log-likelihood the worst. Meanwhile, the Lap2 method is almost as fast as Lap1 while having a log-likelihood value that is very close to AGHQ13. Compared to AGHQ3, Lap2 improves both on the accuracy in terms of the log-likelihood value and the computational efficiency. AGHQ5 gives a log-likelihood value that is the closest to the reference AGHQ13, but the difference to Lap2 is small. The speed improvement of the Laplace-based methods relative to the adaptive quadrature methods is not as great for this three-dimensional model because of the

Table 4
Log-likelihood values and estimation times in seconds for five estimation methods with the 2009 PISA data from Hong Kong, Macao, Shanghai, and Chinese Taipei.

	Lap1	Lap2	AGHQ3	AGHQ5	AGHQ13
Log-likelihood	-627557.7	-627402.9	-627426.6	-627405.3	-627404.7
Estimation time	532.3	608.5	1800.7	6031.8	99081.1

Note. Lap1 = first-order Laplace, Lap2 = second-order Laplace, AGHQ3/AGHQ5/AGHQ13 = adaptive Gauss-Hermite quadrature with 3, 5, or 13 quadrature points.

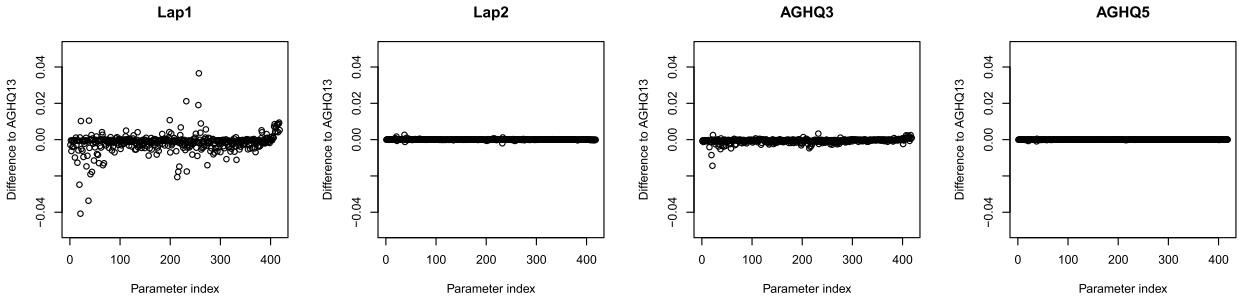

Fig. 3. Differences between parameter estimates obtained from Lap1, Lap2, AGHQ3, and AGHQ5, compared to parameter estimates from AGHQ13.

fewer number of total quadrature points needed with this model compared to a four-dimensional model. Nevertheless, the Lap2 method was almost three times faster than AGHQ3 and almost 10 times faster than AGHQ5 in estimation.

The log-likelihood values suggested that the methods provided slightly different results but this does not necessarily indicate if there are substantial differences in the parameter estimates from the different methods. To illustrate potential differences between the methods, we plotted the differences in the parameter estimates for Lap1, Lap2, AGHQ3 and AGHQ5 when compared to the AGHQ13 method. These differences are provided in Fig. 3, showing that Lap1 has the largest differences to the AGHQ13 method, followed by AGHQ3, Lap2, and AGHQ5. Overall, the differences in parameter estimates are small for all methods, differing at most by 0.0408 in absolute value for Lap1, 0.0027 in absolute value for Lap2, 0.0144 in absolute value for AGHQ3, and 0.0009 in absolute value for AGHQ5. This indicates that all of the methods can be considered sufficiently accurate in terms of parameter recovery in this particular setting, which is a consequence of the fairly large number of item responses by each student.

5. Discussion

Estimation of generalized linear latent variable models is computationally demanding in high dimensions which hinders their usage in many practical situations. In this study we implemented a second-order Laplace approximation method for estimation of generalized linear latent variable models with categorical observed variables and multiple groups. The practical consequences of our results are that the application of generalized linear latent variable models with high dimensionality is possible to do in situations with large sample sizes and with many parameters and that estimation time is reduced compared to alternative methods in other settings. In the numerical illustration, the second-order Laplace approximation method was highly computationally efficient compared to adaptive quadrature with three and five quadrature points per dimension. Meanwhile, the estimation accuracy was improved in relation to the first-order Laplace approximation and adaptive quadrature with three quadrature points while estimation accuracy was almost identical to that attained with five quadrature points. Since the second-order Laplace approximation has the same theoretical error rate as adaptive Gauss-Hermite quadrature with four to six quadrature points (Jin and Andersson, 2020), the second-order Laplace approximation was substantially more efficient than adaptive Gauss-Hermite quadrature at the same level of theoretical accuracy with the examples used in this study.

The results of this study thus imply that the Laplace approximation has a computational efficiency far above that of adaptive quadrature using a number of quadrature points which provides the same theoretical error rate. For the four-dimensional models considered here, the second-order Laplace approximation was also more efficient than adaptive quadrature with three quadrature points, which has a lower theoretical error rate than the second-order Laplace approximation. Generally speaking, the efficiency advantage will be lower with fewer latent variables and the efficiency advantage will be greater with more latent variables. The computational advantage of the second-order Laplace approximation is however also dependent on the complexity of the model structure. The highest efficiency gains are realized when each observed variable is related to only a single latent variable out of many correlated latent variables. The lowest gains are realized when specifying an unrestricted model with uncorrelated latent variables.

Our results can guide the practical use of latent variable models in the following ways. First, compared to the regular Laplace approximation the second-order Laplace approximation is preferred for most situations due to the added estimation accuracy. An exception is for the case of complex models with many indicators where the second-order method may still

be too time-consuming to utilize. With the most commonly used model structures, the second-order Laplace approximation is however fast enough to support up to 12 correlated latent variables (Andersson and Xin, 2021) which should cover most settings in practice. Second, we argue that the second-order Laplace approximation is especially useful for cases when adaptive quadrature with four or more quadrature points is impossible to practically conduct. If adaptive quadrature with many quadrature points is possible to employ, the second-order Laplace approximation becomes less suitable since a higher accuracy can be attained with adaptive quadrature by increasing the number of quadrature points.

Compared to alternatives such as adaptive quadrature and simulation-based methods, the second-order Laplace is attractive because of its high computational efficiency while maintaining a high accuracy. It is also attractive relative to the simulation-based methods due to the highly efficient computations of the log-likelihood and the observed information matrix, which can be extremely time-consuming for simulation-based methods when the sample size is large. Downsides to using the second-order Laplace approximation are that the computational advantage reduces for complex models and that the method requires the computation of a considerable amount of higher-order derivatives that do not easily generalize for different measurement models. It is also not straightforward with the Laplace approximation or adaptive quadrature to utilize the independence structure of the latent variables to improve the efficiency, as is possible with regular numerical quadrature (Gibbons and Hedeker, 1992; Cai, 2010b). However, regular numerical quadrature is unfeasible when the number of correlated latent variables is larger than three.

Future avenues of research include supporting additional types of observed variables combined with the categorical observed variables considered here. For example, providing an efficient yet accurate method that supports combinations of continuous data, count data, ordinal data and nominal data would be ideal to have. In addition, extensions of the approach to support mixture models and additional latent variable distributions beyond the normal distribution are possible.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2023.107710.

References

Andersson, B., Jin, S., 2022. lamle: maximum likelihood estimation of latent variable models using adaptive quadrature and Laplace approximations. https://github.com/bjoernhandersson/lamlepub/releases/tag/v0.1.2-alpha.

Andersson, B., Xin, T., 2021. Estimation of latent regression item response theory models using a second-order Laplace approximation. J. Educ. Behav. Stat. 46 (2), 244–265. https://doi.org/10.3102/1076998620945199.

Ayala, R., 2009. The Theory and Practice of Item Response Theory, Methodology in the Social Sciences. The Guilford Press, New York, NY.

Berndt, E.R., Hall, B.H., Hall, R.E., Hausman, J.A., 1974. Estimation and inference in nonlinear structural models. Ann. Econ. Soc. Meas. 3 (4), 653–665.

Bianconcini, S., 2014. Asymptotic properties of adaptive maximum likelihood estimators in latent variable models. Bernoulli 20 (3), 1507–1531. https://doi.org/10.3150/13-BEJ531.

Bianconcini, S., Cagnone, S., 2012. Estimation of generalized linear latent variable models via fully exponential Laplace approximation. J. Multivar. Anal. 112, 183–193. https://doi.org/10.1016/j.jmva.2012.06.005.

Bock, R.D., 1972. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika 37 (1), 29–51. https://doi.org/10.1007/BF02291411.

Bock, R.D., Aitkin, M., 1981. Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. Psychometrika 46 (4), 443–459. https://doi.org/10.1007/BF02293801.

Cagnone, S., Monari, P., 2013. Latent variable models for ordinal data by using the adaptive quadrature approximation. Comput. Stat. 28 (2), 597–619. https://doi.org/10.1007/s00180-012-0319-z.

Cai, L., 2010a. Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. J. Educ. Behav. Stat. 35 (3), 307–335. https://doi.org/10.3102/1076998609353115.

Cai, L., 2010b. A two-tier full-information item factor analysis model with applications. Psychometrika 75 (4), 581–612. https://doi.org/10.1007/s11336-010-9178-0.

Cho, A.E., Wang, C., Zhang, X., Xu, G., 2021. Gaussian variational estimation for multidimensional item response theory. Br. J. Math. Stat. Psychol. 74 (S1), 52–85. https://doi.org/10.1111/bmsp.12219.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc., Ser. B, Methodol. 39 (1), 1–38. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.

Gibbons, R.D., Hedeker, D.R., 1992. Full-information item bi-factor analysis. Psychometrika 57 (3), 423–436. https://doi.org/10.1007/BF02295430.

Gilbert, P., Varadhan, R., 2019. numDeriv: accurate numerical derivatives, r package version 2016.8-1.1. https://CRAN.R-project.org/package=numDeriv.

Huber, P., Ronchetti, E., Victoria-Feser, M.-P., 2004. Estimation of generalized linear latent variable models. J. R. Stat. Soc., Ser. B, Stat. Methodol. 66 (4), 893–908. https://doi.org/10.1111/j.1467-9868.2004.05627.x.

Hui, F.K.C., Warton, D.I., Ormerod, J.T., Haapaniemi, V., Taskinen, S., 2017. Variational approximations for generalized linear latent variable models. J. Comput. Graph. Stat. 26 (1), 35–43. https://doi.org/10.1080/10618600.2016.1164708.

Jin, S., Andersson, B., 2020. A note on the accuracy of adaptive Gauss–Hermite quadrature. Biometrika 107 (3), 737–744. https://doi.org/10.1093/biomet/asz080.

Jin, S., Noh, M., Lee, Y., 2018. H-likelihood approach to factor analysis for ordinal data. Struct. Equ. Model. 25 (4), 530–540. https://doi.org/10.1080/10705511.2017.1403287.

Jin, S., Vegelius, J., Yang-Wallentin, F., 2020. A marginal maximum likelihood approach for extended quadratic structural equation modeling with ordinal data. Struct. Equ. Model. 27 (6), 864–873. https://doi.org/10.1080/10705511.2020.1712552.

Joe, H., 2008. Accuracy of Laplace approximation for discrete response mixed models. Comput. Stat. Data Anal. 52 (12), 5066–5074. https://doi.org/10.1016/j.csda.2008.05.002.

Muraki, E., 1992. A generalized partial credit model: application of an EM algorithm. Appl. Psychol. Meas. 16 (2), 159–176. https://doi.org/10.1177/014662169201600206.

Muthen, B., Lehman, J., 1985. Multiple group IRT modeling: applications to item bias analysis. J. Educ. Behav. Stat. 10 (2), 133–142. https://doi.org/10.3102/10769986010002133.

Niku, J., Brooks, W., Herliansyah, R., Hui, F.K.C., Taskinen, S., Warton, D.I., 2019. Efficient estimation of generalized linear latent variable models. PLoS ONE 14 (5), 1–20. https://doi.org/10.1371/journal.pone.0216129.

Nocedal, J., Wright, S., 2006. Numerical Optimization. Springer-Verlag, New York, NY.

Noh, M., Lee, Y., 2007. REML estimation for binary data in GLMMs. J. Multivar. Anal. 98 (5), 896–915. https://doi.org/10.1016/j.jmva.2006.11.009.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rabe-Hesketh, S., Skrondal, A., Pickles, A., 2004. Generalized multilevel structural equation modeling. Psychometrika 69 (2), 167–190. https://doi.org/10.1007/BF02295939.

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. J. Comput. Graph. Stat. 9 (1), 141–157. https://doi.org/10.2307/1390617.

Samejima, F., 1969. Estimation of latent ability using a response pattern of graded scores. Psychometrika 34 (Suppl 1), 1–97. https://doi.org/10.1007/BF03372160.

Schilling, S., Bock, R.D., 2005. High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. Psychometrika 70 (3), 533–555. https://doi.org/10.1007/s11336-003-1141-x.

Schleicher, A., Zimmer, K., Evans, J., Clements, N., 2009. Pisa 2009 assessment framework: key competencies in reading, mathematics and science. OECD Publishing (NJ1).

Shun, Z., 1997. Another look at the salamander mating data: a modified Laplace approximation approach. J. Am. Stat. Assoc. 92 (437), 341–349. https://doi.org/10.1080/01621459.1997.10473632.

Thomas, N., 1993. Asymptotic corrections for multivariate posterior moments with factored likelihood functions. J. Comput. Graph. Stat. 2 (3), 309–322.

Wang, W.-C., Chen, P.-H., Cheng, Y.-Y., 2004. Improving measurement precision of test batteries using multidimensional item response models. Psychol. Methods 9 (1), 116–136. https://doi.org/10.1037/1082-989X.9.1.116.

Zhu, J., Eickhoff, J., Yan, P., 2005. Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. Biometrics 61 (3), 674–683. https://doi.org/10.1111/j.1541-0420.2005.00343.x.