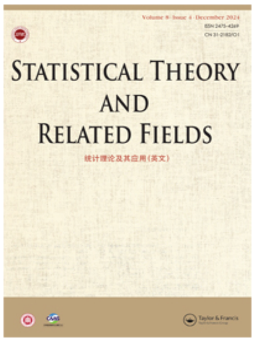


Statistical Theory and Related Fields

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tstf20

An efficient PG-INLA algorithm for the Bayesian inference of logistic item response models

Xiaofan Lin & Yincai Tang

To cite this article: Xiaofan Lin & Yincai Tang (21 Jan 2025): An efficient PG-INLA algorithm for the Bayesian inference of logistic item response models, Statistical Theory and Related Fields, DOI: 10.1080/24754269.2024.2442174

To link to this article: https://doi.org/10.1080/24754269.2024.2442174

Published online: 21 Jan 2025.

STATISTICAL THEORY AND RELATED FIELDS
https://doi.org/10.1080/24754269.2024.2442174

Taylor & Francis
Taylor & Francis Group

An efficient PG-INLA algorithm for the Bayesian inference of logistic item response models

Xiaofan Lin and Yincai Tang

KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai, People's Republic of China

ABSTRACT
In this paper, we propose a Bayesian PG-INLA algorithm which is tailored to both one-dimensional and multidimensional 2-PL IRT models. The proposed PG-INLA algorithm utilizes a computationally efficient data augmentation strategy via the Pólya-Gamma variables, which can avoid low computational efficiency of traditioanl Bayesian MCMC algorithms for IRT models with a logistic link function. Meanwhile, combined with the advanced and fast INLA algorithm, the PG-INLA algorithm is both accurate and computationally efficient. We provide details on the derivation of posterior and conditional distributions of IRT models, the method of introducing the Pólya-Gamma variable into Gibbs sampling, and the implementation of the PG-INLA algorithm for both one-dimensional and multidimensional cases. Through simulation studies and an application to the data analysis of the IPIP-NEO personality inventory, we assess the performance of the PG-INLA algorithm. Extensions of the proposed PG-INLA algorithm to other IRT models are also discussed.

ARTICLE HISTORY
Received 14 January 2024
Revised 4 June 2024
Accepted 1 December 2024

KEYWORDS
Item response theory; two-parameter logistic model; Pólya-Gamma; Gibbs sampler; integrated nested Laplace approximation

1. Introduction

Item response theory (IRT) (see Liu et al., 2018; F.M. Lord, 1980; Reckase, 2009), which often involves large-scale item level data that could be used to measure many latent variables, is important in data analysis in education, psychology, and other social science disciplines. IRT models have been extensively studied; see, for instance, in Chen and Zhang (2020), Embretson and Reise (2000), Thissen and Wainer (2001) and van der Linden and Hambleton (1997), and others. Most of the current IRT models developments use the logistic link and the probit link function. There are numurous algorithms for estimating the IRT model parameters. Two fundamental mechanisms most frequently used for IRT models are the variants of Expectation–Maximization (EM) algorithms (see Bock & Aitkin, 1981; Dempster et al., 1977) and the Markov Chain Monte Carlo (MCMC) methods (see Albert & Chib, 1993; Béguin & Glas, 2001; Cai, 2010a; Edwards, 2010; Jiang & Templin, 2018) such as the Gibbs algorithm and the Metropolis–Hastings (MH) algorithm.

The EM algorithm is a classical approach to obtain the maximum likelihood estimation for IRT models. The computational burden of the EM algorithm increases even when the number of latent traits K is only moderately large, as the computational difficulty of evaluating K-dimensional numerical integrals in the E step increases exponentially with K. Thus, based on the EM algorithm, various methods have been widely proposed, mainly including the quasi-Monte Carlo EM (QMCEM) algorithm (Niederreiter, 1978), the stochastic EM (StEM) algorithm (Celeux & Diebolt, 1985; Ip, 2002) and the Monte Carlo EM (MCEM) algorithm (Song & Lee, 2005). The QMCEM algorithm is more often applied to high-dimensional integration, but it is relatively low in efficiency in estimation when compared with some fully Bayesian methods such as the Metropolis–Hastings Robbins–Monro (MH-RM) algorithm (Cai, 2010a, 2010b).

The Gibbs algorithm and the MH algorithm are the two frequently employed Bayesian sampling algorithms. While the MH algorithm substitutes a proposal distribution for the true conditional distribution to realize the MCMC process, the Gibbs sampling is utilized when the full conditional posterior distributions of parameters can be easily sampled (Lynch, 2007). Because the full conditional posterior distributions of the parameters of the two-parameter logistic (2-PL) IRT model (F. Lord et al., 1968) have closed forms, Jiang and Templin (2018) proposed a PG-MCMC algorithm using pure Gibbs sampling to achieve data augmentation by introducing the Pólya-Gamma variables. However, the algorithm has a lengthy running time.

The maximum marginal likelihood estimation (MMLE), which can be obtained by the EM algorithm, the QMCEM algorithm and the MH-RM algorithm, remains a tricky numerical issue. The most difficult challenge

CONTACT Yincai Tang yctang@stat.ecnu.edu.cn KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai 200062, People's Republic of China

comes from having to approximate difficult high-dimensional intractable integrals in the likelihood function for the item parameters. In order to simplify the estimation process of the item parameters in the IRT models, we propose using the integrated nested Laplace approximation (INLA) method (Rue et al., 2009; Sara & Andrea, 2020). The INLA, which was proposed by Rue et al. (2009), combines Laplace approximation with numerical integration methods. INLA achieves the accuracy of MCMC and the computational efficiency of variational Bayes (VB) (Jaakkola & Jordan, 2000), and has been widely used to approximate the posterior distribution in recent years. For the IRT model, there is few literature about the application of INLA in this field. Mair and Gruber (2022) proposed a more general explanatory additive IRT model that uses INLA to estimate the parameters in the model, building a modern Bayesian framework in a fast and accurate way. Murphy (2021) extended the dyadic IRT model to multiple group designs. Due to computational constraints, he used the fast INLA method to perform Bayesian inference.

In this paper, we propose the PG-INLA algorithm, which combines and takes full advantage of the data augmentation strategy of Pólya-Gamma with the INLA, tailored specifically for the 2-PL IRT model. In fact, drawing 50000 points from $PG(1, 3.2)$ is roughly 15 times faster than drawing the same number of points from truncated normal distribution $TN(0, 1)$, according to a small-scale simulation by Polson et al. (2013). On the other hand, INLA is quick even for large, complex models and does not experience slow convergence and subpar mixing performance because of its deterministic nature (Sara & Andrea, 2020), which is usually a major problem that prevents the adoption of MCMC for such large complex models as the multidimensional 2-PL IRT models discussed in this paper.

The remainder of the paper is structured as follows. The 2-PL IRT model and the key idea of related algorithms are presented in Section 2. Then in Section 3, the PG-INLA algorithms are proposed in one-dimensional and multi-dimensional cases. Section 4 conducts some simulation studies as a way to illustrate the accuracy and effectiveness of the PG-INLA algorithm. A real application of the proposed PG-INLA algorithm is presented in Section 5. Section 6 provides some concluding remarks and possible extensions.

2. 2-PL IRT model and related algorithms

Assume that N individuals respond to J items, with Y_{ij} representing the reaction from individual i to item j and $Y = (Y_{ij})_{N\times J}$ representing the binary data matrix in which all the elements are either 0 or 1. The 2-PL IRT model is often used to model binary data which has the form of

$$P(Y_{ij} = 1 \mid \boldsymbol{a}_j, \boldsymbol{\theta}_i, d_j) = \frac{\exp(\boldsymbol{a}_j^\top \boldsymbol{\theta}_i + d_j)}{1 + \exp(\boldsymbol{a}_j^\top \boldsymbol{\theta}_i + d_j)}, \tag{1}$$

when an item j is parameterized in discrimination/difficulty form, $\boldsymbol{a}_j$ is the discrimination parameter, d_j is the difficulty parameter, and $\boldsymbol{\theta}_i$ is the continuous ability parameter (latent trait) of individual i. Both $\boldsymbol{a}_j$ and $\boldsymbol{\theta}_i$ have dimension K, and if $K \geq 2$, then we obtain a multidimensional 2-PL (M2PL) IRT model (Reckase, 2009).

2.1. INLA algorithm

In contrast to the MCMC, INLA is a fast, accurate and less computationally demanding method for performing approximate Bayesian inference in latent Gaussian models (Rue et al., 2009). A latent Gaussian model is of the form

$$\begin{aligned} \boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\beta} &\sim \prod_{i=1}^{N} \pi(y_i \mid x_i, \boldsymbol{\beta}), \\ \boldsymbol{x} \mid \boldsymbol{\beta} &\sim N(\boldsymbol{\mu}(\boldsymbol{\beta}), Q(\boldsymbol{\beta})^{-1}), \\ \boldsymbol{\beta} &\sim \pi(\boldsymbol{\beta}), \end{aligned} \tag{2}$$

where $\boldsymbol{y}$ is a vector of observed data, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_M)^\top$ is a vector of hyperparameters, and $\boldsymbol{x}$ represents the latent Gaussian field with mean $\boldsymbol{\mu}(\boldsymbol{\beta})$ and precision matrix $Q(\boldsymbol{\beta})$. In many cases, it is assumed that observations $\{y_i\}_{i=1}^N$ belong to an exponential family with mean $u_i = g^{-1}(\eta_i)$, where η_i is the linear predictor and g^{-1} is the link function. To derive the marginal approximate posterior distribution of each component in $\boldsymbol{x}$ and $\boldsymbol{\beta}$, INLA combines analytical approximations and numerical integration effectively.

Specifically, the core steps of the INLA algorithm for the latent Gaussian models have the following three steps. For more details on the INLA method, the readers may refer to Rue et al. (2009) and Sara and Andrea (2020).

Step 1. Derive the nested marginal posterior distribution for each hyperparameter, i.e. $\pi(\beta_m \mid \boldsymbol{y})$, $m = 1, 2, \ldots, M$, which is

$$\pi(\beta_m \mid \boldsymbol{y}) = \int \pi(\boldsymbol{\beta} \mid \boldsymbol{y}) \, \mathrm{d}\boldsymbol{\beta}_{-m}, \tag{3}$$

where $\boldsymbol{\beta}_{-m}$ is $\boldsymbol{\beta}$ excluding β_m. That is, $\pi(\beta_m \mid \boldsymbol{y})$ is approximated by approximating $\pi(\boldsymbol{\beta} \mid \boldsymbol{y})$ and integrating out $\boldsymbol{\beta}_{-m}$. More elaborately, the posterior density of the hyperparameter is approximated by using a Gaussian approximation for the posterior of the latent field, $\tilde{\pi}_G(\boldsymbol{x} \mid \boldsymbol{\beta}, \boldsymbol{y})$, evaluated at the posterior mode, $\boldsymbol{x}^*(\boldsymbol{\beta}) = \arg\max_{\boldsymbol{x}} \pi_G(\boldsymbol{x} \mid \boldsymbol{\beta}, \boldsymbol{y})$ (Moraga, 2020; Rue et al., 2009),

$$\tilde{\pi}(\boldsymbol{\beta} \mid \boldsymbol{y}) \propto \left. \frac{\pi(\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{y})}{\tilde{\pi}_G(\boldsymbol{x} \mid \boldsymbol{\beta}, \boldsymbol{y})} \right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\beta})}. \tag{4}$$

Then, INLA constructs the following nested approximations

$$\tilde{\pi}(\beta_m \mid \boldsymbol{y}) = \int \tilde{\pi}(\boldsymbol{\beta} \mid \boldsymbol{y}) \, \mathrm{d}\boldsymbol{\beta}_{-m}. \tag{5}$$

Step 2. Similarly, derive the nested marginal posterior distributions for each element x_i of the latent field $\boldsymbol{x}$, which can be written as

$$\pi(x_i \mid \boldsymbol{y}) = \int \pi(x_i \mid \boldsymbol{\beta}, \boldsymbol{y}) \pi(\boldsymbol{\beta} \mid \boldsymbol{y}) \, \mathrm{d}\boldsymbol{\beta}, \tag{6}$$

by combining analytical approximations to the full conditional distributions $\pi(x_i \mid \boldsymbol{\beta}, \boldsymbol{y})$ and $\pi(\boldsymbol{\beta} \mid \boldsymbol{y})$ and numerical integration with respect to $\boldsymbol{\beta}$. Thus, this nested formula can be used to approximate $\pi(x_i \mid \boldsymbol{y})$. There are usually three methods to approximate $\pi(x_i \mid \boldsymbol{\beta}, \boldsymbol{y})$: a Gaussian, a Laplace, and a simplified Laplace approximation. Combined with Step 1, then we obtain the following nested approximations

$$\tilde{\pi}(x_i \mid \boldsymbol{y}) = \int \tilde{\pi}(x_i \mid \boldsymbol{\beta}, \boldsymbol{y}) \tilde{\pi}(\boldsymbol{\beta} \mid \boldsymbol{y}) \, \mathrm{d}\boldsymbol{\beta}. \tag{7}$$

Step 3. Perform numerical integration. The nested approximations (5) and (7) can be integrated numerically with respect to $\boldsymbol{\beta}$

$$\tilde{\pi}(\beta_m \mid \boldsymbol{y}) = \sum_k \tilde{\pi}((\boldsymbol{\beta})_k \mid \boldsymbol{y}) \times \Delta_k, \tag{8}$$

$$\tilde{\pi}(x_i \mid \boldsymbol{y}) = \sum_l \tilde{\pi}(x_i \mid (\boldsymbol{\beta})_l^*, \boldsymbol{y}) \tilde{\pi}((\boldsymbol{\beta})_l^* \mid \boldsymbol{y}) \times \Delta_l^*, \tag{9}$$

where Δ_k and Δ_l^* denote the area weight corresponding to $(\boldsymbol{\beta})_k$ and $(\boldsymbol{\beta})_l^*$, respectively.

For the 2-PL IRT model, each Y_{ij} belongs to the binomial distribution given $\boldsymbol{\theta}_i$ $(i = 1, 2, \ldots, N)$

$$\begin{aligned} Y_{ij} &\sim \mathrm{Bin}(1, p_{ij}), \\ p_{ij} &= \frac{\exp(\boldsymbol{a}_j^\top \boldsymbol{\theta}_i + d_j)}{1 + \exp(\boldsymbol{a}_j^\top \boldsymbol{\theta}_i + d_j)}. \end{aligned} \tag{10}$$

Therefore, the linear predictor is $\eta_{ij} = d_j + \boldsymbol{a}_j^\top \boldsymbol{\theta}_i$ and g^{-1} is the logit function. In fact, the likelihood for $\boldsymbol{y}$ depends on $\boldsymbol{x}_j = (d_j, \boldsymbol{a}_j)$ only through the linear predictor η_{ij}. The `R-INLA` package provides defaults as in this case with no integrals as there are no hyperparameters, which simplifies the problem.

Given $\boldsymbol{\theta}_i, i = 1, 2, \ldots, N$, the 2-PL IRT model is a latent Gaussian model. So we can easily fit the 2-PL IRT model using the INLA algorithm. First we need to obtain the `formula` object in `R` according to the form of η_{ij}. For each item j in this 2-PL IRT model, the linear predictor is $\eta_{ij} = d_j + \boldsymbol{a}_j^\top \boldsymbol{\theta}_i$, so the `formula` is $\boldsymbol{y}_j \sim 1 + \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2 + \cdots + \boldsymbol{\theta}_K$, with given $\boldsymbol{\theta}_k = (\theta_{1k}, \theta_{2k}, \ldots, \theta_{Nk})^\top, k = 1, 2, \ldots, K$. And then the latent Gaussian fields, i.e. the intercept d_j and the fixed effects $a_{j1}, a_{j2}, \ldots, a_{jK}$, $j = 1, 2, \ldots, J$, are assigned Gaussian priors $N(0.01, 10)$. Finally, the INLA would follow the above three steps to obtain the final parameter estimates. A minimum of $K(K-1)/2$ restrictions must be placed on the elements of discrimination parameters $\boldsymbol{a}_j$ in order to identify the model. Either multivariate constraints must be applied (Lawley & Maxwell, 1962), or constraints can be made by setting some elements of discrimination parameters $\boldsymbol{a}_j$ to 0 (McDonald, 2013).

2.2. Data augmentation with Pólya-Gamma random variable

This section introduces the idea of the data augmentation method, which is particularly useful for the Bayesian inference of logistic regression. Provided that the number of successes $y_i (i = 1, 2, \dots, N)$ is binomial with success probability $p = \frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})}$ and the number of trials n_i. Polson et al. (2013) demonstrated that the likelihood contribution of observation j in the logistic regression model can be expressed as

$$\begin{aligned} L_i(\boldsymbol{\beta}) &= \frac{\{\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})\}^{y_i}}{\{1+\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})\}^{n_i}} \\ &\propto \exp\left(\kappa_i \boldsymbol{x}_i^\top \boldsymbol{\beta}\right) \int_0^\infty \exp\left(-\frac{\omega_i (\boldsymbol{x}_i^\top \boldsymbol{\beta})^2}{2}\right) p(\omega_i; n_i, 0)\, \mathrm{d}\omega_i, \end{aligned} \tag{11}$$

where $\kappa_i = y_i - \frac{n_i}{2}$, $\boldsymbol{x}_i$ is a vector of predictors for observation i, ω_i is a Pólya-Gamma random variable with parameters $(n_i, 0)$ and $p(\omega_i \mid n_i, 0)$ is its density.

If a random variable γ has a Pólya-Gamma distribution with parameters $b > 0$ and $c \in \mathbb{R}$, which is denoted by $\gamma \sim \mathrm{PG}(b, c)$, then it is equivalent to an infinite weighted sum of gamma random variables

$$\gamma \overset{\mathrm{d}}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{G(b, 1)}{(k-0.5)^2 + c^2/(4\pi^2)}, \tag{12}$$

where $G(b, 1)$ represents the gamma distribution with parameters b and 1.

From Equation (11), we can consider ω_i as an augmented random variable for the data y_i. Biane et al. (1999) proved that if we have a prior distribution $p(\boldsymbol{\beta})$ for $\boldsymbol{\beta}$, then conditioning on a set of Pólya-Gamma random variables $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_N)^\top$, the conditional posterior density of $\boldsymbol{\beta}$ is

$$\begin{aligned} p(\boldsymbol{\beta} \mid \boldsymbol{\omega}, \boldsymbol{y}) &\propto p(\boldsymbol{\beta}) \prod_{i=1}^{N} \exp\left\{\kappa_i \boldsymbol{x}_i^\top \boldsymbol{\beta} - \frac{\omega_i (\boldsymbol{x}_i^\top \boldsymbol{\beta})^2}{2}\right\} \\ &\propto p(\boldsymbol{\beta}) \prod_{i=1}^{N} \exp\left\{-\frac{w_i}{2}\left(\boldsymbol{x}_i^\top \boldsymbol{\theta} - \frac{\kappa_i}{\omega_i}\right)^2\right\} \\ &\propto p(\boldsymbol{\beta}) \exp\left\{-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta})^\top \Omega (\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta})\right\}, \end{aligned} \tag{13}$$

where $\boldsymbol{y} = (y_1, y_2, \dots, y_N)^\top$, $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_N)^\top$, $\boldsymbol{z} = \left(\frac{\kappa_1}{\omega_1}, \frac{\kappa_2}{\omega_2}, \dots, \frac{\kappa_N}{\omega_N}\right)^\top$, $\boldsymbol{X} = (\boldsymbol{x}_1^\top, \boldsymbol{x}_2^\top, \dots, \boldsymbol{x}_N^\top)^\top$ and $\Omega = \mathrm{diag}\,(\omega_1, \omega_2, \dots, \omega_N)$. Furthermore, if the prior distribution $p(\boldsymbol{\beta})$ is set as $N(\boldsymbol{b}, \boldsymbol{B})$, where $\boldsymbol{b}$ is the mean vector and $\boldsymbol{B}$ is the covariance matrix, then Gibbs sampling below can be used to draw a posterior sample from the marginal posterior distribution of $\boldsymbol{\beta}$

$$\begin{aligned} \omega_i \mid \boldsymbol{\beta} &\sim \mathrm{PG}(n_i, \boldsymbol{x}_i^\top \boldsymbol{\theta}), \\ \boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{\omega} &\sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*). \end{aligned} \tag{14}$$

According to the conjugacy of the normal distribution between the prior distribution and the posterior distribution, it can be shown that $\boldsymbol{\Sigma}^* = (\boldsymbol{X}^\top \Omega \boldsymbol{X} + \boldsymbol{B}^{-1})^{-1}$ and $\boldsymbol{\mu} = \boldsymbol{\Sigma}^* (\boldsymbol{X}^\top \boldsymbol{\kappa} + \boldsymbol{B}^{-1}\boldsymbol{b})$, where $\boldsymbol{\kappa} = \left(y_1 - \frac{n_1}{2}, y_2 - \frac{n_2}{2}, \dots, y_N - \frac{n_N}{2}\right)^\top$ (Zeithammer & Lenk, 2006). This forms the basis for subsequent applications involving the IRT models.

3. PG-INLA algorithm for 2-PL IRT models

The PG-INLA algorithm we proposed in this paper is tailored to the 2-PL IRT models including both the one-dimensional and multidimensional ones. Similar to the EM algorithm, the core idea of this PG-INLA algorithm is also a two-step iteration: PG step and INLA step. In the PG step, given the estimates of the item parameters $\boldsymbol{a}_j$ and d_j, which are obtained from the previous INLA step, we use a computationally efficient data augmentation strategy to estimate the individual parameters $\boldsymbol{\theta}_i$ via the Pólya-Gamma distribution (Jiang & Templin, 2018) and Gibbs sampling. And if we know the estimates of the individual parameters $\boldsymbol{\theta}_i (i = 1, 2, \dots, N)$ from the PG step, then we can use the INLA algorithm to estimate the item parameters $\boldsymbol{a}_j$ and d_j easily and accurately. The iteration stops when the following conditions are met

$$\sqrt{\sum_{j=1}^{J}\sum_{k=1}^{K}\left(a_{jk}^{(t)}-a_{jk}^{(t-1)}\right)^2/M}\le\varepsilon,$$
$$\sqrt{\sum_{i=1}^{N}\sum_{k=1}^{K}\left(\theta_{ik}^{(t)}-\theta_{ik}^{(t-1)}\right)^2/(N\times K)}\le\varepsilon, \tag{15}$$

where M is the number of non-zero elements of $\boldsymbol{a}_j$ and ε is a positive number given in advance that is small enough, and we set it to 0.01 in the subsequent simulation studies.

3.1. One-dimensional 2-PL IRT model

For one-dimensional and multidimensional 2-PL IRT models, the Gibbs sampling part of the PG-INLA algorithm will be different. Bayesian inference of the one-dimensional 2-PL IRT models depends on the MCMC process, where parameter blocks are exactly used instead of directly sampling from the overall joint likelihood (Jiang & Templin, 2018). Blocks of parameters are converted into complete conditional forms in order to build Gibbs samplers for the one-dimensional 2-PL IRT models, as shown in Equation (14) (Junker et al., 2017). However, since the conditional posterior distributions lack expressions of closed form, the samplers for $\boldsymbol{\theta}_i$ cannot adopt the Gibbs samplers without the use of the Pólya-Gamma data augmentation strategy.

According to Equations (11) and (13), we can derive the Gibbs sampling process for estimating θ_i for the one-dimensional 2-PL IRT model using the Pólya-Gamma variables. Using the notations in Section 2, we denote y_{ij} as the actual response of individual i to item j and $\boldsymbol{\omega}_i=(\omega_{i1},\omega_{i2},\dots,\omega_{iJ})^\top$. For each θ_i, the conditional posterior distribution could be rewritten as follows

$$p(\theta_i\,|\,\boldsymbol{\omega}_i,a_j,d_j,\boldsymbol{y}_i)\propto p(\theta_i)\exp\left\{-\frac{1}{2}(\boldsymbol{z}_i-A_1\theta_i)^\top\boldsymbol{\Omega}_i(\boldsymbol{z}_i-A_1\theta_i)\right\}, \tag{16}$$

where $p(\theta_i)$ is $N(0,1)$, $\boldsymbol{\Omega}_i=\operatorname{diag}(\omega_{i1},\omega_{i2},\dots,\omega_{iJ})$, $A_1=(a_1,a_2,\dots,a_J)^\top$, $\boldsymbol{z}_i=\left(\frac{\kappa_{i1}-d_1\omega_{i1}}{\omega_{i1}},\frac{\kappa_{i2}-d_2\omega_{i2}}{\omega_{i2}},\dots,\frac{\kappa_{iJ}-d_J\omega_{iJ}}{\omega_{iJ}}\right)^\top$ and $\kappa_{ij}=y_{ij}-\frac{1}{2}$. With Equations (16) and (14), the use of Gibbs sampling becomes possible by the new forms of normal conditionals for θ_i as

$$\theta_i\,|\,\boldsymbol{\omega}_i,a_j,d_j,\boldsymbol{y}_i\sim N(V_1(A_1^\top\boldsymbol{\Omega}_i\boldsymbol{z}_i),V_1), \tag{17}$$

where $V_1=(A_1^\top\boldsymbol{\Omega}_iA_1+1)^{-1}$. And fro Equation (14), the conditional posterior distribution for ω_{ij} is

$$\omega_{ij}\,|\,\theta_i\sim \mathrm{PG}(1,a_j\theta_i+d_j). \tag{18}$$

Given the item parameters estimated from the previous INLA step, we can use Equations (17) and (18) to excute the Gibbs sampler to obtain the estimates of individual paramters θ_i. This PG-INLA algorithm is shown in Algorithm 1.

3.2. Multidimensional 2-PL IRT model

The multidimensional 2-PL model can incorporate multiple latent trait variables with the one-dimensional 2-PL model. The difference of parameter estimation of one-dimensional and multidimensional 2-PL IRT models is mainly reflected in the identifiability of the model and the sampling of covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{\theta}_i$. When constructing multidimensional item response theory (MIRT) models (Reckase, 2009), it is common to set each item to measure only a small number of latent traits, and the discrimination parameter $\boldsymbol{a}_j$ in each item that does not measure latent traits is set to 0. Thus, such models become identifiable by these non-estimated parameters.

In our PG-INLA algorithm, when the model is multidimensional, in order to consider the correlation information of the latent traits of each dimension, we consider adding the covariance matrix $\boldsymbol{\Sigma}$ update iterations in Gibbs sampling. Conjugate inverse Wishart distribution is often employed in a typical Gibbs sampler to draw samples from a covariance matrix. However, since we are estimating the individual parameters given the item parameters and the identifiability of the model, the current estimation on the individual parameters needs to satisfy that the

Algorithm 1: PG-INLA algorithm procedure (one-dimensional).

input: The observed data $\boldsymbol{y}$, the highest number of iteration times T before the PG-INLA algorithm converges, the number of Gibbs sampling iteration times S and burn.in times m, and the initial value $(\theta_i^{(0)}, a_j^{(0)}, d_j^{(0)})^\top$.
output: The posterior samples of $(\theta_i, a_j, d_j)^\top$.

```
for t ← 0 to T do
    PG step
    for i ← 1 to N do
        parallel computation for each i for s ← 1 to S do
            θ_i^(s−1) ← θ_i^(t) for j ← 1 to J do
                ω_ij^(s) ← f(ω_ij | θ_i^(s−1), a_j^(t−1), d_j^(t−1), y);
            end
            θ_i^(s) ← f(θ_i | ω_ij^(s), {a_j^(t−1), j = 1, 2, ..., J}, {d_j^(t−1), j = 1, 2, ..., J}, y);
        end
        θ_i^(t) ← Σ_{s=m+1}^{S} θ_i^s / (S − m)
    end
    INLA step for j ← 1 to J do
        a_j^(t) ← f(a_j | {θ_i^(t), i = 1, 2, ..., N}, y);
        d_j^(t) ← f(d_j | {θ_i^(t), i = 1, 2, ..., N}, y);
    end
    if Condition (15) is met, then
        (θ_i^(t), a_j^(t), d_j^(t))^T
    end
end
```

elements on the diagonal of the covariance matrix are 1. Therefore, the covariance matrix is switched into a correlation matrix with this identifiability condition, thus the inverse Wishart distribution is no longer suitable. An alternative method is to draw the sample covariance matrix from the inverse Wishart distribution firstly, and then convert it into a sample correlation matrix (Imai & Dyk, 2005). However, Lynch (2007) showed that the converting approach is not precise when the sample size is small or the off-diagonal elements are large.

In this paper, we use the MH algorithm proposed by Jiang and Templin (2018) instead of the previous methods mentioned to sample the correlation parameters $\sigma_{kk'}$, $k \neq k'$, $k, k' = 1, 2, \ldots, K$, where $\Sigma = (\sigma_{kk'})_{K\times K}$ is the prior covariance matrix. The transition kernel for correlation $\sigma_{kk'}^{(r)}$ is the symmetric distribution $N(\sigma_{kk'}^{(r-1)}, 0.05)$, where r indicates current step of sampling iteration. As can be seen from the simulation studies that follow, this approach guarantees the speed of the algorithm as well as a high estimation accuracy.

Similar to the Gibbs sampling process for the one-dimensional 2-PL IRT model, based on the above discussion, if $\boldsymbol{\theta}_i$ is assigned a prior $N(\mathbf{0}, \boldsymbol{\Sigma})$, then we can easily obtain the Gibbs sampling process for the M2PL IRT models as follows

$$\begin{gathered}\omega_{ij} \mid \boldsymbol{\theta}_i \sim \mathrm{PG}(1, \boldsymbol{a}_j^\top \boldsymbol{\theta}_i + d_j),\\ \boldsymbol{\theta}_i \mid \boldsymbol{\omega}_i, \sigma_{kk'}, \boldsymbol{a}_j, d_j, \boldsymbol{y}_i \sim N(V_2(A_2^\top \Omega_i \boldsymbol{z}_i), V_2),\\ \sigma_{kk'} \mid \boldsymbol{\theta}_i \sim N(\sigma_{kk'}, 0.05),\end{gathered} \tag{19}$$

where $V_2 = (A_2^\top \Omega_i A_2 + \Sigma^{-1})^{-1}$ and $A_2 = (\boldsymbol{a}_1^\top, \boldsymbol{a}_2^\top, \ldots, \boldsymbol{a}_J^\top)^\top$. The whole PG-INLA algorithm for the M2PL IRT models is presented in Algorithm 2.

4. Simulation study

To demonstrate the effectiveness of the suggested PG-INLA algorithm in parameter estimation, we do some simulation studies in this section. In practice, the EM algorithm is generally effective with 1–3 latent traits, but

Algorithm 2: PG-INLA algorithm procedure (multidimensional).

input : The observed data $\boldsymbol{y}$, the highest number of iteration times T before the PG-INLA algorithm converges, the number of Gibbs sampling iteration times S and burn.in times m, the initial value $(\boldsymbol{\theta}_i^{(0)\top}, \boldsymbol{a}_j^{(0)\top}, d_j^{(0)})^\top$, and Σ^0.

output: The posterior samples of $(\boldsymbol{\theta}_i^\top, \boldsymbol{a}_j^\top, d_j)^\top$ and Σ.

```
for t ← 0 to T do
    PG step for i ← 1 to N do
        parallel computation for each i for s ← 1 to S do
            θ_i^(s−1) ← θ_i^(t) for j ← 1 to J do
                ω_ij^(s) ← f(ω_ij | θ_i^(s−1), a_j^(t−1), d_j^(t−1), y);
            end
            Σ^(s) ~ f(Σ | {θ_i, i = 1, 2, ..., N})
            θ_i^(s) ← f(θ_i | ω_ij^(s), Σ^(s), {a_j^(t−1), j = 1, 2, ..., J}, {d_j^(t−1), j = 1, 2, ..., J}, y);
        end
        θ_i^(t) ← Σ_{s=m+1}^S θ_i^s / (S − m)
    end
    INLA step for j ← 1 to J do
        a_j^(t) ← f(a_j | {θ_i^(t), i = 1, 2, ..., N}, y);
        d_j^(t) ← f(d_j | {θ_i^(t), i = 1, 2, ..., N}, y);
    end
    if Condition (15) is met then
        (θ_i^(t)⊤, a_j^(t)⊤, d_j^(t))⊤ and Σ^(t)
    end
end
```

methods such as the QMCEM algorithm and the MH-RM algorithm should be used when the dimensions are 3 or more. So our simulation studies consist of the following two parts: (1) in the one-dimensional situation, we compare the PG-INLA algorithm with the EM algorithm to show the accuracy and efficacy of the proposed method, and (2) in the second part, the proposed PG-INLA is compared with the QMCEM algorithm and the MH-RM algorithm to show that the PG-INLA algorithm can also be applied to the multidimensional IRT model effectively.

The `R` programming language is used throughout these simulation studies to create data, construct algorithms, define and execute functions, and estimate model parameters. And we use the `mirt` package for parameter estimation for the EM algorithm, the QMCEM algorithm, and the MH-RM algorithm, which are all marginal maximum likelihood algorithms. The `mirt` package provides a toolkit for various `mirt`-related estimate tasks and has extensive citations in many published works (see Eckes & Baghaei, 2015; Matlock et al., 2018; Zhang et al., 2020). More specifically, all the tolerance levels (stop criteria) in the `mirt` are set to 0.01 by default. The PG-INLA algorithm is carried out at a Linux server with 64-core processors, where the functions for the INLA algorithm are provided by the `R-INLA` package.

The point estimates produced by the Bayesian frameworks are referred to as the posterior means throughout all the simulation studies. Due to the extremely weak correlation between the samples of each $\boldsymbol{\theta}_i$, $i = 1, 2, \ldots, N$ and ω_{ij}, $i = 1, 2, \ldots, N, j = 1, 2, \ldots, J$, the Gibbs sampler iteration number is set to 100, where the first 50 iterations are burned. In fact, the final accuracy of the three parameters is close to the setting of the Gibbs sampling with 5000 iterations.

4.1. One-dimensional 2-PL IRT model

We examine the impact of parameter estimation for the EM algorithm and the PG-INLA algorithm when 1000 individuals respond to 40 items in the simulation study of the one-dimensional 2-PL IRT model (i.e. $N = 1000, J = 40$).

Following the notations in Equations (16)–(18), the true individual parameter θ_i is generated by the standard normal distribution $N(0, 1)$, and the true item parameters a_j and d_j are yielded by the distributions $\mathrm{LN}(0.3, 0.2)$ and $N(0, 1)$, respectively. This data generation mechanism is suggested in many papers (see Feinberg & Rubright, 201606; Harwell & Baker, 1991; Jiang & Templin, 2018, for some examples). To compare the estimation accuracy we compute the root mean squared errors (RMSEs) for the estimates of three parameters over 100 replications. For example, the RMSE $(\hat{\boldsymbol{\theta}})$ is defined as

$$\mathrm{RMSE}\ (\hat{\boldsymbol{\theta}}) = \sqrt{\frac{1}{100}\frac{1}{N}\sum_{n=1}^{100}\sum_{i=1}^{N}\left(\hat{\theta}_i^{(n)} - \theta_i^*\right)^2}, \tag{20}$$

where $\hat{\theta}_i^{(n)}$ is the estimate of θ_i from the nth replications and θ_i^* denotes the true value. And similarly, we can define the RMSEs for the item parameters a_j and d_j.

Results are summarized in Table 1 and Figures 1 and 2. From Table 1, the RMSEs between the two algorithms are all very close, which indicates that the parameter estimation from the PG-INLA algorithm is comparable to that from the EM algorithm. Furthermore, the PG-INLA outperforms the EM algorithm in the estimation of discrimination parameters a_j.

Figure 1 shows the boxplots of the mean squared errors (MSEs) for individual parameters θ_i and item parameters a_j and d_j over 100 replications. For instance, the MSE for the estimate of the parameter θ_1 is

$$\mathrm{MSE}\ (\hat{\theta}_1) = \frac{1}{100}\sum_{n=1}^{100}\left(\hat{\theta}_1^{(n)} - \theta_1^*\right)^2, \tag{21}$$

where $\hat{\theta}_1^{(n)}$ is the estimate of θ_1 from the nth replications and θ_1^* is the true value. The left, middle, and right panels of Figure 1 correspond to the three parameters, individual paramteres θ_i, discrimination parameters a_j and difficulty parameters d_j, respectively. In the left panel, the boxplots with labels 'EM.θ' and 'PGINLA.θ' are based on the MSEs for $\theta_1, \theta_2, \ldots, \theta_N$ from the EM algorithm and the PG-INLA algorithm, respectively. Similar labels are given for the boxplots in the middle and right panels. We see that the MSEs of the parameter estimates based on the two algorithms are quite close.

Figure 2 shows the one-dimensional density plots of the real value of θ_i, the estimated value of the EM algorithm and the PG-INLA algorithm, from which it can be seen that the overall distribution of the estimated results of the PG-INLA algorithm is very close to the real distribution. According to Figures 1 and 2, for all the model parameters, the point estimates given by the PG-INLA algorithm and that given by the EM algorithm are almost the same.

4.2. Multidimensional 2-PL IRT model

For multidimensional situations, we compare the proposed algorithm with the QMCEM algorithm and the MH-RM algorithm implemented in the `mirt` package, in two settings: (1) $K = 2, J = 80$ and $N = 1000$; and (2) $K = 4$,

Table 1. The RMSEs for the estimates of the parameters θ_i, a_j and d_j obtained from the EM algorithm and the PG-INLA algorithm.

	EM	PG-INLA
RMSE($\hat{\boldsymbol{\theta}}$)	0.277	0.287
RMSE($\hat{\boldsymbol{a}}$)	0.127	0.114
RMSE($\hat{\boldsymbol{d}}$)	0.079	0.085

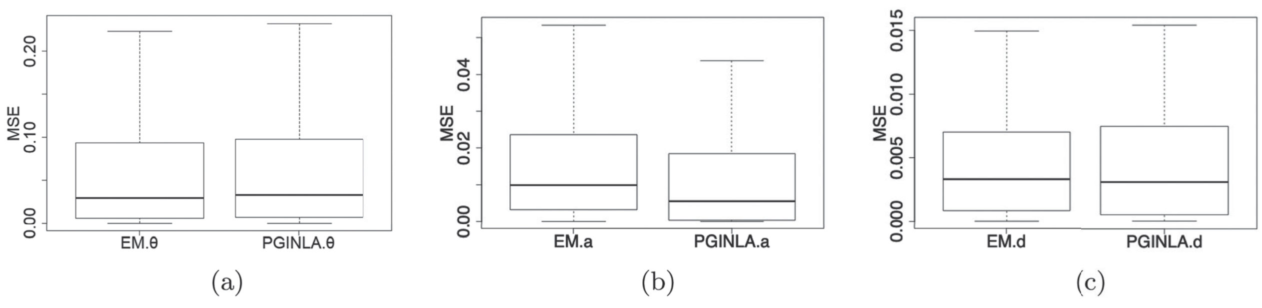


Figure 1. Boxplots of MSEs of the estimated parameters for the EM algorithm and the PG-INLA algorithm: (left) MSEs of individual parameters θ_i; (middle) MSEs of discrimination parameter a_j; (right) MSEs of difficulty parameters d_j. (a) θ_i. (b) a_j and (c) d_j.

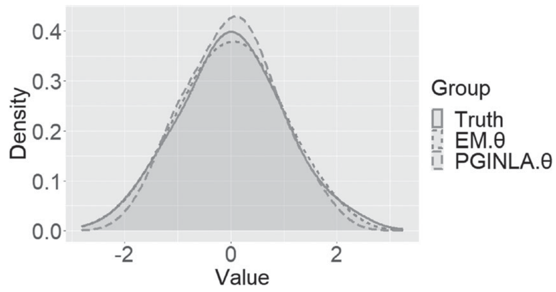


Figure 2. Density plots of the true value of θ_i, the estimated value of the EM algorithm and the PG-INLA algorithm.

Table 2. The RMSE for $\boldsymbol{a}_j$, $\boldsymbol{d}_j$, $\boldsymbol{\theta}_i$ and $\sigma_{kk'}$, $k \neq k'$ for $K = 2$ and $K = 4$.

	QMCEM	MH-RM	PG-INLA	QMCEM	MH-RM	PG-INLA
Algorithm		$K = 2$			$K = 4$	
RMSE($\hat{\boldsymbol{\theta}}$)	0.301	0.293	0.286	0.326	0.311	0.272
RMSE($\hat{\boldsymbol{a}}$)	0.124	0.083	0.101	0.130	0.073	0.084
RMSE($\hat{d}$)	0.098	0.087	0.134	0.085	0.138	0.147
RMSE($\hat{\sigma}$)	0.091	0.037	0.003	0.097	0.052	0.005
Elapsed time (s)	171.56	61.04	392.78	257.33	156.21	469.92

$J = 160$ and $N = 1000$. And we consider a straightforward confirmatory design with 40 items to measure each latent trait. In other words, items 1-40 measure the first latent trait, items 41–80 measure the second latent trait, and so on. The distributions to generate the true non-zero discrimination parameters a_{jk} and the difficulty parameters d_j are the same as those in one-dimensional simulation. The true individual parameters $\boldsymbol{\theta}_i$ are from a multivariate normal distribution where the mean vector is $\mathbf{0}$ and the true correlation among the latent trait dimensions are all set to 0.6.

The three algorithms are compared using 100 replications of each setting. Table 2 presents the RMSE results of the three algorithms for all parameters in two different dimensions. For example, the RMSE for $\boldsymbol{a}$ is computed as

$$\text{RMSE}\,(\hat{\boldsymbol{a}}) = \sqrt{\frac{1}{100}\frac{1}{M}\sum_{n=1}^{100}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\hat{a}_{jk}^{(n)} - a_{jk}^{*}\right)^2}, \tag{22}$$

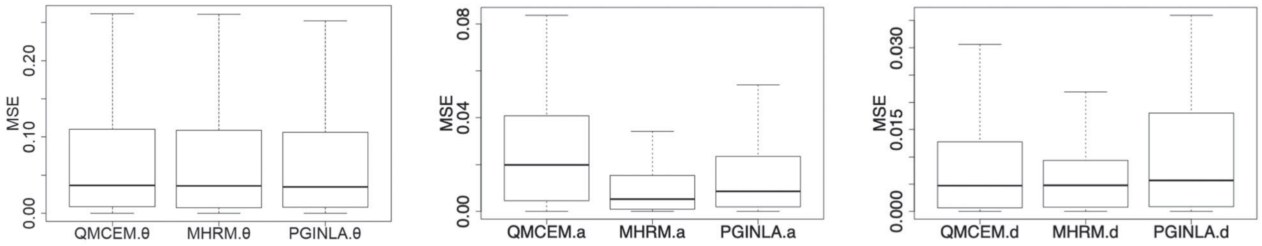
where M is the number of non-zero $\boldsymbol{a}_j, j = 1, 2, \ldots, J$, $\hat{a}_{jk}^{(n)}$ is the estimate of nth replication and a_{jk}^{*} is the true value. For the two different dimensions ($K = 2$ and $K = 4$), the RMSEs of the three algorithms on the four parameters are close. Particularly, the PG-INLA algorithm obtains more accurate results of individual parameters $\boldsymbol{\theta}_i$ and Σ than those of the QMCEM algorithm and the MH-RM algorithm.

The boxplots of the MSEs for individual parameters $\boldsymbol{\theta}_i$, item parameters $\boldsymbol{a}_j$ and d_j for $K = 2$ and $K = 4$ are presented in Figures 3 and 4, respectively. The corresponding MSE is calculated in a similar way as Equation (21). Figures 3(a) and 4(a), Figures 3(a) and 4(a) as well as Figures 3(c) and 4(c) correspond to the three parameter, $\boldsymbol{\theta}_i$, $\boldsymbol{a}_j$ and d_j, respectively. In each panel, the meaning of label is similar to that of label in a one-dimensional simulation study. The boxplots of $\boldsymbol{a}_j$ contains only non-zero elements. The PG-INLA algorithm estimation for the item parameter $\boldsymbol{a}_j$ and d_j is always better than one of the QMCEM algorithm and the MH-RM algorithm. The robustness of PG-INLA algorithm on individual parameter $\boldsymbol{\theta}_i$ estimation is implied by the fact that the MSEs of PG-INLA algorithm on individual parameters $\boldsymbol{\theta}_i$ does not increase as the dimensionality of latent traits rises.

In general, the PG-INLA algorithm does produce superior or comparable accuracy compared with other two algorithms. However, from the time taken by the three algorithms to reach the converged state, as shown in Table 2, we see the PG-INLA algorithm costs the most. On the other hand, Table 2 also indicates that when the MIRT model becomes more complex, the PG-INLA algorithm shows the feature of scalability, as the time increase is relatively the least among the three algorithms.

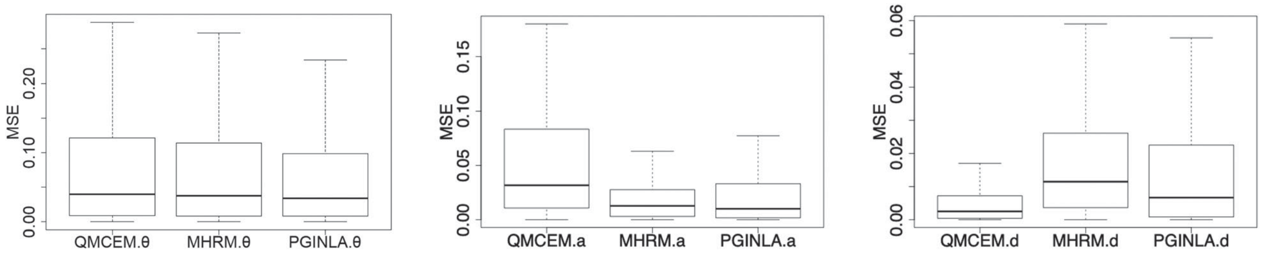
5. Application

We further illustrate the performance of the proposed PG-INLA algorithm through an application to a personality assessment dataset based on an International Personality Item Pool-Neuroticism, Extraversion & Openness (IPIP-NEO) personality inventory (Johnson, 2014). This is a free public domain version of the popular and widely used NEO personality inventory (Costa & McCrae, 2008), which is applied to assess the Big Five personality latent traits



(a) $\boldsymbol{\theta}_i$ (b) $\boldsymbol{a}_j$ (c) d_j

Figure 3. Boxplots of MSEs of the estimated parameters for the QMCEM algorithm, MH-RM algorithm and PG-INLA algorithm when the latent dimension is 2: (left) MSEs of individual parameters $\boldsymbol{\theta}_i$; (middle) MSEs of non-zero discrimination parameter $\boldsymbol{a}_j$; (right) MSEs of difficulty parameters d_j. (a) $\boldsymbol{\theta}_i$. (b) $\boldsymbol{a}_j$ and (c) d_j.



(a) $\boldsymbol{\theta}_i$ (b) $\boldsymbol{a}_j$ (c) d_j

Figure 4. Boxplots of MSEs of the estimated parameters for the QMCEM algorithm, MH-RM algorithm and PG-INLA algorithm when the latent dimension is 4: (left) MSEs of individual parameters $\boldsymbol{\theta}_i$; (middle) MSEs of non-zero discrimination parameter $\boldsymbol{a}_j$; (right) MSEs of difficulty parameters d_j. (a) $\boldsymbol{\theta}_i$. (b) $\boldsymbol{a}_j$ and (c) d_j.

($K = 5$): neuroticism (N), extraversion (E), openness to experience (O), agreeableness (A), and conscientiousness (C) (see Chen et al., 2017; Zhang et al., 2020).

The dataset includes 20993 people and 300 items and we use a subset of it, which contains information from 3000 participants who have completed all of the 300 items. Chen et al. (2017) conducted a structured latent factor analysis using this dataset of 7325 individuals who completed all the items. According to the identifiability theory provided by Chen et al. (2017), the measurement design matrix has a simple structure, which is a safe design for model identification. Each item in this design only evaluates one latent trait, and each trait is evaluated by 60 items. All of the items are on a five-category rating scale, and in order to fit a M2PL IRT model, we binarize them by merging categories $\{1, 2, 3\}$ and $\{4, 5\}$ which was also suggested by Chen et al. (2017).

To further demonstrate the accuracy and efficacy of the PG-INLA algorithm, we also reproduce the results of Chen et al. (2017) as a comparison. The results are summarized by Figures 5 and 6. Figure 5 shows the boxplots of the 300 estimated unconstrained loadings, each of the five factors has 60 unconstrained loadings. The left panel is the results from the PG-INLA algorithm and the right panel is the results from the structured latent factor analysis. We find that the estimated unconstrained loadings of both methods are very close, ignoring the difference in sample sizes as the sample sizes in both methods are very large. In particular, the medians of the boxplots corresponding to each latent factor are almost identical and the majority of the estimated unconstrained loadings fall between 0.5 and 2.5. Figure 6 presents the factor correlation matrix calculated with the estimated factor scores, with the left panel from the proposed PG-INLA algorithm and the right panel from the structured latent factor analysis, which shows no much difference. Both of them are largely consistent with the research on the Big Five personality traits by Digman (1997). The convergence time of the PG-INLA algorithm and the structured latent factor analysis are 1124.86 s and 802.73 s, respectively. From the results above, we see that the results obtained by the PG-INLA algorithm are close to those from Chen et al. (2017).

6. Conclusion

In this paper, we propose a PG-INLA algorithm for estimating 2-PL IRT model parameters. Thanks to the beneficial properties of the Pólya-Gamma distributions, the Gibbs part of the proposed algorithm could have a faster sampling. And due to the quick approximation nature of the INLA algorithm itself, the PG-INLA algorithm can converge with far fewer iterations and have higher efficiency than some fully Bayesian algorithms. Furthermore,

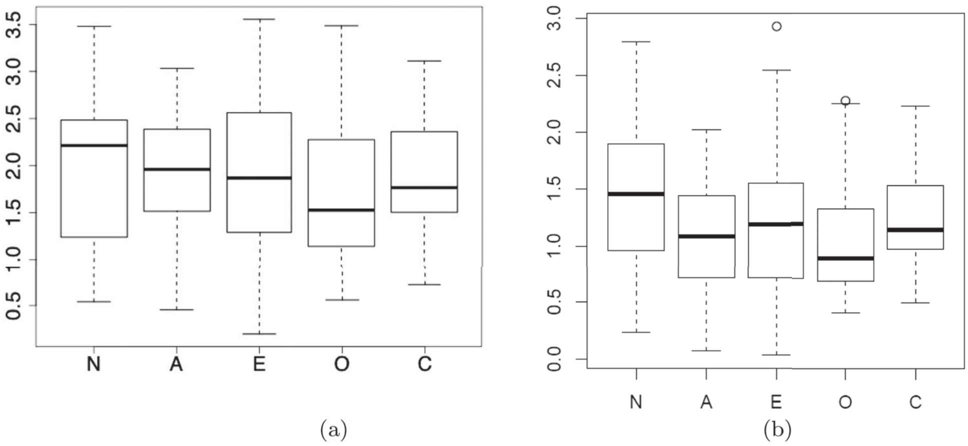


Figure 5. The boxplots of the 60 estimated unconstrained loadings for each of the five factors for the IPIP-NEO dataset: (left) the estimated unconstrained loadings from PG-INLA algorithm; (right) the estimated unconstrained loadings from Chen et al. (2017). (a) PG-INLA and (b) Chen et al.'s algorithm.

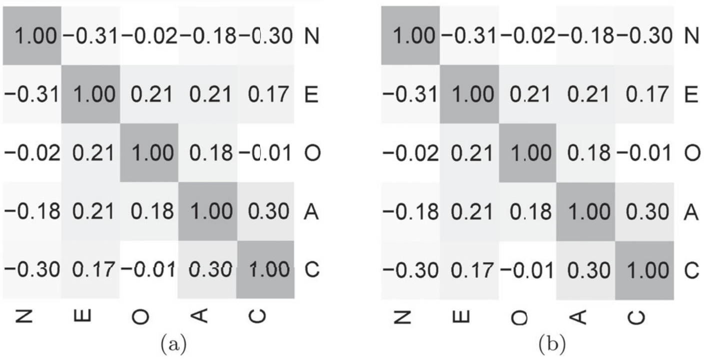


Figure 6. The factor correlation matrix between the estimated factor scores for the IPIP-NEO dataset: (left) the factor correlation matrix from PG-INLA algorithm; (right) the factor correlation matrix from Chen et al. (2017). (a) PG-INLA and (b) Chen et al.'s aogorithm.

the PG-INLA algorithm developed is not only virtually tuning-free and computationally efficient but also produces estimation that closely resembles the MMLE, which is obtained by the EM algorithm, the QMCEM algorithm and the MH-RM algorithm. Our simulation studies show that the PG-INLA algorithm is comparable to these popular MMLE algorithms. In particular, the PG-INLA algorithm outperforms the QMCEM algorithm and the MH-RM algorithm when the dimensionality of the latent traits space is high. Based on these evidences, the proposed PG-INLA algorithm has the potential to be used in research for the 2-PL IRT models.

The subsequent research on the PG-INLA algorithm will be expanded in the following directions in the future. First of all, the 2-PL IRT models used in this paper are for binary response data, where ordinal response data are usually required because many measurement designs have more than two levels of response. The ordinal response data could be fitted by the graded response models (GRM) (Samejima, 1968), the generalized partial credit models (GPCM) (Muraki, 1992) and others. It is possible that the PG-INLA algorithm can be tailored to handle ordinal IRT models, like the GRMs or the GPCMs, as Polson et al. (2013) demonstrated that the Pólya-Gamma strategy can be extended to a multinomial regression model. Secondly, the three-parameter logistic (3-PL) and four-parameter logistic (4-PL) IRT models are also popular. For example, the 3-PL model has one more guessing parameter than the 2-PL model. However, the PG-MCMC algorithm cannot be fully used in 3-PL IRT model estimation as its posterior distributions are not in closed form Jiang and Templin (2018). Nevertheless, if we can use the INLA algorithm to estimate the guessing parameter simultaneously, then the PG-INLA algorithm can be easily generalized to 3-PL IRT models because of the benefits of keeping the posterior distribution in closed form. Finally, we might be able to explore the PG-INLA algorithm application on more complex models, such as the ones with multiple-choice items.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The research is partially supported by the National Natural Science Foundation of China [grant number 12271168] and the 111 Project of China [grant number B14019].

ORCID

Yincai Tang http://orcid.org/0000-0001-6756-6461

References

Albert, J., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*(422), 669–679. https://doi.org/10.1080/01621459.1993.10476321

Béguin, A., & Glas, C. A. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, *66*(4), 541–561. https://doi.org/10.1007/BF02296195

Biane, P., Pitman, J., & Yor, M. (1999). Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. *Bulletin of the American Mathematical Society*, *38*(4), 435–465. https://doi.org/10.1090/bull/2001-38-04

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

Cai, L. (2010a). High-dimensional exploratory item factor analysis by a Metropolis–Hastings Robbins–Monro algorithm. *Psychometrika*, *75*(1), 33–57. https://doi.org/10.1007/s11336-009-9136-x

Cai, L. (2010b). Metropolis–Hastings Robbins–Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, *35*(3), 307–335. https://doi.org/10.3102/1076998609353115

Celeux, G., & Diebolt, J. (1985). The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly*, *2*, 73–82.

Chen, Y., Li, X., & Zhang, S. (2017). Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *Journal of the American Statistical Association*, *115*(532), 1756–1770. https://doi.org/10.1080/01621459.2019.1635485

Chen, Y., & Zhang, S. (2020). Estimation methods for item factor analysis: An overview. In Y. Zhao & D. G. Chen (Eds.), *Modern statistical methods for health research* (pp. 329–350). Springer.

Costa, P. T., & McCrae, R. R. (2008). The revised NEO personality inventory (NEO-PI-R). In G. J. Boyle, G. Matthews & D. H. Saklofske (Eds.), *The Sage handbook of personality theory and assessment*: *Personality measurement and testing* (Vol. 2, pp. 179–198). Sage Publications, Inc.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, *39*(1), 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Digman, J. M. (1997). Higher-order factors of the big five. *Journal of Personality and Social Psychology*, *73*(6), 1246–1256. https://doi.org/10.1037/0022-3514.73.6.1246

Eckes, T., & Baghaei, P. (2015). Using testlet response theory to examine local dependence in C-tests. *Applied Measurement in Education*, *28*(2), 85–98. https://doi.org/10.1080/08957347.2014.1002919

Edwards, M. C. (2010). A Markov chain Monte Carlo approach to confirmatory item factor analysis. *Psychometrika*, *75*(3), 474–497. https://doi.org/10.1007/s11336-010-9161-9

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates Publishers.

Feinberg, R., & Rubright, J. (2016). Conducting simulation studies in psychometrics: Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice*, *35*(2), 36–49. https://doi.org/10.1111/emip.2016.35.issue-2

Harwell, M. R., & Baker, F. B. (1991). The use of prior distributions in marginalized Bayesian item parameter estimation: A didactic. *Applied Psychological Measurement*, *15*(4), 375–389. https://doi.org/10.1177/014662169101500409

Imai, K., & Dyk, D. A. V. (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, *124*(2), 311–334. https://doi.org/10.1016/j.jeconom.2004.02.002

Ip, E. H. (2002). On single versus multiple imputation for a class of stochastic algorithms estimating maximum likelihood. *Computational Statistics*, *17*(4), 517–524. https://doi.org/10.1007/s001800200124

Jaakkola, T. S., & Jordan, M. I. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing*, *10*(1), 25–37. https://doi.org/10.1023/A:1008932416310

Jiang, Z., & Templin, J. L. (2018). Gibbs samplers for logistic item response models via the Pólya gamma distribution: A computationally efficient data-augmentation strategy. *Psychometrika*, *84*(2), 358–374. https://doi.org/10.1007/s11336-018-9641-x

Johnson, J. A. (2014). Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, *51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003

Junker, B. W., Patz, R. J., & VanHoudnos, N. M. (2017). Markov Chain Monte Carlo for item response models. In Wim J. van der Linden (Ed.), *Handbook of item response theory: Volume two: Statistical tools* (pp. 271–312). Chapman and Hall/CRC.

Lawley, D. N., & Maxwell, A. E. (1962). Factor analysis as a statistical method. *Journal of the Royal Statistical Society. Series D*, *12*(3), 209–229.

Liu, Y., Magnus, B., O'Connor, H., & Thissen, D. (2018). Multidimensional item response theory. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing* (Vol. 2–2, pp. 445–493). John Wiley and Sons, Ltd.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.

Lord, F., Novick, M., & Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability*. Addison-Wesley.

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. Springer.

Mair, P., & Gruber, K. (2022). Bayesian explanatory additive IRT models. *British Journal of Mathematical and Statistical Psychology*, *75*(1), 59–87. https://doi.org/10.1111/bmsp.v75.1

Matlock, K. L. , Turner, R. C., & Gitchel, W. D. (2018). A study of reverse-worded matched item pairs using the generalized partial credit and nominal response models. *Educational and Psychological Measurement*, *78*(1), 103–127. https://doi.org/10.1177/0013164416670211

McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press.

Moraga, P. (2020). *Geospatial health data: Modeling and visualization with R-INLA and shiny* (1st ed.). CRC Press.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*(2), 159–176. https://doi.org/10.1177/014662169201600206

Murphy, J. P. (2021). Explanatory item response models for dyadic data from multiple groups. *Sociological Methodology*, *51*(1), 112–145. https://doi.org/10.1177/0081175020967392

Niederreiter, H. (1978). Quasi-Monte Carlo methods and pseudo-random numbers. *Bulletin of the American Mathematical Society*, *84*(6), 957–1041. https://doi.org/10.1090/bull/1978-84-06

Polson, N. G. , Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *Journal of the American Statistical Association*, *108*(504), 1339–1349. https://doi.org/10.1080/01621459.2013.829001

Reckase, M. D. (2009). Multidimensional item response theory. *Handbook of Statistics*, *26*, 607–642. https://doi.org/10.1016/S0169-7161(06)26018-8

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Samejima, F. (1968). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, *34*(1), 1–97.

Sara, M., & Andrea, R. (2020). *Integrated nested Laplace approximations (INLA)*. https://dx.doi.org/10.1002/9781118445112.stat08212

Song, X. Y., & Lee, S. Y. (2005). A multivariate probit latent variable model for analyzing dichotomous responses. *Statistica Sinica*, *15*(3), 645–664.

Thissen, D., & Wainer, H. (2001). *Test scoring*. Routledge.

van der Linden, W. J., & Hambleton, R. K. (1997). Item response theory: Brief history, common models, and extensions. In Wim J. van der Linden (Ed.), *Handbook of modern item response theory* (pp. 1–28). Springer.

Zeithammer, R., & Lenk, P. J. (2006). Bayesian estimation of multivariate-normal models when dimensions are absent. *Quantitative Marketing and Economics*, *4*(3), 241–265. https://doi.org/10.1007/s11129-005-9006-5

Zhang, S., Chen, Y., & Liu, Y. (2020). An improved stochastic EM algorithm for large-scale full-information item factor analysis. *British Journal of Mathematical and Statistical Psychology*, *73*(1), 44–71. https://doi.org/10.1111/bmsp.v73.1