Journal of Mathematical Psychology 82 (2018) 12–25

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Slice–Gibbs sampling algorithm for estimating the parameters of a multilevel item response model

Jing Lu [1], Jiwei Zhang [1], Jian Tao *

School of Mathematics and Statistics, KLAS, Northeast Normal University, Changchun, China

HIGHLIGHTS

- We present a fully Bayesian approach using a slice–Gibbs sampler to estimate all of the parameters in the multilevel IRT framework.
- Two methods of model assessment are proposed to compare the goodness of fit between the models.
- Empirical simulation results show that the proposed method has some real advantages in parameter recovery and model fit.
- The significance of our findings is illustrated with an application to a real data set.

ARTICLE INFO

Article history:
Received 15 October 2016
Received in revised form 10 September 2017

Keywords:
Markov chain Monte Carlo methods
Multilevel model
Item response theory (IRT)
Two-parameter logistic model
Slice–Gibbs sampling

ABSTRACT

In a fully Bayesian framework, a novel slice–Gibbs algorithm is developed to estimate a multilevel item response theory (IRT) model. The advantage of this algorithm is that it can recover parameters well based on various types of prior distributions of the item parameters, including informative and non-informative priors. In contrast to the traditional Metropolis–Hastings (M–H) within Gibbs algorithm, the slice–Gibbs algorithm is faster and more efficient, due to its drawing the sample with acceptance probability as one, rather than tuning the proposal distributions to achieve the reasonable acceptance probabilities, especially for the logistic model without conjugate distribution. In addition, based on the Markov chain Monte Carlo (MCMC) output, two model assessment methods are investigated concerning the goodness of fit between models. The information criterion method on the basis of marginal likelihood is proposed to assess the different structural multilevel models, and the cross-validation method is used to evaluate the overall multilevel IRT models. The feasibility and effectiveness of the slice–Gibbs algorithm are investigated in simulation studies. An application using a real data involving students' mathematics test achievements is reported.

1. Introduction

Multilevel analysis has become increasingly popular in educational and psychological assessments (e.g. Goldstein, 2011; Goldstein & McDonald, 1987; Longford & Muthén, 1992; Muthén, 1989; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). The rapid growth of multilevel item response theory (IRT) modeling has resulted in two realizations. First, the interaction between a person and item is constructed by the IRT model, rather than the traditional aggregate test level performance, assuming all questions contributed equally to the understanding of students' abilities. Compared with the traditional method, the IRT model provides a more nuanced view of the information for each question provided to a student. Second, in the field of social sciences, a critical issue that is ignored by IRT models concerns unobserved heterogeneity of item responses due to the hierarchical structure of data (e.g., data from students nested within schools/classes or measurements nested within individuals). The multilevel IRT modeling accounts for both population heterogeneity and makes valid inferences from item response data that has a nested or hierarchical structure. The multilevel IRT models for binary response data have also been increasingly used to handle a hierarchical structure (e.g. Adams, Wilson, & Wu, 1997; Fox, 2010; Fox & Glas, 2001; Kamata, 2001; Maier, 2001; Mislevy & Bock, 1989; Natesan, 2007).

A variety of estimation procedures have been proposed to estimate the parameters of multilevel IRT models. The multilevel logistic regression models belong to one class of generalized linear

* Corresponding author.
E-mail addresses: luj282@nenu.edu.cn (J. Lu), zhangjw713@nenu.edu.cn (J. Zhang), taoj@nenu.edu.cn (J. Tao).

[1] The MATLAB code for the slice–Gibbs sampling algorithms can be found in the online supplements. The data in Section 7 are not made available due to a confidentiality agreement with NENU/BNU. Jing Lu and Jiwei Zhang are co-first authors. They contributed equally to this work.

https://doi.org/10.1016/j.jmp.2017.10.005
0022-2496/

mixed models (GLMMs) (e.g. Clayton, 1996; Hedeker & Gibbons, 1994; McCulloch & Searle, 2001). Marginal maximum likelihood (MML) (Bock & Aitkin, 1981) is one of the most important methods for estimating GLMMs. However, the complex dependency structures result in the nesting of integrals in multilevel IRT models, and most estimation equations do not have a closed form. Thus, this approach requires a numerical or Monte Carlo integration or an integral approximation method that includes Laplace approximation (Raudenbush, Yang, & Yosef, 2000), adaptive quadrature (Bock & Schilling, 1997; Rabe-Hesketh, Skrondal, & Pickles, 2002, 2005), and Monte Carlo integration (Kuk, 1999; Skaug, 2002). Therefore, MML leads to a complicated estimation procedure. Moreover, it is difficult to incorporate uncertainty (standard errors) into parameter estimates (Patz & Junker, 1999). A fully Bayesian procedure remains straightforward for complex dependency structures in multilevel IRT models and allows for complete uncertainty calculations (Tsutakawa & Soltys, 1988). Studies focusing on multilevel models from a Bayesian perspective usually use a probit link function in their formulation (Albert, 1992; Béguin & Glas, 2001; Fox, 2005; Fox & Glas, 2001; Hoijtink, 2000; Tanner & Wong, 1987). Also popular in Bayesian multilevel approach studies is the cumulative logit function (Browne & Draper, 2006; Gilks & Wild, 1992; Zeger & Karim, 1991).

In this paper, an efficient slice–Gibbs sampling algorithm in a fully Bayesian framework is proposed to estimate multilevel two-parameter logistic models. The slice–Gibbs algorithm, as the name suggests, can be conceived of an extension of Gibbs algorithm. The sampling process consists of two parts. One part is the slice algorithm (Bishop, 2006; Damien, Wakefield, & Walker, 1999; Neal, 2003), which samples the logistic IRT model from the truncated full conditional posterior distribution by introducing the auxiliary variables. The other part is Gibbs algorithm which updates the linear multilevel structure parameters based on the sampled values from IRT model, where the linear multilevel models are assumed to follow normal distributions (Béguin & Glas, 2001; Fox, 2005, 2010; Fox & Glas, 2001). Additionally, the Gibbs algorithm part implements sampling by using conjugate prior and greatly increases efficiency (Albert, 1992; Tanner & Wong, 1987). The motivation for the approach is manifold. First, using the slice sampling method to estimate the two-parameter logistic model (Lord, 1980; Tsutakawa, 1984; van der Linden & Hambleton, 1997) has the advantage of a flexible prior distribution being introduced to obtain samples from the full conditional posterior distributions rather than being restricted to using the conjugate distributions, which is required in Gibbs sampling (Gelfand & Smith, 1990; Geman & Geman, 1984) and limited using the normal ogive framework (Albert, 1992; Béguin & Glas, 2001; Fox, 2005, 2010; Fox & Glas, 2001; Tanner & Wong, 1987). It allows informative priors and non-informative priors of the item parameters. Second, the Metropolis–Hasting algorithm (Chen, Shao, & Ibrahim, 2000; Chib & Greenberg, 1995; Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953; Tierney, 1994) depends on the proposal distributions and variances (tuning parameters) and is sensitive to step size. If the step size is too small, the chain will take longer to traverse the target density. If the step size is too large, there will be inefficiencies due to a high rejection rate. However, the slice sampler automatically tunes the step size to match the local shape of the target density and draws the samples with acceptance probability equal to one. Thus, it is easier and more efficient to implement.

The rest of this paper is organized as follows. In Section 2, the theoretical foundation of the slice–Gibbs sampling algorithm is presented. A general multilevel two-parameter logistic model and model identifiability are presented to explain the binary multilevel data in Section 3. A detailed slice–Gibbs algorithm is given in Section 4. In Section 5, two methods of model assessment are proposed to compare the goodness of fit between the models by using two different parameter structures. In Section 6, four simulated examples that focus on the performance of parameter recovery, the flexibility and sensitivity of prior distributions for the slice sampler, the results of comparing with the M–H within Gibbs algorithm, and information criteria and cross-validation log-likelihood to assess model fit are given. In Section 7, the results of model assessment and performance of the slice–Gibbs sampler in practical situations are shown by an empirical example. Finally, some concluding remarks are presented in Section 8.

2. Theoretical foundation of the slice–Gibbs algorithm

The motivation for the slice sampling algorithm is that we can use the auxiliary variable approach to sample from posterior distributions arising from Bayesian non-conjugate models. The theoretical basis for this algorithm is described below:

Suppose that we wish to simulate values from a density function $q(x)$ (the target density) given by $q(x) \propto \psi(x)\prod_{i=1}^{N} l_i(x)$ which is not possible to sample directly, where $\psi(x)$ is a known density from which samples can be easily drawn and $l_i(x)$ are non-negative invertible functions that do not have to be density functions. We introduce the auxiliary variables $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)$, and each element of the vector from $(0, +\infty)$. The inequalities $\delta_i < l_i(x)$ are established, and the joint density can be written as

$$q(x, \delta_1, \ldots, \delta_N) \propto \psi(x)\prod_{i=1}^{N} I\left\{\delta_i < l_i(x)\right\}. \tag{1}$$

It is easy to show that if the auxiliary variables are integrated out, we obtain the marginal distribution $q(x)$,

$$\begin{aligned} q(x) &= \int_0^{l_1(x)} \cdots \int_0^{l_N(x)} q(x, \delta_1, \ldots, \delta_N) d\delta_N \cdots d\delta_1 \\ &\propto \psi(x)\int_0^{l_1(x)} \cdots \int_0^{l_N(x)} 1 d\delta_N \cdots d\delta_1 = \psi(x)\prod_{i=1}^{N} l_i(x). \end{aligned} \tag{2}$$

Then we obtain the set $\Delta_{\delta_i} = \{x \,|\delta_i < l_i(x)\}$ based on the invertible property of the function $l_i(x)$, which is called the "slice" defined by δ_i (Neal, 2003). We simulate values from the slice sampler by repeatedly sampling from the full conditional distributions, proceeding as follows at iteration t:

- Sample $\delta_i^{(t)} \sim U(0, l_i(x^{(t-1)}))$, $i = 1, \ldots, N$.
- Sample $x^{(t)} \sim \Delta_{\delta_i} = \{x \,\big|\, \delta_i^{(t)} < l_i(x)\}$, thereby deriving a horizontal "slice" under the density function.

A Markov chain based on the slice sampler can be constructed by sampling points alternately from uniform distribution under the density curve and only concerning the horizontal "slice" defined by the current sample points.

The slice sampler shows the better performance, even if we have misspecified priors and proposal distributions or variances in the Metropolis–Hasting (M–H) algorithm. Moreover, more flexible specification of priors can be adopted without using conjugate priors, which is the limitation of Gibbs sampling, and it is not sensitive to the specified priors that will be conducted in the simulation. Furthermore, it avoids retrospective tuning in the M–H algorithm if we do not know how to choose a proper tuning parameter or if no value for the tuning parameter is appropriate. Thus, this approach is more efficient than simple M–H sampling. Considering the target density $q(x) \propto \psi(x)l(x)$, we select $\psi(x)$ as a special proposal density to generate candidate x^* and we have acceptance probability $\alpha(x^{(t)}, x^*) = \min\{1, \frac{l(x^*)}{l(x^{(t)})}\}$. Next, we sample uniform random variable δ if $\delta < \frac{l(x^*)}{l(x^{(t)})}$, the chain will move on.