Brazilian Journal of Probability and Statistics
2013, Vol. 27, No. 2, 245–262
DOI: 10.1214/11-BJPS172

CADEM: A conditional augmented data EM algorithm for fitting one parameter probit models

C. L. N. Azevedo[a] **and D. F. Andrade**[b]

[a]*University of Campinas*
[b]*Federal University of Santa Catarina*

Abstract. In this article we develop an estimation method based on the augmented data scheme and EM/SEM (Stochastic EM) algorithms for fitting one-parameter probit (Rasch) IRT (Item Response Theory) models. Instead of using the S steps of the SEM algorithm, that is, instead of simulating values for the unobserved variables (augmented data and the latent traits), we consider the conditional expectations of a set of unobserved variables on the other set of unobserved variables, the current estimates of the parameters and the observed data, based on the full conditional distributions from the Gibbs sampling algorithm. Our method, named the CADEM algorithm (conditional augmented data EM), presents straightforward E steps, which avoid the need to evaluate the usual integrals, also facilitating the M steps, without the need to use numerical methods of optimization. We use the CADEM algorithm to obtain both maximum likelihood estimates and maximum a posteriori estimates of the difficulty parameters for the one-parameter probit (Rasch) model. Also, we obtain estimates for the latent traits, based on conditional expectations. In addition, we show how to calculate the associated standard errors. Some directions are provided to extend our approach to other IRT models. In this respect, we perform a simulation study to compare the estimation methods. The results indicated that our approach is quite comparable to the usual marginal maximum likelihood (MML) and Gibbs sampling methods (GS) in terms of parameter recovery. However, CADEM is as fast as MML and as flexible as GS.

1 Introduction

IRT consists of a set of measurement models which have been increasingly applied in many fields. Two important aspects concerning item response models (IRM) are the large number of quantities to be estimated and their complex mathematical structures. Due to these aspects, the use of an appropriate estimation method plays a crucial role. Many works have been devoted to proposing and comparing estimation methods. The marginal likelihood based methods, that is, marginal maximum likelihood (MML) and marginal maximum a posteriori (MMAP), are probably the

Key words and phrases. Item response models, maximum likelihood, Bayesian estimates, augmented data, EM algorithm.
Received May 2011; accepted August 2011.

most popular methods for item parameter estimation. They produce consistent estimates in many situations [see Bock and Aitkin (1981) and Mislevy (1986)]. However, they are not appropriate or even able to handle more complex IRT models, such as longitudinal models [see Andrade and Tavares (2005)], multilevel models [see Fox and Glas (2001)] or asymmetric models [see Azevedo et al. (2011) and Bazan et al. (2006)]. On the other hand, the MCMC based methods, which include the full Gibbs sampling [see Albert (1992)] and Metropolis–Hasting within Gibbs sampling [see Patz and Junker (1999a)], are able to handle complex methods even though they are computationally intensive. Recently, Fox (2003) proposed the use of the SEM algorithm for fitting a multilevel IRT model. Even though this method requires less time than the traditional MCMC methods, it still demands a considerable amount of time.

The main goal of this work is to propose an EM/SEM based algorithm to replace the S steps in the SEM algorithm by the calculation of conditional expectations of the augmented data, given the latent traits and the latent traits given the augmented data. Then, the M steps are carried out straightforwardly. Our method is not only very fast (as fast as MML and MMAP methods), but also can handle complex models. Due to the conditional structure based on augmented data, our algorithm is named CADEM (conditional augmented data EM algorithm). In addition, we show how the standard errors can be easily calculated by using the Louis identity [see Louis (1982)], and how it is possible to fit other IRT models. We also show how to obtain Bayesian estimates (maximum of the posterior distribution) through the CADEM algorithm. The results of the simulation study indicate that the CADEM recovers all parameters as well as the usual methods. However, it has the advantage of being as fast as MML and MMAP and as flexible as MCMC methods.

The article is organized as follows. After the introduction, we present the CADEM algorithm. In the subsequent section we present the Bayesian extension of the CADEM, while in the following one we discuss how CADEM can be used to fit more complex IRT models. Then, a simulation study is presented, to compare all estimation methods considered. Finally, we make some comments and conclusions in the last section and show some additional calculations in the Appendix.

2 CADEM algorithm

The two most typical approaches for parameter estimation in IRT are the marginal maximum likelihood—marginal Bayesian methods and Bayesian estimation using MCMC algorithms [see Bock and Aitkin (1981), Mislevy (1986), Albert (1992), Patz and Junker (1999a) and Patz and Junker (1999b)]. While the former is limited due to the use of numerical methods of integration and maximization, the latter demands a large amount of computational processing time. Marginal methods are much faster than MCMC algorithms, even though the latter are much more flexible than the former. Our approach combines the advantages of these two classes of methods.

The EM algorithm is a method that allows obtaining maximum likelihood estimates (or the maximum of posterior distributions) in the presence of missing data; see Dempster et al. (1977). The idea underlying the EM algorithm is to maximize, concerning the parameters, the expectation of the log-likelihood concerning the missing data given the observed data and provisional estimates of the parameters. The goal of our work is to develop a kind of EM algorithm by considering the latent traits and an augmented data set as the unobserved data.

In this paper, we consider the situation where a set of n examinees (students, patients, schools) is submitted to a measurement instrument (cognitive test, clinical evaluation, questionnaire) composed of I items. Let Y_{ij}, the answer of examinee j to item i, be a Bernoulli random variable, with 1 indicating a correct answer and 0 otherwise. To model such probability, we consider the one-parameter probit model (1PP), that is,

$$P_{ij} = P(Y_{ij} = 1|\theta_j, b_i) = \Phi(\theta_j - b_i), \tag{2.1}$$

where b_i is the difficulty parameter and $\Phi(\cdot)$ stands for the cumulative normal function. The reader is referred to Baker and Kim (2004) for further details and interpretations. Model (2.1) is also called the Rasch model, even though in this case the cumulative distribution function of the standard logistic distribution is used instead of the probit. Our main interest lies in making inferences about $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$ and $\mathbf{b} = (b_1, \ldots, b_I)'$.

To define the CADEM, it is convenient to consider the augmented data for the 1PP model proposed by Albert (1992), that is, $Y_{ij} = I_{(Z_{ij} \geq 0)}$, where

$$Z_{ij}|(\theta_j, b_i) \sim N(\theta_j - b_i, 1). \tag{2.2}$$

Furthermore, let $\mathbb{I}_j$ and $\mathcal{N}_i$ be the indexes which represent the set of items answered by (or presented to) examinee j and the set of examinees that answered item i, respectively. Also, it is assumed that $\theta_j \sim N(0, 1)$, mutually independent. Following the EM algorithm nomenclature, let $\mathbf{Y}_{..} = (Y_{11}, \ldots, Y_{1I}, \ldots, Y_{n1}, \ldots, Y_{nI})'$ be the incomplete data set while $(\mathbf{Z}'_{..}, \boldsymbol{\theta}', \mathbf{Y}'_{..})$ is the complete data set, $\mathbf{Z}_{..} = (Z_{11}, \ldots, Z_{1I}, \ldots, Z_{n1}, \ldots, Z_{nI})'$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$. That is, we consider that $\mathbf{W} = (\mathbf{Z}'_{..}, \boldsymbol{\theta}')'$ are the unobserved data. Under the usual assumptions of conditional independence, the augmented data likelihood is given by

$$\begin{aligned} L(\mathbf{b}, \boldsymbol{\theta}, \mathbf{z}_{..}|\mathbf{y}_{..}) &\propto p(\mathbf{z}_{..}|\boldsymbol{\theta}, \mathbf{b}, \mathbf{y}_{..})p(\boldsymbol{\theta}) \\ &= \left\{\prod_{j=1}^{n}\prod_{i\in\mathbb{I}_j} p(z_{ij}|\theta_j, b_i, y_{ij})\right\}\left\{\prod_{j=1}^{n} p(\theta_j)\right\} \\ &= \left\{\prod_{j=1}^{n}\prod_{i\in\mathbb{I}_j} \exp\{-0.5(z_{ij} - \theta_j + b_i)^2\}\mathbb{1}_{(y_{ijk}, z_{ijk})}\right\} \\ &\quad \times \left\{\prod_{j=1}^{n} \exp\{-0.5(\theta_j^2)\}\right\}, \end{aligned} \tag{2.3}$$

where $\mathbb{1}_{(y_{ijk}, z_{ijk})} = \mathbb{1}_{(y_{ij}=1)}\mathbb{1}_{(z_{ij}\geq 0)} + \mathbb{1}_{(y_{ij}=0)}\mathbb{1}_{(z_{ij}<0)}$ and $\mathbb{1}$ is the usual indicator function. Therefore, the augmented log-likelihood is given by

$$\begin{aligned} l^c &= l(\mathbf{b}, \boldsymbol{\theta}, \mathbf{z}_{..}|\mathbf{y}_{..}) \\ &= \left\{\sum_{j=1}^{n}\sum_{i\in\mathbb{I}_j}\{-0.5(z_{ij}^2 - 2z_{ij}\theta_j + \theta_j^2 + 2z_{ij}b_i - 2\theta_j b_i + b_i^2)\}\right\} \qquad (2.4) \\ &\quad - \left\{\sum_{j=1}^{n}\{0.5(\theta_j^2)\}\right\}. \end{aligned}$$

Within the framework of the EM algorithm, it is necessary to evaluate the expectation $\mathbb{E}(\mathbf{W}|\mathbf{b})$ from (2.4), that is, to calculate $\mathbb{E}[\ln p(\mathbf{Z}_{..}, \boldsymbol{\Theta}, \mathbf{b}|\mathbf{y}_{..})|\mathbf{b}]$. However, obtaining the distribution of $(\mathbf{Z}_{..}, \boldsymbol{\Theta}|\mathbf{y}_{..}, \mathbf{b})$ can be quite complicated. On the other hand, the distributions $Z_{ij}|(\boldsymbol{\theta}, \mathbf{b}, \mathbf{y}_{..})$ and $\theta_j|(\mathbf{z}_{..}, \mathbf{b}, \mathbf{y}_{..})$ are known and easy to handle. From Albert (1992), it follows that

$$\theta_j|(\mathbf{z}_{.j}, \mathbf{b}, \mathbf{y}_{.j}) \sim N(\widetilde{\theta}_j\widehat{\psi}_{\theta_j}, \widehat{\psi}_{\theta_j}), \qquad (2.5)$$

where

$$\widetilde{\theta}_j = \sum_{i\in\mathcal{I}_j}(z_{ij} + b_i), \qquad \widehat{\psi}_{\theta_j} = (1 + I_j)^{-1},$$

where I_j is the number of items presented to/answered by examinee j, and $Z_{ij}|(\theta_j, \mathbf{b}, y_{ij}) \sim N(\theta_j - b_i, 1)\mathbb{1}_{(y_{ijk}, z_{ijk})}$. Furthermore, from Albert (1992) and Liu et al. (1998), it follows that

$$\widehat{Z}_{ij} = \mathbb{E}[Z_{ij}|(\theta_j, b_i, y_{ij})] = \begin{cases} \theta_j - b_i + \dfrac{\phi(\theta_j - b_i)}{1 - \Phi(-\theta_j + b_i)}, & \text{if } Y_{ij} = 1, \\ \theta_j - b_i - \dfrac{\phi(\theta_j - b_i)}{\Phi(-\theta_j + b_i)}, & \text{if } Y_{ij} = 0. \end{cases} \qquad (2.6)$$

$$\widehat{\theta}_j = \mathbb{E}[\Theta_j|(z_{ij}, b_i, y_{ij})] = \frac{\sum_{i=1}^{I}(z_{ij} + b_i)}{I + 1}, \qquad (2.7)$$

where ϕ and Φ stand for the density and the cumulative distribution function of the standard normal distribution. The idea is to calculate the expectations (2.6) and (2.7) instead of $\mathbb{E}[\mathbf{W}|\boldsymbol{\vartheta}]$. That is, instead of calculating the expectation of all unobserved variables given the data and current estimates of the parameters (based on the joint distribution of $\mathbf{W}$), the conditional expectations of one set of unobserved variables (augmented data or latent traits) are calculated given the other set (latent traits or augmented data), the data and current estimates of the parameters. That is, we are treating a specific set of missing variables, in each iteration of the EM algorithm, as parameters of interest. This may be viewed as a slight modification in the original EM algorithm. Therefore, the conditional expectation of the log-likelihood is given by

$$
\begin{aligned}
E^c &= \mathcal{E}_{(\mathbf{Z},\mathbf{\Theta}|\mathbf{b},\mathbf{y})}(l(\mathbf{b},\boldsymbol{\theta},\mathbf{z}_{..}|\mathbf{y}_{..})) \\
&\approx \left\{\sum_{j=1}^{n}\sum_{i\in\mathcal{I}_j}\{-0.5(\widehat{Z}_{ij}^2 - 2\widehat{Z_{ij}\theta_j} + \widehat{\theta}_j^2 + 2\widehat{Z}_{ij}b_i - 2\widehat{\theta}_j b_i + b_i^2)\}\right\} \qquad (2.8)\\
&\quad - \left\{\sum_{j=1}^{n}\{0.5(\widehat{\theta}_j^2)\}\right\}.
\end{aligned}
$$

It is necessary to maximize (2.8) with respect to $\mathbf{b}$. The following result summarizes our approach.

Result 2.1. The conditional EM augmented data (CADEM) algorithm for the one-parameter probit model can be expressed as follows:

E-steps: Proceed as follows:

E-step 1: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\boldsymbol{\theta}^{(t)}$, evaluate (2.6).

E-step 2: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\mathbf{Z}^{(t+1)}$, obtained in E-step 1, evaluate (2.7).

M-step: Given the current expectations $\widehat{\mathbf{W}}^{(k)(t+1)}$, $k = 1, 2$, update $\mathbf{b}$ by maximizing (2.8), that is, evaluate

$$
\widehat{b}_i^{(t+1)} = \frac{\sum_{j=1}^{n}\widehat{\theta}_j^{(t+1)} - \sum_{j=1}^{n}\widehat{Z}_{ij}^{(t+1)}}{n_i + 1},
$$

where n_i is the number of examinees that answered item i.

It is straightforward to verify the E-steps, by using the definitions of expectations of the normal and truncated-normal distributions. The M-step is obtained by deriving (2.8) with respect to the vector $\mathbf{b}$ and equating it to $\mathbf{0}$. Therefore, we can see that this procedure avoids not only the need to use numerical integration methods to evaluate the conditional expectations, it also avoids the need for numerical methods in the M-step, as in the Bock and Aitkin approach; see Bock and Aitkin (1981). So the CADEM algorithm is an alternative to the pseudo EM algorithm proposed by Bock and Aitkin (1981).

The standard errors for difficulty parameters can be obtained by using The Louis identity [see Louis (1982)], considering that the latent traits are known and the augmented data are the unobserved data, for the sake of simplicity. That is,

$$
-\frac{\partial^2 l^c}{\partial b_i^2} = \mathbb{E}_{F(Z,\Theta)}\left(\frac{-\partial^2 l^c}{\partial b_i^2}\right) - \mathcal{V}ar_{F(Z,\Theta)}\left(-\frac{\partial l^c}{\partial b_i}\right), \qquad (2.9)
$$

where $F_{(Z,\Theta)}$ is a convenient distribution, which is a function of (Z, Θ) (see the Appendix for more details). The standard errors associated with the latent trait estimates can be calculated by using the distribution (2.5), the current expectations of $\mathbf{Z}_{..}$ and the following relationship:

$$
\begin{aligned}
\mathcal{V}ar(\Theta_j|\mathbf{b},\mathbf{y}_{..}) &= \mathcal{E}_Z(\mathcal{V}ar_{\Theta|Z}(\Theta_j|\mathbf{z}_{..},\mathbf{b},\mathbf{y}_{..})) \\
&\quad + \mathcal{V}ar_Z(\mathcal{E}_{\Theta|Z}(\Theta_j|\mathbf{z}_{..},\mathbf{b},\mathbf{y}_{..})).
\end{aligned} \tag{2.10}
$$

In summary, we are replacing the conditional expectations $\mathcal{E}(\Theta_j|\mathbf{y},\widehat{\mathbf{b}})$, $\mathcal{E}(Z_{ij}|\mathbf{y},\widehat{\mathbf{b}})$ and $\mathcal{E}(\Theta_j Z_{ij}|\mathbf{y},\widehat{\mathbf{b}})$, which must be calculated in the original EM algorithm, by $\mathcal{E}(\Theta_j|\mathbf{y},\widehat{\mathbf{b}},\mathbf{z}_{.j})$, $\mathcal{E}(Z_{ij}|\mathbf{y},\widehat{\mathbf{b}},\theta_j)$ and $\mathcal{E}(Z_{ij}\mathcal{E}(\Theta_j|\mathbf{y},\widehat{\mathbf{b}},Z_{ij}))|(\mathbf{y},\widehat{\mathbf{b}}))$. See the Appendix for more details.

2.1 Bayesian estimation

Like the EM algorithm [see Dempster et al. (1977)], the CADEM algorithm can be also used to obtain the maximum of the posterior distributions of interest. To accomplish that, we assume that the prior distribution of the difficulty parameters is given by

$$
p(\mathbf{b}|\boldsymbol{\eta}_b) = \prod_{i=1}^{I} p(b_i|\boldsymbol{\eta}_b) \propto \prod_{i=1}^{I} \exp\left\{-\frac{(b_i-\mu_b)^2}{2\psi_b}\right\}, \tag{2.11}
$$

where $\boldsymbol{\eta}_b = (\mu_b,\psi_b)'$. Therefore, from (2.3), (2.4) and (2.11), the posterior and log-posterior distribution of $(\mathbf{Z},\boldsymbol{\theta},\mathbf{b})$ are given, respectively, by

$$
\begin{aligned}
p(\mathbf{b},\boldsymbol{\theta},\mathbf{z}_{..}|\mathbf{y}_{..}) &\propto p(\mathbf{z}_{..}|\boldsymbol{\theta},\mathbf{b},\mathbf{y}_{..})p(\boldsymbol{\theta})p(\mathbf{b}) \\
&= \left\{\prod_{j=1}^{n}\prod_{i\in\mathcal{I}_j} p(z_{ij}|\theta_j,b_i,y_{ij})\right\}\left\{\prod_{j=1}^{n} p(\theta_j)\right\}\left\{\prod_{i=1}^{I} p(b_i)\right\} \\
&= \left\{\prod_{j=1}^{n}\prod_{i\in\mathcal{I}_j} \exp\{-0.5(z_{ij}-\theta_j+b_i)^2\}\mathbb{1}_{(y_{ijk},z_{ijk})}\right\} \\
&\quad \times \left\{\prod_{j=1}^{n}\exp\{-0.5(\theta_j^2)\}\right\}\left\{\prod_{i=1}^{I}\exp\left\{-\frac{(b_i-\mu_b)^2}{2\psi_b}\right\}\right\}
\end{aligned} \tag{2.12}
$$

and

$$
\begin{aligned}
p^c &= \ln[p(\mathbf{b},\boldsymbol{\theta},\mathbf{z}_{..}|\mathbf{y}_{..})] \\
&= \left\{\sum_{j=1}^{n}\sum_{i\in\mathcal{I}_j}\{-0.5(z_{ij}^2-2z_{ij}\theta_j+\theta_j^2+2z_{ij}b_i-2\theta_j b_i+b_i^2)\}\right\} \\
&\quad - \left\{\sum_{j=1}^{n}\{0.5(\theta_j^2)\}\right\} - \left\{\sum_{i=1}^{I}\left\{\frac{(b_i-\mu_b)^2}{2\psi_b}\right\}\right\} + \text{const.}
\end{aligned} \tag{2.13}
$$

On the other hand, as in (2.8), the conditional expectation of the log-posterior can be suitably approximated by

$$
\begin{aligned}
Ep^c &= \mathcal{E}_{(\mathbf{Z},\boldsymbol{\Theta}|\mathbf{b},\mathbf{y})}(p(\mathbf{b},\boldsymbol{\theta},\mathbf{z}_{..}|\mathbf{y}_{..})) \\
&\approx \left\{\sum_{j=1}^{n}\sum_{i\in\mathcal{I}_j}\{-0.5(\widehat{Z}_{ij}^2 - 2\widehat{Z_{ij}\theta_j} + \widehat{\theta}_j^2 + 2\widehat{Z}_{ij}b_i - 2\widehat{\theta}_j b_i + b_i^2)\}\right\} \qquad (2.14) \\
&\quad - \left\{\sum_{j=1}^{n}\{0.5(\widehat{\theta}_j^2)\}\right\} - \left\{\sum_{i=1}^{I}\left\{\frac{(b_i-\mu_b)^2}{2\psi_b}\right\}\right\}.
\end{aligned}
$$

Thus, it is necessary to maximize (2.14) with respect to $\mathbf{b}$. This leads to Result 2.1 with the M-Step replaced by the following M-Step.

M-step: Given current expectations $\widehat{\mathbf{W}}^{(k)(t+1)}$, $k = 1, 2$, update $\mathbf{b}$ by maximizing (2.14), that is, evaluate

$$
\widehat{b}_i^{(t+1)} = \frac{\sum_{j=1}^{n}\widehat{\theta}_j^{(t+1)} - \sum_{j=1}^{n}\widehat{Z}_{ij}^{(t+1)} - \mu_b/\psi_b}{n_i + 1/\psi_b}.
$$

The standard errors for the latent traits can be calculated by using formula (2.10). For the difficulty parameters, it is necessary to use the log-posterior (2.13) instead of the log-likelihood (2.4). For details, see the Appendix.

3 CADEM extensions

It is straightforward to use other latent trait distributions within the CADEM structure, as long as they permit a stochastic representation in terms of the symmetric normal distribution. This is required to make the E-Steps easy to implement. For example, the Student t, centered skew normal or finite mixture of normals [see Azevedo et al. (2011)] can be considered. The M-steps concerning the estimation of the parameters of those distributions should also be included. Also, the multiple group framework can be easily implemented, it is only necessary to include additional M-steps concerning the population parameters, like the usual MML estimation [see Bock and Zimowski (1997)]. In addition, the multidimensional one-parameter probit model can also be estimated by using CADEM with a few modifications.

3.1 Other latent trait distributions

Let us assume that $\theta_j \sim t_\nu$ (Student t with ν degrees of freedom), then we can write

$$
\begin{aligned}
\theta_j|t_j &\sim N(0, t_j), \\
t_j &\sim \mathrm{IG}\left(\frac{\nu}{2}, \frac{\nu}{2}\right),
\end{aligned} \qquad (3.1)
$$

where IG(α, β) represents an inverse-gamma distribution with parameters α and β. Assuming the degrees of freedom ν as known, for the sake of simplicity, the CADEM can be applied straightforwardly. The M-step remains the same. However, it is necessary to consider an additional E-step, as follows:

E-steps: Proceed as follows:

E-step 1: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\boldsymbol{\theta}^{(t)}$, evaluate (2.6).

E-step 2: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\mathbf{Z}^{(t+1)}$ and $\mathbf{t}^{(t)}$, obtained in E-step 1 and in the former iteration, respectively, evaluate

$$\widehat{\theta}_j = \mathbb{E}[\Theta_j|(\mathbf{z}_{\cdot j}, \mathbf{b}, t_j, \mathbf{y}_{\cdot j})] = \frac{\sum_{i=1}^{I}(z_{ij} + b_i)}{I + t_j},$$

E-step 3: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\mathbf{Z}^{(t+1)}$ and $\boldsymbol{\theta}^{(t+1)}$, obtained in E-step 1 and in the E-step 2, respectively, from

$$T_j|(\cdot) \sim \mathrm{IG}\left(\frac{1+\nu}{2}, \frac{\theta_j^2 + \nu}{2}\right)$$

evaluate

$$\widehat{t}_j = \mathbb{E}[T_j|(\theta_j)] = \frac{\theta_j^2 + \nu}{\nu - 1}.$$

The assumption of known degrees of freedom can be relaxed. In this case, the M-step must be modified. However, the estimation of the degrees of freedom in the Student t-distribution is not an easy task. The likelihood, for some data sets, can be ill-behaved. In this case, the maximum likelihood estimates could be biased and a Bayesian approach with a suitable prior distribution could be more appropriate. We think that Jeffreys prior, as in Fonseca et al. (2008), would be more appropriate than uniform, exponential or gamma priors. The sensitivity to prior choices depends on the number of examinees and on the number of items.

Now, let us assume that $\theta_j \sim \mathrm{SN}_C(0, 1, \gamma_\theta)$, where $\mathrm{SN}_C(0, 1, \gamma_\theta)$ stands for a centered skew-normal distribution with zero mean, unity variance and asymmetry coefficient γ_θ [see Azevedo et al. (2011)]. Following those authors, we have

$$\theta_j|(t_j, \gamma_\theta) \sim N(\tau_\theta t_j + \alpha_\theta, \sigma_\theta^2), \tag{3.2}$$

$$T_j \sim \mathrm{HN}(0, 1), \tag{3.3}$$

where HN$(0, 1)$ stands for a standard normal distribution truncated to the left of zero and

$$\tau_\theta = \sqrt{\varsigma_\theta}\delta_\theta,$$
$$\sigma_\theta^2 = \varsigma_\theta(1 - \delta_\theta^2).$$

For the sake of simplicity, if we consider γ_θ as known, the CADEM can be applied straightforwardly. The M-step remains the same. However, it is necessary to modify the E-steps as follows:

E-steps: Proceed as follows:

E-step 1: It remains the same

E-step 2: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\mathbf{Z}^{(t+1)}$ and $\mathbf{t}^{(t)}$, obtained in E-step 1 and in the former iteration, respectively, evaluate

$$\widehat{\theta}_j = \mathbb{E}[\Theta_j|(\mathbf{z}_{\cdot j}, \mathbf{b}, t_j, \mathbf{y}_{\cdot j}, \gamma_\theta)] = \frac{\sum_{i=1}^{I}(z_{ij} + \tau_\theta t_j + \alpha_\theta + b_i)}{I + \sigma_\theta^2},$$

E-step 3: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\mathbf{Z}^{(t+1)}$ and $\boldsymbol{\theta}^{(t+1)}$, obtained in E-step 1 and in the E-step 2, respectively, from $T_j|(\cdot) \sim \mathrm{HN}(\widehat{t}_j, \widehat{\psi}_j)$, where

$$\widehat{t}_j = \frac{\tau_\theta(\theta_j - \alpha_\theta)}{(\tau_\theta)^2 + (\sigma_\theta^2)},$$

$$\widehat{\psi}_j = \frac{(\sigma_\theta^2)}{(\tau_\theta)^2 + (\sigma_\theta^2)},$$

evaluate

$$\widehat{t}_j = \mathbb{E}[T_j|(\theta_j, \gamma_\theta)] = \widehat{t}_j + \sqrt{\widehat{\psi}_j}\frac{\phi(-\widehat{t}_j/\sqrt{\widehat{\psi}_j})}{1 - \Phi(-\widehat{t}_j/\sqrt{\widehat{\psi}_j})}.$$

The assumption of the known asymmetry coefficient can be relaxed. In this case, the M-step must be modified. The estimation of the asymmetry coefficient in the skew normal distribution is not an easy task either. The likelihood, for some data sets, can also be ill-behaved. In this case, the maximum likelihood estimate of the asymmetry parameter may not exist and a Bayesian approach with a suitable prior could be more appropriate. We think that the approach and priors considered in Azevedo et al. (2011) and Azevedo et al. (2011) would be suitable, due to the results obtained by the authors. They did not observe significant difference in the results obtained using different priors

3.2 Multiple group one-parameter probit model

In the multiple group model, the selected groups of respondents are of specific interest, such that group-specific population distributions need to be defined. Let us define $k = 1, \ldots, K$, the index for groups, and let us assume that

$$\theta_{jk} \sim N(\mu_{\theta_k}, \psi_{\theta_k}).$$

In this case, the E-steps and M-step need to be modified, that is,

E-steps: Proceed as follows:

E-step 1: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\boldsymbol{\theta}^{(t)}$, evaluate (2.6), with the index k representing the groups.

E-step 2: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\mathbf{Z}^{(t+1)}$, obtained in E-step 1 and in the former iteration, respectively, evaluate

$$\widehat{\theta}_j = \mathbb{E}[\Theta_j|(\mathbf{0}_{\cdot jk}, \mathbf{b}, \mu_{\theta_k}, \psi_{\theta_k}, \mathbf{y}_{\cdot jk})] = \frac{\sum_{i=1}^{I}(z_{ijk} + \mu_{\theta_k} + b_i)}{I + \psi_{\theta_k}}.$$

M-steps: Proceed as follows:

Given the current expectations $\widehat{\mathbf{W}}^{(k)(t+1)}$, $k = 1, 2$, update $\mathbf{b}$ by maximizing (2.14), that is, evaluate

M-step 1:

$$\widehat{b}_i^{(t+1)} = \frac{\sum_{j=1}^{n} \widehat{\theta}_j^{(t+1)} - \sum_{j=1}^{n} \widehat{Z}_{ij}^{(t+1)} - \mu_b/\psi_b}{n_i + 1/\psi_b}.$$

M-step 2: To update μ_{θ_k} and ψ_{θ_k}, $k = 1, \dots, K$, that is, to evaluate

$$\begin{aligned} \widehat{\mu}_{\theta_k} &= \frac{1}{n_k} \sum_{j=1}^{n_k} \widehat{\theta}_{jk}, \\ \widehat{\psi}_{\theta_k} &= \frac{1}{n_k} \sum_{j=1}^{n_k} (\widehat{\theta}_{jk} - \widehat{\mu}_{\theta_k})^2. \end{aligned} \tag{3.4}$$

3.3 Multidimensional one-parameter probit model

Let us suppose that

$$P(Y_{ij} = 1|\boldsymbol{\theta}_j, b_i) = \Phi\left(\sum_{p}^{P} \theta_{jp} - b_i\right),$$

that is, we have the multidimensional probit model [see Reckase (2009) and Beguin and Glas (2001)]. In this case, we can assume that

$$\boldsymbol{\theta}_j \sim N_P(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Psi}_{\boldsymbol{\theta}}),$$

with general structures for both $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$ [see Beguin and Glas (2001)]. In this case, both the E-steps and M-step must be modified as follows:

E-steps: Proceed as follows:

E-step 1: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\boldsymbol{\theta}^{(t)}$, evaluate

$$\widehat{Z}_{ij} = \mathbb{E}[Z_{ij}|(\boldsymbol{\theta}_j, b_i, y_{ij})]$$

$$= \begin{cases} \sum_{p=1}^{P} \theta_{jp} - b_i + \dfrac{\phi(\sum_{p=1}^{P} \theta_{jp} - b_i)}{1 - \Phi(-\sum_{p=1}^{P} \theta_{jp} + b_i)}, & \text{if } Y_{ij} = 1, \\ \sum_{p=1}^{P} \theta_{jp} - b_i - \dfrac{\phi(\sum_{p=1}^{P} \theta_{jp} - b_i)}{\Phi(-\sum_{p=1}^{P} \theta_{jp} + b_i)}, & \text{if } Y_{ij} = 0. \end{cases}$$

E-step 2: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\mathbf{Z}^{(t+1)}$ obtained in E-step 1, evaluate

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_j &= \mathbb{E}[\boldsymbol{\Theta}_j | (\mathbf{0}_{\cdot j}, b_i, \boldsymbol{\mu}_\theta, \boldsymbol{\Psi}_\theta, \mathbf{y}_{\cdot j})] \\ &= (\mathbf{I}_P + \boldsymbol{\Psi}_\theta^{-1})^{-1} \left[\sum_{i=1}^{I} (b_i \mathbf{1}_P + z_{ij} \mathbf{1}_P) + \boldsymbol{\Psi}_\theta^{-1} \boldsymbol{\mu}_\theta \right]. \end{aligned}$$

M-steps: Proceed as follows:

Given the current expectations $\widehat{\mathbf{W}}^{(k)(t+1)}$, $k = 1, 2$, update $\mathbf{b}$ by maximizing (2.14), that is, evaluate

M-step 1:

$$\widehat{b}_i^{(t+1)} = \frac{\sum_{j=1}^{n} \widehat{\theta}_j^{(t+1)} - \sum_{j=1}^{n} \widehat{Z}_{ij}^{(t+1)} - \mu_b/\psi_b}{n_i + 1/\psi_b}.$$

M-step 2: Update $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and $\psi_{\boldsymbol{\theta}}$, that is, evaluate

$$\widehat{\boldsymbol{\mu}}_\theta = \frac{1}{n} \sum_{j=1}^{n} \widehat{\boldsymbol{\theta}}_j,$$

$$\widehat{\boldsymbol{\Psi}}_\theta = \frac{1}{n} \sum_{j=1}^{n} (\widehat{\boldsymbol{\theta}}_j - \widehat{\boldsymbol{\mu}}_\theta)(\widehat{\boldsymbol{\theta}}_j - \widehat{\boldsymbol{\mu}}_\theta)'.$$

3.4 Multidimensional multiple group one-parameter probit model

Let us assume we have several groups and we are considering a multidimentional structure for latent traits, that is,

$$P(Y_{ijk} = 1 | \boldsymbol{\theta}_{jk}, b_i) = \Phi\left(\sum_{p}^{P} \theta_{jkp} - b_i \right), \qquad k = 1, \ldots, K.$$

One can combine the algorithms presented in subsections (3.2) and (3.3) in order to fit this model through CADEM.

3.5 Rasch models

Let us, in any of the above models, replace the probit link by a logit link, that is, in this case, we consider the usual Rasch models. Since the standard logistic distribution can be suitably approximated by a Student t-distribution with $\nu =$

7.581 degrees of freedom [see Chen and Dey (1998)], a hierarchical structure can be considered for the augmented data. This is equivalent to considering that

$$\begin{aligned} Z_{ij}|(\theta_j, b_i, t_i) &\sim N(\theta_j - b_i, t_i), \\ T_i &\sim \mathrm{IG}\left(\frac{7.581}{2}, \frac{7.581}{2}\right). \end{aligned} \tag{3.5}$$

Therefore, CADEM can be used, replacing the E-steps concerning the augmented data by the following E-Step.

E-steps: Proceed as follows:

E-step 1: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\boldsymbol{\theta}^{(t)}$ and $\mathbf{t}^{(t)}$, evaluate

$$\widehat{Z}_{ij} = \mathbb{E}[Z_{ij}|(\theta_j, b_i, y_{ij}, t_i)] = \begin{cases} \theta_j - b_i + \sqrt{t_i}\dfrac{\phi((\theta_j - b_i)/\sqrt{t_i})}{1 - \Phi((-\theta_j + b_i)/\sqrt{t_i})}, & \text{if } Y_{ij} = 1, \\ \theta_j - b_i - \sqrt{t_i}\dfrac{\phi((\theta_j - b_i)/\sqrt{t_i})}{\Phi((-\theta_j + b_i)/\sqrt{t_i})}, & \text{if } Y_{ij} = 0. \end{cases}$$

E-step 2: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\mathbf{Z}^{(t+1)}$ and $\mathbf{t}^{(t)}$, obtained in E-step 1 and in the former iteration, respectively, evaluate

$$\widehat{\theta}_j = \mathbb{E}[\Theta_j|(\mathbf{z}_{\cdot j}, \mathbf{b}, \mathbf{y}_{\cdot j})] = \frac{\sum_{i=1}^{I}(z_{ijk} + b_i)}{I/t_i + 1}.$$

E-step 3: Given the current estimates of $\mathbf{b}^{(t)}$ and current expectations of $\mathbf{Z}^{(t+1)}$ and $\boldsymbol{\theta}^{(t)}$, obtained in E-step 1 and E-step 2, respectively, evaluate

$$\widehat{T}_i = \mathbb{E}[T_i|(\mathbf{z}_{i\cdot}, b_i, \theta_j, \mathbf{y}_{i\cdot})] = \frac{\sum_{j=1}^{n}(z_{ij} - \theta_j + b_i)^2 + 7.581/2}{n + 7.581/2 - 2}.$$

The extension of CADEM to fit other IRT models certainly deserves more investigation. For instance, in the family of the two- and three- parameter and polytomous models [see Nering and Ostini (2010)] it would be necessary to calculate the expectation of products of the unobserved variables. In addition, due to the more complex mathematical structure of polytomous models, it would be more difficult to find suitable augmented data structures. However, this is far beyond the scope of this paper.

4 Simulation

In order to compare our approach with the usual ones, we performed a simulation study. In this effort, we compared the CADEM with marginal maximum likelihood (MML), marginal maximum a posterior (MMAP) and full Gibbs sampling

Table 1 *Description of the estimation methods*

Method	Item Parameter	Latent trait
CADEM ML	Maximum Likelihood CADEM	
CADEM MAP	Maximum a posterior CADEM	
FGS	Full Gibbs sampling	
MML	MML	–
MMAP (FGS prior)	MMAP with CADEM/FGS prior	–
MMAP (BILOG prior)	MMAP with Bilog-MG prior	–
MMAP MAP (FGS prior)	MMAP with CADEM/FGS prior	MAP
MMAP EAP (FGS prior)	MMAP with CADEM/FGS prior	EAP
MMAP MAP (BILOG prior)	MMAP with Bilog-MG prior	MAP
MMAP EAP (BILOG prior)	MMAP with Bilog-MG prior	EAP

algorithm (FGS). Also, we considered the expectation a posteriori (EAP) and maximum a posterior estimates (MAP) to the latent traits, using the MMAP estimates of the difficulty parameters. Also, for MMAP estimation we considered different prior distributions. One of them is the same prior used in the CADEM and FGS, that is, $N(0,9)$. Another is the default prior used by the commercial package Bilog-MG (see http://www.ssicentral.com/irt/index.html). Table 1 describes the estimation methods used.

We considered a situation where $n = 500$ examinees answer a test of $I = 30$ items. The latent traits of the examinees were simulated from a $N(0,1)$ distribution. The item difficulty parameters were chosen ranging from $(-3,3)$, that is, we intend to cover the whole latent trait scale, concerning the simulated values. We generated a set of $R = 50$ replicas, that is, responses of the examinees to the items. In each one of the replicas we used each one of the estimation methods described in Table 1.

To compare the performance of the estimation methods, we considered some appropriate statistics. Let ϑ_l be an element of (θ_j, b_i), where l is a convenient index (i or j) and $\widehat{\vartheta}_{lr}$ its respective estimate obtained in the replica r, $r = 1, \dots, R$. Define also $\widehat{\widehat{\vartheta}}_l = \frac{1}{R}\sum_{r=1}^{R}\widehat{\vartheta}_{lr}$. The aforementioned statistics are as follows:

- Corr: correlation between $\widehat{\widehat{\vartheta}}_l$ and ϑ_l.
- MeanSE: mean of the standard errors over the parameters.
- Bias: bias of the estimates: $(\widehat{\widehat{\vartheta}}_l - \vartheta_l)$.
- Var: variance of the estimates: $\frac{1}{R}\sum_{r=1}^{R}(\widehat{\vartheta}_{lr} - \widehat{\widehat{\vartheta}}_l)^2$.
- RMSE: square root of the mean square error (MSE): $\sqrt{\frac{1}{R}\sum_{r=1}^{R}(\widehat{\vartheta}_{lr} - \vartheta_l)^2}$.
- AVRB: absolute value of the relative bias: $\frac{|\widehat{\widehat{\vartheta}}_l - \vartheta_l|}{|\vartheta_l|}$.

Tables 2 and 3 present the statistics concerning the latent traits and difficulty parameter estimates, respectively. The results are quite similar, indicating that the

Table 2 *Results for latent trait estimation*

Method	Var	MeanSE	Bias	RMSE
CADEM ML	0.076	0.229	0.037	0.298
CADEM MAP	0.076	0.229	0.037	0.298
FGS	0.081	0.298	0.034	0.299
MML MAP	0.079	0.296	0.036	0.299
MML EAP	0.081	0.298	0.033	0.300
MMAP MAP (FGS prior)	0.078	0.294	0.036	0.299
MMAP EAP (FGS prior)	0.080	0.297	0.033	0.299
MMAP MAP (BILOG prior)	0.078	0.294	0.036	0.299
MMAP EAP (BILOG prior)	0.081	0.297	0.033	0.299

Table 3 *Results for difficulty parameter estimation*

Method	Var	MeanSE	Bias	RMSE
CADEM ML	0.007	0.079	0.033	0.088
CADEM MAP	0.006	0.079	0.032	0.088
FGS	0.007	0.082	0.032	0.088
MML	0.007	0.081	0.030	0.088
MMAP (FGS prior)	0.007	0.080	0.029	0.087
MMAP (BILOG prior)	0.007	0.080	0.029	0.087

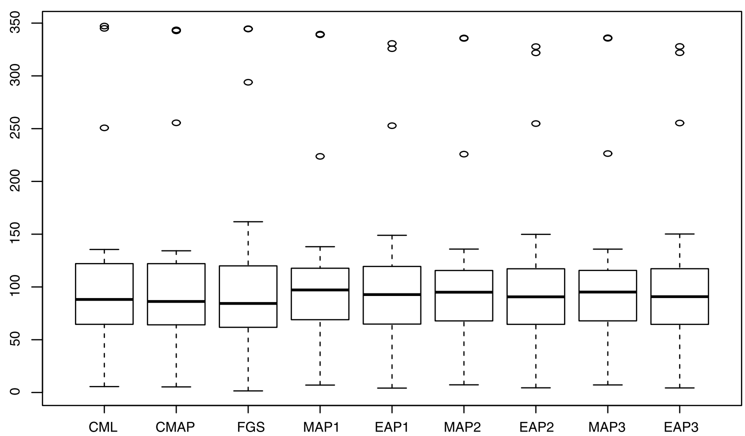

Figure 1 *AVRB for the latent traits estimates.*

estimation methods recovered the parameters with the same accuracy. Figures 1 and 2 suggest the same conclusions. Table 4 indicates that MVM and MMAP (including EAP/MAP) methods were the fastest, followed by CADEM, even though the difference is quite small. As expected, the FGS required more time than the

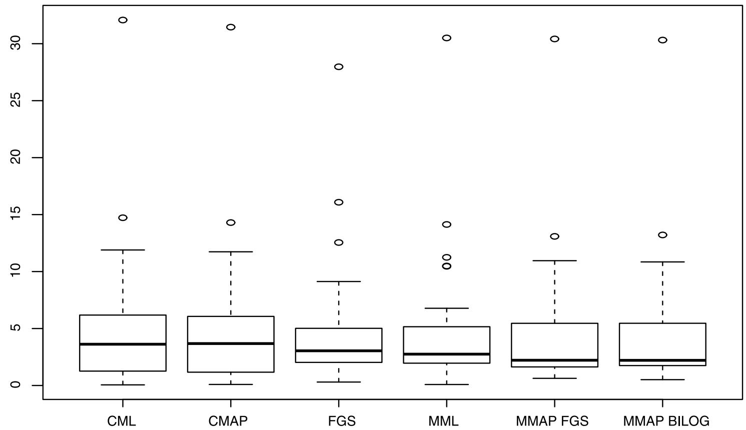

Figure 2 *AVRB for the difficulty parameter estimates.*

Table 4 *Spent time (in seconds), achieved precision for the item estimation (API), number of required iterations for the item estimation (NRII), achieved precision for the latent trait estimate (APLT)*

Method	ST	PAI	NRII	APLT
CADEM ML (CML)	2.818	0.000	93.920	–
CADEM MAP (CMAP)	2.833	0.000	93.840	–
FGS	292.816	0.000		–
MML MAP	0.368	0.000	80.300	0.000
MML EAP	0.327	0.000	80.300	–
MMAP MAP with FGS prior	0.363	0.000	79.940	0.000
MMAP EAP with FGS prior	0.332	0.000	79.940	–
MMAP MAP with BILOG prior	0.363	0.000	79.920	0.000
MMAP EAP with BILOG prior	0.332	0.000	79.920	–

others to estimate the parameters. In conclusion, all methods behaved in a similar way. However, CADEM is faster than the MCMC algorithms and, as shown before, more flexible to estimate IRT models than the marginal likelihood based methods. Therefore, we claim that CADEM is an interesting alternative to fit IRT models.

5 Comments and conclusions

A new estimation method, named CADEM, was proposed for fitting one parameter (Rasch) models. We showed that it recovers the parameters equally as well as the usual methods. The CADEM structure can be extended to handle more complex IRT models, such as multiple groups, two- and three- parameters and multidimensional models, since an augmented data framework, concerning the IRF, is available. CADEM has the advantage of being as fast as MML and MMAP methods and as flexible as MCMC based methods. In addition, CADEM considers the uncertainty of latent trait estimates in the difficulty parameter estimation. As was shown, CADEM can be easily extended to incorporate prior information. Non-normal distributions for the latent traits can be also considered since they admit a stochastic representation in terms of the symmetric normal distribution. More investigation is necessary to improve the standard error calculations as well as to extend CADEM to estimate more complex IRT models.

Appendix

To calculate the standard errors associated with the difficulty parameter estimates $\mathrm{SE}(b_i)$, first, from (2.4), notice that

$$\frac{\partial l^c}{\partial b_i} = \sum_{j=1}^{n} (z_{ij} - \theta_j - b_i), \tag{A.1}$$

$$\frac{-\partial^2 l^c}{\partial b_i^2} = n_i. \tag{A.2}$$

For the sake of simplicity, we use (A.1) and (A.2) in (2.9), by calculating the expectations and variance through the distribution $Z_{ij}|\theta_j, y_{ij}$. Therefore, it follows that

$$\mathrm{SE}(b_i) = \left(n_i - \sum_{j=1}^{n} \mathcal{V}ar(Z_{ij}|\widehat{\theta}_j, \widehat{b}_i, y_{ij}) \right)^{-1}.$$

For the standard errors associated with the latent traits, we use formula (2.10), replacing the expectation and variance $\mathcal{E}_Z$ and $\mathcal{V}ar_Z$ by $\mathcal{E}_{Z|\widehat{\theta}}$ and $\mathcal{V}ar_{Z|\widehat{\theta}}$, respectively. That is,

$$\mathcal{V}ar(\Theta_j|\mathbf{y}) = (1 + \mathcal{I}_j)^{-2} \left(\sum_{i=1}^{I} (\mathcal{V}ar(Z_{ij}|\widehat{\theta}_j, \widehat{b}_i, y_{ij})) \right) + (1 + \mathcal{I}_j)^{-1}.$$

For the Bayesian calculations, it is only necessary to consider the derivative of the log of (2.11), which is not presented here.

Acknowledgments

We would like to thank the CNPq (National Council for Scientific and Technological Research), from Brazil, for financial support, through the project 479802/2009-2 and to the CENAPAD (National Center for High Performance), from Brazil, for the computational support that made the simulation studies and data analysis possible.

References

Albert, J. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational and Behavioral Statistics* **17**, 251–269.

Azevedo, C. L. N., Bolfarine, H. and Andrade, D. F. (2011). Bayesian inference for a skew-normal IRT model under the centred parameterization. *Computational Statistics & Data Analysis* **55**, 353–365.

Andrade, D. F. and Tavares, H. R. (2005). Item response theory for longitudinal data: Population parameter estimation. *Journal of Multivariate Analysis* **51**, 1–22. MR2164119

Azevedo, C. L. N., Bolfarine, H. and Andrade, D. F. (2011). Parameter recovery for a skew normal IRT model under a Bayesian approach: Hierarchical framework, prior and kernel sensitivity and sample size. *Journal of Statistical Computation and Simulation*. To appear, DOI:10.1080/00949655.2011.591798.

Baker, F. B. and Seock-Ho, K. (2004). *Item Response Theory: Parameter Estimation Techniques*, 2nd ed. New York, NY: Marcel Dekker. MR2086862

Bazan, J. L., Branco, M. D. and Bolfarine, H. (2006). A skew item response model. *Bayesian Analysis* **1**, 861–892. MR2282209

Béguin, A. A. B. and Glas, C. A. W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika* **66**, 541–561. MR1961913

Bock, D. R. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika* **46**, 433–459. MR0668311

Bock, D. R. and Zimowski, M. F. (1997). The multiple groups IRT. In *Handbook of Modern Item Response Theory* (W. J. van der Linden and R. K. Hambleton, eds.). New York, NY: Springer-Verlag.

Chen, Ming-Hui and Dey, D. K. (1998). Bayesian modeling of correlated binary responses via scale mixture of multivariate normal link functions. *Sankhya* **60**, 322–343. MR1718864

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society., Ser. B* **39**, 1–38. MR0501537

Fonseca, T., Ferreira, M. A. and Migon, H. S. (2008). Objective Bayesian analysis for the Student-*t* regression model. *Biometrika* **95**(2), 325–333. MR2521587

Fox, J.-P. (2003). Stochastic EM for estimating the parameters of a multilevel IRT model. *British Journal of Mathematical and Statistical Psychology* **56**, 65–81. MR2101794

Fox, J.-P. and Glas, C. A. W. (2001). Bayesian Estimation of a Multilevel IRT Model using Gibbs Sampling. *Psychometrika* **66**, 269–286. MR1836937

Liu, C., Rubin, D. B. and Wu, Y. N. (1998). Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika* **85**, 755–770. MR1666758

Lous, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Ser. B* **44**, 226–233. MR0676213

Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika* **2**, 177–195. MR0848104

Nering, M. L. and Ostini, R. (2010). *Handbook of Polytomous Item Response Theory Models*. London: Routledge Academic.

Patz, J. R. and Junker, B. W. (1999a). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics* **24**, 146–178.

Patz, J. R. and Junker, B. W. (1999b). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics* **24**(4), 342–366.

Reckase, M. D. (2009). *Multidimensional Item Response Theory*. New York, NY: Springer-Verlag.

Department of Statistics
Institute of Mathematics,
Statistics and Computer Science
University of Campinas
Campinas/SP, 13083–859
Brazil
E-mail: cnaber@ime.unicamp.br

Department of Informatica and Statistics
Federal University of Santa Catarina
Florianopólis/SC, 57072–970
Brazil
E-mail: dandrade@ine.ufsc.br