RT-MAE 2008-13

AN ESTIMATION METHOD FOR LATENT TRAIT AND POPULATION PARAMETERS IN NRM

by

Caio Lucidius Naberezny Azevedo
and
Dalton Francisco de Andrade

Palavras-Chave: Nominal response model, latent traint, population parameter, MCMC simulation, Metropolis-Hastings.

Classificação AMS: QA 277.24

- Novembro de 2008 -

An estimation method for latent trait and population parameters in nominal response model

Caio Lucidius Naberezny Azevedo[1] and Dalton Francisco de Andrade[2]

[1] Department of Statistics, Institute of Math and Statistics, University of São Paulo and Centro de seleção e de promoção de eventos, University of Brasília, Brazil
[2] Department of Informatic and Statistics, Federal University of Santa Catarina, Brazil

Abstract

The nominal response model(NRM) was proposed by (Bock, 1972) in order to improve the latent trait (ability) estimation in multiple choice tests with nominal items. When the item parameters are known, expectation a posterior or maximum a posterior methods are commonly employed to estimate the latent traits, under symmetric normality assumption of the latent trait distribution. However, when this item set is presented to a new group of examinees, it is not only necessary to estimate their latent traits but also the population parameters of this group. This article has two main purposes: first, to develop a Monte Carlo Markov Chain algorithm to estimate both latent traits and population parameters concurrently. Second, to compare, in the latent trait recovering, the performance of this method with three other methods: maximum likelihood (ML), expectation a posterior (EAP) and maximum a posterior (MAP). The comparisons are performed by varying the total number of items (NI), the number of categories (NC) and the values of the mean and the variance of the latent trait distribution. The results showed that MCMC+MH outperforms the other methods concerning the latent traits estimation as well as it recoveries properly the population parameters. Furthermore, we found that NI accounts for the highest percentage of the variability in the accuracy of latent trait estimation.

Classificação AMS: QA 277.24

keywords: nominal response model, latent trait, population parameter, MCMC simulation, Metropolis-Hastings

1 Introduction

The nominal response model (NRM), was proposed by Bock (1972) in order to improve the latent trait estimation in nominal tests. Let us suppose that a test consisting of I items with h_i categories each is administered to n examinees, and a random variable Y_{ijh} which indicates the category chosen by subject j to item i, by assuming value 1 for this category and 0 for the all remaining ones. The NRM, which represents such probability, is given by

$$P_{ijh} = P(Y_{ijh} = 1|\theta_j, \zeta_i) = \frac{\exp\left[a_{ih}\left(\theta_j - b_{ih}\right)\right]}{\sum_{s=1}^{h_i} \exp\left[a_{is}\left(\theta_j - b_{is}\right)\right]} = \frac{\exp\left(d_{ih} + a_{ih}\theta_j\right)}{\sum_{s=1}^{h_i} \exp\left(d_{is} + a_{is}\theta_j\right)} \tag{1}$$

where

θ_j : latent trait of subject j,

ζ_i : $(d, a) = (d_{i1}, \ldots, d_{ih_i}, a_{i1}, \ldots, a_{im_i})$,

a_{ih} : slope (discrimination) parameter of the category h of item i,

b_{ih} : difficult parameter of the category h of item i ,

$d_{ih} = -a_{ih}b_{ih}$: intercept parameter of the category h of item i.

Discussion about the interpretations of the model and the item parameters can be found in Bock (1972), Ayala (1992), DeMars (2003), Azevedo (2003) and Baker and Kim (2004), for example.

Most of the articles that deal with estimation in NRM are concerned with item parameter estimation, under different conditions. For instance, see Ayala and Sava-Bolesta (1999), Bolt et al. (2001), Wollack et al. (2002) and DeMars (2003). Latent traits estimation are discussed in Ayala (1989), Ayala (1992) and Baker and Kim (2004). The latter discusses also the estimation of the population parameters via marginal maximum likelihood (MML), but not jointly with the latent traits. The first two aforementioned articles use EAP method to estimate the latent traits, while Baker and Kim (2004) consider ML, EAP and MAP. In both EAP and MAP methods it is assumed a standard normal distribution for the latent traits.

In this work we are concerning with the situation where the item parameters in the NRM are known in some metric, see for example Andrade and Tavares (2005), and we want to estimate the latent traits and the population parameters of a group of examinees, different from that one used to calibrate the item parameters. In this case, the population parameters are free to be estimated.

This paper has two goals: first, to develop a Metropolis-Hastings within Gibbs sampling algorithm to estimate jointly the latent traits and the population parameters. Second, to compare, in the latent trait recovering, the performance of this method with three others: maximum likelihood (ML), expectation a posterior (EAP) and maximum a posterior (MAP). The comparisons are performed by varying the total number of items (NI), the number of categories (NC) and the values of the mean and the variance of the latent trait distribution. The NI and NG are known to have influence in the estimation accuracy, according to Ayala and Sava-Bolesta (1999), Wollack et al. (2002) and DeMars (2003). Furthermore, we want to verify the impact of different values of the mean and variance of the latent trait distribution. We believe that the latent traits will be better estimated when one uses information about the population parameters.

In Section 2 we present the MCMC algorithm. In Section 3 we perform a simulation study to compare the aforementioned methods and in Section 4 we outline some comments and conclusions.

2 MCMC estimation and the other methods

MCMC algorithms are powerful tools to make bayesian inference, see Gamerman & Lopes (2006), for details. One of the most used algorithm of this class is the Gibbs sampling one. This procedure calculates, empirically, joint posterior distributions trough the so-called full conditional distributions. In many situations it is not possible to obtain these distributions analytically. This is the case of NRM. A way of avoiding such problem is to use some auxiliary algorithm as the Metropolis-Hastings or the adaptive rejection, see Patz and Junker (1999) and Chen et al. (2000), for example. Wollack et al. (2002) proposed a MCMC Gibbs sampling with adaptive rejection sampling algorithm to fit the NRM under standard normal latent distribution by using WinBugs package. On the other hand, Patz and Junker (1999) developed a Metropolis-Hastings within Gibbs sampling algorithm for the one, two and three parameter logistic models and for the generalized partial credit model. We considered such approach for the NRM, henceforth MCMC+MH approach.

First, let us calculate the joint posterior distribution. Considering the usual assumptions of conditional independence, the likelihood, for the latent trait of subject j, is given by

$$L(\theta_j|\boldsymbol{y}_{.j.},\boldsymbol{\zeta}) = \prod_{i=1}^{I}\prod_{h=1}^{m_i} P_{ijh}^{y_{ijh}}, \tag{2}$$

where P_{ijh} is as described in (1), $\boldsymbol{y}_{.j.} = (y_{1j1},\ldots,y_{1jm_i},\ldots,y_{Ij1},\ldots,y_{Ijm_i})$ and $\boldsymbol{\zeta} = (\zeta_1,\ldots,\zeta_I)$. For bayesian inference we assume the following prior

$$\begin{aligned} p(\boldsymbol{\theta},\boldsymbol{\eta}_\theta) &= \{p(\boldsymbol{\theta}|\boldsymbol{\eta}_\theta)\}\, p(\boldsymbol{\eta}_\theta) \\ &= \left\{\prod_{j=1}^{n} p(\theta_j|\boldsymbol{\eta}_\theta)\right\} p(\mu_\theta)p(\psi_\theta), \end{aligned}$$

where $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_n)$ and $\boldsymbol{\eta}_\theta = (\mu_\theta,\psi_\theta)$. The prior for θ_j is assumed to be

$$\theta_j|\boldsymbol{\eta}_\theta \sim N(\mu_\theta,\psi_\theta). \tag{3}$$

For the population parameters, natural choices, which lead to conditional conjugate families, see Gelman (2006), are:

$$\mu_\theta \sim N(\mu_\mu, \psi_\mu) \tag{4}$$
$$\psi_\theta \sim IG(\nu_0/2, \kappa_0/2). \tag{5}$$

Thereafore, from (2), (3), (4) and (5) it follows that the joint posterior distribution is given by:

$$\begin{aligned} p(\theta, \eta_\theta | y_{...}) \propto & \left\{ \prod_{i=1}^{I} \prod_{h=1}^{m_i} P_{ijh}^{y_{ijh}} \right\} \left\{ \prod_{j=1}^{n} \exp\left[-\frac{(\theta_j - \mu_\theta)^2}{2\psi_\theta} \right] \right\} \psi_\theta^{-n/2} \\ \times & \left\{ \exp\left[-\frac{(\mu_\theta - \mu_\mu)^2}{2\psi_\mu} \right] \right\} \left\{ \psi_\theta^{-(\nu_0/2+1)} \exp\left(-\frac{\kappa_0}{2\psi_\theta} \right) \right\}. \end{aligned} \tag{6}$$

The distribution (6) has an intractable form. Also, the full conditional distribution of the latent traits is not known. However, the full distributions of the population parameters are known and easy to sample from. Therefore, a hybrid MCMC algorithm can be used to simulate from (6). This algorithm is composed by a Metropolis-Hastings within Gibbs sampling step for the latent traits and two Gibbs sampling steps to estimate η_θ. A kernel function, see Patz and Junker (1999), is necessary to drawn from the latent traits. Following Patz and Junker (1999) we choose:

$$q(\theta_j^{(t)}, \theta_j^{(t-1)}) \sim N(\theta_j^{(t-1)}, \sigma_\theta^2),$$

where $\theta_j^{(t)}$ is the current simulated value of θ and $\theta_j^{(t-1)}$ is the simulated value of the former iteration. Denoting (.) the set of all other parameters, the hybrid MCMC algorithm (MCMC+MH) is as follows:

1. Attempt to draw $\theta_j^{(t)} | (.) \sim p(\theta_j | \zeta^{(t-1)}, y_{...})$ (full conditional distribution), for $j = 1, \ldots, n$ mutually independent:

   (a) Draw $\theta_j^{(*)} \sim N(\theta_j^{(t-1)}, \sigma_\theta^2)$

   (b) Accept $\theta_j^{(t)} = \theta_j^{(*)}$ with probability

$$\pi_j\left(\theta_j^{(t-1)},\theta_j^{(*)}\right)=\min\left\{\frac{L(\theta_j^{(*)}|y_{.j.},\zeta)exp\left\{-\frac{\left(\theta_j^{(*)}-\mu_{\theta_j}\right)^2}{2\sigma^2_{\theta_j}}\right\}}{L(\theta_j^{(t-1)}|y_{.j.},\zeta)exp\left\{-\frac{\left(\theta_j^{(t-1)}-\mu_{\theta_j}\right)^2}{2\sigma^2_{\theta_j}}\right\}},1\right\}$$

otherwise, set $\theta_j^{(t)}=\theta_j^{(t-1)}$.

2. Draw $\mu_\theta^{(t)}$ from $\mu_\theta|(.)\sim N(\widehat{\psi}_\mu^{(t)}\widehat{\mu}_\theta^{(t)},\widehat{\psi}_\mu^{(t)})$, where:

$$\begin{aligned}\widehat{\mu}_\theta^{(t)} &= \frac{1}{\psi_\theta^{(t-1)}}\sum_{j=1}^{n}\left(\theta_j^{(t)}\right)+\frac{\mu_\mu}{\psi_\mu}\\ \widehat{\psi}_\mu^{(t)} &= \left(\frac{n}{\psi_\theta^{(t-1)}}+\frac{1}{\psi_\mu}\right)^{-1}\end{aligned}$$

3. Draw $\psi_\theta^{(t)}$ from $\psi_\theta|(.)\sim N(\widehat{\nu}^{(t)},\widehat{\kappa}^{(t)})$, where:

$$\begin{aligned}\widehat{\nu}^{(t)} &= \frac{1}{2}\left[\sum_{j=1}^{n}\left(\theta_j^{(t)}-\mu_\theta^{(t)}\right)^2+\nu_0\right]\\ \widehat{\kappa}^{(t)} &= \frac{n+\kappa_0}{2}.\end{aligned}$$

Giving suitable starting values $\left(\theta^{(0)},\eta^{(0)}\right)$, the iteration of the three above steps comprises the MCMC+MH algorithm.

The MV, EAP and MAP are well described in the literature, see for example, Baker and Kim (2004). Therefore they will be not presented in this work. In the next section we will present a simulation study.

3 Simulation study

To compare the performance of the four estimation methods we conducted a Monte Carlo simulation study, according to Harwell et al. (1988). That is, considering replication of data sets and using appropriate statistics to measure the goodness of the estimates. The number of replication was R = 20. This choice was based on previous works, see Ayala and Sava-Bolesta (1999), DeMars (2003) and Wollack et al. (2002). The factors (number of levels) considered were: the number of items (NI) (20, 30, 40), the number of categories per item

(NC) (4, 5), the values of the population means (-2, 0, 2) and the values of the population variance (0.5, 1, 1.5). Hence, there are $3 \times 2 \times 3 \times 3 = 54$ combinations. Once that our interest lies on the main factors (and not in the possible interactions) and also because the large time demanding by the MCMC methods, we considered a fractional factorial experiment using 18 of the 54 combinations. These combinations were generated in order to ensure the estimability of the main effects and they are presented in Table 1. In all simulations we set n = 600 examinees.

Tables 6 and 7 present the items used to built the 4 alternatives and 5 alternatives tests, respectively. The 20 items tests were built using the 20 first items and in the 40 item tests we considered all the items. The 30 items tests were built considering the 21 first items and the items: 23, 25, 27, 29, 32, 34, 36, 38 and 40. This allows to have tests in which the items range properly in terms of difficult and discrimination. That is, it is possible to cover the latent trait range and also to discriminate between examinees with different abilities.

One of the most important aspects of the MCMC methods is to verify the convergence of the simulated values to the posterior densities of interest. There are several suggestions in the literature but no agreement about the most suitable one, see Gamerman and Lopes (2006), for example. In this work we considered: the monitoring of the chains generate by three different set of starting values, trace plots and Geweke statistics. In the first starting values set the latent traits were drawn from a N(0,1) distribution and μ_θ and ψ_θ were fixed equal to 0 and 1, respectively. In the second set, the latent traits were all fixed to 0, μ_θ was drawn from N(0,1) and ψ_θ was drawn from U(0,2). Finally, in the third set, the standardized scores were used as the latent traits starting values. For μ_θ and ψ_θ we calculated the sample mean and variance of the standardized scores plus values generated from a N(0,1) distribution. We simulated a set of responses for the combination $(NI = 20, NC = 4, \mu_\theta = -2, \psi_\theta = 1.5)$ and applied the convergence assessment procedures aforementioned. The Geweke statistics were calculated considering independent samples of size 100 after 2000 iterations. They showed that the convergence occurred for all observed parameters (the population parameters and some latent traits randomly chosen). Also the observation of trace plots indicated that a burn-in of 2000 is enough to drawn from the posterior densities. The Figure 1 shows the autocorrelations and trace plots of the generated chains for one latent trait (randomly chosen), the mean and the variance obtained from different starting values. It shows that the autocorrelations become negligible after a lag of 20. Therefore we decided to consider a bur-in of 2000, simulating more 18000 values after that and retaining every 20 values. Hence we have 900 values to estimate the posterior densities.

Table 1: Level factor combinations of the fractional factorial design

NI	NG	μ_θ	ψ_θ	NI	NG	μ_θ	ψ_θ	NI	NG	μ_θ	ψ_θ
20	4	-2.0	1.5	30	4	-2.0	1.0	40	4	-2.0	0.5
20	4	2.0	0.5	30	4	2.0	1.5	40	4	2.0	1.0
20	4	0.0	1.0	30	4	0.0	0.5	40	4	0.0	1.5
20	5	-2.0	1.5	30	5	-2.0	1.0	40	5	-2.0	0.5
20	5	2.0	0.5	30	5	2.0	1.5	40	5	2.0	1.0
20	5	0.0	1.0	30	5	0.0	0.5	40	5	0.0	1.5

To compare the performance of the estimation methods we considered the following decomposition

$$
\begin{aligned}
MSE &= BIAS^2 + Variance \\
\frac{1}{R}\sum_{r=1}^{R}\left(\theta_j - \widehat{\theta}_{jr}\right)^2 &= \left(\theta_j - \widehat{\overline{\theta}}_j\right)^2 + \frac{1}{R}\sum_{r=1}^{R}\left(\widehat{\theta}_{jr} - \widehat{\overline{\theta}}_j\right)^2,
\end{aligned}
$$

where $\widehat{\theta}_{jr}$ is the estimate of the latent trait of examinee j obtained from the data set r and $\widehat{\overline{\theta}}_j = \frac{1}{R}\sum_{j=1}^{R}\widehat{\theta}_{jr}$. The four statistics used to evaluate the accuracy of the latent trait estimation were:

Corr mean of the correlation between $\overline{\theta}_j$ and θ_j among the examinees .

Bias $\frac{1}{n}\sum_{j=1}^{n}\left(\theta_j - \widehat{\overline{\theta}}_j\right)$.

Var $\frac{1}{n}\sum_{j=1}^{n}\frac{1}{R}\sum_{r=1}^{R}\left(\widehat{\theta}_{jr} - \widehat{\overline{\theta}}_j\right)^2$.

MSE square root of $\frac{1}{n}\sum_{j=1}^{n}\frac{1}{R}\sum_{r=1}^{R}\left(\theta_j - \widehat{\theta}_{jr}\right)^2$.

Table 2 presents the aforementioned statistics for each estimation method for all level factors combination. The results indicate that the MCMC performs equally or better than the other methods. When the true population parameters match with that considered in the usual methods, the four estimation procedures produce quite similar results. However, when this is not true, the MCMC produce more accurate results. This is clear when the true population parameters are far from (0,1). Hence we concluded that estimating the population parameters improved the latent traits estimates. Also, the MCMC+MH approach presented the most accurate results.

Figures 2, 3 and 4 present the MSE per latent trait ranges for different combinations of level factors. One can see that, even though the MCMC+MH was the most accurate estimation procedure, the MSE of the estimates of the other methods are smaller than that of the MCMC+MH ones, for some latent trait ranges.

Therefore the MCMC+MH is not uniformly better than the other three methods (concerning the latent trait values). Also, we can notice that the highest MSE occur for the more extreme latent trait values. Furthermore, when the population parameters are far from (0,1) values, the MSE tend to be higher. In general, one can say that MCMC+MH algorithm outperforms the other methods.

An ANOVA was calculated for the fractional factorial design by considering $\ln(MSE)$ as the response variable, see Ayala and Sava-Bolesta (1999) and DeMars (2003). Since that there are many observations (number of examinees) by each combination of factor levels, the statistic ω^2 was considered instead of the F one. This allows to evaluate the contribution of each main effect in the difference of the accuracy of the estimates, see Ayala and Sava-Bolesta (1999) and DeMars (2003), for example. Table 3 presents such results. We see that the NI accounts for the highest percentage of the difference in the $\ln(MSE)$, while the other factors (NC, mean and variance) have a small impact. The remaining effects, including possible interactions and other factors which could be considered as the sample size ratio, see DeMars (2003), have, all together, a reasonable influence. However, in the design that was considered it is not possible to identify the contribution of each one. By inspecting Table 4, one can see that the population parameters were all well recovered by the MCMC+MH approach. Table 5 presents the number of iterations necessary to obtain the convergence in the iterative process for MV and MAP as well as the precision achieved. Also, the time spent by the estimation process is displayed for the two former methods and MCMC+MH algorithm. It is clear that MAP requires less iterations and then, it is less time demanding. Also, we can see that the MCMC+MH algorithm spent much more time than the other methods.

Table 2: Statistics for the latent trait estimation

NI	NC	μ_θ	ψ_θ	E.M.	Corr	Var	Bias	RMSE
20	4	-2.0	1.5	MV	0.988	0.571	-0.134	0.813
				MAP	0.980	0.101	0.405	0.648
				EAP	0.981	0.105	0.388	0.633
				MCMC+MH	0.989	0.159	-0.010	0.462
20	4	2.0	0.5	MV	0.987	0.240	0.041	0.509
				MAP	0.990	0.110	-0.322	0.491
				EAP	0.990	0.113	-0.312	0.486
				MCMC+MH	0.990	0.092	0.003	0.369
20	4	0.0	1.0	MV	0.995	0.164	-0.001	0.417
				MAP	0.995	0.113	0.000	0.379
				EAP	0.995	0.114	-0.001	0.379
				MCMC+MH	0.995	0.115	-0.001	0.379
20	5	-2.0	1.5	MV	0.980	1.416	-0.113	1.234
				MAP	0.977	0.113	0.474	0.729
				EAP	0.978	0.116	0.455	0.712
				MCMC+MH	0.987	0.177	0.004	0.502
20	5	2.0	0.5	MV	0.987	0.233	0.031	0.499
				MAP	0.987	0.124	-0.353	0.531
				EAP	0.986	0.126	-0.345	0.526
				MCMC+MH	0.987	0.098	0.003	0.391
20	5	0.0	1.0	MV	0.994	0.182	-0.003	0.440
				MAP	0.995	0.118	0.003	0.395
				EAP	0.995	0.120	0.001	0.395
				MCMC+MH	0.995	0.120	0.001	0.395
30	4	-2.0	1.0	MV	0.991	0.260	-0.065	0.539
				MAP	0.987	0.086	0.304	0.499
				EAP	0.987	0.088	0.287	0.487
				MCMC+MH	0.991	0.108	0.002	0.380
30	4	2.0	1.5	MV	0.995	0.241	0.052	0.517
				MAP	0.983	0.073	-0.284	0.511
				EAP	0.983	0.074	-0.273	0.502
				MCMC+MH	0.992	0.104	-0.013	0.376
30	4	0.0	0.5	MV	0.995	0.100	-0.008	0.324
				MAP	0.995	0.080	-0.002	0.299
				EAP	0.995	0.081	-0.006	0.300
				MCMC+MH	0.995	0.066	-0.002	0.295
30	5	-2.0	1.0	MV	0.986	0.519	-0.059	0.747
				MAP	0.988	0.098	0.346	0.538
				EAP	0.988	0.101	0.327	0.525
				MCMC+MH	0.992	0.127	-0.001	0.404
30	5	2.0	1.5	MV	0.995	0.206	0.045	0.476
				MAP	0.990	0.084	-0.279	0.490
				EAP	0.989	0.084	-0.273	0.487
				MCMC+MH	0.995	0.112	0.004	0.371
30	5	0.0	0.5	MV	0.993	0.121	-0.008	0.358
				MAP	0.993	0.093	-0.001	0.326
				EAP	0.993	0.094	-0.006	0.326
				MCMC+MH	0.993	0.075	-0.001	0.319

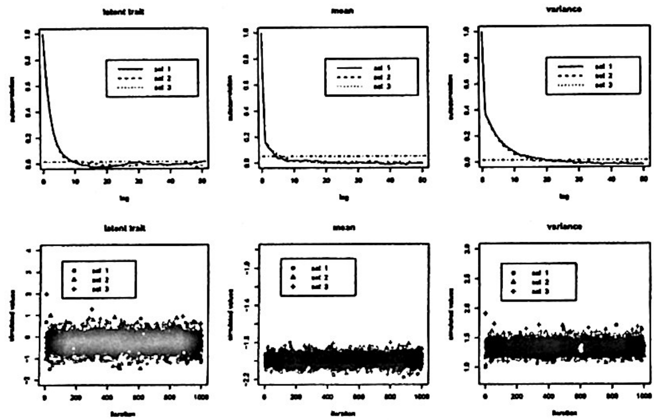

Figure 1: Autocorrelations and trace plots for chains generated from different starting values.

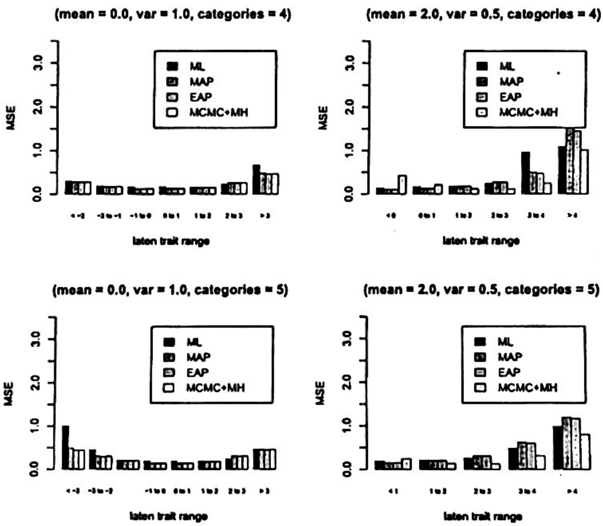

Figure 2: MSE per latent trait range for some 20 items tests.

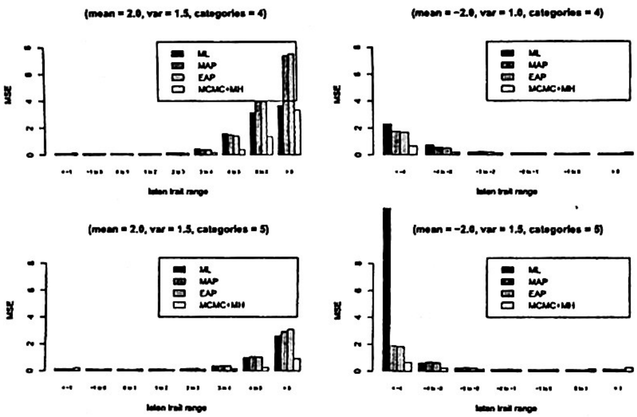

Figure 3: MSE per latent trait range for some 30 items tests.

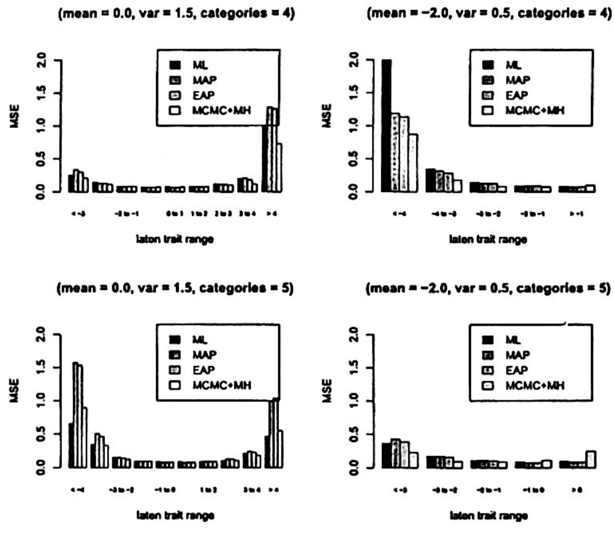

Figure 4: MSE per latent trait range for some 40 items tests.

Continuation of Table 2

NI	NC	μ_θ	ψ_θ	E.M.	Corr	Var	Bias	RMSE
40	4	-2.0	0.5	MV	0.994	0.130	-0.035	0.374
				MAP	0.991	0.070	0.193	0.355
				EAP	0.991	0.072	0.179	0.348
				MCMC+MH	0.991	0.066	0.000	0.298
40	4	2.0	1.0	MV	0.995	0.155	0.032	0.412
				MAP	0.990	0.065	-0.212	0.393
				EAP	0.990	0.066	-0.202	0.387
				MCMC+MH	0.994	0.078	-0.008	0.316
40	4	0.0	1.5	MV	0.999	0.082	0.000	0.295
				MAP	0.998	0.063	0.003	0.285
				EAP	0.998	0.064	0.002	0.284
				MCMC+MH	0.998	0.069	0.001	0.280
40	5	-2.0	0.5	MV	0.992	0.146	-0.031	0.396
				MAP	0.990	0.079	0.232	0.395
				EAP	0.990	0.082	0.215	0.386
				MCMC+MH	0.991	0.075	0.003	0.320
40	5	2.0	1.0	MV	0.997	0.118	0.024	0.354
				MAP	0.995	0.072	-0.210	0.379
				EAP	0.994	0.072	-0.202	0.375
				MCMC+MH	0.996	0.083	0.005	0.312
40	5	0.0	1.5	MV	0.998	0.095	-0.005	0.316
				MAP	0.997	0.072	0.001	0.311
				EAP	0.997	0.073	-0.001	0.310
				MCMC+MH	0.998	0.079	-0.002	0.304

Table 3: Anova for the latent trait estimation

Source of variation	SS	df	MS	F	ω^2
Number of items	161.64	2	80.82	1713.28	0.22
Number of categorias	9.50	1	9.50	201.33	0.01
μ_θ	41.37	2	20.68	438.47	0.06
ψ_θ	24.70	2	12.35	261.80	0.03
Other effects	271.88	10	27.19	576.36	0.36
Error	509.08	10792	0.05	-	-
Total corrected	746.28	-	-	-	-

Table 4: Results of the population parameter estimation

Items	Cat	true μ_θ	true ψ_θ	Statistic	est. μ_θ	est. ψ_θ
20	4	-2	1.5	mean	-2.010	1.516
				meanse	0.055	0.110
				var	< 0.001	0.003
20	4	2.0	0.5	mean	2.003	0.506
				meanse	0.034	0.042
				var	< 0.001	0.001
20	4	0.0	1.0	mean	0.000	0.998
				meanse	0.044	0.069
				var	< 0.001	0.001
20	5	-2.0	1.5	mean	-1.996	1.481
				meanse	0.055	0.111
				var	< 0.001	0.003
20	5	2.0	0.5	mean	2.003	0.503
				meanse	0.035	0.043
				var	< 0.001	0.001
20	5	0.0	1.0	mean	0.001	0.987
				meanse	0.044	0.070
				var	< 0.001	0.001
30	4	-2.0	1.0	mean	-1.998	0.992
				meanse	0.044	0.071
				var	< 0.001	0.001
30	4	2.0	1.5	mean	1.987	1.460
				meanse	0.052	0.100
				var	< 0.001	0.002
30	4	0.0	0.5	mean	-0.002	0.496
				meanse	0.032	0.035
				var	< 0.001	0.001
30	5	-2.0	1.0	mean	-2.001	1.026
				meanse	0.045	0.075
				var	< 0.001	0.001
30	5	2.0	1.5	mean	2.004	1.505
				meanse	0.052	0.100
				var	< 0.001	0.001
30	5	0.0	0.5	mean	-0.001	0.502
				meanse	0.033	0.037
				var	< 0.001	0.000

Continuation of Table 4

NI	NC	μ_θ	ψ_θ	Statistic	est. μ_θ	est. ψ_θ
40	4	-2.0	0.5	mean	-2.000	0.502
				meanse	0.032	0.037
				var	< 0.001	< 0.001
40	4	2.0	1.0	mean	1.992	0.980
				meanse	0.043	0.066
				var	< 0.001	0.001
40	4	0.0	1.5	mean	0.001	1.487
				meanse	0.051	0.093
				var	< 0.001	0.001
40	5	-2.0	0.5	mean	-1.997	0.510
				meanse	0.033	0.039
				var	< 0.001	0.001
40	5	2.0	1.0	mean	2.005	1.019
				meanse	0.043	0.068
				var	< 0.001	0.001
40	5	0.0	1.5	mean	-0.002	1.472
				meanse	0.051	0.093
				var	< 0.001	0.001

mean = mean of the estimates, meanse = mean of the standard errors, var = variance of the estimates

Table 5: Spent time (ST), achieved precision (AP) ad number of required iterations (NRI) in the simulation study

Items	Cat	μ_θ	ψ_θ	MV			MAP			MCMC
				PA	NRI	ST	PA	NRI	ST	ST
20	4	-2.0	1.5	0.000009	52.00	6.65	0.000009	41.10	5.30	1539.69
20	4	2.0	0.5	0.000607	97.55	10.43	0.000009	48.15	5.22	1316.58
20	4	0.0	1.0	0.000009	52.00	6.65	0.000009	41.10	5.30	1539.69
20	5	-2.0	1.5	0.000007	27.20	4.43	0.000008	29.40	4.76	1810.29
20	5	2.0	0.5	0.003033	100.00	20.24	0.000010	70.25	14.27	2392.73
20	5	0.0	1.0	0.000008	25.30	3.56	0.000008	31.65	4.38	1653.41
30	4	-2.0	1.0	0.000332	89.20	15.30	0.000008	31.45	5.55	2095.55
30	4	2.0	1.5	0.000122	99.90	29.92	0.000010	76.95	23.11	3511.32
30	4	0.0	0.5	0.000008	23.50	4.87	0.000008	36.65	7.47	2396.07
30	5	-2.0	1.0	0.000008	24.80	6.49	0.000008	30.65	7.73	3028.68
30	5	2.0	1.5	0.000063	74.70	9.01	0.000009	36.95	4.52	1475.86
30	5	0.0	0.5	0.000007	19.90	5.05	0.000008	31.10	7.74	2848.61
40	4	-2.0	0.5	0.000031	71.15	19.39	0.000009	32.10	8.90	3221.85
40	4	2.0	1.0	0.001557	100.00	27.32	0.000009	80.80	22.21	3276.54
40	4	0.0	1.5	0.000025	69.55	22.28	0.000009	54.40	17.29	3708.44
40	5	-2.0	0.5	0.000007	20.60	8.84	0.000008	26.25	11.16	4719.00
40	5	2.0	1.0	0.000009	76.85	28.34	0.000009	59.10	21.91	4282.20
40	5	0.0	1.5	0.000009	49.55	20.81	0.000009	42.30	18.08	4807.37

Table 6: Item parameters used in the simulation studies : 4 alternatives

Item	b	a	Item	b	a	Item	b	a	Item	b	a
1	-4.00	-0.50	11	-0.80	-0.60	21	-4.00	-0.60	31	-0.80	-0.70
1	-3.80	-0.20	11	-0.40	-0.70	21	-3.80	-0.20	31	-0.40	-0.70
1	-3.40	0.10	11	-0.20	0.40	21	-3.40	0.10	31	-0.20	0.40
1	-3.00	0.60	11	0.16	0.90	21	-2.50	0.70	31	0.16	1.00
2	-3.20	-1.20	12	-0.60	-1.70	22	-3.20	-1.10	32	-0.60	-1.60
2	-2.90	-0.40	12	-0.20	-0.30	22	-2.80	-0.40	32	-0.20	-0.30
2	-2.80	0.20	12	0.10	0.80	22	-2.40	0.20	32	0.10	0.80
2	-2.68	1.40	12	0.47	1.20	22	-2.24	1.30	32	0.47	1.10
3	-3.00	-0.60	13	-0.20	-0.50	23	-3.00	-0.70	33	-0.20	-0.60
3	-2.80	-0.70	13	0.10	-0.20	23	-2.80	-0.70	33	0.10	-0.20
3	-2.50	0.40	13	0.30	0.10	23	-2.50	0.40	33	0.30	0.10
3	-2.37	0.90	13	0.79	0.60	23	-1.97	1.00	33	0.79	0.70
4	-2.80	-1.70	14	0.10	-1.20	24	-2.80	-1.60	34	0.10	-1.10
4	-2.60	-0.30	14	0.40	-0.40	24	-2.60	-0.30	34	0.40	-0.40
4	-2.40	0.80	14	0.80	0.20	24	-2.40	0.80	34	0.80	0.20
4	-2.05	1.20	14	1.11	1.40	24	-1.71	1.10	34	1.11	1.30
5	-2.50	-0.50	15	0.40	-0.60	25	-2.50	-0.60	35	0.40	-0.70
5	-2.20	-0.20	15	0.80	-0.70	25	-2.20	-0.20	35	0.80	-0.70
5	-1.90	0.10	15	1.00	0.40	25	-1.90	0.10	35	1.00	0.40
5	-1.74	0.60	15	1.42	0.90	25	-1.44	0.70	35	1.42	1.00
6	-2.00	-1.20	16	0.60	-1.70	26	-2.00	-1.10	36	0.60	-1.60
6	-1.80	-0.40	16	1.00	-0.30	26	-1.80	-0.40	36	1.00	-0.30
6	-1.60	0.20	16	1.20	0.80	26	-1.60	0.20	36	1.20	0.80
6	-1.42	1.40	16	1.74	1.20	26	-1.18	1.30	36	1.74	1.10
7	-1.80	-0.60	17	1.00	-0.50	27	-1.80	-0.70	37	1.00	-0.60
7	-1.60	-0.70	17	1.20	-0.20	27	-1.60	-0.70	37	1.20	-0.20
7	-1.40	0.40	17	1.60	0.10	27	-1.40	0.40	37	1.60	0.10
7	-1.11	0.90	17	2.05	0.60	27	-0.92	1.00	37	2.05	0.70
8	-1.50	-1.70	18	1.20	-1.20	28	-1.50	-1.60	38	1.20	-1.10
8	-1.30	-0.30	18	1.60	-0.40	28	-1.30	-0.30	38	1.60	-0.40
8	-1.00	0.80	18	2.00	0.20	28	-1.00	0.80	38	2.00	0.20
8	-0.79	1.20	18	2.37	1.40	28	-0.66	1.10	38	2.37	1.30
9	-1.20	-0.50	19	1.40	-0.60	29	-1.20	-0.60	39	1.40	-0.70
9	-1.00	-0.20	19	2.00	-0.70	29	-1.00	-0.20	39	2.00	-0.70
9	-0.80	0.10	19	2.20	0.40	29	-0.80	0.10	39	2.20	0.40
9	-0.47	0.60	19	2.68	0.90	29	-0.39	0.70	39	2.68	1.00
10	-1.00	-1.20	20	1.60	-1.70	30	-1.00	-1.10	40	1.60	-1.60
10	-0.80	-0.40	20	2.00	-0.30	30	-0.80	-0.40	40	2.00	-0.30
10	-0.40	0.20	20	2.60	0.80	30	-0.40	0.20	40	2.60	0.80
10	-0.16	1.40	20	3.00	1.20	30	-0.13	1.30	40	3.00	1.10

Table 7: Item parameters used in the simulation studies : 5 alternatives

Item	b	a	Item	b	a	Item	b	a	Item	b	a
1	-4.20	-0.40	11	-1.00	-0.90	21	-4.00	-0.50	31	-1.00	-0.90
1	-3.90	-0.20	11	-0.70	-0.20	21	-3.80	-0.20	31	-0.80	-0.70
1	-3.60	-0.10	11	-0.40	-0.20	21	-3.60	-0.10	31	-0.50	0.20
1	-3.40	0.10	11	-0.20	0.40	21	-3.40	0.10	31	-0.20	0.40
1	-3.00	0.60	11	0.16	0.90	21	-2.50	0.70	31	0.16	1.00
2	-3.40	-0.80	12	-0.60	-1.20	22	-3.20	-1.00	32	-0.60	-1.00
2	-3.00	-0.50	12	-0.20	-0.60	22	-3.00	-0.60	32	-0.20	-0.60
2	-2.70	-0.30	12	0.00	-0.20	22	-2.70	-0.30	32	0.00	-0.30
2	-2.40	0.20	12	0.10	0.80	22	-2.40	0.60	32	0.10	0.80
2	-2.68	1.40	12	0.47	1.20	22	-2.24	1.30	32	0.47	1.10
3	-3.20	-0.90	13	-0.40	-0.40	23	-3.20	-0.90	33	-0.40	-0.50
3	-3.00	-0.20	13	-0.20	-0.20	23	-3.00	-0.70	33	-0.20	-0.20
3	-2.80	-0.20	13	0.00	-0.10	23	-2.80	0.20	33	0.00	-0.10
3	-2.50	0.40	13	0.30	0.10	23	-2.50	0.40	33	0.30	0.10
3	-2.37	0.90	13	0.79	0.60	23	-1.97	1.00	33	0.79	0.70
4	-3.00	-1.20	14	0.00	-0.80	24	-3.00	-1.00	34	0.10	-1.00
4	-2.77	-0.60	14	0.10	-0.50	24	-2.80	-0.60	34	0.40	-0.60
4	-2.50	-0.20	14	0.40	-0.30	24	-2.60	-0.30	34	0.60	-0.30
4	-2.40	0.80	14	0.80	0.20	24	-2.40	0.80	34	0.80	0.60
4	-2.05	1.20	14	1.11	1.40	24	-1.71	1.10	34	1.11	1.30
5	-2.50	-0.40	15	0.40	-0.90	25	-2.80	-0.50	35	0.10	-0.90
5	-2.20	-0.20	15	0.60	-0.20	25	-2.30	-0.20	35	0.30	-0.70
5	-2.00	-0.10	15	0.80	-0.20	25	-2.10	-0.10	35	0.60	0.20
5	-1.90	0.10	15	1.00	0.40	25	-1.90	0.10	35	1.00	0.40
5	-1.74	0.60	15	1.42	0.90	25	-1.44	0.70	35	1.42	1.00
6	-2.20	-0.80	16	0.40	-1.20	26	-2.30	-1.00	36	0.40	-1.00
6	-1.90	-0.50	16	0.70	-0.60	26	-2.00	-0.60	36	0.90	-0.60
6	-1.80	-0.30	16	1.00	-0.20	26	-1.80	-0.30	36	1.00	-0.30
6	-1.60	0.20	16	1.20	0.80	26	-1.60	0.60	36	1.20	0.80
6	-1.42	1.40	16	1.74	1.20	26	-1.18	1.30	36	1.74	1.10
7	-2.00	-0.90	17	0.90	-0.40	27	-2.20	-0.90	37	0.90	-0.50
7	-1.80	-0.20	17	1.10	-0.20	27	-2.00	-0.70	37	1.10	-0.20
7	-1.60	-0.20	17	1.40	-0.10	27	-1.80	0.20	37	1.30	-0.10
7	-1.40	0.40	17	1.60	0.10	27	-1.40	0.40	37	1.60	0.10
7	-1.11	0.90	17	2.05	0.60	27	-0.92	1.00	37	2.05	0.70
8	-1.50	-1.20	18	1.20	-0.80	28	-1.80	-1.00	38	1.00	-1.00
8	-1.30	-0.60	18	1.60	-0.50	28	-1.50	-0.60	38	1.20	-0.60
8	-1.10	-0.20	18	1.80	-0.30	28	-1.30	-0.30	38	1.40	-0.30
8	-1.00	0.80	18	2.00	0.20	28	-1.00	0.80	38	2.00	0.60
8	-0.79	1.20	18	2.37	1.40	28	-0.66	1.10	38	2.37	1.30
9	-1.50	-0.40	19	1.40	-0.90	29	-1.40	-0.50	39	1.40	-0.90
9	-1.30	-0.20	19	1.80	-0.20	29	-1.10	-0.20	39	1.60	-0.70
9	-1.10	-0.10	19	2.00	-0.20	29	-1.00	-0.10	39	2.00	0.20
9	-0.80	0.10	19	2.20	0.40	29	-0.80	0.10	39	2.20	0.40
9	-0.47	0.60	19	2.68	0.90	29	-0.39	0.70	39	2.68	1.00
10	-1.00	-0.80	20	1.60	-1.20	30	-1.00	-1.00	40	1.60	-1.00
10	-0.80	-0.50	20	2.00	-0.60	30	-1.20	-0.60	40	2.00	-0.60
10	-0.60	-0.30	20	2.30	-0.20	30	-0.90	-0.30	40	2.20	-0.30
10	-0.40	0.20	20	2.60	0.80	30	-0.40	0.60	40	2.60	0.80
10	-0.16	1.40	20	3.00	1.20	30	-0.13	1.30	40	3.00	1.10

4 Concluding remarks

The MCMC+MH procedure that was proposed recoveries properly both latent traits and population parameters. Its computational implementation is straightforward and it can be extended to other situations such as, when the item parameters are unknown and for a different latent trait distributions. It is clear that estimating the population parameters provides a better latent traits estimates. Furthermore, we notice that the number of items accounts for the highest percentage in the variability of the accuracy of the latent traits estimation. The other factors, number of categories and the values of the item parameters, account for a small percentage of these variability.

Acknowledgements

This research was supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) from Brazil, through a Master's program scholarship. It is part of the Master's Dissertation written by the first author under the guidance of the second.

References

D. F. Andrade and H. R. Tavares. Item response theory for longitudinal data: population parameter estimation. *Journal of multivariate analysis*, 95(1):1–22, 2005. 2

R. J. De Ayala. A comparison of the nominal response model and three-parameter logistic model in computerized adaptive testing. *Educational and psychological measurement*, 49:789–805, 1989. 2

R. J. De Ayala. The nominal response model in computerized adaptive testing'. *Applied psychological measurement*, 16(4):327–343, 1992. 2

R. J. De Ayala and M. Sava-Bolesta. Item parameter recovery for the nominal response model. *Applied psychological measurement*, 23(1):3–19, 1999. 2, 5, 8

C. L. N. Azevedo. *Métodos de estimação na Teoria da Resposta ao Item*. Dissertação de Mestrado, IME-USP, 2003. 2

F. B. Baker and Seock-Ho Kim. *Item response theory : Parameter Estimation Techniques*. Marcel Dekker, Inc, New York, NY, second edition, 2004. 2, 5

D. R. Bock. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1):29–51, 1972. 1, 2

D. M. Bolt, A. S. Cohen, and J. A. Wollack. A mixture item response model for multiple choice data. *Journal of Educational and Behavorial Statistics*, 26:381–409, 2001. 2

Ming-Hui Chen, Qi-Man Shao, and J. G. Ibrahim. *Monte Carlo Methods in Bayesian Computation.* Spring-Verlag, New York, first edition, 2000. 3

C. E. DeMars. Sample size and the recovery of nominal response model item parameters. *Applied Psychological Measurement*, 27(4):275–288, 2003. 2, 5, 8

D. Gamerman and H. F. Lopes. *Markov Chain Monte Carlo : Stochastic simulation for bayesian inference.* Chapman & Hall/CRC, London, second edition, 2006. 6

A. Gelman. Prior distribution for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533, 2006. 4

M. R. Harwell, F. B. Baker, and M. Zwarts. Item parameter estimation via marginal maximum likelihood and an em algorithm: A didactic. *Journal of Educational Statistics*, 13(3):243–271, 1988. 5

J. R. Patz and B. W. Junker. A straightforward approach to markov chain monte carlo methods for item response models. *Journal of Educational and Behavorial Statistics*, 24(2):146–178, 1999. 3, 4

J. A. Wollack, D. M. Bolt, A. S. Cohen, and Young-Sun Lee. Recovery of item parameters in the nominal response model: A comparison of marginal maximum likelihood estimation and markov chain monte carlo estimation. *Applied Psychological Measurement*, 26(3):339–352, 2002. 2, 3, 5

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2008-01 - CABRAL, C.R.B.; BOLFARINE, H. Bayesian Inference for the Skew Student-t-Normal Model. 2008. 13p.(RT-MAE-2007-01)

2008-02 - CABRAL, C.R.B.; BOLFARINE, H. Bayesian Density Estimation Using Skew Student-t-Normal Mixtures. 2008. 18p.(RT-MAE-2008-02)

2008-03 - FARIAS, R.B.A.; ARENAS, G.M.; PATRIOTA, A.G. Redução de modelos na presença de parâmetros de perturbação. 2008. 19p. (RT-MAE-2008-03)

2008-04 - MIRANDA, J.C. A Mixed Hammerstein Integral Equation. 2008. 10p. (RT-MAE-2008-04)

2008-05 - WECHSLER, S.; PEREIRA, C.A.B.; MARQUES, P.C.F. Birnbaum's Theorem Redux. 05p. 2008. (RT-MAE-2008-05)

2008-06 - BUENO, V.C.; ALVAREZ, N.G.G. Optimal warranty time under a externsion of the general failure model. 23p. 2008. (RT-MAE-2008-06)

2008-07 - FIGUEIREDO, C.C.; BOLFARINE, H.; SANDOVAL, M.C.; C.R.O.P. LIMA On the skew-normal calibration model. 30p. 2008. (RT-MAE-2008-07)

2008-08 - NELFI, A.G.G.; Tempo ótimo de burn-in usando uma regra de parada monótona. 39p. 2008. (RT-MAE-2008-08)

2008-09 - NELFI, A.G.G.; Tempo ótimo de garantia usando uma regra de parada monótona. 17p. 2008. (RT-MAE-2008-09)

2008-10 - BONASSI, F.V.; STERN, R.B.; WECHSLER, S. The Gambler's Fallacy: A Bayesian Approach. 08p. 2008. (RT-MAE-2008-10)

2008-11 - BUENO, V.C. Characterizing a hot standby redundancy through compensator transform. 11p. 2008. (RT-MAE-2008-11)

2008-12 - DINIZ, I.C., MIRANDA, J.C.S. Poissonian thee Constructed from Independent Poisson Point. 16p. 2008. (RT-MAE-2008-12)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05314-970 - São Paulo, Brasil