

**SENSITIVITY ANALYSIS OF PRIOR SPECIFICATION FOR THE
PROBIT-NORMAL IRT MODEL: AN EMPIRICAL STUDY ***

JORGE L. BAZAN

Department of Sciences, Pontifical Catholic University of Peru.
jlbazan@pucp.edu.pe

HELENO BOLFARINE

Department of Statistics, University of São Paulo.
hbolfar@ine.usp.br

ROSELI APARECIDA LEANDRO

Department of Exact Sciences, University of São Paulo - ESALQ.
rleandr@carpa.ciagri.usp.br

ABSTRACT

Item response theory (IRT) models a set of dichotomous multivariate responses corresponding to I items of a test applied to n subjects. IRT is widely used in several evaluation systems considering frequentist methodology. In this paper we consider the normal ogive or probit-normal model and present the Bayesian estimation procedures considering MCMC methodology for simulation from the posterior distribution of the latent variables. We illustrate the interpretation of model parameters considering an application to fourteen items of a mathematical test for sixth grade students (small sample) using the WinBUGS package. A sensitivity analysis for the prior distribution is considered. The results indicate that the probit-normal model is insensitive to the prior specifications for the difficulty and discrimination parameters in the literature. The priors considered lead to similar posterior distributions and fit.

Key words

Probit-normal model, item response theory, Bayesian estimation, sensitivity analysis.

*Correspondence should be sent to Jorge Luis Bazán, Av. Universitaria Cuadra 18 San Miguel Lima 32-Per. Departamento de Ciencias. Pontificia Universidad Católica del Perú

1. INTRODUCCION

Item response theory (IRT) considers models for describing how the probability of correctly answering an item of a test depends on subjects ability or latent proficiency and also parameters related to the questions being answered. Several references in recent statistical literature can be associated to such models. Specially we mention Baker and Kim (2004), van der Linden and Hambleton (1997) and in Portuguese, the excellent summary by Andrade et al. (2000). Several characterizations for such models have been developed in the last 40 years, specially with respect to the status of the latent variables (Borsboom et al. 2003) - latent parameters in other contexts (for example Baker and Kim, 2004) -, and with respect to classical or Bayesian approaches. In this paper, we consider the characterization due to Holland and Rosenbaum (1986) and Bartholomew and Knoot (1999) and the Bayesian approach in Albert (1992).

Bayesian estimation is not new in IRT models and recently can be distinguished in estimation with and without Markov chain Monte Carlo (MCMC). MCMC simulation methods (Chen et al. 2000, Patz and Junker, 1999) is a method of simulating random samples from any theoretical multivariate distribution - in particular, from the multivariate posterior distribution that is the focus of Bayesian inference - so that features of the theoretical distributions can be estimated by corresponding features of the random sample.

The most widely used IRT models are the dichotomous item response models. Two models, logistic-normal model and probit-normal model, described in the following section, are part of these models. Educational evaluation studies conducted in Brazil and others Latinoamerican countries typically is based on the logit-normal model. This may be due to the fact that this model is implemented in commercial packages and use of the estimation procedures is more complex in probit-normal models and this may explain the very limited use of this model within the classical approach. On the other hand, the use of strict Bayesian approaches in logit-normal model is somewhat more complex (see Rupp et al., 2004). In Brazil, Assuno (1999) implemented the approach by Patz and Junker for MCMC estimation of logistic-normal model. With the use of MCMC techniques, the probit-normal model became more popular and several extensions were proposed in the literature including: Three parameter models (Sahu, 2002, Glas and Meijer, 2003), Multidimensional models (Bguin and Glas, 2001; Jackman, 2001; Linardakis and Dellaportas, 2002), Multilevel models (Fox and Glas, 2001), Testlet (Wang, et al., 2003), Item confirmatory factor analysis model (Segall, 2003) and Measurement error model (Fox and Glas, 2003).

An MCMC implementation for probit-normal model using directly the likelihood

function is implemented in Spiegelhalter et al. (1996) via *adaptive rejection sampling (ARS) scheme* (Gilks and Wild, 1992) or in Spiegelhalter et al. (2003) via the *slice sampling* (Neal, 2003). Another MCMC implementation, introduced in Albert (1992), is based on the Gibbs sampling (GS) scheme since a *data augmented likelihood*, named *data augmented Gibbs sampling (DAGS) scheme* using auxiliary latent variables. Although Bayesian estimation for the probit-normal model has been developed in the literature, very few papers discuss in detail the problem of prior specification for the item parameters. Some results in this direction are found in Albert and Ghosh (2000) and Ghosh et al. (2000) where it is discussed theoretically the consequences of using precise prior distributions for the item parameters. However, it seems not to exist major empirical studies about the effect of different priors specification in the estimation of the parameters for the probit-normal model. In this paper, we present a review of the two main procedures for Bayes estimation via MCMC of the probit-normal model described above. By considering DAGS scheme implemented in WinBUGs, an empiric sensitivity study of the model with respect to prior specification for the item parameters is conducted to a data set with small number of items and examinees.

The paper is organized as follows. In Section 2, data description related to a particular data set with small samples and small number of items is conducted. In Section 3, where we discuss methodology, the dichotomous item response models is presented and a literature revision of their characteristics and estimation approaches is conducted. Also, in Section 3, a Bayesian estimation via MCMC for probit-normal model using the ARS and DAGS schemes are presented. Prior specification is also discussed. Section 4, termed Results, deals with an application to a real data set coming from a mathematical test and sensitivity analysis for different priors in item parameters is conducted. In Section 5, a final discussion is presented.

2. DATA DESCRIPTION

We illustrate the Bayesian approaches to probit-normal IRT models using a data set obtained when applying 14 free items of the Peruvian Primary School Mathematical Test to sixth grade students (UMC, 2001). The UMC (Unidad de Medicin de la Calidad Educativa) in Per has developed language and mathematics test to assess, regularly, the educational progress in several grades. The Math Test was developed by the UMC to the national evaluation in 1998 and corresponds to multiple choice items with four alternatives and, additionally, the Math Test is formed with independent items corresponding to different tasks with different definitions. Given the latent ability U , it is considered that the correct responses to the items are independent. Furthermore, the autocorrelations within individual responses seem to be low, which provides additional support for the assumption

of local independence.

In the empirical study considered in this paper, the 14 free items of the Math Test were applied to 131 students of high social-economical status. Item response data are available from the authors upon request. The summary statistical scores for Mathematics Test are presented in Table 1.

TABLE 1

Summary statistical scores of 14 items of the Mathematics Test applied to 131 students from Peruvian schools

Statistic	value	Statistic	value
Mean	10.840	Median	11
Variance	3.432	Std. Dev.	1.853
Skew	-0.795	Kurtosis	0.449
Minimum	5	Maximum	14
Alpha	0.481	Mean P	0.774

The scores present a mean of 11 points and a standard deviation of almost 2 points. Descriptive statistics, indicate that the scores present negative asymmetry with the presence of high scores. The test presents a medium reliability given by the alpha of Cronbach of 0.48, and present a mean proportion of item of 0.774, indicating an easy Test.

3. METHODS

3.1. DICHOTOMOUS ITEM RESPONSE MODEL

We consider that

$$Y_{ij}|u_i, \eta_j \sim \text{Bernoulli}(p_{ij}), \quad (1)$$

where Y_{ij} are the dichotomous response corresponding to subject i on item j , $\eta_j = (a_j, b_j)'$, are the item parameters, where a_j corresponds to item discrimination and b_j corresponds to item difficulty and u_i is the value corresponding to latent variable U_i associated to subjects i , describing its ability in answering the test with I items and p_{ij} is the probability of correct answer for subject i in test j , $i = 1, \dots, n$, $j = 1, \dots, I$. Further, let

$$p_{ij} = P(Y_{ij} = 1 | u_i, \eta_j) = F_{ij}(m_{ij}), \quad (2)$$

with $m_{ij} = a_j u_i - b_j$, a linear function of u_i , $i = 1, \dots, n$, $j = 1, \dots, I$. The link function $F_{ij}(\cdot)$ is typically known as the *item response function* or *item characteristic*

curve and satisfies the property of latent monotonicity (strictly nondecreasing function of U_i). The function F_{ij} typically is the same for all i and j and the most used ones are the cumulative distributions of the normal and the logistic distributions. Further, negative values of a_j are not expected and b_j and U_i take real values.

IRT models considered typically satisfies the conditional independence property, that is, for subject i , the responses Y_{ij} corresponding to items $j = 1, \dots, I$, are conditionally independent given the values of latent variables U_i , $i = 1, \dots, n$. Further, it is considered independence between responses from different subjects. Under the above assumptions, the joint distributions of $\mathbf{Y} = (\mathbf{Y}'_1, \dots, \mathbf{Y}'_n)'$ with $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iI})'$ given the vector of latent variables $\mathbf{U} = (U_1, \dots, U_n)'$ and the vector of item parameters $\boldsymbol{\eta} = (\eta_1, \dots, \eta_I)'$ can be written as

$$p(\mathbf{Y} = \mathbf{y} | \mathbf{U} = \mathbf{u}, \boldsymbol{\eta}) = \prod_{i=1}^n \prod_{j=1}^I F_{ij}(m_{ij})^{y_{ij}} (1 - F_{ij}(m_{ij}))^{1-y_{ij}} \quad (3)$$

The first IRT model was formally introduced by Lord (1952) and considers $F_{ij}(\cdot) = \Phi(\cdot)$, $i = 1, \dots, n$, and $j = 1, \dots, I$, with $\Phi(\cdot)$ the cumulative function of the standard normal distribution. The model is known as the normal ogive model. Birnbaum (1968) considered $F_{ij}(\cdot) = L_{ij}(\cdot)$ with $L_{ij}(m_{ij}) = e^{m_{ij}} / (1 + e^{m_{ij}})$, denoted the cumulative function of the logistic distribution. This model is known as the logistic model with two parameters. As a special case of this model ($a_j = 1$) we have the Rash model (Fisher and Molenaar, 1995).

3.1.1. CHARACTERISTICS

The dichotomous item response model presented in (1), (2) and (3) involves a total of $n + 2I$ unknown parameters being thus overparameterized. On the other hand, for a fixed number of items, item parameters are known as structural parameters and the latent variables are known as incidental parameters, because they increase with n , the sample size and because the analysis is generally focused on the item parameters. The model is also unidentifiable, since it is preserved under a special class of transformations of the parameters (see Albert, 1992) so that maximum likelihood estimates may not be unique. One way of contouring such difficulties is to impose restrictions on the item parameters as considered, for example, in Bock and Aitkin (1981). Another way follows by specifying a distribution for the latent variables. Lord and Novik (1968, Chap. 16, Albert, 1992) consider

$$U_i \stackrel{iid}{\sim} N(\mu, \sigma^2), \quad i = 1, \dots, n. \quad (4)$$

This assumption establishes that it is believed that the latent variables are well behaved and that are a random sample from this distribution but as in Tsutakawa (1984), we consider in this paper that $\mu = 0$ and $\sigma^2 = 1$.

The IRT model that follows by considering (1)-(4) is denominated by Albert (1992) and Lord and Novick (1968) as the normal ogive model. In Bartholomew and Knoot (1999), the model with normal link and normal latent variable is denoted probit-probit and the model with logistic link and normal latent variable as logit-probit model. We find it more appropriate using the notation probit-normal and logit-normal models, respectively.

3.1.2. ESTIMATION

The estimation problem in IRT models can be grouped into three categories: frequentist (or classical) estimation, Bayesian without MCMC and Bayesian with MCMC estimation.

Classical estimation has been dominated by the likelihood approach. Using the former methodology, several approaches have been proposed such as joint likelihood, marginal likelihood and conditional likelihood (Andrade et al., 2000, Baker and Kim, 2004). However, the maximum likelihood estimators for the structural parameters are not consistent in the presence of incidental parameters (Neyman and Scott, 1948). The most used approach is the marginal likelihood approach using the pseudo-EM-algorithm with Gaussian quadrature for approximating the integrals needed for implementing the E step of the algorithm involved in the estimation of the item parameters (Bock and Aitkin, 1981). This estimation procedure is implemented in ad-hoc software, as, for example, the program BILOG (Mislevy and Bock, 1990), that uses “heuristic” restriction of the model (Du Toit, 2003). The estimation of ability parameters is performed in a second stage with item parameters replaced by estimates computed previously. Limitations of this methodology are discussed in Patz and Junker (1999) and Sahu (2002). To just name a few, we point out the impossibility of estimating U_i for extreme scores, limitations on the statement of regularity conditions that justify the asymptotic theory under which inference for model parameters and model checking are based (Froelich, 2000). In practical situations this estimation procedure defines the evaluation paradigm as is the case with the BILOG software which incorporate their restrictions of computation as criterion in evaluation area. Specifically, the fact that estimation is done in two step implies that evaluation is also done in two steps: *the algorithm defines the evaluation scheme and not the evaluation scheme defines the algorithm*. Although separate estimation of part of the parameters considering known the remaining parameters is possible, inference based on asymptotic arguments, specially for non *i.i.d* observations, may results on non consistent estimation (Patz and Junker, 1999).

In the case of non MCMC Bayesian estimation, Bayesian marginal estimation is used with maximum and expected a posteriori estimates for latent variables

considering hierarchical models or not (see Baker and Kim, 2004; Kim et al. 1994). One characteristic of classical (non MCMC Bayesian) approaches is that they are typically based on the logistic link function. A review of classical and non MCMC Bayesian approaches can be found in Baker and Kim (2004) and Andrade et al. (2000). A review of Bayesian approaches using MCMC can be found in Patz and Junker (1999). The approach for the probit-normal model is described in the following section.

3.2. BAYESIAN INFERENCE VIA MCMC FOR THE PROBIT-NORMAL MODEL

Let $D_{obs} = \mathbf{Y}$, the observed data. Hence, the likelihood function for the probit-normal model is given by

$$L(\mathbf{u}, \boldsymbol{\eta} | D_{obs}) = \prod_{i=1}^n \prod_{j=1}^I \Phi(m_{ij})^{y_{ij}} (1 - \Phi(m_{ij}))^{1-y_{ij}},$$

where $\Phi(\cdot)$ is the cumulative function of the standard normal distribution.

3.2.1. PRIOR SPECIFICATION

We consider the following general class of prior distributions:

$$\pi(\mathbf{u}, \boldsymbol{\eta}) = \prod_{i=1}^n g_{1i}(u_i) \prod_{j=1}^I g_{2j}(\eta_j)$$

where $g_{1i}(u_i) = \phi(u_i; 0, 1)$, $i = 1, \dots, n$, are a standard normal prior distribution for the latent variables u_i , and $g_{2j}(\eta_j) = g_{21j}(a_j)g_{22j}(b_j)$, $j = 1, \dots, I$, are priors for item parameter, in which $g_{21j}(\cdot)$ should be proper to guarantee proper posterior distributions for $(\mathbf{u}, \boldsymbol{\eta})$ (see Albert and Ghosh, 2000, Ghosh et al. 2000).

Following proposals typically considered (see Rupp et al. 2004), we take the following priors for item parameters a_j and b_j , $g_{21j}(a_j) = \phi(a_j; \mu_a, s_a^2)$, $j = 1, \dots, I$ and $g_{22j}(b_j) = \phi(b_j; 0, s_b^2)$, $j = 1, \dots, I$ normal distribution, so that $g_{2j}(\eta_j) = \phi_2(\eta_j; \mu_\eta, \Sigma_\eta)$, $j = 1, \dots, I$, is a bivariate normal distribution with mean vector $\mu_\eta = (\mu_a, 0)'$ and covariance matrix $\Sigma_\eta = \begin{pmatrix} s_a^2 & 0 \\ 0 & s_b^2 \end{pmatrix}$.

Albert and Ghosh (2000) mention that the choice of a proper prior distribution on the latent trait resolves particular identification problems, and, further, informative prior distributions placed for a_j and b_j can be used to reflect the prior belief that the values of the item parameters are not extreme (in the frontier of the parametric space). In the common situation where little prior information is available about the difficulty parameters, one can chose s_b^2 to be a large value.

This choice will have a modest effect on the posterior distribution for non extreme data, and will result in a proper posterior distribution when extreme data (where students are observed to get correct or incorrect answers to every item) is observed (Albert and Ghosh, 2000), also, Sahu (2002) states that larger values of the variance led to unstable estimates. This priors is denominated as vague priors in Roberts (2001). In Table 2 it is shown some priors considered in the literature for the item parameter $\eta_j = (a_j, b_j)$ in the probit-normal model. We observed that $N(0,1)I(0,)$ is the notation for a normal distribution with mean 0 and variance 1 truncated for negative values.

TABLE 2

Prior specification for item parameters $\eta_j = (a_j, b_j)$ in the probit-normal model in the literature

Prior	Author	a_j prior	b_j prior
A	Johnson and Albert (1999)	$N(2,1)$	$N(0,1)$
B	Congdon (2001)	$N(1,1)$	$N(0,1)$
C	Albert and Ghosh (2000)	$N(0,1)$	$N(0,1)$
D	Sahu (2002); Albert and Ghosh (2000)	$N(0,1)$	$N(0,10000)$
E	Spiegelhalter et al. (1996)	$N(0,1)I(0,)$	$N(0,10000)$
F	Sahu (2002); Patz and Junker (1999)	$N(1,0.5)I(0,)$	$N(0,2)$

Priors *A*, *B* and *C* are precise and priors *D* and *E* are vague priors assigned to the difficulty parameters. Priors *E* and *F* assigned to the discrimination parameters are truncated priors.

3.2.2. MCMC USING ADAPTATIVE REJECTION SAMPLING (ARS) SCHEME

By considering the likelihood function and prior specification, a joint posterior distribution is given by:

$$f(\mathbf{u}, \boldsymbol{\eta} | D_{obs}) \propto \prod_{i=1}^n \prod_{j=1}^I \Phi(m_{ij})^{y_{ij}} (1 - \Phi(m_{ij}))^{1-y_{ij}} \prod_{j=1}^I \phi(u_i; 0, 1) \prod_{i=1}^n \phi_2(\eta_j; \mu_\eta, \Sigma_\eta)$$

or

$$f(\mathbf{u}, \boldsymbol{\eta} | D_{obs}) \propto \prod_{i=1}^n \prod_{j=1}^I \Phi(m_{ij})^{y_{ij}} (1 - \Phi(m_{ij}))^{1-y_{ij}} \times \\ \exp \left[-\frac{1}{2} \left(\sum_{i=1}^n u_i^2 + \frac{1}{s_b^2} \sum_{j=1}^I b_j^2 + \frac{1}{s_a^2} \sum_{j=1}^I (a_j - \mu_a)^2 \right) \right]$$

In this way, the probit-normal model can be fitted using MCMC, for example in WinBUGS (see LSAT example in Spiegelhalter et al., 1996). As the joint posterior distributions above are complex to be dealt with, note that all full conditional distributions are non-standard. Hence straightforward implementation of the Gibbs sampler using standard sampling distributions is not possible. However, all the full conditional distributions for the probit-normal model are log-concave (log of the density is concave) according to (Sahu, 2002). Exact sampling from one dimensional log-concave distributions can be performed using rejection sampling, even when the normalizing constants are unknown (Gilks and Wild, 1992). These authors also develop an adaptative rejection sampling (ARS) scheme. ARS dynamically constructs two envelopes (one lower and one upper) for the distribution to be sampled from using successive evaluations of the density at the rejected points. The algorithm stops when one proposed point has been accepted. This procedure is default in WinBUGS 1.4 for the probit-normal model. In the case of restricted range, WinBUGS 1.4 also can use the *slice sampling* (Neal, 2003), which is better to have not crashed when you try to run with the code in the appendix.

Markov chain sampling methods that adapt to characteristics of the distributions being sampled can be constructed using the principle that one can sample from a distribution by sampling uniformly from the region under the plot of its density function. A Markov chain that converge to this uniform distribution can be constructed by alternating uniform sampling in the vertical direction with uniform sampling from the horizontal “slice” defined by the current vertical position or, more generally, with some update that leaves the uniform distribution over this slice invariant. Since this implementation is broadly used in literature, in the paper we described the DAGS scheme with more generality in following section.

3.2.3. MCMC USING DATA AUGMENTED GIBBS SAMPLING (DAGS) SCHEME

An alternative way of writing the probit-normal model with I items and n subjects taking the test follows by considering that

$$Z_{ij} = m_{ij} + e_{ij}, \quad (5)$$

$$e_{ij} \sim N(0, 1), \quad (6)$$

$$y_{ij} = \begin{cases} 1, & Z_{ij} > 0; \\ 0, & Z_{ij} \leq 0. \end{cases} \quad (7)$$

It can be shown after some algebraic manipulations that this expressions recover the model formulated above. It follows basically by noticing that $p_{ij} = P[Y_{ij} = 1] = P[Z_{ij} > 0] = \Phi(m_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, I$, which shows that the linear

normal structure for the auxiliary latent variable yields a model equivalent to the probit-normal model.

Using the new formulation, considering $D = (\mathbf{Z}, \mathbf{y})$ as the complete data with $\mathbf{Z} = (Z_{11}, \dots, Z_{nI})'$, the ‘‘complete data’’, it follows that the likelihood function of the augmented data is given by

$$L(\mathbf{u}, \boldsymbol{\eta} | D) = \prod_{i=1}^n \prod_{j=1}^I \phi(Z_{ij}; m_{ij}, 1) I(Z_{ij}, y_{ij}),$$

in which $I(Z_{ij}, Y_{ij}) = I(Z_{ij} > 0)I(y_{ij} = 1) + I(Z_{ij} \leq 0)I(y_{ij} = 0)$, $i = 1 \dots, n$, $j = 1, \dots, I$, is an indicator variable taking the value one if its argument is true, and the value zero otherwise.

Thus, the complete joint posterior distribution is given by

$$f(\mathbf{u}, \boldsymbol{\eta} | D) \propto \exp \left[-\frac{1}{2} \left(\sum_{i=1}^n \sum_{j=1}^I (Z_{ij} - m_{ij})^2 + \sum_{i=1}^n u_i^2 + \frac{1}{s_b} \sum_{j=1}^I b_j^2 + \frac{1}{s_a} \sum_{j=1}^I a_j^2 \right) \right] I(Z_{ij}, y_{ij})$$

so that the conditional complete distributions for the probit-normal model are given by:

- $\pi(Z_{ij} | u_i, \boldsymbol{\eta}_j, D_{obs}) \propto \phi(Z_{ij}; m_{ij}, 1) I(Z_{ij}, y_{ij})$, $i = 1 \dots, n$, $j = 1, \dots, I$.
- $\pi(u_i | \mathbf{Z}_i, \boldsymbol{\eta}, D_{obs}) \propto \phi(u_i; m_{u_i}, v_{u_i})$, $i = 1 \dots, n$
with $m_{u_i} = \frac{\sum_{j=1}^I a_j (Z_{ij} + b_j)}{\sum_{j=1}^I a_j^2 + 1}$, $v_{u_i} = \frac{1}{\sum_{j=1}^I a_j^2 + 1}$, $i = 1 \dots, n$.
- $\pi(\boldsymbol{\eta}_j | \mathbf{u}, \mathbf{Z}_j, D_{obs}) \propto \phi_2(\boldsymbol{\eta}_j; \mathbf{m}_{\boldsymbol{\eta}_j}, \mathbf{v}_{\boldsymbol{\eta}_j})$, $j = 1, \dots, I$,
with $\mathbf{m}_{\boldsymbol{\eta}_j} = [\mathbf{W}'\mathbf{W} + \Sigma_\eta^{-1}]^{-1} [\mathbf{W}'\mathbf{Z}_j + \Sigma_\eta^{-1} \boldsymbol{\mu}_\eta]$, $\mathbf{v}_{\boldsymbol{\eta}_j} = [\mathbf{W}'\mathbf{W} + \Sigma_\eta^{-1}]^{-1}$, with
 $\boldsymbol{\mu}_\eta = \begin{pmatrix} \mu_a \\ 0 \end{pmatrix}$, $\Sigma_\eta = \begin{bmatrix} S_a^2 & 0 \\ 0 & S_b^2 \end{bmatrix}$ e $\mathbf{W} = (\mathbf{u}, -1)$ with $\mathbf{W}'_i = (u_i, -1)$, $i = 1 \dots, n$.

With the above conditional distributions, the DAGS can be easily implemented. Routines in R (see the MCMC package by Martin and Quinn, 2003), Matlab (see Johnson and Albert, 1999) are available in the Web. In WinBUGS the implementation of the above procedure is not direct since it requires a correct specification for the indicator variables (See Appendix).

3.2.4. MODEL COMPARISON AND SENSITIVITY ANALYSIS

Model checking, or assessing the fit of a model, is a crucial part of any statistical analysis. Before drawing any conclusions from the application of a statistical

model to a data set, an investigator should assess the fit of the model to make sure that the model adequately explains the important features of the data set (Stern and Sinharay, 2005).

In Bayesian inference, a researcher can check the fit of the model in one of three ways (Gelman et al., 1996): (1) checking that the posterior inferences are reasonable, given the substantive context of the model; (2) examining the sensitivity of inferences to reasonable changes in the prior distribution and the likelihood; and (3) checking that the model can explain the data, or in other words that the model is capable of generating data like the observed data. In this paper we made a sensitivity analysis of inferences to different specifications in the prior distribution of the item parameter in the probit-normal model.

In sensitivity analysis, several probability models are fit to the same data set. To compare the models corresponding to the six different prior distributions, we computed the posterior expected deviance ($Dbar$), the deviance of the posterior means ($Dhat$), the effective number of parameters ρ_D and deviance information criterion (DIC), as presented by Spiegelhalter et al. (2002). $Dbar$, is the posterior mean of the deviance that is defined as $-2 * \log(\text{likelihood})$. $Dhat$ is a point estimate of the deviance obtained by substituting in the posterior means, estimates of parameters in the model. ρ_D is given by $\rho_D = Dbar - Dhat$. DIC is given by $DIC = Dbar + pD = Dhat + 2 * pD$. The model with the smallest DIC is estimated to be the model that would best predict a replicate dataset of the same structure as that currently observed.

They claim that the DIC as implemented in the WinBUGS software can be used to compare complex models and large differences in the criterion can be attributed to real predictive differences in the models, although the approach has received critics from several authors (see discussion in Spiegelhalter et al. 2002). In hierarchical modeling with auxiliary latent variables as in the probit-normal model, the “likelihood” or “model complexity” is not unique so that the deviance (or DIC and consequently ρ_D) of a model with latent variables is not unique and can be calculated in several ways (Delorio and Roberts, 2002). With auxiliary latent variables, WinBUGS uses a complete likelihood of the observed variable and the auxiliary latent variable introduced (as fixed effects and random effects in hierarchical modeling) to obtain posterior distributions for the parameters of interest. When this is the case, marginal DIC s for the observed variable (fixed effect) and auxiliary latent variable (random effect) are presented. For a proper comparison of the proposed models, we consider marginal DIC for the observed variables because the focus of the analysis is in $p(\mathbf{y}|\mathbf{u}, \boldsymbol{\eta})$ and although auxiliary random variables are introduced (in two steps) they are not the focus of the

analysis.

4. RESULTS

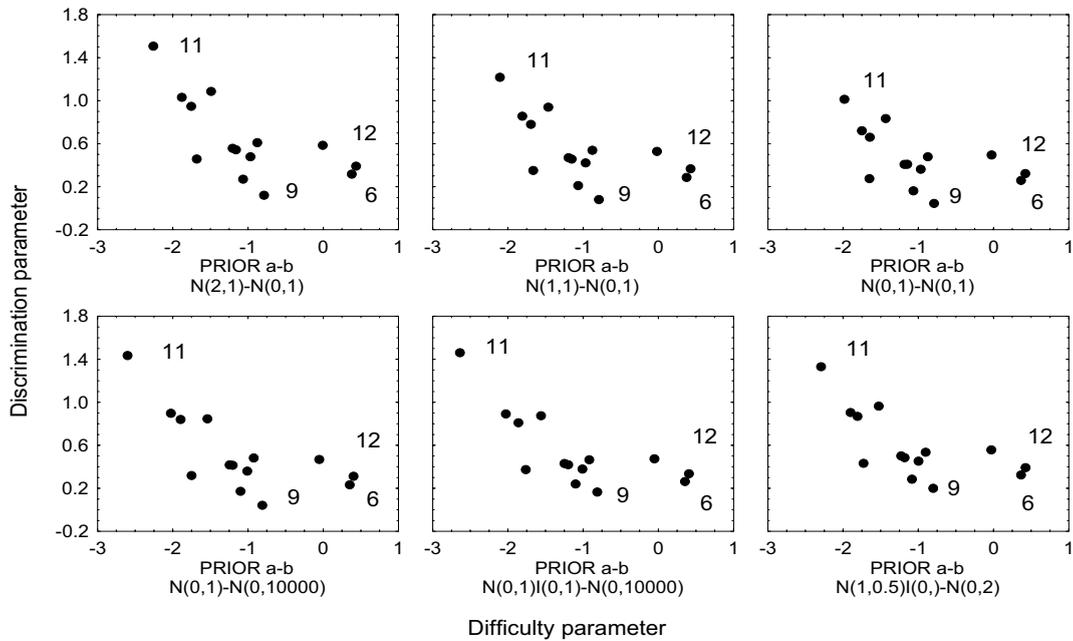
4.1. A EMPIRICAL SENSITIVITY ANALYSIS TO PRIORS OF ITEM PARAMETERS

In order to evaluate the sensitivity of the Bayesian Estimation for the probit-normal model by considering different priors to item parameters (see priors A to F in Table 2), we conduct an analysis using the data set described above. The prior specification, starting values to define the initial state of the Markov Chain, and convergence diagnostics for the Markov chain are discussed by implementing MCMC using DAGS scheme.

Bayesian estimation procedures based in MCMC was implemented in WinBUGS. Chains with 50000 iterations were generated considering $\text{thin}=1, 5, 10$ and discarding the 500 first iterations as in Albert (1992), so that effective sample sizes were 49500, 9900 and 4950, respectively. When using MCMC, the sampled values for initial iterations of the chains are discarded because of their dependence on the starting state and to guarantee proper convergence. Also, in this IRT model, presence of autocorrelation between chain values is expected when latent variables are introduced (Chen et al. 2000). Due to it, thin values up to 10 are recommended.

FIGURE 1

Posterior mean of discrimination and difficulty parameters for different priors in Math Test (14 items applied to 131 examinees) under the probit-normal IRT model. The numbers indicate the correspond item and the order of the priors specification is A, B and C up and D, E, and F



Estimates of item discrimination (a_j) and difficulty parameters (b_j) based in the posterior mean for different priors given in Table 2 are depicted in Figure 1. According to the Figure 1, the probit-normal model is insensitive to the prior specifications for the difficulty and discrimination parameters in the literature. The order of the item considering the media posterior of the parameters of discrimination and difficulty is similar between different specification priors. Moreover, a one-way analysis of variances for the means of posterior coefficient of variation for discrimination and difficulty parameters in the test (see Table 3), are not significant in relation to the different priors used (discrimination parameters: $F(5, 78) = 1.014$; $p < 0.415$, difficulty parameters: $F(5, 78) = 0.631$; $p < 0.676$) and the analysis of the DIC in Table 3, confirms that all priors lead to similar fit. That is, by considering any prior specification (A to F) in the average the ordering of the items is similar. It means that the probit-normal model is insensitive to the use of non-informative priors. Easy items as discriminators are identified similarly by the different priors specified.

TABLE 3

Summary of posterior coefficients of variation (c.v.) for item parameters and fitness criteria for Math Test under probit-normal IRT model with different priors.

prior	a_j posterior c.v.		b_j posterior c.v.		Criterion of Fit			
	mean	sd	mean	sd	Dbar	Dhat	ρ_D	DIC
A	0.590	0.344	-1.895	6.747	1445.00	1352.90	92.11	1537.11
B	0.698	0.549	-0.999	1.827	1453.16	1368.36	84.80	1537.96
C	0.890	1.033	-0.979	1.286	1461.09	1383.61	77.48	1538.57
D	0.901	1.090	-1.099	0.646	1453.22	1375.48	77.75	1530.97
E	0.532	0.114	-1.101	0.634	1450.14	1369.86	80.28	1530.43
F	0.481	0.111	-1.048	1.101	1446.60	1358.26	88.34	1534.94

Another aspect of interest is to evaluate the differences that can exist in the estimated values of the item parameters to the different prior specifications by considering the value of the posterior standard deviation. That is, the "scale" in which the values of item parameters are presented. It is possible to make comparisons for different priors by considering plotting the means vs standard deviations for posterior coefficients of variation for item parameters of Table 3. Great coefficient of variation is indicative of large standard deviation. For the discrimination and difficulty parameter, the estimates with different priors are organized in an horizontal axis, going from the smallest to the greatest mean for the posterior coefficient of variation and the vertical axis from the smallest standard deviations to the greatest standard deviation for the posterior coefficients of variation. Clearly, priors with greater restrictions in discrimination parameter, as it is the case of the truncated priors *F* and *E*, lead to more restricted estimates. This result suggest that priors with hyperparameter that admits negative values for the discrimination parameter (mean equal to zero in prior *C* and *D*) results in minor restricted estimates (in a scale major).

With respect to the difficulty parameter, the estimates with different priors are organized from the least (in order, priors *D* and *E*) to the most precise (least vague) prior (in order, prior *A*, *B* and *C*). A large mean and standard deviation are obtained for prior *A* because the estimates of coefficient of variation for the difficulty parameter of item 3 is large. The major value in the estimative relative to the item 3 follows because the estimates (means) are closed to zero and then to a small standard deviations follow a great coefficients of variation. This fact implies that for prior *A* estimates for difficulty parameter are closed to zero.

Although the priors considered lead to similar posterior distribution and fit, we find it more appropriate to use priors E and F because it results in greater precision to the scale in the which discrimination and difficulty estimates are showed. Since priors with large variances for the difficulty parameter can lead to greater values to the estimate in a larger scale, we considered prior F to be more appropriate. This fact is noted in Sahu (2002), where other choices for the hyper parameters were also investigated, and similar conclusions are observed.

Further studies were performed with the prior F for parameter estimation in the Math Test data set. In this case, we consider burn-in of 1000 and the chain run for 10000 more iterations after the burn-in with $thin = 10$ to reduce chain dependence. In this part of the work we decided to use five parallel chains. Convergence was diagnosed via visualization of trace plot using Gelman and Rubin's statistics (Gelman and Rubin,1992) present in the software R at package CODA (Best et al. 1997).

The average bur-in time takes about 90 seconds and to generate about 10000 further iterations for each chain, takes about 957 seconds on a AMD Duron(tm) Processor with 112 MB RAM). Figure 2 depicts chain history and the empirical density for item 12 and the mean and standard deviations for the latent variable describing subject's ability.

FIGURE 2

Chain history and posterior density for item parameters of item 12 and posterior mean and posterior standard deviation of the latent variable for Math the Test under the probit-normal IRT model

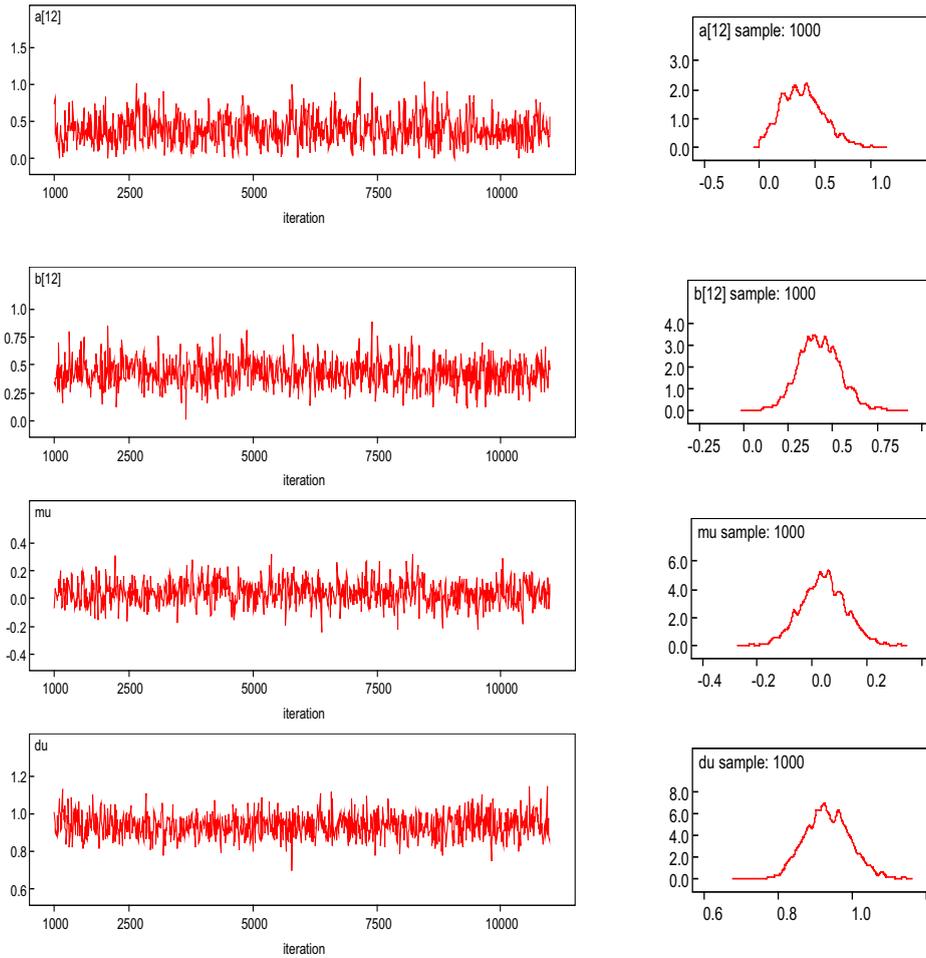


FIGURE 3

Box-plots for discrimination (a) and difficulty (b) item parameters for the Math Test under the probit-normal IRT model

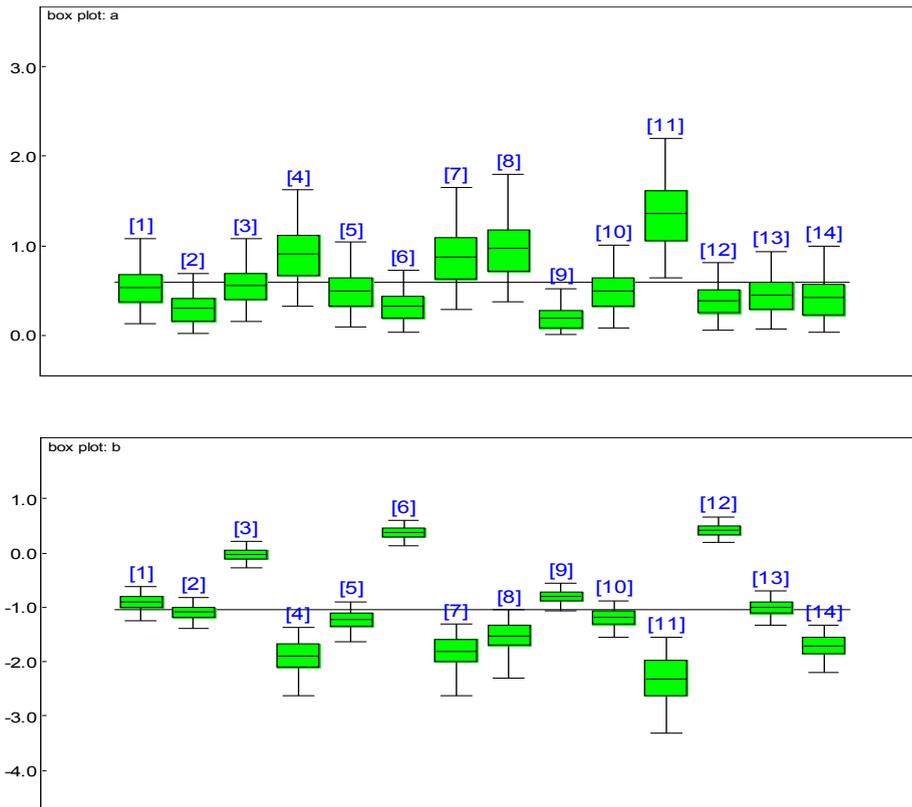


TABLE 4

Summary of posterior inference for the parameters for the Math Test under the probit-normal IRT model

item	discrimination parameter					difficulty parameter				
	mean	sd	$P_{2.5}$	median	$P_{97.5}$	mean	sd	$P_{2.5}$	median	$P_{97.5}$
1	0.54	0.24	0.13	0.52	1.09	-0.90	0.16	-1.25	-0.90	-0.60
2	0.29	0.18	0.03	0.27	0.70	-1.09	0.15	-1.37	-1.08	-0.80
3	0.55	0.24	0.16	0.53	1.09	-0.03	0.13	-0.27	-0.03	0.23
4	0.91	0.33	0.33	0.88	1.61	-1.91	0.33	-2.62	-1.88	-1.36
5	0.49	0.24	0.09	0.48	1.02	-1.23	0.18	-1.62	-1.22	-0.90
6	0.32	0.18	0.03	0.30	0.73	0.37	0.11	0.16	0.37	0.61
7	0.88	0.34	0.29	0.84	1.60	-1.82	0.33	-2.63	-1.78	-1.28
8	0.97	0.35	0.39	0.92	1.80	-1.53	0.31	-2.30	-1.50	-1.03
9	0.20	0.14	0.01	0.18	0.51	-0.80	0.13	-1.06	-0.80	-0.57
10	0.49	0.23	0.08	0.48	1.02	-1.19	0.17	-1.55	-1.18	-0.87
11	1.35	0.41	0.64	1.32	2.18	-2.31	0.46	-3.32	-2.25	-1.53
12	0.39	0.19	0.06	0.38	0.80	0.43	0.12	0.21	0.42	0.67
13	0.45	0.23	0.05	0.43	0.93	-1.01	0.16	-1.35	-1.00	-0.71
14	0.41	0.26	0.03	0.37	0.97	-1.71	0.23	-2.20	-1.70	-1.32
	<i>u</i> mean					<i>u</i> s.d				
	0.94	0.06	0.82	0.93	1.06	0.04	0.08	-0.12	0.04	0.21

Estimates of item discrimination and difficulty parameters for the probit-normal model are presented in Table 4. Item 11 is the most discriminative while item 9 is the least. Also, item 11 is the easiest while item 12 is the most difficult.

Item 11 says: “Luisa, Dora and Mary bought some cloth. Luisa bought half of a meter, Dora bought 75 centimeters and Mary bought fifty centimeters. Which ones did buy the same quantity of cloth?”. Item 12 say: “A recipient receive 4.5 liters of water by minute. How many liters of water will have the recipient after of one and half hour?”. On the other hand, item 6 say: “It solves the following operations decimal: $0.75 - 0.2 + 1.2 - 0.30$ ”. Clearly, for the formulated items, (latent) ability required to correctly solving item 11 is smaller than that required for solving item 12. For a given ability value, a student has greater probability of success with item 11 than with item 12. Hence, according to the probit-normal model, item 11 is easier than 12.

It is also worth noticing with respect to item 11 that a small change in student's ability (knowing the meaning of half meter, for instance) results in greater probability of success for this item. On the other hand, with respect to item 9, it can be depicted that small changes in student's ability (knowledge of decimal places, for instance) is not directly translated into a greater increase in the probability of correctly answering the item. Thus, item 11 seems to present greater discriminatory power because it allows to better distinguish between students that know or don't know some specific knowledge about the items. Box-plots for the item parameters based in posterior inference are presented in Figure 3 indicating the variation in posterior distribution for item parameters. This results is impossible with frequentist approaches and clearly shows the advantage in using Bayesian estimation. For further details on the interpretation of item parameters see Johnson and Albert (1999).

With respect to the latent variable ability, we have that its posterior mean is 0.94 which indicates that the group of students present negative asymmetry, which was also found with the observed scores. Hence, estimated abilities are connected with the observed scores.

5. DISCUSSION

One of the objectives of this paper is to contribute with some literature connected to the use of probit-normal models in the item response theory literature to the Latinoamerican statistical community. This model was chosen because it allows extensions to three parameter models, multidimensional models, multilevel models, inclusion of predictor variables and confirmatory factor analysis models. The paper discusses two procedures using MCMC methodology to make inference on item and ability model parameters by implementing the DAGS scheme.

The methodology was applied to the data set that resulted from a mathematical test applied to 131 sixth grade students from Peruvian schools. A sensitivity analysis by checking model adequacy that follows by using a series of prior distributions is also conducted. It includes the specification of vague prior distributions for the difficulty parameters and precise priors for the discrimination parameters.

The priors specified in Table 2 were suggested in previous studies by other authors. The different priors considered lead to similar estimates of the DIC (described in Spiegelhalter et al., 2002) leading to the conclusion that the Bayesian analysis for the data set under consideration is not sensible to the priors considered. Item most discriminative and most difficult items are equally identified by the different prior specification.

However, by considering the “scale” to the estimative of item parameters we consider that the prior considered in Sahu (2002), which specifies that $a_j \sim N(1; 0.5)I(0, \cdot)$ and $b_j \sim N(0, 2)$, $j = 1 \dots$, seems to us the most adequate one since the values of the standard deviations for item parameters are most precise than for the other vague priors.

The issue of sample size and its effect on item parameter estimation has been well studied (e.g., Swaminathan and Gifford, 1983) in the context of classical estimation. In general, large sample sizes are needed to estimate IRT item parameters in classical estimation. The issue that needs to be addressed is that of estimating or calibrating items using a small sample of examinees. Swaminathan and Gifford (1982, 1985, 1986) and Mislevy (1986) have shown that by incorporating prior information about item parameters, to different IRT models using logistic links, not only can item parameters be estimated more accurately but the estimation can be carried out with smaller sample sizes. In these lines, the present study shows that the probit-normal IRT model is insensitive to non-informative different priors considered and indicate that Bayesian estimation procedures show considerable promise when dealing with a small number of examinees in the samples, as suggested by Swaminathan et al. (2003). We direct to Swaminathan et al. (2003) on the explanation of how to use informative priors in the analysis of IRT models. Informative priors can be obtained, for example, in national and international evaluations since that pilot studies are considered. In future works it can be interesting to compare results of informative and non-informative priors. In more general situations, pilot studies are not available and one application seems to be required. In this case, the work shows that it is possible to use non-informative priors since the ordering in the item parameters is the same and then the conclusions about of the items expected to be the same.

REFERENCES

ALBERT, J. H. (1992). “Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling”. *Journal of Educational Statistics*, **17**: 251 - 269.

ALBERT, J.H. and GHOSH, M. (2000). “Item response modeling”. *Generalized Linear Models: A Bayesian Perspective* (D. Dey, S. Ghosh Mallick, eds.), Marcel-Dekker, New York, 173-193.

ANDRADE, D. F., TAVARES, H. R. and VALLE, R. C. (2000). *Introdução à Teoria da resposta ao Item : Conceitos e Aplicações*. 14º SINAPE : Caxambu, MG.

- ASSUNÇÃO, C. (1999). *Estimao dos parmetros de modelos da teoria de resposta ao item e aplicaes*. Mestrado em Estatstica - UFMG.
- BAKER, F.B. and KIM, S.-H. (2004). *Item Response Theory - Parameter Estimation Techniques (2nd ed.)*. New York: Marcel Dekker, Inc.
- BARTHOLOMEW, D.J. and KNOOT, M. (1999). *Latent variable models and factor analysis*. (2nd ed.). London: Arnold. (Kendalls Library of Statistics 7).
- BÉGUIN, A. A. and GLAS, C. A. W. (2001). “MCMC estimation of multidimensional IRT models”. *Psychometrika*, **66**: 541-562.
- BEST N. G., COWLES M. K. and VINES, S, K, (1997). *CODA: Convergence diagnosis and output analysis software for Gibbs sampling output, Version 0.4*. [Computer program]. MRC Biostatistics Unit, Cambridge University.
- BIRNBAUM, A. (1968). “Some latent trait models and their use in inferring an examinee’s ability”. In F. M. Lord; M. R. Novick. *Statistical Theories of Mental Test Scores*. Reading, MA : Addison-Wesley.
- BOCK, R. D. and AITKIN, M. (1981). “Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm”. *Psychometrika* **64**: 153-168.
- BORSBOOM, D. , MELLENBERGH, G. J. and van HEERDEN, J.(2003). “The theoretical status of latent variables”. *Psychological Review*, **110**: 203-219.
- CONGDON P (2001). *Bayesian Statistical Modelling*. John Wiley & Sons, Chichester, UK.
- CHEN, M.H., SHAO, Q. M. and IBRAHIM, J. G (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer Verlag.
- DELORIO, M. and ROBERT, C. P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin and van der linde (2002). *Journal Royal Statistics Society. Serie B*. **64**: 629-630.
- DU TOIT, M. (2003). *IRT from SSI*. Scientific Software International.
- FISCHER, G. and MOLENAAR, I. (1995). *Rasch models. Foundations, recent*

development, and applications. The Netherlands: Springer-Verlag.

FOX, J. P. and GLAS, C. A.W. (2001). “Bayesian estimation of a multilevel IRT model using Gibbs sampling”. *Psychometrika*, **66**: 271-288.

FOX, J. P. and GLAS, C. A.W. (2003). “Bayesian Modeling of measurement error in predictor variables using item response theory”. *Psychometrika*, **68**: 169-191.

FROELICH, A. G. (2000). *Assesing unidimensionality of test item and some asymptotics of parametric item response theory*. Unpublished doctoral dissertation. University of Illinois at Urbana-Champaign, Departament of Statistics.

GELMAN, A., MENG, X.-L. and STERN, H. (1996). “Posterior predictive assessment of model fitness via realized discrepancies”. *Statistica Sinica*, **6**: 733-807.

GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**: 457-472.

GLAS, C. A. W. and MEIJER, R. (2003). “A Bayesian approach to person fit analysis in item response theory models”. *Applied Psychological Measurement*, **27**: 217-233.

GHOSH, M., GHOSH, A., CHEN, MING-HUI and AGRETI, A. (2000). “Non-informative priors for one parameter item response models”. *Journal of Statistical Planning and Inference*. **88**: 99-115.

GILKS, W. R. and WILD, P. (1992). “Adaptive rejection sampling for Gibbs sampling”. *Applied Statistics*, **41**: 337-348.

HOLLAND, P. and ROSENBAUM, P. (1986). “Conditional association and unidimensionality in monotone latent variable models”. *The Annals of Statistics*. **14**: 1523-1543.

JACKMAN, S. (2001). “Multidimensional analysis of roll call data via Bayesian simulation: identification, estimation, inference and model checking”. *Political Analysis*, **9**: 227-242.

JOHNSON, V. and ALBERT, J. (1999). *Ordinal data modeling*. New York, MA: Springer-Verlag.

KIM, S., COHEN, A. S., BAKER, F. B. and SUBKOVIK, M. J.; LEONARD, T. (1994). "An investigation of hierarchical Bayes procedures in item response theory". *Psychometrika*, **59**: 405-421.

LINARDAKIS, M. and DELLAPORTAS, P. (2002). "An approach to multidimensional item response modeling". E. I. George. (Ed.). *Bayesian Methods with applications to science policy and official statistics*. 331-340.

LORD, F. M. (1952). *A theory of test scores*. New York: Psychometric Society.

LORD, F. and NOVICK, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

MARTIN, A.D. and QUINN, K. M. (2002). MCMCpack.
scythe.wustl.edu/mcmcpack.html.

MISLEVY, R. J. (1986). "Bayes modal estimation in item response models". *Psychometrika*, **51**: 177-195.

MISLEVY, R. J. and BOCK, R.D. (1990). *BILOG: Maximum likelihood item analysis and test scoring with logistic models for binary items* [Computer program]. Chicago: Scientific Software International.

NEAL, RADFORD. (2003). "Slice Sampling". *The Annals of Statistics*, **31**, No. **3**: 705-767.

NEYMAN, J. and E. SCOTT (1948). "Consistent estimates based on partially consistent observations". *Econometrica* **16**: 1-32.

PATZ, R. J. and JUNKER, B. W. (1999). "A straightforward approach to Markov Chain Monte Carlo methods for item response models". *Journal of Educational and Behavioral Statistics*, **24**: 146-178.

ROBERT, C. P.(2001). *The Bayesian Choice: from decision-theoretic foundations to computational implementation*. 2nd ed. New York: Springer-Verlag.

RUPP, A., DEY, D. K. and ZUMBO, B. (2004). "To Bayes or Not to Bayes, from Whether to When: Applications of Bayesian methodology to item response modeling". *Structural Equations Modeling*, **11**: 424-451

SAHU, S. K. (2002). “Bayesian estimation and model choice in item response models”. *Journal of Statistical Computation and Simulation*, **72**: 217-232.

SEGALL, D. (2003). “Confirmatory item factor analysis using Markov Chain Monte Carlo estimation with applications to Online Calibration in CAT”. Preprint.

SPIEGELHALTER, D. J., THOMAS, A., BEST, N. G. and GILKS, W.R. (1996). *BUGS 0.5 examples* (Vol. 1 Version i). Cambridge, UK: University of Cambridge.

SPIEGELHALTER D.J., THOMAS, A. and BEST N. G. (2003). *WinBUGS version 1.4* [Computer program]. Imperial College; Medical Research Council Biostatistics Unit. Institute of Public Health, Cambridge University.

SPIEGELHALTER D. J., BEST N. G., CARLIN B. P. and van DER LINDE A. (2002). Bayesian measures of model complexity and fit (with discussion). *J. Roy. Statist. Soc. B.* **64**: 583-640.

STERN, H. S. and SINHARAY, S. (2005). “Bayesian model checking and model diagnostics” in *Bayesian thinking: modeling and computation, Handbook of Statistics*, **25**, eds. D.K. Dey and C. R. Rao, Elsevier: Amsterdam, 171-192.

SWAMINATHAN, H. and GIFFORD, J. A. (1982). “Bayesian estimation in the Rasch model”. *Journal of Educational Statistics*, **7**: 175-191.

SWAMINATHAN, H. and GIFFORD, J. A. (1983). “Estimation of parameters in the three-parameter latent trait model”. In D. Weiss (Ed.), *New horizons in testing* 13-30. New York: Academic Press.

SWAMINATHAN, H. and GIFFORD, J. A. (1985). “Bayesian estimation in the two-parameter logistic model”. *Psychometrika*, **50**: 175-191.

SWAMINATHAN, H. and GIFFORD, J. A. (1986). “Bayesian estimation in the three-parameter logistic model”. *Psychometrika*, **51**: 581-601.

SWAMINATHAN, H., HAMBLETON, R., SIRECI, S., XING, D., and RIVAZI, S. (2003). “Small Sample Estimation in Dichotomous Item Response Models: Effect of Priors Based on Judgmental Information on the Accuracy of Item Parameter Estimates”. *Applied Psychological Measurement*, **27**: 27-51.

TSUTAKAWA, R. M. (1984). "Estimation of two-parameter logistic item response curves". *Journal of Educational Statistics*, **9**: 263-276.

UMC (2001). *Análisis de ítems de las pruebas Crecer 1998. Resultados de lógico matemática en sexto grado de primaria*. Boletín 13. Unidad de Medición de Calidad Educativa. Ministerio de Educación del Perú. Available from http://www.minedu.gob.pe/umc/publicaciones/boletines/Boletin_13.pdf

van DER LINDEN, W.J. and HAMBLETON, R. K. (1997). *Handbook of Modern Item response Theory*. New York: Springer-Verlag.

WANG, X., BRADLOW, E. T. and WAINER, H. (2003). "A General Bayesian Model for Testlets: Theory and Applications". *Applied Psychological Measurement*, **26**: 109-128.

APPENDIX: WINBUGS CODE TO PROBIT-NORMAL IRT MODEL

```

model
{ #likelihood function
  for (i in 1 : n) {
    for (j in 1 : I) {
      m[i,j] <- a[j]*u[i] - b[j]
      z[i,j] ~ dnorm(m[i,j],1)I(lo[y[i,j]+1],up[y[i,j]+1])
    }
  }
  #priors F for item parameters
  for (j in 1:I) {
    b[j] ~ dnorm(0,0.5);
    a[j] ~ dnorm(1,2)I(0,);
  }
  #prior for latent variable
  for (i in 1:n) { u[i] ~ dnorm(0,1) }

  # auxiliary latent variable
  lo[1] <- -50; lo[2] <- 0; \# i.e., z| y=0 ~ N(m,1)I(-50,0)
  up[1] <- 0; up[2] <- 50; \# i.e., z| y=1 ~ N(m,1)I(0,50)

  # mean and standard deviation for latent variable
  mu<-mean(u[])
  du<-sd(u[])
}

```

```

list(n=131, I=14, y= structure(.Data =c(
1,1,0,1,1,0,1,1,1,1,1,0,0,1,
1,1,1,1,1,1,1,1,1,1,1,0,1,1,
...
...
1,0,0,1,1,1,1,1,1,1,1,1,1,1,
1,1,0,1,0,0,1,1,0,0,1,0,1,1),
.Dim=c(131,14))) )

```

```

Inits list(b = c(0,0,0,0,0,0,0,0,0,0,0,0,0,0),
a=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1))

```

In WinBUGS 1.4 it is possible to change the sampling methods. The sampling methods are held in Updater/Rsrc/Methods.odc and can be edited. For example, if there are problems with WinBUGS's adaptive rejection sampler (DFreeARS), then the method "UpdaterDFreeARS" for "log concave" could be replaced by "UpdaterSlice" (normally used for "real non linear"). This is suggested in the WinBugs site

(<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml#problems>) and helps to avoid Trap message. When the correction is used in the Script example the default method for Z is UpdaterSlice. Interval and not UpdaterDFreeARS. Interval.