Bayesian Analysis (2006) **1**, Number 4, pp. 861–892

# A Skew Item Response Model

Jorge L. Bazán*, Márcia D. Branco †, and Heleno Bolfarine‡

**Abstract.** We introduce a new skew-probit link for item response theory (IRT) by considering an accumulated skew-normal distribution. The model extends the symmetric probit-normal IRT model by considering a new item (or skewness) parameter for the item characteristic curve. A special interpretation is given for this parameter, and a latent linear structure is indicated for the model when an augmented likelihood is considered. Bayesian MCMC inference approach is developed and an efficiency study in the estimation of the model parameters is undertaken for a data set from (Tanner 1996, pg. 190) by using the notion of effective sample size (ESS) as defined in Kass et al. (1998) and the sample size per second (ESS/s) as considered in Sahu (2002) The methodology is illustrated using a data set corresponding to a Mathematical Test applied in Peruvian schools for which a sensitivity analysis of the chosen priors is conducted and also a comparison with seven parametric IRT models is conducted. The main conclusion is that the skew-probit item response model seems to provide the best fit.

**Keywords:** link skew-probit, item response theory, Bayesian estimation, probit-normal model, skew-normal distribution

## 1 Introduction

Item Response Theory (IRT) is concerned with modeling the relationship between multivariate responses and the abilities (or hypothesized traits) of $n$ individuals that are submitted to a test with $k$ items. The model involves latent variables that explains the abilities of the individuals and a set of parameters associated with the items under consideration. Although IRT models can be used in more general contexts (van der Linden and Hambleton 1997), in this paper we are interested in modeling dichotomous item responses by modeling the probability of correct response, namely, $p_{ij}$, as $p_{ij} = F(m_{ij})$, where $F$ is called the item characteristic curve (ICC). Moreover, $m_{ij} = a_j u_i - b_j$, $i = 1, .., n$, $j = 1, ..., k$, where $a_j$ and $b_j$ are parameters associated with the items (denominated discrimination and difficulty parameters, respectively), and $u_i$ is the value of the latent variable $U_i$ (or parameter) associated with the individual ability $i$. In the context of generalized linear models $F^{-1}(.)$ is called the link function. Two special cases follow by considering that $F(.) = \Phi(.)$, or $F(.) = L(.)$, where $\Phi(.)$ is the cumulative distribution function (cdf) of the standard normal distribution and $L(.)$ is the cdf of the standard logistic distribution. Such models are usually called the probit IRT model or ogive normal model (Albert 1992; Albert and Ghosh 2000) and logit IRT-model (see Birnbaum 1968), respectively. A special feature of both models is

*Department of Statistics. University of São Paulo, Brazil, http://www.ime.usp.br/~jbazan/

†Department of Statistics. University of São Paulo, Brazil, http://www.ime.usp.br/~mbranco/index-ing.html

‡Department of Statistics. University of São Paulo, Brazil, http://www.ime.usp.br/~hbolfar/

 ba0004

the symmetric nature of the probit and logit link functions and the corresponding ICCs. However, as emphasized in Chen et al. (1999), symmetric links do not always provide good fits for some data sets. This is specially true, when the probability of a given binary response approaches zero at a different rate than it approaches one. Misspecification of the link function can yield substantial bias in the mean response estimates (see Czado and Santner 1992). Samejima (1997) also indicated the necessity of departures from normal assumptions in developing psychometric theories and methodologies. As a follow-up, Samejima (2000) proposed a family of models, called the *logistic positive exponent family*, which provides asymmetric ICCs and includes the logistic model as a special case. It is also pointed out in that paper that asymmetric ICCs are more appropriate for modeling human item response behavior. A variety of skew links have been proposed for the binary regression models, but hardly one has been used in IRT models. One example of a skew link is the generalized logit link considered in Prentice (1976) that was used by Samejima (2000) to propose the logistic positive exponent model, where the ICC curve is given by $F(.) = \text{Ł}(.)^{\varepsilon_j}$, with Ł$(.)$ as defined above and $\varepsilon_j > 0$ is the skewness parameter associated with the $j$-th item. However, it is not of our knowledge any applications or Bayesian estimation methodologies developed for this model. In fact, a recent review paper by Rupp et al. (2004) reports no Bayesian skew modelling to IRT models.

Moreover, as García-Pérez (1999) points out, works on IRT models have almost exclusively been focused on the development and comparison of parameter estimation techniques and the study of the effects of the characteristics of the data sets (sample size, test length, distribution of the true abilities) and violations of model assumptions (excluding the mathematical form of the ICC) on the capability of available algorithms to recover the generating parameters. No one seems to have questioned if the mathematical form of the ICC can be derived from psychological theory of performance in objective *testing* as opposed to adopting a convenient function that the data are forced fitting to it.

We propose here an asymmetric link function by using the skew-normal distribution (Azzalini 1985) and consequently a new ICC. Azzalini (2005) and Genton (2004) present recent reviews on the most recent and important results related to skew-normal models. In spite of that, this new ICC is not derived from psychological theory motivation, the ICC that we will consider is a generalization of the probit ICC, by introducing a skewness parameter associated to the item that can be interpreted as penalization parameter and can play an important role in testing. It defines, what we call a skew-probit ICC and it includes the probit link as special case. This more general model is flexible enough to allow using symmetric (probit) and asymmetric (skew-probit) ICCs for the $nk$ items.

The skew link proposed in this paper use the skew-normal family of distributions. The approach is characterized by: a) probabilities are defined by considering the cdf evaluated at a linear predictor; b) the asymmetry parameter is associated with the distribution defining the cdf and is independent of the linear predictor and c) a latent linear structure is not necessary for model formulation. The skew-probit link in Chen et al. (1999) also follow properties (a) and (b) above, but no (c). Their propose are based on a latent linear structure, and is a special case of a more general formulation based in the class of the scale mixtures of elliptical distributions.

The asymmetric probit link proposed in Chen et al. (1999) is obtained, using auxiliary latent variables by considering

$$y_i = \begin{cases} 1, & z_i > 0 \\ 0, & z_i \leq 0, \end{cases}$$

where

$$z_i = \eta_i + \lambda v_i + w_i,$$

with $v_i$ independent of $w_i$ and $v_i \sim HN(0,1)$ and $w_i \sim N(0,1)$. Consequently,

$$p_i = \int_0^\infty \Phi(\eta_i + \lambda v_i) g(v_i) dv_i.$$

where $\Phi(.)$ denote the cdf of a standard normal distribution and $g(.)$ is the probability density function (pdf) given by $g(x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x > 0$, corresponding to the half normal distribution. The second expression corresponds to the stochastic representation of the skew-normal distribution given in Chen et al. (1999) and Branco and Dey (2002), and it is a necessary condition in the formulation of the model.

In this paper, we adopt the Bayesian view. Our point is made since several researchers demonstrated that accurate estimation of the item parameters in small samples can only be accomplished through a Bayesian approach (see, for example, Swaminathan et al. 2003. Given the peculiarities of IRT models, maximum likelihood totally relies on large sample theory, which even for a large number of examinees, is complicated by the presence of incidental parameters. Researches using such an approach typically do separate estimation for item and ability parameters. However, there is no way to jointly evaluate estimates precision (Patz and Junker 1999). Such problems do not occur with the Bayesian approach in which, for a large number of examinees, the prior distribution has little effect on the posterior distribution (Sinharay and Johnson 2003). Computation is developed by using the MCMC methodology and the WinBUGS software, which can be used for simulating from the posterior distributions of item parameters and latent variables. Bayesian model fitting is also implemented for the logistic positive exponent model proposed by Samejima (1997) and Samejima (2000) for the data set under study, which seems to be a first attempt in that direction.

In Section 2 we introduce the skew-probit IRT model by considering a skew-probit item characteristic curve. In Section 3, we present Bayesian estimation based in a data augmentation approach and prior specification for the MCMC implementation. In Section 4 we present a study on the estimation efficiency of item parameters in the skew-probit IRT model. Two examples are given in Section 5, including a sensitivity analysis to prior choice for the item parameters, illustrating the usefulness of the approach in comparison with others parametric IRT models. Our analysis also includes the logistic positive exponent model (Samejima 1997, 2000). To choose the model that fits the data better, we consider the deviance information criterion ($DIC$) as presented in

Spiegelhalter et al. (2002). Finally, we discuss possible extensions of the model proposed in the sense of the type of asymmetry considered.

# 2 The skew-probit IRT model

## 2.1 The skew-normal distribution

As considered in Azzalini (1985), a random variable $Z$ follows a standard skew-normal distribution if its probability density function (pdf) is given by

$$\phi_{SN}(z;\lambda) = 2\Phi(z)\Phi(\lambda z).$$

where $\phi(.)$ and $\Phi(.)$ denote, respectively, the standard normal pdf and cumulative distribution function (cdf). We use the notation $Z \sim SN(\lambda)$. The parameter $\lambda$ controls skewness, which is positive when $\lambda > 0$ and negative when $\lambda < 0$. The standard normal distribution is recovered with $\lambda = 0$. Some important properties of the skew-normal distribution are given in the Appendix A0. The skew-normal cumulative distribution function is denoted by $\Phi_{SN}(z;\lambda)$ and is obtained explicitly next.

**Proposition 1.** *The cumulative distribution of $Z \sim SN(\lambda)$ is given by*

$$\Phi_{SN}(z;\lambda) = 2\Phi_2\Big(\begin{pmatrix} z \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix}\Big), \tag{1}$$

*where $\Phi_2(.)$ denotes the cumulative distribution function(cdf) of the bivariate standard normal distribution with correlation coefficient $-\delta$, which, to simplify notation, we denote by $\Phi_2((z,0)^T, -\delta)$.*

*Proof.* It follows directly of the fact that the standard bivariate normal distribution with correlation $\rho$ evaluated at the point $(h,k)$ can be written as (Parrish and Bargmann 1981) $\Phi_2(h,k;\rho) = \int_{-\infty}^{h} \phi(w)\Phi\Big(\frac{k-\rho w}{\sqrt{1-\rho^2}}\Big)dw.$

This result indicates that the skew-normal distribution evaluated at a point $z$ can also be obtained by considering the bivariate standard normal distribution with correlation $-\delta$ evaluated at the point $(z,0)$. This is important because there are several efficient algorithms for computing integrals related to the bivariate normal distribution (see Genz 1992, 1993). Another algorithm to compute the cdf of the skew-normal distribution is based on the use of Owens function (see Azzalini 1985; Dalla-Valle 2004) and is available for R and Matlab programs. More generally, a random variable $X$ follows a skew-normal distribution with location parameter $\mu$ and scale parameter $\sigma^2$, if the density function of $X$ is given by

$$f_X(x) = \phi_{SN}(\,x\,;\,\mu,\sigma^2,\lambda) = \frac{2}{\sigma}\phi_1\left(\frac{x-\mu}{\sigma}\right)\Phi_1\left(\lambda\frac{x-\mu}{\sigma}\right), \tag{2}$$

with the notation $X \sim SN(\mu, \sigma^2, \lambda)$ used in this paper. If $\lambda = 0$, the density of $X$ in (2) reduces to the density of the $N(\mu, \sigma^2)$.

*Remark 1.* By applying the properties of the skew-normal distribution and using variable transformation it follows that, if $Z \sim SN(\mu, \sigma, \lambda)$, then $Z^* = aZ + b \sim SN(a\mu + b, a^2\sigma^2, sign(a)\lambda)$, with $sign(.)$ as defined before.

## 2.2 The model

Formally, the *skew-probit IRT model* is defined by considering that

$$\begin{aligned} Y_{ij}|u_i, a_j, b_j, \lambda_j &\sim Bern(p_{ij}) \\ p_{ij} &= \Phi_{SN}(m_{ij}; \lambda_j) \\ m_{ij} &= a_j u_i - b_j, \end{aligned} \tag{3}$$

where $Bern$ denotes the Bernoulli distribution, $i = 1, \ldots, n$, and $j = 1, \ldots, k$. Moreover, the skew-probit IRT model satisfies the latent conditional independence principle, which considers that for the i-*th* examinee, $Y_{ij}$ are conditionally independent given $U_i$, $j = 1, \ldots, k$. It is also considered that responses from different examinees are independent.

In the following, we use the notation $\mathbf{a} = (a_1, \ldots, a_k)^T$, $\mathbf{b} = (b_1, \ldots, b_k)^T$, $\lambda = (\lambda_1, \ldots, \lambda_k)^T, \mathbf{y} = (y_{11}, \ldots, y_{kn})^T$. Let $D_{obs} = \mathbf{y}$ to denote the observed data, so that the likelihood function for the skew-probit IRT model is given by

$$L(\mathbf{u}, \mathbf{a}, \mathbf{b}, \lambda|D_{obs}) = \prod_{i=i}^{n}\prod_{j=1}^{k}[\Phi_{SN}(m_{ij}; \lambda_j)]^{y_{ij}}[1 - \Phi_{SN}(m_{ij}; \lambda_j)]^{1-y_{ij}}. \tag{4}$$

Using the representation of the skew-normal distribution, given in Proposition 1, we can write

$$L(\mathbf{u}, \mathbf{a}, \mathbf{b}, \lambda|D_{obs}) = \prod_{i=i}^{n}\prod_{j=1}^{k}\left[2\Phi_2\left[(m_{ij}, 0)^T; -\delta_j\right]\right]^{y_{ij}}\left[1-2\Phi_2\left[(m_{ij}, 0)^T; -\delta_j\right]\right]^{1-y_{ij}}. \tag{5}$$

The skew-probit IRT model involves a total of $n+3k$ unknown parameters being thus overparameterized. Moreover, for a fixed number of items, item parameters are considered structural parameters and the latent variables (abilities) are incidental parameters, that is, their number increases as the sample size $n$ increases. The model is also unidentifiable, since it is preserved under a special class of transformations of the parameters (see Albert 1992). Such aspects make it a temerity using maximum likelihood estimation in IRT models. One way to contour such difficulties is to impose restrictions on the item parameters and abilities as considered, and obtain the marginal likelihood to analyze item parameters estimation, for example, in Bock and Aitkin (1981). Another way follows by specifying a distribution for the latent variables. As frequently considered in the literature (see, for example, Patz and Junker 1999), we take $U_i \sim N(0,1)$, $i = 1, \ldots, n$. This assumption establishes that it is believed that the latent variables are well behaved and that the abilities are a random sample from this distribution and, additionally, this establishes a metric for the abilities estimates. Hence, the model defined by (4) or (5) also can be denominated skew-probit normal IRT model and includes the called probit-normal IRT model (Albert 1992; Albert and Ghosh 2000) as a particular case ($\lambda = 0$).

## 2.3 The skew-probit item characteristic curve

The skew-probit item characteristic curve (ICC) is given by

$$p_{ij} = P[Y_i = 1 \mid u_i, a_j, b_j, \lambda_j] = \Phi_{SN}[m_{ij}; \lambda_j] = 2\Phi_2\big[(m_{ij}, 0)^T; -\delta_j\big], \tag{6}$$

where $p_{ij}$ is the conditional probability of a correct response for item $j$ given the latent variable $u_i$ corresponding to the i-*th* examinee, $\delta_j = \frac{\lambda_j}{(1+\lambda_j^2)^{1/2}}$ and $m_{ij} = a_j u_i - b_j, i = 1, \ldots, n$, and $j = 1, \ldots, k$.

In the above expression, the probability $p_{ij}$ is expressed as a function of the quantity $u_i$ and the parameters $\eta_j = (a_j, b_j)^T$ and $\lambda_j$, which are parameters associated with the item $j$. Note that, for $\lambda_j = 0$ expression (5) reduces to $p_{ij} = \Phi(m_{ij})$, as considered in the probit model (Albert 1992).

Motivated by considering the shape of the ICCs for different values of $\lambda$ (see Figure 1), $\lambda$ can be interpreted as a *penalization parameter*. Hence, when an item has associated $\lambda > 0$, we say that the probability of correct response is penalized for low values of the latent variable $U$. A fixed and positive change on the latent variable implies positive but smaller (bigger) changes in the probability of success for lower (higher) values of the latent variable $U$. On the other hand, when an item has associated $\lambda < 0$, we say that the item is penalized for high values of the latent variable $U$. In this case, a positive and fixed change of the latent variable $U$ results in positive but smaller (bigger) changes in the probability of success for higher (lower) values of the latent variable. The interpretation is the same when the parameter $\delta$ is used.

Note also that

$$h(u) \doteq \frac{dp_{ij}}{du_i} = a_j \frac{d\Phi(a_j u_i - b_j; \lambda)}{du_i} = a_j \phi(a_j u_i - b_j; \lambda),$$

which corresponds to the pdf of the skew-normal distribution evaluated at $a_j u_i - b_j$ multiplied by $a_j$. Hence, as $|\lambda|$ (or $|\delta|$) grows, asymmetry in the IRT model also grows, so that $\lambda$ can also be considered as an asymmetry parameter. In the case of $a = 1$ and $b = 0$ we obtain that $h(U) \sim SN(\lambda)$, that is, the standard skew-normal distribution. Note also that when $\lambda = 0$, the estimation of $a$ and $b$ is the same as in the probit-normal model. Moreover, when $\lambda \neq 0$, the interpretation of $a$ and $b$ is the same as in the probit IRT model. As a consequence of the properties of the skew-normal cdf, it can be verified that the ICC is a monotone increasing function of the quantity $u_i$ which is considered as a latent variable. This means that the skew-probit IRT model is an unidimensional monotone latent variable model (Junker and Ellis 1997).

## 2.4 Data augmentation approach

Our approach is motivated by the method of Albert and Chib (1993), where the underlying latent variable follows a normal distribution. Here, the underlying latent variable follows a standard skew-normal distribution (Azzalini 1985). The next result is an extension of a result in Albert (1992) for the probit IRT model.

**Proposition 2.** *The skew-probit IRT model, involving $k$ items and $n$ examinees, as given in (3) is equivalently defined by considering that*

$$Y_{ij} = \begin{cases} 1, & Z_{ij} > 0; \\ 0, & Z_{ij} \leq 0, \end{cases} \tag{7}$$

*where* $Z_{ij} \sim SN(m_{ij}, 1, -\lambda_j)$, $j = 1, \ldots, k$ and $i = 1, \ldots, n$.

*Proof.* The proof uses the fact that $1 - \Phi_{SN}(z; -\lambda) = \Phi_{SN}(-z; \lambda)$ (see Property E in Azzalini 1985) and is similar to that given in Albert (1992).

Clearly, in the special case where $\lambda_j = 0$, $j = 1, \ldots, k$, the corresponding result in Albert (1992) for the symmetric probit IRT model follows. The latent variables $Z_{ij}$ are introduced to avoid working with a Bernoulli type likelihoods. Furthermore, notice that the skewness parameter with the auxiliary latent variable is the opposite of the skewness parameter of the ICC. In the following, we use the notation $\mathbf{z} = (z_{11}, \ldots, z_{kn})^T$ and $\mathbf{D}_1 = (\mathbf{z}^T, \mathbf{y}^T)^T$. The *complete data likelihood function* for the skew-probit IRT model is given by

$$L(\mathbf{u}, \mathbf{a}, \mathbf{b}, \lambda | D_1) \propto \prod_{i=1}^{n} \prod_{j=1}^{k} \phi_{SN}(Z_{ij}; m_{ij}, 1, -\lambda_j) p(y_{ij} \mid z_{ij}), \tag{8}$$

where $p(y_{ij} \mid z_{ij}) = I(z_{ij} > 0)I(y_{ij} = 1) + I(z_{ij} \leq 0)I(y_{ij} = 0)]$ is the conditional likelihood and $I$ is the usual indicator function, $j = 1, \ldots, k$, and $i = 1, \ldots, n$. Note that, if $\lambda_j = 0$, the likelihood function above is similar to the one given in Albert (1992). An alternative way of writing the skew-probit IRT model is presented next by considering the asymmetry parameter $\delta = (\delta_1, \ldots, \delta_k)^T$.

**Proposition 3.** *The skew-probit IRT model considered in Proposition 2 can be equivalently defined by considering the auxiliary latent variables*

$$Z^*_{ij} \sim N(-\delta_j v_{ij} + m_{ij}\ ,\ 1 - \delta_j^2) \qquad \textit{and} \qquad V_{ij} \sim HN(0, 1),$$

*$j = 1, \ldots, k$ and $i = 1, \ldots, n$. Denoting $\mathbf{z}^* = (z^*_{11}, \ldots, z^*_{kn})^T$ and $\mathbf{v} = (v_{11}, \ldots, v_{kn})^T$, the complete data likelihood function corresponding to $D_2 = (\mathbf{z}^*, \mathbf{v}, \mathbf{y})$, is given by*

$$L(\mathbf{u}, \mathbf{a}, \mathbf{b},\ \delta | D_2) \propto \prod_{i=1}^{n} \prod_{j=1}^{k} \phi(Z^*_{ij}\ ; -\delta_j v_{ij} + m_{ij}\ , 1 - \delta_j^2) \phi(V_{i,j}\ ; 0, 1) I(v_{ij} > 0) p(y_{ij} \mid z^*_{ij}), \tag{9}$$

*where $p(y_{ij} \mid z^*_{ij}) = I(z^*_{ij} > 0) I(y_{ij} = 1) + I(z^*_{ij} \leq 0) I(y_{ij} = 0)$ is the conditional likelihood and $I$ is the usual indicator function, $j = 1, \ldots, k$ and $i = 1, \ldots, n$.*

*Proof.* From Remark 1 in Section 2.1 we have that $Z_{ij}$ in Proposition 2 can be expressed as $Z_{ij} = m_{ij} + e_{ij}$ where $e_{ij} \sim SN(0, 1, -\lambda)$. Now considering the stochastic representation given in **?** (Property 4 in the Section 2.1) for $e_{ij}$ we have that:

$$Z_{ij} = m_{ij} - \delta V_{ij} - (1 - \delta^2)^{1/2} W_{ij}$$

where $V_{ij} \sim HN(0, 1)$ and $W_{ij} \sim N(0, 1)$. It then follows from Property 5 in Section 2.1 that $Z^*_{ij} = Z_{ij} \mid v_{ij} \sim N(-\delta_j v_{ij} + m_{ij}\ ,\ 1 - \delta_j^2)$. The proof is completed since $p(y_{ij} \mid z_{ij, v_{ij}})$ is one if $Z^*_{ij}$ obeys the constraint imposed by the observed value $y_{ij}$, and $v_{ij} > 0$ is always true.

# 3 Bayesian estimation

## 3.1 Priors specification

Prior specification is an important step in Bayesian analysis. It is more important for small sample sizes where the posterior distribution represents more of a compromise between the observed data and previous personal opinion. For large sample sizes, it has less importance since the data typically dominates the posterior (information) distribution.

In the IRT literature there seems to be consensus with respect to the prior for $\mathbf{U}$, but different priors have been investigated for the item parameters (see Rupp et al. 2004; **?**). Empirical evidence (Patz and Junker 1999, among others) seems to indicate the presence of posterior correlation between item parameters. However, it seems difficult to assign dependent priors for those parameters, being a specially hard task thinking about values for the correlations for such priors, even if a multivariate normal prior is specified. Hence, we prefer using independent and common priors for $\mathbf{a}$, $\mathbf{b}$ and $\lambda$ and let such correlations be only data dependent. That is, the prior we consider can be written as

$$\pi(\mathbf{u}, \mathbf{a}, \mathbf{b}, \lambda) = \prod_{i}^{n} \phi(u_i; 0, 1) \prod_{j}^{k} \pi_1(a_j) \pi_2(b_j) \pi_3(\lambda_j). \tag{10}$$

Although some authors as Albert (1992) and, more recently, Fox and Glas (2001, 2003), use improper noninformative priors for the parameters $a$ and $b$ of the type $\pi_1(a_j)\pi_2(b_j) =$

$I(a_j > 0)$, we prefer using informative priors on the discrimination parameters $a_j$ since the existence of the joint posterior distribution is not guaranteed when an improper prior is used. Considering the results of Albert and Ghosh (2000) and Ghosh et al. (2000), the distribution of the discrimination parameter must be proper to guarantee a proper joint posterior distribution.

Several informative distributions for $a_j$ have been proposed in the literature. To mention just a few, a) Bradlow et al. (1999) and Johnson and Albert (1999) use $N(\mu_a, \sigma_a^2)$ with or without hyper parameters distributions specified to $\mu_a$ and $\sigma_a^2$, respectively; b) Kim et al. (1999) and Patz and Junker (1999) use $LN(\mu_a, \sigma_a^2)$ with or without hyper parameters distributions specified for $\mu_a$ and $\sigma_a^2$, respectively, where LN(.) is the log-normal distribution and c) Spiegelhalter *et al.* (1996) and Sahu (2002) use $N(\mu_a, \sigma_a^2)I(0,)$ that is the Half-normal distribution with known values for $\mu_a$ and $\sigma_a^2$ and, finally, d) Swaminathan and Gifford (1985) use $IG(m, n)$, the inverted gamma distribution with (known) hyperparameters $m$ and $n$. We consider in this paper the specifications in b) and c) above since $a_j > 0$ and also for conjugation reasons.

When independent informative priors are considered for the item parameters, it is usually assigned the $N(\mu_b, \sigma_b^2)$ for $b_j$, $j = 1, \ldots, k$. Moreover, in the common situation where little prior information is available about the difficulty parameter, one can chose $\sigma_b^2$ to be large. As is mentioned in Albert and Ghosh (2000) and Sahu (2002) in the probit IRT model, this choice will have a modest effect on the posterior distribution for non extreme data, and it will result in a proper posterior distribution when extreme data (all items are correct or all items are incorrect) is observed. Thus, vague priors can be admissible for the difficulty parameter.

In this paper, we consider $\mu_a, \mu_b, \sigma_a^2, \sigma_b^2$, to be known. In more general situations, the prior structure needs to be enlarged so that hyper prior information can also be considered for those parameters. Two different parameterizations will be considered, namely, the *lambda parameterization*, where a prior distribution is considered for the $\lambda_j$ parameters, and the *delta parameterization*, where a prior distribution is considered for $\delta_j = \frac{\lambda_j}{(1+\lambda_j^2)^{1/2}}$, $j = 1, \ldots, k$, taking values in the interval $[-1, 1]$. Since under the *delta parameterization* the parametric space is bounded, a natural specification of a vague prior is to consider that $\delta_j \sim U(-1, 1)$. It is not difficult to see that it is equivalent to assuming that $\lambda_j \sim T(0, 0.5, 2)$, where $T(\mu, \sigma^2, \nu)$ denotes the Student-t distribution with location $\mu$, scale $\sigma^2$ and $\nu$ degrees of freedom (for details see Branco and Rodriguez, 2005). Other specifications can be considered as a normal distribution restricted to the interval $[-1, 1]$, given by $\delta_j \sim N(0, 1)I(-1, 1)$. We consider more convenient to use proper priors for $\delta$ and $\lambda$.

## 3.2 MCMC implementation

Considering the likelihood function in (4) or (5) and the general prior specification given in (10), we point out that to implement a Bayesian estimation procedure involving a Bernoulli likelihood can be complicated. Since the integrals involved to obtain the marginal posterior distributions are difficult to deal with, two approaches based on data augmentation were introduced in the Section 2.4. This approach allows us the implementation of Markov Chain Monte Carlo methods which make it simple to implement efficient sampling from the marginal posterior distributions.

By considering this latent structure, the full conditionals for the skew-probit IRT model and the Bayesian inference via MCMC follows without complications, similarly as reported in Johnson and Albert (1999). Note that some of the full conditionals can not be directly sampled from, requiring algorithms such as the Metropolis-Hastings (Chib and Greenberg 1995).

Moreover, to implement the Bayesian approach in WinBUGS considering directly the likelihood function in (4) it is necessary to have the cdf of the skew-normal distribution, which is not yet implemented in the software. To overcome this difficulty we can use a second augmented likelihood function that consider extra latent variables by modifying the latent variable $Z_{ij}$, $j = 1, \ldots, k$, and $i = 1, \ldots, n$.

In the remainder of this paper we develop a computational procedure for the skew-probit IRT model based in the second augmented likelihood function, given in Proposition 3. Hierarchically, the full likelihood specification for the *delta parameterization* is given as follows:

$$Z^*_{ij}|v_{ij}, y_{ij}, a_j, b_j, \delta_j \sim N(a_j u_i - b_j - \delta_j v_{ij}, 1 - \delta_j^2) I(z^*_{ij}, y_{ij});$$

$$V_{ij} \sim HN(0,1);$$

$$U_i \sim N(0,1);$$

$$a_j \sim \pi_1(\mu_a, \sigma_b);$$

$$b_j \sim \pi_2(\mu_b, \sigma_b^2);$$

and

$$\delta_j \sim \pi_3(.).$$

For the *lambda parameterization* it is only necessary to specify a prior distributions for $\lambda_j$ and $\delta_j = \frac{\lambda_j}{\sqrt{1+\lambda_j^2}}$.

All of the full conditional distributions for implementing the Gibbs sampler are straightforward to derive and to sample from (see Bazán et al. 2004b). Moreover, note that when $\delta_j = 0$ or $\lambda_j = 0$ , the hierarchical structure of the augmented likelihood corresponding to the probit-normal model follows by eliminating the second and fifth lines in the above hierarchy. The program code in Winbugs used in this application is presented in Appendix A.1.

## 4 A study on the efficiency in the estimation of the parameters of the skew-probit IRT model

In this section we investigate the efficiency in the estimation of the parameters of the skew-probit IRT model using a data set previously analyzed in the literature. The aims of the study are a) to evaluate the behavioral of the autocorrelation of the model parameters when a data augmentation approach is considered, and b) to evaluate the performance of the *delta* and *lambda parameterizations* introduced in the Section 3.2 for the penalization parameter. We consider a data set previous analyzed in (Tanner 1996, pg. 190) which includes $k = 6$ items and $n = 39$ examinees.

Typically, in MCMC, the sampled values for initial iterations of the chains are discarded because of their dependence on starting states. In addition, the presence of autocorrelations within chain values is expected when latent variables are introduced in the model Chen et al. (2000), being thus important to perform autocorrelation analyzes.

In order to make fair comparisons between efficiency in the estimation of the parameters we use the notion of effective sample size (ESS) as defined in Kass et al. (1998) and the effective sample size per second (ESS/s). Sahu (2002) has used ESS/s to compare two different estimation approaches with the same model and points out that the ratio $EF_{12} = \frac{ESS/s_1}{ESS/s_2}$, can be considered as one measure of efficiency between different estimation approaches, such that large values of $EF_{12}$ indicates that estimation is more efficient under approach 1 than under approach 2. We propose using such index for comparing the efficiency in estimating different parameters within the same model and the same parameter in different models.

For each parameter, ESS is defined as the number of MCMC samples drawn, $B$, divided by the autocorrelation parameter time, $\gamma = 1 + 2\sum_{s=1}^{\infty} \rho_s$, where $\rho_s$ is the autocorrelation at lag $s$. In estimating $\gamma$, Sahu (2002) recommended using only the first (largest) autocorrelation estimate $(\widehat{\rho}_1)$ but we use instead $\widehat{\rho}_s$ with $s = 49$, since it is found that autocorrelation estimates between successive parameter draws become negligible for lags greater than 50 (see similar result to the two-parameter logistic (2PL) IRT model in Patz and Junker 1999, pg. 165.

We use this concept in studying efficiency in estimation for the skew-probit and probit IRT models. This concept can also be used for the comparison of the efficiency in the estimation of the parameters of the skew-probit IRT model when the *delta* or *lambda* parameterizations introduced in Section 3.2 are considered. Priors $a_j \sim N(1, 0.5)I(a_j > 0)$, $b_j \sim N(0, 2)$, and $\delta_j \sim U(-1, 1)$ [or $\lambda_j \sim T(0, 0.5, 2)$] are considered in such study. We consider, as in Spiegelhalter et al. (1996), initial values $a_j = 1$, and $b_j = 0$, $j = 1 \ldots, k$. Initial values for the penalization parameter $\delta_j$ can be randomly generated, but we prefer using fixed ones, and we propose as initial value, $\delta_j = 0$ (or $\lambda_j = 0$). Initial values for the latent variables $U_i$, $Z_{ij}$ and $V_{ij}$ are randomly generated by considering the corresponding distributions. We work, as in Sahu (2002), with run lengths of 5000 iterations after discarding 1000 initial iterations, so that the effective sample size is 4000.

Using the efficiency concept defined before on Table 1, Figure 2 depicts the error bar of the efficiency ($EF_{12}$) of each parameter for each two of the models considered. By considering the results on Table 1 and Figure 2, we note that:

- To any parametrization of the skew-probit model, the estimation is more efficient for the $u$ parameters, followed by the $a$ parameters, which are followed by the $b$ parameters, and, finally, $\delta$ following the $\lambda$ parameters.
- In the skew-probit model, the *delta parameterization* is more efficient than the *lambda parameterization*, for the estimation of all parameters.
- When a probit IRT model is appropriate for a data set (as is the case with the data set analyzed) it is noted to exist a reduction in the efficiency of the estimation of the $a$, $b$ and $u$ parameters when the skew-probit IRT model is considered, especially when using the *lambda parameterization* .

| IRT Model | Parameter | ESS | | ESS/s | |
|---|---|---|---|---|---|
| | | Mean | sd | Mean | sd |
| skew-probit | a | 260.7 | 125.9 | 3.7 | 1.8 |
| (delta parametrization) | b | 157.6 | 20.9 | 2.2 | 0.3 |
| | $\delta$ | 137.3 | 16.7 | 1.9 | 0.2 |
| | $\lambda$ | 121.4 | 13.7 | 1.7 | 0.2 |
| | u | 582.9 | 165.5 | 8.2 | 2.3 |
| | Total | 425.3 | 245.6 | 6.0 | 3.5 |
| skew-probit | a | 182.8 | 93.8 | 2.3 | 1.1 |
| (lambda parametrization) | b | 100.2 | 9.7 | 1.2 | 0.1 |
| | $\delta$ | 76.4 | 7.6 | 0.9 | 0.1 |
| | $\lambda$ | 71.1 | 8.4 | 0.9 | 0.1 |
| | u | 509.4 | 199.9 | 6.3 | 2.5 |
| | Total | 356.4 | 254.4 | 4.4 | 3.1 |
| probit | a | 307.9 | 127.5 | 9.1 | 3.7 |
| | b | 905.2 | 338.8 | 26.6 | 10.0 |
| | u | 783.5 | 236.5 | 23.0 | 7.0 |
| | Total | 741.9 | 287.7 | 21.8 | 8.5 |

Table 1: Performance of the parameter estimation for Tanners data set: mean and standard deviation (sd) of the estimated ESS, ESS/s, for each type of parameter for the skew-probit (*delta* or *lambda parameterization*) and probit IRT models over all the corresponding parameters.

Chen et al. (2000) mentioned that when the sample size $n$ is large ($n \geq 50$), slow convergence associated with Albert-Chib's algorithm (data augmentation approach) may occur. Slow convergence of the chain corresponding to the penalization parameter is detected, speciality with the *lambda parameterization*. Some algorithms to improve convergence of the Gibbs sampler in the second data augmentation approach are suggested in Chen et al. (2001) and must be explored in future works.

We have found the presence of autocorrelation in the skew-probit model, and hence, thin values up to 100 are recommended. The impact of the initial parameter vector and the Markov nature of the algorithm are reduced by having a burn in period and by thinning the chain.Thinning the chain is typically used to increase the independence of sequential values. Consequently, some authors as Jackman (2004) consider, with probit IRT model, a large number of iterations for inference based on the joint posterior density. The paper recommends running half a million iterations, retaining only every thousandth iteration so as to produce an approximately independent sequence of sampled values from the joint posterior density. However other authors as Gelman et al. (1996) consider that thinning is not necessary, that is, the whole sample can be used.

# 5 Applications

## 5.1 Application 1: Math Data set

Model checking or assessing the fit of a model implies seeking independent empirical evidence supporting some model prediction and it is a crucial part in any statistical analysis. In IRT modeling, model checking is an underdeveloped area and, moreover, no universal model-checking tools seems to be available in this field, especially from the classical point of view. This may be due to the complexity of the IRT models, which does not involve, for example, observed covariates but unobserved latent variables (or trait).

To compare IRT models including the skew-probit model, we computed the posterior expected deviance ($Dbar$), the deviance of the posterior means ($Dhat$), the effective number of parameters $\rho_D$ and the deviance information criterion ($DIC$), as presented in Spiegelhalter et al. (2002). $Dbar$ is the posterior mean of the deviance that is defined as $-2log(likelihood)$, $Dhat$ is a point estimate of the deviance obtained by substituting posterior means estimates for model parameters, $\rho_D$ is given by $\rho_D = Dbar - Dhat$ and $DIC$ is given by $DIC = Dbar + \rho_D = Dhat + 2\rho_D$. The model with the smallest DIC is expected to be the model that would best predict a replicated data set of the same structure as that currently observed. In the presence of auxiliary latent variables (in a data augmentation scheme) marginal DICs for the observed variables must be considered since the focus of the analysis is $p(\mathbf{y}|\mathbf{u}, \mathbf{a}, \mathbf{b}, \delta)$. DIC is a hierarchical modeling generalization of the AIC (Akaike Information Criterion) that has been proved to work somewhat efficiently in different complex models. The advantages of the DIC is that it reduces each model to a single number summary and the models to be compared need not be nested. Also DIC is readily available within WinBUGS for all models used. For other proposals see, for example, Sinharay and Johnson (2003) and Sahu (2002).

In addition, other important aspects of Bayesian modeling follows by examining the sensitivity of inferences to reasonable changes in the prior distribution and the likelihood (Gelman et al. 1996). This is specially valid when small samples are available.

In order to evaluate the sensitivity of the Bayesian estimation in the skew-probit IRT model by considering different priors to item parameters suggested in the Section 3.1, we conduct an analysis using a data set with questions from the Mathematical Test applied in Peruvian schools. Specifically, the data set corresponds to the application of 14 items of multiple choice of the Mathematical Test available for download at `http://www.minedu.gob.pe/umc/` applied to 131 students of high social-economical status.

The Mathematical 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.

Motivated by the asymmetric nature of the observed scores Figure (3, we explore the possibility of using the skew-probit IRT model for the Math Test data. However, it is fair to point out that, so far, there is no clear indication that asymmetrically distributed scores should imply in asymmetric ICC curves. It seems to us that even symmetric models can produce asymmetrically distributed scores, specially when $p_{ij}$ is close to zero (or one) and $n$ is not large. However, as shown in Chen et al. (1999), asymmetric ICCs seem adequate when the rate at which $p_{ij}$ approaches one (zero) is not the same as it approaches zero (one). Then, it is legitimate to consider a general specification to $p_{ij}$ by considering the skew-probit IRT model and obtain more information on some items by including the parameter $\delta$, if it is the case, or to obtain the confirmation that other items are correctly specified with symmetric IRT models as the probit IRT model.

In our point of view, external factors, such as economical status (which seems to be the case with the Math Test data), may play a role in conjunction with individual trace (or ability), which is hard to assess.

We present next a study on the fit of the parametric IRT models discussed earlier with the Math Data. When considering the model $p_{ij} = F(a_j u_i - b_j)$ with $F(.)$ as in the probit (P) or logit (L) links we have the two-parameters models. If $a_j = 1$, $j = 1, \ldots, n$ then we have the one-parameter models, and if we write $p_{ij} = c_j + (1 - c_j)F(a_j u_i - b_j)$, where $c_j$, $j = 1, \ldots, n$, is the guessing parameter, then we have three-parameters models. They are denoted, respectively, by 1P, 1L, 2P, 2L, 3P and 3L, when it is the case. Moreover, we implement the Bayesian approach for the logistic positive exponent model proposed by Samejima (2000). This model, which can produce asymmetric ICCs, is named here *generalized logistic model* and is denoted by 2GL. To the best of our knowledge, this seems the first attempt in fitting this model to a data set.

In order to develop a sensitivity study for the IRT models listed above we chose to work with the skewness parameter $\delta$ ($\delta$ parameterization) because it offers the same interpretation as the parameter $\lambda$ in a reduced scale and also is more efficient as showed in Section 4. We use two different combinations of prior distributions for the item parameters as suggested in Section 3.1. Hence, we have the following prior specifications:

Prior A: $a_j \sim N(1, 0.5)I(a_j > 0)$, $b_j \sim N(0, 2)$, $\delta_j \sim U(-1, 1)$.
Prior B: $a_j \sim N(1, 0.5)I(a_j > 0)$, $b_j \sim N(0, 2)$, $\delta_j \sim N(0, 1)I(-1, 1)$ .
Prior C: $a_j \sim LN(0, 0.5)$, $b_j \sim N(0, 2)$, $\delta_j \sim U(-1, 1)$.
Prior D: $a_j \sim LN(0, 0.5)$, $b_j \sim N(0, 2)$, $\delta_j \sim N(0, 1)I(-1, 1)$ .

Following suggestions in Sahu (2002), prior A was used with models 1P, 1L, 2P, 2L and 2GL. In addition, we consider $c_j \sim Beta(12.5, 37.5)$, where $Beta(r, s)$, denotes the beta distribution with parameters $r$ and $s$, which leads to more precise prior information, and hence the marginal densities of the parameters $c_j$ become more peaked and concentrated around the prior mean $r/(r+s) = 0.25$ that corresponds to the quantity $\frac{1}{4}$ in a multiple choice type test with 4 alternatives. It serves as a very good guess to the $c_j$ parameters and leads to greater efficiency than that with other elections of the hyperparameters $r$ and $s$ Sahu (2002).

With model 2GL we consider $e_j$ independent and $e_j \sim N(0, 1)I(e_j > 0)$. Also, for the sake of making comparisons, priors A and C are used with model 2P and the estimation is implemented by using an augmented likelihood, which is denoted by 2Pa. The skew-probit model is also implemented using the augmented likelihood. As remarked in Section 4, a large number of iterations must be generated with the skew-probit model due to the presence of high autocorrelations for some parameters so that the MCMC procedure becomes slow. For inference purposes and comparisons of all proposed models, we generated 202000 iterations and discard the 2000 initial ones. Using a thin of 100, an effective size of 2000 was considered. Estimates of model parameters are computed from these iterations. Several criteria computed using the CODA package, including the ones proposed by Geweke (1992) were used to evaluate convergence.

DIC values shown in Table 2 to the eight parametric IRT models in analysis seem to indicate that the skew-probit model (model (8)), improve any proposed model including the corresponding symmetric ones (1P, 1L, 2P,2L, 3P, 3L) and also the GL model ( models (1) to (7)) for any of the prior distributions considered. Hence, we expect that ICC estimates are more precise with the asymmetric probit models, especially when the $p_{ij}$'s are close to one (or zero). Note also that the best fit follows with priors A and C, which assign the uniform prior to $\delta$. Moreover, for the probit IRT model (model (4)), we observed that the MCMC approach based in the augmented likelihood took less time than the MCMC approach using the original likelihood (A similar result is observed in Sahu (2002), with model 3P). Therefore, for the SP IRT models we do not implement the MCMC approach using the original likelihood. Note that time of convergence with the skew-probit IRT model is about four times slower than that with the probit IRT model. However, the probit IRT model presents slower convergence than the *logistic positive exponent* model proposed by Samejima (1997, 2000), but both seem to converge faster than the 3L model, as implemented here. Note that $\rho_D$ is negative for the SP model. It can happen when the likelihood function is not log concave, which seems to be the case here.

| | Models | Type | Time | Parameters | *Dbar* | *Dhat* | $\rho_D$ | *DIC* |
|---|---|---|---|---|---|---|---|---|
| Symmetric | (1) | 1L | 53 | 145 | 1467 | 1372 | 94.46 | 1561 |
| | (2) | 1P | 42 | 145 | 1460 | 1340 | 120.3 | 1581 |
| | (3) | 2L | 78 | 159 | 1461 | 1378 | 82.74 | 1544 |
| | (4) | 2P | 71 | 159 | 1447 | 1359 | 88.41 | 1536 |
| | | 2Pa A prior | 26 | 159 | 1447 | 1359 | 88.32 | 1535 |
| | | 2Pa C prior | 27 | 159 | 1447 | 1356 | 91.65 | 1539 |
| | (5) | 3P | 87 | 173 | 1443 | 1356 | 87.8 | 1531 |
| | (6) | 3L | 148 | 173 | 1464 | 1384 | 80.08 | 1544 |
| Asymmetric | (7) | 2GL | 115 | 173 | 1465 | 1426 | 38.28 | 1503 |
| | (8) | SP prior A | 110 | 173 | 1328 | 1365 | -37.3 | 1290 |
| | | SP prior B | 105 | 173 | 1346 | 1364 | -17.78 | 1328 |
| | | SP prior C | 104 | 173 | 1335 | 1361 | -26.16 | 1308 |
| | | SP prior D | 98 | 173 | 1353 | 1359 | -6.527 | 1346 |

Time in seconds to run 2000 iterations in a Pentium IV with 1800 MHZ and 256 Ram.

Table 2: Results comparing the skew-probit with other parametric IRT models using DIC

We also obtained the quantity reported in the MC error column in Table 3 is an estimate of the Monte Carlo standard error of the mean ($\frac{\sigma}{\sqrt{n}}$), where the batch means method outlined in Roberts (1996) is used to estimate $\sigma$. Coefficients of variation index of the MC errors associated with the posterior means of the 14 $\delta$s was obtained. The corresponding values were 15.70, 15.88, 19.62 and 19.30 for A, B, C and D priors, respectively. By considering this coefficient, priors A and B showed to be more precise in estimating $\delta$. Hence, the truncated normal prior for the $a$ parameters seems to be more precise than the log-normal prior. In summary, prior A seems to present a better fit to the data set under consideration and also more precise parameter estimates. Estimates of item discrimination and difficulty parameters for the probit and skew-probit IRT models using prior A are presented in Figure 4. The two types of parameters are equally interpretable under both models: item 11 is the most discriminating while Item 9 is the least; also, item 11 is the easiest while item 12 is the most difficult. Note that the skew-probit IRT model offers the same general conclusions about difficulty and discrimination parameters as the probit IRT model.

Figure 5 depicts differences between probit and skew probit IRT models, which are expected when asymmetry parameter estimates are different from zero. This is the case for items 11 and 4, which present penalty parameter estimates ($\delta$ or $\lambda$) large and negative (negative asymmetry on the item characteristic function), while discrimination and difficulty estimates differ in the two models. In the special case of items 11 and 4, the difference between models as consequence of the asymmetry parameter affects the difficulty parameter. The other items present penalty parameter estimates around zero indicating that a probit IRT model is adequate for explaining their behavior.

In addition, we explore the situation when a less informative prior distribution is

| | Prior A | | Prior B | | Prior C | | Prior D | |
|---|---|---|---|---|---|---|---|---|
| | mean | MC error | mean | MC error | mean | MC error | mean | MC error |
| $\delta_1$ | -0.05 | 0.016 | -0.04 | 0.013 | -0.04 | 0.018 | -0.05 | 0.012 |
| $\delta_2$ | -0.04 | 0.015 | -0.05 | 0.015 | -0.03 | 0.016 | -0.03 | 0.011 |
| $\delta_3$ | 0.01 | 0.012 | 0.02 | 0.013 | 0.00 | 0.016 | -0.01 | 0.013 |
| $\delta_4$ | -0.12 | 0.017 | -0.10 | 0.019 | -0.13 | 0.018 | -0.11 | 0.019 |
| $\delta_5$ | -0.02 | 0.015 | -0.06 | 0.014 | -0.06 | 0.014 | -0.03 | 0.014 |
| $\delta_6$ | 0.02 | 0.014 | 0.01 | 0.012 | 0.01 | 0.011 | 0.00 | 0.012 |
| $\delta_7$ | -0.15 | 0.019 | -0.07 | 0.013 | -0.15 | 0.013 | -0.12 | 0.015 |
| $\delta_8$ | -0.08 | 0.014 | -0.08 | 0.017 | -0.07 | 0.018 | -0.09 | 0.013 |
| $\delta_9$ | -0.03 | 0.015 | -0.04 | 0.014 | -0.01 | 0.011 | -0.02 | 0.010 |
| $\delta_{10}$ | -0.01 | 0.017 | -0.04 | 0.011 | -0.06 | 0.016 | -0.04 | 0.012 |
| $\delta_{11}$ | -0.31 | 0.020 | -0.32 | 0.018 | -0.29 | 0.022 | -0.31 | 0.018 |
| $\delta_{12}$ | 0.01 | 0.012 | 0.01 | 0.014 | 0.01 | 0.012 | 0.01 | 0.014 |
| $\delta_{13}$ | -0.01 | 0.014 | -0.04 | 0.013 | -0.03 | 0.016 | -0.03 | 0.014 |
| $\delta_{14}$ | -0.09 | 0.018 | -0.07 | 0.013 | -0.07 | 0.014 | -0.07 | 0.012 |

Table 3: Estimates to the posterior means and MC errors for $\delta$ (item) parameters under the four prior scenarios in the skew-probit IRT model

considered for the difficulty parameters by considering prior E: $a_j \sim N(1, 0.5)I(a_j > 0)$, $b_j \sim N(0, 10000)$, $\delta_j \sim U(-1, 1)$. In this case, the observed DIC was 1261 but the observed variability coefficient index for the the MC error was 16.67. On a preliminary analysis, **?**, using prior scenarios A and E specified for parameters $a$ and $b$ in the probit IRT model, indicated that the probit IRT model is insensitive to the prior specification. This result is not observed for the skew-probit IRT model, which is sensitive to the specification of more diffuse priors to $b$. It was also noted that with priors A and E, estimates for parameters $a$ and $b$ present somewhat strong correlations, but, estimates for $\delta$ (or $\lambda$) under the two priors are much less correlated. Moreover, estimates for $\delta$ (or $\lambda$) seem to depend more closely on the prior considered.

## 6 Application 2: Weight Data set

In this application we report results of a study involving a data set with 2141 female teens and 15 items. The data set corresponds to a weight's perception scale applied in an epidemiological study about dietary disturbances in Metropolitan Lima (Per) (see Martínez et al. 2003). We use the same final priors considered in the Math Data set. It takes Winbugs about 157 seconds to run 1000 iterations in an Pentium IV with 1800 MHZ and 256 memory Ram. As with the Math Data set, we implement the probit and skew-probit IRT models by considering the same $A$ (final) priors. For inference purposes and comparisons of all proposed models, we generated 22000 iterations, discarding the 2000 initial ones. Using a thin of 10, an effective size of 2000 was considered. Estimates of model parameters are computed from these iterations. Several criteria computed using the CODA package, including the ones proposed in Geweke (1992) were used to evaluate convergence.

The DIC values for the PN and SPN are 31028.8 and 28315.5, respectively, indicating the skew-probit IRT model as the model presenting the best fit. Hence, we expect that ICC estimates are more precise with the asymmetric probit models. The values of $\rho_D$ are 1689.46 and 2322.430, respectively, for the PN and SPN models. We found that only the item 7 has a significant asymmetry (Figure 6). Also, for this item, clearly the ICCs for the probit and skew-probit models are different. Figure 7 depicts the ICCs considering the PN and SPN models to these items, clearly showing differences among them.

## 7 Final discussion

The type of asymmetry considered in this paper is associated with the probabilities $p_{ij}$ and not with the latent variable because we are motivated to assume that some probabilities of correct response to several individuals in different items don't show symmetry around 0.5, independently of the distributional shape of the trait.

In the literature, different forms have been considered to specify skewness in IRT models: a) modifying the distributional shape of the trait $u$, that is, consider that $u$ takes an asymmetric distribution function or assuming a nonlinear function in $u$ as, for example, a power function such as $\beta_0 + \beta_1 u + \beta_2 u^2 + \beta_3 u^3$ (Fleishman 1978), or b) considering that $F(.)$ corresponds to an asymmetric distribution function as proposed by Samejima (1997, 2000). However, the effect of the distributional shape of the trait $u$ on item parameter estimation is unclear. Kirisci et al. (2001) found that the main effects of the distribution of $U$ was not significant for all parameters investigated in the three-parameters logistic model contradicting the results of Ree (1979), Stone (1992) and Swaminathan and Gifford (1983). They argue that, applied to the case they study, the effect of non-normality of $u$ in item estimation is minimized for longer tests and larger samples.

This article presents a new asymmetric IRT model by considering a new asymmetric ICC when the cumulative distribution of the standard skew-normal distribution (Azzalini 1985) is considered. As such, it extends the work of Albert (1992) for asymmetric IRT models, including the symmetric probit-normal model as a special case. Moreover, a general representations for the likelihood of the data was also provided, which seems not to be the case with other symmetric models in the literature. We introduced asymmetry in the two-parameter (2P) IRT model and not in the three-parameter (3P) IRT model, because several generalizations and applications are possible for the 2P model (see, for example, Bradlow et al. 1999, Béguin and Glas 2001, Fox and Glas 2001, Linardakis and P. Dellaportas 2002, Fox and Glas 2003, and Jackman 2004). Moreover, important routines in Bayesian inference are available for the 2P model in the web (Martin and Quinn 2006; Johnson and Albert 1999; Jackman 2004). However, it seems to us that it is possible to extend the results in the paper for the (3P) model and to a skew-Rasch item response model. Extensions for testlet type models (Bradlow et al. 1999) seem also to be straightforward. We expect that the proposed SP models can be used also in clinical contexts since probabilities near zero and one are expected in latent traces that are not cognitive.

Two data augmentation approaches are proposed for implementing Bayesian estimation by using the MCMC methodology in the SP model. When using data augumentation it is likely to have high autocorrelations, which may require large number of iterations for adequately assessing variability of parameter estimates (sample means, sample modes, quantiles). This can be achieved by carefully subsampling from the Markov chains, which also is a useful tool to minimize storage requirements. The notions of effective sample size (ESS), as defined in Kass et al. (1998), and the sample size per second (ESS/s), as considered in Sahu (2002), are considered in studying estimation efficiency, specially in the presence of high autocorrelation. As has been demonstrated by several researches (Swaminathan et al. 2003), accurate estimation of item parameters in small samples, particularly in the two and three parameters models, can only be adequately accomplished through a Bayesian approach.

Comparisons of symmetric and asymmetric IRT models are presented by using the deviance information criterion (DIC) described in Spiegelhalter et al. (2002). Finally, from the point of view of the test designer, the presence of a new item parameters that can explain the asymmetric behavior of the ICC in terms of variations on the probability of success for different ability levels can be used on the development of more precise tests for the estimation of examinees ability.

The skew link proposed in this paper is related with the skewed- probit link proposed by Chen et al. (1999), where it is considered the skew-normal distribution defined in a more general context in citetBrancoDey and Sahu et al. (2003). A common aspect of these two links is that the asymmetry parameter is associated with the distribution of $F(.)$ and not with the linear predictor. However the formulation of both models is different, since a latent linear structure is not used in the definition of the link function proposed, as is the case with the approach in Chen et al. (1999). In addition, the link function we propose presents a closed form expression for the cumulative distribution function $F(.)$.

Finally, the results in the Proposition 3 is general and can be extended to mixed models for binary response (de Boeck and Wilson 2004) by considering $m_{ij} = \mathbf{x}'_{ij}\beta + \mathbf{t}'_{ij}\mathbf{b}_i$ where $i = 1, \ldots, n$, denote subjects (level-2 units) and $j = 1, \ldots, n_i$, are repeated observations (level-1 units) nested within each subject, with $t_{ij}$ denoting the $r \times 1$ vector of covariates associated with the random effects $b_i$, $\mathbf{x}_{ij}$ denotes the $p \times 1$ vector of covariates or fixed effect and $\beta$ is a vector of regression coefficients. In addition, if we write the latent stochastic representation of the link function as $Z_{ij} = m_{ij} + e_{ij}$, where $e_{ij}$ is a random error, assumed to be distributed according to the skew-normal distribution with parameters vector $\theta = (\mu, \sigma^2, \lambda)$, where $\mu$ is the location parameter, $\sigma^2$ is the scale parameter and $\lambda$ is the asymmetry parameter, we obtain a generalized skew-probit link that has as a particular case the skew-probit link due to Chen et al.

(1999) when $\mu = 0$ and $\sigma^2 = 1 + \lambda^2$. This extension is discussed in detail in Bazán et al. (2005).

# Appendix

## A0: Properties of the standard skew-normal distribution

Considering $Z \sim SN(\lambda)$, the following properties are readily established (see Azzalini, 1985 and Henze, 1986):

1. The mean and variance of $Z$ are given, respectively, by
$$E[Z] = \sqrt{\frac{2}{\pi}}\delta \;\text{ and }\; Var[Z] = 1 - \frac{2}{\pi}\delta^2.$$
where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1, 1]$;

2. The skewness and kurtosis indexes of $Z$ are given, respectively, by
$$\gamma = \frac{\sqrt{2}(4-\pi)}{\pi\sqrt{\pi}} sign(\delta)(1 - 2\frac{\delta^2}{\pi})^{-3/2} \qquad \text{and } \;\kappa = \frac{8(\pi-3)\delta^4}{(\pi - 2\delta^2)^2},$$
implying that $-0.9953 < \gamma < 0,9953$ and $0 < \kappa < 0.8692$, where $sign(.)$ is a sign function taking value 1 when $\delta$ is positive and $-1$ otherwise;

3. The density of $Z$ is log concave;

4. An important stochastic representation was presented by Henze (1986). Let $V \sim HN(0, 1)$ and $W \sim N(0, 1)$ be the standard half normal and the standard normal distributions, respectively. If $V$ and $W$ are independent random variables, then the marginal distribution of $Z = \delta V + (1 - \delta^2)^{1/2} W$ is $SN(\lambda(\delta))$, with $\lambda = \frac{\delta}{(1-\delta^2)^{1/2}}$.

5. The stochastic representation given in 4. can be rewritten hierarchically by considering that the distribution of $Z|V$ ($Z$ given $V$) is a normal distribution with mean $\delta v$ and variance $1-\delta^2$, resulting in the following hierarchical representation: $Z|V \sim N(\delta v, 1 - \delta^2)$ and $V \sim HN(0, 1)$, leading to the marginal distribution of $Z$ given above.

## A1: Program

We describe in the sequel a program in Winbugs used to implement the data augumentation approach described in the paper.

```
model{
for (i in 1:n) { for (j in 1:k) {
    m[i,j] <- a[j]*u[i] - b[j]
```

```
        muz[i,j]<-m[i,j]-delta[j]*V[i,j]
   Zs[i,j] ~ dnorm(muz[i,j],preczs[j])I(lo[y[i,j]+1],up[y[i,j]+1])
   V[i,j] ~ dnorm(0,1)I(0,)          }
                   }
    #priors for item parameters
     for (j in 1:k) {
    #A prior: Sahu (2002)
          a[j] ~ dnorm(1,2)I(0,)
          b[j] ~ dnorm(0,0.5)
          delta[j]  ~ dunif(-1,1)
    #B prior : Sahu (2002)
        #a[j] ~ dnorm(1,2)I(0,)
        #b[j] ~ dnorm(0,0.5)
        #delta[j] ~ dnorm(0,1)I(-1,1)
    #prior C: Patz e Junker (1999)
        # a[j] ~ dlnorm(0,2)
        # b[j] ~ dnorm(0,0.5)
        # delta[j]  ~ dunif(-1,1)
    #prior D Patz e Junker (1999)
        # a[j] ~ dlnorm(0,2)
        # b[j] ~ dnorm(0,0.5)
        # delta[j] ~ dnorm(0,1)I(-1,1)

    preczs[j]<- 1/(1-pow(delta[j],2))
    lambda[j]<-delta[j]*sqrt(preczs[j])
           }
#latent variable prior
          for(i in 1:n){ u[i]~ dnorm(0,1)   }

lo[1]<- -50; lo[2]<- 0 ## Zs*|y=0~N(-delta*V+m,1-delta^2)I(-50,0)
up[1]<- 0;  up[2]<-50  ## Zs*|y=1~N(-delta*V+m,1-delta^2)I(0,50)

           mu<-mean(u[])
           du<-sd(u[])
        }

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),
delta=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0))

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

# References


Albert, J. H. (1992). "Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling." *Journal of Educational Statistics*, 17: 251–269. 861, 865, 866, 867, 868, 878

Albert, J. H. and Chib, S. (1993). "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association*, 88: 669–679. 867

Albert, J. H. and Ghosh, M. (2000). "Item response modeling." In D. Dey, S. G. and Mallick (eds.), *Generalized Linear Models: A Bayesian Perspective*, 173–193. New York: Addison–Wesley. 861, 866, 869

Azzalini, A. (1985). "A class of distributions which includes the normal ones." *Scandinavian Journal of Statistics*, 12: 171–178. 862, 864, 867, 878

— (2005). "The Skew-normal Distribution and Related Multivariate Families." *Scandinavian Journal of Statistics*, 32: 159–188. 862

Bazán, J., Bolfarine, H., and Branco, M. (2004b). "A new family of asymmetric models for Item Response Theory: a Skew-Normal IRT Family." *Department of Statistics, University of São Paulo*, (Technical Report RT-MAE-2004-17). 870

— (2005). "A general skew-probit link for binary response." In *Proceedings of the 9th School of Regression Models*, 267–281. São Pedro, Brazil: Associação Brasileira de Estatstica. 880

Béguin, A. A. and Glas, C. A. W. (2001). "Estimation of multidimensional IRT models." *Psychometrika*, 66: 541–562. 878

Birnbaum, A. (1968). "Some latent trait models and their use in inferring an examinee's ability." In Lord, F. M. and Novick, M. R. (eds.), *Statistical Theories of Mental Test Scores*, 395–479. Massachusetts: Addison Wesley. 861

Bock, R. D. and Aitkin, M. (1981). "Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm." *Psychometrika*, 46: 443–459. 865

Bradlow, E., Wainer, H., and Wang, X. (1999). "A Bayesian Random Effects Model for Testlets." *Psychometrika*, 64: 153–168. 869, 878, 879

Branco, M. and Dey, D. K. (2002). "Regression model under skew elliptical error distribution." *The Journal of Mathematical Sciences*, 1: 151–169. 863

Chen, M.-H., Dey, D. K., and Shao, Q. M. (1999). "A New Skewed Link Model for Dichotomous Quantal Response Data." *Journal of the American Statistical Association*, 94: 1172–1186. 862, 863, 874, 879

— (2001). "Bayesian Analysis of Binary Data Using Skewed Logit Models." *Calcutta Statistical Association Bulletin*, 51: 201–202. 873

Chen, M.-H., Shao, Q. M., and Ibrahim, J. G. (2000). *Monte Carlo methods in Bayesian Computation*. New York: Springer-Verlag. 871, 872

Chib, S. and Greenberg, E. (1995). "Understanding the Metropolis-Hastings Algorithm." *The American Statistician*, 49: 327–335. 870

Czado, C. and Santner, T. J. (1992). "The Effect of Link Misspecification on Binary Regression Inference." *Journal of Statistical Planning and Inference*, 33: 213–231. 862

Dalla-Valle, A. (2004). "The skew-normal distribution." In Genton, M. G. (ed.), *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, 3–24. CRC, Boca Raton: Chapman & Hall. 864

de Boeck, P. and Wilson, M. (eds.) (2004). *Explanatory Item Response Models : A Generalized Linear and Nonlinear Approach*. New York: Springer-Verlag. 879

Fleishman, A. (1978). "A method for simulation nonnormal distributions." *Psychometrika*, 43: 521–532. 878

Fox, J. P. and Glas, C. A. W. (2001). "Bayesian Estimation of a Multilevel IRT Model Using Gibbs Sampling." *Psychometrika*, 66: 269–286. 868, 878

— (2003). "Bayesian Modeling of Measurement Error in Predictor Variables Using Item Response Theory." *Psychometrika*, 68: 169–191. 868, 878

García-Pérez, M. (1999). "Fitting logistic IRT models: small wonder." *The Spanish Journal of Psychology*, 2: 74–94. 862

Gelman, A., Meng, A., and Stern, H. S. (1996). "Posterior predictive assessment of model fitness via realized discrepancies." *Statistica Sinica*, 6: 733–807. 873, 874

Genton, M. G. (ed.) (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. CRC Boca Raton:: Chapman & Hall. 862

Genz, A. (1992). "Numerical Computation of Multivariate Normal Probabilities." *Journal of Computational and Graphical Statistics*, 1: 141–150. 864

— (1993). "Comparison of Methods for the Computation of Multivariate Normal Probabilities." *Computing Sciences and Statistics*, 25: 400–405. 864

Geweke, J. (1992). "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments." In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds.), *Bayesian Statistics. Proceedings of the Fourth Valencia International Meeting*, volume 4, 169–193. Oxford: Clarendon Press. 875, 878

Ghosh, M., Ghosh, A., Chen, M.-H., and Agresti, A. (2000). "Noninformative priors for one parameter item response models." *Journal of Statistical Planning and Inference*, 88: 99–115. 869

Jackman, S. (2004). "Bayesian Analysis for Political Research." *Annual Review of Political Science*, 7: 483–505. 873, 878

Johnson, V. and Albert, J. (1999). *Ordinal data modeling*. New York: Springer-Verlag. 869, 870, 878

Junker, B. W. and Ellis, J. L. (1997). "A characterization of monotone unidimensional latent variable models." *Annals of Statistics*, 25: 1327–1343. 867

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. (1998). "Markov Chain Monte Carlo in Practice: A Roundtable Discussion." *The American Statistician*, 52: 93–100. 861, 871, 879

Kim, S., Cohen, A. S., Baker, F. B., Subkoviak, M. J., and Leonard, T. (1999). "An investigation of hierarchical Bayes procedures in item response theory." *Psychometrika*, 59: 405–421. 869

Kirisci, L., Hsu, T., and Yu, L. (2001). "Robustness of Item Parameter Estimation Programs to Assumptions of Unidimensionality and Normality." *Applied Psychological Measurement*, 25: 146–162. 878

Linardakis, M. and P. Dellaportas, P. (2002). "An approach to multidimensional item response modeling." In *Bayesian Methods with Applications to Science Policy and Oficial Statistics*, 331–340. 878

Martin, A. D. and Quinn, K. M. (2006). "Markov chain Monte Carlo (MCMC) Package, R." Technical report. 878

Martínez, P., Zusman, L., Hartley, J., Morote, R., and Calderón, A. (2003). "Estudio epidemiológico de los trastornos alimentarios y factores asociados en Lima Metropolitana." *Revista de Psicologa PUCP*, 21: 233–269. 877

Parrish, R. and Bargmann, R. (1981). "A Method for Evaluation of Cumulative Probabilities. of Bivariate Distributions Using the Pearson Family." *Statistical Distributions in Scientific Work*, 5: 241–257. 864

Patz, R. J. and Junker, B. W. (1999). "A straightforward approach to Markov chain Monte Carlo methods for item response." *Journal of Educactional and Behavioral Statistics*, 24: 146–178. 863, 866, 868, 869, 871

Prentice, R. L. (1976). "A generalization of the probit and logit methods for dose-response curves." *Biometrika*, 32: 761–768. 862

Ree, M. J. (1979). "Estimating item characteristic curves." *Applied Psychological Measurement*, 3: 371–385. 878

Roberts, G. O. (1996). "Markov chain concepts related to sampling algorithms." In W. R. Gilks, S. R. and Spiegelhalter, D. J. (eds.), *Markov chain Monte Carlo in Practice*, 45–58. CRC, Boca Raton: Chapman & Hall. 876

Rupp, A., Dey, D., and Zumbo, B. (2004). "To Bayes or Not to Bayes, From Whether to When: Applications of Bayesian Methodology to Modeling." 11: 424–451. 862, 868

Sahu, S. K. (2002). "Bayesian estimation and Model Choice in Item Response Models." *Journal of Statistical Computation and Simulation*, 72: 217–232. 861, 869, 871, 873, 875, 879

Sahu, S. K., Dey, D. K., and Branco, M. (2003). "A New Class. of Multivariate Skew Distributions with Applications to. Bayesian Regression Models." *The Canadian Journal of Statistics*, 29: 217–232. 879

Samejima, F. (1997). "Departure from normal assumptions: A promise for future psychometrics with substantive mathematical modelling." *Psychometrika*, 62:4: 471–493. 862, 863, 875, 878

— (2000). "Logistic positive exponent family of models: Virtue of asymmetric item characteristic curves." *Psychometrika*, 65:3: 319–335. 862, 863, 875, 878

Sinharay, S. and Johnson, M. S. (2003). "Simulation studies applying posterior predictive model checking for assessing fit of the common item response theory models." *Educational Testing Service*, (ETS RR-03-26). 863, 873

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linden, A. (2002). "Bayesian measures of model complexity and fit." *Journal Royal Statistical Society*, 64:3: 583–639. 864, 873, 879

Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W. R. (1996). *BUGS 0.5 Examples Volume 1*. Cambridge: UK: MRC. Biostatistics Unit. 871

Stone, C. A. (1992). "Recovery of Marginal Maximum Likelihood Estimates in the Two-Parameter Logistic Response Model: An Evaluation of MULTILOG." *Applied Psychological Measurement*, 16: 1–16. 878

Swaminathan, H. and Gifford, J. (1983). "Estimation of parameters in the three-parameter latent trait model." In Weiss, D. J. (ed.), *New horizons in testing: Latent Trait Theory and Computerized Adaptive Testing*, 13–30. Paris: Academic Press. 878

— (1985). "Bayesian estimation in the two-parameter logistic model." *Psychometrika*, 50: 349–364. 869

Swaminathan, H., Hambleton, R. K., Dehui, X., and Saba, M. R. (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*, 27:1: 27–51. 863, 879

Tanner, M. (1996). *Tools for Statistical Inference*. New York: Springer-Verlag. 861, 871

van der Linden, W. J. and Hambleton, R. K. (1997). *Handbook of Modern Item response Theory*. New York: Springer-Verlag. 861

**Acknowledgments**

We thank Arturo Calderón for providing the Weight data set. Authors also acknowledges partial financial support from CNPq and Capes-Brasil. We are grateful to an Associate Editor for valuable comments and suggestions.

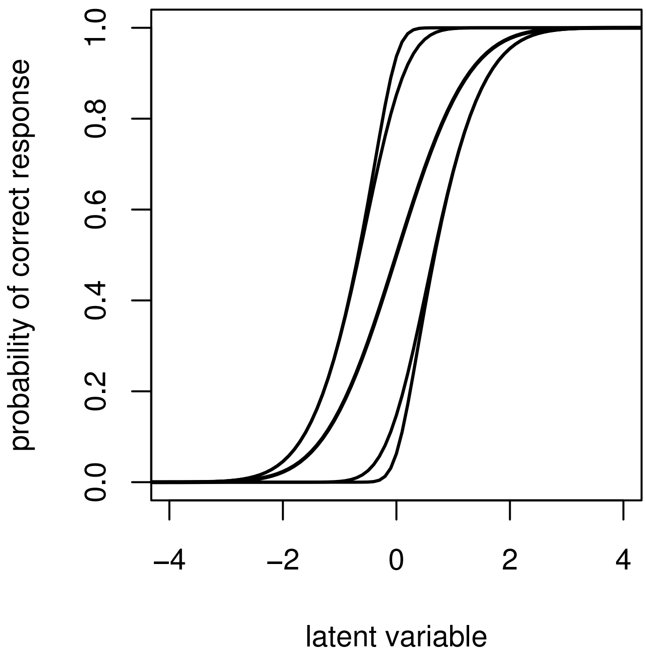



Figure 1: ICCs for different values of the latent variable associated with individual ability $U$, and considering item parameters $a = 1, b = 0$ and $\lambda = -2, -1, 0, 1, 2$ in the skew-probit IRT model. Note that when $\lambda > 0$, the probability of success has a slow growth for low values of the latent variable $U$. On the other hand, when $\lambda < 0$, the probability of success has a quick growth for low values of the latent variable $U$.

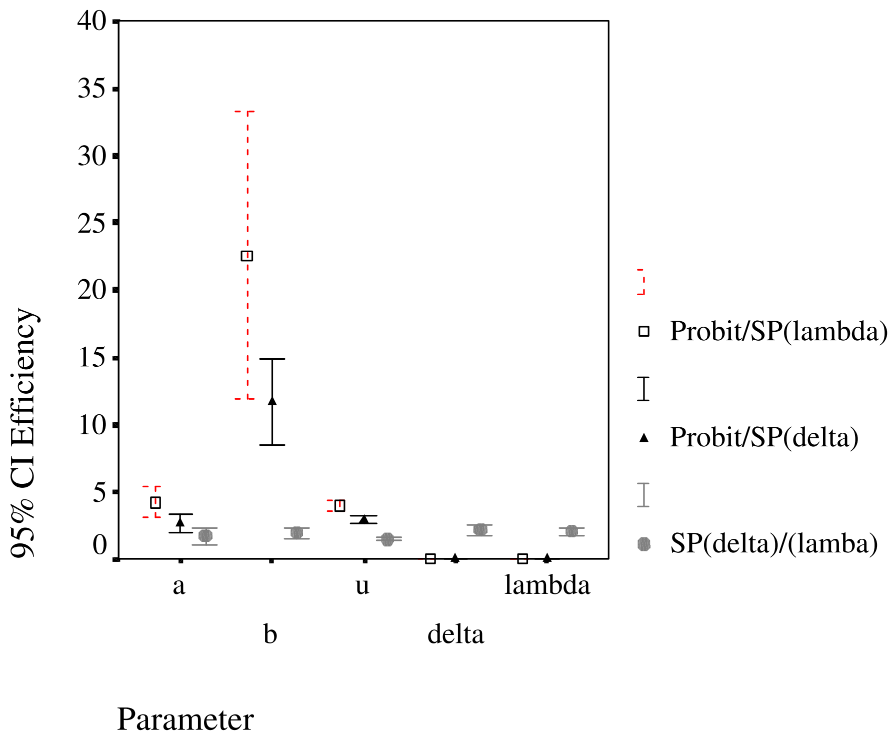



Figure 2: Performance of parameter estimates for Tanners data, using 95 % confidence intervals for the efficiency ($EF_{12}$) between probit an skew-probit models with $\lambda$ (SP(lambda)) and $\delta$ (SP(delta)) parameterizations.

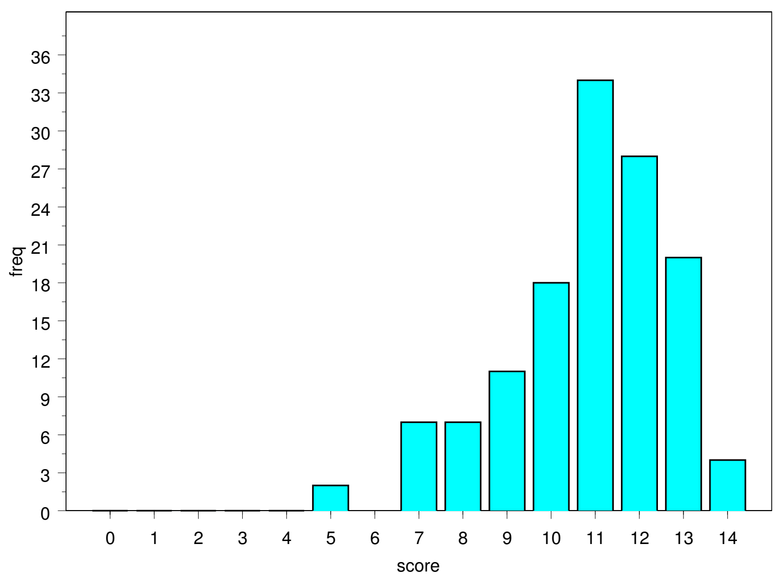
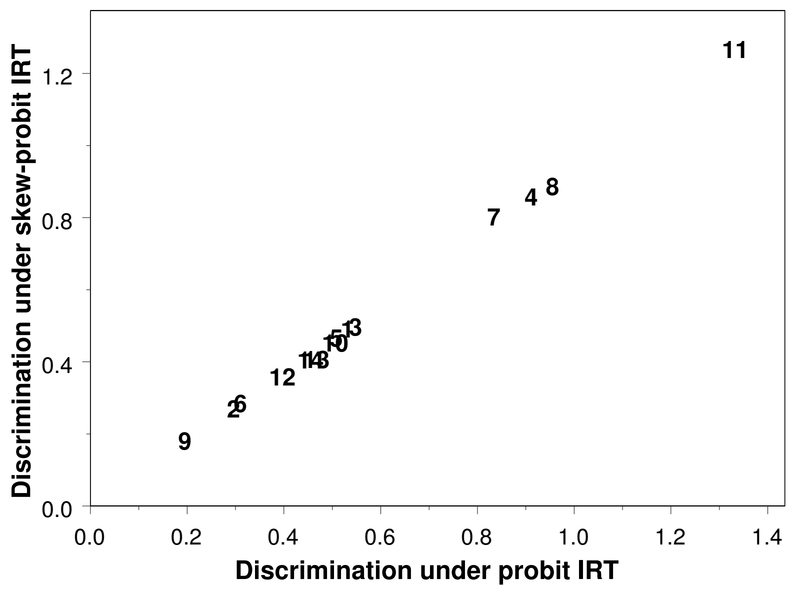



Figure 3: Histogram for the observed scores of the 131 sixth grade students for a mathematical test (Math Data) with $k$=14 items. The sample mean value is 10.84, the standard deviation is 0.449 and the sample skewness index is $-0.804$).



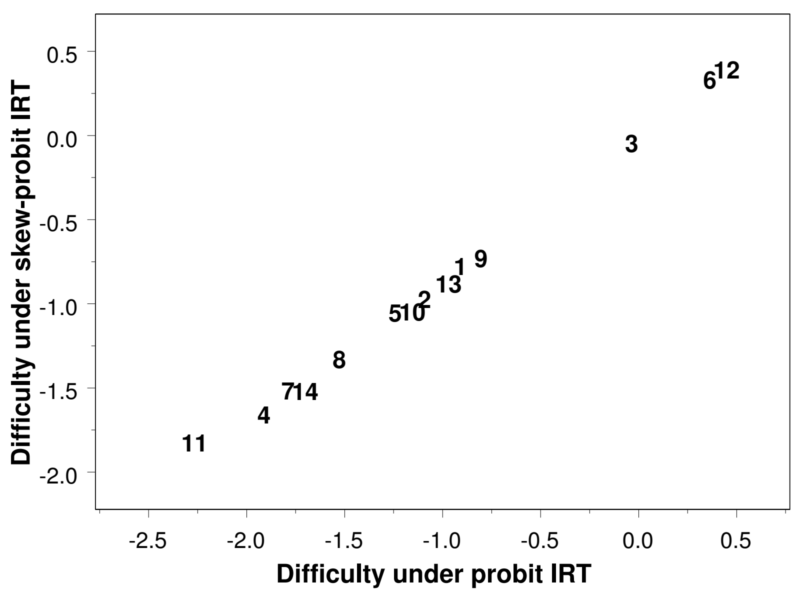




Figure 4: Discrimination and difficulty parameter estimates under probit and skew-probit IRT models. Note that the parameters are equally interpretable under both models: item 11 is the most discriminating while item 9 is the least; also, item 11 is the easiest while item 12 is the most difficult.

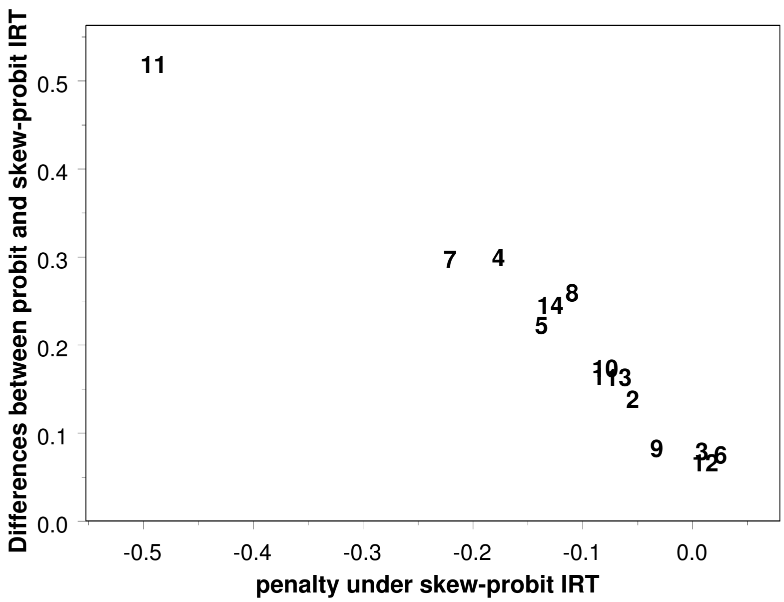



Figure 5: Differences between the probit and skew-probit IRT models (sum of absolute values of differences in $a_j$ and $b_j$ parameters) for the estimated $\delta$ parameter. As expected, the estimated difficulty and discrimination parameters in the probit and skew-probit IRT models are approximately equal when the asymmetry parameter is close to zero

.

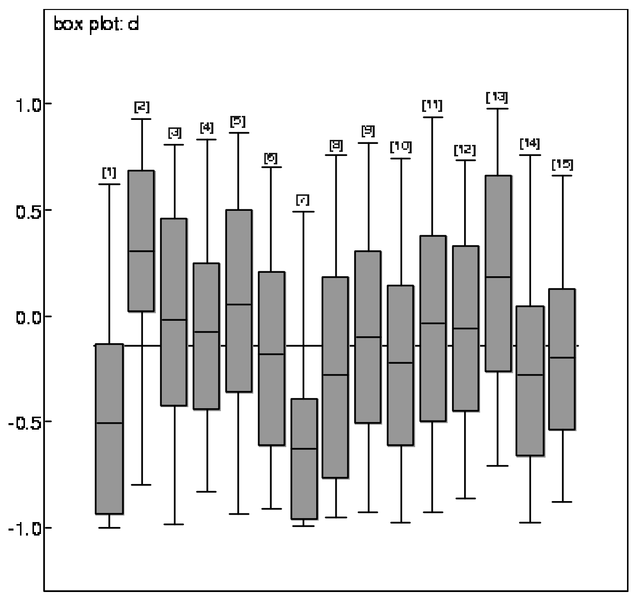



Figure 6: Box-Plots for the $\delta$ parameters of the 15 items weight's perception scale under the skew-probit IRT model

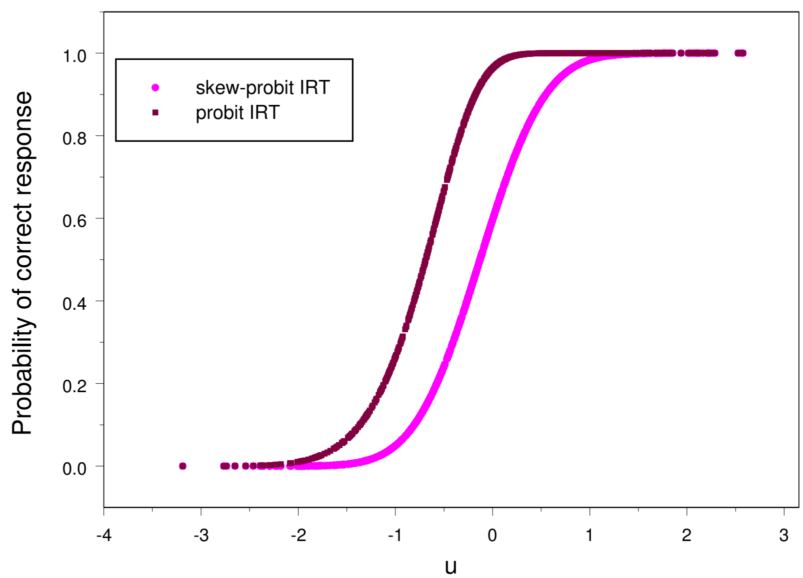



Figure 7: ICCs for item 7 for the weights data set, under probit and skew-probit IRT models

.