# A new family of asymmetric models for item response theory:

# A Skew-Normal IRT family

Jorge Luis Bazán, Heleno Bolfarine, Marcia D'Elia Branco

Department of Statistics

University of São Paulo

October 04, 2004

Correspondence should be sent to Marcia D´Elia Branco E-mail: mbranco@ime.usp.br

Caixa postal 66281, CEP 05315-970

Sâo Paulo - S.P. - Brazil.

# A new family of asymmetric models for item response theory: A Skew-Normal IRT family

Abstract

Normal assumptions for the latent variable and symmetric item characteristics curves have been used in the last 50 years in many psychometric methods for item-response theory (IRT) models. This paper introduces a new family of asymmetric models for item response theory, namely the skew-normal item- response theory (SN-IRT) model. This family extends the ogive normal (symmetric probit-normal) model by considering: a) an accumulated skew-normal distribution for the item characteristic curve and b) skew-normal distributions are assumed as priors for latent variables for modeling individuals' ability. Four models compose the SN-IRT family: skew-probit-skew-normal, skew-probit-normal, probit-skew-normal and probit-normal models as a particular case. Hence, the SN-IRT is a more flexible model for fitting data sets with dichotomous responses. Bayesian inference methodology using two data augmentation approaches for implementing the MCMC methodology is developed. Model selection between symmetric and asymmetric models is considered by using the deviance information criterion (DIC) and the expected AIC and expected BIC and by using latent residuals. The proposed penalization (asymmetry) parameter is interpreted in the context of a particular data set related to a mathematical test. Suggestions for use the in news applications of skew probit propose in the paper are discussed.

Key words: skew probit link, item response theory, Bayesian estimation, normal ogive model, skew-normal distribution, binary regression

The Item Response Theory (IRT) for multivariate dichotomous responses resulting from $n$ individuals evaluated in a test with $I$ items, considers a latent variable $U$ that explains (is associated to) individuals ability and a set of parameters associated to the items under consideration. Different characterizations have been considered in the literature over the last 40 years regarding the status of the latent variables (Borboom et al., 2003), such as latent parameter and latent variable; with Bayesian and classical interpretations (see Rupp et al., 2004). In this paper we consider the characterization due to Holland & Rosembaum (1986) and Bartholomew & Knoot (1999) and the Bayesian interpretation given by Albert (1992).

IRT models analyze the probability of correctly answering the items of a test as a function of a linear relationship involving item parameters and examinees abilities. Formally, let $y_{ij}$ the dichotomous (or binary) response corresponding to the i*th* individual, $i$=1,...,$n$, on the j*th* item, $j$=1,...I, which takes the value 1 if the response is correct and 0 otherwise. It is considered that: $Y_{ij} \sim Bern(p_{ij})$ (Bern: Bernoulli distribution) with $p_{ij} = P(Y_{ij} = 1 \mid u_i, \eta_j)$ being the probability that the i*th* examine is able to answer the j*th* item correctly. In fact, $p_{ij}$ is the conditional probability of correct response given the i*th* ability value $u_i$ and j*th* item parameters $\eta_j = (\alpha_j, \beta_j)$. It is considered that $p_{ij} = F(m_{ij})$, where, the function $F$ links the probability $p_{ij}$ with the linear function $m_{ij} = \alpha_j (u_i - \beta_j)$. In the literature $F$ is known as the item response curve or the item characteristic curve (ICC), it is common for all $i$, and all $j$, and satisfy the latent monotonicity property (Holland and Rosembaum, 1986). Moreover, the IRT model satisfies the latent conditional independence principle (known as "independence local" in the psychometric literature), which considers that for the i*th* examinee $Y_{ij}$ is conditionally independent given $u_i$,. It is also considered that responses from different individuals are also independent. Considering the above assumptions for the IRT model, the multivariate joint density of $\boldsymbol{Y} = (\boldsymbol{Y}_1^{\boldsymbol{T}}, \cdots, \boldsymbol{Y}_{\boldsymbol{n}}^{\boldsymbol{T}})^{\boldsymbol{T}}$, with $\boldsymbol{Y}_{\boldsymbol{i}}^{\boldsymbol{T}} = (\boldsymbol{Y}_{i1}, \cdots, \boldsymbol{Y}_{\boldsymbol{iI}})$, given the vector of latent variables $\boldsymbol{u}^{\boldsymbol{T}} = (u_1, \cdots, u_n)^T$ and the item parameter vector $\boldsymbol{\eta}^{\boldsymbol{T}} = (\eta_j, \cdots, \eta_I)$ can be written as

$$\boldsymbol{p}(\boldsymbol{Y} \mid \boldsymbol{u}, \eta) = \prod_{i=1}^{n} \prod_{j=1}^{I} \left[\boldsymbol{F}(\boldsymbol{m}_{ij})\right]^{Y_{ij}} \left[1 - \boldsymbol{F}(\boldsymbol{m}_{ij})\right]^{1-Y_{ij}}, \; i = 1, \ldots, n, \; j = 1, \ldots, I \qquad (1)$$

An assumption added to the model is that $U_i \sim N(\mu, \sigma^2)$, which establishes that the latent variables associated with the individuals taking the test are *well behaved* and that their abilities are a random sample from this distribution (Albert, 1992). Specification of values for $\mu$ and $\sigma^2$ (as in Albert (1992), who considers $\mu = 0$ and $\sigma^2 = 1$) or specification of distributions for $\mu$ and $\sigma^2$ (as in Patz & Junker, 1999) solves the identifiability problem for the IRT model. The model is not identifiable since it is possible to preserve the model (likelihood) by conveniently transforming the parameters. As also pointed out in Albert & Ghosh (2000) this model involves $n + 2I + 2$ unknown parameters which means that it is overparameterized. Other characteristics of IRT models are listed, for example, in Albert & Ghosh (2000), Rupp et al. (2004), Patz & Junker (1999) and Bazán et al. (2004).

Item-Response Theory is a set of models used for modelling variables related to human behavior, replacing observed scores for the items of a test. Since these variables are not observable, they are assumed as latent variables. Statistical assumptions in modeling academic achievement and other variables associated with human behavior are based on the normality assumption of the distribution of the scores. Several authors have questioned this assumption (see Samejima, 1997 and Micceri, 1989) since it is somewhat restrictive for modelling human behavior. Micceri (1989) presents examples of situations where the latent variables can be assumed not normally distributed. In an investigation of the distributional characteristics of 440 large-sample achievements and psychometric measures, Micceri (1989) found that 15.2% of the distributions had both tails with weights at or about the Normal, 49.1% of the distributions had at least one extremely heavy tail, and 18% had both tail with weights less than the normal distribution. Among ability measures, the percentages were similar with 19.5% having both tails with weights at or about the Normal, 57.6% having at least one heavy tail, and 22.9% having both tails less than the Normal. Amongst psychometric measures, 13.6% had tail weights near the Normal, 65.6% had at least one moderately heavy tail, and 20.8% had both tail weights less than the Normal.

Lord considered the first IRT model in 1952, in which $p_{ij} = F(m_{ij}) = \Phi(m_{ij}.)$ where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution. This model is known in the literature as the ogive normal model. In 1968 Birbaum introduced the logistic model, in which $F(.) = L(.)$ is the cumulative distribution of the standard logistic distribution, known as the two parameter logistic model. A special feature of both models is the symmetric nature of the link function $F^{-1}$ or ICC $F$. By considering normally distributed latent variables a convenient nomenclature that we follow in this paper for those models are probit-normal and logit-normal models, respectively (see Bazán et al., 2004). This notation allows distinguishing the link function and the distribution associated with the latent variables.

Chen et al. (1999) emphasizes that commonly used symmetric links for binary response data models, such as logit and probit links, do not always provide the best fit available for a given dataset. In this case the link could be misspecified, which can yield substantial bias in the mean response estimates (see Czado & Santner, 1992). In particular, when the probability of a given binary response approaches 0 at a different rate than it approaches 1, a symmetric link function, is inappropriate.

Samejima (2000) proposed a family of models, called the logistic positive exponent family, which provides asymmetric ICC and includes the ICC with logit link as a particular case. She considers that asymmetric ICCs are more appropriate for modelling human item response behavior. Also Samejima (1997) points out that there should be restrictions in using statistical theories and methods developed for something other than human behavior, in particular, those based on normal assumptions. One necessity under these circumstances is a departure from normal assumptions in developing psychometric theories and methodologies.

In this paper a new asymmetric ICC curve $F$ is assumed, by considering the cumulative distribution function of the standard skew-normal distribution (Azzalini, 1985). Moreover, a new latent variable distribution is assumed considering the skew-normal distribution. Considering simultaneously or separately, asymmetric ICC or asymmetric latent variables, a new family of IRT models can be formed. The new family defined is called as the Skew Normal Item Response Theory (SN-IRT) familiy and includes the symmetric probit-normal model as a special case, and hence extending the usual symmetric IRT models. According to the methodology considered in this paper, four models can be proposed for a data set under consideration: 1) the symmetrical ogive normal, namely, the "probit-normal (PN) model", 2) the skew-normal ICC model, namely the “skew probit- normal (SPN) model”, 3) the ogive normal model with asymmetric latent variable, namely, the “probit-skew normal (PSN) model”, and 4) the model that considers both types of asymmetry, asymmetrical ICC and asymmetry in the latent variable, namely, the “skew probit-skew normal (SPSN) model”.

The skew-normal distribution is an important asymmetric distribution with the normal distribution as a special case. Recent developments of asymmetric-normal models in the statistical literature include Azzalini & Dalla Valle (1996), Azzalini & Capitanio (1999) and Sahu et al. (2003). This distribution has been considered in the psychometric context in Arnold (1993). The possibility of considering asymmetric ICC has been previously formulated in Samejima (1997) and a particular skew-normal distribution has been used as a link function in Chen et al. (1999) for dichotomous quantal response data in regression models. (see also Chen, 2004)

The paper is organized as follows. Section 2 introduces the skew-normal item response theory model by considering the asymmetry parameter in the item characteristics curve, called here penalty parameter and by considering asymmetry in the latent variable. In the third Section, maximum likelihood fitting for the SN-IRT model is discussed. A Bayesian estimation approach is developed in Section 4 by using the MCMC methodology for simulating from the posterior distributions of item parameters and latent variables. Two data augmentation approaches are considered. In Section 5 an application with a data set from a mathematical test is reported comparing the presented models and also interpreting the proposed penalization (asymmetry) parameter.

## 2. The Skew-normal item response model

### 2.1 A new asymmetric item characteristic curve: the skew probit ICC

A new item characteristic curve IRT model is defined considering that the conditional probability $p_{ij}$ of a correct response for item $j$, given the value $u_i$ of the latent variable corresponding to the $i$*th* individual, is given by

$$p_{ij} = \Phi_{SN}(m_{ij};\lambda_j) = 2\,\Phi_2\left((m_{ij},0)^T, -d_j\right),\ \mathrm{i} = 1,\ldots,n\,,\ j = 1,\ldots,I \qquad (2)$$

with $-\infty < \lambda_j < \infty$, $d_j = \dfrac{\lambda_j}{\left(1+\lambda_j^{\,2}\right)^{1/2}}$, $\left| d_j \right| < 1$ , $i = 1,\ldots,n$ , $j = 1,\ldots,I$.

Where $\Phi_{SN}$ (.) denotes the cumulative distribution function of the standard skew-normal distribution (Azzalini, 1985) and $\Phi_2$(.) is the cumulative distribution function of the bivariate normal distribution (see as the last expression is deduced in Appendix A), with $\lambda_j$ the parameter of asymmetry of the skew-normal distribution and $d_j$ the correlation coefficient in the bivariate normal distribution. Note that $d_j$ is a reparametrization (1-1 transformation) of $\lambda_j$, so, it also seen as a parameter of asymmetry.

In the above expression, the probability $p_{ij}$ is expressed as a function of the quantity $u_i$ and the parameters $\eta_j = (\alpha_j, \beta_j)$ and of $\lambda_j$ (or $d_j$), which are parameters associated with item $j$. For $\lambda_j$=0, $p_{ij}$ reduces to $p_{ij} = \Phi(m_{ij})$, and the symmetric ICC probit follows.

As a consequence of the properties of the cumulative skew-normal distribution, it can be verified that the ICC skew-probit is a monotone increasing function of the quantity $u_i$, which is considered as a unidimensional latent variable. This means that the SN-IRT models are unidimensional monotone latent variable models (Junker & Ellis, 1997).

Figure 1 shows skew-probit ICCs for different values of a latent variable $U$ when fixing the item parameters $\alpha$=1 and $\beta$=0 and varying asymmetric parameter values $d$. Six different ICCs are considered for $d = -0.9, -0.7, -0.5, 0.5, 0.7, 0.9$ for comparison with $d$=0. For $d = 0$ (or $\lambda$=0), the ICC is symmetric and for $d > 0$ (or $\lambda > 0$) the ICC presents positive asymmetry, and for $d < 0$ (or $\lambda < 0$), its presents negative asymmetric. The ICC skew-probit is asymmetric as in the case of the family of positive exponent logistic models considered in Samejima (1997, 2000), which also considers an asymmetric item parameter called acceleration parameter. The ICC that we propose is an extension of the probit ICC, by directly introducing an asymmetry parameter. The asymmetric part is contained in the cumulative distribution of the standard normal distribution $\Phi(\lambda_j z)$. Further comments considering asymmetric ICC will be given in Section 4.3.

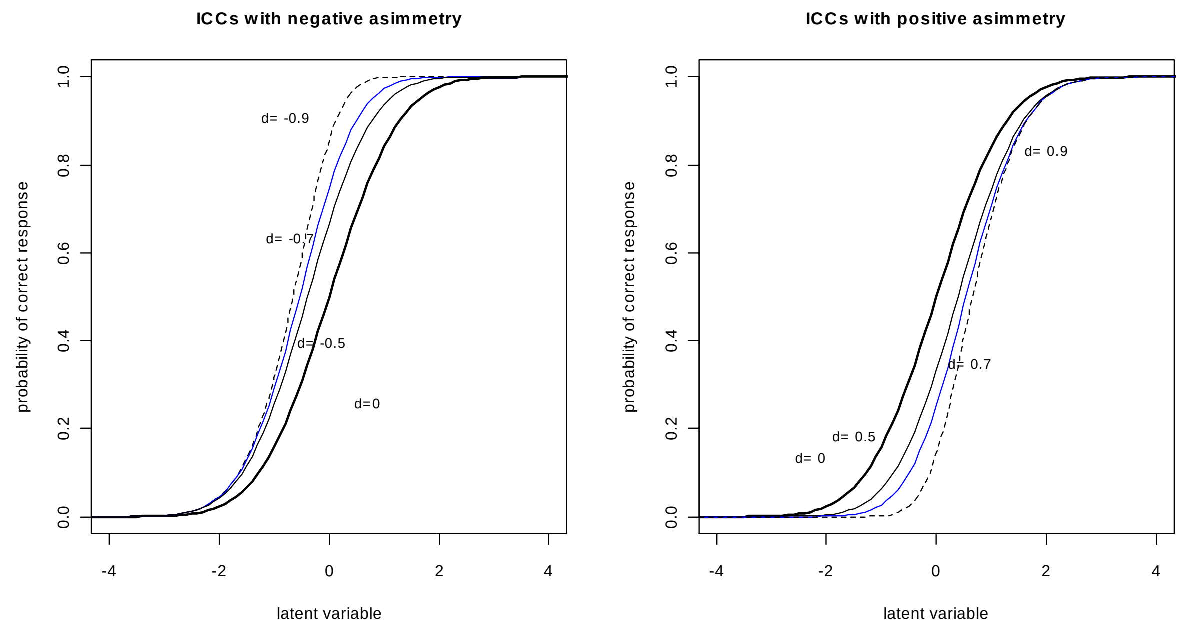


FIGURE 1. *Skew-probit ICCs for ($\alpha$=1, $\beta$= 0) and different values of the asymmetry parameter d*

The distance between the probabilities associated with the two ICCs (skew-probit and probit) at a point $z$, that is, $|\Phi(z) - \Phi(z, \lambda_j)|$, is at most $arctan(\mid \lambda_j \mid)/\pi$ (see Property 6, Appendix A). This means that if the model presents increasing positive asymmetry, then the probability of a correct response for considering a skew-probit ICC is diminished with respect to the symmetric probit ICC. On the other hand, if¸ the model presents increasing negative asymmetry; the skew-probit ICC presents increasing probabilities of a correct response with respect to the symmetric probit ICC. Such changes in the probability of a correct response occur due to the changes in the shape of the characteristic curve. However, the changes are not uniform as the latent variable changes. For highly negative values of the latent variable there is little change on the probability of correct response. On the other hand, for highly positive values, there are great changes on the probability of correct response. The item asymmetry parameter can be psychometrically interpreted as penalty (reward) of the probability of correct response. Hence, an item with negative asymmetry parameter penalizes (rewards) students with larges (smaller) levels of the latent variable and an item with positive asymmetry parameter rewards (penalizes) individuals with larger (smaller) levels of the latent variable (see Figure 1). Hence, we call the asymmetry parameter $d$ as the *item penalization parameter*.

Figure 2 presents six different characteristic curves assuming different values for $\lambda_j$, the asymmetry parameter. The first three curves consider additional variations in the parameter $\beta_j$ with $\alpha_j$ fixed. The last three curves consider variations in the parameter $\alpha_j$ with $\beta_j$ fixed. The parameter $\beta_j$ is called the item intercept or item difficulty parameter. This parameter controls item difficulty levels. If we fix the parameters $\alpha_j$ and $\beta_j$ and increment the value of $\beta_j$ , the basic form of the characteristic curve does not change but is translated to the right. Curves *C1*, *C2* and *C3* in Figure 2 correspond, respectively, to ($\alpha_j = 1$, $\beta_j = -1$), ($\alpha_j = 1$, $\beta_j = 0$), and ($\alpha_j = 1$, $\beta_j = 1$).

An item of a test with a highly negative value of $\beta_j$ (curve *C1*) corresponds to an easy item in which individuals with smaller averages in the latent variable presents relatively low probability of correct response. In contrast, an item with a large value of $\beta_j$ (curve *C3*) is difficult since individuals with large levels of the latent variable presents relatively low probability of correct response. The behavior of this parameter for the skew-probit ICC can be observed in Figure 2 for different values of the asymmetry parameter.

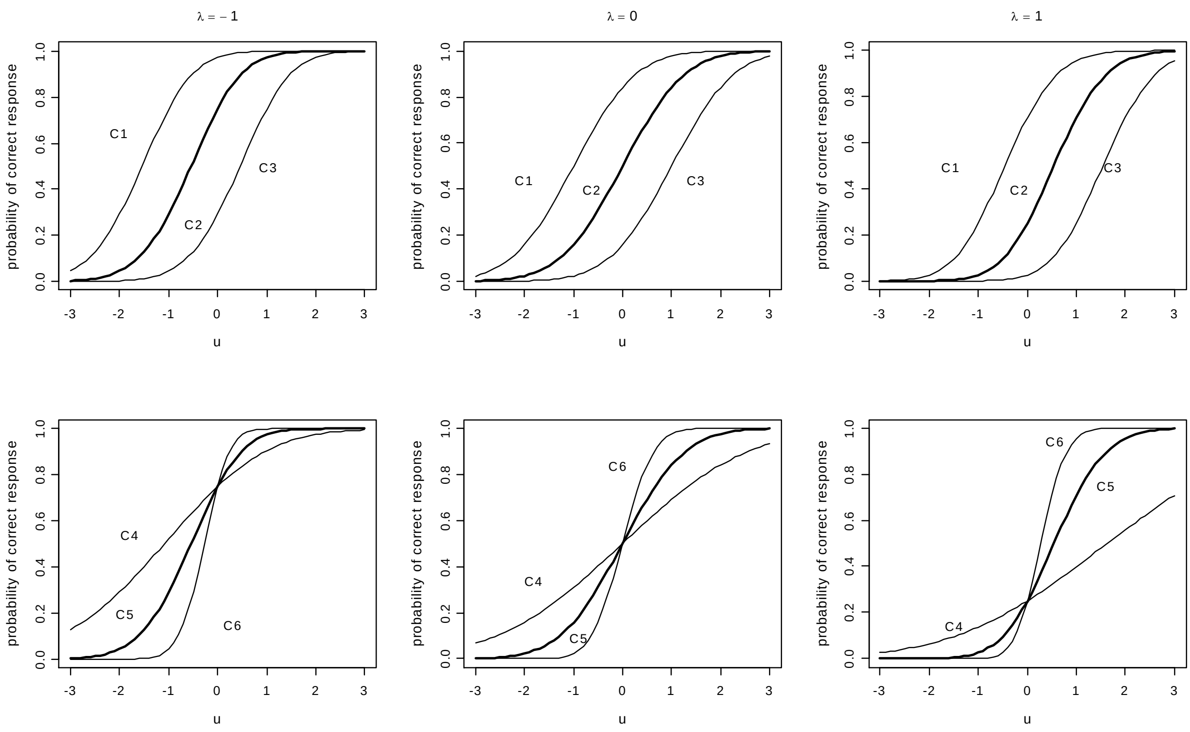


FIGURE 2. *Skew-probit ICCs for different values of the (α , β) item parameters and the penalization parameter* λ

The parameter $\alpha_j$ is called slope or discrimination parameter. This parameter controls the slope of the response function of an item. If we consider an ICC with $\beta_j$ and $\lambda_j$ fixed, the curve is steeper for increasing values of $\alpha_j$. The curves *C4*, *C5* and *C6* in Figure 2 correspond, respectively, to ($\alpha_j$ = 0.5, $\beta_j$ = 0), ($\alpha_j$ = 1, $\beta_j$ = 0) and ($\alpha_j$ = 2, $\beta_j$ = 0). A steeper item response curve corresponds to an item that highly discriminates students of smaller and greater levels of the latent variable. The probability of correct response changes rapidly for higher values of the latent variable in an interval containing zero. This means that the probability of correct response changes just a small amount when the latent variable goes from a student with lower latent variable values to a student with higher latent variable values. An item with a small value of $\alpha_j$ is a relatively poor discriminator between students for changing values of the latent variable.

In summary, the interpretation of the item parameters ($\alpha_j$ , $\beta_j$), is the same for probit ICC and skew-probit ICC in the usual probit-normal model and the SN-IRT model.

Following different proposals in the literature, we reparameterized the model introduced by considering $a = \alpha$ and $b = -\alpha\beta$ such that $m_{ij} = a_j u_i - b_j$, with $\eta_j = (a_j, b_j)$ the item parameter corresponding to j*th* item. According to Cook et al. (2002) and Baker (1992) this parameterization resulted in more stable computations. We use the notation $\boldsymbol{a} = (a_1, \cdots, a_I)^T$ and $\boldsymbol{b} = (b_1, \cdots, b_I)^T$.

2.2 Asymmetrically distributed latent variable

A new distribution for the latent variable $U_i$ corresponding to i*th* individual can be defined considering

$$U_i \sim SN(\mu,\ \sigma^2,\ \kappa),\ \ i = 1, \ldots, n, \tag{3}$$

which denotes the skew-normal distribution with location parameter $-\infty < \mu < \infty$, scale parameter $\sigma^2 > 0$ and asymmetry parameter $-\infty < \kappa < \infty$. See Appendix A for some properties of this distribution. Its density function is denoted by $\phi_{SN}(u; \mu, \sigma^2, \kappa)$.

Hence, we are considering that the latent variable follows a skew-normal distribution with common asymmetric parameters $\kappa$ for the individuals. It is our opinion that asymmetry does not to change from individual to individual since it is a property of the distribution of the latent variable.

Figure 3 shows the density functions of latent variables for different values of hyperparameters $\mu$ and $\sigma^2$ and for the asymmetry parameter $\kappa$. The three curves on the right side are examples of positive asymmetry parameter $\kappa$, modeling latent variables concentrated on lower values. The three curves on the left side are examples of negative asymmetry parameter $\kappa$ modeling latent variables concentrated on higher values. As a reference, in all cases, the *N*(0,1) curve is also presented.

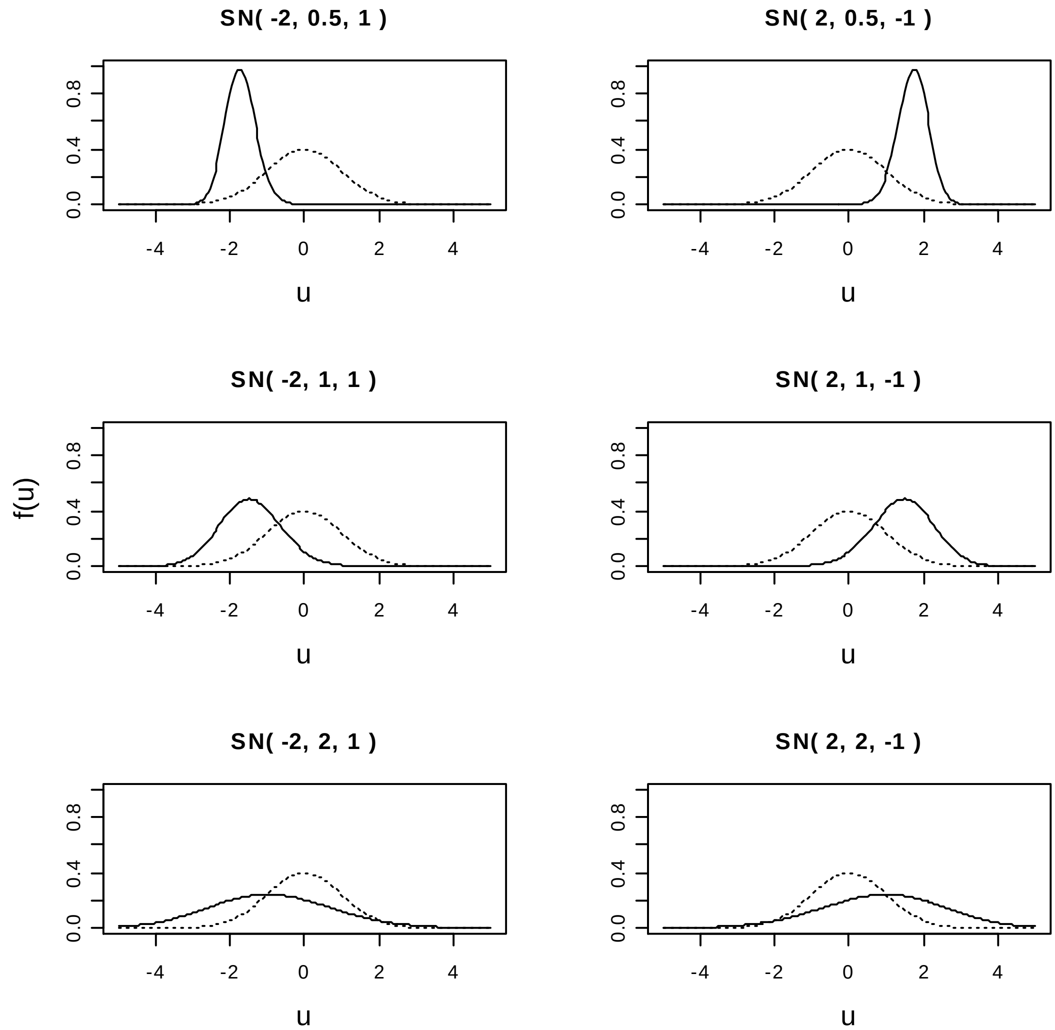


FIGURE 3. *Different density functions describing latent variables skew-normally distributed* $SN(\mu, \sigma^2, \kappa)$

Note that the curves showed in Figure 3 may correspond to the distribution of latent variables corresponding to human behavior in different cases, as observed in Micceri (1989). Examples of such behavior are depression (see Riddle et al. 2002) and anxiety (see Zaider et al., 2003) where certain asymmetry is expected considering a non-clinic population. Moreover, in educational contexts several predictor variables related to school proficiency can be asymmetrically distributed, as noted in Hashimoto (2002) and Vianna (2003). Then, the skew normal distribution is a flexible model for modeling of latent variables and accommodates the normal distribution as an especial case.

## 2.3 The Skew-Normal Item-Response Theory family

Formally, the SN-IRT family is defined by considering the specification of the news asymmetric ICCs and asymmetry distributions for the latent variable. Four SN-IRT models are possible by considering two or just one type of asymmetry on the specification of the ICC curve and on the specification of the distribution of the latent variable. They are:

a) The skew-probit skew-normal (SPSN) model,

b) The skew-probit-normal (SPN) model in that $\kappa=0$;

c) The probit-skew-normal (PSN) model in that $d_j=0$;

d) The probit-normal (PN) model in that $\kappa=0$ and $d_j=0$.

Note that the PN model is a special case of the SN-IRT model. An interesting aspect of the models formulated above are the flexibility in detecting items specified according to ICCs with skew-probit links and items specified according to ICCs with probit links. As such, the SN-IRT family presents flexible models for mathematical modelling of the psychological and educational behavior, based in *deductive processes* (Samejima, 1997) and fulfill the search for models that fits the behavior in question theoretically. *Inductive processes* (Samejima, 1997) as nonparametric IRT are just important as deductive processes and dynamic uses of both processes enable us to simulate human behavior. Besides, the link function studied in this paper can be also, considered for regression models with multivariate binary outcomes (see Chib & Greenberg, 1998) in which $m_{ij}=X_{ij}^{\mathrm{T}}\beta_{\mathrm{j}}$ with $\beta_{\mathrm{j}}$ a vector of unknown parameters, and $X_{\mathrm{ij}}$ a vector of covariates.

### 3. Maximum Likelihood Fitting

Let $\boldsymbol{D_{obs}} = \boldsymbol{y}$ denotes the observed data, so that the likelihood function for the SN-IRT family is given by

$$\boldsymbol{L}(\mu,\eta,\lambda \mid \boldsymbol{D_{obs}}) = \prod_{i=1}^{n}\prod_{j=1}^{I}\left[\Phi_{SN}(m_{ij};\lambda_j)\right]^{y_{ij}}\left[1-\Phi_{SN}(m_{ij};\lambda_j)\right]^{1-y_{ij}} . \tag{4}$$

Using the equivalent representation for the cumulative distribution function of the skew-normal distribution (see appendix A), we can be writing

$$\boldsymbol{L}(\mu,\eta,\boldsymbol{d} \mid \boldsymbol{D_{obs}}) = \prod_{i=1}^{n}\prod_{j=1}^{I}\left[2\Phi_2\left[(m_{ij},0)^T;-d_j\right]\right]^{y_{ij}}\left[1-2\Phi_2\left[(m_{ij},0)^T;-d_j\right]\right]^{1-y_{ij}} . \tag{5}$$

As in the case of the probit-normal model, computing the maximum likelihood estimators or the observed Fisher information matrix using the likelihood above is simplified if we have available procedures for computing bivariate integrals. Joint maximum likelihood (JML) estimates similar to the ones used with the probit-normal model can be implemented for the estimation of $\boldsymbol{u},\eta$ and $\lambda$ (or $\boldsymbol{d}$). Estimates can be obtained by iteratively maximizing the likelihood over either item parameters or latent variables while treating the other group of parameters as fixed at their current values (Baker, 1992). However, limitations of this method are well known (see Baker, 1992, and Bock & Aitkin, 1981). Assuming the latent variable or the item parameters known, we can also implement separated maximum likelihood estimators for the latent variables and item parameters.

Another procedure based on the "divide-and-conquer" strategy (Patz & Junker, 1999) for estimating only the item parameters $\eta$ and $\lambda$ (or $\boldsymbol{d}$) is the marginal maximum likelihood (MML) approach. The procedure can be developed by implementing an EM type algorithm as considered in Bock & Aitkin (1981) treating the latent variables as missing data and requiring the specification of a distribution for the latent variable ***U***. As we propose to consider a skew-normal distribution for the latent variable with common asymmetry parameters for individuals, i.e. $U_i \sim$ SN $(\mu, \sigma^2, \kappa)$, $i = 1, \ldots, n$, it follows that the marginal likelihood function for the SN-IRT model with interest on the item parameters can be written as:

$$\boldsymbol{L}(\boldsymbol{\eta}, \boldsymbol{\lambda} \mid \boldsymbol{D_{obs}}) = \int_{\Re^n} \prod_{i=1}^{n} \prod_{j=1}^{I} \left[\Phi_{SN}(m_{ij}; \lambda_j)\right]^{y_{ij}} \left[1 - \Phi_{SN}(m_{ij}; \lambda_j)\right]^{1-y_{ij}} \phi_{SN}(u; \mu, \sigma^2, \kappa) du \,. \qquad (6)$$

Estimation procedures based on the EM algorithm can be implemented by using Gaussian quadrature as a way of dealing with the multiple integral in the likelihood. The same approach can be used with the likelihood expressed in terms of the truncated bivariate normal distribution parameterized with $d_j$. However, as mentioned in Patz & Junker (1999) as the complexity of the model increases, as in the SN-IRT model situation, using the EM algorithm becomes less straightforward and is difficult to incorporate uncertainty (standard errors) into the item estimation of the parameters and in calculations of uncertainty (standard errors) for inferences on examinees (individuals), and there is no way of assessing the extent to which standard errors for examinee inferences are overly optimistic because of this. Chen et al. (1999) discuss conditions on existence of maximum likelihood estimators in probit models with skew-normal links, which can be extended to the models considered in this paper.

## 4. Bayesian Estimation

### 4.1 Priori Specification

Considering the likelihood function above, it is possible to implement a Bayesian estimation procedure by incorporating prior distributions for $\boldsymbol{u}$, $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ (or ***d***). For the SN-IRT familiy, we consider the following general class of indepents prior distributions:

$$\pi(\boldsymbol{u}, \boldsymbol{\eta}, \boldsymbol{\lambda}) = \prod_{i=1}^{n} g_{1i}(u_i) \prod_{j=1}^{I} g_{ej}(\eta_j) g_{3j}(\lambda_j) \,, \qquad (7)$$

where $g_{2j}(\eta_j) = g_{21j}(a_j)g_{22j}(b_j)$ in which $g_{21j}(.)$ should be proper to guarantee a proper posterior distribution (see Albert & Ghosh, 2000 & Ghosh et al. 2000). Following proposals usually considered (see Rupp et al., 2004), we take $g_{21j}(.)$ as the density of the N($\mu_a, s_a^2$), $j = 1, \ldots, I$, and $g_{22j}(.)$ as the density of N(0, $s^2_b$), $j = 1, \ldots, I$, the normal distribution, so that $g_{2j}(.)$ is the density of the $N_2(\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\eta)$, with mean vector $\boldsymbol{\mu}_\eta^T = (\mu_a, 0)^T$ and variance-covariance matrix

$$\boldsymbol{\Sigma}_\eta = \begin{pmatrix} S_a^2 & 0 \\ 0 & S_b^2 \end{pmatrix}.$$

Additionally, we consider $g_{1i}(.)$ as the density of the $SN(u_i ; \kappa)$, $i = 1, \ldots, n$, and $g_{3j}(.)$ as the density of the $SN(\lambda_j; \omega)$, $j = 1, \ldots, I$, where $\mu_a$, $s_a^2$, $s^2_b$, $\kappa$ and $\omega$ are known. In more general situations where $\kappa$ and $\omega$ are also unknown, the prior structure needs to be enlarged so that prior information are also considered for those hyperparameters. A possible extension of the model follows by considering $g_{1i}(.)$ as the density of the $SN(u_i ; \mu, \sigma^2, \kappa)$, $i = 1, \ldots, n$, with adequate specification of the hyperparameters $\mu$, $\sigma^2$ or by specifying hyperpriors for them.

Using the Bernoulli likelihood type and the prior distribution above it is possible to obtain posterior distributions and implement a Bayesian estimation procedure using the WinBUGS scheme (Spiegelhalter et al., 2003). However, such an approach is complicated because the integrals involved to obtain the marginal posterior distributions involve the cumulative function of the skew-normal distribution, which is not available in WinBUGS. We use instead an approach based on data augmentation as considered in Albert (1992) and Sahu (2002) for the PN model. This approach allows the implementation of Markov chain Monte Carlo methods, which simplify efficient sampling from the marginal posterior distributions.

## 4.2 Data augmentation approach

Using auxiliary latent variables, where these underlying variables have a standard skew-normal distribution, motivates our approach. The result that we present next is an extension of a similar result in Albert (1992) for the case of the PN model and is obtained using Property 6 in Appendix A. It can be shown that the skew-probit link in the SN-IRT family, involving $I$ items and $n$ individuals, with $Y_{ij} \sim Bern(p_{ij})$ and $p_{ij} = P(Y_{ij} = 1 | U_i = u_i, \eta_j, \lambda_j) = \Phi_{SN}(m_{ij}; \lambda_j)$ in which $m_{ij} = a_j u_i - b_j$, is equivalent to considering that

$$y_{ij} = \begin{cases} 1, & Z_{ij} > 0 \\ 0, & Z_{ij} \leq 0 \end{cases}, \quad \text{where } Z_{ij} \sim SN(m_{ij}, 1, -\lambda j), \; i = 1, \ldots, n, \; j = 1, \ldots, I.$$

Clearly, in the special case of $\lambda_j = 0$, $j = 1, \ldots, I$, the corresponding result in Albert (1992) for the symmetric PN model follows. The auxiliary latent variables $Z_{ij}$ are introduced to avoid working with Bernoulli type likelihoods. As a consequence of property 5 in the Appendix A it follow that the asymmetry parameter with the auxiliary latent variable is the opossit (in sign) of the asymmetry parameters with the ICC. In the following, we use the notation $\boldsymbol{Z} = (\boldsymbol{Z}_1^T, \cdots, \boldsymbol{Z}_n^T)^T$, with $\boldsymbol{Z}_i^T = (Z_{i1}, \cdots, Z_{iI})$ the vector of auxiliary latent variables. The next result follows from the previous results. The complete-data likelihood function for the SN-IRT models with skew-probit link and $\boldsymbol{D} = (\mathbf{Z}, \boldsymbol{y})$ is given by

$$\boldsymbol{L}(\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\lambda} | \boldsymbol{D}) = \prod_{i=1}^{n} \prod_{j=1}^{I} \phi(Z_{ij}; m_{ij}, 1, -\lambda_j) I(Z_{ij}, y_{ij}) \qquad (8)$$

where $m_{ij} = a_j u_i - b_j$, and $I(Z_{ij}, y_{ij}) = I(Z_{ij} > 0) I(y_{ij} = 1) + I(Z_{ij} \leq 0) I(y_{ij} = 0)$, $j = 1, \ldots, I$ and $i = 1, \ldots, n$, is used on the derivation of the conditional likelihood $p(y_{ij} | Z_{ij})$. Note that, if $\lambda_j = 0$ then the likelihood function above is equal to the one given in Albert (1992). Hence, the PN model is a reduced model in the SN-IRT family.

Considering the SN-IRT family with complete likelihood and prior distributions above for $\boldsymbol{u}$, $\eta$ and $\lambda$, the full posterior distribution corresponding to $\boldsymbol{u}$, $\eta$ ,$\lambda$ and $\boldsymbol{Z}$ is given by

$$\boldsymbol{f}(\boldsymbol{u},\eta,\lambda,\boldsymbol{Z}) \propto \prod_{i=1}^{n}\prod_{j=1}^{I}\phi_{SN}\left(Z_{ij};m_{ij},1,-\lambda_j\right)I(Z_{ij},y_{ij})\prod_{i=1}^{n}\phi_{SN}(u_i;\kappa)\prod_{j=1}^{I}\phi_2(\eta_j:\mu_\lambda,\Sigma\eta)\phi_{SN}(\lambda_j:\omega) \quad (9)$$

This distribution has an intractable form and will be very difficult to simulate from it. Therefore, a Gibbs sampling algorithm will be used where the three steps of the original algorithm by Albert (1992) are extended to four steps. Each step consists of sampling from the posterior distribution of one of the four parameter vectors $\boldsymbol{u}$, $\eta$, $\lambda$ and $\boldsymbol{Z}$ conditionally on all other parameters. This full conditional distribution should be tractable and easy to simulate from. The main steps for posterior in the Equation 8 are:

Step 1 *Sampling* $\boldsymbol{Z}$. Given $\boldsymbol{u}$, $\eta$ and $\lambda$ the variables $Z_{ij}$ are independent, and $Z_{ij} \mid \boldsymbol{u}, \eta, \lambda$ is distributed as the $SN(m_{ij}, 1, -\lambda_j)$ truncated at the left by 0 if $y_{ij}=1$ and truncated at the right by 0 if $y_{ij}=0$, for $i = 1, \ldots, n$ and $j = 1, \ldots, I$.

Step 2 *Sampling* $\boldsymbol{u}$. The latent variables are independent given $\eta$ ,$\lambda$ and $\boldsymbol{Z}$, with

$\pi(u_i \mid Z_j,,\eta,\lambda,D_{obs}) \propto \phi(u_i : m_{u_i}, v_{u_i})\psi(u_i)$, $i = 1, \ldots, n$,

where $m_{u_i} = \dfrac{\sum_{j=1}^{I} a_j\left(Z_{ij\,ij} + b_j\right)}{1+\sum_{j=1}^{I} a_j^2}$, $v_{u_i} = \dfrac{1}{1+\sum_{j=1}^{I} a_j^2}$, and $\psi(u_i) = \prod_{j=1}^{I}\Phi(-\delta_j Z_{ij} - m_{ij})\Phi(ku_i)$,

Step 3. *Sampling* $\eta$ . The classical item parameters are independent given $\boldsymbol{u}$ , $\lambda$ and $\boldsymbol{Z}$ with

$\pi(\eta_j \mid u, Z_j, \lambda_j, D_{obs}) \propto \phi_2(\eta_j; m_{\eta_j}, v_{\eta_j})\psi(\eta_j)$ for $j = 1, \ldots, I$,

with $\psi(\eta_j) = \prod_{i=1}^{n}\Phi\left(-\lambda_j Z_{ij} - W_i^T\eta_j\right)$, mean vector $m_{\eta_j} = \left[W^TW + \Sigma_{\eta_j}^{-1}\right]^{-1}\left[W^TZ_j + \Sigma_{\eta_j}^{-1}\mu_{\eta_j}\right]$ and covariance matrix $v_{\eta_j} = \left[W^TW + \Sigma_{\eta_j}^{-1}\right]^{-1}$ where $W = (W_1^T, \cdots, W_n^T)^T$, $W_i^T = (u_i, -1)$, $i = 1, \ldots, n$.

Step 4. *Sampling* $\lambda$. The asymmetry parameters are independent given $\boldsymbol{u}$,$\eta$ and $\boldsymbol{Z}$ with

$\pi(\lambda_j \mid u, Z_j, \eta_j, D_{obs}) \propto \phi(\lambda_j)\psi(\lambda_j)$ in which $\psi(\lambda_j) = \prod_{i=1}^{n} \Phi(-\delta_j Z_{ij} - m_{ij})\Phi(\omega\lambda_j)$, $i$ = 1, . . ., $n$ and $j$ = 1, . . . , $I$.

Note that some of the full conditionals cannot be directly sampled. For example, the case of the full conditional posterior for the parameter $\eta_j$ requires algorithms such as the Metropolis-Hastings (Chib & Greenberg, 1995). Further, if $\kappa$=0 and $\lambda_j$ = 0, $j$ = 1, . . . , $I$, then it follows that the full conditional distributions become the ones in Albert (1992).

### 4.3 An alternative data augmentation approach

To overcome the difficulties described above we propose to incorporate extra latent variables by modifying the auxiliary latent variable $Z_{ij}$, $j$ = 1, . . . , $I$ and $i$ = 1, . . . , $n$. These variables are considered next. Further, we consider the SN-IRT models in terms of the asymmetry parameter $d_j$, $j$ = 1, . . . , $I$, taking values in the interval (−1, 1), so that we can consider a uniform prior in (−1, 1) for $g_{3\,j}$ (.) and hence hyperparameters are not necessary. To consider a uniform prior for $d_j$ it is equivalent to consider a Student-t distribution for $\lambda_j$ with location 0, scale 1/2 and 2 degree of freedom (to see this, use the variable transformation $\lambda = \frac{d}{\sqrt{1-d^2}}$).

For $n$ examinees responding to $I$ items of a test it is known (Albert, 1992) that the probit link in the PN model should be rewritten as

$$Z_{ij} = m_{ij} + e_{ij}\text{, with } e_{ij} \sim N(0,1) \text{ and } y_{ij} = \begin{cases} 1, & Z_{ij} > 0 \\ 0, & Z_{ij} \leq 0 \end{cases},\ i = 1, \ldots, n \text{ and } j = 1, \ldots, I.$$

It follows that $p_{ij} = P(Y_{ij} = 1 \mid U_i = u_i, \eta_j) = \Phi(m_{ij})$. This representation shows a linear structure in the auxiliary latent variable $Z_{ij}$ normally distributed, which produces an equivalent model with a probit link. Further, the error $e_{ij}$ in the linear structure introduced are latent residuals independent and identical distributed (see Albert & Chib, 1995), a fact that can be used for model checking. Similarly, we define a linear error structure for the SPN and SPSN models with skew-probit link by considering that.

$$Z_{ij} = m_{ij} + e_{ij}\text{, with } e_{ij} \sim SN(0,1, -\lambda_j) \text{ and } y_{ij} = \begin{cases} 1, & Z_{ij} > 0 \\ 0, & Z_{ij} \le 0 \end{cases},\ i = 1, \ldots, n \text{ and } j = 1, \ldots, I.$$

Notice that $p_{ij} = P(Y_{ij} = 1 | U_i = u_i, \eta_j, \lambda_j) = \Phi_{SN}(m_{ij}; \lambda_j)$. Using the stochastical representation for a skew-normal distribution (Henze, 1986, Property 7 in the Appendix A) we can write $e_{ij} = -d_j V_{ij} - (1-d_j^2)^{1/2} W_{ij}$, with $V_{ij} \sim HN(0,1)$, the half normal distribution, and $W_{ij} \sim N(0,1)$, the standard normal distribution. By considering this stochastical representation, the skew probit link that we propose is more general than the probit link and is different from a skew link as given in Chen et al. (1999) (see also Chen, 2004) where $e_{ij} = W_{ij} + d_j V_{ij}$ take a different skew normal distribution class (see Sahu et al. 2003). It follows that the conditional distribution of $e_{ij} | V_{ij} = v_{ij}$ is a normal distribution with mean $-d_j v_j$ and variance $1 - d_j^2$ (see Property 8 in Appendix A). Moreover simulation of $Z_{ij}$ in the lineal structure should be considered in two steps. First simulate $V_{ij} \sim HN(0,1)$ and then simulate the conditional $Zs_{ij} \equiv Z_{ij} \mid V_{ij} = v_{ij} \sim N(m_{ij} - d_j v_{ij}, 1 - d_j)$. This defines an important hierarchical representation of the skew-normal distribution similar to the one derived for the Student-t distribution (see Arellano et al.,1994). In a similar way, the latent residuals in the probit link, the latent residuals $e_{ij}$ in the skew probit link are independent and identically distributed and also can be used for model checking.

We consider now the new complete data likelihood function involving the conditional auxiliary latent variables $\boldsymbol{Zs} = (\boldsymbol{Zs}_1^T, \cdots, \boldsymbol{Zs}_n^T)^T$ with $\boldsymbol{Zs}_i^T = (Zs_{i1}, \cdots, Zs_{iI})$, $i=1,...,n,$ and

$\boldsymbol{V} = (\boldsymbol{V}_1^{\boldsymbol{T}}, \cdots, \boldsymbol{V}_{\boldsymbol{n}}^{\boldsymbol{T}})^{\boldsymbol{T}}$ with $\boldsymbol{V}_{\boldsymbol{i}}^{\boldsymbol{T}} = (V_{i1}, \cdots, V_{iI})$, $i$ = 1, . . . , $n$. The new complete-data likelihood function for the SN-IRT models with $\boldsymbol{D}$ = (**Zs** **,*V*, *y***) is given by

$$\boldsymbol{L}(\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\lambda} \mid \boldsymbol{D}) = \prod_{i=1}^{n} \prod_{j=1}^{I} \phi\left(Zs_{ij}; -d_j V_{ij} + m_{ij}, 1 - d_j^2\right) I(Zs_{ij}, y_{ij}) \phi\left(V_{ij}; 0,1\right) I(V_{ij} > 0), \qquad (10)$$

where $m_{ij} = a_j u_i - b_j$ , and $I(Zs_{ij}, y_{ij}) = I\left(Zs_{ij} > 0\right) I\left(y_{ij} = 1\right) + I\left(Zs_{ij} \leq 0\right) I\left(y_{ij} = 0\right)$, $j$ = 1, . . . , $I$ and $i$ = 1, . . . , $n$, is used in the derivation of the conditional likelihood $p\left(y_{ij} \mid Zs_{ij}\right)$. Note that if $d_j$=0 then the likelihood function above is similar to the one given in Albert (1992), and $V_{\text{ij}}$ is not necessary.

Considering the SN-IRT models with new complete likelihood in the Equation 9 and prior distributions for $\boldsymbol{u}$, $\eta$ and $\boldsymbol{d}$ given above, the full posterior distribution corresponding to $\boldsymbol{u}$, $\eta$ , $\boldsymbol{d}$ and ***Z, V*** is given by

$$\boldsymbol{f}(\boldsymbol{u}, \boldsymbol{\eta}, \boldsymbol{d}, \boldsymbol{Zs}, \boldsymbol{V}) \propto \prod_{i=1}^{n} \prod_{j=1}^{I} \phi\left(Zs_{ij}; -d_j V_{ij} + m_{ij}, 1\right) I(Zs_{ij}, y_{ij}) \phi\left(V_{ij}; 0,1\right) I(V_{ij} > 0) \times$$

$$\prod_{i=1}^{n} \phi_{SN}(u_i, \kappa) \prod_{j=1}^{I} \phi_2(\eta_j : \mu_\eta, \Sigma_\eta) \qquad (10)$$

Although this density has an intractable form, we can simulate from it using a direct Gibbs sampling algorithm with five steps. Each step consists of sampling from the posterior of one the four-parameter vectors $\boldsymbol{u}$, $\eta$, ***d, Z,*** and ***V*** conditionally on all other parameters. This fully conditional distribution should be tractable and easy to simulate from. Some algebraic manipulations yield:

Step 1 *Sampling* ***Zs***. Given $\boldsymbol{u}$, $\eta$, ***d,*** and ***V***, the variables $Zs_{ij}$ are independent and distributed according to the $N(m_{ij} - d_j v_{ij}, 1 - d_j^2)$, $j$ = 1, . . . , $I$ and i = 1, . . . , $n$; distribution truncated at the left by 0 if $y_{ij}$=1 and truncated at the right by 0 if $y_{ij}$=0.

Step 2 *Sampling* $\boldsymbol{V}$. Given $\boldsymbol{u}$, $\eta$, and $\boldsymbol{d}$, the variables $V_{ij}$ are independent and distributed according to the $HN(d_j(m_{ij} - v_{ij}), 1 - d_j^2)$, $j = 1, \ldots, I$ and i = 1, . . . , $n$; distribution truncated at the left by 0.

Step 3 *Sampling* $\boldsymbol{u}$. The latent variables $u_i$ are independent given $\boldsymbol{Zs}$, $\boldsymbol{V}$, $\eta$ and $\boldsymbol{d}$ with

$\pi(u_i \mid Zs_j, V_j, \eta, d, D_{obs}) \propto \phi(u_i : m_{u_i}, v_{u_i})\Phi(ku_i)$, $i = 1, \ldots, n$,

where $m_{u_i} = \dfrac{\sum_{j=1}^{I} a_j \left(Zs_{ij} + d_j V_{ij} + b_j\right)}{1 + \sum_{j=1}^{I} a_j^2 - d_j^2}$, $v_{u_i} = \dfrac{1 - d_j^2}{1 + \sum_{j=1}^{I} a_j^2 - d_j^2}$ ;

Step 4. *Sampling* $\boldsymbol{\eta}$ . Given $\boldsymbol{Zs}$, $\boldsymbol{V}$, $\boldsymbol{u}$, and $\boldsymbol{d}$, the item parameters $\eta_j$ are independent and distributed according to the $N_2(m_{n_j}, v_{n_j})$, $j = 1, \ldots, I$, with mean vector

$\boldsymbol{m}_{\eta_j} = \left\lfloor \boldsymbol{W}^T \boldsymbol{W} + \Sigma_{\eta j}^{-1} \right\rfloor^{-1} \left\lfloor \boldsymbol{W}^T \Sigma_{\eta j}^{-1} \left(\boldsymbol{Zs}_j + \boldsymbol{d}_j \boldsymbol{V}_j\right) + \Sigma_{\eta j}^{-1} \mu_{\eta_j} \right\rfloor$ and covariance matrix

$\boldsymbol{v}_{\eta_j} = \left\lfloor \boldsymbol{W}^T \boldsymbol{W} + \Sigma_{\eta j}^{-1} \right\rfloor^{-1}$, where $W = (W_1^T, \cdots, W_n^T)^T$, $W_i^T = (u_i, -1)$, $i = 1, \ldots, n$:.

Steps 5. *Sampling* $\boldsymbol{d}$. Given $\boldsymbol{Zs}$, $\boldsymbol{V}$, $\boldsymbol{u}$, and $\eta$, the asymmetry parameters $d_j$ are independent and distributed according to the $\boldsymbol{N}\left(\dfrac{\boldsymbol{m}_{ij} - \boldsymbol{Zs}_{ij}}{\boldsymbol{V}_{ij}}, \dfrac{1}{\boldsymbol{V}_{ij}^2}\right)$, $j = 1, \ldots, I.$

We call attention to the fact that the conditional distributions of $\boldsymbol{Zs}$, $\boldsymbol{u}$ and $\eta$ given the other parameters are as given in Johnson & Albert (1999) for the special case of the symmetric PN model.

## 5. An application

We illustrate the Bayesian approaches to SN-IRT family developed in the paper, using the data set corresponding to the Mathematics Test applied in Peruvian schools with the goal to estimate the item parameter, included the new asymmetry parameter and to show advantages of the SPN model with respect to the PN model to capture more information on the item, in the presence of asymmetric scores.

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 the MCMC algorithm for the second augmentation approach above. The MCMC procedure is based on results of Proposition 6 and implemented on WinBUGS software.

We also present comparison of symmetrical and asymmetrical IRT models by using the Deviance Information Criterion (DIC) described in Spiegelhalter et al. (2002), a Expected Akaike Information Criterion (AIC) and Expected Bayesian Information (Schwarz) Criterion (EBIC) as suggested in Carlin & Louis (2000) and Brooks (2002) and sum-of-squared-latent residuals.

Spiegelhalter et al. (2002) claim that the DIC as implemented in the WinBUGS can be used to compare complex models (see Johnson, 2003, for example) and large differences in the criterion can be attributed to real predictive differences in the models, although there are critics to this approach. In hierarchical modelling with auxiliary latent variables as in the SN-IRT family, the likelihood (or "model complexity") is not unique so that the deviance (and also DIC and $\rho_D$ which are based on it) of a model with latent variables is not unique and can be calculated in several ways (see 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 DICs for the observed variables (fixed effects) and auxiliary latent variables (random effects) are presented. For a proper comparison of the proposed models, we considered marginal DIC for the observed variable because the focus of the analysis is in $p(y \mid \boldsymbol{u}, \boldsymbol{\eta})$ and although auxiliary random variables are introduced (in two steps) they are not the focus of the analysis. On the other hand, EAIC and EBIC are criterions proposed in Carlin & Louis (2000) and Brooks (2002) that penalizes the Posterior expected deviance by using $2p$ and $p\log n$, respectively, where as usual $p$ is the number of parameters in the model, and $n$ is the number of datapoints. As in Yan et al. (2003) we used the posterior sum of squares of latent residuals (SSR=$\sum_{i=1}^{n}\sum_{j=1}^{I} e_{ij}^2$) for the data set as a global discrepancy measure for models comparison.

## 5.1 Implementing the MCMC Algorithm

*Prior specification*

As have been mentioned, proper priors for $a_j$ and $b_j$ guarantee that the complete posterior for the model is proper. Albert & Ghosh (2000) mentions that the choice of proper prior distributions for the latent trait resolves particular identification problems, further, informative prior distributions placed on $a_j$ 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, we can choose $S_b^2$ to be large. 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 & Ghosh, 2000). Furthermore Sahu (2002) states that larger values of the variance led to unstable estimates. In Bazán et al. (2004) it is compared the use and perfomance of six different priors in literature for discrimination and difficulty parameters in the PN model. A sensitivity analysis by checking model adequacy that follows by using a series of prior distributions is conducted. It includes the specification of vague prior distributions for the difficulty parameters and precise parameters for the discrimination parameters. The priors specified 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 sensitive to the priors considered. However, the priors considered in Sahu (2002), which specifies that $a_j$ ~N(1; 0.5)I(0; ) and $b_j$ ~N(0; 2), $j$ = 1,...,$I$, seems to us the most adequate because it results in great precision for discrimination and difficulty parameter estimates.

For the models with link skew-probit, i.e., SPN and SPSN models, priors were specified for $d_j$ and not (directly) for $\lambda_j$, $j = 1, \ldots, I$, . The prior specified for $d_j$ is U(−1, 1).

For the models with asymmetry in the latent variable, i.e. SPSN and PSN models, we can consider $U_i \sim SN(\kappa)$, $i$=1, ... ,$n$. We consider with the SPSN model used for comparison, the values suggested by the scores in the test which for data set is question leads to $\mu_\kappa$ =-0.804 and $\sigma_\kappa^2$=6.329 . This specification is based in that the distribution of the scores in the test is a gross approximation to the distribution of the latent variable.

*Initial values*

We considered, as in Spiegelhalter (1996), initial values 1 and 0 for the item parameters $a_j$ and $b_j$, $j = 1, \ldots, I$, respectively. For the SPSN model we consider value 0 for the parameter$\kappa$. For the SPN and SPSN models we propose as initial values $d_j = 0$ for the asymmetry parameters because it corresponds to the mean/expected value of the uniform distribution on (−1, 1). Initial values for the latent variable $U_i$ and auxiliary latent variables corresponding to the different models (as $V_{ij}$ and $W_{ij}$), are considered as generated from standard normal distributions.

*Markov chain convergence*

Model SPSN is the more general. In the Math data set it involves 42 item parameters and 131 individual traits for the 131 individuals in the sample, but we can be interested only in the mean and standard deviation of the latent traits. Because a great number of chains must be generated for the different parameters, the MCMC procedure becomes slow (with the algorithm considered, PN model takes about 1 minute to run 1000 iterations on a Pentium IV Processor with 256 MB RAM). Informally, under the SPN model it takes twice the time it takes under the PN model; under SPSN model it takes about 1.5 times it takes under the PN model, and the generation was fastly complicated under PSN model. The time needed to run the chains for each model is also related to the presence of structures of the latent variables and mixtures (Chen et al. 2000), and to sample size (Sahu, 2002), which may significantly affect time of execution of SN-IRT models.

When using MCMC, the sampled values for initial iterations of the chains are discarded because of their dependence on starting states. Also, with SN-IRT models, presence of autocorrelations between chain values is expected when latent variables are introduced (Chen et al. 2000). Due to it, thin values up to 100 are recommended. Several criteria computed using the CODA package in the WinBUGS program, including the ones proposed by Geweke (1992) was used for to convergence analysis. An alternative to considerer is generate a great number of iterations and uses large values of thin. For example, Jacmank (2004) consider for the PN model, run half a million of iterations retain only every thoushandth iterations so as to produce an approximately independent sequence of sampled values from the joint posterior density.

Also Chen et al. (2000) mentioned that when the sample size n is large, ($n \geq 50$) slow converge of the Albert-Chib algorithm (data augmentation approach) may occur. Slow-converge of the chain corresponding to the asymmetry parameter is detected. Some algorithms to improve converge of the Gibbs sampler in the second data augmentation approach is suggested in Chen et al. (2001). Here we consider a large number of iterations, a total of 204000 iterations. Starting with a burn-in of 4000 iteration and them using thin=100, a sample size of 2000 is obtained.

5.2 Item Parameter Estimation: The Elementary-School Mathematics Test Example

In educational evaluation research it is typical to detect differences in scholar performance due to social-economic status. As an example, in a study conducted in Peru, Bazán et al. (2001) report differences observed in a mathematical test for sixth grade students in favor of students with higher social-economic status. Test scores show a negative asymmetric distribution since most of the students with higher social-economic status tend to obtain higher scores in the test.

In this application, 14 items of the Peruvian Elementary School Mathematical Test (UMC, 2001) were applied to 131 students of high social-economical status. Item response vectors are available from authors upon request. The distribution for the scores presents a mean of 10.84, a median of 11 and a standard deviation of 1.859. The skew and kurtosis indexes are estimated as -0.804 and 0.449, respectively. The test presents a regular reliability index given by Cronbach´s alpha of 0.48, and presents a mean proportion of items of 0.774, indication of being an easy test. It is shown that the scores present negative asymmetry in the behavior of sixth grade students for the mathematical test; hence item-test regression for different items does not present symmetric form. Although an item-test regression is not a close approximation to an item characteristic curve (Lord & Novick, 1968, p. 363), it may indicate a possible form for the true item characteristic curve. This justifies exploring SN-IRT models for this data set. This data set has been analyzed in Bazan et al., 2004 using a PN model.

To illustrate the utility of SN-IRT models, SPN and SPSN models are fit and compared with the fitting of the PN model. The main goal is to show advantages of the proposed models that consider a new item parameter, which is able to extract more information from the items.

Table 1.

Comparing of the PN, SPN and SPSN models using different criterion

| Criterion | Probit-Normal model | Skew-probit-Normal model | Skew-probit-Skew normal model |
|---|---|---|---|
| Number of parameters | 159 | 173 | 174 |
| Posterior expected deviance | 1446.17 | 1317.28 | 1321.83 |
| Deviance of the posterior means | 1358.13 | 1365.11 | 1363.35 |
| Effective number of parameters | 88.04 | -47.83 | -41.52 |
| DIC | 1534.21 | 1269.44 | 1280.32 |
| Expected AIC | 1750.17 | 1663.28 | 1669.83 |
| Expected BIC | 1942.21 | 1881.85 | 1889.66 |
| SSR posterior mean | 1853 | 1362 | 1351 |

From the posterior expected deviance, DIC, EAIC, EBIC values shown in Table 1, we see that the SPN model improves the corresponding symmetric PN model and asymmetric SPSN model. This later model also presents better fit than that of the PN model. This result is also observed by considering the sum-of-squared-error (SSR) of the latent residuals (see figure 4), which SPSN model can be alternative for the SPN model for the data set. In summary, the SPN model presents the best for the data set. Spiegelhalter et al. (2002) mentions that $p_D$ in the table can be negative and an explication for this fact is that it can indicate conflicting information between prior and data. This problem can be important in SN-IRT models when prior information is not available. Informative prior elicitation using historical data, as proposed by Chen, et al. (2001) and model sensitivity to choose of priors can be explored in subsequent studies.

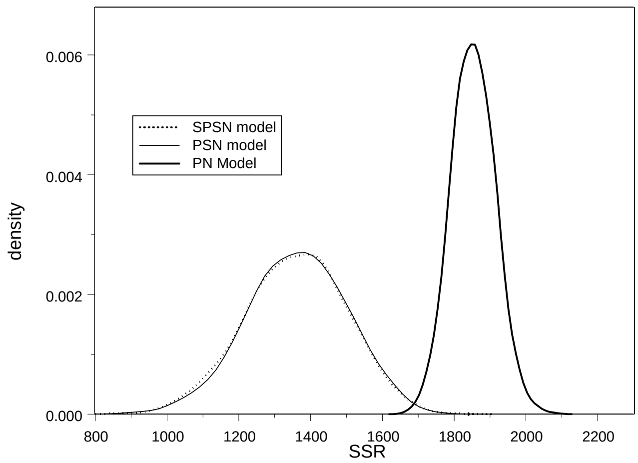


FIGURE 4. *Sum-of-squared-latent residuals* (*SSR*) *for PN, SPN and SPSN models*

Tables 2 and 3 illustrate the behavior of item parameter estimates for probit-normal and skew-probit-normal models. Estimates of item discrimination and difficulty parameters for the two models present similar behavior as are presented in the Table 2. The two types of parameters are equally interpretable under both models. 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. So, the skew probit-normal model is a model that offers the same conclusions about difficulty and discrimination parameters as the probit-normal model.

Table 3 shows estimated parameter values corresponding to the asymmetry parameter $d_j$ instead of the penalty parameter $\lambda_j$. We prefer to present inference on $d_j$ because it offers the same interpretation in simples scale. Figure 5 presents the histogram of the posterior distributions of parameter $d$, which indicates the presence of asymmetrical and symmetrical densities.

It is important say that by considering the notion of effective sample size (ESS) as used in Sahu (2002) we find that the $a$ parameters have best convergence. Convergence for the b parameters seems less precise, an indication that a larger chain is needed. This seems also to be the case with parameters $d$ and $\lambda$ although in a somewhat minor degree. There is indication that better convergence follows with the $d$ parametrization than with the $\lambda$ parametrization. In summary, for the asymmetry parameter, it seems better to work with $d$ parametrization, but big variability is observed.

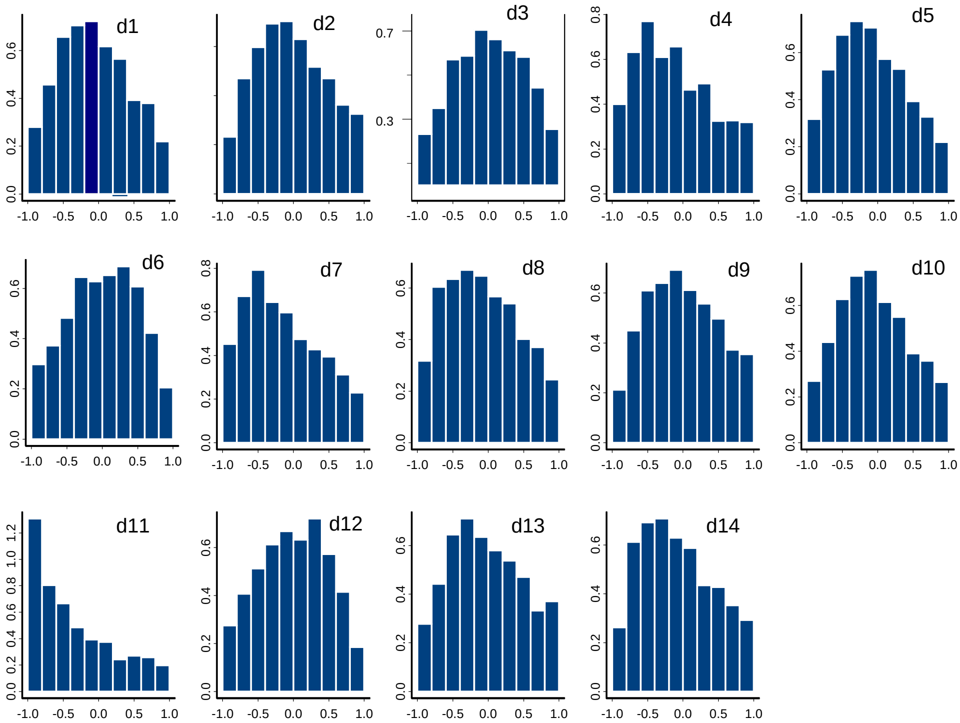


FIGURE 5. *Histograms of the posterior distribution for parameter d*

In Table 3 it is presented differences between posterior mean for the $a_j$ and $b_j$ parameters in the probit-normal and skew probit-normal models. As expected, the difficulty and discrimination parameters in the probit-normal model and skew-probit-normal model are approximately equal when the asymmetry parameter is close to zero. As an estimate of penalty parameter $d_j$ it seems more convenient to consider the posterior median instead of the posterior mean by reasons of asymmetry.

**I**tems 11, 4, and 7 are the ones that present penalty parameter estimates $d_j$ larger and negative (negative asymmetry on the item characteristic function), while discrimination and difficulty parameters differ in the two models. The posterior probability of negative values for the asymmetry parameter was computed for items 11, 4, and 7, resulting in the following values: 0.732, 0.614 and 0,632, respectively, which seems to indicate that these parameters are not equal to zero. Hence, the assumption of a skew symmetric ICC seems adequate.

In the special case of items 11, 4, and 7, the difference between models as consequence of the asymmetry parameter affects the difficulty parameter. Also observe that items 3 and 6 present positive penalty parameter estimates indicating that they present ICC with positive asymmetry but their value are not significantly different from zero, and also that the other items can be modeled correctly by considering a probit-normal model with symmetrical ICC.

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 4 says: "Pepe divides a number by 17, obtaining a quotient of 9 and a residual of 2. Which is the number he used?". We believe that these items penalize students with better knowledge while rewarding those with less knowledge. Students with better knowledge have very little differences on their probability of success for an item, but for students with less knowledge a little change on the terms used on the text of the item can produce a significant change on the probability of success. This type of analysis cannot be done with the probit-normal model and show possible advantages on psychometric interpretation of the skew models proposed.

As Albert & Ghosh (2000) mention, it can be hard to interpret the difficulty parameter $b_j$ since it is not expressible on the probability scale. An alternative measure of difficulty is the *unconditional probability of correct response* $p_j$ (see Appendix B), which is the probability that a randomly chosen individual from the population obtains a correct response to the j*th* question and can be interpreted as the proportion of student in the population that will correctly answer the j*th* item. Table 4 presents the unconditional probability of correct response under the PN and SPN models and the observed proportion of students in the sample, which correctly answer the j*th* item. The unconditional probability is obtained using the expression in the Appendix B and is estimated using the posterior median for item parameters.

Table 4.

Comparison of the estimated proportion of correct response under PN and SPN with the observed proportion in the sample

| Item | *d* estimated | Proportion Observed | Proportion Estimated | | Estimated-Observed | |
|---|---|---|---|---|---|---|
| | | | Under PN model | Under SPN model | Under PN model | Under SPN model |
| 1 | -0.084 | 0.794 | 0.786 | 0.789 | 0.0075 | 0.0052 |
| 2 | -0.053 | 0.855 | 0.851 | 0.854 | 0.0042 | 0.0007 |
| 3 | 0.010 | 0.519 | 0.510 | 0.517 | 0.0092 | 0.0022 |
| 4 | -0.175 | 0.931 | 0.918 | 0.923 | 0.0134 | 0.0078 |
| 5 | -0.136 | 0.870 | 0.863 | 0.872 | 0.0075 | -0.0014 |
| 6 | 0.027 | 0.366 | 0.361 | 0.361 | 0.0053 | 0.0055 |
| 7 | -0.219 | 0.924 | 0.914 | 0.921 | 0.0102 | 0.0026 |
| 8 | -0.108 | 0.878 | 0.860 | 0.870 | 0.0175 | 0.0077 |
| 9 | -0.031 | 0.786 | 0.782 | 0.786 | 0.0044 | 0.0004 |
| 10 | -0.084 | 0.863 | 0.854 | 0.861 | 0.0091 | 0.0015 |
| 11 | -0.495 | 0.931 | 0.914 | 0.929 | 0.0173 | 0.0021 |
| 12 | 0.007 | 0.351 | 0.347 | 0.349 | 0.0036 | 0.0024 |
| 13 | -0.072 | 0.824 | 0.818 | 0.823 | 0.0066 | 0.0010 |
| 14 | -0.134 | 0.947 | 0.943 | 0.948 | 0.0040 | -0.0012 |

The unconditional probability of correct responses under PN an SPN models can be considered as the estimated population proportion of correct responses and can be compared with the proportion observed proportion in the sample. The comparison of the differences (estimated-observed) with the penalty parameter is shown in the Figure 6. The solid circle corresponds to differences under the SPN model, and the transparent circle corresponds to differences under the PN model. Note that the distance between the differences in the proportion of correct response under PN and SPN models are higher when the penalty parameter is higher as is the case of item 11. Thus, there is strong evidence that for some items of the data set, the SPN model is more appropriate.

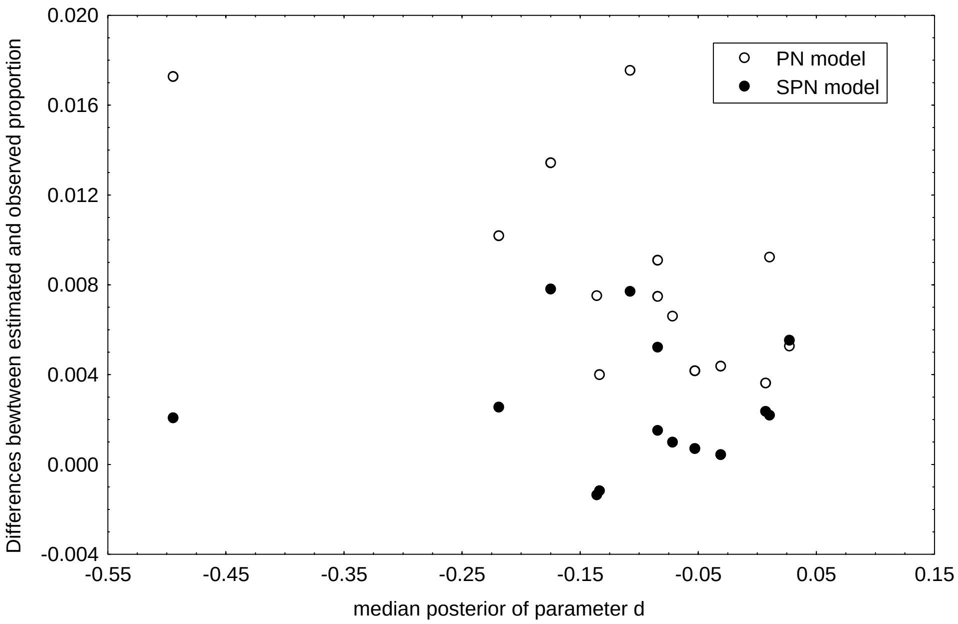


FIGURE 6. *Difference comparisons (estimated-observed) under PN and SPN models for the penalty parameter d*

## Discussion

This article proposes a new asymmetrical item response theory model, namely, the skew-normal item response theory model (SN-IRT) which considers a new asymmetric item characteristic curve by considering the cumulative distribution of the standard skew-normal distribution (Azzalini, 1985), but also considering the standard skew-normal model for the distribution of the latent variable. This new IRT model is denominated SN-IRT model. This extends the work of Albert (1992) to fit asymmetrical IRT models and includes the symmetric normal ogive or probit-normal model as a special case. Two data augmentation approaches are proposed by implementing a Bayesian estimation approach by using the MCMC methodology for simulating from the posterior distribution of item parameters and latent variables. In the first, MCMC methodology can be implemented by using the Metropolis-Hasting algorithm describes in Chib & Grenberg (1995).

In the second data augmentation approach MCMC methodology can be implemented by using simple Gibbs Sampling algorithms. Another contribution of this article is the investigation of model comparisons procedures. Comparison of symmetrical and asymmetrical IRT models are studied by using the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002), Expected AIC and expected BIC (Carlin & Louis, 2000 and Brooks, 2002). We also introduce latent residuals for the models and global discrepancy measures as the posterior sum of squares of the latent residuals, which can be used for model comparisons.

Finally, we give some interpretation to the new item parameter proposed called penalization parameter in the context of a data set from a mathematical test. It is also shown that the SPN model seems to present the better fit for the observed data. Although the HPD intervals in Table 3 indicate that the penalization parameters are not different from zero, so that a PN model would be more adequate to fit the data, we prefer to base our conclusion on which model to choose by using criteria such as DIC, expected AIC and expected BIC for the comparison between estimated and expected proportion of correct responses. This is due to the fact that the estimates of penalization parameters present large variances, perhaps consequences of too small sample size. Extensions to consider a model SN-IRT multidimensional model, hierarchical SN-IRT model, SN-IRT multinivel model can be studied in future developments.

Furthermore, the new skew-probit link proposed in the paper, based on the cumulative distribution of the standard skew-normal distribution (Azzalini, 1985), can be used in the context of binary and binomial regression models and extensions. Adaptations of the algorithms proposed on the paper for these cases are easy to be done. Other extensions of the skew probit link for ordinal responses as in Johnson (2003) are also to be studied in future development.

## 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. & Chib, S. (1995). Bayesian residual analysis for binary response regression models. Biometrika, 82, 747-759.

Albert, J.H. & Ghosh, M. (2000). Item response modeling. In Generalized Linear Models: A Bayesian Perspective (D. Dey, S. Ghosh & Mallick, eds.), Marcel-Dekker, New York, 173-193.

Andrade, D. F., Tavares, H. R., & Valle, R. C. (2000). Introducão a Teoria da resposta ao Item: Conceitos e Aplicações. 14o SINAPE: Caxambu, MG.

Arellano-Valle, R., Bolfarine, H., & Iglesias, P. (1994). Preditivistic interpretation to the multivariate t distribution. Test, 3, 1-16.

Arnold, B.C., Beaver, R. J., Groeneveld, R. A., & Meeker, W. Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. Psychometrika, 58, 471-478.

Azzalini A. (1985). A class of distributions which includes the normal ones. Scand. J. Statistical, 12, 171-178.

Azzalini A. & Capitanio A. (1999). Statistical applications of the multivariate skewnormal distributions. J. R. Statist. Soc. B, 61, 579-602.

Azzalini A. & Dalla Valle A. (1996). The multivariate skew-normal distribution. Biometrika, 83, 715-726.

Baker, F.B. (1992). Item Response Theory - Parameter Estimation Techniques. New York: Marcel Dekker, Inc.

Bartholomew,D.J., & Knoot, M. (1999). Latent variable models and factor analysis. (2nd ed.). London: Arnold. (Kendalls Library of Statistics 7.

Bazán, J., Espinosa G.,& Farro Ch. (2002). Rendimiento y actitudes hacia la matemática en el sistema escolar peruano. In Rodriguez, J. , Vargas, S. (eds.). Análisis de los Resultados y Metodología de las Pruebas Crecer 1998. Documento de trabajo 13. Lima: MECEP-Ministerio de Educación. 55-70.

Bazán, J., Bolfarine, H., & Aparecida, R. (2004a). Bayesian estimation via MCMC for probit-normal model in item response theory. 26p. Technical report (RT-MAE-2004-15).. Department of Statistics. University of São Paulo

Bazán, J., Bolfarine, H., & Branco, M. (2004b). A skew item response model. ISBA 2004 World Meeting. Viña del Mar. Chile, May 23-27, 2004. ISBA (International Society for Bayesian Analysis).

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika 64,153-168.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J.(2003). The Theoretical Status of Latent variables. Psychological Review, 110, 203-219.

Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin and van del Linde (2002). Journal Royal Statistical Society, Series B, 64,3, 616-618..

Chen, M.-H. (2004) Skewed link models for categorical response data, in Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality, Genton, M. G., Ed., Chapman & Hall / CRC, Boca Raton, FL, pp. 131-151.

Chen, M-H, Dey, D. K., & Shao, Q-M. (1999). A new skewed link model for dichotomous quantal response data. Journal of the American Statistical Association, 94, 448, 1172-1186.

Chen, M-H, Shao, Q. M, & Ibrahim, J. G (2000). Monte Carlo Methods in Bayesian Computation. New York: Springer Verlag.

Chen, M-H,,Dipack, K.D., & Shao, Q-M. (2001). Bayesian analysis of binary data using skewed logit models. Calcutta Statistical Association Bulletin, 51, 201-202.

Chib, S. & Greenberg, E. (1995). Undestanding the Metropolis-Hastings algorithm. The American Statistician, 49, 327-335.

Chib, S. & Greenberg, E. (1998). Analysis of multivariate probit models. Biometrika, 85,2, 347-361.

Cook, S., Barnard, Y. L., Rubin, D.B., Coleman, M.J., Matthysse, S., Levy, D.L., & Holzman, P.S. (2002). Working Memory Impairments in Schizophrenia Patients: A Bayesian Bivariate IRT Analysis. Gatsonis, C., Kass, R., Carriquiry, A., Gelman, A., Higdon, D., Pauler, D., Verdinelli, I. (Eds.). Case Studies in Bayesian Statistics. Springer- Verlag New York, Inc. 193-206.

Czado, C., & Santner, T. J. (1992). The effect of link misspecifiaction on binary regression inference. Journal of Statistical Planning and Inference, 33, 213-231.

Dalla Valle, A. (2004) The skew-normal distribution, in Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality, Genton, M. G., Ed., Chapman & Hall / CRC, Boca Raton, FL, pp. 3-24.

Delorio, M. & Robert, C.P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin and van del Linde (2002). Journal Royal Statistical Society, Series B, 64,3, 629-630.

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. Statistical Science, 7, 457-472.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.). Bayesian Statistics, 4, (pp 169-193). Oxford: Oxford University Press.

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

Hashimoto, .Y. (2002). Motivation and willingness to communicate as predictors of reported l2 use: the japanesse esl context. Second Language Studies, *20*(2), Spring 2002, pp. 29-70.

Henze, N. (1986). A probabilistic representation of the ”skew-normal” distribution. Scandinavian Journal Statistical. 13, 271-275.

Holland, P., & Rosenbaum, P. (1986). Conditional Association and Unidimensionality in Monotone Latent variable models. The Annals of Statistics. 14 1523-1543.

Jackman, S. (2004). Bayesian analysis for political research. Annual Review Political Science, 7, 483-505.

Johnson, V., & Albert, J. (1999). Ordinal Data Modeling. New York, MA: Springer-Verlag.

Johnson, T. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. Psychometrika, 68, 563-583.

Junker, B. W., & Ellis, J. L. (1997). A characterization of monotone unidimensional latent variable models. Annals of Statistics, 25, 1327-1343.

Liseo, B. & Loperfido, N. (2003). A Bayesian interpretation of the multivariate skew-normal distribution. Statistics & Probability Letters, 61, 395-401.

Lord, F., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Adisson-Wesley.

Micceri, T. (1989). The Unicorn, The Normal Curve, and Other Improbable Creatures Psychological Bulletin, *105*, 1, 156-166

Patz, R. J., & Junker, B. W. (1999). A straighforward approach to Markov Chain Monte Carlo methods for item response models. Journal of Educational and Behavioral Statistics, 24, 146-178.

Riddle, A S., Blais, M. R. &Hess, U. (2002). A Multi-Group Investigation of the CES-D's Measurement Structure Across Adolescents, Young Adults and Middle-Aged Adults.. Centre Interuniversitaire de recherche eit analysis des organizations. Scientific Series 2002s-36. Montreal

Rupp, A., Dey, D. K., & Zumbo, B. To Bayes or Not to Bayes, from Whether to When: Applications of Bayesian Methodology To Item Response Modeling. Structural Equations Modeling (to appear).

Sahu, S. K. (2002). Bayesian Estimation and Model Choice in Item Response Models. Journal of Statistical Computation and Simulation, 72, 217-232.

Sahu, S. K., Dey, D.E., Branco, M. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. The Canadian Journal of Statistics, 29,217-232.

Samejima, F. (1997). Departure from normal assumptions: a promise for future psychometrics with substantive mathematical modeling. Psychometrika, 62, 4,471-493.

Samejima, F. (2000). Logistic positive exponent family of models: virtue of asymmetric item characteristics curves. Psychometrika, 65, 3,319-335.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W.R.(1996). BUGS 0.5 examples (Vol. 1 Version i). Cambrigde, UK: University of Cambride.

Spiegelhalter D., Thomas, A. Best N. & (2003). WinBUGS version 1.4 [Computer program]. Imperial College & Medical Research Council Bioestatistics Unit. Institute of Public Health, Cambridge University.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal Royal Statistical Society, Series B, 64,3, 583-639.

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

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

Vianna, H. (2003). Avaliações Nacionais em Larga Escala: análise e propostas. Estudos em Avaliação Educacional, 27, 41-76.

Yan, D., Mislevy, R. J. & Almond, R. G. (2003). Design and analysis in a cognitive assessment. ETS Research Report 03-32, Educational Testing Service, Princeton, NJ.

Zaider, T. I. , Heimberg, R. G., Fresco, D.M., Schneier, F. R. & Liebowitz, M.R. (2003). Evaluation of the Clinical Global Impression Scale among individuals with social anxiety disorder. Psychological Medicine, *33*, 611–622.

Table 2.

Posterior mean, standard deviation (sd) and 90 % HPD intervals for the common parameters under the IRT Probit-normal and Skew-probit-normal models

| | Probit-Normal Model | | | | | Skew Probit-Normal Model | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | mean | sd | median | HPD Lower | HPD Upper | mean | sd | median | HPD Lower | HPD Upper |
| $a_1$ | **0.535** | 0.228 | 0.520 | 0.149 | 0.880 | **0.491** | 0.222 | 0.469 | 0.147 | 0.860 |
| $a_2$ | **0.300** | 0.181 | 0.280 | 0.007 | 0.557 | **0.269** | 0.172 | 0.245 | 0.001 | 0.499 |
| $a_3$ | **0.552** | 0.228 | 0.531 | 0.197 | 0.912 | **0.496** | 0.209 | 0.480 | 0.141 | 0.809 |
| $a_4$ | **0.915** | 0.348 | 0.889 | 0.332 | 1.462 | **0.856** | 0.351 | 0.824 | 0.252 | 1.375 |
| $a_5$ | **0.512** | 0.249 | 0.493 | 0.117 | 0.912 | **0.465** | 0.231 | 0.440 | 0.093 | 0.824 |
| $a_6$ | **0.314** | 0.166 | 0.299 | 0.026 | 0.551 | **0.284** | 0.154 | 0.267 | 0.040 | 0.524 |
| $a_7$ | **0.838** | 0.341 | 0.801 | 0.307 | 1.375 | **0.801** | 0.341 | 0.755 | 0.251 | 1.339 |
| $a_8$ | **0.959** | 0.350 | 0.928 | 0.350 | 1.470 | **0.885** | 0.331 | 0.858 | 0.341 | 1.401 |
| $a_9$ | **0.199** | 0.140 | 0.171 | 0.001 | 0.393 | **0.180** | 0.130 | 0.158 | 0.000 | 0.355 |
| $a_{10}$ | **0.496** | 0.237 | 0.479 | 0.112 | 0.879 | **0.451** | 0.233 | 0.430 | 0.053 | 0.782 |
| $a_{11}$ | **1.323** | 0.394 | 1.296 | 0.662 | 1.915 | **1.266** | 0.435 | 1.232 | 0.559 | 1.928 |
| $a_{12}$ | **0.387** | 0.196 | 0.369 | 0.043 | 0.668 | **0.357** | 0.179 | 0.340 | 0.047 | 0.616 |
| $a_{13}$ | **0.457** | 0.225 | 0.432 | 0.096 | 0.801 | **0.405** | 0.212 | 0.379 | 0.048 | 0.726 |
| $a_{14}$ | **0.445** | 0.276 | 0.411 | 0.001 | 0.825 | **0.404** | 0.257 | 0.364 | 0.002 | 0.753 |
| $b_1$ | **-0.900** | 0.153 | -0.895 | -1.124 | -0.630 | **-0.780** | 0.421 | -0.817 | -1.449 | -0.123 |
| $b_2$ | **-1.083** | 0.144 | -1.080 | -1.308 | -0.840 | **-0.976** | 0.426 | -1.043 | -1.613 | -0.285 |
| $b_3$ | **-0.026** | 0.127 | -0.028 | -0.217 | 0.195 | **-0.049** | 0.394 | -0.055 | -0.691 | 0.567 |
| $b_4$ | **-1.902** | 0.338 | -1.861 | -2.409 | -1.374 | **-1.662** | 0.581 | -1.701 | -2.591 | -0.690 |
| $b_5$ | **-1.232** | 0.178 | -1.218 | -1.524 | -0.952 | **-1.057** | 0.458 | -1.124 | -1.798 | -0.337 |
| $b_6$ | **0.373** | 0.121 | 0.371 | 0.181 | 0.570 | **0.328** | 0.392 | 0.347 | -0.304 | 0.933 |
| $b_7$ | **-1.779** | 0.317 | -1.746 | -2.267 | -1.278 | **-1.519** | 0.561 | -1.578 | -2.397 | -0.574 |
| $b_8$ | **-1.517** | 0.294 | -1.476 | -1.953 | -1.045 | **-1.332** | 0.512 | -1.396 | -2.060 | -0.417 |
| $b_9$ | **-0.794** | 0.131 | -0.790 | -1.010 | -0.592 | **-0.731** | 0.406 | -0.777 | -1.330 | -0.079 |
| $b_{10}$ | **-1.179** | 0.176 | -1.166 | -1.467 | -0.898 | **-1.051** | 0.441 | -1.112 | -1.709 | -0.321 |
| $b_{11}$ | **-2.290** | 0.440 | -2.236 | -3.025 | -1.663 | **-1.829** | 0.719 | -1.862 | -2.945 | -0.586 |
| $b_{12}$ | **0.424** | 0.123 | 0.418 | 0.226 | 0.627 | **0.388** | 0.386 | 0.405 | -0.199 | 1.021 |
| $b_{13}$ | **-0.995** | 0.156 | -0.988 | -1.230 | -0.740 | **-0.884** | 0.438 | -0.934 | -1.526 | -0.174 |
| $b_{14}$ | **-1.726** | 0.239 | -1.705 | -2.111 | -1.388 | **-1.522** | 0.502 | -1.612 | -2.306 | -0.736 |
| $u$ mean | **0.045** | 0.087 | 0.045 | -0.105 | 0.1819 | **0.036** | 0.087 | 0.039 | -0.109 | 0.178 |
| $u$ sd | **0.933** | 0.062 | 0.933 | 0.872 | 1.039 | **0.931** | 0.060 | 0.930 | 0.828 | 1.027 |

Table 3.

Posterior mean, sd and 90 % HPD intervals for the *d* parameters under the fitted skew probit-normal model

| | Posterior Statistics | | | | | Difference between PN and SPN models | |
|---|---|---|---|---|---|---|---|
| | mean | sd | median | Lower HPD | Upper HPD | *Mean difference in a* | *Mean difference in b* |
| d1 | -0.058 | 0.486 | **-0.084** | -0.772 | 0.823 | 0.044 | -0.120 |
| d2 | -0.021 | 0.499 | **-0.053** | -0.787 | 0.843 | 0.031 | -0.108 |
| d3 | 0.018 | 0.489 | **0.010** | -0.802 | 0.781 | 0.056 | 0.023 |
| d4 | -0.110 | 0.525 | **-0.175** | -0.892 | 0.788 | 0.059 | -0.240 |
| d5 | -0.089 | 0.488 | **-0.136** | -0.896 | 0.665 | 0.046 | -0.175 |
| d6 | 0.006 | 0.488 | **0.027** | -0.849 | 0.730 | 0.030 | 0.044 |
| d7 | -0.145 | 0.516 | **-0.219** | -0.975 | 0.640 | 0.037 | -0.260 |
| d8 | -0.077 | 0.506 | **-0.108** | -0.854 | 0.767 | 0.074 | -0.185 |
| d9 | -0.002 | 0.501 | **-0.031** | -0.850 | 0.778 | 0.019 | -0.064 |
| d10 | -0.048 | 0.488 | **-0.084** | -0.850 | 0.740 | 0.044 | -0.128 |
| d11 | -0.334 | 0.554 | **-0.495** | -1.000 | 0.571 | 0.057 | -0.461 |
| d12 | -0.006 | 0.484 | **0.007** | -0.737 | 0.826 | 0.031 | 0.036 |
| d13 | -0.028 | 0.510 | **-0.072** | -0.834 | 0.830 | 0.052 | -0.111 |
| d14 | -0.074 | 0.505 | **-0.134** | -0.864 | 0.741 | 0.042 | -0.204 |

## Appendix A

The skew-normal distribution

As considered in Azzalini (1985), 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

$$f_X(x) = \frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x-\mu}{\sigma}\right),$$

where $\phi(.)$ and $\Phi(.)$ denote, respectively, the density and distribution function of the standard normal distribution, with the notation $X \sim SN(\mu,\sigma^2, \lambda)$ used in this paper. The density above is denoted by $\phi_{SN}(x;\mu,\sigma^2,\lambda)$. Note that if $\lambda = 0$, the density of $X$ above reduces to the density of the $N(\mu,\sigma^2)$. In the special case of $\mu = 0$ and $\sigma^2 = 1$, which we denote as $X \sim SN(\lambda)$, it is called the standard skew-normal distribution.

The random variable $Z = (X - \mu)/\sigma$ is distributed according to the standard skew-normal distribution with density function given by $f_Z(z) = 2\phi(z)\Phi(\lambda z)$ represented by $\phi_{SN}(z; \lambda)$. The cumulative distribution function (cdf) of $Z$, denoted by $\Phi_{SN}(z; \lambda)$, is:

$$\Phi_{SN}(z;\lambda) = \int_{-\infty}^{z}\phi_{SN}(u;\lambda)du = \int_{-\infty}^{z}2\phi(u)\Phi(\lambda u)du = 2\,\Phi_2\left((z,0)^T,\mathbf{0},\Omega\right),$$

where straightforward algebraic manipulations yield the expression on the right with $d = \frac{\lambda}{\left(1+\lambda^2\right)^{1/2}}$, $|d| < 1$ and $\Phi_2(.)$ denotes the distribution function of the bivariate standard normal distribution with mean vector $\mathbf{0}=(0,0)^T$ and correlation matrix $\Omega=\begin{pmatrix}1 & -d\\ -d & 1\end{pmatrix}$. For simplicity, we denote $2\,\Phi_2\left((z,0)^T,\mathbf{0},\Omega\right)$ by $2\,\Phi_2\left((z,0)^T,-d\right)$. This result indicates that the cdf of the skew-normal distribution evaluated at a point $z$ can also be obtained by considering the bivariate standard normal cumulative distribution with mean vector $\mathbf{0}=(0,0)^T$ and correlation coefficient −d evaluated at the point ($z$, 0). This result is important since several computational algorithms are available for computing integrals related to the cumulative distribution of the bivariate normal distribution. Another algorithm for the cdf of skew-normal distribution based in the use Owen´s function (Azzalini, 1985, Dalla Valle, 2004) is available for R and Matlab program in http://tango.stat.unipd.it/SN/

Note that if $\lambda$=0, then $\Phi_{SN}(z;\lambda)$=$\Phi(z)$, the cdf of standard normal distribution. Some important properties of the skew-normal distribution are the following (see Azzalini 1985, Henze,1986)

1. If $X \sim SN(\mu, \sigma^2, \lambda)$, $E(X) = \mu + \sqrt{\frac{2}{\pi}}\frac{\lambda}{\sqrt{1+\lambda^2}}$ and $Var(X) = \left(1 - \frac{2\lambda^2}{\pi(1+\lambda^2}\right)\sigma^2$;
2. If $Z \sim SN(\lambda)$, the asymmetry and kurtosis indexes are given by $\gamma = \frac{1}{2}(4-\pi)sig(\lambda)\left[\frac{E^2[Z]}{Var[Z]}\right]^{3/2}$ and $\kappa = 2(\pi-3)\left[\frac{E^2[Z]}{Var[Z]}\right]^2$, implying that

$-0.9953<\gamma<0.9953$ and $0<\kappa<0.8692$ ;

3. If $Z \sim SN(\lambda)$, it follows that − Z is $SN(-\lambda)$ ;
4. The density of $Z$ is strongly unimodal, i.e. $\log\phi(z;\lambda)$ is a concave function of $z$ ;
5. $1 - \Phi(z;-\lambda) = \Phi(-z;\lambda)$ and $\Phi(z;1) = \Phi^2(z)$ ;
6. $\sup_Z |\Phi(z) - \Phi(z;\lambda)| = \arctan|\lambda|/\pi$,
7. An important stochastic representation (Henze, 1986) states that, if $V \sim HN(0, 1)$ the standard half normal distribution (see, Johnson et al. 1994) and $W \sim N(0, 1)$ are independent random variables then, the marginal distribution of $Z = dV + (1-d^2)^{1/2}W$, is $SN(\lambda)$, with $\lambda = \frac{d}{\sqrt{1-d^2}}$.

8. Considering the stochastic representation in Property 7, the conditional distribution $Z|V=v$ is a normal distribution with mean $dv$ and variance $1-d^2$, i.e. $Z|V=v \sim N(dv, 1-d^2)$. This represents an important hierarchical representation of the skew-normal distribution similar to the one derived for the Student-t distribution.

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

## Appendix B

Unconditional Probability of correct response under SN-IRT models

The following property of the SN-IRT family its a generalizations with respect to the usual property for symmetric probit-normal models.

Property 1. Consider the skew-probit skew-normal (SPSN)-IRT model with item parameters $\eta_j = (a_j, b_j)$ and $d_j$ and denoting the conditional probability of correct response with a latent variable *u,* as $p_{ij}$ = $P(Y_{ij} = 1 \mid u_i, \eta_j)$ = $p_j(u)$ , then, the unconditional probability of correct response for the item j is

$$p_j = \int_{-\infty}^{+\infty} p_j(u)\phi_{SN}(u;\kappa)du = 4\Phi_3\left[\begin{pmatrix} -b_j \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1+a_j^2 & -d_j & a_j\kappa \\ -d_j & 1 & 0 \\ a_j\kappa & 0 & 1+k^2 \end{pmatrix}\right], j = 1, \ldots, I.$$

*Proof*. The proof uses the following property of the normal distribution given by Liseo & Loperfido (2003) that is useful for computations with the skew-normal distribution. If $Z \sim N_k(\mu_1, \Sigma)$, *u* is an m-dimensional vector and *A* a full rank matrix, with $m \leq k$, then,

$\boldsymbol{E}_Z\left[\Phi_{\boldsymbol{m}}(\boldsymbol{u} + \boldsymbol{A}\boldsymbol{Z}; \mu_2, \Omega)\right] = \Phi_{\boldsymbol{m}}(\boldsymbol{u}; \mu_2 - \boldsymbol{A}\mu_1, \Omega + \boldsymbol{A}\Sigma\boldsymbol{A}^{\boldsymbol{T}})$. In particular, if $Z \sim N(0, 1)$, it follows that $E(\Phi(hZ + k)) = \Phi(\frac{k}{\sqrt{1+h^2}})$ (see Azzalini, 1985).

Similar arguments as in the proof given in Albert (1992) to the case of the usual symmetric probit-normal model can be used.

Two special cases are of interest. If we consider asymmetry only for the ICCs ($\kappa = 0$), the model becomes the skew-probit normal (SPN) IRT model, then,

$$p_j = \int_{-\infty}^{+\infty} p_j(u)\phi_{SN}(u;\kappa)du = 2\Phi_2\left[\begin{pmatrix}-b_j\\0\end{pmatrix};\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1+a_j^2 & -d_j\\-d_j & 1\end{pmatrix}\right] = \Phi_{SN}\left(\frac{-b_j}{\sqrt{1+a^2}};\, 0,1,\frac{-d_j}{\sqrt{1+a^2+d_j^2}}\right),$$

$j = 1, \ldots, I$.

On the other hand, considering only asymmetry for the latent variable, (i.e $d_j = 0$) , the models become the probit skew-normal (PSN) IRT model, and we obtain

$$p_j = \int_{-\infty}^{+\infty} p_j(u)\phi_{SN}(u;\kappa)du = 2\Phi_2\left[\begin{pmatrix}-b_j\\0\end{pmatrix};\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1+a_j^2 & a_j\kappa\\a_j\kappa & 1+k^2\end{pmatrix}\right],\quad j = 1, \ldots, I.$$

The two expressions above present generalizations with respect to the usual symmetric probit-normal model because if there is no asymmetry in the item characteristics curve and in the latent variable we obtain that the conditional probability of obtaining a correct response for item $j$ is $p_j = \Phi\left(\frac{-b_j}{\sqrt{1+a_j^2}}\right)$ as presented in Albert (1992). This is the probability that a randomly chosen individual from the population obtains a correct response for the item $j$ ($j = 1, \ldots, I$).