RT-MAE 2005-6

A SKEW ITEM RESPONSE MODEL

by

Jorge L. Bazán,
Márcia D'Elia Branco
and
Heleno Bolfarine

Palavras-Chave: link skew-probit, item response theory, Bayesian estimation, probit-normal model, skew-normal distribution.
Classificação AMS: 62F15; 62P15; 62E25.

- Fevereiro de 2005 -

A SKEW ITEM RESPONSE MODEL

Jorge L. Bazán, Márcia D'. Branco, Heleno Bolfarine*
Department os Statistics
University of São Paulo

December 01, 2004

Abstract

We introduce a new skew-probit link for binary response by considering an accumulated skew-normal distribution. Thus, an skew-item response model, that extends the symmetric probit-normal model, is derived by considering a new item 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. A Bayesian MCMC inference approach is developed and the study of efficiency in the estimation of the parameters of the model is undertaken for data set from Tanner (1996, p. 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 the data set corresponding to the Mathematics Test applied in Peruvian schools and a sensitivity analysis for the chosen priors is conducted.

KEY-WORDS: link skew-probit, item response theory, Bayesian estimation, probit-normal model, skew-normal distribution.

[1]Correspondence should be sent to Heleno Bolfarine e-mail: hbolfar@ime.usp.br. Caixa postal 66281, CEP 05315-970.

1 Introduction

Item Response Theory (IRT) is used to model a set of multivariate dichotomous responses from n individuals that are submitted to a test with k items. This type of model involves latent variables that explains individuals´ ability and a set of parameters associated with the items under consideration. It operates by modelling the probability of success (correct response) p_{ij} as $p_{ij} = F(m_{ij})$, where F is called the item characteristic curve (ICC), F^{-1} is called the link function and $m_{ij} = a_j u_i - b_j, i = 1, .., n,\ j = 1, ..., k$, where a_j and b_j are parameters associated to the items and u_i is a parameter associated with the individual abilities.

One special case follows by considering that $F(.) = \Phi(.)$, where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution. In this case the model is usually called the probit-normal IRT model or the ogive normal model (Albert, 1992; Bazan *et al.*, 2004). A logit link function has also been considered in the literature (see Birbaum, 1968). A special feature of both models is the symmetric nature of the link function and the ICC.

However, Chen *et al.* (1999) emphasize that symmetric links, such as logit and probit links, do not always provide the best fit available for a given data set. In this case the link could be misspecified, which can yield substantial bias in the mean response estimates (see Czado and Santner, 1992). This is true, when the probability of a given binary response approaches 0 at a different rate than it approaches 1. Also, Samejima (1997) indicated the necessity of a departure from normal assumptions in developing psychometric theories and methodologies. Thus, Samejima (2000) proposed a family of models, called the logistic positive exponent family, which provides asymmetric item characteristic curves and includes the logistic model, and suggested that asymmetric ICCs are more appropriate for modelling human item response behavior.

We propose an asymmetric link function by using the skew normal distribution (Azzalini, 1985). It define a skew- probit IRT model that includes the probit-normal model as special case. Recent developments of asymmetric-normal models in statistical literature can be seen in Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999), Sahu *et al.* (2003) and Genton (2004).

The paper is organized as follows. Section 2 introduces the skew-probit IRT model by considering a skew-probit link for the item characteristics curve of the IRT. In the third section, a Bayesian estimation approach, based in two data augmentation approaches are considered. In the fourth section, a study on the efficiency in the estimation of the parameters in the skew-probit IRT model is presented. Computation is developed by using the MCMC methodology for simulating from the posterior distributions of item parameters and latent variables. In section 5 an example of possible application is presented illustrating the usefulness of the approach and the more general class of models 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

$$\phi(z,\lambda) = 2\Phi(z)\Phi(\lambda z).$$

where $\phi(.)$ and $\Phi(.)$ denote the standard normal pdf and cumulative distribution function (cdf), respectively. We use the notation $Z \sim SN(\lambda)$. The parameter λ controls skewness, which is positive when $\lambda > 0$ and negative when $\lambda < 0$. The standard normal distribution is recovered with $\lambda = 0$. The skew normal cdf is defined by

$$\Phi(z,\lambda) = \int_{-\infty}^{z} 2\phi(u)\Phi(\lambda u)du.$$

Some important properties of the skew-normal distribution are the following (see Azzalini 1985, Henze,1986):

1. The mean and variance are given by

$$E[Z] = \sqrt{\frac{2}{\pi}}\frac{\lambda}{\sqrt{1+\lambda^2}}; \qquad Var[Z] = 1 - \frac{2\lambda^2}{\pi(1+\lambda^2)}.$$

2. The asymmetry and kurtosis indexes are given by

$$\gamma = \frac{1}{2}(4-\pi)sign(\lambda)\left[\frac{E^2[Z]}{Var[Z]}\right]^{3/2}; \qquad \kappa = 2(\pi-3)\left(\frac{E^2[Z]}{Var[Z]}\right)^2,$$

   implying that $-0.9953 < \gamma < 0,9953$ and $0 < \kappa < 0.8692$. where $sign(.)$ is a sign function that takes value 1 when λ is positive and -1 in otherwise.

3. The density of Z is strongly unimodal, i.e. $log\phi(z;\lambda)$ is a concave function of z.

4. An important stochastic representation (Henze, 1986) states that, if $X \sim HN(0,1)$ and $Y \sim N(0,1)$, the standard half normal distribution and standard normal respectively, are independent random variables then, the marginal distribution of $Z = \delta X + (1-\delta^2)^{1/2}Y$ is $SN(\lambda(\delta))$, with $\lambda = \frac{\delta}{(1-\delta^2)^{1/2}}$.

5. The stochastic representation in 4 can be represented hierarchically by considering that, the conditional distribution of $Z|X$ is a normal distribution with mean δx and variance $1-\delta^2$, i.e. $Z|X \sim N(\delta x, 1-\delta^2)$. and $X \sim HN(0,1)$. This represents an important hierarchical representation of the skew-normal distribution similar.

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

$$\Phi(z;\lambda) = 2\Phi_2\left(\begin{pmatrix} z \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix}\right), \tag{1}$$

where $\delta = \frac{\lambda}{(1+\lambda^2)^{1/2}}$ and $\Phi_2(.)$ denotes the cumulative distribution function of the bivariate standard normal distribution with correlation coefficient $-\delta$, which for easiness of notation, we denote as $\Phi_2((z,0)^T, -\delta)$.

Proof. Use directly the result of Parrish and Bergmann (1981) that consider that the standard normal bivariate distribution with correlation ρ evaluated at the points (h,k) can be written as $\Phi_2(h,k;\rho) = \int_{-\infty}^{h} \phi(w)\Phi\left(\frac{k-\rho y}{\sqrt{1-\rho^2}}\right)dw$.

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 a the point $(z,0)$. This result is important since several efficient computational algorithms are available for computing integrals related to the bivariate normal distribution (see Genz 1992, 1993). Another algorithm for the cdf of the skew normal distributions is based in the use Owen´s function (Azzalini, 1985 and Dalla Valle, 2004) and are available for R and Matlab programs.

More generally, a random variable X follows a skew-normal distribution with location parameter μ and scale parameter σ^2, if the density function of X is given by

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

with the notation $X \sim SN(\mu, \sigma^2, \lambda)$ used in this paper. The density (2) is denoted by $\phi(x;\mu,\sigma^2,\lambda)$. 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 $Z^* = aZ + b \sim SN(a\mu + b, a^2\sigma^2, sign(a)\lambda)$, where $sign(.)$ was defined before.

2.2 The Asymmetrical item characteristic curve

The item characteristic curve (ICC) for the conditional probability p_{ij} of a correct response for an item j given the latent variable U_i corresponding to the ith examinee is given by

$$p_{ij} = P[y_i = 1 \mid u_i, a_j, b_j, \lambda_j] = \Phi[m_{ij}; \lambda_j] = 2\Phi_2\left[(m_{ij}, 0)^T; -\delta_j\right] \tag{3}$$

with $\delta_j = \frac{\lambda_j}{(1+\lambda_j^2)^{1/2}}, i = 1,\ldots,n$ 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 λ_j, which are parameters associated with item j. Note that for $\lambda_j = 0$, (3) reduces to $p_{ij} = \Phi(m_{ij})$, as considered in the probit-normal model.

Figure 1 shows ICCs for different values of u and considering $a = 1, b = 0$ fixed with varying asymmetric parameter values. Five different characteristic curves are considered for $\lambda_j = -2, -1, 0, 1, 2$. In the ICC of the skew probit IRT model, when $\lambda > 0$, the probability of success starts with a slow growth for low values of the latent variable u. On the other hand, when $\lambda < 0$ the probability of success starts with a quick growth for low values of the latent variable u. By considering this fact, λ can be interpreted as a *penalization parameter*.
Thus, when an item has associated $\lambda > 0$ we say that the probability of correct response is penalized for low values of the latent variable. A fixed change on the latent variable implies in smaller (bigger) changes in the probability of success for low (high) values of the latent variable. 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. A fixed change of the latent variable results in smaller (bigger) changes in the probability of success for high (low) values of the latent variable. The interpretation is the same when the parameterization $\delta = \frac{\lambda}{(1+\lambda)^{1/2}}$ is used.

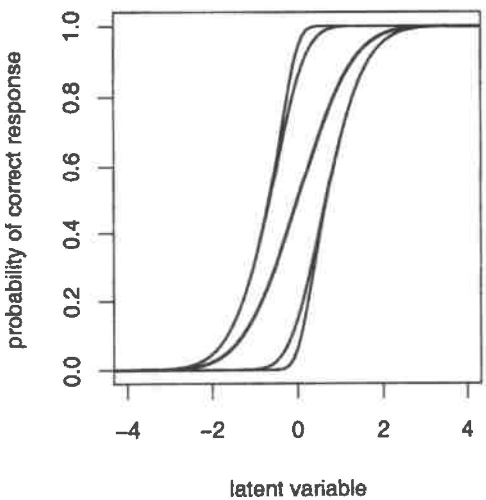

Figure 1: ICC for $a = 1, b = 0$ and different values of the asymmetry parameter $\lambda = -5, -2, 0, 2, 5$

As a consequence of the properties of the cumulative skew-normal distribution, 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).

2.3 Skew-probit IRT Model

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

$$y_{ij}|u, a_j, b_j, \lambda_j \sim Bern(p_{ij}) \tag{4}$$

(Bern: Bernoulli distribution) with p_{ij} defined in (2), $i = 1, \dots, n$ and $j = 1, \dots, k$. Moreover, the IRT model satisfies the latent conditional independence principle, which considers that for the ith examinee, y_{ij} are conditionally independent given u_i, $j = 1, \dots, k$. It is also considered that response from different examinees are also independent.

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

$$L(\boldsymbol{u}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\lambda}|D_{obs}) = \prod_{i=i}^{n}\prod_{j=1}^{k} \Phi_{SN}(m_{ij}; \lambda_j)^{y_{ij}}(1 - \Phi_{SN}(m_{ij}; \lambda_j))^{1-y_{ij}}, \tag{5}$$

and using the representation of the cumulative distribution function of the skew-normal distribution in (8), we can write

$$L(\boldsymbol{u}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\lambda}|D_{obs}) = \prod_{i=i}^{n}\prod_{j=1}^{k} \left[2\Phi_2[(m_{ij}, 0)^T; -\delta_j]\right]^{y_{ij}} \left[1 - 2\Phi_2[(m_{ij}, 0)^T; -\delta_j]\right]^{1-y_{ij}}. \tag{6}$$

Note that the skew-probit IRT model defined in (5) or (6) includes the probit-normal IRT model (Albert, 1992; Bazan *et al.* 2004) as particular case when $\lambda = 0$ is considered. The skew-probit model involves a total of $n+3k$ unknown parameters being thus overparametrized. Also, for a fixed number of items, item parameters are structural parameters and the latent variables are known as incidental parameters. In IRT, generally the analyzes is focused on item parameters and the number of parameters increases when n increases.The model is also unidentifiable,since it is preserved under a special class of transformations of the parameters (see Albert, 1992). One way of contouring such difficulties is to impose restrictions on the item parameters as considered, for example, in Bock and Aitkin (1981). Another way follows by specifying a distribution for the latent variables. As in Bazán *et al.*, (2004), we consider $U_i \sim N(0, 1)$,$i = 1, \dots, n$. This assumption establishes that it is believed that latent variable are well behaved and that the abilities are a random sample from this distribution.

2.4 Data augmentation approach

Our approach is motivated by using the latent variable method of Albert and Chib (1993), where the underlying latent variable has a skew distribution such as the so called standard skew-normal distribution (Azzalini, 1985). The result we present next is an extension of a similar result in Albert (1992) for the case of the Skew-probit IRT model.

Proposition 2 *The Skew-probit IRT model, involving k items and n examinees, with $y_{ij} \sim Ber(p_{ij})$ and $p_{ij} = \Phi(m_{ij}, \lambda_j)$ in which $m_{ij} = a_j u_i - b_j$, 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 $1 - \Phi(z; -\lambda) = \Phi(-z; \lambda)$ (see property E in Azzalini, 1985) and is similar to that given in Albert (1992).

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

$$L(\boldsymbol{u}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\lambda}|D) = \prod_{i=1}^{n} \prod_{j=1}^{k} \phi_{SN}(Z_{ij}; m_{ij}, 1, -\lambda_j) I(Z_{ij}, y_{ij}), \tag{8}$$

where $m_{ij} = a_j u_i - b_j$, and $I(Z_{ij}, y_{ij}) = I(Z_{ij} > 0)I(y_{ij} = 1) + I(Z_{ij} \leq 0)I(y_{ij} = 0)$, $j = 1, \ldots, k$ and $i = 1, \ldots, n$ correspond for the conditional density $p(y_{ij}|Z_{ij})$. use for derivation of (8). Note that, if $\lambda_j = 0$ then the likelihood function (14) is similar to the one given in Albert (1992).

An alternative way of writing the skew probit IRT model is presented next.

Proposition 3. *The skew-normal IRT model, with k items and n subjects, with $y_{ij} \sim Ber(p_{ij})$ and $p_{ij} = \Phi_{SN}(m_{ij}, \lambda_j)$ in which $m_{ij} = a_j u_i - b_j$, is equivalent to considering that*

$$y_{ij} = \begin{cases} 1, & Z^*_{ij} > 0; \\ 0, & Z^*_{ij} \leq 0, \end{cases},$$

with $Z^*_{ij} \sim N(-\delta_j X_{ij} + m_{ij}\ ,\ 1 - \delta_j^2)$, and $X_{ij} \sim HN(0, 1)$, $j = 1, \ldots, k$ and $i = 1, \ldots, n$.

Proof Considering $Z^*_{ij} = Z_{ij}|X_{i,j}$ and the stochastic representation for the skew-normal distribution (Henze, 1986) for Z_{ij}, where $Z_{ij} \sim SN(m_{ij}\ , 1, -\lambda_j)$, $j = 1, \ldots, k$, $i = 1, \ldots, n$, as given in Proposition 2. Then, the proof follows by using Remark 1 with the skew-normal distribution.

We consider now the complete data likelihood function involving the latent data $\boldsymbol{Z}^* = (Z^*_{11}, \ldots, Z^*_{kn})^T$. Let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k)^T$. The likelihood function of

$(u^T, a^T, b^T, \delta^T)^T$ with augmented data $D = (Z^*, X, y)$ is given by

$$L(u, a, b,\ \delta|D) = \prod_{i=1}^{n}\prod_{j=1}^{k} \phi(Z_{ij}^*\ ; -\delta_j X_{ij} + m_{ij}\ , 1-\delta_j^2) I(Z_{ij}^*\ , y_{ij}) \phi(X_{i,j}\ ; 0, 1) I(X_{ij} > 0) \tag{9}$$

in which $m_{ij} = a_j u_i - b_j$ and $I(c, d)$ is equal to 1 when $\{c > 0, d = 1\}$ or $\{c \leq 0, d = 0\}$ and zero otherwise.

3 Bayesian Estimation

3.1 Prior Distribution

To implement a Bayesian estimation procedure we have that incorporating prior distributions for u, a, b and λ. We consider the following general class of independent prior distributions:

$$\pi(u, a, b, \lambda) = \pi_1(u)\pi_2(a)\pi_2(b)\pi_4(\lambda). \tag{10}$$

Since independent priors are considered in (10) we can choose priors for $\pi_1(.)$, $\pi_3(.)$ and $\pi_3(.)$ given in the probit-normal model. Thus, following proposals usually considered for the probit-normal model (see Sahu, 2002; and Bazan *et al.*, 2004), we take conjugate normal priors, that is,

- $U_i \overset{iid}{\sim} N(0, 1), i = 1, \ldots, n,$
- $a_j \overset{iid}{\sim} N(\mu_a, \sigma_a^2) I(a_j > 0), j = 1, \ldots, k$, and
- $b_j \overset{iid}{\sim} N(0, \sigma_b^2), j = 1, \ldots, k.$

By considering the results of Albert and Ghosh, (1999) and Ghosh *et al.* (2001) the distribution of the discrimination parameter must be proper to guarantee a proper posterior distribution and values of μ_a and σ_a^2 should be informative. Already, in the common situation where little prior information is available about the difficulty parameter, one can chose σ_b^2 to be a large value. As is mentioned in Albert and Ghosh (2000) and Sahu (2002) for the probit-normal model, 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 examinees are observed to get correct or incorrect response to every item) is observed. Thus, vague priors can be admissible for the difficulty parameter. We consider μ_a, s_a^2, s_b^2, as known values. In more general situations, the prior structure needs to be enlarged so that hyper prior information can be also considered for those parameters.

We can use the following prior for a and b parameters: a) $a_j \sim N(0, 1) I(a_j > 0)$ and $b_j \sim N(0, 10000)$, as in Spiegelhalter *et al.* (1996) or b) $a_j \sim N(1, 0.5) I(a_j > 0)$ and $b_j \sim N(0, 2)$ as in Sahu (2002). Although the use of another prior available in the literature lead to similar posterior analyzes for the probit normal model, Bazán *et al.* (2004) find that the priors above are more appropriate because they result in greater precision for discrimination estimates and difficulty parameters.

Further, we can consider the skew probit IRT model in terms of $\delta_j = \frac{\lambda_j}{(1+\lambda_j^2)^{1/2}}$, $j = 1, \ldots, k$, taking values in the interval $(-1, 1)$ this define the *delta parametrization* of the skew-probit IRT model, so that we can consider that. In this case

- $\delta_j \sim U(-1, 1)$.

The prior above is equivalently to $\lambda_j \sim T(0, 0.5, 2)$, where $T(\mu, \sigma^2, \nu)$ denote the student-t distribution with location μ, scale σ^2 and ν degrees of freedom. When considered this prior, this define the *lambda parametrization* of the skew-probit IRT model.

3.2 MCMC implementation

Using the likelihood in (5) or (6) and the prior distribution in (10) to implement a Bayesian estimation procedure involving a Bernoulli likelihood it is complicated since the integrals involved to obtain the marginal posterior distributions are difficult to deal with. Two approaches based on data augmentation as considered in Albert (1992) was introduced in the Section 2.4. This approach allows the implementation of Markov Chain Monte Carlo methods which simplify efficient sampling from the marginal posterior distributions.

By considering this latent structure, the full conditionals for the skew probit IRT model using the first augmented likelihood function in Section 2.4, Bayesian inference via MCMC follows without complications as reported in Johnson and Albert (2000). Moreover some of the full conditionals can not be directly sampled from, requiring algorithms such as the Metropolis-Hastings (Chib and Greenberg, 1995). To overcome the difficulties described above we can use the 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 section we develop a computational procedure for the skew probit IRT model based in this second augmented likelihood function. Two parameterization can be used. The full likelihood specification for the *delta parameterization* is given as follows:

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

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

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

$$b_j \sim N(0, \sigma_b^2);$$

and

$$\delta_j \sim U[-1, 1].$$

The full likelihood specification for the *lambda parameterization* is given as follows:

$$Z^*_{ij}|y_{ij}, a_j, b_j, \lambda_j \sim N(m_{ij} - \frac{\lambda_j}{\sqrt{1+\lambda_j^2}} X_{ij}, \frac{1}{1+\lambda_j^2}) I(Z^*_{ij}, y_{ij});$$

$$X_{ij} \sim HN(0,1);$$
$$a_j \sim HN(\mu_a, \sigma_b);$$
$$b_j \sim N(0, \sigma_b^2);$$
and
$$\lambda \sim T(\mu, \sigma^2, v).$$

This hierarchical structures can be easily introduced in WinBugs software. Note that all of the full conditional distributions for Gibbs sampling are straightforward to derive and to sample from. Also note that when $\delta_j = 0$, the hierarchical structure of the augmented likelihood corresponding to the probit-normal follows by eliminating the second and fifth lines in the above hierarchy.

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

We consider in this section an investigation on 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 parameters of the model when a data augmentation approach is considered, and b) to evaluate the use of the *delta* and *lambda parameterizations* introduced in the Section 3.2 for the penalization parameter. We consider a reanalysis of the data set taken from Tanner (1996, page 190) which includes $k = 6$ items and $n = 39$ examinees.

When using MCMC, the sampled values for initial iterations of the chain are discarded because of their dependence on starting states. But, presence of autocorrelations between chain values is expected when latent variables are introduced (Chen *et al.*, 2000) as is the case of the data augmentation approach. As such it, is convenient to perform an analysis of the autocorrelations for the parameters in the skew-probit IRT model introduced. 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 sample size per second (ESS/s) as considered in Sahu (2002), that is ESS divided by the running time. ESS is defined for each parameter as the number of MCMC samples drawn, B, divided by the parameter´s autocorrelation time, $\gamma = 1 + 2\sum_{s=1}^{\infty} \rho_s$, where ρ_s is the autocorrelation at lag s. Estimation of γ using sample autocorrelations is problematic because fewer MCMC samples are used in estimating ρ_s as s increases. We use $s = 49$. Thus, Sahu (2002) points out that the ratio of the means of ESS/s of the parameters analyzed can be considered one measure of efficiency in its estimation. We use this concept for studying efficiency in estimation for the skew-probit and probit-normal models. This concept also can be used for the comparison of the efficiency in the estimation of the parameters of the skew-probit model when *delta* or *lambda parameterization* introduced in Section 3.2 is considered.

For this study, it is considered the 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 $\lambda_j \sim T(0, 0.5, 2)$]. We consider, as in Spiegelhalter (1996), initial values $a_j = 1$, and $b_j = 0$, $j = 1 \ldots, k$. Initial values for the asymmetry parameter δ_j can be randomly generated, but on Winbugs it is preferable to choose fixed ones. We propose as initial value, $\delta_j = 0$ or $\lambda_j = 0$. Initial values for U_i the latent variables and auxiliary latent variables such as Z_{ij} and X_{ij} are randomly generated by considering the corresponding distributions.

In Table 1 we report the mean and standard deviation of the estimated of ESS and ESS/s for each type of parameter for the skew-probit (*delta* or *lambda parameterization*) and probit normal models over the number corresponding parameters.

Table 1: Performance in the parameter estimation for the Tanner´s data example (Std: standard deviation)

Model	Parameter	ESS		ESS/s	
		Mean	Std.	Mean	Std.
Skew-probit	a	260.7	125.9	3.7	1.8
(delta parametrization)	b	157.6	20.9	2.2	0.3
	δ	137.3	16.7	1.9	0.2
	λ	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
	δ	76.4	7.6	0.9	0.1
	λ	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 normal	a	307.9	127.5	3.8	1.6
	b	905.2	338.8	11.2	4.2
	u	783.5	236.5	9.7	2.9
	Total	741.9	287.7	9.2	3.5

Figures 2 and 3 shows the error bar of the ESS/s of each parameter for each model. Using the concept of efficiency defined above and forming the ratios for each parameters in the models we note that:

- They are a ranking of efficiency in the estimation of the parameters in the models. The estimation is more efficient for the u parameters, followed by the a parameters, which are followed by the b parameters, and, finally, δ following the λ parameters.
- In the skew-probit model, the *delta parameterization* is two times more efficient

in the estimation of the δ and the λ parameters than the *lambda parameterization.*

- When a probit-normal 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 b parameter when the skew-probit model is considered, especially when using the *lambda parameterization.*

Other data sets (the known LSAT example and the Math example in Section 5) present similar behavior as described above.

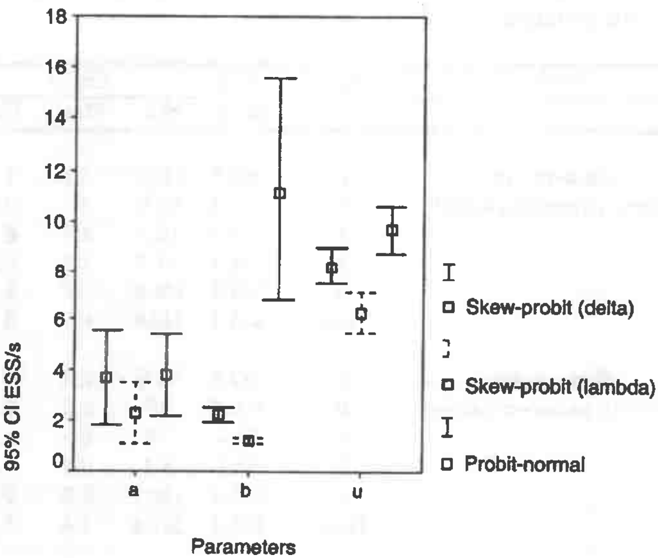

Figure 2: Performance of the parameter estimates a and b for the Tanner´s data example using 95 % Confidence interval of sample size per second (ESS/s)

By considering the presence of autocorrelation in the Skew-probit model, thin values up to 100 are recommended. Consequently a large number of iterations are necessary for inference based on the joint posterior density. This kind of behavior was observed for other models as well. Jackman (2004) considering the Probit-normal model, run 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.

In addition, 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)

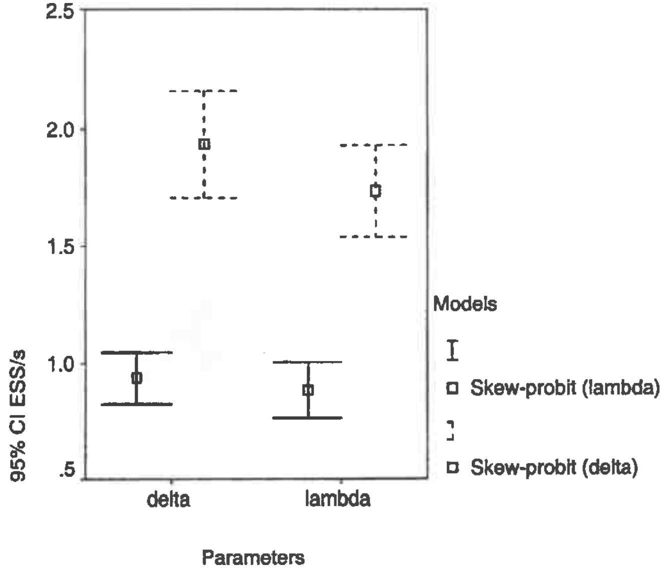

Figure 3: Performance of the parameter estimates δ for the Tanner´s data example using 95 % Confidence interval of sample size per second (ESS/s)

may occur. Slow convergence of the chain corresponding to the asymmetry parameter is detected speciality in 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 subsequent works.

5 An application

We illustrate the Bayesian approaches developed in this paper for the Skew-probit IRT model, using the data set corresponding to the Mathematics Test applied in Peruvian schools. The data set is describe in Bazán *et al.* (2004) and corresponds for the application of 14 items of the Mathematical Test available for download at `http://www.minedu.gob.pe/umc/` applied to 131 students of high social-economical status.

Figure 4 shows how scores present negative asymmetry in the behavior of sixth grade students for the mathematical test. The sample skew indexes is -0.804. The distribution of the scores presents a mean value of 10.84 and a standard deviation of 0.449. By considering this asymmetry we can explore the use of the Skew-probit IRT model for this data set.

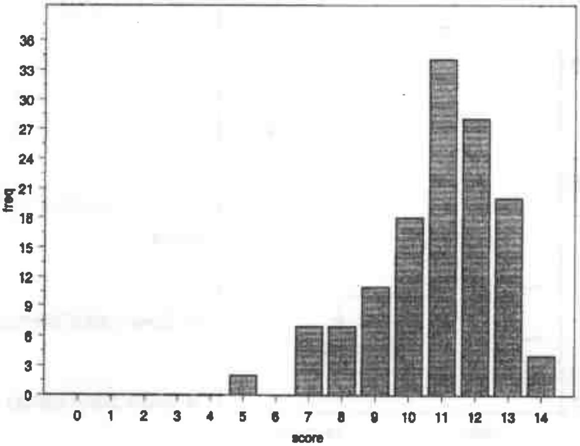

Figure 4: Histogram for scores corresponding to the mathematics test for n=131 individuals and k=14 items

In order to for use the skew-probit model we use two different combinations of prior distribution for the item parameters as suggested in Section 3.1. They are:
Priors A: $a_j \sim N(0,1)I(a_j > 0)$, $b_j \sim N(0,10000)$, $\delta_j \sim U(-1,1)$ and
Priors B: $a_j \sim N(1,0.5)I(a_j > 0)$, $b_j \sim N(0,2)$, $\delta_j \sim U(-1,1)$.

The MCMC procedure described in Section 3 was implemented using `Winbugs` software. The Skew-probit model involves 42 item parameters: $(a_j, b_j, \lambda_j(\delta_j)), j = 1, \ldots 14$. The approach also allows us obtaining individual traits for the 131 examinees in the sample $(U_i, i = 1, \ldots, 131)$, but we are interested only in the mean and standard deviation of the latent traits. As remarked in Section 4, since a great number of chains must be generated for the model due to the presence of autocorrelation, the MCMC procedure becomes slow. For inference purposes about the four proposed models, we generate 304000 iterations and discard the 4000 initial ones. Using a thin of 150 an effective size of 2000 was considered. The *lambda parameterization* was used.

Several criteria computed using the CODA package, including the ones proposed by Geweke (1992) were used to evaluating convergence. Estimates of model parameters are computed from these iterations.

To compare the probit-normal and skew-probit model, we computed the posterior expected deviance ($Dbar$), the deviance of the posterior means ($Dhat$), the effective number of parameters ρ_D and deviance information criterion (DIC), as presented by Spiegelhalter *et al.* (2002). $Dbar$, is the posterior mean of the deviance that is defined as -2* log(likelihood). *Dhat* is a point estimate of the deviance obtained by substituting in the posterior means estimates of model parameters. ρ_D is given by $\rho_D = Dbar - Dhat$. DIC is given by $DIC = Dbar + \rho_D = Dhat + 2 * \rho_D$. The model with the smallest DIC is estimated to be the model that would best predict a replicated dataset of the same structure as that currently observed. As Bazán *et al.* (2004) observed, in the presence of auxiliary latent variables, marginal DICs for the observed variables must be considered since the focus of the analysis is $p(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\delta})$.

From DIC values shown in Table 2, we see that skew-probit model, improve the corresponding symmetric probit-normal model for any of the prior distributions considered. Spiegelhalter *et al.* (2002) mentions that ρ_D can be negative and it can indicate conflict between prior and data. This problem can be important in the skew-probit models when prior information is not available. Informative prior elicitation using historical data, as proposed by Chen, *et al.* (2001) can be explored in subsequent studies.

Table 2: Comparison of the Skew-probit and probit-normal IRT models using DIC

models	parameters	*Dbar*	*Dhat*	ρ_D	*DIC*
Probit-normal	159	1446.17	1358.13	88.04	1534.21
Skew-probit (A priors)	173	1315.88	1370.39	-54.511	1261.37
Skew-probit (B priors)	173	1330.97	1367.5	-36.528	1294.44

On a preliminary analysis, Bazán *et al.* (2004), using prior scenarios A and B specified for the a and b parameters in the probit-normal model, indicated that the probit-normal model is insensitive to the prior specification. This result is not observed for the skew-probit model, which is sensitive for the specification of more diffuse priors.

Table 3 shows estimated correlations between the posterior means of the parameter corresponding to the two priors considered. Note that all item parameters are correlated under prior scenario B. Note that for priors A and B, parameters a and b present somewhat strong correlation among themselves. But it is important to observe that δ (or λ) in the two priors are not correlated. Estimates of δ or λ are different when the priors A and B are considered. The use of prior A implies that the posterior mean of δ (or λ) is closed to zero. We prefer to present inference results for prior B. Additionally, we choose to work with asymmetry parameter δ because it offers the same interpretation as the parameter λ in a simpler scale.

Estimates of item discrimination and difficulty parameters for the probit-normal and skew-probit models using prior B are presented in Figures 5. 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 model is a model that offers the same conclusions about difficulty and discrimination parameters as the Probit-normal model. In Figure 6 we present differences between probit-normal and skew probit-normal models (sum of absolute value of differences in a_j and b_j parameters) for the

Table 3: Correlation of posterior means for item parameters under the two prior scenarios

		A prior				B prior			
		a	b	δ	λ	a	b	δ	λ
A prior	a	1							
	b	-0.71	1						
	δ	-0.31	-0.32	1					
	λ	-0.23	-0.39	0.92	1				
B prior	a	0.99	-0.70	-0.31	-0.22	1			
	b	-0.64	0.99	-0.38	-0.45	-0.65	1		
	δ	-0.90	0.88	0.09	0.02	-0.87	0.83	1	
	λ	-0.91	0.74	0.27	0.19	-0.87	0.67	0.97	1

new item parameter δ_j.

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. This is not the case with items 11 and 4, which present penalty parameter estimates (δ or λ) large and negative (negative asymmetry on the item characteristic function), while discrimination and difficulty parameters 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-normal model is adequate for explaining the behavior of the items.

6 Discussion

This article proposes a new asymmetrical item response theory model. It considers a new asymmetric item characteristic curve by considering the cumulative distribution of the standard skew-normal distribution (Azzalini, 1985). This extends the work of Albert (1992) to fit asymmetrical IRT models and includes the symmetric 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. 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) is considered for the study of efficiency in the estimation of the parameters of the model introduced. We call attention to the importance of investigating such quantities when the data augmentation approach are used in the skew-probit model and there is strong evidence of autocorrelation in the parameter estimated that follow from the generated chains. However, ergodicity of the Markov chain follows from the regularity conditions required for the convergence of the Markov chain. But, one consequence

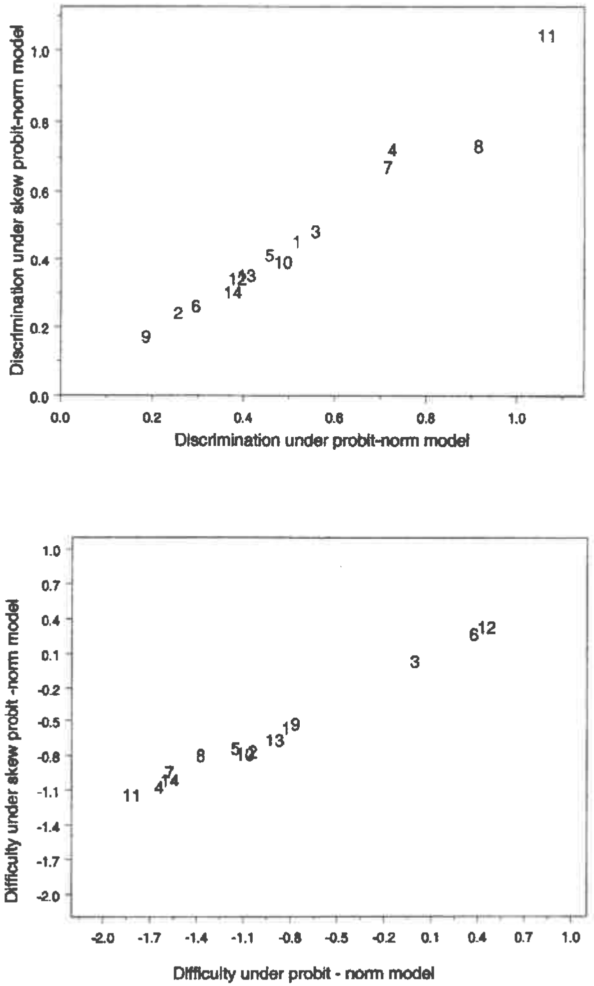

Figure 5: Discrimination and Difficulty parameter under probit-norm and skew probit-norm models

of this is that large number of iterations are necessary for assessing the variability of the estimates (sample mean, sample mode, quantile). This can be achieved by careful subsampling from the Markov chains, which also is a useful tool to minimize storage requirements. With the data set used in the Section 5, it was not possible to detect asymmetry when a vague prior was considered for the parameter b, as was the case with more precise priors. Subsequent studies on the study of sensitivity of

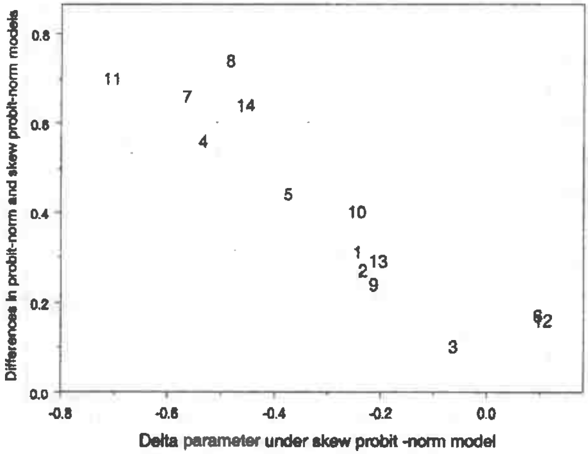

Figure 6: Differences in probit-norm and skew probit-norm models for delta parameter

the posterior of δ (or λ) with respect to prior choice is called for.
Another contribution of this article is the introduction of a skew probit link with a general representations for the likelihood of the data, which is not the case in Chen et al (1999). Comparison of symmetrical and asymmetrical 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 ICC in term 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 examines's ability. An example is give for illustration.

References

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

Albert, J. H. and Chib, S. (1993). Bayesian Analysis of binary and polytomous response data. *Journal of the American Statistical Association*, *88*, 669-679.

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

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

Azzalini A. and Capitanio A. (1999). Statistical applications of the multivariate skew normal distributions. *J. R. Statist. Soc. B, 61*, 579-602.

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

Bazán, J., Bolfarine, H., and 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.

Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Infering an Examinee's Ability. In F. M. Lord & M. R. Novick. *Statistical Theories of Mental Test Scores.* eading, MA : Addison-Wesley.

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

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*, 448, 1172-1186.

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

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

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

Czado, C., and 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.

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

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics, 1*, 141-150.

Genz, A. (1993). Comparison of methods for the computation of multivariate normal probabilities. *Computing Science and Statistics, 25*, 400-405.

Geweke, J. (1992). Evaluating the accuraty 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., and Chen, Ming-Hui & Agresti, A. (2000). Noninformative priors for one parameter item response models.*Journal of Statistical Planning and Inference. 88*, 99-115.

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

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

Johnson, V., and Albert, J. (2000). *Ordinal Data Modeling.* New York, MA: Springer-Verlag.

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

Kass, R. E., Carlin, B. P., Gelman, A. and Neal, R. (1998). Markov chain Monte Carlo in practice: A roundtable discussion. *American Statistician, 52*, 93-100.

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.

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., and Branco, M. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*, Vol. 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., and Gilks, W.R.(1996). *BUGS 0.5 examples* (Vol. 1 Version i). Cambrigde, UK: University of Cambride.

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

Tanner, M. A. (1996). *Tools for Statistical Inference.* Springer-Verlag: Heildelberg.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2005-01 - DE SOUZA BORGES,W., GUSTAVO ESTEVES, L., WECHSLER, S. Process Parameters Estimation in The Taguchi On-Line Quality Monitoring Procedure For Attributes. 2005.19p. (RT-MAE-2005-01)

2005-02 - DOS ANJOS,U.,KOLEV, N. Copulas with Given Nonoverlapping Multivariate Marginals. 2005.09p. (RT-MAE-2005-02)

2005-03 - DOS ANJOS,U.,KOLEV, N. Representation of Bivariate Copulas via Local Measure of Dependence. 2005.15p. (RT-MAE-2005-03)

2005-04 - BUENO, V. C., CARMO, I. M. A constructive example for active redundancy allocation in a k-out-of-n:F system under dependence conditions. 2005. 10p. (RT-MAE-2005-04)

2005-05 - BUENO, V. C., MENEZES, J.E. Component importance in a modulated Markov system. 2005. 11p. (RT-MAE-2005-05)

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

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05311-970 - São Paulo, Brasil