Bayesian estimation for Skewed Logistic

IRT models

BOLFARINE, Heleno; BAZAN, Jorge L.

January, 5, 2008

Abstract

A Bayesian Inference approach using MCMC is developed for two skewed Logistic Item Response Theory (IRT) models, namely, the Logistic Positive Exponent (LPE) model proposed by SAMEJIMA (1997, 2000) and the one named Reflection LPE model. The methodology developed is Bayesian and makes use of Markov chain Monte Carlo (MCMC) generation schemes for its implementation. It is illustrated using a data set corresponding to a Mathematical Test applied in Peruvian schools where comparisons with other parametric IRT models are also conducted. The main conclusion is that the item response model implemented seems to provide the best fit.

Keywords: asymmetrical link, binary response, Bayesian estimation, logistic positive exponent, model comparison, Skewed, logistic, item response.

\footnotesize

\footnotesize
\textsuperscript{1}Department of Statistics. University of São Paulo. Brazil. Caixa Postal 66281. CEP 05508-090
\textsuperscript{2}Department of Sciences. Pontifical Catholic University of Perú.
1 Introduction

Item Response Theory (IRT) models to dichotomous item responses assumes that the sequence of binary random variables \( \{Y_{ij} : 1 \leq i \leq n ; 1 \leq j \leq k \} \) associated with item responses are mutually independent. It is considered that \( Y_{ij} = 1 \) if subject \( i \) correctly answers item \( j \) and \( Y_{ij} = 0 \) otherwise. The response pattern of person \( i \) is written as \( Y_i = (Y_{i1}, \ldots, Y_{ik}) \). It is also assumed that the probability of the event \( Y_{ij} = 1 \) (correct response), namely, \( p_{ij} \), can be written as

\[
p_{ij} = P[Y_{ij} = 1 \mid \theta_i, a_j b_j] = F(m_{ij}),
\]

where \( F \) is called the item characteristic curve (ICC), and

\[
m_{ij} = a_j(\theta_i - b_j), \quad i = 1, \ldots, n, \quad j = 1, \ldots, k
\]

is linear predictor where \( a_j \) and \( b_j \) are parameters associated with the items (denomin-inated discrimination and difficulty parameters, respectively), and \( \theta_i \) is the value of the latent variable associated with the ability for individual \( i \).

Two known cases of ICCs follow by considering in (1) the cumulative distribution function (cdf) of the standard normal distribution and cdf of the standard logistic distribution. Such models are usually called the ogive normal IRT model and Logistic IRT model, respectively. In the context of generalized linear models, the inverse function of \( F(\cdot) \) in (1) is called the link function. by empathizing the correspondents link considered, the models can be named also as probit and logit IRT models, respectively. A special feature of both models is the symmetric nature of the probit and logit link or of the corresponding ICCs used. These models are also named point-symmetric (SAMEJIMA, 1997) models.

However, as emphasized in CHEN et al. (1999) in the context of binary regression, symmetric links do not always provide good fits for some data sets. This is specially true, when the probability of a given binary response approaches zero at a different rate than it approaches one. As pointed out by several authors, misspecifi-
cation of the link function can yield substantial bias in the mean response estimates (CZADO and SANTNER, 1992).

Moreover, as GARCIA-PEREZ (1999) points out, works on IRT models have almost exclusively been focused on the development and comparison of parameter estimation techniques and the study of the effects of the characteristics of the data sets (sample size, test length, distribution of the true abilities) and violations of model assumptions (excluding the mathematical form of the ICC) on the capability of available algorithms to recover the generating parameters. No one seems to have questioned if the mathematical form of the ICC can be derived from psychological theory of performance in objective testing as opposed to adopting a convenient function that the data are forced fitting to it. Using such arguments, SAMEJIMA (1997) has indicated the necessity of considering departures from normal (ability) assumptions in developing psychometric theories and methodologies, indicating that asymmetric ICCs are more appropriate for modeling human item response behavior. Thus, in successive papers, SAMEJIMA (1972, 1997, 2000) has presented the derivation of a Skewed Logistic IRT model, the one named logistic positive exponent (LPE) family, which consider asymmetric ICCs which can be appropriate since that symmetric ICC treats both correct and incorrect answers symmetrically, which results in a logical contradiction in ordering examinees on the ability scale (SAMEJIMA, 1997, p. 2).

The LPE model has a high level degree of substantive validity and inner consistency in ordering individuals (SAMEJIMA, 1997) and the point-symmetric (logistic) model is treated as one of the infinitely many models of the family but LPE seems to provide more appropriate models for human behavior than those on error distributions. Thus, the LPE model includes the logistic IRT model as special case so that it is a more flexible model to allow using symmetric (logit) and asymmetric (skew-logit) ICCs (links) for the items in the test. In addition, as proved in SAMEJIMA (2000), the contradiction in the rank order of response patterns does not exist in LPE. Moreover, it is not of our knowledge applications of estimation method-
ologies (either Bayesian or classical) developed for \textit{LPE} IRT models. It is specially true since that RUPP et al. (2004) reports the absence of Bayesian estimation for the Skewed Logistic IRT models.

In this paper we introduce Bayesian estimation to the \textit{LPE} model. In addition another IRT model is introduced. As we will see later, this new model is a reflection of the \textit{LPE} model and is named here as \textit{RLPE}. Both models can be considered as Skewed Logistic IRT models and have as its main characteristic a point-asymmetric ICC. As also we will see later, both model are derived by considering two links proposed by PRENTICE (1976).

Bayesian estimation will be developed by using the MCMC methodology and the WinBUGS software which can be used for simulating from the posterior distributions of item parameters and latent variables.

The objective of this paper is to offer a clear presentation of Bayesian estimation via MCMC for the two Skewed Logistic IRT models considered. The paper is organized as follows. In Section 2 we introduce the \textit{LPE} IRT model by considering a particular skew-logit link but also a new Skewed Logistic IRT model is proposed by considering another skew-logit link. In Section 3, we deal with Bayesian inference for the models considered. In section 4 we illustrate the methodology with simulated data sets since SAMEJIMA (2000). An example is given in Section 5 illustrating the usefulness of the approach in comparing it with other parametric IRT models to a real data set. To choose the model that fits the data better, we consider the deviance information criterion \textit{(DIC)} as presented in SPIEGELHALTER et al. (2002) and other models comparison criteria. Finally, we discuss possible extensions of the model proposed.
2 Models

2.1 Some Asymmetric ICCs

A variety of asymmetric ICCs or asymmetric links have been proposed for the binary regression models (see BAZAN, BOLFARINE and BRANCO, 2006), but hardly the two have been used in IRT models (skew-probit model proposed by Bazan, BRANCO, BOLFARINE, 2006, and the LPE model proposed by SAMEJIMA, 1997). However, there are several ways to obtain asymmetric ICCs to IRT models by taking in (1) the cdf of an asymmetric distribution. A very popular example of this situation is the well known and widely used complementary log-log link, where the cdf of the Gumbel distribution is considered. But the cdf of other distributions such as the Weibull and log-normal distributions can also be used to define new IRT models. In these cases, the cdf is completely specified and it does not depend on any unknown parameter and no relationship between them and the usual symmetric links are established being this a nuisance restriction. Other less restrictive ICCs to IRT models can be obtained when considering the following cdfs

\[
F_1(x) = 1 - (1 + e^x)^{-\lambda} \quad \text{and} \quad F_2(x) = (1 + e^{-x})^{-\lambda}, \quad \lambda > 0. \tag{3}
\]

In spite of the fact that there is no well agreed name for the first cdf, ACHEN (2002) has named it as the standard scobit distribution and the second cdf corresponds to the standard Burr type II distribution (JOHNSON et al., 1994).

Corresponding links using distributions \( F_1(.) \) and \( F_2(.) \) were proposed in PRENTICE (1976) and were popularized in the statistical literature by ARANDA-ORDAZ (1981) and in the econometric literature by NAGLER (1994) and ACHEN (2002). These links are skewed modifications of the logit link and are here termed \textit{scobit} and \textit{power logit}, respectively, and include the logit link as special case by considering the parameter \( \lambda = 1 \).

In general, if we define \( Y = \mu + \sigma X \), we say that a variable \( Y \sim Burr - II(\mu, \sigma) \) (“\( \sim \)” meaning “distributed as”) or \( Y \sim Scobit(\mu, \sigma) \). The corresponding probability
density function (pdf) are given by
\[ f_1(x) = \frac{\lambda \exp(x - \mu)}{\sigma[1 + \exp(x - \mu)]^{\lambda+1}} \]
and
\[ f_2(x) = \frac{\lambda \exp(x - \mu)}{\sigma[1 + \exp(x - \mu)]^{\lambda+1}} , \]
respectively, where \( \mu \) is a location parameter and \( \sigma \) is a scale parameter. For example, if \( Y \sim Burr - II(\mu, \sigma) \) then \( \frac{Y-\mu}{\sigma} \sim Burr - II(0,1) \).

Note that \( F_1(-y) \neq 1 - F_1(y) \) or \( F_2(-y) \neq 1 - F_2(y) \) and then \( F_1 \) and \( F_2 \) are not point-symmetric but \( F_1(-y) = 1 - F_2(y) \) and thus, the Burr-II and the Scobit distributions are distinct, though closely related since one is the reflection of the other.

2.2 LPE and the Reflection of LPE IRT models

Notice that the ICCs arising from the several links proposed above may be new IRT models but they are not derived from psychological theory of performance as GARCIA-PEREZ (1999) points out. The one model derived from a psychological theory (that is, originated from psychological motivation) is the LPE IRT model (see SAMEJIMA, 1997, 2000). The LPE IRT model can be obtained by considering the power logit link or, equivalently, when the Burr type II distribution \( F_2(.) \) is considered in (1). Moreover, the formulation of the LPE IRT model essentially implies that the Logistic IRT model is nested within the LPE IRT model. However, another interesting IRT model can be obtained when the Scobit link or the Scobit distribution \( F_1(.) \) is considered. This other IRT model is nominated here as the reflection of the LPE IRT model, namely the RLP E IRT model. In general, we say that LPE and RLP E IRT models are Skewed Logistic IRT models and are obtained by considering.

\[ p_{ij} = P[Y_{ij} = 1 | \theta_i, a_j, b_j] = F_{\lambda_j}(m_{ij}), \quad (4) \]
where \( F_{\lambda_j} \) is the cdf \( F_2(.) \) (or \( F_1(.) \)) in (2) indexed with \( \lambda_j \) and evaluated at \( m_{ij} \). When \( F_2(.) \) is considered we have the LPE model and when \( F_1(.) \) is considered we
have the RLPE model.

The LPE IRT model can also be obtained within SAMEJIMA’s formulation, by considering \( p_{ij} = L(m_{ij})^{\lambda_j} \), with \( L(.) \) the cdf of the standard logistic function and \( \lambda_j > 0 \) is the shape parameter associated with the \( j \)-th item, which provides asymmetric ICCs and includes the Logit IRT model as a special case when \( \lambda_j = 1 \). In these formulations, the LPE model is as a generalization of the logit link, which follows by introducing a shape parameter associated to the item that is, it interpreted as a penalization item parameter and can play an important role in testing as has been mentioned by SAMEJIMA (1997, 2000).

Figure 1: Probability curves for \( \lambda = 0.4, 0.6, 1, 2, 8 \) in LPE (up) and RLPE (below) models considering different ranges for \( \theta \) and \( a = 1, b = 0 \)

Figure 1 depicts different probability curves or ICC’s of the LPE and RLPE with logistic IRT models by using different values for \( \theta \) for an item with \( a = 1 \) and \( b = 0 \). For \( \lambda = 1 \), the ICC corresponds to the logistic IRT model and for \( \lambda < 1 \) (or \( \lambda > 1 \)) the ICC is generally above (below) the ICC corresponding to the logistic IRT model for a range of ability values. Note also that to each value of \( \lambda \), RPLE ICC is a reflection of the LPE ICC.
In \( LPE \), if \( \lambda < 1 \), then even individuals with very low ability levels have substantially high probabilities to pass the item. In this case, if a test consists of items with common values of \( a \), \( \lambda < 1 \) and different values of \( b \), individuals failing to solve easier items are penalized. On the other hand, when \( \lambda < 1 \), then even individuals with high ability levels have a substantially low probability to pass the item. In this case, if a test consists of items with common values of \( a \), \( \lambda > 1 \) and different values of \( b \), individuals succeeding in solving more difficult items are rewarded.

In \( RLP E \), if \( \lambda < 1 \), then individuals presenting very low ability levels have substantially low probabilities of failing the item. In this case, if a test consists of items with common \( a \) values, \( \lambda < 1 \) and different \( b \)'s, individuals with low abilities correctly solving easier items are rewarded. Alternatively, if \( \lambda < 1 \), then individuals with high ability levels have a substantially high probability of failing the item. In this case, if a test consists of items with common \( a \) values, \( \lambda > 1 \) and different \( b \) values, credits are given to fail in solving more difficult items.

3 Inference

The likelihood function for the Skewed Logistic IRT class (or family) of models indexed by \( \lambda_j \) is given by

\[
L(\beta, \theta | y, X) = \prod_{i=1}^{n} \prod_{j=1}^{k} \left[ F_{\lambda_j}(m_{ij}) \right]^{y_{ij}} \left[ 1 - F_{\lambda_j}(m_{ij}) \right]^{1-y_{ij}},
\]

where \( \beta = (a', b')', a = (a_1, \ldots, a_n)' \), \( b = (b_1, \ldots, b_n)' \), \( m_{ij} \) is the latent linear predictor in (2) and \( F_{\lambda_j}(m_{ij}) \) is the cdf \( F_2(.) \) (or \( F_1(.) \)) in (4), indexed with \( \lambda_j \) and evaluated at \( m_{ij} \).

In this section we present a complete-data likelihood function for the Skewed Logistic IRT models, so that we start with an important alternative representation.
Proposition. The Skewed Logistic models, as defined before, can be equivalently written as

\[ y_{ij} = I(s_{ij} > 0) = \begin{cases} 1, & s_{ij} > 0; \\ 0, & s_{ij} \leq 0, \end{cases} \quad i = 1, \ldots, n, \quad j = 1, \ldots, k, \]  

where \( s_{ij} \sim \text{Scobit}(m_{ij}, 1) \) if the LPE model is defined, or \( s_{ij} \sim \text{Burr-II}(m_{ij}, 1) \) if the RLPE model is defined, with \( I(.) \) as the usual indicator function.

Proof. Note that to the LPE model

\[ P(Y_{ij} = 1) = P(S_{ij} > 0) = 1 - P(S_{ij} \leq 0) = 1 - F_1(-m_{ij}) \]

\[ = 1 - [1 - (1 + \exp(-m_{ij})^{-\lambda_j})] = [1 + \exp(-m_{ij})^{-\lambda_j}] = F_2(m_{ij}), \]

and,

\[ P(Y_{ij} = 0) = P(S_{ij} < 0) = F_1(-m_{ij}) = 1 - [1 + \exp(-m_{ij})^{-\lambda_j}] = 1 - F_2(-m_{ij}). \]

A similar proof can be presented for the RLPE model. □

The latent variables \( s_{ij} \)'s are introduced to avoid working with Bernoulli type likelihoods and this representation shows a latent linear structure producing equivalent models for the LPE or RLPE classes. Therefore, the complete-data likelihood function for the Skewed Logistic IRT model is given by

\[ L(\alpha, b, \theta | s, y) = \prod_{i=1}^{n} \prod_{j=1}^{k} f^*_\lambda(s_{ij})p(y_{ij} | s_{ij}), \]  

where \( p(y_{ij} | s_{ij}) = I(s_{ij}, y_{ij}) = I(s_{ij} > 0)I(y_{ij} = 1) + I(s_{ij} \leq 0)I(y_{ij} = 0), \)

\( i = 1, \ldots, n, j = 1, \ldots, k, \) and \( f^*_\lambda \) is the pdf of the distribution corresponding to the reflection of \( F_\lambda \) considered in (5). That is, we use the pdf of the Scobit distribution if LPE model is assumed or the pdf of the Burr-II distribution if the RLPE is assumed. In both cases, when \( \lambda_j = 1 \), the corresponding result for the \( L(.) \) model follows, similarly as the result presented in Albert (1992) for the probit IRT model. The result in (7) can also be obtained by considering a latent linear
structure for the Skewed Logistic IRT model, namely

\[ s_{ij} = m_{ij} + e_{ij}, \quad e_{ij} \sim F^\ast(.) \]

that is, \( e_{ij} \) is the equation error distributed as the standard reflection distribution considered in (5). Note that the errors \( e_i \) are independent and are “latent data” residuals (ALBERT and CHIB, 1995), when considering \( e_{ij} = s_{ij} - m_{ij} \), which is estimated using the generated data, they can also be used for model checking. To understand how the observations \( y_{ij} \) change the distribution of these residuals, we consider the posterior distribution of \( e_{ij} \) conditional on \( a_j, b_j, \theta_i \) and \( s_{ij} \), that is,

\[ e_{ij}^\ast = e_{ij} | a_j, b_j, \theta_i, y_{ij}, s_{ij} \]

Note that in the Skewed Logistic IRT model, the parameters \( \lambda \) and \( a, b, \theta \) have quite different meaning. On one hand, \( \lambda \) is a vector of structural parameters associated with the choice of the link function. On the other hand, traditional IRT parameters of the Logistic IRT model \( a, b, \theta \) are a vector of parameters inherent to the observed data and not depending on model choice (for a discussion, see, for example, TAYLOR and SIQUEIRA, 1996). By considering this fact, two scenarios can be considered. The first scenario is one in which \( \lambda \) and the traditional IRT parameters are jointly estimated; in the second scenario only traditional parameters \( a, b, \theta \) are allowed to vary and \( \lambda \) is fixed at its “true” value \( \lambda_0 \). As in TAYLOR and SIQUEIRA (1996), we shall refer to these two scenarios as the unconditional and conditional ones, respectively.

Inference with the conditional scenario is easier to be implemented from both maximum likelihood (ML) and Bayesian approaches because it corresponds to a particular \( L \) model by considering a fixed value \( \lambda_0 \) for the parameter \( \lambda \). However, conditions shall be imposed for the existence of the ML estimators and the posterior distribution of \( a, b, \theta \) under improper uniform priors.

Regarding the unconditional approach, computing the ML estimators using the versions of the likelihood functions given in Section 2 is not simple and is necessary to develop new conditions for the existence of the ML estimators. Additionally, it
is important to study the propriety of the posterior distribution under improper uniform priors for \( b, \theta \) and \( \lambda \) but considering proper priors for \( a \). These aspects are being directed to another paper. In this paper as well as in other papers dealing with IRT modeling, the existence of the ML estimators is assumed and proper priors are considered to the parameters in the model.

### 3.1 A Bayesian approach

In this paper, we adopt mainly the Bayesian view. Our point is made since several researchers demonstrated that accurate estimation of the item parameters in small samples can only be accomplished through a Bayesian approach (see, for example, SWAMINATHAN et al., 2003). Given the peculiarities of IRT models, maximum likelihood totally relies on large sample theory, which even for a large number of examinees, can be complicated by the presence of incidental parameters. Researches using such an approach typically do separate estimation for item and ability parameters. However, there is no way to jointly evaluate estimates precision (Patz and Junker, 1999). Because of this, an EM type algorithm as the one given by BOCK and AITKEN (1981) is preferable. Such problems do not occur with the Bayesian approach in which, for a large number of examinees, the prior distribution has little effect on the posterior distribution (SINHARAY and JONHSON, 2003).

**Prior specification**

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

In the IRT literature there seems to be consensus with respect to the prior for \( \theta \), that is, it is usually assumed that \( \theta_i \sim N(0,1) \) for \( i = 1, \ldots, n \), but different priors
have been investigated for the traditional item parameters $a_j$ and $b_j$ (see RUPP et al., 2004). Empirical evidence (see PATZ and Junker, 1999, among others) seems to indicate the presence of posterior correlation between item parameters. However, it seems difficult to assign dependent priors for those parameters, being a specially hard task thinking about values for the correlations for such priors, even if a multivariate normal prior is specified. Hence, we prefer using independent and common priors for $a$, $b$ and $\lambda$ and let such correlations be only data dependent. That is, the prior we consider can be written as

$$\pi(\theta, a, b, \lambda) = \prod_{i}^{n} \phi(\theta_i) \prod_{j}^{k} \pi_1(a_j) \pi_2(b_j) \pi_3(\lambda_j). \quad (9)$$

where $\phi(.)$ is the pdf of the normal distribution with mean $\mu$ and variance $\sigma^2$ and $\pi_1(.)$, $\pi_2(.)$, $\pi_3(.)$ are the prior pdf for parameters $a_j$, $b_j$ and $\lambda_j$, respectively.

Although some authors as ALBERT (1992) and Fox and GLASs (2001, 2003), use improper noninformative priors for the parameters $a_j$ and $b_j$ of the type $\pi_1(a_j) \pi_2(b_j) = I(a_j > 0)$, we prefer using informative priors on the discrimination parameters $a_j$ since the existence of the joint posterior distribution is not guaranteed when an improper prior is used. Considering the results of ALBERT and GHOSH (1999), the distribution of the discrimination parameter must be proper to guarantee a proper joint posterior distribution.

Several informative distributions for $a_j$ have been proposed in the literature. To mention just a few, a) BRADLOW et al. (1999) and ALBERT (1992) use the $N(\mu_a, \sigma_a^2)$ with or without hyperparameter distributions specified for $\mu_a$ and $\sigma_a^2$, respectively; b) KIM et al. (1994) and PATZ and Junker (1999) use the $LN(\mu_a, \sigma_a^2)$ with or without hyperparameter distributions specified for $\mu_a$ and $\sigma_a^2$, respectively, where $LN(.)$ is the log-normal distribution and c) SPIEGELHALTER et al.(1996) and SAHU (2002) use the $HN(\mu_a, \sigma_a^2)$ where $HN(.)$ is the Half-normal distribution with known values for $\mu_a$ and $\sigma_a^2$ and, finally, d) SWAMINATHAN and GIFFORD (1985) use the $IG(m, n)$, the inverted gamma distribution with (known) hyperparameters $m$ and $n$. We consider in this paper the specifications in b) above since
$a_j > 0$ and also for conjugational reasons.

When independent informative priors are considered for the item parameters, it is usually assigned the $N(\mu_b, \sigma^2_b)$ for $b_j, j = 1, \ldots, k$. Moreover, in the common situation where little prior information is available about the difficulty parameter, one can choose $\sigma^2_b$ to be large. As is mentioned in ALBERT and GHOSH (1999) in the Logistic IRT model, this choice will have a modest effect on the posterior distribution for non extreme data, and it will result in a proper posterior distribution when extreme data (all items are correct or all items are incorrect) is observed. Thus, vague priors can be admissible for the difficulty parameter.

In this paper, we consider $\mu, \sigma^2, \mu_a, \sigma^2_a, \mu_b, \sigma^2_b,$ to be known. In more general situations, the prior structure needs to be enlarged so that hyper prior information can also be considered for those parameters. The same prior distribution is considered for the $\lambda_j$ parameters. We consider the prior specified in the example 5.7 of CARLIN and LOUIS (2000) in the context of binary regression using the power link. That is $\lambda_j \sim \text{Gamma}(0.25, 0.25)$.

**MCMC Bayesian estimation**

Considering the likelihood function in (5) or (7) and the general prior specification given in (9) Bayesian estimation can be implemented by considering implementation of Markov Chain Monte Carlo methods which make it simple to implement efficient sampling from the marginal posterior distributions.

In the first case, the specification by considering hierarchical structures is easy in WinBugs (see appendix) but not when the model is implemented using other programs. We point out that to implement a Bayesian estimation procedures involving a Bernoulli type likelihood can be complicated since the integrals involved to obtain the marginal posterior distributions with likelihood (5) are difficult. On the other hand, in the second case, by considering the latent structure in (7), and the approach based on data augmentation that was introduced in the Section 3,
the full conditionals for the Skewed Logistic IRT model and the Bayesian inference via MCMC follows without complications, similarly as reported in ALBERT (1992) when the ordinary model is probit model is implemented. Note that some of the full conditionals can not be directly sampled from, requiring algorithms such as the Metropolis-Hastings. However, to implement the Bayesian approach in WinBUGS considering directly the likelihood function in (7) it is necessary to have the cdf of the Burr or Scobit distribution which, to the best of our knowledge, is not yet implemented with the software.

In the remainder of this paper we develop a computational procedure for the Skewed Logistic IRT model based on the original likelihood function in (5). Hierarchically, the full likelihood specification is given as follows:

\[ y_{ij} | a_j, b_j, \theta_j, \lambda_j \sim \text{Ber}(F_\lambda(a_j \theta_i - b_j)); \]  
\[ \theta_i \sim N(\mu, \sigma^2); \]  
\[ a_j \sim \pi_1(\mu_a, \sigma_a); \]  
\[ b_j \sim \pi_2(\mu_b, \sigma_b^2); \]  
and  
\[ \lambda_j \sim \pi_3(.). \]

Note that when \( \lambda_j = 1 \), the hierarchical structure of the likelihood corresponding to the \( L \) model follows by eliminating the fifth lines in the above hierarchy. The program code in WinBUGS used in this application is presented in the Appendix.

### 3.2 Models comparison criteria using MCMC Outputs

A variety of methodologies exist to compare alternative Bayesian model fits but the principal criteria used in those works are the Deviance Information Criterion (DIC) proposed by SPIEGELHALTER et al. (2002), and the Expected Information
Criteria corresponding to Akaike (\(EAIC\)) and Schwarz or Bayesian (\(EBIC\)) as proposed in CARLIN and LOUIS (2000) and BROOKS (2002). The criteria are based on the Posterior mean of the deviance: 
\[
E[D(a, b, \lambda, \theta)] ,
\]
where
\[
D(a, b, \lambda, \theta) = -2 \ln(p(y|a, b, \lambda, \theta)) = -2 \sum_{i=1}^{n} \ln P(Y_{ij} = y_{ij}|a, b, \lambda, \theta),
\]
which is also a measure of fit that can be approximated by using the MCMC output, by considering
\[
D_{\text{bar}} = \frac{1}{G} \sum_{i=1}^{G} D(a^g, b^g, \lambda^g, \theta^g),
\]
where the index \(g\) represents the \(g\)-th realization of a total of \(G\) realizations, and is the Bayesian deviance.

\(EAIC\), \(EBIC\) and \(DIC\) can be estimated using the MCMC output by considering
\[
\widehat{EAIC} = D_{\text{bar}} + 2p,
\]
\[
\widehat{EBIC} = D_{\text{bar}} + p \log N,
\]
and
\[
\widehat{DIC} = D_{\text{bar}} + \widehat{\rho}_D = 2D_{\text{bar}} - D_{\text{hat}},
\]
respectively, where \(p\) is the number of parameters in the model, \(N\) is the total number (that is, \(N = k \times n\)) of observations and \(\rho_D\), the effective number of parameters, is defined as
\[
\rho_D = E[D(a, b, \lambda, \theta)] - D[E(a), E(b), E(\lambda), E(\theta)],
\]
where \(D[E(a), E(b), E(\lambda), E(\theta)]\) is the deviance of posterior mean obtained when considering the mean values of the generated posterior means of the model parameters, which is estimated by
\[
D_{\text{hat}} = D\left( \frac{1}{G} \sum_{i=1}^{G} a^g, \frac{1}{G} \sum_{i=1}^{G} b^g, \frac{1}{G} \sum_{i=1}^{G} \lambda^g, \frac{1}{G} \sum_{i=1}^{G} \theta^g \right).
\]

Given the comparison of two alternative models, the model that fits better a data set is the model with the smallest value of the Posterior Mean of the Deviance,
DIC, EBIC and EAIC. In EAIC and EBIC, $2p$ and $p \log N$ are fixed to penalize the posterior mean of the deviance whereas in IRT models, $p$ is the number of parameters of the model (in the Skewed Logistic IRT model, $p = 3k + n$) and $N$ is the total number of observations. Moreover, as there is no consensus in the use of the Deviance of the posterior mean (see discussion in SPIEGELHALTER et al., 2002), the use of more than one criteria seems more appropriate to perform models comparison.

4 Illustration on simulated data

In order to evaluate the Bayesian estimation of the Skewed Logistic IRT models presented and evaluate the performance of models comparison criteria we reply the estimation showed in Table 2 in SAMEJIMA (2000). Table 2 in SAMEJIMA (2000) shows, considering Maximum Likelihood (ML), the Estimates of $\theta$ based on 32 Response Pattern of 5 Dichotomous Items Following $2L$, $2P$ and $LPE$ model with $\lambda = 2$, $a_j = 1$ for All Items and $b = c(-3.0, -1.5, 0.0, 1.5, 3.0)$, respectively.

Since that ML estimation is not possible to extreme cases (Response Pattern 1,1,1,1,1 and 0,0,0,0,0), in order to be comparable the results we conduct Bayesian estimation for the other Response Patterns. In addition we consider a vague prior for $\theta$ that is $\theta_j \sim N(0, 1000)$ and use the posterior mean of $\theta$ as estimates.

By considering the sum of squares of the differences of the estimates, namely

$$\sum_{i=1}^{30} (\theta_{i,ML} - \theta_{i,B})^2$$

we found the minimum differences between the estimates of $\theta$ under ML and Bayesian approaches (the values were 0.044, 0.099, 0.321 for Probit, Logistic and LPE IRT models, respectively).

However, SAMEJIMA (1997) showed no comparisons of the models proposed con-
sidering any of the existing criteria. In Table 1 we show results for several Models Comparison Criteria discussed in Section 3.2, including the RLPE model.

Table 1: Models Comparison Criteria for abilities under Logistic, Probit, LPE and RLPE IRT models based on 32 Response Patterns of 5 Dichotomous Items given by SAMEJIMA (2000)

<table>
<thead>
<tr>
<th>Models</th>
<th>$D_{bar}$</th>
<th>$D_{hat}$</th>
<th>$\rho_D$</th>
<th>DIC</th>
<th>EAIC</th>
<th>EBIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic</td>
<td>470.4</td>
<td>439.919</td>
<td>30.525</td>
<td>501.0</td>
<td>530.4</td>
<td>535.7</td>
</tr>
<tr>
<td>Probit</td>
<td>555.1</td>
<td>525.064</td>
<td>29.996</td>
<td>585.1</td>
<td>615.1</td>
<td>620.3</td>
</tr>
<tr>
<td>LPE</td>
<td>556.9</td>
<td>526.374</td>
<td>30.488</td>
<td>587.4</td>
<td>616.9</td>
<td>622.1</td>
</tr>
<tr>
<td>RLPE</td>
<td>557.3</td>
<td>526.229</td>
<td>31.046</td>
<td>588.3</td>
<td>617.3</td>
<td>622.6</td>
</tr>
</tbody>
</table>

By considering this criteria, there is strong indication that Logistic IRT model is the better model for the abilities. A simple correlation of the abilities estimated under the models with the observed scores showed that $2L$ present the better correlation coefficient (0.99).

5 Application to real data

We illustrate the Bayesian approach developed in this paper for the application of the proposed Skewed Logistic IRT model using a data set corresponding to real data. We consider an analysis on the response pattern obtained by the application of a Mathematical Test to fourth grade students of the rural Peruvian Elementary Schools. Item response vectors are available from authors upon request and correspond to response of 974 students to 18 items qualified as binary responses (correct or incorrect). The scores present a mean of 8.27, a median of 8 and a standard deviation of 4.20. The skewness and kurtosis indexes are estimated as -0.075 and -0.836, respectively. The test presents a regular reliability index given by Cronbach’s alpha of 0.83, and presents a mean proportion of items of 0.449.
The Mathematical Test is formed with independent items corresponding to different tasks with different definitions. Given the latent ability $\theta$, it is considered that the correct responses to the items are independent. Furthermore, the autocorrelations within individual responses seems to be low, which provides additional support for the assumption of local independence.

We present next a study on the fit of the parametric IRT models discussed earlier with the Math Data. Symmetric $L$-IRT models with one-parameter, two-parameters and three-parameters are considered and denoted by $1L$, $2L$ and $3L$, respectively. Moreover, we implement the Bayesian approach for the Skewed Logistic IRT models as discussed in section 3.

Table 2: Results comparing the Skewed Logistic IRT models with other IRT models

<table>
<thead>
<tr>
<th>models</th>
<th>$p$</th>
<th>$Dbar$</th>
<th>$Dhat$</th>
<th>$\rho_D$</th>
<th>DIC</th>
<th>EAIC</th>
<th>EBIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symmetric $1L$</td>
<td>992</td>
<td>17243</td>
<td>16488.9</td>
<td>754.1</td>
<td>17997.1</td>
<td>19227</td>
<td>26936.6</td>
</tr>
<tr>
<td>2L</td>
<td>1010</td>
<td>16888</td>
<td>16030.7</td>
<td>857.3</td>
<td>17745.3</td>
<td>18908</td>
<td>26757.5</td>
</tr>
<tr>
<td>3L</td>
<td>1028</td>
<td>17006.2</td>
<td>16148.9</td>
<td>857.3</td>
<td>17863.5</td>
<td>19062.2</td>
<td>27051.6</td>
</tr>
<tr>
<td>Asymmetric LPE</td>
<td>1028</td>
<td>16854.9</td>
<td>16221.9</td>
<td>633</td>
<td>17488</td>
<td>18910.9</td>
<td>26900.3</td>
</tr>
<tr>
<td>RLPE</td>
<td>1028</td>
<td>16832.9</td>
<td>17553.3</td>
<td>-720.4</td>
<td>16112.5</td>
<td>18888.9</td>
<td>26878.3</td>
</tr>
</tbody>
</table>

Several criteria computed using the CODA package, including the ones proposed by GEWEKE (1992), were used to evaluate convergence.

DIC values shown in Table 1 seem to indicate that the Skewed Logistic IRT models, improve any other proposed model including the corresponding symmetric ones ($1L$, $2L$ and $3L$). Hence, we expect that ICC estimates are more precise with the Skewed Logistic IRT models.

By considering the $RLPE$ model and according to Figure 2 and Table 2, we found that items 14, 5, 8, 17, 4, 18 and 9 present significant asymmetries (that is, the corresponding HPD intervals does not include $\lambda = 1$). Moreover, for these items, clearly the ICCs for the $L$ and $RLPE$ IRT models are different. Figure 3
Figure 2: Box-Plots for the $\lambda$ parameters for the 18 items Math data set under the RLPE IRT model 

depicts the ICCs fit considering the $L$ and RLPE IRT models to the item 14, clearly showing differences among them. That is, to values of abilities greater than $-1.5$, the probability of correct response to the item is greater with the RLPE IRT model or, equivalently, to higher ability values, there exists a penalty for failing to solve an easy item (SAMEJIMA, 2000).

6 Final Discussion

This article proposes a new ICC to Item Response Theory, named Reflection of LPE or RLPE model and implement a Bayesian approach to the already studied LPE model. The two models are named Skewed Logistic IRT models and is proved that one is the reflection of the other. This corresponds to the proposal of SAMEJIMA
Figure 3: ICC to the Item 14 under the \textit{RLPE} (\(\lambda = 0,11\)) and \textit{L} IRT model (1997, 2000) to the fitting of asymmetrical IRT models and includes the symmetric logistic IRT model as a special case. A data augmentation approach is proposed to implement Bayesian estimation by using the MCMC methodology. The MCMC methodology can be implemented by using the Metropolis-Hasting algorithm. Several models comparison procedures are used to compare the symmetrical and asymmetrical IRT models (DIC, EAIC and EBIC). We also introduce latent residuals for the models and global discrepancy measures as the posterior sum of squares of the latent residuals. All these measurement shown that the Skewed Logistic IRT model presents better fit than the usual Logistic IRT model for the observed data. Moreover, to several items there is clear indication that the penalization parameter \(\lambda\), in the \textit{RLPE} IRT model, is different from one, indicated by the credibility interval, suggesting that asymmetric ICCs are more adequate. This parameter has a helpful interpretation in the context of the mathematical test. Moreover, the penalization parameter is conceptually different from the guessing parameter in the three-parameters model, which is valid to the case of items with multiple choices (not considered in the example studied). Extensions of the model proposed will be the subject of future developments as for example to consider the inclusion of the new
Table 3: $\lambda$ parameters in $RLPE$ model

<table>
<thead>
<tr>
<th>parameter</th>
<th>mean</th>
<th>median</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{14}$</td>
<td>0.120</td>
<td>0.096</td>
<td>0.0119</td>
<td>0.372</td>
</tr>
<tr>
<td>$\lambda_{17}$</td>
<td>0.190</td>
<td>0.172</td>
<td>0.035</td>
<td>0.463</td>
</tr>
<tr>
<td>$\lambda_{8}$</td>
<td>0.248</td>
<td>0.170</td>
<td>0.032</td>
<td>0.993</td>
</tr>
<tr>
<td>$\lambda_{4}$</td>
<td>0.344</td>
<td>0.240</td>
<td>0.036</td>
<td>1.252</td>
</tr>
<tr>
<td>$\lambda_{5}$</td>
<td>0.381</td>
<td>0.298</td>
<td>0.039</td>
<td>1.236</td>
</tr>
<tr>
<td>$\lambda_{7}$</td>
<td>0.388</td>
<td>0.226</td>
<td>0.028</td>
<td>1.483</td>
</tr>
<tr>
<td>$\lambda_{18}$</td>
<td>0.399</td>
<td>0.345</td>
<td>0.081</td>
<td>1.142</td>
</tr>
<tr>
<td>$\lambda_{16}$</td>
<td>0.482</td>
<td>0.333</td>
<td>0.054</td>
<td>1.959</td>
</tr>
<tr>
<td>$\lambda_{15}$</td>
<td>0.486</td>
<td>0.412</td>
<td>0.156</td>
<td>1.319</td>
</tr>
<tr>
<td>$\lambda_{6}$</td>
<td>0.523</td>
<td>0.393</td>
<td>0.098</td>
<td>1.725</td>
</tr>
<tr>
<td>$\lambda_{12}$</td>
<td>0.538</td>
<td>0.419</td>
<td>0.037</td>
<td>1.616</td>
</tr>
<tr>
<td>$\lambda_{9}$</td>
<td>0.600</td>
<td>0.537</td>
<td>0.170</td>
<td>1.414</td>
</tr>
<tr>
<td>$\lambda_{3}$</td>
<td>0.718</td>
<td>0.554</td>
<td>0.160</td>
<td>2.153</td>
</tr>
<tr>
<td>$\lambda_{11}$</td>
<td>1.034</td>
<td>0.798</td>
<td>0.259</td>
<td>3.515</td>
</tr>
<tr>
<td>$\lambda_{13}$</td>
<td>1.241</td>
<td>1.000</td>
<td>0.309</td>
<td>3.482</td>
</tr>
<tr>
<td>$\lambda_{1}$</td>
<td>1.460</td>
<td>1.178</td>
<td>0.357</td>
<td>4.087</td>
</tr>
<tr>
<td>$\lambda_{10}$</td>
<td>1.925</td>
<td>1.720</td>
<td>0.259</td>
<td>5.079</td>
</tr>
<tr>
<td>$\lambda_{2}$</td>
<td>2.089</td>
<td>1.84</td>
<td>0.691</td>
<td>4.835</td>
</tr>
</tbody>
</table>

parameter in one-parameter and testlet models (BRADLOW, et al., 1999). Extensions to more general models such as Multidimensional, hierarchical and Multilevel Skewed Logistic IRT models will be the subject of future developments. A another direction is to conduct a sensitivity analysis of prior specification for the Skewed Logistic IRT model as the one conduced in BAZAN, BOLFARINE and LEANDRO (2006).
References


**Program**

We describe in the sequel a program in WinBUGS used to implement the data augmentation approach described in the paper.

```plaintext
model{
  for (i in 1:n) { for (j in 1:k )
    { y[i,j]~dbern(p[i,j])
      m[i,j]<-a[j]*(theta[i]-b[j])
      #LPE Model
      logist[i,j]<-exp(m[i,j])/(1+exp(m[i,j]))
      p[i,j]<-pow(logist[i,j],lambda[j])
      #RLPE Model
      p[i,j]<-1-pow(1+exp(m[i,j]),-lambda[j])
    }
  }
  #abilities priors
  for (i in 1:n) { theta[i]~dnorm(0,1)}
  #items priors
  for (j in 1:k) {
    # usual priors
    b[j]~dnorm(0,1)
    a[j]~dlnorm(0,0.5)
    #CARLIN and Lois (2000, Example 5.7)
    lambda[j]~dgamma(0.25,0.25) }
}
```