



Item Response Theory in WinBUGS using *BayesianModeling* software

Jorge L. Bazán

Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, São Carlos-SP, Brasil -
jlbazan@icmc.usp.br

Francisco Torres-Avilés*

Universidad de Santiago de Chile, Departamento de Matemática y Ciencia de la Computación, Santiago,
Chile - francisco.torres@usach.cl

Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, São Carlos-SP, Brasil -
Postdoctoral Position

Abstract

This article aims to provide an introduction on the theory and practical analytic skills in fitting several Item Response Theory (IRT) models. IRT theory is explained with special emphasis on Bayesian approaches using MCMC methods. A worked example is used to illustrate our proposal. This example contains step-by-step guides on using the statistical languages R and WinBUGS to fit the IRT models. IRT model codes are obtained considering *BayesianModeling* software (Bazán, 2011). With this work, we want to encourage to professionals from the different areas of knowledge to apply IRT models in the re-analysis of existing data and for its application in future research.

Keywords: Item Response Theory; WinBUGS; Bayesian inference; Link function; MCMC methods.

1. Introduction

Since 90's, Bayesian methods have become popular in social sciences, health sciences, education and psychology due to its flexibility in accommodating different models to any situation in data analysis, when constraints must be considered to guarantee identifiability properties. One of the main applications where these methodologies has been widely used is Item Response Theory (IRT). IRT models are mostly used to model multivariate dichotomous responses resulting from n individuals, after answering a questionnaire (test or survey) containing k items. These models consider a unidimensional latent variable θ associated to individual abilities and a set of parameters associated to the items, which may represent the probability that the i -th examinee is able to answer the j -th item correctly.

For inference purposes when the Bayesian perspective must be implemented, it is common to use Markov chain Monte Carlo (MCMC) techniques to estimate the parameters associated to the IRT models (Fox, 2010). MCMC methods offer many advantages, including its ease implementation and the availability of free software. In contrast, MCMC algorithms can be quite sophisticated and their proper use require careful attention to several facets of implementation, for example, a heavy computational demand (Kim and Bolt, 2007). Furthermore, the use of BUGS based softwares allows to implement adequate and specific codes in order to fit traditional and new variations of IRT models, not available in other specialized software packages. Our proposal is to present the *BayesianModeling* software (Bazán, 2011) to generate syntax of several IRT models under a Bayesian approach, using MCMC methods. Therefore, the syntax can be executed in Bugs based programs such as OpenBUGS, WinBUGS or in R, the latter using the libraries R2WinBUGS, R2OpenBUGS, BRugs or rbugs.

BayesianModeling may be thought for practitioners that, given a data set, they wish to know the syntax of diverse IRT models in bugs format, usually non available in other specialized statistical programs including the program R. This program write two files: a bugs model file for each one of the available models, considering adequate priors, lists with sensible starting values and size of the data set, and finally a data set file in rectangular format, both readable in WinBUGS or OpenBUGS. Information about of the software can be downloaded for free from <http://argos.pucp.edu.pe/~jlbazan/software.html>.

2. Item Response Theory

We will discuss potential models that can be implemented in our program. Let consider data collected from n individuals who answered k different items of a specific test. A traditional Two-Parameter IRT model with a binary response, is a system where each person (\mathbf{Y}, Θ_i) has an unidimensional monotone latent variable model, defined by the following general hierarchical expression, for $i = 1 \dots, n, j = 1, \dots, k$

$$\begin{aligned} Y_{ij} | \theta_i, \eta_j &\sim \text{Bernoulli}(p_{ij}) \\ p_{ij} &= P(Y_{ij} = 1 | \theta_i, \eta_j) = F(m_{ij}) \\ m_{ij} &= a_j(\theta_i - b_j), \end{aligned} \tag{1}$$

where

- Y_{ij} represents the response variable which model the binary response of the person i that answer to the item j . The items have binary outcomes, i.e., the items are scored as 1 if correct and 0 if not.
- $\eta_j = (a_j, b_j)$ are two parameters that represent, respectively, the discrimination and the difficulty of the item j .
- θ_i is the value of the latent variable Θ_i for the person i , and some occasions it is interpreted as the latent or random ability of the person i .
- p_{ij} is the conditional probability given $\Theta_i = \theta_i$ to answer correctly to item j .
- F is called the item characteristic curve (ICC); and,
- m_{ij} is a latent linear predictor associated with the latent trait of person i and the parameters of item j .

The IRT models satisfies the property of latent conditional independence, that is, for a person i , the response Y_{ij} to the different items are conditionally independent given the latent variable $\Theta_i, i = 1 \dots, n$. Independence between persons and one-dimensional latent trait are also assumed. In addition, IRT models satisfies the property of latent monotonicity, because $F(m_{ij})$ is a strictly non decreasing function of $\theta_i, i = 1 \dots, n$ and $j = 1, \dots, k$, and $F^{-1}(\cdot)$ is called the link function.

On the other hand, the parameters associated to difficulty b_j and discrimination a_j represent the location and slope of the item respectively. Values as $a_j < 0$ are not expected. The parametric space for the parameter b_j is arbitrary and to be the same as θ_i , hence it usually takes values onto the real line.

A common parameterization for the latent linear predictor is $m_{ij} = a_j(\theta_i - b_j)$. The latter is very important from a computational point of view, since it facilitates the computational time of convergence. When this parameterization is used, the parameter of difficulty can be obtained from the quotient $\frac{b_j}{a_j}$. Generally, this

parameterization is preferred in the Bayesian Inference and also in Bayesian Modeling.

In The Two-Parameter IRT model, the conjoint density of the vector of multivariate responses $Y = (Y_1, \dots, Y_n)'$, with $Y_i = (Y_{i1}, \dots, Y_{ik})$ given the vector of latent variables $\Theta = (\Theta_1, \dots, \Theta_n)'$ and the vector of parameters of the items $\eta = (\eta_1, \dots, \eta_k)'$ can be written as,

$f(\mathbf{Y} | \Theta, \eta) = \prod_{i=1}^n \prod_{j=1}^k F(m_{ij})^{y_{ij}} (1 - F(m_{ij}))^{1-y_{ij}}$ The proof of this result is direct by considering the latent conditional independence.

The first IRT binary model was introduced by Lord (1952) with an ICC given by $F(x) = \Phi(x)$, being $\Phi(\cdot)$ the cdf of a standard normal variable. This model is known in the psychometric literature as *normal ogive model* which corresponds in the context of Generalize Linear Models to a probit link function and this is generally called as *2P model*.

On the other hand, Birbaum (1968) considered a ICC given by $F(x) = L(x)$, where $L(x) = \frac{e^x}{1 + e^x}$ denotes the cdf of a logistic variable, which induces to work with a *logit link function*. This model is known as the logistic model and it is known as *2L model*.

2.1. Some Particular cases

The IRT model allows for a wide class of formulations, which depends on the ICC. Its simplest version could take $a_j = 1$ and consider an ICC of the form $P(Y_{ij} = 1 | \theta_i, b_j) = F(\theta_i - b_j)$. This is called as one-parameter

IRT model and when probit or logit links are considered we have *1P* or *1L* IRT model respectively. In general, we can consider an ICC of the form $P(Y_{ij} = 1 | u_i, a_j, b_j, c_j) = c_j + (1 - c_j)F(a_j(\theta_i - b_j))$, where the parameter $c_j \in [0, 1]$ indicates the probability that individuals with very low ability answer correctly the item j by random chance. $F(\cdot)$ is the distribution function. This is known as the three-parameter IRT model. If $c_j = 0$, the model is reduced to the two-parameters IRT model. Again, when probit or logit links are considered we have *3P* or *3L* IRT model respectively.

From here, we can consider the following potential models,

- **The IRT model with logit link:**

The IRT model with logit link or logistic model is probably the model more used in IRT. The version of three parameters for this model establishes that the probability that the person i hit the item j is given by,

$P(Y_{ij} = 1 | \theta_i, a_j, b_j, c_j) = c_j + (1 - c_j) \frac{1}{1 + e^{-D a_j (\theta_i - b_j)}}$, where usually is assumed that $D = 1$ although some authors consider also the value $D = 1.7$ for approximating this model to the normal ogive model.

The model with $c_j = 0$ and $a_j = 1$ is known as Rasch model but it has own interpretations and derivations (see for example Fischer and Molenaar, 1995).

- **IRT Models with asymmetric links**

In the traditional IRT models, the ICC are generally considered symmetrical; this is the case of the logistic and normal ogive models. However, Samejima (2000), Bazán et al. (2006) and Bolfarine and Bazán (2011) proposed asymmetric ICC's considering a new parameter of item that controls the shape of the curve. This asymmetry is necessary in many cases for a better fitting of answers with a low proportion of 0s or 1's. Three of these models are,

- **The skew normal ogive model** The skew normal ogive model was proposed by Bazán et al. (2006) assuming that the probability of success is given by,

$$p_{ij} = P(Y_{ij} = 1 | \theta_i, \eta_j) = \Phi_{SN}(m_{ij}; \lambda_j),$$

$$i = 1 \dots, n, j = 1, \dots, k$$

where $\lambda_j \in R$ is a parameter of asymmetry, $m_{ij} = a_j(\theta_i - b_j)$ is the latent linear predictor and Φ_{SN} denotes the cdf of a skew normal distributions with density function $\phi_{SN}(x; \lambda) = 2\Phi(x)\Phi(\lambda x)$, being ϕ the pdf of a standard normal variable.

- **The LPE and RLPE models** Logistic positive exponent (LPE) was initially proposed by Samejima (2000). A reversal version, called Reflection Logistic positive exponent (RLPE) was developed by Bolfarine and Bazán (2010). The models studied in Bolfarine and Bazán et al. (2010) assume that the probability of correct response is given by

$$p_{ij} = P(Y_{ij} = 1 | \theta_i, a_j, b_j) = F_{\lambda_j}(m_{ij}),$$

$$i = 1 \dots, n, j = 1, \dots, k$$

where F_{λ_j} is the cdf of the logistic distribution indexed by the parameter $\lambda_j > 0$ and evaluated in m_{ij} . For LPE model $F_{\lambda_j} = L(m_{ij})^{\lambda_j}$ and for RLPE model $F_{\lambda_j} = 1 - L(-m_{ij})^{\lambda_j}$.

For both models λ_j can also interpret like a parameter of penalty or bonus in a similar way to the case of the model of skew normal ogive model given by Bazán et al. (2006).

For both asymmetric cases, extensions considering one-parameter or three-parameters are possible. More details of these models classes of models can be reviewed in Bazán, et al. (2006), Bolfarine and Bazán (2010) and Bolfarine and Bazán (2011).

3. Bayesian estimation in IRT

Let Y_{ij} be the dichotomous variable corresponding to the response of the i th individual, $i = 1, \dots, n$, on the j th item, $j = 1, \dots, k$, that is, Y_{ij} takes the value 1 if the response is correct and 0, otherwise. A three-parameter skew IRT model may follow the hierarchical formulation in (1) with a general form of $p_{ij} = c_j + (1 - c_j)F_{\lambda_j}(m_{ij})$.

The probability $p_{ij} = P(Y_{ij} = 1 | \theta_i, \eta_j)$ is the *conditional probability of correct response* given the i -th ability value θ_i and the j -th item parameter $\eta_j = (a_j, b_j, c_j, \lambda_j)$ where a_j, b_j, c_j, λ_j are named discrimination, difficulty, guessing and penalization item parameter respectively. Moreover, $F_{\lambda_j}(\cdot)$ denotes the cdf of the determined distribution indexed by λ_j and m_{ij} is the *latent linear component* relating θ_i and the item parameter η_j . $F_{\lambda_j}(\cdot)$ is denoted by an asymmetric cdf which includes some symmetric cases when the $\lambda_j = 0$. Hence, considering the Bernoulli distribution for the response variable, the likelihood function for IRT model in the three-parameter IRT model is given by

$$L(\mathbf{a}, \mathbf{b}, \mathbf{c}, \lambda, \theta, | \mathbf{y}) = \prod_{i=1}^n \prod_{j=1}^k \{F_{\lambda} [a_j(\theta_i - b_j); c_j]\}^{y_i} \{1 - F_{\lambda} [a_j(\theta_i - b_j); c_j]\}^{1-y_i}$$

where $F_{\lambda} [a_j(\theta_i - b_j); c_j]$ is the asymmetric cdf indexed by the parameter λ associated with the asymmetric ICC. Our particular cases, the Logistic (logit), Normal ogive (probit), LPE and RLPE use the form of this likelihood function to perform the inference procedures.

As it is known, in the Bayesian Inference the parameters of interest are assumed as random variables, therefore, we need to elicitate a priori probability distributions that reflect our previous knowledge over them. Combining the likelihood function and the prior distributions we will be able to obtain the posteriori distributions of the parameters of interest.

Bayesian Inference in IRT models is facilitated with the use of different methods MCMC implemented in WinBUGS or OpenBUGS software. For traditional IRT models, Bugs codes are available in Curtis (2010), Fox (2010), Bazán et al. (2010). Even when there exist Bayesian packages developed in Matlab, for the models presented here, there are no programs or codes to be implemented in WinBUGS or OpenBUGS. This is facilitated using BayesianModeling.

Vague proper prior distributions are considered in BayesianModeling, assuming independent distributions with big values for the variance. Prior distributions for traditional IRT models are discussed by Bolfarine and Leandro (2006) or Fox (2010). There is consensus about the prior specification for latent trait, where $\theta_i \sim N(0, 1)$ is assumed for $i = 1, \dots, n$. However, for the item parameter, there are several proposals. Here, we assume independent priors, that is,

$$\pi(\theta, \mathbf{a}, \mathbf{b}, \mathbf{c}, \lambda) = \prod_i^n \phi(\theta_i) \prod_j^k \pi_1(a_j)\pi_2(b_j)\pi_3(c_j)\pi_4(\lambda_j).$$

where $\phi(\cdot)$ is the pdf of a normal distribution and $\pi_1(\cdot), \pi_2(\cdot), \pi_3(\cdot), \pi_4(\cdot)$ correspond to the prior distributions of item parameters a_j, b_j, c_j and λ_j , respectively. In the special case of small sample size, we suggest the use of the following prior specification,

- $a_j \sim HN(\mu_a, \sigma_a^2)$ with $\mu_a = 1$ and $\sigma_a^2 = 0.5$ where $HN(\cdot)$ correspond to the positive normal or Half-normal distribution
- $b_j \sim N(\mu_b, \sigma_b^2)$ with $\mu_b = 0$ and $\sigma_b^2 = 2$
- $c_j \sim Beta(5, 17)$
- $\lambda_j \sim gamma(0.25, 0.25)$ for LPE and RLPE models or $\lambda_j = \frac{\delta_j}{\sqrt{1-\delta_j^2}}$ with $\delta_j \sim Unif(-1, 1)$, where $-1 \leq \delta_j \leq 1$ for Skew-Probit or Ogive skew normal model..

We will consider the traditional models to implement them in BayesianModeling program, classified as,

- Symmetric: logit (1L, 2L, 3L), probit (1P, 2P, 3P).


```

library(R2WinBUGS)
datos <- read.table("F:/MILUS/MathData.dat", header=TRUE, sep=" ", na.strings="NA",
dec=".", strip.white=TRUE)
n=nrow(datos)
k=ncol(datos)
data<-list(y=as.matrix(datos),n=131,k=14)
inits <- function(){ list(a=rep(1,k),b=rep(0,k),delta=rep(0,k),theta=rep(n,0.5))}
out<-bugs(data,inits,parameters.to.save=c("a","b","delta"),
model.file="F:/MILUS/modelirt.txt", n.chains=1, n.iter=24000,
n.burnin=4000, program="WinBUGS")

```

The results obtained after fitting this model using a specific type of data, are showed in Bazán et al. (2006). For more details about the software, including the use of different binary response models, review the work developed by Bazán (2011).

5. Discussion

Traditional IRT models have been widely used for several applications in science. Generally, their main characteristic is related to assuming that probability of correct response can be linked considering symmetric behaviors, such as logit or probit. However, alternative models with asymmetric links have been proposed recently. From a Bayesian point of view, traditional and new models can easily be implemented to existing BUGS codes.

We discussed recent models focussed in relaxing the assumption of symmetry in the probability of a dichotomous response. Furthermore, we present a Bayesian program to fit a wide class of models, which we call *Bayesian Modeling*. Even when this version can be considered as a beta version, the program seems to be a preliminary useful tool to fit a variety of statistical models, with clear applications in Social Sciences.

References

- Albert, J. H. (1992) Bayesian estimation of normal ogive item response functions using Gibbs sampling. *Journal of Educational Statistics*, 17, 251–269.
- Bazán, J. L. (2011). *Bayesian Modeling Users Guide*. Sciences Department. Pontificia Universidad Católica del Perú. Available at <http://argos.pucp.edu.pe/~jlbazan/software.html>
- Bazán, J. L., Branco, M. D., & Bolfarine, H. (2006). A skew item response model. *Bayesian Analysis*, 1, 861–892.
- Bazán, J. L., Bolfarine, H., & Branco, M. D. (2006). A generalized skew probit class link for binary regression. Technical report (RT-MAE-2006-05). Department of Statistics. University of So Paulo. URL <http://argos.pucp.edu.pe/~jlbazan/download/gspversion14.pdf>
- Bazán, J. L., Bolfarine, H., & Branco, M. D. (2010). A framework for skew-probit links in Binary regression. *Communications in Statistics - Theory and Methods*, 39, 678–697.
- Bazán, J. L., Bolfarine, H., & Leandro, A. R. (2006). Sensitivity analysis of prior specification for the probit-normal IRT model: an empirical study. *Estadística, Journal of the Inter-American Statistical Institute* 58(170-171), 17-42. URL <http://argos.pucp.edu.pe/~jlbazan/download/bazanestadistica.pdf>.
- Bazán, J. L., Valdivieso, L., & Calderón, A. (2010). Enfoque bayesiano en modelos de Teoría de Respuesta al Item. *Reporte de Investigación. Serie B. Nro 25*. Departamento de Ciencias. PUCP. URL <http://argos.pucp.edu.pe/~jlbazan/download/Reportef27.pdf>.
- Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In FM Lord, MR Novick (eds), *Statistical Theories of Mental Test Scores*. New York, Addison-Wesley.
- Bolfarine, H., & Bazán, J. L. (2010). Bayesian estimation of the logistic positive Exponent IRT Model. *Journal of Educational Behavioral Statistics*, 35, 6, 693–713.
- Fischer, G., & Molenaar, I. (1995). *Rasch Models Foundations, recent development, and applications*. The Netherlands, Springer-Verlag.
- Fox, J. P. (2010). *Bayesian Item Response Modeling: Theory and Applications*. New York, Springer.
- Kim, J-S., & Bolt, D. M. (2007). Markov chain Monte Carlo estimation of item response models. *Educational Measurement: Issues and Practice*, 26, 38–51.
- Lord, F. M. (1952). *A theory of test scores*. New York, Psychometric Society.
- Samejima, F. (2000). Logistic positive exponent family of models: Virtue of asymmetric item characteristic curves. *Psychometrika*, 65, 3, 319–335.