Bayesian item response modeling: an application for universities admission tests

Javier Martínez.
Scientific Computing & Statistics Dept., Universidad Simón Bolívar, Venezuela.
javiermartinez@usb.ve

Irene García Mosquera.
Mathematics & Informatics Dept., Universitat de les Illes Balears, Spain.
irene.garcia@uib.es

Abstract

We introduce a novel application that we have developed with Shyny on the problem of bayesian estimating dificulties, discrimination and chance to guess right answers of the issues from the admission test to universities. Our application implements a bayesian item response model, and we evaluated the performance of the model on a real data set. It consists on the responses that gave the students attending to the admission test to Universidad Simón Bolívar at Venezuela on 2012. Results were compared with those obtained through a standard method based on a classical statistical approach. In addition, we present a generalized regression model Gamma with log link to make a prospective analysis on the index of progress of that group of students.

Why are the admission tests important?

- Most four-year colleges or universities consider applicants scores on admission tests when deciding whom to accept.
- The relevance of test scores in the admission process varies from college to college and depends on the approach and policies of each institution.
- Admission tests apply a common standard to everyone. This helps to colleges evaluate and compare the preparation of students who come from different high schools.
- Admission tests like SAT evaluate the reading, writing and math skills that the student will need in college.
- In addition, the scores on the tests can be used to place students in classes that are in the right level for them. They can also identify students who may benefit from specific advisors or academic support in college

Item response models

Whereas classical test theory focuses on the test as a whole, item response theory (IRT) shifts its focus to the individual items (questions) themselves [1].

The item response theory (IRT) refers to a diverse family of models. Each model is designed to represent the relation between an individual's item response and an underlying latent trait.

In the IRT, the underlying trait is denoted by θ and represents the performance of the subject.

We focus on IRT models for dichotomously scored items (correct-incorrect) and we considerer two types of models:

- The 2 parameter model (2P) uses both item difficulty (b) and item discrimination (a), where a measures the extent that the item differentiates from low to top examinees. The higher values are better.
- The 3 parameter model (3P) uses item difficulty, item discrimination and the extent which candidates can guess the correct answer (c). Typically c is focuses on $1/k$ where k is the number of anwers options.

Specifically, for the 3P model, the probability that the i-th subject responds correctly the j-th item can be written as:

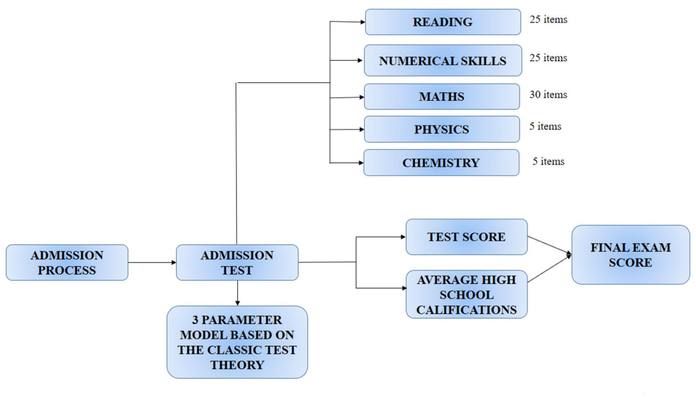
$$Pr(Y_{ij} = 1|\, \theta_i, a_j, b_j) = c_j + (1 - c_j) \cdot F(a_j \theta i - bj),$$

where F denotes the canonical member of the location-scale. Often, F is assumed to be a standard normal (our case) or standard logistic distribution function.

Admission test to Simon Bolivar University on 2012

The 80% of careers at Universidad Simon Bolivar (USB) are from engineering and sciences, and the total qualification of the admission process is the sum of 25% from the average notes that students get at high school and 75% from the admission test.

The test consists of 90 items from 4 areas (see Figure). Each item has only one right answer among 5 possibilities. One point is added for each right answer whereas 0.25 points is deducted for each wrong answers



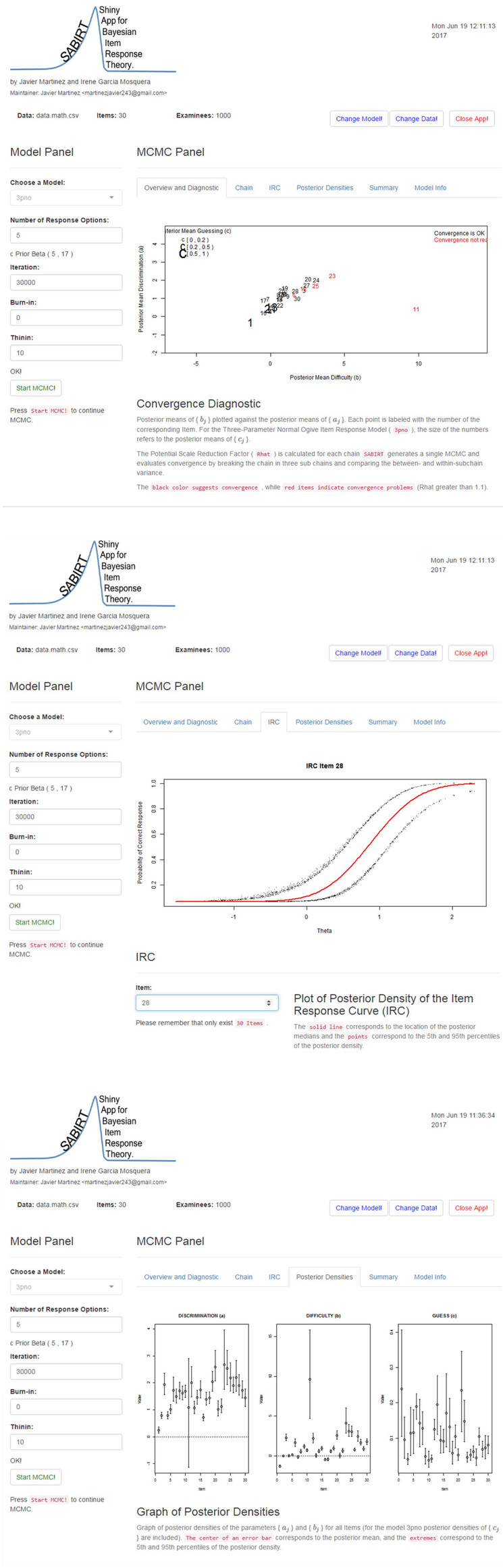
To fit a response item model with three parameters, the USB used the program PARAM-3PL (Calibration Software for the 3 Parameter Logistic IRT Model) created by [2] that estimates the parameters using a classical approach. The outcome of the program measures the quality of items.

The Technical Commission warned about some calibration errors in estimating difficulty and discrimination of items in Mathematics, Physics and Chemistry. Therefore, for the cohort of 2012 we proposed a Bayesian fit of the parameters in order to analyze the quality of the items. Besides, the positive predictability of the admission process was evaluated through the performance of the students after the first scholar year.

Shiny App for Bayesian IRT (SABIRT)

SABIRT is a Shiny application intended for the making of an MCMC estimation and model-fit of the item response models designed by [3] (2Pno) and [4] (3pno). To study convergence we used the potential scale reduction factor ($\hat{R}$) proposed by [5].

The outcome are the items parameters (difficulties and discrimination for 2pno, and additionally the chance to guess the right answers for 3pno) and also the latent abilities of each examinee.



Prospective analysis

One important issue for the Technical Commission is the assessment of the performance of the accepted students through the first scholar year. This was made by computing the Index of progress (IP), that index takes into account the notes that students get in the subjects studied as well as his or her advances during the course. Specifically:

$$IP = \left(\frac{\sum_i (SS_i \cdot CUS_i)}{TT} \right) \cdot \left(\frac{TPCU}{TMCP} \right),$$

where SS_i is the score on the subject i. CUS_i is the credit units of the subject i. TT is the total of taken subjects credit units. $TPCU$ is the total of passed credit units. $TMCP$ is the total of mandatory credits in the period.

Since the index values show a Gamma distribution, a generalized regression model Gamma with log link was fitted [6]. The most relevant variables to explain the Index of progress was the average notes from high school and the note in mathematics from the admission test.

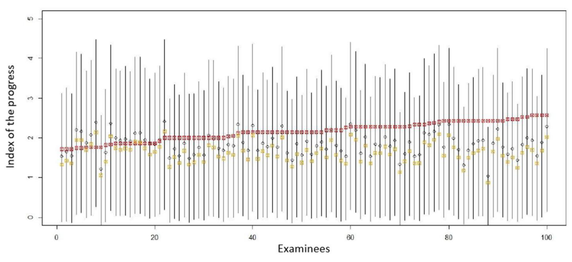


Figure 1: Prediction of the Index of progress for 100 examinees. The vertical line represents the 95% prediction interval. The red dots are the observed value. The orange dots are the bayesian prediction, and the black dots are the classic prediction.

Final remarks

- SABIRT helps the not familiar users with R or MCMC to obtain bayesian estimations of the item's quality in a dichotomous test (discrimination, difficulty and guessing).
- In the case of examinees on 2012 of the USB, the bayesian parameters improve the estimates of quality items. However, the predictibility of the admission test was low. It is necessary to take into account more explanatory variables for the prediction of the index of progress.

References

[1] DeMars, C. (2010). Item response theory.*New York-Oxford*.

[2] Lawrence, R. (2005). PARAM-3PL calibration software fot the 3 parameter logistic irt model (freeware). *http://edres.org/irt/param*.

[3] Johnson, V. & Albert, J. (1999). Ordinal data modeling. *New York. Springer*.

[4] Beguin, A. & Glas, C. (2001). MCMC estimation and some model-fit analysis of multidimentional IRT models. *Psychometrika*. **66**, pp. 541562.

[5] Gelman, A., Carlin, J., Stern, H. & Rubin, B. (2004). Bayesian data analysis. *New York. Chapman & Hall*.

[6] Ibrahin, J., Chen, M. & Sinha, D. (2015). Bayesian survival analysis. *Springer*.

Acknowledgments

The research reported has been partially supported by "La Caixa" grant for attendance at conferences, the Spanish Ministry of Economy and Competitiveness and European Regional Development Fund project, DPI2015-67082-P (MINECO/FEDER)(GC and JCP) and the S1 program to support the novel researcher of the Deanery of Research and Development in the Universidad Simón Bolívar, March 2017.

Obra Social "la Caixa"