Generalising IRT to Discriminate Between Examinees

Ahcène Boubekki
ULB Darmstadt/TU Darmstadt/DIPF
boubekki@dipf.de

Ulf Brefeld
TU Darmstadt/DIPF
brefeld@cs.tu-darmstadt.de

Thomas Delacroix
Telecom Bretagne
thomas.delacroix@telecom-bretagne.eu

ABSTRACT

We present a generalisation of the IRT framework that allows to discriminate between examinees. Our model therefore introduces examinee parameters that can be optimised with Expectation Maximisation-like algorithms. We provide empirical results on PISA data showing that our approach leads to a more appropriate grouping of PISA countries than by test scores and socio-economic indicators.

1. INTRODUCTION

Developments in Psychometrics have led to a multitude of logistic models, ranging from simple classical test theory to sophisticated multidimensional generalizations (e.g., [2]). Usually, these generalizations focus on items and the success of solving an item depends on a particular set of skills. On the contrary, examinees are only represented by their ability although, according to the original theoretical IRT problem, items and examinees are supposed to be treated symmetrically.

In this paper, we propose to balance this asymmetry by including a discrimination parameter for examinees. We present a *homographic* parametrization that preserves symmetry and allows to derive characteristics of examinees. We report on empirical results on PISA 2012 data showing that the use of *examinee discrimination parameters* reveals insights that cannot be identified with traditional approaches.

2. A SYMMETRIC AND LOGISTIC MODEL

The traditional 1PL model [5] is given by

$$IRF_{1PL}(i,j) = \frac{1}{1+e^{\theta_i+\beta_j)}}, \tag{1}$$

where the real numbers θ and β represent the examinee's ability and the item difficulty, respectively. These parameters can be related to the score x_i and the rate of success of the question a_j by using the transformations $\beta_j = \log\left(\frac{1-a_j}{a_j}\right)$ and $\theta_i = \log\left(\frac{1-x_i}{x_i}\right)$. Note that x_i and a_j are real numbers bounded by 0 and 1. After substitution, the model can be expressed as

$$IRF_{1PL}(i,j) = \frac{a_j x_i}{a_j x_i + (1-a_j)(1-x_i)}. \tag{2}$$

A similar transformation can be applied to the 2PL [1], where $\alpha_j = b_j$ are non negative real numbers called *item discrimination*,

$$\begin{aligned} IRF_{2PL}(i,j) &= \frac{1}{1+e^{\alpha_j(\theta_i+\beta_j)}} \\ &= \frac{(a_j x_i)^{b_j}}{(a_j x_i)^{b_j} + ((1-a_j)(1-x_i))^{b_j}}. \end{aligned} \tag{3}$$

The multidimensional two-parameter logistic model (M2PL) [2] splits the items in k different skills. The examinee has an ability parameter for each skill that is affected by a skill discrimination parameter. The ability is now a vector of real numbers $\boldsymbol{\theta}_i = (\theta_{i,1}, ..., \theta_{i,k})$ and the item discrimination a vector of non-negative real numbers $\boldsymbol{\alpha}_j = (\alpha_{j,1}, ..., \alpha_{j,k})$,

$$\begin{aligned} IRF_{M2PL}(i,j) &= \frac{1}{1+e^{\boldsymbol{\alpha}_j\boldsymbol{\theta}_i}+\beta_j} \\ &= \frac{a_j \mathbf{x}_i^{b_j}}{a_j \mathbf{x}_i^{b_j} + (1-a_j)(\mathbf{1}-\mathbf{x}_i)^{b_j}}. \end{aligned} \tag{4}$$

The appealing use of *item discrimination parameters* can be translated to examinees, for instance to distinguish between a regular scholarly student and a talented, yet slacking one. Let us introduce an *examinee discrimination parameter* denoted by the non-negative real number y_i that acts as the analogue of its peer b_j. The discrimination parameters will also be decoupled from the other item or examinee parameter. This assures the identifiability of the model. The resulting model is called the *Symmetric Logistic Model* (SyLM) and given by

$$\begin{aligned} IRF_{SyLM}(i,j) &= \frac{1}{1+e^{b_j\theta_i+y_i\beta_j}} \\ &= \frac{a_j^{y_i} x_i^{b_j}}{a_j^{y_i} x_i^{b_j} + (1-a_j)^{y_i}(1-x_i)^{b_j}}. \end{aligned} \tag{5}$$

At first sight, the logistic parametrization of the SyLM appears as a special case of the M2PL by setting $\beta_j = 0$ and renaming the parameters, however, the homographic parameterization renders them intrinsically different. Actually, SyLM is closer to the 2PL as it does not subdivide items into skills although a multidimensional extension could be easily derived. For lack of space, we will thus only compare SyLM to the 1PL and 2PL.

Table 1: Synthetic results

Model	Param.	log.Lik	AIC	BIC
1PL	Log.	-3847.1	8504.3	10100.1
1PL	Hom.	-3836.6	8483.2	10079.0
2PL	Log.	-3809.2	8478.5	10172.8
2PL	Hom.	-3724.3	8308.7	10002.9
SyLM	Log.	-3809.2	9238.5	12430.1
SyLM	Hom.	-3455.5	8531.1	11722.6

3. EMPIRICAL EVALUATION

3.1 Synthetic Comparison

For each approach, logistic and homographic parameterizations are tested. Parameters are inferred by a Maximum Likelihood [4] algorithm supported by a Newton-Raphson optimization. The dataset consists of the results to the first Mathematic booklet of PISA 2012 study in France (380 examinees, 25 items). For items having two degrees of success, both cases are considered as a success. Similarly, answers entered as "not reached" or "NA" are considered as failures.

Although the results shown in Table 1 should be independent of the paramtetrization, estimations using homographic parametrizations produce better results throughout all settings. As expected, the additional parameters brought into the optimization by SyLM are crucial for the information criteria. However, comparing SyLM with the 1PL shows SyLM as the winner in two out of three cases. The decrease of the log-likelihood exceeds the increase of the AIC due to the significantly higher number of parameters.[1] The difference is even stronger for BIC and increases with the number of samples, hence naturally penalizing SyLM.

3.2 PISA Analysis

We now analyse the PISA 2012 ranking [3] and its associated country clustering with SyLM. The original grouping is based on the scores in the different tests and on social and economical variables of the countries. We focus on four pairs of countries/economies and shown in Table 2. Although Shanghai and Singapore are not reported similar, we study them together as they are the top ranked and the only ones without a similar peer. Our analysis is again performed on the Mathematics test. For each country, booklets are analyzed separately before the results are merged.

For the the twelve countries listed in Table 2, Figure 1 focuses on the distribution of *examinee's discrimination* given the *examinee's ability*. The coloring indicates the ratio of pupils having a high or a low *normalized*[2] discrimination given the fact that they have a low or a high *normalized* ability. We consider values below .25 as a low *normalized* characteristic and above .75 as a high one.

Although Switzerland and Japan are in the same PISA group, their figures are very different. The Japanese distribution is closer to the other Asiatic countries while the Swiss is similar to the German one. The geographic argument holds for Brazil and Argentina but not for USA and Russia, which are geographically and culturally very different. Again the two neighbors Canada and USA produce very different results. While the distribution for USA is closer to the British one, the Canadian one shows very different. Based on our results, an improved clustering can be proposed. Shanghai, Singapore and Japan constitute the first group; Switzerland, Germany the second. Great Britain, the USA and Russia form the third group while Brazil and Argentina make a group of their own. Canada and France remain outsiders.

[1]The 2PL counts $N+2M$ parameters, SyLM has $2N+2M$.
[2]Data is normalized by $y_i \to \frac{y_i}{1+y_i}$ and $\theta_i \to \frac{1}{1+e^{\theta_i}} = x_i$.

Table 2: PISA country grouping

QCN Shanghai	CHE Switzerland	GER Germany
SGP Singapore	JPN Japan	CAN Canada
FRA France	USA USA	BRA Brazil
GBR Great Britain	RUS Russia	ARG Argentina

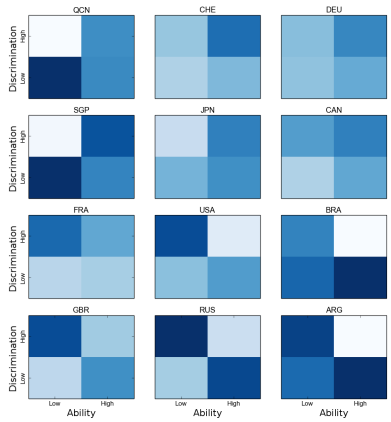

Figure 1: SyLM results for PISA

4. CONCLUSION

We proposed the Symmetric Logistic Model as a generalization of the Rasch model. Our approach can be interpreted as a symmetric 2PL at the cost of additional parameters. Empirically, our Symmetric Logistic Model showed that the PISA grouping of countries based on score and socio-economic backgrounds is suboptimal. More appropriate groups could be formed by taking examinee discrimination parameters into account.

5. REFERENCES

[1] A. Birnbaum. Some latent trait models and their use in inferring an examinee's ability. In *Statistical Theories of Mental Test Scores*. Addison-Wesley, 1968.

[2] R. L. McKinley and M. D. Reckase. An extension of the two-parameter logistic model to the multidimensional latent space. Technical report, DTIC Document, 1983.

[3] OECD. *PISA 2012 Results in Focus: What 15-year-olds know and what they can do with what they know.* OECD Publishing, 2013.

[4] N. Rose. Maximum likelihood and Bayes modal ability estimation in two-parametric IRT models: Derivations and implementation. *(Schriften zur Bildungsf.)*, 2010.

[5] N. Verhelst and C. Glas. The one parameter logistic model. In G. Fischer and I. Molenaar, editors, *Rasch Models*, pages 215–237. Springer New York, 1995.