Book of Short Papers SIS 2020

Editors: Alessio Pollice, Nicola Salvati and Francesco Schirripa Spagnolo

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891910776

Teaching statistics: an assessment framework based on Multidimensional IRT and Knowledge Space Theory

Un modello di valutazione della conoscenza per l'insegnamento della statistica: l'esperienza del progetto Erasmus+ ALEAS

Cristina Davino[a], Rosa Fabbricatore[a], Carla Galluccio[b], Daniela Pacella[a], Domenico Vistocco[a], Francesco Palumbo[a]

[a] University of Naples Federico II
[b] University of Florence

Abstract In recent years, there is an increasing need for technology-based platforms to assist traditional learning methodologies. However, it is challenging to set up a common assessment framework to evaluate user knowledge. To address this issue, we propose an approach to teaching undergraduate statistics that makes use of the psychometric Item Response Theory based on latent class categorization to evaluate the user ability based on the European learning outcomes - the Dublin descriptors. Additionally, we enclose the user assessment workflow in a formalized structure using the principles of Knowledge Space Theory to track the current user knowledge state adaptively. The methodological framework serves as a base for the app developed within the ALEAS ERASMUS+ Project.

Abstract *Di recente, è stata evidenziata una maggiore necessità di piattaforme tecnologiche come ausilio alle metodologie di apprendimento tradizionali. Tuttavia, è difficile istituire un quadro di valutazione comune per le conoscenze degli utenti. Pertanto, proponiamo un approccio all'insegnamento della statistica universitaria che si avvalga della teoria psicometrica Item Response Theory basata sulla categorizzazione in classi latenti, allo scopo di valutare la capacità dell'utente sulla base di obiettivi di apprendimento europei, i descrittori di Dublino. Inoltre, racchiudiamo il flusso di valutazione degli utenti in una struttura formalizzata utilizzando i principi della Knowledge Space Theory per tracciare in modo adattivo l'attuale stato della conoscenza dell'utente.*

Key words: Item Response Theory; Knowledge Space Theory; Statistics; Technology-enhanced learning; Intelligent tutoring systems

1 Introduction

The problem of assessing the student on a multidimensional level is one of the most challenging tasks in education (Deonovic et al., 2018). The student behaviour and responses should be evaluated on the one hand in comparison with the student population, on the other hand in comparison with the individual performance changes over time, so to measure the learning outcomes.

Recently, there is rising attention towards teaching statistics using technology. Although there are several technologies developed for teaching and assessing statistics, there are not many applications specifically targeting undergraduate students in non-scientific courses. Recently, Lopez Lamezon and colleagues demonstrated the advantages of using a virtual environment for teaching statistics in Medicine degree courses although not grounding their learning assessment upon specific educational theories (López Lamezón et al., 2018).

A complete assessment framework should be able to:

- detect the overall ability level of the student sample and to adapt to it;
- personalize the student experience selecting the most appropriate set of topics, questions and items to present according to each student's knowledge progress.

Concerning the first requirement, we exploit the Item Response Theory (IRT) (Rasch, 1960a). IRT allows modelling the probability that each student answers correctly to a particular item given her or his ability. It also allows estimating the difficulty of each item given the sample of answers. We categorize the items of each statistics subtopic according to the Dublin descriptors of the learning outcomes. Concerning the second requirement, we resort to Knowledge Space Theory (KST) (Doignon and Falmagne, 1985), widely used in the field of expert systems to the end of organizing the full knowledge required to master a specific subject into a direct acyclic graph structure. Nodes of the graph represents a competency. Each student is classified into a state and the system decides the following direction on her/his learning path. The proposal describes the core of ALEAS (Adaptive LEArning in Statistics), a learning platform developed as the intellectual output of the homonymous Erasmus+ project. ALEAS aims to be a complement to traditional teaching methodologies. It focuses on the assessment phase, primarily for university students enrolled in non-scientific degree curricula.

Over the knowledge structure for basic statistics knowledge, a directed acyclic graph consisting of several statistics subtopics is defined. Students' performance assessment is based on the multidimensional latent class IRT model with dichotomous items (Bartolucci, 2007).

The Dublin descriptors are the framework for qualifications of higher education courses in the European area; they qualify the expected outcome of any learning process. Our proposal focuses on the assessment of hard skills, and in particular statistics, and hence we consider the following three descriptors (Gudeva et al., 2012):

- *Knowledge and understanding (K)*: the ability to demonstrate knowledge and understanding with a theoretical, practical and critical perspective on the topic;
- *Applying knowledge and understanding (A)*: the ability to apply the knowledge identifying, analysing and solving problems sustaining an argument;

- *Making judgments (J)*: the ability to gather, evaluate and present information exercising appropriate judgment.

2 Methodology

A description of both the general KST framework and the IRT methodology will be provided in the following subparagraphs.
KST is a methodological framework that analyzes the performance over time of individual students (Doignon and Falmagne, 1985, 2016). It is particularly useful for the current knowledge assessment of the students, and also to adaptively select the most tailored items to present to the student given his or her ability. KST formalizes a structure of the domain into nodes of a directed acyclic graph aiming to assess each learner according to which subsets of the domain they master. The use of KST induces a twofold issue:

- to construct the structure of the knowledge domain investigating the relationship among the different subsets of the domain;
- to select the most appropriate questions that allow assessing the current knowledge subsets the student has mastered.

Several algorithms implement the progression rule from one to the next knowledge state. The probabilistic model considers each knowledge state having a likelihood function whose parameters are updated according to the student answers; while the expert model that we proposed for the ALEAS platform aims to construct a knowledge graph with the help of experts. We divided statistical knowledge into nodes, and we set constraints to the student progress path. In particular, to progress from one node to the other, a student has to obtain mastery - considered as the highest level of ability - in all the required nodes. The level of ability is evaluated through the multidimensional IRT model described below (Fabbricatore et al., 2019).
IRT is a model-based theory that aiming at estimating the probability that each student will answer correctly to each item of a set. IRT models founded on the idea that the probability that an individual provides a certain response to a certain item can be described as a function of the person's latent trait, plus one or more parameters that characterize a specific item. A latent trait is typically described by a continuous normal probability distribution (Bartolucci et al., 2015).

According to the traditional logistic IRT models with dichotomous items, the probability that the s-th subject (with $s = 1,\dots,S$) will respond correctly to the item i (with $i = 1,\dots,I$) in the most general form (4-parameter logistic model; 4PL) can be formalized as following:

$$P(X_{si} = 1|\theta_s, \Gamma_i) = c_i + \frac{1 - d_i - c_i}{1 + e^{a_i(\theta_s - b_i)}} \tag{1}$$

where X_{si} is the response of the s-th subject on the i-th item with realization $x_{si} \in [0,1]$, $\theta_s \in \mathbb{R}$ is the ability of the s-th subject, and $\Gamma_i = (a_i, b_i, c_i, d_i)$ is the whole set of parameters of the i-th item. In particular, $a_i \in \mathbb{R}$ is the item discrimination parameter,

$b_i \in \mathbb{R}$ represents the item difficulty, and $c_i, d_i \in \mathbb{R}$ are the guessing and the ceiling error parameters respectively (Noventa et al., 2019). So, logistic IRT models are based on the idea that the correct response probability follows a logistic curve called the Item Characteristic Curve, and it is always directly proportional to the ability of the student (Rasch, 1960a; Birnbaum, 1969). Each item has its specific curve, which is defined by a set of parameters, between one and four. The four parameters refer to: *(i)* discriminating power (slope), *(ii)* item difficulty (location), *(iii)* guessing (lower asymptote) and *(iv)* ceiling (upper asymptote). Reduced models can be obtained constraining one or more parameters. This is the case of 2 parameter logistic IRT ($c_i, d_i = 0$ in equation 1) and 1 parameter logistic IRT ($a_i = 1$ and $c_i, d_i = 0$ in equation 1) models. In particular, only the error parameters (c_i, d_i) are considered for the model with 2 parameters. Such parameters take into account the case incidence on the probability of the correct answer. Models with 1 parameter also assume that all items have the same discriminative power (Rasch model; Rasch (1960b)).

In ALEAS, to perform the categorization of the items we adopted the model with 2 parameters, disregarding the guessing and the ceiling parameters. The choice was made taking into account that each item has four possible answers, lowering the impact of guessed answers. The model is applied to student responses, where we only take into account if the given response is correct or incorrect.

3 The proposed model for item and user categorization

The extension of traditional IRT models proposed by Bartolucci (2007) with the introduction of multidimensional latent class IRT models provides a suitable statistical tool to pursue our goals. Introducing the concepts of multidimensionality and discreteness of the latent traits, these models allow releasing both the IRT constraints of unidimensionality and continuous nature of the latent trait. In this view, we can simultaneously consider more latent traits each of which is represented by a discrete distribution with $\xi_1, \dots, \xi_k$ support points defining k latent classes with weights $\pi_1, \dots, \pi_k$. Latent class weights π_c (with $c = 1, \dots, k$) represent the probability that a subject belongs to class c, and can be expressed as:

$$\pi_c = p(\Theta_s = \xi_c) \tag{2}$$

with $\sum_{c=1}^{k} \pi_c = 1$ and $\pi_c \geq 0$, where Θ_s $(s = 1, \dots, S)$ represents the discrete random variable of the latent trait of the sth subject. Multidimensionality, in this model, is introduced through a simple formulation in which each item is related only to one latent trait. In particular, items are divided in different subsets I_d (with $d = 1, \dots, D$) based on D different dimensions, which, in our model, are represented by the three Dublin descriptors considered: K (knowledge and understanding), A (Applying knowledge and understanding), and J (making judgments).

Once the model has been defined, the next step is to estimate the parameters. We exploit a Maximum Marginal Likelihood approach making use of the Expectation-Maximization (EM) algorithm (Bartolucci et al., 2014).

In particular, the model tries to estimate (Expectation step) the probability of each individual belonging to one of the latent classes given his or her vector of responses; during the Maximization step it maximizes the log-likelihood that the individual belongs to that latent class. These two steps are repeated until convergence.

In order for the parameters of the model to be identified, a full matrix of individuals and responses must be supplied without missing values. Additionally, the model requires the constraint that, for each latent trait, one discriminating index is equal to 1 and one difficulty parameter is equal to 0 (Bartolucci et al., 2014).

Each user will thus be categorized into a latent class according to their response configuration. The number of latent classes k for the data can be estimated using information criteria such as the Bayesian Information Criterion or the Akaike information criterion (Gnaldi, 2017), or they can be defined according to the theoretical framework. For the present work, we established the number of classes to be $k = 3$, consistently with the number of dimensions defined by the Dublin descriptors, and hence with the theorized levels of ability. The model is applied using the R package `MultiLCIRT` (Bartolucci et al., 2014). The IRT model is applied separately for every node of the KST. The flow of information within each of the knowledge states of the KST is shown in Figure 1.

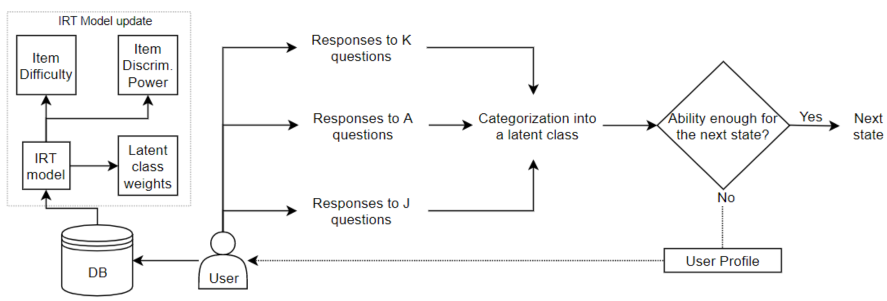

Fig. 1 Flow of information within a KST node and progress using the ability evaluated by the IRT model.

4 Conclusion

We proposed a methodological framework for the development of an intelligent tutoring system for assessing knowledge of statistics. The model targets undergraduate students and assesses them on a multidimensional level. To do so, we have presented a possible integration of a classical statistical and psychometric methodology - the

Item Response Theory, the Dublin descriptors of learning outcomes and a computational knowledge formalization - the Knowledge Space Theory for the structure of the user model.

References

Bartolucci, F. (2007). A class of multidimensional irt models for testing unidimensionality and clustering items. *Psychometrika 72*(2), 141.

Bartolucci, F., S. Bacci, and M. Gnaldi (2014). `MultiLCIRT`: An R package for multidimensional latent class item response models. *Computational Statistics & Data Analysis 71*, 971–985.

Bartolucci, F., S. Bacci, and M. Gnaldi (2015). *Statistical analysis of questionnaires: A unified approach based on R and Stata*. Chapman and Hall/CRC.

Birnbaum, A. (1969). Statistical theory for logistic mental test models with a prior distribution of ability. *Journal of Mathematical Psychology 6*(2), 258–276.

Deonovic, B., M. Yudelson, M. Bolsinova, M. Attali, and G. Maris (2018). Learning meets assessment. *Behaviormetrika 45*(2), 457–474.

Doignon, J.-P. and J.-C. Falmagne (1985). Spaces for the assessment of knowledge. *International journal of man-machine studies 23*(2), 175–196.

Doignon, J.-P. and J.-C. Falmagne (2016). Knowledge spaces and learning spaces. *New Handbook of Mathematical Psychology 2*, 274–321.

Fabbricatore, R., C. Galluccio, C. Davino, D. Pacella, D. Vistocco, and F. Palumbo (2019, September). The effects of attitude towards statistics and math knowledge on statistical anxiety: A path model approach. In M. Carpita and L. Fabbris (Eds.), *Statistics for Health and Well-being*, pp. 97–100.

Gnaldi, M. (2017). A multidimensional irt approach for dimensionality assessment of standardised students tests in mathematics. *Quality & Quantity 51*(3), 1167–1182.

Gudeva, L. K., V. Dimova, N. Daskalovska, and F. Trajkova (2012). Designing descriptors of learning outcomes for higher education qualification. *Procedia-Social and Behavioral Sciences 46*, 1306–1311.

López Lamezón, S., R. Rodríguez López, L. M. Amador Aguilar, and L. M. Azcuy Lorenz (2018). Social significance of a virtual environment for the teaching and learning of descriptive statistics in medicine degree course. *Humanidades Médicas 18*(1), 50–63.

Noventa, S., A. Spoto, J. Heller, and A. Kelava (2019). On a generalization of local independence in item response theory based on knowledge space theory. *Psychometrika 84*(2), 395–421.

Rasch, G. (1960a). Probabilistic models for some intelligence and attainment tests. studies in mathematical psychology. *Copenhagen: Danmarks Paedagogiske Institut*.

Rasch, G. (1960b). Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests.