ISCTE IUL

DOCTORAL PROGRAM IN COMPLEXITY SCIENCES

Evolution of Computerized Adaptive Testing

Harvey Restrepo
29 January, 2018

harvey_restrepo@iscte-iul.pt

The context and the objective

One of the great objectives of science is to rigorously evaluate the knowledge, abilities, aptitudes and attitudes that a person, or a group of people, have. Of special interest is the relationship between observable responses and latent variables that could predict non-obvious personality traits, responses to certain stimuli, or different levels of sensory perception in specific contexts.

Computerized Adaptive testing (CAT) systems based on Item Response Theory (IRT) has made an enormous advance in testing better and in a more efficient way. Despite its growing popularity in several disciplines, in each CAT evaluation session, there is a set of records of the answers that the examinee delivers and that records, with the current technological development, finally could be analyzed in real time to do better assessments.

Abstract

“Daring to measure is an act of honesty. However, one thing is to measure and another, very different, to measure well.” HR

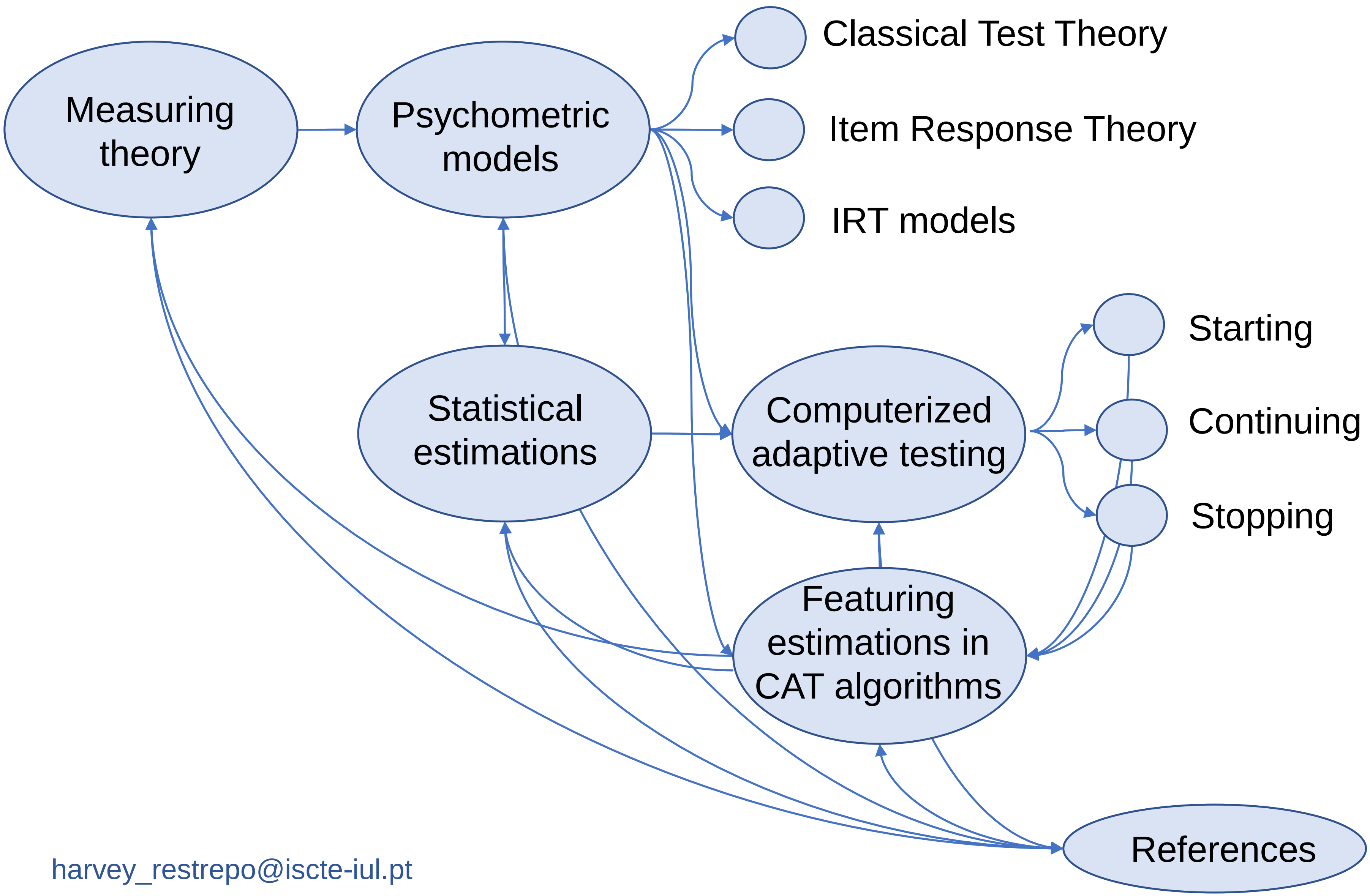
This literature review shows the evolution of the Computerized Adaptive Testing (CAT) and the development phases of the Theory of measurement in its two main theoretical frameworks: Classical Test Theory (CTT) and Item Response Theory (IRT). Likewise, the most relevant equations are analyzed to perform the statistical estimations of the information provided by each item, as well as its extension to a CAT process for developing dynamic evaluations. In the last section, the limitations of the analytical methods that underlie the current algorithms are shown and some lines of research are identified, mainly focused in facing the problems of the analysis and the recognition of complex patterns in a nonlinear CAT process.

Keywords: CAT, measuring complexity, adaptive testing, large-scale assessment, Item response theory.

Content

harvey_restrepo@iscte-iul.pt

State of the art

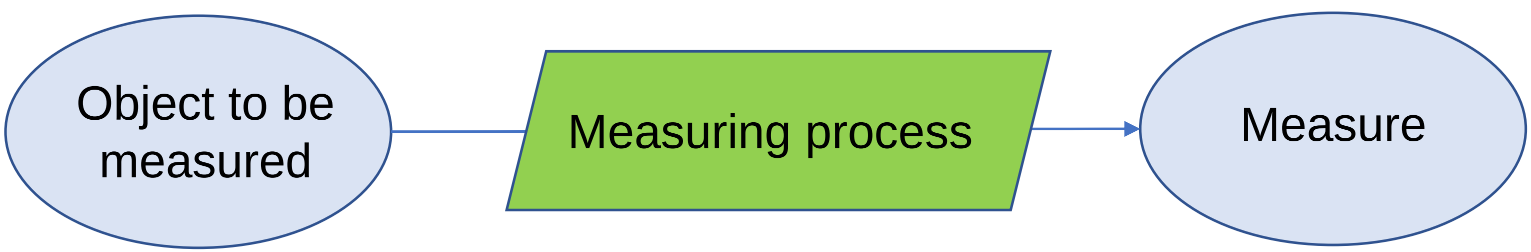


harvey_restrepo@iscte-iul.pt

1. Measuring theory

"Measuring human being means measure complexity in its deepest state." HR

This first part of the paper shows the evolution of CAT, explaining the Theory of measurement development in its two main theoretical frameworks: Classical Test Theory (CTT) and Item Response Theory (IRT).



harvey_restrepo@iscte-iul.pt

2. Psychometric models

Spearman showed the randomness of the observed scores as the interaction of two independent sources:

1. the 'true score' of the construct, and
2. the measurement error, as shown in equation (1):

$$S_T = S_O + \varepsilon \ (1)$$

From the definition of equation (1), it is possible to construct a set of mathematical models that provide estimates of the true score S_T when considering the error term as a random variable.

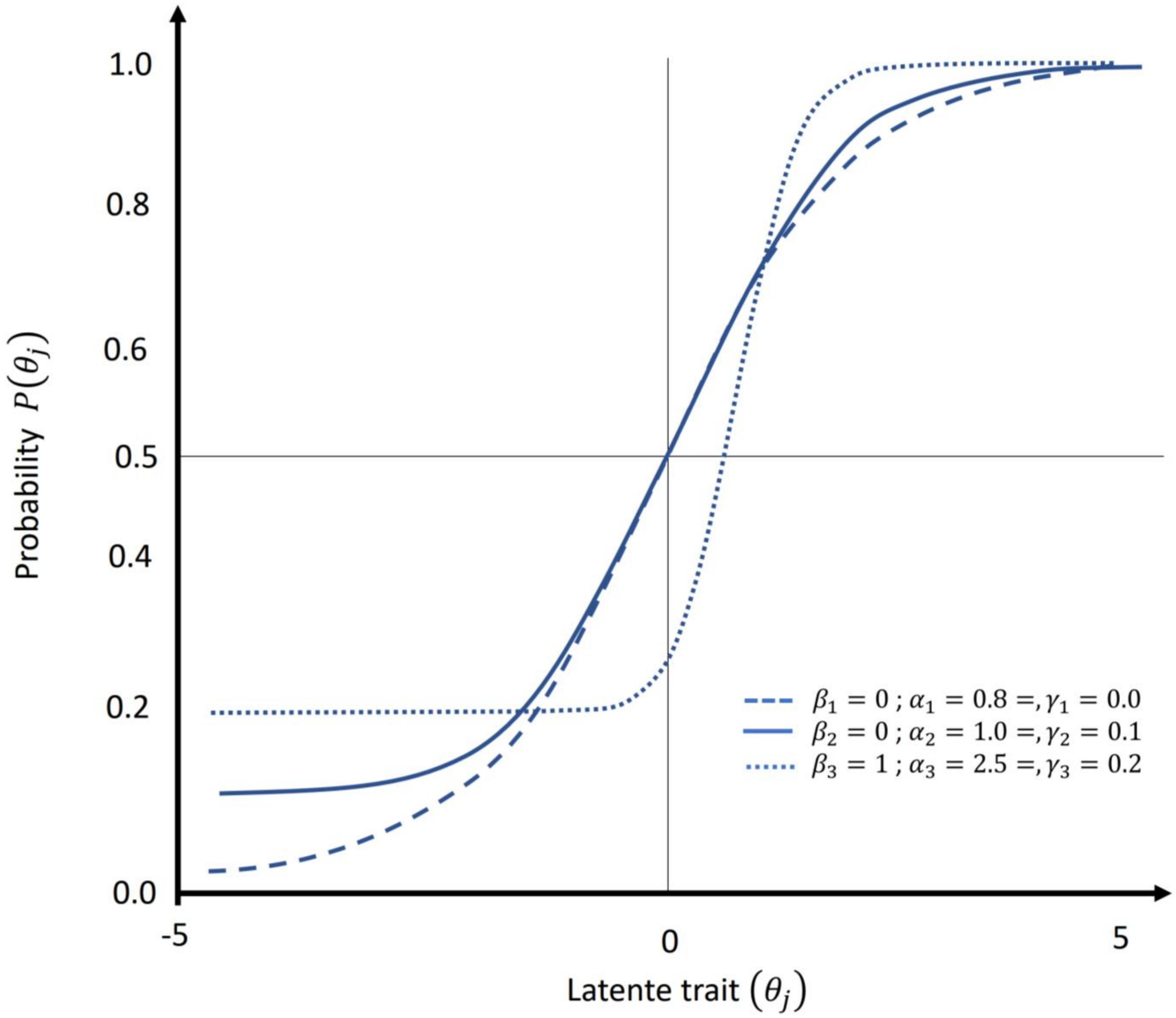
harvey_restrepo@iscte-iul.pt

2. Psychometric models

The IRT is defined on a mathematical framework that enables the design of various models and offer the possibility of estimating one (1-PL), two (2-PL) or three (3-PL) parameters to characterize the equation that models the behavior of the items. In all cases, the logistic equation estimates the probability that a subject responds correctly to a stimulus based on its level of latent trait and the psychometric properties of the item:

$$P(\theta_j) = 1 + (1 - \gamma_i)\frac{e^{[\alpha_i(\theta_j - \beta_i)]}}{1 + e^{[\alpha_i(\theta_j - \beta_i)]}}, \text{with } \theta_j, \alpha_i, \beta_j \in (-\infty, \infty)\ y\ \gamma_j \in [0, 1]\ (2)$$

2. Psychometric models



3. Statistical estimations

Since the objective of the evaluation is to obtain a quantifiable representation of the latent trait of an examinee, impossible to observe directly, it is necessary to know the parameters of the items used in the measurement of the trait to incorporate them into the likelihood function given by the equation (4):

$$L(\boldsymbol{y}|\theta) = \prod_{i=1}^{n} P(y_i|\theta) \quad (4)$$

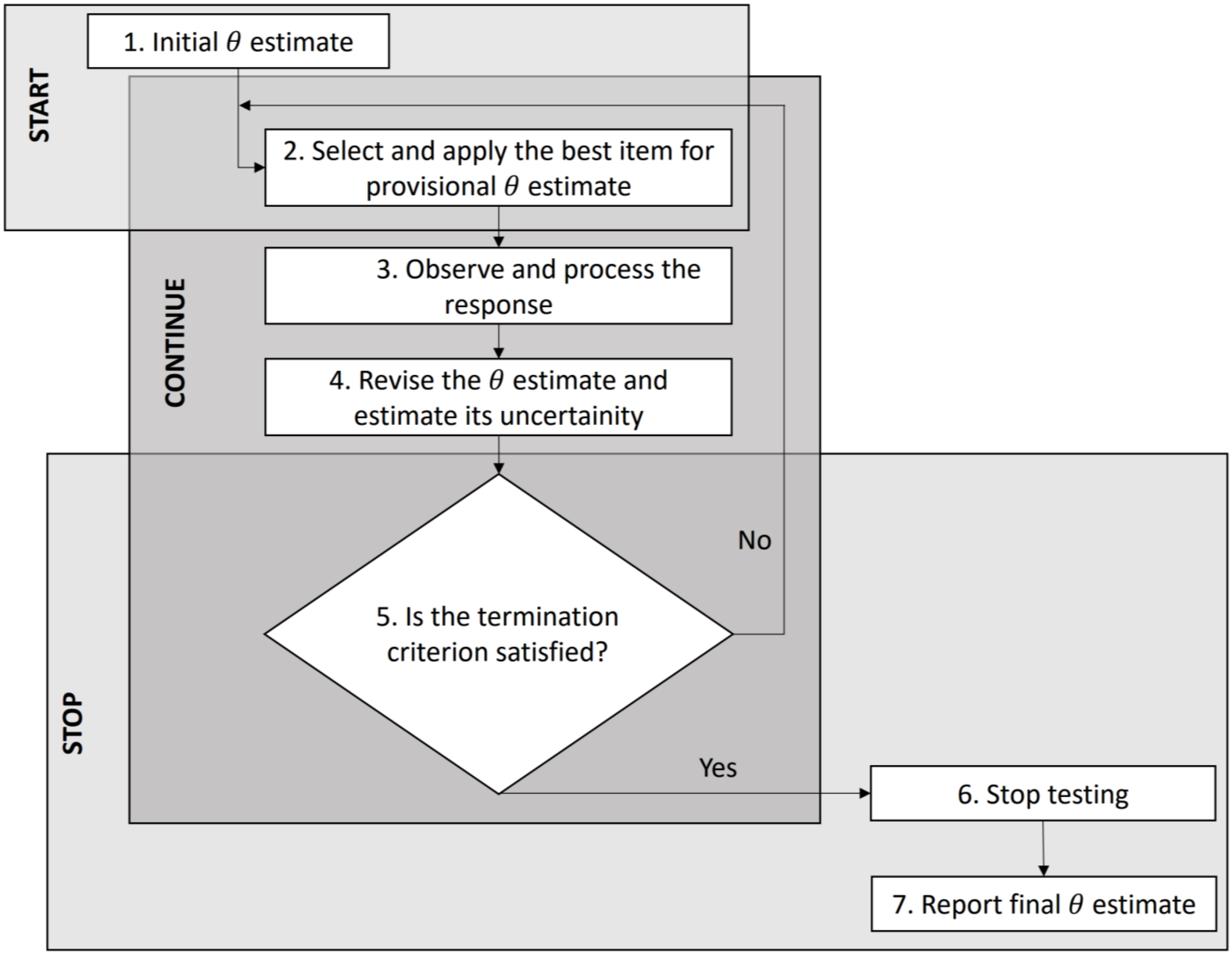
the sample variance $\sigma_e^2(\hat{\theta}_j)$ of θ can be obtained as follows:

$$\sigma_e^2(\hat{\theta}_j) = \left[\sum_{i=1}^{n} \frac{[P_i'(\theta_j)]^2}{P_i(\theta_j)[1 - P_i(\theta_j)]}\right]^{-1} \quad (5)$$

where $P_i(\theta_j)$ is the probability that a latent feature would be expressed in the item $i = 1, \ldots, n$ and $P_i'(\theta_j)$ is the first derivative of $P_i(\theta_j)$ with respect to θ.

harvey_restrepo@iscte-iul.pt

4. Computerized adaptive testing



Graph based on the model of Pickett y Croudace (2016).

harvey_restrepo@iscte-iul.pt

5. Featuring estimations in CAT algorithms

The methodologies for selecting the items and the estimation methods of latent traits used in the CAT algorithm are still under development, especially in their formal analytical foundations. Van der Linden (2010) have pointed out that, in both the Bayesian and the likelihood-based methods, the estimation of the latent trait lead to identical asymptotic results and converge to the true value of θ. However, it is not possible to keep the asymptotic behavior in the practical implementation of CAT when using short instruments.

Future work

Main research lines are related with analysis of the patterns, optimization tools and information theory focused on modelling those observed variables related with complex traits of the human being, such as:

- Strange attractors
- Nonlinear trajectories
- Stochastic process
- Adaptive computing
- Evolutionary algorithms