Bayesian Item Response Theory

Modeling sample bias in personality test development

Background

Alva Labs is a Stockholm based startup in HR tech. Our mission is to put best practice recruitment processes in the hands of hiring managers. Read more about us here: https://www.alvalabs.io/

A core part of our product is our personality test. It is built using Item Response Theory (IRT) to enable adaptive testing and continuous development (Fox, 2010). The Graded Response Model (GRM) was proposed by Samejima (1969) to model ordinal response data within the framework of IRT. We implemented the GRM in the probabilistic programming language PyMC3 (Salvatier, Wiecki & Fonnesbeck, 2016) to enable Bayesian inference using gradient-based MCMC and Variational approximation.
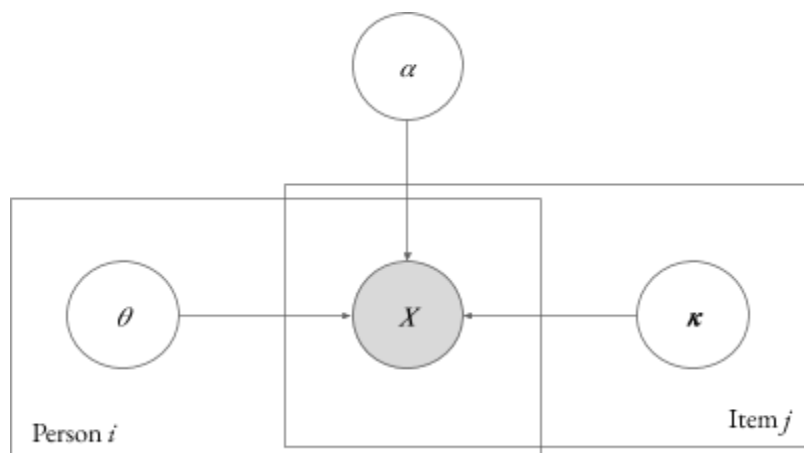

Figure 1: The Graded Response Model

After training the model, we use the posterior means of the item parameters α and κ in our scoring algorithm to estimate person trait θ during the adaptive test.

A challenge in our applied work is sample bias. We collect personality test data from our customers, where a majority of the candidates are below 30 years old for example. At the same time we want results to be applicable to a wide population, Swedish working adults, where the age distribution is more spread out. How can we adjust our model to generalize to this population?

Project objectives

- Propose a method to control for sampling bias. For example, introducing a latent regression in the GRM with demographic variables as covariates for θ, and applying poststratification to adjust estimates of κ.
- Implement model in probabilistic programming language (e.g. PyMC3, Stan, Pyro, TensorFlow Probability/Edward, Figaro)
- Simulation study to validate method

Requirements

- Bayesian Learning (732A91)
- Advanced Machine Learning (732A96)
- Some experience in programming, preferably in Python

Data

The data to be used in the project are in tabular form, accessed through Google BigQuery with SQL syntax. There are about 130,000 responses from 5,000 individuals to 300 questions in our database today. The demographic information collected through the platform are age, gender, primary language and education level.

Contact information

Morgan Pihl
0739461008
morgan@alvalabs.io

References

Fox, J-P. (2010). *Bayesian Item Response Modeling: Theory and Applications*. Springer Science.

Salvatier J., Wiecki T.V., Fonnesbeck C. (2016) Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2:55. DOI: 10.7717/peerj-cs.55

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 34(4), 100.