Using Item Response Theory in a Political Setting

Rebekah Fearnhead Supervisor: Gabriel Wallin

STOR-i, Lancaster University

1. Introduction

- Item Response Theory (IRT) models are commonly used in areas of cognitive and behavioural measurement.
- They are typically used with categorical data and are probabilistic models for individuals' responses to a set of items.
- These classify individuals into groups based on 'traits' where everyone with the same trait behaves in the same way.

2. IRT Models

We can first look at the model in the context of educational testing:

- Suppose we have N individuals taking a test with J questions.
- We represent individual i's response to item j as a binary random variable Y_{ij} where a value of 1 indicates a correct response, and $Y_{ij} = 0$ otherwise.
- The IRT model assumes that the $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ})^\top$ are independent and aims to model the joint distribution of this random vector by assuming one latent variable for each individual, denoted by θ_i.
- θ_i represents the individual's level for the ability (latent trait) which is being measured in the test and we assume these latent trait levels completely characterise each individual's response patterns.

- The two-parameter logistic (2PL) model is a common form of the IRT model.
- The Item Response Function (IRF) for the model is:

$$g_j(\theta|\pi_j) := P(Y_{ij} = 1|\theta_i = \theta) = \frac{\exp(d_j + a_j\theta)}{1 + exp(d_j + a_j\theta)}$$

- In the testing setting:
  - d_j is the easiness parameter with a higher value meaning individuals are more likely to get item j correct.
  - a_j is the discrimination parameter with $a_j > 0$ meaning higher latent trait levels lead to a higher chance of the individual getting the item correct.

3. Adding Covariates

Sometimes covariates of individuals can also be collected, for example age or gender and these can be incorporated into the model.

- If these are viewed to affect the responses for some of the items, then a Multiple Indicators, Multiple Causes (MIMIC) model can be used.
- This is used to study Differential Item Functioning (DIF) where items may function differently for one group of individuals compared to another.

- Assume that the respondents are from $K + 1$ unobserved groups, and represent the group membership using a latent variable, $\xi_i \in \{0, 1, \ldots, K\}$.
- The DIF-effect parameter can be represented as $\delta_{j\xi_i}$.
- The latent classes, ξ_i can be assumed to follow a categorical distribution.
- The 2PL model becomes:

$$g_j(\theta|\pi_j) = \frac{\exp(d_j + a_j\theta + \delta_{j\xi_i})}{1 + exp(d_j + a_j\theta + \delta_{j\xi_i})}.$$

4. Model Selection

- As both the latent classes ξ_i, and the latent variables θ_i, are unobserved, the inference on the proposed model is based on the marginal likelihood with both ξ_i and θ_i marginalised out.
- An L_1 regularised estimator can be used learn the sparsity pattern of the DIF-effect parameters, and estimate the unknown model parameters.
- The L_1 regularisation term behaves similar to Lasso regression as it tends to shrink some of the DIF-effect parameters to zero.
- The tuning parameter λ can be selected by using the Bayesian Information Criterion (BIC) with grid search approach.

References

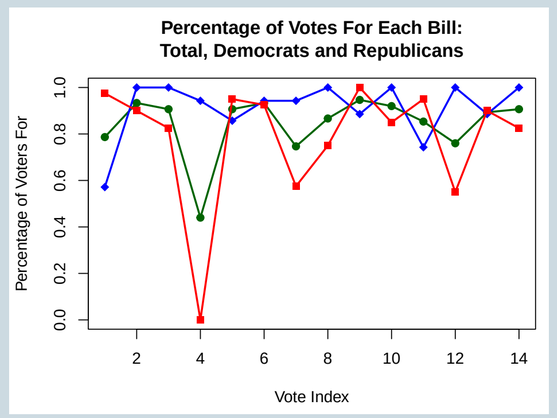
Yunxiao Chen, Xiaoou Li, Jingchen Liu, and Zhiliang Ying. Item Response Theory – A Statistical Framework for Educational and Psychological Measurement, 2021.

Jeffrey B. Lewis, Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet. Voteview: Congressional Roll-Call Votes Database., 2021.

Gabriel Wallin, Yunxiao Chen, and Irini Moustaki. DIF Analysis with Unknown Groups and Anchor Items. *Psychometrika*, 89(1):267–295, 2024

5. Political Setting

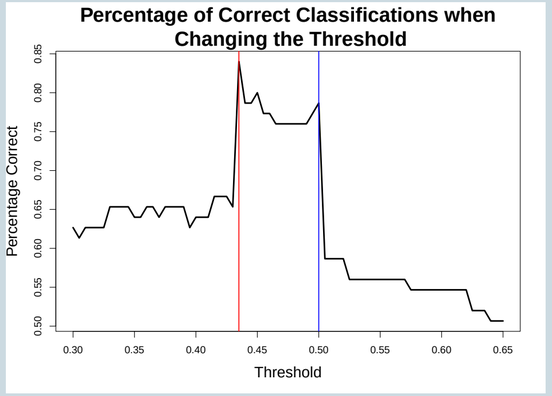
- This method can also be used in other areas, for example with US Senate voting data.
- The data below contains votes from 14 bills voted on in the 116th Congress.
  - There were 75 senators who voted in all 14 of the bills: 35 were Democrats, and 40 were Republicans.
  - This graph shows the percentage of Democrats (blue), Republicans (red) and total senators (green) who voted for each of the bills.



- We wish to see if an IRT model can successfully classify which party each of the senators belong to based on their voting patterns.

6. Setting the Threshold

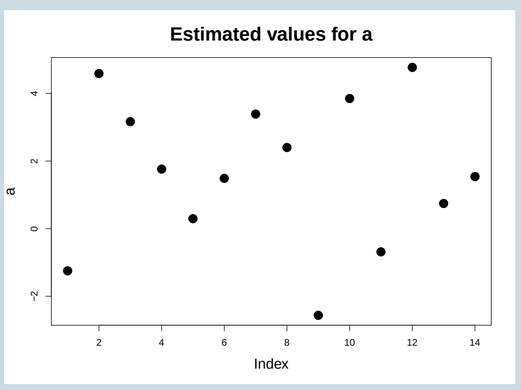
- After fitting the model, we need to be to able use this to classify the senators into their parties.
- This can be done by setting a threshold for the value of the normalised probability needed for an individual to be assigned to the DIF group.
- As we know the true parties of each of the individuals, we can use this to see which threshold correctly classifies the most senators.

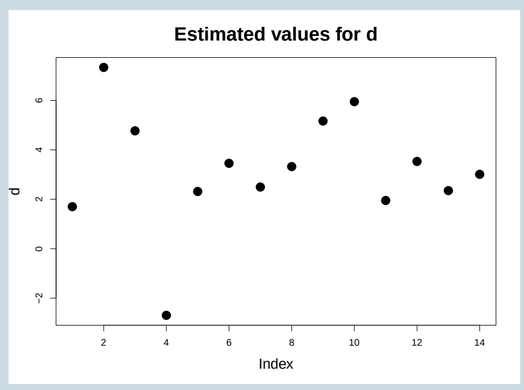


  - With a threshold of 0.435 (red), 84% of the individuals were correctly classified which is the highest percentage achieved with any threshold.
  - Usually a threshold of 0.5 (blue) is used and for this data it leads to a correct classification of 79%.

- Using a threshold of 0.5, the model classifies 47 of the senators as Democrats and 28 as Republicans.
- Out of the 35 Democrats, 33 are correctly classified by the model (94.3%) but out of the 40 Republicans, only 26 are correctly classified (65%).

7. Estimated Parameters





- Unlike in the testing setting, the values of a_j do not have to be positive. In this setting, θ_i represents the ideological position of each senator.
- This means that when a_j has a large absolute value, members of one party are much more likely to vote for (or against) a bill than the other party.
- In this setting a high d_j means that item j is more likely to be voted for by any of the senators.
- The only negative d_j value, d_4 occurs for the bill where fewer than half the senators voted for it.

8. Further Research

- The performance of other IRT models, such as Rasch or Probit models could be compared to the 2PL model.
- The number of groups could be increased as factors such as age or State represented could also affect voting patterns.
- This method could be tested in other areas such as marketing or psychology.