ITEM RESPONSE THEORY

The Item response theory (IRT) model: - has a long history since the 1920s. The IRT model is more popular for generation of psychometric tests and adaptive tests (where the difficulty level of the successive questions adaptively increases or decreases as the student gives correct or incorrect answers to questions of a given difficulty level). Examples of such standardised tests include GMAT and GRE and SAT.

The basic idea in an IRT model is this: - there are some latent (hidden) traits such as a **question's difficulty, student's ability, discrimination, chance factor etc.** Now from the student's correct or incorrect responses to the test questions, we can get a good idea of these latent variables.

Overview of CAT

There are 4 IRT models based on increased complexity: 1 parameter IRT model (called 1 PL model) to 4 parameter IRT (or 4 PL model).

1PL IRT model

The 1PL (or 1 parameter) IRT model (**also known as the Rasch model**)-

As per the 1PL model, the probability P_{ij} of the ith user correctly answering the jth question is given as: $logit(P_{ij}) = i — j$

where the logit function is given by $logit(x) =(1+(-x))-1$ Where

i is a learner (student), j is a question, ϑ_i is the ability of the learner, β_j is the difficulty level of the question.

Using the 1PL IRT model, we can predict the ability level ϑ_i of a learner, given the data about the learner's response to each attempted question.

2PL IRT model

We explore the 2 parameter (2PL) IRT model in a similar way. The 2PL model adds an additional factor, discrimination, to the 1PL model. The discrimination parameter indicates how good the question is in discriminating between students of differing abilities.

As per the 2PL IRT model, the probability P_{ij} of the jth user responding correctly to the ith question is given as follows:

$logit\ (P_{ij}) = a_i\ (\vartheta_j — \beta_i\ )$ where

i is a question, j is a learner student, θ_j is the ability of the learner, a_i is the discrimination, β_i the difficulty level of the question.

3PL IRT model

The 3PL IRT model is more complex than the 1PL and 2PL models. It adds an additional factor, chance or guess, to the 2PL model. The guess parameter indicates how good the question is in discriminating between students of differing abilities.

As per the 3PL IRT model, the probability P_{ij} of the jth user responding correctly to the ith question is given as follows:

$P_{ij} = c_j + (1\text{-}c_j)\ logit\ (a_i\ (\vartheta_j — \beta_i\ ))$ where

i is a question and j are a learner student, θ_j is the ability of the learner, ai is the discrimination, β_i the difficulty level of the question and c_j is the chance parameter, indicating the probability that the student will guess the answer to the question.

4PL IRT model

The 4PL IRT model adds an additional factor to the 3PL model, called d, which is the upper asymptotic limit on the chance of getting the answer correct.

As per the 4PL IRT model, the probability Pij of the jth user responding correctly to the ith question is given as follows:

$P\_{ij} = c\_j + (d\_j - c\_j)\ \mathrm{logit}\ (a\_i\ (\vartheta\_j — \beta\_i\ ))$

where

i is a question, j is a learner student, θ_j is the ability of the learner, a_i is the discrimination, β_i is the difficulty level of the question, c_j is the chance parameter, indicating the probability that the student will guess the answer to the question, d_j is the upper asymptotic limit

Therefore, by training any of these models (1PL to 4PL IRT model) on the student's test performance data, we can find the values of the variables such as discrimination level, difficulty level etc.

One can then plot the **item characteristic curve-ICC** (student's ability level on X axis vs probability of getting the answer right on Y axis).

Most used IRT Models:

In Brief a Mathematical formulation behind the scenes:

The simplest IRT model is often called the Rasch model or the one-parameter logistic model (1PL). According to the Rasch model, an individual's response to a binary item (i.e., right/wrong, true/false, agree/disagree) is determined by the individual's trait level and the difficulty of the item. One way of expressing the Rasch model is in terms of the probability that an individual with a particular trait level will correctly answer an item that has a particular difficulty. This is often presented as:

So, refers to the probability (P) that subject s will respond to item i correctly. The vertical bar in this statement indicates that this is a "conditional" probability. The probability that the subject will correctly respond to the item depends on (i.e., is conditional upon) the subject's trait level (θs) and the item's difficulty (βi). In an IRT analysis, trait levels and item difficulties are usually scaled on a standardized metric, so that their means are 0 and the standard deviations are 1. Consider these examples in terms of a mathematics test.

- What is the probability that an individual who has an above-average level of math ability (say, a level of math ability that is 1 standard deviation above the mean, $\theta s = 1$) will correctly answer an item that has a relatively low level of difficulty (say, $\beta i = -.5$)?

This indicates that there is a .82 probability that the individual will correctly answer the item. In other words, there is a high likelihood (i.e., greater than an 80% chance) that this individual will answer correctly. This should make intuitive sense because an individual with a high level of ability is responding to a relatively easy item. This is how IRT Model works with different formulas behind the scene for estimating parameters.

A few IRT packages (in Python) include the following:

- https://colab.research.google.com/drive/1dBcpXxHuc9YXv9yGllxlahx585hEmdbn?usp=sharing **CATSIM**
- https://github.com/nd-ball/py-irt **py-irt-Bayesian IRT Model in Python**
- https://github.com/lukas-a-olson/item-response-theory **Implementing Bayesian IRT Model and MCMC algorithm(Marcov chain Monte carlo)**
- https://github.com/pluralsight/irt_parameter_estimation **Implementing IRT Parameter Estimation**
- https://github.com/ckyeungac/DeepIRT **Deep learning Based IRT Model**
- https://github.com/mhw32/variational-item-response-theory-public

IRT (In R Studio) include the following:

- https://bookdown.org/bean_jerry/using_r_for_social_work_research/item-response-theory.html **MIRT**
- https://www.youtube.com/watch?v=EZ8OxIUx3oM **ltm**
-

Public Datasets for student's learning:

A few public datasets of student performance in tests (on which the IRT models may be trained and tested) are as follows:

1. ASSISTment 2009–2010 dataset https://sites.google.com/site/assistmentsdata/home
2. 2. ASSISTment 2015 dataset https://sites.google.com/site/assistmentsdata/home/2015-assistments-skill-builder- data

3. 3. KDD Cup 2010 cognitive tutor dataset https://pslcdatashop.web.cmu.edu/KDDCup/rules_data_format.jsp
4. 4. Synthetic-5 dataset Simulated data used in the DKT paper: Github has the dataset https://github.com/chrispiech/DeepKnowledgeTracing
5. 5. OLI Engineering statics 2011 dataset https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=507

Some References/Useful Resources:

- https://in.sagepub.com/sites/default/files/upm-binaries/18480_Chapter_13.pdf
- https://quantdev.ssri.psu.edu/tutorials/introduction-irt-modeling
- https://www.metheval.uni-jena.de/irt/VisualIRT.pdf
- https://towardsdatascience.com/a-bayesian-approach-to-rasch-models-item-response-theory-cc08805cbb37
- https://mc-stan.org/users/documentation/case-studies/tutorial_twopl.html
- https://telrp.springeropen.com/articles/10.1186/s41039-020-00132-w
- https://www.youtube.com/user/karoncook
- https://www.youtube.com/watch?v=HoMVasu2tg8