IRT workshop

Christopher David Desjardins

5 February 2014

Overview

1. Review

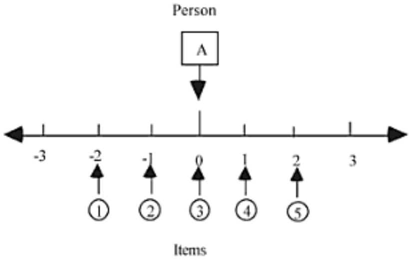
2. One-Parameter Model

- Structure of seminar and course assignments (i.e. short paper and presentation
- Quickly and conceptually introduced IRT
- Quickly introduced R
- Today: One-parameter and Rasch models

Proposed topics

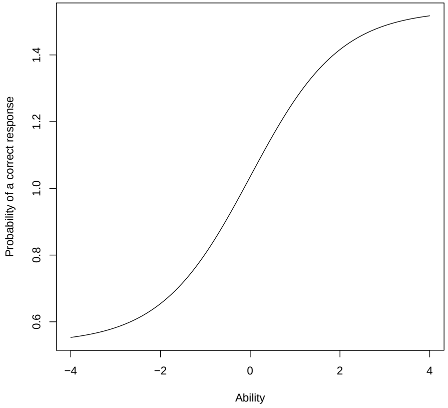
1-PL and Rasch model (5/2)	Differential Item Function (19/3)
2-PL and 3-PL models (19/2)	Latent class analysis (2/4)
Polytomous models (5/3)	Scaling, linking & Presentations (16/4)

Remaining classes are from 1300 - 1500 in Aragata 14.



- Assume we are measuring **reading proficiency**
- The **items** are located at: -2, 1, 0, 1, 2. (δ_j)
- **Person** A is located at 0. (θ_A)
- Which items would we expect Person A to respond correctly to? Which should be the easiest? Which the hardest?
- Which items would we have a more difficult time of knowing whether Person A would get them correct or not?

Conceptually (cont.)



This is known as an item characteristic curve, an item curve, or an **item response function (IRF)**. What kind of model should we fit?

Rasch model

- The logistic model
  $p(x = 1|z) = \frac{e^z}{1+e^z}$

Rasch model

- The logistic model

  $p(x = 1|z) = \frac{e^z}{1+e^z}$

- The logistic regression model

  $p(x = 1|x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$, so $z = \beta_0 + \beta_1 x$

Rasch model

- The logistic model
  $p(x = 1|z) = \frac{e^z}{1+e^z}$
- The logistic regression model
  $p(x = 1|x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$, so $z = \beta_0 + \beta_1 x$
- The Rasch model
  $p(x_j = 1|\theta, \delta_j) = \frac{e^{\theta-\delta_j}}{1+e^{\theta-\delta_j}}$, so $z = \theta - \delta_j$
- The probability of responding correctly to item j is a function of the **distance between the person and the location of item j (i.e. the item's difficulty)**!

Assume a person with an ability score of 1 encounters an item with a difficulty of 1.5. What is the probability they would get it correct if we assume the Rasch model is true?

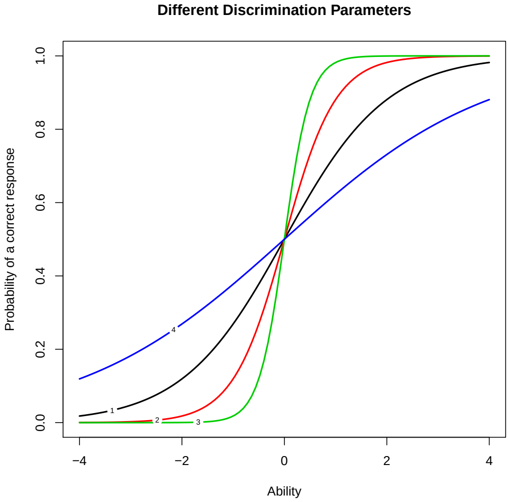
```
rasch <- function(person, item) {
    exp(person - item)/(1 + exp(person - item))
}
rasch(person = 1, item = 1.5)

## [1] 0.3775
```

The One-Parameter model

- The One-Parameter model
  $p(x_j = 1|\theta, \delta_j) = \frac{e^{\alpha(\theta-\delta_j)}}{1+e^{\alpha(\theta-\delta_j)}}$
- This parameter α is known as item discrimation and this affects the slope of the IRF.
- Higher α means we are doing a better job discriminating among respondents located at different abilities
- There is no item subscript in the one-parameter model!
- What does α equal in the Rasch model?

Varying α



Rasch vs. One-PL and Assumptions

- Many consider the Rasch model to be the true model and should be the only model to construct an instrument.
- Reality, Rasch is just the One-PL where $\alpha = 1$.

Rasch vs. One-PL and Assumptions

- Many consider the Rasch model to be the true model and should be the only model to construct an instrument.
- Reality, Rasch is just the One-PL where $\alpha = 1$.
- Assumptions
  - Unidimensionality - i.e. one underlying trait relating items.
  - Local independence - i.e. only thing that relates the items is the underlying construct.
  - The data follow either the model, data are S-shaped and appropriate, i.e. functional-form is correct.
  - *Consequence*: These models assume α is the same for all items and *c* the guessing parameter, is set to 0.

Ability estimates

1. First, calculate probability of each response for a respondent.
2. Second, determine probability of the response pattern (this is the product of #1 bc of local independence).
3. Third, repeat #1 and #2 for θ -3 to 3 (or -4 to 4).
4. Fourth, select the θ with the highest likelihood of producing the pattern.

Example: Assume that $\theta = 2.5$ on an instrument with 3 items.

- Assume the pattern of 011, where $\delta_1 = 2$, $\delta_2 = 1.2$, and $\delta_3 = 2.5$.
- Step 1 - Calculate the probabilities

  $$p(x_1 = 0|\theta = 2.5, \delta_1 = 2) = 1 - \frac{e^{2.5-2}}{1+e^{2.5-2}} = 1 - 0.622 = 0.378$$

  $$p(x_2 = 1|\theta = 2.5, \delta_2 = 1.2) = \frac{e^{2.5-1.2}}{1+e^{2.5-1.2}} = 0.786$$

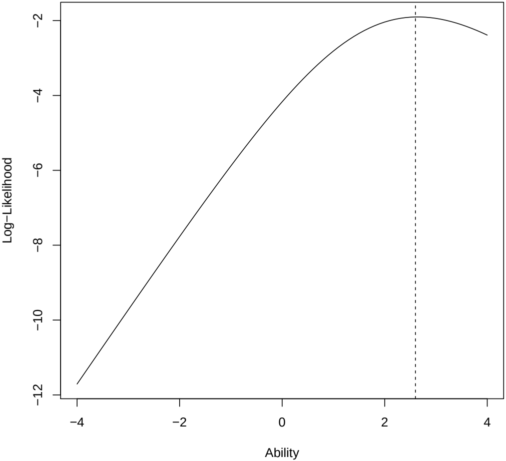
  $$p(x_3 = 1|\theta = 2.5, \delta_2 = 2.5) = \frac{e^{2.5-2.5}}{1+e^{2.5-2.5}} = 0.500$$

- Step 2 - Find the likelihood

  $p(x_1 = 0|\theta = 2.5, \delta_1 = 2) * p(x_2 = 1|\theta = 2.5, \delta_2 = 1.2) * p(x_3 = 1|\theta = 2.5, \delta_2 = 2.5) = 0.378 * 0.786 * .500 = \mathbf{0.148}$.

- Typically report, the log of the likelihood = log(0.148) = -1.91

Log-likelihood function for 011 for $\theta = -4$ to 4



Do this in R

```
person <- seq(from= -4, to = 4, by = .1)
item1 <- 2; item2 <- 1.2; item3 <- 2.5
LogLiks <- NULL
for(i in 1:length(person)){
p1 <- 1-rasch(person = person[i],item = item1)
p2 <- rasch(person = person[i],item = item2)
p3 <- rasch(person = person[i],item = item3)
LogLiks[i] <- log(p1*p2*p3)
}
plot(LogLiks person,type = "l",xlab = "Ability",
ylab = "Log-Likelihood")
abline(v = person[which.max(LogLiks)],lty=2)
```

Few more things about the log-likelihood function

- We want the maximum of this function, i.e. the maximum likelihood estimate (MLEs).
- MLEs doesn't exist for the pattern 000 or 111, therefore need to use Bayesian estimation.

Standard errors of estimate

- The standard error of estimate (SEE) represents our degree of uncertainty about the location of a person.
- The larger the SEE, the more uncertain we are.
- This is the same thing as a standard error in statistics and you could use it to create 95% confidence intervals around person and item parameters.

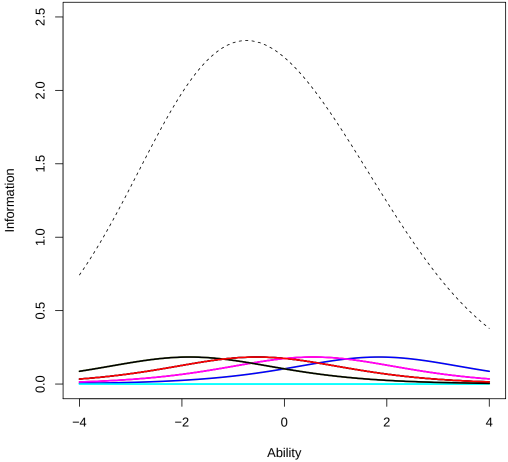
  $\hat{\theta} \pm 1.96 * SE(\hat{\theta})$ is a 95% confidence interval for θ.

  Recall how to interpret this!

Information

- Information (precision) is the inverse of the SEE.
- The smaller the SEE, the more information we have about a person's location.
- This tell us, essentially, where we can get precise estimates of person location.
- With the 1-PL model
  - Item maximum information at it's location.
  - Item information function is unimodel and symmetric at δ
  - All items on the instrument provide the maximum amount of information of $\alpha^2 0.25$.
- Because items are independent, the total information for a test (or instrument) is the sum of the item information functions!
- This is what we'll use to build a test or an instrument.

Information Function - Test and Item



Next time

- Please look for a data set that you could use for your class project. If you need help, please ask!
- Please read chapters 3 & 4 in Baker's IRT book.
- Optional: Chapters 5 & 6 in de Ayala. Read 3 & 4, if you're interested in the different estimation techniques.