EDUC 829 Week 11:  
Intro to IRT

## Plan for today

* Introduce the 2PL IRT model
* Assignment 3 grading will not be completed until next week
* Who would rather submit a paper for the final project?

## IRT vs CTT

* CTT makes strong assumptions, not suitable for categorical data
* IRT provides a model for each item
  + The model for the overall test is built up from the item-level model
  + Leads to better approaches to reliability, test scoring, test assembly, …

## IRT vs Factor analysis

* IRT is factor analysis for categorical data
  + Sometimes called item factor analysis
* However, IRT usually assumes unidimensionality
* There is mIRT and exploratory IRT, but that is not the standard stuff

## Other perspectives

* In IRT a peron’s true score is defined independently of the specific test (Hambleton & Jones)
* Imagine two non-parallel maths test (one is easier)
  + In CTT, a person’s true score must differ over tests (by definition)
  + In IRT, a person can have the same true score on the different tests
* Important in education – we can define math “ability” independently of the specific math test given

## IRT terminology

* “Ability” = factor = true score = latent trait
* “Item difficulty” ≈ proportion of people who get an item incorrect (who do not endorse the item)
* “Item discrimination” ≈ strength association between item and trait (factor loading)

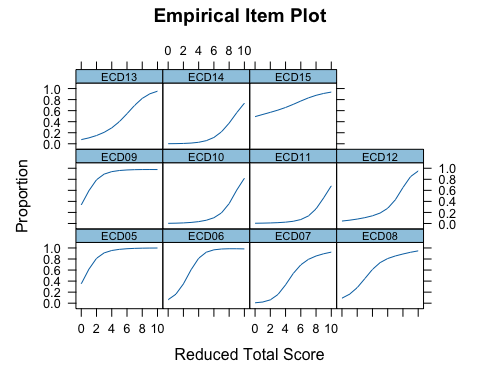
## IRT for binary data

library(mirt)

Loading required package: stats4

Loading required package: lattice

ecdi\_learning <- read.csv("ECDI\_learning.csv")  
empirical\_plot(ecdi\_learning, which.items = 1:11, smooth = T)



* Work on intuitive understanding before getting into model assumptions

# Item response functions (IRFs)

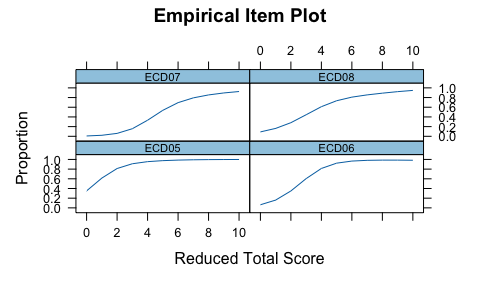
## Item response function

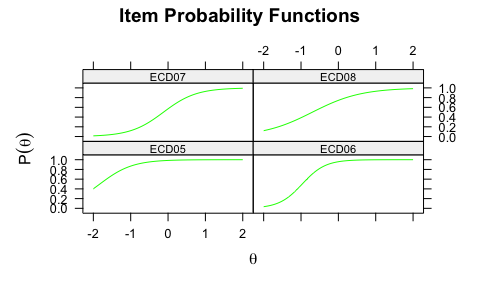
* is the latent trait (analagous to or )
  + Assume is standardized (, )
* is the probability of answering item “correctly”, treated as a function of the

## The logistic / logit functions

* Many IRT models use the a logistic model for
* Logistic maps a variable on to the unit interval (0,1)
* Its inverse, the logit, maps from the unit interval back to

## The logistic / logit functions





## The two-parameter logistic (2PL) model

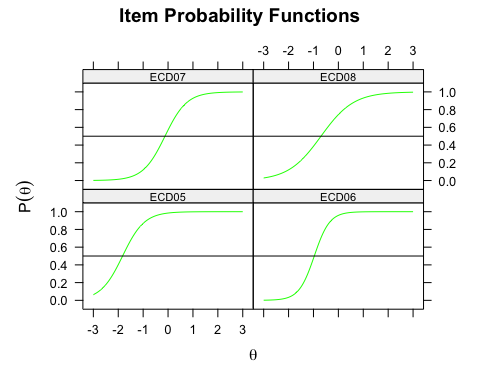
* Logistic formulation
* Logit formulation

## The item difficulty,

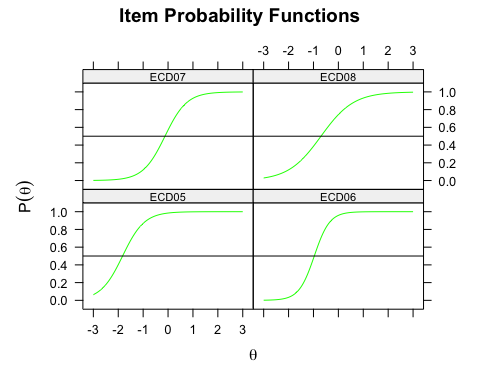
* If ,
* Interpretation: is the value of at which the proability of endorsing an item = 1/2.
* Respondents with ability above the difficulty level of the item have probability > 1/2 of answering the item correctly, and conversely.

## The item difficulty,

* The value of where the curve intersects



## Match the values to the curves



* Difficulty parameters:

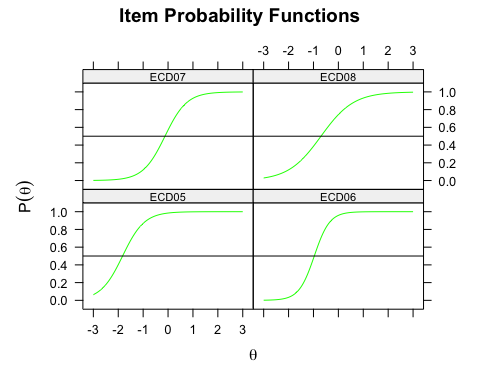
[1] -0.967 -1.826 -0.698 -0.124

## The item disrcimination,

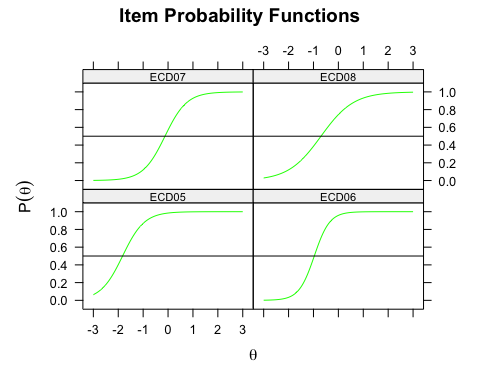
* Rate of change in :
* Interpretation:
  + Slope of IRF at is
  + Items that are more strongly associated with the trait are more “discriminating”

## The item disrcimination,

* The slope of at



## Match the values to the curves

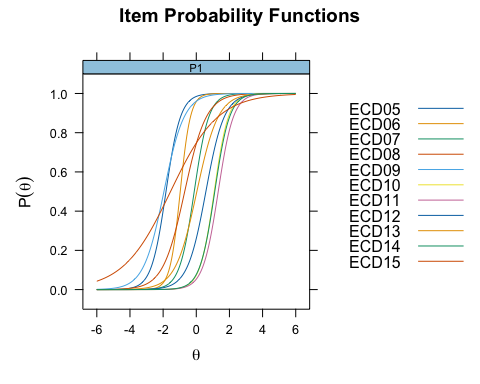


* Discrimination parameters:

[1] 1.527 2.282 3.240 2.269

## IRFs for the ECDI data

#load(ECDI\_learning.RData")  
library(mirt)  
fit.2pl <- mirt(ecdi\_learning, verbose = F)  
plot(fit.2pl, type = "trace", facet = F)



## Summary

* IRT models are usually defined via their IRFs
* The IRF relates the probability of endorsing an item to the trait being measured
* In the 2PL model, the IRF is a logistic function with 2 parameters
  + The difficultly is the level of the trait required to have prob ≥ 1/2 of endorsing the item
  + The discrimination describes how strongly the item is related to the trait (like a factor loading)

# Item and test information

* The precision with which we can estimate

## Information and reliability

* In IRT, the concept of (Fisher) information takes a central role
* Information is the precision with which we can esimate given the observed response data
* Interpretation: minimum zero, big values are good!
* In IRT, information takes the central role rather than reliability

## Item and test info

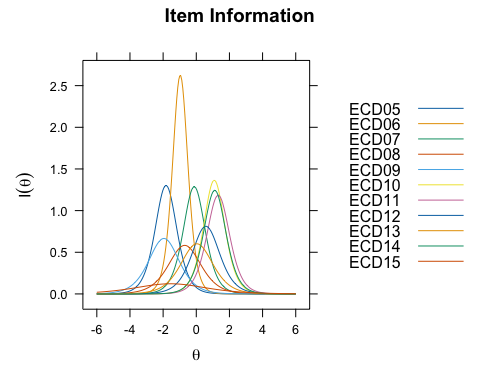
* The item information function (IIF) is the precision that results when estimating the latent trait using a single item
* In practice, we would never use only a single item on a test
* But, we can build up the information function of the entire test from that of each individual item
* So, we start with the IIF and then use that to get the test information function (TIF)

## Item information function (IIF)

* For the 2PL, the IIF is:
* Very similar to slope!
* Interpretation:
  + The info for an item is maximized when
  + The amount info at the maximum is

## Item information function (IIF)

plot(fit.2pl, type = "infotrace", facet = F)



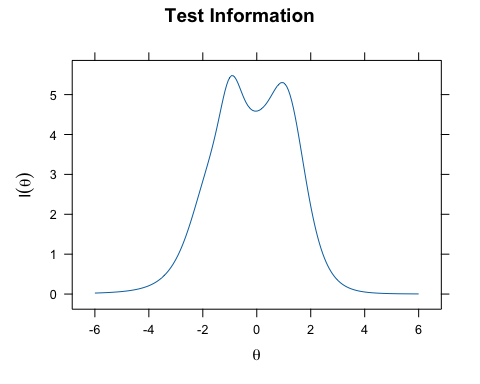
* Important idea: items are more or less informative for different values of

## Test information function (TIF)

* The TIF is just the IIFs aggregated to the test level
* Interpretation: The information provided by a test is just the sum of the information of its individual items!
* Note: this requires an assumption called conditional independence we will discuss next week

## Test information function (TIF)

plot(fit.2pl, type = "info", facet = F)



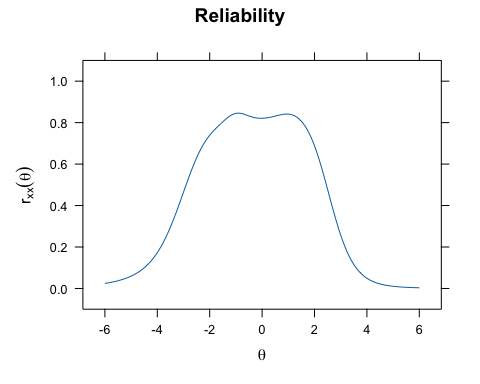
* Important idea: Tests are more or less informative for different values of

## What about reliability?

* Information is not easy to interpret
* Used mainly for comparisons among different tests
* To report out for a single test, more usual to use reliability, but now as a function of
* See Nicewander reading for derivation

## Reliability function

plot(fit.2pl, type = "rxx", facet = F)



* Important idea: Tests are more or less reliable for different values of

## Marginal reliability

* Sometimes it is still desirable to have a single number summary for reliability
  + e.g., APA journals require reliability is reported for each Measure
* For this, we can take the average of the reliability function
* Usually called marginal IRT reliability, but could also be called average reliability
* Sort of defeats the purpose of IRT…

## Marginal reliability

* IRT based reliability

marginal\_rxx(fit.2pl)

[1] 0.8148199

* Compared to CTT:

psych::alpha(ecdi\_learning)$total[1]

raw\_alpha  
 0.7800469

## Summary

* In IRT, info replaces reliability as main concept
* Test information is the sum of items’ information
* Unlike CTT, information varies depending on the level of the trait being measured
* For 2PL:
  + Items are most informative at their level of difficulty
  + The amount of info provided is given by the discrimination
* Info can be converted to reliability (function or average)

## Code for As 4

load("ECDI\_learning.RData")  
library(mirt)  
  
# Fit model  
fit.2pl <- mirt(ecdi\_learning, verbose = F)  
  
# Model params  
coef(fit.2pl, IRTpars = T, simplify = T)  
  
# Plots  
plot(fit.2pl, type = "trace", facet = F) # IRFs  
plot(fit.2pl, type = "infotrace", facet = F) # IIFs  
plot(fit.2pl, type = "info") # TIF  
plot(fit.2pl, type = "rxx") # Reliability  
  
# Marginal reliability  
marginal\_rxx(fit.2pl)

## Wrap up

* Assignment 4 is due next week.
* We will wrap up discussion of IRT
* No additional readings assigned
  + If there is anything else you wanted to talk about this semester, mention in your Readings email today or tomorrow
  + If not just send an email saying “no questions”
* Will leave class time next week to discuss plans for final project (also today!)