Week 4 Video 4

Knowledge Inference:
Item Response Theory

Item Response Theory

- A classic approach for assessment, used for decades in tests and some online learning environments

- In its classical form, has some key limitations that make it less useful for assessment in online learning
  - But variants such as ELO and CDM address some of those limitations

Key goal of IRT

- Measuring how much of some latent trait a person has

- How intelligent is Bob?
- How much does Bob know about snorkeling?
  - SnorkelTutor

Typical use of IRT

- Assess a student's current knowledge of topic X
- Based on a sequence of items that are *dichotomously scored*
  - E.g. the student can get a score of 0 or 1 on each item

Key assumptions

- There is only one latent trait or skill being measured per set of items
  - This assumption is relaxed in the extension Cognitive Diagnosis Models (CDM) (Henson, Templin, & Willse, 2009)

- No learning is occurring in between items
  - E.g. a testing situation with no help or feedback

Key assumptions

- Each learner has ability θ
- Each item has difficulty b and discriminability a
- From these parameters, we can compute the probability P(θ) that the learner will get the item correct

Note

- The assumption that all items tap the same latent construct, but have different difficulties
- Is a very different assumption than is seen in PFA or BKT

The Rasch (1PL) model

- Simplest IRT model, very popular
- Mathematically the same model (with a different coefficient), but some different practices surrounding the math (that are out of scope for this course)
- There is an entire special interest group of AERA devoted solely to the Rasch model (RaschSIG) and modeling related to Rasch

The Rasch (1PL) model

- No discriminability parameter

- Parameters for student ability and item difficulty

The Rasch (1PL) model

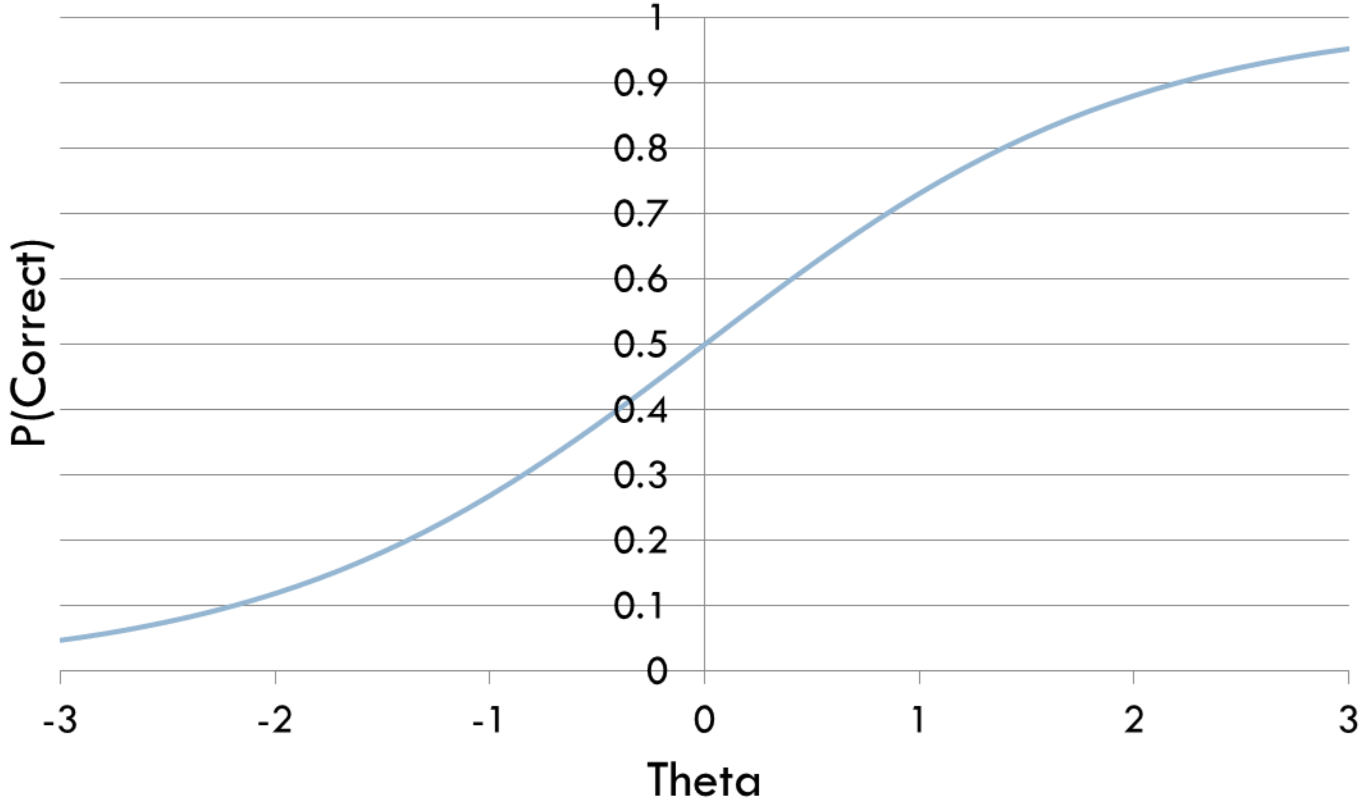
- Each learner has ability θ

- Each item has difficulty b

$$P(\theta) = \frac{1}{1 + e^{-1(\theta - b)}}$$

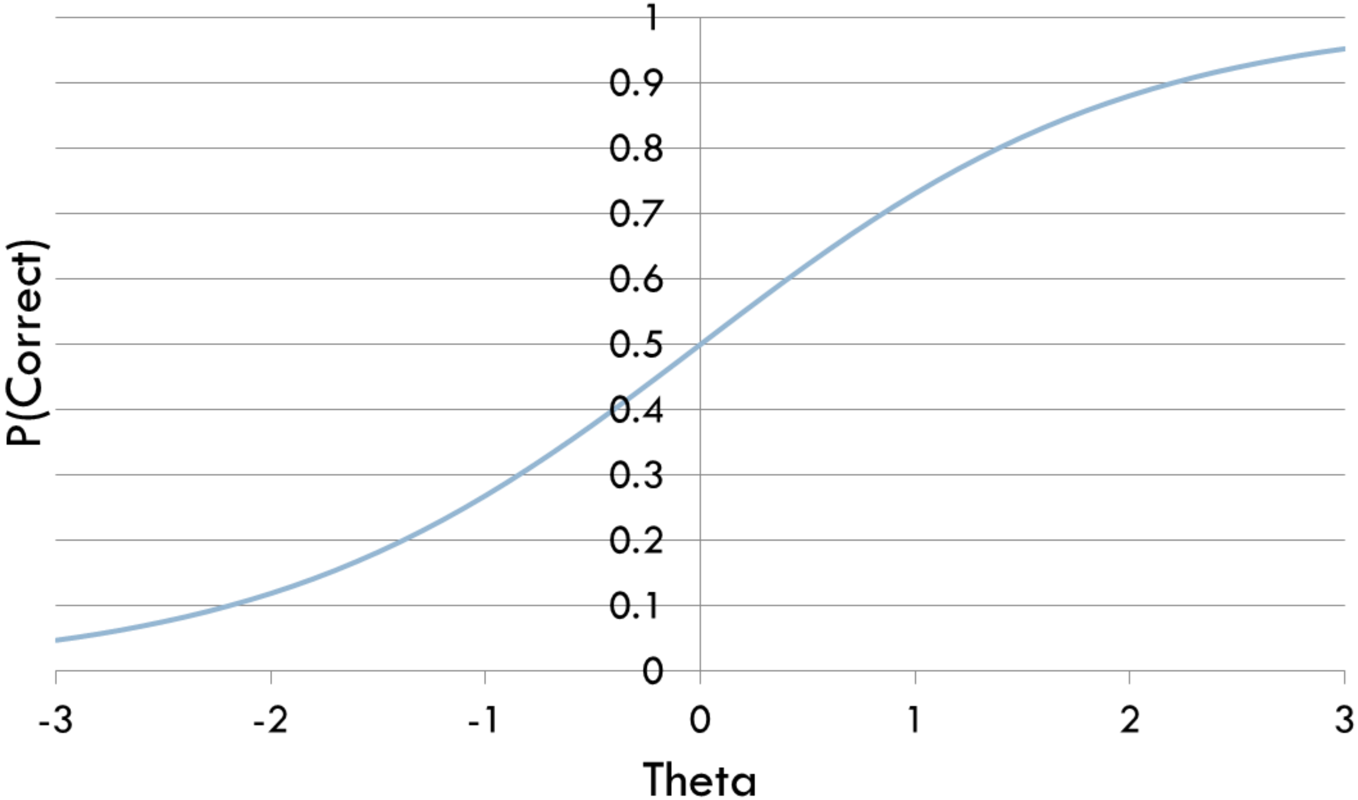
Item Characteristic Curve

- A visualization that shows the relationship between student skill and performance

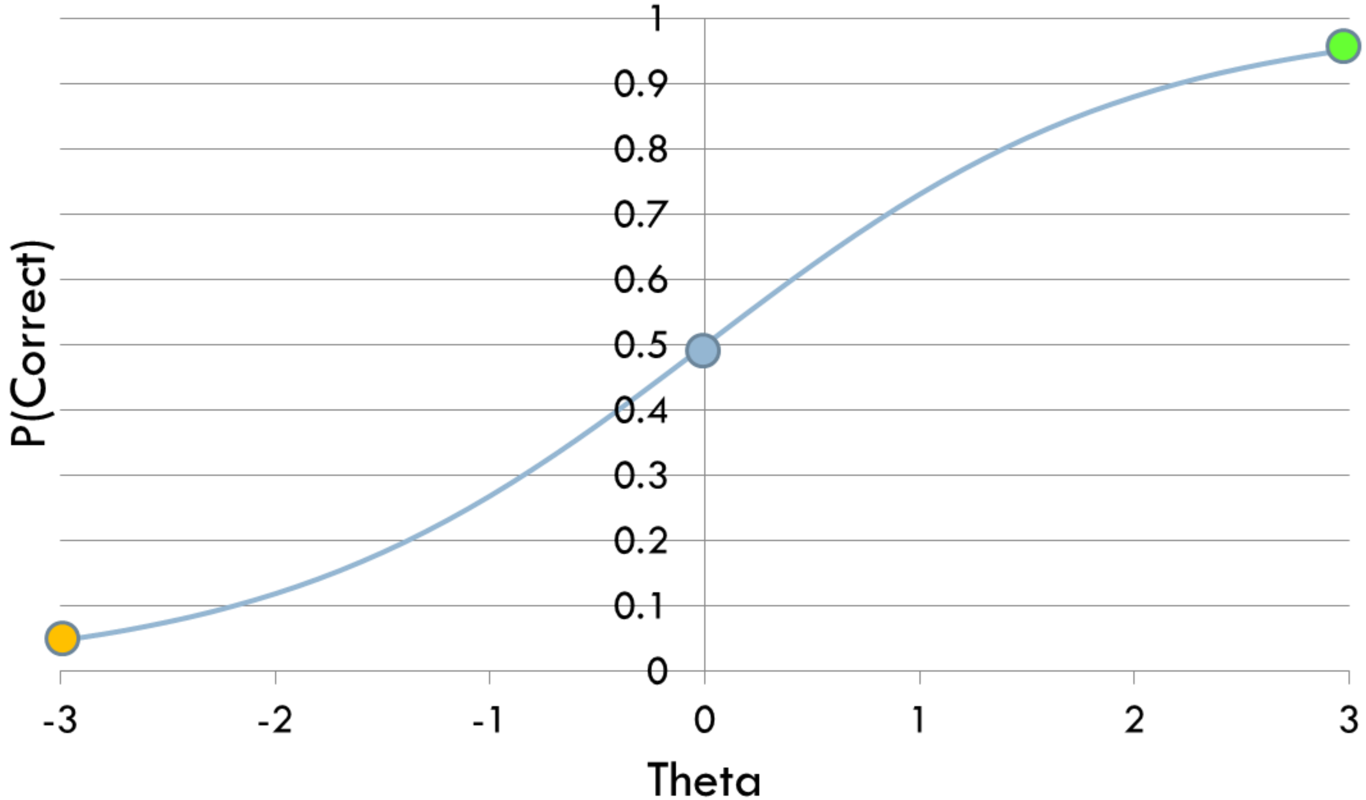


As student skill goes up, correctness goes up

- This graph represents b=0
- When θ=b (knowledge=difficulty), performance = 50%

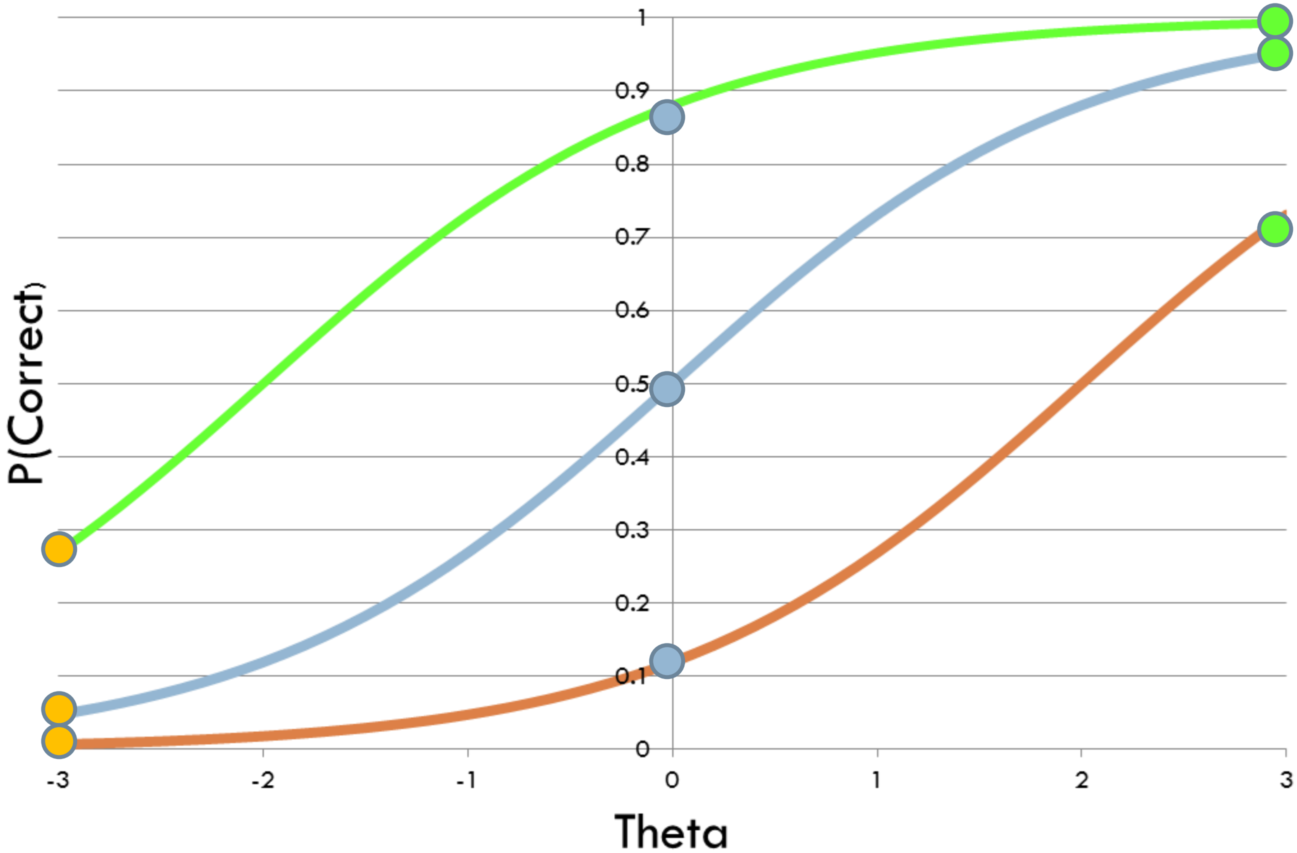


As student skill goes up, correctness goes up



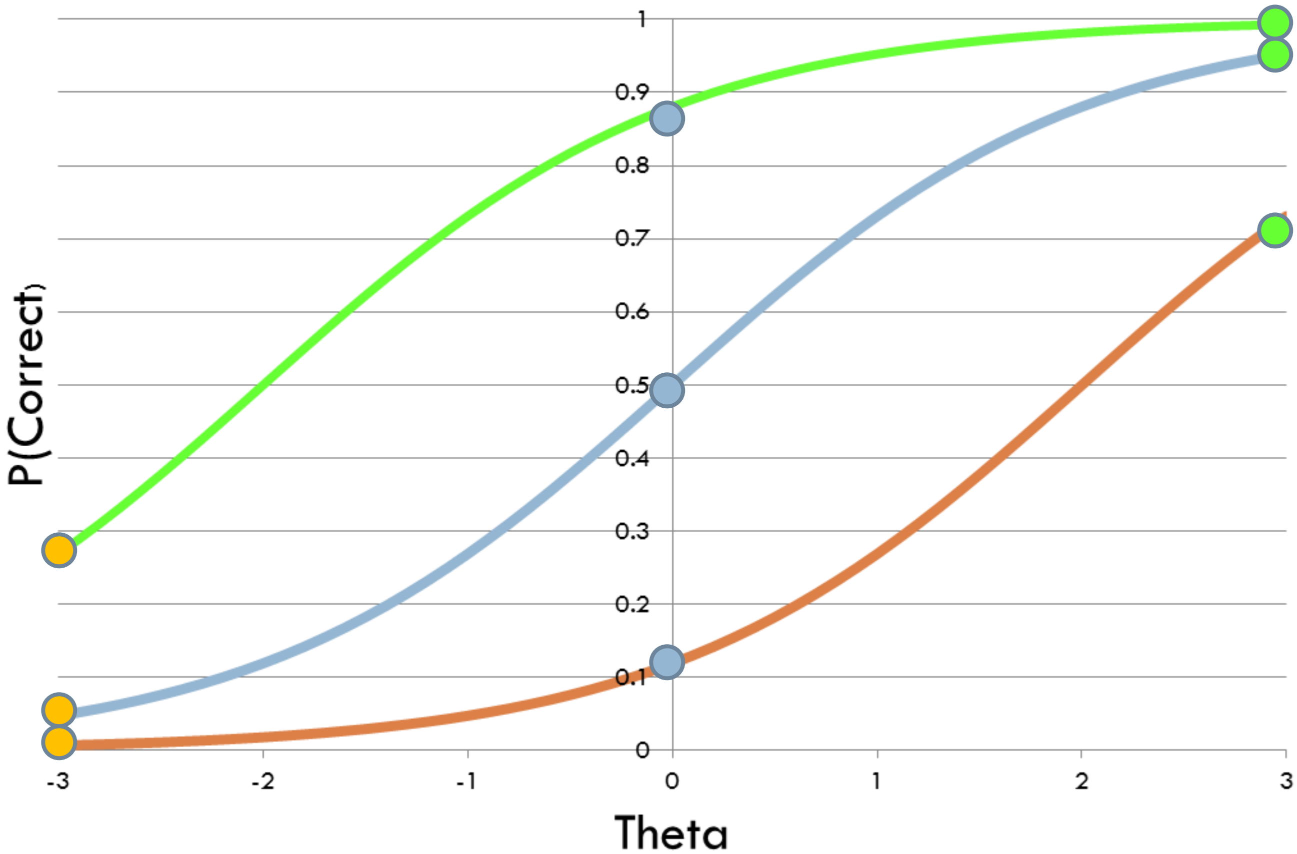
Changing difficulty parameter

- Green line: b=-2 (easy item)
- Orange line: b=2 (hard item)



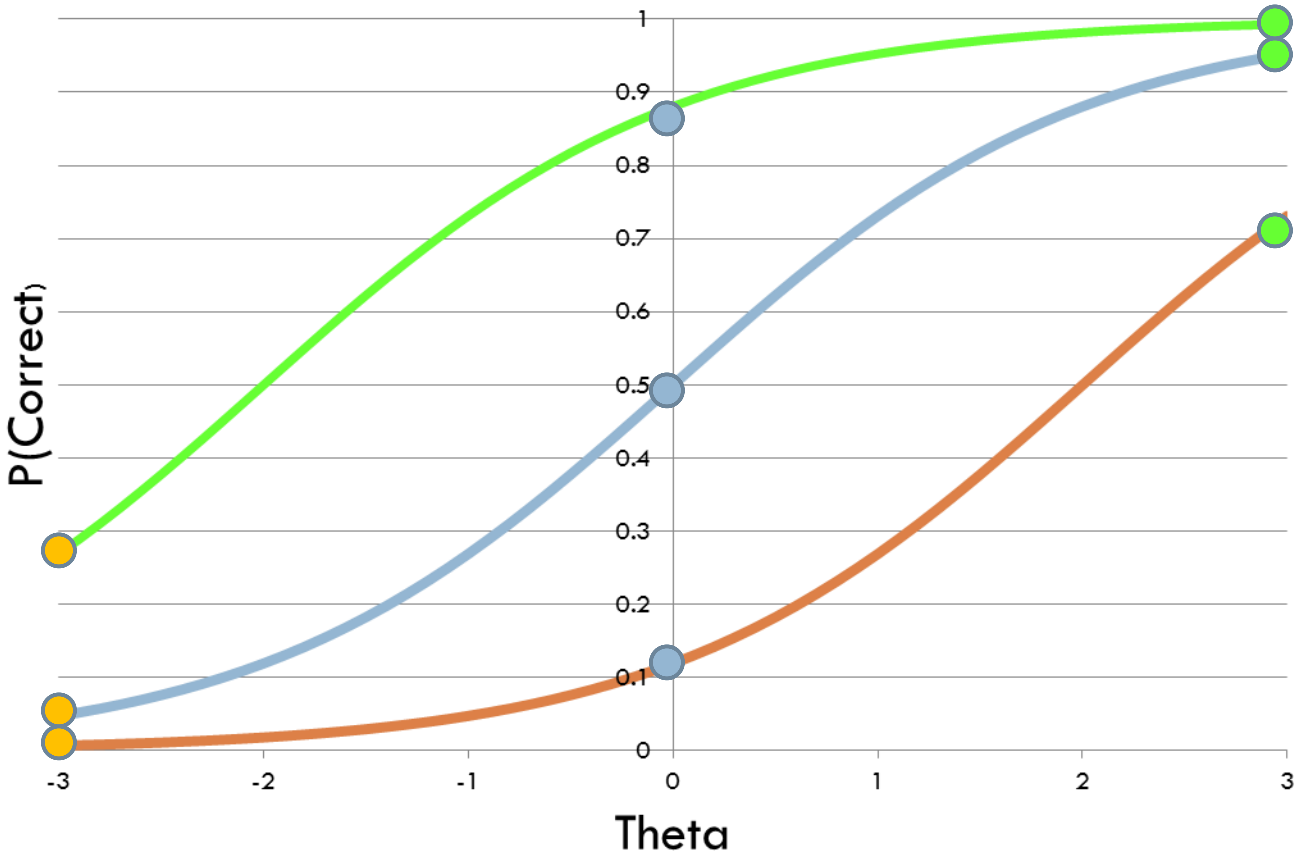
Note

- The good student finds the easy and medium items almost equally difficult



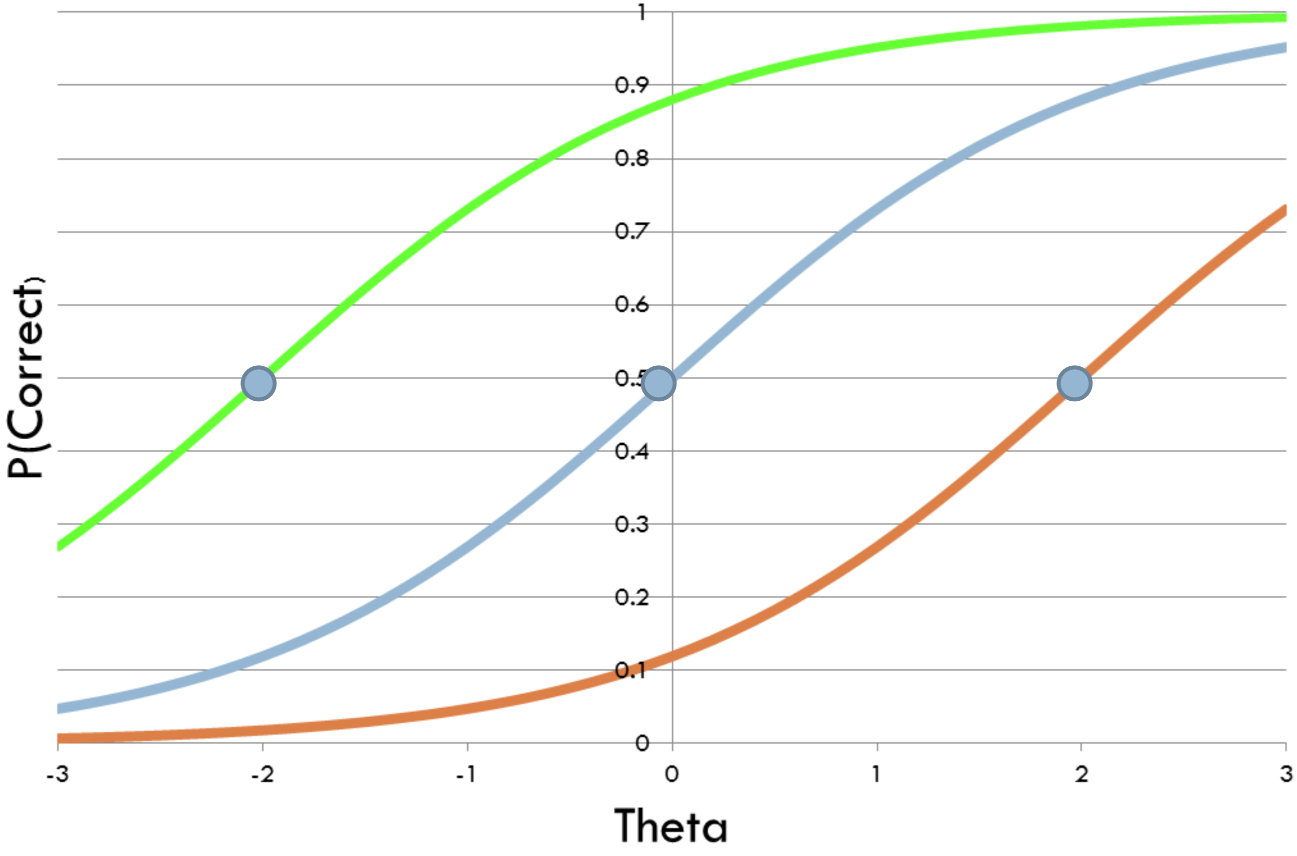
Note

- The weak student finds the medium and hard items almost equally hard



Note

- When b=θ
- Performance is 50%



The 2PL model

- Another simple IRT model, very popular
- Discriminability parameter a added

$$P(\theta) = \frac{1}{1 + e^{-1(\theta - b)}}$$

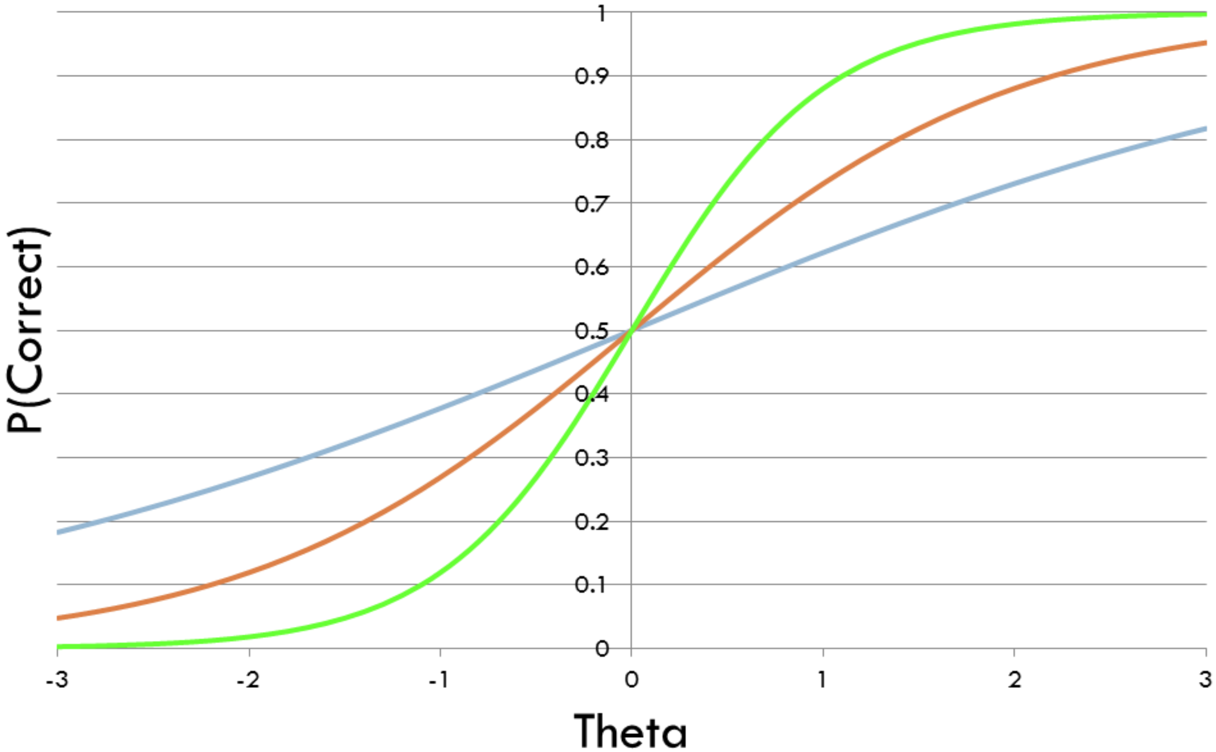
Rasch

$$P(\theta) = \frac{1}{1 + e^{-a(\theta - b)}}$$

2PL

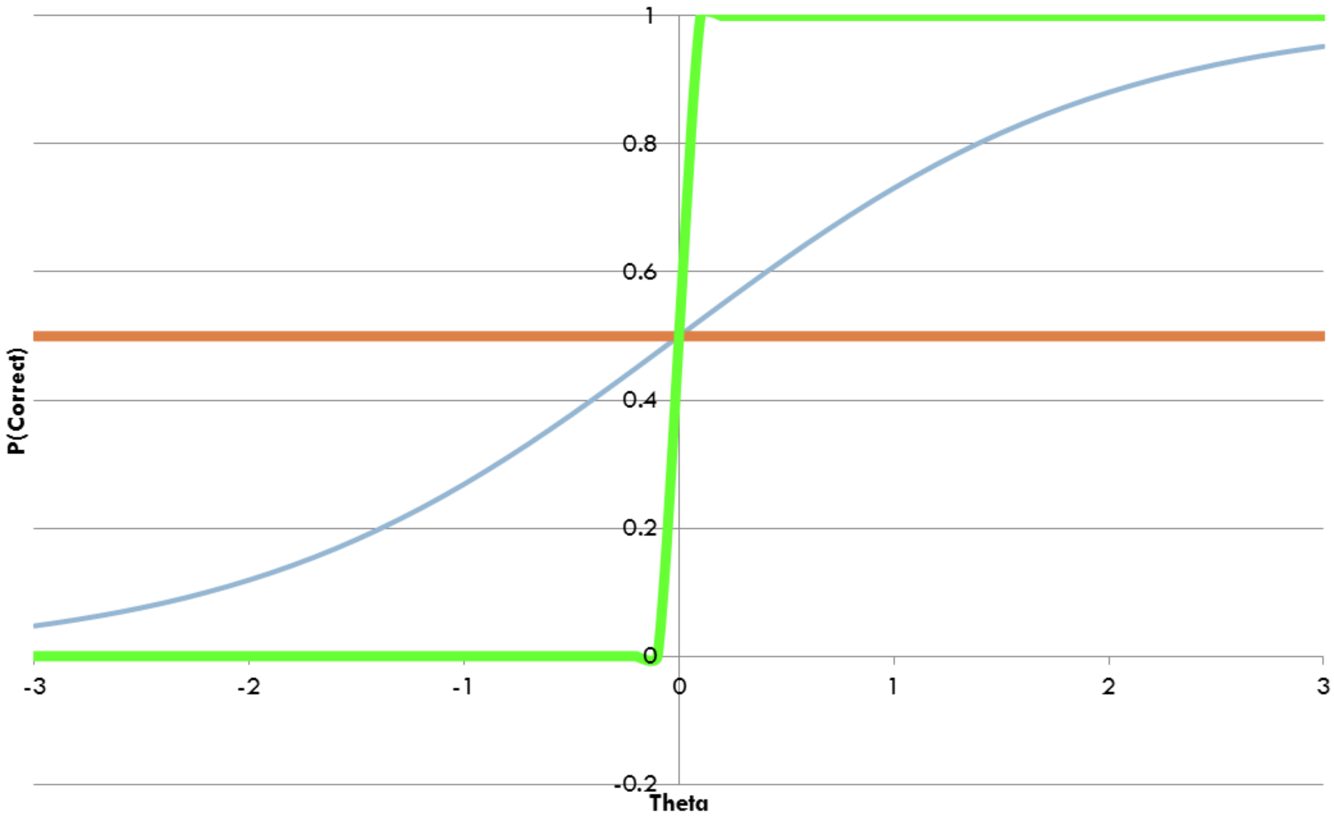
Different values of a

- Green line: a = 2 (higher discriminability)
- Blue line: a = 0.5 (lower discriminability)



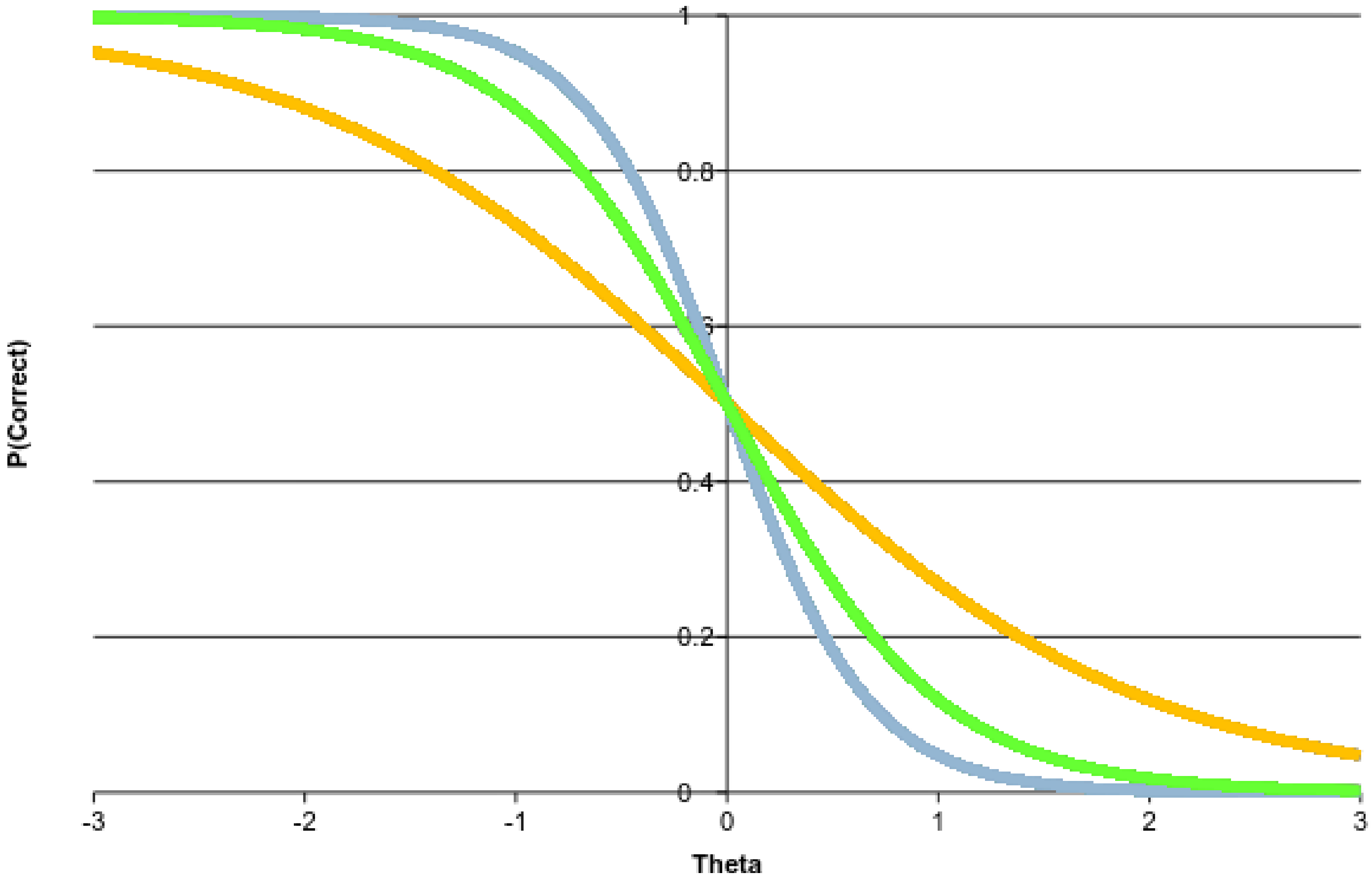
Extremely high and low discriminability

- a=0
- a approaches infinity



Model degeneracy

- a below 0…



The 3PL model

- A more complex model
- Adds a guessing parameter c

The 3PL model

$$P(\theta) = c + (1 - c)\frac{1}{1 + e^{-a(\theta - b)}}$$

- Either you guess (and get it right)
- Or you don't guess (and get it right based on knowledge)

Fitting an IRT model

- Can be done with Expectation Maximization
  - As discussed in previous lectures

- Estimate knowledge and difficulty together
  - Then, given item difficulty estimates, you can assess a student’s knowledge in real time

Uses…

- IRT is used quite a bit in computer-adaptive testing

- Not used quite so often in online learning, where student knowledge is changing as we assess it
  - For those situations, BKT and PFA are more popular

ELO (Elo, 1978; Pelanek, 2016)

- A variant of the Rasch model which can be used in a running system
- Continually estimates item difficulty and student ability, updating both every time a student encounters an item

ELO (Elo, 1978; Pelanek, 2016)

$$\theta_{i+1} = \theta_i + K\,(c - P(c))$$

$$b_{i+1} = b_i + K\,(c - P(c))$$

- Where K is a parameter for how strongly the model should consider new information

Next Up

- Advanced BKT