Useful Concepts in IRT

PSY 600K
Dr. Alyssa Gibbons
May 3, 2018

Agenda

- A few more IRT concepts:
  - Local independence
  - Item parameter invariance
  - Information
  - Linking IRT and CTT
  - Polytomous models

Local Independence

- (most) IRT models require that the test be unidimensional – or **sufficiently unidimensional**.
  - Judgment call – a multidimensional test with a large general factor may be "sufficiently unidimensional," or you can divide a multidimensional test into unidimensional subtests.
- If our test is really unidimensional, the items are **locally independent**.
  - The latent trait accounts for **all** of the observed relationship between any two item responses.
  - At a given level of θ, the items are **independent.**
    - I.e., uncorrelated!

Local Independence

- How does that work?
  - Key words: **at a given level of θ.**
  - If we don't know θ, we expect the items to be correlated!
  - Items are only independent *locally* – that is, at a particular value of θ.
- This is analogous (conceptually) to the idea of item uniquenesses being uncorrelated with factor scores and with each other.
  - All of the common variance is accounted for by the common factor (or latent trait); any other variance in item responses is unique.
- Need to test local independence – if it does not hold, there may be another latent trait affecting item responses.

Item Parameter Invariance

- In CTT, all of our parameter estimates are linked to our specific sample of test-takers and items.
  - E.g., item-total correlations or factor loadings depend on the other items in the test.
  - Item difficulty depends on the group ability level.
- In IRT, that's not the case.
  - We estimate item parameters for each item individually based on θ estimates.
  - So item parameters should be **invariant** across different combinations of items and populations of test-takers.
  - We do get the θ estimates from the items (it's an iterative process), but they also aren't linked to a specific set of items.
    - We can *equate θ metrics* across groups of examinees.

Estimating θ

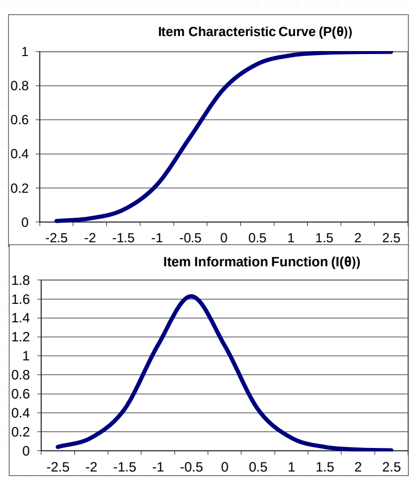
- We can estimate item parameters easily if we know participants' θ values…
  - …. And we can estimate θ if we know the item parameters!
- So how do we start?
  - Choose arbitrary start values for ability parameters.
  - Use these values to estimate the item parameters; then use these item parameters to estimate the ability parameters.
  - Keep going back and forth until these estimates converge.
- If we use **maximum likelihood estimation**, lots of items, and lots of participants, we can usually get good estimates this way.
  - One reason why IRT is so data-intensive!

Information

- Obviously, we want items that give us a lot of information.
  - Conceptually, information is the opposite of uncertainty.
- In IRT, information has a specific technical meaning:
  - The inverse (reciprocal) of the squared standard error of θ.
- But in IRT, this value is **not the same** at all values of θ!
  - It's an **information function** or curve.
  - Calculus-based, not a straightforward hand calculation.
  - $I(\theta)$ is highest at b – remember that we measure best around the item difficulty parameter.

Interpreting Information Curves

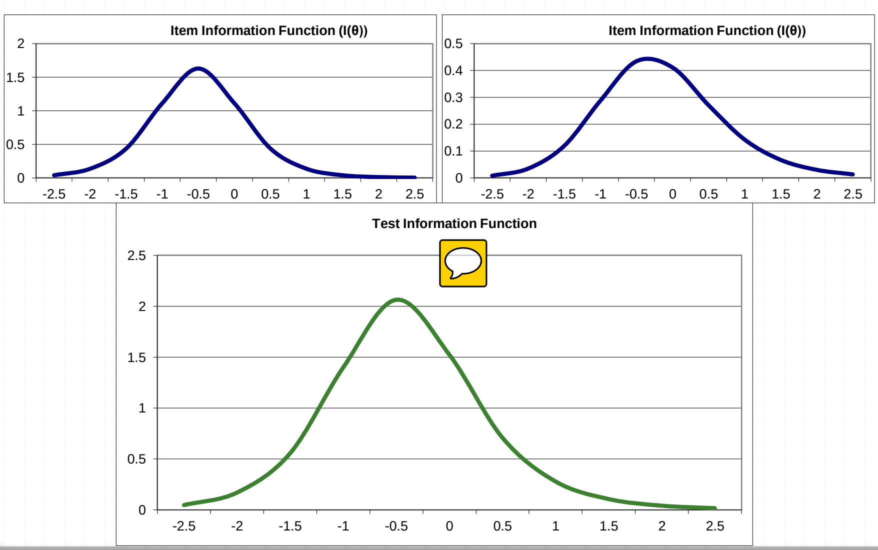
- IRT software will estimate these.
  - $1/SE(\theta)^2$ for lots of different values of θ.
- Shows you not only *how* precise your θ estimates are, but *where* they are precise.



Test Information Curves

- We can add the item information curves together to get a **test information function**.
  - Information for one item is **independent** of information for the others!
- Why is this useful?
  - We can identify the contribution of each item to the total information in the test.
  - We can choose items to cover the range of θ we are concerned about. No single item measures across the construct; but we can choose the combination we want.

Test Information Curves



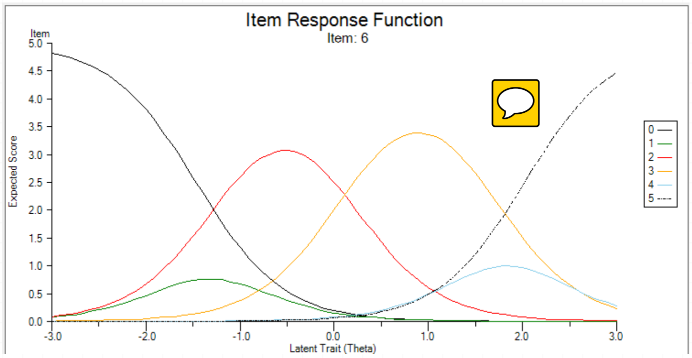
Linking CFA and IRT

- We've talked about IRT as an alternative to CFA…
- … but CFA models can be **reparameterized** to be equivalent to IRT models!
  - Both are fundamentally concerned with estimating the underlying latent construct.
  - See McDonald (1999) for details.
- This work is the basis for some modern IRT approaches.
  - E.g., multidimensional IRT models.

Polytomous Models

- What to do when you have more than 2 response options?
  - **Polytomous IRT models**
- Again, lots of choices:
  - Graded response model
  - Generalized partial credit model
  - Partial credit model
  - Rating scale model
  - (ordered from most flexible to most restrictive).

Graded Response Model Example



- The item has one overall discrimination parameter **and** a difficulty parameter for every "threshold" (# of response options – 1).

What Can You Do With a Polytomous Model?

- Use more of the information in rating-scale type items!
  - Not dichotomizing – but note the response is still seen as a categorical variable.
- Get some insight into the response scale.
  - Are the responses really ordered as we think they should be?
- Estimate **ideal point** models.
  - When you can have too much of a good thing – e.g., conscientiousness.

Questions?

Please do course evals! (online)
For next time: Final exam
Read: Look over the final exam, posted later today!
Lab Friday: Intro to IRT in Mplus