Item Response Theory
Session 12, Lecture 10
11/29/06

Outline

• Item Response Theory vs. Classical Test Theory
• Basic assumptions and concepts
• Three main types of models
• Examples from the literature
Classical vs. Item Response Theory

**CTT (Test-oriented)**
- Indices like reliability are group-specific
- Scores are test-specific
- Contribution of item measured using other items (e.g., item-total correlations)
- SEM for a fixed set of items is constant

**IRT (Item-oriented)**
- Indices are invariant
- Scores are test-independent
- Contribution of item measured independent of other items.
- SEM for a fixed set of items varies with theta
Assumptions of IRT

• **Unidimensionality**  a “one factor” model

• **Local Independence**  
  Take an individual with a given value for theta. 
  Give her two questions of specific (same) difficulty. 
  Does knowing how she did on the first tell you anything about how she’ll do in the second beyond what you knew from her value of theta? NO!

• **Invariance** – item attributes are constant across different subgroups/populations. *More on this in a bit…*

---

Item Characteristic Curves / Item Traces

plots the probability of responding correctly to an item as a function of the latent trait (denoted by $\theta$) underlying performance on the items on a test

- Pr that a random person with score $\theta$ will respond correctly to a specific item.
- Pr that a specific examinee will correctly answer a randomly chosen item with a given trace.
- NOT Pr that a specific examinee answers a specific item correctly.
1-Parameter Logistic Model (AKA Rasch Model)

- $\theta$ is value for latent trait
- $i$ indexes items
- $b_i$ is the difficulty of the item

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}$$

Ask yourself: Why is it $-b_i$?

Examples of 1-Parameter Logistic Models

Here, assuming that the trait is $\sim N(0,1)$, then if $b=0$, then one would expect a person with a $\theta$ Z-score of 0 would have a 50% chance of getting it right.

If $b=2$, you’d have to have a $\theta$ Z-score of 2, to have that same chance of being right.
2 - Parameter Logistic Model

• D - constant: (1.701) a scaling factor introduced to make the logistic function as close as possible to the normal curve.

\[ P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \]

• \( a_i \) - item discrimination

Examples of 2 - Parameter Logistic Model

The steeper the curve, the better the item is at discriminating between different values of \( \theta \).
3 - Parameter Logistic Model

$c_i$ is the chance-level parameter

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{D_a_i(\theta - b_i)}}{1 + e^{D_a_i(\theta - b_i)}}$$

*Mnemonic: $c$ is the probability that a chicken would get the right answer.*

Examples of 3 - Parameter Logistic Model
Summary

\[ b \text{ (difficulty)} \]
\[ \text{moves curve left-right} \]

\[ a \text{ (discrim.)} \]
\[ \text{makes curve shallow/steep} \]

\[ c \text{ (guessing)} \]
\[ \text{moves curve up or down} \]

Information Content

With a scale with three items like these: good at estimating theta’s around 2, not good at estimating estimating lower theta’s.
Information is a function of $\theta$, the discrimination parameter.

Information Function

\[ I(\theta) = \sum_i \frac{\left[p_i(\theta)\right]^2}{p_i(\theta)q_i(\theta)} \]

\[ \sigma_{\hat{\theta}\theta}^2 = \frac{1}{I(\theta)} \]

Information is maximized (and SE is minimized) when $\theta = b$. 
Computer Adaptive Testing

• Starts with an average item (b=0)

• Based on previous items, “zeros in” on an estimate of test-taker’s theta that is as precise as possible (by giving items with high information at the current estimate of theta).

• When the desired precision is reached, the test stops.

• Result: as brief a test as possible BUT
  • Scores among test-takers are measured on the same scale
  • All test-takers scores are measured with the same precision

Assessment of Self-Reported Physical Activity in Patients With Chronic Pain: Development of an Abbreviated Roland-Morris Disability Scale

Michael W. Stroud, * Patrick E. McKnight,† and Mark P. Jensen *


Can a long scale on pain-related disability be reduced without loss of information?
Sample Characteristics

988 patients with chronic pain (425 men, 573 women) who were screened for possible treatment at the University of Washington Multidisciplinary Pain Program.

Average age of study patients was 43.5 years, mostly white 84.4%, 59.5% reported that they were unemployed as a result of pain. 56.7% were receiving financial compensation for their pain.

Roland-Morris Disability Scale (18-item version)

1. I stay at home most of the time because of my back.
2. I walk more slowly than usual because of my back.
3. Because of my back, I am not doing any of the jobs that I usually do around the house.
4. Because of my back, I use a handrail to get upstairs.
5. Because of my back, I lie down to rest more often.
6. Because of my back, I have to hold on to something to get out of an easy chair.
7. Because of my back, I try to get other people to do things for me.
8. I get dressed more slowly than usual because of my back.
9. I only stand up for short periods of time because of my back.
10. Because of my back, I try not to bend or kneel down.
11. I find it difficult to get out of a chair because of my back.
12. My back is painful almost all the time.
13. I find it difficult to turn over in bed because of my back.
14. I have trouble putting on my socks (or stockings) because of the pain in my back.
15. I sleep less well because of my back.
16. I avoid heavy jobs around the house because of my back.
17. Because of my back pain, I am more irritable and bed tempered with people than usual.
18. Because of my back, I go upstairs more slowly than usual.
Criteria for Keeping Items

- Want high discrimination (a)
- BUT, even highly discriminant items will have low information for thetas that are far from that item’s b.
- So, you want a good range of b’s, as well.
Estimating Theta

Ok, given test responses and item characteristics, how do you estimate someone’s theta? Must be done iteratively.

\[
\hat{\theta}_{s+1} = \hat{\theta}_s + \frac{\sum_i -a_i [u_i - P_i(\hat{\theta}_s)]}{\sum_i a_i^2 P_i(\hat{\theta}_s)Q_i(\hat{\theta}_s)}
\]

Estimating Theta

Suppose your test has these three items, and someone got the first and third correct, but the second wrong.
Start with a guess about their ability: 1

\[ \hat{\Theta}_{s+1} = \hat{\Theta}_s + \frac{\sum_i -a_i [u_i - P_i(\hat{\Theta}_s)]]}{\sum_i a_i^2 P_i(\hat{\Theta}_s) Q_i(\hat{\Theta}_s)} \]

<table>
<thead>
<tr>
<th>Item</th>
<th>U</th>
<th>P</th>
<th>Q</th>
<th>A*(u-P)</th>
<th>A<em>A</em>(PQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.88</td>
<td>.12</td>
<td>.119</td>
<td>.105</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>.77</td>
<td>.23</td>
<td>-.922</td>
<td>.255</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>.5</td>
<td>.5</td>
<td>.4</td>
<td>.160</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td></td>
<td></td>
<td>-.403</td>
<td>.520</td>
</tr>
</tbody>
</table>

\[ \Delta \hat{\Theta} = -.403 / .520 = -.773 \quad \text{New } \hat{\Theta} = .227 \]

New guess about their ability: .277

\[ \hat{\Theta}_{s+1} = \hat{\Theta}_s + \frac{\sum_i -a_i [u_i - P_i(\hat{\Theta}_s)]]}{\sum_i a_i^2 P_i(\hat{\Theta}_s) Q_i(\hat{\Theta}_s)} \]

<table>
<thead>
<tr>
<th>Item</th>
<th>U</th>
<th>P</th>
<th>Q</th>
<th>A*(u-P)</th>
<th>A<em>A</em>(PQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.77</td>
<td>.23</td>
<td>.227</td>
<td>.175</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>.57</td>
<td>.43</td>
<td>-.681</td>
<td>.353</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>.35</td>
<td>.65</td>
<td>.520</td>
<td>.146</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td></td>
<td></td>
<td>.066</td>
<td>.674</td>
</tr>
</tbody>
</table>

\[ \Delta \hat{\Theta} = .066 / .674 = .097 \quad \text{New } \hat{\Theta} = .324 \]

you’d use a computer to do this; continue until score stops changing