Properties of Item Response Models

Measurement Methods

Lecture 20

Chapter 13

Today's Class

- What's done when using IRT models.
- How to get estimates of ability in:
  - Linear factor models.
  - IRT models.
- How to bound ability estimates in:
  - Linear factor models.
  - IRT models.
  - These do differ.
- The cult status of one certain IRT model.

Properties of Item Response Models

How IRT Models are Used – Two Steps

- **Two phases of work in IRT:**

1. Fitting the model
   - Determine whether or not the data are unidimensional/homogenous.
   - Estimate the item parameters.
   - Assessing goodness of fit
   - Assess the items as relatively good or bad indicators of the factor attribute/latent trait they measure in common.

2. Estimation of ability of any given examinee
   - Estimation of measurement error.
   - Employment of certain general invariance properties of common factor/latent trait models to get comparable results out of subsets of respondents [comparing populations]
     - Or subsets of items [test equating].

Concept of a Score

- **Latent traits (common factor or attribute)** are measured by infinite sets of "similar" items (not same as true score).

- **True score** is the expected value of the test score for a given value of the respondent's latent trait, i.e. E (Y | F = f) :
  - Which is equal to $\Sigma_j (\lambda_j f) + \Sigma_j \mu_j$ if linear common factor model is assumed or $\Sigma_j \phi_j(f)$ where $\phi_j()$ is the normal-ogive, the logistic or possibly some other link function)
    - See formula (13.3) and (13.4).

More Scoring

- **Formula test scores** refers to any quantity calculated from the item scores including:
  - Weighted sum of the item scores (gives different weights on different items).
  - Nonlinear function of the item scores (e.g. log-odds).
  - Simple sum score (sum of item scores).
  - Relative score (mean of item scores).

Test Characteristic Function/Curve

- The Test Characteristic Curve (TCC) represents the relationship between true score and latent trait and is equal to sum of item characteristic function/curve [see formula (13.3) and (13.4)].

- It is a glorified Item Characteristic Curve (only for test score).

Information and the Linear Item Response Model

- In linear common factor model, $X_j = \lambda_j f + \mu_j + E_j$, *f* is the value of the latent trait of a given respondent
  - Defined by an infinite behavior domain from which the *m* items are drawn.

- The best possible unbiased estimator of *f* is:

$$\tilde{f} = \sum_{j=1}^{m} w_j \left(x_j - \mu_j \right)$$

where $w_j = \left[\frac{1}{\sum \left(\frac{\lambda_j^2}{\psi_j^2} \right)} \right] \left(\frac{\lambda_j^2}{\psi_j^2} \right)$

Note this is the information of item j

- This is the maximum likelihood estimator.

Properties of the Estimator

1. It maximizes the likelihood of the item scores of the respondents and makes the observed scores more probable than any other value of *f*.

2. It minimizes the sum of squares of the differences between the item scores and their common parts, divided by their unique variances to give most weight to the best indicators.

3. It is a sufficient statistic, that is, it uses all the information of the data that can be used for the estimation of *f*.

4. It is asymptotically efficient, that is, in sufficiently large sets of items it has a smaller error of estimate than any other weighted sum of item-scores.

Wrapping Up

- The common factor model is a great tool for data analysis.
- There are many variations of the model that are used in research.
- Please be sure to read Ch. 9 carefully – this is the most important chapter in the entire book.

Variance of the Estimate

- The variance of error in using to estimate *f* (measurement error variance) is equal to:

$$\frac{1}{\sum_{j=1}^{m}\left(\frac{\lambda_j^2}{\psi_j^2}\right)}$$

- Reciprocal of this error variance is known as **test information** and is equal to:

$$\sum \frac{\lambda_j^2}{\psi_j^2}$$

More Information

- The **test information** and is equal to the sum of item information
  - This is because of local independence principle.

- This implies each item makes a separate contribution to the reduction to measurement error.

Notes About Error Variance

- The smaller the error variance, the greater is the information given by the estimator about the location of the respondent's attribute value.
- The item with the largest ratio of the squared factor loading to the unique variance makes the largest contribution to error reduction, and so gives most information.
- An item with zero loading contributes no information, that is, no error reduction.
- If we add indefinitely more items with nonzero information, the error variance reduces to zero.

Relative Efficiency of Two Tests

Relative Efficiency of Two Tests

- The **relative efficiency** of two tests, A and B, measuring the same attribute (*f*) is defined by the ratio of their error variances.
- Or equivalently, the ratio of their information functions:

  $RE(f,A,B) = I_A(f)/I_B(f) = Var(\varepsilon_B) / Var(\varepsilon_A)$

- If the first test (A) is less efficient (gives less information, or has a larger error of measurement), the ratio is less than one.

Information and Measurement Error in Item Response Models

Setup for Information

- For either logistic or normal-ogive item response models:

Probability of a response pattern given F:

$$P(X_j = x_j \mid F = f) = P_j^{x_j} Q_j^{(1-x_j)}$$

where

- $P_j = P(X_j = 1 \mid F = f) = 1/ (1+\exp(-1.701 * (b_j F + a)))$ [for logistic function],

-or-

- $P_j = P(X = 1 \mid F = f) = \Phi(Z)$, $\Phi( )$ is the area to the left of Z in standard table of the normal curve [for normal-ogive function]

- Qj = 1- Pj.

- If $X_j = 1$, $P(X_j = x_j \mid F = f) = P(X_j = 1 \mid F = f) = P_j$ and
- If $X_j = 0$, $P(X_j = x_j \mid F = f) = P(X_j = 0 \mid F = f) = Q_j$ and

Maximum Likelihood Estimation

- Given the examinee's response pattern, and regarding *f* as to be estimated, we need an expression for the likelihood function
  - The likelihood of the data for varying values of the person parameter *f*.

- Maximum likelihood estimation method leads us to choose an estimate of *f* at which this function has its maximum value
  - The value of *f* that makes the observed response pattern most probable.

Log Likelihood

- The logarithm of the likelihood is equal to:

$$\ln\left(P(X_1 = x_1, \ldots, X_m = x_m\right) = \sum_j [x_j \ln P_j + (1 - x_j) \ln Q_j]$$

Test Information About *f*

- Test information about *f* in the response pattern (using either logistic or normal-ogive functions) again is the reciprocal of the error of measurement and is given by:

$$I(f) = \frac{1}{\mathrm{Var}(\tilde{\varepsilon})} = \sum_j \frac{[P_j'(f)]^2}{[P_j(f)Q_j(f)]}$$

- The right hand side of the equation is item information where the numerator is the gradient of logistic or normal-ogive function.

Specific Model Item Information

- Specifically, for logistic models, the item information is:
- One-parameter: $(1.701)^2 P_j Q_j$
  - Here information is a function of the data only – the sum score is a sufficient statistic for *f*.
- Two-parameter: $(1.701)^2 b_j^2 P_j Q_j$ (b_j: item discrimination in Rod's parameterization)
- Three-parameter $\frac{(1.701)^2 b_j^2 Q_j}{P_j}[(P_j - c_j)^2/(1-c_j)^2]$ (c_j: guessing parameter)

Note About Information For Binary Items

- In the linear models, the item supplies the same information, and hence the measurement error variance at is equal for every value of the attribute (*f*).

- In the nonlinear item response models, the information from each item, and hence the error variance, becomes a function of the attribute.
  - Some tests measure some skill levels better than others.

Distinctions

- A distinction is made between the test information function and the test score (or formula score) information function.

- Test information function is the information in the entire response pattern and corresponds to the precision of maximum likelihood estimator.

- Test score information function is the information in the test score (e.g. sum or mean) and corresponds to the (generally reduced) precision we get when we summarize the data by test score, and transform it to estimate *f*.
  - Computation for test score information function will not be examined.

Properties of IRT Models

Review of Common IRT Models

- **The One-, Two-, and Three-Parameter Item Response Models**
- One parameter model: assumes common values for item discrimination and different values of item difficulty, no guessing parameter (i.e. zero lower asymptote).

- Two parameter model: assumes different values for item discrimination and different values of item difficulty, no guessing parameter (i.e. zero lower asymptote).

- Three parameter model: assumes different values for item discrimination and different values of item difficulty and introduce guessing parameter (i.e. non-zero lower asymptote).

Specific Objectivity

- **Specific Objectivity** refers to a comparison of attributes independent of the items chosen from a specified set of them, and a comparison of item difficulties independent of the examinees chosen from a specified population.

- The property of specific objectivity defines a class of models called *Rasch type* models.
  - *Rasch* models have attained cult-like status for some people (not for me, but check out www.rasch.org if you want to see what I mean).

- The one parameter model for binary data is an example of such models.

Properties of the 1PL or Rasch Model

- One-parameter logistic model has two properties:

(a) The sum score is a sufficient statistic for f .

(b) Comparisons of sub-populations are specifically objective
  - Independent of the item or items used for the comparison.

Specific Objectivity and Other Models

- Two parameter logistic model also has sufficient statistic but is not specifically objective.
  - It depends on the items (and item parameters of the model).
- Three parameter logistic model does not have sufficient statistics.
- All the three normal-ogive models do not have sufficient statistics.

Next Time

- Review session – bring your questions.