Item Response Models

Measurement Methods

Lecture 19

Chapter 12

Today's Class

- Item Response Models
- Item Factor Models
- Estimation
- Local independence revisited.
- How IRT and FA are nearly the same.

Limitations to Applying the Factor Model to Binary Data

Major Limitations

- Recall the linear factor model (or common factor model – here with a single factor):

$$X_j = \mu_j + \lambda_j F + E_j$$

- When items are binary, the formula represent the probability of passing an item for an examinee with a given value of F.
- A small enough value of F can result in negative probability.
- A large enough value of F can give probability greater than unity.

More Limitations

- Errors of the total test score from a set of binary items are not independent of the test score.
  - Because the mean and variance are related in binary (or categorical) items.
- The linear factor model assumes error of estimate of the factor score is constant over all values of F
  - This cannot true for binary items.

So, I Have Binary Items, What Do I Do?

- Two approaches to applying the factor model to binary data:

  1. Use of nonlinear functions suitable for conditional probabilities which make up a general class of **item response models** (also referred as *item response theory* or *latent trait theory*).

  2. Extension of the theory underlying biserial and tetrachoric correlation coefficients which is referred as **item factor analysis**.

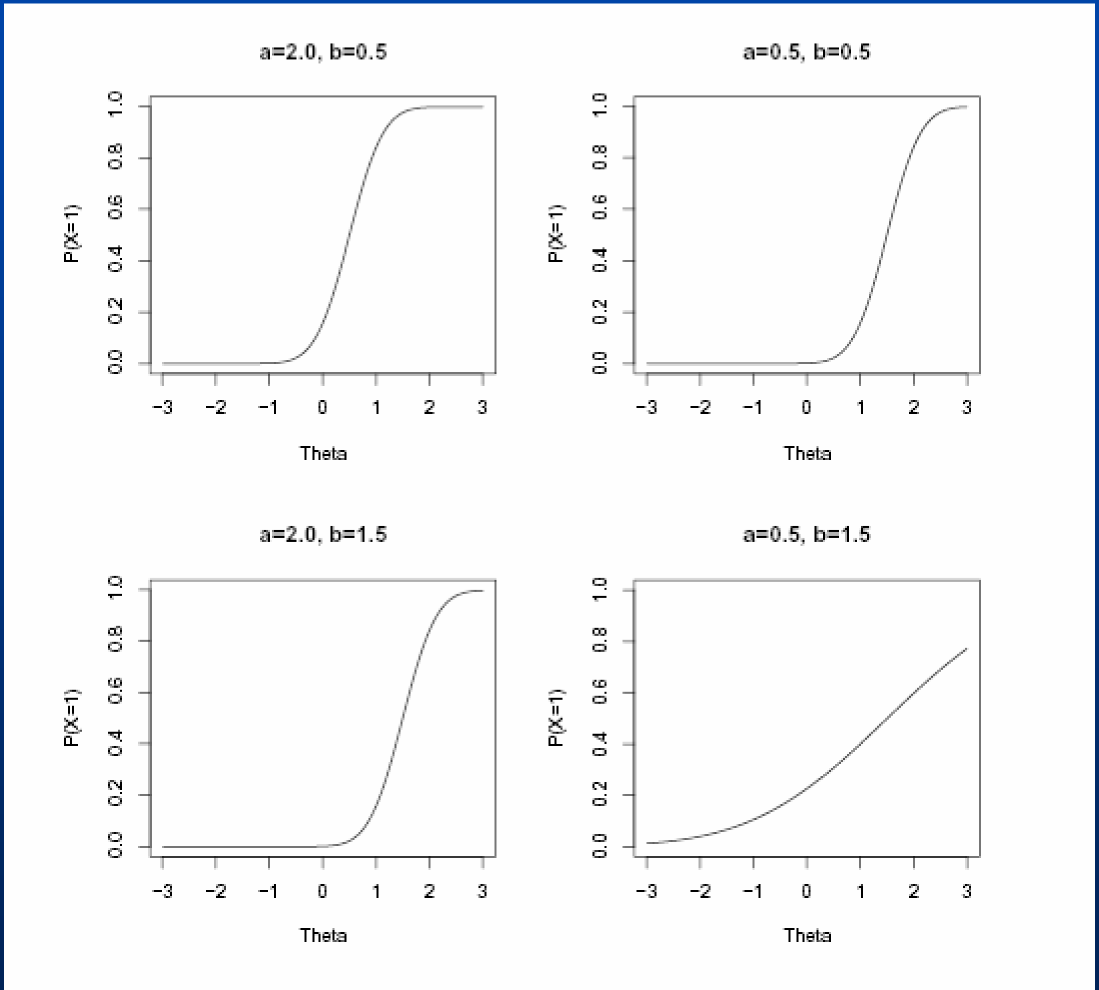
- As we will see, both methods can be (approximately) equated.

Item Response Functions (Approach #1)

The Item Characteristic Curve (ICC)

- A main component of Item Response Models is the ***Item Characteristic Curve (ICC)***
  - Also called the item response function.
  - Also called the item response curve.
- It refers to the mean/expected value of an item scores as a function of the attribute(s) measured by the test.
- For binary items it is the conditional probability of the keyed response as a function of the attribute.

Item Response Curve Example



Properties of Item Response Functions

- Three properties required to represent the conditional probability Prob(X = 1 | F = *f*):

1. It should be bounded above by one and below by zero.
   - Note: This assumption is relaxed for certain IRT models (called the 3PL and 4PL).

2. It should be smooth and monotone-increasing.
   - Note: This assumption is relaxed in non-parametric IRT.

3. It should approach horizontal asymptotes at each extreme value of F.

Link Functions

- Two commonly used item response functions are (1) **logistic function** and (2) **normal-ogive function**.
- These functions can be regarded as **link function** linking the probability of the keyed response to a linear function of the explanatory variable.
- The following formulation of the model is known as **response function parameterization**.

Logistic Link Function

- Logistic Link Function:

$$P\left(X_j = 1 \mid F = f\right) = \frac{1}{\exp\left(-1.701(bf + a)\right)}$$

- a = item difficulty (larger value means easier item)
- b = item discrimination (larger value means larger discriminating power)
- 1.701 is introduced so to get comparable result from normal-ogive function.
- F = latent trait

Rod v. Everyone Else in IRT

- The typical logistic link function is given by:

$$P\left(X_j = 1 \mid F = f\right) = \frac{1}{\exp\left(-1.701a(f-b)\right)}$$

- b = item difficulty (larger value means easier item)
- a = item discrimination (larger value means larger discriminating power)
- 1.701 is introduced so to get comparable result from normal ogive function.
- Rod calls this “Lord’s Parameterization”

Normal Ogive Link Function

- As an alternative, the normal ogive link function can be used:

$$P\left(X_j = 1 \mid F = f\right) = \Phi\left(bf + a\right)$$

- Here, the Phi function is the probability of an observation falling below Z in a normal distribution:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx$$

More About the Normal Ogive

- The normal ogive function is a means to the end, to transform a linear function of f, a, and b into a probability.

- All the parameters of the model (a, b, and f) retain the same meaning as previously.

- The difference in the curve between the normal ogive and the logistic (using 1.701 as a multiplier) is very small (less than 0.001 across the space of f).

Effect of *b* and *a* on ICC

- *b* measures item discriminating power and is also the slope parameter which controls the steepness of the slope of the curve at its point of inflexion (which is the point where it changes its direction of curvature and is the point corresponding to a probability of .5) [see Figure 12.3 p.253].

- *a* measures item difficulty level and is the intercept parameter which controls the position of the point of inflexion of the curve (at fixed value of *b*) [see Figure 12.2 p.253].

- The larger the value of a, the easier the items is.

- Position of inflexion is $-a/b$.

Family of Item Response Models

- The models you have just seen are often called the two-parameter item response models (2PL for logistic).

- There are other models with more parameters and fewer parameters.

- One-parameter family = assume ***b*** to a common value for all items and allow only ***a*** to vary from item to item (called the Rasch model).

- Two-parameter family = ***a*** and ***b*** are different for different items.

Three Parameter Models

- Three-parameter family = a pseudo-guessing parameter, ***c***, is introduced.

- For the logistic function, the 3PL IRF is:

$$P\left(X_j = 1 \mid F = f\right) = c + (1-c)\left[\frac{1}{\exp\left(-1.701a(f-b)\right)}\right]$$

- As f becomes a large negative number, the curve will approach the nonzero lower asymptote ***c***.

- This model becomes difficult to estimate without some trick applied (or use of Bayesian estimators).

- The model can be parameterized similarly for the normal ogive IRF.

Local Independence Revisited

Two Principles of Local Independence

- Rod denotes two different takes on LI, with their differences reflected in the differing estimation algorithms for IRT models:
  - Strong LI (also called full LI).
  - Weak LI (also called bivariate LI).

Strong Local Independence

- In a subpopulation in which *r* latent traits $F_1, \ldots, F_r$ take fixed values $f_1, \ldots, f_r$, the responses fare conditionally independent.

- The principle of local independence states what is meant by latent traits, that is, $F_1, \ldots, F_r$ are latent traits if the item responses are independent in a subpopulation in which $F_1, \ldots, F_r$ are fixed.

Strong LI

- Strong local independence states that, for all items:

$$\text{Prob}(X_1 = x_1; X_2 = x_2; \ldots, X_m = x_m \mid F_1 = f_1, F_2 = f_2, \ldots, F_r = f_r) =$$

$$\text{Prob}(X_1 = x_1 \mid F_1 = f_1, F_2 = f_2, \ldots, F_r = f_r) \text{ x}$$
$$\text{Prob}(X_2 = x_2 \mid F_1 = f_1, F_2 = f_2, \ldots, F_r = f_r) \text{ x}$$
$$\ldots$$
$$\text{Prob}(X_m = x_m \mid F_1 = f_1, F_2 = f_2, \ldots, F_r = f_r)$$

Weak Local Independence

- The weak (or bivariate) principle of local independence states that pairs of items *j*, *k* are uncorrelated in a subpopulation in which the *r* latent traits are fixed.

- Strong principle of local independence implies weak principle of local independence but NOT the reverse.

- The covariance between all pairs of items is zero when *r* latent traits are fixed.

- This is also the same principle we have used to define common factors.

Weak LI

- Weak local independence states that, for all ***pairs of items***:

$$\text{Prob}(X_j = x_j; X_k = x_k | F_1 = f_1,\ F_2 = f_2,\ \ldots,\ F_r = f_r) =$$

$$\text{Prob}(X_j = x_j \mid F_1 = f_1,\ F_2 = f_2,\ \ldots,\ F_r = f_r) \text{ x}$$
$$\text{Prob}(X_k = x_k \mid F_1 = f_1,\ F_2 = f_2,\ \ldots,\ F_r = f_r) \text{ x}$$

LI and Estimation Methods

- These two principles affect the types of estimation methods used in IRT.

- Estimation methods using strong principle of local independence are known as ***full information methods***, which use the distribution of item responses
  - The population probabilities and sample frequencies of the 2^m response patterns.
  - Bock & Aitkin's Marginal Maximum Likelihood method.

Weak LI Estimation

- Estimation methods using weak principle of local independence are known as ***bivariate information methods***, which use just the probabilities/sample frequencies of the keyed responses to pairs of items.
- This is the method used in NOHARM (normal ogive by harmonic analysis robust method) .

Additional Notes on LI

- Notice latent traits are the precise counterparts in nonlinear models for common factors as in linear models.
  - They are defined by the property that they account for association between item responses and they are interpreted as the attributes the items measure in common.

- In general, the principle of local independence is <u>not</u> itself a testable assumption.
  - It states what is meant by latent traits or common factors.

- When LI is violated, it has been theorized that additional factors are not accounted for in the analysis (see the work of Stout and his students).
  - To this end the DIMTEST program was developed to “test” for unidimensionality based on the LI assumption.

Item Factor Analysis
(Approach #2)
-or-
Meet the New Boss, Same as the Old Boss

Item Factor Analysis

- An alternative *parameterization* of the model in terms of underlying quantitative "response tendencies" – **common factor parameterization**.

- Each binary item has associated with it an "underlying" quantitative response tendency X_j^* and a threshold value τ_j, such that:

  If $X_j^* > \tau_j$ then $X_j = 1$

  If $X_j^* \le \tau_j$ then $X_j = 0$

Underlying Response Model

- The underlying response tendencies, X_1^*, X_2^*, … , X_m^* fit a factor analysis model, say Spearman single factor model (mean omitted – see below):

$$X_j^* = \lambda_j F + E_j^*$$

- With uncorrelated unique parts E*.
- We impose a scale on each X_j^* so that it is standardized:
  - With mean zero (hence no μ_j).
  - With variance one, so $\lambda_j^2 + \psi_j^2 = 1$.

Building on the Previous Model

- F and each E_j* have a normal distribution.
- Each X_j* has a normal distribution.
- This leads to, for item *j*,

$$P\left(X_j = 1 \mid F = f\right) = P\left(X_j^* > \pi \mid F = f\right) = \Phi\left(\left[\frac{\lambda_j}{\sqrt{1-\lambda_j^2}}\right] f - \left[\frac{1}{\sqrt{1-\lambda_j^2}}\right] \pi_j\right)$$

- Larger value of λ_j means larger discriminating power.
- The larger the value of τ_j, the more difficult the item.

Meet the New Boss…

- The item factor parameterization is an equivalent model (or nearly equivalent if logistic link is used) to those described in Approach #1.

- What is important to take from this is that:
  - If you know Factor Analysis
  - If you understand link functions
  - Then you know the basics of Item Response Theory!!!!!!!

- Please don't say, "I am not an IRT person" if you do factor analysis (or any other latent variable technique).
  - You can refer to factor loadings and item discrimination synonymously.

…Same as the Old Boss

- We can relate our new item factor analysis parameterization to the IRT parameterization:

	Item Factor Analysis	Rod's IRT
Item Discrimination	$\lambda_j = \left[\frac{b_j}{\sqrt{1+b_j^2}} \right]$	$b_j = \left[\frac{\lambda_j}{\sqrt{1-\lambda_j^2}} \right]$
Item Difficulty	$\pi_j = \left[\frac{-a_j}{\sqrt{1+b_j^2}} \right]$	$a_j = \left[\frac{-\tau_j}{\sqrt{1-\lambda_j^2}} \right]$

So Why Use One Parameterization over the Other?

- The common factor parameters are most useful in a preliminary examination of the structure of the data.
  - Because we can use established factor-analytic criteria for judging the sizes of the factor loadings.

- The response function parameterizations are useful in applications of a fitted model because they generally simplify computations.

Interpretation of Item Factor Parameters

- One can interpret the common factor parameters in relation to classical item analysis:

1. Probability of the keyed response to each item, π_j, (or proportion of respondents give "keyed" response) is related to the threshold, τ_j, given by $\Phi(-\tau j)$.

2. The factor loading, λ_j, of the item is the product-moment correlation between X_j* and F
   - Which is the biserial correlation between binary X_j and F.

3. The product of the factor loadings between any pair of items (j and k) gives the model estimate of the tetrachoric correlation between the items:

$$\hat{\rho}^*_{jk} = \hat{\lambda}_j \hat{\lambda}_k$$

Item Parameter Estimation

Item Parameter Estimation

- Item parameters can be estimated by several methods:

1. Heuristic Method:

- Substitute sample tetrachoric correlation for population tetrachorics and sample proportions giving keyed responses for population probabilities, we can get consistent estimates of the threshold (τj) and factor loading (λj) parameters of the item factor model by fitting the linear common factor model to the tetrachoric coefficients using ULS.

  – Applying inverse Φ to sample proportions to give estimates of τj.

Deficiencies of Heuristic Methods

- Deficiencies of heuristic methods:
  - Does not give standard errors for the parameter estimates (done via ULS).
  - Does not give a statistical basis for judging goodness of fit.
  - Cannot handle the three-parameter model with a pseudo-guessing parameter.

Bivariate Information Methods

2. Bivariate information methods:
   – e.g. Weighted least squares method by Christoffersson (1975) and Muthen (1978) which gives chi-square and standard errors of parameter estimates.

   – NOHARM method (by McDonald, 1980) gives ULS (unweighted least squares) fits a polynomial to the data that is close to the normal-ogive which yields a fitted item covariance matrix, discrepancy matrix (departures from pairwise independence of the items after their associations are accounted for by the model)
      - Chi-square and standard errors are not provided in NOHARM.

Full Information Methods

3. Full information methods

- e.g. Bock and Aitkin's marginal maximum likelihood method which also gives chi-square and standard errors of parameter estimates which uses the frequencies of all the response patterns.
  - This method is limited in applications because the total number of response patterns – 2^m – grows rapidly with the number of items.

- BILOG

Remarks About Estimation

- Population from which a sample gives the estimated item parameters are known as the **<u>calibration or reference population</u>**.
- This population determines the origin and unit of measurement.
- The mean of F is zero and the variance of F is unity in this population.
- The estimation of the item parameters are referred as **<u>item calibration</u>**.
- The estimated parameters are then used to estimate the latent traits of new examinees.

Wrapping Up

- Item response models have much in common with linear factor models.
- This lecture was an introduction to the item response models.
- The next lecture will focus on a few properties of item response models.
  - Where they slightly differ from linear factor models due to the link function.

Next Time

- Chapter 13 – Properties of Item Response Models.