Multidimensional Item Response Theory

Lecture #12
ICPSR Item Response Theory Workshop

Overview

- Basics of MIRT
  - Assumptions
  - Models
  - Applications

- Guidance about estimating MIRT

MULTIDIMENSIONAL ITEM RESPONSE THEORY (MIRT) MODELS

Why MIRT?

- Many of the more sophisticated approaches to IRT (including MIRT) exist because of “problems” with certain kinds of test data

- By “problems” we basically mean: the data don’t meet the assumptions of the IRT model
  - Looking at it from the more common educational measurement point-of-view

- Or…theory posits the existence of more than one trait

Multidimensional IRT (MIRT)

- Unidimensionality is reasonable for many educational tests, but not for all
  - Personality measures
  - Achievement tests with multiple areas

- If violations of unidmensionality are found:

  (1) delete items

  (2) assume model robustness and carry on

  (3) form clusters and scale separately (e.g., SATs)

  (4) use a MIRT model.

Goals of MIRT

- Represent people with a set of ability scores
  - One for each trait or dimension believed to be generating the data

- Represent each test item by a set of item parameters relating each trait to the item

Types of MIRT Models

- **Compensatory** – allow high ability on one dimension to compensate for low ability on other dimensions
  - Much more common
  - Additive in the logit

- **Non-Compensatory** – no such compensation; High probability means high ability for all dimensions
  - Infrequently used, often difficult to estimate
  - Non-additive (think latent variable interactions)

Difference in Models

- **Compensatory** – linear terms are summed;
  - high theta on one or more dimensions can “save” you from low theta(s) on other dimensions

- **Non-Compensatory** – linear terms are multiplied
  - Theta must be high for all dimensions to maintain a high probability of success

MIRT Applications

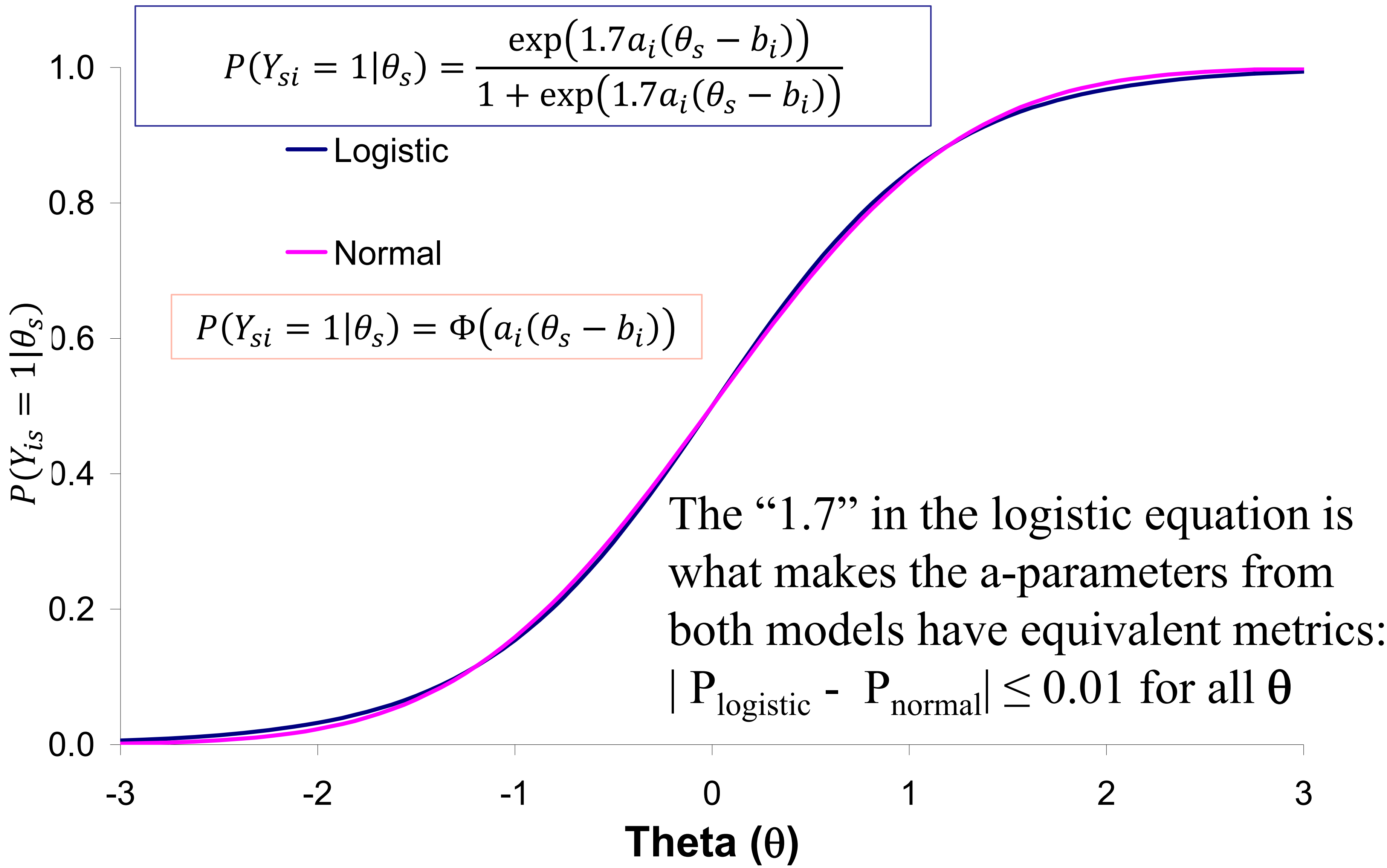
- Multi-trait feedback in Education
  - One potentially useful application: using MIRT models to derive scores on subscales within a test

- Personality Assessment
  - Tests usually designed to measure more than one (possibly correlated) latent traits at one time

Different MIRT Approaches

- Main Differences: specification of the non-linear link in the MIRT model
  - Normal Ogive MIRT (McDonald)
    - Uses a cumulative normal density to model the probability of success on an item
  - Logistic MIRT (Reckase)
    - Uses the more-familiar logistic function to model the probability of success on an item

Equivalence of the Normal Ogive and Logistic Metrics



$$P(Y_{si} = 1|\theta_s) = \frac{\exp\big(1.7a_i(\theta_s - b_i)\big)}{1 + \exp\big(1.7a_i(\theta_s - b_i)\big)}$$

$$P(Y_{si} = 1|\theta_s) = \Phi\big(a_i(\theta_s - b_i)\big)$$

The "1.7" in the logistic equation is what makes the a-parameters from both models have equivalent metrics: $|P_{logistic} - P_{normal}| \leq 0.01$ for all θ

“Slope/Intercept” Form

- McDonald presents the normal ogive MIRT model in slope-intercept form:

$$P(Y_{si} = 1|\boldsymbol{\theta}_s) = \Phi(\beta_0 + \boldsymbol{\beta}^T\boldsymbol{\theta}_s)$$
$$= \Phi(\beta_0 + \beta_1\theta_{s1} + \cdots + \beta_k\theta_{sk})$$

- MIRT may be thought of as a special case of “item” factor analysis
  - The purpose of the normal-ogive approach to MIRT in slope-intercept form is make this clear

- MIRT is basically a non-linear factor analysis, where the slope parameters are analogous to factor loadings

Logisitic Approach

- Reckase presents the logistic MIRT model in slope-intercept form:

$$P(Y_{si}=1|\boldsymbol{\theta}_s)=c_i+(1-c_i)\frac{\exp\left(\boldsymbol{a}_i^T\boldsymbol{\theta}_s-d_i\right)}{1+\exp\left(\boldsymbol{a}_i^T\boldsymbol{\theta}_s-d_i\right)}$$

- $\mathbf{a}^T$ is a transposed vector of discrimination parameters, one for each dimension
- d is the intercept parameter, related to difficulty
- c is the guessing parameter, same as always
- The θ is the vector of person latent traits

Assumptions

- Monotonically increasing relationship between trait and probability of endorsement or correct response

- The function relating trait and item response is "smooth," meaning derivatives are defined (makes the model both functional and estimable)

- Local Independence of item responses (conditional on θ, a vector of latent traits)

Person Parameters

- θ is a vector of latent traits, one for each dimension
  - The # of dimensions is based on theory
    - Don't do exploratory analyses (although a lot of people in Education think that is MIRT)
  - The θ dimensions are associated
    - Covariances (correlations) are estimated

Discrimination

- **a** is a vector of item slopes, one for each dimension measured by an item
  - Same interpretation as the unidimensional IRT model
  - Unless there is "simple structure," an item will be more discriminating for combinations of dimensions that for a single dimension
- Discriminating power of an item for the most discriminating combinations of dimensions (where all θ are close to b):

$$\eta_{MDISC} = \sqrt{\sum_{k=1}^{p} a_k^2}$$

$$p = \#\text{ of dimensions}$$

Multidimensional Item Difficulty

- The value “d” from the model is an intercept,
- Like –ab, except that there is more than one a-parameter (still just one b, though). The equivalent to the b-parameter is:

$$b_{MDIFF} = \frac{-d}{MDISC}$$

Lower Asymptote

- The "c" parameter has just the same interpretation as from unidimensional models
  - It is now the lower asymptote in a multidimensional space

- c = probability of success for an examinee who is low for all thetas

Other Statistics

- Test Charactertic Curve is now thought of as a **Test Characteristic Surface**, but is computed in the same way (i.e., by summing ICCs)

- Test Information is also computed in the same way
  - The function becomes a surface, not just a curve

Re-parameterized Model

- Reckase (1985, 1997)

$$P_{ij} = c_j + (1 - c_j)\frac{e^{DZ_j(\underline{\theta}_i)}}{1 + e^{DZ_j(\underline{\theta}_i)}}$$

$$\text{where } Z_j(\underline{\theta}_i) = \sum_{m=1}^{M} a_{jm}\theta_{im} + b_j^* + \eta_j$$

$$\text{and} \quad b_j^* = -\eta_j b_j$$

The result is an Item Characteristic ***Surface***, instead of a Curve, with these parameters:

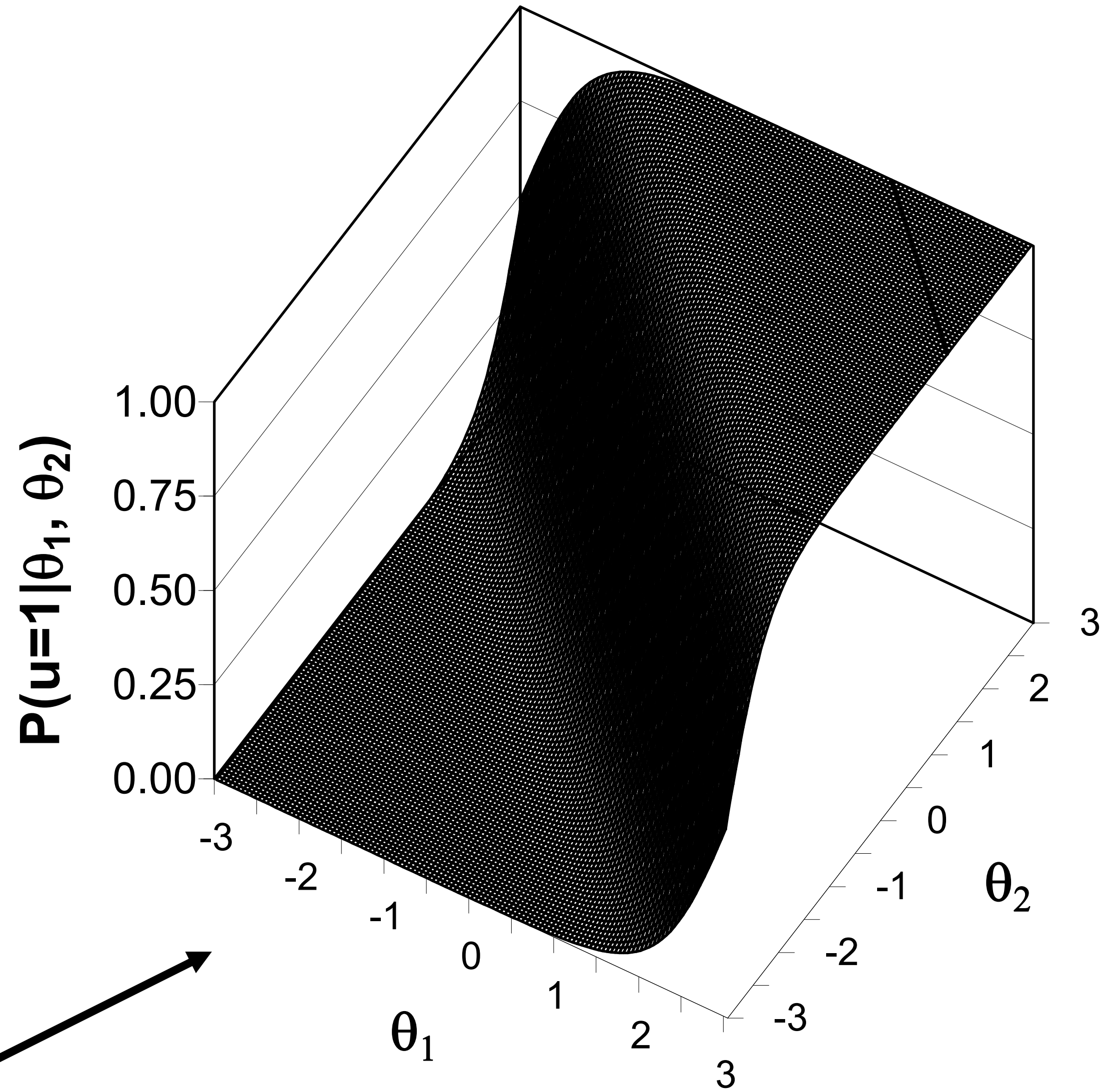
1 θ per dimension

1 ***a*** per dimension

1 ***b***

1 ***c***

a1	***= 1.5***
a2	***= 1.5***
b	***= 1.0***
c	***= 0.0***



This is a 2-d MIRT model solution… any more and you couldn't graph it!

The result is an Item Characteristic ***Surface***, instead of a Curve, with these parameters:

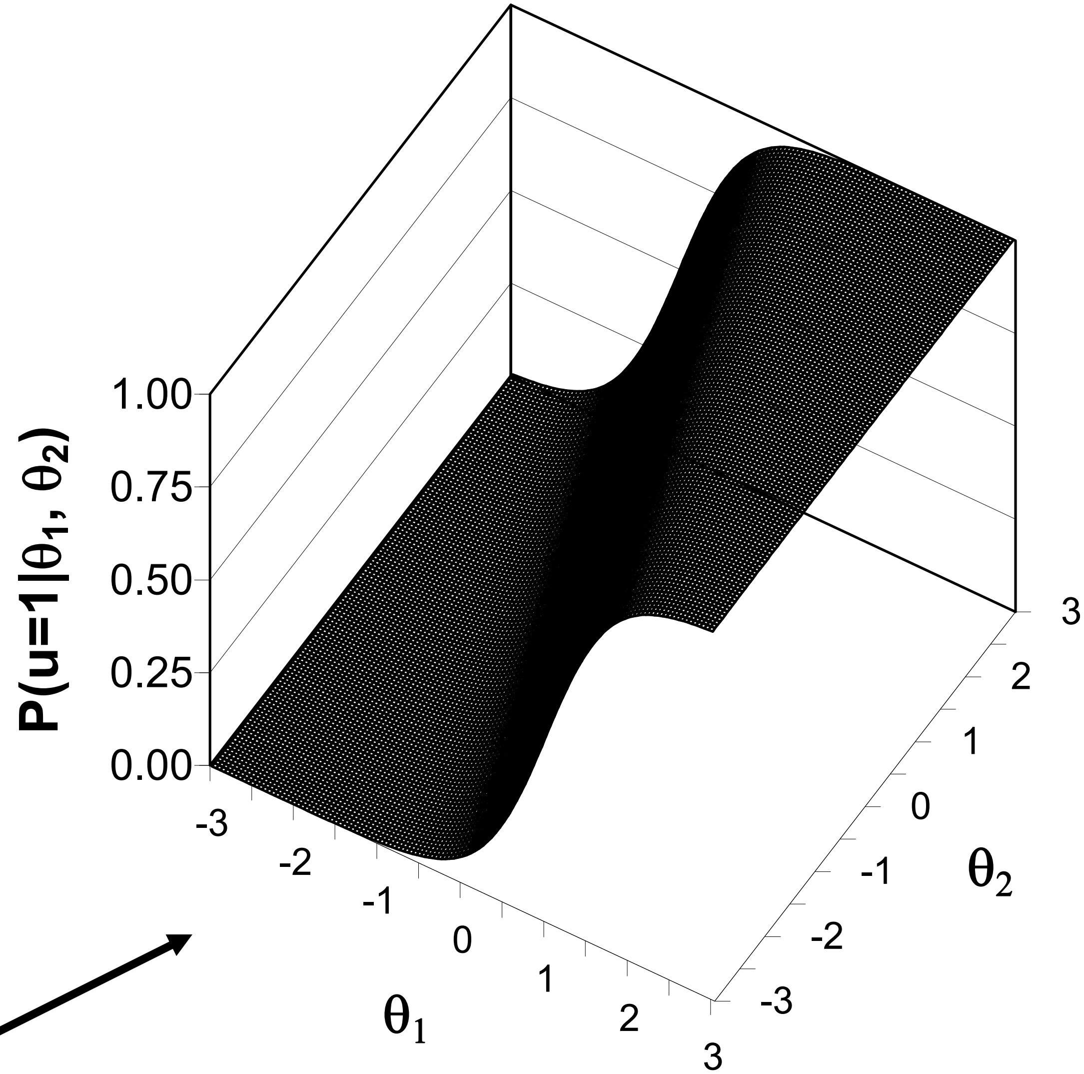
1 θ per dimension

1 ***a*** per dimension

1 ***b***

1 ***c***

a1	***= 0.5***
a2	***= 1.5***
b	***= 1.0***
c	***= 0.0***



This is a 2-d MIRT model solution… any more and you couldn't graph it!

The result is an Item Characteristic ***Surface***, instead of a Curve, with these parameters:

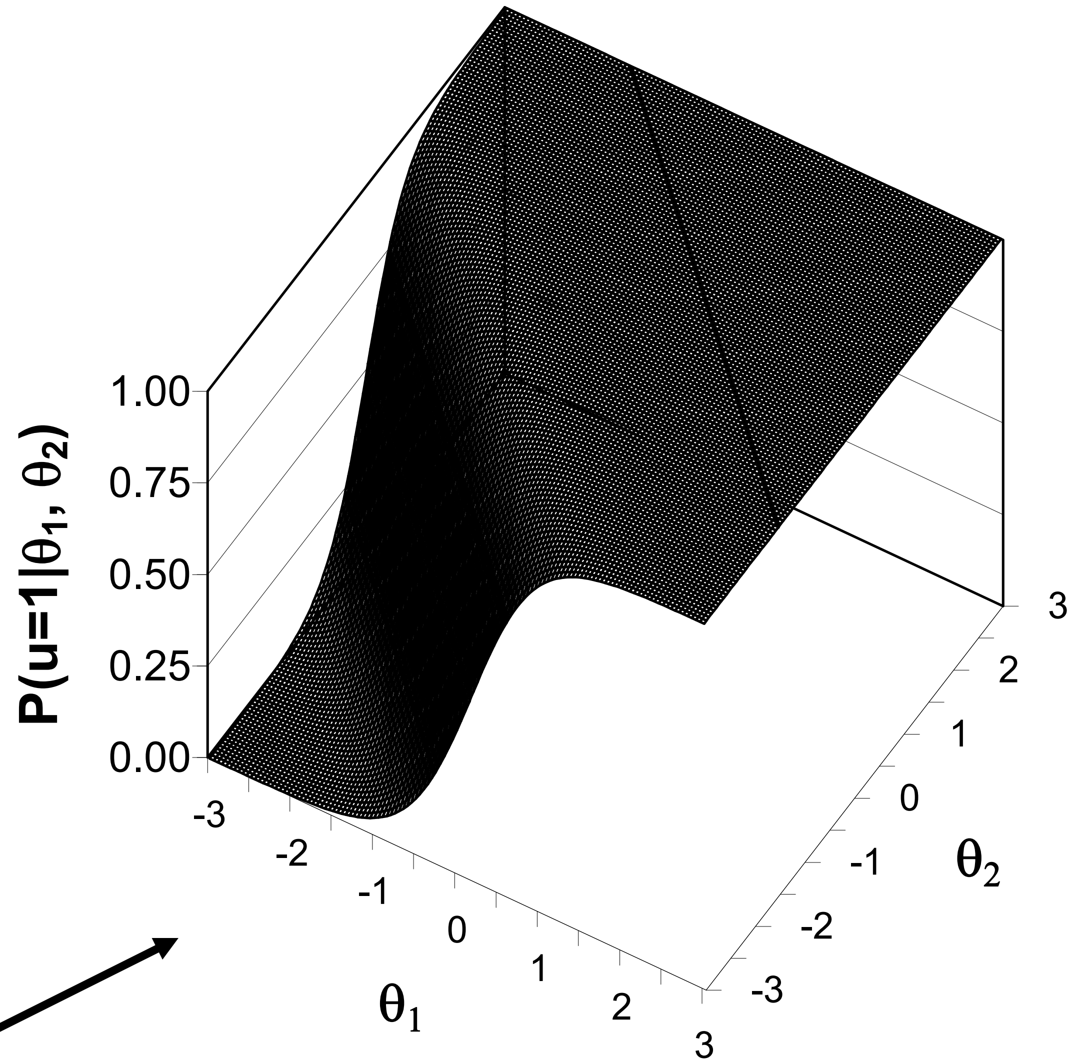
1 θ per dimension

1 ***a*** per dimension

1 ***b***

1 ***c***

a1	***= 1.5***
a2	***= 1.5***
b	***= 0.0***
c	***= 0.0***



This is a 2-d solution…
any more and you couldn't graph it!

The result is an Item Characteristic ***Surface***, instead of a Curve, with these parameters:

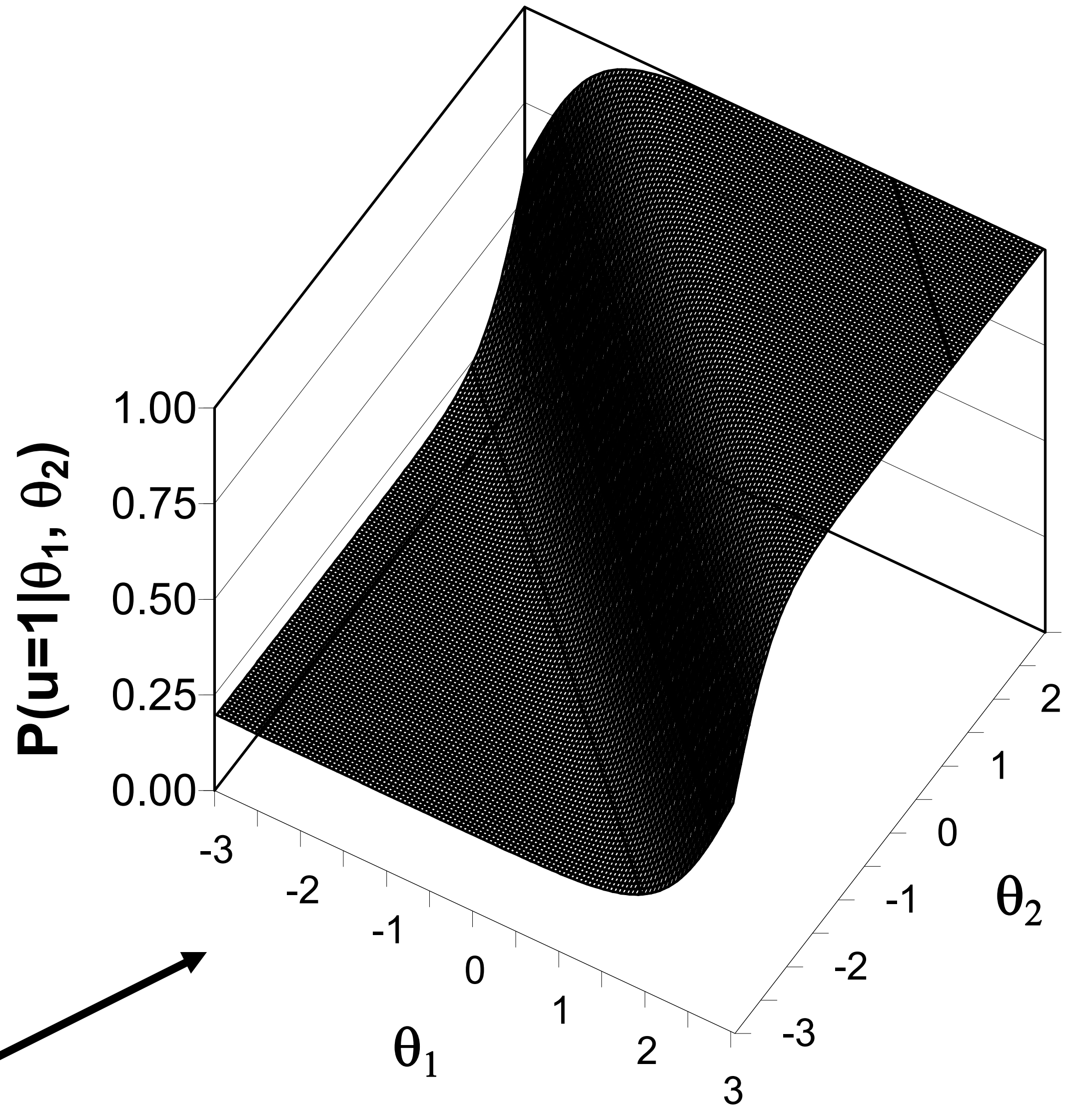
1 θ per dimension

1 ***a*** per dimension

1 ***b***

1 ***c***

a1	***= 1.5***
a2	***= 1.5***
b	***= 1.0***
c	***= 0.2***



This is a 2-d solution…
any more and you couldn't graph it!

ESTIMATION OF MIRT MODELS

MIRT Model Estimation

- MIRT model estimation via marginal maximum likelihood (MML) is difficult to impossible depending upon the size of
  - Your data
  - The number of thetas
- The issue is with the marginalization (integration); each theta exponentially increases the number of quadrature points needed
- Furthermore, sample sizes needed for good convergence are likely to be well beyond reasonable
- So…approximations must be used
  - Limited information
- …or MCMC with the ability to "tighten" priors for misbehaving parameters

MIRT Software: Mplus

- Mplus is the most comprehensive package to estimate MIRT models
  - Default estimator is a limited information estimator (weighted least squares mean and variance; WLSMV)
    - More commonly known as diagonally weighted least squares
  - MML is available
    - Integration can be by Monte Carlo (speeds the estimator)
  - MCMC is also available

MIRT Software: NOHARM

- Normal Ogive Harmonic Analysis (Robust Method): NOHARM
  - Fits the Normal Ogive MIRT model
  - Authors: McDonald & Fraser (1988)
    - Can be contacted for a copy of the software
  - Provides multidimensional item parameter estimates, but does not provide multidimensional person parameter estimates

- One of original programs
  - Uses limited information (but differently from Mplus)

MIRT Software

- TESTFACT
  - Part of the SSI suite of IRT software
  - Performs factor analysis as part of testing dimensionality, fits MIRT models
  - Person and item parameters provided
  - Used for exploratory analyses :: BAD!

CONCLUDING REMARKS

Concluding Remarks

- MIRT (and IRT in general) can be thought of as simply a non-linear factor analysis
  - More on this later in the afternoon

- MIRT can be used for tests where the unidimensionality assumption is not viable
  - Either by theory
  - Or by empirical findings

- MIRT is an extension of CFA with multiple factors
  - Just for categorical data

Up Next…

- Model-based Diagnostic Assessment

- Conclusions, Discussion, Questions