PEARSON

An Introduction to Multidimensional Item Response Theory

Psychometrics conf: 2012

Ou Zhang

Preview and Traditional Unidimensinoal IRT

There are two fundamental assumptions of unidimensional IRT models:

- Local independence
  - The probability of responding to an item is statistically independent of responding to any other item while conditioned on ability within test (Hambleton & Swaminathan, 1985).
- Unidimensionality
  - There is only a single ability underlying the difference in person responses to items (Embretson & Reise, 2000).

PEARSON

What is Multidimensionality?

- Although unidimensional IRT has been widely used in the educational measurement, research shows that the unidimensional assumptions are often difficult to meet in real world contexts (Ackerman, 1994; Reckase, 1985).

- In the real world situation, test items may require <u>more than one ability/trait</u> or hypothetical construct to solve it.

PEARSON

Example of Multidimensional Situation

The table below gives information about the planets: their periods of revolution about the Sun and rotation about their axes.

Planet	Mean distance from the Sun (million kilometers)	Period of revolution (Earth time)	Period of rotation (Earth time)
Mercury	58	88 days	59 days
Venus	108	225 days	243 days
Earth	150	365 days	23.9 hs
Mars	228	687 days	24.6 hs

Which planet has the longest year in Earth time?

A. Mercury

B. Venus

C. Earth

D. Mars

Example adapted from (Reckase, 2009; p. 76)

How many skills and knowledge that are needed to arrive at the correct answer?

- Reading skill
- Math skill
- Astronomy knowledge

PEARSON

Why do we care about multidimensionality in measurement?

- It's important to measure what you intend to measure.
  - If the assumptions of UIRT violated, item parameter estimates will be biased, and the standard errors associated with ability estimates will be too small.
- Test fairness—Multidimensionality can cause DIF, Item bias, etc.
- Some tests have designs that lead one to expect multidimensionality.
- Some tests are modelled as unidimensional, but results are reported as subscore or composite score.
- Multiple dimensions may be useful diagnostically.

PEARSON

Multidimensional Item Response Theory Model (MIRT)

In general, MIRT models are developed and classified as two types:

- Compensatory Models and
- Partially Compensatory (or non-compensatory) Models

The probabilities of a person's item response are determined differently among the ability dimensions in these two types of models (Min, 2003).

PEARSON

Multidimensional Item Response Theory Model

- Compensatory MIRT model (McKinley & Reckase, 1983)

$$P(x_{ij}=1\,|\,\boldsymbol{\theta}_j,\mathbf{a}_\mathbf{i},d_i)=\frac{e^{D(\mathbf{a'_i}\boldsymbol{\theta}_\mathbf{j}+d_i)}}{1+e^{D(\mathbf{a'_i}\boldsymbol{\theta}_\mathbf{j}+d_i)}}$$

$\boldsymbol{\theta}_s$ represents **multiple** ability parameters associated with each respondent,
$\mathbf{a}_i$ represents **multiple** discrimination parameters associated with each item,
and d_i represents an item's location on an item response **surface**.

In a compensatory MIRT model, a respondent with high amount of one dimension can compensate for low amounts in another. This is reflected by the additive nature of the model.

Example of a 2-dimension 2PL compensatory MIRT model

$$P_{ij} = \frac{1.0}{1.0 + e^{-1.7(a_{1i}\theta_{1j} + a_{2i}\theta_{2j} + d_i)}}$$

Discrimination Parameters

Latent ability composite

Difficulty Parameter

Multidimensional Item Response Theory Model (cont.)

- Noncompensatory MIRT model (Sympson, 1978)

$$P(x_{ij}=1|\boldsymbol{\theta}_j,\mathbf{a_i},\mathbf{b_i})=\prod_{\ell=1}^{m}\frac{e^{Da_{i\ell}(\theta_{j\ell}-b_{i\ell})}}{1+e^{Da_{i\ell}(\theta_{j\ell}-b_{i\ell})}}$$

- In noncompensatory MIRT model, probability of a correct response is largely governed by respondent's lowest ability dimension. This is reflected by the multiplicative nature of the model.

- Being high on one dimension doesn't compensate for being low on another.

PEARSON

Example of a 2-dimension 2PL non-compensatory MIRT model

$$P_{ij} = \left[\frac{1.0}{1.0 + e^{-1.7(a_{1i}\theta_{1j} - b_{1i})}} \right] \left[\frac{1.0}{1.0 + e^{-1.7(a_{2i}\theta_{2j} - b_{2i})}} \right]$$

This model is essentially the product of two unidimensional 2PL models

The Generalized Model

Let $f_1 = a_1(\theta_1 - b_1)$ and $f_2 = a_2(\theta_2 - b_2)$

Then the compensatory model can be written as:

$$P_c = \frac{e^{(f_1 - f_2)}}{1 + e^{(f_1 - f_2)}}$$

and the noncompensatory model can be written as:

$$P_{nc} = \frac{e^{(f_1 - f_2)}}{[1 + e^{(f_1 - f_2)}][e^{f_1} + e^{f_2}]}$$

PEARSON

The Generalized Model (cont.)

The generalized model can then be expressed as:

$$P_g = \frac{e^{(f_1 - f_2)}}{[1 + e^{(f_1 - f_2)}] + \mu\,[e^{f_1} + e^{f_2}]}$$

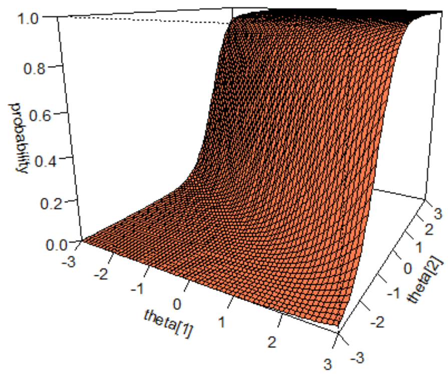
Where, if
$\mu = 0$, P_g is the compensatory model,
and if
$\mu = 1$, P_g is the noncompensatory model

PEARSON

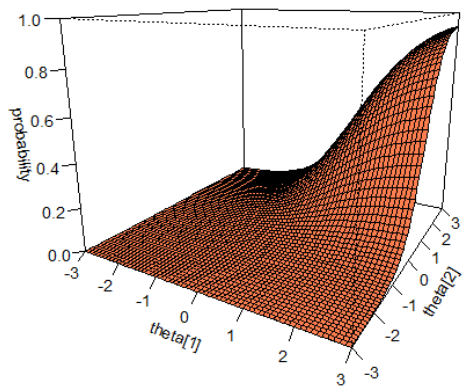
Graphical Presentation of MIRT Model- Item Response Surface

Only 2-dimension MIRT models can be graphically represented

•(Compensatory vs non-compensatory)



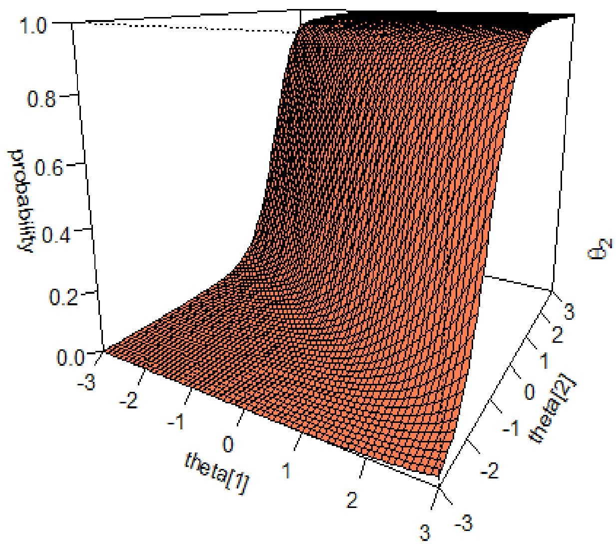
$$a_1 = .8, a_2 = 1.4, d = -.3$$

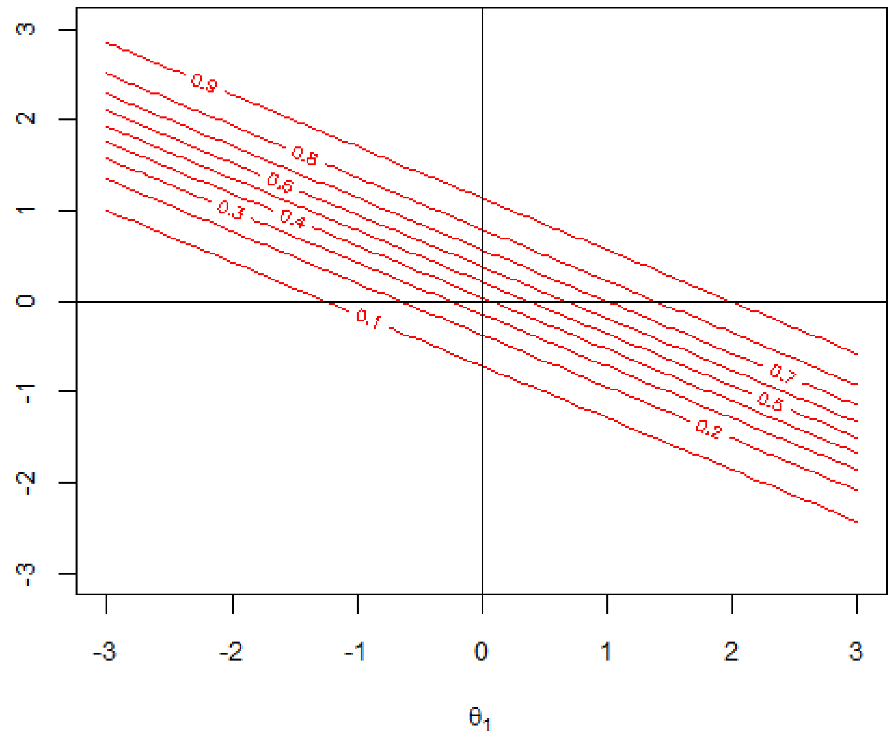


$$a_1 = .8, a_2 = 1.2, b_1 = .7, b_2 = .8$$

PEARSON

Graphical Presentation of MIRT Model-Equiprobable Contour Plot





$$a_1 = .8, a_2 = 1.4, d = -.3$$

Three Key Characteristics of MIRT Model and their Characteristics in the Item Arrow Plot

Discrimination, Difficulty and Direction of the Item for the Best Measurement

$$MDISC_i = \sqrt{\sum_{k=1}^{m} a_{ki}^2}$$

Total discrimination power for an item.

$$MDIFF_i = \frac{-d_i}{MDISC_i}$$

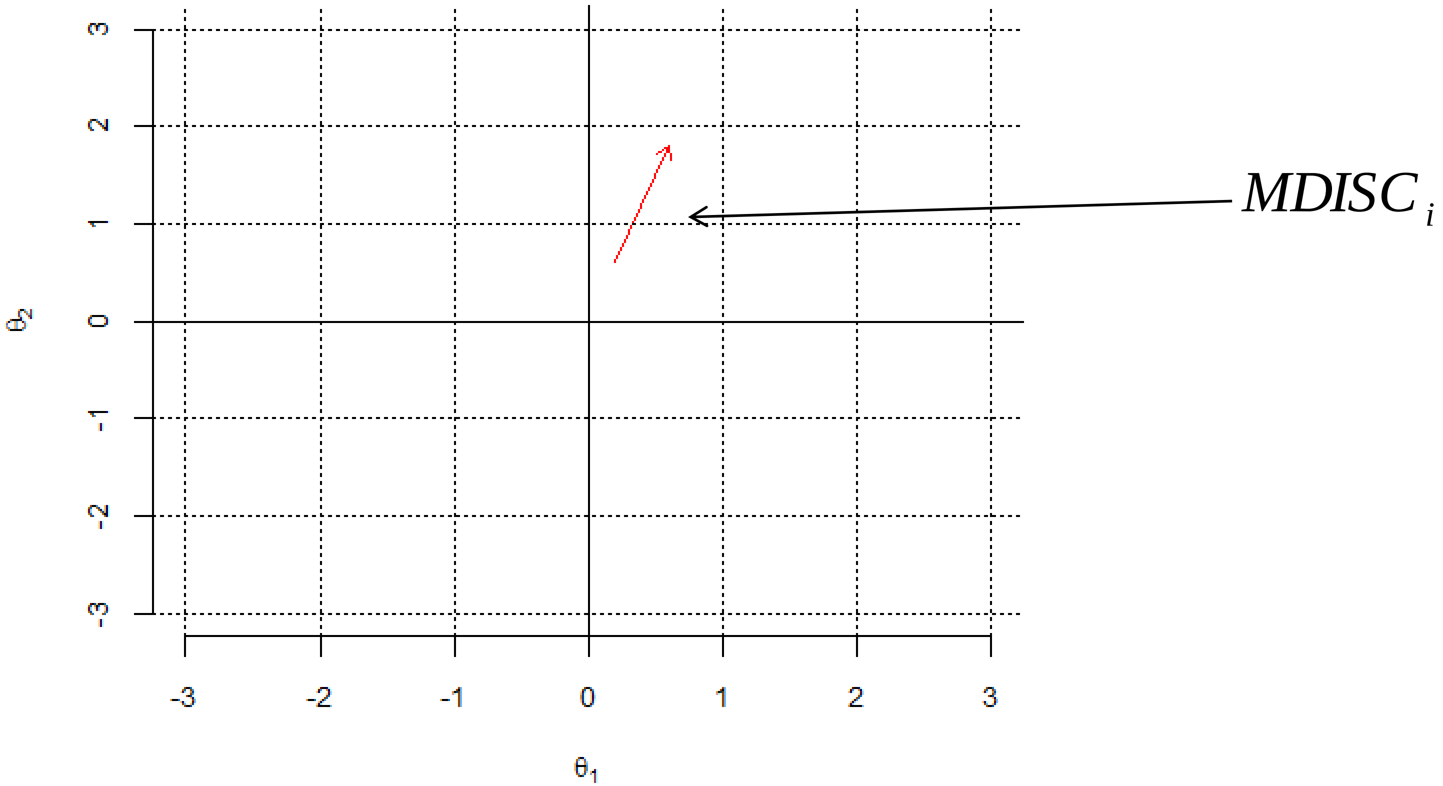
Difficulty of an item. Represents the location of item in dimensional space. Positive = harder, negative = easier. Magnitude reflects distance from origin necessary for 50% probability of correct response.

$$\alpha_{ik} = \arccos\left(\frac{a_{ik}}{MDISC_i}\right)$$

α_{ik} is the direction of the best discrimination in the dimensional space and the angle from the kth dimension

PEARSON

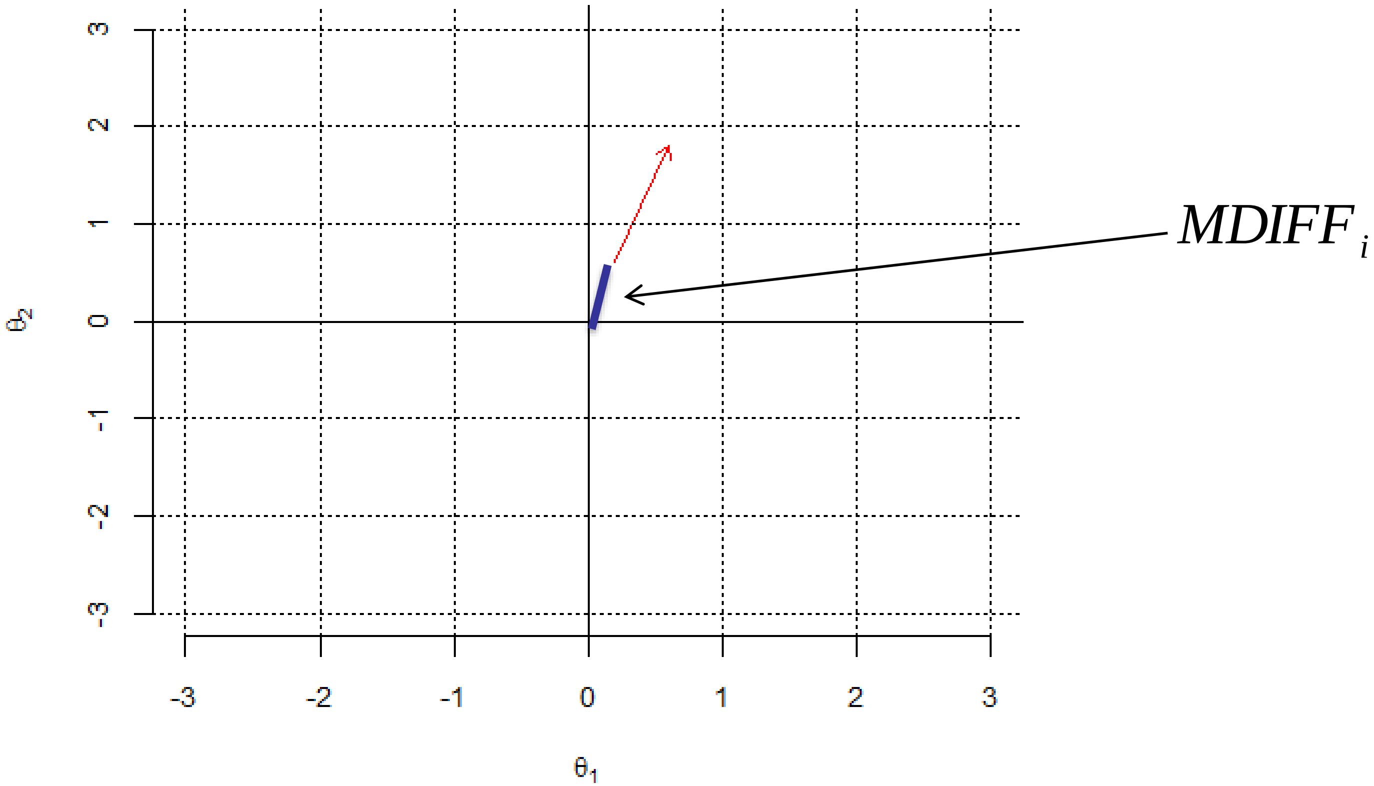
Graphical Presentation of MIRT Model-Item Arrow Plot



$$a_1 = .4,\ a_2 = 1.2,\ d = -.8$$

PEARSON

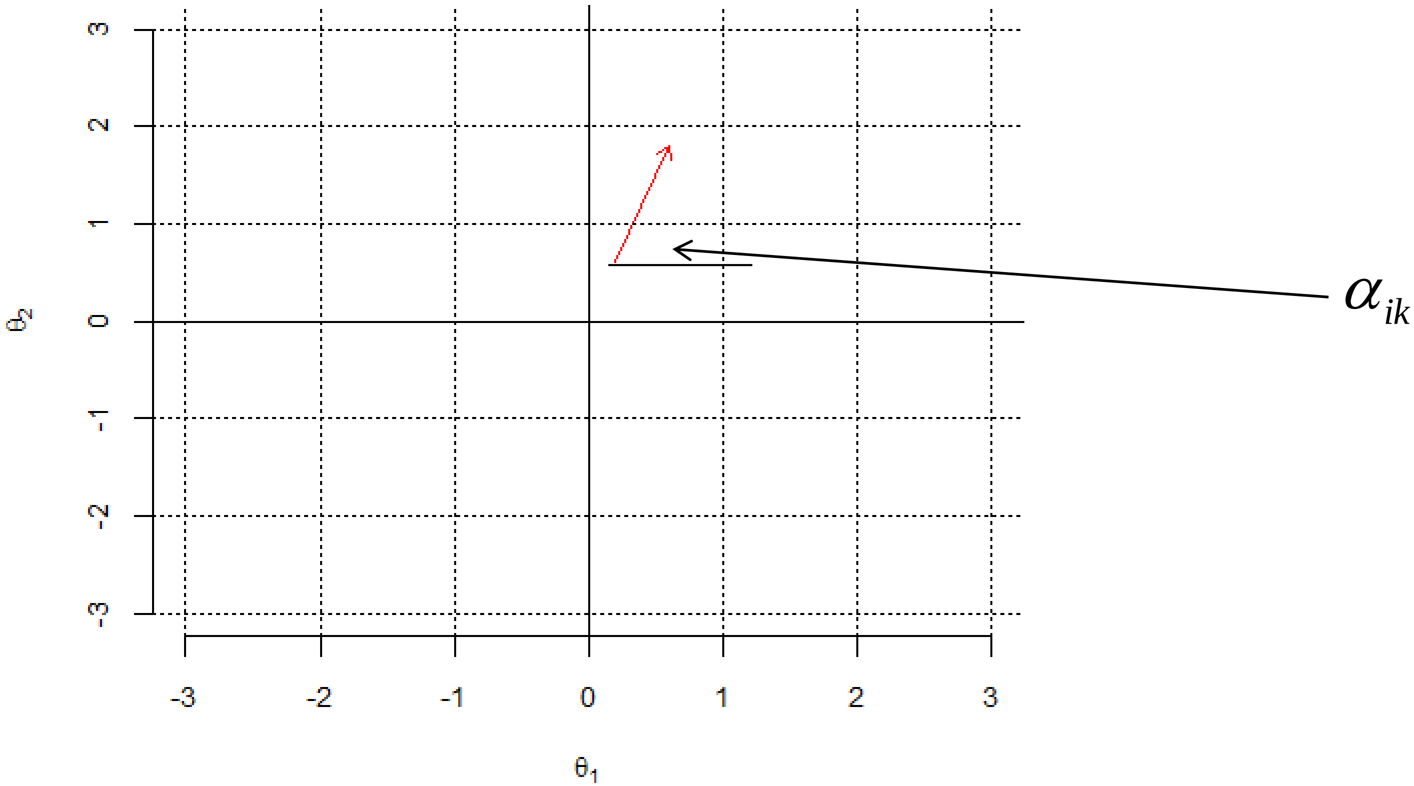
Graphical Presentation of MIRT Model-Item Arrow Plot



$$a_1 = .4, a_2 = 1.2, d = -.8$$

PEARSON

Graphical Presentation of MIRT Model-Item Arrow Plot

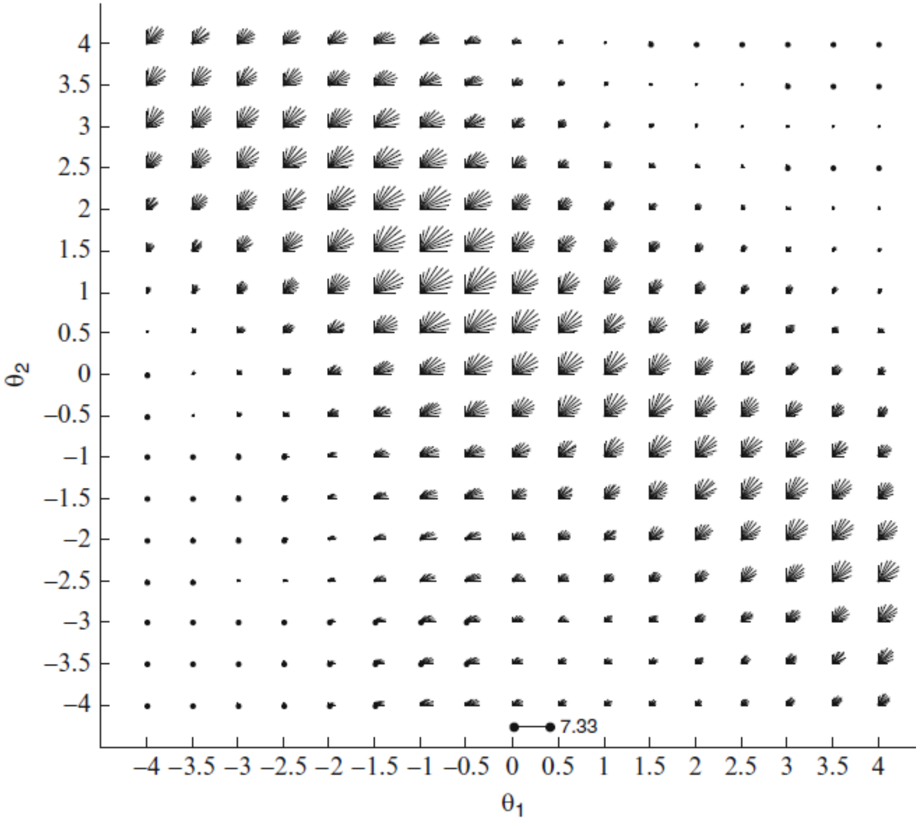


$$a_1 = .4,\ a_2 = 1.2,\ d = -.8$$

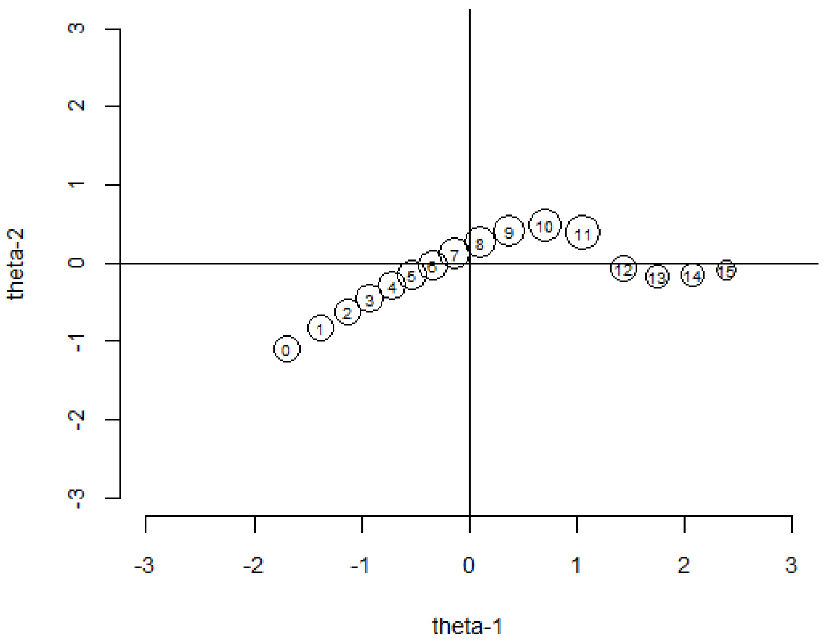
PEARSON

Graphical Presentations of Multidimensional Tests

- Clam Shell Plot



Centroid Plot



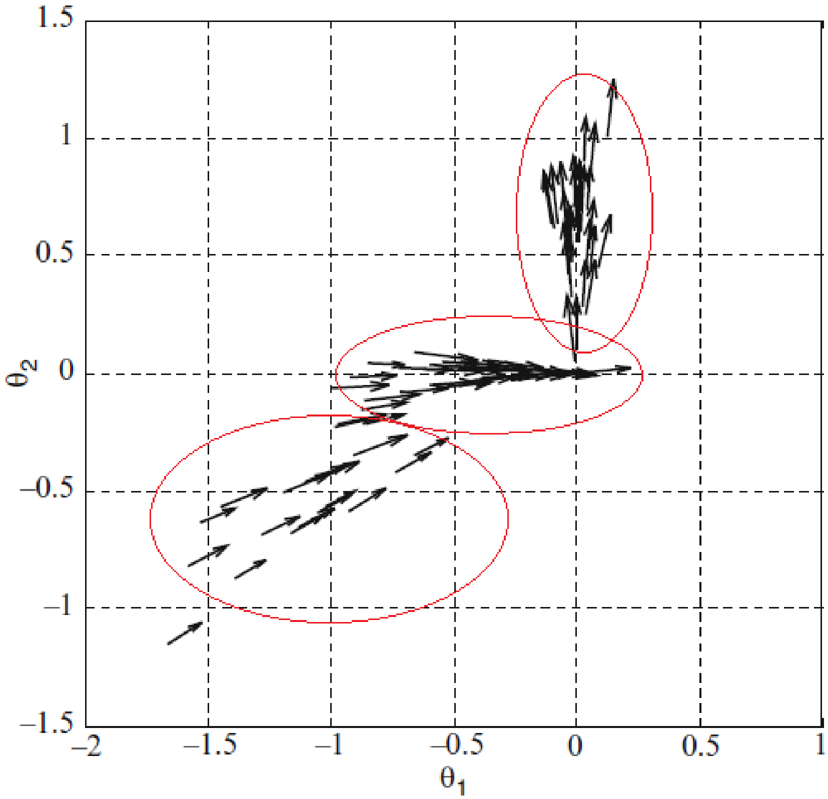
Determining Dimensionality

- Nonparametric Method:
  - DIMTEST (Stout, 1999)
    - Determine Unidimension or Multidimension?
  - DETECT (Zhang & Stout, 1999)
    - Determine number of dimensions needed to represent the relationships in the item-score matrix.
- Nonlinear Factor Analysis (Parallel Analysis)
  - EFA (Eigenvalue > 1.0) or Scree Plot
  - Chi-square test in fit of models with m and m+1 dimensions.
- Principal Component Analysis
  - Eigenvalue > 1.0

PEARSON

Determining Item Structure (cont.)

- Hierarchical Agglomerative Cluster Analysis (Roussos, Stout, and Marden, 1998)



PEARSON

Computer Programs for Estimating MIRT Parameters

- Dimensionality Check
  - DIMPACK 1.0 (DIMTEST, DETECT, HCA/CCPROX)
- TESTFACT (Bock, Gibons, Schilling, Muraki, Wilson, & Wood, 2003)
  - MML
  - Tetrachorical Correlation
- NOHARM (Fraser, 1998)
  - Weighted Least Square
- ConQuest (Wu, Adams, & Wilson, 1997)
  - Rasch Feature and Multidimensional Random Coefficient Multinomial Logit Model (MRCML)
- BMIRT (Bayesian Multivariate Item Response Theory, Yao, 2003)
- IRTPRO

PEARSON

Computer Programs for Estimating MIRT Parameters (cont.)

- SAS PROC NLMIXED
- GLLAMM (within Stata)
- “mirt” R-Package (Phil Chalmers, 2012)

PEARSON

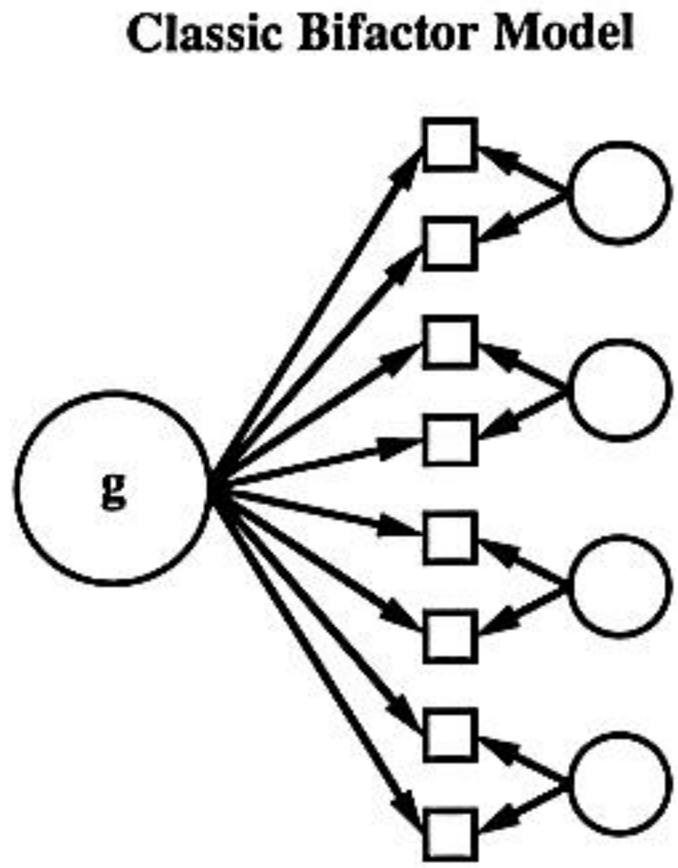
Special MIRT Model:Bi-Factor Model

- In the bi-factor model, each item response is a function of the primary trait and one of the secondary traits.

- The secondary traits are orthogonal to the primary trait and to each other.

- Among all bi-factor cases, testlet is the most popular special case of the bi-factor model.

- Test items are often grouped into clusters, or testlets, centered around a <u>common stimulus</u>.



PEARSON

Bi-Factor Model: Testlet Example

- Example:
- A 40-year-old male presents with the sudden onset of a severe headache localizing toward the occiput and neck. There is an associated defect in vision, along with unilateral numbness and weakness. His temperature is 37.5ºC (99.5ºF). On physical examination the neck is stiff when bending forward, and a Kernig sign is present.
- **Q1: The most appropriate initial step is**

  (A) cerebral arteriography
  (B) complete skull radiographs
  (C) CT scan of the head
  (D) electroencephalography
  (E) lumbar puncture

- **Q2: The most likely diagnosis is**

  (A) cerebral aneurysm with hemorrhage
  (B) conversion reaction
  (C) occipital brain tumor
  (D) subdural hematoma
  (E) vertebrobasilar insufficiency

Stimulus

Testlet items

Key References

- Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, *4*, 255-278.
- Ackerman, T. A. (1996). *Graphical representation of multidimensional item response theory analyses.* Applied Psychological Measurement 20, 311–329.
- Fraser, C., & McDonald, R. P. (1988). NOHARM: Least squares item factor analysis. *Multivariate Behavioral Research*, *23*, 267-269.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.
- Roussos, L., Stout, W., & Marden, J. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement*, *35*, 1–30.

PEARSON

Thank you!

PEARSON