



What have we learned thus far...

Victor Duran Undergraduate Psychology Student Universidade Federal da Bahia, Brazil Supervisor: Prof. Dr. Igor Menezes

First of all

- This presentation is largely based on:
 - M.D. RECKASE (2009) Multidimensional Item Response Theory: Statistics for Social and Behavioral Sciences. New York, NY: Springer.

Summary

- Introduction to IRT
 - Advantages
 - Limitations
- MIRT
 - Solutions
 - Empirical data
 - Models
 - Compensatory

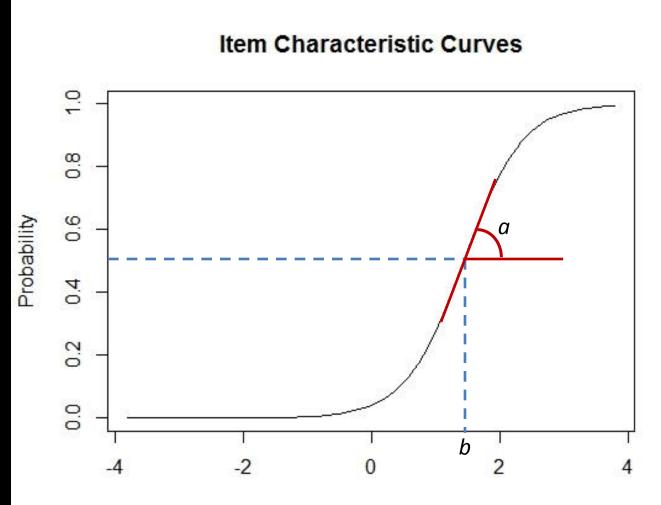
- Partially compensatory
- Item Information
 Surface.
- Advantages of using MIRT

- Various advantages:
 - Sample independent parameters and estimates
 - Increased precision in estimates
 - Standard metric of estimates
 - Between subjects
 - Between tests
 - Simpler composition of tests for specific audiences
 - Item banks

Computerised Adaptive Testing

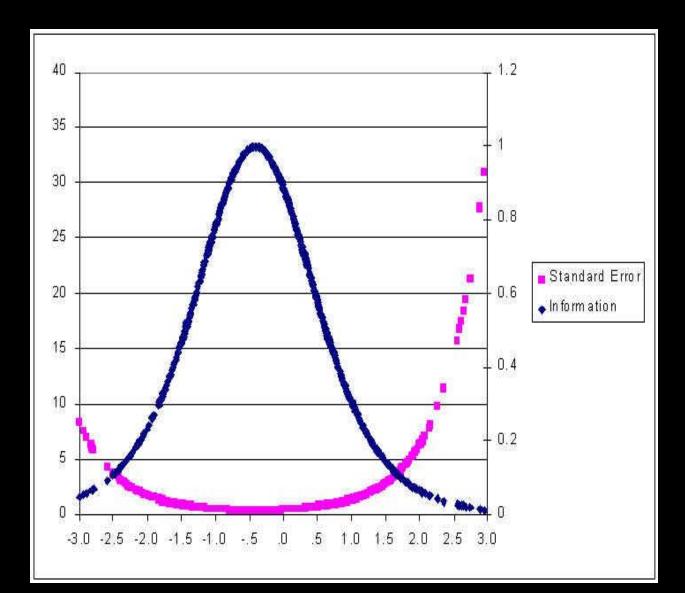
- Advancements from IRT models
 - More information on item level
 - Proportion correct (p value): a single value that describes the item.
 - Item difficulty: summarized as *b* at 50% probability of scoring correctly, but provides information throughout the ability spectrum.

- Advancements from IRT models
 - More information on item level
 - **Discrimination in CTT :** biserial (or point-biserial) correlation between item score and total score.
 - Item discrimination: slope of the tangent line to ICC summarized as *a* at its maximum point.
 - Obtainable at any θ value.



Ability

- Advancements from IRT models
 - More information on item level
 - Standard Error of Measurement (CTT): a single measure for the whole test.
 - Item and Test Information Function: inversely related to error, again a function of θ .

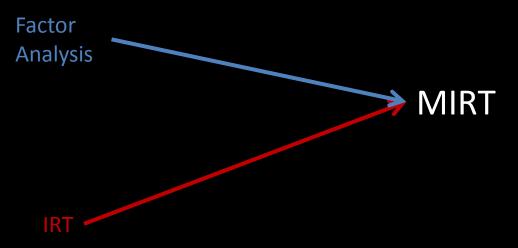


- Advancements from IRT models
 - More information on item level
 - Estimates for a range of ability traits (theta)
 - Difficulty (p-value vs b)
 - Discrimination (bisserial correlation vs *a*)
 - Information (SEM vs IIF)
 - Consequence: sample independence

- Unidimensionality assumption
 - Sample independence is only true (and relevant) if the assessed trait explains enough response variance.
 - What to do with inherently multidimensional constructs?
 - Personality
 - Executive Functions
 - Intelligence

- What to do with inherently multidimensional constructs?
 - Separate the test into **unidimensional subtests**?
 - Issues:
 - Each subtest has to be considered valid on its own, which results in very long and tiresome instruments.
 - What to do with items that load in more than one dimension?
 - Thats when MIRT comes in!

- Historical background
 - Earliest MIRT models date from 1970s
 - Reckase (1972), Mulaik (1972), Sympson (1978) and Whitely (1980).



- Construct dimensionality
- Item loadings
- Item parameters
- Probability of scores as a result of the interaction between items and person abilities

- Example (Reckase, 2009, p.80)
 - Math items with two factors:
 - Arithmetic problem solving (θ_1)
 - Algebraic symbol manipulation (θ_2)
- 1. A survey asked a sample of people which of two products they preferred. 50% of the people said they preferred Product A best, 30% said they preferred Product B, and 20% were undecided. If 1,000 people preferred Product A, how many people were undecided?
 - A. 200
 - **B.** 400
 - C. 800
 - D. 1,200
 - E. 2,000

• Observed data:

 Table 1. Proportions of correct responses to Item 1 for 4,114 participants (Reckase, 2009, p.81)

Midpoints in θ_2									
Midpoints in $oldsymbol{ heta}_1$	-1.75	-1.25	-0.75	-0.25	0.25	0.75	1.25	1.75	
-1.25	0.20		0.09						
-0.75	0.06	0.18	0.39	0.47	0.19	0.67			
-0.25	0.18	0.25	0.30	0.45	0.54	0.50	0.61	0.82	
0.25	0.19	0.40	0.39	0.53	0.45	0.46	0.77	0.57	
0.75	0.24	0.34	0.49	0.53	0.50	0.65	0.76	0.71	
1.25	0.30	0.35	0.54	0.55	0.47	0.63	0.78	0.55	
1.75	0.51	0.55	0.57	0.62	0.60	0.71	0.71	0.65	

• Observed data:

 Table 1. Proportions of correct responses to Item 1 for 4,114 participants (Reckase, 2009, p.81)

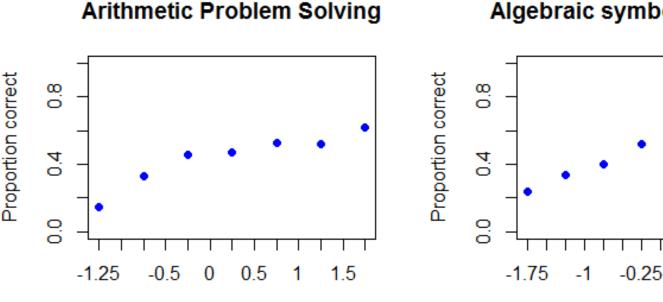
	Midpoints in θ_2								
Midpoints in $oldsymbol{ heta}_1$	-1.75	-1.25	-0.75	-0.25	0.25	0.75	1.25	1.75	
-1.25	0.20		0.09						
-0.75	0.06	0.18	0.39	0.47	0.19	0.67			
-0.25	0.18	0.25	0.30	0.45	0.54	0.50	0.61	0.82	
0.25	0.19	0.40	0.39	0.53	0.45	0.46	0.77	0.57	
0.75	0.24	0.34	0.49	0.53	0.50	0.65	0.76	0.71	
1.25	0.30	0.35	0.54	0.55	0.47	0.63	0.78	0.55	
1.75	0.51	0.55	0.57	0.62	0.60	0.71	0.71	0.65	
Proportions in θ_2	0.24	0.34	0.40	0.52	0.46	0.60	0.73	0.66	

• Observed data:

 Table 1. Proportions of correct responses to Item 1 for 4,114 participants (Reckase, 2009, p.81)

	Midp	points	in θ_2						
Midpoints in $oldsymbol{ heta}_1$	-1.75	-1.25	-0.75	-0.25	0.25	0.75	1.25	1.75	Proportions in ${m heta}_1$
-1.25	0.20		0.09						0.15
-0.75	0.06	0.18	0.39	0.47	0.19	0.67			0.33
-0.25	0.18	0.25	0.30	0.45	0.54	0.50	0.61	0.82	0.46
0.25	0.19	0.40	0.39	0.53	0.45	0.46	0.77	0.57	0.47
0.75	0.24	0.34	0.49	0.53	0.50	0.65	0.76	0.71	0.53
1.25	0.30	0.35	0.54	0.55	0.47	0.63	0.78	0.55	0.52
1.75	0.51	0.55	0.57	0.62	0.60	0.71	0.71	0.65	0.62
Proportions in θ_2	0.24	0.34	0.40	0.52	0.46	0.60	0.73	0.66	

If we assumed the data to be Unidimensional \bullet - We could only model θ_1 or θ_2



θ1

Algebraic symbol manipulation

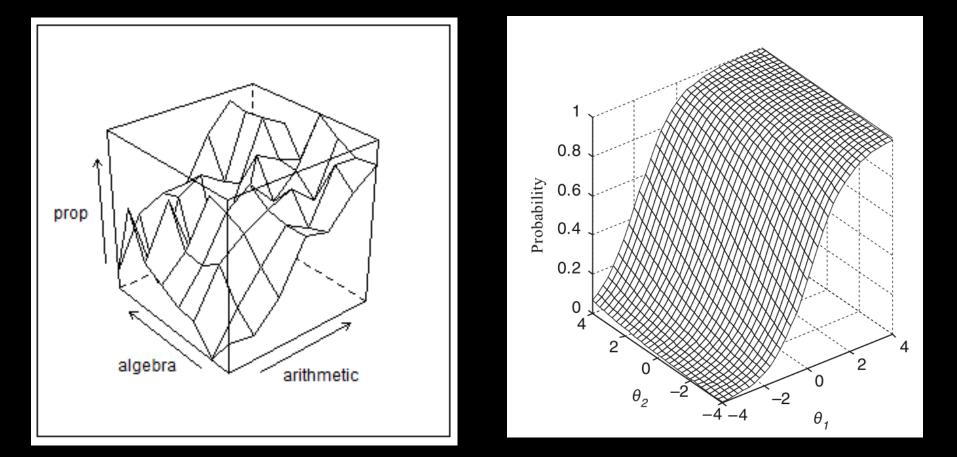
 θ_2

5 0

1.5

- Displays of MIRT models
 - Surfaces
 - Examples in two dimensions only
 - Two coordinate axes are necessary to describe the ability level.
 - θ_1 and θ_2
 - A third axis to represent the proportion of correct responses or the probability $P(1|\theta_{j1}, \theta_{j2})$.

Observed data and Modelled surface



- Assumptions
 - Monotonicity assumption
 - Local Independence assumption
 - The term "local" understood as: in that position of the θ_m space
- Major groups of models
 - Compensatory
 - Non-compensatory (partially compensatory)

- Models
 - Compensatory
 - Dimensions combine linearly to produce probability of scoring (endorsing) the item
 - High scores in a dimension can *compensate* lower scores in other dimensions.
 - By the time of Reckase's book, only compensatory models were defined for polytomous items.
 - Simpler and most common on literature (Reckase, 2009).

Models

- Non-compensatory
 - Also called *partially compensatory* models
 - Each dimension is treated separately and the final estimated probability is the product of the individual probabilities
 - Hence, results are a nonlinear combination of the thetas.
 - A very low probability will never be compensated by a higher ability level on another factor.

• MIRT Models (dichotomous data):

Compensatory models

Multidimensional extension of the 2PLM

$$P(U_{ij} = 1 | \theta_j, a_i, b_i) = \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}}$$

Unidimensional 2PLM

Where, i = 1, 2, ..., #items j = 1, 2, ..., # participants $U_{ij} = score \ of \ person \ j \ on \ item \ i$ $a_i = discrimination \ for \ item \ i$ $b_i = difficulty \ for \ item \ i$ $\theta_i = location \ of \ subject \ j \ on \ \theta$

• MIRT Models (dichotomous data):

Compensatory models

Where,

U_{ij}

Multidimensional extension of the 2PLM

$$P(U_{ij} = 1 | \theta_j, a_i, b_i) = \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}} \quad \text{Unidimensional}$$

$$= 1, 2, \dots, \# \text{ participants}$$

$$= 1, 2, \dots, \# \text{ participants}$$

$$= 3 \text{ correof person j on item } i$$

$$= 3 \text{ discrimination for item } i$$

$$= 1 \text{ location of subject j on } \theta$$

$$P(U_{ij} = 1 | \theta_j, a_i, d_i) = \frac{e^{a_i \theta'_j + d_i}}{1 + e^{a_i \theta'_j + d_i}} \quad \text{Multidimensional}$$

Multidimensional extension of the 2PL

$$P(U_{ij} = 1 | \boldsymbol{\theta}_j, \boldsymbol{a}_i, d_i) = \frac{e^{a_i \theta'_j + d_i}}{1 + e^{a_i \theta'_j + d_i}}$$

Multidimensional extension of the 2PL

$$P(U_{ij} = 1 | \boldsymbol{\theta}_j, \boldsymbol{a}_i d_i) = \frac{e^{\boldsymbol{a}_i \boldsymbol{\theta}_j' + d_i}}{1 + e^{\boldsymbol{a}_i \boldsymbol{\theta}_j' + d_i}}$$

Multidimensional extension of the 2PL

$$P(U_{ij} = 1 | \boldsymbol{\theta}_j, \boldsymbol{a}_i, \boldsymbol{d}_i) = \frac{e^{\boldsymbol{a}_i \boldsymbol{\theta}_j' + \boldsymbol{d}_i}}{1 + e^{\boldsymbol{a}_i \boldsymbol{\theta}_j' + \boldsymbol{d}_i}}$$

Both $\boldsymbol{\theta}$ and \boldsymbol{a} are now $a_i = [a_{i1} a_{i2} \dots a_{im}]$ 1 x m vectors. $a_i = [a_{i1} a_{i2} \dots a_{im}]$ m = # dimensions $\boldsymbol{\theta}_j = [\theta_{j1} \theta_{j2} \dots \theta_{jm}]$

A discrimination statistic (MDISC or A_i) that summarizes the a_i vector is available:

$$A_i = \sqrt{\sum_{k=1}^m a_{ik}^2}.$$

Multidimensional extension of the 2PL

$$P(U_{ij} = 1 | \boldsymbol{\theta}_j, \boldsymbol{a}_i | \boldsymbol{d}_i) = \frac{e}{1+e}$$

This parameter is defined as an intercept, or a location parameter.

Note that in this first generalisation, even though *d* originates from the product of *a* and *b*, it is not a vector but a **scalar**.

Derivation of d: $a(\theta - b)$ $a\theta - ab$ Let, d = -ab $a\theta + d$

Multidimensional extension of the 2PL

$$P(U_{ij} = 1 | \theta_j, a_i, d_i) = \frac{e^{a_i \theta'_j + d_i}}{1 + e^{a_i \theta'_j + d_i}}$$

If - *d* is divided by an element of a_i , we obtain a measure of difficulty associated with that dimension.
A summary "difficulty" (MDIFF or *b*) for the whole
Derivation of *d*:
 $a(\theta - b)$
 $a\theta - ab$
Let,
 $d = -ab$
 $a\theta + d$

A summary "difficulty" (MDIFF or *b*) for the whol item is obtained by:

$$b = \frac{-d}{\sqrt{\mathbf{a}\mathbf{a}'}} = \frac{-d}{\sqrt{\sum_{\nu=1}^{m} a_{\nu}^2}}.$$

Multidimensional extension of the 2PL

$$P(U_{ij} = 1 | \boldsymbol{\theta}_j, \boldsymbol{a}_i d_i) = \frac{e^{\boldsymbol{a}_i \boldsymbol{\theta}_j' + d_i}}{1 + e^{\boldsymbol{a}_i \boldsymbol{\theta}_j' + d_i}}$$

The exponent is a linear combination of discrimination (a_i) and θ values.

$$\mathbf{a_i}\boldsymbol{\theta_j}' + \underline{d_i} = \underline{a_{i1}}\theta_{j1} + \underline{a_{i2}}\theta_{j2} + \dots + \underline{a_{im}}\theta_{jm} + \underline{d_i} = \sum_{\ell=1}^m a_{i\ell}\theta_{j\ell} + d_i.$$
Ist Dimension

The higher the result of the sum, the higher the probability of correct response. Each a_i works as a weight for the total sum. Hence, the more discriminative the item is on a particular dimension, the more influence it should have on the outcome.

Compensatory feature

$$\mathbf{a_i}\boldsymbol{\theta_j}' + d_i = a_{i1}\theta_{j1} + a_{i2}\theta_{j2} + \dots + a_{im}\theta_{jm} + d_i = \sum_{\ell=1}^m a_{i\ell}\theta_{j\ell} + d_i.$$
1st Dimension

- Let the exponent be equal to a given value k

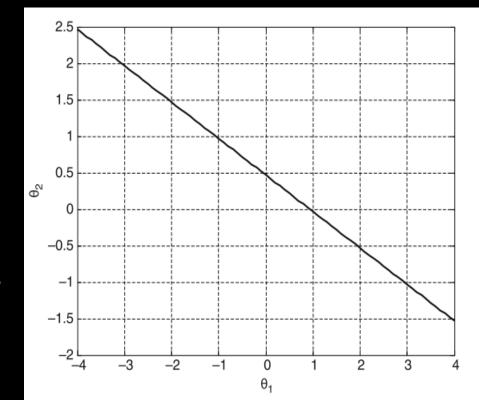
• Then, all θ vectors that satisfy $k = a_i \theta'_j + d_i$ fall in a straight line.

- Compensatory feature
 - $k = a_i \theta'_j + d$
 - For example, let k = 0
 - Let item *i* be defined in two dimensions and have parameters $a_i = [.75 \ 1.5]$ and $d_i = -.7$
 - The exponent then becomes:
 - $k = .75\theta_1 + 1.5\theta_2 .7 = 0$

- Compensatory feature
 - Solving for θ_2

•
$$\theta_2 = -.5\theta_1 + \frac{.7}{1.5}$$

- Plot of theta vectors that yield exponents of k = 0 for a test item with parameters $a_1 = .75$, $a_2 = 1.5$ and d = -.7



Compensatory feature

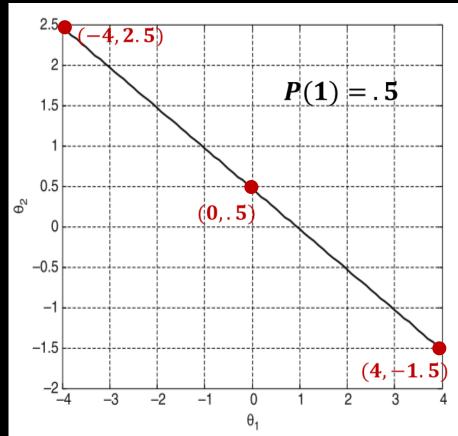
•
$$P(1) = \frac{e^0}{1+e^0} = \frac{1}{2} = 0.5$$

$$\bullet \theta = [0.5]$$

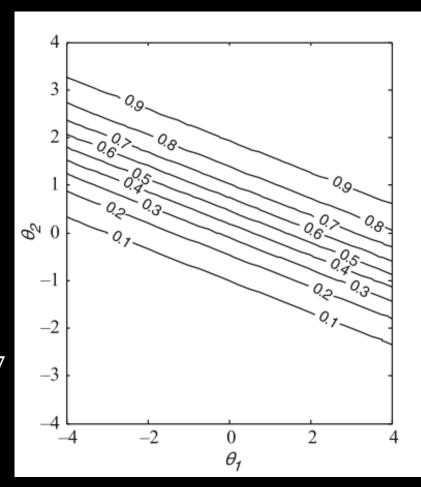
•
$$\theta = [-4 \ 2.5]$$

• $\theta = [4 - 1.5]$

 Different combinations of \(\theta_1\) and \(\theta_2\) yield the same probability of correct response



- Compensatory Feature
 - The same can be repeated for several different probability lines
 - Probability contours
 - This item has parameters $a_1 = .5$, $a_2 = 1.5$ and d = -.7

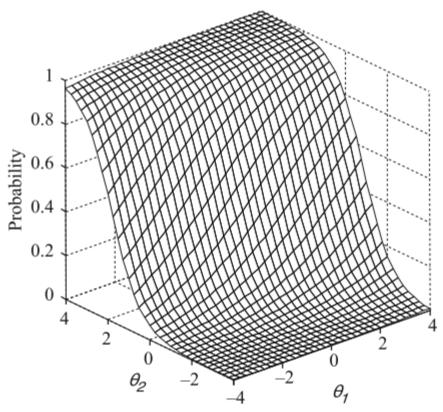


Item Response Surface

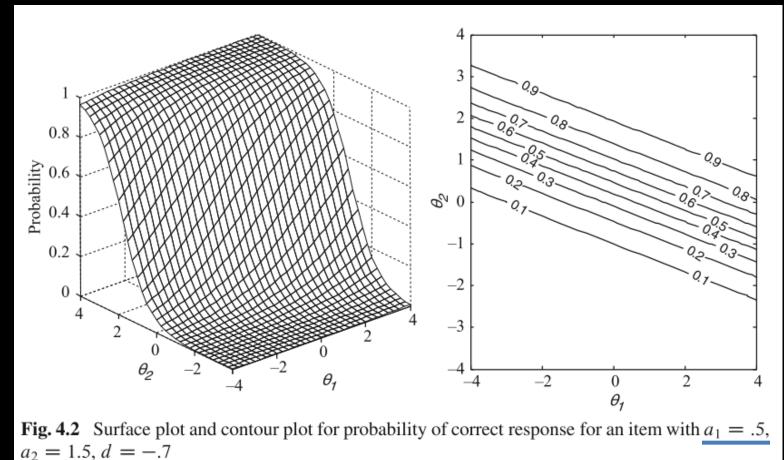
– IRS

And now for all possible probability values.

- Item parameters: $a_1 = .5$, $a_2 = 1.5$ and d = -.7



• Both plots:



• A different example. What has changed?

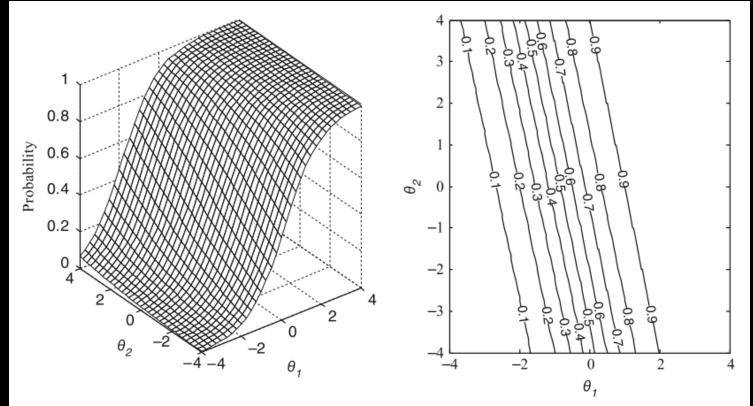


Fig. 4.3 Surface plot and contour plot for the probability of correct response for an item with $a_1 = 1.2, a_2 = .3, d = 1$

- A quick glance at the M3PLM
 - The lower asymptote

$$-P(U_{ij}=1|\boldsymbol{\theta}_{j},\boldsymbol{a}_{i},c_{i},d_{i}) \neq c_{i}+(1-c_{i})\frac{e^{a_{i}\boldsymbol{\theta}_{j}^{\prime}+d_{i}}}{1+e^{a_{i}\boldsymbol{\theta}_{j}^{\prime}+d_{i}}}$$

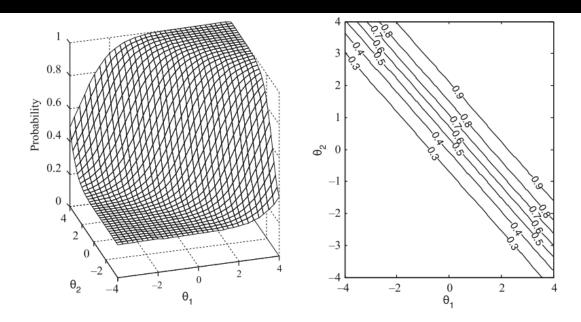


Fig. 4.5 Surface plot and contour plot for probability of correct response for an item with $a_1 = 1.3$, $a_2 = 1.4$, d = -1, c = .2

- Non-compensatory models
 - The rationale behind:
 - Suppose the introductory item measured, instead:
 - Arithmetic problem solving (θ_1)
 - Reading skill (θ_2)
 - A person with very low reading skills attempts the example item. Even with extremely high mathematic skills, an individual would not obtain success on the item.

Non-compensatory models

– Mathematical model:

$$P(U_{ij} = 1 | \boldsymbol{\theta_j}, \mathbf{a_i}, \mathbf{b_i}, c_i) = c_i + (1 - c_i) \left(\prod_{\ell=1}^{m} \frac{e^{1.7a_{i\ell}(\theta_{j\ell} - b_{i\ell})}}{1 + e^{1.7a_{i\ell}(\theta_{j\ell} - b_{i\ell})}} \right)$$

– Sympson (1978).

Non-compensatory models

– Mathematical model:

$$P(U_{ij} = 1 | \boldsymbol{\theta}_{\mathbf{j}}, \mathbf{a}_{\mathbf{i}} (\mathbf{b}_{\mathbf{j}}, c_i) = c_i + (1 - c_i) \left(\prod_{\ell=1}^{m} \frac{e^{1.7a} (\theta_{\ell} - b_{\ell})}{1 + e^{1.7a} (\theta_{\ell} - b_{\ell})} \right)$$

- The b_i also becomes a $1 \ x \ m$ vector
- Each unidimensional probability is calculated and their overall product is the estimated probability of scoring the item.
- Model was originally devised with a lower asymptote

- Non-compensatory models
 - Probability contours
 - For the case where c = 0 Consider this simpler model:

$$k = \prod_{\ell=1}^{m} p_{\ell}$$

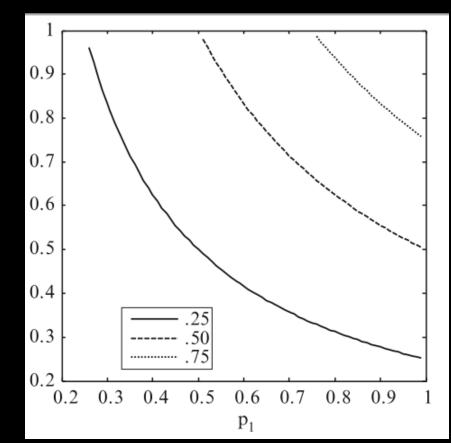
- p_{ℓ} is the probability of scoring correctly in each dimension (ℓ).
- For the simpler two dimensional case: $k = p_1 p_2$
 - This yields the following plot:

Non-compensatory models

- Probability contours (for k = .25, .50 and .75)

Mathematicaly, these are hyperbolas

Not yet a function of θ_j



- Non-compensatory models
 - Probability contours (for k = .25, .50 and .75)

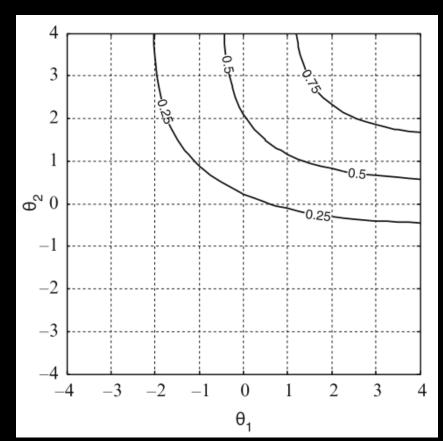
Now as a function of θ_i

Interesting feature: the .5 probability curve asymptotes to the values of b_i

If, say, $\theta_1 = b_{i1}$ then the unidimensional $P(1|\theta_1, b_i) = .5$.

Since you are multiplying probabilities [0,1], this becomes the highest possible probability value, for a subject with this specific θ_1 ,

$$c_i = 0, \quad a_{i1} = .7, \quad a_{i2} = 1.1, \\ b_{i1} = -.5 \text{ and } b_{i2} = .5$$



- Non-compensatory models
 - Probability contours (for k = .25, .50 and .75)

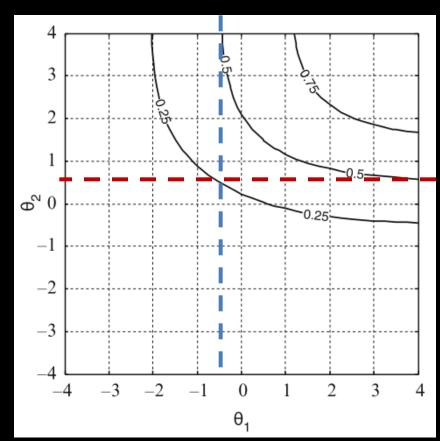
Now as a function of θ_i

Interesting feature: the .5 probability curve asymptotes to the values of b_i

If, say, $\theta_1 = b_{i1}$ then the unidimensional $P(1|\theta_1, b_i) = .5$.

Since you are multiplying probabilities [0,1], this becomes the highest possible probability value, for a subject with this specific θ_1 ,

$$c_i = 0, \quad a_{i1} = 7, \quad a_{i2} = 1, 1, \\ b_{i1} = -.5 \text{ and } b_{i2} = .5$$



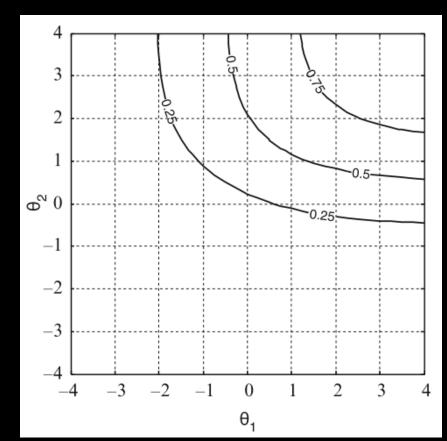
- Non-compensatory models
 - Probability contours (for k = .25, .50 and .75)

Now as a function of θ_i

Interesting feature: the .5 probability curve asymptotes to the values of b_i

The probability of correct response for an item that follows this model can never be greater than the probability for the component with the lowest probability

$$c_i = 0$$
, $a_{i1} = .7$, $a_{i2} = 1.1$,
 $b_{i1} = -.5$ and $b_{i2} = .5$



- Non-compensatory models
 - Item Response Surface (IRS):

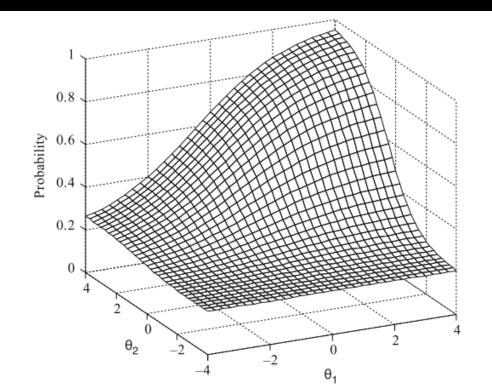


Fig. 4.9 Item response surface for the partially compensatory model when $a_1 = .7$, $a_2 = 1.1$, $b_1 = -.5$, $b_2 = .5$, and c = .2

- Non-compensatory models
 - Item parameters as number of dimensions increase.
 - In UIRT: if $\theta_j = b_i$, then P(1) = .5
 - In Compensatory MIRT: if θ_j is the 0-vector and d = 0, then P(1) = .5
 - For Partially Compensatory models this is not true
 - For m = 2, if $\boldsymbol{\theta}_{i} = \boldsymbol{b}_{i}$, then P(1) = .25
 - For m = 3, if $\theta_{i} = b_{i}$, then P(1) = .125
 - For any m, if $\boldsymbol{\theta}_{j} = \boldsymbol{b}_{i}$, then $P(1) = .5^{m}$

- Non-compensatory models
 - Item parameters as number of dimensions increase.
 - P(1) = .5 for the case where $\theta_j = 0$, all $a_i = .588$ (because of *D* constant) and all b_i s are equal.

Number of dimensions	<i>b</i> -parameter
1	0
2	88
3	-1.35
4	-1.66
5	-1.91
6	-2.10

- When to use each type of models?
 - Ultimately, the fit to the data will define
 - For positively correlated dimensions expect little difference.
 - According to Reckase (2009) there are few studies that compare fit from both models.

- Multidimensional Item Information Function:
 - From UIRT: item information relates to the slope parameter.
 - The bigger the slope the higher the information function.
 - The same happens in MIRT
 - Issue: at each point of the IRS a different slope exists depending on the chosen direction.

- Multidimensional Item Information Function:
 - One solution: the Item Information Surface is displayed across different angles.

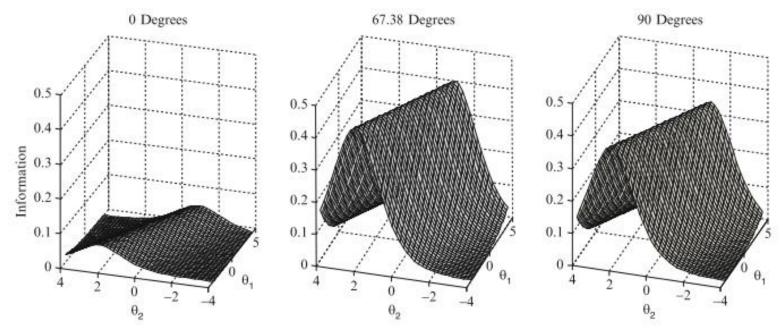


Fig. 5.6 Information surfaces for a M2PL test item with $a_1 = .5$, $a_2 = 1.2$, and d = -.6 in three directions

- Multidimensional Item Information Function:
 - A grid of points in the θ space, where the information is depicted in different angles by small intervals.

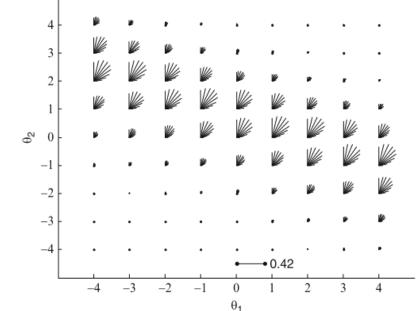


Fig. 5.7 Information for a M2PL test item with $a_1 = .5$, $a_2 = 1.2$, and d = -.6 at equally spaced points in the θ -space for angles from 0° to 90° at 10° intervals

- Applications:
 - Test length reduction with MCAT
 - By allowing a single item to provide information for more than one dimension, test length can be reduced significantly. Specially using MCAT.
 - Differential Item Functioning
 - Identifying the extent to which the underlying element is causing unexpected invariance.
 - Progressess on fields such as abnormal response patterns.
 - An underlying trait could be modelled

Thank You!

• For further questions please e-mail me at: victorduran89@gmail.com

References

- Reckase M., D. (1972). Development and application of a multivariate logistic latent trait model. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.
- Reckase, M., D. (2009). *Multidimensional Item Response Theory: Statistics for Social and Behavioral Sciences*. New York, NY: Springer.
- Makransky, G., Mortensen, E. L., & Glas, C. a W. (2013). Improving personality facet scores with multidimensional computer adaptive testing: an illustration with the NEO PI-R. *Assessment*, 20(1), 3–13. doi:10.1177/1073191112437756.

References

- Mulaik S., A. (1972). A mathematical investigation of some multidimensional Rasch models for psy- chological tests. Paper presented at the annual meeting of the Psychometric Society, Princeton, NJ.
- Sympson J., B. (1978). Amodel for testingwith multidimensional items. InWeiss DJ (ed) Proceedings of the 1977 Computerized Adaptive Testing Conference, University of Minnesota, Minneapolis
- Whitely S., E. (1980). Multicomponent latent trait models for ability tests. Psychometrika 45: 479–494