Assessment, analysis and interpretation of Patient-Reported Outcomes (PROs)

Day 4

Summer school in Applied Psychometrics

Peterhouse College, Cambridge

12th to 16th September 2011

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

This course is prepared by

Anna Brown, PhD ab936@medschl.cam.ac.uk
Jan Stochl, PhD js883@cam.ac.uk
Tim Croudace, PhD tjc39@cam.ac.uk
(University of Cambridge, department of Psychiatry)
Jan Boehnke, PhD boehnke@uni-trier.de
(University of Trier, Department of Clinical Psychology and Psychotherapy)

The course is funded by the ESRC RDI and hosted by
The Psychometrics Centre

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Anna Brown

11. MEASUREMENT PRECISION

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Measurement precision

- Score precision level must be known in order to produce informed judgements for decision-making purposes
  - IRT family of models provides methods that address precision of measurement conditionally
  - enables clearer decisions to be made for an individual (such as deciding upon significance of any change occurring in scores, for instance in response to treatment)

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Standard error of measurement - CTT

- Reliability is defined as proportion of variance due to true score

$$\rho = \sigma_T^2/\sigma_Y^2 = (\sigma_Y^2 - \sigma_E^2)/\sigma_Y^2$$

$$\sigma_E^2 = \sigma_Y^2 - \sigma_Y^2\rho$$

$$\sigma_E^2 = \sigma_Y^2 (1 - \rho)$$

- Classical formula for SEm

$$\text{SE}m = \text{SD}(y)\sqrt{1-\rho}$$

  – with reliability 0.75, SEm is a half of score's SD
  – Sem is constant for all scores

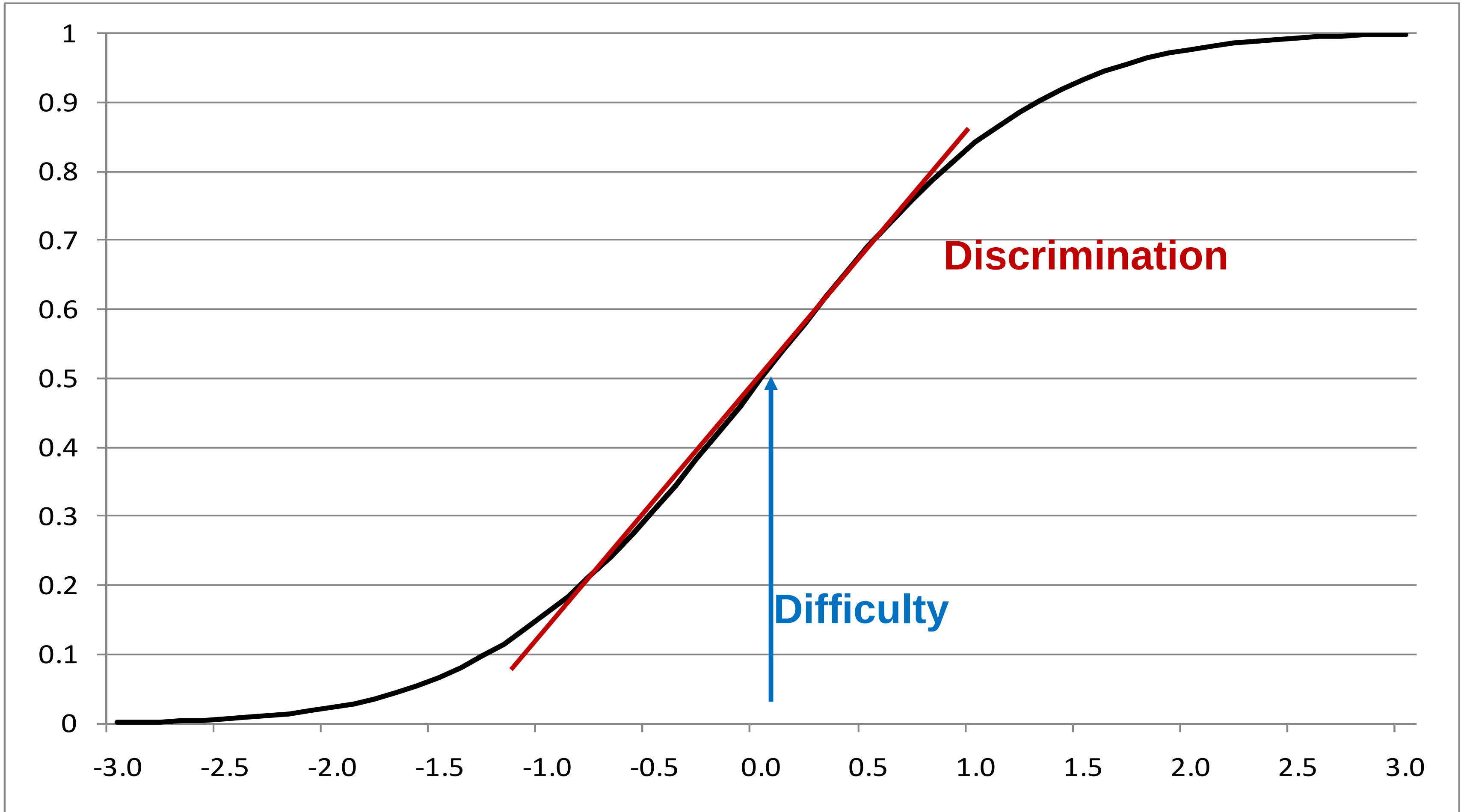
UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Measurement precision in IRT

- Items have different discrimination power depending on the latent score
- Items discriminate better around their difficulty parameter
  - An easy item is useless at discriminating between persons of high standing on the trait (they all will endorse it)
  - A difficult item is useless at discriminating between persons of low standing (they all will not endorse it)
- In contrast with CTT, IRT addresses the issue of varying precision of measurement for different levels of the latent trait

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Example IRF

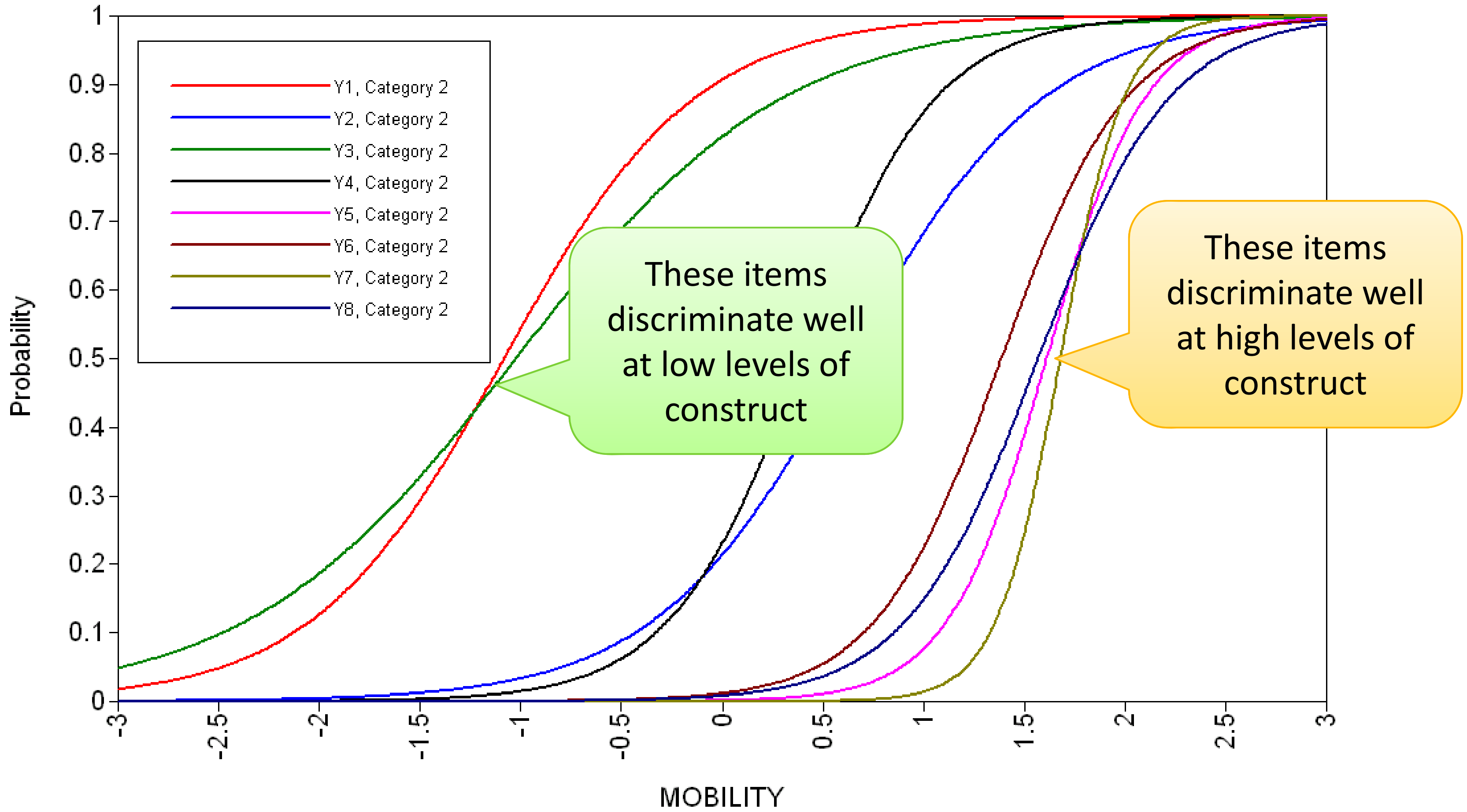


UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Gradient and derivative

- The concept of the gradient of a function z=f(x)
  - change in z corresponding to a an increase in x
  - slope of a local tangent to the curve at each point
  - item discrimination parameter in 2PL model reflects the slope of a tangent at the curve inflection point (item difficulty)
- Derivative f'(x) is a relative change in f(x) when x increases by an infinitely small amount
- In the multidimensional case, we consider directional derivative
  - Typically, we are interested in directions corresponding to measured traits

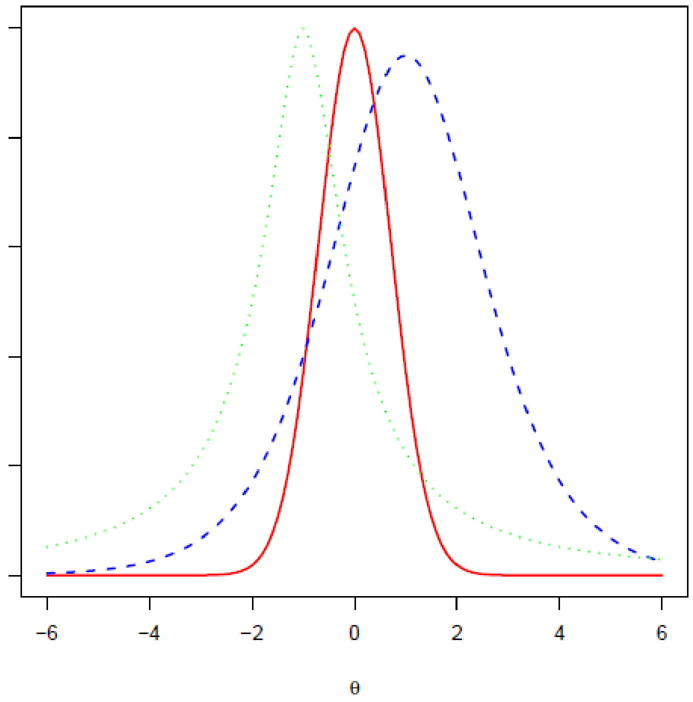
Example IRFs



UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Standard error in IRT

- IRT score is estimated from a likelihood distribution
  - Likelihoods from highly discriminating items are peaked
- Many scores are possible – but not all are equally likely
  - Look at the likelihood distributions for three different persons
- SD of the likelihood distribution is the **SE of the estimate**
- SE will depend on the person's score



UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

SE of the EAP estimator

- EAP provides an easy way to estimate the SD of the likelihood numerically
  - Several points on the grid are taken for finding a likelihood mean
  - For those points, deviations from the mean are then computed to estimate the SE
  - Direct measure of “width” of the likelihood

Fisher information

- How to compute SE for ML and MAP?
- Fisher information provides analytical computation
  - Fisher information describes the “peakiness” of a function
    - equals to an inverse of the second derivative of the **log**likelihood function

$$\mathcal{I}\left(\boldsymbol{\theta}\right) = -\frac{\partial^2 \log l\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}^2}$$

  - The greater the information about a score, the smaller the error

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

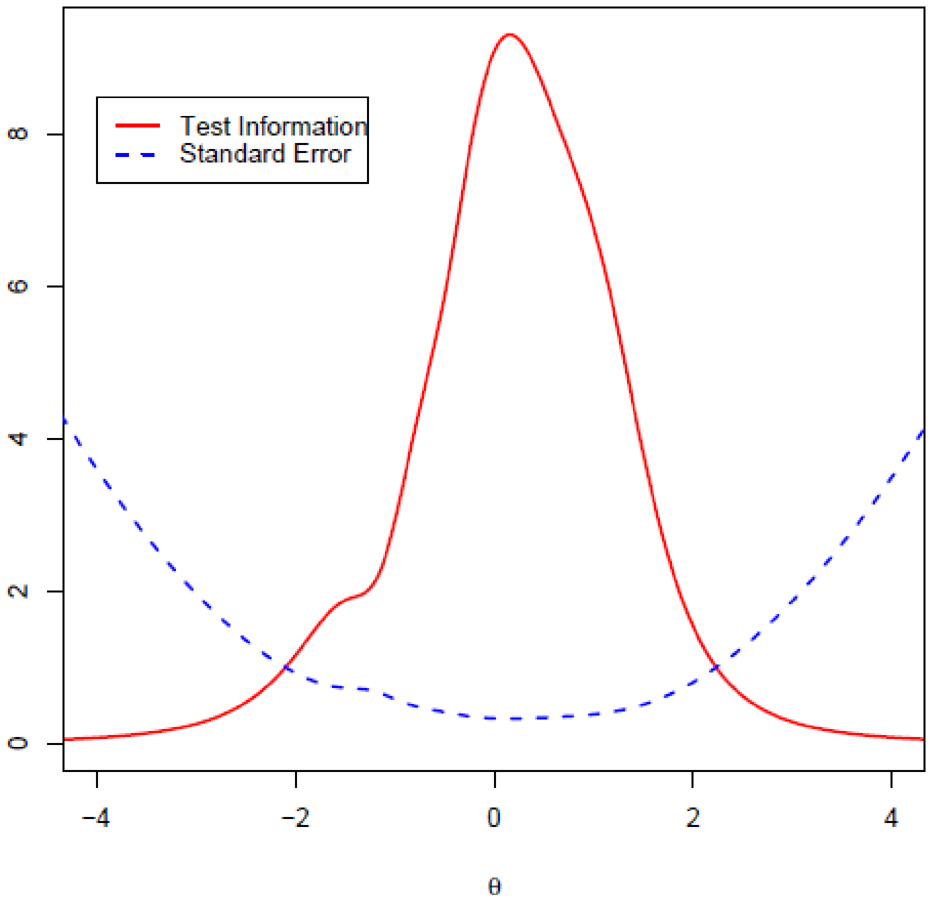
Information and Standard Error

- Standard error should be the focus of investigation, not the information as such
- Standard error (SE) is an estimate of measurement precision at a given theta
- Error of measurement is related to information
  - Error variance of a score estimate is the reciprocal of the Fisher information

$$SE^2\left(\theta\right) = \frac{1}{\mathcal{I}\left(\theta\right)}$$

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Information & Standard Errors



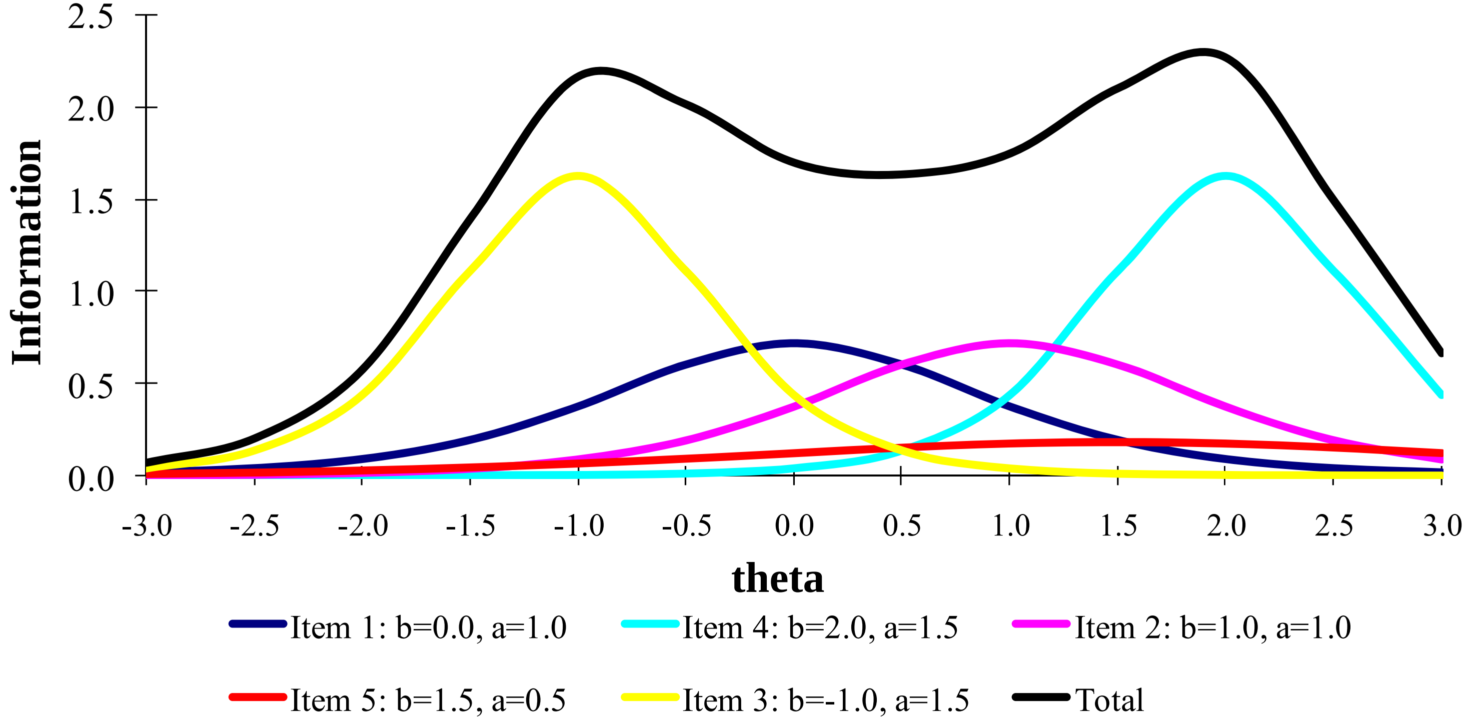
UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Test information function (TIF)

- Information provided by the test about a score is compiled of independent pieces of information provided by the test's items
- Likelihood of responses to all items is the product of likelihoods of responses to each item
  - *Log*likelihood of a response pattern is a sum of items' *log*likelihoods
  - Additive property of information
- Test information is the sum of all item information functions
  - Providing that the local independence holds!

$$I(\theta) = \sum_{i=1}^{p} I_i(\theta)$$

IIFs and TIF



UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Item Information Function (IIF):

$$I_i(\theta) = \frac{\left[P'_i(\theta)\right]^2}{P_i(\theta)\left[1 - P_i(\theta)\right]}$$

- The amount of information the item provides about the latent trait
- Analytical expressions for derivatives of both logistic and normal-ogive functions are easy to derive
  - From known likelihood expressions
- Then they can be substituted in the formula

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

IIFs for logistic models

- For 2PL model (remember constant D=1.7?)

$$I_i(\theta) = [1.7a_i]^2 P_i(\theta)[1 - P_i(\theta)]$$

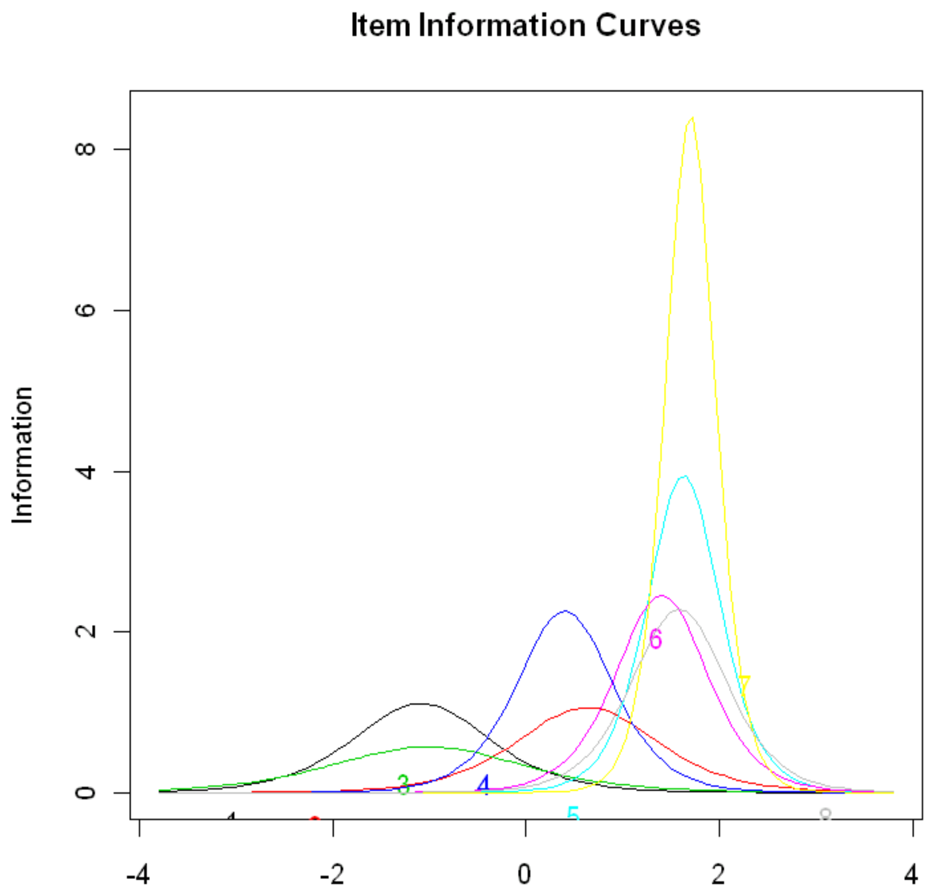
- For 1PL model

$$I_i(\theta) = [1.7a]^2 P_i(\theta)[1 - P_i(\theta)]$$

- For normal ogive models, expressions are slightly more complicated but results will be nearly the same

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

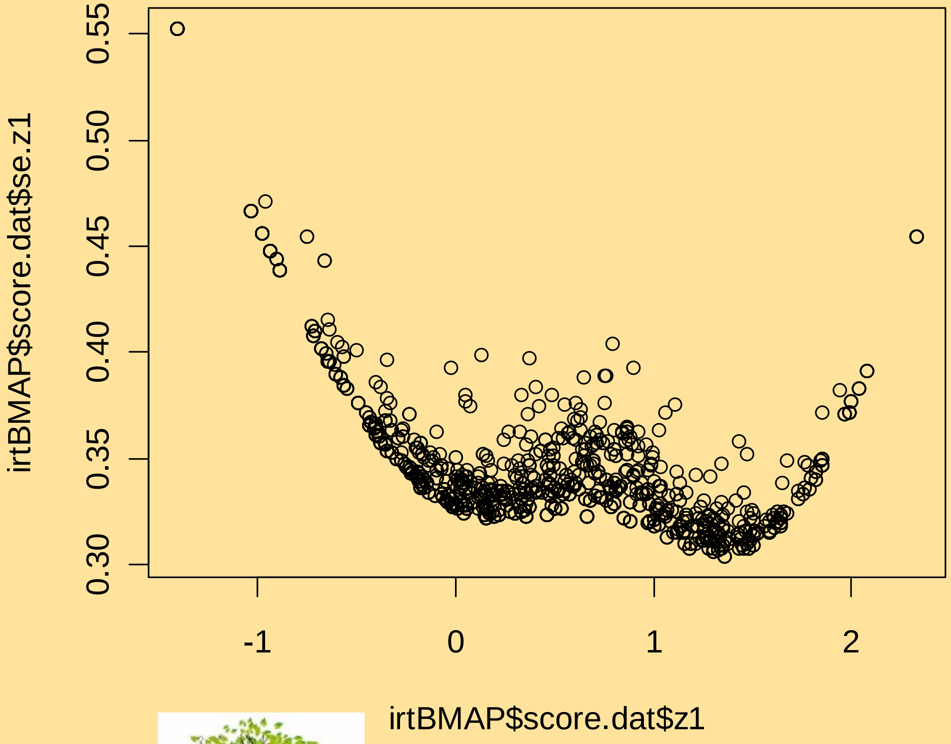
Examples of IIFs



UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Practical – information and SE

- Obtain and assess item information curves for the GHQ facet Anxiety in R software
  - We have estimated the corresponding model and produced scores for individuals already
- Obtain and assess test information curves
- Compute standard errors for each individual and plot them against individual theta scores

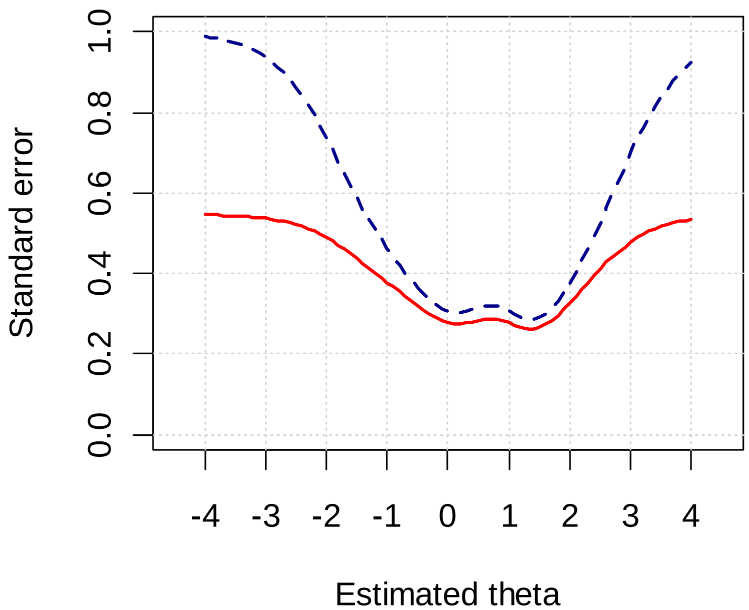


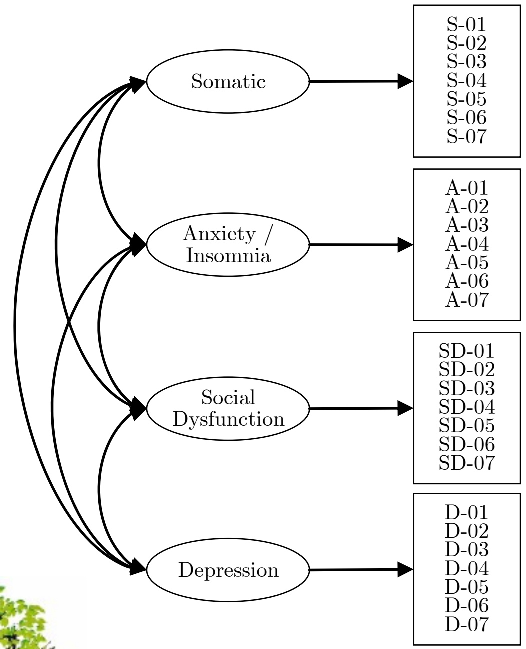
Posterior information

- When a Bayesian estimator is used, knowledge of prior distribution contributes to the information about the score
- The prior distribution's density is treated as another "likelihood" distribution
- Known density formula can be differentiated to obtain Fisher information
  - Univariate normal adds 1 to the ML test information
  - Multivariate normal adds $\left[\boldsymbol{\Sigma}^{-1}\right]_a$ (the diagonal element of the inverted latent trait correlation matrix corresponding to trait a)
  - Multivariate normal with correlated traits adds more information

Benefits of multivariate prior

- Estimating IRT scores for correlated traits with multivariate prior is better than estimating them univariately
- Related traits add information
- Here are the SE functions for subscale Anxiety of GHQ





UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Information for multidimensional models

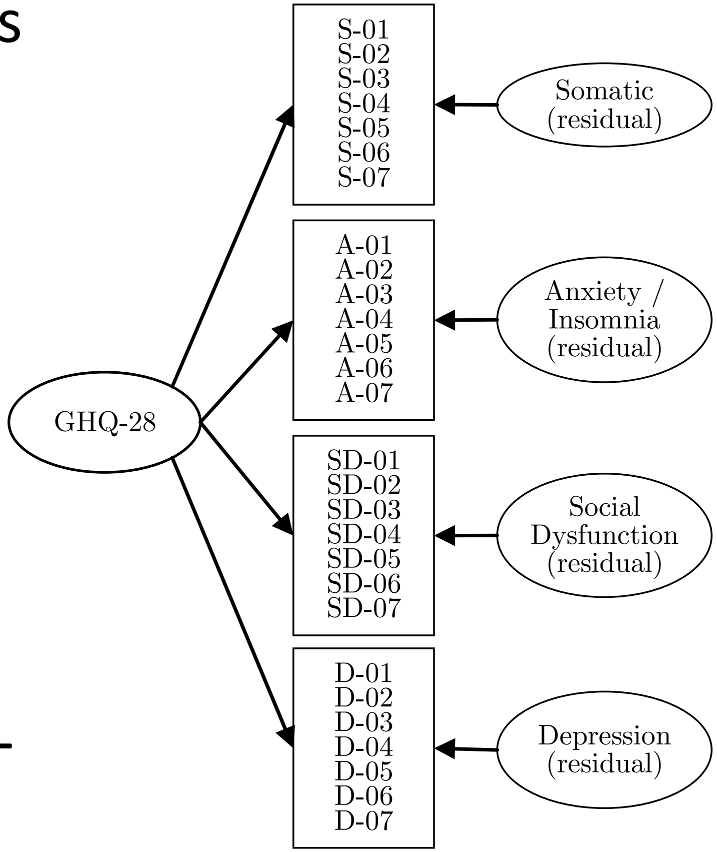
- Item information can be computed for multidimensional models as well
  - General theory is available in textbooks, though largely useless without somebody doing maths for your specific model
  - Special cases – bifactor model, hierarchical model - derivations are available from Anna
- Test information for multidimensional cases can be difficult to summarise
  - Too many variables are involved
  - Sample-based summary is feasible, will discuss next
- Posterior information can be easily computed and added

Special case – bifactor models

- Each item is influenced by two factors – general and specific
- Item information for the general factor

$$\mathcal{I}_i^g\left(\mathbf{g}^*\right) = \frac{\left(\beta_{0i}\right)^2\left[\phi\left(\alpha_i + \beta_{0i}g + \beta_{ai}s_a\right)\right]^2}{\Phi\left(\alpha_i + \beta_{0i}g + \beta_{ai}s_a\right)\left[1 - \Phi\left(\alpha_i + \beta_{0i}g + \beta_{ai}s_a\right)\right]}$$

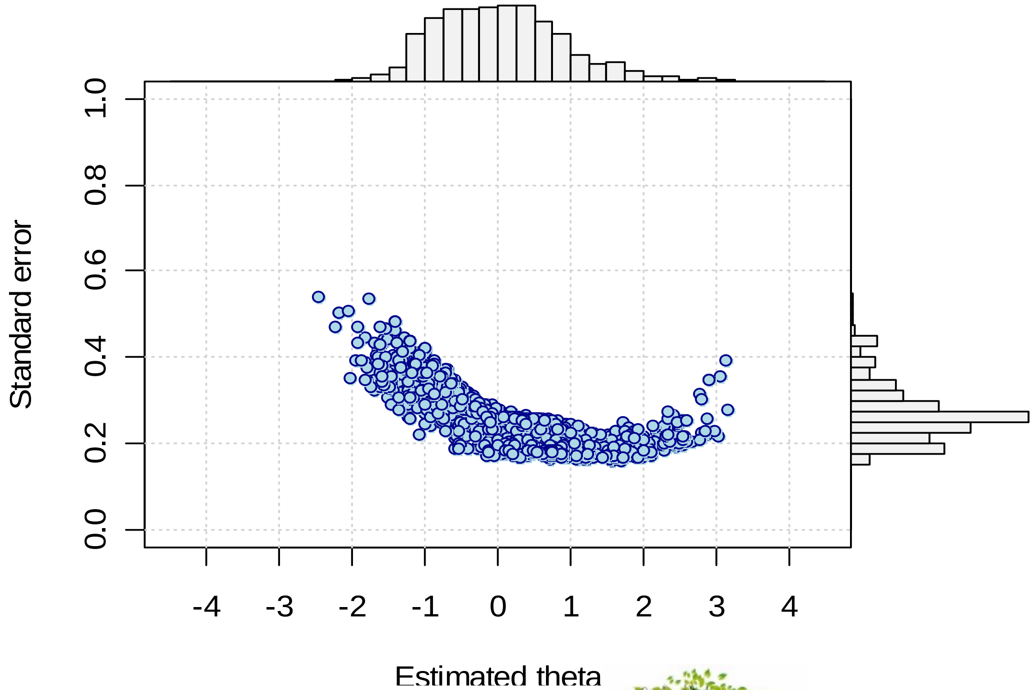
  - is a function of item slopes on the general factor
  - is also conditional on specific factors
- Test information is conditional on ALL specific factors, and is very hard to summarise



UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Illustration - GHQ data

- Bifactor IRT model
  - multidimensional information for the general factor fully conditioned on specific factors



UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Reliability in IRT

- Test reliability is defined as the proportion of variance in the test scores due to the true score
- This can be easily extended to IRT
  - True variance is the latent trait variance
  - Score variance is the sum of the latent trait variance and the error variance
  - Error variance σ_{error} is the squared SE, or reciprocal of test information

$$\sigma^2_{error}(\theta) = SE^2(\theta) = \frac{1}{I(\theta)}$$

  - Hence reliability will depend on the score

$$\rho = \frac{\sigma^2 - \sigma^2_{error}}{\sigma^2} = 1 - \sigma^2_{error} = 1 - \frac{1}{I(\boldsymbol{\theta})}$$

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Marginal IRT reliability

- Single summary index of reliability might be desirable in applications
  - Errors can be summarised across the latent trait
  - More meaningful when the SE function is relatively uniform

$$\rho = \frac{\sigma^2 - \overline{\sigma}^2_{error}}{\sigma^2} = 1 - \frac{\overline{\sigma}^2_{error}}{\sigma^2}$$

- IRT theoretical reliability
  - Squared SEs are averaged across the latent trait (integration is required)
- IRT empirical reliability
  - Squared SEs are averaged across estimated values in the sample

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{j=1}^{N}\left(\hat{\theta}_j - \overline{\hat{\theta}}\right)^2 \qquad \overline{\hat{\sigma}}^2_{error} = \frac{1}{N}\sum_{j=1}^{N} SE^2\left(\hat{\theta}_j\right)$$

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Illustration – GHQ data

Model	Estimator	Reliability estimates					
		S	A	SD	D	*General*	$\left[\text{corr}(\theta,\hat{\theta})\right]^2$
CTT (alpha)	Sum score	.809	.878	.816	.890	.926	.760
Unidimensional	MAP					.899	.824
	EAP					.908	
Correlated traits	MAP	.817	.871	.666	.639	-	-
Hierarchical	MAP	.813	.867	.665	.607	.891	.852
Bifactor	MAP	-	-	-	-	.902	.869

Scores were estimated using Bayesian EAP and MAP estimators with standard normal prior according to their respective model

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Assessments of measurement precision – summary

- For each IRT estimated score, standard error of measurement is computed
  - For EAP method, it is the standard deviation of the likelihood function for the response pattern
  - For MAP method, it is the reciprocal of the square root of the test information function
- SE depends on the estimated score (a function)
- A summary index can be also computed
  - Marginal reliability