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

Day 3

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

Jan Boehnke

8. UNIDIMENSIONAL IRT MODELS FOR ORDINAL DATA

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Polytomous IRT

- Many of the instruments in use have polytomous items
- as well as it is in CTT this is advantageous for IRT models:
  - every item thereby covers a range of the latent trait
  - and this heightens measurement precision

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Polytomous IRT

- Basically, every polytomous item can be dichotomized repeatedly:
  - every item with g categories
  - will be decomposed in g-1 dichotomous items

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Polytomous IRT

- This might be thought of as transforming the item "Were you limited in doing vigorous activities" (with not limited / limited a little / limited a lot) into two questions:
  - "Were you limited a little in doing..." (Yes / No) – measuring the transition from the lowest to the middle category
  - "Were you limited a lot in doing..." (Yes / No) – measuring the transition from the middle to the top category

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Polytomous IRT

- To do this a) more efficiently and/or b) more correctly with regard assumptions of IRT several models have been proposed
  - (Generalized) Partial Credit Model, (G)PCM: covered in a second in ltm
  - Graded Response Model (GRM): used by Mplus and covered tomorrow

Generalized Partial Credit Model

- The model is: $$P_{ix}(\theta) = \frac{\exp \sum_{s=0}^{x} a_i(\theta - b_{is})}{\sum_{r=0}^{m} \left[\exp \sum_{s=0}^{r} a_i(\theta - b_{is}) \right]}$$

- Easier to see step by step (assume 3 categories):
  - Probability of completing 0 steps

$$P_{i0}(\theta) = \frac{\exp[0]}{\exp[0] + \exp\left[0 + a_i(\theta - b_{i1})\right] + \exp\left[0 + a_i(\theta - b_{i1}) + a_i(\theta - b_{i2})\right]}$$

  - Probability of completing 1 step

$$P_{i0}(\theta) = \frac{\exp\left[0 + a_i(\theta - b_{i1})\right]}{\exp[0] + \exp\left[a_i(\theta - b_{i1})\right] + \exp\left[0 + a_i(\theta - b_{i1}) + a_i(\theta - b_{i2})\right]}$$

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

The Partial Credit logic

- Created specifically to handle items that require logical steps, and partial credit can be assigned for completing some steps (common in mathematical problems)
- Completing a step assumes completing **all steps** below
- Computing probability of response to each category is direct ("divide-by-total"):
  - Probability of responding in category x (completing x steps) is associated with ratio of
    - odds of completing all steps before and including this one, and
    - odds of completing all steps
  - Each step's odds are modelled like in binary logistic models
    - For an item with m+1 response categories, m *step difficulty* parameters $b_1 \dots b_m$ are modelled

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Polytomous data set

- Reading data for polytomous example:

```
GHQ28poly <- read.table(file.choose(),
  header=TRUE, sep="\t", na.strings="NA",
  dec=".", strip.white=TRUE)

Anxiety.poly<-GHQ28poly[,8:14]
```

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Estimating the GPCM

- The (G)PCM is estimated in ltm using the gpcm() command:

```
gpcm(data, constraint = c("gpcm", "1PL", "rasch"), IRT.param = TRUE,
    start.val = NULL, na.action = NULL, control = list())
```

- PCM assumes that items differ only in their difficulty and their threshold spacing:

```
ResultPCM<-gpcm(Anxiety.poly,
  constraint=c("rasch"))
```

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Estimating the GPCM

- The (G)PCM is estimated in ltm using the gpcm() command:

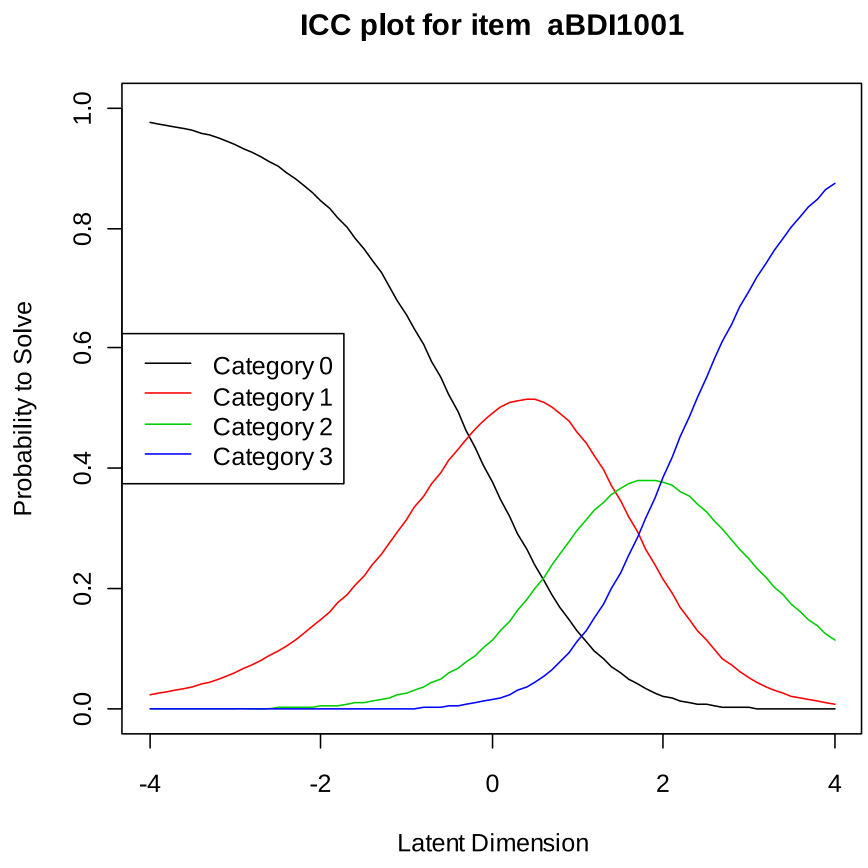
```
gpcm(data, constraint = c("gpcm", "1PL", "rasch"), IRT.param = TRUE,
    start.val = NULL, na.action = NULL, control = list())
```

- GPCM assumes that items differ in their difficulty, threshold spacing and their discrimination:

```
ResultGPCM<-gpcm(Anxiety.poly,
  constraint=c("gpcm"))
```

Interpretation

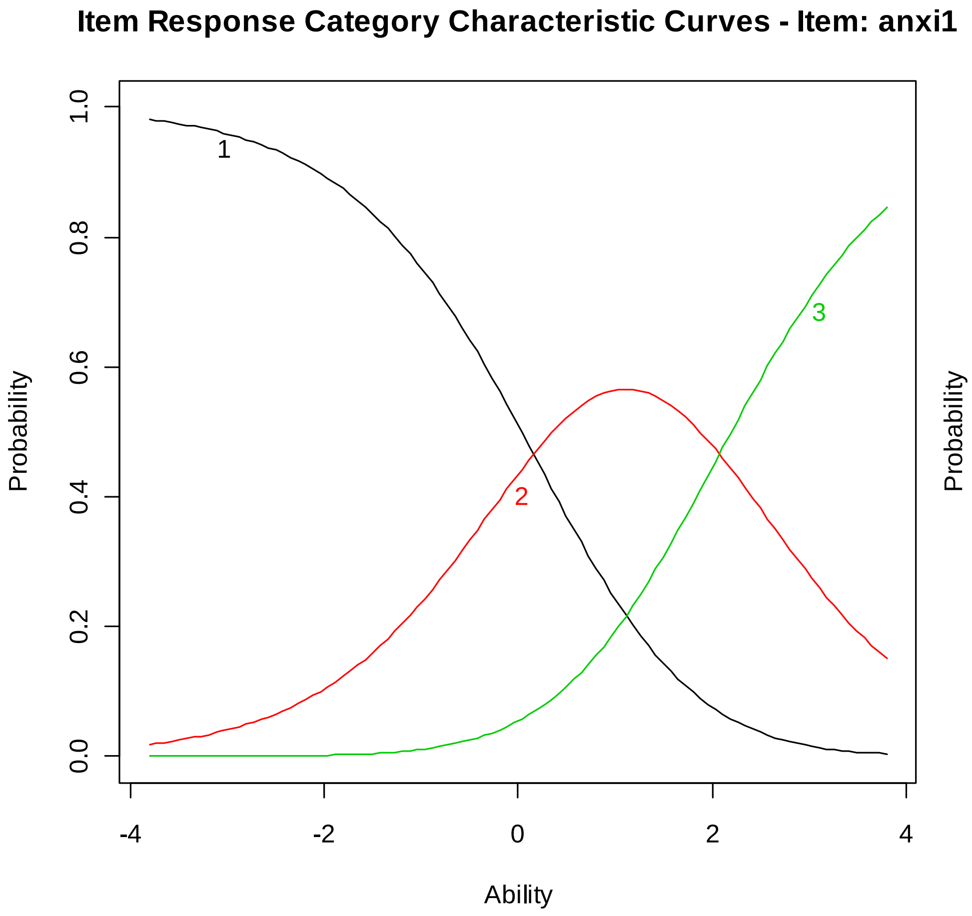
- Step difficulty parameters have an easy graphical interpretation – they are points where the category lines cross
- Relative step difficulty reflects how easy it is to make transition from one step to another
  - Step difficulties do not have to be ordered
  - "Reversal" happens if a category has lower probability than any other at all levels of the latent trait

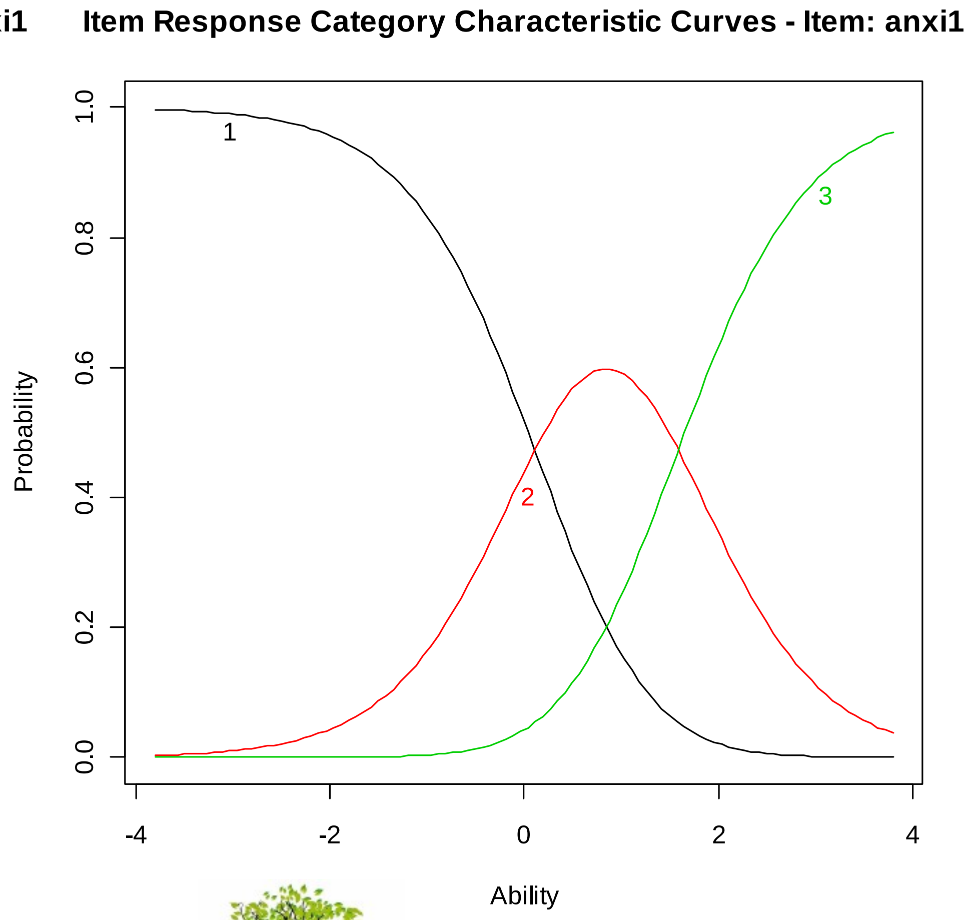


UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

PCM vs. GPCM

(again use `plot()`)





UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

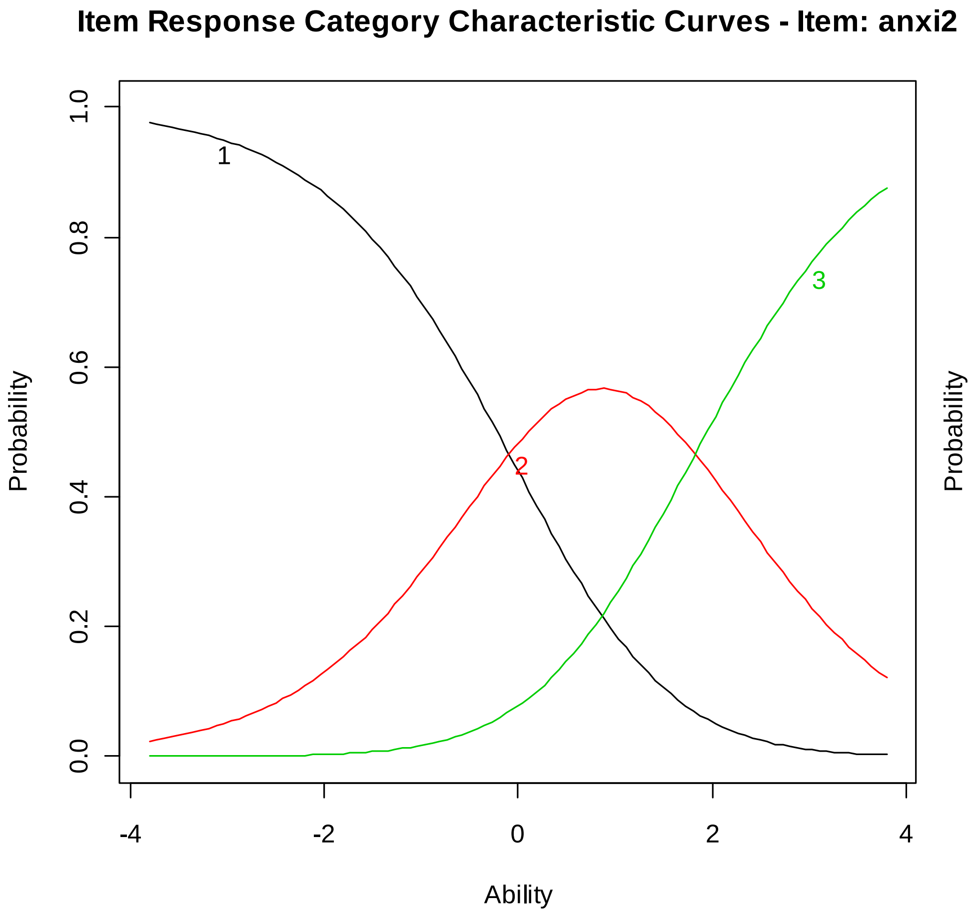
Visual inspection of ICCs

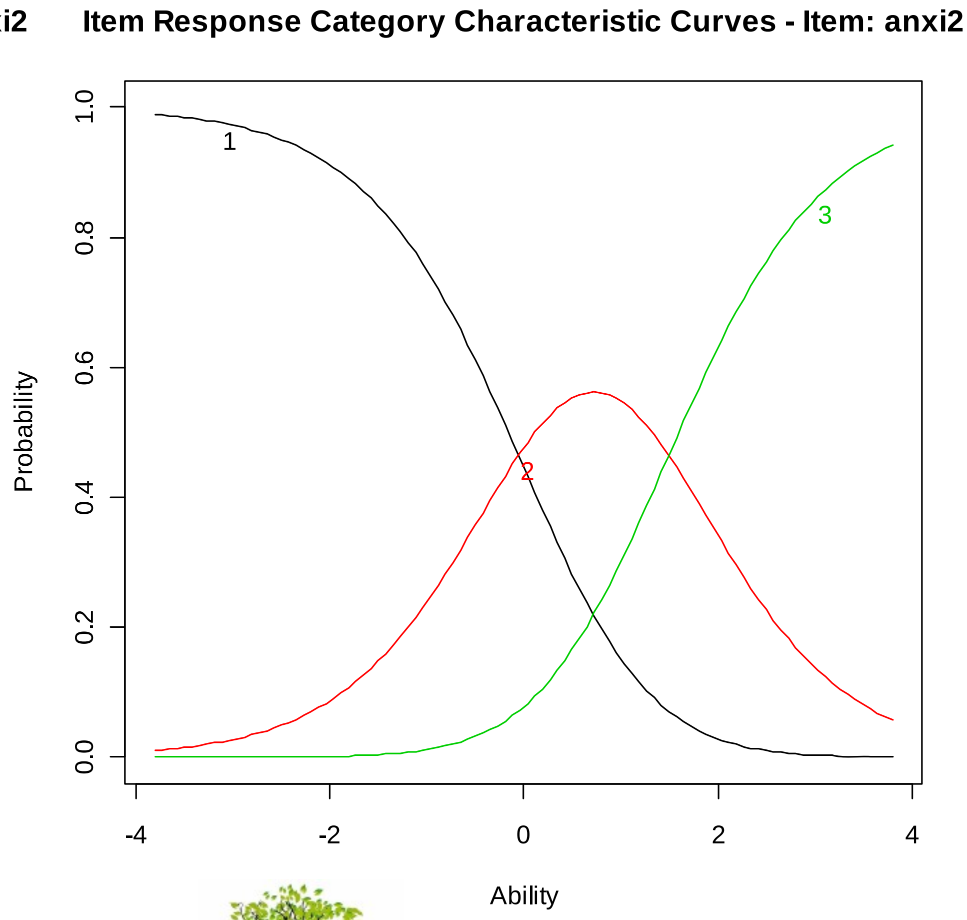
- Usefulness of visual inspection:
  - model assumptions: can be used to identify deviations from monotonicity / scalability
  - scale development: informs on the use of the scale by the respondents (e.g. ordinal format really accurat?)
  - scale development: (other way round) needed number of categories overall

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

PCM vs. GPCM

(again use `plot()`)

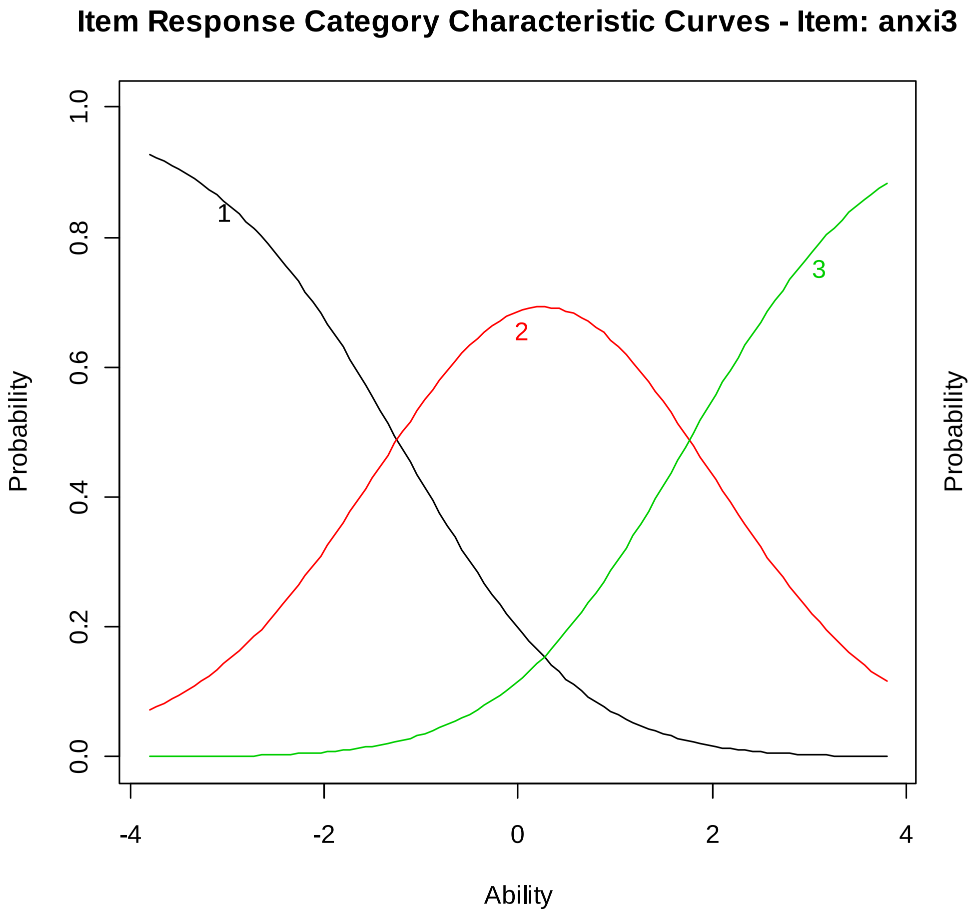


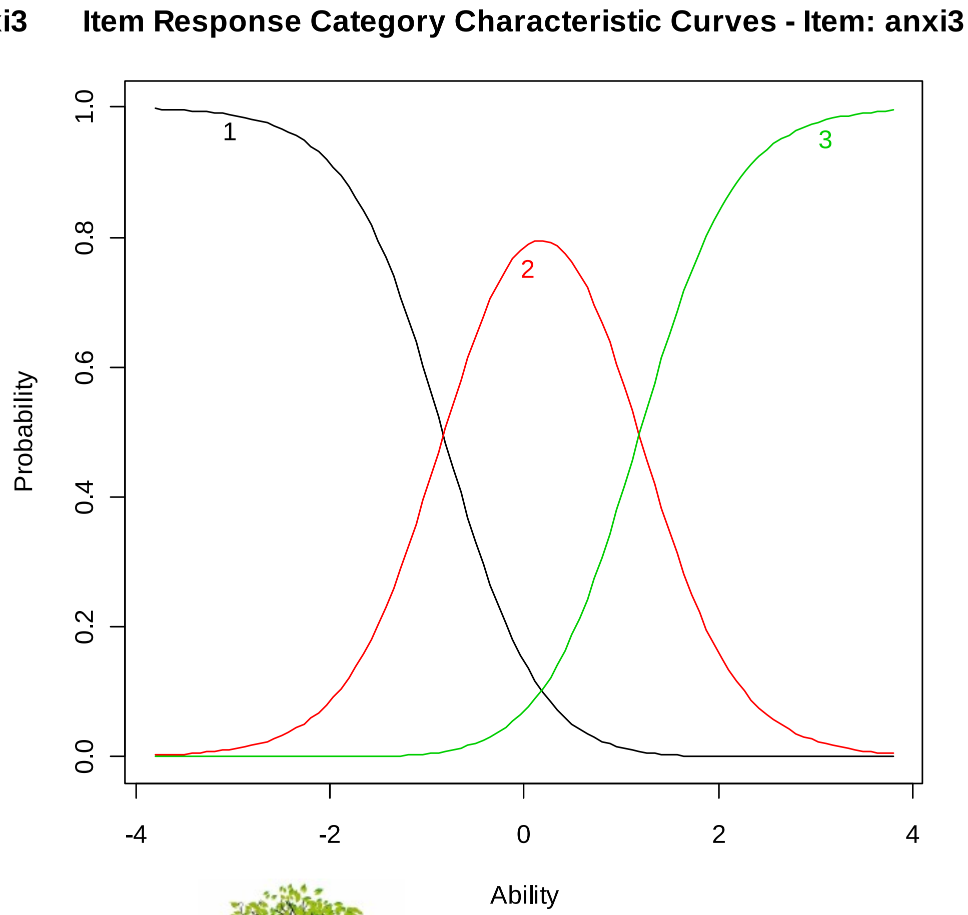


UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

PCM vs. GPCM

(again use `plot()`)

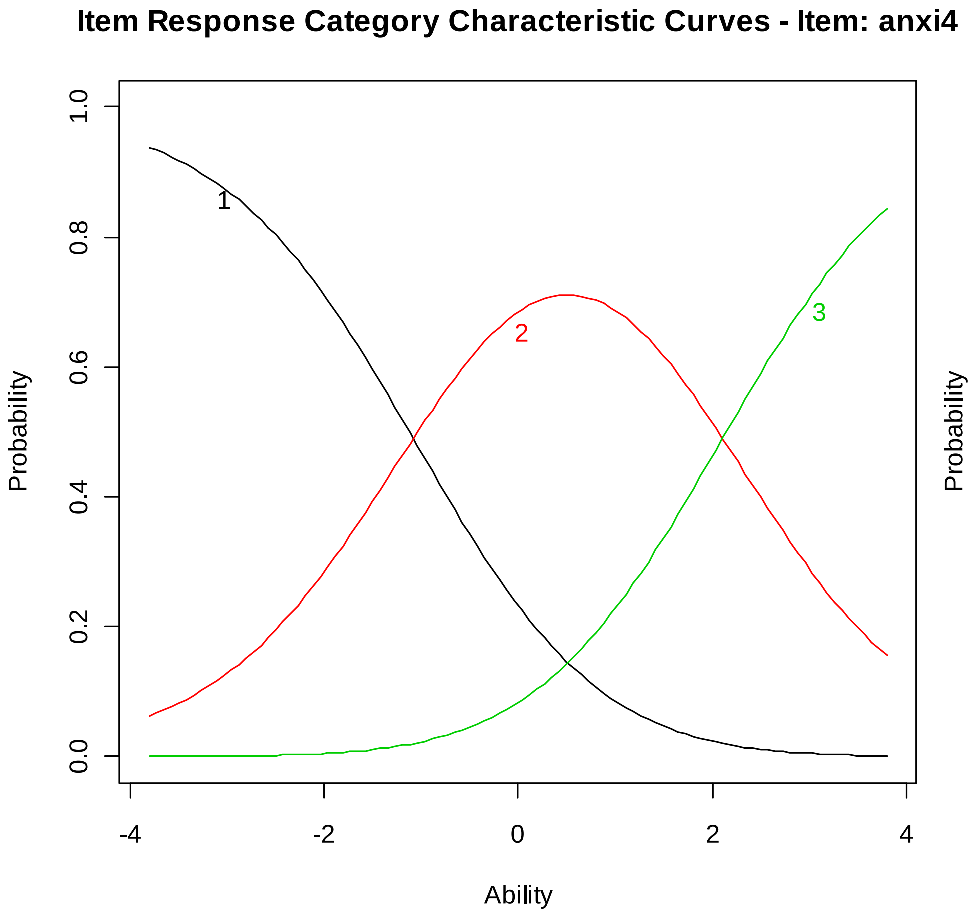


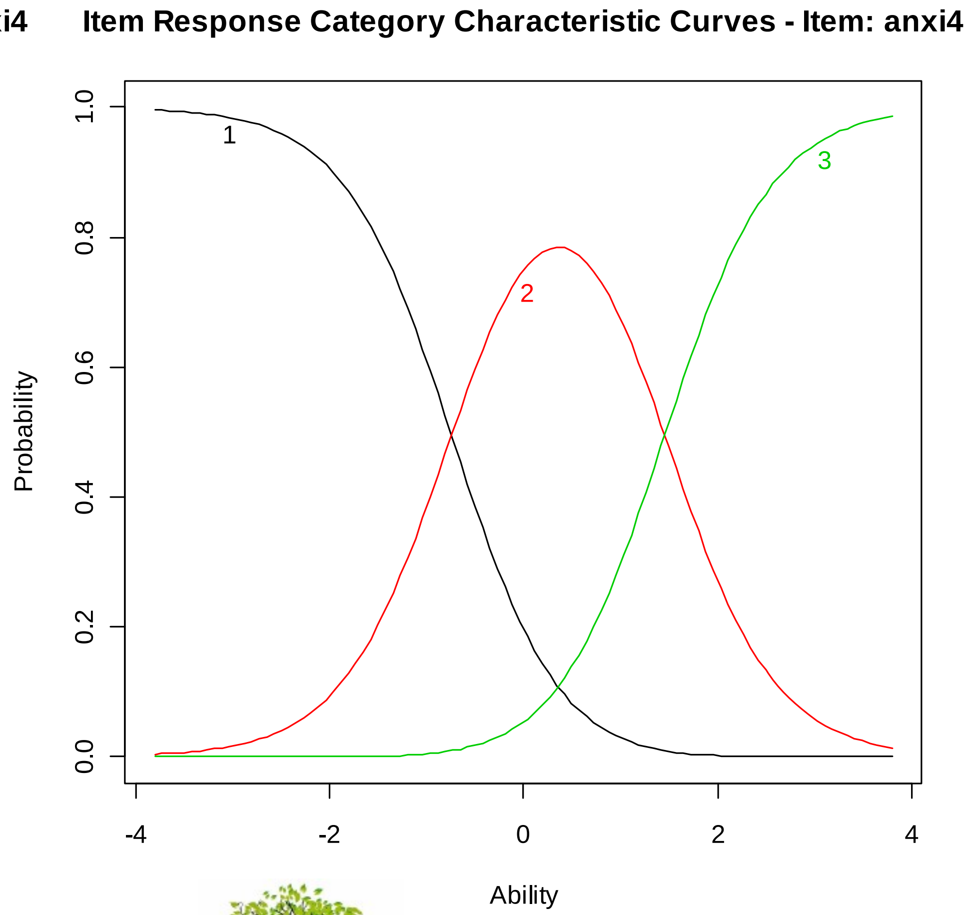


UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

PCM vs. GPCM

(again use `plot()`)

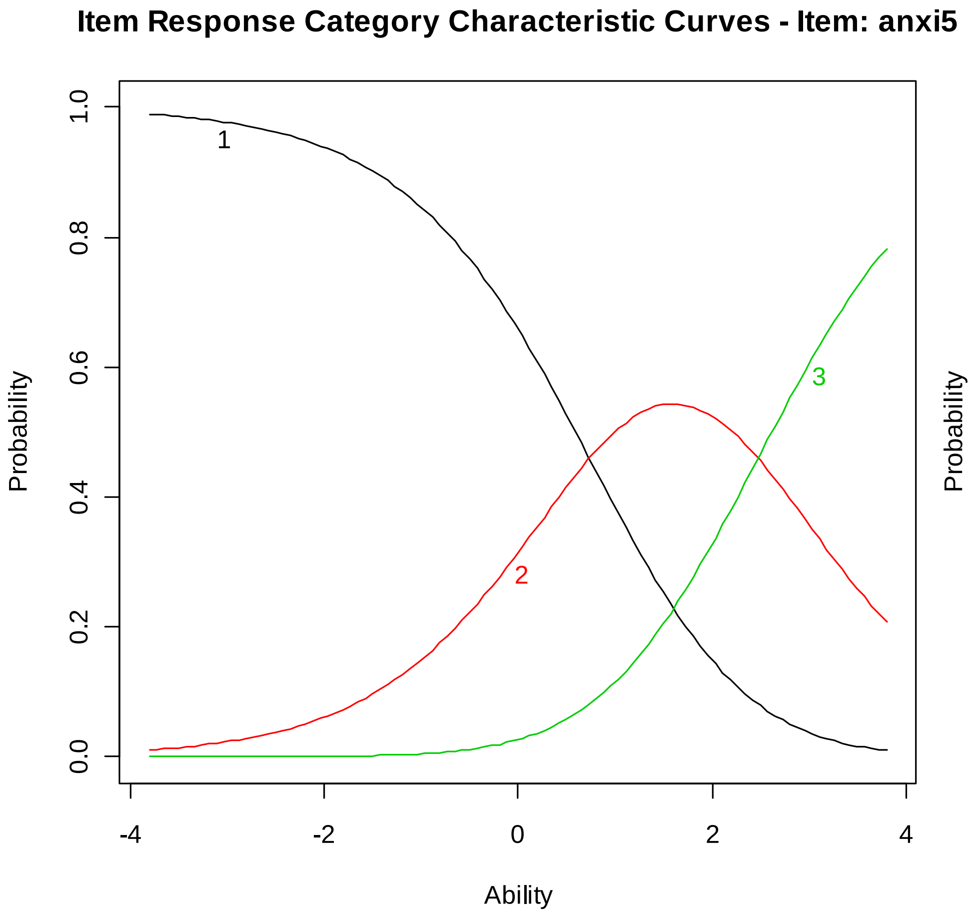


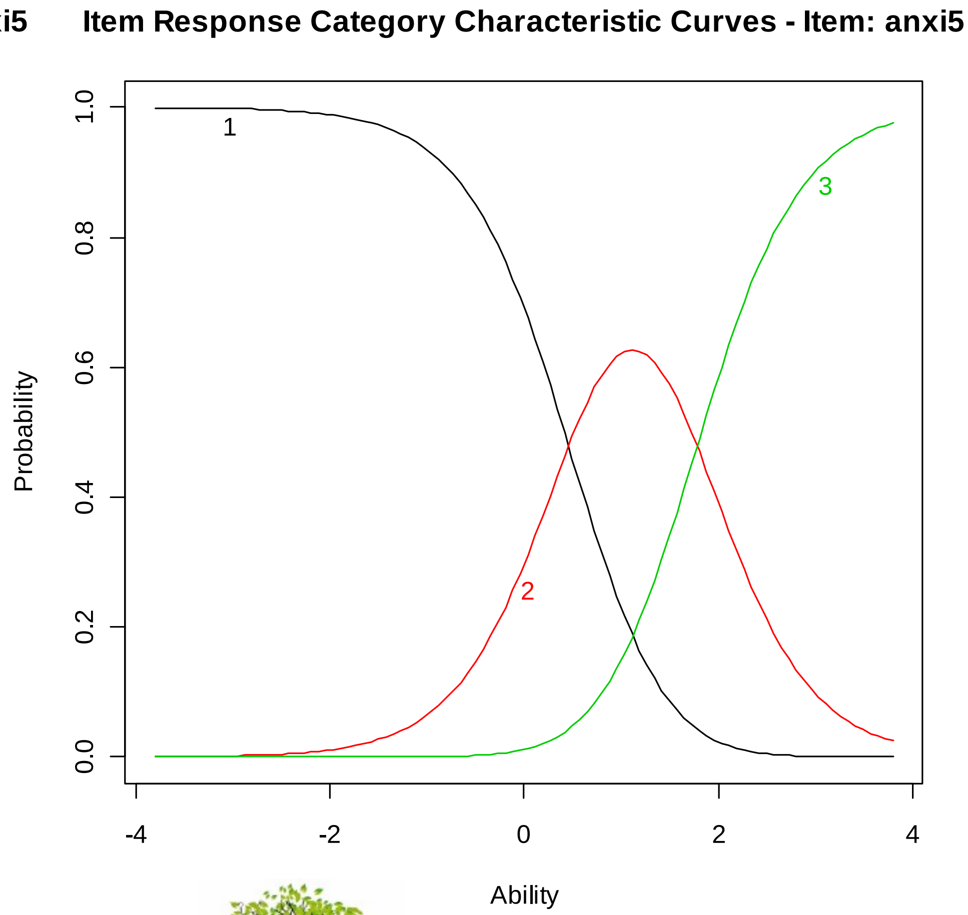


UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

PCM vs. GPCM

(again use `plot()`)

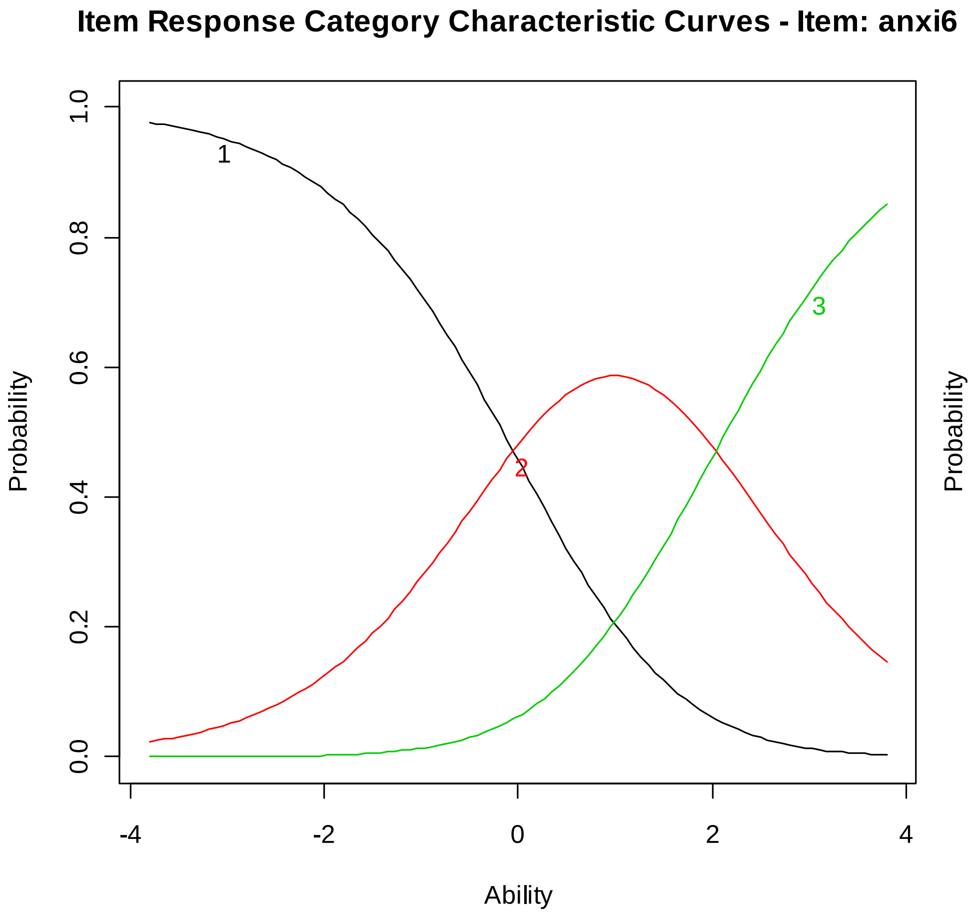


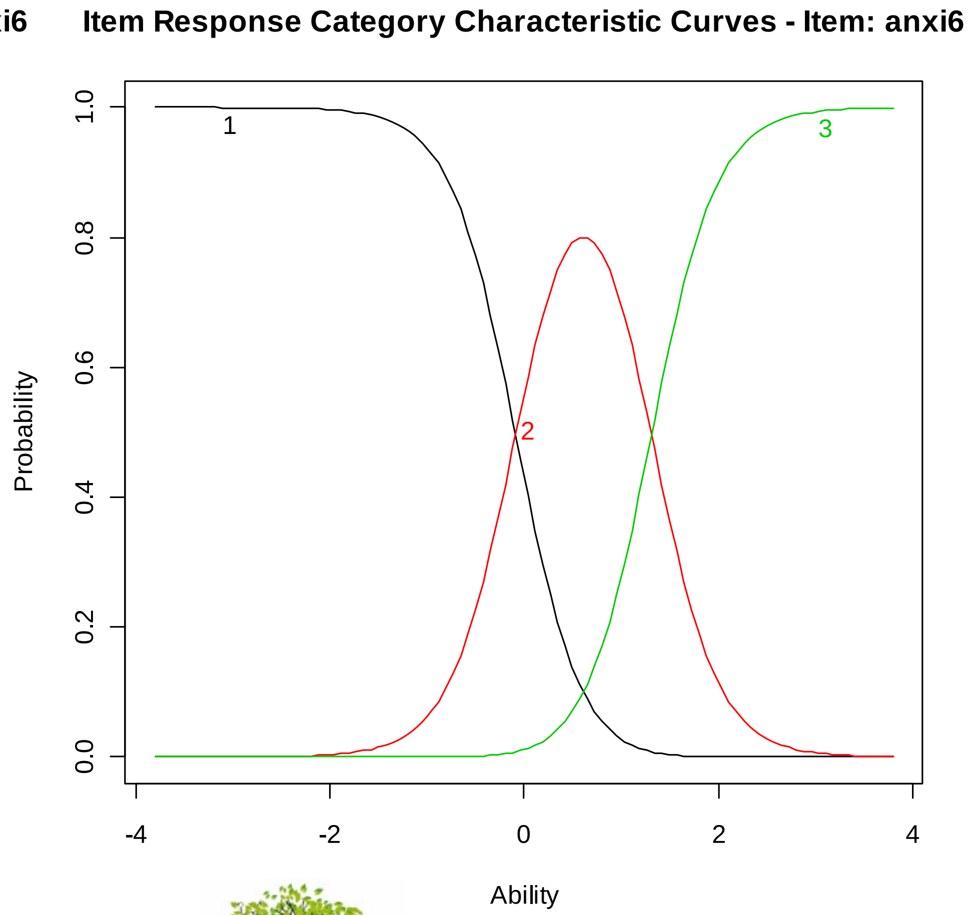


UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

PCM vs. GPCM

(again use `plot()`)

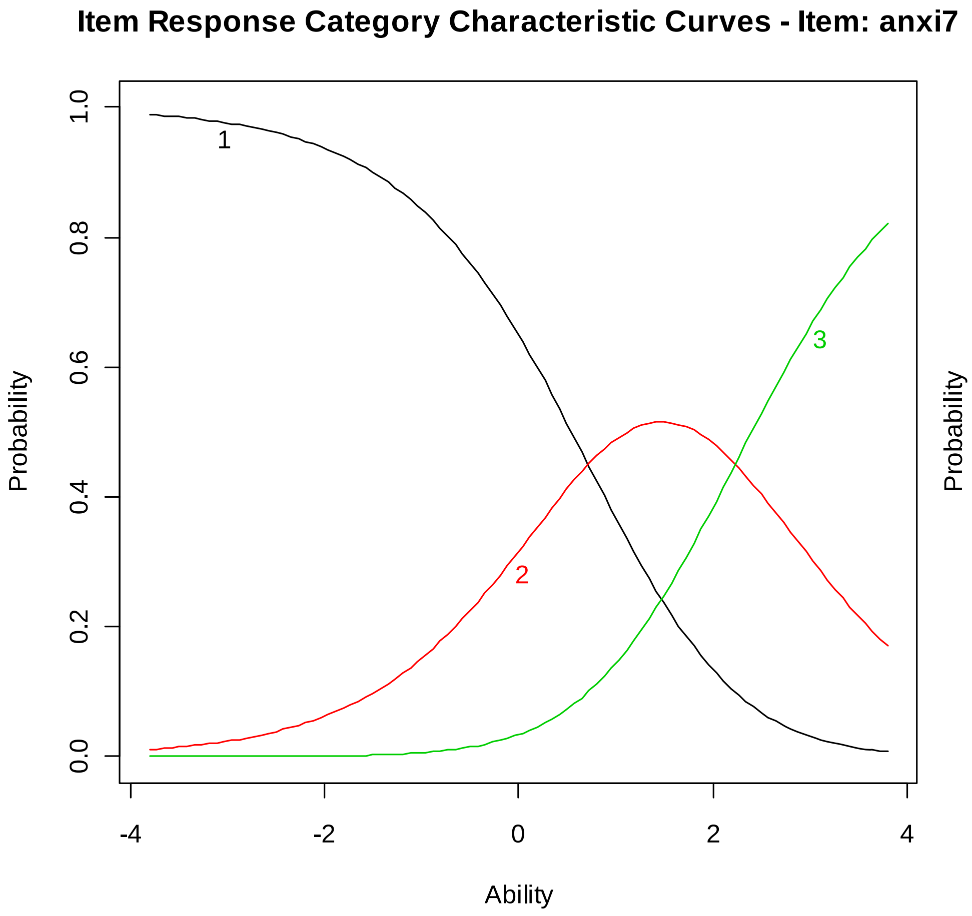


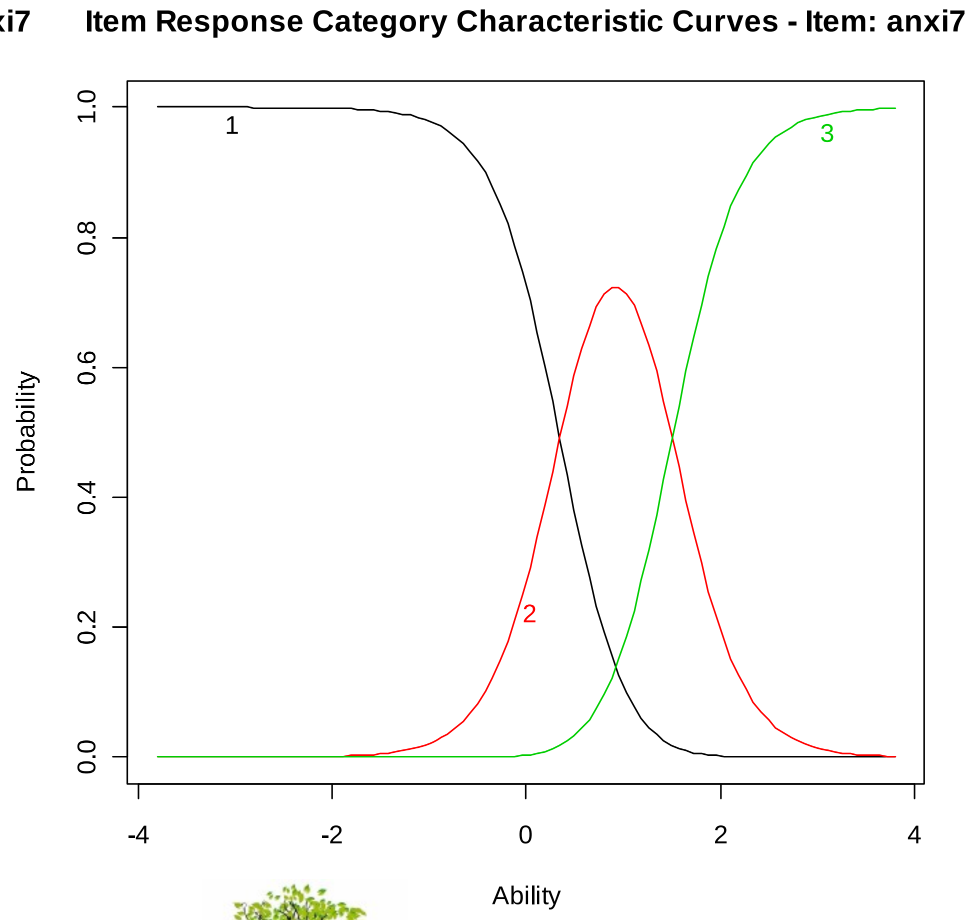


UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

PCM vs. GPCM

(again use `plot()`)





UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

The Graded Response Model (GRM)

- Extension of the 2PL model to handle multiple response categories that are logically ordered
- GRM is a model specified to estimate the probability of scoring into a specific category *or above*
- for a given item i, its item parameters and the ability of a person

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

The Graded Response Model (GRM)

- Computing probability of response to each category requires a 2-step process:
  - First, probability of responding in or above category x, Px*, is computed
  - These are simple 2PL curves reflecting the dichotomy
  - Second, probability of responding in category x equals the difference Px* - Px+1*

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

The Graded Response Model

- Let $x = 0,1,\ldots, m_i$ be a category number
- Then
  - the probability of responding in the lowest category or above is 1 ($P^*_0=1$)
  - Probability of responding in the highest category is $P_{mi}= P^*_{mi}$
- Probability of responding in any intermediate category x is

  $P_x= P^*_{mx}- P^*_{mx+1}$
- Probability of falling in the category x or above is

$$P^*_{ix}(\theta) = \frac{e^{Da_i(\theta - b_{ix})}}{1 + e^{Da_i(\theta - b_{ix})}}$$

- Item has one discrimination (a_i) and m_i threshold parameters (b_{ix})

UNIVERSITY OF CAMBRIDGE The Psychometrics Centre

Estimating the GRM

- The GRM is estimated in ltm using the grm() command:

```
grm(data, constrained = FALSE, IRT.param = TRUE, Hessian = FALSE,
    start.val = NULL, na.action = NULL, control = list())
```

- can also be constrained to items having the same discriminations / slopes:

```
ResultGRM1<-grm(Anxiety.poly, constrained=TRUE)
```

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Estimating the GRM

- The GRM is estimated in ltm using the grm() command:

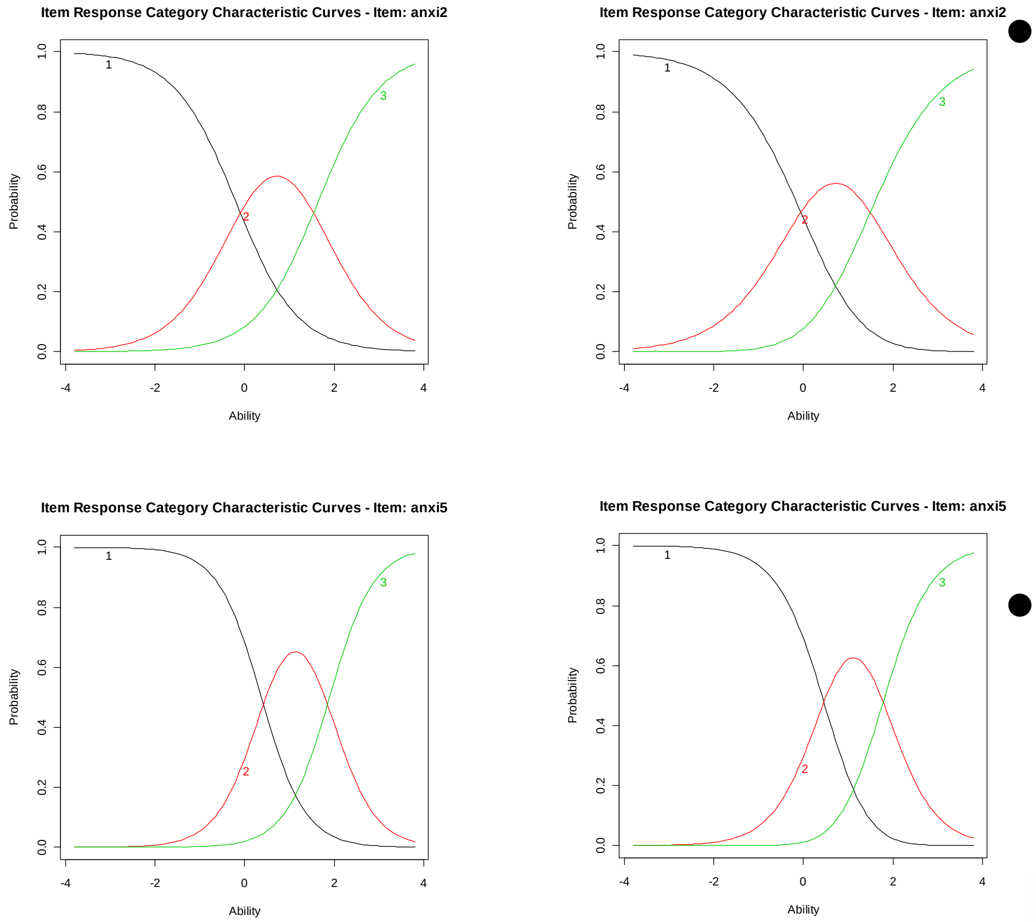
```
grm(data, constrained = FALSE, IRT.param = TRUE, Hessian = FALSE,
    start.val = NULL, na.action = NULL, control = list())
```

- or with free discriminations as well:

```
ResultGRM2<-grm(Anxiety.poly)
```

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

GRM vs. GPCM



- despite the differences in interpretation of the curves (conceptually important!)
- results are visually often very similar

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

GRM vs. GPCM

- Both widely applicable to questionnaire data
  - Items can have different discriminations
  - Items can have different number of categories
  - Category thresholds can be spaced at any intervals
    - Do not have to worry about whether distance between "never" and "rarely" is the same as between "sometimes" and "often"
  - Category thresholds have to be ordered (reasonable assumption for questionnaires using rating scales)

GRM vs. GPCM

- GRM might have slight computational advantage when there are no responses in a given category – the cumulative probability can nevertheless be determined
- GPCM logic of item parameters being that point of the continuum, where adjacent categories have the same probabilities to be scored in maybe more intuitive
  - (than in GRM: the point on the continuum where the probability of choosing this or a higher category is .50)

Testing models

- For the PCM which like the 1PL in the dichotomous case deals only with the persons patterns, also the GoF test is possible (description see above)

```
TestPCM<-GoF.gpcm(ResultPCM,B=499)
```

Testing models

`TestPCM`

Parametric Bootstrap Approximation to Pearson chi-squared Goodness-of-Fit Measure

Tobs: 3133.78

data-sets: 500

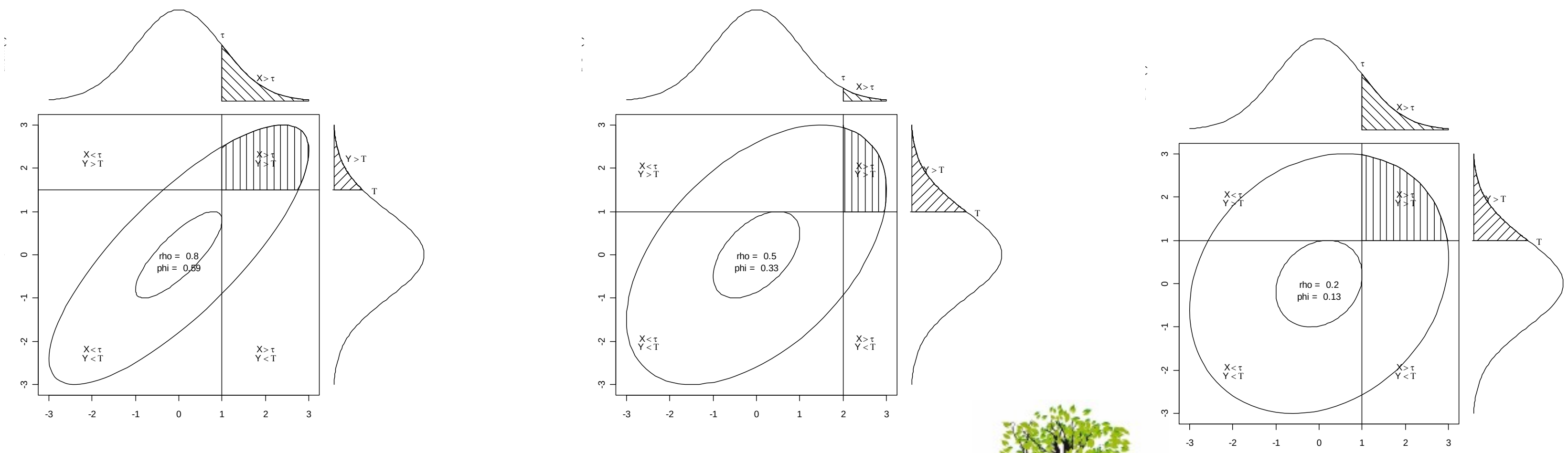
p-value: 0.006

- The PCM does not predict the observed response patterns adequately

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Testing unidimensionality

- Testing unidimensionality of polytomous items in `ltm` not possible
- therefore parallel analysis based on the polychoric correlations between the items



UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Testing unidimensionality

(e.g. Hayton, Allen & Scarpello (2004). *Organizational Research Methods*, *7*, 191 -205.)

- Calculate polychoric correlations in observed data, perform FA/PCA and save eigenvalues
- Simulation:
  1. simulate data set with same properties (N, number of items, categories per item) – but with random items so that any ρ(i1, i2) has an expectancy of 0
  2. Calculate polychoric correlations in observed data, perform FA/PCA and save eigenvalues
- Repeat these steps; compare the observed and quantiles of simulated eigenvalues: how many of the observed eigenvalues are above their respective simulated quantiles? – These indicate factors that do not contain only random variation
- Depending on quantile, a high number of simulated data sets is needed (e.g. 95th with B = 100 only 5 eigenvalues are used to estimate the quantile – not very stable)

Testing unidimensionality

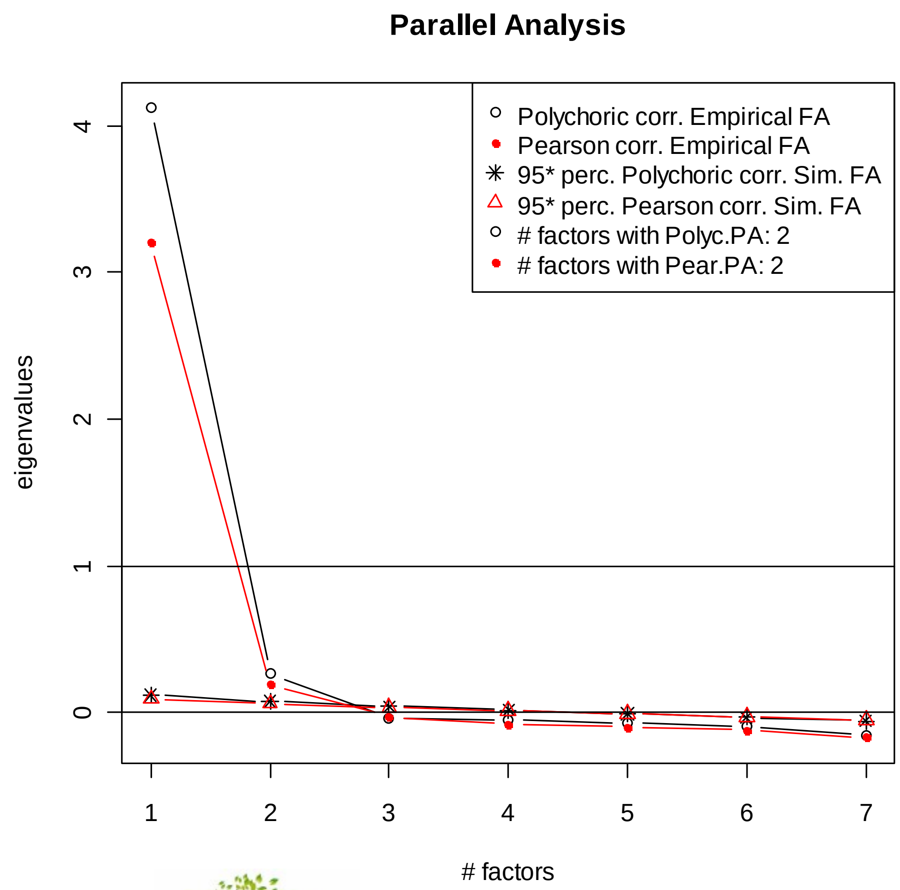
- Since this test takes a while (about 35min), here only syntax and results:

```
library(random.polychor.pa)

Anxiety.polychor<-
  random.polychor.pa(nvar=7,n.ss=2901,
  nstep=3,nrep=500,Anxiety.poly.pa,
  q.eigen=.95)
```

Testing unidimensionality

- .95-quantile of 1st and 2nd factor lower for the simulated data
- therefore two factors might be considered
- nevertheless: difference for 2nd factor very small



UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Testing GPCM vs. PCM

```
anova(ResultPCM,ResultGPCM)
```

Likelihood Ratio Table

	AIC	BIC	log.Lik	LRT	df	p.value
ResultPCM	32772.61	32856.23	-16372.30		14	
ResultGPCM	31611.15	31736.57	-15784.57	1175.46	21	<0.001

- GPCM again provides better fit

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Comparison in information criteria GPCM vs. PCM

- GPCM provides more parsimonious fit than PCM
- GRM with free parameters more parsimonious fit than constrained GRM
- models in principle comparable on information criteria but decision should better by guided by theoretical reasons

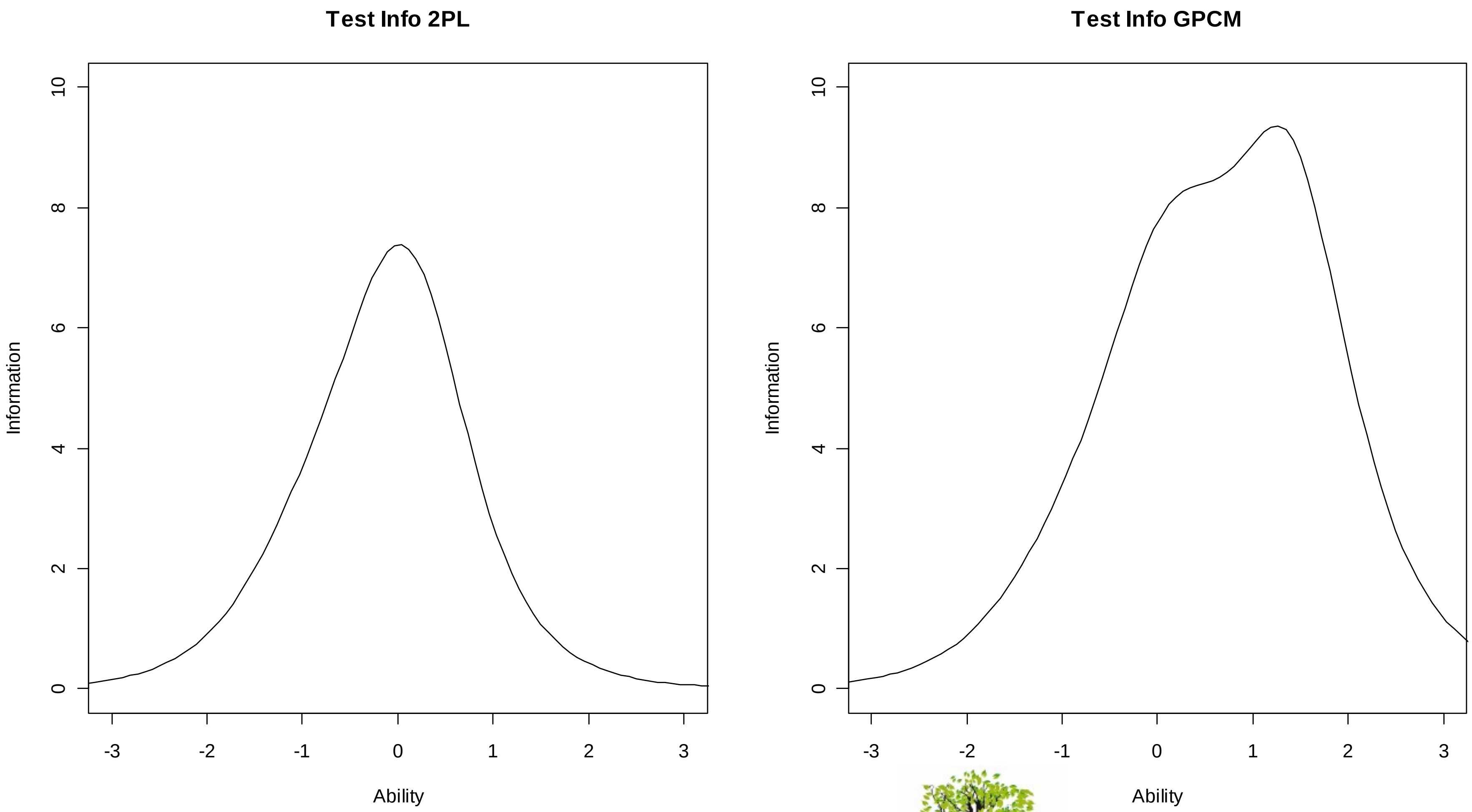
	PCM	GPCM	GRM; constrained	GRM; free
LogLike	-16372	-15784	-15928	-15798
AIC	32772	31611	31887	31639
BIC	32856	31736	31977	31765

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Item information function

```
par(mfrow=c(1,2))
plot(Result2PL,type="IIC",items=
 0,xlim=c(-3,3),ylim=c(0,10),
 main="Test Info 2PL")
plot(ResultGPCM,type="IIC",items
 =0,xlim=c(-3,3),ylim=c(0,10) ,
 main="Test Info GPCM")
```

Item information function



UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

How to choose from the many available IRT models?

- Is data binary, polytomous, or mixed?
- How large is sample size?
  - smaller samples, less complex models
- How do model fit statistics compare?
  - Model fit results should be influential in model selection
- How much experience do I or my colleagues have with IRT models?
  - Or, can I get technical help?

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

How to choose from the many available IRT models?

- When deciding especially between 1PL, 2PL and 3PL:
  - every parameter included should have a substantive meaning that also can be linked to theory
  - "c" in cognitive tests maybe guessing; in symptom checklists maybe base-rate; etc.

Rasch vs. 2PL or 3PL Model?
(or PC vs. GR and GPCM?)

- This comparison has been of interest for many years, and generated quite emotional debate.
- Rasch model has many desirable properties
  - estimation of parameters is straightforward,
  - sample size does not need to be big,
  - number of items correct is the sufficient statistic for person's score,
  - measurement is completely additive,
  - specific objectivity.
- But your data might not fit the Rasch model…

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Rasch vs. 2PL or 3PL Model? (Cont.)

- Two-parameter logistic model is more complex
  - Often fits data better than the Rasch model
  - Requires larger samples (500+)
- Three-parameter logistic model is even more complex
  - Fits data where guessing is common better
  - Estimation is complex and estimates are not guaranteed without constraints
  - Sample needs to be large in applications.

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Choice of model must be pragmatic

- Desirable measurement properties of the Rasch model may make it a target model to achieve when constructing measures
  - Rasch maintained that if items have different discriminations, the latent trait is not unidimensional
- However, in many applications it is impossible to change the nature of the data
  - Take school exams with a lot of varied curriculum content to be squeezed in the test items
- There must be a pragmatic balance between the parsimony of the model and the complexity of the application

Nominal responses

- What about items where ordering of categories does not make sense or is not obvious?
  - Distractor alternatives in multiple choice cognitive items
    - Of course simple correct/incorrect scoring will do in most cases but some distracters can be “more correct than others” and therefore provide useful information
  - Questionnaire items with response options that are not rating scale (e.g. possible alternatives for attitudes or behaviours)
    - In a measure of risk for bulimia: “***I prefer to eat***”

(a) at home alone - (b) at home with others – (c) in a restaurant – (d) at a friend’s house – (e) doesn’t matter

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Nominal response model

- Bock (1972) proposed another "divide-by-total" model

$$P_{ix}(\theta) = \frac{\exp(a_{ix}\theta - c_{ix})}{\sum_{x=0}^{m} \exp(a_{ix}\theta - c_{ix})}$$

- Notice that:
  - Each category has its own discrimination parameter a_x (and these can be positive and negative)
  - Each category has its own intercept parameter c_x
  - To identify the model, constraints on a_x and c_x must be set

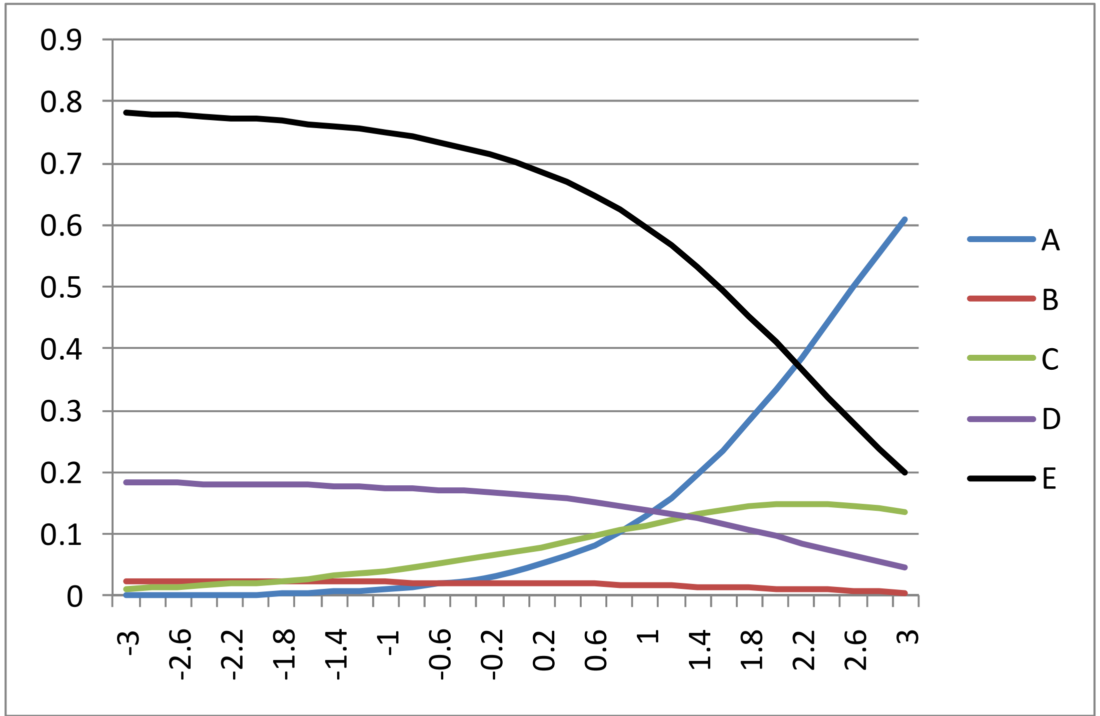
UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Nominal response curves

- ***"I prefer to eat"***

  (a) at home alone *(b) at home with others* *(c) in a restaurant*

  (d) at a friend's house *(e) doesn't matter*



UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre