Whats' beyond Concerto: An introduction to the R package *catR*

Session 4:

Overview of polytomous IRT models

The Psychometrics Centre, Cambridge, June 10th, 2014

Outline:

1. Introduction

Current version of *catR* is version 3.0 (available online)

Up to version 2.6, only dichotomous IRT models were embedded into *catR*

Version 3.0 holds now most known polytomous IRT models

Goal of this session:

1. To briefly describe the general framework for polytomous IRT models
2. To set up the notations and IRT models that are embedded in *catR*

2. General notations

With dichotomous items, only two responses are possible

With polytomous items, more than two responses are allowed

For item j, set $g_j + 1$ as the number of possible responses

Responses are coded as $k \in \{0, 1, ..., g_j\}$

Settiong $g_j = 1$ yields a dichotomous item

Item responses can be

- ordinal: e.g., ("never", "sometimes", "often", "always")
- nominal: e.g., color, political affiliation, etc.

$\boldsymbol{p}_j$ is the set of item parameters (depending on the model)

$P_{jk}(\theta) = Pr(X_j = k|\theta, \boldsymbol{p}_j)$ is the probability of answering response k to item j, given proficiency θ and item parameters

2. General notations

Two main categories of polytomous IRT models (Thissen & Steinberg, 1986):

1. Difference models: probabilities $P_{jk}(\theta)$ are set as differences between cumulative probabilities

$$P^*_{jk}(\theta) = Pr(X_j \geq k|\theta, \boldsymbol{p}_j)$$

2. Divide-by-total models: probabilities $P_{jk}(\theta)$ are set as ratios of values divided by the sum of these values:

$$P_{jk}(\theta) = \frac{t_{jk}}{\sum_{l=0}^{g_j} t_{jl}}$$

Notations to be used come from Embretson and Reise (2000)

2. General notations

Most-known difference models:

- Graded response model (GRM; Samejima, 1969)
- Modified graded response model (MGRM; Muraki, 1990)

Most-known difference models:

- Partial credit model (PCM; Masters, 1982)
- Generalized partial credit model (GPCM; Muraki, 1992)
- Rating scale model (RSM; Andrich, 1978)
- Nominal response model (NRM; Bock, 1972)

All these models are available in *catR*

3. Difference models

With difference models, probabilities $P_{jk}(\theta)$ are set as differences between cumulative probabilities

$$P^*_{jk}(\theta) = Pr(X_j \geq k|\theta, \boldsymbol{p}_j)$$

that is, the probability of selecting response in $(k,\ k+1,\ ...,\ g_j)$

Graded response model (GRM; Samejima, 1969):

$$P^*_{jk}(\theta) = \frac{\exp\left[\alpha_j\left(\theta - \beta_{jk}\right)\right]}{1 + \exp\left[\alpha_j\left(\theta - \beta_{jk}\right)\right]}$$

with $P^*_{j0}(\theta) = 1$ and $P^*_{j,g_j}(\theta) = P_{j,g_j}(\theta)$

α_j discrimination (slope) parameter and β_{jk} threshold (intercept) parameters of cumulative probabilities

Only g_j thresholds β_{j1}, ..., β_{j,g_j} are necessary to distinguish the $g_j + 1$ response categories

3. Difference models

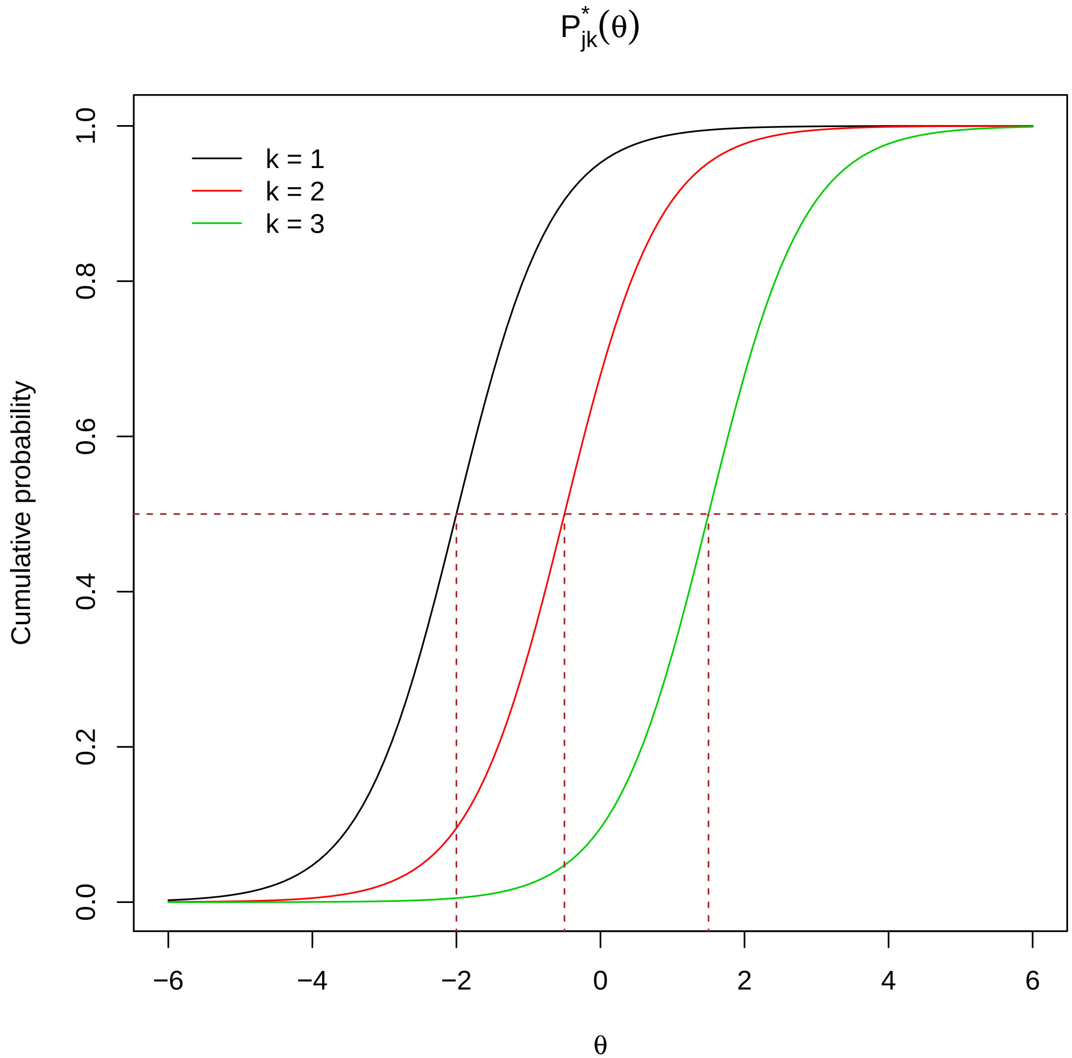
Response category probabilities $P_{jk}(\theta)$ are found back as follows (with $0 < k < g_j$):

$$
\begin{aligned}
P_{j,g_j}(\theta) &= P^*_{j,g_j}(\theta) \\
P_{jk}(\theta) &= P^*_{jk}(\theta) - P^*_{j,k+1}(\theta) \\
P_{j0}(\theta) &= 1 - P^*_{j1}(\theta)
\end{aligned}
$$

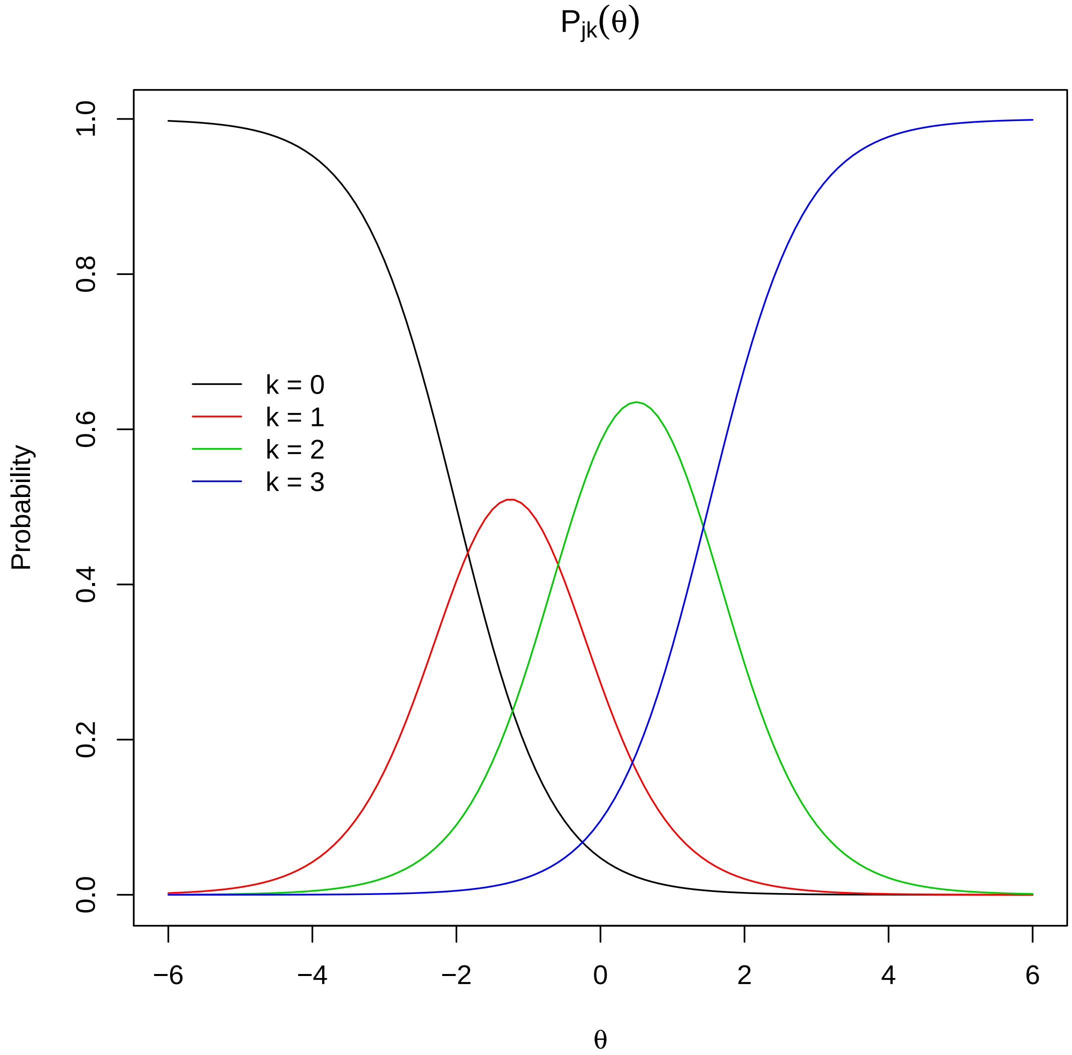
Illustration with

- 4 response categories (i.e. $g_j = 3$),
- $\alpha_j = 1.5$,
- $\beta_{j1} = -0.5$,
- $\beta_{j2} = 0.5$, and
- $\beta_{j3} = 1.2$

3. Difference models



3. Difference models



3. Difference models

With GRM, threshold parameters β_{jk} may vary across items

$\Rightarrow$ items may not have the same number of response categories g_j

Modified graded response model (MGRM; Muraki, 1990): modification of GRM to allow for common thresholds across all items

With MGRM, thresholds β_{jk} are split in two components: b_j (general intercept parameter for item j) and c_k (threshold parameter between categories k and $k+1$, common to all items)

Cumulative probability under MGRM:

$$P^*_{jk}(\theta) = \frac{\exp\left[\alpha_j\left(\theta - b_j + c_k\right)\right]}{1 + \exp\left[\alpha_j\left(\theta - b_j + c_k\right)\right]}$$

3. Difference models

With GRM, all items have the same number of categories (since all c_k parameters are equal across items)

MGRM applicable for Likert scales with same response categories across items, such as e.g.

Never - Rarely - Sometimes - Often - Always

4. Divide-by-total models

With divide-by-total models, probabilities $P_{jk}(\theta)$ are set directly as ratios of values divided by the sum of these values (across response categories)

Partial credit model (PCM; Masters, 1982):

$$P_{jk}(\theta) = \frac{\exp \sum_{t=0}^{k} (\theta - \delta_{jt})}{\sum_{r=0}^{g_j} \exp \sum_{t=0}^{r} (\theta - \delta_{jt})}$$

with

$$\sum_{t=0}^{0} (\theta - \delta_{jt}) = 0 \quad \text{or} \quad \exp \sum_{t=0}^{0} (\theta - \delta_{jt}) = 1$$

4. Divide-by-total models

Detailed equations with three response categories (i.e. $g_j = 2$ and $k \in \{0, 1, 2\}$):

$$\begin{aligned}\sum_{r=0}^{g_j} \exp \sum_{t=0}^{r} (\theta - \delta_{jt}) &= \sum_{r=0}^{2} \exp \sum_{t=0}^{r} (\theta - \delta_{jt}) \\ &= 1 \\ &\quad + \exp\left(\theta - \delta_{j1}\right) \\ &\quad + \exp\left(\theta - \delta_{j1} + \theta - \delta_{j2}\right)\end{aligned}$$

Denominator is then equal to

$$\sum_{r=0}^{g_j} \exp \sum_{t=0}^{r} (\theta - \delta_{jt}) = 1 + \exp\left(\theta - \delta_{j1}\right) + \exp\left(2\,\theta - \delta_{j1} - \delta_{j2}\right)$$

4. Divide-by-total models

Response category probabilities are then

$$P_{j0}(\theta) = \frac{1}{1 + \exp\left(\theta - \delta_{j1}\right) + \exp\left(2\,\theta - \delta_{j1} - \delta_{j2}\right)},$$

$$P_{j1}(\theta) = \frac{\exp\left(\theta - \delta_{j1}\right)}{1 + \exp\left(\theta - \delta_{j1}\right) + \exp\left(2\,\theta - \delta_{j1} - \delta_{j2}\right)},$$

and

$$P_{j2}(\theta) = \frac{\exp\left(2\,\theta - \delta_{j1} - \delta_{j2}\right)}{1 + \exp\left(\theta - \delta_{j1}\right) + \exp\left(2\,\theta - \delta_{j1} - \delta_{j2}\right)},$$

Idea: to allow for partial credit in the item response (e.g., false - incomplete - almost correct - correct responses)

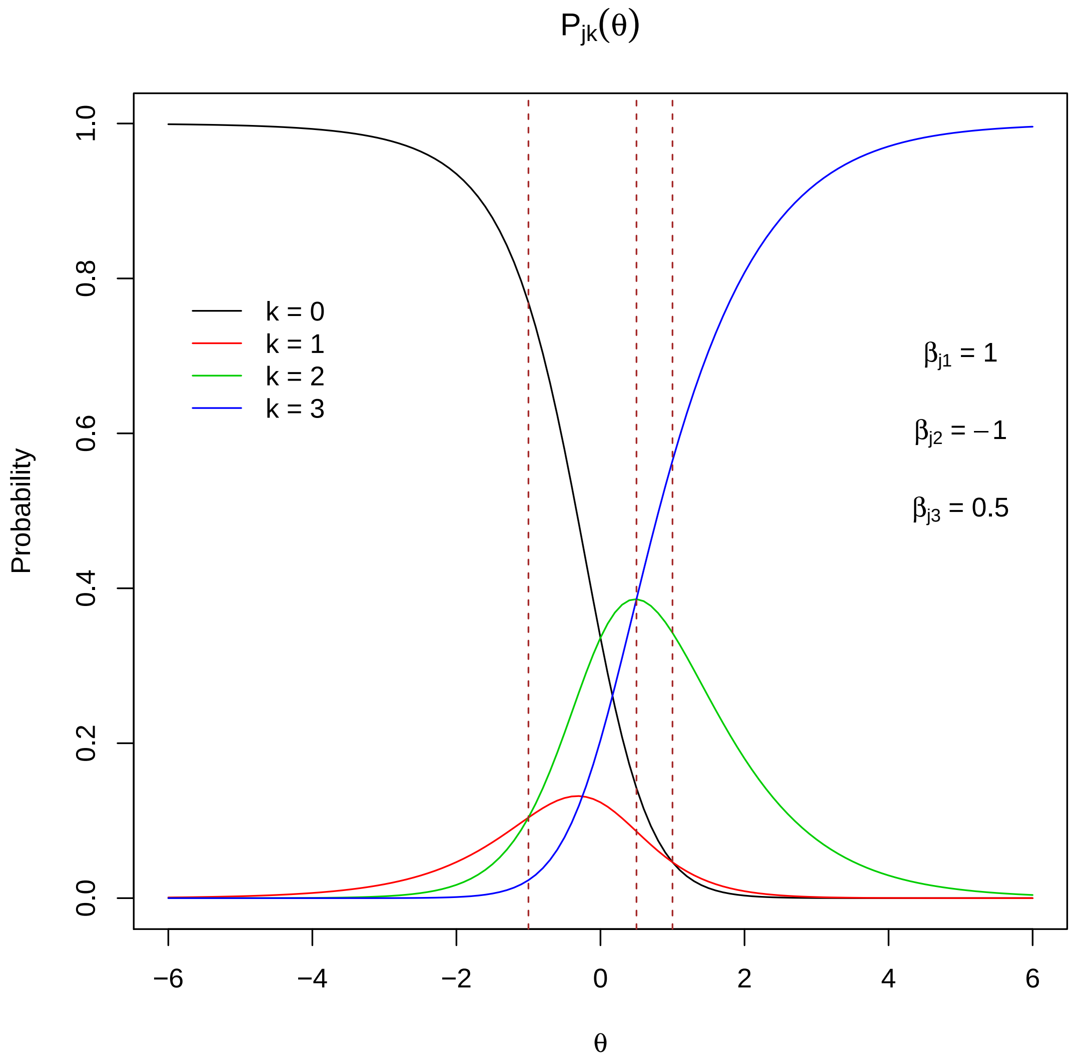
4. Divide-by-total models

Only g_j thresholds $(\delta_{j1}, ..., \delta_{j,g_j})$ are necessary

Illustration with

- $g_j = 3$ (i.e. four response categories)
- $(\delta_{j1}, \delta_{j2}, \delta_{j3}) = (1, -1, 0.5)$

4. Divide-by-total models



4. Divide-by-total models

Generalized partial credit model (GPCM; Muraki, 1992) extends the PCM by introducing a discrimination parameter α_j for the item:

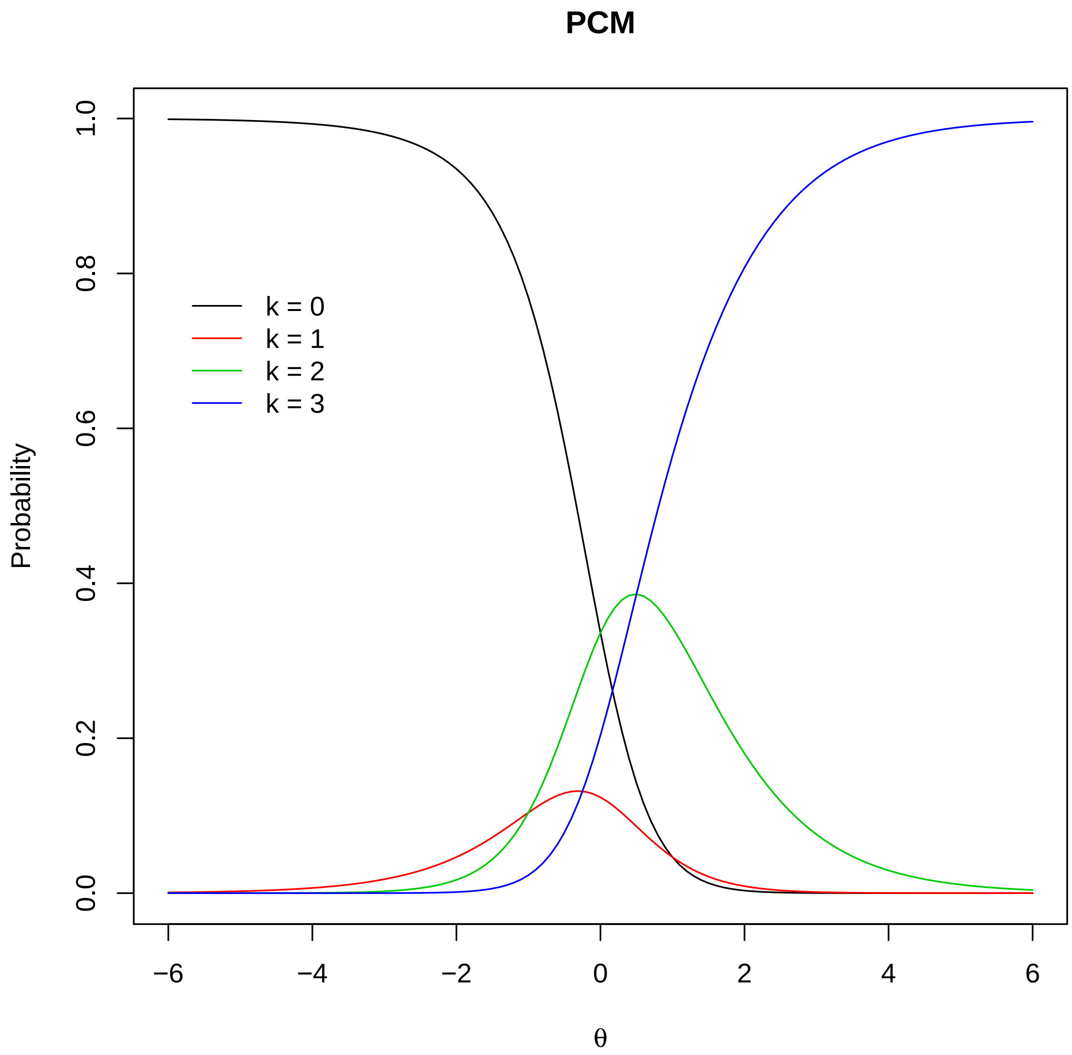
$$P_{jk}(\theta) = \frac{\exp \sum_{t=0}^{k} \alpha_j (\theta - \delta_{jt})}{\sum_{r=0}^{g_j} \exp \sum_{t=0}^{r} \alpha_j (\theta - \delta_{jt})}$$

with

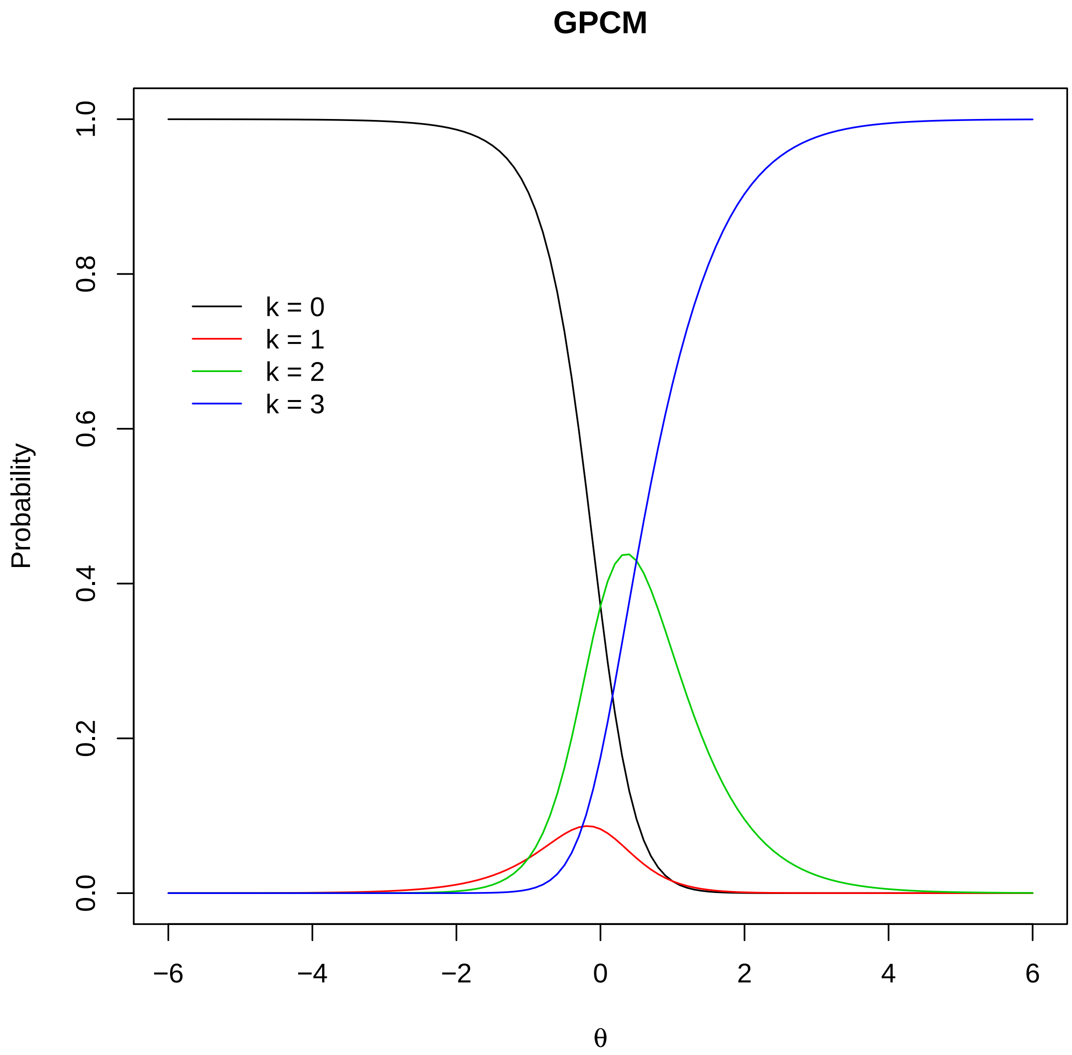
$$\sum_{t=0}^{0} \alpha_j (\theta - \delta_{jt}) = 0$$

Illustration with same parameters as PCM, i.e. $(\delta_{j1}, \delta_{j2}, \delta_{j3}) = (1, -1, 0.5)$, and with $\alpha_j = 1.5$

4. Divide-by-total models



4. Divide-by-total models



4. Divide-by-total models

PCM allows for different thresholds δ_{jt} across the items, and also different response categories

Rating scale model (RSM; Andrich, 1978) is a modification of PCM to allow equal thresholds across items

With RSM, thresholds δ_{jt} are split in two components, λ_j (general intercept parameter) and δ_t (threshold parameter common to all items):

$$P_{jk}(\theta) = \frac{\exp \sum_{t=0}^{k} [\theta - (\lambda_j + \delta_t)]}{\sum_{r=0}^{g_j} \exp \sum_{t=0}^{r} [\theta - (\lambda_j + \delta_t)]}$$

with

$$\sum_{t=0}^{0} [\theta - (\lambda_j + \delta_t)] = 0$$

4. Divide-by-total models

Finally, Nominal response model (NRM; Bock, 1972) specifies response category probabilities as follows:

$$P_{jk}(\theta) = \frac{\exp\left(\alpha_{jk}\,\theta + c_{jk}\right)}{\sum_{r=0}^{g_j} \exp\left(\alpha_{jr}\,\theta + c_{jr}\right)}$$

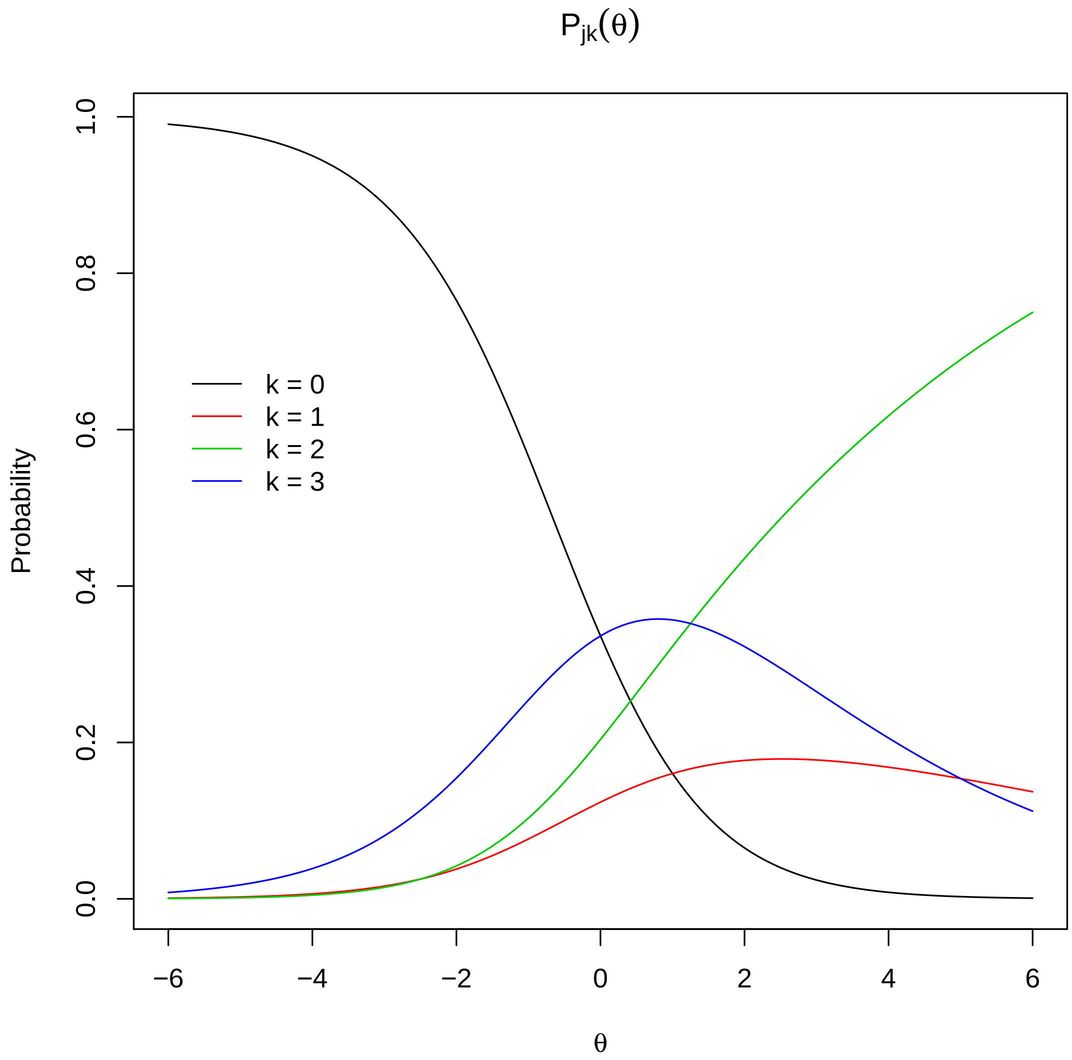
with

$$\alpha_{j0}\,\theta + c_{j0} = 0$$

Illustration with four categories (i.e.$g_j = 3$) and

- $(\alpha_{j1}, c_{j1}) = (1, -1)$,
- $(\alpha_{j2}, c_{j2}) = (1.2, -0.5)$,
- $(\alpha_{j3}, c_{j3}) = (0.8, 0)$

4. Divide-by-total models



5. Summary

Polytomous IRT models extend dichotomous IRT models by allowing more than two possible responses

Response categories can be the same across all items or they may vary

Polytomous IRT models: GRM, MGRM, PCM, GPCM, RSM, NRM, ...

MGRM and RSM are convenient only if the same response categories (as they share common threshold parameters for all items)

Extensions of algorithms for dichotomous IRT model calibration to the polytomous framework exist

5. Summary

Each model can be fully described with specific sets of parameters for each item j:

- $(\alpha_j, \beta_{j1}, ..., \beta_{j,g_j})$ for GRM
- $(\alpha_j, b_j, c_1, ..., c_g)$ for MGRM
- $(\delta_{j1}, ..., \delta_{j,g_j})$ for PCM
- $(\alpha_j, \delta_{j1}, ..., \delta_{j,g_j})$ for GPCM
- $(\lambda_j, \delta_1, ..., \delta_g)$ for RSM
- $(\alpha_{j1}, c_{j1}, ..., \alpha_{j,g_j}, c_{j,g_j})$ for NRM

These orders of parameters are used in *catR*...

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561-573.

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29-51.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement, 14*, 59-71.

References

Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores.* Psychometrika monograph (Vol. 17).

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika, 51*, 567-577.