A Unified Framework for Visualization and Inference in Item Response Theory Models

Theoretical Background and Implementation in R

Basil Abou El-Komboz

A Unified Framework for Visualization and Inference in Item Response Theory Models

Theoretical Background and Implementation in R

Basil Abou El-Komboz

Master's thesis supervised by

Prof. Dr. Helmut Küchenhoff
Ludwig-Maximilians-Universität München

Prof. Dr. Achim Zeileis
Universität Innsbruck
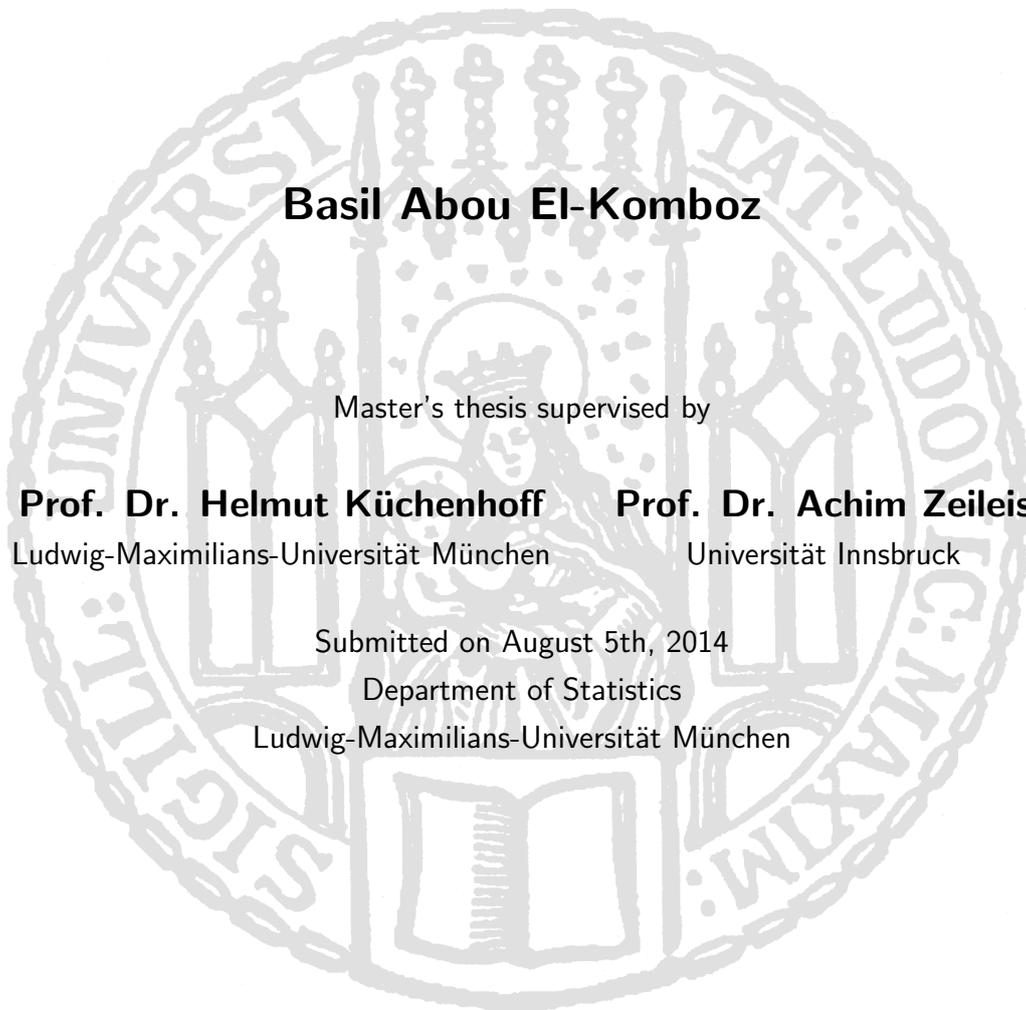
Acknowledgments

This Master's thesis was only possible with the support of several people whom I would like to thank here.

First of all, I would like to thank my supervisors Prof. Dr. Helmut Küchenhoff and Prof. Dr. Achim Zeileis. I am very grateful to Prof. Dr. Helmut Küchenhoff for offering his supervision and therefore making this work possible. His benevolent feedback encouraged me and increased the quality of this thesis further. I am very grateful to Prof. Dr. Achim Zeileis for his excellent mentoring. The continuous and immediate feedback and the personal discussions positively shaped this thesis and also my view on other things.

I am indebted to Jessica Hey for her valuable comments while proofreading this thesis.

Last but not least, I would like to thank my friends and my family for helping me in balancing work and life.

Summary

A unified framework for visualization and inference in item response theory (IRT) models is developed within this Master's thesis and implemented in the R package psychotools (Zeileis et al., 2014).

For this purpose, a theoretical framework is established in a first step by introducing the generalized partial credit model (GPCM, Muraki, 1992), one of the most general parametric IRT models for polytomous items. In addition, the relations to several other popular IRT models, existing parametrizations and parameter estimation approaches as well as the issue of parameter identifiability are discussed. In a second step, four contextually different structural components of IRT models are identified based on the GPCM: Person parameters, item discrimination parameters, item location parameters and absolute or relative item threshold parameters. For each of these structural components, a suitable representation is developed and implemented in the R package psychotools. Starting with the estimated parameters of the IRT models already implemented in the package, the computation of each structural component and its variance-covariance matrix are derived and additionally implemented. In a third step, several established visualization techniques and tools for inference are implemented on top of the structural components thus providing a unified framework for visualization and inference in IRT models.

Advantages and possibilities of the provided framework such as numerical and graphical model comparisons, model selection and hypotheses tests as well as applications when detecting differential item functioning are illustrated in several examples. Limitations and directions for further research are pointed out in the final discussion of this Master's thesis.

Contents

List of Figures

List of Tables

List of Code Segments

1. Introduction

In psychological testing, a series of items is typically administered to subjects to measure psychological, i.e., non-observable constructs like abilities or attitudes. The results of such assessments are then used in a variety of situations like "screen[ing] applicants for jobs [..., ...] counsel[ing ...] individuals for educational, vocational, and personal counseling purposes, [... or ...] diagnos[ing] and prescrib[ing] psychological and physical treatments in clinics and hospitals [...]." (Aiken, 1994, p. 11–12).

Item response theory (IRT) and the various statistical models subsumed under this theoretical framework provide means to develop, assess and validate the items used in psychological testing in the first place. This is done by probabilistic modeling of the subjects' responses to the administered items as a function of characteristics of the items and the subjects. The parameter estimates and various test statistics of a fitted IRT model then allow conclusions about the properties of the administered items as well as the attitudes or abilities of the tested subjects. Depending on the specific formulation of item and subject characteristics, the type of response observed and whether or not a parametric form of the response curve is assumed, a variety of different IRT models can be distinguished (for an overview see, e.g., Fischer & Molenaar, 1995; Van der Linden & Hambleton, 1997).

To carry out an IRT analysis, several add-on packages for the R system for statistical computing (R Core Team, 2014) exist. An up-to-date overview can be found on the Comprehensive R Archive Network task view (Mair, 2014) on Psychometric Models and Methods. A recent review about available packages which are accompanied by a peer-reviewed publication is given by Rusch et al. (2013). One well-known package for "computing Rasch models [i.e., a certain class of IRT models,] and several extensions" (Mair & Hatzinger, 2007, p. 1) is the R package eRm (Mair et al., 2014). It provides the functionality to fit a series of related IRT models based on an "unified CML [i.e., conditional maximum likelihood] approach" (Mair & Hatzinger, 2007, p. 4). While this approach is elegant as a general framework is established, the resulting functions to compute the various IRT models are rather slow due to the computational overhead. In addition, the framework used only comprises a small number of IRT models and cannot

be easily extended. The R package psychotools (Zeileis et al., 2014) on the other hand follows a different approach: It provides fast and highly specialized implementations of various IRT models. The model-fitting functions are specifically designed for a certain IRT model and can be used as building blocks for further, more complex, psychometric methods like Rasch trees (Strobl et al., 2013) or Rasch mixtures (Frick et al., 2012). While this approach avoids the computational overhead which is present in more general approaches, a unified theoretical and computational framework for tasks like inference or visualizing a fitted IRT model is missing.

The motivation of this Master's thesis is to fill this gap in the R package psychotools and thereby make a synthesis between the two approaches illustrated above, i.e., a top-down approach like in the R package eRm which provides an elegant but limited framework and a bottom-up approach like in the R package psychotools which (so far) only provides fast and highly specialized model-fitting functions but no unified theoretical and computational framework. To achieve the aforementioned goal, the theoretical background and the generalized partial credit model by Muraki (1992) are first introduced in Chapter 2. In Chapter 3, the unifying structural components of different IRT models are identified based on the generalized partial credit model as theoretical framework and a suitable representation of these components is developed and implemented in the R package psychotools. Based on these structural components, tools for visualization (Chapter 4) and inference (Chapter 5) are developed and also implemented in the R package psychotools. Overall, a theoretical and computational framework for visualization and inference in IRT models is provided which is detached from a specific model and can be easily extended. Throughout this Master's thesis, the usage and the advantages of the provided framework are illustrated within several application examples. The data set used in these application examples is introduced in more detail in the following.

The Verbal Aggression Data

The example data set used in this Master's thesis was collected by Vansteelandt (2000) in a study on verbally aggressive behaviors. It consists of the responses of 316 subjects to 24 items. In each item, one out of four frustrating situations was presented to a subject which was asked to judge on a three point likert scale with response categories "yes", "perhaps" and "no" whether it would react with a specific verbal aggressive behavior in the given situation. The four frustrating situations used were missing a bus because it fails to stop, missing a train because a clerk gave faulty information, standing in front of a

grocery store which just closed when one was about to enter and being disconnected in a phone call by the operator after the last ten cents have been used up. The possible verbal aggressive behaviors resulted from a combination of two behavioral modes (wanting or doing something) and three verbally aggressive responses (cursing, scolding or shouting). The factorial combination of the two behavioral modes, the three verbally aggressive responses and the four frustrating situations make up the given 24 items. In the following, a version of this data set available in the R package psychotools is used. In addition to the responses on the three point likert scale, this data set contains a dichotomized version of the responses with the categories "yes" and "perhaps" merged together. In a first step, the R package and the data set are loaded in the following:

```
> library("psychotools")
> data("VerbalAggression", package = "psychotools")
```

In a second step, a subset of the verbal aggression data consisting only of the items 13-18 is extracted and stored in a R object named dat. Within this object, the dichotomized responses are stored in an element named dich and the original responses on the three point likert scale are stored in an element named poly:

```
> dat <- data.frame(dich = rep(NA, nrow(VerbalAggression)),
+                    poly = rep(NA, nrow(VerbalAggression)))
> dat$dich <- VerbalAggression$resp2[, 13:18]
> dat$poly <- VerbalAggression$resp[, 13:18]
```

The items of this subset all contain the same frustrating situation of standing in front of a grocery store which just closed when one was about to enter and will be used in the following illustrations. In addition, the gender and an anger score measuring the momentarily anger of each subject is extracted and stored in corresponding elements named gender and anger:

```
> dat$gender <- VerbalAggression$gender
> dat$anger <- VerbalAggression$anger
```

In a last step, custom labels describing the behavioral mode and the verbal aggressive response posed within an item are set for readability:

```
> lbs <- c("Want-Curse", "Do-Curse",
+   "Want-Scold", "Do-Scold", "Want-Shout", "Do-Shout")
> colnames(dat$dich) <- colnames(dat$poly) <- lbs
```

2. A Unified Framework: The Generalized Partial Credit Model

Focusing on parametric IRT models for ordered polytomous items, the *generalized partial credit model (GPCM)* by Muraki (1992) is one of the most general models. The GPCM,

$$P(X_{ij} = x_{ij}|\theta_i, \alpha_j, \boldsymbol{\delta}_j) = \frac{\exp\left[\sum_{k=1}^{x_{ij}} a_j(\theta_i - \delta_{jk})\right]}{\sum_{\ell=1}^{p_j} \exp\left[\sum_{k=1}^{\ell} a_j(\theta_i - \delta_{jk})\right]}, \tag{2.1}$$

describes the probability that subject i with ability θ_i chooses one of the p_j ordered categories of item j. Subjects are characterized by a single parameter θ_i in the GPCM, i.e., a uni-dimensional latent trait is assumed. Items however are characterized by two types of parameters: the item discrimination parameter α_j and the absolute item threshold parameters δ_{jk} (with $k = 1, \ldots, p_j$). While the item discrimination parameter α_j describes the steepness of the category response curves, i.e., the impact an increase in the latent trait has on the probability of choosing a certain category k on item j, the absolute item threshold parameters δ_{jk} indicate the locations on the ability axis where the probability of choosing category k is equal to the probability of choosing category $k-1$, i.e., the intersection of the category response curves of two adjacent categories. This is illustrated in Figure 2.1 where the category response curves, i.e., the predicted probabilities under the GPCM, for two items with three categories, item discrimination parameters $\alpha_1 = 1.3$ and $\alpha_2 = 1.6$ and absolute item threshold parameters $\boldsymbol{\delta}_1 = (-1.2, 1.2, 1.8)^\top$ and $\boldsymbol{\delta}_2 = (0.5, -1, 1.5)^\top$ are depicted.

As can be seen from Figure 2.1, even in the case of unordered absolute item threshold parameters $\boldsymbol{\delta}_j$ as in item two, the absolute item threshold parameters $\boldsymbol{\delta}_j$ still indicate the locations of the intersections of the category response curves of two adjacent categories. In this case, however, not every category has a region on the latent trait axis where this category is the single most probable category.
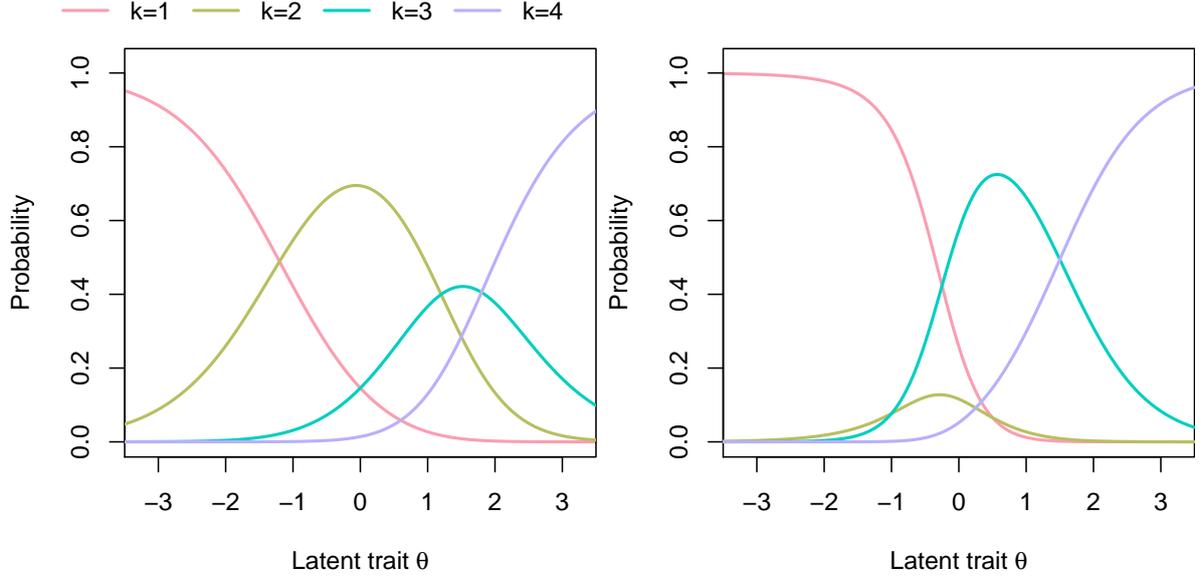
Figure 2.1. Category response curves under the GPCM for two polytomous items. Item 1 (left figure) has characteristics $\alpha_1 = 1.3$ and $\boldsymbol{\delta}_1 = (-1.2, 1.2, 1.8)^\top$, item 2 (right figure) has characteristics $\alpha_2 = 1.6$ and $\boldsymbol{\delta}_2 = (0.5, -1, 1.5)^\top$.

2.1. Related Models and Other Parametrizations

With certain restrictions on the parameters of the GPCM from Equation (2.1), several popular IRT models result as special cases of this very general model. For polytomous items, the partial credit model by Masters (PCM, 1982),

$$P(X_{ij} = x_{ij} | \theta_i, \boldsymbol{\delta}_j) = \frac{\exp\left[\sum_{k=1}^{x_{ij}} (\theta_i - \delta_{jk})\right]}{\sum_{\ell=1}^{p_j} \exp\left[\sum_{k=1}^{\ell} (\theta_i - \delta_{jk})\right]}, \tag{2.2}$$

results as special case when the item discrimination parameters α_j are restricted to one. For dichotomous items, the GPCM from Equation (2.1) specializes to the 2-PL or Birnbaum model (Birnbaum, 1968). If additionally the item discrimination parameters α_j are restricted to one, the popular 1-PL or dichotomous Rasch model (Rasch, 1960) results.

Sometimes, the absolute item threshold parameters δ_{jk} of the GPCM from Equation (2.1) or the PCM from Equation (2.2) are reparametrized into an item-specific
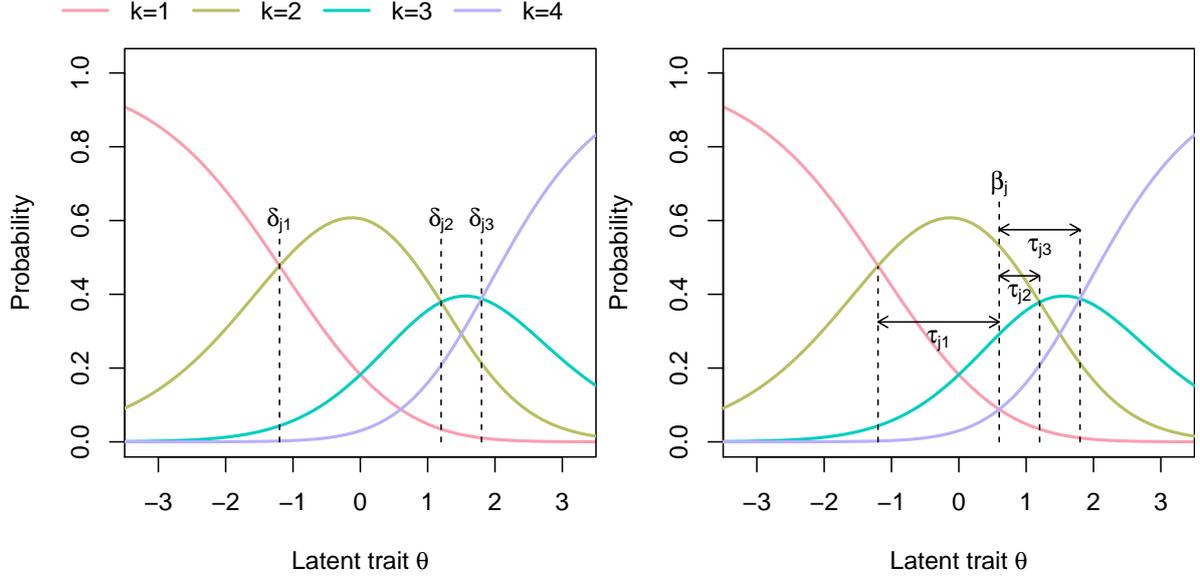
Figure 2.2. Illustration of the absolute (left figure) and relative (right figure) item threshold parametrization in the GPCM for an item j with four categories.

location parameter β_j,

$$\beta_j = \frac{1}{p_j} \sum_{k=1}^{p_j} \delta_{jk}, \qquad (2.3)$$

and several category-specific "relative" item threshold parameters τ_{jk},

$$\tau_{jk} = \delta_{jk} - \beta_j. \qquad (2.4)$$

This reparametrization, labeled as "rating [scale] formulation" by Van der Ark (2001, p. 275), is illustrated along with the usual parametrization in Figure 2.2. Whereas the absolute item threshold parameters δ_{jk} indicate the "absolute" location of the intersection of two adjacent categories on the theta axis (left figure), the relative item threshold parameters τ_{jk} indicate the location of this intersection "relative" to the item location parameter β_j, which describes the "center" of an item on the theta axis (right figure).

A very similar but slightly different parametrization has been suggested by Andrich (1978) as rating scale model (RSM),

$$P(X_{ij} = x_{ij} | \theta_i, \beta_j, \boldsymbol{\tau}) = \frac{\exp\left[\sum_{k=1}^{x_{ij}} (\theta_i - (\beta_j + \tau_k))\right]}{\sum_{\ell=1}^{p} \exp\left[\sum_{k=1}^{\ell} (\theta_i - (\beta_j + \tau_k))\right]}, \qquad (2.5)$$

and similarly for the GPCM by Muraki (1992, p. 164) as GPCM-RSM,

$$P(X_{ij} = x_{ij}|\theta_i, \alpha_j, \beta_j, \boldsymbol{\tau}) = \frac{\exp\left[\sum_{k=1}^{x_{ij}} a_j\left(\theta_i - (\beta_j + \tau_k)\right)\right]}{\sum_{\ell=1}^{p_j} \exp\left[\sum_{k=1}^{\ell} a_j\left(\theta_i - (\beta_j + \tau_k)\right)\right]}. \tag{2.6}$$

The subtle difference between the reparametrization discussed before and the parametrization in the RSM from Equation (2.5) and the GPCM-RSM from Equation (2.6) is that the relative item threshold parameters τ_k are assumed to be identical over the items in the latter two models, while they can vary between the items in the "rating [scale] formulation" (Van der Ark, 2001, p. 275) from Equation (2.4).

Two other parametrizations exist which are less frequently used and which have no such intuitive interpretation as the two parametrizations discussed above, but should nevertheless briefly be mentioned here. Both parametrizations can be seen as "cumulative" parametrizations. The first is cumulative in the absolute item threshold parameters δ_{jk}. Instead of the absolute item threshold parameters δ_{jk}, cumulative absolute item threshold parameters,

$$\eta_{j\ell} = \sum_{k=1}^{\ell} \delta_{jk}, \tag{2.7}$$

are considered. This parametrization is discussed in, e.g., Andersen (1977); Wilson & Masters (1993); Fischer & Ponocny (1994); Adams et al. (2012) and the cumulative absolute item threshold parameters η_{jk} are sometimes called item-category parameters (see, e.g. Punzo, 2008, p. 5).

The second "cumulative parametrization" is cumulative in the relative item threshold parameters τ_k (RSM) or τ_{jk} (G/PCM). Similar to the absolute item threshold parameters δ_{jk} and the item-catergory parameters η_{jk}, a cumulative parametrization,

$$\kappa_{(j)\ell} = \sum_{k=1}^{\ell} \tau_{(j)k} \tag{2.8}$$

exists here too (see, e.g., Andrich, 1978; Fischer & Molenaar, 1995). Table A.1 in Appendix A gives an overview of the different parametrizations discussed above and their relations to each other.

2.2. Parameter Estimation

Several approaches have been suggested to estimate the person and item parameters in IRT models (for an overview see, e.g., Baker & Kim, 2004). In the following, three frequently used likelihood-based approaches are briefly illustrated and the issue of parameter identifiability is discussed. Under the assumption of independence of the responses of a given subject to different items (local independence) and the assumption of independence of the responses of different subjects, the joint likelihood of the GPCM is given as

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\theta} | \boldsymbol{x}) &= \prod_{i=1}^{n} \prod_{j=1}^{m} P(x_{ij} | \theta_i, \alpha_j, \boldsymbol{\delta}_j) \\
&= \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{\exp\left[\sum_{k=1}^{x_{ij}} a_j(\theta_i - \delta_{jk})\right]}{\sum_{\ell=1}^{p_j} \exp\left[\sum_{k=1}^{\ell} a_j(\theta_i - \delta_{jk})\right]} \\
&= \frac{\exp\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{x_{ij}} a_j(\theta_i - \delta_{jk})\right]}{\prod_{i=1}^{n}\prod_{j=1}^{m}\sum_{\ell=1}^{p_j} \exp\left[\sum_{k=1}^{\ell} a_j(\theta_i - \delta_{jk})\right]}.
\end{aligned} \tag{2.9}
$$

In the joint maximum likelihood approach (JML), the joint likelihood given in Equation (2.9) is maximized and both types of parameters are estimated together. As the number of person parameters to estimate increases with the sample size, a major drawback of this approach is that the item parameter estimators are not consistent (Andersen, 1973; Gosh, 1995).

To overcome this problem, a distribution $f(\theta_i | \boldsymbol{\xi})$ is assumed for the person parameters in the marginal maximum likelihood approach (MML). Hence, instead of viewing them as fixed quantities as in the JML approach, the person parameters are regarded as random quantities coming from a certain distribution f with parameters $\boldsymbol{\xi}$ (if a parametric distribution is assumed, see below). In a second step, the marginal likelihood is computed by integrating over the the person parameters. For the GPCM from Equation (2.1) with the joint likelihood from Equation (2.9) the marginal likelihood follows as

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\xi} | \boldsymbol{x}) = \prod_{i=1}^{n} \int \prod_{j=1}^{m} P(x_{ij} | \theta_i, \alpha_j, \boldsymbol{\delta}_j) f(\theta_i | \boldsymbol{\xi}) d\theta_i$$

$$= \prod_{i=1}^{n} \int \prod_{j=1}^{m} \frac{\exp \left[\sum_{k=1}^{x_{ij}} a_j(\theta_i - \delta_{jk}) \right]}{\sum_{\ell=1}^{p_j} \exp \left[\sum_{k=1}^{\ell} a_j(\theta_i - \delta_{jk}) \right]} f(\theta_i | \boldsymbol{\xi}) d\theta_i$$

$$= \prod_{i=1}^{n} \int \frac{\exp \left[\sum_{j=1}^{m} \sum_{k=1}^{x_{ij}} a_j(\theta_i - \delta_{jk}) \right]}{\prod_{j=1}^{m} \sum_{\ell=1}^{p_j} \exp \left[\sum_{k=1}^{\ell} a_j(\theta_i - \delta_{jk}) \right]} f(\theta_i | \boldsymbol{\xi}) d\theta_i. \qquad (2.10)$$

The marginal likelihood is then maximized with respect to the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\delta}$ and $\boldsymbol{\xi}$. Typically, a parametric distribution $f(\theta | \boldsymbol{\xi})$ is assumed for the person parameters (for example a standard-normal distribution as in, e.g., Bock & Lieberman, 1970; Bock & Aitkin, 1981) but there are also suggestions for non- or semi-parametric approaches (see, e.g., Heinen, 1996). The disadvantages of the MML approach are on the one hand the necessity and possible misspecification of the distributional assumption for the person parameters θ_i, and, on the other hand, that no analytical solution exists for the integral in the marginal maximum likelihood function and hence numerical approximations have to be used when maximizing the marginal likelihood.

In addition to the JML and MML approach, Andersen (1972) suggested a conditional maximum likelihood approach (CML). In this approach, the person parameters are eliminated out of the joint likelihood by conditioning on the sum scores $s_i = \sum_{j=1}^{m} x_{ij}$ as sufficient statistic for the person parameter. The resulting conditional likelihood is then maximized with respect to the item parameters. It can be shown that under certain regularity conditions, the resulting item parameter estimators are "asymptotically efficient, and [...] the loss of information [by conditioning on the total scores of the persons] becomes negligible when [... the sample size approaches infinity]" (Molenaar, 1995, p. 47). The person parameters are then estimated in a second step by plugging the estimated item parameters into the joint likelihood and maximizing it with respect to the person parameters. A disadvantage of this approach is that the uncertainty associated with the estimation of the item parameters is not considered when the person parameters are estimated. Another disadvantage is that this approach is only applicable to models

where a sufficient statistic for the person parameter exists. Especially for models where the item discrimination parameter α_j is not fixed to unity, like in the GPCM, this is not the case. For these models, a MML approach is recommended (see, e.g., Baker & Kim, 2004).

In the R package psychotools, a CML approach is implemented for fitting dichotomous Rasch, rating scale and partial credit models. For each of these three models, a highly-specialized model-fitting function exists which is called in the following based on the previously extracted subset of the verbal aggression data:

```
> rmmod <- raschmodel(dat$dich)
> rsmod <- rsmodel(dat$poly)
> pcmod <- pcmodel(dat$poly)
```

In the next section, the issue of parameter identifiability is discussed in more detail.

2.3. Parameter Identifiability

The item and person parameters in IRT models are typically not identifiable, i.e., no unique solution exists when the joint, conditional or marginal likelihood is maximized to estimate the parameters (for a more formal definition of parameter identifiability, see, e.g., Casella & Berger, 2001, p. 24). This can be easily seen, e.g., in the dichotomous Rasch model,

$$P(X_{ij} = x_{ij}|\theta_i, \delta_j) = \frac{\exp(\theta_i - \delta_j)}{1 + \exp(\theta_i - \delta_j)}.$$

If one replaces the parameters θ_i and δ_j with parameters $\tilde{\theta}_i = \theta_i + c$ and $\tilde{\delta}_j = \delta_j - c$ where c is an arbitrary selected constant c, the probability distribution described by the model remains the same. Hence, the parameters in the dichotomous Rasch model are only identified up to a certain constant c and some restriction is necessary to ensure that a unique solution exists when estimating the parameters of the model. The exact definition of the necessary restriction depends on which IRT model and estimation technique is used precisely. In the following, the restriction(s) necessary in the Rasch, RSM and PCM, estimated via the conditional maximum likelihood approach as implemented in the R package psychotools are discussed in more detail. As shown above, the parameters in the dichotomous Rasch model are only identified up to a constant, i.e., a single restriction is necessary which fixes the origin of the scale (for more details, see Fischer, 1981). In the CML approach, this means that typically one or more of the m absolute

item threshold parameters δ_j are fixed at a certain value. In the implementation of the dichotomous Rasch model in the R package psychotools, it is the first absolute item threshold parameter δ_1 which is set to zero. But this is arbitrary and as Eggen & Verhelst (2006) discussed, any restriction of the form

$$d_0 + \sum_{j=1}^{m} d_j \delta_j = 0 \quad \text{with } d_j \in \mathbb{R} \text{ and } \sum_{j=1}^{m} d_j \neq 0 \qquad (2.11)$$

could be used to ensure parameter identifiability in the dichotomous Rasch model. As Kopf et al. (2013, p. 4) pointed out, the general form depicted in Equation (2.11) includes all of the typically used restrictions in the dichotomous Rasch model. For example, the restriction used in the model-fitting function `raschmodel()` is a special case and results with $d_0 \equiv 0$ and $\boldsymbol{d} = (d_1, d_2, \ldots, d_m)^T = (1, 0, 0, \ldots, 0)^\top$.

Similar as in the dichotomous Rasch model, only $\sum_{j=1}^{m}(p_j - 1) - 1$ of the $\sum_{j=1}^{m}(p_j - 1)$ absolute item threshold parameters δ_{jk} can be freely estimated in the PCM depicted in Equation (2.2) and it is again a single restriction necessary such that a set of unique parameter estimates exists. In the implementation of the PCM in the R package psychotools, a CML approach is used which estimates the cumulative absolute item threshold parameters η_{jk} (see Section 2.1). As for the absolute item threshold parameters δ_{jk}, a single restriction is necessary and it is again the first cumulative absolute item threshold parameter η_{11} which is set to zero to ensure parameter identifiability, but this is, of course, again arbitrary and other restrictions could be used instead.

In the RSM from Equation (2.5) with its two types of item parameters, two restrictions are necessary, one for the item location parameters β_j and one for the relative item threshold parameters τ_k. If only one restriction is used, e.g., for the item location parameters β_j, a shift of the person parameters θ_i by a constant c can be still captured by a similar shift of the relative item threshold parameters τ_k. Hence, only $(m-1)+(p-2)$ item parameters can be freely estimated in the RSM (with p as the number of categories). In the implemented CML approach in the R package psychotools, it is the first item-specific parameter (the term item location parameter is avoided here as the estimated parameters cannot be interpreted as the center of an item due to the cumulative relative item threshold parametrization used in the implementation in the function `rsmodel()`) and the first cumulative relative item threshold parameter κ_1 which is set to zero, but this is, of course, again arbitrary and other restrictions could be used instead.

The relation of the two restrictions necessary in the RSM and the single restriction in the dichotomous Rasch model or the PCM can be seen by the following reformulation

of the number of free parameters in the PCM:

$$\underbrace{\sum_{j=1}^{m}(p_j - 1) - 1}_{\text{Free absolute item threshold parameters}} = \sum_{j=1}^{m}(p_j - 1) - m + m - 1 \tag{2.12}$$

$$= (m - 1) + \sum_{j=1}^{m}(p_j - 1) - m$$

$$= \underbrace{(m - 1)}_{\text{Free item-specific parameters}} + \underbrace{\sum_{j=1}^{m}(p_j - 2)}_{\text{Free relative item threshold parameters}}.$$

This reformulation is interesting in two aspects. First, it relates two perspectives to each other which are present in all three models and are also reflected by the different parametrizations mentioned in Section 2.1, but which are rarely discussed explicitly: An absolute perspective reflected by the absolute item threshold parameters δ_{jk} and a relative perspective reflected by the relative item threshold parameters τ_{jk}. While the RSM in Equation (2.5) is formulated in the relative perspective and the PCM in Equation (2.2) is formulated in the absolute perspective, the transformations shown in Table A.1 easily allow to convert the parameters from one perspective to the parameters in another perspective. These two perspectives can be also found in the dichotomous Rasch model. As there are two categories per item, only a single relative item threshold parameter τ_1 exists which is always (implicitly) set to zero. Under this condition, the absolute item threshold parameters δ_j and the item location parameters β_j are equivalent in this model.

The second interesting aspect to note in the reformulation shown in Equation (2.12) is that in the relative item threshold parametrization, the single restriction placed upon the relative item threshold parameters in the RSM generalizes to m restrictions in the PCM, i.e., one for each set of item-specific relative threshold parameters $\boldsymbol{\tau_j} = (\tau_{j1}, \tau_{j2}, \ldots, \tau_{jp_j})^\top$.

12

3. Structural Components and Their Implementation

In this chapter, the unifying structural components of different IRT models are first identified based on the GPCM as one of the most general parametric IRT models for ordered polytomous items. For each identified structural component, a suitable representation and implementation in the R package psychotools is then discussed in more detail. Overall, a unified framework for a wide class of IRT models is constructed and implemented for the three IRT models of the R package psychotools. Several application examples of this framework are illustrated in the following and in addition in the Chapters 4 and 5.

3.1. Structural Components of the GPCM and Related IRT Models

The two main "ingredients" of the GPCM (and hence of all the related IRT models, see Section 2.1) are the subjects, characterized by the person parameters, and the items, characterized by the item discrimination parameters and either by a set of (cumulative) absolute item threshold parameters or by a single item location parameter and a set of (cumulative) relative item threshold parameters (see Section 2.1 and Table A.1 for more details on the different parametrizations). As discussed in Chapter 2, the item discrimination parameter modulates the impact of the latent trait on the probability of choosing a certain category k on item j while the item location parameter as well as the item threshold parameters ([cumulative] absolute or [cumulative] relative) characterize positions on the latent trait axis. Because of these contextual differences, the item discrimination parameters will be viewed and represented separately in the following. In addition, the item location parameters will also be viewed and represented separately from the item threshold parameters (either [cumulative] absolute or [cumulative] relative) as they provide a characterization of an item by a single parameter in contrast to the item threshold parameters which characterize an item in more detail.

Based on these considerations, four structural components will be differentiated from

Table 3.1. The four contextually differentiated structural components of the GPCM (and hence of all related IRT models) and their suggested corresponding R classes.

Description	Formal representation	Corresponding R class
Person parameters	θ_i	`personpar`
Item discrimination parameters	α_j	`discrpar`
Item location parameters	β_j	`itempar`
Item threshold parameters	δ_{jk}, η_{jk}, $\tau_{(j)k}$ or $\kappa_{(j)k}$	`threshpar`

here on: Person parameters, item discrimination parameters, item location parameters and item threshold parameters in one of the four forms discussed in Section 2.1. Each of the four structural components will be represented by a R class together with a generic function of the same name to extract a particular component of a fitted model object. Table 3.1 summarizes the outlined framework. The implementation of this framework in the R package psychotools will be discussed in the following.

3.2. An Implementation in the R Package psychotools

Until now, three IRT models are implemented in the R package psychotools: The dichotomous RM, the RSM and the PCM. All three IRT models are special cases of the GPCM and are estimated with the CML approach (see Chapter 2 for more details) through highly-specialized model-fitting functions `raschmodel()`, `rsmodel()` and `pcmodel()`. To implement the unified framework outlined above in the R package psychotools, the R classes for each structural component are implemented in a first step. In a second step, specific methods for the generic functions of these classes are implemented for each of the three IRT models. These methods allow the extraction of a structural component from a fitted RM, RSM or PCM object. The structure of the four implemented classes as well as some details on the implemented model-specific extractor methods are discussed in the following separately for each of the four structural components.

3.2.1. Person Parameters

Person parameters $\boldsymbol{\theta}$ are represented by the R class `personpar`. Objects of this class consist of a named numeric vector with the estimated person parameters $\hat{\boldsymbol{\theta}}$ of a dichotomous RM, RSM or PCM. In addition, a label referring to the underlying IRT model and, if requested, the estimated variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$ are attached as attributes. The interface of the generic function `personpar()` is shown in code segment 3.1. Besides

```
personpar(object, ref = NULL, vcov = TRUE,
  start = NULL, tol = 1e-6, ...)
```

Code segment 3.1. Interface of the generic function `personpar()`.

a fitted model object of class `raschmodel`, `rsmodel` or `pcmodel`, a restriction for the estimated absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ can be specified in the argument `ref` (see below for more details). With the argument `vcov`, the estimation of the variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$ of the estimated person parameters $\hat{\boldsymbol{\theta}}$ can be turned on or off. With the arguments `start` and `tol`, starting values and the precision when estimating the person parameters $\boldsymbol{\theta}$ can be specified.

As discussed in Chapter 2, under a CML approach, the person parameters $\boldsymbol{\theta}$ are typically estimated by maximizing the joint likelihood with the estimated item parameters given. This approach is implemented in the three model-specific `personpar()` methods `personpar.raschmodel()`, `personpar.rsmodel()` and `personpar.pcmodel()`. The joint likelihoods of the three IRT models as a function of the person parameters $\boldsymbol{\theta}$ and conditioned on the estimated absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ are depicted in the following. For the dichotomous RM the joint likelihood is given as

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x},\hat{\boldsymbol{\delta}}) = \frac{\exp\left(\sum_{i=1}^{n}\sum_{j=1}^{m}(\theta_i - \hat{\delta}_j)\right)}{\prod_{i=1}^{n}\prod_{j=1}^{m}\left(1 + \exp\left(\theta_i - \hat{\delta}_j\right)\right)},$$

for the RSM it is given as

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x},\hat{\boldsymbol{\delta}}) = \frac{\exp\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{x_{ij}}(\theta_i - \hat{\delta}_{jk})\right)}{\prod_{i=1}^{n}\prod_{j=1}^{m}\sum_{\ell=1}^{p}\exp\left(\sum_{k=1}^{\ell}(\theta_i - \hat{\delta}_{jk})\right)}$$

with $\hat{\delta}_{jk} = \hat{\beta}_j + \hat{\tau}_k$ and for the PCM it is given as

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x},\hat{\boldsymbol{\delta}}) = \frac{\exp\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{x_{ij}}(\theta_i - \hat{\delta}_{jk})\right)}{\prod_{i=1}^{n}\prod_{j=1}^{m}\sum_{\ell=1}^{p_j}\exp\left(\sum_{k=1}^{\ell}(\theta_i - \hat{\delta}_{jk})\right)}.$$

If argument `vcov` is `TRUE` (the default), one of the above likelihood functions is maximized with the R-internal function `nlm()`, i.e., a newton-type algorithm is used and the variance-covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ of the estimated person parameters $\hat{\boldsymbol{\theta}}$ is numerically approximated (see the help page of the function `nlm()` for more details). If no variance-covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ is requested by the user, i.e., argument `vcov` is `FALSE`, the equations given in Hoijtink & Boomsma (1995, p. 55, Eq. 4.8) and Andersen (1995, p. 286, Eq. 15.43) are used and solved with the R-internal function `uniroot()` with respect to each $\theta_i, i = 1, \ldots, m-1$. This root-searching approach is typically faster than the maximization of the joint likelihood but no estimation of the variance-covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ is available.

Several utility functions are provided for extracted `personpar` objects: A `coef()` method allows the extraction of the estimated person parameters $\hat{\boldsymbol{\theta}}$ without any additional attributes, a `vcov()` method extracts the estimated variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$ of a `personpar` object. A `print()` method provides an overview of the estimated person parameters $\hat{\boldsymbol{\theta}}$.

Hence, the person parameters $\boldsymbol{\theta}$ of the PCM fitted in the previous chapter can be estimated by a call to the generic function `personpar()`:

```
> pp <- personpar(pcmod)
```

As the argument `vcov` is set to `TRUE` per default, the variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$ of the estimated person parameters $\hat{\boldsymbol{\theta}}$ is numerically approximated. It can be easily extracted with a call to the provided `vcov()` method:

```
> vcpp <- vcov(pp)
```

3.2.2. Item Discrimination Parameters

Item discrimination parameters $\boldsymbol{\alpha}$ are represented by the R class `discrpar`. Objects of this class consist of a named numeric vector with the estimated item discrimination

```
discrpar(object, ref = NULL, alias = TRUE, vcov = TRUE, ...)
```

Code segment 3.2. Interface of the generic function `discrpar()`.

parameters $\hat{\boldsymbol{\alpha}}$ of a fitted model object. A label referring to the underlying IRT model, the restriction used, if chosen, the removed aliased parameters and, if requested, the estimated variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}}$ are attached as attributes. The interface of the generic function `discrpar()` is shown in code segment 3.2. As before, the first argument is a fitted model object of class `raschmodel`, `rsmodel` or `pcmodel`. In the second argument `ref`, a restriction to be applied to the estimated item discrimination parameters $\hat{\boldsymbol{\alpha}}$ can be specified. As the item discrimination parameters $\boldsymbol{\alpha}$ are fixed to unity in all three IRT models, this argument is currently not used. With the argument `alias`, the user can choose whether aliased, i.e., restricted, item discrimination parameters $\boldsymbol{\alpha}$ should be included in the result object. The argument `vcov` provides control over the attachment of the estimated variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}}$ of the estimated item discrimination parameters $\hat{\boldsymbol{\alpha}}$.

The model-specific `discrpar()` methods are rather simple as the item discrimination parameters $\boldsymbol{\alpha}$ in all three IRT models of the R package **psychotools** are fixed to unity (see Section 2.1). Hence, these functions only set up a numeric vector of length m filled with integers of value 1. If `vcov` is `TRUE`, a numeric matrix of dimension $m \times m$ is constructed and filled with zeros as the item discrimination parameters $\boldsymbol{\alpha}$ are fixed quantities without any random variation. As for the other structural components, various utility functions (`coef()`, `vcov()`, ...) are provided for extracted `discrpar` objects.

3.2.3. Item Location Parameters

Item location parameters $\boldsymbol{\beta}$ are represented by the R class `itempar`. The structure of objects of this class as well as the interface of the generic function is identical to that of the objects of R class `discrpar`. To specify a restriction to be used in `itempar()` methods, the argument `ref` can be either a numeric vector of item indices or a character vector of item labels which specifies the items to be used as restriction, i.e., any restriction of the form depicted in Equation (2.11) with $d_0 \equiv 0$ and any real-valued vector $\boldsymbol{d}$ with $d_j \in \{0, 1\}, j = 1, \ldots m$ can be specified through the argument `ref`. In addition, an arbitrary contrast matrix can be specified in the argument `ref` (see below for more details). Hence, it is possible to extract the estimated item location parameters $\hat{\boldsymbol{\beta}}$ from a fitted model object under any arbitrary restriction and therefore separating the

estimated item location parameters $\hat{\boldsymbol{\beta}}$ from the specific characteristics of the estimation approach used in a model-fitting function.

For the dichotomous RM, those item location parameters $\tilde{\hat{\boldsymbol{\beta}}}$ which are adjusted with respect to the restriction specified in the argument `ref`, are derived based on the estimated item location parameters $\hat{\boldsymbol{\beta}}$ within the model-specific method `itempar.raschmodel()` as

$$\tilde{\hat{\beta}}_j = \hat{\beta}_j - \frac{1}{|\texttt{ref}|} \sum_{k \in \texttt{ref}}^{m} \hat{\beta}_k. \tag{3.1}$$

In addition, the adjusted variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\tilde{\hat{\boldsymbol{\beta}}}}$ is derived with the multivariate delta method (see, e.g., Casella & Berger, 2001, p. 61) as

$$\hat{\boldsymbol{\Sigma}}_{\tilde{\hat{\boldsymbol{\beta}}}} = \boldsymbol{C} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}} \boldsymbol{C}^\top \tag{3.2}$$

where

$$(\boldsymbol{C})_{ab} = \frac{\partial \tilde{\hat{\beta}}_a}{\partial \hat{\beta}_b} = \begin{cases} 1 & \text{if } a = b \wedge b \notin \texttt{ref} \\ 1 - \frac{1}{|\texttt{ref}|} & \text{if } a = b \wedge b \in \texttt{ref} \\ -\frac{1}{|\texttt{ref}|} & \text{if } a \neq b \wedge b \in \texttt{ref} \\ 0 & \text{else} \end{cases} \qquad a = 1, \ldots, m, \quad b = 1, \ldots, m. \tag{3.3}$$

If a contrast matrix $\boldsymbol{C}$ is specified in the argument `ref`, the estimated item location parameters $\hat{\boldsymbol{\beta}}$ are directly adjusted by applying the specified contrast matrix $\boldsymbol{C}$, i.e.,

$$\tilde{\hat{\boldsymbol{\beta}}} = \boldsymbol{C} \hat{\boldsymbol{\beta}},$$

and the estimated (or derived) variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}$ is adjusted as shown in Equation (3.2) with the user-specified contrast matrix $\boldsymbol{C}$ plugged in.

For the RSM and the PCM, no direct estimates of the item location parameters $\boldsymbol{\beta}$ exist in the implementation of these models in the R package psychotools. The model-fitting function `rsmodel()` for the RSM estimates an item-specific parameter (labeled as $\boldsymbol{\xi}$ in the following to avoid confusion), but this parameter is estimated under the cumulative relative item threshold parametrization (see Section 2.1 for more details) with restrictions such that it cannot be reasonably interpreted as a characterization of the center of an item. With respect to the PCM, the model-fitting function `pcmodel()` estimates cumulative absolute item threshold parameters $\hat{\boldsymbol{\eta}}$ which also cannot be reasonably interpreted as a characterization of the center of an item. To nevertheless provide a char-

acterization of an item by a single parameter for these models, item location parameters $\boldsymbol{\beta}$ as introduced in Equation (2.3), i.e.,

$$\hat{\beta}_j = \frac{1}{p_j} \sum_{k=1}^{p_j} \hat{\delta}_{jk},$$

are computed as "mean" absolute item threshold parameters in the model-specific `item-par()` methods of the RSM and the PCM. As discussed in Chapter 2 and illustrated in Figure 2.2, these parameters characterize the position of the center of a polytomous item on the latent dimension.

Based on the estimated item-specific parameters $\hat{\boldsymbol{\xi}}$ and the estimated cumulative relative item threshold parameters $\hat{\boldsymbol{\tau}}$, the item location parameters $\boldsymbol{\beta}$ as formulated in Equation (2.3) are computed in the RSM as

$$\hat{\beta}_j = \frac{1}{p} \sum_{k=1}^{p} \left(\hat{\xi}_j + (\hat{\kappa}_k - \hat{\kappa}_{k-1}) \right)$$

$$= \hat{\xi}_j + \frac{1}{p} \sum_{k=1}^{p} (\hat{\kappa}_k - \hat{\kappa}_{k-1})$$

$$= \hat{\xi}_j + \frac{\hat{\kappa}_p}{p}$$

with $\kappa_0 \equiv 0$ for notational purposes. Their variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}$ is derived with the multivariate delta method as

$$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}} = \boldsymbol{C} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\kappa}}} \boldsymbol{C}^\top$$

where

$$\boldsymbol{C} = (\boldsymbol{C_1} \ \ \boldsymbol{C_2})$$

with

$$(\boldsymbol{C_1})_{ab} = \frac{\partial \hat{\beta}_a}{\partial \hat{\xi}_b} = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{else} \end{cases} \qquad a, b = 1, \ldots, m,$$

and

$$(\boldsymbol{C_2})_{ab} = \frac{\partial \hat{\beta}_a}{\partial \hat{\kappa}_b} = \begin{cases} \frac{1}{p} & \text{b = p} \\ 0 & \text{else} \end{cases} \qquad a = 1, \ldots, m, \quad b = 1, \ldots, p,$$

19

and $\hat{\mathbf{\Sigma}}_{\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\kappa}}}$ is given as

$$
\hat{\mathbf{\Sigma}}_{\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\kappa}}} =
\begin{pmatrix}
0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\
0 & \hat{\sigma}^2_{\hat{\xi}_2} & \hat{\sigma}_{\hat{\xi}_2,\hat{\xi}_3} & \dots & 0 & \hat{\sigma}_{\hat{\xi}_2,\hat{\kappa}_2} & \hat{\sigma}_{\hat{\xi}_2,\hat{\kappa}_3} & \dots & \hat{\sigma}_{\hat{\xi}_2,\hat{\kappa}_p} \\
0 & \hat{\sigma}_{\hat{\xi}_3,\hat{\xi}_2} & \hat{\sigma}^2_{\hat{\xi}_3} & \dots & 0 & \hat{\sigma}_{\hat{\xi}_3,\hat{\kappa}_2} & \hat{\sigma}_{\hat{\xi}_3,\hat{\kappa}_3} & \dots & \hat{\sigma}_{\hat{\xi}_3,\hat{\kappa}_p} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\
0 & \hat{\sigma}_{\hat{\kappa}_2,\hat{\xi}_2} & \hat{\sigma}_{\hat{\kappa}_2,\hat{\xi}_3} & \dots & 0 & \hat{\sigma}^2_{\hat{\kappa}_2} & \hat{\sigma}_{\hat{\kappa}_2,\hat{\kappa}_3} & \dots & \hat{\sigma}_{\hat{\kappa}_2,\hat{\kappa}_p} \\
0 & \hat{\sigma}_{\hat{\kappa}_3,\hat{\xi}_2} & \hat{\sigma}_{\hat{\kappa}_3,\hat{\xi}_3} & \dots & 0 & \hat{\sigma}_{\hat{\kappa}_3,\hat{\kappa}_2} & \hat{\sigma}^2_{\hat{\kappa}_3} & \dots & \hat{\sigma}_{\hat{\kappa}_3,\hat{\kappa}_p} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \hat{\sigma}_{\hat{\kappa}_p,\hat{\xi}_2} & \hat{\sigma}_{\hat{\kappa}_p,\hat{\xi}_3} & \dots & 0 & \hat{\sigma}_{\hat{\kappa}_p,\hat{\kappa}_2} & \hat{\sigma}_{\hat{\kappa}_p,\hat{\kappa}_3} & \dots & \hat{\sigma}^2_{\hat{\kappa}_p}
\end{pmatrix}. \tag{3.4}
$$

The zeros in the estimated variance-covariance matrix $\hat{\mathbf{\Sigma}}_{\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\kappa}}}$ of Equation (3.4) arise from the restrictions $\xi_1 \equiv 0$ and $\kappa_1 \equiv 0$ used in the model-fitting function `rsmodel()` to ensure parameter identifiability (see Section 2.3 for more details).

In the PCM, item location parameters $\boldsymbol{\beta}$ as introduced in Equation (2.3) are computed from the estimated cumulative absolute item threshold parameters $\hat{\boldsymbol{\eta}}$ as

$$
\begin{aligned}
\hat{\beta}_j &= \frac{1}{p_j} \sum_{k=1}^{p_j} \left(\hat{\eta}_{jk} - \hat{\eta}_{j(k-1)} \right) \\
&= \frac{\hat{\eta}_{jp_j}}{p_j}
\end{aligned}
$$

with $\eta_{j0} \equiv 0$ for notational purposes. Their variance-covariance matrix $\hat{\mathbf{\Sigma}}_{\hat{\boldsymbol{\beta}}}$ is derived with the multivariate delta method as

$$
\hat{\mathbf{\Sigma}}_{\hat{\boldsymbol{\beta}}} = \boldsymbol{C} \hat{\mathbf{\Sigma}}_{\hat{\boldsymbol{\eta}}} \boldsymbol{C}^\top
$$

where

$$
(\boldsymbol{C})_{ab} = \frac{\partial \hat{\beta}_a}{\partial \hat{\eta}_b} =
\begin{cases}
\frac{1}{p_j} & \text{if } \hat{\eta}_b = \hat{\eta}_{ap_a} \\
0 & \text{else}
\end{cases}
\qquad a = 1, \dots, m, \quad b = 1, \dots, \sum_{j=1}^{m} p_j.
$$

After computing the item location parameters $\hat{\boldsymbol{\beta}}$ in the RSM and the PCM, the restriction specified in the argument `ref` is applied. As in the function `itempar.raschmodel()`, this is done by applying the transformation given in Equation (3.1) to the computed item location parameters $\hat{\boldsymbol{\beta}}$. Their adjusted variance-covariance matrix $\hat{\mathbf{\Sigma}}_{\hat{\boldsymbol{\beta}}}$ is derived by the

multivariate delta method as shown in Equation (3.2) with the contrast matrix given in Equation (3.3). If an arbitrary contrast matrix C is specified in the argument `ref`, the adjustment is identical to that in the function `itempar.raschmodel()` described before.

With the implemented interface for the item location parameters β, the same restriction can be easily applied to the estimated item location parameters $\hat{\beta}$ of the different IRT models. For example, the estimated item location parameters $\hat{\beta}$ of the fitted dichotomous Rasch model, the fitted RSM and the fitted PCM with the first item location parameter β_1 restricted to zero, can be extracted with the following call to the generic function `itempar()`:

```
> iprm <- itempar(rmmod, alias = FALSE, ref = 1)
> iprsm <- itempar(rsmod, alias = FALSE, ref = 1)
> ippcm <- itempar(pcmod, alias = FALSE, ref = 1)
```

By setting the argument `alias` to `FALSE`, the aliased, i.e., restricted item location parameter β_1 is omitted from the result object. Based on the extracted estimated item location parameters $\hat{\beta}$, a numerical comparison of the three IRT models fitted in the previous chapter is easily obtained by, e.g., binding them to a single matrix:

```
> print(cbind(RM = iprm, RSM = iprsm, PCM = ippcm), digits = 5)

                 RM      RSM      PCM
Do-Curse    0.79149 0.69856 0.72758
Want-Scold 1.33350 1.08845 1.08268
Do-Scold    2.26493 1.79282 1.69005
Want-Shout 2.28973 1.85656 1.82664
Do-Shout    3.87750 3.17708 2.96675
```

As can be seen from the results, the estimated item location parameters $\hat{\beta}$ of the RSM and the PCM are relatively similar, whereas the estimated item location parameters $\hat{\beta}$ of the dichotomous Rasch model are substantially higher for almost all items.

3.2.4. Item Threshold Parameters

Item threshold parameters δ are represented by the R class `threshpar`. In contrast to the other structural components, estimated item threshold parameters $\hat{\delta}$ are represented as a named list. This format is chosen because the number of estimated item threshold parameters $\hat{\delta}$ can in some models, e.g., in the PCM, vary per item. A label referring to the

```
threshpar(object, type = c("mode", "mean", "median"), ref = NULL,
  alias = TRUE, relative = FALSE, cumulative = FALSE, vcov = TRUE, ...)
```

Code segment 3.3. Interface of the generic function `threshpar()`.

underlying IRT model, the restriction used, if chosen, the removed aliased parameters, if requested, the estimated (and adjusted) variance-covariance matrix $\hat{\Sigma}_{\hat{\delta}}$, the type of the extracted estimated item threshold parameters $\hat{\delta}$ and the information whether they are relative or not and cumulative or not are attached as attributes to the named list of class `threshpar`.

The interface of the generic function `threshpar()` is shown in code segment 3.3. As in the other generic functions, the first argument `object` is a fitted model object of class `raschmodel`, `rsmodel` or `pcmodel`. With the argument `type`, item threshold parameters $\hat{\delta}$ based on different definitions can be extracted. Within this Master's thesis only item threshold parameters $\hat{\delta}$ of `type mode` are considered. These correspond to the item threshold parameters $\hat{\delta}$ discussed in Chapter 2 and illustrated in Figure 2.2. As before, the argument `alias` allows to choose whether the aliased, i.e., restricted parameters should be included in the result object. With the logical argument `relative`, the extraction of absolute (`FALSE`) or relative (`TRUE`) item threshold parameters can be requested. With the logical argument `cumulative`, cumulative item threshold parameters $\hat{\eta}$ as discussed in Section 2.1 can be requested. The argument `vcov` controls whether the adjusted variance-covariance matrix $\hat{\Sigma}_{\hat{\delta}}$ is attached as attribute. The adjustment with respect to the restriction specified in the argument `ref` is based on the multivariate delta method. The specification of the argument `ref` is dependent on the underlying IRT model and will be discussed in the following together with the details of the three model-specific methods `threshpar.raschmodel()`, `threshpar.rsmodel()` and `threshpar.pcmodel()`.

In the dichotomous Rasch model as implemented in the R package `psychotools` and discussed in Section 2.3, the estimated absolute item threshold parameters $\hat{\delta}$ directly correspond to the estimated item location parameters $\hat{\beta}$. Therefore, the estimated (and possibly adjusted) item location parameters $\hat{\beta}$ are returned if absolute item threshold parameters δ are requested. The adjustment and the specification of the argument `ref` is identical to that in the function `itempar.raschmodel()` and described in Section 3.2.3. If relative item threshold parameters τ are requested, `itempar.raschmodel()` returns a numeric vector of length m filled with zeros as there is only one relative item threshold parameter τ_1 which is (arbitrarily) restricted to zero. Other restrictions can be applied

by supplying a contrast matrix $\boldsymbol{C}$ in the argument `ref`. The relative item threshold parameters $\boldsymbol{\tau}$ are then adjusted by multiplying the unmodified parameter vector with the supplied contrast matrix. The variance-covariance matrix $\boldsymbol{\Sigma_\tau}$ of the relative item threshold parameters $\boldsymbol{\tau}$ is always a $m \times m$ matrix filled with a zero as there is no relative item threshold parameter to be estimated in the dichotomous Rasch model. In addition, cumulative or non-cumulative absolute or relative item threshold parameters are identically in the dichotomous Rasch model as there is only one parameter per item.

In the implementation of the RSM and the PCM in the R package psychotools, the absolute or relative item threshold parameters are not estimated directly and therefore the following steps are carried out in the functions `threshpar.rsmodel()` and `threshpar.pcmodel()` to arrive at the parameters requested by the user:

1. Based on the estimated parameters, absolute or relative item threshold parameters and their variance-covariance matrix are computed.

2. The restriction specified in the argument `ref` is applied on the computed parameters and their variance-covariance matrix is adjusted accordingly.

3. If requested, cumulative absolute or relative item threshold parameters and their variance-covariance matrix are computed.

These steps are explained in more detail in the following separately for the RSM and the PCM.

For the RSM, absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ are computed based on the estimated item-specific parameters $\hat{\boldsymbol{\xi}}$ and the estimated cumulative relative item threshold parameters $\hat{\boldsymbol{\kappa}}$ as

$$\hat{\delta}_{jk} = \hat{\xi}_j + (\hat{\kappa}_k - \hat{\kappa}_{k-1})$$

with $j = 1, \ldots, m$, $k = 1, \ldots, p$ and $\kappa_0 \equiv 0$ for notational purposes. Their variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\delta}}}$ is derived with the multivariate delta method as

$$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\delta}}} = \boldsymbol{C} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\kappa}}} \boldsymbol{C}^\top$$

where

$$\boldsymbol{C} = (\boldsymbol{C_1} \ \ \boldsymbol{C_2})$$

with

$$(\boldsymbol{C_1})_{ab} = \frac{\partial \hat{\delta}_a}{\partial \hat{\xi}_b} = \begin{cases} 1 & \text{if } \hat{\delta}_a \in \{\hat{\delta}_{b1}, \hat{\delta}_{b2}, \ldots, \hat{\delta}_{bp}\} \\ 0 & \text{else} \end{cases} \qquad a = 1, \ldots, m \cdot p, \quad b = 1, \ldots m,$$

and

$$
(\boldsymbol{C_2})_{ab} = \frac{\partial \hat{\delta}_a}{\partial \hat{\kappa}_b} = \begin{cases} 1 & \text{if } \hat{\delta}_a \in \{\hat{\delta}_{1b}, \ldots, \hat{\delta}_{mb}\} \\ -1 & \text{if } \hat{\delta}_a \in \{\hat{\delta}_{1(b-1)}, \ldots, \hat{\delta}_{m(b-1)}\} \wedge b > 1 \\ 0 & \text{else} \end{cases}
$$

with $a = 1, \ldots, m \cdot p, b = 1, \ldots, p$ and $\hat{\boldsymbol{\Sigma}}_{\hat{\xi}, \hat{\kappa}}$ given as in Equation (3.4). Relative item threshold parameters $\hat{\boldsymbol{\tau}}$ are computed based on the estimated cumulative relative item threshold parameters $\hat{\boldsymbol{\kappa}}$ as

$$
\hat{\tau}_k = \hat{\kappa}_k - \hat{\kappa}_{k-1}
$$

with $k = 1, \ldots, p$ and $\kappa_0 \equiv 0$ for notational purposes. Their variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\kappa}}$ is derived with the multivariate delta method as

$$
\hat{\boldsymbol{\Sigma}}_{\hat{\tau}} = \boldsymbol{C} \hat{\boldsymbol{\Sigma}}_{\hat{\kappa}} \boldsymbol{C}^\top
$$

where

$$
\hat{\boldsymbol{\Sigma}}_{\hat{\kappa}} = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ 0 & \hat{\sigma}^2_{\hat{\kappa}_2} & \hat{\sigma}_{\hat{\kappa}_2, \hat{\kappa}_3} & \ldots & \hat{\sigma}_{\hat{\kappa}_2, \hat{\kappa}_p} \\ 0 & \hat{\sigma}_{\hat{\kappa}_3, \hat{\kappa}_2} & \hat{\sigma}^2_{\hat{\kappa}_3} & \ldots & \hat{\sigma}_{\hat{\kappa}_3, \hat{\kappa}_p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \hat{\sigma}_{\hat{\kappa}_p, \hat{\kappa}_2} & \hat{\sigma}_{\hat{\kappa}_p, \hat{\kappa}_3} & \ldots & \hat{\sigma}^2_{\hat{\kappa}_p} \end{pmatrix}
$$

and

$$
(\boldsymbol{C})_{ab} = \frac{\partial \hat{\tau}_a}{\partial \hat{\kappa}_b} = \begin{cases} 1 & \text{if } a = b \\ -1 & \text{if } a = b - 1 \wedge b > 1 \\ 0 & \text{else} \end{cases} \qquad a, b = 1, \ldots, p.
$$

For both, absolute and relative item threshold parameters a restriction identical to that of the item location parameters $\boldsymbol{\beta}$ can be specified through the argument `ref`, i.e., `ref` can be again either a character vector of absolute or relative item threshold parameter labels or a numeric vector with absolute or relative item threshold parameter position indices. In both cases, the particular parameters are adjusted as the item location parameters $\boldsymbol{\beta}$ in the dichotomous Rasch model, i.e., a transformation identically to that in Equation (3.1) is applied on the computed absolute or relative item threshold parameters and their variance-covariance matrix is adjusted by the multivariate delta method. For this purpose, a contrast matrix structurally identical to that in Equation (3.3) is used. Additionally, as in the `itempar()` methods, the user can again specify an arbitrary contrast matrix $\boldsymbol{C}$ which is then directly used to transform the computed absolute or relative

item threshold parameters and their variance-covariance matrix. If cumulative absolute or relative item threshold parameters are requested by the user, the computed and transformed absolute or relative item threshold parameters and their variance-covariance matrix are in a last step adjusted by a block-diagonal contrast matrix

$$
\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C_1} & \dots & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \vdots & \ddots & & \ddots & \vdots \\ \boldsymbol{0} & & \boldsymbol{C_j} & & \boldsymbol{0} \\ \vdots & \ddots & & \ddots & \vdots \\ \boldsymbol{0} & \dots & \boldsymbol{0} & \dots & \boldsymbol{C_m} \end{pmatrix} \tag{3.5}
$$

with

$$
\boldsymbol{C_j} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{p \times p}. \tag{3.6}
$$

The block-diagonal contrast matrix $\boldsymbol{C}$ implements a cumulative sum over the absolute or relative item threshold parameters of an item j (see Table A.1 for an overview of the different transformations).

For the PCM, absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ are computed based on the estimated cumulative absolute item threshold parameters $\hat{\boldsymbol{\eta}}$ as

$$
\hat{\delta}_{jk} = \hat{\eta}_{jk} - \hat{\eta}_{j(k-1)} \tag{3.7}
$$

with $j = 1, \dots, m$, $k = 1, \dots, p_j$ and $\eta_{j0} \equiv 0$ for notational purposes. Their variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\delta}}}$ is derived with the multivariate delta method as

$$
\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\delta}}} = \boldsymbol{C} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\eta}}} \boldsymbol{C}^{\top}
$$

where

$$
\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\eta}}} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & \hat{\sigma}^2_{\hat{\eta}12} & \hat{\sigma}_{\hat{\eta}12,\hat{\eta}13} & \dots & \hat{\sigma}_{\hat{\eta}12,\hat{\eta}mp_m} \\ 0 & \hat{\sigma}_{\hat{\eta}13,\hat{\eta}12} & \hat{\sigma}^2_{\hat{\eta}13} & \dots & \hat{\sigma}_{\hat{\eta}13,\hat{\eta}mp_m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \hat{\sigma}_{\hat{\eta}mp_m,\hat{\eta}12} & \hat{\sigma}_{\hat{\eta}mp_m,\hat{\eta}13} & \dots & \hat{\sigma}^2_{\hat{\eta}mp_m} \end{pmatrix} \tag{3.8}
$$

and C is given as block-diagonal matrix

$$
C = \begin{pmatrix}
C_1 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & & \ddots & \vdots \\
0 & & C_j & & 0 \\
\vdots & \ddots & & \ddots & \vdots \\
0 & \cdots & 0 & \cdots & C_m
\end{pmatrix}
$$

with

$$
(C_j)_{ab} = \frac{\partial \hat{\delta}_{ja}}{\partial \hat{\eta}_{jb}} = \begin{cases}
1 & \text{if } a = b \\
-1 & \text{if } a = b - 1 \wedge b > 1 \\
0 & \text{else}
\end{cases} \qquad a, b = 1, \ldots, p_j.
$$

Based on the transformations discussed in Section 2.1 and shown in Equation (2.3), Equation (2.4) and Equation (3.7), relative item threshold parameters $\hat{\boldsymbol{\tau}}$ are computed in the PCM from the estimated cumulative absolute item threshold parameters $\hat{\boldsymbol{\eta}}$ as

$$
\begin{aligned}
\hat{\tau}_{jk} &= \hat{\delta}_{jk} - \hat{\beta}_j \\
&= \hat{\delta}_{jk} - \frac{1}{p_j} \sum_{k=1}^{p_j} \hat{\delta}_{jk} \\
&= \left(\hat{\eta}_{jk} - \hat{\eta}_{j(k-1)} \right) - \frac{1}{p_j} \sum_{k=1}^{p_j} \left(\hat{\eta}_{jk} - \hat{\eta}_{j(k-1)} \right) \\
&= \left(\hat{\eta}_{jk} - \hat{\eta}_{j(k-1)} \right) - \frac{\hat{\eta}_{jp_j}}{p_j}
\end{aligned}
$$

with $j = 1, \ldots, m$, $k = 1, \ldots, p_j$ and $\eta_{j0} \equiv 0$ for notational purposes. Their variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\tau}}$ is derived with the multivariate delta method as

$$
\hat{\boldsymbol{\Sigma}}_{\hat{\tau}} = C \hat{\boldsymbol{\Sigma}}_{\hat{\eta}} C^\top
$$

where $\hat{\Sigma}_{\hat{\eta}}$ is given as in Equation (3.8) and C is given as block-diagonal matrix

$$
C = \begin{pmatrix}
C_1 & \dots & 0 & \dots & 0 \\
\vdots & \ddots & & \ddots & \vdots \\
0 & & C_j & & 0 \\
\vdots & \ddots & & \ddots & \vdots \\
0 & \dots & 0 & \dots & C_m
\end{pmatrix}
$$

with

$$
(C_j)_{ab} = \frac{\partial \hat{\tau}_{ja}}{\partial \hat{\eta}_{jb}} = \begin{cases}
1 & \text{if } a = b \wedge a < p_j \\
1 - \frac{1}{p_j} & \text{if } a = b \wedge a = p_j \\
-1 & \text{if } a = b - 1 \wedge 1 < b < p_j \\
0 & \text{else}
\end{cases} \qquad a, b = 1, \dots, p_j.
$$

In the case of absolute item threshold parameters $\boldsymbol{\delta}$, a restriction identical as above for the RSM can be specified through the argument `ref` in the function `itempar.pcmodel()`. In the case of relative item threshold parameters $\boldsymbol{\tau}$, m restrictions have to be specified as was discussed in Section 2.3. The argument `ref` in this case can be either again a single character vector of relative item threshold parameter labels or a single numeric vector with relative item threshold parameter position indices or a list with m different character or numeric vectors. In the case a single restriction was specified, this restriction is used for all m sets of relative item threshold parameters $\boldsymbol{\tau}_j$. As before in the RSM, the absolute or relative item threshold parameters are adjusted by applying a transformation identically to that in Equation (3.1) and their variance-covariance matrix is again adjusted by the multivariate delta method with a contrast matrix structurally identical to that in Equation (3.3). Additionally, the user can again specify an arbitrary contrast matrix C which then will be used instead. If cumulative absolute or relative item threshold parameters are requested by the user, these parameters and their variance-covariance matrix are computed as in the RSM, i.e., by applying a block-diagonal contrast matrix similar to that shown in Equation (3.5) and Equation (3.6) on the computed and adjusted absolute or relative item threshold parameters.

Similar as the estimated item location parameters $\hat{\boldsymbol{\beta}}$, the estimated absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ can be extracted from a fitted RSM and a fitted PCM with the following calls to the generic function `threshpar()`:

```
> atprsm <- threshpar(rsmod, relative = FALSE, ref = 1)
> atppcm <- threshpar(pcmod, relative = FALSE, ref = 1)
```

As before, the first absolute item threshold parameter δ_1 is restricted to zero and this time is included in the result object. The extracted estimated absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ can be again used for a numerical comparison of the two IRT models. The `coef()` method allows the extraction of the estimated absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ in a matrix format which can be easily binded together to provide a convenient summary of the two IRT models:

```
> print(cbind(coef(atprsm, type = "matrix"),
+              coef(atppcm, type = "matrix")), digits = 5)

             C1      C2      C1      C2
Want-Curse 0.00000 1.5038 0.00000 1.5779
Do-Curse   0.69856 2.2024 0.65786 2.3752
Want-Scold 1.08845 2.5923 1.12687 2.6164
Do-Scold   1.79282 3.2966 1.92399 3.0340
Want-Shout 1.85656 3.3604 1.91094 3.3202
Do-Shout   3.17708 4.6809 3.26369 4.2477
```

As can be seen from the results, there is a rather strong variation in the estimated absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ between the two IRT models. E.g., the first absolute item threshold parameter $\hat{\delta}_{21}$ of the second item "Do-Curse" is smaller for the PCM compared to the RSM where instead the second absolute item threshold parameter $\hat{\delta}_{22}$ is larger for the PCM compared to the RSM. This variation in the estimated absolute item threshold parameters $\hat{\delta}_{jk}$ between the two IRT models can be used as a first indication when selecting an appropriate IRT model.

A comparison of the estimated relative item threshold parameters $\hat{\boldsymbol{\tau}}$ of the two models allows a more direct assessment whether the more restrictive threshold parametrization in the RSM compared to the PCM is appropriate for the present data set (see Section 2.1 for a more detailed discussion of the different parametrizations and assumptions). As the provided interface allows the extraction of the estimated relative item threshold parameters $\hat{\boldsymbol{\tau}}$ under arbitrary restrictions, such a comparison is easily possible. For this purpose, the estimated relative item threshold parameters $\hat{\boldsymbol{\tau}}$ are in a first step extracted from the fitted model objects with the following calls to the generic function `threshpar()`:

```
> rtprsm <- threshpar(rsmod, ref = 1, relative = TRUE)
> rtppcm <- threshpar(pcmod, ref = 1, relative = TRUE)
```

As before, the same restriction is applied for both models by appropriately setting the `ref` argument. In the RSM, this means that the first relative item threshold parameter τ_1 is set to zero. In the PCM, as was described before, this means that each item-specific relative item threshold parameter τ_{j1} is set to zero. A numerical comparison of the two IRT models is again easily created by binding the extracted parameter matrices of the two models together:

```
> print(cbind(coef(rtprsm, type = "matrix"),
+             coef(rtppcm, type = "matrix")), digits = 5)

            C1     C2 C1       C2
Want-Curse   0 1.5038  0 1.57787
Do-Curse     0 1.5038  0 1.71731
Want-Scold   0 1.5038  0 1.48949
Do-Scold     0 1.5038  0 1.10998
Want-Shout   0 1.5038  0 1.40926
Do-Shout     0 1.5038  0 0.98399
```

The numerical comparison of the estimated relative item threshold parameters $\hat{\tau}_{jk}$ indicates a rather strong variation in the item-specific estimates of the PCM compared to the global estimate of the RSM. Hence, the more flexible parametrization of the PCM compared to the RSM might be more appropriate for the present subset of the verbal aggression data. In the following chapters, the question whether the RSM or the PCM is more appropriate for the present subset of the verbal aggression data will be pursued further and additional tools will be presented which facilitate the process of selecting an appropriate IRT model.

An unified framework for IRT models was introduced in this chapter and implemented in the R package psychotools. This framework provides an extensible tool to extract the different parameters of IRT models together with their variance-covariance matrices independent of the restriction used when estimating them. For the person parameters $\boldsymbol{\theta}$, a maximum likelihood estimation procedure was additionally implemented for the three IRT models of the R package psychotools. As was shown in several application examples, a numerical comparison of IRT models based on their estimated parameters is easily possible with the provided framework. In addition, it is the foundation for several additional visual and inferencial tools introduced in the following chapters.

4. Visualization of IRT Models

This chapter focuses on the visualization of IRT models. Based on the available literature, existing R packages for IRT modeling and the theoretical background presented in Chapter 2, several more established visualization techniques for IRT models are in a first step discussed in the following Section 4.1. In a second step, an implementation of the discussed visualization techniques which is built upon the unified framework established in the previous Chapter 3 is provided in Section 4.2. Several application examples illustrate the advantages of the provided implementation in more detail in the following.

4.1. Strategies to Visualize IRT Models

The most frequently used visualization of IRT models is the category response curve visualization exemplarily illustrated in Figure 2.1 and Figure 2.2 for a GPCM. This type of visualization of an IRT model can be found in nearly all popular IRT text books (e.g., Hambleton et al., 1991; Van der Linden & Hambleton, 1997; De Boeck & Wilson, 2004; Nering & Ostini, 2010) under various labels, e.g., as trace lines, item or category operating curves, item or category characteristic curves or item or category response curves which is also the term used in this Master's thesis. It is also implemented in several R packages like the R package eRm (Mair et al., 2014) or the R package ltm (Rizopoulos, 2013). As already discussed in Chapter 2, the probability of choosing a category k of an item j as predicted under a certain IRT model is illustrated in this type of visualization. The visualized probabilities directly result from the specific model equation, e.g., for the GPCM, the probabilities are computed as shown in Equation 2.1. Hence, the category response curve visualization illustrates all available information as it not only uses the point estimates of the item parameters but also the assumed structure of the underlying IRT model. In addition, it is available for every IRT model. If a comparison of different items is intended, a matrix approach is often used with the category response curve visualization discussed before. In this approach, the category response curves of different items are arranged in a matrix such that each cell represents the category response curves of a certain item as predicted under a certain IRT model. This is illustrated in Figure 4.1
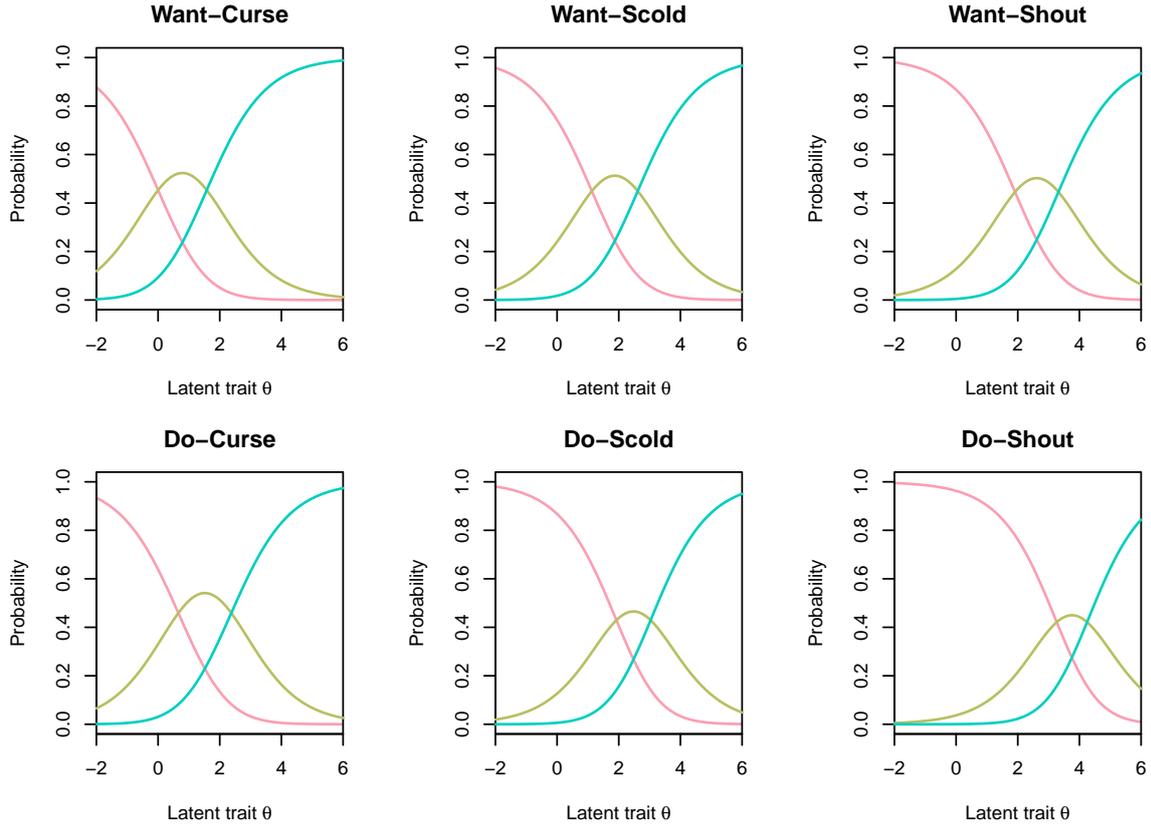
Figure 4.1. Visualization of the category response curves under a PCM fitted to the items 13-18 of the verbal aggression data in a matrix of "curve plots".

for a PCM fitted to the items 13-18 of the verbal aggression data.

In a related visualization technique used by Masters & Wright (1997) besides the category response curve visualization, each category is visualized only by a single region instead of a whole curve as in the category response curve visualization. The region of a category marks the area on the dimension of the latent trait where this category is the most probable chosen category. Such a "region plot" is illustrated in Figure 4.2 for the items and the PCM already illustrated in Figure 4.1. As the regions are completely determined by the estimated absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ and the information concerning the specific probabilities of choosing some category k is dismissed, no knowledge of the underlying IRT model is necessary in this type of visualization. The only necessity are the estimated absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ of an IRT model. In addition, by dismissing some information, multiple items can be illustrated more compactly and a comparison between different items is much easier. The region plot visualization is related to the "effect plots" suggested by Fox & Hong (2009) in the context

31
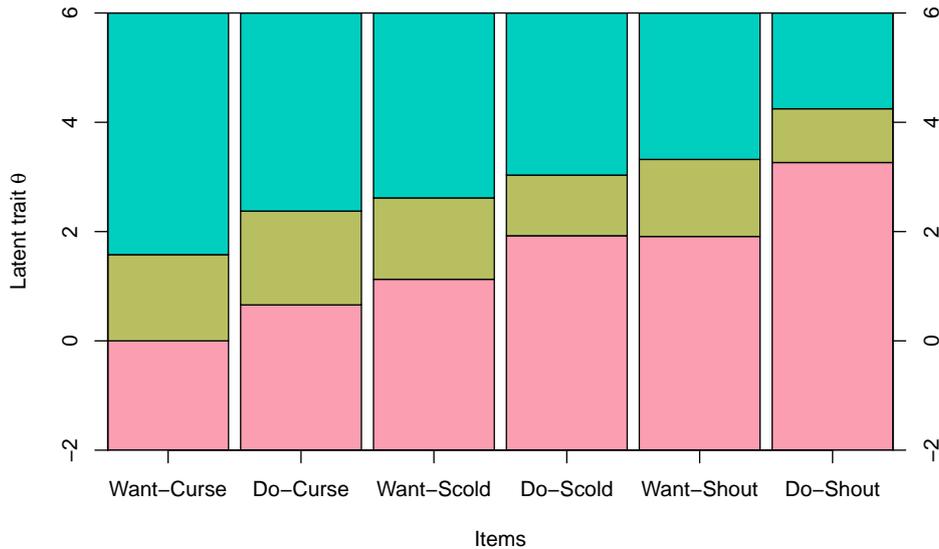
Figure 4.2. Visualization of the estimated absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ of a PCM fitted to the items 13-18 of the verbal aggression data in a "region plot".

of multinomial and a proportional-odds logit models where the proportions of choosing some option of a categorical item is visualized along the linear predictor of a multinomial or proportional-odds logit model.

Another strategy to visualize IRT models are "profile plots". In this approach, each item is visualized only by its estimated item location parameter $\hat{\beta}_j$. A profile is displayed by connecting the individual point estimates by a dashed line thus facilitating the recognition of differences between the items. Hence, as in the region plot visualization, all information concerning the underlying IRT model is dismissed. In addition, each item is solely represented by a single point estimate. This type of visualization is available for all IRT models where an item can be represented by a single parameter. As was discussed in Chapter 2, this is the case for the GPCM and all related models when a relative item threshold parametrization is used. Such a "profile plot" of the PCM already illustrated in Figure 4.1 and Figure 4.2 is illustrated in Figure 4.3.

Besides the introduced approaches, few alternative visualization techniques for IRT models can be found in the literature or other R packages for IRT modeling. One more established alternative is the joint visualization of person and item parameters. In this approach, the distribution of the estimated person parameters $\hat{\boldsymbol{\theta}}$ is visualized against the locations of the absolute item threshold parameters $\hat{\boldsymbol{\delta}}$. This type of visualization can be found, e.g., in Andrich (2013) and is also implemented in the R package eRm (Mair et al., 2014). An example of this type of visualization as implemented in the R
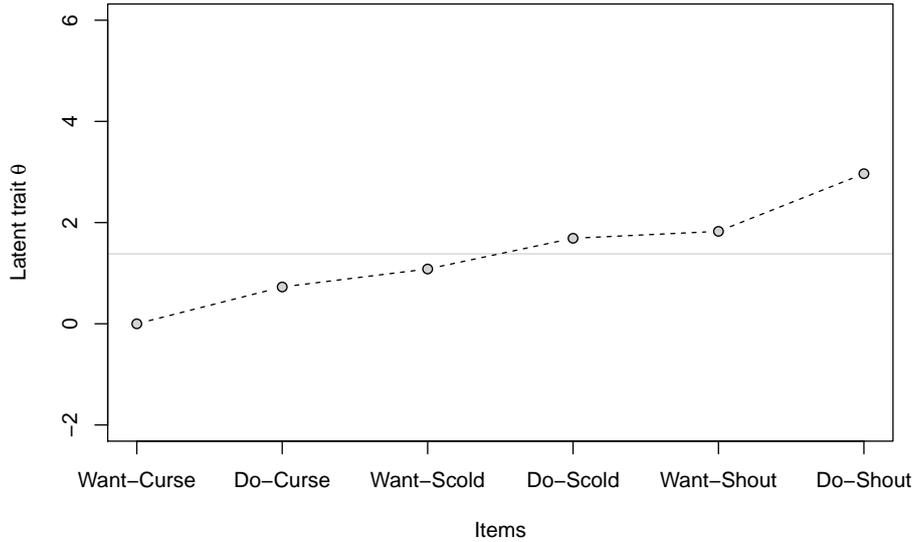
Figure 4.3. Visualization of the estimated item location parameters $\hat{\boldsymbol{\beta}}$ of a PCM fitted to the items 13-18 of the verbal aggression data in a "profile plot".

package eRm (Mair et al., 2014) is illustrated in Figure 4.4 for the PCM already shown in the figures before and will be called "person-item plot" from here on. As the region and profile plots, person-item plots dismiss the structure of the underlying IRT model and only visualize point estimates or their frequencies.

A more indirect visualization technique for IRT models is the illustration of the "information" a category, an item or a whole test provides under a certain IRT model with respect to the estimation of the person parameters $\boldsymbol{\theta}$. The term "information" here denotes the Fisher information, i.e., the negative expectation of the second derivative of the likelihood function of a certain IRT model. A formal definition of the information of a category, an item or a test under the GPCM introduced in Section 2 can be found in Muraki (1993). This type of illustration is also implemented in the R package eRm (Mair et al., 2014) and will be called "information plot" from here on. It is exemplarily illustrated in Figure 4.5 for the PCM already shown in the figures before. Similar to the item or category response curve visualization introduced before, the visualization of the information is dependent on the structure of the underlying IRT model and is available for all IRT models.

Two things can be concluded from the above introduction and illustration of several existing visualization techniques for IRT models: First, they differ in their degree of abstractness and the information visualized, and second, different structural elements are necessary to create them. All visualization techniques except the item or category oper-
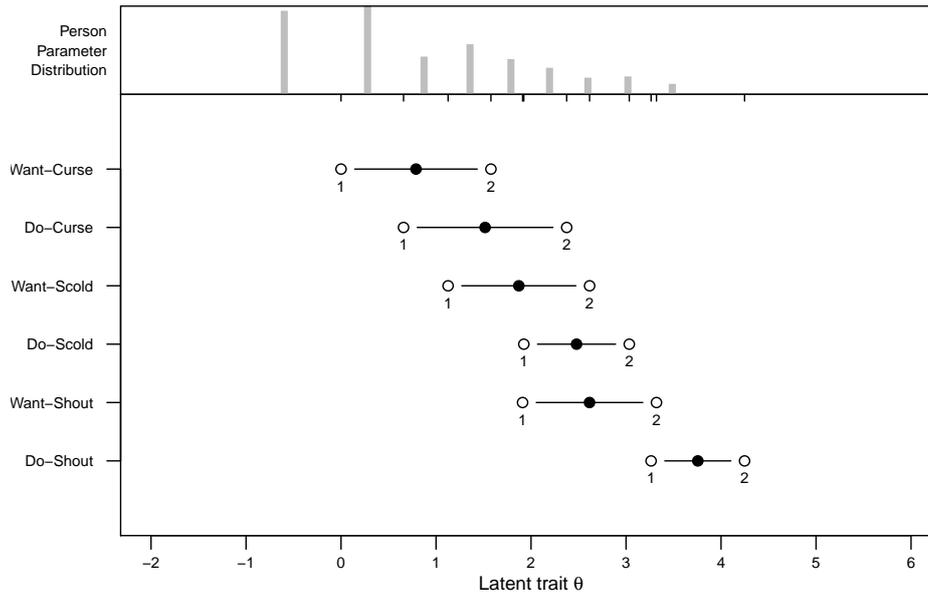
Figure 4.4. Joint visualization of the estimated person parameters $\hat{\boldsymbol{\theta}}$ and the estimated absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ of a PCM fitted to the items 13-18 of the verbal aggression data in a "person-item plot" as implemented in the R package eRm (Mair et al., 2014).
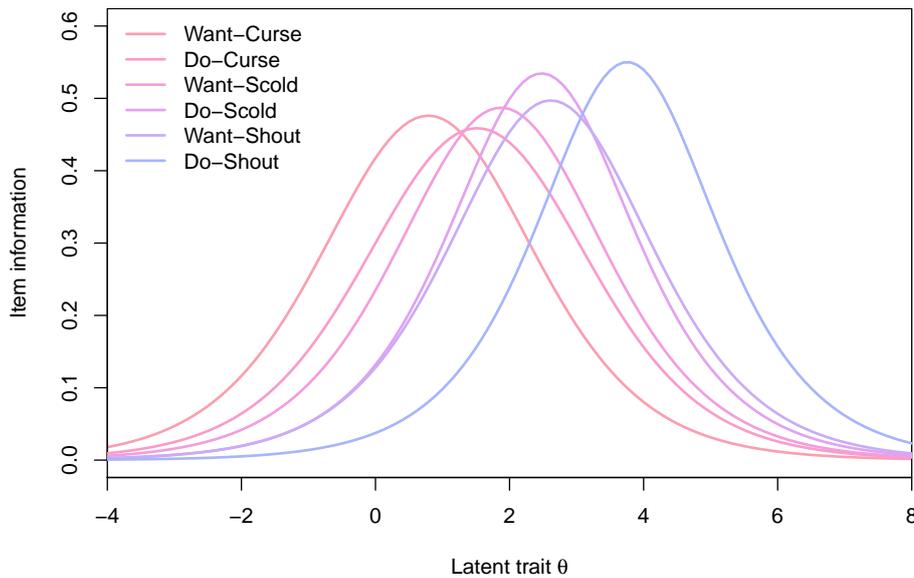
Figure 4.5. Visualization of the item information under a PCM fitted to the items 13-18 of the verbal aggression data in an "information plot".

ating curves and the visualization of the category, item or test information dismiss the structure of the underlying IRT model and only depend on one or more of the identified

Table 4.1. Summary of the different visualization techniques, the necessary structural components and the name of the implemented R function to create them.

| Visualization technique	Necessary structural components	Implemented R function
Item or category response curve (matrix)	Model structure, absolute item threshold parameters $\boldsymbol{\delta}$	`curveplot()`
Region plot	Absolute item threshold parameters $\boldsymbol{\delta}$	`regionplot()`
Profile plot	Depending on the visualized profile either item location parameters $\boldsymbol{\beta}$, absolute item threshold parameters $\boldsymbol{\delta}$ or item discrimination parameters $\boldsymbol{\alpha}$	`profileplot()`
Person-item plot	Person parameters $\boldsymbol{\theta}$ and absolute item threshold parameters $\boldsymbol{\delta}$	`piplot()`
Information plot	Fisher information under a specific IRT model	`infoplot()`

structural components of an IRT model. Based on the unified framework introduced in Chapter 3, an implementation of these visualization techniques independent of a certain IRT model is suggested in the next Section 4.2. For the item or category operating curves and the visualization of the information, an additional step is necessary to provide such a model-independent implementation. A solution for this problem together with an implementation of these visualization techniques is also discussed in the next Section 4.2.

4.2. An Implementation Based on the Unified Framework

Beginning from the category response curve visualization either for a single item as illustrated in Figure 2.1 and Figure 2.2 or for multiple items as illustrated in Figure 4.1, the above introduced visualization techniques successively reduce the amount of information illustrated and also the elements necessary to create them. This is summarized in Table 4.1 where for each visualization technique the necessary structural components and, in addition, the name of the implemented R function to create a specific visualization are depicted.

As can be seen from Table 4.1, all visualization techniques except the item or category response curve visualization and the visualization of the information of a category, item

```
curveplot(object, ref = NULL, items = NULL, names = NULL, layout = NULL,
  xlim = NULL, ylim = c(0, 1), col = NULL, lty = NULL, main = NULL,
  xlab = "Latent trait", ylab = "Probability", add = FALSE, ...)
```

Code segment 4.1. Interface of the function `curveplot()`.

or test only depend on one or two of the identified structural components. With the provided unified framework discussed in Section 3.2, an implementation independent of the underlying IRT model is possible for these visualization techniques. For the item or category response curve visualization, the structure of the underlying IRT model is necessary to compute the probabilities illustrated in this type of visualization. For the visualization of the information, the Fisher information under a certain IRT model is necessary. Solutions implemented for these two visualization techniques and more details on the implemented functions to create the different types of visualization are discussed in the following sections separately for each visualization technique.

4.2.1. Item or Category Response Curves (Curve Plots)

Item or category response curves have been implemented in the function `curveplot()`. The interface of this function is shown in code segment 4.1.

In the first argument `object`, a fitted model object of class `raschmodel`, `rsmodel` or `pcmodel` can be specified for which item or category response curves should be visualized. Hence, the function `curveplot()` is implemented model-independent. To nevertheless compute model-dependent probabilities as needed in the item or category response curve visualization, a method of the generic function `predict()` for the fitted model given in the argument `object` has to exist, which is called internally in the function `curveplot()`. This method should be able to compute the category-specific probabilities for a given value of the person parameter θ and a given restriction for the absolute item threshold parameters δ which can be specified in the argument `ref`. Such `predict()` methods have been implemented within this Master's thesis for the dichotomous Rasch model, the RSM and the PCM in the R package **psychotools**.

With the function `curveplot()`, both, a visualization of the item or category response curves of a single item as well as a matrix of the item or category response curves of multiple items as shown in Figure 4.1 can be created. This can be controlled with the argument `items` where either position indices or names of items of the data set used to fit the model can be specified. For example, the category response curves only of the

```
regionplot(object, thresharg = list(type = NULL, ref = NULL,
  alias = TRUE), names = NULL, main = NULL, xlab = "Items",
  ylab = "Latent trait", ylim = NULL, off = 0.1, col = NULL, ...)
```

<div align="center">Code segment 4.2. Interface of the function <code>regionplot()</code>.</div>

13th item of the verbal aggression data under the previously estimated PCM can be visualized with the following call:

```
> curveplot(pcmod, items = 1)
```

The category response curve visualization of all six items which is illustrated in Figure 4.1 can be created with the following call:

```
> curveplot(pcmod, ref = 1, items = 1:6,
+    layout = matrix(1:6, ncol = 3, nrow = 2, byrow = FALSE))
```

In the previous call, the arrangement of the category response curves of the six items visualized was modified by setting the `layout` argument. In this argument, a matrix with successive increasing integers can be specified and the category response curves of different items will be arranged in a matrix with the same dimensions and in the order as the specified integer values are located within the given matrix. In addition, the first estimated absolute item threshold parameter $\hat{\delta}_{11}$ was set to zero by using the argument `ref`. To increase the readability, additional arguments to control further aspects like the axis labels have been omitted in the illustrated call. A detailed description of these arguments can be found on the help site of the function `curveplot()`. With the argument `add`, the category response curves of different models can be overlayed. This is an advantage of the model-independent implementation of the different visualization techniques and will be illustrated in more detail in Section 4.3.

4.2.2. Region Plots

Region plots have been implemented in the function `regionplot()`. The interface of this function is shown in code segment 4.2.

As discussed before, the region plot visualization is only dependent on the absolute item threshold parameters δ. Hence, the function `regionplot()` is independent of a certain IRT model and only relies on a method for the generic function `threshpar()` for the model specified in the first argument `object`. Such methods have been implemented

```
profileplot(object, what = c("items", "thresholds", "discriminations"),
  paramarg = list(type = NULL, ref = NULL, alias =  TRUE), index = TRUE,
  names = NULL, main = NULL, abbreviate = FALSE, ref = TRUE, col = NULL,
  pch = NULL, cex = 1, refcol = "lightgray", linecol = "black", lty = 2,
  ylim = NULL, xlab = NULL, ylab = NULL, add = FALSE, ...)
```

Code segment 4.3. Interface of the function `profileplot()`.

for the three IRT models available in the R package **psychotools** (see Section 3.2 for more details). Several settings of the internal call of the function `threshpar()` can be specified in the argument `thresharg`. Various aspects concerning the graphical appearence of a region plot can be controlled by the additional arguments of the function `regionplot()`. A detailed description of these arguments can be found on the help site of this function. The region plot shown in Figure 4.2 can be created with the following call:

```
> regionplot(pcmod, names = lbs, thresharg = list(ref = 1),
+    ylab = expression(paste("Latent trait ", theta)),
+    ylim = c(-1.5, 2.5), col = c("#E495A5", "#ABB065", "#39BEB1"))
```

As before, the first estimated item threshold parameter $\hat{\delta}_{11}$ has been restricted to zero by setting the argument `ref`. In addition, several visual aspects like the axis labels, the axis range, the item labels and the colors of the illustrated regions have been adjusted in the above call.

4.2.3. Profile Plots

Profile plots have been implemented in the function `profileplot()`. The interface of this function is shown in code segment 4.3.

The function `profileplot()` is very general. Depending on the specification of the argument `what`, a profile plot as shown in Figure 4.3 can be created, but the function also allows to create profiles of other structural components, e.g., a profile plot of the discrimination parameters $\boldsymbol{\alpha}$ or a profile plot of the absolute item threshold parameters $\boldsymbol{\delta}$. The requested parameters are retrieved inside the function `profileplot()` by a call to one of the generic functions `itempar()`, `threshpar()` or `discrpar()`. Hence, independent of the requested profiles to visualize, the function `profileplot()` is also implemented model-independent as long as methods for the generic functions `itempar()`, `threshpar()` and `discrpar()` exist for the model object specified in the argument

```
piplot(object, ref = NULL, items = NULL, xlim = NULL, names = NULL,
  labels = TRUE, main = "Person-Item Map", xlab = "Latent trait",
  abbreviate = FALSE, cex.axis = 0.8, cex.text = 0.5,
  cex.points = 1.5, ...)
```

Code segment 4.4. Interface of the function `piplot()`.

`object`. The argument `paramarg` of the function `profileplot()` allows to adjust certain settings in the call of these functions (see Section 3.2 for a more detailed description of these arguments). Several further arguments allow to modify the visual appearance of the created profile plot. As before, a detailed description of these arguments can be found on the help site of the function `profileplot()`. Examples of other profile plots are illustrated in the following Section 4.3.

4.2.4. Person-Item Plots

Person-item plots have been implemented in the function `piplot()`. The interface of this function is shown in code segment 4.4.

As listed in Table 4.1, person parameters θ and absolute item threshold parameters δ are necessary to create an person-item plot as shown in Figure 4.4. These structural components are retrieved from the fitted model object specified in the first argument `object` by internal calls to the functions `personpar()` and `threshpar()`. Hence, as the other implemented graphical functions, the function `piplot()` is also model-independent. A restriction to be used for the absolute item threshold parameters δ, both when estimating the person parameters θ and in the call of the generic function `threshpar()`, can be specified with the argument `ref`. The items for which the absolute item threshold parameters δ should be visualized can be specified with the argument `items`. The specification of this argument is identical as for the function `curveplot()` (see Section 4.2.1 for more details). The visual appearance of a person-item plot can be adjusted by several additional arguments. A detailed description of these arguments can be again found on the help site of the function `piplot()`.

4.2.5. Information Plots

Information plots have been implemented in the function `infoplot()`. The interface of this function is shown in code segment 4.5.

As the function `curveplot()`, the function `infoplot()` is implemented model-inde-

```
infoplot(object, what = c("categories", "items", "test"), ref = NULL,
   items = NULL, names = NULL, layout = NULL, xlim = NULL, ylim = NULL,
   col = NULL, lty = NULL, lwd = NULL, main = NULL, legend = TRUE,
   xlab = "Latent trait", ylab = "Information", add = FALSE, ...)
```

Code segment 4.5. Interface of the function `infoplot()`.

pendent. This is again possible by relying on the existence of a model-specific `predict()` method which should be able to compute the category, item and test information for the fitted model specified in the first argument `object`. Such `predict()` methods have been implemented for the three IRT models of the R package psychotools according to the formal definitions given by Muraki (1993). Similar as in the other plotting functions, a restriction to be used for the absolute item threshold parameters δ can be specified with the argument `ref`. If a visualization of the item or category information is requested, a subset of items can be selected by appropriately setting the argument `items`. Similar as in the function `curveplot()`, the argument `layout` allows the modification of the arrangement of the visualized item or category information curves. If it is set to `NULL`, the information curves of different items are overlayed. This is illustrated in more detail in the following Section 4.3. As in the other functions, the visual appearance of the information plots can be adjusted by several additional arguments which are described in more detail on the corresponding help site of the function `infoplot()`.

4.3. Advantages and Application Examples

Based on the proposed unified framework for IRT models in Chapter 3, implementations for all popular visualization techniques have been suggested in the previous sections which do not depend on a specific IRT model. The only necessary condition for an IRT model to be visualized by the introduced graphical functions are the existence of model-specific methods for the generic functions `itempar()`, `threshpar()`, `discrpar()` and `personpar()`. In addition, a model-specific `predict()` method is necessary for the item or category response curve visualization and the visualization of the category, item or test information. As discussed before, all these methods have been implemented for all IRT models of the R package psychotools within this Master's thesis.

A major advantage of the implemented visualization approach is its extensibility to new IRT models. These can be directly visualized by the introduced graphical functions as soon as methods for the generic functions of the various structural components

(or the generic function `predict()`) exist. Hence, also the IRT models implemented in other R packages, e.g., in the R package eRm (Mair et al., 2014) or in the R package ltm (Rizopoulos, 2013), can be visualized by the introduced graphical functions as soon as methods for the generic functions `itempar()`, `threshpar()`, `discrpar()` and `personpar()` are provided for the models of these packages. Extensibility is also possible with respect to the visualization techniques provided. Any imaginable visualization of the different structural components can be easily set up in a R function and if it is build on the generic extractor functions of these components, it will be immediately applicable to all models where methods for these functions exist. For example, a histogram of the person parameters similar to that in the person-item plot illustrated in Figure 3.2 is easily implemented model-independent in the following function `personhist()` which is built around the R-internal function `hist()` and only relies on a method to the generic function `personpar()` for the fitted model specified in the first argument `object`:

```
> personhist <- function (object, ...) {
+     pp <- personpar(object)
+     hist(pp, ...)
+ }
```

Another advantage of the implemented visualization approach is the possibility of graphical comparisons of different IRT models. By abstracting the visualization techniques from a specific IRT model and relying on shared structural components, the visualization of different IRT models can be overlayed in the same illustration and hence compared between each other. Based on the fitted dichotomous Rasch model, the fitted RSM and the fitted PCM, this is illustrated with some examples in the following.

The category response curves predicted under the RSM and the PCM for a single item, e.g., the first item, can be easily compared by two successive calls to the function `curveplot()`:

```
> curveplot(rsmod, items = 1, ref = 1)
> curveplot(pcmod, items = 1, ref = 1, lty = 2, add = TRUE)
```

The result is illustrated in Figure 4.6. To make the results comparable to previous illustrations, the argument `ref` was again set to one, i.e., the first absolute item threshold parameter δ_{11} was restricted to zero. For readability, additional arguments set to modify the visual appearance and add a legend to the result have been omitted in the above calls. As can be seen in Figure 4.6, the predicted probabilities for choosing the different
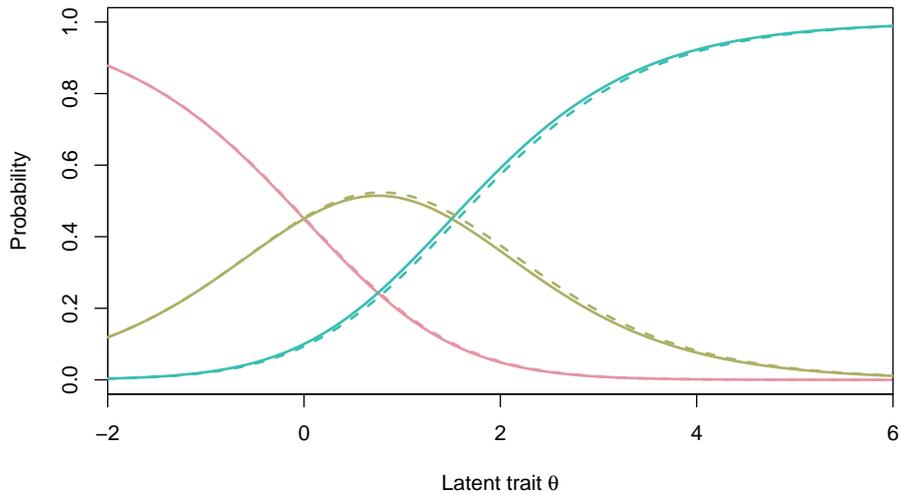
Figure 4.6. Visualization of the category response curves of the 13th item as predicted under a RSM and a PCM each fitted to the items 13-18 of the verbal aggression data.

categories under the two models are almost identical for the 13th item of the verbal aggression data.

While a model comparison with the category response curve visualization in the provided implementation is only possible for a single item, the profile plot visualization allows to compare IRT models with respect to multiple items. For example, the estimated item location parameters $\hat{\boldsymbol{\beta}}$ of all three models for the items 13-18 of the verbal aggression data can be easily compared with the following calls to the function `profileplot()`:

```
> profileplot(rmod, what = "items",
+   paramarg = list(ref = 1), col = "green")
> profileplot(rsmod, what = "items",
+   paramarg = list(ref = 1), col = "blue", add = TRUE)
> profileplot(pcmod, what = "items",
+   paramarg = list(ref = 1), col = "red", add = TRUE)
```

To provide the comparability of the resulting Figure 4.7 to previous illustrations and also to provide the comparability between the different models, the same restriction, i.e., restricting the first item location parameter β_1 to zero, was applied to all three IRT models by setting the `ref` argument to one. Additional arguments which modify the visual appearance have been again omitted for readability in the previous calls. While the estimated absolute item location parameters $\hat{\boldsymbol{\beta}}$ of the items 13 and 14 are nearly
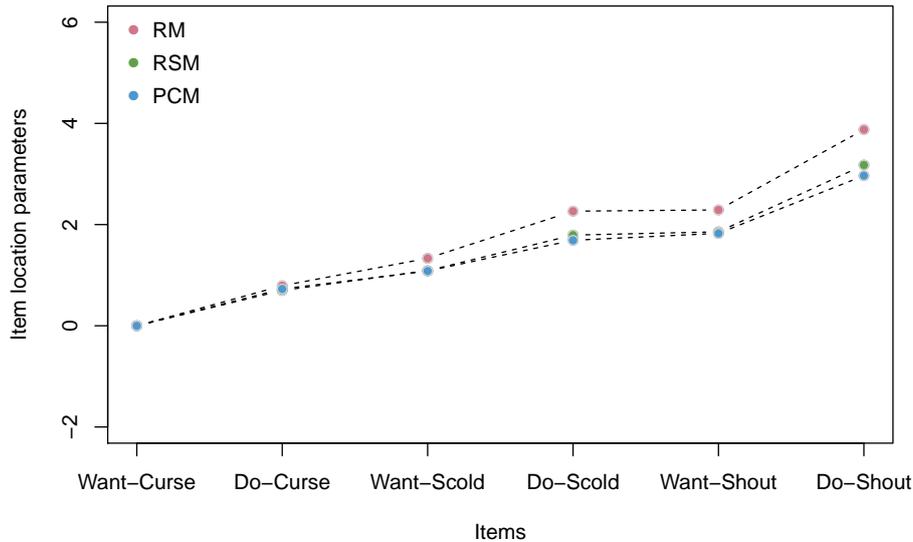
Figure 4.7. Graphical comparison of the estimated item location parameters $\hat{\boldsymbol{\beta}}$ of a dichotomous Rasch Model, a RSM and a PCM fitted to the items 13-18 of the verbal aggression data.

identical between all three IRT models, the estimated item location parameters $\hat{\boldsymbol{\beta}}$ of the remaining items 15-18 are substantially higher in the dichotomous Rasch model compared to the RSM and the PCM. In contrast, the estimates of the RSM and the PCM are rather similar. Figure 4.7 is the graphical pendant to the numerical model comparison based on the estimated item location parameters $\hat{\boldsymbol{\beta}}$ shown in Chapter 3.

A more detailed comparison between the RSM and the PCM can be achieved by a profile plot of the estimated absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ for these two models. Such a plot can be created by the following two calls to the function `profileplot()`:

```
> profileplot(rsmod, what = "threshold",
+   paramarg = list(ref = 1), col = c("blue", "blue"))
> profileplot(pcmod, what = "threshold",
+   paramarg = list(ref = 1), col = c("red", "red"), add = TRUE)
```

The result is illustrated in Figure 4.8 and as before, the same restriction was applied to both models to make the results comparable. Additional arguments which modify the visual appearance have been again omitted for readability. In this more detailed comparison, subtle differences can be seen between the RSM and the PCM. While in the RSM, the distance between the absolute item threshold parameters $\boldsymbol{\delta}$ of an item are assumed to be equal, these can vary between the items in the PCM (see Section 2.1 for

43
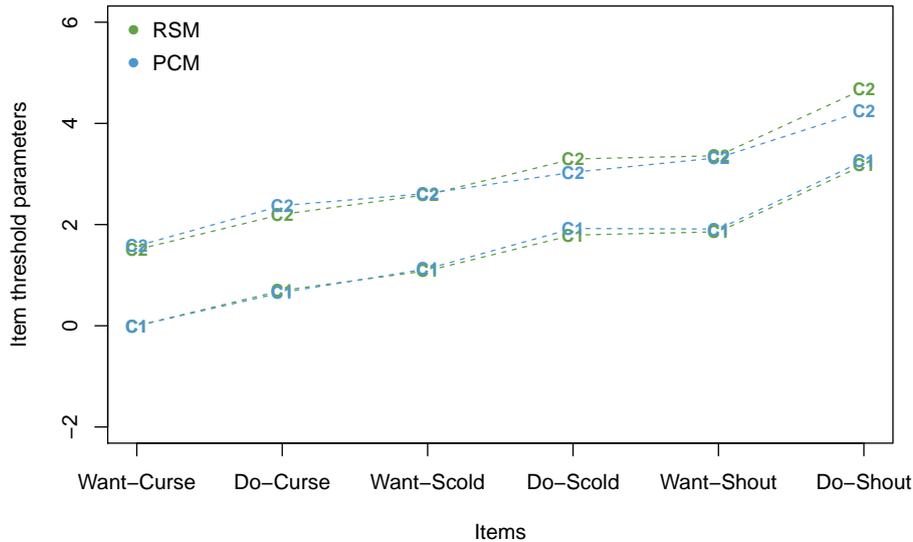
Figure 4.8. Profile plots of the estimated absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ of the RSM and the PCM fitted to the items 13-18 of the verbal aggression data.

a more detailed discussion of the different parametrizations and assumptions). Hence, the variation between the profiles of the estimated absolute item threshold parameters $\hat{\boldsymbol{\delta}}$ between the two IRT models illustrated in Figure 4.8 provides insights with respect to violations of the assumptions and adequacy of the RSM for the analyzed data.

With the implemented visualization of the information of a category, item or test additional insights concerning the response format can be achieved. For example, the information of the items 13-18 of the verbal aggression data with respect to the estimation of the person parameter θ under the dichotomous Rasch model and the PCM can be compared with the following two calls:

```
> infoplot(rmmod, what = "items", lty = 1, xlim = c(-4, 8))
> infoplot(pcmod, what = "items", lty = 2, add = TRUE)
> legend(x = "topright", legend = c("RM", "PCM"), bty = "n", lty = 1:2)
```

The result is illustrated in Figure 4.9. As before, additional arguments which modify the visual appearance have been again omitted for readability in the previous function calls. As can be seen, the item information is generally much higher under the PCM compared to the dichotomous RM. This is a consequence of the underlying polytomous response format of the items under the PCM which allows to capture the response of a subject in more detail compared to the dichotomous response format of the items under the dichotomous RM.
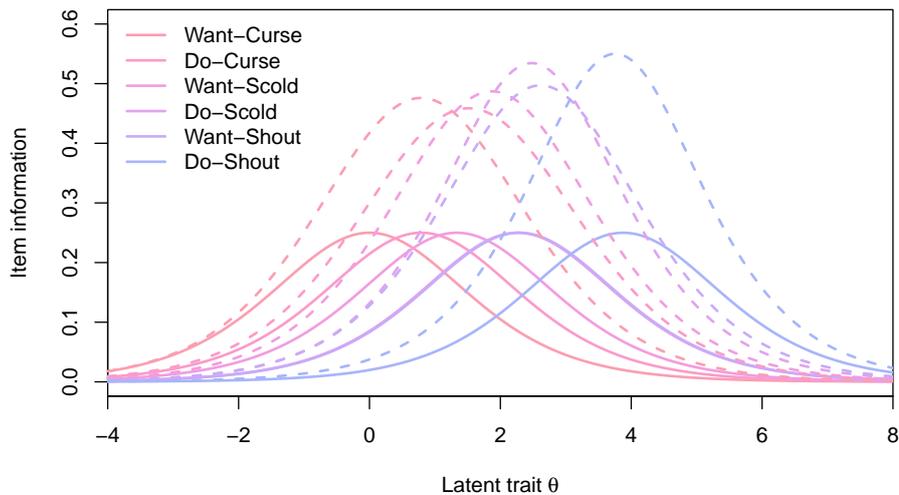
Figure 4.9. Visualization of the item information under a dichotomous RM and a PCM fitted to the items 13-18 of the verbal aggression data.

The category information of the items 13-18 of the verbal aggression data under a PCM can be visualized with the following call:

```
> infoplot(pcmod, what = "categories", items = 1:6, xlim = c(-4, 8)
+   layout = matrix(1:6, ncol = 3, nrow = 2))
```

The result is illustrated in Figure 4.10. By setting the `layout` argument, an arrangement of the category information curves of the different items similar to the arrangement of the category response curve matrix illustrated in Figure 4.1 is achieved. As can be seen from Figure 4.10, the amount of information the middle category provides is directly associated with the width of the region on the dimension of the latent trait where this category is the most probable category.

Overall, the provided and illustrated model-independent implementation of several more established visualization techniques for IRT models offers an helpful visual tool in the process of model selection and also item selection. As was illustrated, different models can be easily compared with respect to their parameters and hence violations of the underlying model assumptions can be identified. In addition, different items can be compared with respect to their properties reflected through their item parameters or their information. As was discussed in the beginning of this section, the provided toolkit can be easily extended to new IRT models and new visualization techniques can be added by relying on the provided unified framework for IRT models discussed in the previous Chapter 3.
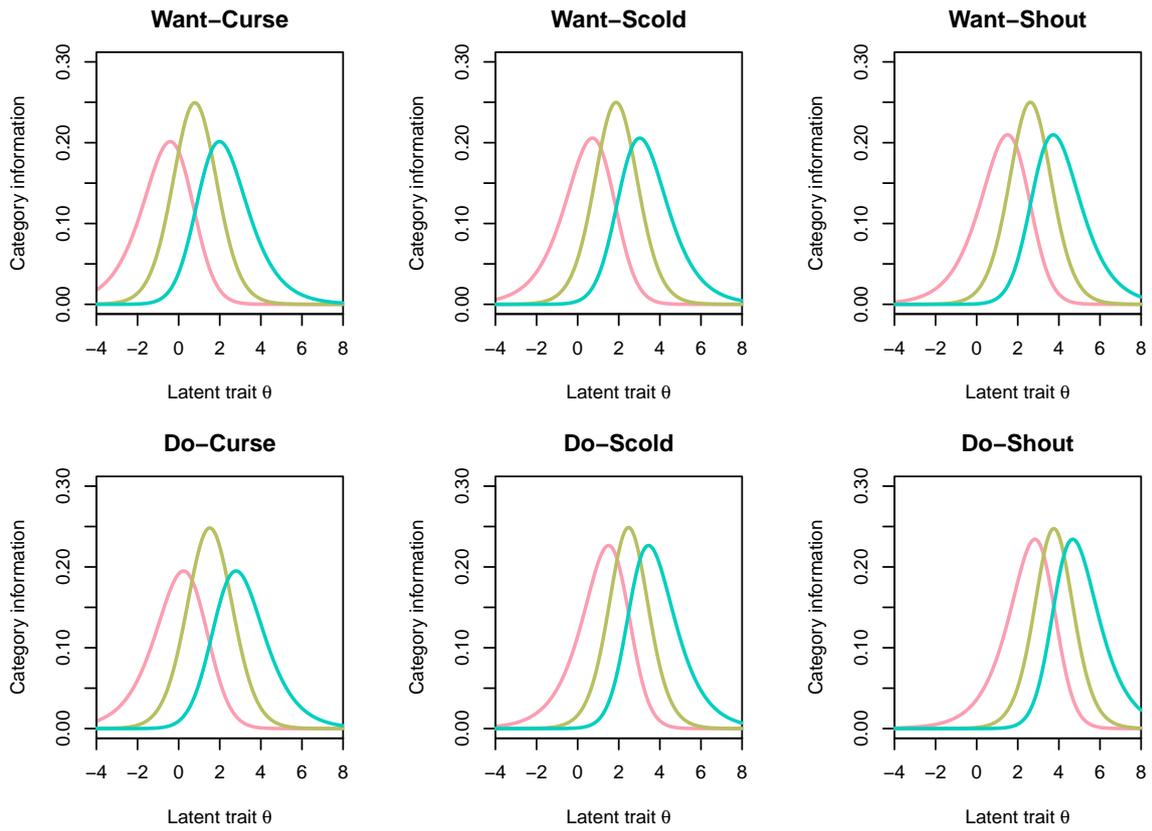
Figure 4.10. Visualization of the category information under a PCM fitted to the items 13-18 of the verbal aggression data.

5. Inference in IRT Models

In this chapter, tools and strategies for statistical inferences in IRT models are discussed and illustrated within several application examples. Most of the presented tools and strategies are based and possible through the unified framework introduced in Chapter 3. Table 5.1 summarizes the illustrated functions. The remaining chapter is organized as following: Section 5.1 focuses around the issue of model selection, Section 5.2 covers the topics Wald coefficient tests, Wald confidence intervals, linear hypotheses and hence, is centered around the estimated model parameters. Section 5.3 illustrates and points out further applications in the context of testing for differential item functioning, i.e., identifying measurement differences in the items of a psychological questionnaire.

Table 5.1. Summary of the illustrated methods and functions for statistical inferences in IRT models. The methods in the first block are the foundation for the generic functions in the second block.

Function	Description
`coef()`	Extract the estimated coefficients of a structural component or a fitted IRT model.
`estfun()`	Extract the individual contributions to the score function of a fitted IRT model.
`logLik()`	Extract the value of the conditional log likelihood at the estimated parameter values.
`vcov()`	Extract the estimated variance-covariance matrix of a structural component or a fitted IRT model.
`AIC()`, `BIC()`	Compute the AIC/BIC information criterion for a fitted IRT model.
`coeftest()`	Partial Wald test of model parameters.
`confint()`	Confidence intervals for one or more parameters of a fitted IRT model.
`lrtest()`	Likelihood ratio test of two nested IRT models.
`linearHypothesis()`	Wald test of linear hypotheses.

5.1. Model Selection

The process of statistical model selection typically involves the computation of one or more information criteria for several related models. In a second step, the model with the smallest value in one or more of the computed information criteria is selected as favorite model. The two most widely used information criteria for this purpose are the Akaike information criterion (AIC),

$$AIC = -2 \cdot \log\left(\mathcal{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{x})\right) + 2 \cdot k,$$

and the Bayesian (or Schwarz) information criterion (BIC),

$$BIC = -2 \cdot \log\left(\mathcal{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{x})\right) + \log(n) \cdot k.$$

Both, the AIC and the BIC, reduce a statistical model to a single number which tries to reflect the goodness of fit of the model to the data represented by two times the negative value of the maximized log-likelihood $\mathcal{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{x})$ while at the same time penalizing for the complexity of the model reflected by some function of the number of estimated parameters k. In the AIC, it is two times the number of estimated parameters k whereas in the BIC, it is $\log(n)$ times the number of estimated parameters k. Hence, given a sample size larger then 100, the AIC tends to prefer more complex models compared to the BIC.

For the three IRT models implemented in the R package psychotools, the AIC and the BIC are computed based on the conditional likelihood $\mathcal{L}_C(\hat{\boldsymbol{\beta}}|\boldsymbol{x}, \boldsymbol{s})$ (see Chapter 2 for more details). With model-specific methods for the generic functions `logLik()` and `nobs()`, both information criteria can be easily computed with the built-in R functions `AIC()` and `BIC()`. Such methods have been implemented for the three IRT models of the R package psychotools. Hence, the AIC of the fitted RSM and the fitted PCM can be computed with the following call to the generic function `AIC()`:

```
> AIC(rsmod, pcmod)

      df      AIC
rsmod  6 1329.798
pcmod 11 1337.672
```

Equivalently, the BIC of these models can be computed with a call to the generic function `BIC()`:

```
> BIC(rsmod, pcmod)

      df      BIC
rsmod  6 1352.333
pcmod 11 1378.985
```

Both information criteria favor the more parsimonious RSM compared to the more flexible PCM for the items 13-18 of the verbal aggression data. As expected, this preference is stronger in the more conservative BIC compared to the more liberal AIC. Overall, the above results indicate that the numerical differences between the estimated parameters of the two hierarchically related models discussed in Chapter 3 and visualized in Figure 4.8 of Chapter 4 are negligible.

With the implemented methods for the generic function `logLik()`, a variety of tools provided by other R packages are readily available. For example, a more formal comparison of the two hierarchically related models based on an asymptotic (conditional) likelihood ratio test (LRT) is directly possible by using the function `lrtest()` of the R package lmtest (Zeileis & Hothorn, 2002):

```
> library("lmtest")
> lrtest(rsmod, pcmod)

Likelihood ratio test

Model 1: rsmod
Model 2: pcmod
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   6 -658.90
2  11 -657.84  5 2.1264     0.8314
```

Similar as the two information criteria, the result of the asymptotic LRT indicates that the more parsimonious RSM is sufficient to adequately describe the responses of the 316 subjects to the items 13-18 of the verbal aggression data.

5.2. Hypotheses Tests and Confidence Intervals

On the level of the estimated parameters of an IRT model, Wald coefficient tests, Wald confidence intervals and linear hypotheses are available as inferential tools. All of these tools are founded on the asymptotic normality of the estimated model parameters given in the (conditional) likelihood framework. Based on the implemented methods and functions shown in Table 5.1, these inferencial tools can be easily applied as shown in the following.

Wald confidence intervals and Wald coefficient tests are founded on the provided `coef()` and `vcov()` methods for all structural components discussed in Chapter 3 and are readily available through the application of several built-in or add-on R functions. For example, Wald confidence intervals for the estimated item location parameters $\hat{\boldsymbol{\beta}}$ of the fitted dichotomous Rasch model can be computed by the built-in R function `confint()` as shown in the following. The significance level is set to $\alpha = 0.05$ per default. By setting the argument `alias` to `FALSE`, the aliased, i.e., restricted, item location parameter β_1 is not included in the output.

```
> confint(itempar(rmmod, ref = 1, alias = FALSE))

              2.5 %    97.5 %
Do-Curse    0.4072442 1.175733
Want-Scold  0.9319176 1.735087
Do-Scold    1.8246728 2.705187
Want-Shout  1.8482788 2.731184
Do-Shout    3.2962717 4.458728
```

Wald coefficient tests are available by the add-on function `coeftest()` of the R package lmtest (Zeileis & Hothorn, 2002). The significance level is again set to $\alpha = 0.05$ per default and the null and alternative hypothesis tested for each estimated item location parameter $\hat{\beta}_j$ is

$$H_0 : \hat{\beta}_j = 0 \quad \text{vs.} \quad H_1 : \hat{\beta}_j \neq 0.$$

The aliased, i.e., restricted, item location parameter β_1 is again not included in the output by setting the argument `alias` to `FALSE`.

```
> coeftest(itempar(rmmod, ref = 1, alias = FALSE))

z test of coefficients:
```

```
          Estimate Std. Error z value  Pr(>|z|)
Do-Curse    0.79149    0.19605  4.0372 5.408e-05 ***
Want-Scold  1.33350    0.20489  6.5083 7.603e-11 ***
Do-Scold    2.26493    0.22463 10.0832 < 2.2e-16 ***
Want-Shout  2.28973    0.22523 10.1660 < 2.2e-16 ***
Do-Shout    3.87750    0.29655 13.0754 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the provided `coef()` and `vcov()` methods for the different structural components, the add-on function `linearHypothesis()` of the R package `car` (Fox & Weisberg, 2011) allows to test arbitrary linear hypotheses of the estimated or computed model parameters. This offers interesting possibilities. For example, the question whether the more parsimonious RSM is sufficient compared to the more flexible PCM to adequately describe the responses of the 316 subjects to the items 13-18 of the verbal aggression data can be checked with the following linear hypothesis on the estimated relative item threshold parameters $\hat{\tau}_{jk}$ of the fitted PCM:

$$H_0 : \boldsymbol{C}\hat{\boldsymbol{\tau}} = \boldsymbol{0} \quad \text{vs.} \quad H_1 : \boldsymbol{C}\hat{\boldsymbol{\tau}} \neq \boldsymbol{0} \tag{5.1}$$

with

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}, \quad \hat{\boldsymbol{\tau}} = \begin{pmatrix} \hat{\tau}_{12} \\ \hat{\tau}_{22} \\ \hat{\tau}_{32} \\ \hat{\tau}_{42} \\ \hat{\tau}_{52} \\ \hat{\tau}_{62} \end{pmatrix} \quad \text{and} \quad \boldsymbol{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

More specifically, the above formulated linear hypothesis tests whether the estimated item-specific relative threshold parameters $\hat{\tau}_{jk}$ of the fitted PCM are the same across all items, i.e., whether a single, global, relative item threshold parameter as assumed in the RSM is sufficient. The hypothesis is formulated under the restriction that the first relative item threshold parameter τ_{j1} is restricted to zero for all items. With the implemented interface for the relative item threshold parameters $\boldsymbol{\tau}$, the following call to the generic function `threshpar()` extracts the estimated relative item threshold parameters $\hat{\boldsymbol{\tau}}$ from the fitted PCM under this restriction. The aliased, i.e., restricted, relative item threshold parameters τ_{j1} are again omitted from the result object by setting the

argument `alias` to `FALSE`:

```
> tp <- threshpar(pcmod, type = "mode",
+                     relative = TRUE, ref = 1, alias = FALSE)
```

To test the hypothesis stated in Equation (5.1), the contrast matrix C is constructed in a first step with the following call:

```
> C <- cbind(1, diag(-1, nrow = 5, ncol = 5))
```

In a second step, the R package `car` is loaded and the linear hypothesis is tested with a call to the function `linearHypothesis()`:

```
> library("car")
> linearHypothesis(tp, hypothesis.matrix = C, rhs = rep.int(0, 5))

Linear hypothesis test

Hypothesis:
Want - Curse - C2 - Do - Curse - C2 = 0
Want - Curse - C2 - Want - Scold - C2 = 0
Want - Curse - C2 - Do - Scold - C2 = 0
Want - Curse - C2 - Want - Shout - C2 = 0
Want - Curse - C2 - Do - Shout - C2 = 0

Model 1: restricted model
Model 2: tp

  Df  Chisq Pr(>Chisq)
1
2  5 2.1839     0.8232
```

As in the LRT computed in the previous section, the result indicates that the more parsimonious RSM is sufficient to adequately describe the responses of the 316 subjects to the items 13-18 of the verbal aggression data.

5.3. Testing for DIF

Besides the previously discussed inferencial tools and strategies, the unified framework for IRT models introduced in Chapter 3 provides the foundation for further applications in the context of testing for differential item functioning. The term differential item functioning (DIF) here denotes differences in the measurement properties of an item of a psychological questionnaire between two or more groups of subjects. Due to reasons of fairness, such differences are tried to be avoided. In an IRT modeling context, DIF is typically operationalized as differences in the estimated item parameters between two or more groups of subjects and various statistical procedures have been suggested to detect DIF (for a review see, e.g., Holland & Wainer, 1993).

5.3.1. Global Likelihood Ratio Tests

One more popular statistical procedure to detect DIF is the likelihood ratio test (LRT). It provides a global DIF detection procedure testing whether there is DIF between two or more groups in any of the items under observation. Based on the provided `logLik()` method, the LRT can be easily computed. This is illustrated in more detail in the following example where the items 13-18 of the verbal aggression data are tested for DIF between the group of male and female subjects. In a first step, a dichotomous Rasch model is fitted to the responses of all subjects and separately to the responses of male and female subjects:

```
> rmall <- raschmodel(dat$dich)
> rmmale <- raschmodel(subset(dat, gender == "male")$dich)
> rmfemale <- raschmodel(subset(dat, gender == "female")$dich)
```

The LRT test statistic

$$LRT = -2 \cdot \log \left(\frac{\mathcal{L}_C(\hat{\boldsymbol{\beta}}^{(a)}|\boldsymbol{x}^{(a)}, \boldsymbol{s})}{\mathcal{L}_C(\hat{\boldsymbol{\beta}}^{(m)}|\boldsymbol{x}^{(m)}, \boldsymbol{s}) \cdot \mathcal{L}_C(\hat{\boldsymbol{\beta}}^{(f)}|\boldsymbol{x}^{(f)}, \boldsymbol{s})} \right)$$

for the above situation is easily computed with the provided `logLik()` method:

```
> (LRT <- as.vector(- 2 * (logLik(rmall) -
+                         (logLik(rmmale) + logLik(rmfemale)))))

[1] 13.99331
```

The corresponding p-value is obtained by the function `pchisq()`:

```
> pchisq(LRT, df = 5, lower.tail = FALSE)
```

```
[1] 0.01565199
```

As the p-value of the computed LRT is below the significance level of 5%, the null hypothesis of no DIF between male and female subjects in any of the six items is rejected. Whereas the LRT as global DIF detection procedure does indicate *if* there is DIF in any of the six items of the verbal aggression data set, it does not indicate *which* items are affected by the DIF. For this purpose, an item-wise DIF detection procedure as introduced in the next section can provide further insights.

5.3.2. Item-wise Wald Tests and Anchor Selection

In contrast to the LRT as global DIF detection procedure, the classical Wald test with test statistic

$$T_j = \frac{\hat{\beta}_j^{(m)} - \hat{\beta}_j^{(f)}}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_j^{(m)}) + \widehat{\mathrm{Var}}(\hat{\beta}_j^{(f)})}} \tag{5.2}$$

for the j-th item in the above situation provides an item-wise DIF detection procedure. However, before individual items can be tested for DIF with the Wald test from Equation (5.2), a common measurement scale for the estimated item location parameters $\hat{\boldsymbol{\beta}}$ of both groups has to be established. This is done by imposing the same restriction in both groups. In the dichotomous Rasch model, this is also known as anchor selection and several strategies have been suggested to select the items forming the anchor (for a recent review and assessment, see Kopf et al., 2014). Both, the process of establishing a common measurement scale as well as computing the item-wise Wald tests can be done with the implemented interface for the item location parameters $\boldsymbol{\beta}$ as illustrated in the following.

Assuming the second and third item of the analyzed subset of the verbal aggression data have been detected as the final set of anchor items, a common measurement scale, i.e., imposing the same restriction on the estimated item location parameters $\hat{\boldsymbol{\beta}}$ of both groups is easily done with the following two calls to the generic function `itempar()`:

```
> ipmale <- itempar(rmmale, ref = c(2, 3))
> ipfemale <- itempar(rmfemale, ref = c(2, 3))
```

In a second step, the test statistic T_j from Equation (5.2) and corresponding p-values are computed for all m items simultaneously with the following two commands:

```
> waldtests <- (coef(ipmale) - coef(ipfemale)) /
+     sqrt(diag(vcov(ipmale)) + diag(vcov(ipfemale)))
> pvalues <- 2*pnorm(abs(waldtests), lower.tail = FALSE)
```

A summary of the test statistics T_j and their corresponding p-values is easily created by, e.g., binding the two objects together as shown in the following command:

```
> cbind("Test Statistic" = waldtests, "P-Value" = pvalues)
```

```
           Test Statistic   P-Value
Want-Curse    2.254109205 0.0241893
Do-Curse     -0.849600648 0.3955472
Want-Scold    0.849600648 0.3955472
Do-Scold     -0.443541733 0.6573739
Want-Shout    2.503493232 0.0122974
Do-Shout      0.007902345 0.9936949
```

With the second and third item as anchors, the first and fifth item of the verbal aggression data subset have been identified as showing DIF between the group of male and female subjects in the item-wise Wald tests. However, three limitations of this result should be noted: First, due to the single restriction necessary when estimating the dichotomous Rasch model, only five of the six DIF tests can be meaningfully interpreted (for more details, see Kopf et al., 2014), secondly, even though multiple item-wise Wald tests have been computed in the above example, no correction for multiple testing was applied, and thirdly, the result of the item-wise DIF test is dependent on the chosen anchor items. In the previous example, the second and third item have been arbitrarily selected as anchor items. Based on the findings of Kopf et al. (2013, 2014), several more systematic and established anchor selection strategies have been implemented in the function `anchor()`. The function `anchortest()` combines the functionality of `anchor()` and additionally provides the results of item-wise Wald tests implemented in a similar style as shown above but leaving the first final anchor item with the lowest rank out and also allowing to apply a correction for the p-values. Hence, in the following call to the function `anchortest()`, the anchor items are first selected according to the recommendations in Kopf et al. (2013, 2014), i.e., based on the mean p-value threshold selection

criterion in combination with the iterative-forward anchor class and in a second step, item-wise Wald tests are computed to test for DIF between the group of male and female subjects. In addition, the *p*-values of the item-wise Wald tests are adjusted according to the correction suggested by Holm (1979):

```
> anchortest(rmfemale, rmmale, class = "forward",
+            method = "MPT", adjust = "holm")

Anchor items:
[1] 6 4 2


Anchored item parameters:
Want-Curse_1   Do-Curse_1 Want-Scold_1   Do-Scold_1 Want-Shout_1
 -2.60329872  -1.52326466  -1.08032960  -0.02619854  -0.37042813
Want-Curse_2   Do-Curse_2 Want-Scold_2   Do-Scold_2 Want-Shout_2
 -1.59174180  -1.59174179  -0.75634267  -0.09009544   0.92504657


Ranking order:
[1] 3 6 4 2 1 5


Criterion values (not sorted):
[1] 4 2 2 2 4 2


Final DIF tests:


        Simultaneous Tests for General Linear Hypotheses


Linear Hypotheses:
              Estimate Std. Error z value Pr(>|z|)
Want-Curse == 0 -1.01156    0.41330  -2.448   0.0575 .
Do-Curse == 0    0.06848    0.31736   0.216   1.0000
Want-Scold == 0 -0.32399    0.40725  -0.796   1.0000
Do-Scold == 0    0.06390    0.30662   0.208   1.0000
Want-Shout == 0 -1.29547    0.44377  -2.919   0.0175 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- holm method)
```

As can be seen in the above output, the second, fourth and sixth item have been identified as anchor items according to the mean p-value threshold selection criterion in combination with the iterative-forward anchor class. Based on these anchor items, only the sixth item is identified as showing DIF between the group of male and female subjects, i.e., showing a p-value below the significance level of 5%. This result is in accordance with the results of the LRT and the item-wise Wald tests reported above.

5.3.3. Trees and Generalized M-Fluctuation Tests

A further application of the provided methods and functions for statistical inference in IRT models can be found in the DIF detection procedures suggested by Strobl et al. (2013) and Abou El-Komboz et al. (2014). In contrast to the LRT and the item-wise Wald tests illustrated above, no pre-specification of the groups tested for DIF is required in these procedures as they are detected in a data-driven way by relying on the *model-based recursive partitioning* (MOB) algorithm suggested by Zeileis et al. (2008). This algorithm allows to recursively identify DIF groups based on a number of given binary, categorical or numerical covariates. A crucial aspect of this algorithm are the individual contributions to the score function of an IRT model. These individual score contributions are repeatedly ordered with respect to the available covariates and a suitable test of the class of generalized M-fluctuation tests for parameter instability (Zeileis & Hornik, 2007) is then applied to identify groups with DIF. The derivation of the individual score contributions of the three IRT models of the R package psychotools can be found in Appendix B. Their computation has been implemented within model-specific methods for the generic function `estfun()`. The generic function is provided by the R package sandwich (Zeileis, 2006). After loading this package, the individual score contributions of the fitted dichotomous Rasch model from above can be easily extracted and ordered with respect to the continuous covariate `anger` by the following command:

```
> library("sandwich")
> orderedscores <- estfun(rmall)[order(dat$anger),]
```

The DIF detection procedures suggested by Strobl et al. (2013) and Abou El-Komboz et al. (2014) which are based on the above discussed `estfun()` methods are implemented in the related R package psychotree (Zeileis et al., 2014). After loading this package, the function `raschtree()` allows the application of the DIF detection procedure suggested by Strobl et al. (2013) in the context of the dichotomous Rasch model. This is illustrated in more detail in the following. In a first step, the R package psychotree is loaded:

```
> library("psychotree")
```

With the following call to the function `raschtree()`, the items 13-18 of the verbal aggression data set are tested for DIF with respect to the two covariates `gender` and `anger`:

```
> rt <- raschtree(dich ~ gender + anger, data = dat)
```

A visualization of the resulting Rasch tree is obtained with the following call to the `plot()` method where custom item labels are set for readability and the abbreviation of the item labels is turned off:

```
> plot(rt, tp_args = list(names = lbs, abbreviate = FALSE))
```
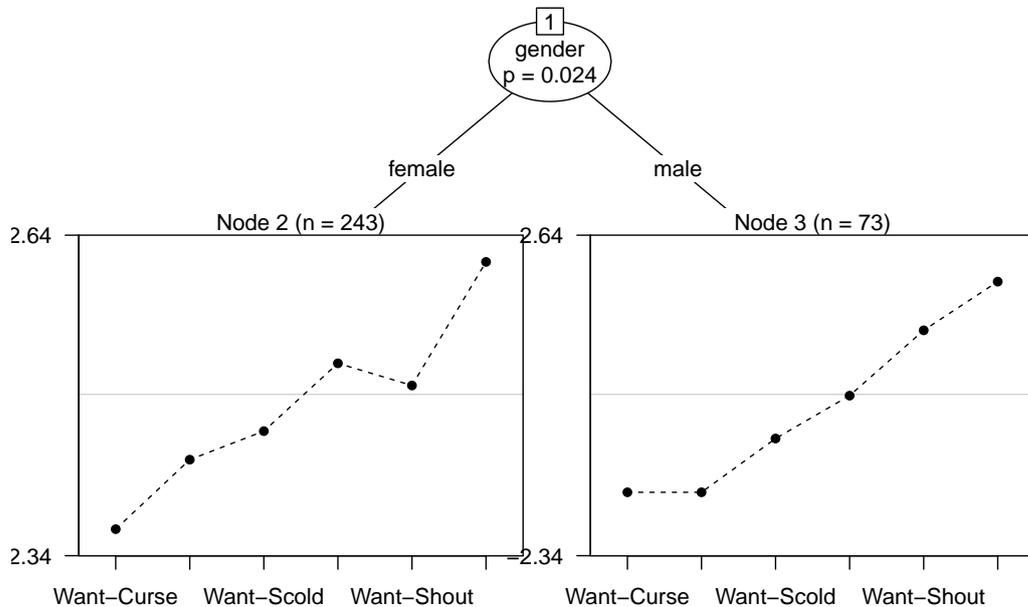

Figure 5.1. Visualization of a Rasch tree (Strobl et al., 2013) fitted to the items 13-18 of the verbal aggression data.

The resulting Rasch tree is illustrated in Figure 5.1. As there is more than one terminal node, the null hypothesis of no DIF in any of the items 13-18 of the verbal aggression data is rejected. In contrast to the previous approaches, both covariates `gender` and `anger` have been used simultaneously here to identify DIF but again, only the covariate `gender` was identified to be associated with DIF.

6. Discussion and Outlook

The objective of this Master's thesis was to develop a unified framework for visualization and inference in IRT models which combines the advantages of highly specialized model-fitting functions as implemented, e.g., in the R package psychotools, with the advantages of a general framework as implemented, e.g., in the R package eRm. Based on the GPCM as theoretical framework, four contextually different structural components of IRT models have been identified and implemented in a suitable representation in the R package psychotools. Several established visualization techniques and tools for inference have been developed on the foundation provided by these structural components. Overall, the structural components, the visualization techniques and the tools for inference form the computational part of the suggested unified framework for visualization and inference in IRT models.

By separating the tools for visualization and inference from a specific IRT model, the suggested framework easily allows to compare related IRT models either numerically or graphically as was illustrated in several application examples throughout this Master's thesis. In addition, the implemented tools for inference provide the foundation for further applications in related areas, e.g., when detecting DIF.

With the GPCM as theoretical foundation, a wide class of parametric IRT models often labeled as "divide-by-total models" (Sijtsma & Hemker, 2000, p. 404) or "adjacent-category models" (Mellenbergh, 1995, p. 93) are covered by the suggested framework. However, two other classes of parametric IRT models also discussed in the literature, the "cumulative probability models" and the "continuation-ratio models" (Mellenbergh, 1995, p. 93), are not covered. Although the same structural components are existent in the models of these model classes, the parametrized response and therefore also the interpretation of the structural components differs. In cumulative probability models, the cumulative probability of choosing a category k *or higher* is described by the model equation and hence, the interpretation of the structural components therefore also refers to these cumulative probabilities. In continuation-ratio models, the conditional probability of choosing a category k *given a category k or higher* is described by the model equation and hence, the interpretation of the structural components here refers

to these conditional probabilities (see Mellenbergh, 1995, for more details). Because of these distinctions between the different model classes, cumulative probability models and continuation-ratio models can not be easily incorporated in the suggested framework.

All three IRT models of the R package psychotools have been integrated into the suggested framework by providing extractor functions for the various structural components. A crucial aspect here has been the issue of parameter identifiability which was discussed in more detail in this thesis for the CML approach used for estimating these models. As was briefly pointed out, often a MML approach instead of the discussed CML approach is used in more complex IRT models where the item discrimination parameters are not fixed to unity. In the MML approach, the issue of parameter identifiability becomes more complex as not only the location but also the scale of the estimated model parameters has to be fixed. The integration of IRT models estimated with a MML approach into the suggested framework therefore provides a challenge for future research.

The suggested unified framework can be extended in two ways and thus provides additional opportunities for further research: More IRT models related to the GPCM, e.g., from other R packages, can be integrated by providing methods for the generic functions of the various structural components. After providing these methods, all existing tools for visualization and inference are directly available for these models. Another opportunity for further research is the implementation of new tools for visualization and inference which can be implemented on top of the structural components and are then directly applicable to a wide class of IRT models.

A. Existing Parametrizations in the GPCM

The following Table A.1 gives an overview of the different parametrizations available in the GPCM (and hence all related models) and the transformations necessary to convert one type of parameter into another type of parameter.

Table A.1. Overview of the different parametrizations in the GPCM (and hence all related IRT models) as well as their relations to each other. The direction of the relations is from column to row, i.e., the operation in cell ij gives the transformation necessary to be applied on the type of parameter in column j to arrive at the type of parameter in row i. Often, an additional type of parameter is necessary to transform one type of parameter into another type of parameter.

	Item Location Parameter	Absolute Item Threshold Parameter	Cumulative Absolute Item Threshold Parameter	Relative Item Threshold Parameter	Cumulative Relative Item Threshold Parameter
	β_j	δ_{jk}	η_{jk}	$\tau_{(j)k}$	$\kappa_{(j)k}$
β_j	Identity	$\frac{1}{p_j} \sum_{k=1}^{p_j} \delta_{jk}$	$\frac{1}{p_j}(\sum_{k=2}^{p_j}(\eta_{jk} - \eta_{j(k-1)}) + \eta_{j1})$	$\delta_{jk} - \tau_{(j)k}$	$\delta_{jk} - (\kappa_{(j)k} - \kappa_{(j)k-1})$
δ_{jk}	$\beta_j + \tau_{(j)k}$	Identity	$\eta_{jk} - \eta_{j(k-1)}$	$\beta_j + \tau_{(j)k}$	$\beta_j + (\kappa_{(j)k} - \kappa_{(j)k-1})$
η_{jk}	$k \cdot \beta_j + \sum_{\ell=1}^{k} \tau_{(j)\ell}$	$\sum_{\ell=1}^{k} \delta_{j\ell}$	Identity	$k \cdot \beta_j + \sum_{\ell=1}^{k} \tau_{(j)\ell}$	$k \cdot \beta_j + \kappa_{(j)k}$
$\tau_{(j)k}$	$\delta_{jk} - \beta_j$	$\delta_{jk} - \beta_j$	$(\eta_{jk} - \eta_{j(k+1)}) - \beta_j$	Identity	$\kappa_{(j)k} - \kappa_{(j)k-1}$
$\kappa_{(j)k}$	$\sum_{\ell=1}^{k} \delta_{j\ell} - k \cdot \beta_j$	$\sum_{\ell=1}^{k} \delta_{j\ell} - k \cdot \beta_j$	$\eta_{jk} - k \cdot \beta_j$	$\sum_{\ell=1}^{k} \tau_{(j)\ell}$	Identity

B. Individual Score Contributions

Based on the individual contributions to the conditional likelihood, the individual contributions to the (conditional) score function of the dichotomous Rasch model, the RSM and the PCM are derived in the following. The computation of these individual score contributions of a fitted model object has been implemented in model-specific methods for the generic function `estfun()` in the R package psychotools. An application of these methods within the context of detecting DIF was illustrated in Chapter 5.

The contribution of subject i to the conditional log likelihood $\Psi_{C,i}(\boldsymbol{\beta}|\boldsymbol{x}_i, s_i)$ of the dichotomous Rasch model is given as

$$
\begin{aligned}
\Psi_{C,i}(\boldsymbol{\beta}|\boldsymbol{x}_i, s_i) &= \log(\mathcal{L}_{C,i}(\boldsymbol{\beta}|\boldsymbol{x}_i, s_i)) \\
&= \log\left(\frac{\exp\left(-\sum_{j=1}^{m} x_{ij} \cdot \beta_j\right)}{\gamma_{s_i}(\boldsymbol{\beta})}\right) \\
&= -\sum_{j=1}^{m} x_{ij} \cdot \beta_j - \log(\gamma_{s_i}(\boldsymbol{\beta}))
\end{aligned}
$$

where $\boldsymbol{x}_i$ denotes the response vector and s_i the raw score of subject i and γ_{s_i} denotes the elementary symmetric function of order s_i. The individual score contributions $\psi_{C,i}(\boldsymbol{\beta}|\boldsymbol{x}_i, s_i)$ to the item location parameter β_j of the j-th item then follow as

$$
\begin{aligned}
\psi_{C,i}(\boldsymbol{\beta}|\boldsymbol{x}_i, s_i)_j &= \frac{\partial \Psi_{C,i}(\boldsymbol{\beta}|\boldsymbol{x}_i, s_i)}{\partial \beta_j} \\
&= \frac{\partial - \sum_{j=1}^{m} x_{ij} \cdot \beta_j - \log(\gamma_{s_i}(\boldsymbol{\beta}))}{\partial \beta_j} \\
&= -x_{ij} - \frac{1}{\gamma_{s_i}(\boldsymbol{\beta})} \cdot \frac{\partial \gamma_{s_i}(\boldsymbol{\beta})}{\partial \beta_j}.
\end{aligned}
$$

The contribution of subject i to the conditional log likelihood $\Psi_{C,i}(\boldsymbol{\beta}, \boldsymbol{\tau}|\boldsymbol{x}_i, s_i)$ of the

RSM is given as

$$
\begin{aligned}
\Psi_{C,i}(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{x}_i, s_i) &= \log(\mathcal{L}_{C,i}(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{x}_i, s_i)) \\
&= \log \left(\frac{\exp\left(-\sum_{j=1}^{m} (x_{ij} \cdot \beta_j + \sum_{k=1}^{x_{ij}} \tau_k) \right)}{\gamma_{s_i}(\boldsymbol{\beta}, \boldsymbol{\tau})} \right) \\
&= -\sum_{j=1}^{m} (x_{ij} \cdot \beta_j + \sum_{k=1}^{x_{ij}} \tau_k) - \log(\gamma_{s_i}(\boldsymbol{\beta}, \boldsymbol{\tau})).
\end{aligned}
$$

The individual score contributions $\psi_{C,i}(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{x}_i, s_i)$ to the item location parameter β_j of the j-th item then follow as

$$
\begin{aligned}
\psi_{C,i}(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{x}_i, s_i)_j &= \frac{\partial \Psi_{C,i}(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{x}_i, s_i)}{\partial \beta_j} \\
&= \frac{\partial - \sum_{j=1}^{m} (x_{ij} \cdot \beta_j + \sum_{k=1}^{x_{ij}} \tau_k) - \log(\gamma_{s_i}(\boldsymbol{\beta}, \boldsymbol{\tau}))}{\partial \beta_j} \\
&= -x_{ij} - \frac{1}{\gamma_{s_i}(\boldsymbol{\beta}, \boldsymbol{\tau})} \cdot \frac{\partial \gamma_{s_i}(\boldsymbol{\beta}, \boldsymbol{\tau})}{\partial \beta_j}.
\end{aligned}
$$

Equivalently, the individual score contributions $\psi_{C,i}(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{x}_i, s_i)$ to the relative item threshold parameter τ_k of the k-th category then follow as

$$
\begin{aligned}
\psi_{C,i}(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{x}_i, s_i)_k &= \frac{\partial \Psi_{C,i}(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{x}_i, s_i)}{\partial \tau_k} \\
&= \frac{\partial - \sum_{j=1}^{m} (x_{ij} \cdot \beta_j + \sum_{k=0}^{x_{ij}} \tau_k) - \log(\gamma_{s_i}(\boldsymbol{\beta}, \boldsymbol{\tau}))}{\partial \tau_k} \\
&= -n_k - \frac{1}{\gamma_{s_i}(\boldsymbol{\beta}, \boldsymbol{\tau})} \cdot \frac{\partial \gamma_{s_i}(\boldsymbol{\beta}, \boldsymbol{\tau})}{\partial \tau_k}
\end{aligned}
$$

where n_k denotes the number of times subject i has chosen category k or higher.

The contribution of subject i to the conditional log likelihood $\Psi_{C,i}(\boldsymbol{\delta} | \boldsymbol{x}_i, s_i)$ of the

PCM is given as

$$\Psi_{C,i}(\boldsymbol{\delta}|\boldsymbol{x}_i, s_i) = \log(\mathcal{L}_{C,i}(\boldsymbol{\delta}|\boldsymbol{x}_i, s_i))$$

$$= \log\left(\frac{\exp\left(-\sum_{j=1}^{m}\sum_{k=0}^{x_{ij}}\delta_{jk}\right)}{\gamma_{s_i}(\boldsymbol{\delta})}\right)$$

$$= -\sum_{j=1}^{m}\sum_{k=0}^{x_{ij}}\delta_{jk} - \log(\gamma_{s_i}(\boldsymbol{\delta}))$$

with $\delta_{j0} \equiv 0$ for notational purposes.

The individual score contributions $\psi_{C,i}(\boldsymbol{\beta}, \boldsymbol{\tau}|\boldsymbol{x}_i, s_i)$ to the absolute item threshold parameter δ_{jk} of the j-th item and the k-th category then follow as

$$\psi_{C,i}(\boldsymbol{\delta}|\boldsymbol{x}_i, s_i)_{jk} = \frac{\partial\Psi_{C,i}(\boldsymbol{\delta}|\boldsymbol{x}_i, s_i)}{\partial\delta_{jk}}$$

$$= \frac{\partial - \sum_{j=1}^{m}\sum_{k=0}^{x_{ij}}\delta_{jk} - \log(\gamma_{s_i}(\boldsymbol{\delta}))}{\partial\delta_{jk}}$$

$$= -I_{[x_{ij}\geq k]}(x_{ij}) - \frac{1}{\gamma_{s_i}(\boldsymbol{\delta})} \cdot \frac{\partial\gamma_{s_i}(\boldsymbol{\delta})}{\partial\delta_{jk}}$$

with $I_{[x_{ij}\geq k]}$ as an indicator function returning one if subject i has chosen category k or higher on item j and zero otherwise.

Bibliography

Abou El-Komboz, B., Zeileis, A., & Strobl, C. (2014). Detecting Differential Item and Step Functioning with Rating Scale and Partial Credit Trees. Technical Report 152, Ludwig-Maximilians-Universität München.

Adams, R. J., Wu, M. L., & Wilson, M. (2012). The Rasch Rating Model and the Disordered Threshold Controversy. *Educational and Psychological Measurement, 72*(4), 547–573.

Aiken, L. R. (1994). *Psychological Testing and Assessment* (8 ed.). Needham, Heights, Massachusetts: Allyn and Bacon.

Andersen, E. (1995). Polytomous Rasch Models and their Estimation. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch Models: Foundations, Recent Developments, and Applications* (pp. 271–291). New York: Springer.

Andersen, E. B. (1972). The Numerical Solution of a Set of Conditional Estimation Equations. *Journal of the Royal Statistical Society. Series B (Methodological), 34*(1), 42–54.

Andersen, E. B. (1973). *Conditional Inference and Models for Measuring.* Copenhagen: Mentalhygiejnisk Forlag.

Andersen, E. B. (1977). Sufficient Statistics and Latent Trait Models. *Psychometrika, 42*(2), 69–81.

Andrich, D. (1978). A Rating Formulation for Ordered Response Categories. *Psychometrika, 43*(2), 561–573.

Andrich, D. (2013). An Expanded Derivation of the Threshold Structure of the Polytomous Rasch Model that Dispels any 'Threshold Disorder Controversy'. *Educational and Psychological Measurement, 73*(1), 553–580.

Baker, F. B. & Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques.* (2 ed.). New York: CRC Press.

Birnbaum, A. (1968). Some Latent Trait Models and their Use in Inferring an Examinee's Ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical Theories of Mental Test Scores* (pp. 353–370). Reading, MA: Addison-Wesley.

Bock, R. & Aitkin, M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. *Psychometrika*, *46*(4), 443–459.

Bock, R. & Lieberman, M. (1970). Fitting a Response Model for Dichotomously Scored Items. *Psychometrika*, *35*(2), 179–197.

Casella, G. & Berger, R. (2001). *Statistical Inference* (2 ed.). Duxbury: Thomson Press.

De Boeck, P. & Wilson, M. (2004). *Explanatory Item Response Models. A Generalized Linear And Nonlinear Approach.* New York: Springer.

Eggen, T. J. & Verhelst, N. D. (2006). Loss of Information in Estimating Item Parameters in Incomplete Designs. *Psychometrika*, *71*(2), 303–322.

Fischer, G. (1981). On the Existence and Uniqueness of Maximum-Likelihood Estimates in the Rasch Model. *Psychometrika*, *46*(1), 59–77.

Fischer, G. H. & Molenaar, I. W. (1995). *Rasch Models: Foundations, Recent Developments, and Applications.* New York: Springer.

Fischer, G. H. & Ponocny, I. (1994). An Extension of the Partial Credit Model with an Application to the Measurement of Change. *Psychometrika*, *59*(2), 177–192.

Fox, J. & Hong, J. (2009). Effect Displays in R for Multinomial and Proportional-Odds Logit Models: Extensions to the effects Package. *Journal of Statistical Software*, *32*(1), 1–24.

Fox, J. & Weisberg, S. (2011). *An R Companion to Applied Regression* (2 ed.). Thousand Oaks CA: Sage.

Frick, H., Strobl, C., Leisch, F., & Zeileis, A. (2012). Flexible Rasch Mixture Models with Package psychomix. *Journal of Statistical Software*, *48*(7), 1–25.

Gosh, M. (1995). Inconsistent Maximum Likelihood Estimators for the Rasch Model. *Statistics & Probability Letters*, *23*, 165–170.

Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of Item Response Theory.* Measurement Methods for the Social Science. SAGE Publications.

Heinen, T. (1996). *Latent Class and Discrete Latent Trait Models: Similarities and Differences.* Thousend Oaks, CA: Sage.

Hoijtink, H. & Boomsma, A. (1995). On Person Parameter Estimation in the Dichotomous Rasch Model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch Models: Foundations, Recent Developments, and Applications* (pp. 53–68). New York: Springer.

Holland, P. W. & Wainer, H. (1993). *Differential Item Functioning.* Hillsday, New Jersey: Lawrence Erlbaum Associates.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure Based on a Modified Bonferroni Test. *Scandinavian Journal of Statistics*, *6*, 65–70.

Kopf, J., Zeileis, A., & Strobl, C. (2013). Anchor Methods for DIF Detection: A Comparison of the Iterative Forward, Backward, Constant and All-Other Anchor Class. Technical Report 141, Ludwig-Maximilians-Universität München.

Kopf, J., Zeileis, A., & Strobl, C. (2014). Anchor Selection Strategies for DIF Analysis: Review, Assessment, and New Approaches. *Educational and Psychological Measurement*, *1*(1), 1–35.

Mair, P. (2014). ctv: Infrastructure for Psychometric Modeling. http://CRAN.R-project.org/package=ctv. R package version 0.8-0.

Mair, P. & Hatzinger, R. (2007). Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R. *Journal of Statistical Software*, *20*(9), 1–20.

Mair, P., Hatzinger, R., & Maier, M. (2014). eRm: Extended Rasch Modeling. http://CRAN.R-project.org/package=eRm. R package version 0.15-4.

Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika*, *47*(2), 149–174.

Masters, G. N. & Wright, B. D. (1997). The Partial Credit Model. In W. J. v. d. Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 101–122). New York: Springer.

Mellenbergh, G. J. (1995). Conceptual Notes on Models for Discrete Polytomous Item Responses. *Applied Psychological Measurement*, *19*(1), 91–100.

Molenaar, I. W. (1995). Estimation of Item Parameters. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch Models: Foundations, Recent Developments, and Applications* (pp. 3–14). New York: Springer.

Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement, 16*(2), 159–176.

Muraki, E. (1993). Information Functions of the Generalized Partial Credit Model. *Applied Psychological Measurement, 17*(4), 351–363.

Nering, M. L. & Ostini, R. (2010). *Handbook of Polytomous Item Response Theory Models*. New York: Routledge.

Punzo, A. (2008). Remarks on the Existence of CML Estimates for the PCM by means of the R-package eRm. Technical Report 1, Universitá Degli Studi Di Bergamo.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. R version 3.1.1.

Rasch, G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests*. Kopenhagen: Danmarks Paedogogiske Institut.

Rizopoulos, D. (2013). ltm: An R Package for Latent Variable Modeling and Item Response Analysis. http://CRAN.R-project.org/package=ltm. R package version 1.0-0.

Rusch, T., Mair, P., & Hatzinger, R. (2013). Psychometrics with R: A Review of CRAN Packages for Item Response Theory. Technical Report 2013/2, Universität Wien.

Sijtsma, K. & Hemker, B. T. (2000). A Taxonomy of IRT Models for Ordering Persons and Items Using Simple Sum Scores. *Journal of Educational and Behavioral Statistics, 25*(4), 391–415.

Strobl, C., Kopf, J., & Zeileis, A. (2013). Rasch Trees: A New Method for Detecting Differential Item Functioning in the Rasch Model. *Psychometrika, 1*(1), 1–28.

Van der Ark, L. A. (2001). Relationships and Properties of Polytomous Item Response Theory Models. *Applied Psychological Measurement, 25*(3), 273–282.

Van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer.

Vansteelandt, K. (2000). *Formal Models for Contextualized Personality Psychology.* PhD thesis, K.U. Leuven, Belgium.

Wilson, M. & Masters, G. (1993). The Partial Credit Model and Null Categories. *Psychometrika*, *58*, 87–99.

Zeileis, A. (2006). Object-Oriented Computation of Sandwich Estimators. *Journal of Statistical Software*, *16*(9), 1–16.

Zeileis, A. & Hornik, K. (2007). Generalized M-Fluctuation Tests for Parameter Instability. *Statistica Neerlandica*, *61*(4), 488–508.

Zeileis, A. & Hothorn, T. (2002). Diagnostic Checking in Regression Relationships. R News, *2*(3), 7–10.

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based Recursive Partitioning. *Journal of Computational and Graphical Statistics*, *17*(2), 492–514.

Zeileis, A., Strobl, C., Wickelmaier, F., & Abou El-Komboz, B. (2014). psychotools: Infrastructure for Psychometric Modeling. `http://CRAN.R-project.org/package=psychotools.` R package version 0.2-0.

Zeileis, A., Strobl, C., Wickelmaier, F., Kopf, J., & Abou El-Komboz, B. (2014). psychotree: Recursive Partitioning Based on Psychometric Models. `http://CRAN.R-project.org/package=psychotree.` R package version 0.13-0.

Declaration of Authorship

I hereby declare in accordance with section §13(8) of the study and examinations regulations of the Ludwig-Maxmilians-Universität München for the Master's programm Statistics with focus on economic and social science, that I wrote this Master's thesis by myself without auxiliary means or sources other than those indicated. All parts taken from published and unpublished scripts are indicated as such. This Master's thesis has not been previously presented as an examination paper in this or any other form.

Munich, August 5th, 2014,

_____

(Basil Abou El-Komboz)