pcIRT: An **R** Package for Polytomous and Continuous Rasch Models

Christine Hohensinn
University of Vienna

Abstract

A key task in psychological assessment is the scaling of new psychological tests and questionnaires. Item response theory (IRT) is a well-established framework for this area of research. At present, IRT comprises a great variety of models, but to date, relatively little attention has been paid to the scaling of nominal and continuous data. The R package **pcIRT** provides functions for estimating IRT models for polytomous (nominal) and continuous data – the multidimensional polytomous Rasch model (Rasch 1961) and the continuous rating scale model (Müller 1987). Both models are extensions of the dichotomous logistic Rasch model (Rasch 1980) and retain its key feature of the separability of structural and nuisance parameters. The multidimensional polytomous Rasch model is suitable for nominal data under the assumption of a multidimensional space for the response categories, and the continuous rating scale model is a direct extension of the rating scale model developed by Andrich (1978) for continuous data.

Keywords: multidimensional polytomous Rasch model, continuous rating scale model, item response theory.

1. Introduction

Instruments such as tests and questionnaires that measure a latent trait are well-established research methods in the social sciences. Item response theory (IRT) provides a range of statistical models for scaling these measurement tools. IRT is the predominant scaling method used in large-scale assessments administered in an educational context (cf. OECD 2012; Olson, Martin, and Mullis 2008) and is also an important scaling method for psychological tests and questionnaires. To date, the majority of IRT models have focused on dichotomous and ordinal data. However, new and innovative response formats make it necessary to provide scaling methods for other data types as well. The newly developed package **pcIRT** (Hohensinn 2018) – written in R (R Core Team 2018) – makes scaling methods for polytomous (nominal) and

continuous data available to the user. More specifically, package **pcIRT** provides easy-to-use functions for estimating the multidimensional polytomous Rasch model (MPRM), which handles polytomous (nominal) data. In addition, the package allows the estimation of the continuous rating scale model (CRSM), which scales continuous data. Package **pcIRT** is available from the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/package=pcIRT`.

IRT comprises a great variety of models. A majority of them have already been implemented in various R packages. An up-to-date overview of R packages for IRT measurement is provided by the CRAN Task View on Psychometric Models and Methods (Mair 2018).

The main characteristic of all IRT models is that they model the relationships between manifest behavior on test tasks or questionnaire items and latent traits. Thus, IRT is based on statistical models obtained by estimating the parameters of an item and of the respondents' characteristics (Baker and Kim 2004). More basically, in the (dichotomous logistic) Rasch model (Rasch 1980), the probability of a response of respondent v to item i, $X_{vi} = 1$, is given by the following logistic function:

$$\mathsf{P}(X_{vi} = 1 | \theta_v, \beta_i) = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)}, \tag{1}$$

where β_i is interpreted as the difficulty of task or question i and θ_v represents the ability or characteristic of the measured latent trait of respondent v.

The main characteristic of the Rasch model is the existence of sufficient statistics for the parameters and, consequently, the ability to estimate the structural (item) parameters separately from the nuisance (person) parameters. This separate estimation is realized through the derivation of the conditional likelihood for the Rasch model. In the conditional likelihood, the person parameters are conditioned out. The estimation of item parameters using the conditional likelihood is well known as conditional maximum likelihood (CML) estimation (see Baker and Kim 2004). Several extensions of the Rasch model given in Equation 1 have been derived that retain this key feature, whereas other IRT models, e.g., the 3-PL model developed by Birnbaum (1968), are no longer models of the exponential family. Like the **eRm** package (Maier, Hatzinger, and Maier 2015), the **pcIRT** package is based on CML estimation. Mair and Hatzinger (2007) emphasize the advantages of extended Rasch models belonging to the exponential family and of CML estimation in a very detailed way; please refer to their paper for an extensive explanation.

Extensions of the dichotomous logistic Rasch model that retain the characteristic of being a member of the exponential family can be divided into two broad categories: first, models for polytomous and continuous data, and second, multidimensional models that assume more than one latent trait θ. Among the larger set of R packages for IRT models, the **eRm** package (Maier *et al.* 2015) is the only one that implements CML estimation and includes functions for extended Rasch models that maintain parameter separability. Note that the **eRm** package offers a variety of functions for estimating unidimensional Rasch models for both dichotomous and polytomous data. By contrast, the **pcIRT** package presented here offers functions for estimating the multidimensional polytomous Rasch model (MPRM) and a (unidimensional) extension of the Rasch model for continuous data – the continuous rating scale model (CRSM).

In its conception, the MPRM is an exceptional model. Its main difference from the nominal response model developed by Bock (1972) (included, for example, in the **mcIRT** package, Reif

2014) is its assumption of a multidimensional latent space. Furthermore, its primary difference from all "typical" multidimensional IRT models is that the MPRM assumes a latent trait for each response category instead of latent traits for subsets of items.

Other R packages such as **TAM** (Kiefer, Robitzsch, and Wu 2015) and **mirt** (Chalmers 2012) allow the estimation of multidimensional IRT models but use marginal maximum likelihood (MML) estimation instead of a CML estimation procedure.

Moreover, package **pcIRT** allows the MPRM to be estimated in an easy-to-use way while still allowing the user to set a variety of constraints in the design matrix. In addition, the package offers functions for estimating the CRSM and the Rasch model. This paper focuses on describing the MPRM and its implementation in greater detail, with illustration through examples.

2. Multidimensional polytomous Rasch model

When one discusses multidimensional IRT (MIRT) models, one is most often referring to models in which a specific kind of MIRT model is included. These models have the common feature that an item is assigned to a specific latent trait or to more than one trait (for an overview of multidimensional IRT models, see Reckase 2009). The multidimensional polytomous Rasch model (MPRM), which was formulated by Georg Rasch himself (Rasch 1961), has a completely different model formulation. It is also a multidimensional model, in that it assumes a multidimensional latent trait space. However, in contrast to other well-known MIRT models, the MPRM assigns response categories to latent traits instead of items. Huang and Mislevy (2010) demonstrated how the use of the MPRM provides additional information about a multiple-choice achievement test: They analyzed a multiple-choice mechanics test in physics and aimed at measuring both the competence level and the (mis)conceptions of students who are learning physics. Therefore, the distractors for the items were developed to represent different fundamental misconceptions (e.g., an Aristotelian instead of a Newtonian view). The MPRM assumes a latent dimension θ for each response category – that is, for each conceptual approach in physics, in this application. The results provide details regarding the difficulties associated with each distractor and information about the (mis)conceptions to which each student clings.

To emphasize the differences between the MPRM and "typical" MIRT models, as a first step, the very general model formulation developed by Adams, Wilson, and Wang (1997) is presented which is given by[1]

$$\mathsf{P}(X_{vi} = h|\theta_{vd}, \eta_k, \mathrm{b_{ihd}}, \mathrm{a_{ihd}}) = \frac{\exp(\sum_{d=1}^{D} \mathrm{b_{ihd}}\theta_{vd} - \sum_{k=1}^{K} \mathrm{a_{ihd}}\eta_k)}{\sum_{j=0}^{m} \exp(\sum_{d=1}^{D} \mathrm{b_{ihd}}\theta_{vd} - \sum_{k=1}^{K} \mathrm{a_{ihd}}\eta_k)}. \tag{2}$$

The probability of responding to item i in category h depends on the latent trait $\boldsymbol{\theta}_d = (\theta_{1d}, \ldots, \theta_{Nd})$ and the item components η_k. v indicates the person, and d indicates the latent trait. Now, for deriving the MPRM from the general model given in Equation 2, the design matrices $\mathbf{A}$ and $\mathbf{B}$ must be specified in the following way: Each response category corresponds to a different underlying latent trait d. Consequently, the number of latent dimensions d must

---

[1]Equation 2 is expressed according to Rost and Carstensen (2002), with slightly modified indices for consistency with the other model formulas given in this paper. In Rost and Carstensen (2002), a formalized derivation of the MPRM from the general MIRT model given in Equation 2 can be found.

	θ_1	θ_2	θ_3
$h = 1$	1	0	0
$h = 2$	0	1	0
$h = 3$	0	0	1

Table 1: An example of a design matrix **B** for a certain item i is given. Each response category h is assigned to a different latent dimension θ_d in the MPRM.

	η_1	η_2	η_3	η_4	η_5	η_6	η_7	η_8	η_9	η_{10}	η_{11}	η_{12}
$h = 1$	1	0	0	0	0	0	0	0	0	0	0	0
$h = 2$	0	1	0	0	0	0	0	0	0	0	0	0
$h = 3$	0	0	1	0	0	0	0	0	0	0	0	0

Table 2: An example of a design matrix **A** for the three response categories of item 1 is given. Each response category h is assigned to a different item component parameter η_k in the MPRM.

be equal to the number of response categories h, that is, $d = h$. Furthermore, all items in a data set must have the same number of categories. Suppose that we have one item i associated with three response categories $h = 1, 2, 3$; then, $\boldsymbol{\theta}_v = (\theta_{v1}, \theta_{v2}, \theta_{v3})$. Consequently, **B** is a three-dimensional diagonal matrix that maps each response category h exclusively to one θ_d, as shown in Table 1.

A design matrix **A** expresses the relations between each response category h and the item component parameters $\boldsymbol{\eta}$. In the MPRM, an item parameter is assigned to each item for each category, which means that they are item category parameters. Continuing with the example of three response categories $h = 1, 2, 3$, four items $i = 1, \ldots, 4$ are now assumed. This results in a total of 12 item component parameters η_k. Table 2 shows the design matrix **A** for item 1. There are three item parameters for item $i = 1$: η_1 for category 1, η_2 for category 2 and η_3 for category 3. Obviously, η_4, η_5 and η_6 are the item parameters for item 2, and so on. For the MPRM, the term "item category parameter" expresses the meaning of the parameters more precisely than the general term "item component parameter". Therefore, these parameters will henceforth be referred to as the item category parameters β_{ih}. For example, item component parameter η_2 which is assigned to item 1 and category 2 is indicated as β_{12}. Restricting the general MIRT model expressed in Equation 2 in this way leads to the MPRM (cf. Andersen 1995):

$$\mathsf{P}(X_{vi} = h | \theta_{vh}, \beta_{ih}) = \frac{\exp\left(\theta_{vh} + \beta_{ih}\right)}{\sum_{j=1}^{m} \exp\left(\theta_{vj} + \beta_{ij}\right)}. \tag{3}$$

The probability that a person v responds to item i in category h depends on a person parameter, θ_{vh}, and the item parameter, β_{ih}.

The MPRM described thus far suffers from over-parameterization. This is avoided by imposing the following constraints (see Fischer 1974):

$$\sum_{h=1}^{m} \theta_{vh} = 0, \qquad \beta_{im} = 0, \quad i = 1, \ldots, r, \qquad \sum_{i=1}^{r} \beta_{ih} = 0, \tag{4}$$

where m is the highest response category, $h = 1, \ldots, m$, and r is the number of items, $i = 1, \ldots, r$. The item component parameters for the highest category are set to zero. In the example given above, this reduces the number of estimated η_k parameters to eight (see Table 2) and the number of estimated θ_d parameters for person v to two (see Table 1).

As a consequence, it follows that the MPRM reduces to the dichotomous logistic Rasch model in the case of only two categories.

In contrast to most other MIRT models, the MPRM belongs to the family of Rasch models, which means that sufficient statistics for the parameters are available. For the MPRM given in Equation 3, the vector

$$\mathbf{y}_v = \left(\sum_{i=1}^{r} x_{vi1}, \ldots, \sum_{i=1}^{r} x_{vim} \right) \tag{5}$$

that sums all responses of person v for each category h provides sufficient statistics for the person parameter vector of person v, $\boldsymbol{\theta}_v$. By virtue of the existence of sufficient statistics $\mathbf{y}_v$, the conditional likelihood for the data set $\mathbf{X}$ can be derived as follows (Andersen 1973):

$$\begin{aligned} \mathrm{cL}\left(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r | \mathbf{X}\right) &= \mathrm{cL}\left(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r | \mathbf{Y}\right) \\ &= \exp\left[\sum_{i=1}^{r} \sum_{h=1}^{m} \beta_{ih} y_{ih} \right] \prod_{v=1}^{n} \gamma^{-1}\left(\mathbf{Y}, \exp(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k) \right), \end{aligned} \tag{6}$$

with $\mathbf{Y}$ based on the data set $\mathbf{X}$ according to Equation 5. γ is the combinatoric function that yields the number of all possible response vectors according to a given marginal response vector $\mathbf{y}_v$. Andersen (1972) developed an algorithm for the recursive calculation of γ that is computationally efficient. The conditional likelihood expressed in Equation 6 enables parameter estimation using the CML method.

Of course, the MPRM is primarily a model for scaling nominal data, as it has been used, for instance, by Huang and Mislevy (2010). However, the MPRM also offers other interesting applications. Andersen (1973) and Fischer (1974), in particular, note the possibility of testing whether the assumed multidimensional space $\boldsymbol{\Theta}$ can be reduced to a unidimensional one. This test for unidimensionality is performed by setting the item category parameters $\beta_{i1}, \ldots, \beta_{im}$ for an item i as linearly dependent: $\beta_{ih} = \beta_i \cdot \phi_h$. In this manner, one can test whether the assumption of ordinal scoring (an assumption that is always made, e.g., with regard to rating scales for questionnaires) holds and, furthermore, whether the scoring function of an ordinal scale can be estimated. This opens up an important field of application in the context of test and questionnaire construction.

3. Implementation of the MPRM in the pcIRT package

3.1. General

Like most R packages, **pcIRT** uses S3 classes. The `MPRM` function enables the user to set constraints on the model in a flexible way by specifying a design matrix through the argument `desmat` as well as the arguments `ldes` and `lp`. How the restrictions on this model can be set will be explained in greater detail and illustrated in Section 4.2. Furthermore, in Section 2, the unidimensionality test for the MPRM was mentioned. This test is implemented in an extended form. In its original conception, the test assumes the linear dependency $\beta_{ih} = \beta_i \phi_h$ for $i = 1, \ldots, r$ and $h = 1, \ldots, m$. `MPRM` also allows the setting of linearly dependent parameters for only a subset of user-specified items i and categories h. This leads to far greater flexibility in testing hypotheses regarding the unidimensionality of different response categories. The

resulting conditional likelihood of the adapted MPRM is derived as follows:

$$\mathrm{cL}\Big(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r | \mathbf{X}\Big) = \mathrm{cL}\Big(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r | \mathbf{Y}\Big) =$$

$$\exp\Big[\Big(\sum_{i=1}^{r}\sum_{h=1}^{m}\beta_{ih}y_{ih}\Big)^{(1-c)} + \Big(\sum_{i=1}^{r}\sum_{j=1}^{h}\beta_i\phi_h y_{ih}\Big)^c\Big]\prod_{v=1}^{n}\gamma^{-1}\Big(\mathbf{Y},\exp(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_r)\Big), \quad (7)$$

where c is an indicator variable. Thus, $c = 1$ indicates a given specification of the linearly dependent parameters. As in Equation 6, $\mathbf{Y}$ is the sufficient statistic for the person parameters and is defined in Equation 5.

This conditional likelihood is maximized by the `optim` function from the base R **stats**, using the BFGS optimizer by default. It is possible to choose another available optimizer using the `control` argument.

The functions for estimating the model are supplemented with several additional functions that allow the user to check the fit of the model. A function for generating data sets in accordance with the MPRM is also included.

3.2. Accuracy of parameter estimation

The described functions of the **pcIRT** package were evaluated through simulation studies. The results of one of these simulations are presented here to illustrate the accuracy of parameter estimation.

Considering that the main focus of this paper is on the MPRM, a comparison of the parameter estimation accuracy of this model offered by the **pcIRT** package with that of the **TAM** package[2] is presented. For this purpose, 1000 data sets were simulated with the number of items $r = 15$, the number of categories $m = 3$ and the sample size $n = 1000$. The β_{ih} were randomly drawn from $\mathcal{N}(0, 1.5)$ once and then fixed (parameters $\in (-2.41; 4.01)$). The θ_{vh} were drawn from $\mathcal{N}(0, 1)$. The MPRM was estimated using the **pcIRT** package and the **TAM** package. Figure 1 shows the differences between the fixed β_{ih} values and the estimated ones (the means and the quantiles at 0.05 and 0.95 are displayed). β_{121} and β_{142} show a larger difference compared with the other parameters. This is because these parameters were those with the highest absolute values. The estimates produced by both packages are very similar so that the two facets of the plot looks almost identical.

According to Figure 1, the mean square errors of parameter estimation are 0.01297 for the **pcIRT** package and 0.01285 for the **TAM** package. The median time required to estimate the MPRM was 2.55 seconds for the **TAM** package and 33.89 seconds for the **pcIRT** package.

4. Application of the `MPRM` function

This section illustrates the application of the MPRM with examples.

4.1. Estimation of the MPRM

As noted in Section 2, the `MPRM` function treats the input data as nominal. Thus, the main field of application of the MPRM is the scaling of nominal data. The application of Huang

---

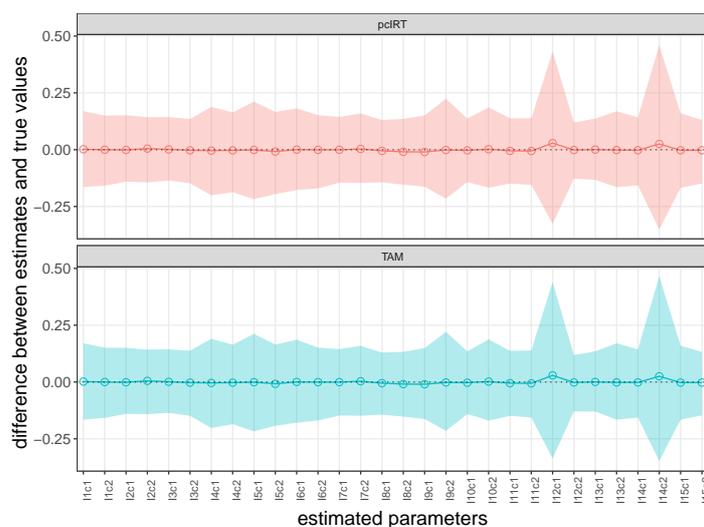[2]Thanks to Alexander Robitzsch for his help in defining the design matrices in package **TAM**.

Figure 1: Parameter estimation accuracy is compared between packages **pcIRT** and **TAM**. The labels identify the parameters, e.g., the parameter β_{11} for item 1 and category 1 is labeled as I1c1.

and Mislevy (2010) described above illustrates this main purpose very well. Therefore, a different example was chosen here to provide insight into the variety of possible applications. The example concerns a fundamental problem in test and questionnaire construction: Very often, scoring functions are determined without further empirical evaluation. Subsequently, the MPRM is used to assess and, finally, to estimate the appropriate scoring weights of the response categories. This example follows an approach described in Fischer (1974).

Intelligence and competence tests are often administered as speed-and-power tests by additionally recording the time taken by the examinee to give a response. The purpose of this analysis is to determine whether a fast solution expresses a different competence or facet of ability than a slow solution. If this is true, then a fast solution would be linked to a different latent trait than a slow solution, and a scoring method that awards additional points for fast solutions (which is done in some intelligence tests) would be inappropriate. Otherwise, if all three categories represent only graduated levels of the same competency, such awarding of points would lead to a higher measurement precision. However, in either case, the selection of an appropriate scoring function for the response categories is crucial. This question is studied using the MPRM.

The data set `reason` is used, which contains the responses and response times for each of eleven items on the reasoning test "META" (Gatternig and Kubinger 1994). The respondents are asked to solve encoding tasks without time pressure. The data set was obtained from a low-stakes test situation and includes observations of 404 students. After excluding three respondents who quit the test and 21 persons who apparently only clicked through the test, data from 380 students are analyzed. As usual, the items are scored as "not solved" ($X_{vi} = 0$) or "solved" ($X_{vi} = 1$). In addition to the item responses, the response times in seconds for each item are stored.

```
R> data("reason", package = "pcIRT")
R> head(reason.test[, 1:11], 3)
```

```
      I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11
[1,]   1  0  0  0  0  1  0  0  0   0   0
[2,]   1  0  0  0  0  0  0  0  1   0   0
[3,]   1  0  0  0  0  0  0  0  0   0   0

R> head(reason.test[, 12:17], 3)

        BT1    BT2    BT3    BT4    BT5  BT6
[1,] 105.8 300.0 255.2 289.7 300.0 63.7
[2,]  54.7 166.3 136.8  89.0 108.2 61.0
[3,] 110.2 151.4 182.0 156.9 144.4 38.1
```

First, the item responses are recoded to a 3-categorical format. Based on the initial findings for items 1 to 4, fast solutions with response times of up to 180 seconds ($x_{vi} = 1 \,\&\, b_{vi} \leq 180$) are labeled as 0, slow solutions ($x_{vi} = 1 \,\&\, b_{vi} > 180$) are labeled as 1, and incorrect responses are labeled as 2. For items 5 to 11, the threshold for the response times is set to 130 seconds instead.

```
R> reason.testR <- reason.test
R> itpos <- 1:11
R> tipos <- itpos + 11
R> co <- rep(c(180, 130), c(4, 7))
R> for (i in 1:length(itpos)) {
+    reason.testR[reason.test[, itpos[i]] == 1 &
+      reason.test[, tipos[i]] <= co[i], itpos[i]] <- 0
+    reason.testR[reason.test[, itpos[i]] == 1 &
+      reason.test[, tipos[i]] > co[i], itpos[i]] <- 1
+    reason.testR[reason.test[, itpos[i]] == 0, itpos[i]] <- 2
+  }
```

The MPRM takes as input data sets **X** in a wide format; that is, each column represents an item, and each row represents an observation. The category levels must be consecutive numbers, and all items must have the same number of response categories. Now, the MPRM is estimated for the eleven items using the recoded data.

The generic summary function displays the convergence and estimated item category parameters.

```
R> MPRM.res <- MPRM(reason.testR[, 1:11])
R> summary(MPRM.res)

 Call:  MPRM(data = reason.testR[, 1:11])

Function calls:
function gradient
      85        29
Convergence:            convergence
```

```
Deviance:              3023.463
Number of Parameters:         20


---------------------------------------------------------
Parameter estimates:
        beta I1    beta I2   beta I3  beta I4     beta I5    beta I6
cat 1 -3.258005 0.01287835 1.0428143 1.016488 -0.004341561 -0.7922870
cat 2 -1.880010 0.13279345 0.5030431 0.693977 -0.068718977 -0.6560851
cat 3  0.000000 0.00000000 0.0000000 0.000000  0.000000000  0.0000000
        beta I7   beta I8    beta I9   beta I10  beta I11
cat 1  1.124702 0.6170842 -0.2543161 -0.1916351 0.6866170
cat 2  1.444394 0.3100316 -0.4439317 -0.2201612 0.1846675
cat 3  0.000000 0.0000000  0.0000000  0.0000000 0.0000000
          SE I1     SE I2     SE I3     SE I4     SE I5     SE I6
cat 1 0.1513772 0.1735940 0.2410501 0.2355404 0.1752277 0.1497027
cat 2 0.1688102 0.1584189 0.1681306 0.1773452 0.1514670 0.1414774
cat 3        NA        NA        NA        NA        NA        NA
          SE I7     SE I8     SE I9    SE I10 SE I11
cat 1 0.2364460 0.2088610 0.1681042 0.1678936     NA
cat 2 0.2255492 0.1617454 0.1425058 0.1480273     NA
cat 3        NA        NA        NA        NA    NA
```

R> iccplot(MPRM.res, items = 4)

Finally, 20 item parameters are estimated; the parameters of the last (eleventh) item and those for the highest category are fixed for normalization reasons (in accordance with Equation 4). The highest category is the reference category and is set to 0. For the fixed parameters, no SE is estimated, and the SE is therefore reported as NA in the summary. The normalization constraints mentioned above must be taken into account when interpreting the parameter estimations. For example, when comparing the different category parameters of item 1, one should be aware that these parameter estimations are affected by the other item category parameters because of the normalization within a category to a sum of zero. Thus, for example, item 1 seems to be a rather easy item in general. That is, a response in the category "fast solution" is more likely than one in the category "slow solution". Similarly, both "solved" categories are more likely than the (reference) category "not solved". By contrast, for item 4, it appears distinctly more difficult to solve the item quickly than to solve it with a longer response time. The task of item 4 therefore seems to be considerably more difficult than that of item 1. Figure 2 displays the item characteristic curves for item 4. The plots show the variations in response probability for a certain response category depending on the location with respect to the latent traits. For instance, the ICC here shows that the response probability for category 3 increases with low latent trait locations in the first two latent trait dimensions θ_1 and θ_2. This trait location corresponds to a high position in θ_3. These plots make it clear that the person parameters are interdependent. This is already obvious from the normalization constraints of the person parameters (cf. 4).

In the next step, the person parameters θ_{vh} are estimated for all observed raw score vectors except those with extreme scores (that is, at least one observation and at most $r - 1$ observations in each response category).
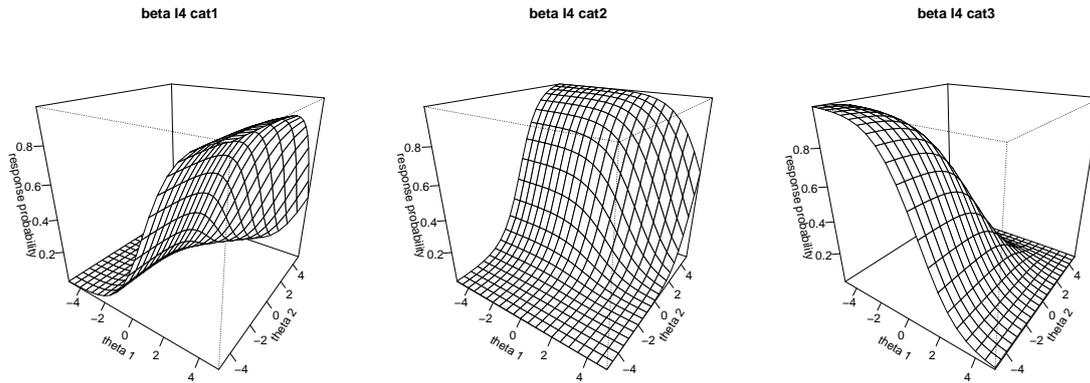
Figure 2: The ICCs for all three response categories of item 4 are shown.

```
R> MPRM.pp <- person_par(MPRM.res)
R> head(MPRM.pp$ptable, 3)


      pers.par.cat1 pers.par.cat2 pers.par.cat3 SE cat 1  SE cat 2
1|1|9     -1.058391    -0.6310490     1.6894399 1.446805 1.1688397
1|2|8     -1.191389    -0.1326835     1.3240726 1.397705 0.6738535
1|3|7     -1.234189     0.1811570     1.0530319 1.370542 0.5202010
```

The output displays one part of a table providing the person parameter estimates for the observed raw score vectors. The row names indicate the raw score vectors. According to the normalization constraints given in Equation 4, the person parameters must be interpreted as intra-individual tendencies to choose a response category. For instance, respondents with a raw score vector of $\boldsymbol{x}_{vi} = (1, 1, 9)$ have a strong tendency to respond in category 3, with $\hat{\theta}_3 = 1.69$.

Next, it is tested whether the categorization of fast and slow solutions is a gradual one or whether it is a qualitative difference. In the former case, the β_{ih} parameters should be linearly dependent for a given item i. This is tested by means of a likelihood-ratio test comparing the likelihood of the MPRM to the likelihood of a restricted MPRM that includes the linear dependency $\beta_{ih} = \beta_i \cdot \phi_h$ described in Section 2. To test this general unidimensional hypothesis, the dLRT function is applied.

```
R> dLRT(MPRM.res)


emp Chi2:           16.04115
df:                  9
p-value:             0.066


R> cor(MPRM.res$itempar[1, ], MPRM.res$itempar[2, ])

[1] 0.9427702
```

The unidimensional model does fit the data set with a type I risk of $\alpha = 0.05$. The correlation between the response categories, $\mathsf{COR}(\hat{\beta}_{i1}, \hat{\beta}_{i2}) = 0.943$, is high.

4.2. Estimation of a constrained MPRM

It is illustrated how the user can easily set constraints on the MPRM. To explain the relevant functions, the example from Section 4.1 is continued. The example as presented thus far, a case of three-categorical scoring with bonus points for a fast solution, is appropriate for a reasoning test. Now, using a constrained MPRM, the optimal scoring should be determined empirically. For this purpose, constraints are set on the MPRM such that each item category parameter of category 2 is linearly dependent on the corresponding parameter of category 1: $\beta_{i2} = \beta_{i1} \cdot \phi_h$. Furthermore, the scoring parameter is set equal for all categories h, resulting in one general ϕ. In essence, this is a unidimensional polytomous model with a scoring parameter to be estimated.

To obtain this specification, three arguments must be passed to the function `MPRM`: First, the design matrix must be built. The easiest way to achieve this is by using the automatically built design matrix for the estimated general MPRM (from the estimation presented in Section 4.1) and applying the desired modifications.

```
R> design <- MPRM.res$design
```

Each column of the design matrix represents an item category parameter that must be estimated, whereas each row displays the resulting item category parameters (including fixed parameters or parameters resulting from normalization). For the general model, the resulting number of item category parameters is 33 (11 items multiplied by 3 categories). Because of the normalization constraints (see Equation 4), 20 item category parameters are estimated in the general MPRM. Now, in this application, only the category parameters for category 1 and the scoring parameter ϕ – that is, a total of 11 parameters – need to be estimated. In the first step, the item category parameters of categories 2 and 1 are set equal in the design matrix for all items:

```
R> s1 <- seq(2, 30, by = 3)
R> s2 <- seq(1, 20, by = 2)
R> for (s in 1:length(s1)) {
+    design[s1[s], s2[s]] <- 1
+  }
R> design[32, s2]<- -1
R> designC <- design[, -c(2, 4, 6, 8, 10, 12, 14, 16, 18, 20)]
```

Modifying the design matrix in the presented way and using it in the function `MPRM` would lead to a constrained MPRM, with the item category parameters of category 2 set equal to those of category 1. In our example, however, we wish to estimate a constrained MPRM by setting the item category parameters to be linearly dependent (not equal). Therefore, exactly which item category parameters should be set as linearly dependent must be specified. This is achieved by means of the argument `ldes`. For this purpose, a vector with a length equal to the number of resulting item category parameters is created. Each position represents the corresponding item category parameter. For example, in the present case, the item category parameter in position 5 in the vector represents the parameter for item 2 and category 2 (β_{22}). Now, in this vector, the position of the item category parameter on which an item category parameter is linearly dependent is plugged into that dependent item category parameter. For example, if item category parameter β_{22} is estimated as linearly dependent on β_{21}, the

position of item category parameter β_{21} (which is 4) must be input into the vector at the position of item category parameter β_{22} (which is 5).

The vector then looks as shown in the code output below:

```
R> ldc <- rep(0, nrow(designC))
R> ldc[c(2, 5, 8, 11, 14, 17, 20, 23, 26, 29)] <- c(1, 4, 7, 10, 13, 16, 19,
+    22, 25, 28)
R> ldc

 [1] 0  1  0  0  4  0  0  7  0  0 10  0  0 13  0  0 16  0  0 19  0  0 22  0
[25] 0 25  0  0 28  0  0  0  0
```

Finally, the user uses the argument `lp` to specify exactly how many scoring parameters ϕ need to be estimated and which of these scoring parameters are equal. In our example, only one ϕ should be estimated for all items. The three arguments `desmat`, `ldes` and `lp` are passed to the `MPRM` function.

```
R> lpc <- rep(1, 10)
R> MPRM.unid <- MPRM(reason.testR[, 1:11], desmat = designC, ldes = ldc,
+    lp = lpc)
R> c(MPRM.unid$linpar, MPRM.unid$linpar_se)

 [1] 0.6093301  0.03896333
```

The code output displays $\hat{\phi}$ with its standard error. This result yields the empirically derived optimal scoring proportions of $(0; 0.6; 1)$ for the `reason` data set. Thus, an appropriate scoring approach for the test is to award 2 points for a fast and correct solution, 1.2 points for a slow and correct solution, and 0 points for an incorrect response.

5. Future plans and outlook

There are plans to add additional possibilities for assessing the fit of item and response patterns – that is, item and person fit indices for the models implemented thus far. Another major feature planned for implementation is the extension of the model estimation for handling values that are missing by design. In analogy to other extensions of the Rasch model, the likelihood will be modified such that the MPRM can be estimated even if not all items were administered to all examinees (which is often the case in large-scale assessments, for example).

References

Adams RJ, Wilson M, Wang W (1997). "The Multidimensional Random Coefficients Multinomial Logit Model." *Applied Psychological Measurement*, **21**(1), 1–23. doi: 10.1177/0146621697211001.

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

Andersen EB (1973). "Conditional Inferences for Multiple-Choice Questionnaires." *British Journal of Mathematical and Statistical Psychology*, **26**(1), 31–44. `doi:10.1111/j.2044-8317.1973.tb00504.x`.

Andersen EB (1995). "Polytomous Rasch Models and Their Estimation." In GH Fischer, IW Molenaar (eds.), *Rasch Models. Foundations, Recent Developments, and Applications*, chapter 15, pp. 271–292. Springer-Verlag, New York.

Andrich D (1978). "A Rating Formulation for Ordered Response Categories."

Baker FB, Kim SH (2004). *Item Response Theory. Parameter Estimation Techniques*. 2nd edition. Dekker, New York.

Birnbaum A (1968). "Some Latent Trait Models and Their Use in Infering an Examinee's Ability." In FM Lord, MR Novick (eds.), *Statistical Theories of Mental Test Scores*, pp. 395–479. Addison-Wesley, Reading/Mass.

Bock RD (1972). "Estimating Item Parameters and Latent Ability When Responses Are Scored in Two or More Nominal Categories." *Psychometrika*, **37**(1), 29–51. `doi:10.1007/bf02291411`.

Chalmers RP (2012). "**mirt**: A Multidimensional Item Response Theory Package for the R Environment." *Journal of Statistical Software*, **48**(6), 1–29. `doi:10.18637/jss.v048.i06`.

Fischer GH (1974). *Einführung in die Theorie psychologischer Tests*. Hans Huber, Stuttgart. ISBN 3456800398. [Introduction to the Theory of Psychological Tests].

Gatternig J, Kubinger KD (1994). *Erkennen von Metaregeln*. Swets, Frankfurt.

Hohensinn C (2018). *pcIRT: IRT Models for Polytomous and Continuous Item Responses*. R package version 0.2.3, URL `https://CRAN.R-project.org/package=pcIRT`.

Huang CW, Mislevy RJ (2010). "An Application of the Polytomous Rasch Model to Mixed Strategies." In ML Nering, R Ostini (eds.), *Handbook of Polytomous Item Response Theory Models*, chapter 9, pp. 211–228. Taylor & Francis Group, New York.

Kiefer T, Robitzsch A, Wu M (2015). *TAM: Test Analysis Modules*. R package version 1.6-0, URL `https://CRAN.R-project.org/package=TAM`.

Maier P, Hatzinger R, Maier MJ (2015). "**eRm**: Extended Rasch Modeling." URL `http://erm.R-forge.R-project.org/`.

Mair P (2018). *CRAN Task View: Psychometric Models and Methods*. Version 2018-01-26, URL `https://CRAN.R-project.org/view=Psychometrics`.

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). `doi:10.18637/jss.v020.i09`.

Müller H (1987). "A Rasch Model for Continuous Ratings." *Psychometrika*, **52**(2), 165–181. doi:10.1007/bf02294232.

OECD (2012). *PISA 2009 Technical Report*. PISA. OECD Publishing. ISBN 9789264040182. doi:10.1787/9789264167872-en.

Olson JF, Martin MO, Mullis IVS (2008). *TIMSS 2007 Technical Report*. TIMSS & PIRLS International Study Center. ISBN 1889938505.

Rasch G (1961). "On General Laws and the Meaning of Measurement in Psychology." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 4, pp. 321–333. University of California Press, Berkeley, CA.

Rasch G (1980). *Probabilistic Models for Some Intelligence and Attainment Tests. (Reprint from 1960)*. University of Chicago Press, Chicago.

R Core Team (2018). R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org.

Reckase M (2009). *Multidimensional Item Response Theory*. Springer-Verlag, New York. ISBN 9780387899756.

Reif M (2014). **mcIRT**: *IRT Models for Multiple Choice Items*. R package version 0.40, URL https://github.com/manuelreif/mcIRT.

Rost J, Carstensen CH (2002). "Multidimensional Rasch Measurement via Item Component Models and Faceted Designs." *Applied Psychological Measurement*, **26**(1), 42–56. doi:10.1177/0146621602026001003.

Affiliation:

Christine Hohensinn
Department of Psychological Assessment and Applied Psychometrics
Faculty of Psychology
University of Vienna
1010 Vienna, Austria
E-mail: christine.hohensinn@univie.ac.at