Extended Rasch Modeling: The R Package eRm

Patrick Mair
Wirtschaftsuniversität Wien

Reinhold Hatzinger
Wirtschaftsuniversität Wien

Abstract

This package vignette is an update of the **eRm** papers by published in a special issue on Psychometrics in the Journal of Statistical Software and in Psychology Science (Mair and Hatzinger 2007a,b). Since the publication of these papers various extensions and additional features have been incorporated into the package. We start with a methodological introduction to extended Rasch models followed by a general program description and application topics. The package allows for the computation of simple Rasch models, rating scale models, partial credit models and linear extensions of these. The incorporation of such linear structures allows for modeling the effects of covariates and enables the analysis of repeated categorical measurements. The item parameter estimation is performed by means of CML, for the person parameters we use ordinary ML. The estimation routines work for incomplete data matrices as well. Based on these estimators, item-wise and global goodness-of-fit statistics are described and various plots are presented.

Keywords: eRm package, Rasch model, LLTM, RSM, LRSM, PCM, LPCM, CML estimation.

1. Introduction

Rost (1999) claimed in his article that "even though the Rasch model has been existing for such a long time, 95% of the current tests in psychology are still constructed by using methods from classical test theory" (p. 140). Basically, he quotes the following reasons why the Rasch model (RM) is being rarely used: The Rasch model in its original form (Rasch 1960), which was limited to dichotomous items, is arguably too restrictive for practical testing purposes. Thus, researchers should focus on extended Rasch models. In addition, Rost argues that there is a lack of user-friendly software for the computation of such models. Hence, there is a need for a comprehensive, user-friendly software routine. Corresponding recent discussions can be found in Kubinger (2005) and Borsboom (2006).

In addition to the RM, the models that can be computed by means of the **eRm** package are: the linear logistic test model (Scheiblechner 1972), the rating scale model (Andrich 1978), the linear rating scale model (Fischer and Parzer 1991), the partial credit model (Masters 1982), and the linear partial credit model (Glas and Verhelst 1989; Fischer and Ponocny 1994). These models and their main characteristics are presented in Section 2.

Concerning parameter estimation, these models have an important feature in common: Separability of item and person parameters. This implies that the item parameters β can be estimated without estimating the person parameters achieved by conditioning the likelihood on the sufficient person raw score. This conditional maximum likelihood (CML) approach is described in Section 3.

Several diagnostic tools and tests to evaluate model fit are presented in Section 4.

In Section 5, the corresponding implementation in R (R Development Core Team 2007) is described by means of several examples. The **eRm** package uses a design matrix approach which allows to reparameterize the item parameters to model common characteristics of the items or to enable the user to impose repeated measurement designs as well as group contrasts. By combining these types of contrasts one allows that the item parameter may differ over time with respect to certain

subgroups. To illustrate the flexibility of the **eRm** package some examples are given to show how suitable design matrices can be constructed.

2. Extended Rasch models

2.1. General expressions

Briefly after the first publication of the basic Rasch Model (Rasch 1960), the author worked on polytomous generalizations which can be found in Rasch (1961). Andersen (1995) derived the representations below which are based on Rasch's general expression for polytomous data. The data matrix is denoted as $\mathbf{X}$ with the persons in the rows and the items in the columns. In total there are $v = 1, ..., n$ persons and $i = 1, ..., k$ items. A single element in the data matrix $\mathbf{X}$ is indexed by x_{vi}. Furthermore, each item I_i has a certain number of response categories, denoted by $h = 0, ..., m_i$. The corresponding probability of response h on item i can be derived in terms of the following two expressions (Andersen 1995):

$$P(X_{vi} = h) = \frac{\exp[\phi_h(\theta_v + \beta_i) + \omega_h]}{\sum_{l=0}^{m_i} \exp[\phi_l(\theta_v + \beta_i) + \omega_l]} \tag{1}$$

or

$$P(X_{vi} = h) = \frac{\exp[\phi_h \theta_v + \beta_{ih}]}{\sum_{l=0}^{m_i} \exp[\phi_l \theta_v + \beta_{il}]}. \tag{2}$$

Here, ϕ_h are scoring functions for the item parameters, θ_v are the uni-dimensional person parameters, and β_i are the item parameters. In Equation 1, ω_h corresponds to category parameters, whereas in Equation 2 β_{ih} are the item-category parameters. The meaning of these parameters will be discussed in detail below. Within the framework of these two equations, numerous models have been suggested that retain the basic properties of the Rasch model so that CML estimation can be applied.

2.2. Representation of extended Rasch models

For the ordinary Rasch model for dichotomous items, Equation 1 reduces to

$$P(X_{vi} = 1) = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)}. \tag{3}$$

The main assumptions, which hold as well for the generalizations presented in this paper, are: uni-dimensionality of the latent trait, sufficiency of the raw score, local independence, and parallel item characteristic curves (ICCs). Corresponding explanations can be found, e.g., in Fischer (1974) and mathematical derivations and proofs in Fischer (1995a).

For dichotomous items, Scheiblechner (1972) proposed the (even more restricted) linear logistic test model (LLTM), later formalized by Fischer (1973), by splitting up the item parameters into the linear combination

$$\beta_i = \sum_{j=1}^{p} w_{ij} \eta_j. \tag{4}$$

Scheiblechner (1972) explained the dissolving process of items in a test for logics ("Mengenrechentest") by so-called "cognitive operations" η_j such as negation, disjunction, conjunction, sequence, intermediate result, permutation, and material. Note that the weights w_{ij} for item i and operation j have to be fixed a priori. Further elaborations about the cognitive operations can be found in
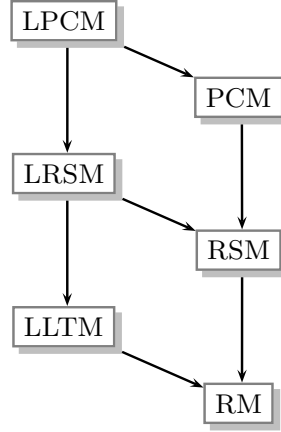
Figure 1: Model hierarchy

Fischer (1974, p. 361ff.). Thus, from this perspective the LLTM is more parsimonous than the Rasch model.

Though, there exists another way to look at the LLTM: A generalization of the basic Rasch model in terms of repeated measures and group contrasts. It should be noted that both types of reparameterization also apply to the linear rating scale model (LRSM) and the linear partial credit model (LPCM) with respect to the basic rating scale model (RSM) and the partial credit model (PCM) presented below. Concerning the LLTM, the possibility to use it as a generalization of the Rasch model for repeated measurements was already introduced by Fischer (1974). Over the intervening years this suggestion has been further elaborated. Fischer (1995b) discussed certain design matrices which will be presented in Section 2.3 and on the basis of examples in Section 5.

At this point we will focus on a simple polytomous generalization of the Rasch model, the RSM (Andrich 1978), where each item I_i must have the same number of categories. Pertaining to Equation 1, ϕ_h may be set to h with $h = 0, ..., m$. Since in the RSM the number of item categories is constant, m is used instead of m_i. Hence, it follows that

$$P(X_{vi} = h) = \frac{\exp[h(\theta_v + \beta_i) + \omega_h]}{\sum_{l=0}^{m} \exp[l(\theta_v + \beta_i) + \omega_l]},$$
(5)

with k item parameters $\beta_1, ..., \beta_k$ and $m+1$ category parameters $\omega_0, ..., \omega_m$. This parameterization causes a scoring of the response categories C_h which is constant over the single items. Again, the item parameters can be split up in a linear combination as in Equation 4. This leads to the LRSM proposed by Fischer and Parzer (1991).

Finally, the PCM developed by Masters (1982) and its linear extension, the LPCM (Fischer and Ponocny 1994), are presented. The PCM assigns one parameter β_{ih} to each $I_i \times C_h$ combination for $h = 0, ..., m_i$. Thus, the constant scoring property must not hold over the items and in addition, the items can have different numbers of response categories denoted by m_i. Therefore, the PCM can be regarded as a generalization of the RSM and the probability for a response of person v on category h (item i) is defined as

$$P(X_{vih} = 1) = \frac{\exp[h\theta_v + \beta_{ih}]}{\sum_{l=0}^{m_i} \exp[l\theta_v + \beta_{il}]}.$$
(6)

It is obvious that (6) is a simplification of (2) in terms of $\phi_h = h$. As for the LLTM and the

LRSM, the LPCM is defined by reparameterizing the item parameters of the basic model, i.e.,

$$\beta_{ih} = \sum_{j=1}^{p} w_{ihj}\eta_j. \tag{7}$$

These six models constitute a hierarchical order as displayed in Figure 1. This hierarchy is the base for a unified CML approach presented in the next section. It is outlined again that the linear extension models can be regarded either as generalizations or as more restrictive formulations pertaining to the underlying base model. The hierarchy for the basic model is straightforward: The RM allows only items with two categories, thus each item is represented by one parameter β_i. The RSM allows for more than two (ordinal) categories each represented by a category parameter ω_h. Due to identifiability issues, ω_0 and ω_1 are restricted to 0. Hence, the RM can be seen as a special case of the RSM whereas, the RSM in turn, is a special case of the PCM. The latter model assigns the parameter β_{ih} to each $I_i \times C_h$ combination.

To conclude, the most general model is the LPCM. All other models can be considered as simplifications of Equation 6 combined with Equation 7. As a consequence, once an estimation procedure is established for the LPCM, this approach can be used for any of the remaining models. This is what we quote as *unified CML approach*. The corresponding likelihood equations follow in Section 3.

2.3. The concept of virtual items

When operating with longitudinal models, the main research question is whether an individual's test performance changes over time. The most intuitive way would be to look at the shift in ability θ_v across time points. Such models are presented e.g. in Mislevy (1985), Glas (1992), and discussed by Hoijtink (1995).

Yet there exists another look onto time dependent changes, as presented in Fischer (1995b, p 158ff.): The person parameters are fixed over time and instead of them the item parameters change. The basic idea is that one item I_i is presented at two different times to the same person S_v is regarded as a pair of *virtual items*. Within the framework of extended Rasch models, any change in θ_v occuring between the testing occasions can be described without loss of generality as a change of the item parameters, instead of describing change in terms of the person parameter. Thus, with only two measurement points, I_i with the corresponding parameter β_i generates two virtual items I_r and I_s with associated item parameters β_r^* and β_s^*. For the first measurement point $\beta_r^* = \beta_i$, whereas for the second $\beta_s^* = \beta_i + \tau$. In this linear combination the β^*-parameters are composed additively by means of the real item parameters β and the treatment effects τ. This concept extends to an arbitrary number of time points or testing occasions.

Correspondingly, for each measurement point t we have a vector of *virtual item parameters* $\boldsymbol{\beta}^{*(t)}$ of length k. These are linear reparameterizations of the original $\boldsymbol{\beta}^{(t)}$, and thus the CML approach can be used for estimation. In general, for a simple LLTM with two measurement points the design matrix $\boldsymbol{W}$ is of the form as given in Table 1.

The parameter vector $\boldsymbol{\beta}^{*(1)}$ represents the item parameters for the first test occasion, $\boldsymbol{\beta}^{*(2)}$ the parameters for the second occasion. It might be of interest whether these vectors differ. The corresponding trend contrast is η_{k+1}. Due to this contrast, the number of original β-parameters is doubled by introducing the $2k$ virtual item parameters. If we assume a constant shift for all item parameters, it is only necessary to estimate $\hat{\boldsymbol{\eta}}' = (\hat{\eta}_1, ..., \hat{\eta}_{k+1})$ where $\hat{\eta}_{k+1}$ gives the amount of shift. Since according to (4), the vector $\hat{\boldsymbol{\beta}}^*$ is just a linear combination of $\hat{\boldsymbol{\eta}}$.

As mentioned in the former section, when using models with linear extensions it is possible to impose group contrasts. By doing this, one allows that the item difficulties are different across subgroups. However, this is possible only for models with repeated measurements and virtual items since otherwise the introduction of a group contrast leads to overparameterization and the group effect cannot be estimated by using CML.

		η_1	η_2	$\ldots$	η_k	η_{k+1}
Time 1	$\beta_1^{*(1)}$	1	0	0	0	0
	$\beta_2^{*(1)}$	0	1	0	0	0
	$\vdots$			$\ddots$		$\vdots$
	$\beta_k^{*(1)}$	1	0	0	1	0
Time 2	$\beta_{k+1}^{*(2)}$	1	0	0	0	1
	$\beta_{k+2}^{*(2)}$	0	1	0	0	1
	$\vdots$			$\ddots$		$\vdots$
	$\beta_{2k}^{*(2)}$	1	0	0	1	1

Table 1: A design matrix for an LLTM with two timepoints.

Table 2 gives an example for a repeated measurement design where the effect of a treatment is to be evaluated by comparing item difficulties regarding a control and a treatment group. The number of virtual parameters is doubled compared to the model matrix given in Table 1.

			η_1	η_2	$\ldots$	η_k	η_{k+1}	η_{k+2}
Time 1	Group 1	$\beta_1^{*(1)}$	1	0	0	0	0	0
		$\beta_2^{*(1)}$	0	1	0	0	0	0
		$\vdots$			$\ddots$		$\vdots$	$\vdots$
		$\beta_k^{*(1)}$	1	0	0	1	0	0
	Group 2	$\beta_{k+1}^{*(1)}$	1	0	0	0	0	0
		$\beta_{k+2}^{*(1)}$	0	1	0	0	0	0
		$\vdots$			$\ddots$		$\vdots$	$\vdots$
		$\beta_{2k}^{*(1)}$	1	0	0	1	0	0
Time 2	Group 1	$\beta_1^{*(2)}$	1	0	0	0	1	0
		$\beta_2^{*(2)}$	0	1	0	0	1	0
		$\vdots$			$\ddots$		$\vdots$	$\vdots$
		$\beta_k^{*(2)}$	1	0	0	1	1	0
	Group 2	$\beta_{k+1}^{*(2)}$	1	0	0	0	1	1
		$\beta_{k+2}^{*(2)}$	0	1	0	0	1	1
		$\vdots$			$\ddots$		$\vdots$	$\vdots$
		$\beta_{2k}^{*(2)}$	1	0	0	1	1	1

Table 2: Design matrix for a repeated measurements design with treatment and control group.

Again, η_{k+1} is the parameter that refers to the time contrast, and η_{k+2} is a group effect within measurement point 2. More examples are given in Section 5 and further explanations can be found in Fischer (1995b), Fischer and Ponocny (1994), and in the software manual for the LPCM-Win program by Fischer and Ponocny-Seliger (1998).

By introducing the concept of virtual persons, **eRm** allows for the computation of the linear logistic test model with relaxed assumptions (LLRA Fischer 1977). Corresponding explanations will be given in a subsequent version of this vignette.

3. Estimation of item and person parameters

3.1. CML for item parameter estimation

The main idea behind the CML estimation is that the person's raw score $r_v = \sum_{i=1}^{k} x_{vi}$ is a sufficient statistic. Thus, by conditioning the likelihood onto $\boldsymbol{r}' = (r_1, ..., r_n)$, the person parameters $\boldsymbol{\theta}$, which in this context are nuisance parameters, vanish from the likelihood equation, thus, leading to consistently estimated item parameters $\hat{\boldsymbol{\beta}}$.

Some restrictions have to be imposed on the parameters to ensure identifiability. This can be achieved, e.g., by setting certain parameters to zero depending on the model. In the Rasch model one item parameter has to be fixed to 0. This parameter may be considered as baseline difficulty. In addition, in the RSM the category parameters ω_0 and ω_1 are also constrained to 0. In the PCM all parameters representing the first category, i.e. β_{i0} with $i = 1, \ldots, k$, and one additional item-category parameter, e.g., β_{11} have to be fixed. For the linear extensions it holds that the β-parameters that are fixed within a certain condition (e.g. first measurement point, control group etc.) are also constrained in the other conditions (e.g. second measurement point, treatment group etc.).

At this point, for the LPCM the likelihood equations with corresponding first and second order derivatives are presented (i.e. *unified CML equations*). In the first version of the **eRm** package numerical approximations of the Hessian matrix are used. However, to ensure numerical accuracy and to speed up the estimation process, it is planned to implement the analytical solution as given below.

The conditional log-likelihood equation for the LPCM is

$$\log L_c = \sum_{i=1}^{k} \sum_{h=1}^{m_i} x_{+ih} \sum_{j=1}^{p} w_{ihj} \eta_j - \sum_{r=1}^{r_{max}} n_r \log \gamma_r. \tag{8}$$

The maximal raw score is denoted by r_{max} whereas the number of subjects with the same raw score is quoted as n_r. Alternatively, by going down to an individual level, the last sum over r can be replaced by $\sum_{v=1}^{n} \log \gamma_{r_v}$. It is straightforward to show that the LPCM as well as the other extended Rasch models, define an exponential family (Andersen 1983). Thus, the raw score r_v is minimally sufficient for θ_v and the item totals $x_{.ih}$ are minimally sufficient for β_{ih}.

Crucial expressions are the γ-terms which are known as *elementary symmetric functions*. More details about these terms are given in the next section. However, in the **eRm** package the numerically stable *summation algorithm* as suggested by Andersen (1972) is implemented. Fischer and Ponocny (1994) adopted this algorithm for the LPCM and devised also the first order derivative for computing the corresponding derivative of $\log L_c$:

$$\frac{\partial \log L_c}{\partial \eta_a} = \sum_{i=1}^{k} \sum_{h=1}^{m_i} w_{iha} \left(x_{+ih} - \epsilon_{ih} \sum_{r=1}^{r_{max}} n_r \frac{\gamma_r^{(i)}}{\gamma_r} \right). \tag{9}$$

It is important to mention that for the CML-representation, the multiplicative Rasch expression is used throughout equations 1 to 7, i.e., $\epsilon_i = \exp(-\beta_i)$ for the person parameter. Therefore, ϵ_{ih} corresponds to the reparameterized item $\times$ category parameter whereas $\epsilon_{ih} > 0$. Furthermore, $\gamma_r^{(i)}$ are the first order derivatives of the γ-functions with respect to item i. The index a in η_a denotes the first derivative with respect to the ath parameter.

For the second order derivative of $\log L_c$, two cases have to be distinguished: the derivatives for the off-diagonal elements and the derivatives for the main diagonal elements. The item categories with respect to the item index i are coded with h_i, and those referring to item l with h_l. The second order derivatives of the γ-functions with respect to items i and l are denoted by $\gamma_r^{(i,l)}$. The

corresponding likelihood expressions are

$$\frac{\partial \log L_c}{\partial \eta_a \eta_b} = - \sum_{i=1}^{k} \sum_{h_i=1}^{m_i} w_{ih_i a} w_{ih_i b} \epsilon_{ih_i} \sum_{r=1}^{r_{max}} n_r \frac{\log \gamma_{r-h_i}}{\gamma_r}$$

$$- \sum_{i=1}^{k} \sum_{h_i=1}^{m_i} \sum_{l=1}^{k} \sum_{h_l=1}^{m_l} w_{ih_i a} w_{lh_l b} \left[\epsilon_{ih_i} \epsilon_{lh_l} \left(\sum_{r=1}^{r_{max}} n_r \frac{\gamma_r^{(i)} \gamma_r^{(l)}}{\gamma_r^2} - \sum_{r=1}^{r_{max}} n_r \frac{\gamma_r^{(i,l)}}{\gamma_r} \right) \right] \tag{10}$$

for $a \neq b$, and

$$\frac{\partial \log L_c}{\partial \eta_a^2} = - \sum_{i=1}^{k} \sum_{h_i=1}^{m_i} w_{ih_i a}^2 \epsilon_{ih_i} \sum_{r=1}^{r_{max}} n_r \frac{\log \gamma_{r-h_i}}{\gamma_r}$$

$$- \sum_{i=1}^{k} \sum_{h_i=1}^{m_i} \sum_{l=1}^{k} \sum_{h_l=1}^{m_l} w_{ih_i a} w_{lh_l a} \epsilon_{ih_i} \epsilon_{lh_l} \sum_{r=1}^{r_{max}} n_r \frac{\gamma_{r-h_i}^{(i)} \gamma_{r-h_l}^{(l)}}{\gamma_r^2} \tag{11}$$

for $a = b$.

To solve the likelihood equations with respect to $\hat{\eta}$, a Newton-Raphson algorithm is applied. The update within each iteration step s is performed by

$$\hat{\boldsymbol{\eta}}_s = \hat{\boldsymbol{\eta}}_{s-1} - \mathbf{H}_{s-1}^{-1} \boldsymbol{\delta}_{s-1}. \tag{12}$$

The starting values are $\hat{\boldsymbol{\eta}}_0 = \mathbf{0}$. $\mathbf{H}_{s-1}^{-1}$ is the inverse of the Hessian matrix composed by the elements given in Equation 10 and 11 and $\boldsymbol{\delta}_{s-1}$ is the gradient at iteration $s-1$ as specified in Equation 9. The iteration stops if the likelihood difference $\left| \log L_c^{(s)} - \log L_c^{(s-1)} \right| \leq \varphi$ where φ is a predefined (small) iteration limit. Note that in the current version (0.14-4) $\mathbf{H}$ is approximated numerically by using the **nlm** Newton-type algorithm provided in the **stats** package. The analytical solution as given in Equation 10 and 11 will be implemented in the subsequent version of **eRm**.

3.2. Mathematical properties of the CML estimates

A variety of estimation approaches for IRT models in general and for the Rasch model in particular are available: The *joint maximum likelihood* (JML) estimation as proposed by Wright and Panchapakesan (1969) which is not recommended since the estimates are not consistent (see e.g. Haberman 1977). The basic reason for that is that the person parameters $\boldsymbol{\theta}$ are nuisance parameters; the larger the sample size, the larger the number of parameters.

A well-known alternative is the *marginal maximum likelihood* (MML) estimation (Bock and Aitkin 1981): A distribution $g(\theta)$ for the person parameters is assumed and the resulting situation corresponds to a mixed-effects ANOVA: Item difficulties can be regarded as fixed effects and person abilities as random effects. Thus, IRT models fit into the framework of *generalized linear mixed models* (GLMM) as elaborated in de Boeck and Wilson (2004). By integrating over the ability distribution the random nuisance parameters can be removed from the likelihood equations. This leads to consistent estimates of the item parameters. Further discussions of the MML approach with respect to the CML method will follow.

For the sake of completeness, some other methods for the estimation of the item parameters are the following: Anderson, Li, and Vermunt (2007) propose a Pseudo-ML approach, Molenaar (1995) and Linacre (2004) give an overview of various (heuristic) non-ML methods, Bayesian techniques can be found in Baker and Kim (2004, Chapter 7), and for nonparameteric approaches it is referred to de Leeuw and Verhelst (1986).

However, back to CML, the main idea behind this approach is the assumption that the raw score r_v is a minimal sufficient statistic for θ_v. Starting from the equivalent multiplicative expression of

Equation 1 with $\xi_v = \exp(\theta_v)$ and $\epsilon_i = \exp(-\beta_i)$, i.e.,

$$P(X_{vi} = 1) = \frac{\xi_v \epsilon_i}{1 + \xi_v \epsilon_i}, \tag{13}$$

the following likelihood for the response pattern $\boldsymbol{x}_v$ for a certain subject v results:

$$P(\boldsymbol{x}_v | \xi_v, \boldsymbol{\epsilon}) = \prod_{i=1}^{k} \frac{(\xi_v \epsilon_i)^{x_{vi}}}{1 + \xi_v \epsilon_i} = \frac{\theta_v{}^{r_v} \prod_{i=1}^{k} \epsilon_i{}^{x_{vi}}}{\prod_{i=1}^{k}(1 + \xi_v \epsilon_i)}. \tag{14}$$

Using the notation $\boldsymbol{y} = (y_1, \ldots, y_k)$ for all possible response patterns with $\sum_{i=1}^{k} y_i = r_v$, the probability for a fixed raw score r_v is

$$P(r_v | \xi_v, \boldsymbol{\epsilon}) = \sum_{\boldsymbol{y} | r_v} \prod_{i=1}^{k} \frac{(\xi_v \epsilon_i)^{x_{vi}}}{1 + \xi_v \epsilon_i} = \frac{\theta_v{}^{r_v} \sum_{\boldsymbol{y} | r_v} \prod_{i=1}^{k} \epsilon_i{}^{x_{vi}}}{\prod_{i=1}^{k}(1 + \xi_v \epsilon_i)}. \tag{15}$$

The crucial term with respect to numerical solutions of the likelihood equations is the second term in the numerator:

$$\gamma_r(\epsilon_i) \equiv \sum_{\boldsymbol{y} | r_v} \prod_{i=1}^{k} \epsilon_i{}^{x_{vi}} \tag{16}$$

These are the *elementary symmetric functions* (of order r). An overview of efficient computational algorithms and corresponding simulation studies can be found in Liou (1994). The **eRm** package uses the summation algorithm as proposed by Andersen (1972).

Finally, by collecting the different raw scores into the vector $\boldsymbol{r}$ the conditional probability of observing response pattern $\boldsymbol{x}_v$ with given raw score r_v is

$$P(\boldsymbol{x}_v | r_v, \boldsymbol{\epsilon}) = \frac{P(\boldsymbol{x}_v | \xi_v, \boldsymbol{\epsilon})}{P(r_v | \xi_v, \boldsymbol{\epsilon})}. \tag{17}$$

By taking the product over the persons (independence assumption), the (conditional) likelihood expression for the whole sample becomes

$$L(\boldsymbol{\epsilon} | \boldsymbol{r}) = P(\boldsymbol{x} | \boldsymbol{r}, \boldsymbol{\epsilon}) = \prod_{v=1}^{n} \frac{\prod_{i=1}^{k} \epsilon_i{}^{x_{vi}}}{\gamma_{r_v}}. \tag{18}$$

With respect to raw score frequencies n_r and by reintroducing the β-parameters, (18) can be reformulated as

$$L(\boldsymbol{\beta} | \boldsymbol{r}) = \frac{\exp\left(\sum_{i=1}^{k} x_{+i} \beta_i\right)}{\prod_{r=0}^{k} \gamma_r^{n_r}}, \tag{19}$$

where x_{+i} are the item raw scores. It is obvious that by conditioning the likelihood on the raw scores $\boldsymbol{r}$, the person parameters completely vanished from the expression. As a consequence, the parameters $\hat{\boldsymbol{\beta}}$ can be estimated without knowledge of the subject's abilities. This issue is referred as *person-free item assessment* and we will discuss this topic within the context of specific objectivity in the next section.

Pertaining to asymptotical issues, it can be shown that under mild regularity conditions (Pfanzagl 1994) the CML estimates are consistent for $n \to \infty$ and k fixed, unbiased, asymptotically efficient, and normally distributed (Andersen 1970). For the computation of a Rasch model, comparatively small samples are sufficient to get reliable estimates (Fischer 1988). Whether the MML estimates are unbiased depends on the correct specification of the ability distribution $g(\theta)$. In case of an incorrect assumption, the estimates are biased which is surely a drawback of this method. If $g(\theta)$ is specified appropriately, the CML and MML estimates are asymptotically equivalent (Pfanzagl 1994).

Fischer (1981) elaborates on the conditions for the existence and the uniqueness of the CML estimates. The crucial condition for the data matrix is that $\boldsymbol{X}$ has to be *well-conditioned*. To introduce this issue it is convenient to look at a matrix which is *ill-conditioned*: A matrix is ill-conditioned if there exists a partition of the items into two nonempty subsets such that all of a group of subjects responded correctly to items $i+1, \ldots, k$ ($\boldsymbol{X}_2$) and all of all other subjects failed for items $1, \ldots, i$ ($\boldsymbol{X}_3$), i.e.,

$$\boldsymbol{X} = \left(\begin{array}{c|c} \boldsymbol{X}_1 & \boldsymbol{X}_2 \\ \hline \boldsymbol{X}_3 & \boldsymbol{X}_4 \end{array} \right) = \left(\begin{array}{ccc|ccc} & & & 1 & \ldots & 1 \\ & \boldsymbol{X}_1 & & \vdots & \ddots & \vdots \\ & & & 1 & \ldots & 1 \\ \hline 0 & \ldots & 0 & & & \\ \vdots & \ddots & \vdots & & \boldsymbol{X}_4 & \\ 0 & \ldots & 0 & & & \end{array} \right)$$

Thus, following the definition in Fischer (1981): $\boldsymbol{X}$ will be called *well-conditioned* iff in every possible partition of the items into two nonempty subsets some subjects has given response 1 on some item in the first set and response 0 on some item in the second set. In this case a unique solution for the CML estimates $\hat{\boldsymbol{\beta}}$ exists.

This issue is important for structurally incomplete designs which often occur in practice; different subsets of items are presented to different groups of persons $g = 1, \ldots, G$ where $G \leq n$. As a consequence, the likelihood values have to be computed for each group separately and the joint likelihood is the product over the single group likelihoods. Hence, the likelihood in Equation 19 becomes

$$L(\boldsymbol{\beta}|\boldsymbol{r}) = \prod_{g=1}^{G} \frac{\exp\left(\sum_{i=1}^{k} x_{+i}\beta_i\right)}{\prod_{r=0}^{k} \gamma_{g,r}{}^{n_{g,r}}} \tag{20}$$

This also implies the necessity to compute the elementary symmetric functions separately for each group. The **eRm** package can handle such structurally incomplete designs.

From the elaborations above it is obvious that from an asymptotical point of view the CML estimates are at least as good as the MML estimates. In the past, computational problems (speed, numerical accuracy) involved in calculating the elementary symmetric functions limited the practical usage of the CML approach (see e.g. Gustafsson 1980). Nowadays, these issues are less crucial due to increased computer power.

In some cases MML estimation has advantages not shared by CML: MML leads to finite person parameters even for persons with zero and perfect raw score, and such persons are not removed from the estimation process (Molenaar 1995). On he other hand the consideration of such persons does not seem meaningful from a substantial point of view since the person parameters are not reliable anymore – for such subjects the test is too difficult or too easy, respectively. Thus, due to these covering effects, a corresponding ability estimation is not feasible. However, if the research goal is to find ability distributions such persons should be regarded and MML can handle this.

When estimates for the person parameters are of interest some care has to be taken if the CML method is used since person parameters cancel from the estimation equations. Usually, they are estimated (once having obtained values for the item parameters) by inserting $\hat{\boldsymbol{\beta}}$ (or equivalently $\hat{\boldsymbol{\epsilon}}$) into Equation 14 and solving with respect to $\boldsymbol{\theta}$. Alternatively, Bayesian procedures are applicable (Hoijtink and Boomsma 1995). It is again pointed out that each person in the sample gets an own parameter even though limited by the number of different raw scores.

3.3. CML and specific objectivity

In general, the Rasch model can be regarded as a measurement model: Starting from the (nomi-

nally scaled) 0/1-data matrix $\boldsymbol{X}$, the person raw scores r_v are on an ordinal level. They, in turn, are used to estimate the item parameters $\boldsymbol{\beta}$ which are on an interval scale provided that the Rasch model holds.

Thus, Rasch models allow for comparisons between objects on an interval level. Rasch reasoned on requirements to be fulfilled such that a specific proposition within this context can be regarded as "scientific". His conclusions were that a basic requirement is the "objectivity" of comparisons (Rasch 1961). This claim contrasts assumptions met in *classical test theory* (CTT). A major advantage of the Rasch model over CTT models is the *sample independence* of the results. The relevant concepts in CTT are based on a linear model for the "true score" leading to some indices, often correlation coefficients, which in turn depend on the observed data. This is a major drawback in CTT. According to Fischer (1974), sample independence in IRT models has the following implications:

- The person-specific results (i.e., essentially $\boldsymbol{\theta}$) do not depend on the assignment of a person to a certain subject group nor on the selected test items from an item pool Ψ.

- Changes in the skills of a person on the latent trait can be determined independently from its base level and independently from the selected item subset $\psi \subset \Psi$.

- From both theoretical and practical perspective the requirement for representativeness of the sample is obsolete in terms of a true random selection process.

Based on these requirements for parameter comparisons, Rasch (1977) introduced the term *specific objectivity*: *objective* because any comparison of a pair of parameters is independent of any other parameters or comparisons; *specifically objective* because the comparison made was relative to some specified frame of reference (Andrich 1988). In other words, if specific objectivity holds, two persons v and w with corresponding parameters θ_v and θ_w, are comparable independently from the remaining persons in the sample and independently from the presented item subset ψ. In turn, for two items i and j with parameters β_i and β_j, the comparison of these items can be accomplished independently from the remaining items in Ψ and independently from the persons in the sample.

The latter is crucial since it reflects completely what is called sample independence. If we think not only of comparing β_i and β_j but rather to estimate these parameters, we achieve a point where specific objectivity requires a procedure which is able to provide estimates $\hat{\boldsymbol{\beta}}$ that do not depend on the sample. This implies that $\hat{\boldsymbol{\beta}}$ should be computable without the involvement of $\boldsymbol{\theta}$. CML estimation fulfills this requirement: By conditioning on the sufficient raw score vector $\boldsymbol{r}$, $\boldsymbol{\theta}$ disappears from the likelihood equation and $L(\boldsymbol{\beta}|\boldsymbol{r})$ can be solved without knowledge of $\boldsymbol{\theta}$. This issue is referred to as *separability of item and person parameters* (see e.g. Wright and Masters 1982). Furthermore, separability implies that no specific distribution should be assumed neither for the person nor for the item parameters (Rost 2000). MML estimation requires such assumptions. At this point it is clear that CML estimation is the only estimation method within the Rasch measurement context fulfilling the requirement of *person-free item calibration* and, thus, it maps the epistemological theory of specific objectivity to a statistical maximum likelihood framework. Note that strictly speaking any statistical result based on sample observations is sample-dependent because any result depends at least on the sample size (Fischer 1987). The estimation of the item parameters is "sample-independent", a term indicating the fact that the actually obtained sample of a certain population is not of relevance for the statistical inference on these parameters (Kubinger 1989, p. 23).

3.4. Estimation of person parameters

CML estimation for person parameters is not recommended due to computational issues. The **eRm** package provides two methods for this estimation. The first is ordinary ML where the CML-based item parameters are plugged into the joint ML equation. The likelihood is optimized with respect

to $\boldsymbol{\theta}$. Andersen (1995) gives a general formulation of this ML estimate with $r_v = r$ and $\theta_v = \theta$:

$$r - \sum_{i=1}^{k} \sum_{h=1}^{m_i} \frac{h \exp(h\theta + \hat{\beta}_{ih})}{\sum_{l=0}^{m_i} \exp(h\theta_v + \hat{\beta}_{il})} = 0 \tag{21}$$

Warm (1989) proposed a weighted likelihood estimation (WLE) which is more accurate compared to ML. For the dichotomous Rasch model the expression to be solved with respect to $\boldsymbol{\theta}$ is

$$P(\theta_v | \boldsymbol{x}_v, \hat{\boldsymbol{\beta}}) \propto \frac{exp(r_v \theta_v)}{\prod_i (1 + exp(\theta_v - \hat{\beta}_i))} \sum_i p_{vi}(1 - p_{vi}) \tag{22}$$

Again, the item parameter vector $\hat{\boldsymbol{\beta}}$ is used from CML. This approach will implemented in a subsequent **eRm** version. Additional explanations and simulation studies regarding person parameter estimation can be found in Hoijtink and Boomsma (1995).

4. Testing extended Rasch models

Testing IRT models involves two parts: First, item- and person-wise statistics can be examined; in particular item-fit and person-fit statistics. Secondly, based on CML properties, various model tests can be derived (see Glas and Verhelst 1995a,b).

4.1. Item-fit and person-fit statistics

Commonly in IRT, items and persons are excluded due to item-fit and person-fit statistics. Both are residual based measures: The observed data matrix $\mathbf{X}$ is compared with the model probability matrix $\mathbf{P}$. Computing standardized residuals for all observations gives the $n \times k$ residual matrix $\mathbf{R}$. The squared column sums correspond to item-fit statistics and the squared row sums to person-fit statistics both of which are χ^2-distributed with the corresponding degrees of freedom. Based on these quantities unweighted (*outfit*) and weighted (*infit*) mean-square statistics can also be used to evaluate item and person fit (see e.g. Wright and Masters 1982).

4.2. A Wald test for item elimination

A helpful implication of CML estimates is that subsequent test statistics are readily obtained and model tests are easy to carry out. Basically, we have to distinguish between test on item level and global model tests.

On item level, sample independence reflects the property that by splitting up the sample in, e.g., two parts, the corresponding parameter vectors $\hat{\boldsymbol{\beta}}^{(1)}$ and $\hat{\boldsymbol{\beta}}^{(2)}$ should be the same. Thus, when we want to achieve Rasch model fit those items have to be eliminated from the test which differ in the subsamples. This important issue in test calibration can be examined, e.g., by using a graphical model test. Fischer and Scheiblechner (1970) propose a $N(0, 1)$-distributed test statistic which compares the item parameters for two subgroups:

$$z = \frac{\beta_i^{(1)} - \beta_i^{(2)}}{\sqrt{Var_i^{(1)} - Var_i^{(2)}}} \tag{23}$$

The variance term in the denominator is based on Fisher's function of "information in the sample". However, as Glas and Verhelst (1995a) point out discussing their Wald-type test that this term can be extracted directly from the variance-covariance matrix of the CML estimates. This Wald approach is provided in **eRm** by means of the function `Waldtest()`.

4.3. Andersen's likelihood-ratio test

In the **eRm** package the likelihood ratio test statistic LR, initially proposed by Andersen (1973) is computed for the RM, the RSM, and the PCM. For the models with linear extensions, LR has

to be computed separately for each measurement point and subgroup.

$$LR = 2 \left(\sum_{g=1}^{G} \log L_c(\hat{\boldsymbol{\eta}}_g; \boldsymbol{X}_g) - \log L_c(\hat{\boldsymbol{\eta}}; \boldsymbol{X}) \right) \tag{24}$$

The underlying principle of this test statistic is that of *subgroup homogeneity* in Rasch models: for arbitrary disjoint subgroups $g = 1, ..., G$ the parameter estimates $\hat{\boldsymbol{\eta}}_g$ have to be the same. LR is asymptotically χ^2-distributed with *df* equal to the number of parameters estimated in the subgroups minus the number of parameters in the total data set. For the sake of computational efficiency, the **eRm** package performs a person raw score median split into two subgroups. In addition, a graphical model test (Rasch 1960) based on these estimates is produced by plotting $\hat{\boldsymbol{\beta}}_1$ against $\hat{\boldsymbol{\beta}}_2$. Thus, critical items (i.e. those fairly apart from the diagonal) can be identified and eliminated. Further elaborations and additional test statistics for polytomous Rasch models can be found, e.g., in Glas and Verhelst (1995a).

4.4. Nonparametric ("exact") Tests

Based on the package **RaschSampler** by Verhelst, Hatzinger, and Mair (2007) several Rasch model tests as proposed by (Ponocny 2001) are provided.

4.5. Martin-Löf Test

Applying the LR principle to subsets of items, Martin-Löf (1973, see Glas and Verhelst 1995a) suggested a statistic to evaluate if two groups of items are homogeneous, i.e., to test the unidimensionality axiom.

5. The eRm package and application examples

The underlying idea of the **eRm** package is to provide a user-friendly flexible tool to compute extended Rasch models. This implies, amongst others, an automatic generation of the design matrix **W**. However, in order to test specific hypotheses the user may specify **W** allowing the package to be flexible enough for computing IRT-models beyond their regular applications. In the following subsections, various examples are provided pertaining to different model and design matrix scenarios. Due to intelligibility matters, the artificial data sets are kept rather small. A detailed description in German of applications of various extendend Rasch models using the **eRm** package can be found in Poinstingl, Mair, and Hatzinger (2007).

5.1. Structure of the eRm package

Embedding **eRm** into the flexible framework of R is a crucial benefit over existing stand-alone programs like WINMIRA (von Davier 1998), LPCM-WIN (Fischer and Ponocny-Seliger 1998), and others.

Another important issue in the development phase was that the package should be flexible enough to allow for CML compatible polytomous generalizations of the basic Rasch model such as the RSM and the PCM. In addition, by introducing a design matrix concept linear extensions of these basic models should be applicable. This approach resulted in including the LLTM, the LRSM and the LPCM as the most general model into the **eRm** package. For the latter model the CML estimation was implemented which can be used for the remaining models as well. A corresponding graphical representation is given in Figure 2.

An important benefit of the package with respect to linearly extended models is that for certain models the design matrix $\boldsymbol{W}$ can be generated automatically (LPCM-WIN, Fischer and Ponocny-Seliger 1998) also allows for specifying design matrices but in case of more complex models this can become a tedious task and the user must have a thorough understanding of establishing proper
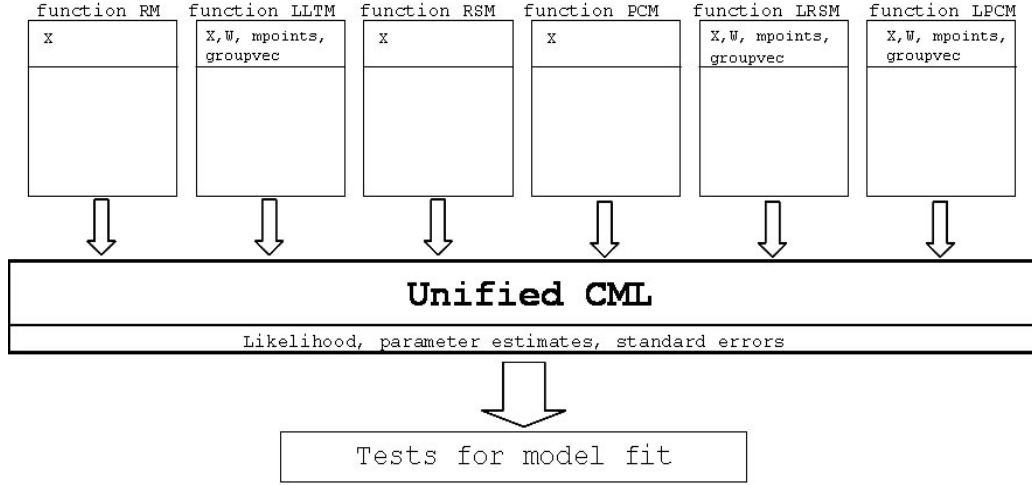
Figure 2: Bodywork of the **eRm** routine

design structures). For repeated measurement models time contrasts in the **eRm** can be simply specified by defining the number of measurement points, i.e., `mpoints`. To regard group contrasts like, e.g., treatment and control groups, a corresponding vector (`groupvec`) can be specified that denotes which person belongs to which group. However, W can also be defined by the user.

A recently added feature of the routine is the option to allow for structurally missing values. This is required, e.g., in situations when different subsets of items are presented to different groups of subjects as described in Section 3.2. These person groups are identified automatically: In the data matrix X, those items which are not presented to a certain subject are declared as `NA`s, as usual in R.

After solving the CML equations by the Newton-Raphson method, the output of the routine consists of the "basic" parameter estimates $\hat{\eta}$, the corresponding variance-covariance matrix, and consequently the vector with the standard errors. Furthermore, the ordinary item parameter estimates $\hat{\beta}$ are computed by using the linear transformation $\hat{\beta} = W\hat{\eta}$. For ordinary Rasch models these basic parameters correspond to the item easiness. For the RM, the RSM, and the PCM, however, we display $\hat{\eta}$ as $-\hat{\eta}$, i.e., as difficulty. It has to be mentioned that the CML equation is solved with the restriction that one item parameter has to be fixed to zero (we use $\beta_1 = 0$). For the sake of interpretability, the resulting estimates $\hat{\beta}$ can easily be transformed into "sum-zero" restricted $\hat{\beta}^*$ by applying $\hat{\beta}_i^* = \hat{\beta}_i - \sum_i \hat{\beta}_i/k$. This transformation is also used for the graphical model test.

5.2. Example 1: Rasch model

We start the example section with a simple Rasch model based on a 100×30 data matrix. First, we estimate the item parameters using the function `RM()` and then the person parameters with `person.parameters()`.

```
> library(eRm)
> data(raschdat1)
> res.rasch <- RM(raschdat1)
> pres.rasch <- person.parameter(res.rasch)
```

Then we use Andersen's LR-test for goodness-of-fit with mean split criterion:

```
> lrres.rasch <- LRtest(res.rasch, splitcr = "mean", se = TRUE)
> lrres.rasch
```

```
Andersen LR-test:
LR-value: 30.288
Chi-square df: 29
p-value:  0.4
```

We see that the model fits and a graphical representation of this result (subset of items only) is given in Figure 3 by means of a goodness-of-fit plot with confidence ellipses.

```
> plotGOF(lrres.rasch, beta.subset=c(14,5,18,7,1), tlab="item", conf=list(ia=FALSE,col="blue",lty
```
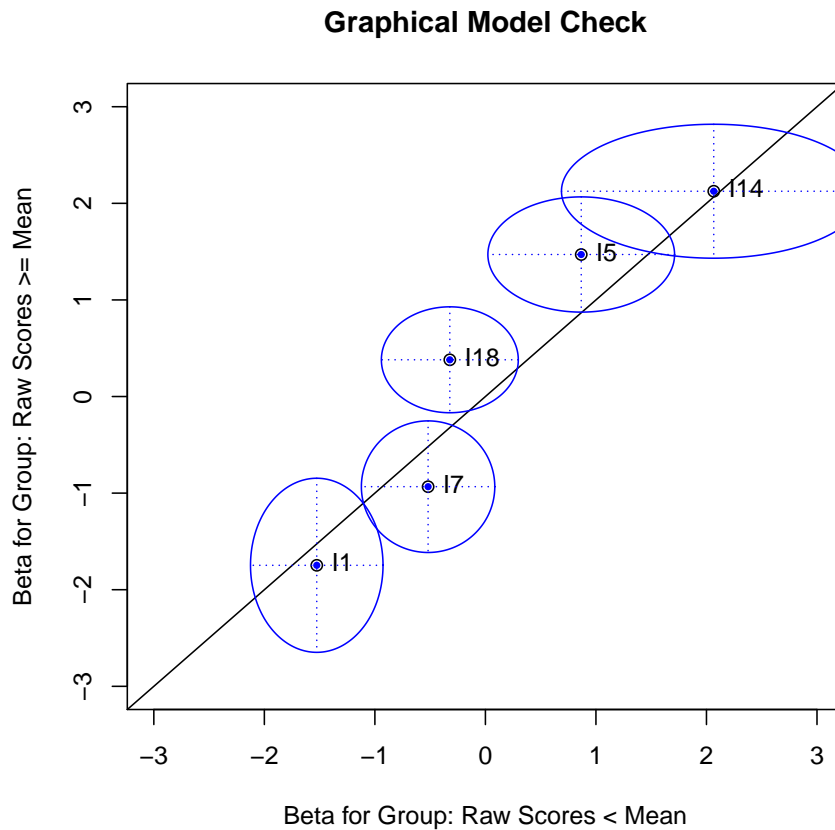
Graphical Model Check

Figure 3:   Goodness-of-fit plot for some items with confidence ellipses.

To be able to draw confidence ellipses it is needed to set `se = TRUE` when computing the LR-test.

5.3. Example 2: LLTM as a restricted Rasch model

As mentioned in Section 2.2, also the models with the linear extensions on the item parameters can be seen as special cases of their underlying basic model. In fact, the LLTM as presented below and following the original idea by Scheiblechner (1972), is a restricted RM, i.e. the number of estimated parameters is smaller compared to a Rasch model. The data matrix $\mathbf{X}$ consists of $n = 15$ persons and $k = 5$ items. Furthermore, we specify a design matrix $\mathbf{W}$ (following Equation 4) with specific weight elements w_{ij}.

```
> data(lltmdat2)
> W <- matrix(c(1,2,1,3,2,2,2,1,1,1),ncol=2)
```

```
> res.lltm <- LLTM(lltmdat2, W)
> summary(res.lltm)


Results of LLTM estimation:

Call:  LLTM(X = lltmdat2, W = W)

Conditional log-likelihood: -31.65225
Number of iterations: 7
Number of parameters: 2


Basic Parameters eta with 0.95 CI:
      Estimate Std. Error lower CI upper CI
eta 1   -0.098      0.313   -0.711    0.516
eta 2    0.114      0.478   -0.823    1.051


Item Easiness Parameters (beta) with 0.95 CI:
         Estimate Std. Error lower CI upper CI
beta I1     0.130      1.098   -2.022    2.283
beta I2     0.033      1.302   -2.519    2.584
beta I3     0.016      0.651   -1.259    1.292
beta I4    -0.179      1.184   -2.500    2.141
beta I5    -0.081      0.903   -1.850    1.688
```

The `summary()` method provides point estimates and standard errors for the basic parameters and for the resulting item parameters. Note that item parameters in **eRm** are always estimated as easiness parameters according to equations 1 and 2 but not 3. If the sign is switched, the user gets difficulty parameters (the standard errors remain the same, of course). However, all plotting functions `plotGOF`, `plotICC`, `plotjointICC`, and `plotPImap`, as well as the function `thresholds` display the difficulty parameters. The same applies for the basic parameters η in the output of the RM, RSM, and PCM.

5.4. Example 3: RSM and PCM

Again, we provide an artificial data set now with $n = 300$ persons and $k = 4$ items; each of them with $m+1 = 3$ categories. We start with the estimation of an RSM and, subsequently, we calculate the corresponding category-intersection parameters using the function `thresholds()`.

```
> data(pcmdat2)
> res.rsm <- RSM(pcmdat2)
> thresholds(res.rsm)


Design Matrix Block 1:
   Location Threshold 1 Threshold 2
I1  1.60712     0.59703     2.61721
I2  1.92251     0.91242     2.93260
I3  0.00331    -1.00678     1.01340
I4  0.50743    -0.50266     1.51752
```

The location parameter is basically the item difficulty and the thesholds are the points in the ICC plot given in Figure 4 where the category curves intersect:

```
> plotICC(res.rsm, mplot=TRUE, legpos=FALSE,ask=FALSE)
```
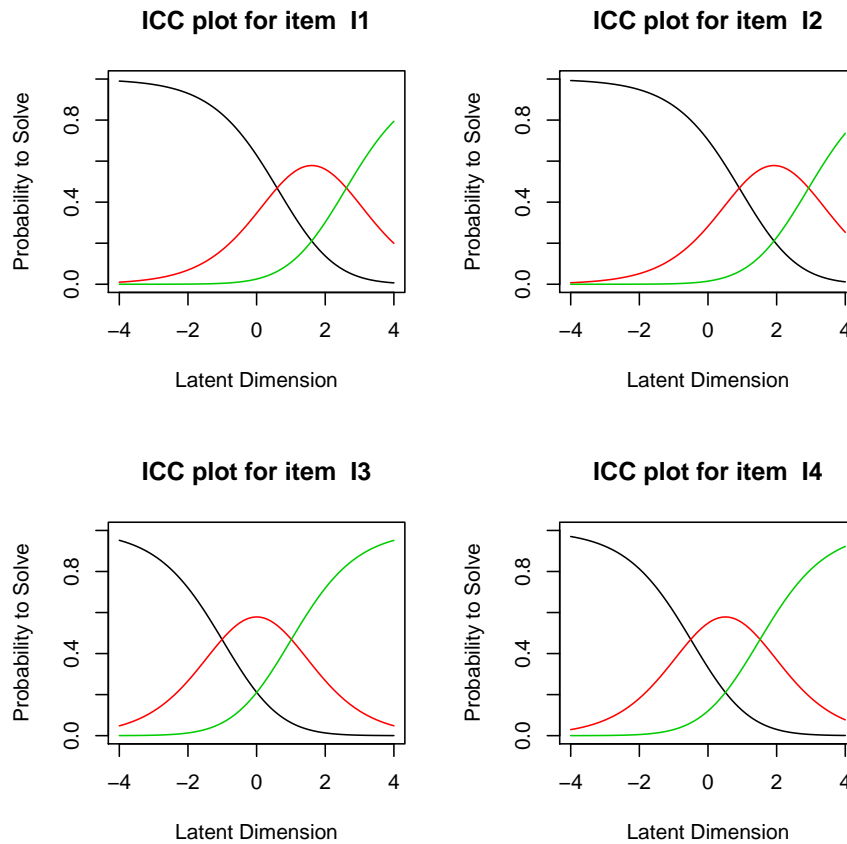
Figure 4:   ICC plot for an RSM.

The RSM restricts the threshold distances to be the same across all items. This strong assumption can be relaxed using a PCM. The results are represented in a person-item map (see Figure 5).

```
> res.pcm <- PCM(pcmdat2)
> plotPImap(res.pcm, sorted = TRUE)
```

After estimating the person parameters we can check the item-fit statistics.

```
> pres.pcm <- person.parameter(res.pcm)
> itemfit(pres.pcm)


Itemfit Statistics:
     Chisq  df p-value Outfit MSQ Infit MSQ Outfit t Infit t
I1 225.617 255   0.907      0.881     0.885    -1.55   -1.53
I2 215.948 255   0.964      0.844     0.903    -1.81   -1.26
I3 179.811 255   1.000      0.702     0.713    -3.55   -3.92
I4 214.473 255   0.969      0.838     0.809    -1.88   -2.54
```

A likelihood ratio test comparing the RSM and the PCM indicates that the PCM provides a better fit.

Person–Item Map
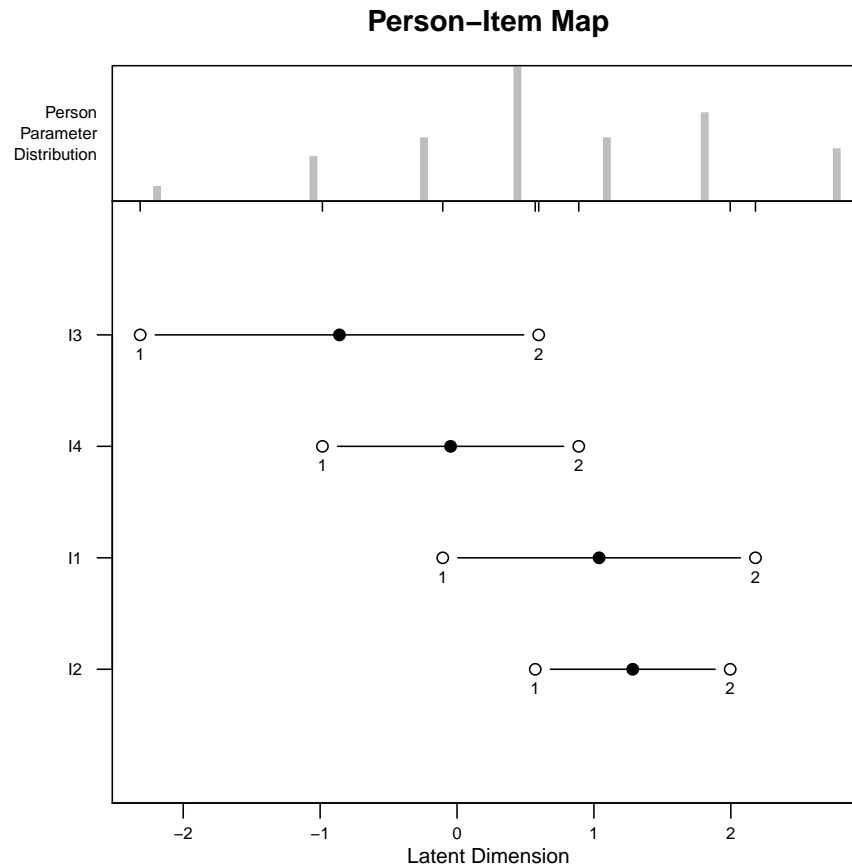

Figure 5: Person-Item map for a PCM.

```
> lr<- 2*(res.pcm$loglik-res.rsm$loglik)
> df<- res.pcm$npar-res.rsm$npar
> pvalue<-1-pchisq(lr,df)
> cat("LR statistic: ", lr, "  df =",df, "  p =",pvalue, "\n")

LR statistic:  11.69992   df = 3   p = 0.00848509
```

5.5. An LPCM for repeated measurements in different groups

The most complex example refers to an LPCM with two measurement points. In addition, the hypothesis is of interest whether the treatment has an effect. The corresponding contrast is the last column in $\mathbf{W}$ below.

First, the data matrix $\mathbf{X}$ is specified. We assume an artificial test consisting of $k = 3$ items which was presented twice to the subjects. The first 3 columns in $\mathbf{X}$ correspond to the first test occasion, whereas the last 3 to the second occasion. Generally, the first k columns correspond to the first test occasion, the next k columns for the second, etc. In total, there are $n = 20$ subjects. Among these, the first 10 persons belong to the first group (e.g., control), and the next 10 persons to the second group (e.g., treatment). This is specified by a group vector:

```
> data(lpcmdat)
> grouplpcm <- rep(1:2, each = 10)
```

Again, W is generated automatically. In general, for such designs the generation of W consists first of the item contrasts, followed by the time contrasts and finally by the group main effects except for the first measurement point (due to identifiability issues, as already described).

```
> reslpcm <- LPCM(lpcmdat, mpoints = 2, groupvec = grouplpcm, sum0 = FALSE)
> model.matrix(reslpcm)
```

	eta 1	eta 2	eta 3	eta 4	eta 5	eta 6	eta 7	eta 8	eta 9	eta 10
I1.c1 t1 g1	0	0	0	0	0	0	0	0	0	0
I1.c2 t1 g1	1	0	0	0	0	0	0	0	0	0
I1.c3 t1 g1	0	1	0	0	0	0	0	0	0	0
I2.c1 t1 g1	0	0	1	0	0	0	0	0	0	0
I2.c2 t1 g1	0	0	0	1	0	0	0	0	0	0
I2.c3 t1 g1	0	0	0	0	1	0	0	0	0	0
I3.c1 t1 g1	0	0	0	0	0	1	0	0	0	0
I3.c2 t1 g1	0	0	0	0	0	0	1	0	0	0
I3.c3 t1 g1	0	0	0	0	0	0	0	1	0	0
I1.c1 t1 g2	0	0	0	0	0	0	0	0	0	0
I1.c2 t1 g2	1	0	0	0	0	0	0	0	0	0
I1.c3 t1 g2	0	1	0	0	0	0	0	0	0	0
I2.c1 t1 g2	0	0	1	0	0	0	0	0	0	0
I2.c2 t1 g2	0	0	0	1	0	0	0	0	0	0
I2.c3 t1 g2	0	0	0	0	1	0	0	0	0	0
I3.c1 t1 g2	0	0	0	0	0	1	0	0	0	0
I3.c2 t1 g2	0	0	0	0	0	0	1	0	0	0
I3.c3 t1 g2	0	0	0	0	0	0	0	1	0	0
I1.c1 t2 g1	0	0	0	0	0	0	0	0	1	0
I1.c2 t2 g1	1	0	0	0	0	0	0	0	2	0
I1.c3 t2 g1	0	1	0	0	0	0	0	0	3	0
I2.c1 t2 g1	0	0	1	0	0	0	0	0	1	0
I2.c2 t2 g1	0	0	0	1	0	0	0	0	2	0
I2.c3 t2 g1	0	0	0	0	1	0	0	0	3	0
I3.c1 t2 g1	0	0	0	0	0	1	0	0	1	0
I3.c2 t2 g1	0	0	0	0	0	0	1	0	2	0
I3.c3 t2 g1	0	0	0	0	0	0	0	1	3	0
I1.c1 t2 g2	0	0	0	0	0	0	0	0	1	1
I1.c2 t2 g2	1	0	0	0	0	0	0	0	2	2
I1.c3 t2 g2	0	1	0	0	0	0	0	0	3	3
I2.c1 t2 g2	0	0	1	0	0	0	0	0	1	1
I2.c2 t2 g2	0	0	0	1	0	0	0	0	2	2
I2.c3 t2 g2	0	0	0	0	1	0	0	0	3	3
I3.c1 t2 g2	0	0	0	0	0	1	0	0	1	1
I3.c2 t2 g2	0	0	0	0	0	0	1	0	2	2
I3.c3 t2 g2	0	0	0	0	0	0	0	1	3	3

The parameter estimates are the following:

```
Results of LPCM estimation:

Call:  LPCM(X = lpcmdat, mpoints = 2, groupvec = grouplpcm, sum0 = FALSE)

Conditional log-likelihood: -103.7225
Number of iterations: 37
Number of parameters: 10
```

```
Basic Parameters eta:
               eta 1      eta 2      eta 3      eta 4      eta 5      eta 6
Estimate -0.4615900 -1.60959 -0.5713666 -0.8388422 -1.739492 -0.7232787
Std.Err   0.7346554  1.19433  0.6232625  0.9854643  1.438174  0.6534176
               eta 7      eta 8      eta 9     eta 10
Estimate -0.7096129 -1.209864 -0.2014868 1.094043
Std.Err   0.9862220  1.414803  0.2608239 0.387040
```

Testing whether the η-parameters equal 0 is mostly not of relevance for those parameters referring to the items (in this example $\eta_1, ..., \eta_8$). But for the remaining contrasts, $H_0 : \eta_9 = 0$ (implying no general time effect) can not be rejected ($p = .44$), whereas hypothesis $H_0 : \eta_{10} = 0$ has to be rejected ($p = .004$) when applying a z-test. This suggests that there is a significant treatment effect over the measurement points. If a user wants to perform additional tests such as a Wald test for the equivalence of two η-parameters, the `vcov` method can be applied to get the variance-covariance matrix.

6. Additional topics

This section will be extended successively with new developments and components which do not directly relate to the modeling core of **eRm** but may prove to be useful add-ons.

6.1. The eRm simulation module

A recent **eRm** development is the implementation of a simulation module to generate 0-1 matrices for different Rasch scenarios. In this article we give a brief overview about the functionality and for more detailed descriptions (within the context of model testing) it is referred to Mair (2006) and Suárez-Falcón and Glas (2003).

For each scenario the user has the option either to assign $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ as vectors to the simulation function (e.g. by drawing parameters from a uniform distribution) or to let the function draw the parameters from a $N(0, 1)$ distribution. The first scenario is the simulation of Rasch homogenous data by means of the function `sim.rasch()`. The parameter values are plugged into equation 3 and it results the matrix $\mathbf{P}$ of model probabilites which is of dimension $n \times k$. An element p_{vi} indicates the probability that subject v solves item i. In a second step the matrix $\mathbf{P}$ has to be transformed into the 0-1 data matrix $\mathbf{X}$. The recommended way to achieve this is to draw another random number $p_{vi}^{\star}$ from a uniform distribution in $[0; 1]$ and perform the transformation according to the following rule:

$$x_{vi} = \begin{cases} 1 & \text{if } p_{vi}^{\star} \leq p_{vi} \\ 0 & \text{if } p_{vi}^{\star} > p_{vi} \end{cases}$$

Alternatively, the user can specify a fixed cutpoint $p^{\star} := p_{vi}^{\star}$ (e.g. $p^{\star} = 0.5$) and make the decision according to the same rule. This option is provided by means of the `cutpoint` argument. Caution is advised when using this deterministic option since this leads likely to ill-conditioned data matrices.

The second scenario in this module regards the violation of the parallel ICC assumption which leads to the two-parameter logistic model (2-PL) proposed by Birnbaum (1968):

$$P(X_{vi} = 1) = \frac{\exp(\alpha_i(\theta_v - \beta_i))}{1 + \exp(\alpha_i(\theta_v - \beta_i))}. \tag{25}$$

The parameter α_i denotes the item discrimination which for the Rasch model is 1 across all items. Thus, each item score gets a weight and the raw scores are not sufficient anymore. The function for simulating 2-PL data is `sim.2pl()` and if $\boldsymbol{\alpha}$ is not specified by the user by means of the argument `discrim`, the discrimination parameters are drawn from a log-normal distribution. The reasons

for using this particular kind of distribution are the following: In the case of $\alpha_i = 1$ the ICC are Rasch consistent. Concerning the violations, it should be possible to achieve deviations in both directions (for $\alpha_i > 0$). If $\alpha_i > 0$ the ICC is steeper than in the Rasch case and, consequently, if $\alpha_i < 1$ the ICC is flatter. This bidirectional deviation around 1 is warranted by the lognormal distribution $LN(\mu, \sigma^2)$ with $\mu = 0$. Since it is a logarithmic distribution, α_i cannot be negative. The degrees of model violation can be steered by mea ns of the dispersion parameter σ^2. A value of $\sigma^2 = .50$ already denotes a strong violation. The lower σ^2, the closer the values lie around 1. In this case the α_i are close to the Rasch slopes.

Using the function `sim.xdim()` the unidimensionality assumptions is violated. This function allows for the simulation of multidimensional Rasch models as for instance given Glas (1992) and Adams, Wilson, and Wang (1997). Multidimensionality implies that one single item measures more than one latent construct. Let us denote the number of these latent traits by D. Consequently, each person has a vector of ability parameters $\boldsymbol{\theta}_v$ of length D. These vectors are drawn from a multivariate normal distribution with mean $\boldsymbol{\mu} = \mathbf{0}$ and VC-matrix $\boldsymbol{\Sigma}$ of dimension $D \times D$. This matrix has to be specified by the user with the argument `Sigma`. In order to achieve strong model violations, very low correlations such as .01 should be provided. To specify to which extend item i is measuring each of the D dimensions, a corresponding vector of weights $\mathbf{z}_i$ of length D is defined. If the resulting $k \times D$ matrix $\mathbf{Z}$ is not provided by the user, `sim.xdim()` generates $\mathbf{Z}$ such that each $\mathbf{z}_i$ contains only nonzero element which indicates the assigned dimension. This corresponds to the *between-item multidimensional model* (Adams *et al.* 1997). However, in any case the person part of the model is $\mathbf{z}_i^T \boldsymbol{\theta}_v$ which replaces θ_v in Equation 3.

Finally, locally dependent item responses can be produced by means of the function `sim.locdep()`. Local dependence implies the introduction of pair-wise item correlations δ_{ij}. If these correlations are constant across items, the argument `it.cor` can be a single value δ. A value $\delta = 0$ corresponds to the Rasch model whereas $\delta = 1$ leads to the strongest violation. Alternatively, for different pair-wise item correlations, the user can specify a VC-matrix Δ of dimension $k \times k$. The formal representation of the corresponding IRT model is

$$P(X_{vi} = 1 | X_{vj} = x_{vj}) = \frac{\exp(\theta_v - \beta_i + x_{vj}\delta_{ij})}{1 + \exp(\theta_v - \beta_i + x_{vj}\delta_{ij})}. \tag{26}$$

This model was proposed by Jannarone (1986) and is suited to model locally dependent item responses.

7. Discussion and outlook

Here we give a brief outline of future **eRm** developments. The CML estimation approach, in combination with the EM-algorithm, can also be used to estimate *mixed Rasch models* (MIRA). The basic idea behind such models is that the extended Rasch model holds within subpopulations of individuals, but with different parameter values for each subgroup. Corresponding elaborations are given in Rost and von Davier (1995).

In Rasch models the item discrimination parameter α_i is always fixed to 1 and thus it does not appear in the basic equation. Allowing for different discrimination parameters across items leads to the two-parameter logistic model as given in Equation 25. In this model the raw scores are not sufficient statistics anymore and hence CML can not be applied. 2-PL models can be estimated by means of the **ltm** package (Rizopoulos 2006). However, Verhelst and Glas (1995) formulated the one parameter logistic model (OPLM) where the α_i do not vary across the items but are unequal to one. The basic strategy to estimate OPLM is a three-step approach: First, the item parameters of the Rasch model are computed. Then, discrimination parameters are computed under certain restrictions. Finally, using these discrimination weights, the item parameters for the OPLM are estimated using CML. This is a more flexible version of the Rasch model in terms of different slopes.

To conclude, the **eRm** package is a tool to estimate extended Rasch models for unidimensional

traits. The generalizations towards different numbers of item categories, linear extensions to allow for introducing item covariates and/or trend and optionally group contrasts are important issues when examining item behavior and person performances in tests. This improves the feasibility of IRT models with respect to a wide variety of application areas.

References

Adams RJ, Wilson M, Wang WC (1997). "The multidimensional random coefficients multinomial logit model." *Applied Psychological Measurement*, **21**, 1–23.

Andersen EB (1970). "Asymptotic properties of conditional maximum likelihood estimators." *Journal of the Royal Statistical Society, Series B*, **32**, 283–301.

Andersen EB (1972). "The numerical solution of a set of conditional estimation equations." *Journal of the Royal Statistical Society, Series B*, **34**, 42–54.

Andersen EB (1973). "A goodness of fit test for the Rasch model." *Psychometrika*, **38**, 123–140.

Andersen EB (1983). "A General Latent Structure Model for Contingency Table Data." In H Wainer, S Messik (eds.), "Principals of Modern Psychological Measurement," pp. 117–138. Erlbaum, Hillsdale, NJ.

Andersen EB (1995). "Polytomous Rasch Models and their Estimation." In G Fischer, I Molenaar (eds.), "Rasch models: Foundations, recent developments, and applications," pp. 271–292. Springer, New York.

Anderson C, Li Z, Vermunt J (2007). "Estimation of models in the Rasch family for polytomous items and multiple latent variables." *Journal of Statistical Software*, **20**(6). URL http://www.jstatsoft.org/v20/i06/.

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

Andrich D (1988). *Rasch Models for Measurement (Sage University paper series on quantitative applications in the social sciences)*. Sage, Newbury Park, CA.

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

Birnbaum A (1968). "Some latent trait models and their use in inferring an examinee's ability." In FM Lord, MR Novick (eds.), "Statistical theories of mental test scores," pp. 395–479. Addison-Wesley, Reading, MA.

Bock RD, Aitkin M (1981). "Marginal maximum likelihood estimation of item parameters: as application of an EM algorithm." *Psychometrika*, **46**, 443–459.

Borsboom D (2006). "The attack of the psychometricians." *Psychometrika*, **71**, 425–440.

de Boeck P, Wilson M (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer, New York.

de Leeuw J, Verhelst N (1986). "Maximum likelihood estimation in generalized Rasch models." *Journal of educational statistics*, **11**, 183–196.

Fischer GH (1973). "The linear logistic test model as an instrument in educational research." *Acta Psychologica*, **37**, 359–374.

Fischer GH (1974). *Einführung in die Theorie psychologischer Tests [Introduction to Mental Test Theory]*. Huber, Bern.

Fischer GH (1977). "Linear Logistic Trait Models: Theory and Application." In H Spada, WF Kempf (eds.), "Structural Models of Thinking and Learning," pp. 203–225. Huber, Bern.

Fischer GH (1981). "On the existence and uniqueness of maximum-likelihood estimates in the Rasch model." *Psychometrika*, **46**, 59–77.

Fischer GH (1987). "Applying the principles of specific objectivity and of generalizability to the measurement of change." *Psychometrika*, **52**, 565–587.

Fischer GH (1988). "Spezifische Objektvität: Eine wissenschaftstheoretische Grundlage des Rasch-Modells. [Specific objectivity: An epistemological foundation of the Rasch model.]." In K Kubinger (ed.), "Moderne Testtheorie," pp. 87–111. Beltz, Weinheim.

Fischer GH (1995a). "Derivations of the Rasch Model." In G Fischer, I Molenaar (eds.), "Rasch Models: Foundations, Recent Developments, and Applications," pp. 15–38. Springer, New York.

Fischer GH (1995b). "Linear Logistic Models for Change." In G Fischer, I Molenaar (eds.), "Rasch Models: Foundations, Recent Developments, and Applications," pp. 157–180. Springer, New York.

Fischer GH, Parzer P (1991). "An extension of the rating scale model with an application to the measurement of change." *Psychometrika*, **56**, 637–651.

Fischer GH, Ponocny I (1994). "An extension of the partial credit model with an application to the measurement of change." *Psychometrika*, **59**, 177–192.

Fischer GH, Ponocny-Seliger E (1998). *Structural Rasch Modeling: Handbook of the Usage of LPCM-WIN 1.0*. ProGAMMA, Groningen.

Fischer GH, Scheiblechner HH (1970). "Algorithmen und Programme für das probabilistische Testmodell von Rasch. [Algorithms and programs for Rasch's probabilistic test model.]." *Psychologische Beiträge*, **12**, 23–51.

Glas CAW (1992). "A Rasch Model with a Multivariate Distribution of Ability." In M Wilson (ed.), "Objective Measurement: Theory into Practice, Volume 1," pp. 236–258. Ablex, Norwood, NJ.

Glas CAW, Verhelst N (1989). "Extensions of the partial credit model." *Psychometrika*, **54**, 635–659.

Glas CAW, Verhelst N (1995a). "Testing the Rasch model." In G Fischer, I Molenaar (eds.), "Rasch Models: Foundations, Recent Developments, and Applications," pp. 69–96. Springer, New York.

Glas CAW, Verhelst N (1995b). "Tests of Fit for Polytomous Rasch Models." In G Fischer, I Molenaar (eds.), "Rasch Models: Foundations, Recent Developments, and Applications," pp. 325–352. Springer, New York.

Gustafsson J (1980). "Testing and obtaining fit of data to the Rasch model." *British Journal of Mathematical and Statistical Psychology*, **33**, 205–233.

Haberman SJ (1977). "Maximum likelihood estimates in exponential response models." *The Annals of Statistics*, **5**, 815–841.

Hoijtink H (1995). "Linear and Repeated Measures Models for the Person Parameter." In G Fischer, I Molenaar (eds.), "Rasch Models: Foundations, Recent Developments, and Applications," pp. 203–214. Springer, New York.

Hoijtink H, Boomsma A (1995). "On person parameter estimation in the dichotomous Rasch model." In G Fischer, I Molenaar (eds.), "Rasch Models: Foundations, Recent Developments, and Applications," pp. 53–68. Springer, New York.

Jannarone RJ (1986). "Conjunctive item response theory model kernels." *Psychometrika*, **51**, 357–373.

Kubinger KD (1989). "Aktueller Stand und kritische Würdigung der Probabilistischen Testtheorie. [Current status and critical appreciation of probabilistic test theory]." In K Kubinger (ed.), "Moderne Testtheorie: Ein Abriss samt neuesten Beiträgen," pp. 19–83. Beltz, Weinheim.

Kubinger KD (2005). "Psychological test calibration using the Rasch model: Some critical suggestions on traditional approaches." *International Journal of Testing*, **5**, 377–394.

Linacre JM (2004). "Estimation Methods for Rasch Measures." In EV Smith Jr, RM Smith (eds.), "Introduction to Rasch Measurement," pp. 25–48. JAM Press, Maple Grove, MN.

Liou M (1994). "More on the computation of higher-order derivatives of the elementary symmetric functions in the Rasch model." *Applied Psychological Measurement*, **18**, 53–62.

Mair P (2006). *Simulation Studies for Goodness-of-Fit Statistics in Item Response Theory.* Master's thesis, Department of Psychology, University of Vienna.

Mair P, Hatzinger R (2007a). "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 (2007b). "CML based estimation of extended Rasch models with the eRm package in R." *Psychology Science*, **49**, 26–43.

Masters GN (1982). "A Rasch model for partial credit scoring." *Psychometrika*, **47**, 149–174.

Mislevy RJ (1985). "Estimation of latent group effects." *Journal of the American Statistical Association*, **80**, 993–997.

Molenaar I (1995). "Estimation of Item Parameters." In G Fischer, I Molenaar (eds.), "Rasch models: Foundations, recent developements, and applications," pp. 39–51. Springer, New York.

Pfanzagl J (1994). "On Item Parameter Estimation in Certain Latent Trait Models." In G Fischer, D Laming (eds.), "Contributions to Mathematical Psychology, Psychometrics, and Methodology," pp. 249–263. Springer, New York.

Poinstingl H, Mair P, Hatzinger R (2007). *Manual zum Softwarepackage eRm: Anwendung des Rasch-Modells.* Pabst Science Publishers, Lengerich.

Ponocny I (2001). "Nonparametric goodness-of-fit tests for the Rasch model." *Psychometrika*, **66**, 437–460.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rasch G (1960). *Probabilistic Models for some Intelligence and Attainment Tests.* Danish Institute for Educational Research, Copenhagen.

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

Rasch G (1977). "On specific objectivity: An attempt at formalising the request for generality and validity of scientific statements." *Danish Yearbook of Philosophy*, **14**, 58–94.

Rizopoulos D (2006). "**ltm**: An R package for latent variable modeling and item response theory analyses." *Journal of Statistical Software*, **17**(5), 1–25. URL http://www.jstatsoft.org/v17/i05/.

Rost J (1999). "Was ist aus dem Rasch-Modell geworden? [What Happened with the Rasch Model?]." *Psychologische Rundschau*, **50**, 140–156.

Rost J (2000). "The Growing Family of Rasch Models." In A Boomsma, M van Duijn, T Snijders (eds.), "Essays on item response theory," pp. 25–42. Springer, New York.

Rost J, von Davier M (1995). "Polytomous Mixed Rasch Models." In G Fischer, I Molenaar (eds.), "Rasch Models: Foundations, Recent Developments, and Applications," pp. 371–382. Springer, New York.

Scheiblechner H (1972). "Das Lernen und Lösen komplexer Denkaufgaben. [The learning and solving of complex reasoning items.]." *Zeitschrift für Experimentelle und Angewandte Psychologie*, **3**, 456–506.

Suárez-Falcón JC, Glas CAW (2003). "Evaluation of global testing procedures for item fit to the Rasch model." *British Journal of Mathematical and Statistical Society*, **56**, 127–143.

Verhelst N, Glas CAW (1995). "The one parameter logistic test model." In G Fischer, I Molenaar (eds.), "Rasch Models: Foundations, Recent Developments, and Applications," pp. 215–238. Springer, New York.

Verhelst N, Hatzinger R, Mair P (2007). "The Rasch sampler." *Journal of Statistical Software*, **20(4)**, 1–14.

von Davier M (1998). *WINMIRA: A Windows program for mixed Rasch models.* IPN, Kiel.

Warm TA (1989). "Weighted likelihood estimation of ability in item response theory." *Psychometrika*, **54**, 427–450.

Wright BD, Masters GN (1982). *Rating scale analysis: Rasch measurement.* Mesa Press, Chicago.

Wright BD, Panchapakesan N (1969). "A procedure for sample-free item analysis." *Educational and Psychological measurement*, **29**, 23–48.