Journal of Mathematical Psychology 122 (2024) 102876

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Invariance of comparisons: Separation of item and person parameters beyond Rasch models

Gerhard Tutz

Ludwig-Maximilians-Universität München, Akademiestraße 1, 80799 München, Germany

ARTICLE INFO

Keywords:
Rasch model
Specific objectivity
Invariance of comparisons
Latent trait models

ABSTRACT

The Rasch model is the most prominent member of the class of latent trait models that are in common use. The main reason is that it can be considered as a measurement model that allows to separate person and item parameters, a feature that is referred to as invariance of comparisons or specific objectivity. It is shown that the property is not an exclusive trait of Rasch type models but is also found in alternative latent trait models. It is distinguished between separability in the theoretical measurement model and empirical separability with empirical separability meaning that parameters can be estimated without reference to the other group of parameters. A new type of pairwise estimator with this property is proposed that can be used also in alternative models. Separability is considered in binary models as well as in polytomous models.

1. Introduction

The binary Rasch model (Rasch, 1961) is one of the cornerstones of modern item response theory and has been extended to a whole family of models, see for example, Andrich (2016), von Davier (2016), von Davier and Carstensen (2007), Fischer and Molenaar (2012), Masters (1982), Rasch (1960). The main advantage of the Rasch model is that it allows measurement of latent traits to be independent of the measurement instrument, which is considered an advantage over other models as, for example, the normal-ogive model considered by Lord (1952).

With reference to stimuli rather than items Rasch (1961) formulated the requirements for comparing individuals and comparing stimuli by stating:

> The comparison between two stimuli should be independent of which particular individuals were instrumental for the comparison; and it should also be independent of which other stimuli within the considered class were or might also have been compared.
>
> Symmetrically, a comparison between two individuals should be independent of which particular stimuli within the class considered were instrumental for the comparison; and it should also be independent of which other individuals were also compared, on the same or on some other occasion (Rasch, 1961), p.331.

The strength of the Rasch model is that it allows for comparisons of item difficulties (person abilities) that are independent of the selection of persons (items). Rasch used the term *specific objectivity* for measurements that allow for the comparison of subjects or objects without reference to the instrument (within a given well-defined frame of reference), and considered it as general scientific principle to obtain generalizable measurements. The property has also be referred to as *parameter separability* or *invariance of comparisons*.

Although there are stringent mathematical arguments why the Rasch model, which uses logistic item characteristic curves, allows for the separation of parameters it is hardly convincing that the normal-ogive model, which in applications typically shows very similar results and fits, should not allow for invariant comparisons. It is demonstrated that parameter separability is possible in a much wider class of parametric models, although not by means of conditional likelihood as in the Rasch model. When investigating separability it is distinguished between separability as a property of the probabilistic measurement model and empirical separability as an estimation problem. While the former is easily seen to hold for more general models than the Rasch model separate estimation of parameters is less straightforward.

In Section 2 invariance in the Rasch model is briefly considered. In Section 3 invariance is investigated for the wider class of monotone homogeneity models and an estimator is derived that separates parameters. In Section 3.5 further properties of the estimator are considered and a motivation as a smoothing method is given. It is also investigated what independence of parameter estimates means, which is often misunderstood. Section 4 it devoted to polytomous models, in particular separability in the graded response model, which is not a member of the Rasch family, is considered.

E-mail address: tutz@stat.uni-muenchen.de.

https://doi.org/10.1016/j.jmp.2024.102876
Received 5 June 2024; Received in revised form 4 August 2024; Accepted 21 August 2024
Available online 28 August 2024

2. Invariance of comparisons in the binary Rasch model

Let $Y_{pi} \in \{0,1\}$ denote the response of person p on item i. The binary Rasch model can be given by

$$P(Y_{pi}=1|\theta_p,\delta_i) = \frac{\exp(\theta_p-\delta_i)}{1+\exp(\theta_p-\delta_i)}, i=1,\dots,I, p=1,\dots,P,$$

where θ_p is the ability of person p, and δ_i is the difficulty of item i. A key property concerning parameter separability can be derived by considering odds. The odds of a response $Y_{pi}=1$ are given by

$$\gamma_{pi} = \frac{P(Y_{pi}=1|\theta_p,\delta_i)}{P(Y_{pi}=0|\theta_p,\delta_i)} = \exp(\theta_p-\delta_i).$$

Then the odds ratio for two persons p_1, p_2 is given by

$$\frac{\gamma_{p_1 i}}{\gamma_{p_2 i}} = \frac{\exp(\theta_{p_1})}{\exp(\theta_{p_2})} = \frac{\theta^*_{p_1}}{\theta^*_{p_2}}, \quad (1)$$

where $\theta^*_p = \exp(\theta_p)$. That means comparison of persons can be carried out independently of the items involved, person parameters can be separated from item parameters. Rasch (1961) preferred the parameterization θ^*_p, which yields the ratio $\theta^*_{p_1}/\theta^*_{p_2}$ when comparing odds of two items. Here, we will mostly use the parameterization θ_p, which is linked to differences of parameters rather than the ratio since $\exp(\theta_{p_1})/\exp(\theta_{p_2}) = \exp(\theta_{p_1}-\theta_{p_2})$

Since the model is symmetric in the parameters one obtains a similar result for the comparison of two items i, j,

$$\frac{\gamma_{pi}}{\gamma_{pj}} = \frac{\exp(-\delta_i)}{\exp(-\delta_j)} = \frac{\delta^*_i}{\delta^*_j}, \quad (2)$$

where $\delta^*_i = \exp(-\delta_i)$, which does not depend on the person. Thus, comparisons of items can be carried out independent of the persons involved.

Eqs. (1) and (2) show that parameters can be separated by using odds, however odds that are not directly observed. In general, in probabilistic models inference tools are needed that approximate unobservable terms. In Rasch models a possible path to *empirical* separation of parameters is based on exploiting that the total scores are sufficient statistics. Let $S_p = Y_{p+} = \sum_{i=1}^I y_{pi}$ denote the number of items solved by person p, that is, the total score of person p. Then one can derive that

$$P(Y_{p1}=y_{p1},\dots,Y_{pI}=y_{pI}|S_p=s) = \frac{e^{-\sum_{i=1}^I y_{pi}\delta_i}}{\gamma_s(\boldsymbol{\delta})}, \quad (3)$$

where the functions $\gamma_s(\boldsymbol{\delta}) = \sum_{a_i\in\{0,1\},\sum_i a_i=s} e^{-\sum_i a_i\delta_i}$ are the so-called symmetric functions of order s, depending on $\boldsymbol{\delta}^T = (\delta_1,\dots,\delta_I)$ only. Since the conditional probability given in (3) is a function of item parameters only it can be used to estimate item parameters irrespective of the persons involved by maximizing the conditional likelihood $L_c(\boldsymbol{\delta}) = \prod_p P(Y_{p1}=y_{p1},\dots,Y_{pI}=y_{pI}|\sum_{i=1}^I Y_{pi}=y_{p+})$. For the asymptotic distribution and necessary and sufficient conditions for the existence of estimates see Andersen (1977), Pfanzagl (1994) and Fischer (1981). For the loss of information in conditional maximum likelihood, see Eggen (2000), Eggen and Verhelst (2006).

3. Comparisons in binary latent trait models

Let us consider the more general class of monotone homogeneity models (MH models), which comprises models of the form

$$\pi_{pi}(\theta_p,\delta_i) = P(Y_{pi}=1|\theta_p,\delta_i) = F(\theta_p-\delta_i), \quad (4)$$

where $F(.)$ is a strictly monotone distribution function also called response function. The models contain one parameter per person, θ_p, and one parameter per item, δ_i. They are homogeneous since the item characteristic functions all have the same form. They are monotone since the probability of success increases monotonically with increasing person parameter. The binary Rasch model is contained as the special case where $F(.)$ is the logistic function $F(\eta) = \exp(\eta)/(1+\exp(\eta))$.

For a MH model one obtains a form of separability of parameters when considering two items i, j. From (4) one can immediately derive

$$F^{-1}(\pi_{pi}(\theta_p,\delta_i)) - F^{-1}(\pi_{pj}(\theta_p,\delta_j)) = \delta_j - \delta_i, \quad (5)$$

which is independent of θ_p. Thus a transformation of the involved probabilities yields a function that does not contain person parameters, and which can be used to compare item parameters. Since the function $F^{-1}(.)$ is the quantile function differences of item parameters reflect the differences of quantiles of the distribution function $F(.)$, which do not depend on the person parameters. In a similar way persons can be compared without reference to item parameters.

3.1. Invariance in terms of the model

The separability property (5) can be formulated in a more general way. For a uni-dimensional latent trait model with response probabilities $\pi_{pi} = P(Y_{pi}=1)$ for person p and item i *separability of item parameters from person parameters* (*invariance of comparison of item parameters*, *specific objectivity for the comparison of items*) holds if a parameterization (θ_p,δ_i) and a transformation function C_{it} called item comparator exist such that $C_{\text{it}}(\pi_{pi},\pi_{pj})$ is equal to the difference of item parameters for all items i, j. For the parameterization that is assumed to exist one has

$$C_{\text{it}}(\pi_{pi}(\theta_p,\delta_i),\pi_{pj}(\theta_p,\delta_j)) = \delta_i - \delta_j. \quad (6)$$

The definition of separability only assumes that a parameterization exists in order to make the definition independent of the specific parameterization that is used in the formulation of the model. For example, in the Rasch model the parameterization θ_p, δ_i provides such a parameterization but the parameterization θ^*_p, δ^*_i does not. Fischer (1995) used a similar equation when investigating invariance in the Rasch model. He assumed for a fixed parameterization $C(\rho_{pi}(\theta_p,\delta_i),\rho_{pj}(\theta_p,\delta_i)) = V(\delta_{pi},\delta_{pj})$, where $V(.)$ is an additional function and ρ_{pi} a "reaction parameter". The definition used here considers the response probability as reaction parameter, which seems quite natural since it definitely determines the response. A general function $V(.)$ seems not necessary if one does not consider a specific parameterization but assumes the existence of a parameterization. In addition, an unspecified function $V(.)$ could be very difficult while differences are easy to handle and suffice for the models considered here. Although a similar definition can be given concerning the invariance of comparison of person parameters in the following we focus on the invariance of comparison of item parameters.

It follows from (5) that in any monotone homogeneity model the comparison of item parameters is invariant and does not depend on the person parameter. The Rasch model is just a special case but by far not the only model. It is to be emphasized that the invariance property considered here is a property of the probabilistic measurement model but is not directly observable. However, it can be considered as the essential property that is needed to also obtain empirical invariance, that is, invariance referring to estimation.

3.2. Empirical invariance

In the Rasch model empirical separation of parameters is usually obtained by exploiting that the total scores are sufficient statistics. This works only since the item response curves are logistic, for any other response function total scores are not sufficient statistics. The deeper reason is that binary responses are members of the exponential family and logits are linked to the natural parameter in exponential families, see, for example, McCullagh and Nelder (1989).

Conditional maximum likelihood estimation with the conditioning on sufficient statistics is not an option if the response function is not the logistic function. Although in monotone homogeneity models the function $F^{-1}(\pi_{pi}(\theta_p,\delta_i)) - F^{-1}(\pi_{pi}(\theta_p,\delta_j))$ does not depend on person parameters it is not obvious how this property can be exploited in estimation since replacing $\pi_{pi}(\theta_p, \delta_i)$ by observations, that is, by 0 or 1, yields $-\infty$ and ∞ when building $F^{-1}(Y_{pi})$.

In the following an estimation method is proposed that uses pseudo observations. Let us consider the pseudo observations

$$Y^*_{pi} = Y_{pi}(1-2\gamma) + \gamma = \begin{cases} \gamma & Y_{pi} = 0 \\ 1-\gamma & Y_{pi} = 1, \end{cases}$$

where $0 < \gamma < 0.5$ is a fixed value. For small values of γ the pseudo observations approximate the original values, and in the extreme case are identical, $\lim_{\gamma\to 0} Y^*_{pi} = Y_{pi}$. Replacing probabilities by pseudo observations yields

$$F^{-1}(Y_{pi}(1-2\gamma)+\gamma) = \begin{cases} F^{-1}(\gamma) =: \gamma_0 & Y_{pi} = 0 \\ F^{-1}(1-\gamma) =: \gamma_1 & Y_{pi} = 1. \end{cases}$$

Although $\lim_{\gamma\to 0} F^{-1}(\gamma) = -\infty$, $\lim_{\gamma\to 0} F^{-1}(1-\gamma) = \infty$, for values $\gamma > 0$ one obtains finite values for γ_0 and γ_1. The empirical analogue to (5) when replacing probabilities by pseudo observations is

$$\hat\delta_j(p) - \hat\delta_i(p) = F^{-1}(Y_{pi}(1-2\gamma)+\gamma) - F^{-1}(Y_{pj}(1-2\gamma)+\gamma) = \\ = \begin{cases} 0 & Y_{pi} = Y_{pj} \\ \gamma_1 - \gamma_0 & Y_{pi} = 1, Y_{pj} = 0 \\ \gamma_0 - \gamma_1 & Y_{pi} = 0, Y_{pj} = 1 \end{cases}$$

It is an empirical approximation of $F^{-1}(\pi_{pi}(\theta_p, \delta_i)) - F^{-1}(\pi_{pi}(\theta_p, \delta_j))$, which is the theoretical difference for person p. From this representation an estimate of the difference $\delta_{ji} := \delta_j - \delta_i$ is derived by summing over the contributions of all persons,

$$\hat\delta_{ji} = \hat\delta_j - \hat\delta_i = \{(\gamma_1-\gamma_0)n^{(i,j)}_{10} + (\gamma_0-\gamma_1)n^{(i,j)}_{01}\}/n(Y_{pi} \neq Y_{pj}) \\ = \gamma_{10}(n^{(i,j)}_{10} - n^{(i,j)}_{01})/n(Y_{pi} \neq Y_{pj})$$

where $\gamma_{10} = \gamma_1 - \gamma_0$, $n^{(i,j)}_{10}$ is the number of persons with response $Y_{pi} = 1, Y_{pj} = 0$, $n^{(i,j)}_{01}$ is the number of persons with response $Y_{pi} = 0, Y_{pj} = 1$, and $n(Y_{pi} \neq Y_{pj}) = n^{(i,j)}_{10} + n^{(i,j)}_{01}$. The values correspond to the entries of the contingency table

	0	1	
0	$n^{(i,j)}_{00}$	$n^{(i,j)}_{01}$	$n^{(i,j)}_{0+}$
1	$n^{(i,j)}_{10}$	$n^{(i,j)}_{11}$	$n^{(i,j)}_{1+}$
	$n^{(i,j)}_{+0}$	$n^{(i,j)}_{+1}$	p

An alternative representation of the estimator which uses only the sums of persons that solved specific items is given by

$$\hat\delta_{ji} = \gamma_{10}\sum_{p=1}^{P}(Y_{pi} - Y_{pj})/n(Y_{pi} \neq Y_{pj}) = \gamma_{10}(Y_{+i} - Y_{+j})/n(Y_{pi} \neq Y_{pj}),$$

where $Y_{+i} = \sum_p Y_{pi}$ are the number of persons that solved item i. If $F(.)$ is a symmetric function, for example the normal distribution, one has $\gamma_0 = -\gamma_1$, and the estimator simplifies to $\hat\delta_j - \hat\delta_i = 2\gamma_1(Y_{+i} - Y_{+j})/n(Y_{pi} \neq Y_{pj})$.

Since parameters are only defined up to an additive constant one can set $\delta_1 = 0$. That yields a simple estimator of δ_i by using $\hat\delta^{(s)}_i = \gamma_{10}(Y_{+i} - Y_{+1})/n(Y_{pi} \neq Y_{p1})$. A disadvantage is that it uses only the item pairs $(1,2), (1,3), \dots, (1, I)$. An estimator that uses all the differences and is called the *pairwise separation estimator* is given by

$$\hat\delta_i = \sum_{j=1}^{I}(\hat\delta_{ij} - \hat\delta_{1j})/I,$$

where $\hat\delta_{11} := 0$. It is an average across all estimators that use the estimated differences to a fixed item with the constraint $\delta_1 = 0$. This can be seen by considering a fixed anchor item j. Then $\hat\delta_{ij}$ is an estimator of $\delta_i - \delta_j$. It aims at estimating the differences between item i and item j. If one sets $\hat\delta_i = \hat\delta_{ij}$ for all i implicitly the parameter $\hat\delta_j$ is set to zero. In order to set $\hat\delta_1 = 0$ one has to subtract $\hat\delta_{1j}$.

The estimate contains the scaling factor γ_{10} but ratios $\hat\delta_i/\hat\delta_j$ do not depend on the scaling, which means that the estimates change by a factor if different scaling factors are used. The scaling factor is linked to the unit of measurement that is used. In the general measurement model $P(Y_{pi} = 1|\theta_p, \delta_i, \alpha) = F(\alpha(\theta_p - \delta_i))$ parameters are identifiable if one chooses a specific value for α, for example $\alpha = 1$, and fixes the location, for example by choosing $\delta_1 = 0$ or $\sum_i \delta_i = 0$. The former choice corresponds to the (arbitrary) choice of a measurement unit, the latter to the choice of the origin of the measurement, which is on an interval scale. More precisely, the scales for person and item measurement are interval scales with the same unit of measurement and with possibly different zero points, see also Fischer (1995). The choice of the scaling factor γ_{10} corresponds to the choice of the measurement unit but not on the scale of the parameters θ_p, δ_i. In order to link it to the choice of the unit on the (θ_p, δ_i)-scale we use a data-based approach. It selects the parameter γ_{10} that yields the best fit of the data when the measurement unit has been fixed. We use $\alpha = 1$, which is a common choice. The advantage is that the resulting estimates can be compared to estimates obtained by alternative methods that use $\alpha = 1$. The γ_{10} obtained in this way is not an estimator of the measurement unit, which can be chosen arbitrarily, it is just an adaptation of the estimator to common choice of units.

The data-based approach to selecting the scale parameter is the following. For fixed γ_{10} and the resulting item parameter estimates $\hat\delta_i(\gamma_{10})$ person parameters can be estimated by maximizing the log-likelihood function

$$l_{\gamma_{10}}(\theta_p) = \sum_{i=1}^{I} Y_{pi}\log(\pi_{pi}(\theta_p, \hat\delta_i(\gamma_{10}))) + (1-Y_{pi})\log(1-\pi_{pi}(\theta_p, \hat\delta_i(\gamma_{10}))),$$

yielding $\hat\theta_p(\gamma_{10})$. Maximization is simple since it is a one-dimensional maximization problem. In a second step the goodness-of-fit of the resulting estimate $\pi_{pi}(\hat\theta_p(\gamma_{10}), \hat\delta_i(\gamma_{10}))$ is investigated by considering loss functions that reflect the differences between observations and estimated probabilities, $L(Y_{pi}, \pi_{pi}(\hat\theta_p(\gamma_{10}, \hat\delta_i(\gamma_{10}))))$. Candidates are the quadratic loss $L_Q(Y_{pi}), \hat\pi_{pi}) = 2(Y_{pi} - \hat\pi_{pi})^2$ and the Kullback–Leibler loss $L_{KL}(Y_{pi}, \hat\pi_{pi}) = -(Y_{pi}\log(\hat\pi_{pi}) - (1-Y_{pi})\log(1-\hat\pi_{pi}))$, where minimization of the latter corresponds to maximum likelihood estimation. The final estimator is obtained by using that scaling parameter that minimizes

$$\text{Loss}(\gamma_{10}) = \sum_{i=1}^{I}\sum_{p=1}^{P} L(Y_{pi}, \pi_{pi}(\hat\theta_p(\gamma_{10}, \hat\delta_i(\gamma_{10}))))$$

with respect to γ_{10}.

For illustration we consider the estimates obtained for the normal-ogive model, in which $F(.)$ is the normal distribution function. The item parameters for $I = 6$ items were $0, -1.5, -1, 0.5, 1.2, 1.5$. Person parameters were drawn from a standardized normal distribution. Fig. 1 shows the box plots of estimates (200 repetitions) and the kernel density estimate of item parameter 5 for $P = 100$ (first row) and $P = 300$ (second row), the latter is obtained by using the function *density* from R. It is seen that the estimates approximate the true values rather well. It demonstrates that item parameters can be estimated separately also for the normal-ogive model, for which no sufficient statistics exist. Fig. 2 shows the quadratic and the Kullback–Leibler loss functions used to select the scale parameter for the last of the simulated data sets. It is seen that the both loss functions select very similar scale parameters. Although the values of the functions differ the minima for both functions are close to 0.18. In the simulations the Kullback–Leibler loss has been used.

3.3. Separation of parameters in the binary Rasch model

Let us again consider the binary Rasch model. As already mentioned in Section 2 it is special among monotone homogeneity models since it allows for estimation of item parameters by using sufficient statistics for person parameters. The existence of sufficient statistics can be used

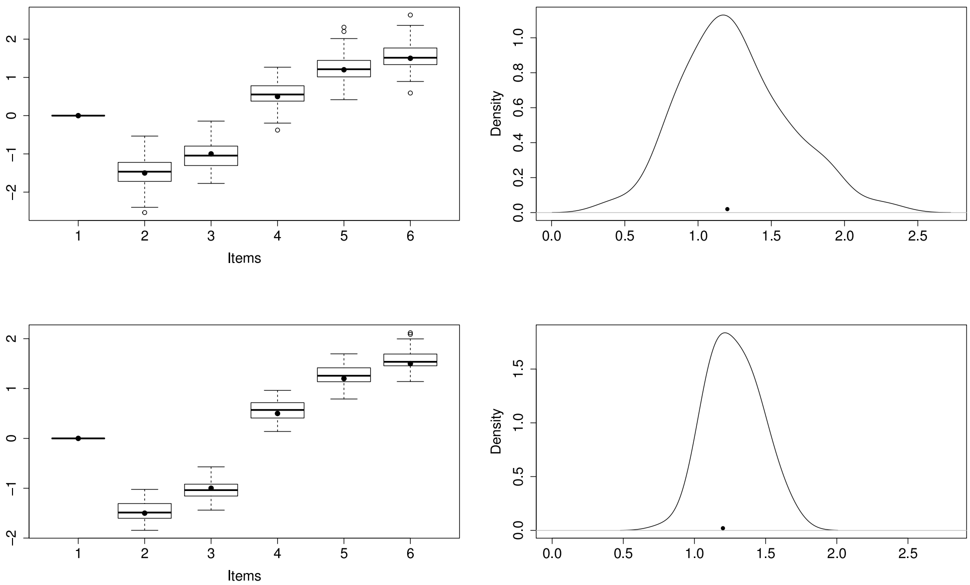

Fig. 1. Left: Box plots for estimates for six items with dots indicating the true values assuming a normal-ogive model; right: density of estimates for item 5. First row: (P = 100), second row: $P = 300$.

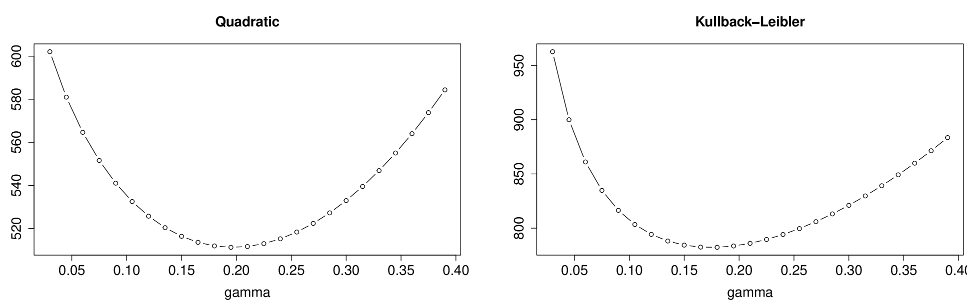

Fig. 2. Loss functions for the estimation of the scaling factor γ_{10} for one simulated data set.

to maximize the conditional log-likelihood yielding the *conditional estimates*. It also allows to construct further estimates of item parameters that do not depend on the selected persons.

An alternative estimation method that is strongly linked to the separability seen in Eq. (2) is conditional likelihood estimation for pairs of items. Under the condition $Y_{pi} + Y_{pj} = 1$ solving the conditional likelihood for observations on items i, j only yields the estimator

$$\frac{e^{-\hat{\delta}_i}}{e^{-\hat{\delta}_j}} = \frac{N_i}{N_j},$$

where N_i denotes the number of persons that solved item i and exactly one of the two items i, j. It can be seen as an empirical version of Eq. (2) and is referred to as *pairwise conditional estimation*. Item parameters are estimated without reference to the persons involved. A corresponding property holds for the comparison of persons but is not explicitly given. For more on pairwise estimates, see Zwinderman (1995), von Davier (2016).

Both methods conditional estimation and pairwise conditional estimation exploit the existence of sufficient statistics given the total scores, in the latter case under the condition that only two items are considered. In Fig. 3 the estimators are compared to the pairwise separation estimator with the same item parameters as in Fig. 1 and the assumption that the Rasch model holds. It is seen that the conditional estimates have smaller variance than the other two estimators. The pairwise conditional estimator and the pairwise separation estimator show almost the same performance. This is also supported by computing the average absolute deviations of estimates from the true values, which was 0.231 for the conditional estimator, 0.268 for the pairwise separation estimator, and 0.252 for the pairwise conditional estimator. Similar pictures are obtained if the number of persons is larger (see appendix).

3.4. Non-symmetric response models

The strength of the pairwise separation estimator is that it provides estimates also when the response function is not the logistic function. That includes response functions that are not symmetric. Non-symmetric skewed response functions are not so often used since item parameter estimates are not so easily obtained as in the Rasch model although there is no compelling reason why item response functions should be symmetric. Advantages of skewed response functions have been outlined by Samejima (2000), Bazán et al. (2006) and, more recently, by Bolt and Liao (2022).

As examples of skewed response functions we consider the maximum value distribution function (Gumbel distribution function) $F(\eta) = \exp(-\exp(-\eta))$ and the minimum value distribution (Gompertz distribution function) $F(\eta) = 1 - \exp(-\exp(\eta))$. Figs. 4 and 5 show the boxplots of estimates when using the pairwise separation estimator (parameters are the same as in Fig. 1). Fig. 4 shows the estimates for the maximum value model ($P = 100$) and the density of estimates of item 4 (last simulation). Fig. 5 shows the estimates for the minimum value model with $P = 100$ on the left and P = 50 on the right hand side. It is seen that the estimator approximates the true parameters rather well.

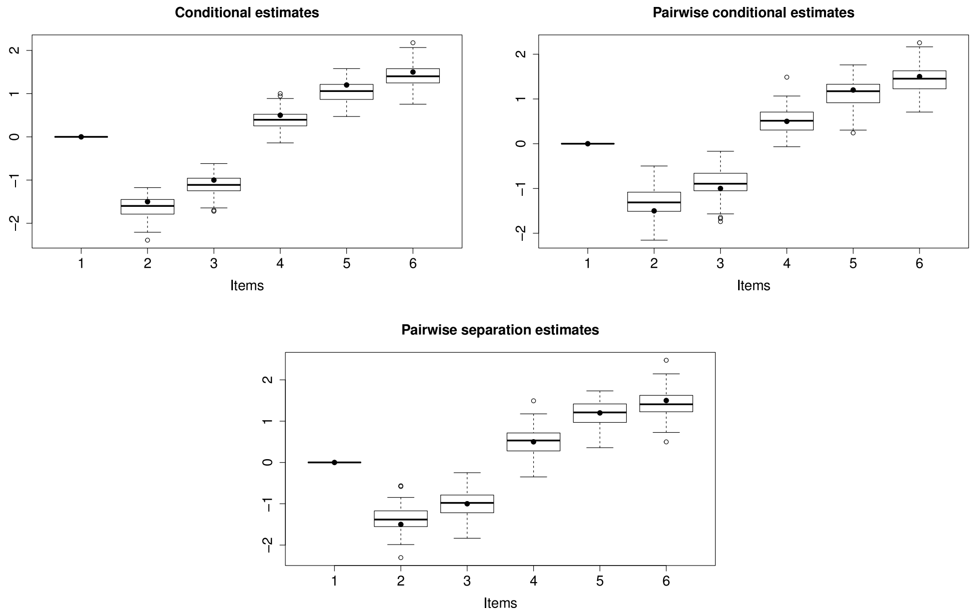

Fig. 3. Box plots of parameter estimates for conditional, pairwise conditional and separation estimators in the Rasch model with $P = 100$.

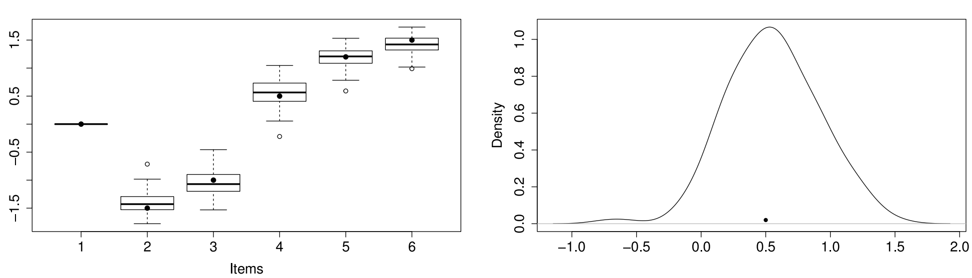

Fig. 4. Left: Box plots for estimates for six items with dots indicating the true values assuming a maximum value model; right: density of estimates for item 4 $P = 100$.

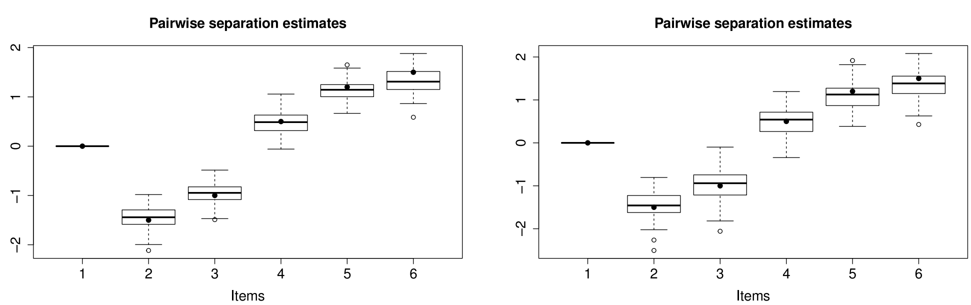

Fig. 5. Box plots for estimates for six items with dots indicating the true values assuming a minimum value model $P = 100$ (left), P = 50 (right).

3.5. Further issues

Independence of estimates

In Rasch models item parameters are often considered to not depend on persons when maximizing the conditional likelihood given person sums since the conditional likelihood does not contain person parameters. However, this is a misinterpretation. Although the formulae do not contain the person parameters it does not mean that the choice of persons has no impact on the estimates, in particular on their accuracy. It matters, for example, if persons are drawn from a population covering a wide range of abilities or from a population that contains only persons with high (or low) abilities (to be demonstrated in the following).

Also the separation estimator is constructed by using equations that do not contain the person parameters. Nevertheless, at some point person parameter estimates are used, however only to obtain the scaling parameter since only proportions $\hat{\delta}_i/\hat{\delta}_j$ are scale free. To obtain the scaling parameter γ_{10} it is not necessary to use all persons, almost the same estimates are found if only a subset of persons is used to find the scaling parameter. But, as in conditional estimators the selection of the sub population has an impact on the *accuracy* of estimates.

Table 1 illustrates the effect of the choice of the person population. Person parameters have been drawn from several distributions, the standard normal distribution ($\theta_p \sim \mathrm{N}(0,1)$), which covers a wide range of person parameters, the χ^2-distribution with one degree of freedom ($\theta_p = \tilde{\theta}_p^2$ with $\tilde{\theta}_p \sim \mathrm{N}(0,1)$), which means that all person parameters are positive, the non-central χ^2-distribution with non-centrality parameter 1 ($\theta_p = \tilde{\theta}_p^2$ with $\tilde{\theta}_p \sim \mathrm{N}(1,1)$), and the non-central χ^2-distribution with non-centrality parameter 1.5 ($\theta_p = \tilde{\theta}_p^2$ with $\tilde{\theta}_p \sim \mathrm{N}(1.5,1)$). In particular the latter scenarios mean that only rather large person parameters are in the sample. The item parameters are the same as in Fig. 1. As

Table 1
Deviations of estimates of item parameters from true values for several estimation methods.

		pairwise separation	conditional	pairwise conditional
P = 80	standard normal	0.298	0.257	0.290
	chi-squared	0.300	0.258	0.289
	nonc chi-squared, $N(1,1)^2$	0.315	0.303	0.346
	nonc chi-squared, $N(1.5,1)^2$	0.368	0.369	0.452
P = 100	standard normal	0.268	0.231	0.252
	chi-squared	0.276	0.240	0.265
	nonc chi-squared, $N(1,1)^2$	0.277	0.253	0.291
	nonc chi-squared, $N(1.5,1)^2$	0.316	0.311	0.365
P = 200	standard normal,	0.196	0.154	0.170
	chi-squared	0.201	0.166	0.180
	nonc chi-squared, $N(1,1)^2$	0.203	0.182	0.200
	nonc chi-squared, $N(1.5,1)^2$	0.224	0.213	0.231

measure of accuracy we use the absolute deviation $|\delta_i - \hat{\delta}_i|$ averaged across all items. Results are given for several sample sizes. It is seen that all estimates perform poorer when the mean of the distribution of the person parameters gets larger. This includes the estimators that use the conditional likelihood approach. In particular the pairwise conditional estimator suffers strongly if the available persons are from distributions with larger means. The separation estimator is relatively stable and performs well also in these cases although for Rasch models as simulated here the conditional estimator shows the best performance. The results demonstrate that also conditional estimators of item parameters cannot be considered as being independent of the choice of persons although person parameters are not involved in the estimation equations.

Smoothing and the separation estimator

The separation estimator can be seen as an estimator based on smoothing techniques. Methods as smooth kernel estimators for densities and distribution functions have a long tradition, see, for example, Wand and Jones (1995), Simonoff (1996).

Let us first consider the general case of a random variable Y. For a fixed observation y_0 of the random variable the contribution to the density estimate is the kernel function $g((y-y_0)/h)$, where $g(.)$ is a fixed density function and h is the window width. The contribution to the estimated distribution function is $G((y-y_0)/h)$, where $G(.)$ is the distribution function corresponding to $g(.)$. Alternatively one can consider estimates of the survivor function $S(y) = P(Y \geq y)$, which are used in the representation of the binary response models considered here.

An un-smoothed step function estimate of $P(Y \geq y)$ when y_0 has been observed is the function $S(y) = 1 - H(y - y_0)$, where $H(.)$ is the Heaviside function $H(x) = 1$ if $x \geq 0$ and $H(x) = 0$ otherwise. To obtain a smooth estimator one can replace the Heaviside function by a smooth approximation. Any distribution function that is very steep at zero could be used. The choice corresponds to the choice of the kernel in density estimation. For simplicity we choose the function $F_\alpha(\eta) = 1 - F(-\alpha\eta)$, where $F(.)$ is the response function of the item response model (centred at zero such that $F(0) = .5$) and $\alpha > 0$ is a large constant. Since the function is centred at zero for increasing α the function $F_\alpha(.)$ becomes the Heaviside function for all values unequal zero. Smooth estimates result if

$P(Y \geq y)$ is estimated by $F(-\alpha(y-y_0))$ and

$F^{-1}(P(Y \geq y))$ is estimated by $F^{-1}(F(-\alpha(y-y_0))) = -\alpha(y-y_0)$.

When applied to the responses Y_{pi} one obtains

$$F^{-1}(P(Y_{pi} \geq y)) - F^{-1}(P(Y_{pi} \geq y)) =$$
$$= F^{-1}(F(-\alpha(y-y_{pi}))) - F^{-1}(F(-\alpha(y-y_{pj}))) =$$

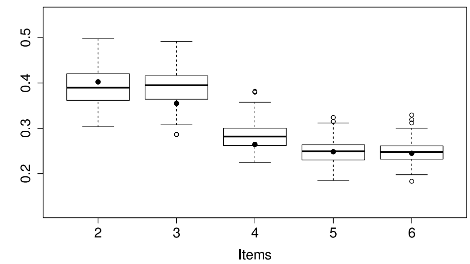

Fig. 6. Box plots of bootstrap estimators of standard errors, dots denote the "true" standard errors (estimated from 300 drawings).

$$= \alpha(y_{pi} - y_{pj}) = \begin{cases} 0 & y_{pi} = y_{pj} \\ \alpha & y_{pi} = 1, y_{pj} = 0 \\ -\alpha & y_{pi} = 0, y_{pj} = 1 \end{cases}$$

This is equivalent to the estimate $\hat{\delta}_j - \hat{\delta}_i$ for one person with the smoothing parameter α corresponding to γ_{10}.

Standard errors

A way to obtain estimates of standard errors is bootstrapping (Davison & Hinkley, 1997; Efron & Tibshirani, 1994). For a given data set consisting of $P \times I$ observations, the estimation procedure is carried out repeatedly for data that are obtained by drawing persons from the data set with replacement. The variation of the estimates is used to compute realistic standard errors of parameters.

For illustration we consider the approximation of the true standard errors in the scenario given in Fig. 1. Estimates of the true standard errors are computed from 300 repetitions. To obtain bootstrap standard errors, we used 50 repetitions of bootstrap estimates. Each bootstrap estimate is based on $n_{\text{boost}} = 200$ drawings with replacement. Fig. 6 shows the box plots of bootstrap repetitions. It is seen that bootstrap standard errors approximate the true errors rather well.

4. Polytomous models

Separability of parameters can also be investigated in models with more than two response categories. Let polytomously scored items have values $Y_{pi} \in \{0,1,\dots,k\}$, $p = 1,\dots,P$, $i = 1,\dots,I$, with $\{0,1,\dots,k\}$ representing the coding of categories.

Several polytomous latent trait models have been considered in the literature, in particular ordinal models as the graded response model (Samejima, 1995, 2016) and the partial credit model (Masters, 1982; Masters & Wright, 1984) are widely used. Separability of parameters is especially interesting for the graded response model since, in contrast to the partial credit model, no sufficient statistics exist that could be used to construct conditional log-likelihood estimators.

4.1. Invariance in terms of the model

Polytomous models specify the response probabilities

$$\pi_{pir}(\theta_p, \boldsymbol{\delta}_i) = P(Y_{pi} = r|\theta_p, \boldsymbol{\delta}_i), \quad r = 0,\dots,k,$$

as functions of person parameters θ_p and a vector of item parameters $\boldsymbol{\delta}_i$ (for simplicity the number of response categories is the same for all items). The latter usually has the form $\boldsymbol{\delta}_i^T = (\delta_{i1},\dots,\delta_{ik})$. The total vector of response probabilities for person p and item i is denoted by $\boldsymbol{\pi}_{pi}(\theta_p, \boldsymbol{\delta}_i)^T = (\pi_{pi0}(\theta_p, \boldsymbol{\delta}_i),\dots,\pi_{pik}(\theta_p, \boldsymbol{\delta}_i))$. A widely used polytomous model is the graded response model (Samejima, 1995, 2016), which specifies the *cumulative* probabilities

$$\pi_{pi}(r) = P(Y_{pi} \geq r|\theta_p, \boldsymbol{\delta}_i) = F(\theta_p - \delta_{ir}), \quad r = 1,\dots,k,$$

yielding the probabilities $\pi_{pir}(\theta_p, \boldsymbol{\delta}_i) = \pi_{pi}(r) - \pi_{pi}(r+1)$.

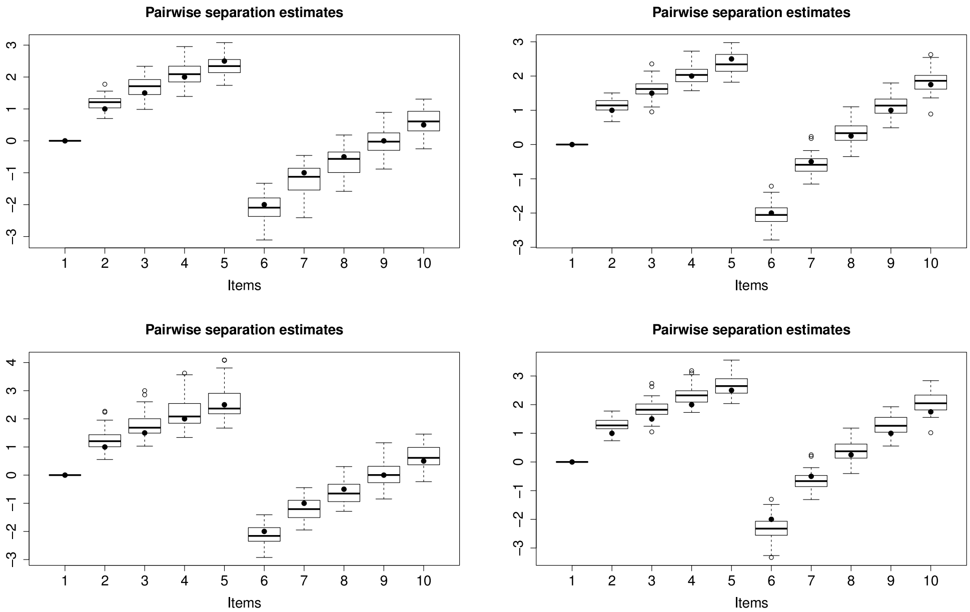

Fig. 7. Cumulative logit model, $P = 100, I = 4, k = 5$, only first two items, left: $(\delta_{11}, \dots, \delta_{15}) = ((0, 1, 1.5, 2, 2.5))$, other items are shifted versions, right: second item modified, first row shows the averaged estimator, second row shows the estimator with anchor item 1.

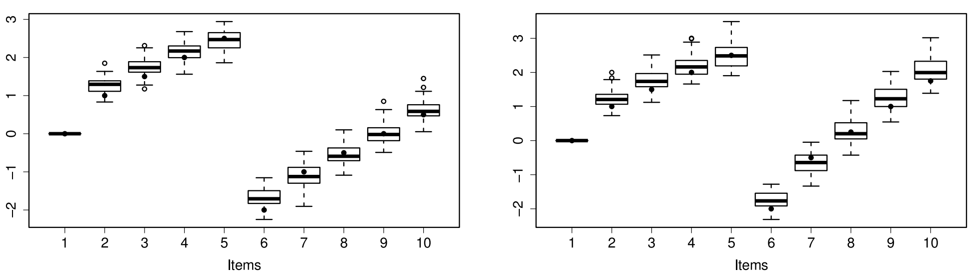

Fig. 8. Cumulative minimum value model, $P = 100, I = 4, k = 5$, only first two items, left: $(\delta_{11}, \dots, \delta_{15}) = ((0.01.01.52.02.5))$, other items are shifted versions, right:second item modified.

When considering invariance it is useful to distinguish between person and item parameters since the former are uni-dimensional while the latter are multi-dimensional.

For a polytomous latent trait model *invariance of comparison of person parameters* (*specific objectivity for the comparison of persons*) holds if a parameterization and at least one transformation function called comparator C_{pers} exist such that for all persons p_1, p_2 and all items

$$C_{\text{pers}}(\boldsymbol{\pi}_{p_1 i}(\theta_{p_1}, \boldsymbol{\delta}_i), \boldsymbol{\pi}_{p_2 i}(\theta_{p_2}, \boldsymbol{\delta}_i)) = \theta_{p_1} - \theta_{p_2}.$$

If the cumulative model is the data generating model for each $r \geq 1$

$$F^{-1}(\pi_{p_1 i}(r)) - F^{-1}(\pi_{p_2 i}(r)) = \theta_{p_1} - \theta_{p_2} \tag{7}$$

holds. Since the cumulative probabilities are simple functions of the vector of probabilities it is obvious that several comparator functions C_{pers} exist. Eq. (7) shows that k functions can be used to compare person parameters.

For a polytomous latent trait model *invariance of comparison of item parameters* (*specific objectivity for the comparison of items*) holds if a parameterization and a set of transformation functions (comparators) $C_{\text{it},i}$ exist such that for two items i, j the differences of item parameters are *uniquely* determined in the form

$$C_{\text{it},i}(\boldsymbol{\pi}_{pi}(\theta_p, \boldsymbol{\delta}_i), \boldsymbol{\pi}_{pj}(\theta_p, \boldsymbol{\delta}_j)) = \delta_{iq} - \delta_{jr},$$

where $i \in \text{Comp}$. In contrast to the comparison of person parameters it is not sufficient to postulate the existence of just one comparator function. Since each item has several parameters more than one comparator function is needed to determine all differences uniquely. Typically different sets of comparator functions can be found that determine all differences. One could also consider minimal sets of comparator functions but they would only be of theoretical interest.

If the cumulative model is the data generating model for any $r, q > 1$

$$F^{-1}(\pi_{pi}(r)) - F^{-1}(\pi_{pj}(q)) = \delta_{jq} - \delta_{ir} \tag{8}$$

holds, which defines the comparator functions. With the constraint $\delta_{11} = 0$ the set of functions

$$F^{-1}(\pi_{p1}(r)) - F^{-1}(\pi_{p1}(1)) = \delta_{11} - \delta_{1r}, \quad r > 1$$
$$F^{-1}(\pi_{p1}(1)) - F^{-1}(\pi_{pi}(q)) = \delta_{iq} - \delta_{11}, \quad i > 1, q = 1, \dots, k, \tag{9}$$

forms a sufficiently large (and minimal) set of comparator functions such that all differences and therefore item parameters are uniquely determined.

4.2. Empirical invariance

For the investigation of empirical invariance it is helpful to consider split variables, which are defined by

$$Y_{pi}(r) = \begin{cases} 1 & Y_{pi} \geq r \\ 0 & Y_{pi} < r. \end{cases}$$

The split variables are binary variables that partition the categories into the subsets $\{0, \dots, r-1\}$ and $\{r, \dots, k\}$. With response categories $\{0, 1, \dots, k\}$ one has k split variables $Y_{pi}(1), \dots, Y_{pi}(k)$. Split variables are the empirical analogues to the cumulative probabilities, $Y_{pi}(r)$ is

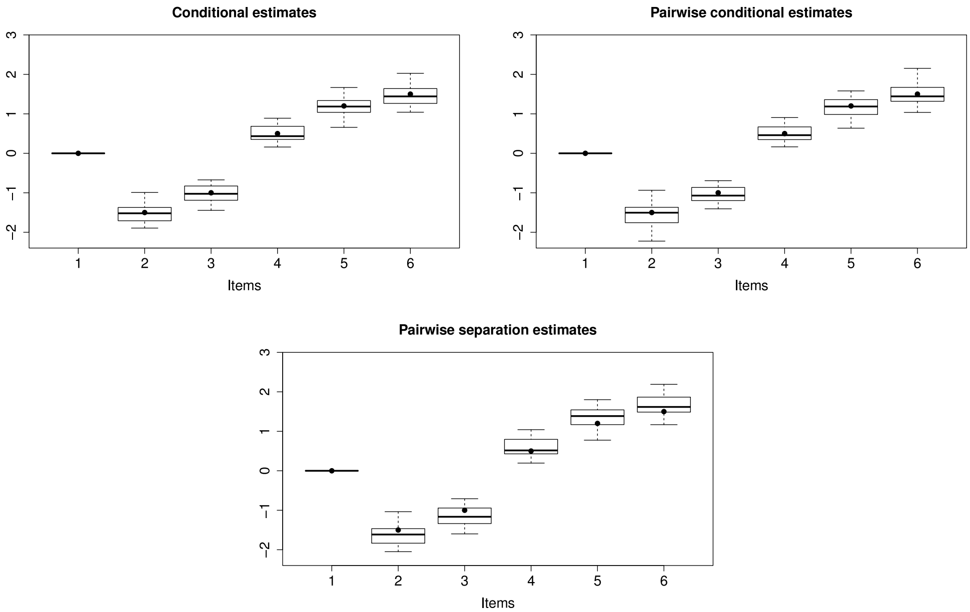

Fig. 9. Box plots of parameter estimates for conditional, pairwise conditional and separation estimators in the Rasch model with $P = 300$.

an observation of the binary model $P(Y_{pi}(r) = 1) = P(Y_{pi} \geq r)$. If the cumulative model holds the probability is given by $P(Y_{pi} \geq r) = F(\theta_p - \delta_{ir})$, which yields the binary models

$$P(Y_{pi}(r) = 1) = F(\theta_p - \delta_{ir}), r = 1, \dots, k. \quad (10)$$

The link between split variables and the cumulative model is even stronger since the cumulative model holds if models (10) hold simultaneously for $r = 1, \dots, k$, see also Tutz (2020).

For the cumulative model the theoretical invariance of items is formulated in Eq. (8). The empirical analogue is obtained by replacing the cumulative probabilities by their observational counterparts. However, as for binary models instead of using the observations themselves one uses pseudo observations, which are slightly distorted original observations. Thus, instead of $Y_{pi}(r)$ the pseudo observation $Y_{pi}(r)^* = Y_{pi}(r)(1-2\gamma) + \gamma$ is used yielding

$$F^{-1}(Y_{pi}(r)^*) - F^{-1}(Y_{pj}(q)^*) = \begin{cases} 0 & Y_{pi}(r) = Y_{pj}(q) \\ \gamma_1 - \gamma_0 & Y_{pi}(r) = 1, Y_{pj}(q) = 0 \\ \gamma_0 - \gamma_1 & Y_{pi}(r) = 0, Y_{pj}(q) = 1, \end{cases}$$

where $\gamma_0 = F^{-1}(\gamma), \gamma_1 = F^{-1}(1-\gamma)$. It is an estimator of $\delta_{jq} - \delta_{ir}$. From this one can construct the estimator

$$\begin{aligned} &\hat{\delta}_{jq} - \hat{\delta}_{ir} = \\ &= \frac{(\gamma_1 - \gamma_0)n(Y_{pi}(r) = 1, Y_{pj}(q) = 0) + (\gamma_0 - \gamma_1)n(Y_{pi}(r) = 0, Y_{pj}(q) = 1)}{n(Y_{pi}(r) \neq Y_{pj}(q))} \\ &= \frac{(\gamma_1 - \gamma_0)n(Y_{pi}(r) = 1, Y_{pj}(q) = 0) - n(Y_{pi}(r) = 0, Y_{pj}(q) = 1)}{n(Y_{pi}(r) \neq Y_{pj}(q)),} \end{aligned} \quad (11)$$

where $n(Y_{pi}(r) = r, Y_{pj}(q) = s)$ is the number of observations with $Y_{pi}(r) = r, Y_{pj}(q) = s$. Using the reduced set of differences given in Eq. (9), which means setting $i = r = 1$ and $\delta_{11} = 0$, yields

$$\hat{\delta}_{iq} = \frac{(\gamma_1 - \gamma_0)n(Y_{p1}(1) = 1, Y_{pi}(q) = 0) - n(Y_{p1}(1) = 0, Y_{pi}(q) = 1)}{n(Y_{p1}(1) \neq Y_{pj}(q))}. \quad (12)$$

The choice of the scaling parameter $\gamma_{10} = \gamma_1 - \gamma_0$ is again based on minimization of a loss function. The quadratic loss has the form $L_Q(\boldsymbol{y}_{pi}, \boldsymbol{\pi}_{pi}(\hat{\theta}_p, \hat{\boldsymbol{\delta}}_i)) = \sum_r (y_{pir} - \boldsymbol{\pi}_{pir}(\hat{\theta}_p, \hat{\boldsymbol{\delta}}_i))^2$, where $\boldsymbol{y}_{pi}^T = (y_{pi0}, \dots, y_{pik})$, $y_{pir} = 1$ if $Y_{pi} = r$, $y_{pir} = 0$ otherwise, and $\boldsymbol{\pi}_{pir}^T = (\pi_{pi0}, \dots, \boldsymbol{\pi}_{pik})$, $\pi_{pir} = P(Y_{pi} = r)$. The Kullback–Leibler loss is given by $L_{KL}(\boldsymbol{y}_{pi}, \boldsymbol{\pi}_{pi}(\hat{\theta}_p, \hat{\boldsymbol{\delta}}_i)) = \sum_r y_{pir} \log(y_{pir}/\boldsymbol{\pi}_{pir}(\hat{\theta}_p, \hat{\boldsymbol{\delta}}_i))$.

The estimator (12) does not use the available information efficiently since it uses only separator functions that refer to item 1 as anchor item (by choosing $i = r = 1$ in Eq. (11)). More efficient estimates are obtained by using all items as anchor items one at a time. Thus the final estimator is constructed as an average over these estimates. It can be computed in an easy way by swapping items. For each pairs of items $(1, i)$ one computes the estimator for "new" observations $\tilde{Y}_{pi}$, where $\tilde{Y}_{p1} = Y_{pi}$, $\tilde{Y}_{pi} = Y_{p1}$, and $\tilde{Y}_{pj} = Y_{pj}$ for $j \neq 1, j \neq i$. This yields I sets of estimated item parameters that are averaged.

For illustration Fig. 7 shows the resulting estimates for the logistic cumulative model with $k = 5$, $I = 4$, $P = 100$ (only first two items shown). The item parameters are given by $(\delta_{11}, \dots, \delta_{15}) = ((0, 1, 1.5, 2, 2.5))$. The other items are shifted versions (−2 for item 2, −1.5 for item 3, −3 for item 4). In the right picture item 2 is modified by $(\delta_{21}, \dots, \delta_{25}) = 1.5(\delta_{11}, \dots, \delta_{15})$. The first row shows the resulting box plots for the averaged estimator, in the second row only item 1 is considered as anchor item. It is seen that in particular for the averaged estimator the true parameters (dots) are approximated rather well. The corresponding results when fitting the model with the minimum value distribution as response function (averaged estimator) are given in Fig. 8.

5. Concluding remarks

It has been demonstrated that parameters are separable in a wider class of models. It holds for binary as well as polytomous latent trait models. To this end an estimator is derived that uses pseudo observations and can be derived as a smoothing estimator. We do not claim that this is the only possible or best estimator. There might be much better ones that estimate parameters without reference to the other group of parameters.

The basis for success in deriving such an estimator is that separability as a property of the theoretical measurement model holds. If it holds it should be possible to derive a corresponding estimator. The main reason why only Rasch models are considered to allow for separability is that (conditional) maximum likelihood estimation has been considered the only option to generate estimators that separate parameters. However, there is no general reason why maximum likelihood estimation has to be used. Restriction to maximum likelihood estimation as a principle to obtain empirical separability is an unnecessarily limited view.

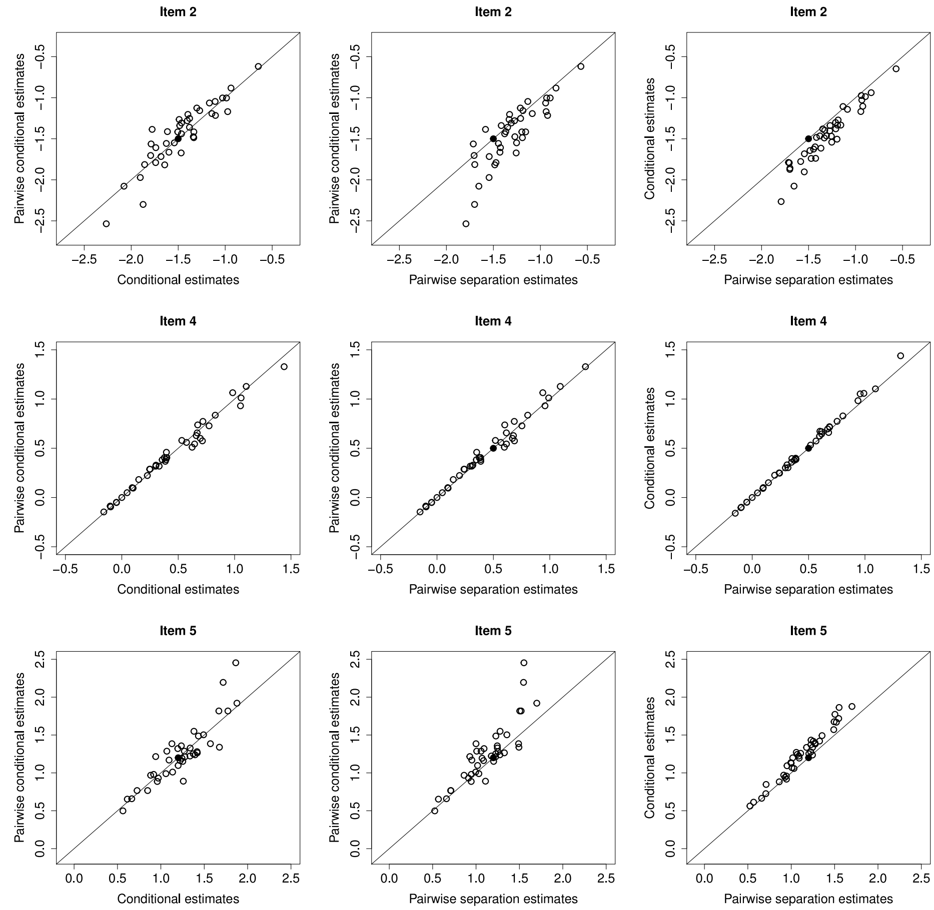

Fig. 10. Parameter estimates for conditional, pairwise conditional and separation estimators for three items, scenario as in Fig. 3.

Separability in the measurement model is based on the existence of comparator functions of the form $C(\boldsymbol{\pi}_{pi}, \boldsymbol{\pi}_{qj})$ with two probability vectors as arguments. More general functions with more arguments could be used to separate parameters. For example, Irtel (1995) proposed a function that uses four arguments to compare three subjects in a model with slopes. His function could be seen as a comparator function in the sense used here. However, he considered theoretical comparability only and did not propose an estimator of parameters.

It should be mentioned that it has been shown before that specific objectivity is not an exclusive trait of Rasch type models but is also found in alternative latent trait models. The isotonic ordinal probabilistic (ISOP) model proposed by Scheiblechner (1995) and extended in Scheiblechner (2007) has the property of ordinal specific objectivity, which is a generalization of Rasch's concept of specific objectivity. The ISOP model depends on order relations only and in the case of binary data it is equivalent to Mokken's doubly monotone model (Mokken, 1971). The main difference to the concepts considered here is that ISOP type models are nonparametric models in which relevant relations are ordinal including the property of ordinal specific objectivity. The models considered here are parametric monotone homogeneity models, for which a stronger form of specific objectivity that does not only refer to ordering holds.

For the separation estimator software will be made available on Github. Conditional likelihood estimators in binary Rasch models have been obtained by using the eRm package (Mair et al., 2009).

CRediT authorship contribution statement

Gerhard Tutz: Methodology.

Acknowledgements

Thanks to Norman Verhelst, Pascal Jordan and Clemens Draxler for their constructive and helpful comments on an earlier draft of the paper.

Appendix

Some more simulation results are given that demonstrate the efficiency of the separation estimator. Fig. 9 shows the box plots of estimates for the scenario considered in Fig. 3 but with larger number of person parameters.

Fig. 10 shows the estimates of item parameters for three items in a pairwise comparison. The scenario is the same as in Fig. 3, only the estimates for the first 40 simulations are shown. It is seen that estimates hardly differ for items in the middle of the item range ($\delta_4 = 0.5$). Deviations are found for items with small or large item parameters ($\delta_2 = -1.5, \delta_5 = 1.2$). In some simulations the pairwise conditional estimator yields extreme values.

Fig. 11 shows the box plots of estimates for the binary Rasch model with $I = 18$ items, $P = 100$ persons drawn from $N(0, 1)$. Estimates are rather similar with stronger variability seen for the pairwise conditional estimator.

Fig. 12 compares the performance of the estimators when persons are drawn from a χ-squared distribution (binary Rasch model, persons drawn from χ_1^2, $P = 100, I = 18$). It is seen that in particular the conditional pairwise estimator sometimes yields estimates far from the true values.

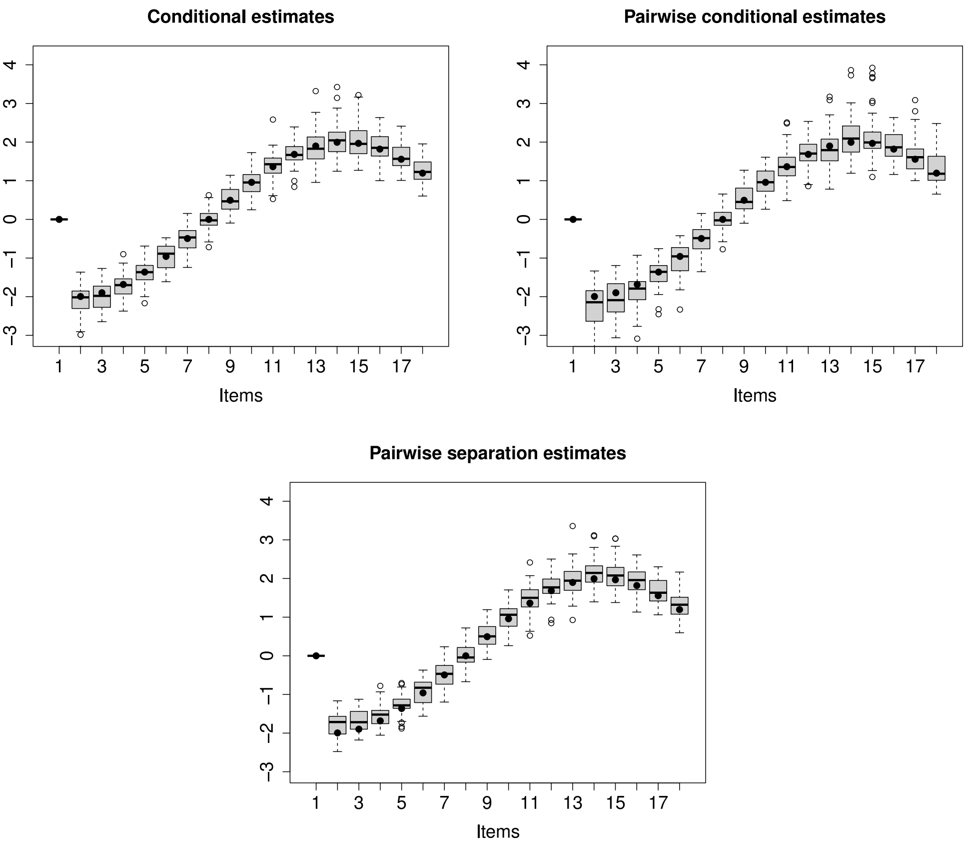

Fig. 11. Box plots for estimators; generating model is the binary Rasch model, persons drawn from $N(0,1)$ ($P = 100, I = 18$), dots indicate the true parameters.

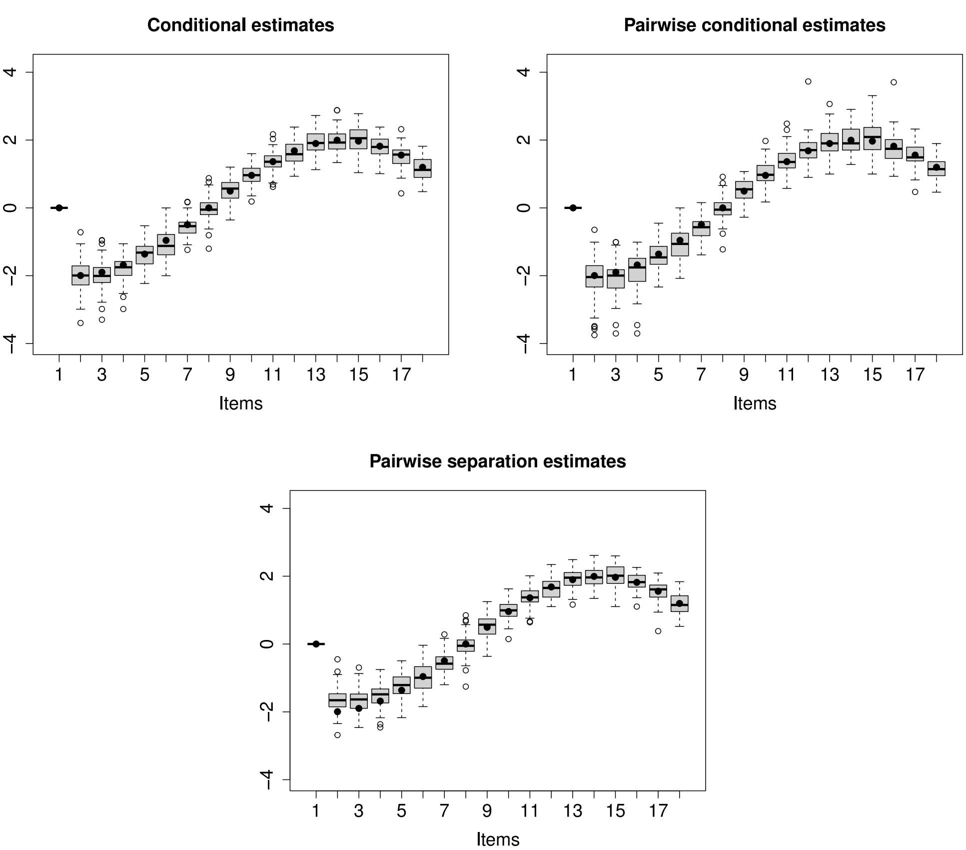

Fig. 12. Box plots for estimators; generating model is the binary Rasch model, persons drawn from χ_1^2 ($P = 100, I = 18$), dots indicate the true parameters. Conditional estimate did not exist in any of the simulations.

References

Andersen, E. (1977). Sufficient statistics and latent trait models. *Psychometrika.*, *42*, 69–81.

Andrich, D. (2016). Rasch rating-scale model. In W. Van der Linden (Ed.), *Handbook of modern item response theory* (pp. 75–94). Springer.

Bazán, J. L., Bolfarine, H., & Branco, M. D. (2006). A skew item response model. *Bayesian Analysis*, *1*(4), 861–892.

Bolt, D. M., & Liao, X. (2022). Item complexity: A neglected psychometric feature of test items? *Psychometrika.*, 1–19.

von Davier, M. (2016). Rasch model. In W. van der Linden (Ed.), *Handbook of item response theory*: *Vol. 1*, (pp. 59–76). Chapman and Hall/CRC.

von Davier, M., & Carstensen, C. H. (2007). *Multivariate and mixture distribution Rasch models*. Springer Science+ Business Media, LLC New York, NY.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*: *Vol. 57*, CRC Press.

Eggen, T. J. (2000). On the loss of information in conditional maximum likelihood estimation of item parameters. *Psychometrika.*, *65*(3), 337–362.

Eggen, T. J., & Verhelst, N. D. (2006). Loss of information in estimating item parameters in incomplete designs. *Psychometrika.*, *71*, 303–322.

Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika.*, *46*(1), 59–77.

Fischer, G. H. (1995). Derivations of the Rasch model. In G. H. Fischer, & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 15–38). Springer.

Fischer, G. H., & Molenaar, I. W. (2012). *Rasch models: Foundations, recent developments, and applications*. Springer.

Irtel, H. (1995). An extension of the concept of specific objectivity. *Psychometrika.*, *60*(1), 115–118.

Lord, F. (1952). A theory of test scores. *Psychometric Monographs*, (7).

Mair, P., Hatzinger, R., & Maier, M. J. (2009). Extended Rasch modeling: The R package eRm. *PDF-Dateianhang zum Programmpaket eRm*, 1–24.

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

Masters, G. N., & Wright, B. (1984). The essential process in a family of measurement models. *Psychometrika.*, *49*, 529–544.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). New York: Chapman & Hall.

Mokken, R. J. (1971). *A theory and procedure of scale analysis*. Berlin: Walter de Gruyter.

Pfanzagl, J. (1994). On item parameter estimation in certain latent trait models. In G. H. Fischer, & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 249–263). Springer.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

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*: *Vol. 4*, (pp. 321–333).

Samejima, F. (1995). Acceleration model in the heterogeneous case of the general graded response model. *Psychometrika.*, *60*(4), 549–572.

Samejima, F. (2000). Logistic positive exponent family of models: Virtue of asymmetric item characteristic curves. *Psychometrika.*, *65*(3), 319–335.

Samejima, F. (2016). Graded response model. In W. Van der Linden (Ed.), *Handbook of item response theory* (pp. 95–108).

Scheiblechner, H. (1995). Isotonic ordinal probabilistic models (ISOP). *Psychometrika.*, *60*, 281–304.

Scheiblechner, H. (2007). A unified nonparametric IRT model for d-dimensional psychological test data (d-ISOP). *Psychometrika.*, *72*(1), 43–67.

Simonoff, J. (1996). *Smoothing methods in statistics*. New York: Springer-Verlag.

Tutz, G. (2020). On the structure of ordered latent trait models. *Journal of Mathematical Psychology*, *96*, Article 102346.

Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*. London: Chapman & Hall.

Zwinderman, A. H. (1995). Pairwise parameter estimation in Rasch models. *Applied Psychological Measurement*, *19*(4), 369–375.