A Comparison of Estimation Methods for the Rasch Model

Alexander Robitzsch[1,2,‡]

[1] IPN – Leibniz Institute for Science and Mathematics Education, Kiel, Germany
[2] Centre for International Student Assessment (ZIB), Kiel, Germany

Abstract The Rasch model is one of the most prominent item response models. In this article, different item parameter estimation methods for the Rasch model are compared through a simulation study. The type of ability distribution, the number of items, and sample sizes were varied. It is shown that variants of joint maximum likelihood estimation and conditional likelihood estimation are competitive to marginal maximum likelihood estimation. However, efficiency losses of limited-information estimation methods are only modest. It can be concluded that in empirical studies using the Rasch model, the impact of the choice of an estimation method with respect to item parameters is almost negligible for most estimation methods. Interestingly, this sheds a somewhat more positive light on old-fashioned joint maximum likelihood and limited information estimation methods.

Key words: Rasch model, item parameter estimation, maximum likelihood estimation, item response model

1 Rasch Model

The Rasch model [9, 18] is likely the most important item response model. It is of interest to select appropriate estimation methods in a diverse applications. A variety of estimation methods has been proposed. In this article, a comprehensive comparison of different estimation methods for the Rasch model is conducted. We manipulate the factors test length (i.e., number of items), sample size, and type of ability of distribution.

For a number of items X_i ($i = 1, \ldots, I$) and a random variable θ (ability), the item response function for the Rasch model is given as

---

[‡] Alexander Robitzsch
IPN – Leibniz Institute for Science and Mathematics Education, Olshausenstrasse 62, D-24118 Kiel, Germany, e-mail: robitzsch@leibniz-ipn.de

$$P(X_i = 1|\theta; b_i) = \Psi(\theta - b_i) \quad , \quad \theta \sim F \tag{1}$$

where Ψ is the logistic link function, b_i is the item difficulty, and F is some distribution for ability θ. In addition, items X_i are assumed to be locally independent, that is $P(X_1, \ldots, X_I|\theta) = \prod_{i=1}^{I} P(X_i|\theta)$. Importantly, the sum score $S = \sum_{i=1}^{I} X_i$ is a sufficient statistic for θ if maximum likelihood (ML) estimation is employed. Hence, all items are equally weighted in θ, which eases the interpretation of Rasch model parameters. Moreover, because in the Rasch model, only a single parameter is estimated per item, low sample sizes are required for reliable estimation.

2 Estimation Methods for the Rasch Model

A variety of estimation methods has been proposed for the Rasch model [16]. In the Rasch model, item parameters $\mathbf{b} = (b_1, \ldots, b_I)$ and distribution parameters of F are estimated. Assume that item responses x_{pi} are available for persons $p = 1, \ldots, P$ and items $i = 1, \ldots, I$. Denote by $\mathbf{x}_p$ the vector of item responses and by s_p the sum score of person p.

In marginal maximum likelihood estimation (MML; [4]), latent variables θ are integrated out by posing some distributional assumption $G_{\boldsymbol{\gamma}}$ for θ, where distribution parameters $\boldsymbol{\gamma}$ are simultaneously estimated with $\mathbf{b}$. The log-likelihood function $l(\mathbf{b}, \boldsymbol{\gamma})$ is maximized. The likelihood contribution for person p is given by $l_p(\mathbf{b}, \boldsymbol{\gamma}) = \log \left[\int \prod_{i=1}^{I} P(X_i = x_{pi}|\theta; b_i) \mathrm{d}G_{\boldsymbol{\gamma}}(\theta) \right]$. If $G_{\boldsymbol{\gamma}}$ differs from the data-generating distribution F, biased item parameters can occur. Frequently, a normal distribution for θ is posed (MML-N), and a standard deviation σ is estimated. The integral in the likelihood function is evaluated by numerical integration. Alternatively, a multinomial distribution for θ can be estimated. This approach starts with a fixed grid of θ points $\theta_1, \ldots, \theta_C$ and estimates probabilities $\gamma_c = P(\theta = \theta_c)$. A log-linear smoothing of these probabilities has been proposed in the so-called general diagnostic model (MML-LM; [19, 22]). Typically, smoothing is performed for up to three or four moments. In a located latent class model with C classes, the values of the grid points θ_c are estimated in addition to probabilities γ_c (MML-LC; [7, 10]). It has been shown that in the Rasch model with I items, at most $C = I/2$ latent classes can be identified. The MML-LC approach imposes the weakest assumptions about F.

In conditional maximum likelihood estimation (CML; [1]), a conditioning step on the sum score S is performed that eliminates θ from estimation equations. In more detail, $l_p(\mathbf{b}) = \log P(\mathbf{X} = \mathbf{x}_p|S = s_p)$ is evaluated that is independent of θ. In joint maximum likelihood estimation (JML; see [16] for an overview), persons are regarded as fixed effects, and person parameters $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_P)$ are simultaneously estimated with item parameters $\mathbf{b}$. In practice, the estimation JML algorithm alternates between $\boldsymbol{\theta}$ and $\mathbf{b}$ parameter estimation in one iteration. Because the number of estimated parameters grows with sample size, a bias correction for item parameters is required [14, 21]. With obtained item parameters $\hat{b}_i$, the bias-corrected item parameter is computed as $(I-1)/I \cdot \hat{b}_i$. In order to include all persons in the es-

timation (because an ML estimate for θ is not defined for persons with extreme scores $s_p = 0$ or $s_p = I$), weighted likelihood estimation (WLE; [20]) as person parameter estimates. As an alternative to WLE, the ε-algorithm of Bertoli-Bersotti (JMLε; [3]) employs a modified likelihood by replacing the sufficient statistic s_p with $\varepsilon + (s_p - 2\varepsilon)/I$ using an appropriate $\varepsilon > 0$. In penalized JML (PJML; [5]), a ridge penalty term is added to the log-likelihood function. This approach corresponds to assuming a normal prior distribution $\theta \sim N(0, \sigma^2_{\text{prior}})$ with an appropriate choice of the regularization parameter $\sigma_{\text{prior}} > 0$. This approach also circumvents the exclusion of persons with extreme scores from CML. It has been demonstrated that JML and CML can be considered particular variants of MML estmation [13].

Several simpler estimation alternatives (so-called limited in- formation methods) do not rely on the full item response pattern $\mathbf{x}_p$. In pairwise MML (PMML; [15]) person contributions $P(X_i = x_{pi}, X_j = x_{pj})$ are considered by integrating out the latent variable θ as in MML. Typically, a normal distribution is employed. In pairwise CML (PCML; [23]), the conditioning $P(X_i = x_{pi}, X_j = x_{pj})/P(X_i + X_j = x_{pi} + x_{pj})$ is used for optimization that also removes θ from estimation equations as in CML. The row averaging approach (RA; [6]), the eigenvector method (EVM; [12]; see also [2]) as well as the MINCHI method [8] only rely on the evaluation of bivariate frequencies $P(X_i = x, X_j = y)$ $(x, y = 0, 1)$ and do not require assumptions about the distribution F of θ.

3 Simulation Study

3.1 Method

In the simulation study, item response data has been generated for the Rasch model. We varied the number of items ($I = 10$, and 30) and sample sizes ($N = 100, 250, 500$, and $1,000$). We chose I equidistant item parameters in the interval -1.5 and 1.5. Three types of ability distributions were simulated. First, we assumed a normal distribution $N(0, 1)$ (Normal) for θ. Second, we simulated a standardized chi-square (Chi^2) distribution with one degree of freedom. Third, we simulated a located latent class Rasch model with three classes (LC3) and θ points $-0.790, 1.033, 2.248$ with corresponding probabilities $.60, .35$, and $.05$.

As analysis models, we implemented the estimation methods described in Section 2. For MML-LM estimation, we used a log-linear smoothing up to three and four moments. We specified MML-LC with 3, 4, and 5 located latent classes. For JMLε estimation, we tried values $\varepsilon = 0.1, 0.3$, and 0.5. In PJML estimation, we chose normal priors $N(0, \sigma^2_{\text{prior}})$ with $\sigma_{\text{prior}} = 1, 1.5$, and 2.

The whole simulation was carried out in R [17] utilizing the R packages immer, pairwise and sirt. To enable comparisons of estimated item parameters across estimation methods, the set of item parameters were centered after estimation (i.e., they have a mean of 0). In total, 5,000 replications were conducted in each cell

of the simulation design. Bias, standard deviation (SD), and root mean square error (RMSE) were estimated for all item parameters. We consider two summary measures of item parameter recovery. First, the average absolute bias $\text{AAB}(\hat{\mathbf{b}}) = I^{-1} \sum_{i=1}^{I} |\text{Bias}(\hat{b}_i)|$ quantifies the average bias of item parameters. Second, bias and variability is summarized in the average relative RMSE (RRMSE) that is defined as $\text{RRMSE}(\hat{\mathbf{b}}) = \left[\sum_{i=1}^{I} |\text{RMSE}(\hat{b}_i)| \right] / \left[\sum_{i=1}^{I} |\text{SD}_{\text{MML}-\text{N}}(\hat{b}_i)| \right]$, where $\text{SD}_{\text{MML}-\text{N}}$ is the SD of item parameters using MML-N estimation. Hence, MML estimation using the normal distribution serves as the reference method.

3.2 Results

We only report JMLε with $\varepsilon = .3$ and PJML with $\sigma_{\text{prior}} = 1.5$ that performed best on average across conditions for lack of space. We also only state results for MML-LM with smoothing 4 moments (MML-LM4) which was superior to only using three moments). MML-LC is reported for 3 located latent classes (MML-LC3), but the there were only low efficiency losses when using 4 or 5 classes.

The bias (i.e., the MAB) of item parameters was highest for JML with using WLE (JMLW) for short test length ($I = 10$) but vanished in a long test ($I = 30$). However, MML using an incorrect normal distribution (MML-N) produced slightly biased item parameters in the case of non-normal distributions (Chi^2 and LC3). Surprisingly, the normal distributional misspecification in pairwise MML (PMML) had even worse consequences than in MML-N. Bias and RRMSE values were averaged across conditions for each methods and ranked. These ranks are shown in Table 1. Overall, CML, the limited information methods EVM, RA, and CCML as well as MML-LC3 and MML-LM4 performed best in terms of bias. It may also be surprising that MML with located latent classes (MML-LC3) also performs well for continuous ability distributions.

In Table 1, the ranks of estimation methods across all conditions and results for 10 items are shown for the RRMSE. The findings for 30 items were similar but less pronounced. Overall, JML estimation methods performed well, in particular the ε-algorithm JMLε. Notably, MML with more flexible distributions and CML produced low RRMSE values. Interestingly, misspecified MML using a normal distribution (MML-N) outperformed limited information estimators with respect to variability (PMML, CMML, EVM, RA, MINCHI). Hence, the potential bias introduced by MML-N compared to the latter estimation methods can be compensated by smaller variability. It is likely that these findings also transfer to test designs with missing data.

Table 1 Performance ranks and relative RMSE (RRMSE) of item parameters in the Rasch model for different ability distributions and estimation methods as a function of the number of items I and sample size N

	Rank		Relative RMSE											
			Normal, $I = 10$				Chi^2, $I = 10$				LC3, $I = 10$			
			sample size N				sample size N				sample size N			
Method	Bias	RMSE	100	250	500	1,000	100	250	500	1,000	100	250	500	1,000
MML-N	9	6	100.1	100.1	100.0	100.0	100.4	100.8	101.6	103.2	100.1	100.3	100.6	101.3
MML-LM4	6	3	100.2	100.1	100.1	100.1	101.1	100.7	100.3	99.9	100.8	100.1	99.6	99.1
MML-LC3	4	2	100.2	99.7	99.5	99.4	100.8	100.2	99.7	99.3	100.3	99.6	99.1	98.8
CML	1	4	100.2	100.2	100.1	100.1	100.8	100.5	100.1	99.8	100.0	99.9	99.8	99.5
JMLW	12	5	97.0	98.8	102.0	107.8	97.6	98.6	100.7	104.8	97.0	98.7	102.1	108.3
JMLε	7	1	98.2	98.4	98.6	98.8	99.0	99.3	99.5	99.7	98.0	98.2	98.4	98.6
PJML	10	7	99.3	99.5	99.5	99.5	102.6	103.6	104.8	107.1	98.7	99.0	99.2	99.7
PMML	11	8	100.1	100.1	100.0	100.0	100.5	101.0	102.0	104.1	100.4	100.7	101.2	102.2
CCML	5	9	103.2	102.7	102.4	102.5	103.5	102.7	102.4	101.8	103.1	102.8	102.2	102.0
EVM	2	10	104.0	103.5	103.2	103.3	104.3	103.5	103.2	102.5	104.1	103.7	103.1	102.9
RA	3	11	104.1	103.6	103.3	103.4	104.4	103.6	103.3	102.6	104.2	103.8	103.2	103.0
MINCHI	8	12	106.1	104.4	103.9	103.9	106.2	104.3	103.7	103.0	106.1	104.8	103.8	103.4

Note. Bias = average absolute bias (AAB); RMSE = relative root mean square error (RRMSE); Normal = $\theta \sim N(0,1)$; Chi^2 = $\theta \sim \chi^2(df = 1)$ with subsequent transformation such that $E(\theta) = 0$ and $Var(\theta) = 1$; Discrete = discrete ability distribution with 3 support points (see "Method"); MML-N = marginal maximum likelihood estimation (MML) with normal distribution; MML-LM4 = MML with log-linear smoothing up to 4 moments; MML-LC3 = MML with 3 located latent classes; CML = conditional maximum likelihood; JMLW = joint maximum likelihood estimation (JML) with WLE person parameter estimation and bias correction; JMLε = JML with ε-algorithm using $\varepsilon = 0.3$; PJML = penalized maximum likelihood estimation with prior $N(0, 1.5^2)$; PMML = pairwise MML; PCML = pairwise CML; EVM = eigenvector estimation method; RA = row averaging method; MINCHI = Fischer's Minchi estimation method. Ranks smaller than 7 and RRMSE values smaller than 100.5 are colored in gray.

4 Discussion

In this article, we compared several estimation methods for the Rasch model. It has been shown that the choice of the ability distribution impacts estimated item parameters. However, differences between estimation methods are only modest, in particular for longer test lengths. Interestingly, joint maximum likelihood estimation methods outperformed conditional and marginal maximum likelihood as well as limited information estimation methods. Prior distributions for item parameters can further improve estimation in small samples [11].

References

1. Andersen, E.B.: The numerical solution of a set of conditional estimation equations. J. R. Stat. Soc. Series B Stat. Methodol. **34**(1), 42–54 (1972). DOI 10.1111/j.2517-6161.1972.tb00887.x
2. Andrich, D., Luo, G.: Conditional pairwise estimation in the Rasch model for ordered response categories using principal components. J. Appl. Meas. **4**(3), 205–221 (2003)
3. Bertoli-Barsotti, L., Lando, T., Punzo, A.: Estimating a Rasch Model via fuzzy empirical probability functions. In: D. Vicari, A. Okada, G. Ragozini, C. Weihs (eds.) Analysis and modeling of complex data in behavioral and social sciences, pp. 29–36. Springer, Cham (2014). DOI 10.1007/978-3-319-06692-9_4
4. Bock, R.D., Aitkin, M.: Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika **46**(4), 443–459 (1981). DOI 10.1007/BF02293801
5. Chen, Y., Li, X., Zhang, S.: Joint maximum likelihood estimation for high-dimensional exploratory item factor analysis. Psychometrika **84**(1), 124–146 (2019). DOI 10.1007/s11336-018-9646-5
6. Choppin, B.: A fully conditional estimation procedure for Rasch model parameters. Eval. Educ. **9**(1), 29–42 (1982)
7. De Leeuw, J., Verhelst, N.: Maximum likelihood estimation in generalized Rasch models. J. Educ. Behav. Stat. **11**(3), 183–196 (1986). DOI 10.3102/10769986011003183
8. Fischer, G.H.: Rasch models. In: C.R. Rao, S. Sinharay (eds.) Handbook of statistics, Vol. 26: Psychometrics, pp. 515–585 (2007). DOI 10.1016/S0169-7161(06)26016-4
9. Fischer, G.H., Molenaar, I.W.: Rasch models. Foundations, recent developments, and applications. Springer, New York (1995). DOI 10.1007/978-1-4612-4230-7
10. Formann, A.K.: Constrained latent class models: Theory and applications. Brit. J. Math. Stat. Psychol. **38**(1), 87–111 (1985). DOI 10.1111/j.2044-8317.1985.tb00818.x
11. Fox, J.P.: Bayesian item response modeling. Springer, New York (2010). DOI 10.1007/978-1-4419-0742-4
12. Garner, M.: An eigenvector method for estimating item parameters of the dichotomous and polytomous Rasch models. J. Appl. Meas. **3**(2), 107–128 (2002)
13. Holland, P.W.: On the sampling theory foundations of item response theory models. Psychometrika **55**(4), 577–601 (1990). DOI 10.1007/BF02294609
14. Jansen, P.G.W., van den Wollenberg, A.L., Wierda, F.W.: Correcting unconditional parameter estimates in the Rasch model for inconsistency. Appl. Psychol. Meas. **12**(3), 297–306 (1988). DOI 10.1177/014662168801200307
15. Katsikatsou, M., Moustaki, I., Yang-Wallentin, F., Jöreskog, K.G.: Pairwise likelihood estimation for factor analysis models with ordinal data. Comp. Stat. Data An. **56**(12), 4243–4258 (2012). DOI 10.1016/j.csda.2012.04.010
16. Molenaar, I.W.: Estimation of item parameters. In: G.H. Fischer, I.W. Molenaar (eds.) Rasch models. foundations, recent developments, and applications, pp. 39–52. Springer, New York (1995). DOI 10.1007/978-1-4612-4230-7_3
17. R Core Team: *R: A language and environment for statistical computing* (2020). Vienna, Austria. https://www.R-project.org/
18. Rasch, G.: Probabilistic models for some intelligence and attainment tests. Danish Institute for Educational Research, Copenhagen (1960)
19. von Davier, M.: A general diagnostic model applied to language testing data. Brit. J. Math. Stat. Psychol. **61**(2), 287–307 (2008). DOI 10.1348/000711007X193957
20. Warm, T.A.: Weighted likelihood estimation of ability in item response theory. Psychometrika **54**(3), 427–450 (1989). DOI 10.1007/BF02294627
21. Wright, B.D., Douglas, G.A.: Best procedures for sample-free item analysis. Appl. Psychol. Meas. **1**(2), 281–295 (1977). DOI 10.1177/014662167700100216
22. Xu, X., von Davier, M.: *Fitting the structured general diagnostic model to NAEP data.* (Research Report No. RR-08-28). Educational Testing Service (2008). DOI 10.1002/j.2333-8504.2008.tb02113.x
23. Zwinderman, A.H.: Pairwise parameter estimation in Rasch models. Appl. Psychol. Meas. **19**(4), 369–375 (1995). DOI 10.1177/014662169501900406