Consistency of Rasch Model Parameter Estimation: A Simulation Study

Arnold L. van den Wollenberg
University of Nijmegen

Folkert W. Wierda
Postbank

Paul G. W. Jansen
The Netherlands Telecommunications Services

It is shown in this paper that the unconditional or simultaneous maximum likelihood estimation procedure for the one-parameter logistic model gives rise to biased estimators. This bias cannot be removed by a correction factor $(K - 1)/K$ (where K is the number of items), contrary to the contention of several authors. The bias is dependent not only on the number of items, but also on the distribution of the item parameters, which makes correcting for bias practically impossible. Furthermore, it is shown that the minimum chi-square estimation procedure, as introduced by Fischer, results in unbiased estimates. In addition, this method is computationally fast, so that it seems to be a good alternative for CML estimation when the latter method meets practical impediments. *Index terms: Maximum likelihood estimation, conditional; Maximum likelihood estimation, unconditional; Minimum chi-square estimation; One-parameter logistic model; Rasch model.*

As delineated in Jansen, van den Wollenberg, and Wierda (1988), the Rasch model (Rasch, 1960/1980) has a number of attractive properties which are guaranteed only when its parameters are estimated by the conditional maximum likelihood (CML) method. Because the CML method consumes a large amount of computing time, especially when the number of items is large, alternative estimation methods have been used extensively.

The most prominent of these is the simultaneous or unconditional maximum likelihood (UML) method. Estimators obtained under the latter procedure, however, have been proven to be inconsistent. Andersen (1973a) proved that for the case of two items, the bias in the UML estimators can be removed by a correction $(K - 1)/K = 1/2$, where K is the number of items. Haberman (1977) showed that when both K and N tend to infinity, the factor remains correct. Several authors (Fischer, 1974; Gustafsson, 1980; Wright & Douglas, 1977) have since reported that the factor $(K - 1)/K$ seems to work satisfactorily for other numbers of items as well. Recently, however, Jansen et al. (1988) have demonstrated that the simulation study of Wright and Douglas rests on inadequate logic. In addition, the results reported by Fischer (1974) and Gustafsson (1980) seem rather meager.

This implies that a new, systematic inquiry into the consistency question is needed. A simulation study was designed to compare the estimation results of the two maximum likelihood procedures with each other and with the theoretical expectations.

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 12, No. 3, September 1988, pp. 307–313
© Copyright 1988 Applied Psychological Measurement Inc.
0146-6216/88/030307-07$1.60

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

In addition, the minimum chi-square (MINCHI) procedure (Fischer, 1970) was studied. Fischer showed that this method gives consistent estimates and that its results are in good agreement with CML results. The method, however, has received little attention in the literature, because Fischer stated that the distribution of the criterion is unknown, hence statistical testing is not sufficiently substantiated. It will be argued below that as an approximation of CML, MINCHI outperforms UML.

The Minimum Chi-Square Method

As both the unconditional and the conditional methods of estimation are very well known, only the less familiar minimum chi-square method is presented here. This presentation leans heavily on Fischer (1974).

Let the stochastic variable N_{ij} represent the number of persons responding positively to item i and negatively to item j. Given the item parameters ε_i and ε_j, and fixed sum $n = n_{ij} + n_{ji}$, this stochastic variable is a binomial with expectation and variance

$$\mathrm{E}(N_{ij}) = n \frac{\varepsilon_i}{\varepsilon_i + \varepsilon_j} \tag{1}$$

and

$$\mathrm{Var}(N_{ij}) = n \frac{\varepsilon_i \varepsilon_j}{(\varepsilon_i + \varepsilon_j)^2} \quad . \tag{2}$$

Let δ_i equal the reciprocal of ε_i, and consider the deviate $N_{ij}\delta_i - N_{ji}\delta_j$. It is easily shown that this deviate has expectation 0 and variance $(n_{ij} + n_{ji})\delta_i\delta_j$ (Fischer, 1974).

Fischer (1974, pp. 269–270) proposed to minimize the following expression:

$$\sum_{i<j} \frac{(n_{ij}\delta_i - n_{ji}\delta_j)^2}{\delta_i\delta_j(n_{ij} + n_{ji})} \quad . \tag{3}$$

Although Fischer stated only that the individual terms are plausible as they are deviates normed by their standard deviations, it is easily shown that for each fixed pair of items with indices i and j considered alone, the corresponding term of Expression 3 is asymptotically distributed as chi-square with 1 degree of freedom:

$$z_{ij}^2 = \frac{[n_{ij} - \mathrm{E}(N_{ij})]^2}{n^{-1}\mathrm{E}(N_{ij})[n - \mathrm{E}(N_{ij})]} = \frac{(n_{ij}\varepsilon_j - n_{ji}\varepsilon_i)^2}{(n_{ij} + n_{ji})\varepsilon_i\varepsilon_j} = \frac{(n_{ij}\delta_i - n_{ji}\delta_j)^2}{(n_{ij} + n_{ji})\delta_i\delta_j} \quad . \tag{4}$$

When the number of items is larger than 2, the n_{ij} will be interdependent and the sum of chi-squares will not be a chi-square. The problem could be solved if the criterion could be written as a quadratic term. Another possibility would be to remove the interdependence by a proper correction factor and to accordingly adjust the degrees of freedom, as is done in the construction of the Q_1 and Q_2 statistics of van den Wollenberg (1982). Work on this is still in progress.

It is obvious that, given N_{ij} and N_{ji}, the criterion is independent of the person parameters and as such is a conditional method of estimation. Although Fischer argued that the method is mathematically not completely valid, as the n_{ij} are dependent, he found that the method gave approximately the same parameter estimates as CML, but that it consumed less time.

The Simulation Study

Method

In the generation program, item parameters were input as fixed parameters; and by means of the IMSL routine GNORM, person parameters were sampled from the standard normal distribution. By means

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

of these item and person parameters, response probabilities were obtained according to the basic equation of the Rasch model.

The response probabilities were converted into binary manifest responses by means of random numbers sampled from a uniform distribution in the range (0,1). For each person-item combination, such a random number was obtained by the IMSL routine GGUBS. A manifest response was positive when the probability exceeded the random number, and negative otherwise. In the following, analyses are reported that are based on samples of size 4,000, unless stated otherwise.

Item parameters were equally spaced in the range $(-2,2)$, for 2 through 10, 15, and 20 items. For each item parameter configuration, 25 replications were obtained. For the Rasch analyses, the program RADI (Raaymakers & van den Wollenberg, 1979) was used; this program allows for all three estimation methods.

Results

Initially, the CML and UML methods were compared both with each other and with the theoretical (input) parameters. Table 1 reports the ratio of the obtained parameter estimate to the input parameter. For the estimate to be unbiased, this ratio should equal 1. Results are reported for the first item parameter (value 2.00). This was done to avoid possible averaging out, when mean values over all items were taken. In the second column of the table, the inverse of the correction factor $(K - 1)/K$ is given; this is the expected bias.

The following conclusions can be drawn from these results:

1. The CML estimates behave properly. The mean estimates are within the theoretical probability boundaries of the sampling distribution.
2. The UML estimates are clearly biased and the bias depends heavily on the number of items; as the number of items increases, the bias decreases.
3. The bias for the case of two items is in accordance with the proof of Andersen (1973a), who showed that here the correction factor $(K - 1)/K$ must be applied.
4. The correction factor $(K - 1)/K$ is not valid for other numbers of items, although above 10 items the difference practically disappears. The difference between the predicted bias according to the correction factor and the bias observed here is further illustrated in Figure 1.

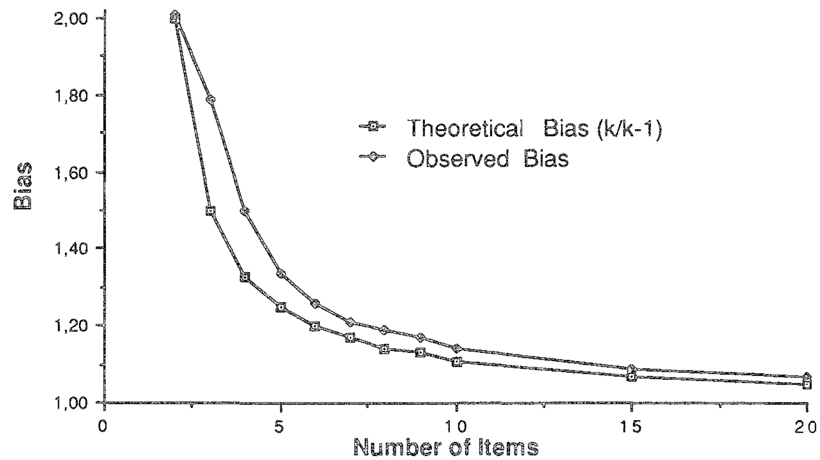
Because the correction factor is invalid, it could be asked whether another correction factor would be

Table 1
Mean Ratio of Estimated Parameters to True Item Parameters for UML and CML Procedures

K	$K/(K-1)$	UML	CML
2	2.00	2.01	1.01
3	1.50	1.79	1.00
4	1.33	1.50	.99
5	1.25	1.34	1.00
6	1.20	1.26	1.00
7	1.17	1.21	1.00
8	1.14	1.19	1.00
9	1.13	1.17	1.00
10	1.11	1.14	1.00
15	1.07	1.09	1.00
20	1.05	1.07	1.00

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Figure 1
Bias in the Unconditional Maximum Likelihood Method:
Parameter Range (−2,2)



feasible. Correction would be possible if the correction factor were solely dependent on the number of items. This, however, is not the case, as may become apparent from the next series of simulations.

For each number of items, different ranges of the item parameters were used. As is obvious from Table 2, the range also influences the amount of bias. With increasing range the bias also increases.

In the two-item case the mean ratio of observed to input parameters again behaves according to the theory. The discrepancy that is found in the case of the range (− 4.00,4.00) is accounted for by random fluctuation. For these extreme parameter values, the standard error of estimate becomes relatively large and deviations from the expected ratio of this magnitude are quite feasible. Given this result, a proper correction factor should not only account for the number of items, but also for the item parameter range.

The situation becomes even more complex when the following point is taken into consideration. In the previous analyses the item parameter distributions were always symmetrical, and it could be asked

Table 2
Estimation Bias for Three Item Parameter Ranges
in the UML Estimation Procedure

		Item Parameter Range		
K	*K*/(*K*−1)	(−2,2)	(−3,3)	(−4,4)
2	2.00	2.01	2.03	1.95
3	1.50	1.79	1.94	1.96
4	1.33	1.50	1.70	1.85
5	1.25	1.34	1.43	1.51
6	1.20	1.26	1.39	1.44
7	1.17	1.21	1.27	1.30
8	1.14	1.19	1.20	1.33
9	1.13	1.17	1.18	1.25
10	1.11	1.14	1.15	1.19
15	1.07	1.09	1.11	1.13
20	1.05	1.07	1.08	1.09

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

whether presence or absence of this characteristic could also influence the bias. In Table 3, results are reported for item parameter configurations with the same ranges as those of Table 2, but with skewed distributions of the item parameters. For $K = 2$ and $K = 3$ no results were obtained for skewed item parameters, as in these cases the restrictions of skewness and range could not be met simultaneously.

It can be seen that the bias is again influenced by this basically irrelevant property of the dataset. It must be noted, however, that although the ranges are equal for the analyses in Tables 2 and 3, the variances of the item parameter distributions are not; a contamination of skewness and variation cannot be completely excluded.

Based on these results, it must be concluded that the correction factor is not valid except for the case of two items. The failure of the correction factor decreases as the number of items increases. But the correction factor itself becomes less important when the number of items increases: The correction factor fails just when it is most needed.

The correction factor $(K - 1)/K$ cannot be substituted for by any other factor, because the bias is dependent not only on the number of items, but also on seemingly irrelevant characteristics of the dataset such as range and skewness of the item parameter distribution. The UML method is biased and there is no obvious way to correct for this bias.

The Minimum Chi-Square Estimation Method

The main arguments in favor of the UML method as opposed to the theoretically superior CML method are computational speed and accuracy. Fischer (1974) claimed that the MINCHI method gives consistent estimators and is computationally simple, giving rise to both speed and accuracy of the computational process. If these claims hold, the MINCHI method could be used as an approximation to the CML method, and thus could supplant the UML method.

In Table 4 the consistency claim is substantiated. The ratios of observed to true parameters are reported for both CML and MINCHI. It is obvious that MINCHI, like CML, gives unbiased estimates. The correspondence between the two methods is high, as may be observed in the last column, where the correlation between the estimated item parameters over the 25 replications is reported. These correlations are to be considered high, as differences in the item parameters are solely due to random fluctuations (the role of random fluctuations is rather small here because the estimates are based on samples of size

Table 3
Estimation Bias for Symmetric and Asymmetric Item Parameter Distributions in UML Estimation

K	$K/(K-1)$	Symmetric (−2,2)	Asymmetric (−3,1)
2	2.00	2.01	----
3	1.50	1.79	----
4	1.33	1.50	1.61
5	1.25	1.34	1.44
6	1.20	1.26	1.42
7	1.17	1.21	1.35
8	1.14	1.19	1.26
9	1.13	1.17	1.23
10	1.11	1.14	1.19
15	1.07	1.09	1.14
20	1.05	1.07	1.10

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Table 4
Estimation Bias for CML and the Minimum Chi-Square Procedures, and Correlation (r) of Item Parameter Estimates

K	Estimation Bias CML	Estimation Bias MINCHI	r
2	1.01	1.01	1.0000
3	1.00	1.00	.9744
4	.99	.98	.9470
5	1.00	1.00	.9623
6	1.00	1.00	.9595
7	1.00	1.00	.9543
8	1.00	1.00	.9645
9	1.00	1.00	.9537
10	1.00	.99	.9582
15	1.00	1.00	.9550
20	1.00	.99	.9764

4,000). The present analyses were also performed on smaller samples (1,000 and 200), and the results were equally satisfactory.

With respect to computing time, the MINCHI method seemed to be at least as fast as (and possibly faster than) the UML method. However, no definite conclusions could be drawn with respect to this point and more systematic comparison is needed.

Discussion and Conclusions

The above simulation studies have shown that the UML method is not consistent and the bias cannot be removed by a simple correction factor. This implies that the UML method is not a good alternative to the CML method, at least when small numbers of items are involved. When the number of items becomes large, the bias becomes relatively small and a correction is no longer needed. In that case the UML method could be used as a fast alternative to the CML method.

The MINCHI method gives consistent estimates and is at least as fast as the UML method in the range studied here. There seems to be no reason to suspect that the method becomes inaccurate with large numbers of items.

Fischer (1974) reported that MINCHI deviates from CML results when the model has been violated. Besides this, attention should be given to the behavior of test statistics under this estimation method. In the present application, test statistics (the conditional likelihood ratio test of Andersen, 1973b, and the Q_1 test of van den Wollenberg, 1982) behaved satisfactorily, but all of the datasets used here conformed to the model; power characteristics were not studied.

UML estimation is definitely not proper with relatively small numbers of items. In this case, MINCHI estimation gives a good approximation of CML and is thus an acceptable alternative. Further study is needed to gain insight into the MINCHI method, its advantages and its limitations.

References

Andersen, E. B. (1973a). *Conditional inference and models for measuring* (Publication No. 5). Copenhagen: Mentalhygiejnisk Forskningsinstitut.

Andersen, E. B. (1973b). A goodness-of-fit test for the Rasch model. *Psychometrika, 38,* 123–140.

Fischer, G. H. (1970). *A further note on estimation in*

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Rasch's measurement model with two categories of answers (Research Bulletin No. 3). Vienna: Psychological Institute, University of Vienna.

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests. Grundlagen und Anwendungen.* Bern: Huber.

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

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

Jansen, P. G. W., van den Wollenberg, A. L., & Wierda, F. W. (1988). Correcting unconditional parameter estimates in the Rasch model for inconsistency. *Applied Psychological Measurement, 12,* 297–306.

Raaymakers, M. H., & van den Wollenberg, A. L. (1979). *RADI: Program for the dichotomous Rasch model* (Internal Report 79 MA 06). Nijmegen: Psychologisch Laboratorium, University of Nijmegen.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research. [Expanded edition, University of Chicago Press, 1980.]

Rasch, G. (1966, July). An informal report on a theory of objectivity in comparisons. In L. J. T. van der Kamp & C. A. J. Vlek (Eds.), *Psychological measurement theory: Proceedings of the NUFFIC (Netherlands Universities Foundation Fund for International Cooperation) international summer session in science at "Het Oude Hof," The Hague* (unpublished).

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

Roskam, E. E., & Jansen, P. G. W. (1984). A new derivation of the Rasch model. In E. Degreef & J. van Buggenhaut (Eds.), *Trends in mathematical psychology* (pp. 293–307). Amsterdam: Elsevier North Holland.

van den Wollenberg, A. L. (1982). Two new test statistics for the Rasch model. *Psychometrika, 47,* 123–140.

Wright, B. D., & Douglas, G. A. (1977). Conditional versus unconditional procedures for sample-free item analysis. *Educational and Psychological Measurement, 37,* 573–586.

Acknowledgments

Preparation of this manuscript was supported by grant 00-56-243 of the Dutch Science Foundation (ZWO).

Author's Address

Send requests for reprints or further information to Arnold L. van den Wollenberg, Department of Psychology, Mathematical Psychology Group, University of Nijmegen, P.O. Box 9104, Nijmegen, The Netherlands.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/