

Pairwise Parameter Estimation in Rasch Models

Aeilko H. Zwinderman

University of Leiden

Rasch model item parameters can be estimated consistently with a pseudo-likelihood method based on comparing responses to pairs of items irrespective of other items. The pseudo-likelihood method is comparable to Fischer's (1974) Minchi method. A simulation study found that the pseudo-likelihood estimates and their (estimated) standard errors were

comparable to conditional and marginal maximum likelihood estimates. The method is extended to estimate parameters of the linear logistic test model allowing the design matrix to vary between persons. *Index terms: item parameter estimation, linear logistic test model, Minchi estimation, pseudo-likelihood, Rasch model.*

For Rasch model item parameter estimation, conditional maximum likelihood (CML) and marginal maximum likelihood (MML) methods are available in many computer programs (for an overview see Glas, 1989). The computational problems encountered in the past can now be handled (e.g., large item sets can be analyzed on personal computers). To avoid lengthy computation times, Fischer (1974; Fischer & Scheiblechner, 1970) investigated alternative estimation methods in the early 1970s. One method compared pairs of item responses irrespective of all other item responses. Because the method yielded a Pearson χ^2 optimization criterion, the method was called Minchi estimation (Fischer, 1974). Van der Linden & Eggen (1986) proposed a similar method based on maximizing the corresponding pseudo-likelihood. This approach also was studied by Arnold & Strauss (1988), Andrich (1988), and Choppin (1968). However, the van der Linden & Eggen (1986) method was rejected because no statistical properties of the estimators could be established, despite the fact that the method was simple, fast, and seemingly accurate.

Properties of the pseudo-likelihood method, such as consistency and the efficiency of the estimators, were investigated here. A simulation study was conducted to compare the pseudo-likelihood estimates to the CML and MML estimates. Also, the pseudo-likelihood method is extended to the linear logistic test model (LLTM; Fischer, 1983) allowing the design matrix to vary between persons.

Pseudo-Likelihood Estimation

Consider a set of k dichotomous items to which a sample of N persons responded. All persons need not respond to all items, but it is assumed here that they have. The probability that person v with trait level θ_v ($v = 1, \dots, N$) responded correctly (or to the keyed alternative) to item i with easiness parameter α_i ($i = 1, \dots, k$) satisfies the Rasch model:

$$p_{vi} = p(Y_{vi} = 1 | \theta_v, \alpha_i) = \frac{\exp(\theta_v + \alpha_i)}{1 + \exp(\theta_v + \alpha_i)} \quad (1)$$

One item parameter is fixed to 0 for identifiability reasons; thus, only $k - 1$ item parameters are estimated. Let f_{ij} be the probability that person v responded correctly to item i ($Y_{vi} = 1$) and incorrectly to item j ($Y_{vj} = 0$) given that v responded correctly to only one of them (Fischer, 1974)

APPLIED PSYCHOLOGICAL MEASUREMENT

Vol. 19, No. 4, December 1995, pp. 369-375

© Copyright 1995 Applied Psychological Measurement Inc.

0146-6216/95/040369-07\$1.60

369

$$f_{ij} = \frac{\exp(\alpha_i)}{\exp(\alpha_i) + \exp(\alpha_j)} \tag{2}$$

The α s can be estimated by maximizing the pseudo-loglikelihood of all pairs of item responses given θ :

$$l(\alpha|\theta) = \sum_{v=1}^N l_v(\alpha|\theta_v) = \sum_{v=1}^N \sum_{i < j} y_{vi}(1-y_{vj}) \log(f_{ij}) + (1-y_{vi})y_{vj} \log(1-f_{ij}) \tag{3}$$

Although Equation 3 appears to be a loglikelihood equation, it is not because the (i, j) combinations are dependent. Therefore, Equation 3 is a pseudo-loglikelihood equation. Standard maximum likelihood theory will not apply to the estimator $\hat{\alpha}$ that maximizes Equation 3. Therefore, whether $\hat{\alpha}$ is consistent must be evaluated. Also, a suitable expression for the covariance matrix of $\hat{\alpha}$ must be identified. Other properties are also important (such as bias and small sample behavior), but are not considered here.

Consistency of the Estimates

The consistency of $\hat{\alpha}$ can be proved by noting that the expectation of $l(\alpha|\theta)$ equals

$$\begin{aligned} E[l(\alpha|\theta)] &= \sum_{v=1}^N \sum_{i < j} E[\log(f_{ij})y_{vi}(1-y_{vj}) + \log(1-f_{ij})(1-y_{vi})y_{vj} | \theta_v] \\ &= \sum_{v=1}^N \sum_{i < j} \log(f_{ij}) E[y_{vi}(1-y_{vj}) | \theta_v] + \log(1-f_{ij}) E[(1-y_{vi})y_{vj} | \theta_v] \\ &= \sum_{v=1}^N \sum_{i < j} \log(f_{ij}) p_{vi}^0(1-p_{vj}^0) + \log(1-f_{ij})(1-p_{vi}^0)p_{vj}^0 \end{aligned} \tag{4}$$

where the superscript 0 denotes that true values are considered. The last step is possible due to local stochastic independence between Y_{vi} and Y_{vj} given θ_v . Equating the derivatives of Equation 4 to 0 with respect to α yields

$$\frac{\partial E[l(\alpha|\theta)]}{\partial \alpha_i} = \sum_{v=1}^N \sum_{i < j} (1-f_{ij})p_{vi}^0(1-p_{vj}^0) - f_{ij}(1-p_{vi}^0)p_{vj}^0 = 0, \tag{5}$$

which is true when for each term

$$\begin{aligned} f_{ij}(1-p_{vi}^0)p_{vj}^0 &= (1-f_{ij})p_{vi}^0(1-p_{vj}^0) \Leftrightarrow \frac{f_{ij}}{(1-f_{ij})} = \frac{1-p_{vj}^0}{p_{vj}^0} \cdot \frac{p_{vi}^0}{1-p_{vi}^0} \\ &\Leftrightarrow \frac{\exp(\alpha_i)}{\exp(\alpha_j)} = \frac{\exp(\alpha_i^0)}{\exp(\alpha_j^0)} \Leftrightarrow \exp(\alpha_i - \alpha_j) = \exp(\alpha_i^0 - \alpha_j^0) \end{aligned} \tag{6}$$

Hence, the expectation of $l(\alpha|\theta)$ is maximal when the differences between the estimated parameters ($\hat{\alpha}_i - \hat{\alpha}_j$) are equal to the differences between the true item parameters ($\alpha_i^0 - \alpha_j^0$). θ cancels when equating $\partial E[l(\alpha|\theta)]/\partial \alpha$ to 0; thus, no assumptions concerning θ or its distribution are necessary. Therefore, the conditioning argument is not further considered here.

By fixing the mean of the estimates to 0, $\hat{\alpha}$ is a consistent estimator of α . This result also applies to CML and MML: only differences between item parameters are estimated consistently by their CML or MML estimators (Glas, 1989). The above conclusion about the consistency of $\hat{\alpha}$ is valid only when $\hat{\alpha}$ exists; it does not ensure uniqueness. As yet, no sufficient and uniqueness conditions for the existence of $\hat{\alpha}$ are known,

but a necessary condition is that there is at least one person responding correctly to item i and incorrectly to item j for all combinations (i, j) .

The pseudo-score function or first-order derivative of $l(\alpha)$ for α_i is

$$U(\alpha_i) = \frac{\partial l(\alpha)}{\partial \alpha_i} = \sum_{v=1}^N \sum_{i < j} y_{vi}(1-y_{vj})(1-f_{ij}) - y_{vj}(1-y_{vi})f_{ij} = \sum_{j \neq i} [n_{ij} - (n_{ij} - n_{ji})f_{ij}], \tag{7}$$

where n_{ij} is the number of persons responding correctly to item i and incorrectly to item j .

Estimates of the Covariance Matrix

It can be shown (Connolly & Liang, 1988, p. 504) that under mild regularity conditions $(\hat{\alpha} - \alpha^0)$ is asymptotically normally distributed with mean $\mathbf{0}$ and covariance matrix

$$\left[\frac{-\partial^2 l(\alpha^0)}{\partial \alpha^2} \right]^{-1} \left\{ \text{cov} \left[\frac{\partial l(\alpha^0)}{\partial \alpha} \right] \right\} \left[\frac{-\partial^2 l(\alpha^0)}{\partial \alpha^2} \right]^{-1}. \tag{8}$$

Because the contributions of the N persons to the score functions are independent of each other and have expectation 0, an estimate of the covariances of the score functions is given by (see Connolly & Liang, 1988, p. 504)

$$\text{cov} \left[\frac{\partial l(\alpha^0)}{\partial \alpha} \right] = \sum_{v=1}^N \left[\frac{\partial l_v(\alpha^0)}{\partial \alpha} \right] \left[\frac{\partial l_v(\alpha^0)}{\partial \alpha} \right]^T. \tag{9}$$

Accordingly, the covariance matrix of $\hat{\alpha}$ can be estimated by substituting $\hat{\alpha}$ in Equations 5 and 6. This results in simple counting and arithmetic. The negative matrix of second-order derivatives of Equation 1 is of order k , but has rank $k - 1$, and has diagonal entries $\sum_j (n_{ij} + n_{ji})f_{ij}(1 - f_{ij})$ and off-diagonal entries $-(n_{ij} + n_{ji})f_{ij}(1 - f_{ij})$. The covariance matrix of the score functions (Equation 9) is also of order k , has rank $k - 1$, and has entries $\sum_v g_{ih}g_{vj}$, where $(h = 1, \dots, k; h \neq i)$

$$g_i = \sum_{h=1 \neq i}^k \left\{ y_{vi}(1-y_{vh}) - [y_{vi}(1-y_{vh}) - (1-y_{vi})y_{vh}] f_{ih} \right\}. \tag{10}$$

An Illustrative Simulation

The pseudo-likelihood estimator was compared with the CML and MML estimators in a simulation study using 13 items and $N = 500$. Item responses were simulated according to a method described elsewhere (Zwinderman, 1991). Item parameters were fixed (see Table 1) and θ s were sampled from the standard normal distribution. For each person-item combination, the response probability was computed according to Equation 1. The simulated observation was set to 1 if this probability was greater than a random number generated from a $[0, 1]$ uniform distribution and 0 otherwise. CML and MML estimates were computed using the RSP program (Glas & Ellis, 1994).

Results are reported in Table 1. The pseudo-likelihood estimates were not very different from the CML and MML estimates, nor were the corresponding standard errors (SEs). In general, CML and MML estimates were very similar to each other, but the pseudo-likelihood estimates were slightly different. However, none of the three sets of estimates was in general best (i.e., none was the closest to the true item parameters). The estimated SEs of the pseudo-likelihood estimates were on average slightly larger, which is probably due to the fact that the covariance matrix in Equation 5 tends to be unstable and underestimates the true covariance matrix.

Table 1
 Pseudo-Likelihood, CML, and MML Estimates of the Simulated Item Parameters

True Item Parameters	Estimated Item Parameters					
	Pseudo		CML		MML	
	$\hat{\alpha}$	SE	$\hat{\alpha}$	SE	$\hat{\alpha}$	SE
-3.0	-2.92	.151	-2.84	.169	-2.83	.168
-2.5	-2.59	.212	-2.64	.159	-2.63	.158
-2.0	-1.83	.136	-1.80	.126	-1.80	.125
-1.5	-1.42	.148	-1.47	.117	-1.47	.117
-1.0	-1.16	.127	-1.16	.111	-1.17	.110
-.5	-.48	.097	-.49	.102	-.50	.102
.0	-.08	.103	-.06	.100	-.06	.099
.5	.59	.104	.60	.102	.60	.101
1.0	.95	.111	.97	.105	.96	.104
1.5	1.50	.094	1.51	.113	1.51	.113
2.0	1.90	.098	1.86	.121	1.82	.121
2.5	2.50	.138	2.49	.142	2.50	.144
3.0	3.03	.174	3.03	.169	3.06	.170

Extension of the Pairwise Estimation Method to the LLTM

In the LLTM (Fischer, 1974, 1983), the item parameters in Equation 1 are modeled as a function of a design matrix \mathbf{X} and basic weights β , where \mathbf{X} is the same for all persons. The model is extended here by allowing the design matrix \mathbf{X} to vary between persons. Inserting into Equation 1 yields

$$p_{vi} = P(Y_{vi} = 1 | \theta_v, \mathbf{X}_{vi}, \beta) = \frac{\exp(\theta_v + \beta' \mathbf{X}_{vi})}{1 + \exp(\theta_v + \beta' \mathbf{X}_{vi})} \tag{11}$$

where \mathbf{X}_{vi} is a vector of length m with known entries and β is a vector of length m with unknown parameters. \mathbf{X} is a vector with the observed values of m covariates, and these vary both between and within persons (from item to item). Thus, Equation 11 is the logistic regression model with random effects, and β may be interpreted as a regression weight.

β can be estimated using CML or MML estimation, but it is difficult. When using CML, the basic symmetric functions and their first- and second-order derivatives (Liou, 1994; Verhelst, Glas, & van der Sluis, 1984) must be evaluated for all contributions to the loglikelihood. These cannot be grouped together in sum score groups as in the ordinary Rasch model. With 10,000 persons responding to 50 items, the CML loglikelihood of the ordinary Rasch model consists of 49 terms, one for each sum score group; however, in the LLTM it consists of $50 \times 10,000$ terms because the design matrix can vary. When using MML, a similar numerical explosion occurs, regarding the evaluation of the integral involved in the MML method. In contrast, the numerical work of the pseudo-likelihood approach remains relatively limited.

The pseudo-loglikelihood is still given by Equation 3 with β substituted for α , but f_{ij} is now defined as:

$$f_{ij} = \frac{\exp(\beta' \mathbf{X}_{vi})}{\exp(\beta' \mathbf{X}_{vi}) + \exp(\beta' \mathbf{X}_{vj})} \tag{12}$$

Equating the derivatives of $E[l(\beta)]$ to 0 with respect to β yields

$$f_{ij}(1 - p_{vi}^0)p_{vj}^0 = (1 - f_{ij})p_{vi}^0(1 - p_{vj}^0) \Leftrightarrow \frac{f_{ij}}{1 - f_{ij}} = \frac{1 - p_{vj}^0}{p_{vj}^0} \frac{p_{vi}^0}{1 - p_{vi}^0} \Leftrightarrow \frac{\exp(\beta' \mathbf{X}_{vi})}{1 + \exp(\beta' \mathbf{X}_{vj})} = \frac{\exp(\beta^0' \mathbf{X}_{vi})}{\exp(\beta^0' \mathbf{X}_{vj})} \Leftrightarrow \beta = \beta^0 \tag{13}$$

for each term of Equation 4, where β^0 refers to the true value. Therefore, $\hat{\beta}$, which maximizes $l(\beta)$, is a consistent estimator of true β , if it exists. Again no uniqueness and sufficient conditions for the existence of $\hat{\beta}$ are known, but in addition to the conditions mentioned before, it is necessary that each covariate X has nonzero positive variance.

The pseudo-score equation for the regression parameter of the p th covariate, $\hat{\beta}_p$, is:

$$U(\beta_p) = \sum_{v=1}^N U_v(\beta_p) = \sum_{v=1}^N \sum_{i < j} (X_{vip} - X_{vjp}) \left[y_{vi}(1 - y_{vj})(1 - f_{ij}) - y_{vj}(1 - y_{vi})f_{ij} \right], \tag{14}$$

where $U(\cdot)$ is defined as in Equation 7.

Because the contributions of the N persons to the score functions are independent and have expectation 0, the covariance matrix of $\hat{\beta}$ can be estimated with the estimator in Equation 8. The negative matrix of second-order derivatives of the pseudo-loglikelihood is of order and rank m with elements $(p, q = 1, \dots, m)$

$$\frac{-\partial^2 l(\beta)}{\partial \beta_p \partial \beta_q} = \sum_{v=1}^N \sum_{i < j} (X_{vip} - X_{vjp})(X_{viq} - X_{vjq}) f_{ij}(1 - f_{ij}) \left[y_{vi}(1 - y_{vj}) + y_{vj}(1 - y_{vi}) \right]. \tag{15}$$

The covariance matrix of the score functions is also of order and rank m and has elements $\sum_v U_v(\beta_p)U_v(\beta_q)$, which unfortunately do not reduce to a simple equation.

An Example

Method. The binary item responses of 209 asthmatic patients to an 11-item questionnaire that measures functional independence were analyzed (Maille, Koning, Zwinderman, & Kaptein, 1994). The ages of the respondents varied from 41 to 87. The mean age was 64, and the standard deviation was 14 years.

According to Glas' R_1 test in the RSP program (Glas & Ellis, 1994), the Rasch model fit the data ($R_1 = 57$, $df = 40$, $p = .04$), except when the sample was split into two groups (persons 65 years or younger and persons older than 65; $R_1 = 62$, $df = 10$, $p < .0001$). Therefore, differential item functioning was investigated using Equation 11. Two sets of covariates (X and Z) were constructed: (1) $x_{vj} = 1$ for item $j + 1$ ($j = 1, \dots, 10$) and 0 otherwise; and (2) z_{vj} , the interaction between x_{vj} and v 's age. The first set of covariates (X) was simply a transformation of the usual item parameters, but normalized so that $\alpha_1 = 0$. The second set indicated whether the item easiness parameters varied over persons of different ages. Age could also be incorporated into Equation 11, but its regression parameter could not be estimated with the pseudolikelihood method due to the conditioning argument in Equation 12 (nor with the CML method).

Results. Table 2 shows the CML and pseudo-likelihood estimates of the regression parameters. As in the simulation, the CML and pseudo-likelihood estimates were similar, as were their SEs. Item 3 ("Have you difficulty walking stairs?") appeared to be more difficult for older persons than for younger [Wald test (Rao, 1973, p. 417): $p < .001$]. For example, the conditional probability of a correct response to Item 3 given a 0 value of the latent functional independence scale ($\theta = 0$) was .27 for a 50-year old and .86 for a 75-year old. It was concluded that difficulties walking stairs were not particularly indicative of the functional limitations experienced by asthmatic patients, but that they reflected the aging process.

Conclusions

Using the pseudo-likelihood method to estimate Rasch model item parameters provides estimators that are consistent and similar in efficiency to CML and MML estimators, at least for the small datasets examined here. The pseudo-likelihood method was developed to solve the estimation problems of the Rasch model, but many have already been solved. Nevertheless, pseudo-likelihood estimation is a powerful tool that may be used for situations in item response theory in which estimation is not yet straightforward, such as in linear logistic models or longitudinal or dynamic logistic test models.

Table 2
 Pseudo-Likelihood and CML Estimates of the Item Parameters, and the Interaction Effect of Age and the Item Parameters ($\hat{\alpha} \times \text{Age}$)

Item	Pseudo				CML			
	$\hat{\alpha}$	SE	$\hat{\alpha} \times \text{Age}$	SE	$\hat{\alpha}$	SE	$\hat{\alpha} \times \text{Age}$	SE
1	0.00	—	0.000	—	0.00	—	0.000	—
2	-.32	.97	.003	.015	-.44	1.01	.004	.016
3	7.01	1.18	-.117	.018**	6.47	1.20	-.110	.019**
4	-.03	.99	-.010	.016	.15	1.02	-.012	.016
5	.52	1.03	-.023	.016	.66	1.04	-.025	.016
6	-.58	1.07	-.010	.017	-.69	1.06	-.009	.017
7	-2.26	1.01	.008	.016	-2.26	1.16	.008	.018
8	-2.25	1.05	.002	.016	-2.34	1.23	.003	.019
9	-4.51	1.59	.026	.023	-4.92	1.59	.032	.020
10	-.57	1.24	-.031	.019	-.97	1.25	-.023	.020
11	-.14	1.36	-.054	.023*	-.07	1.54	-.052	.025*

* $p < .05$ according to the Wald test.

** $p < .001$ according to the Wald test.

The present pairwise estimation method also has limitations. The estimator of the covariance matrix of the estimators (Equation 8) tends to be unstable, because it depends on both the parameter estimates and the data. In LLTMs, the regression parameters of covariates that vary between persons, but are constant within persons, cannot be estimated. Finally, the method does not generalize to other item response theory models. It only generalizes partially to polytomous item responses. For example, the item parameters of the unidimensional polytomous Rasch model with equidistant scoring of the categories (Andersen, 1973) may be estimated consistently and efficiently by comparing all pairs of items i and j with a response to category c on item i and a response not in c to item j . However, the category parameters of this model cannot be estimated by such pairwise comparisons.

References

- Andersen, E. B. (1973). Conditional inference for multiple choice questionnaires. *British Journal of Mathematical and Statistical Psychology*, 26, 31-44.
- Andrich, D. (1988). *Rasch models for measurement*. Beverly Hills: Sage.
- Arnold, B. C., & Strauss, D. (1988). *Pseudolikelihood estimation* (Technical Rep. 164). Riverside: University of California, Department of Statistics.
- Choppin, B. (1968). An item bank using sample-free calibration. *Nature*, 219, 870-872.
- Connolly, M. A., & Liang, K. Y. (1988). Conditional logistic regression models for correlated binary data. *Biometrika*, 75, 501-506.
- Glas, C. A. W. (1989). *Contributions to estimating and testing Rasch models*. Unpublished doctoral dissertation, University of Twente, Enschede, The Netherlands.
- Glas, C. A. W., & Ellis, J. L. (1994). *RSP: The Rasch scaling package* [Computer program]. Groningen: Institute Progamma.
- Fischer, G. H. (1974). *Einfuehrung in die theorie psychologischer tests* [Introduction to the theory of psychological tests]. Bern: Huber.
- Fischer, G. H. (1983). Logistic latent trait models with linear constraints. *Psychometrika*, 48, 3-26.
- Fischer, G. H., & Scheiblechner, H. (1970). Algorithmen und programme für das probabilistischen testmodell von Rasch [Algorithms and programs for the probabilistic test model of Rasch]. *Psychologische Beiträge*, 12, 23-32.
- Liou, M. (1994). More on the computation of higher-order derivatives of the elementary symmetric functions in the Rasch model. *Applied Psychological Measurement*, 18, 53-62.
- Maille, A. R., Koning, C. J. M., Zwiderman, A. H., & Kaptein, A. A. (1994). De ontwikkeling van een kwaliteit van leven vragenlijst voor CARA-patienten [The development of a quality of life questionnaire for

patients with respiratory illness]. *Gedrag en Gezondheid*, 22, 35–43.

Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: Wiley.

van der Linden, W. J., & Eggen, T. J. H. M. (1986). An empirical Bayesian approach to item banking. *Applied Psychological Measurement*, 10, 345–354.

Verhelst, N. D., Glas, C. A. W., & van der Sluis, A. (1984). Estimation problems in the Rasch model: The basic symmetric functions. *Computational Statistics Quarterly*, 1, 245–262.

Zwinderman, A. H. (1991). A generalized Rasch model

for manifest predictors. *Psychometrika*, 56, 589–600.

Acknowledgments

This research was supported by a grant from the Netherlands Organization for Scientific research (NWO).

Author's Address

Send requests for reprints or further information to Aeilko H. Zwinderman, Department of Medical Statistics, University of Leiden, P.O. Box 9604, 2300 RC Leiden, The Netherlands. Email: a.h.zwinderman@medstat.medfac.leidenuniv.nl.