DOCUMENT RESUME

ED 308 246 TM 013 666

AUTHOR	Engelen, R. J. H.
TITLE	Semiparametric Estimation in the Rasch Model. Project Psychometric Aspects of Item Banking No. 14. Research Report 87-1.
INSTITUTION	Twente Univ., Enschede (Netherlands). Dept. of Education.
PUB DATE	87
NOTE	31p.
AVAILABLE FROM	Mediatheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Equations (Mathematics); *Estimation (Mathematics); Foreign Countries; *Latent Trait Theory; Mathematical Models; Maximum Likelihood Statistics; Secondary Education; Secondary School Students; Statistical Analysis
IDENTIFIERS	Netherlands; Parameter Identification; *Rasch Model; *Semiparametric Estimation

ABSTRACT

A method for estimating the parameters of the Rasch model is examined. The unknown quantities in this method are the item parameters and the distribution function of the latent trait over the population. In this sense, the method is equivalent to marginal maximum likelihood estimation. The new procedure is based on a method suggested by J. Kiefer and J. Wolfowitz (1956). Their conclusions are reviewed, and links to the Rasch model are specified. In marginal maximum likelihood estimation, the item parameters are estimated first, and then the prior distribution of the person parameters is estimated using these estimates. The proposed method illustrates that it is possible to estimate these two quantities together and arrive at consistent estimates. Two data tables are provided. (SLD)

Semiparametric Estimation in the Rasch Model

Research Report 87-1

R.J.H. Engelen

Department of Education

Division of Educational Measurement and Data Analysis

University of Twente

Project Psychometric Aspects of Item Banking No.14

Colophon

Typing : L. Padberg

Cover design : M. Driessen, AV-section, University of Twente

Printed by : Central Reproduction Department, University of Twente

Semiparametric Estimation in the

Rasch Model

R. Engelen

Abstract

Various maximum likelihood estimation procedures in the Rasch model are reviewed. It is shown that semiparametric estimation allows simultaneous estimation of the item parameters and the distribution function of ability. Moreover, both the item parameters and the distribution function of ability are estimated consistently.

Key words: Rasch model, Incidental Parameters, Structural Parameters.

Semiparametric Estimation in the Rasch Model

In the framework of latent trait theory, the Rasch (1960) model has become increasingly popular over the past decade. This is not only due to its wide applicability (Fischer, 1977; Lord, 1983), but also to the fact that statistically nice procedures for estimating its parameters have been developed.

The purpose of this paper is to examine yet another method for estimating the parameters of the Rasch model. In contrast with other methods, the unknown quantities in this method are the item parameters and the distribution function of the latent trait over the population. In this sense, the method is equivalent to marginal maximum likelihood estimation where the prior distribution function (of the latent trait in the population) is completely unspecified and has to be estimated from the data.

As an introduction, in the second section, we will shortly review the different maximum likelihood estimation procedures that have previously been used with the Rasch model. These procedures are unconditional maximum likelihood (UML), conditional maximum likelihood (CML) and marginal maximum likelihood (MML). In addition, the new estimation procedure developed in this paper is introduced.

The new estimation procedure is based on a method suggested by Kiefer and Wolfowitz (1956). In the paper, they show that if the number of parameters increases with the number of observations, then UML estimation does not always yield consistent estimates for all parameters. Two kinds of parameters are distinguished, structural and incidental parameters. A parameter that appears in all the distribution functions will be called structural, while all others are called incidental. Within this context Kiefer and Wolfowitz show that, under fairly general conditions, one is able to consistently estimate the structural parameter and the distribution function of the incidental parameters simultaneously. The essence of their paper, and some links to the Rasch model will be given in the third section.

The fourth section serves to check Kiefer and Wolfowitz' conditions in the specific case of the Rasch model.

Finally, a discussion and an example will be given in the last section.

Maximum Likelihood Estimation in the Rasch Model

We assume that a group of V examinees are administered k dichotomously scored items and that all items measure the same unidimensional (latent) trait or ability.

In the Rasch model, the item response function, i.e., the probability that examinee v answers item i correctly, takes a simple form:

$$(1) \qquad P(X_{vi}=1|\theta_v,\sigma_i) = \frac{\exp(\theta_v-\sigma_i)}{1+\exp(\theta_v-\sigma_i)},$$

where θ_v is the person's ability parameter and σ_i is the item difficulty parameter. On the usual conditional independence assumptions, the probability that person v's response pattern is $x_v = (x_{v1},\ldots,x_{vk})$ is given by

$$(2) \qquad P(X_v=x_v|\theta_v,\sigma) = \prod_i \frac{\exp\{x_{vi}(\theta_v-\sigma_i)\}}{1+\exp(\theta_v-\sigma_i)}$$

where $x_{vi} = 1$ if item i is answered correctly and $x_{vi} = 0$ otherwise, and $\sigma = (\sigma_1,\ldots,\sigma_k)$. Hence, the joint probability distribution of the response vectors for all examinees is given by

$$(3) \qquad P(X_1=x_1,\ldots,X_v=x_v|\theta_1,\ldots,\theta_v,\sigma) = \prod_v \prod_i \frac{\exp\{x_{vi}(\theta_v-\sigma_i)\}}{1+\exp(\theta_v-\sigma_i)}.$$

If both item and person-parameters are unknown, then the objective is to estimate both sets of parameters. Some standard methods of estimation will be briefly reviewed; recall that we are dealing with maximum likelihood estimation only.

The first method is unconditional maximum likelihood estimation (UML), where all the parameters are estimated simultaneously, by maximizing the joint likelihood function (3) over all parameters. But this leads to inconsistent item parameter estimates, as was shown by Andersen (1973). This is the link to Kiefer and Wolfowitz (1956), since in the Rasch model we are dealing with structural (item) and incidental (person) parameters.

The second method is conditional maximum likelihood estimation

(CML). In this method, one uses the fact that the total score for person v, i.e., the total number of correct responses, is a sufficient statistic for that person's ability. Therefore, one can condition the likelihood function on the sufficient statistics so that the (incidental) person parameters are no longer part of it. Hence, the likelihood is now a function of the item parameters only, which can be estimated consistently by maximum likelihood. Using these estimates, the person parameters can now be estimated.

The third method is marginal maximum likelihood estimation (MML). In this method, a prior distribution for the person parameters is specified and the (incidental) person parameters are integrated out of the likelihood function, so that again one is able to estimate the item parameters. The specification of the prior distribution can be done in several ways; usually one assumes a special parametric form for the prior distribution, say $f(x|\alpha)$, and the parameter vector α is estimated along with the item parameters.

The new estimation method is based on MML, the only difference being that no assumptions on the prior distributions are made. Hence, this method can be called semiparametric marginal maximum likelihood estimation (Wellner, 1985).

Structural and Incidental Parameters

The notion of incidental and structural parameters was introduced rigorously by Neyman and Scott (1948). Suppose that we have independent random variables X_{ij} ($j=1,\ldots,k$, $i=1,2,\ldots$), such that $X_{i1},\ldots,X_{ik}$ ($i=1,2,\ldots$) are identically distributed with a distribution function of the form $f(x|\sigma,\theta_i)$, where σ and θ_i are possibly vector valued. Then Neyman and Scott (1948) called the parameter σ, upon which all the distributions depend, structural, whereas the parameters $\{\theta_i\}$ are called incidental. They showed that the maximum likelihood estimator of σ, in the presence of the incidental parameters $\{\theta_i\}$, need not be consistent. They also gave the following nice and easy example:

Example: Let X_{ij}, $i=1,\ldots,n$, $j=1,\ldots,k$ be independent random variables such that $X_{i1},\ldots,X_{ik}$ are normally distributed with mean μ_i and variance σ^2 ($i=1,\ldots,n$). Now consider the problem of estimating σ^2 (as $n \to \infty$). The maximum likelihood estimates of the parameters are then given by

$$\mu_i = \sum_j x_{ij}/k \equiv x_{i.}, \text{ for } i = 1,\ldots,n.$$

$$\sigma^2 = \sum_i \sum_j (x_{ij}-x_{i.})^2/nk \equiv S^2 .$$

However, it is well known that $E(S^2) = k\sigma^2/(k-1)$, and hence the maximum likelihood estimate of σ^2 is not consistent.

This problem was investigated, among others, by Kiefer and Wolfowitz (1956) [henceforth denoted as KW (1956)] and Andersen (1973). Note that there is an intuitive explanation for the above phenomenon; with each new group of observations we get additional information about the structural parameter, but we also introduce a new incidental parameter and thereby more 'bias'.

Now KW (1956) proved that the maximum likelihood estimator of the structural parameter is (strongly) consistent, when the (infinitely many) incidental parameters are iid with a common distribution function. Furthermore, this distribution can also be estimated consistently. The following is a heuristic explanation for their results: "a sequence of chance variables is more 'regular' than an arbitrary sequence".

A different route was followed by Andersen (1973), who solved the problem by using and developing the Neyman and Scott method of conditional inference. In this set up, one needs to have sufficient statistics for the incidental parameters to be able to estimate the structural parameter consistently.

It is the purpose of this paper to explore KW (1956) ideas and use their method in one specific case: the Rasch model. Note that all maximum likelihood estimation methods mentioned in the second section fit into KW's (1956) framework; in the Rasch model one has structural and incidental parameters which can be estimated simultaneously (UML), after conditioning on a sufficient statistic (CML), or marginally (MML).

To proof consistency, KW (1956) need the fulfillment of several assumptions. Since their proof is of the Wald-type (i.e., no differentiability assumptions for the underlying density function are assumed), their assumptions and proof are essentially a modification of Wald's (1949) proof of the consistency in the ordinary case (only a finite number of parameters).

Let us review Kiefer and Wolfowitz assumptions briefly. Note that the parameters in KW (1956) are the structural parameter (σ) and the common prior distribution of the incidental parameters $P(\theta)$. KW (1956) consider the densities

$$(4) \qquad \bar{f}(x_i|\sigma,P) = \int f(x_i|\sigma,\theta)\, dP(\theta); \qquad i = 1,2,\ldots$$

Now suppose that Q is a set of priors P such that the true prior P_0 is in Q; the true structural parameter is denoted by σ_0. Next, KW (1956) define a maximum likelihood estimator for (σ_0,P_0) as a pair (σ_n',P_n') such that

$$(5) \qquad \prod_i \bar{f}(x_i|\sigma_n',P_n') \geq \prod_i \bar{f}(x_i|\sigma,P)$$

for all σ and all P. It is KW's (1956) main intention to give conditions such that $\sigma_n' \to \sigma_0$ and $P_n' \to P_0$. In doing so, an important role is played by the identifiability assumption:

If $(\sigma_1,P_1) \neq (\sigma_2,P_2)$ then there is a value y for which

$$(6) \qquad \int_{-\infty}^{y} \bar{f}(x|\sigma_1,P_1)\, d\mu \neq \int_{-\infty}^{y} \bar{f}(x|\sigma_0,P_2)\, d\mu .$$

If these conditions are fulfilled, then KW (1956) are able to proof the consistency of the ML estimates.

There is a problem with the size of the set of priors Q; if Q is too large, the identifiability assumption may not be fullfilled or no prior P_n may be found such that (5) is true, and if Q is too small, the true prior P_0 need not be an element of Q. Therefore, additional conditions on the set of priors Q are needed.

Checking the KW-assumptions in the Rasch Model

In the Rasch model, we consider $\sigma = (\sigma_1,\ldots,\sigma_k)$ (the item parameter) as the 'structural' parameter, whereas θ_v (the person parameters, v=1,2,...) are the incidental parameters. Note that the number of items is usually fixed and small, and the number of examinees is large ; with each new examinee we also introduce a new, unknown person parameter.

Furthermore, in the Rasch model, the vector of item parameters is multidimensional, while in KW (1956) paper the structural parameter is unidimensional. This is no serious problem however.

Note that the Rasch model itself is unidentifiable; adding the same number to both all item and the person-parameters does not change the probability in (3). Therefore, a constraint has to be imposed on the parameters. A common one is $\sum_i \sigma_i = 0$. For our purposes, as will become clear later on, it is easy to take the constraint $\sigma_1 = 0$.

<u>Assumption 1:</u> $f(x|\sigma,\theta_v)$ is a density with respect to a sigma finite measure μ on a Euclidean space of which x is a generic point.

Now, if we take $X = N^k$, $A(X) = (2^N)^k$ the sigma field of all subsets of N^k, and μ as the product of the k counting measures on $(N,2^N)$, this assumption can be shown to be fulfilled for the Rasch model. Let Y_i be Bin $(1,\exp(\theta_v-\sigma_i)/(1+\exp(\theta_v-\sigma_i)))$ and X be $\prod_i Y_i$ distributed. Then X generates a probability measure P on (X,A(X)) with $dP/d\mu = f$ and

$$(7) \qquad f(x_v) = \begin{cases} \prod_i \dfrac{\exp\{x_{vi}(\theta_v-\sigma_i)\}}{1+\exp(\theta_v-\sigma_i)} & \text{if } x_{vi}=0 \text{ or } 1 \quad (i=1,\ldots,k) \\ 0 & \text{otherwise} \end{cases}$$

This shows the validity of assumption 1.

In order to proof the validity of assumption 2, some additional definitions are needed. Note that R^k is the space of values that σ can take and that R is the space of possible values of θ. Denote the set of all cumulative distribution functions G with finite second moment by Q. Later on, it will become clear that more assumptions on Q are needed, but for the time being this definition will do. Let σ_0 and G_0 be the true value of the parameter σ and the true distribution function of θ respectively. Assume that (σ_0,G_0) is in (R^k,Q) and denote this true parameter by $\beta_0 = (\sigma_0,G_0)$.

Now, let $\beta = (\sigma,G)$ be a generic point in R^k x Q. Define

(8) $\bar{f}(x|\beta) = \int_R f(x|\sigma,z)dG(z)$

and

(9) $$\delta(\beta',\beta'') = \delta\{(\sigma',G'),(\sigma'',G'')\} = \sum_i |\arctan \sigma_i' - \arctan \sigma_i''| + \int_R |G'(z)-G''(z)|\exp(-|z|)dz$$

Then, δ is a metric in $R^k \times Q$. The completed space of $R^k \times Q$ ($R^k \times Q$ together with all the limits of its Cauchy sequences in the sense of the metric δ), will be denoted by $\underline{R^k \times Q}$.

Note that we have a well-known connection (Billingsley, 1977) between weak convergence and the metric δ, given by the following lemma:

LEMMA The weak convergence of distribution functions is metrisized by

(10) $d(F,G) = \int | F(x)-G(x)|\exp(-|x|)dx$

i.e., F_n converges weakly to F iff $d(F_n,F) \to 0$.

Now we are ready for assumption 2.

Assumption 2 It is possible to extend the definition of $\bar{f}(x|\beta)$ so that the range of β will be $R^k \times Q$ and so that for any $\beta_1, \beta_2, \ldots$ and β^* in $R^k \times Q$, $\beta_i \to \beta^*$ implies that $\bar{f}(x|\beta_i) \to \bar{f}(x|\beta^*)$, almost sure $\bar{f}(x|\beta_0)$.

First, the definition of f for (σ,θ) in $R^k \times Q$ is completed by $f(x|\sigma,\theta) = 0$ whenever $|\theta| = \infty$ or if $|\prod_i \sigma_i| = \infty$ (one or more of the σ's are plus or minus ∞). Now, note that f is continuous in σ_i (for all i) and that f is bounded between 0 and 1. Furthermore, weak convergence of F_n to F is equivalent with $E\, h(X_n) \to E\, h(X)$ for all bounded measurable functions h (where E stands for expectation). But this means that $(\sigma_i, G_i) \to (\sigma^*, G^*)$ implies $\bar{f}(x|\beta_i) \to \bar{f}(x|\beta^*)$.

Assumption 3: For any β in $R^k \times Q$ and any $\tau > 0$, $w(x|\beta,\tau)$ is a measurable function of x, where $w(x|\beta,\tau) = \sup \bar{f}(x|\beta')$, the supremum taken over all β' in $R^k \times Q$ for which $\delta(\beta,\beta') < \tau$.

Since $\bar{f}(x|\beta) > 0$ only for countable many values of x, $w(x|\beta,\tau)$ can be replaced by $g(x|\beta,\tau)$ where

$$(11) \qquad g(x|\beta,\tau) = \begin{cases} w(x|\beta,\tau) & \text{when } f(x|\beta) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Now $g(x|\beta,\tau)$ is obviously a measurable function of x (Wald, 1949). Let X be a chance variable with density $\bar{f}(x|\beta_0)$; the operator E will always denote expectation under β_0.

<u>Assumption 4:</u> For any β in $R^k \times Q$ we have

$$(12) \qquad \lim_{\tau \downarrow 0} E \log \left[\frac{\{w(X|\beta,\tau)\}}{\{\bar{f}(x|\beta_0)\}} \right]^+ < \infty$$

This assumption is also called the <u>integrability assumption</u>. Since $w(x|\beta,\tau) < 2$ for all τ, it suffices to prove that $E \log[2/\bar{f}(x|\beta_0)]^+ < \infty$. But this is immediately clear since X can only assume a finite number of values.

The only assumption that needs to be verified is assumption 5, the identifiability assumption. This assumption does not hold for the Rasch model without additional conditions.

<u>Assumption 5:</u> If β_1 in $R^k \times Q$ is different from β_0, then, for at least one y,

$$(13) \qquad \int_{-\infty}^{y} \bar{f}(x|\beta_1)d\mu \neq \int_{-\infty}^{y} \bar{f}(x|\beta_0)d\mu$$

Note that in our case μ is a counting measure, so that in fact the integral-sign is just a sum-sign, and even over a finite number of x's. Therefore, we can change the two "integration" parts. Hence, it is sufficient to prove that $\bar{f}(x|\sigma_0,G_0) = \bar{f}(x|\sigma_1,G_1)$ implies $(\sigma_0,G_0) = (\sigma_1,G_1)$. So, from

$$(14) \qquad \int_R \prod_i \frac{\exp\{x_i(\sigma_{i0}-z)\}}{1+\exp(\sigma_{i0}-z)} dG_0(z) = \int_R \prod_i \frac{\exp\{x_i(\sigma_{i1}-z)\}}{1+\exp(\sigma_{i1}-z)} dG_1(z)$$

we want to conclude that $\sigma_0 = \sigma_1$ and $G_0 = G_1$, where (14) holds for all $(x_1,\ldots,x_k)$ in $\{0,1\}^k$.

Recall that the Rascn model is unidentifiable and that therefore a constraint has to be given. A common one is $\sum_i \sigma_i = 0$. For our purposes it is easier to take the constraint $\sigma_1 = 0$, since then (14) greatly simplifies. Now, let us rewrite and simplify (14). To save space, only the left hand siue of the formulas will be given, since the right hand side can then be written down easily with the obvious changes.

$$(15) \qquad \int_R \prod_i \frac{\exp\{x_i(\sigma_{i0}-z)\}}{1+\exp(\sigma_{i0}-z)}\, dG_0(z) =$$

$$= \exp(\sum_i x_i\sigma_{i0}) \int_R \frac{\exp(-\sum_i x_i z)}{\prod_i \{1+\exp(\sigma_{i0}-z)\}}\, dG_0(z) =$$

$$= \exp(\sum_i x_i\sigma_{i0}) \int \frac{\exp(-\sum_i x_i z)}{h(\sigma_0,z)}\, dG_0(z)$$

Now, let $\sum_i x_i = 1$, or in words, one item has been answered correctly and recall that $\sigma_1 = 0$. Since (14) must hold for all x in $\{0,1\}^k$, we find for $x_1 = 1$:

$$(16) \qquad \int \frac{\exp(-z)}{h(\sigma_0,z)}\, dG_0(z) = \int \frac{\exp(-z)}{h(\sigma_1,z)}\, dG_1(z)$$

and for $x_j = 1$ (all $j \neq 1$)

$$\text{(16a)} \qquad \exp(\sigma_{j0}) \int \frac{\exp(-z)}{h(\sigma_0, z)} dG_0(z) =$$

$$= \exp(\sigma_{j1}) \int \frac{\exp(-z)}{h(\sigma_1, z)} dG_1(z)$$

thus we now conclude that $\sigma_{j0} = \sigma_{j1}$ for all $j=2,\ldots,k$.

This holds for every pair of distribution functions G_0 and G_1. So, (15) can be simplified further; the factors in front of the integrals are equal, and the integrand is a function of the observations only through $\sum_i x_i$. Thus we have the following set of equations (for $\sum_i x_i = 0,1,\ldots,n$):

$$\text{(17)} \qquad \int_R \frac{\exp(-\sum_i x_i z)}{\prod_i \{1+\exp(\sigma_i - z)\}} dG_0(z) = \int_R \frac{\exp(-\sum_i x_i z)}{\prod_i \{1+\exp(\sigma_i - z)\}} dG_1(z) \, ,$$

where $\sigma_{i0} = \sigma_{i1} \equiv \sigma_i$ all i).

Substituting $\sum_i x_i = t$, $\exp(-z) = u \in (0,\infty)$, $\exp(\sigma_i) = \xi_i$ and $dG_j(z) = dH_j(u)$ for j=1,2 we can rewrite (17) as

$$\text{(18)} \qquad \int_0^\infty \frac{u^t}{\prod_i (1+\xi_i u)} dH_0(u) = \int_0^\infty \frac{u^t}{\prod_i (1+\xi_i u)} dH_1(u), \quad t=0,1,\ldots,n$$

We want to conclude $H_0(z) = H_1(z)$ for all non negative z, where H_0 and H_1 are both distribution functions on the nonnegative reals. Note that $\xi_1,\ldots,\xi_n$ are fixed but unknown constants.

Now it is time to look at the set Q of distribution functions G. We started by assuming almost nothing about this set Q : The distribution function had only had to have a finite second moment. But to solve (18) we must put more restrictions on Q. On the other hand, we can not put too many restrictions on Q since then we would overidentify the problem. An example for the latter would be to assume that G is a normal distribution. This would enable us to verify assumption 5, but then there may be no connection to the "real world" anymore. One would like, however, Q to be a large class, containing many distributions of different forms (e.g., normal, χ^2, Poisson).

On the other hand, there are only n independent equations in (18), so that one can not expect to find a solution if the prior distribution function has more than n "parameters". Since there are only n items, it is clear that we cannot hope to estimate a distribution function that has more than n "parameters" because we have for each person only n independent observations.

If we take for Q the class of all discrete distribution function with at most m steps, then there are 2m-1 free parameters (m knots and m-1 weights) and thus if we choose m=(n+1)/2 for n odd and m=(n+2)/2 for n even, then we may fulfill (18) in the largest possible class.

Note that (18) is a moment problem on the positive half line.

Karlin and Studden (1966) showed that if one restricts oneself to canonical distribution functions, the solutions to this problem are step functions: a lower and an upper principal representation. The lower principal representation is the interesting one, since the upper one places mass at infinity. This lower principal solution has (n+1)/2 steps at different points for n odd and (n+2)/2 steps at different points with one point set equal to zero for n even. Since the functions $v_t = u^t \prod_i (1+\xi_i u)^{-1}$ form a Tchebycheff system of Type II ($v_t(u)/v_s(u) \to 0$ for t<s if $u \to 0$). This theory can be applied in this special case. For more details on this matter, see also the Leeuw and Verhelst (1986), who reached the same conclusion from a different starting point.

Discussion

So far this method has been called "new". What is so new then about this method? The strong relationship with nonparametric ML has already been noted. The main difference is that in MML the item parameters are estimated first and thereafter, using these estimates the prior distribution of the person parameters is estimated. The method in this paper shows that it is possible to estimate these two quantities together, and that one gets consistent estimates.

An Example

Table 1 shows the observed frequencies of response patterns to five dichotomously scored items. The data consists of the responses on the subtest measurement of the IEA Second Mathematics Study (Pelgrum et al., 1983).

Insert Table 1 about here

For this data, the Andersen's Conditional Likelihood Ratio Test (Andersen, 1973) gave a chi-square of 3.271 with 4 degrees of freedom, so that the data fits the Rasch model pretty well.

For the distribution of the ability, a step function with 3 knots (values of ξ at which the distribution function makes a jump) was chosen. The probability masses at these points are denoted by w.

In table 2, the CML item parameter estimates are given together with the estimates of the item parameters (σ) and the estimates of the parameters of the distribution function, using the method in this paper.

Insert Table 2 about here

Note that the item parameter estimates obtained by the new method are close to the CML item parameter estimates and that the distribution function is a step function with steps at different points.

References

Andersen, E.B. (1973). Conditional inference and models for measuring. Copenhagen: Mentalhygienjnisk Forlag.

Billingsley, P. (1977). Probability and measure. New York: Wiley.

de Leeuw, J., & Verhelst, N. (1986). Maximum likelihood estimation in generalized Rasch models. Journal of Educational Statistics, 11, 183-196.

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

Karlin, S., & Studden, W.J. (1966). Tchebycheff systems: With applications to analysis and statistics. New York: Wiley.

Kiefer, J., & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Annals of Mathematical Statistics, 27 887-903.

Lord, F.M. (1983). Small N justifies Rasch model. In D. Weiss, (Ed.), New horizons in testing. New York: Academic Press.

Neyman, J. & Scott, E.L. (1948). Constistent estimates based on partially consistent observations. Econometrika, 16, 1-32.

Pelgrum, W.J. Eggen, T.J.H.M., & Plomp, Tj. (1983). Tweede wiskunde project: Beschrijving van uitkomsten [Second mathematics project: Description of results]. Enschede, The Netherlands: Department of Education, University of Twente.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Paedagogiske Institut.

Wald, A. (1948). Estimation of a parameter when the number of unknown parameters increases indefinitely with the number of observations. Annals of Mathematical Statistics, 19, 220-227.

Wellner, J.A. (1985). Semiparametric models: Progress and problems. [CWI Newsletter, No. 9]. Amsterdam: Centrum voor Wiskunde en Informatica.

Table 1

Observed Frequencies for the measurement data

Response Pattern	Frequency	Response Pattern	Frequency
00000	14	10001	3
00010	26	10010	9
00011	3	10011	2
00100	4	10110	7
00110	11	10111	2
01000	10	11000	4
01001	2	11010	21
01010	62	11011	20
01011	22	11100	1
01100	3	11101	1
01101	2	11110	46
01110	28	11111	74
01111	18		

Table 2

Item and ability distribution estimates for the measurement data.

CML	$\hat{\sigma}$	$\hat{\xi}$	$\hat{w}$
.849	.851	-.253	.547
-1.075	-1.079	-29.779	.024
.745	.748	2.259	.429
-1.990	-1.995		
1.472	1.475		

Author's Note

The author is indebted to Norman Verhelst (National Institute of Educational Measurement, Arnhem, The Netherlands) for making the necessary computer program available, and Wim J. van der Linden (University of Twente, Enschede, The Netherlands) for his comments on an earlier version of the paper.

The content of this paper is however fully his own responsibility.

Titles of Recent Research Reports

RR-86-1 W.J. van der Linden, The use of test scores for classification decisions with threshold utility

RR-86-2 H. Kelderman, Item bias detection using the loglinear Rasch model: Observed and unobserved subgroups

RR-86-3 E. Boekkooi-Timminga, Simultaneous test construction by zero-one programming

RR-86-4 W.J. van der Linden, & E. Boekkooi-Timminga, A zero-one programming approach to Gulliksen's random matched subtests method

RR-86-5 E. van der Burg, J. de Leeuw, & R. Verdegaal, Homogeneity analysis with k sets of variables: An altenating least squares method with optimal scaling features

RR-86-6 W.J. van der Linden, & T.J.H.M. Eggen, An empirical Bayes approach to item banking

RR-86-7 E. Boekkooi-Timminga, Algorithms for the construction of parallel tests by zero-one programming

RR-86-8 T.J.H.M. Eggen, & W.J. van der Linden, The use of models for paired comparisons with ties

RR-86-9 H. Kelderman, Common Item Equating Using the Loglinear Rasch Model

RR-86-10 W.J. van der Linden, & M.A. Zwarts, Some Procedures for Computerized Ability Testing

RR-87-1 R. Engelen, Semiparametric Estimation in the Rasch Model

RR-87-2 W.J. van der Linden (Ed.), IRT-based Test Construction

Research Reports can be obtained at costs from Mediatheek, Faculteit Toegepaste Onderwijskunde, Universiteit Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

A publication by
the Department of Education
of the University of Twente,
P.O. Box 217,
7500 AE Enschede,
The Netherlands

Department of Education