



Marginal maximum likelihood estimation in polytomous Rasch models using SAS

Karl Christensen, Maja Olsbjerg

► To cite this version:

Karl Christensen, Maja Olsbjerg. Marginal maximum likelihood estimation in polytomous Rasch models using SAS. Annales de l'ISUP, 2013, 57 (1-2), pp.69-84. hal-03615254

HAL Id: hal-03615254

<https://hal.science/hal-03615254>

Submitted on 21 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Marginal maximum likelihood estimation in polytomous Rasch models using SAS

Karl Bang Christensen and Maja Olsbjerg
 Department of Biostatistics, University of Copenhagen

Abstract: The growing use of Patient Reported Outcomes (PRO's) calls for user friendly software to fit IRT models in standard software like SAS or R. This paper describes a SAS macro %*rasch_mml* that fits polytomous Rasch models. The macro estimates item parameters using marginal maximum likelihood (MML) estimation and person locations using both MLE and Warm's Weighted likelihood estimation (WLE). A number of standard graphical presentations that are useful for investigating the properties of items are included: plots of item characteristic curves (ICC's), person-item location maps (Wright maps) comparing estimates of person locations and item locations, and item and test information functions. A graphical goodness-of-fit-test is also produced.

Keywords: Rasch model, marginal maximum likelihood (MML), goodness of fit, SAS.

1. Introduction

Item response theory (IRT) models are statistical models used for situations where several questions (called items) are used for ordering of a group of subjects with respect to a unidimensional latent variable [1]. The main application have been within educational research where the models describe how the probability of a correct answer to an item in a test depends on ability. The models are, however, applicable in any situation where there is a need for describing the location of persons and items on an underlying latent scale. Patient Reported Outcomes (PRO's) like physical functioning and psychological well-being are typical in applications of IRT models. Traditional applications in education have often used dichotomous (correct/incorrect) item scoring, but polytomous items seem to be the standard in PRO's.

With the growing use of PRO's need for user friendly software to fit IRT models is increasing. This paper describes a SAS macro %*rasch_mml* that fits an IRT model, the polytomous Rasch model [2, 3]. The SAS macro is available from

<http://biostat.ku.dk/~kach>

The macro estimates item parameters and plots item characteristic curves. Moreover it estimates person locations and compares them to the item parameters in a Wright map. Item and test information functions are plotted and a graphical test of fit produced.

K.B. Christensen : kach@sund.ku.dk : University Copenhagen
M.Olsbjerg : mola@sund.ku.dk : University Copenhagen

Several SAS macros for Rasch models already exist. Two macros `%anaqol` [4] and `%irtfit` [5] that encompass a wide range of IRT models and a macro `%rasch` [6] for Rasch models.

The SAS macro `%anaqol` computes Cronbachs coefficient alpha [7], several useful graphical representations and estimates the parameters for any of five IRT models (the dichotomous Rasch model [2, 3], the Birnbaum (2PL) model [8], OPLM, the partial credit model [9], and the rating scale model [10]) using marginal maximum likelihood. The SAS macro `%irtfit` produces a variety of indices for testing the fit of IRT models to dichotomous and polytomous item response data. The macro does not perform estimation of item parameters, but requires that the item parameters have been estimated in IRT model software programs. The SAS macro `%rasch` uses conditional maximum likelihood for parameter estimation in dichotomous and polytomous Rasch models and includes simple graphical model fit diagnostics.

No matter what methods are used for estimation of item parameters and person locations a number of standard graphical presentations are useful for evaluating goodness of fit and also for investigating the properties of the items and utilizing the fact that IRT models place item and person parameters on the same scale.

1.1. Requirements

Before ordering of subjects can be done in a meaningful way a number of requirements must be met

- (i) Items should measure only one latent variable.
- (ii) Items should be increase with the underlying latent variable.
- (iii) Items should be sufficiently different to avoid redundancy.
- (iv) Items should function in the same way in any sub population.

These requirements are standard in educational tests where (i) items should deal with only one subject (e.g. not be a mixture of math and language items), (ii) the probability of a correct answer should increase with ability, (iii) items should not ask the same thing twice and (iv) the difficulty of an item should depend only on the ability of the student, e.g. an item should not have features that makes it easier for boys than for girls at the same level of ability.

Let θ denote the latent variable and let $\bar{X} = (X_i)_{i=1, \dots, I}$ denote the vector of item responses. A mathematical formulation of the two first requirements is the following

- (i) θ is a scalar.
- (ii) $\theta \mapsto E(X_i|\theta)$ is increasing for all items i .

since we would expect two similar items to be highly correlated, and even to be have a higher correlation than the underlying latent variable accounts for it is usual to impose the requirement of *local independence*

- (iii) $P(\bar{X} = \bar{x}|\theta) = \prod_{i=1}^I P(X_i = x_i|\theta)$ for all θ .

however, it should be noted that, while they are related, this requirement is differed from the requirement of non-redundancy.

(iv) $P(X_i = x_i|Y, \theta) = P(X_i = x_i|\theta)$ for all items i and all variables Y .

These requirements are referred to as unidimensionality, monotonicity, local independence and absence of differential item functioning (DIF). Fitting observed data to an IRT model enables us to test if these requirements are met.

The SAS macro %**rasch_mml** presented here extends previously implemented macros in that it the output and features are the same for dichotomous and polytomous item response formats. That it presents more graphics, specifically new goodness-of-fit plot where observed item means are compared item means simulated under the model. The macro presents both MLE and WLE of the person locations, and the number of response options is allowed to vary across items. The macro also contributes by presenting several plots of item and test information functions and by presenting person-item location maps (so called Wright maps).

2. The Polytomous Rasch Model

Consider I items, where item i has $m_i + 1$ response categories represented by the numbers $0, \dots, m_i$. Let X_i be the response to item i with realization x_i . For items $i = 1, \dots, I$ the *polytomous Rasch model* is given by probabilities

$$P(X_i = x_i|\theta) = \exp(x_i\theta + \eta_{ix_i})K_i^{-1} \quad (2.1)$$

where $\bar{\eta}_i = (\eta_{ih})_{h=1, \dots, m_i}$ is the vector of item parameters, $\eta_{i0} = 0$ for all i and

$$K_i = K_i(\theta, \bar{\eta}_i) = \sum_{l=0}^{m_i} \exp(l\theta + \eta_{il})$$

is a normalizing constant. An alternative way of parameterizing is in terms of the thresholds $\beta_{ik} = -(\eta_{ik} - \eta_{ik-1})$, for $i = 1, \dots, I$ and $k = 1, \dots, m_i$ that are easily interpreted, since β_{ik} is the location on the latent continuum where scale where the probability, for item i , of choosing category $k-1$ equals the probability of choosing category k . This model was originally proposed by Andersen [11], see also [10]. Masters [9] called this models the *Partial Credit model* and derived the probabilities (2.1) from the requirement that the conditional probabilities $P(X_i = k|X_i \in \{k-1, k\}; \theta)$, for $k = 1, \dots, m_i$ fit a dichotomous Rasch model:

$$P(X_i = k|X_i \in \{k-1, k\}; \theta) = \frac{\exp(\theta - \beta_{ik})}{1 + \exp(\theta - \beta_{ik})}.$$

The thresholds (β 's) are easier to understand, but parameterizing by the η_{ih} 's is sometimes used.

2.1. Likelihood

Using the assumption (iii) of local independence the vector $\bar{X} = (X_i)_{i=1,\dots,I}$ with realization $\bar{x} = (x_i)_{i=1,\dots,I}$

$$\begin{aligned} P(\bar{X} = \bar{x} | \theta) &= \exp\left(\sum_{i=1}^I (x_i \theta + \eta_{ix_i})\right) K(\theta)^{-1} \\ &= \exp(r\theta) \exp\left(\sum_{i=1}^I \eta_{ix_i}\right) K(\theta)^{-1} \end{aligned} \quad (2.2)$$

where $r = \sum_{i=1}^I x_i$ and $K(\theta) = \prod_{i=1}^I K_i(\theta, \bar{\eta}_i)$. By Neyman's factorization theorem it is clear from (2.2) that the sum of item responses $R = \sum_{i=1}^I X_i$ is sufficient for θ . The joint log likelihood for a sample of $v = 1, \dots, N$ persons is given by

$$l(\bar{\eta}_1, \dots, \bar{\eta}_I; \bar{\theta}) = \sum_{v=1}^N R_v \theta_v + \sum_{v=1}^N \sum_{i=1}^I \eta_{ix_{vi}} - \sum_{v=1}^N \log K(\theta_v) \quad (2.3)$$

where $\bar{\theta} = (\theta_1, \dots, \theta_N)^T$. Jointly estimating all parameters from (2.3) does not provide consistent estimates, since the number of parameters increase with the sample size. If our interest is estimating the item parameters the person parameters can be interpreted as incidental or nuisance parameters [12] and for this reason marginal or conditional estimation of item parameters is used.

3. Marginal Maximum Likelihood (MML) estimation of item parameters

This section describes marginal maximum likelihood (MML) estimation of item parameters and its implementation in SAS. Let $\bar{X}_v = (X_{vi})_{i=1,\dots,I}$ denote the response vector of individual v and $\bar{\eta}_i = (\eta_{ih})_{h=1,\dots,m_i}$ the item parameters of item i .

3.1. Marginal likelihood

Assuming that θ has some distribution with density g_ω where ω denotes a set of parameters determining the distribution. Then the marginal log likelihood is

$$l_M(\bar{\eta}_1, \dots, \bar{\eta}_I, \omega) = \sum_{v=1}^N \left(\sum_{i=1}^I \eta_{ix_{vi}} + \log \int \frac{\exp(r_v \theta)}{K(\theta)} g_\omega(\theta) d\theta \right) \quad (3.1)$$

where g_ω is typically chosen to be a normal density. If the distribution is correctly specified MLE of item parameters are consistent [13], but if the distribution is not normal then marginal ML estimates are not consistent. The

integral in (3.1) cannot be calculated explicitly and has to be approximated. The NLMIXED procedure in SAS fits nonlinear mixed models by maximizing an approximation to the likelihood integrated over the random effects. Different integral approximations are available, the principal one being adaptive Gaussian quadrature.

3.1.1. Implementation in SAS

The NLMIXED procedure in SAS can be used to maximize an approximation to the likelihood integrated over the random effects. The procedure uses an adaptive Gaussian quadrature, found to be one of the best methods in a comparison of several different integrated likelihood approximations [14]. The conditional distribution (given the random effects) can be specified to be a general distribution using SAS programming statements. Examples of how to do this for the polytomous Rasch model have been presented earlier [15, 4]. The macro %*rasch_mml* uses NLMIXED and writes threshold parameter estimates $\hat{\beta}_{ih}$ and their asymptotic standard errors in a data set.

3.2. Estimating population parameters

This model can be extended by imposing a linear structure

$$\begin{bmatrix} \theta_1 \\ \vdots \\ \theta_N \end{bmatrix} = \bar{\bar{Y}}\bar{\eta} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{bmatrix} \quad (3.2)$$

given by a design matrix $\bar{\bar{Y}}q$ with one row for each person and a vector $[\epsilon_1 \ \dots \ \epsilon_N]^T \sim MVN(0, \sigma^2 I_N)$ of residuals. Models with this structure are often called *latent regression models* and have been discussed by many authors for dichotomous [16, 17, 18, 19, 20] as well as polytomous [21, 22] items. Estimating the population distribution however does not tell us anything about the location of individuals.

4. Estimation of person locations

There are various ways of estimating the person locations. An important feature of the Rasch model is that the sum score $R = \sum_{i=1}^I X_i$ is sufficient for θ and consequently that the likelihood function for estimating θ_v is proportional to the probabilities

$$\begin{aligned} P(R = r|\theta) &= \sum_{\bar{x} \in A_r} P(\bar{X} = \bar{x}|\theta) \\ &= \exp(r\theta)K(\theta)^{-1} \sum_{\bar{x} \in A_r} \exp\left(\sum_{i=1}^I \sum_{h=0}^{m_i} \eta_{ix_i}\right) \end{aligned}$$

where summation is over the set $A_r = \{\bar{x} | \sum_{i=1}^I x_i = r\}$ of all response vectors with sum r . Now, define the γ -polynomials

$$\gamma_r = \gamma_r(\bar{\eta}_1, \dots, \bar{\eta}_I) = \sum_{\bar{x} \in A_r} \exp \left(\sum_{i=1}^I \sum_{h=0}^{m_i} x_{ih} \eta_{ih} \right)$$

to obtain the expression

$$P(R = r | \theta) = \exp(r\theta) \gamma_r K(\theta)^{-1}. \quad (4.1)$$

Note from this that the normalizing constant $K(\theta)$ can be written as a function of the γ 's

$$K(\theta) = \prod_{i=1}^I \sum_{k=0}^{m_i} \exp(k\theta + \eta_{ik}) = \sum_{r=0}^R \exp(r\theta) \gamma_r.$$

Calculation of the γ 's is thus essential for estimation of the person locations. A recursion formula is described in the following subsection.

4.1. Computation of γ functions

Let $\gamma_r^{(i)}$ denote the γ -polynomial based on the first i items. It is then possible to calculate $\gamma_t^{(i+1)}$ by the recursion formula

$$\gamma_r^{(i+1)} = \sum_x \exp(\eta_{i+1,x}) \gamma_{r-x}^{(i)}$$

since a total score of r on the items $1, \dots, i+1$ must be obtained by scoring x on item $i+1$ and $r-x$ on the items $1, \dots, i$. The values of x in the summation in the formula above must be chosen in such a way that the sum of the first items is at most $\sum_{k=1}^i m_k$ and that item i is at most m_i implying that r can not exceed m_i . That is $\gamma_t^{(i)}$ becomes

$$\gamma_r^{(i+1)} = \sum_{x=r-\sum_{k=1}^i m_k}^{\min(m_{i+1}, r)} \exp(\eta_{i+1,r-x}) \gamma_{r-x}^{(i)}. \quad (4.2)$$

4.2. Maximum likelihood estimation of the person locations

Since the γ 's do not depend on θ (4.1) is an exponential family where the likelihood equation for estimating θ is

$$R = \frac{\partial}{\partial \theta} \left(\log \left[\sum_{r=0}^R \exp(r\theta) \gamma_r \right] \right) = E(R | \theta)$$

and the maximum likelihood estimator (MLE) $\hat{\theta}$ can be obtained by the Newton-Raphson algorithm. The probabilities (4.1) show that the score is increasing as a function of θ . For individuals who have obtained scores of zero or the largest possible score $R = \sum_{i=1}^I m_i$ the probabilities (4.1) attain their maximum when θ is $-\infty$ and ∞ , respectively.

4.3. Bayes modal estimation of the person locations

The Bayes modal Estimator (BME) of θ is obtained by choosing a prior density g for the latent parameter and then maximizing the posterior density

$$g_x(\theta) = \frac{P(X = x|\theta)g(\theta)}{P(X = x)} \propto P(X = x|\theta)g(\theta) \quad (4.3)$$

w.r.t. θ keeping item parameters and the observations fixed. The MLE described above is a special case corresponding to $g_\omega \equiv 1$.

4.4. Weighted likelihood estimation of the person locations

Choosing the prior as the square root of the Fisher information

$$g(\theta) = \sqrt{\mathcal{I}(\theta)}$$

results in the *weighted maximum likelihood* estimator (WLE) [23]. With this prior one obtains an estimator with minimal bias and the same asymptotic distribution as the MLE. The equation to be solved in order to obtain the WLE is

$$R = \frac{\partial}{\partial \theta} \left(\log \left[\sum_{r=0}^R \exp(R\theta) \gamma_r \right] - \frac{1}{2} \log \mathcal{I}(\theta) \right).$$

and the Newton-Raphson algorithm can be used for this.

4.5. Implementation in SAS

The SAS macro %*rasch_mml* uses NLMIXED and writes person locations estimated by maximum likelihood estimation (MLE) and by weighted likelihood estimation (WLE) and their asymptotic standard errors in a data set. Furthermore a copy of input data set with an added column containing the maximum likelihood estimates is created.

5. Simulation

When evaluating model fit it is helpful to compare what has been observed with simulated data that describe what could have been observed under the model. The SAS macro %*rasch_mml* simulates data sets under the model. These are obtained by first simulating N person locations from the empirical distribution of the $\hat{\theta}_v$'s and then simulating item responses. Let $L = \{\hat{\theta}_v | v = 1, \dots, N\}$ denote the set of estimated locations and for $\theta \in L$ define $V_\theta = \{v | \theta_v = \theta\} \subset \{1, \dots, N\}$. Let $N_\theta = \#V_\theta$ denote the number of persons with each estimated

location. First simulate $\theta_1^{(S)}, \dots, \theta_N^{(S)}$ with probabilities $P(\theta_v^{(S)} = \theta) = \frac{N_\theta}{N}$, next simulate $X_{vi}^{(S)}$ using the probabilities

$$P(X_{vi}^{(S)} = x) = \exp(x\theta_v^{(S)} + \hat{\eta}_{ix})K_i^{-1} \quad (5.1)$$

this procedure is repeated a number of times yielding data matrices $(X_{vi}^{(S)})$, $S = 1, 2, \dots$

6. Graphics

A number of graphical representations are made by the SAS macro %*rasch_mml*. These comprise item characteristic curves, item and test information functions, Wright maps and item fit plots.

6.1. Item characteristic curves

Item characteristic curves (ICC's) display the response probabilities along the latent continuum.

6.2. Item and test information functions

In IRT it is standard to evaluating an items contribution in estimation of θ using the *item information function*

$$\mathcal{I}_i(\theta) = V(X_i|\theta) \quad (6.1)$$

that is easily computed by inserting estimates of $\bar{\eta}_i$ in (2.1). The item information is large for values of θ near thresholds. For so that for instance easy items works well in the estimation for people with low values of θ . An overall measure is achieved by summation resulting in the *test information function*

$$\mathcal{I}(\theta) = \sum_{i=1}^I \mathcal{I}_i(\theta) = \sum_{i=1}^I V(X_i|\theta) = V\left(\sum_{i=1}^I X_i \middle| \theta\right). \quad (6.2)$$

where the last equal sign is due to the assumption (iii) of local independence.

6.3. Wright maps

Item and test information functions are useful for evaluating if a set of items is suitable for a specific population. The *Wright map* or *person-item location map* also compares item and person locations.

6.4. Item fit plots

Beyond ensuring that items are well-targeted to the population under consideration, evaluation of model fit is crucial. Many fit statistics exist [24], but the issue of fit should also be addressed graphically. The observed mean score function for item i is

$$\theta \mapsto \frac{1}{N_\theta} \sum_{v \in V_\theta} X_{vi}.$$

The simulated mean score function is obtained as follows by simulating item responses $X_{1i}^{(S)}, \dots, X_{Ni}^{(S)}$ as described in section 5 and calculating

$$\theta \mapsto \frac{1}{N_\theta^{(S)}} \sum_{v \in V_\theta} X_{vi}^{(S)}.$$

where $N_\theta^{(S)} = \#\{v | \hat{\theta}^{(S)} = \theta\}$.

7. Goodness of fit

Infit and outfit tests statistics [25] use standardized differences between observed and expected item responses to evaluate the fit of items. The statistical properties of the outfit and infit test statistics are not known, but transformed measures are often assumed to have approximate scaled chi-squared distributions.

7.1. Infit and outfit test statistics

For an item i we consider the mean $E_{iv} = \sum_{k=0}^{m_i} kP(X_{iv} = k|\theta_v)$, and the variance $V_{iv} = \sum_{k=0}^{m_i} k^2 P(X_{iv} = k|\theta_v) - E_{iv}^2$. Infit and outfit test statistics are based on the standardized residuals

$$Z_{iv} = \frac{X_{iv} - E_{iv}}{\sqrt{V_{iv}}}. \quad (7.1)$$

The infit test statistic for item i is the weighted mean square

$$INFIT_i = \frac{\sum_{v=1}^N V_{iv} Z_{iv}^2}{\sum_{v=1}^N V_{iv}} \quad (7.2)$$

and the outfit test statistic for item i is the unweighted mean square

$$OUTFIT_i = \frac{1}{N} \sum_{v=1}^N Z_{iv}^2 = \frac{1}{N} \sum_{v=1}^N \frac{(X_{iv} - E_{iv})^2}{V_{iv}}. \quad (7.3)$$

Person fit statistics based on standardized residuals have also been suggested. The infit test statistic for person v is $(\sum_{i=1}^I V_{iv} Z_{iv}^2) / (\sum_{i=1}^I V_{iv})$ and the outfit test statistic for person v is $\frac{1}{I} \sum_{i=1}^I Z_{iv}^2 = \frac{1}{I} \sum_{i=1}^I (X_{iv} - E_{iv})^2 / V_{iv}$.

7.2. Other tests

With the simulation feature implemented several other tests are possible. A test of the assumption (iii) of local independence can be based on computing the observed association for an item pair and comparing this to simulated values.

8. The SAS macro

The SAS macro is illustrated using data reported by Pallant and Tennant [26] from the Hospital Anxiety and Depression Scale (HADS). Designed as a brief instrument used to assess symptoms of anxiety and depression [27] the HADS contain 14 items often scored as two seven-item sub scales: 'Depression' (even numbered items) and 'Anxiety' (odd numbered items). To analyze the latter create a SAS data set

```
data inames;
  input item_name $ item_text $ max @@;
  cards;
  AHADS1 anx1 3 AHADS3 anx3 3 AHADS5 anx5 3
  AHADS7 anx7 3 AHADS9 anx9 3 AHADS11 anx11 3
  AHADS13 anx13 3
  ;
run;
```

that describes the items. Here `item_name` is the name of the items, `item_text` are text strings attached to the items, and `max` is the maximum item score for each item. The HADS items are scored 0, 1, 2, 3. Note that the maximum score can vary across items. The macro is called using the statement

```
%rasch_mml(DATA=work.HADS, ITEM_NAMES=inames, OUT=HADSTEST);
```

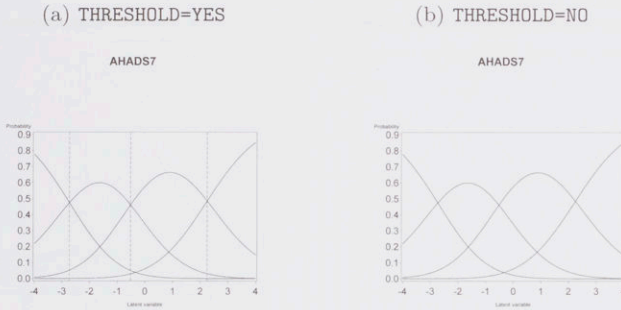
where `DATA=` specifies the data set to be analyzed, `ITEM_NAMES=` is the data set that describes the items and `OUT=` specifies a prefix for all output data sets generated by the macro (default MML). SAS creates a data set `HADSTEST_ipar` that contains item parameter estimates and their standard errors and a data set `HADSTEST_latent` that contains MLE and WLE estimates of person locations and their standard errors.

```
proc print data=hadstest_ipar round noobs; run;
```

yields the (edited) output

Label	Estimate	Standard Error
AHADS1 1	-3.4822	0.3213
AHADS1 2	0.4885	0.1699
AHADS1 3	1.1290	0.2068
:		
AHADS7 1	-2.7261	0.2817
AHADS7 2	-0.5178	0.1635
AHADS7 3	2.2513	0.2237

FIG 1. Item characteristic curves (ICC's) for HADS item 7 plotted with or without the option *THRESHOLD=YES* that creates vertical dashed lines at the thresholds $\beta_{71} = -2.7261$, $\beta_{72} = -0.5178$, and $\beta_{73} = 2.2513$.



these are the threshold parameters (β 's). The data set `HADSTEST_latent` contains estimated person locations for each value of the total score and the data set `HADSTEST_outdata` is a copy of the input data with an added variable `theta_estimate` containing the estimated location. The macro also creates a data set `HADSTEST_logl` with the likelihood value and a data set `HADSTEST_ppar` with the population parameter (the variance σ^2).

Several further options can be specified: The option `ICC=YES` yields a plot of the item characteristic curves for each item. The ICC's for HADS item 7 is shown in Figure 1. An extra option `THRESHOLD=YES` indicates that ICC's are to be plotted together with dashed lines indicating threshold locations.

The option `INFORMATION=YES` plots item and test information functions. For each item the item information function is plotted along with the ICC's (shown for HADS item 7 in Figure 2, (a)) and the test information function is plotted together with all item information functions (Figure 2, (b)).

The option `WRIGHT=YES` makes the macro plot histograms of item and person locations (cf. Figure 3).

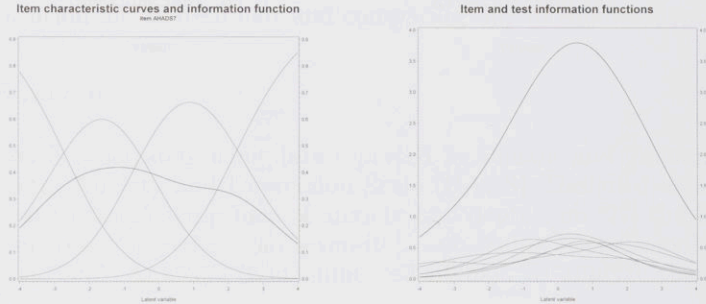
Using the option `FITPLOT=YES` makes the macro plot item means against $\hat{\theta}$ as solid black lines along with item means simulated under the model plotted as gray dashed lines. The default number of simulations is 30, but this can be changed using the `NSIMU=` option. Figure 4 shows an example.

When some person locations appear infrequently a coarsening $L_0 \subset L$, such that $N_{\theta} > \text{CLASS_SIZE}$ for all $\theta \in L_0$ can be used. The default value is `CLASS_SIZE=15`.

Specifying `FITTEST=YES` makes the macro compute the standardized residuals (7.1) along with the INFIT test statistic (7.2) and the outfit test statistic (7.3). The macro then simulates `NSIMU_FITTEST` data sets (default value 100) to provide confidence intervals for these and saves this to a data set `HADSTEST_fit`. The macro also computes the association measure γ for all item pairs and compares this to simulated values (output data set `HADSTEST_ld`). The standardized

FIG 2. Item and test information functions for the HADS items.

- (a) Item information function (HADS item 7) (b) Test information function



residuals (7.1) are saved in the data set `HADSTEST_residuals`. The simulation process can be time consuming.

9. Discussion

Several proprietary software packages for fitting Rasch models exist, the most widely used being RUMM [28], Conquest [29], and WINSTEPS [30]. With the increasing use of IRT and Rasch models in new research areas like PRO's where access to specialized proprietary software is limited it is important to provide implementations in standard software such as R and SAS. The R package `eRm` [31] is a flexible tool for these analyses, as are the SAS macros `%anaqol` [4] and `%rasch` [6].

The SAS macro `%irtfit` encompasses a wide range of IRT models, but does not estimate item parameters. The SAS macro `%anaqol` is very useful, but some features are only available for dichotomous items and the implemented plots of empirical and theoretical ICC's do not show confidence limits. The SAS macro `%rasch` plots observed and expected counts in each score group, and these plots can be interpreted as empirical versions of the item characteristic curves. However when many score groups are small, as is often the case in applications these plots are not helpful. Furthermore the macro `%rasch` does not produce a single item-level goodness-of-fit plot, does not provide item or test information functions. The SAS macro `%rasch_mml` presented here extends previously implemented macros in that it the output and features are the same for dichotomous and polytomous item response formats. That it presents more graphics, specifically new goodness-of-fit plot where observed item means are compared item means simulated under the model. The macro presents both MLE and WLE of the person locations, and the number of response options is allowed to vary across items. The macro also contributes by presenting several plots of item and test information functions and by presenting person-item location maps (so

FIG 3. Person-item location (Wright) map showing the person and item locations on the common scale.

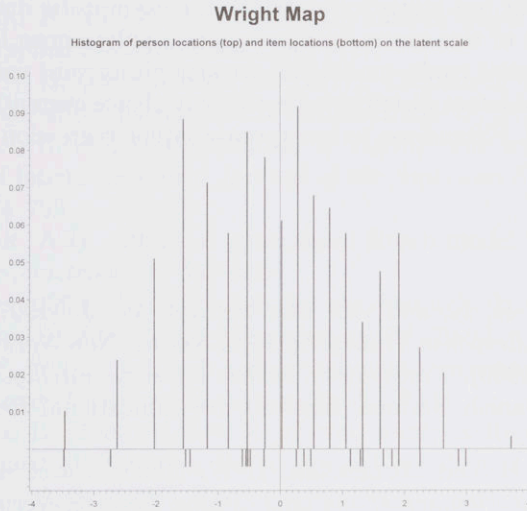
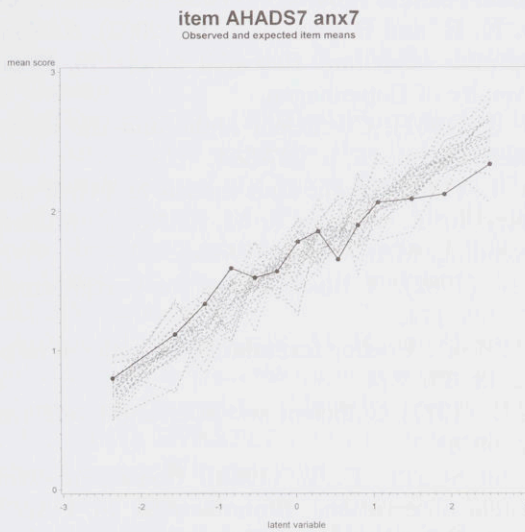


FIG 4. Item fit plot for HADS item 7.



called Wright maps).

The implementation of simulation in the macro appears promising, because it enables the use of fit test statistics for which the asymptotic distribution is not known. Extensions of this macro to more general polytomous IRT models like the generalized partial credit model [32] are straightforward using NLMIXED, however the normal latent distribution is the only choice currently supported by PROC NLMIXED. Extensions to incorporate latent regression is also feasible in NLMIXED.

References

- [1] HAMBLETON, R. K. and VAN DER LINDEN, W. J. (1997). *Handbook of Modern Item Response Theory*. Springer-Verlag, New York.
- [2] RASCH, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish National Institute for Educational Research, Copenhagen.
- [3] FISCHER, G. H. and MOLENAAR, I. W. (1995). *Rasch models: Foundations, recent developments, and applications*. Springer-Verlag, New York.
- [4] HARDOUIN, J.-B. and MESBAH, M. (2007). The SAS Macro-Program %AnaQol to Estimate the Parameters of Item Responses Theory Models. *Communications in Statistics Simulation and Computation*, 36, 437–453.
- [5] BJORNER, J. B., SMITH, K. J., STONE, C. and Sun, X. IRT-FIT: A Macro for Item Fit and Local Dependence Tests under IRT Models. Lincoln RI: QualityMetric Incorporated 2007. [http://outcomes.cancer.gov/areas/measurement/irtfit_macro_users_guide]
- [6] CHRISTENSEN, K. B. and BJORNER, J. B. (2003). *SAS macros for Rasch based latent variable modelling*. Research report 13, Department of Biostatistics, University of Copenhagen.
- [7] CRONBACH L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, vol. 16, p. 297–334.
- [8] BIRNBAUM, A. L. (1968). latent trait models and their use in inferring an examinees ability. In: F. M. Lord & M. R. Novick *Statistical Theories of Mental Test*. Reading, MA: Addison-Wesley.
- [9] MASTERS, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- [10] ANDRICH, D. (1978a). A rating formulation for ordered response categories. *Psychometrika*, 43, 357–374.
- [11] ANDERSEN, E.B. (1977). Sufficient statistics and latent trait models. *Psychometrika*, 42, 69–81.
- [12] NEYMAN, J. and SCOTT, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, 16, 1–32.
- [13] ZWINDERMAN, A. H. and VAN DEN WOLLENBERG, A. L. (1990). Robustness of Marginal Maximum Likelihood Estimation in the Rasch Model. *Applied Psychological Measurement*, 14, 73–81.
- [14] PINHEIRO, J. C. and BATES, D. M. (1995). Approximations to the log-

- likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4, 12–35.
- [15] CHRISTENSEN, K. B. (2006). Fitting Polytomous Rasch Models in SAS. *Journal of Applied Measurement*, 7, 407–417.
- [16] ANDERSEN, E. B. and MADSEN, M. (1977). Estimating the parameters of the latent population distribution. *Psychometrika*, 42, 357–74.
- [17] SANATHANAN, L. and BLUMENTHAL, S. (1978). The logistic model and estimation of latent structure. *Journal of the American Statistical Association*, 73, 794–799.
- [18] ZWINDERMAN, A.H. (1991). A generalized Rasch model for manifest predictors. *Psychometrika*, 56, 589–600.
- [19] KAMATA, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*, 38, 79–93.
- [20] MAIER, K. S. (2001). A Rasch hierarchical measurement model. *Journal of Educational and Behavioral Statistics*, 26, 307–330.
- [21] ANDERSEN, E.B. (1994). *Latent Regression Analysis*. Research Report no. 106, Department of Statistics, University of Copenhagen.
- [22] ZWINDERMAN, A.H. (1997). Response models with manifest predictors. In: van der Linden & Hambleton (editors) *Handbook of modern item response theory*, Berlin: Springer-Verlag, 245–258.
- [23] WARM, T. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427–450.
- [24] GLAS, C. A. W. and VERHELST, N. D. (1995). Tests of Fit for Polytomous Rasch Models. In: Fischer & Molenaar (editors) *Rasch Models - Foundations, Recent Developments, and Applications*. Berlin: Springer-Verlag, 325–352.
- [25] WRIGHT, B. D. and MASTERS, G. N. (1982). *Rating Scale Analysis*. MESA Press, Chicago.
- [26] PALLANT, J. and TENNANT, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, 46, 1–18.
- [27] ZIGMOND, A. S. and SNAITH, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, 67, 361–370.
- [28] ANDRICH, D., LYNE, A., SHERIDAN, B. and LUO, G. (2003). *RUMM 2020*. Perth: RUMM Laboratory.
- [29] WU, M. L., ADAMS, R. J., WILSON, M. R. and HALDANE, S.A. (2007). *ACER ConQuest Version 2: Generalised item response modelling software*. Camberwell: Australian Council for Educational Research.
- [30] LINACRE, J. M. (2011) *WINSTEPS Rasch measurement computer program*. Beaverton, Oregon: Winsteps.com
- [31] MAIR, P. and HATZINGER, R. (2007). Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R *Journal of Statistical Software* 20 9.
- [32] MURAKI, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, 16, 159–176.