Conditional maximum likelihood estimation in polytomous Rasch models using SAS

Karl Bang Christensen
kach@sund.ku.dk
Department of Biostatistics, University of Copenhagen

November 29, 2012

Abstract

IRT models are widely used, but often relies on distributional assumptions about the latent variable. For a simple class of IRT models, the Rasch models, conditional inference is feasible. This enables consistent estimation of item parameters without reference to the distribution of the latent variable in the population. Traditionally, specialized software has been needed for this, but conditional maximum likelihood estimation can be done using standard software for fitting generalized linear models. This paper describes a SAS macro `%rasch_cml` that fits polytomous Rasch models. The macro estimates item parameters using conditional maximum likelihood (CML) estimation and person locations using MLE and Warm's Weighted likelihood estimation (WLE). Graphical presentations are included: plots of item characteristic curves (ICC's) and a graphical goodness-of-fit-test is also produced.

Keywords: Rasch model, conditional maximum likelihood (CML), goodness of fit, SAS.

1 Introduction

Item response theory (IRT) models were developed to describe probabilistic relationships between correct responses on a set of test items and continuous latent traits [1]. In addition to educational and psychological testing, IRT models have been also used in other areas of research, e.g. in health status measurement and evaluation of Patient Reported Outcomes (PRO's) like physical functioning and psychological well-being are typical in applications of IRT models. Traditional applications in education often use dichotomous (correct/incorrect) item scoring, but polytomous items are common in other applications.

Formally IRT models deal with the situation where several questions (called items) are used for ordering of a group of subjects with respect to a unidimensional latent variable. 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 redundance.

(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 $\overline{X} = (X_i)_{i=1,\ldots,I}$ denote the vector of item responses. The two first requirements can be written

  (i) θ is a scalar.

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

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

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

This requirement is related to the requirement of non-redundancy. The firth requirement can be written

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

The requirements (i)- (iv) are referred to as unidimensionality, monotonicity, local independence and absence of differential item functioning (DIF), respectively. Fitting observed data to an IRT

model enables us to test if these requirements are met. Evaluation of model fit is crucial and many fit statistics exist [4], but the issue of fit can also be addressed graphically.

This paper describes a SAS macro `%rasch_cml` that fits an IRT model, the polytomous Rasch model [2, 3]. The SAS macro is available from

$$\texttt{http:\textbackslash\textbackslash biostat.ku.dk\textbackslash\~{}kach.}$$

it estimates item parameters, plots item characteristic curves, estimates person locations, and produces graphical tests of fit.

2 The Polytomous Rasch Model

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

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

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

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

a normalizing constant. An alternative way of parameterizing is in terms of the thresholds $\beta_{ik} = -(\eta_{ik} - \eta_{ik-1})$, for $i = 1, \ldots, I$ and $k = 1, \ldots, m_i$ 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 [5], see also [24]. Masters [23] called this models the *Partial Credit model* and derived the probabilities (1) from the requirement that the conditional probabilities $P(X_i = k | X_i \in \{k - 1, k\}; \theta)$, for $k = 1, \ldots, m_i$ fit a dichotomous Rasch model:

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

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

$$
\begin{aligned}
P(\overline{X} = \overline{x} | \theta) &= \exp(\sum_{i=1}^{I} (x_i \theta + \eta_{ix_i})) K(\theta)^{-1} \\
&= \exp(r\theta) \exp(\sum_{i=1}^{I} \eta_{ix_i}) K(\theta)^{-1} \tag{2}
\end{aligned}
$$

where $r = \sum_{i=1}^{I} x_i$ and $K(\theta) = \prod_{i=1}^{I} K_i(\theta, \overline{\eta}_i)$. By Neyman's factorization theorem it is clear from (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 $v = 1, ..., N$ persons is given by

$$l(\overline{\eta}_1, \ldots, \overline{\eta}_I; \overline{\theta}) = \sum_{v=1}^{N} R_v \theta_v + \sum_{v=1}^{N} \sum_{i=1}^{I} \eta_{ix_{vi}} - \sum_{v=1}^{N} \log K(\theta_v) \tag{3}$$

3

where $\overline{\theta} = (\theta_1, \ldots, \theta_N)^T$. Jointly estimating all parameters from (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 [6].

3 Conditional maximum likelihood estimation

The joint log likelihood function (3) can be written

$$l(\overline{\eta}_1, \ldots, \overline{\eta}_I; \overline{\theta}) \quad = \quad \sum_{v=1}^{N} \theta_v R_v + \sum_{i=1}^{I} \sum_{h=1}^{m_i} C_{ih} - \log K \tag{4}$$

where $K = \prod_{v=1}^{N} \prod_{i=1}^{I} K_i(\theta_v, \eta_i)$, $r_v = \sum_{i=1}^{N} x_{vi}$ is the total score observed for person $v = 1, \ldots, N$ and $(C_{ih})_{i=1,\ldots,I; h=1,\ldots,m_i}$ defined by

$$C_{ih} = \sum_{v=1}^{N} 1_{(X_{vi}=h)}$$

are the item margins. Note that from (??) it can be seen that the total score R_v is sufficient for the person location θ_v and that for each $i = 1, \ldots, I$ the item margin $(C_{ih})_{h=1,\ldots,m_i}$ is sufficient for the item parameter η_i.

Restrictions are needed to ensure that the model (??) is identified since from (1) it is clear that for all (θ, η_i)
$$P(X_{vi} = x_{vi}|\theta, \eta_i) = P(X_{vi} = x_{vi}|\theta^*, \eta_i^*)$$
for (θ^*, η_i^*) defined by $\theta^* = \theta - k$ and

$$\eta_{ih}^* = \eta_{ih} + kh \tag{5}$$

for $h = 1, \ldots, m_i$.

To obtain consistent item parameters estimates marginal or conditional maximum likelihood estimation is used. The marginal approach to item parameter estimation assumes that the latent variables are sampled from a population and introduces an assumption about the distribution of the latent variable. The sufficiency property can also be used to overcome the problem of inconsistency of item parameter estimates. This can be done by conditioning on the sum R_v of the entire response vector $\overline{X}_v = (X_{v1}, \ldots, X_{vk})$ yielding conditional maximum likelihood (CML) inference. For a vector $\overline{X}_v = (X_{v1}, \ldots, X_{vk})$ from the Rasch model, the distribution of the score $R_v = \sum_{i=1}^{k} X_{vi}$ is given by the probabilities

$$\Pr(R_v = r|\theta) \quad = \quad \sum_{\overline{x} \in X^{(r)}} \Pr(\mathbf{X} = \mathbf{x}|\theta)$$

$$= \quad \sum_{\overline{x} \in X^{(r)}} \frac{\exp(r\theta + \sum_{i=1}^{k} \eta_{ix_i})}{\prod_{i=1}^{k} K_i(\eta_i, \theta)}$$

where summation is over the set $X^{(r)}$ of all response vectors $\overline{x} = (x_1, \ldots, x_k)$ with $\sum_{i=1}^{k} x_i = r$. The probability can be written

$$\Pr(R_v = r|\theta) = \frac{\exp(r\theta)}{\prod_{i=1}^{k} K_i(\eta_i, \theta)} \sum_{\overline{x} \in X^{(r)}} \exp(\sum_{i=1}^{k} \eta_{ix_i}).$$

Let the last sum be denoted by

$$\gamma_r = \gamma_r(\eta_1, \ldots, \eta_k) = \sum_{\overline{x} \in X^{(r)}} \exp\left(\sum_{i=1}^{k} \eta_{ix_i}\right).$$

The score is sufficient for θ and the item parameters can be estimated consistently using the conditional distribution of the responses given the scores. The conditional distribution of the vector $\overline{X}_v = (X_{v1}, \ldots, X_{vk})$ of item responses given the score is given by the probabilities

$$\Pr(\overline{X}_v = \overline{x}_v|R_v = r, \Theta_v = \theta_v) = \frac{\exp\left(\sum_{i=1}^{k} \eta_{ix_{vi}}\right)}{\gamma_r}.$$

These do not depend on the value of θ_v and the conditional likelihood function is the product

$$L_C(\eta_1, \ldots, \eta_k) = \prod_{v=1}^{n} \frac{\exp(\sum_{i=1}^{k} \eta_{ix_{vi}})}{\gamma_{r_v}}.$$

Again a linear restriction on the parameters is needed in order to ensure that the model is identified. Maximizing this likelihood yields item parameter estimates which are conditionally consistent. If, for each possible response vector $\overline{x} = (x_1, \ldots, x_k)$, we let $n(\overline{x})$ denote the number of persons with this response vector and for each possible score r and $n(r)$ denote the observed number of persons with this value of the score this likelihood function can be written

$$L_C(\eta_1, \ldots, \eta_k) = \frac{\prod_{\overline{x}} \exp(n(\overline{x}) \sum_{i=1}^{k} \eta_{i,x_i})}{\prod_r \gamma_r^{n(r)}}$$

and, using the indicator functions $(I_{vih})_{v=1,\ldots,n;i=1,\ldots,k;h=1,\ldots,m}$, this likelihood function can be rewritten

$$L_C(\eta_1, \ldots, \eta_k) = \frac{\prod_{v=1}^{n} \exp(\sum_{i=1}^{k} \sum_{h=1}^{m} X_{vih} \eta_{ih})^{n(r)}}{\prod_r \gamma_r}$$

yielding the conditional log likelihood function

$$l_C(\eta_1, \ldots, \eta_k) = \sum_{i=1}^{k} \sum_{h=1}^{m} X_{.ih} \eta_{ih} - \sum_{r=0}^{km} n_r \log(\gamma_r)$$

where $X_{.ih} = \sum_{v=1}^{n} X_{vih}$ are the sufficient statistics for the item parameters. These sufficient statistics, called item margins, are the number of persons giving the response h to item i. The item parameters in this model can be estimated by solving the likelihood equations that equate the sufficient statistics $(X_{.ih})_{i=1,\ldots,k;h=1,\ldots,m}$ to their expected values conditional on the observed value $\overline{r} = (r_1, \ldots, r_n)$ of the vector $\overline{R} = (R_1, \ldots, R_n)$ of scores. These expected values have the form

$$E(X_{.ih}|\overline{R} = \overline{r}) = \sum_{v=1}^{n} \Pr(X_{vi} = h|R_v = r_v)$$

5

and for an item i these can be written in terms of the probabilities of having a score of $r - h$ on the remaining items yielding

$$E(X_{.ih}|\overline{R} = \overline{r}) = \exp(\eta_{ih}) \sum_{r=0}^{km} n(r) \frac{\gamma_{r-h}^{(i)}}{\gamma_r}.$$

Because these likelihood equations have the same form as those in a generalized linear model [7, 8, 9] the item parameters can be estimated using standard software like SPSS [10] or SAS [11].

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

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

where, as before, summation is over the $X^{(r)} = \{\overline{x}| \sum_{i=1}^{I} x_i = r\}$ of all response vectors with sum r. Now, define the γ-polynomials

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

to obtain the expression

$$P(R = r|\theta) = \exp(r\theta) \gamma_r K(\theta)^{-1}. \tag{6}$$

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 what follows. 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 $1, \ldots, i+1$ must be obtained by scoring x on item $i+1$ and $r - x$ on the items $1, \ldots, i$. 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

$$\gamma_r^{(i+1)} = \sum_{x=r-\sum_{k=1}^{i} m_k}^{\min(m_{i+1},r)} \exp(\eta_{i+1,r})\gamma_{r-x}^{(i)}. \tag{7}$$

Person locations can be estimated using maximum likelihood estimation or Bayes modal estimation. A special case of the latter is so-called *weighted likelihood estimation*. Since the γ's do not depend on θ (6) 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 (6) 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 (6) attain their maximum when θ is $-\infty$ and ∞, respectively. 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

$$g_x(\theta) = \frac{P(X = x|\theta)g(\theta)}{P(X = x)} \propto P(X = x|\theta)g(\theta) \tag{8}$$

w.r.t. θ keeping item parameters and the observations fixed. The MLE described above is a special case corresponding to $g_\omega \equiv 1$. 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) [12]. 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.

5   Implementation in SAS

The SAS macro %rasch_cml uses PROC GENMOD to estimate item parameters, and PROC NLMIXED to estimate person locations. It 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.

6 Simulation

Evaluating of model fit can be done by comparing what has been observed with simulated data describing what could have been observed under the model. The SAS macro `%rasch_cml` simulates data sets under the model. These are obtained by first simulating N person scores locations from the empirical score distribution and then simulating item responses. Let L denote the set of possible scores and for $r \in L$ define $V_r = \{v | R_v = r\} \subset \{1, \ldots, N\}$. Let $N_r = \sharp V_r$ denote the number of persons with each score. First simulate $R_1^{(S)}, \ldots, R_N^{(S)}$ with probabilities $P(R_v^{(S)} = r) = \frac{N_r}{N}$, next simulate a data matrix $(X_{vi}^{(S)})$ using the probabilities

$$P(X_{vi}^{(S)} = \; \cdot \; | R_v = R_v^{(S)}).$$

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

7 Graphics

Three graphical representations are made by the SAS macro `%rasch_cml`: item characteristic curves (ICC's) that display the response probabilities along the latent continuum and two item fit plots. Let N_{ihr} denote the number of persons with total score r giving the answer h to item i. For for each combination $(i, h) \in \{1, \ldots, I\} \times \{0, 1, \ldots, m_i\}$ the macro plots the observed proportion

$$r \mapsto \frac{N_{ihr}}{N_r}$$

as solid black dots and the expected proportions (the probabilities)

$$r \mapsto P(X_{vi} = h | R_v = r)$$

as solid blue lines along with 95% confidence limits as dashed green lines. These plots are illustrated in Figure 2 and are closely related to plots of the ICC's

$$\theta_v \mapsto P(X_{vi} = h | \theta_v),$$

because R_v is sufficient for θ_v.

The observed mean score function for item i is

$$\theta \mapsto \frac{1}{N_r} \sum_{v \in V_r} X_{vi}.$$

The simulated mean score function is obtained as follows by simulating item responses $X_{1i}^{(S)}, \ldots, X_{Ni}^{(S)}$ as described in section 6 and calculating

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

where $N_r^{(S)} = \sharp\{v | \sum_{i \in I} X_{vi}^{(S)} = r\}$.

8 The SAS macro

The Hospital Anxiety and Depression Scale (HADS) was designed as a brief instrument used to assess symptoms of anxiety and depression [13] and contains 14 items often scored as two seven-item sub scales: 'Depression' (even numbered items) and 'Anxiety' (odd numbered items). The SAS macro is illustrated using data reported by Pallant and Tennant [14]. The first step is to create a data set

```
data inames;
 input item_name $ item_text $ max group @@;
 cards;
 AHADS1 anx1 3 1   AHADS3 anx3 3 2    AHADS5 anx5  3 3   AHADS7 anx7 3 4
 AHADS9 anx9 3 5   AHADS11 anx11 3 6  AHADS13 anx13 3 7
 ;
run;
```

that describes the items: **item_name** is the name of the items, **item_text** are text strings attached to the items, **max** is the maximum item score for each item, and **group** are integers defining groups of items that have the same item parameters. Thus, all HADS items are scored $0, 1, 2, 3$ and they all have their own vector of item parameters. The macro is called using the statement

```
%rasch_cml(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 (the default value is **CML**).

The SAS macro creates six data sets. The data set **CML_logl** contains the maximum value of the conditional log likelihood function. The data sets **CML_par** and **CML_par_ci** contains item parameter estimates, the difference between them is illustrated by the (edited) output

```
  item      beta1    beta2    beta3
  AHADS1    -3.75     0.17     0.82
  :
  AHADS9    -0.93     1.51     2.43
```

from **CML_par** and

```
                                  Lower     Upper
  item      Label    cat  estimate   CL        CL
  AHADS1    eta11     1       3.75    3.15      4.35
  AHADS1    eta12     2       3.58    2.95      4.21
  AHADS1    eta13     3       2.76    2.12      3.39
  :
  AHADS9    eta71     1       0.93    0.64      1.22
  AHADS9    eta72     2      -0.58   -1.01     -0.14
  AHADS9    eta73     3      -3.00   -3.66     -2.35
  :
  AHADS1    beta11    1      -3.75   -4.35     -3.15
  AHADS1    beta12    2       0.17   -0.14      0.48
  AHADS1    beta13    3       0.82    0.45      1.19
  :
  AHADS9    beta71    1      -0.93   -1.22     -0.64
  AHADS9    beta72    2       1.51    1.13      1.89
  AHADS9    beta73    3       2.43    1.75      3.11
```

from **CML_par_ci**. Note that the threshold parameters (β's) are the same.

The data sets `CML_pp_regr` and `CML_regr` are copies of the input data set with added variables useful for latent regression [15]. The data set `CML_theta` contains MLE and WLE estimates of person locations and their standard errors.

Further options can be specified: `ICC=YES` yields a plot of the item characteristic curves for each item. The ICC's for HADS item 1 is shown in Figure 1. Specifying `plotcat=YES` creates plots
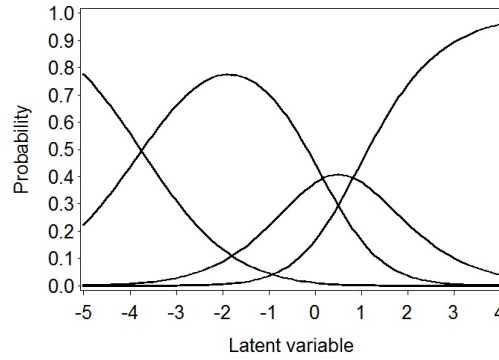

Figure 1: Item characteristic curves (ICC's) for HADS item 1 plotted with option `ICC=YES`. The curves intersect at the thresholds $\beta_{11} = -3.75$, $\beta_{12} = 0.17$, and $\beta_{13} = 0.82$

of observed and expected item category frequencies stratified by the total score. This yields m_i plots for item i as exemplified in Figure 2
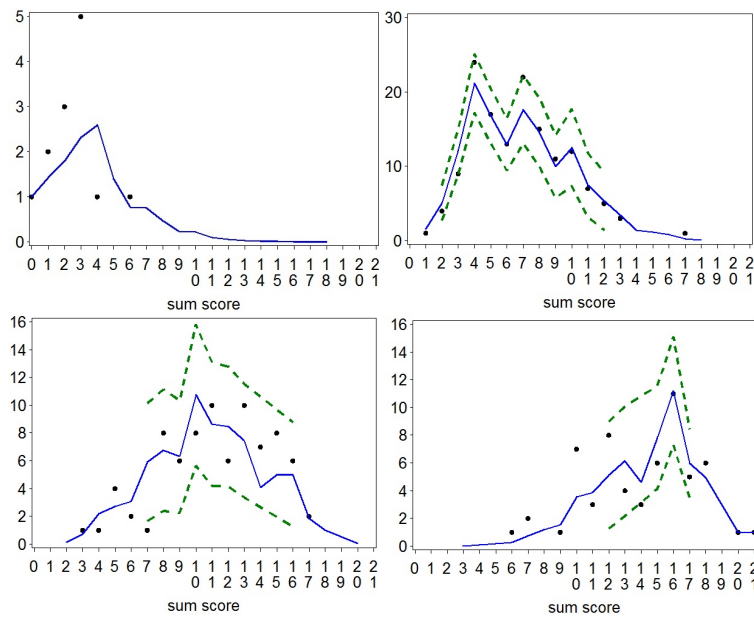

Figure 2: Observed and expected item category frequencies stratified by the total score plotted with option `plotcat=YES`.

Using the option `plotmean=YES` makes the macro plot item means against raw scores 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 3 shows an example.
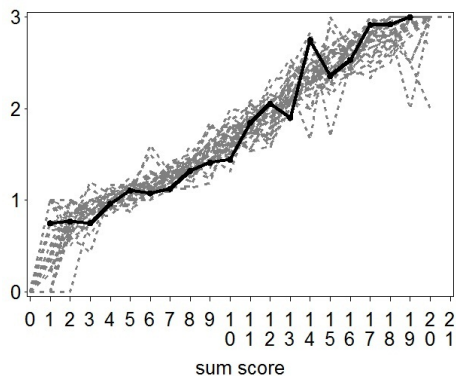

Figure 3: Item fit plot for HADS item 1 plotted with option `plotmean=YES`.

the plot shows the mean scores to be increasing with the total score, cf. requirement (ii), and that the variations observed in the data are well within the range of what would be expected under the model.

9 Discussion

Several proprietary software packages for fitting Rasch models exist, the most widely used being RUMM [16], Conquest [17], and WINSTEPS [18]. With the increasing use of IRT and Rasch models in new research areas where access to specialized proprietary software is limited it is important to provide implementations in standard statistical software such as R and SAS. SAS macros for Rasch models already exist. The macros `%anaqol` [19] and `%irtfit` [20] encompass a wide range of IRT models. The SAS macro `%anaqol` computes Cronbachs coefficient alpha [21], several useful graphical representations and estimates the parameters for any of five IRT models (the dichotomous Rasch model [2, 3], the Birnbaum (2PL) model [22], OPLM, the partial credit model [23], and the rating scale model [24]) 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, it does not perform estimation of item parameters, but require that these have been estimated using other IRT model software programs. The R package eRm [25] is a flexible tool for these analyses, as are the SAS macros `%anaqol` [19] and `%irtfit` [20]. 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.

Because the macro uses the contingency table of item responses no responses must be missing, if the estimation procedure fails to converge a warning or error message is printed. The plots of observed and expected counts in each score group 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. Therefore the macro produces a single item-

11

level goodness-of-fit plot. Furthermore it extends previously implemented macros in that it the output and features are the same for dichotomous and polytomous item response formats and that it presents more graphics, specifically new goodness-of-fit plot where observed item means are compared item means simulated under the model.

References

[1] HAMBLETON, R. K. & 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. & MOLENAAR, I. W. (1995). *Rasch models: Foundations, recent developments, and applications*. Springer-Verlag, New York.

[4] GLAS, C. A. W. & 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, pp. 325–352.

[5] ANDERSEN, E.B. (1977). Sufficient statistics and latent trait models. *Psychometrika*, vol. 42, p. 69–81.

[6] NEYMAN, J. & SCOTT, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, vol. 16, p. 1–32.

[7] KELDERMAN, H. (1984). Loglinear Rasch model tests. *Psychometrika*, vol. 49, p. 223–245.

[8] TJUR, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scandinavian Journal of Statistics*, vol. 9, p. 23–30.

[9] AGRESTI, A. (1993). Computing conditional maximum likelihood estimates for generalized Rasch models using simple loglinear models with diagonals parameters. *Scandinavian Journal of Statistics*, vol. 20, p. 63–71.

[10] TENVERGERT, E., GILLESPIE, M. & KINGMA, J. (1993). Testing the assumptions and interpreting the results of the Rasch model using log-linear procedures in SPSS. *Behaviour Research Methods, Instruments, & Computers*, vol. 25, p. 350–359.

[11] CHRISTENSEN, K. B. (2006). Fitting Polytomous Rasch Models in SAS. *Journal of Applied Measurement*, vol. 7, p. 407–417.

[12] WARM, T. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, vol. 54, p. 427–450.

[13] ZIGMOND, A. S. & SNAITH, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, vol. 67, p. 361–370.

[14] PALLANT, J. & 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*, vol. 46, p. 1–18.

[15] CHRISTENSEN, K. B., BJORNER, J. B., KREINER, S. & PETERSEN, J. H.. (2004). Latent Regression in log linear Rasch models. *Communications in Statistics – Theory and Methods*, vol. 33, p. 1295–1313.

[16] ANDRICH, D., LYNE, A., SHERIDAN, B. & LUO, G. (2003). *RUMM 2020.* Perth: RUMM Laboratory.

[17] WU, M. L., ADAMS, R. J., WILSON, M. R., & HALDANE, S. A. (2007). *ACER ConQuest Version 2: Generalised item response modelling software.* Camberwell: Australian Council for Educational Research.

[18] LINACRE, J. M. (2011) *WINSTEPS Rasch measurement computer program.* Beaverton, Oregon: Winsteps.com

[19] HARDOUIN, J.-B. & MESBAH, M. (2007). The SAS Macro-Program %AnaQol to Estimate the Parameters of Item Responses Theory Models. *Communications in Statistics – Simulation and Computation*, vol. 36, p. 437–453.

[20] BJORNER, J. B., SMITH, K. J., STONE, C., & SUN, X. (2007) IRTFIT: A Macro for Item Fit and Local Dependence Tests under IRT Models. Lincoln RI: QualityMetric Incorporated. [http://outcomes.cancer.gov/areas/measurement/irtfit_macro_users_guide.pdf].

[21] CRONBACH, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, vol. 16, p. 297-334.

[22] BIRNBAUM, A. L. (1968). Latent trait models and their use in inferring an examinee's ability. In: F. M. Lord & M. R. Novick *Statistical Theories of Mental Test.* Reading, MA: Addison-Wesley.

[23] MASTERS, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, vol. 47, p. 149–174.

[24] ANDRICH, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, vol. 43, p. 357–374.

[25] MAIR, P. & HATZINGER, R. (2007). Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R *Journal of Statistical Software*, 20, 9.