LIRT: SAS macros for longitudinal IRT models

Maja Olsbjerg
Karl Bang Christensen

Research Report 14/03

Department of Biostatistics
University of Copenhagen
LIRT: SAS macros for longitudinal IRT models

Maja Olsbjerg  
Section of Biostatistics,  
University of Copenhagen

Karl Bang Christensen  
Section of Biostatistics,  
University of Copenhagen

Abstract

Item response theory models are often applied when a number items are used to measure a unidimensional latent variable. Originally proposed within educational research, they are now also being used when focus is on e.g. physical functioning or psychological well-being. Modern applications often need more general models, typically models for multidimensional latent variables or longitudinal models for repeated measurements. This paper describes a collection of SAS macros that can be used for fitting data to and simulating from a two-dimensional generalized partial credit model. Macros for plotting item characteristic curves and estimating person locations are also included. The macros encompass both Rasch type models and models with item discrimination parameters and handle both dichotomous and polytomous item response formats. The models are sufficiently flexible to accommodate changes in item parameters across time points and local dependence between responses at different time points.

Keywords: Polytomous IRT model, Rasch model, 1PL model, Birnbaum model, 2PL model, Partial Credit model, Generalized Partial Credit model, longitudinal IRT model, marginal maximum likelihood (MML) estimation, item parameter drift, local dependence, SAS macro.
1. Introduction

Item response theory (IRT) models were developed to describe probabilistic relationships between correct responses to a set of test items and continuous latent traits (van der Linden and Hambleton 1997). IRT models were originally developed and used in educational testing, where the models describe how the probability of a correct answer to an item in a test depends on ability, but they are applicable whenever location of persons and items on an underlying latent scale is of interest. Traditional applications in education often use dichotomous (correct/incorrect) item scoring, but polytomous items are common in other applications.

IRT models are increasingly used in health status measurement and evaluation of Patient Reported Outcomes (PRO’s), e.g. physical functioning and psychological well-being (Reeve, Hays, JB, Cook, Crane, Teresi, Thissen, Revicki, Weiss, Hambleton, Liu, Gershon, Reise, Lai, Cella, and Group 2007). The simplest IRT model, the Rasch (1960) model, (Fischer and Molenaar 1995; Christensen, Kreiner, and Mesbah 2013), is increasingly used for validation of measurement instruments (Tennant and Conaghan 2007) and has been shown to be superior to classical approaches (Blanchin, Hardouin, Le Neel, Kubis, Blanchard, Mirallié, and Sébille 2011).

The use of IRT models in new research fields increases the need for implementation in standard statistical software like SAS. Estimation of IRT models using SAS has been the topic of several research papers (Rijmen, Tuerlinckx, De Boeck, and Kuppens 2003; Smits and De Boeck 2003; Nandakumar and Hotchkiss 2012). In particular, implementation of polytomous Rasch models in SAS has been discussed (Christensen 2006), and SAS macros that use these ideas are available (Christensen and Bjorner 2003; Hardouin and Mesbah 2007; Christensen and Olsbjerg 2013; Christensen 2013).

In many applications multidimensional latent variables or repeated measurements of the same latent variable are considered. Longitudinal Rasch models were studied by Pastor and Beretvas (2006), who illustrated how these models can be seen as hierarchical generalized linear models and implemented in the software program HLM (Raudenbush, Bryk, Cheong, and Congdon 2004). HLM uses penalized quasi-likelihood for estimation, but as noted by Pastor and Beretvas, various estimation procedures and software programs for these kinds of models exist. An example is the Random Weights Linear Logistic Test Model (Rijmen and De Boeck 2002), which is a special case of the multidimensional
random coefficients multinomial logit model (Adams, Wilson, and Wang 1997a) implemented in the computer program ConQuest (Wu, Adams, Wilson, and Haldane 2007). For these more general models, implementation in standard software is also useful.

This paper describes four SAS macros %lirt_mml, %lirt_ppar, %lirt_icc and %lirt_simu, all available from

\texttt{biostat.ku.dk/~mola}

The macros constitute a framework for fitting and simulating from two-dimensional generalized partial credit models (GPCM). The macros %lirt_mml and %lirt_ppar estimate the item and person parameters respectively. The macro %lirt_icc plots item characteristic curves (ICC’s) of a given model and finally %lirt_simu generates data sets with item responses simulated from the model. The use of the SAS macros is illustrated using data from a longitudinal study based on a health-related quality of life (HRQoL) questionnaire applied to women screened for breast cancer.

The two most important macros of the four are %lirt_mml and %lirt_ppar. The first of these uses marginal maximum likelihood (MML) estimation (Bock and Aitkin 1981; Thissen 1982; Zwinderman and van den Wollenberg 1990) to estimate item parameters and the parameters of a two-dimensional latent distribution. It is sufficiently flexible to model item parameter drift and local dependence across time points. The second SAS macro, %lirt_ppar, estimates the person parameters and changes in these over time using maximum likelihood estimation (MLE), given the (estimated) item parameters.

\section{2. IRT models for a single time point}

IRT models describe the responses to a set of items $\mathbf{X} = (X_{i})_{i \in I}$ that measures a latent variable $\theta \in \mathbb{R}$. For dichotomously scored items two IRT models have traditionally been applied: the Rasch or 1PL model (Rasch 1960; Fischer and Molenaar 1995) and the Birnbaum or 2PL model (Birnbaum 1968). These, the simplest IRT models, are given by the probabilities

$$P(X_i = x_i|\theta) = \frac{\exp(x_i(\theta + \eta_i))}{1 + \exp(\theta + \eta_i)} \quad (x_i = 0, 1) \quad (1)$$
and
\[ P(X_i = x_i | \theta) = \frac{\exp(\alpha_i(x_i(\theta + \eta_i)))}{1 + \exp(\alpha_i(\theta + \eta_i))} \]  \( (x_i = 0, 1) \)  \( (2) \)
for each item \( i \in I \). The formula (1) defining the Rasch model and the formula (2) defining the Birnbaum model reveal that these are logistic regression models. When \( \alpha_i \) is constant across items the two IRT models are identical.

For ordinal item response formats, i.e. response categories 0, 1, ..., \( m_i \) for item \( i \in I \) the models (1) and (2) can be generalized to
\[ P(X_i = x_i | \theta) = \frac{\exp(x_i \theta + \eta_{ix_i})}{\sum_{k=0}^{m_i} \exp(k \theta + \eta_{ik})} \]  \( (x_i = 0, 1, \ldots, m_i) \)  \( (3) \)
and
\[ P(X_i = x_i | \theta) = \frac{\exp(\alpha_i(x_i(\theta + \eta_{ix_i}))}{\sum_{k=0}^{m_i} \exp(\alpha_i(k \theta + \eta_{ik}))} \]  \( (x_i = 0, 1, \ldots, m_i) \)  \( (4) \)
respectively. For identification purposes \( \eta_{i0} = 0 \) for all \( i \in I \). Note that for \( m_i = 1 \) this corresponds to setting \( \eta_{i0} = 0 \) and \( \eta_{i1} = \eta_i \) in (1) and (2), respectively.

The model (3) generalizing (1) is called the partial credit model (PCM) Masters (1982), though originally proposed without this name by Andersen (1977). The model (4) generalizing (2) is called the generalized partial credit model (GPCM) Muraki (1992).

In these models the parameter \( \alpha_i \) and the vector \( (\eta_{ik})_{k=1,\ldots,m_i} \) are parameters describing the item while \( \theta \) is a parameter describing the person responding. In the psychometric literature \( \alpha_i \) is referred to as an item discrimination parameter. Also in the polytomous case the PCM appears as the special case of the GPCM where the item discrimination is constant across items
\[ \alpha_i = \alpha \quad \text{for all} \quad i \in I. \]  \( (5) \)

An alternative parametrization of the \( \eta \)'s can be obtained by using the so-called item thresholds defined by \( \beta_{i0} = 0 \) and
\[ \beta_{ik} = - (\eta_{ik} - \eta_{ik-1}) \quad (k = 1, \ldots, m_i) \]  \( (6) \)
for each item $i \in I$. These are called thresholds because they correspond to the values of the latent variable for which adjacent categories are equally likely

$$P(X_i = x|\theta = \beta_{ix}) = P(X_i = x - 1|\theta = \beta_{ix}).$$

2.1. Interpretation of item parameters

The above models all specify the conditional probability of each response option given the value $\theta$ of the latent variable. The item parameters are most easily interpreted by plotting these probabilities against $\theta$, that is by plotting the functions

$$\theta \mapsto P(X_i = x|\theta) \quad (\text{for } x = 0, 1, \ldots, m_i).$$

These are called item characteristic curves (ICC’s). Consider two items each with three response options. The first item has discrimination parameter $\alpha = 1$ and the second one $\alpha = 2.5$. The remaining item parameters are given by

$$(\eta_1, \eta_2) = (1, -1) \quad \text{and} \quad (\eta_1, \eta_2) = (0, -1)$$

for the first and second item respectively. According to (6) this corresponds to thresholds

$$(\beta_1, \beta_2) = (-1, 2) \quad \text{and} \quad (\beta_1, \beta_2) = (0, 1).$$

Figure 1 shows the item characteristic curves for both items. We note that the curves are steeper for the item with the highest value of the discrimination parameter $\alpha$.

2.2. Joint likelihood and identification

Common to all IRT models is the technical assumption that items are locally independent, that is the vector $\overline{X} = (X_i)_{i \in I}$ of item responses satisfies
Figure 1: Item characteristic curves (ICC’s) for two items with $(\alpha, \beta_1, \beta_2) = (1, -1, 2)$ and $(\alpha, \beta_1, \beta_2) = (2.5, 0, 1)$ respectively.

\[
P(X = x_i | \theta) = \prod_{i \in I} P(X_i = x_i | \theta) \text{ for all } \theta \in \mathbb{R}. \tag{8}
\]

For a person $v \in \{1, \ldots, N\}$ responding to items in a subset $I_v \subset I$, with observed response vector $\mathbf{x}_v = (x_{vi})_{i \in I_v}$, the contribution to the joint likelihood implied by (8) becomes

\[
L(\alpha, \eta, \theta_v | \mathbf{x}_v) = \prod_{i \in I_v} P(X_{vi} = x_{vi} | \theta_v) \tag{9}
\]

where $\alpha = (\alpha_i)_{i \in I}$ and $\eta = (\eta_i)_{i \in I}$ with $\eta_i = (\eta_{i1}, \ldots, \eta_{im_i})$. Inserting the probabilities (4) in (9) yields

\[
L(\alpha, \eta, \theta_v | \mathbf{x}_v) = \frac{\exp \left( \theta_v \sum_{i \in I_v} \alpha_{vi} + \sum_{i \in I_v} \alpha_{i} \eta_{i} \right)}{\prod_{i \in I_v} \sum_{k=0}^{m_i} \exp(\alpha_{i}(k\theta_v + \eta_{ik}))}. \tag{10}
\]

The model is only identified if restrictions are placed on either item parameters or the latent variable, since the reparametrizations

\[
\alpha^* = \gamma \alpha \quad \text{and} \quad \theta^* = \frac{1}{\gamma} \theta \tag{11}
\]

or

\[
\eta_{ik}^* = \eta_{ik} - \omega k \quad \text{and} \quad \theta^* = \theta + \omega \tag{12}
\]
for arbitrary $\gamma, \omega > 0$ yield the same probabilities. An often used restriction is

$$\prod_{i \in I} \alpha_i = 1$$

(13)

for the discrimination parameters and

$$\sum_{i \in I} \eta_{im} = 0$$

(14)

or equivalently

$$\sum_{i \in I} \sum_{k=0}^{m_i} \beta_{ik} = 0$$

(15)

for the item parameters. Alternatively, restrictions on the latent variable can be imposed. If the mean of $\theta$ is restricted then all $\eta$’s can be estimated, and if the variance of $\theta$ is restricted then all $\alpha$’s can be estimated.

### 2.3. Item parameter estimation

Estimation based on the likelihood (10) leads to inconsistent estimates (Neyman and Scott 1948). In the special case of the PCM, conditional maximum likelihood (CML) estimation (Andersen 1973) can be used for item parameter estimation, but in general, marginal maximum likelihood (MML) estimation (Bock and Aitkin 1981; Thissen 1982; Zwinderman and van den Wollenberg 1990) is used. This estimation method is based on a distributional assumption about the latent variable, typically that $\theta_1, \ldots, \theta_N$ are iid and normally distributed. If the item parameters are restricted as in (13) and (14) then the mean and variance of this normal distribution can be estimated. Alternatively, assuming that $\theta \sim N(0, 1)$ with density $\varphi$, all item parameters can be estimated. In that case the marginal likelihood is given as

$$L_M(\alpha, \eta | x_1, \ldots, x_N) = \prod_{v=1}^{N} \int L(\alpha, \eta, \theta | x_v) \varphi(\theta) d\theta.$$  

(16)

Since random effects models like this are easily implemented in PROC NLMIXED this is the procedure used in the SAS macros. In the implementation we assume that
$\theta \sim N(0, 1)$ and estimate all the discrimination parameters $\bar{\alpha}$ and all the item parameters $\bar{\eta}$. If the restriction (13) is imposed on the discrimination parameters we can assume that $\theta \sim N(0, \sigma^2)$ and from (11) we see that

$$\sigma = \left( \prod_{i \in I} \alpha_i \right)^{\frac{1}{|I|}}.$$

If the restriction (14) is imposed on the threshold parameters, we can assume $\theta \sim N(\mu, 1)$ and from (12) we see that

$$\mu = \frac{1}{\sum_{i \in I} m_i} \left( \sum_{i \in I} \eta m_i \right).$$

### 2.4. Person parameter estimation

It is usually of interest to estimate individual values of the latent variable referred to as the person parameters. This can be done by substituting the item parameters by their MML estimates resulting in the likelihood function

$$L_P(\bar{\theta}) = L(\hat{\alpha}, \hat{\eta}, \bar{\theta} | \bar{x}_1, ..., \bar{x}_N).$$

An estimate of $\bar{\theta} = (\hat{\theta}_v)_{v=1,...,N}$ can then be obtained by numerical optimization of this likelihood.

### 3. Longitudinal IRT models

For time points $t = 1, 2$ let $\bar{X}_t = (X_{it})_{i \in I}$ be a set of items measuring a value $\theta_t \in \mathbb{R}$ of the latent variable. In a situation like this it is natural to assume that $\text{Corr}(\theta_1, \theta_2) > 0$ and thus that item responses from the same person at the two time points are positively correlated. For each person $v \in \{1, ..., N\}$ we are interested in a baseline measurement $\theta_{v1}$, a follow-up measurement $\theta_{v2}$ and measurement of change $\delta_v = \theta_{v2} - \theta_{v1}$.

Assume that person $v$ responds to items in subsets $I_{v1}, I_{v2} \subset I$ at time 1 and 2 respectively. That is we allow for items only to be administered at one time point. We consider
items fitting the polytomous IRT model, the GPCM (4), at both time points

\[ P(X_{it} = x_{it} | \theta_t) = \frac{\exp(\alpha_{it}(x_{it} \theta_t + \eta_{i x_{it}}))}{\sum_{k=0}^{m_i} \exp(\alpha_{it}(k \theta_t + \eta_{ik}))} \] (18)

and assume that local independence (8) holds within each time point \( t \)

\[ P(\mathbf{X}_t = \mathbf{x}_t | \theta_t) = \prod_{i \in I} P(X_{it} = x_{it} | \theta_t) \text{ for all } \theta_t \in \mathbb{R}. \] (19)

To fully specify the model, further assumptions about the dependence structure are needed.

### 3.1. Simple model

The simplest specification of the longitudinal GPCM is obtained by extending the independence assumption (8) to hold for all pairs of items \((i, i')\)

\[ P(X_{i 1} = x_{i 1}, X_{i' 2} = x_{i' 2} | \theta_1, \theta_2) = P(X_{i 1} = x_{i 1} | \theta_1)P(X_{i' 2} = x_{i' 2} | \theta_2) \] (20)

and assuming that the item parameters are constant over time, i.e. that for all \( i \in I \)

\[ \alpha_{i 1} = \alpha_{i 2} \quad \text{and} \quad \eta_{ik 1} = \eta_{ik 2} \quad \text{for all } k = 0, ..., m_i. \] (21)

Based on these assumptions the contribution to the joint likelihood of person \( v \), with observed responses \( \mathbf{x}_v = (\mathbf{x}_{v 1}, \mathbf{x}_{v 2}) \) for vectors \( \mathbf{x}_{v 1} = (x_{vi 1})_{i \in I_{v 1}} \) and \( \mathbf{x}_{v 2} = (x_{vi 2})_{i \in I_{v 2}} \), becomes

\[ L(\alpha, \eta, \theta_{v 1}, \theta_{v 2} | \mathbf{x}_v) = \prod_{t=1}^{2} P(\mathbf{X}_{vt} = \mathbf{x}_{vt} | \theta_{vt}) \]

\[ = \prod_{t=1}^{2} \prod_{i \in I_{vt}} \sum_{k=0}^{m_i} \frac{\exp(\alpha_i(x_{vt i} \theta_{vt} + \eta_{ix_{vt i}}))}{\sum_{k=0}^{m_i} \exp(\alpha_i(k \theta_{vt} + \eta_{ik}))}. \] (22)
Here $\alpha_i = (\alpha_i)_i \in I$ and $\eta_i = (\eta_i, \ldots, \eta_{i,m_i})$ denote the item parameters, which in this case are the same at the two time points. For the special case of the dichotomous Rasch model the longitudinal model implied by the assumptions (20) and (21) was discussed by Andersen (1985) and Embretson (1991).

Assuming that the vectors $\theta_v = (\theta_{v1}, \theta_{v2})$ for $v = 1, \ldots, N$ are iid from a two-dimensional normal distribution $N(\mu, \Sigma)$ item and population parameters can be estimated using MML estimation. Several papers have considered this specification of the model in the special case of the Rasch model Andersen (1977); Embretson (1991); Adams et al. (1997a); Adams, Wilson, and Wu (1997b).

3.2. Allowing for local dependence across time points

When an item is used at both time points the extended assumption of local independence (20) might not be justified. In that case we should only rely on what we know, namely that

$$P(X_{i1} = x_{i1}, X_{i2} = x_{i2} | \theta_1, \theta_2) = P(X_{i1} = x_{i1} | \theta_1)P(X_{i2} = x_{i2} | X_{i1} = x_{i1} ; \theta_2).$$

(23)

Thus, taking account of local dependence is a matter of choosing a suitable model for the conditional probabilities in (23). One option is to stay with the GPCM assuming that

$$P(X_{i2} = x_{i2} | X_{i1} = x_{i1}; \theta_2) = \frac{\exp[\alpha_{i2}^*(x_{i1}) (x_{i2} \theta_2 + \eta_{i,k2}(x_{i1}))]}{\sum_{k=0}^{m_i} \exp[\alpha_{i2}^*(x_{i1}) (k\theta_2 + \eta_{i,k2}(x_{i1}))]}$$

(24)

where $\alpha_{i2}^*(x_{i1})$ and $(\eta_{i,k2}(x_{i1}))_{k=1}^{m_i}$ are item parameters depending on the response observed at time 1. If item $i$ in fact meets the extended assumption of local independence (20) then $\alpha_{i2}^*(x)$ and $\eta_{i,k2}(x) = (\eta_{i,k2}(x))_{k=1}^{m_i}$ are constant across $x \in \{0, 1, \ldots, m_i\}$. Henceforth, we assume that this is the case for some items forming the subset $I_0 \subseteq I$.

3.3. Allowing for item parameter drift

Another way of adding flexibility to (22) is by allowing item parameters to change over time. Clearly change in a person is easiest to evaluate when the measurement instrument
does not change over time. However, it is feasible to evaluate changes $\delta = \theta_2 - \theta_1$ in the latent variable as long as a subset $J_0 \subset I_0$ of the items satisfy (21).

### 3.4. General model

A flexible model that extends (22) to a model with both local dependence across time points and item parameter drift, for some items, can be formulated by specifying the item subsets $I_0$ and $J_0$. Recall that $I_0$ are the items that are locally independent across time points and $J_0$ the items with no item parameter drift. The contribution to the joint likelihood for person $v$ with observed responses $\mathbf{x}_v$ becomes

\[
L(\mathbf{\alpha}_1, \mathbf{\alpha}_2, \mathbf{\eta}_1, \mathbf{\eta}_2, \theta_{v1}, \theta_{v2} | \mathbf{x}_v) = \prod_{i \in I \cap I_{v1}} P(X_{vi1} = x_{vi1} | \theta_{v1}) \prod_{i \in I \cap I_{v2}} P(X_{vi2} = x_{vi2} | \theta_{v2}) \prod_{i \in (I \cap I_{v1}) \cap (I \cap I_{v2})} P(X_{vi2} = x_{vi2} | X_{vi1} = x_{vi1}; \theta_{v2})
\]

\[
= \prod_{i \in I \cap I_{v1}} \frac{\exp(\alpha_{v1}(x_{vi1}\theta_{v1} + \eta_{ix_{vi1}}))}{\sum_{k=0}^{m_i} \exp(\alpha_{v1}(k\theta_{v1} + \eta_{ik1}))} \prod_{i \in I \cap I_{v2}} \frac{\exp(\alpha_{v2}(x_{vi2}\theta_{v2} + \eta_{ix_{vi2}}))}{\sum_{k=0}^{m_i} \exp(\alpha_{v2}(k\theta_{v2} + \eta_{ik2}))} \prod_{i \in (I \cap I_{v1}) \cap (I \cap I_{v2})} \frac{\exp(\alpha^*_2(x_{vi2}\theta_{v2} + \eta^*_ix_{vi2}(x_{vi1})))}{\sum_{k=0}^{m_i} \exp(\alpha^*_2(x_{vi2}(k\theta_{v2} + \eta^*_ik2(x_{vi1})))},
\]

(25)

Restrictions on the parameters are needed in order for the model to be identified.

### 3.5. Item parameter estimation

Joint estimation based on (25) leads to inconsistent estimates. Instead we can turn to MML estimation assuming that
LIRT: SAS macros for longitudinal IRT models

\[
\begin{bmatrix}
\theta_1 \\
\theta_2
\end{bmatrix} \sim N_2 \left( \begin{bmatrix}
\mu_1 \\
\mu_2
\end{bmatrix}, \begin{bmatrix}
\sigma_1^2 & \sigma_1 \sigma_2 \rho \\
\sigma_1 \sigma_2 \rho & \sigma_2^2
\end{bmatrix} \right)
\] (26)

where \( \rho = \text{Corr}(\theta_1, \theta_2) \) represents the latent correlation. An alternative parametrization, obtained by restricting the latent distribution at time 1, leads to a parametrization of the longitudinal GPCM in terms of change in the mean and variance, which is often the main interest. This can be expressed as

\[
\begin{bmatrix}
\theta_1 \\
\theta_2
\end{bmatrix} \sim N_2 \left( \begin{bmatrix}
0 \\
\mu
\end{bmatrix}, \begin{bmatrix}
1 & \sigma \rho \\
\sigma \rho & \sigma^2
\end{bmatrix} \right).
\] (27)

For either of the choices (26) or (27) the item parameters and the parameters of the latent distribution can be estimated by numerical optimization of an approximation to the marginal likelihood

\[
L_M(\bar{\alpha}_1, \bar{\alpha}_2, \bar{\eta}_1, \bar{\eta}_2 | \bar{x}_1, \ldots, \bar{x}_N) = \prod_{v=1}^{N} \int_{\mathbb{R}^2} L(\bar{\alpha}_1, \bar{\alpha}_2, \bar{\eta}_1, \bar{\eta}_2, \bar{\theta} | \bar{x}_v) \varphi_{\mu, \Sigma}(\bar{\theta}) d\bar{\theta}
\] (28)

for the relevant normal density \( \varphi_{\mu, \Sigma} \) of the latent vector \( \bar{\theta} \).

3.6. Estimation of person parameters

It is usually of interest to estimate change in the latent variable at the individual level. As in the unidimensional case this can be done by inserting MML estimates of the item parameters in the likelihood (28) obtaining
\[ L_P(\theta_v | \bar{\theta}_v) = L(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\eta}_1, \hat{\eta}_2, \bar{\theta}_v | \bar{\theta}_v) \]

\[ = \prod_{i \in I \cap I_{v1}} \frac{\exp(\hat{\alpha}_{1i}(x_{v1i}\theta_{v1} + \hat{\eta}_{ix_{v1i}}))}{\sum_{k=0}^{m_i} \exp(\hat{\alpha}_{1i}(k\theta_{v1} + \hat{\eta}_{ik1}))} \]
\[ \times \prod_{i \in J_0 \cap I_{v2}} \frac{\exp(\hat{\alpha}_{12i}(x_{v2i}\theta_{v2} + \hat{\eta}_{ix_{v2i}}))}{\sum_{k=0}^{m_i} \exp(\hat{\alpha}_{12i}(k\theta_{v2} + \hat{\eta}_{ik2}))} \]
\[ \times \prod_{i \in (I \setminus J_0) \cap I_{v2}} \frac{\exp(\hat{\alpha}_{12i}(x_{v1i}\theta_{v1} + \hat{\eta}_{ix_{v1i}}))}{\sum_{k=0}^{m_i} \exp(\hat{\alpha}_{12i}(k\theta_{v1} + \hat{\eta}_{ik1}))} \]
\[ \times \prod_{i \in (I \setminus J_0) \cap I_{v2}} \frac{\exp(\hat{\alpha}_{12i}(x_{v2i}\theta_{v2} + \hat{\eta}_{ix_{v2i}}))}{\sum_{k=0}^{m_i} \exp(\hat{\alpha}_{12i}(k\theta_{v2} + \hat{\eta}_{ik2}))}. \] (29)

Estimation based on (29) can actually be considered MML estimation since a one-point distribution of the item parameters at their estimates is implicitly imposed. However, the estimates obtained by numerical optimization of (29), are referred to as the maximum likelihood estimates (MLE).

4. Implementation in SAS

Our implementation consists of four SAS macros that provides a framework for fitting and comparing different IRT models. The general specification of the longitudinal GPCM represented by (25) can be fitted, and by using different choices of the item subsets \( I_0 \) and \( J_0 \), different models can be specified. The SAS macros handle both the longitudinal case of two time points, as described above, and the case of a single time point. Two of the macros, \%lirt_mml and \%lirt_ppar estimate item and person parameters respectively. The macro \%lirt_mml uses the NLMIXED procedure for MML estimation of the item parameters, and the parameters of the two-dimensional normal distribution (27) assumed for the latent vector. PROC NLMIXED fits nonlinear mixed models (Rijmen et al. 2003; Smits and De Boeck 2003) and is very flexible because the conditional distribution given the random effects can be specified to be a general distribution using SAS programming statements. The NLMIXED procedure maximizes an approximation to the likelihood integrated over the random effects. Different integral approximations are available, the principal one being adaptive Gaussian quadrature. The SAS macro \%lirt_ppar estimates
LIRT: SAS macros for longitudinal IRT models

the latent variable(s) for each person and is also based on PROC NLMIXED. In this case there are no random effects to integrate out because the item parameters are considered fixed and are given in an input data set, that can be specified by the user. The macro %lirt_mml creates this data set. The third macro, %lirt_icc, plots ICC’s for given item parameters. Finally, the macro %lirt_simu simulates responses from a given model within the framework.

The macros take different input data sets and options as listed below. However, one important input data set is similar for all macros, the data set names, which is the key specifier of the model. The structure of this data set determines whether there is item parameter drift and local dependence across time points for any of the items. For %lirt_mml it also specifies which items should be modeled according to the PCM and the GPCM respectively. For %lirt_ppar, %lirt_icc and %lirt_simu additional variables with values of the item parameters are required and for %lirt_simu further information about the potential local dependence structure. Input data sets and options for the macros are listed below.

Input data sets and options for %lirt_mml:

- **data**: data set with item responses. Each person should be represented by one record and each item by one variable.
- **names**: a model-specifying data set with information about the items.
- **dim**: dimension of the latent variable (1 or 2).
- **out**: prefix for output data sets.
- **delete**: option indicating whether temporary data sets created by the macro should be deleted (Y) or not (N).

Input data sets and options for %lirt_ppar:

- **data**: data set with item responses. Each person should be represented by one record and each item by one variable.
• **names**: a model-specifying data set with information about the items and estimates of the item parameters.

• **dim**: dimension of the latent variable (1 or 2).

• **id**: variable in data containing unique person ID’s.

• **out**: prefix for output data sets.

• **delete**: option indicating whether temporary data sets created by the macro should be deleted (Y) or not (N).

---

Input data sets and options for **%lirt_icc**:

• **names**: a model-specifying data set with information about the items and estimates of the item parameters.

• **dim**: dimension of the latent variable (1 or 2).

• **out**: prefix for output data sets.

• **delete**: option indicating whether temporary data sets created by the macro should be deleted (Y) or not (N).

---

Input data sets and options for **%lirt_simu**:

• **names**: a model-specifying data set with information about the items and the dependence structure, and estimates of the item parameters.

• **dim**: dimension of the latent variable (1 or 2).

• **ndata**: number of data sets.

• **npersons**: number of persons.
• **pdata**: data set either specifying the normal distribution of the latent variable or holding values of the latent variables for a number of individuals.

• **out**: prefix for output data sets.

• **delete**: option indicating whether temporary data sets created by the macro should be deleted (Y) or not (N).

The variables required in the data set **names** differ depending on whether one (dim=1) or two (dim=2) time points are considered. It should become clear from the data example below what they should be. Except for the output data sets, all of the data sets created by the macros are given names starting with an underscore. Such data sets are the only ones deleted with the option **delete=Y**.

As for the output data sets an important one of %lirt.mml is **OUT_names** which is in essence a copy of the input data set **names**, but with scores and the corresponding item parameter estimates added. This data set can be used directly as the input data set **names** for the macro %lirt.icc and by adding a couple of variables also for the macro %lirt.simu. Another output data set from %lirt.mml is **OUT_logl** containing the likelihood value for the fitted model. This makes it straightforward to compare various specifications of the model using likelihood ratio tests. Goodness of fit can be further evaluated (graphically) by comparing observed data to data simulated under the model using %lirt.simu. For %lirt.icc the main output is of course plots of the ICC’s. The output data sets of the macros are listed below.

Output data sets of %lirt.mml:

• **OUT_names**: a model-specifying data set with information about the items, including the estimated item parameters. This can be used as the input data set **names** for the macros %lirt.ppar, %lirt.icc and after a slight modification/extension also for %lirt.simu.

• **OUT_disc**: data set with estimated item discriminations.

• **OUT_disc_std**: data set with standardized estimates of item discriminations.
• OUT_ipar: data set with estimated item parameters ($\eta'$s).

• OUT_thres: data set with estimated item thresholds ($\beta'$s).

• OUT_poppar: data set with estimated parameters of the 2-dimensional normal distribution of the latent variable (only available when dim=2).

• OUT_logl: the value of the loglikelihood for the fitted model.

• OUT_conv: data set with the convergence status of the numerical estimation.

Output data sets of %lirt_ppar:

• OUT_ppar: data with estimated person parameters.

• OUT_conv: data set with the convergence status of the numerical estimation.

Output data sets of %lirt_icc:

• OUT_plot: a data set with data points of the ICC’s.

Output data sets of %lirt_simu:

• OUT_simu1, OUT_simu2,...: simulated data sets with item responses. As many as the specified with the option ndata.

• OUT_names: a model-specifying data set with information about the items.
4.1. Additional SAS macros

Some additional SAS macros facilitating the use of the implementation are available. The SAS macro %lirt_split can be used for splitting locally dependent items. More generally it can be used to split any variable in a given input data set according to another. This means that it can be used for recoding items in analyses of differential item functioning (DIF). Two SAS macros %lirt_simu_names and %lirt_pdata, generate input data sets for %lirt_simu. The first one creates the input data set names and is particularly convenient when local dependence across time points is to be specified for some items. The second one creates the input data set pdata specifying the distribution of the latent variable.

5. Example: longitudinal data from the COS-BC

Many women participating in screening mammography experience a false positive result. Most of these women will experience negative psychosocial consequences. The Psychological Consequences Questionnaire (PCQ) Cockburn, De, Hurley, and Clover (1992) is a questionnaire designed to measure psychological consequences of screening mammography. The Consequences Of Screening in Breast Cancer (COS-BC) Brodersen and Thorsen (2008) is an adaptation and translation of this instrument to a Danish setting with a particular focus on improving the relevance for women with false positive screening results. It consists of several subscales measuring among others anxiety, sense of dejection and sleep.

We consider responses to four polytomous items, intended to measure sleep problems, collected at two time points from women who underwent screening for breast cancer. The data set sleep1_2 contains responses to the four sleep items at both time points. The first 10 records look as follows
Note that not everyone participates at time 2. The ID variable idnr illustrates that each person is represented by one record only. The response options are 0, 1, 2 and 3 (where 0 is 'Not at all' and 3 is 'A lot'). The item wording and the marginal item frequencies at the two time points are shown in Table 1.

<table>
<thead>
<tr>
<th>Item Wording (SAS variable name)</th>
<th>Time</th>
<th>Response options</th>
<th>Total</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>not at all</td>
<td>a bit</td>
<td>quite a bit</td>
</tr>
<tr>
<td>I have been awake</td>
<td>1</td>
<td>1096 (85.0%)</td>
<td>112 (8.7%)</td>
<td>57 (4.4%)</td>
</tr>
<tr>
<td>most of the night</td>
<td>2</td>
<td>933 (86.5%)</td>
<td>90 (8.4%)</td>
<td>36 (3.3%)</td>
</tr>
<tr>
<td>It has taken me a long time to fall asleep (fallasl)</td>
<td>1</td>
<td>938 (72.9%)</td>
<td>180 (14.0%)</td>
<td>90 (7.0%)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>852 (81.1%)</td>
<td>114 (10.9%)</td>
<td>52 (5.0%)</td>
</tr>
<tr>
<td>I have slept badly</td>
<td>1</td>
<td>918 (71.2%)</td>
<td>196 (15.2%)</td>
<td>106 (8.2%)</td>
</tr>
<tr>
<td>(sleepba)</td>
<td>2</td>
<td>825 (78.6%)</td>
<td>135 (12.9%)</td>
<td>59 (5.6%)</td>
</tr>
<tr>
<td>I have woken up far too early in the morning (wokenup)</td>
<td>1</td>
<td>1005 (77.7%)</td>
<td>152 (11.7%)</td>
<td>76 (5.9%)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>876 (81.5%)</td>
<td>117 (10.9%)</td>
<td>41 (3.8%)</td>
</tr>
</tbody>
</table>

Table 1: Item wording and marginal distribution at the two time points for the four sleep items.

We start by carrying out one-dimensional analyses of each time point separately, investigating whether the PCM (3) or the GPCM (4) is more appropriate, and as an initial evaluation of whether item parameters are stable across time points. Subsequently, a
longitudinal model is considered.

5.1. Separate analyses for each time point

The data set names specifying the one-dimensional GPCM at time 1 should contain the variables name, max and disc_yn. It can be created using the SAS code

```sas
data names1;
  input name $ max disc_yn $;
datalines;
  awake1  3  Y
  fallasl1 3  Y
  sleepba1 3  Y
  wokenup1 3  Y
 ;
run;
```

where the variable max contains the maximum response option observed in the data set to be analysed. Specifying disc_yn=Y for all items meaning that for they are modeled as GPCM items (a discrimination parameter is included). Note that the values of the variable name are identical to the the names of the variables in sleep1_2 representing the items. The model is fitted using the macro call

```
%lirt_mml( DATA=sleep1_2,
  NAMES=names1,
  DIM=1,
  OUT=gpcm1);
```

The maximum likelihood value of the fitted model is stored in the data set gpcm1_logl. The item parameter estimates are also stored in data sets. The item parameters (the \( \eta \)'s) in gpcm1_ipar, the thresholds (the \( \beta \)'s) in gpcm1_thres and the discrimination parameters (the \( \alpha \)'s) in gpcm1_disc and gpcm1_disc_std. Another important output data set is gpcm1_names collecting all the information about items; the names, response
formats and all the estimated parameters. This data set can be used as input for the macros \%lirt_ppar, \%lirt_icc and \%lirt_simu.

Letting \texttt{disc\_yn=N} for all items specifies the PCM, and a comparison of the estimated likelihood values yields a likelihood ratio test of the PCM against the more general GPCM model. By adjusting the \texttt{names} data set

```plaintext
data names2;
  set names1;
  disc\_yn='N';
run;
```

and fitting the PCM

```plaintext
\%lirt_mml( DATA=sleep1_2,
  NAMES=names2,
  DIM=1,
  OUT=pcm1);
```

we can test whether this simpler model is justified for the items using a likelihood ratio test

```plaintext
proc sql;
  select Value into :_ll_pcm from pcm1_logl;
  select Value into :_ll_gpcm from gpcm1_logl;
quit;

data _lrt;
  lrt=&_ll_pcm-&_ll_gpcm;
  df=3;
  p=1-cdf('chisquare',lrt,df);
run;

proc print data=_lrt round noobs;
run;
```
The PCM is clearly rejected, with a $\chi^2$ value of 145.3 on 3 degrees of freedom and we stay with the GPCM specified by names1. Similar analyses for time point two also rejected the PCM (results not shown). The estimated item parameters are shown in Tables 2 and 3 along with the item parameters estimated at time 2 separately.

<table>
<thead>
<tr>
<th>Item</th>
<th>$\beta_1$</th>
<th>95% CI</th>
<th>est.</th>
<th>std.</th>
<th>est.</th>
<th>95% CI</th>
<th>est.</th>
<th>std.</th>
</tr>
</thead>
<tbody>
<tr>
<td>awake</td>
<td>1.19</td>
<td>(1.08, 1.30)</td>
<td>-0.11</td>
<td>1.22</td>
<td>(1.12, 1.32)</td>
<td>-0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.51</td>
<td>(1.38, 1.64)</td>
<td>0.21</td>
<td>1.62</td>
<td>(1.48, 1.76)</td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.06</td>
<td>(1.86, 2.26)</td>
<td>0.76</td>
<td>2.02</td>
<td>(1.82, 2.23)</td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fallas</td>
<td>0.71</td>
<td>(0.63, 0.80)</td>
<td>-0.59</td>
<td>0.98</td>
<td>(0.89, 1.06)</td>
<td>-0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.19</td>
<td>(1.09, 1.29)</td>
<td>-0.11</td>
<td>1.43</td>
<td>(1.32, 1.54)</td>
<td>-0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.49</td>
<td>(1.37, 1.61)</td>
<td>0.19</td>
<td>1.78</td>
<td>(1.63, 1.93)</td>
<td>0.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sleepba</td>
<td>0.62</td>
<td>(0.55, 0.70)</td>
<td>-0.67</td>
<td>0.85</td>
<td>(0.78, 0.93)</td>
<td>-0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.15</td>
<td>(1.06, 1.24)</td>
<td>-0.15</td>
<td>1.37</td>
<td>(1.27, 1.47)</td>
<td>-0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.57</td>
<td>(1.45, 1.69)</td>
<td>0.27</td>
<td>1.76</td>
<td>(1.62, 1.90)</td>
<td>0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wokenup</td>
<td>1.13</td>
<td>(0.99, 1.28)</td>
<td>-0.16</td>
<td>1.23</td>
<td>(1.08, 1.38)</td>
<td>-0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.35</td>
<td>(1.19, 1.50)</td>
<td>0.05</td>
<td>1.60</td>
<td>(1.42, 1.78)</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.60</td>
<td>(1.41, 1.79)</td>
<td>0.30</td>
<td>1.59</td>
<td>(1.38, 1.81)</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu$</td>
<td>0</td>
<td></td>
<td>-1.30</td>
<td>0</td>
<td></td>
<td>-1.46</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Estimates of centralized and uncentralized threshold parameters for each of the two time points.
Table 3: Estimates of standardized and unstandardized item discrimination parameters for each of the two time points.

<table>
<thead>
<tr>
<th>Item</th>
<th>est.</th>
<th>95% CI</th>
<th>std.</th>
<th>est.</th>
<th>95% CI</th>
<th>std.</th>
</tr>
</thead>
<tbody>
<tr>
<td>awake</td>
<td>α 4.03</td>
<td>(3.23, 4.84)</td>
<td>0.98</td>
<td>5.06</td>
<td>(3.86, 6.26)</td>
<td>0.97</td>
</tr>
<tr>
<td>fallas</td>
<td>α 4.68</td>
<td>(3.74, 5.62)</td>
<td>1.14</td>
<td>6.38</td>
<td>(4.77, 7.99)</td>
<td>1.55</td>
</tr>
<tr>
<td>sleepba</td>
<td>α 6.73</td>
<td>(4.92, 8.54)</td>
<td>1.64</td>
<td>9.18</td>
<td>(5.87, 12.48)</td>
<td>2.23</td>
</tr>
<tr>
<td>wokenup</td>
<td>α 2.25</td>
<td>(1.88, 2.61)</td>
<td>0.55</td>
<td>2.47</td>
<td>(2.01, 2.94)</td>
<td>0.60</td>
</tr>
</tbody>
</table>

%lirt_icc( NAMES=gpcm1_names, 
DIM=1, 
OUT=icc1);

plots the ICC’s at time 1. This plot, together with the one at time 2 is shown in Figure 2. The output data set icc1_plot created by %lirt_icc contains data points of the ICC’s for a wider range of θ than what is plotted by the macro. This makes it easy for the user to adjust graphical features of the plot, outside the macro.

Estimation of the person locations can also be obtained using the data set gpcm1_names. For time 1 the macro call

%lirt_ppar( DATA=sleep1_2 (where=(idnr<110011)), 
NAMES=gpcm1_names, 
ID=idnr, 
DIM=1, 
OUT=pp1);

yields estimates, saved in a data set pp1_ppar, for the first 10 respondents.
Figure 2: Item characteristic curves for the items at time 1 and 2.
An edited version of the output for both time points looks as follows

<table>
<thead>
<tr>
<th>Parameter</th>
<th>TIME 1</th>
<th>Standard</th>
<th>Lower</th>
<th>Upper</th>
<th>TIME 2</th>
<th>Standard</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Error</td>
<td></td>
<td></td>
<td>Estimate</td>
<td>Error</td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta1</td>
<td>1.2328</td>
<td>0.1480</td>
<td>0.9312</td>
<td>1.5343</td>
<td>1.2328</td>
<td>0.1480</td>
<td>0.9312</td>
<td>1.5343</td>
</tr>
<tr>
<td>theta2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta3</td>
<td>1.1398</td>
<td>0.1619</td>
<td>0.8125</td>
<td>1.4671</td>
<td>1.4062</td>
<td>0.1314</td>
<td>1.1385</td>
<td>1.6740</td>
</tr>
<tr>
<td>theta4</td>
<td>0.7253</td>
<td>0.2005</td>
<td>0.3202</td>
<td>1.1305</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta5</td>
<td>1.2802</td>
<td>0.1564</td>
<td>0.9641</td>
<td>1.5963</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta6</td>
<td>1.0407</td>
<td>0.1716</td>
<td>0.6940</td>
<td>1.3874</td>
<td>0.8760</td>
<td>0.1627</td>
<td>0.5445</td>
<td>1.2075</td>
</tr>
<tr>
<td>theta7</td>
<td>1.8476</td>
<td>0.2300</td>
<td>1.3828</td>
<td>2.3123</td>
<td>2.0641</td>
<td>0.2163</td>
<td>1.6235</td>
<td>2.5047</td>
</tr>
<tr>
<td>theta8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that persons with minimum and maximum scores do not get an estimate.

Tests of fit

In the analyses above likelihood ratio tests of the PCM against the GPCM were carried out at each time point separately. Other tests of fit supported by our implementation are tests of DIF which can be done using a (help) macro `%lirt_split`. This macro splits one variable (e.g. an item) in a given input data set according to another variable (e.g. a person factor). Likelihood ratio tests can then be carried out by comparing the likelihood values resulting from fitting models with or without split items using `%lirt_mml`. The use of `%lirt_split` will be demonstrated later in the context of handling local dependence across time points. Moreover, a graphical evaluation of item fit, based on simulations, has been proposed by Christensen and Olsbjerg (2013). It compares observed and simulated item means for the GPCM across values of the sum score and can be based on `%lirt_simu`. We consider the four sleep items at time point 1 and investigate the fit of the item `awake1`. The mean score on this item across values of the sum score at time 1 can be computed using the statements
LIRT: SAS macros for longitudinal IRT models

\%let it1=awake1 fallasl1 sleepba1 wokenup1;

data sleep1_2;
  set sleep1_2;
  score1=sum(of &it1);
run;

proc means data=sleep1_2;
  var awake1;
  class score1;
  output out=means mean=mean;
run;

and a plot of this can be created using the SAS code

axis1 order=0 to 12 by 1 value=(H=2) minor=NONE label=(H=2);
axis2 value=(H=2) minor=NONE label=(H=2 A=90);

proc gplot data=means;
  plot mean*score1 / haxis=axis1 vaxis=axis2;
  symbol v=none i=join w=3 l=1 color=black;
run;

The plot is shown in Figure 3. In order to simulate from the GPCM fitted at time point 1 the macro \%lirt_simu is used. The output data set gpcm1_names holds the information needed about the items and is used directly as the input data set names for \%lirt_simu. We also need information about the respondents which should be given in a data set pdata. This can take two different forms. The first one needs the variables id and theta. That is a unique person ID and the corresponding (estimated) value of the latent variable. All the data sets generated by \%lirt_simu will be based on these fixed values of the latent variable and the item parameters given in names. For the first 10 respondents with an estimate of the latent variable pdata should in this case look as follows
In order for the macro `%lirt_simu` to know that `pdata` is of this form the option `NPERSONS=0` should be used. Then the number of persons in each simulated data set will be the number of persons present in `pdata`. The second possible form of `pdata` corresponds to specifying the latent normal distribution and should contain the variables `parameter` and `estimate`. Specifying the standard normal distribution is done with the simple piece of code

```sas
data pdata;
  input PARAMETER $ ESTIMATE;
datalines;
mu 0
sigma 1
; run;
```

In this case `NPERSONS` tells `%lirt_simu` how many respondents should be drawn from this distribution. For each generated data set new values of the latent variable are drawn. In this example the latter way of specifying `pdata` is the one used. Four data sets can be simulated as follows

<table>
<thead>
<tr>
<th>id</th>
<th>theta</th>
</tr>
</thead>
<tbody>
<tr>
<td>110004</td>
<td>1.1398</td>
</tr>
<tr>
<td>110006</td>
<td>0.7253</td>
</tr>
<tr>
<td>110008</td>
<td>1.2802</td>
</tr>
<tr>
<td>110009</td>
<td>1.0407</td>
</tr>
<tr>
<td>110010</td>
<td>1.8475</td>
</tr>
</tbody>
</table>
This macro call creates the data sets `s_simu1`, `s_simu2`, `s_simu3` and `s_simu4`. An item mean plot based on the first simulated data set can be obtained using the SAS code

```sas
%lirt_simu( NAMES=gpcm1_names,
    DIM=1,
    NDATA=4,
    NPERSONS=1289,
    PDATA=pdata,
    OUT=s);
```

This plot is shown in Figure 3.

A GoF plot including all of the simulated data sets can be obtained with the piece of code

```sas
data gof1;
    set s_simu1;
    score1=sum(of &it1);
run;

proc means data=gof1;
    var awake1;
    class score1;
    output out=means mean=mean;
run;

axis1 order=0 to 12 by 1 value=(H=2) minor=NONE label=(H=2);
axis2 value=(H=2) minor=NONE label=(H=2 A=90);

proc gplot data=means;
    plot mean*score1 / haxis=axis1 vaxis=axis2;
    symbol v=none i=join w=3 l=33 color=grey;
run;
```

This plot is shown in Figure 3.

A GoF plot including all of the simulated data sets can be obtained with the piece of code
data s0; set sleep1_2; dataset=0; run;
data s1; set s_simu1; dataset=1; score1=sum(of &it1); run;
data s2; set s_simu2; dataset=2; score1=sum(of &it1); run;
data s3; set s_simu3; dataset=3; score1=sum(of &it1); run;
data s4; set s_simu4; dataset=4; score1=sum(of &it1); run;
data gof; set s0-s4; run;

proc means data=gof;
  var awake1;
  class score1 dataset;
  output out=means mean=mean;
run;

axis1 order=0 to 12 by 1 value=(H=2) minor=NONE label=(H=2);
axis2 value=(H=2) minor=NONE label=(H=2 A=90);

proc gplot data=means(where=(dataset ne .));
  plot mean*score1 = dataset / haxis=axis1 vaxis=axis2;
  symbol1 v=none i=join w=3 l=1 color=black;
  symbol2 v=none i=join w=3 l=33 color=grey r=4;
run;

The resulting plot is shown in Figure 4.
Figure 4: The GOF plot comparing the observed item means (solid curves) with four curves based on data sets simulated under the model.

5.2. Longitudinal analysis

We now take into account the longitudinal nature of the data by specifying the longitudinal GPCM (18) for the four sleep items

```sas
data names3;
    input name1 $ name2 $ max disc_yn $;
    datalines;
    awake1 awake2 3 Y
    fallas11 fallas12 3 Y
    sleepba1 sleepba2 3 Y
    wokenup1 wokenup2 3 Y
; run;
```

Note that for longitudinal data the variables representing items at time point 1 and 2 are called `name1` and `name2`, respectively. When, for a given record, both of these variables contain the name of an item, it is assumed that there is no parameter drift for that item. Hence, `names3` specifies a model with no item parameter drift, that is $J_0 = I$. It also represents a model with local independence across time points for all items, that is $I_0 = I$. This will become clear later. Models allowing for local dependence across time
points and/or item parameter drift can be specified by changing the names data set. For now we fit the simple longitudinal model with no parameter drift or dependence with the macro call

```plaintext
%lirt_mml( DATA=sleep1_2,
    NAMES=names3,
    DIM=2,
    OUT=lgpcm );
```

Estimates of the item parameters (the \( \eta \)'s) are stored in the data set lgpcm_ipar and the item thresholds (the \( \beta \)'s) in the data set lgpcm_thres. A subset of lgpcm_thres with estimated thresholds for the two items awake and fallsas1 at both time points looks as follows:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>awake1</td>
<td>1.1398</td>
<td>0.04843</td>
<td>1.0448</td>
<td>1.2348</td>
</tr>
<tr>
<td>awake2</td>
<td>1.5458</td>
<td>0.05994</td>
<td>1.4282</td>
<td>1.6633</td>
</tr>
<tr>
<td>awake3</td>
<td>2.0741</td>
<td>0.08748</td>
<td>1.9025</td>
<td>2.2458</td>
</tr>
<tr>
<td>awake4</td>
<td>1.1398</td>
<td>0.04843</td>
<td>1.0448</td>
<td>1.2348</td>
</tr>
<tr>
<td>awake5</td>
<td>1.5458</td>
<td>0.05994</td>
<td>1.4282</td>
<td>1.6633</td>
</tr>
<tr>
<td>awake6</td>
<td>2.0741</td>
<td>0.08748</td>
<td>1.9025</td>
<td>2.2458</td>
</tr>
<tr>
<td>fallsas1</td>
<td>0.7282</td>
<td>0.04134</td>
<td>0.6471</td>
<td>0.8093</td>
</tr>
<tr>
<td>fallsas2</td>
<td>1.2490</td>
<td>0.04940</td>
<td>1.1521</td>
<td>1.3459</td>
</tr>
<tr>
<td>fallsas3</td>
<td>1.5928</td>
<td>0.06061</td>
<td>1.4739</td>
<td>1.7117</td>
</tr>
<tr>
<td>fallsas4</td>
<td>0.7282</td>
<td>0.04134</td>
<td>0.6471</td>
<td>0.8093</td>
</tr>
<tr>
<td>fallsas5</td>
<td>1.2490</td>
<td>0.04940</td>
<td>1.1521</td>
<td>1.3459</td>
</tr>
<tr>
<td>fallsas6</td>
<td>1.5928</td>
<td>0.06061</td>
<td>1.4739</td>
<td>1.7117</td>
</tr>
</tbody>
</table>

The item discriminations (the \( \alpha \)'s) are stored in the data sets lgpcm_disc:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>awake1</td>
<td>4.2444</td>
<td>0.3468</td>
<td>3.5641</td>
<td>4.9246</td>
</tr>
<tr>
<td>awake2</td>
<td>4.2444</td>
<td>0.3468</td>
<td>3.5641</td>
<td>4.9246</td>
</tr>
<tr>
<td>fallsas1</td>
<td>4.9996</td>
<td>0.4089</td>
<td>4.1973</td>
<td>5.8018</td>
</tr>
<tr>
<td>fallsas2</td>
<td>4.9996</td>
<td>0.4089</td>
<td>4.1973</td>
<td>5.8018</td>
</tr>
<tr>
<td>sleepba1</td>
<td>6.3798</td>
<td>0.6002</td>
<td>5.2024</td>
<td>7.5572</td>
</tr>
<tr>
<td>sleepba2</td>
<td>6.3798</td>
<td>0.6002</td>
<td>5.2024</td>
<td>7.5572</td>
</tr>
<tr>
<td>wokenup1</td>
<td>2.3170</td>
<td>0.1569</td>
<td>2.0093</td>
<td>2.6247</td>
</tr>
<tr>
<td>wokenup2</td>
<td>2.3170</td>
<td>0.1569</td>
<td>2.0093</td>
<td>2.6247</td>
</tr>
</tbody>
</table>
and finally the parameters of the latent distribution are stored in the data set `lgpcm_poppar`

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>-0.3428</td>
<td>0.07291</td>
<td>-0.4858</td>
<td>-0.1998</td>
</tr>
<tr>
<td>rho</td>
<td>0.8279</td>
<td>0.01896</td>
<td>0.7907</td>
<td>0.8651</td>
</tr>
<tr>
<td>sigma</td>
<td>1.1219</td>
<td>0.06464</td>
<td>0.9950</td>
<td>1.2487</td>
</tr>
</tbody>
</table>

telling us that the mean decreases over time and that the latent correlation \( \rho = \text{Corr}(\theta_1, \theta_2) = 0.83 \) is substantial. Furthermore, as in the unidimensional case `%lirt_mml` creates a data set \( \text{lgpcm_names} \) containing the names, scores and estimated parameters of the items. By using this data set as input to `%lirt_ppar` individual estimates of the latent variable at both time points can be obtained. For the first 10 respondents this is done using the following macro call

```sas
%lirt_ppar( DATA=sleep1_2 (where=(idnr<110011)),
              NAMES=lgpcm_names,
              DIM=2,
              ID=idnr,
              OUT=pp);
```

The output printed directly from the data set `pp_ppar` created by `%lirt_ppar` looks as follows

<table>
<thead>
<tr>
<th>idnr</th>
<th>theta1</th>
<th>se_theta1</th>
<th>theta2</th>
<th>se_theta2</th>
<th>delta</th>
<th>se_delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>110001</td>
<td>.</td>
<td>.</td>
<td>1.0664</td>
<td>0.1887</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110002</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110003</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110004</td>
<td>1.1997</td>
<td>0.1771</td>
<td>1.2350</td>
<td>0.1750</td>
<td>0.03532</td>
<td>0.2490</td>
</tr>
<tr>
<td>110005</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110006</td>
<td>0.7796</td>
<td>0.2194</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110007</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110008</td>
<td>1.3139</td>
<td>0.1719</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110009</td>
<td>1.0272</td>
<td>0.1928</td>
<td>0.5784</td>
<td>0.2517</td>
<td>-0.4487</td>
<td>0.3170</td>
</tr>
<tr>
<td>110010</td>
<td>1.9114</td>
<td>0.2355</td>
<td>2.0910</td>
<td>0.2911</td>
<td>0.1796</td>
<td>0.3744</td>
</tr>
</tbody>
</table>

As we see an estimate of the change \( \delta = \theta_2 - \theta_1 \) itself and its standard error is provided. In this case the estimated change is unavailable for most individuals, the ones with minimum or maximum scores at one of the time points.
ICC’s based on the estimated item parameters can again be plotted by the macro \texttt{\%lirt\_icc}, this time using the macro call

\begin{verbatim}
\%lirt_icc( NAMES=lgpcm_names,
            DIM=2,
            OUT=icc);
\end{verbatim}

This produces ICC’s for each set of item parameters resulting in 4 curves in this simple model with no item parameter drift.

5.3. Models allowing for item parameter drift

Extending the simple longitudinal GPCM specified by \texttt{names3} can be done by allowing for item parameter drift e.g. for the item \texttt{awake}. This corresponds to specifying the subset \( J_0 = \{ \text{fallas, sleepba, wokenup} \} \). Items appearing in separate records in the \texttt{names} data set will be modeled with their own set of item parameters by the macro \texttt{\%lirt\_mml}. Therefore a model with item parameter drift for the \texttt{awake} item is specified by creating the \texttt{names} data set

\begin{verbatim}
data names4;
   input name1 $ name2 $ max disc_yn $;
datalines;
  awake1 . 3 Y
  . awake2 3 Y
  fallas11 fallas12 3 Y
  sleepba1 sleepba2 3 Y
  wokenup1 wokenup2 3 Y
; run;
\end{verbatim}

We fit the GPCM with the macro call
%lirt_mml(DATA=sleep1_2,
NAMES=names4,
DIM=2,
OUT=drift);

By comparing the likelihood value stored in drift_log1 to that of simple model stored in lgpcm_log1, a likelihood ratio test of item parameter drift for the awake item can be conducted.

5.4. Models allowing for local dependence across time points

A model allowing for local dependence between the items awake1 and awake2, corresponding to \( I_0 = \{ \text{sleepba, fallas1, wokenup} \} \), can be fitted by splitting the item awake2 for dependence. This can be done using the macro %lirt_split that requires the simple data set names3 as input. The macro returns a data set with the original and the split items, the latter named cat0_awake2, cat1_awake2, cat2_awake2 and cat3_awake2 with the prefix indicating the corresponding item response at time 1. Fitting a model including the split items using the macro %lirt_mml requires a names data set with the corresponding item names. A limitation of using the datalines-statement to create this data set is that name2 can only be 8 characters long, which conflicts with the longer item names given by %lirt_split to the split items. The names data set can easily be created without the datalines-statement but here we opt for a bit of data management instead. We need to modify names3 as follows

```sas
data names3;
set names3;
if name2='awake2' then name2='aw2';
run;
```

and to make a copy of the variable awake2 named aw2 in the data set sleep1_2
data sleep1_2;
set sleep1_2;
aw2=awake2;
run;

The macro call

```sas
%lirt_split( DATA=sleep1_2,
    NAMES=names3,
    DIM=2,
    INDEP=awake1,
    DEP=aw2);
```

splits the item(s) in DEP according to the item(s) in INDEP in the order they appear. The macro call creates a data set named `sleep1_2_split` that is a copy of the original data set with the addition of the items `cat0_aw2`, `cat1_aw2`, `cat2_aw2` and `cat3_aw2` resulting from the item split. The first ten records of `sleep1_2_split` are given as

<table>
<thead>
<tr>
<th>idnr</th>
<th>sleepba</th>
<th>fallsa</th>
<th>wokenu</th>
<th>awake1</th>
<th>sleepba2</th>
<th>fallsa2</th>
<th>wokenu2</th>
<th>awake2</th>
<th>aw2</th>
<th>aw2</th>
<th>aw2</th>
<th>aw2</th>
</tr>
</thead>
<tbody>
<tr>
<td>110001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110002</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110003</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110004</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>1</td>
</tr>
<tr>
<td>110005</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110006</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110007</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110008</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110009</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>110010</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>2</td>
</tr>
</tbody>
</table>

Each individual in the data set is represented by exactly one of the items `cat0_aw2`, `cat1_aw2`, `cat2_aw2` and `cat3_aw2` depending on their response to the item `awake1`. Thus a lot of missing values are present in the data set `sleep1_2_split`. The model-specifying data set for the longitudinal GPCM with local dependence between `awake1` and `awake2` can be created as follows
data names5;
    input name1 $ name2 $ max disc_yn $;
    datalines;
    awake1 . 3 Y
    . cat0_aw2 3 Y
    . cat1_aw2 3 Y
    . cat2_aw2 3 Y
    . cat3_aw2 3 Y
    fallasl1 fallasl1 3 Y
    sleepba1 sleepba2 3 Y
    wokenup1 wokenup2 3 Y
    ;
run;

Note that the original item awake2 is not included. The data set names5 holds no information about which item is dependent on which. It only tells the macro %lirt_mml to model awake1 and the split items with their own set of item parameters (recall that each record corresponds to a set of item parameters). The dependence structure has been taken care of in the splitting procedure. Thus, with the dependence assumption built into the data sets sleep1_2_split and names5 we simply fit the model using the macro call

```sas
%lirt_mml( DATA=sleep1_2_split,
           NAMES=names5,
           DIM=2,
           OUT=split);
```

As before the likelihood value, the discrimination parameters and the threshold parameters are stored in data sets. Furthermore a likelihood ratio test for local dependence between the awake items at time point 1 and 2 can be conducted using the likelihood values stored in split_logl and lgpcm_logl respectively.
5.5. Models allowing for item parameter drift and local dependence across time points

More general models can be specified and fitted in order to assess possible item parameter drift and local dependence across time points by adjusting the data set names. This is done by changing the structure of the variables name1 and name2. For instance a model with local dependence across time points for the item awake and item parameter drift for the item sleepba can be specified by creating the names data set

```plaintext
data names6;
  input name1 $ name2 $ max disc_yn $;
datalines;
  awake1 . 3 Y
  . cat0_aw2 3 Y
  . cat1_aw2 3 Y
  . cat2_aw2 3 Y
  . cat3_aw2 3 Y
  fallasl1 fallasl2 3 Y
  sleepba1 . 3 Y
  . sleepba2 3 Y
  wokenup1 wokenup2 3 Y
;
run;
```

which can then be used as input to %lirt_mml.

5.6. Tests of fit

Being able to fit the longitudinal models described in the previous sections provides a flexible tool for detecting item parameter drift and local dependence across time points. Likelihood ratio tests can be carried out by comparing the likelihood values, stored in the output data sets OUT_logl, for nested models.

Individual item fit can also be evaluated graphically by comparing observed and simulated mean scores as illustrated in the unidimensional case, cf. Figure 4. In order to
briefly describe how to simulate in the longitudinal case we return to the simple model from section 5.2 specified by the data set names3. Again two input data sets are needed. The output data set lgpcm_names from %lirt_mml with the estimated item parameters in principle holds the required information. However, because %lirt_simu can also simulate from models with local dependence across time points, a structure that needs to be specified in the names data set, two variables ld_group and ld_item has to be added

```sas
data lgpcm_names;
   set lgpcm_names;
   LD_GROUP=.;
   LD_ITEM=.;
run;
```

In this case the two variables have only missing values. Had the objective been to simulate from the model with local dependence across time points represented by names5 (cf. section 5.4) the variables ld_group and ld_item should again have been added resulting in the data set

<table>
<thead>
<tr>
<th>name1</th>
<th>name2</th>
<th>max</th>
<th>disc yn</th>
<th>ld_item</th>
<th>ld_group</th>
</tr>
</thead>
<tbody>
<tr>
<td>awake1</td>
<td></td>
<td>3</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cat0_aw2</td>
<td></td>
<td>3</td>
<td>Y</td>
<td>awake1</td>
<td>0</td>
</tr>
<tr>
<td>cat1_aw2</td>
<td></td>
<td>3</td>
<td>Y</td>
<td>awake1</td>
<td>1</td>
</tr>
<tr>
<td>cat2_aw2</td>
<td></td>
<td>3</td>
<td>Y</td>
<td>awake1</td>
<td>2</td>
</tr>
<tr>
<td>cat3_aw2</td>
<td></td>
<td>3</td>
<td>Y</td>
<td>awake1</td>
<td>3</td>
</tr>
<tr>
<td>fallas11</td>
<td>fallas12</td>
<td>3</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sleepba1</td>
<td>sleepba2</td>
<td>3</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wokenup1</td>
<td>wokenup2</td>
<td>3</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For a dependent time 2 item in name2 the variable ld_item tells the macro which item it is dependent on. The variable ld_group tells the macro which group of respondents the item (resulting from the splitting procedure) represent. Hence, the values of ld_group are the possible scores on the ld_item.
The other input data set `pdata` can also in the longitudinal case take two different forms. The first one needs the variables `id`, `theta1` and `theta2`. That is, a unique person ID and the corresponding (estimated) values of the latent variable at each time point. All data sets generated by the macro `%lirt_simu` are based on these fixed values of the latent variable and the item parameters given in `names`. For the first 10 respondents in the data set `sleep1_2`, with estimates (based on the simple model) of the latent variable at both time points, `pdata` should look as follows

```
id   theta1   theta2
110004 1.1997 1.2350
110009 1.0272 0.5784
110010 1.9114 2.0910
```

Using this type of data set again needs to be combined with the option `NPERSONS=0`. The other form of `pdata` specifies the latent normal distribution, for the simple model as follows

```
data pdata;
    input PARAMETER $ ESTIMATE;
datalines;
    mu1 0
    mu2 -0.3428
    sigma1 1
    sigma2 1.1219
    rho 0.8279
;
run;
```

Recall that in this case `NPERSONS` tells `%lirt_simu` how many respondents should be drawn from this distribution. A new set of values of the latent variables are drawn from this distribution for each new generated data set. Using the latter version of `pdata`, four data sets can be simulated using the macro call
Note that the macro call is identical to that of the unidimensional case, except for the option \texttt{DIM=2}. Plots of the type shown in Figure 4 can be made using code similar to that of the unidimensional case.

\section*{Discussion}

Unidimensional IRT models are frequently used when items are intended to measure a latent variable. Several specialized software packages for fitting unidimensional IRT models, in particular the Rasch model, are available. The proprietary software packages \textit{RUMM} (Andrich, Sheridan, and Luo 2010) and \textit{WINSTEPS} (Linacre 2011) for fitting and testing Rasch models are widely used. Implementations in more general statistical software such as \textit{SAS} and \textit{R} also exist.

When a set of items is used at several time points to measure change in a latent variable longitudinal extensions of unidimensional IRT models are called for. Formulations of the two-dimensional Rasch model was first discussed by Andersen (1985) and Embretson (1991). They both rely on two assumptions, that of no item parameter drift and that of local independence across time points. However, these assumptions might not be justified and need to be checked. Tests, based on unidimensional IRT models, for detecting violations of these have been suggested by Olsbjerg and Christensen (2013a). Furthermore, longitudinal IRT models allowing for item parameter drift and local dependence have been proposed (Olsbjerg and Christensen 2013b).

Despite the fact that many applications deal with multidimensional or longitudinal data \textit{RUMM} and \textit{WINSTEPS} fit unidimensional models only.

We presented in this paper an implementation in \textit{SAS} consisting of the four main \textit{SAS} macros \texttt{%lirt\_mml}, \texttt{%lirt\_ppar}, \texttt{%lirt\_icc} and \texttt{%lirt\_simu}. The macros constitute a flexible tool for fitting and testing against each other various specifications of the longitudinal
GPCM. The specification of a model is built into a data set names which is a key component of all four macros.

The implementation allows for item parameter drift and local dependence across time points for subsets of items. Being able to actually fit these models makes it easy to test for these two characteristics with likelihood ratio tests, using the macro %lirt_mml, thus adding to existing methods for detection of item parameter drift (Donoghue and Isham 1998; DeMars 2004; Galdin and Laurencelle 2010). The macro %lirt_simu provides another tool for testing the fit since it can be used to compare various features of the observed data with what would be expected under the model. In the example the names data set for the macroo %lirt_simu is created by programming one line in the data set at a time. In a situation where many items are analysed, potentially with drift and local dependence, doing this becomes a bit tedious. Therefore a help macro %lirt_simu_names (not used in the example) facilitating this procedure is available on the website for the macros.

Ways of extending this implementation would be to accommodate other IRT models such as the Rating Scale Model (Andrich 1978) and the Graded Response Model (Samejima 1997). It would also be of interest to fit models for measurements at more than two time points including various specifications of the latent correlation matrix. In that context local dependence structures that go beyond two consecutive time points can be considered.
References


Birnbaum AL (1968). “Latent trait models and their use in inferring an examinee’s ability.” In FM Lord, MR Novick (eds.), *Statistical Theories of Mental Test*. Addison-Wesley, Reading, MA.


LIRT: SAS macros for longitudinal IRT models


11/1  Kreiner, S. Is the foundation under PISA solid? A critical look at the scaling model underlying international comparisons of student attainment.

11/2  Andersen, P.K. Competing risks in epidemiology: Possibilities and pitfalls.

11/3  Holst, K.K. Model diagnostics based on cumulative residuals: The R-package gof.


11/5  Holst, K.K., Budtz-Jørgensen, E. & Knudsen, G.M. A latent variable model with mixed binary and continuous response variables.


12/01 Keiding, N., Hansen, O.K.H., Sørensen, D.N. & Slama, R. The current duration approach to estimating time to pregnancy.

12/02  Andersen, P.K. A note on the decomposition of number of life years lost according to causes of death.
12/03 Martinussen, T. & Vansteelandt, S. A note on collapsibility and confounding bias in Cox and Aalen regression models.


12/05 Martinussen, T. & Pipper, C.B. Estimation of Causal Odds of Concordance based on the Aalen additive model.

12/06 Binder, N., Gerds, T.A. & Andersen, P.K. Pseudo-observations for competing risks with covariate dependent censoring.

12/07 Keiding, N. & Clayton, D. Standardization and control for confounding in observational studies: a historical perspective.

12/08 Hilden, J. & Gerds, T.A. Evaluating the impact of novel biomarkers: Do not rely on IDI and NRI.


13/01 Forman, J.L. & Sørensen M. A new approach to multi-modal diffusions with applications to protein folding.


13/03 Wolbers, M., Koller, M.T., Witteman, J.C.M. & Gerds, T.A. Concordance for prognostic models with competing risks.

13/04 Andersen, P.K., Canudas-Romo, V. & Keiding, N. Cause-specific measures of life lost.

13/05 Keiding, N. Event history analysis.


13/07 Hansen, S., Andersen, P.K. & Parner, E. Events per variable for risk differences and relative risks using pseudo-observations.


13/11 Olsbjerg, M. & Christensen K.B. SAS macro for marginal maximum likelihood estimation in longitudinal polytomous Rasch models.


14/01 Ambrogi, F. & Andersen, P.K. Predicting Smooth Survival Curves through Pseudo-Values.

14/02 Olsbjerg, M. & Christensen, K.B. Modeling local dependence in longitudinal IRT models.

14/03 Olsbjerg, M. & Christensen, K.B. LIRT: SAS macros for longitudinal IRT models.