

A Bayesian IRT model for mixed responses and skew latent variable

Serena Broccoli and Giulia Cavrini

Abstract

The aim of this study is to formulate a flexible Item Response Theory (IRT) based model for mixed responses questionnaires and relaxing the hypothesis of normal distributed latent variable. The new model is a combination of two models already presented in literature, that is, a latent trait model for mixed responses and an IRT model for Skew Normal latent variable. It is developed in a Bayesian framework, a Markov chain Monte Carlo procedure is used to generate samples of the posterior distribution of the parameters of interest. The proposed model is test on a questionnaire composed by 5 discrete items and one continuous to measure HRQoL in children, the EQ-5D-Y questionnaire. A large sample of children collected in the schools was used. In comparison with a model for only discrete responses and a model for mixed responses and normal latent variable, the new model has better performances, in term of deviance information criterion (DIC), chain convergences times and precision of the estimates.

Key words: Bayesian IRT model, Mixed Responses, Skew-Normal Latent Variable

1 Introduction

Item Response Theory (IRT) models are based on the assumption that the observed items of a questionnaire are measuring an underlying (latent) construct, that is, manifest responses are assumed to reflect the location of a given subject on a continuum which is not directly observable (such as intelligence, depressed mood or health-related quality of life). The parameters of these psychometric measurement models are estimated by using statistical techniques derived from the specific assumptions that

Serena Broccoli, Department of Statistics; email: serena.broccoli@gmail.com

Giulia Cavrini, Department of Statistics; email: giulia.cavrini@unibo.it

constitute a particular model. The HRQoL scores are called person parameters, and the tariffs associated with each level of a given item are called item parameters. Item parameters are mathematically treated such as real parameters of a latent trait model in which person parameters are the latent trait.

IRT originated in the 1960s (Lord & Novick, 1968) (Rasch, 1960) (van der Linden & Hambleton, 1997) and over the years a large number of researchers have contributed further to the subject. The general idea is that the probability of a certain answer is a function of a latent trait value (θ), the HRQoL in this case, plus one or more parameters characterizing the particular item. For each item the probability of a certain answer as a function of the latent trait is called Item Characteristic Curve (ICC) or Item Response Function (IRF).

Let θ be a latent variable defined on a scale from minus infinity to plus infinity. It might be a cognitive ability, a physical ability, a skill, knowledge, attitude, satisfaction or, more generally, any unobservable variable that describes a state or condition of a person.

The probability of a “correct response”¹, or more generally the answer coded as 1, for subject i on a dichotomous item j , given the latent trait (usually called person parameter) is equal to

$$P(y_{ij} = 1 / \theta_i) = P_{ij}(\theta_i) = c_j + (1 - c_j) \frac{\exp[a_j(\theta_i - b_j)]}{1 + \exp[a_j(\theta_i - b_j)]}$$

where a_j , b_j and c_j are the item parameters.

b_j represents the item difficulty, that is the value of the latent trait for which the probability of a correct answer is equal to the probability of an incorrect answer. If a high number of subjects answer an item correctly it is considered easier than an item to which only a small number of subjects answer correctly.

a_j is the discrimination parameter of item j . In fact, in many tests the items do not differ only in respect to difficulty but also in respect to discriminating power. It indicates the degree to which an item distinguishes between the different levels of latent variable across the continuum.

c_j is the pseudo-chance level parameter. It is the probability of answering correctly when the latent trait level is very low, that is, the probability of guessing. This parameter is specific for cognitive tests, where there is the possibility that a student answers item correctly because he/she guesses; in the context of HRQoL it is more difficult to imagine the use of this kind of parameter.

The 2 parameter logistic model (2-PLM) is a special case of the 3 parameter logistic model (3-PLM) with $c_j=0$ for all items. The 1 parameter logistic model (1-PLM or Rasch Model) is a special case of the 2 parameter logistic model with $a_j=1$ for all items. The logistic models are valid only for tests composed of dichotomous items. While, the items of a questionnaire frequently have a polytomous form, that is, more than two possible answers. Master in 1982 presented the Partial Credit Model which can be applied in any situation in which performances on an item are recorded in two or more ordered categories.

The mathematical formulation of the PCM comes from an extension of the 1-PLM. The probability of subject i scoring x to item j (item with k_j levels of answer), given the latent variable θ is

¹ It is defined as “correct” because the terminology is largely influenced by educational field, where IRT has been mainly used.

$$P_{ijk}(\theta_i) = \frac{\exp \sum_{t=0}^x (\theta_i - b_{jt})}{\sum_{k=0}^{k_j} \exp \sum_{t=0}^k (\theta_i - b_{jt})}$$

where $b_{j0}=0$ and

$$\sum_{t=0}^0 (\theta_i - b_{jt}) = 0$$

The interpretation of the parameters is similar to the interpretation of the 1-PLM parameters, except for the fact that in the case of the PCM the number of item parameters for each item is greater than one. In fact, an item parameter is estimated for each level of answer, except for the first category which is usually arbitrarily selected to be the reference category. If for item j there are k_j levels of answer, there will be k_j-1 item parameters estimated. They represent the value of the latent variable at which the probability of answering x or $x-1$ is the same.

It is necessary to formulate some assumptions because the IRT holds:

- Unidimensionality: the test of interest must measure only one latent construct, that is, that the covariance among the items can be explained by a single underlying dimension.
- Local independence: item responses are independent given the person parameter. Local dependence can potentially arise among subsets of items that have a similar stem (e.g., a set of items that all refer to someone's experience of physical pain), items that have very similar content, or items that are presented sequentially. If this occurs, one item from the pair should be removed and the scale should be recalibrated.
- Monotonicity of ICCs: item characteristic curves p_{ij} of all items have a continuous, strictly monotone increasing trend for all latent trait parameters θ .
- No possibility of cheating (except for the 3-PLM logistic model): the 1- and 2-PLM models and the PCM do not model the possibility of cheating. It is assumed that it is not possible to cheat.
- Normal distributed latent variable: when we have to make any assumption with respect to the distributional form of a variable, but we do not have any information, a natural choice is a more or less bell-shaped distribution. Traditionally the obvious choice is the normal distribution. Furthermore, to overcome identifiability problems the mean of the latent variable is usually set at 0 and the standard deviation at 1.

The main procedures used to estimate the parameters of IRT models are the Joint Maximum Likelihood (JML) estimation, the Conditional Maximum Likelihood (CML) estimation and the Marginal Maximum (MML) Likelihood estimation. (Details in Baker, 1992). Further to the methods based on the maximization of the likelihood function, Bayesian estimation methods are frequently used.

A wide literature on IRT is available and many IRT models have been developed. In this paper an IRT model for mixed responses (continuous or polytomous) is proposed. The model is developed under the Bayesian framework and it includes the possibility to relax the assumption of Normal distribution of the latent variable. The model arise from the combination of the IRT model for mixed responses presented by Moustaki in 2000 and of the IRT model for skewed latent trait presented by Bazan and colleagues in 2006.

2 Methods

2.1 The model

Suppose a questionnaire is composed of r continuous items, s dichotomous items and c polytomous items, where $r + s + c = q$ (total number of items). Letting w_{ij} with $j=1\dots r$ be the answer of subject i to the continuous item j , v_{ij} with $j=r+1\dots r+s$ being the answer of subject i to the dichotomous item j and t_{ij} with $j=r+s+1\dots r+s+c$ being the answer of subject i to the polytomous item j with ordered categories.

$$\begin{aligned} w_{ij} &\sim N(\mu_{ij}, \sigma^2) & j &= 1, \dots, r & i &= 1, \dots, N \\ v_{ij} &\sim \text{Ber}(\mu_{ij}) & j &= r+1, \dots, r+s & i &= 1, \dots, N \\ t_{ij} &\sim \text{Multinomial}(\mu_{ij}, 1) & j &= r+s+1, \dots, r+s+c & i &= 1, \dots, N \end{aligned}$$

We assume the existence of a latent variable θ_i such that

$$\begin{aligned} \mu_{ij} &= a_j \theta_i - b_j & j &= 1 \dots r \\ \mu_{ij} &= P(v_{ij} = 1) = \frac{\exp[a_j(\theta_i - b_j)]}{1 + \exp[a_j(\theta_i - b_j)]} & j &= r+1 \dots r+s \\ \mu_{ij} &= \left[\mu_{ijk} = \frac{\exp \sum_{x=0}^k (\theta_i - b_{jx})}{\sum_{k=0}^{k_j} \exp \sum_{x=0}^k (\theta_i - b_{jx})}, x = 1, \dots, k_j \right] & j &= r+s+1 \dots r+s+c \end{aligned}$$

It is assumed that θ comes from a asymmetric normal distribution, the Skew Normal distribution (Azzalini, 1985)

$$\theta_i \sim SN(\alpha, \beta, \delta) \quad i = 1, \dots, N$$

Identifiability constraints on the parameters are necessary, in fact., similarly to classical IRT model, the latent variable is assumed to be standardized. A centred parameterization (Azzalini, 1986) of the Skew Normal distribution enables the specification of the mean and the variance of the latent variable, usually equal to 0 and 1 respectively. The problem was to derive the mean and the variance of a “not centred variable” and in order to impose this restriction.

Given θ the skew-normal distributed latent variable with mean μ and variance σ^2 , and $Z \sim SN(0, 1, \delta)$ the standard skew normal distribution, the following relation holds

$$\theta = \alpha + \beta Z = \mu + \sigma \left[\frac{Z - E[Z]}{\sqrt{\text{Var}[Z]}} \right]$$

where

$$E[Z] = \left(\frac{2}{\pi} \right)^{\frac{1}{2}} \delta$$

$$V[Z] = \left(1 - \frac{2}{\pi} \delta^2\right)$$

Imposing $\mu=0$ and $\sigma=1$ and conditioning on $Z \sim SN(0, 1, \delta)$, the specification of a skew latent variable centred on its mean and with unit variance is obtained:

$$\theta = \frac{Z - \left(\frac{2}{\pi}\right)^{1/2} \delta}{\sqrt{\left(1 - \frac{2}{\pi} \delta^2\right)}}$$

Since several statistical software do not allow to define a variable as Skew Normal, the conditional distribution of θ on z will be expressed using its probabilistic representation (Henze, 1986)

$$Z = \delta|W| + \sqrt{1 - \delta^2}V$$

where $|W| \sim HN(0, 1)$ and $V \sim N(0, 1)$ are independent random variables.

Thus

$$\begin{aligned} \theta &= \alpha + \beta(\delta|W| + \sqrt{1 - \delta^2}V) = \\ &= \mu + \sigma \left[\frac{\delta|W| + \sqrt{1 - \delta^2}V - E[Z]}{\sqrt{Var[Z]}} \right] = \\ &= \mu + \frac{\sigma\delta|W|}{\sqrt{Var(Z)}} + \frac{\sqrt{1 - \delta^2}\sigma V}{\sqrt{Var(Z)}} - \frac{E[Z]\sigma}{\sqrt{Var(Z)}} \end{aligned}$$

The conditional distribution of the latent variable then becomes a Normal distribution with mean and variance defined as follow

$$\begin{aligned} \theta|W, \delta &\sim N\left(\mu + \frac{\sigma\delta|W|}{\sqrt{Var(Z)}} + 0 - \frac{E[Z]\sigma}{\sqrt{Var(Z)}}, \frac{(1 - \delta^2)\sigma^2}{Var(Z)}\right) \\ \theta|W, \delta &\sim N\left(\mu + \frac{1}{\sqrt{\left(1 - \frac{2}{\pi}\delta^2\right)}}\delta\sigma|W| + 0 - \frac{\left(\frac{2}{\pi}\right)^{\frac{1}{2}}}{\sqrt{\left(1 - \frac{2}{\pi}\delta^2\right)}}\delta\sigma, \frac{(1 - \delta^2)\sigma^2}{\left(1 - \frac{2}{\pi}\delta^2\right)}\right) \end{aligned}$$

Fixing the unconditional mean of θ equal to 0 and the variance equal to 1 the distribution becomes

$$\theta|W, \delta \sim N\left(\frac{1}{\sqrt{\left(1 - \frac{2}{\pi}\delta^2\right)}}\delta|W| - \frac{\left(\frac{2}{\pi}\right)^{\frac{1}{2}}}{\sqrt{\left(1 - \frac{2}{\pi}\delta^2\right)}}\delta, \frac{(1 - \delta^2)}{\left(1 - \frac{2}{\pi}\delta^2\right)}\right)$$

though it depends on the skewness parameter and the Half Normal random variable.

The marginal joint distribution of the observed variables y_{ij} (w_{ij} , t_{ij} and v_{ij}) is the same as expressed by Moustaky (2000)

$$g(y_i | \theta_i) = \prod_{j=1}^q g(y_{ij} | \theta_i) = \prod_{j=1}^r h(w_{ij} | \theta_i) \prod_{j=r+1}^{r+s} k(v_{ij} | \theta_i) \prod_{j=r+s+1}^{r+s+c} l(t_{ij} | \theta_i)$$

Consequently, the log-likelihood can be expressed as

$$\log L = \sum_{i=1}^n \log f(y_i) = \sum_{i=1}^n \log \int_{-\infty}^{+\infty} g(y_i | \theta) h(\theta) d\theta$$

where $h(\theta)$ is the Skew Normal distribution function of mean 0 and variance 1.

2.2 Estimation Procedure

Having defined the likelihood function, it is possible to develop a Bayesian estimation procedure by incorporating prior distributions for the parameters of interest, which are the person parameter θ_i , the item parameters a (q -dimensional vector) and b ($\sum_{j=1}^q k_j$ - dimensional vector¹).

As a hypothesis the latent variable is supposed to be a skew normal random variable, with mean 0 and variance 1, that can be expressed in its centred parameterization (cp) as

$$\theta_i \sim SN_{cp}(0, 1, \delta) \quad i = 1 \dots N$$

The prior distributions for the item parameters have been chosen as locally non-informative normal priors

$$\begin{aligned} a_j &\sim N(0, 100) \quad j = 1 \dots q \\ b_j &\sim N(0, 100) \quad j = 1 \dots q \end{aligned} \quad (3.28)$$

The last prior distribution must be specified only for continuous item, because they depend not only on item and person parameters but also on the variance of the normal distribution. The conjugate prior for the variance of a normal distribution has been chosen.

$$\sigma_j \sim invgamma(10, 10) \quad j = 1 \dots r$$

The only hyperparameter is the skewness parameter of the Skew Normal distribution. Following Bazan et. al. (2006)

$$\delta \sim U(-1, 1)$$

The posterior distribution of the parameters of the model becomes

$$\begin{aligned} p(\omega_i | y_i) &\propto \prod_{j=1}^r g(w_{ij} | \omega_i) h(a_j) \prod_{j=r+1}^{r+z} g(v_{ij} | \omega_i) \prod_{j=r+z+1}^{r+z+f} g(t_{ij} | \omega_i) \prod_{j=1}^q h(a_j) h(b_j) h(\theta_i | \delta) r(\delta) \end{aligned}$$

With Gibbs sampling procedures, random samples of person and item parameters can be drawn from the full conditionals derived from the posterior distribution specified for this hierarchical model.

3 Application

The IRT model presented in is used principally to obtain a continuum synthetic quality of life score for each child who compiled the EQ-5D-Y questionnaire. It defines the perceived health related quality of life in children and adolescents according to five dimensions: mobility, self care, usual activities, pain/discomfort, and anxiety/depression. Each dimension is described by one specific level of adequacy: no problem, moderate problems or severe problems. The EQ-5D questionnaire also includes a Visual Analogue Scale (VAS), by which respondents can attribute their perceived

¹ With the number of levels of answer (k) for the continuous items fixed equal to 1.

health status a grade ranging from 0 (the worst possible health status) to 100 (the best possible health status). The five items are polytomous items and the VAS score is a continuous item. The assumption of normal distribution of the quality of life score must be relaxed in favour of a skew normal distribution. In fact, few children coming from a school population are expected to exhibit low values of quality of life, thus the latent variable is supposed to be skewed to the right.

Bayesian parameter estimates were obtained using Gibbs sampling algorithms as implemented in the computer program WinBUGS 1.4 (Spiegelhalter, Thomas, Best, & Lunn, 2003). The value taken as the MCMC estimate is the mean over iterations sampled starting with the first iteration following burn-in. The R-Package CODA (Best, Cowles, & Vines, 1995) was used to compute convergence diagnostics.

Three different IRT models were compared: Model 1: Partial Credit Model (5 items used); Model 2: IRT model for mixed responses (5 items and VAS used); Model 3: IRT for mixed responses and skew latent variable (5 items and VAS used)

In order to check if an improvement of the fit measures will occur, the models were compared using the Deviance Information Criterion (DIC), an index proposed by Spiegelhalter, Best, Carlin, & Van der Linde (2002) and implemented in WinBUGS.

Model 2 and Model 3 are able to distinguish between children in terms of HRQoL better than the PCM and they should be preferred when analyzing data from the EQ-5D-Y questionnaire, especially when a large number of children answer “No problem” in all the items.

Still, some other aspects must be taken in account when choosing between the three models, particularly between Model 2 and Model 3 which seem to perform similarly in terms of subject ranking. The performance in chain convergence is one of these aspects. The convergence times could be quite long if a lot of children are in the sample, because an increase in one unit in the sample increases by one unit the number of parameters to be estimated. Model 2 takes a substantially longer time to reach convergence than Model 3. The misspecification of the a priori for the latent variable (assumed to be normal) might be the cause of the Model 2 delay. Secondly, as introduced, the DIC measure for the fit of the model has been used. Table 1 shows the DIC values for Models 1, 2 and 3. Since the lower the DIC value, the higher the performance of the model, the best is Model 3, but the difference between Model 2 and Model 3 is lower than 5 and it could be misleading just to report the model with the lowest DIC (Spiegelhalter, Best, Carlin, & Van der Linde, 2002).

Table 1: Model 1, Model 2 and Model 3 DIC values

<i>Model</i>		<i>DIC</i>
1	PCM	181.1
2	PCM + VAS	133.3
3	PCM + VAS + skewed normal latent variable a priori	129.8

In short, the several aspects evaluated in this section suggest Model 3 as the best of one between the three studied models for analyzing the EQ-5D-Y questionnaire.

4 Conclusion

A flexible IRT model for mixed responses and skew latent variable is presented as well as a Bayesian method for the estimation of the parameters. A comparison between classical IRT models and the model presented on a real dataset shows that the use of all type of items in the questionnaire at the same time and the specification of the right a priori for the latent variable improves the fit of the model and reduces the estimation times.

5 References

1. Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* , 12:171-178.
2. Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica* , 46:199-208.
3. Baker, F. (1992). *Item response theory: parameter estimation techniques*. New York: Dekker.
4. Bazan, J., Branco, M., & Bolfarine, H. (2006). A Skew Item Response Model. *Bayesian Analysis* , 1(4):861-892.
5. Henze, N. (1986). A probabilistic representation of the 'skew-normal' distribution. *Scand. J. Statist.* , 271-275.
6. Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores*. MA: Adisson-Wesley.
7. Moustaki, I., & Knott, M. (2000). Generalized latent trait model. *Psychometrika* , 65(3), 391-411.
8. Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmarks Paedagogiske Institut (Reprinted by University of Chicago Press, 1980).
9. Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology* , 64:583-616.
10. Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2003). *WinBUGS 1.4 User manual*.
11. van der Linden, W., & Hamblenton, R. (1997). *Handbook of modern item response theory*. New York: Springer.