Multidimensional latent class item response models for binary and ordinal polytomous items: theory and application with R software

Silvia Bacci

Dipartimento di Economia - Università di Perugia

1silvia.bacci@stat.unipg.it
1. Introduction

2. Types of parameterizations for IRT models

3. Multidimensional latent class IRT models

4. Package MultiLCIRT

5. References
Starting point

Item Response Theory (IRT) models (Hambleton and Swaminathan, 1985; Van der Linden and Hambleton, 1997) are increasingly used to the assessment of individuals’ latent traits.

They allow us to translate the qualitative information coming from the questionnaire in a quantitative measurement of the latent trait.

IRT models build on the central idea that the probability to provide a certain answer to an item is a function of the person’s position on the latent trait and one or more parameters which characterize the item.
Basic notation

- $i$: generic individual who answers the test items, $i = 1, \ldots, n$
- $j$: generic test item, $j = 1, \ldots, J$
- $Y_{ij}$: response variable for the $j$-th item and $i$-th person
- $y$: generic item response category, $y = 0, \ldots, l_j - 1$ (special case: $l_j = 2$)
- $\Theta$: latent trait or latent variable, which assumes value $\theta_i$ for individual $i$
- $p_{jy}(\theta_i) = p(Y_{ij} = y|\theta_i)$: item response function (IRF) or item response category characteristic curve (IRCCC)
  - conditional probability of responding with category $y$ to item $j$ by an individual $i$ with latent trait level $\theta_i$
  - according to the way $p_{jy}(\theta_i)$ is specified different IRT models are defined
  - in case of dichotomously-scored items, $p_{j1}(\theta_i) = p(Y_{ij} = 1|\theta_i)$ is modeled and it is named item characteristic curve (ICC)
Main assumptions

- Local independence
- Unidimensionality of latent trait
- Monotonicity of the cumulative probabilities

\[ p^*_jy(\theta_i) = p(Y_{ij} \geq y \mid \theta_i), \quad y = 1, \ldots, l_j - 1, \]

As far as dichotomously-scored items this assumption reduces to the monotonicity of \( p_{j1}(\theta_i) \)
What’s about $\Theta$?

- No specific assumption about $\Theta$ is necessary to define the IRT models, however . . .
- According to the estimation method, a parametric assumption could be necessary
- Two main approaches:
  - Fixed effect approach: $\theta_i \ (i = 1, \ldots, n)$ is a fixed but unknown parameter, which is estimated with the other model parameters
  - Random effect approach: $\theta_i$ is a realization of a random variable having a certain distribution in the population from which the observed sample of individuals has been drawn
    - Continuous distribution, typically the normal one (parametric approach)
    - Discrete distribution (semi-parametric approach)
Types of parameterizations for IRT models

**General formulation for IRT models**

\[ g_y[p_j(\theta_i)] = \lambda_j(\theta_i - \beta_{jy}), \quad j = 1, \ldots, J, \ y = 1, \ldots, l_j - 1, \tag{1} \]

- \( p_j(\theta_i) = (p_{j0}(\theta_i), \ldots, p_{j,l_j-1}(\theta_i))' \)
- \( g_y(\cdot) \): link function specific for category \( y \)
  - In case of dichotomously-scored items \( g_y(\cdot) \) reduces to a logit link
- \( \lambda_j \): (item) discriminating parameter
- \( \beta_{jy} \): (item threshold or step or category boundary) difficulty parameter
  - In case of dichotomously-scored items \( \beta_{jy} = \beta_j \)
Classification criteria

**Type of link function** (in case of ordinal responses)

- **Global logit** (or cumulative odds logit) compares the probability that item response is in category $y$ or higher with the probability that it is in a lower category

$$g_y[p_j(\theta_i)] = \log \frac{p(Y_{ij} \geq y|\theta_i)}{p(Y_{ij} < y|\theta_i)} = \log \frac{p_{jy}(\theta_i) + \cdots + p_{j,l-1}(\theta_i)}{p_{j0}(\theta_i) + \cdots + p_{j,y-1}(\theta_i)}$$

- **Local logit** (or adjacent category logits) compares the probability of each category $y$ with the probability of the previous category, $y-1$

$$g_y[p_j(\theta_i)] = \log \frac{p(Y_{ij} = y|\theta_i)}{p(Y_{ij} = y-1|\theta_i)} = \log \frac{p_{jy}(\theta_i)}{p_{j,y-1}(\theta_i)}$$

- **Continuation ratio logit** compares the probability of a response in category $y$ or higher with the probability of the previous category $y-1$

$$g_y[p_j(\theta_i)] = \log \frac{p(Y_{ij} \geq y|\theta_i)}{p(Y_{ij} = y-1|\theta_i)} = \log \frac{p_{jy}(\theta_i) + \cdots + p_{j,l-1}(\theta_i)}{p_{j,y-1}(\theta_i)}$$
Classification criteria

Specification of item discrimination parameters

- $\lambda_{j_1} \neq \lambda_{j_2}$, for all $j_1, j_2 = 1, \ldots, J$ (unconstrained parameters)
- $\lambda_j = 1$, $j = 1, \ldots, J$

Specification of item step difficulty parameters

- $\beta_{jy} = \beta_j + \tau_{jy}$ (unconstrained parameters)
- $\beta_{jy} = \beta_j + \tau_y$ (rating scale parameterization)
### Classification of ordinal polytomous IRT models

<table>
<thead>
<tr>
<th>$\lambda_j$</th>
<th>$\beta_{jy}$</th>
<th>logit link</th>
<th>binary items</th>
</tr>
</thead>
<tbody>
<tr>
<td>free</td>
<td>free</td>
<td>GRM</td>
<td>2PL</td>
</tr>
<tr>
<td>free</td>
<td>constrained</td>
<td>RS-GRM</td>
<td>2PL</td>
</tr>
<tr>
<td>constrained</td>
<td>free</td>
<td>1P-GRM</td>
<td>Rasch</td>
</tr>
<tr>
<td>constrained</td>
<td>constrained</td>
<td>1P-RS-GRM</td>
<td>Rasch</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>local</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GPCM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RS-GPCM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RS-2P-SM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PCM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RSM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRSM</td>
<td></td>
</tr>
</tbody>
</table>

- GRM: Graded Response Model (Samejima, 1969)
- RS-GRM: Rating Scale version of Graded Response Model (Muraki, 1990)
- 1P-GRM: Graded Response Model with fixed $\lambda_j$
- 1P-RS-GRM: Graded Response Model with a Rating Scale param. and fixed $\lambda_j$
- GPCM: Generalized Partial Credit Model (Muraki, 1992)
- RS-GPCM: Rating Scale version of Generalized Partial Credit Model (Muraki, 1992)
- PCM: Partial Credit Model (Masters, 1982)
- RSM: Rating Scale Model (Andrich, 1978)
- 2P-SM: two-parameter Sequential Model (Hemker et al., 2001)
- SRM: Sequential Rasch Model (Tutz, 1990)
- SRSM: Sequential Rating Scale Model (Tutz, 1990)
- 2PL: two-parameter logistic model (Birnbaum, 1968)
- Rasch: Rasch model (Rasch, 1960)
Limits of traditional IRT models

- A same questionnaire is usually used to measure **several latent traits**
- Multidimensional IRT models (Reckase, 2010)
- Normality assumption implies a computationally cumbersome estimation process, mainly in presence of more than one latent trait
- Often, normality of latent trait is not a realistic assumption
- In some contexts (e.g., health care) can be not only more realistic, but also more convenient for the decisional process, to assume that population is composed by **homogeneous classes of individuals** who have very similar latent characteristics (Lazarsfeld and Henry, 1968; Goodman, 1974), so that individuals in the same class will receive the same kind of decision (e.g., clinical treatment)

**Multidimensional Latent Class (LC) IRT models** (Bartolucci, 2007)

For further extensions and generalizations of traditional IRT models see, for instance, Wilson and De Boeck (2004) and Von Davier and Carstensen (2007)
Main features

1. more latent traits $\Theta = (\Theta_1, \ldots, \Theta_s)$ are simultaneously considered \textit{(multidimensionality)}

2. each item measures only one latent trait \textit{(between-item multidimensionality)}

3. these latent traits are represented by a random vector $\Theta$ with a \textbf{discrete distribution} common to all subjects, with support points $\{\xi_1, \ldots, \xi_k\}$ and weights $\{\pi_1, \ldots, \pi_k\}$: each support point identifies a different latent class of individuals

4. the number $k$ of latent classes is the same for each latent trait

5. weights may also depend on individual covariates, that is $\pi_{ci}(x) = p(\Theta = \xi_c | X_{1i} = x_1, \ldots, X_{mi} = x_m)$ \textit{(individual-specific weights)}

6. either free or constrained item parameters may be specified

7. either a global or a local link function may be adopted

Main references: Bartolucci (2007); Bartolucci, Bacci, Gnaldi (2014)
Multidimensional latent class IRT models

- **general formulation** of the class of models

\[
g_y[p_j(\xi_c)] = \lambda_j \left( \sum_{d=1}^{s} \delta_{jd} \theta_d - \beta_{jy} \right)
\]

- \(p_j(\xi_c) = p(Y_{ij} = y|\Theta = \xi_c)\)

- \(\delta_{jd}: \) dummy variable equal to 1 if item \(j\) measures latent trait of type \(d, d = 1, \ldots, s\)

- **manifest distribution** of the full response vector \(Y = (Y_1, \ldots, Y_J)'\)

\[
p(Y = y) = \sum_{c=1}^{k} p(Y = y|\Theta = \xi_c)\pi_c
\]

- \(\pi_c = p(\Theta = \xi_c)\)

- assumption of local independence: \(p(Y = y|\Theta = \xi_c) = \prod_{j=1}^{J} p_j(\xi_c)\)
Multidimensional latent class IRT models

Extension to latent regression

- the above illustrated model does not take into account the possible effect of one or more observed covariates
- we may suppose that the probability to belong to a given latent class is influenced by some individual characteristics: $X_1, \ldots, X_m$ ($m \geq 1$)
- in such a case **individual-specific weights** are defined:
  \[
  \pi_c(x) = p(\Theta = \xi_c | X_{1i} = x_1, \ldots, X_{mi} = x_m)
  \]
- a multinomial logit model is introduced for weights $\pi_{ci}(x)$, $c = 2, \ldots, k$, with respect to $\pi_{1i}(x)$ (or another weight), as follows
  \[
  \log \frac{\pi_{ci}(x)}{\pi_{1i}(x)} = \log \frac{p(\Theta = \xi_c | X_{1i} = x_1, \ldots, X_{mi} = x_m)}{p(\Theta = \xi_1 | X_{1i} = x_1, \ldots, X_{mi} = x_m)} = \zeta_{0c} + \sum_{h=1}^{m} \zeta_{hc} x_{hi}
  \]
  - $\zeta_{hc}$ denotes the effect of covariate $X_h$ on the logit of $\pi_{ci}(x)$ with respect to $\pi_{1i}(x)$
  - $\zeta_{0c}$ is the intercept specific for class $c$
Maximum log-likelihood estimation

Let $i$ denote a generic subject and let $\eta$ the vector containing all the free parameters. The log-likelihood may be expressed as

$$\ell(\eta) = \sum_i \log[p(Y_i = y_i)]$$

- Estimation of $\eta$ may be obtained by the discrete (or LC) MML approach (Bartolucci, 2007)

- $\ell(\eta)$ may be efficiently maximize by the EM algorithm (Dempster et al., 1977)

- The software for the model estimation has been implemented in the R package MultiLCIRT
Main important functions

- **est_multi_poly** estimates models belonging to the class of multidimensional LC IRT models.
- **aggr_data** collapses original records having the same response pattern so as to obtain a matrix with a record for each distinct response configuration (rather than for each statistical unit).
- **search.model** searches for the global maximum of the log-likelihood given a vector of possible number of classes to try for.
- **compare_models** allows to compare nested or non nested models (estimated through **est_multi_poly**) by means of AIC and BIC criteria and also LR test in case of nested models.
- **test_dim** performs an LR test to compare pairs of nested models that differ only in the way items are grouped (e.g., bidimensional versus unidimensional structure of items).
- **class_item** performs a model-based hierarchical clustering, when the grouping structure of items is not a priori known.
Model-based hierarchical clustering

- The proposed model-based hierarchical clustering is based on the likelihood ratio (LR) test, which compares two nested models: a general model with items grouped in \( s \) dimensions (latent traits) is compared with a restricted model with \( s - 1 \) dimensions, all other things being constant (i.e., type of logit, number of latent classes, item parameterization).

- Function `class_item` builds a sequence of nested models: the most general one uses a separate dimension for each item and the most restrictive one uses only one dimension common to all items (uni-dimensional model).

- The clustering procedure performs \( J - 1 \) steps.

- At each step, the LR test statistic is computed for every pair of possible aggregations of items (or groups of items).

- The aggregation with the minimum value of the statistic (or the highest p-value) is then adopted and the corresponding model is fitted before moving to the next step.
Main references


