Multilevel LC models

Silvia Bacci

silvia.bacci@unipg.it
Dipartimento di Economia - Università degli Studi di Perugia (IT)

Outline

Starting point

- Basic LC model and its extensions to account for individual covariates and item characteristics assume the independency of item responses coming from different individuals
- In other words, it is ignored that part of unobserved heterogeneity of item responses due to multilevel structures of data (e.g., students within schools, repeated measurements within individuals)
- When individuals are nested in groups it is reasonable to assume that individuals (e.g., students) sharing a same group (e.g., school) context are more similar than their colleagues belonging to different groups in terms of latent trait levels
- As a consequence, the corresponding item responses cannot be assumed to be independent

- We introduce a multilevel extension of the class of LC models to allow for the group (school) effect
- We also consider the effect of covariates at individual and group levels
- Specifically, we assume the presence of a single latent trait at the school level which affects the abilities (possibly, more than one) at student level
- Individual abilities are measured through a multidimensional version of LC-IRT models based on a 2PL parameterisation (Birnbaum, 1968) for the conditional probability of a certain response given the underlying ability, in the case of binary items

Basic notation

- Y_{hij}: response provided by subject i within group h to item j, with possible values 0 and 1 and $h = 1, \ldots, H$, $i = 1, \ldots, n_h$, and $j = 1, \ldots, r$
- V_{hi}: discrete latent variable with k_V support points, describing the distribution of latent traits measured by the test
- $\boldsymbol{\xi}_v^{(V)}$: vector of support points with elements $\xi_{vd}^{(V)}$ corresponding to the ability level of subjects in latent class v ($v = 1, \ldots, k_V$) with respect to dimension d ($d = 1, \ldots, s$)
- U_h: discrete latent variable with k_U support points, describing the group effect
- $\boldsymbol{W}_h = (W_{h1}, \ldots, W_{hm_U})'$: vector of m_U group level covariates for group h
- $\boldsymbol{X}_{hi} = (X_{hi1}, \ldots, X_{him_V})'$: vector of m_V individual level covariates for subject i in group h

The multilevel LC-2PL model

- The relation between V_{hi} and Y_{hij} is as follows

$$\text{logit}[p(Y_{hij} = 1 \mid V_{hi} = v)] = \gamma_j \left(\sum_{d=1}^{s} \delta_{jd} \xi_{vd}^{(V)} - \beta_j \right)$$

  - β_j: difficulty level of item j
  - γ_j: discriminating level of item j
  - δ_{jd}: indicator for item j measuring the dth latent trait

- The discreteness assumption of V_{hi} implies the following manifest distribution of the full response vector $\boldsymbol{Y}_{hi} = (Y_{hi1}, \dots, Y_{hir})'$

$$p(\boldsymbol{y}_{hi}) = p(\boldsymbol{Y}_{hi} = \boldsymbol{y}_{hi}) = \sum_{v=1}^{k_V} p_v(\boldsymbol{y}_{hi})\pi^{(V)}_{hi,v|u}$$

  - $\boldsymbol{y}_{hi} = (y_{hi1}, \dots, y_{hir})'$: realisation of $\boldsymbol{Y}_{hi}$
  - $p_v(\boldsymbol{y}_{hi}) = p(\boldsymbol{Y}_{hi} = \boldsymbol{y}_{hi} \mid V_{hi} = v) = \prod_{j=1}^{r} p(Y_{hij} = y_{hij} \mid V_{hi} = v), \quad v = 1, \dots, k_V$ (local independence assumption)
  - $\pi^{(V)}_{hi,v|u} = p(V_{hi} = v | U_h = u, \boldsymbol{X}_{hi} = \boldsymbol{x}_{hi})$: weight of the v-th support point, depending on U_h and on the individual configuration of $\boldsymbol{X}_{hi}$

- Weights at individual level, $\pi^{(V)}_{hi,v|u}$, describe the a prior probability that an individual belongs to latent class v
- $\pi^{(V)}_{hi,v|u}$ depend on the latent trait U_h and on the individual configuration of $\boldsymbol{X}_{hi}$ through a multinomial logit parameterisation, as follows:

$$\log \frac{\pi^{(V)}_{hi,v|u}}{\pi^{(V)}_{hi,1|u}} = \psi^{(V)}_{0uv} + \boldsymbol{x}'_{hi}\boldsymbol{\psi}^{(V)}_{1v}, \quad v = 2, \dots, k_V$$

  - $\boldsymbol{\psi}^{(V)}_{1v}$: vector of regression parameters corresponding to the effect of individual covariates $\boldsymbol{X}_{hi}$ on the logit of $\pi^{(V)}_{hi,v|u}$ with respect to $\pi^{(V)}_{hi,1|u}$
  - $\psi^{(V)}_{0uv}$: intercept specific for individuals of class v that belong to a group of type u

- We also introduce weights $\pi_{hu}^{(U)} = p(U_h = u | \boldsymbol{W}_h = \boldsymbol{w}_h)$ associated to the support points for U_h
- Weights at group level, $\pi_{hu}^{(U)}$, describe the a prior probability that a group belongs to latent class (or type) u
- $\pi_{hu}^{(U)}$ depend on the group covariate configuration $\boldsymbol{W}_h = \boldsymbol{w}_h$ through a similar multinomial logit parameterisation, as follows:

$$\log \frac{\pi_{hu}^{(U)}}{\pi_{h1}^{(U)}} = \psi_{0u}^{(U)} + \boldsymbol{w}_h' \boldsymbol{\psi}_{1u}^{(U)}, \quad u = 2, \ldots, k_U$$

  - $\boldsymbol{\psi}_{1u}^{(U)}$: vector of regression parameters corresponding to the effect of group covariates $\boldsymbol{W}_h$ on the logit of $\pi_{hu}^{(U)}$ with respect to $\pi_{h1}^{(U)}$
  - $\psi_{0u}^{(U)}$: intercept specific for groups within type u

The log-likelihood function

- For given k_U and k_V, the model parameters may be estimated by maximising the (incomplete) log-likelihood

$$\ell(\boldsymbol{\theta}) = \sum_{h=1}^{H} \log \sum_{u=1}^{k_U} \pi_{hu}^{(U)} \prod_{i=1}^{n_h} \sum_{v=1}^{k_V} \pi_{hi,v|u}^{(V)} \prod_{j=1}^{r} p(y_{hij}|V_{hi} = v),$$

where $\boldsymbol{\theta}$ is the vector containing all the free parameters

- In order to maximise the log-likelihood $\ell(\boldsymbol{\theta})$, we make use of the EM algorithm (Dempster, Laird and Rubin, 1977), which is implemented in the `R` package named `MultiLCIRT` (Bartolucci, Bacci and Gnaldi, 2014)

The EM algorithm

- The EM algorithm is based on the maximisation of the following complete log-likelihood

$$\ell^*(\boldsymbol{\theta}) = \sum_{h=1}^{H} \ell^*_{1h}(\boldsymbol{\theta}) + \ell^*_{2h}(\boldsymbol{\theta}) + \ell^*_{3h}(\boldsymbol{\theta})$$

with

$$\begin{aligned}
\ell^*_{1h}(\boldsymbol{\theta}) &= \sum_{i=1}^{n_h}\sum_{j=1}^{J}\sum_{v=1}^{k_V} z_{hiv} \log p(y_{hij}|V_{hi}=v), \\
\ell^*_{2h}(\boldsymbol{\theta}) &= \sum_{i=1}^{n_h}\sum_{u=1}^{k_U}\sum_{v=1}^{k_V} z_{hu} z_{hiv} \log \pi^{(V)}_{hi,v|u}, \\
\ell^*_{3h}(\boldsymbol{\theta}) &= \sum_{u=1}^{k_1} z_{hu} \log \pi^{(U)}_{hu},
\end{aligned}$$

  - z_{hiv}: indicator function for subject i being in latent class v ($V_{hi}=v$)
  - z_{hu}: indicator function for cluster h being of typology u ($U_h = u$)
  - $z_{hu}z_{hiv}$ is equal to 1 if both conditions are satisfied and to 0 otherwise

Based on function $\ell^*(\boldsymbol{\theta})$, the EM algorithm alternates the following two steps until convergence in $\ell(\boldsymbol{\theta})$

- E-step. It consists of computing the expected value of the complete log-likelihood $\ell^*(\boldsymbol{\theta})$. This is equivalent to computing the posterior expected values of the indicator variables
- M-step. It consists of updating the model parameters by maximising the expected value of $\ell^*(\boldsymbol{\theta})$ obtained at the E-step. As an explicit solution does not exist for the model parameters, iterative optimisation algorithms of Newton-Raphson type are used. The resulting estimate of $\boldsymbol{\theta}$ is used to update the expected value of $\ell^*(\boldsymbol{\theta})$ at the next E-step and so on

- Note that only the ability parameters $\boldsymbol{\xi}_v^{(V)}$ are estimated, whereas parameters $\xi_u^{(U)}$ are estimated as average of the $\boldsymbol{\xi}_v^{(V)}$ with suitable weights:

$$\hat{\xi}_u^{(U)} = \frac{1}{ns} \sum_{d=1}^{s} \sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{v=1}^{k_V} \hat{\xi}_{dv} \hat{\pi}_{hi,v|u}$$

- Parameters $\xi_u^{(U)}$ are interpretable as the effect on the individuals' abilities for groups of type u
- After parameter estimation, each subject i can be allocated to one of the k_V latent classes on the basis of the response pattern $\boldsymbol{y}_i$ she/he provided, her/his covariates $\boldsymbol{x}_i$, and the typology of group she/he belongs to
- Similarly, each group h can be allocated to one of the k_U latent classes
- In both cases, the most common approach is to assign the subject and the group to the class with the highest posterior probability

Data description

- Nationally representative sample of 27,592 middle school students within 1,305 Italian schools belonging to the 21 Italian Regions
- Each student answers to tests on Language and Mathematics abilities, collected in June 2009 by the National Institute for the Evaluation of the Education System (INVALSI)
- Three latent traits: Reading comprehension (V_1), Grammar (V_2), and Mathematics (V_3)
- V_1 is measured by 30 binary items, V_2 is measured by 10 binary items, and V_3 is measured by 27 binary items
- Individual level covariate: student's gender
- Group level covariate: school geographic area (North-East, North-West, Centre, South, Islands)

Model selection

- In the LC setting, the number of support points, that is, k_V and k_U, has to be a priori fixed
- We adopt a widespread classification of students into three groups (i.e., basic, advanced, and proficient), so that $k_V = 3$
- Given the value of k_V, we choose the number of school types relying on the Bayesian Information Criterion (BIC) computed on multilevel LC models without item parameterization

Table : Log-likelihood, number of parameters and BIC values for $k_U = 1, \ldots, 6$ latent classes for the INVALSI Test; in boldface is the smallest BIC value.

k_U	ℓ	#par	*BIC*
1	-531346.4	205	1064688
2	-530195.9	212	1062455
3	-529947.7	219	1062027
4	-529829.2	226	1061858
5	-529782.3	233	**1061833**
6	-529766.5	240	1061869

After the selection of k_V and k_U, alternative models with different item parameterisations are considered

Table : Model selection: log-likelihood and BIC values for the 1PL and 2PL model with covariates; in boldface is the smallest BIC value.

model	ℓ	#par	BIC
1PL	-533994.6	105	1069011
2PL	-530039.6	169	**1061724**

- In conclusion, we select a model with $k_V = 3$ latent classes of students, $k_U = 5$ latent classes of schools, and with a 2PL parameterisation

Distribution of the latent traits

Table : Students' level: distribution of the ordered estimated abilities $\hat{\boldsymbol{\xi}}_v^{(V)}$ for the three involved dimensions within classes, together with the average weights ($\hat{\bar{\pi}}_v^{(V)}$).

	V1	V2	V3	$\hat{\bar{\pi}}_v^{(V)}$
Class 1 (worst performers)	-0.643	1.322	1.214	0.174
Class 2 (intermediate performers)	0.657	2.302	1.671	0.428
Class 3 (best performers)	1.988	3.564	2.214	0.398

Table : School level: distribution of the estimated average abilities $\hat{\xi}_u^{(U)}$, for $u = 1, \ldots, 5$, and the average weights ($\hat{\bar{\pi}}^{(U)}$).

	Type 1	Type 2	Type 3	Type 4	Type 5
$\hat{\xi}_u^{(U)}$	1.009	1.531	1.594	2.046	2.451
$\hat{\bar{\pi}}^{(U)}$	0.081	0.081	0.370	0.351	0.116

Effects of covariates

- Effect of gender: females have a greater tendency than males to be grouped into classes 2 and 3 (intermediate and best performers) with respect to class 1 (worst performers)
- Effect of geographic area:

Table : Estimated conditional probabilities $\hat{\pi}_{hu}^{(U)}$ to belong to the five types of schools given the geographic area.

	Type 1	Type 2	Type 3	Type 4	Type 5
NE	0.042	0.000	0.498	0.387	0.073
NW	0.036	0.000	0.620	0.329	0.014
Centre	0.031	0.044	0.358	0.472	0.095
South	0.115	0.113	0.239	0.341	0.192
Islands	0.180	0.243	0.140	0.232	0.206

  - The great majority of the Italian schools tends to be classified into average and high (Types 3 and 4) attainment schools
  - Schools from the South and the Islands have a relatively high probability to be classified among the best schools (Type 5) and, at the same time, among the worst schools (Type 1 schools)

Conclusions

- The proposed framework allows us for assessing the relationship between latent classes of individuals (i..e, students) and groups (i.e., schools) and observed characteristics, and establishing the ways observed characteristics are related to unobserved groupings, accounting at the same time for the multilevel structure of data
- The adopted approach is an extension of the basic LC model, by accounting for (i) the multilevel structure of the data, (ii) the effects of observed covariates at the students' and school levels, (iii) the presence of more than one latent trait at students' level, and (iv) item characteristics affecting the observed item responses
- At the various levels of the hierarchy, the approach permits the combined use of information derived from observed group membership (i.e., examinees' gender and school geographic area) and unobserved groupings (i.e., latent classes of examinees and type of school)

R package MultiLCIRT

- Multilevel LC models may be estimated through R package MultiLCIRT
- The function for the multilevel LC model estimation is `est_multi_poly_clust`
- This function is very similar to `est_multi_poly`

Specification and estimation of multilevel LC models

- Model estimation

```
> out <- est_multi_poly_clust(S, kU=kU, kV=kV,
W=W, X=X, clust=clust, start=1, link =1,disc=1,difl=0,
multi = multi,output=TRUE)
```

  - `S`: data matrix
  - `kU`: number of latent classes at cluster (school) level
  - `kV`: number of latent classes at individual (students') level (or ability level)
  - `W`: matrix of covariates that affects the weights at cluster level
  - `X`: matrix of covariates that affects the weights at individual level
  - `clust`: vector of cluster (school) indicator for each individual unit (student)

- `start`: method of initialization of the algorithm (0 = deterministic, 1 = random, 2 = arguments given as input)
- `link`: type of link function
- `disc`: indicator of constraints on the discriminating indices
- `difl`: indicator of constraints on the difficulty levels
- `multi`: matrix specifying the multidimensional structure of the model

- Output
  - `out$Th`: estimated matrix of latent trait levels for each dimension and latent class at individual level
  - `out$Piv`: matrix of weights for each profile at individual level
  - `out$DeV` and `out$DeU`: estimated regression coefficients of covariates at individual level and cluster level, respectively
  - `out$seDeV` and `out$seDeU`: standard errors of the estimated regression coefficients
  - `out$La`: a posterior probabilities of belonging to every cluster level latent class for every cluster unit (school)
  - All outputs shown for the LC models (with and without covariates) are still valid

Main references

- Fox J. P. (2005). Multilevel IRT using dichotomous and polytomous response data. Br. J. Math. Stat. Psychol., 58:145?172
- Goldstein H. (2011). Multilevel statistical models. Wiley, Hoboken
- Gnaldi M., Bacci S., Bartolucci F. (2015). A multilevel finite mixture item response model to cluster examinees and schools, Advances in Data Analysis and Classification, doi: 10.1007/s11634-014-0196-0
- Vermunt J. K. (2003). Multilevel latent class models. Sociol. Methodol., 33:213?239