MULTILEVEL CROSS-CLASSIFIED LATENT CLASS MODELS

Silvia Columbu [1], Nicola Piras [1] and Jeroen K. Vermunt[2]

[1] Department of Mathematics and Computer Science, University of Cagliari, (e-mail: silvia.columbu@unica.it, nicola.piras97@unica.it)

[2] Department of Methodology and Statistic, Tilburg University, (e-mail: J.K.Vermunt@tilburguniversity.edu)

ABSTRACT: We propose an extension of latent class models to deal with multilevel cross-classified data structures, where each observation is considered simultaneously nested within two groups, such as for instance, children within both schools and neighborhoods. We show how such a situation can be dealt with by having a separate set of mixture components for each of the crossed classifications. Unfortunately, given the intractability of the derived loglikelihood, the EM algorithm can no longer be used in the estimation process. We therefore propose an approximate estimation of this model using a stochastic version of the EM algorithm similar to Gibbs sampling.

KEYWORDS: Latent class, cross-classified, Stochastic EM

1 Introduction

Latent class analysis (LCA) is a popular model-based approach for data clustering of units on the basis of observations arising from a set of categorical indicators. When the data have a multilevel hierarchical structure with units nested within higher level observations, such as children nested within schools, a possible extension (Laird, 1978) discussed in Vermunt, 2003 and Vermunt, 2008 takes two levels of clustering with separate latent variables for lower-level units and higher-level ones. Sometimes data have a cross-classified structure with units grouped within multiple higher level units, for example, children can be considered nested within both schools and neighborhoods. In this contribution we propose to extend Multilevel Latent Class analysis to handle cross-classification. Given the untractability of the derived likelihood the standard EM algorithm can not be applied in the estimation, and we propose to use a stochastic version of the EM algorithm that can handle the hierarchy of units but also their double cross-classification, similar to what done in Keribin *et al.*, 2015 for coclustering.

2 Model definition

Let Y_{ijkq} be the response on categorical indicator (or item) i ($i = 1,\ldots,I$) of individual or first level unit j ($j = 1,\ldots,n_{kq}$) belonging simultaneously to the group level units k ($k = 1,\ldots,K$) and q ($q = 1,\ldots,Q$). We denote with X_{jkq}, W_k and Z_q the discrete latent variables respectively for membership of level-1 units and for the two group level units. A particular latent class will be indicated with ℓ ($\ell = 1,\ldots,L$), for level-1 units, h ($h = 1,\ldots,H$) and r ($r = 1,\ldots,R$) for level-2 units. For ease of notation, we focus on binary indicators and denote with $\pi_{i|\ell}$ the probability distribution parameters of each item within the first level latent class.
The data model consists of two parts, described through two separate equations, one for the level-2 cross-classified (or higher level) units and one for the level-1 (or lower level) units. Each of the two equations is a mixture of probabilities. The model for the higher part is described, in the complete data form, by

$$\begin{aligned} P(\mathbf{Y}_{kq}, W_k = h, Z_q = r) &= P(W_k = h, Z_q = r) P(\mathbf{Y}_{kq} | W_k = h, Z_q = r) \\ &= P(W_k = h, Z_q = r) \prod_{j=1}^{n_{kq}} P(\mathbf{Y}_{jkq} | W_k = h, Z_q = r) \\ &= P(W_k = h) P(Z_q = r) \prod_{j=1}^{n_{kq}} P(\mathbf{Y}_{jkq} | W_k = h, Z_q = r). \end{aligned}$$

We assumed independence of observations within a combination of groups given their belonging to the cross-classified latent classes, and also marginal independence of the two higher level latent classes W_k and Z_q.
The second part models the density of observations conditionally to their simultaneous belonging in higher level cross-classified latent classes, that is:

$$P(\mathbf{Y}_{jkq} | W_k = h, Z_q = r) = \sum_{\ell=1}^{L} P(X_{jkq} = \ell | W_k = h, Z_q = r) \prod_{i=1}^{I} P(Y_{ijkq} | X_{jkq} = \ell),$$

in which we have assumed the local independence of indicators within latent classes.

3 Parameters' Estimation

The estimation of model parameters $\boldsymbol{\theta} = \{\pi_{\ell|hr}, \pi_h, \pi_r, \pi_{i|\ell}\}$, requires the maximization of the observed likelihood of the model in the form

$$L(\boldsymbol{\theta}; \mathbf{y}) = \sum_{h_1=1}^{H} \sum_{h_2=1}^{H} \cdots \sum_{h_K=1}^{H} \sum_{r_1=1}^{R} \sum_{r_2=1}^{R} \cdots \sum_{r_Q=1}^{R} \prod_{k=1}^{K} P(W_k = h_k) \prod_{q=1}^{Q} P(Z_q = r_q) \times$$

$$\prod_{j=1}^{n_{kq}}\left[\sum_{\ell=1}^{L}P(X_{jkq}=\ell|W_k=h_k,Z_q=r_q)\prod_{i=1}^{I}P(Y_{ijkq}|X_{jkq}=\ell)\right].$$

The presence of a double missing data structure at higher level, with W_k and Z_q unobserved, causes that the likelihood cannot factorize as a product of the mixing probabilities as for standard LC and multilevel LC models. The likelihood becomes easily untractable and standard EM algorithms cannot be directly applied for its maximization. We propose to consider a Stochastic version of the algorithm with the inclusion of a Gibbs sampling scheme between the E and the M step. The Stochastic step consists in the consecutive sampling from marginal posterior distributions of higher level and lower level latent classes, which reduces the computational burden.

E and S step

After initialization of $\pi_h = P(W_k = h)$, $\pi_r = P(Z_q = r)$, $\pi_{\ell|hr} = P(X_{jkq} = \ell|W_k = h, Z_q = r)$ and $\pi_{i|\ell}$ iterate the following sampling steps

1) Draw $\mathbf{w}^{(t)}$ from a Multinomial distribution with probabilities

$$P(W_k=h|\mathbf{y}_k,\mathbf{z}^{(t-1)})=\frac{\pi_h P(\mathbf{Y}_k|\mathbf{z}^{(t-1)},W_k=h)}{P(\mathbf{Y}_k|\mathbf{z}^{(t-1)})},$$

$$P(\mathbf{Y}_k|\mathbf{z},W_k=h)=\prod_{q_k=1}^{Q_K}\prod_{r=1}^{R}\left[\prod_{j=1}^{n_{kq}}P(\mathbf{Y}_{jkq}|W_k=h,Z_q=r)\right]^{z_q^r};$$

2) Draw $\mathbf{z}^{(t)}$ from a Multinomial distribution with probabilities

$$P(Z_q=r|\mathbf{y}_q,\mathbf{w}^{(t)})=\frac{\pi_r P(\mathbf{Y}_q|\mathbf{w}^{(t)},Z_q=r)}{P(\mathbf{Y}_q|\mathbf{w}^{(t)})},$$

$$P(\mathbf{Y}_q|\mathbf{w},Z_q=r)=\prod_{k_q=1}^{K_Q}\prod_{h=1}^{H}\left[\prod_{j=1}^{n_{kq}}P(\mathbf{Y}_{jkq}|W_k=h,Z_q=r)\right]^{w_k^h};$$

3) Draw $\mathbf{x}^{(t)}$ from a Multinomial distribution with probabilities

$$P(X_{jkq}=\ell|\mathbf{y}_{jkq},\mathbf{w}^{(t)},\mathbf{z}^{(t)})=\frac{\left[\pi_{\ell|h,r}P(\mathbf{Y}_{jkq}|X_{jkq}=\ell)\right]^{w_{jk}^h z_{jq}^r}}{P(\mathbf{Y}_{jkq})},$$

where w_k^h, z_q^r, w_{jk}^h, z_{jq}^r and x_{jkq}^ℓ are all binary indicators of units' membership at different levels, in particular w_{jk}^h, z_{jq}^r are the expansion of higher level latent class indicators over the first level units j.

M step

$$\pi_h = \frac{\sum_{k=1}^{K} w_k^{h(t)}}{K}, \quad \pi_r = \frac{\sum_{q=1}^{Q} z_q^{r(t)}}{Q},$$

$$\pi_{\ell|hr} = \frac{\sum_{j=1}^{n} w_{jk}^{h(t)} z_{jq}^{r(t)} x_{jkq}^{\ell(t)}}{\sum_{j=1}^{n} w_{jk}^{h(t)} z_{jq}^{r(t)}}, \quad \pi_{i|\ell} = \frac{\sum_{j=1}^{n} x_{jkq}^{\ell(t)} y_{ijkq}}{\sum_{j=1}^{n} x_{jkq}^{\ell(t)}}.$$

Final estimates are calculated as the mean over the total number of iterations, burn-in period excluded.

Results from simulation studies with data generated under varying scenarios, prove that the estimators have satisfactory finite sample properties. In figure 1 is reported the error resulting from the estimation of $\pi_{\ell|h=1,r=1}$ over 50 binary simulated datasets with fixed number of classes L=4, H=R=2. Two scenarios of moderate increasing separation have been compared. It emerges that the average across replications is close to the true value, with an improvement with the increase of the number of groups. Similar results are observed for the other first-level and distribution parameters. Almost no error is observed for high-level latent class parameters. In the implementation of the SEM-Gibbs 150 iterations have been considered, including 50 burn-in. These are sufficient for convergence.

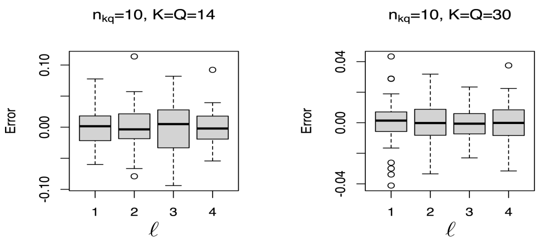

Figure 1. *Error on the estimation of* $\pi_{\ell|h=1,r=1}$.

References

KERIBIN, C., BRAULT, V., CELEUX, G., & GOVAERT, G. 2015. Estimation and selection for the latent block model on categorical data. *Statistics and Computing.*, **25**(6), 1201–1216.

LAIRD, N. 1978. Nonparametric Maximum-Likelihood Estimation of a Mixture Distribution. *Journal of the American Statistical Association.*, **73**, 805–811.

VERMUNT, J. K. 2003. Multilevel latent class models. *Sociological Methodology.*, **33**, 213–239.

VERMUNT, J. K. 2008. Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research.*, **17**, 33–51.