SDS
Statistica e
Data Science

Palermo, 11-12 April 2024

Proceedings of the Statistics and Data Science 2024 Conference

New perspectives on Statistics and Data Science

Edited by

Antonella Plaia – Leonardo Egidi
Antonino Abbruzzo

Proceedings of the SDS 2024 Conference
Palermo, 11-12 April 2024
Edited by: Antonella Plaia – Leonardo Egidi - Antonino Abbruzzo
-
Palermo: Università degli Studi di Palermo.

ISBN Ebook (Pdf)
978-88-5509-645-4

Log-likelihood approximation in Stochastic EM for Multilevel Latent Class Models

Silvia Columbu, Nicola Piras and Jeroen K. Vermunt

Abstract Multilevel cross-classified Latent Class Models are an extension of standard latent class for handling data in which each observation is simultaneously nested within two groups. The likelihood associated to the model is untractable and approximation methods such as stochastic versions of the EM can be applied. The knowledge of a final estimate of the log-likelihood can be helpful in the evaluation of parameter estimates and for selection purposes. We propose two alternative log-likelihood approximation procedures and test their performances in the Hierarchical Multilevel Latent Class Model for which a finite estimate of the likelihood is provided through a special version of the EM.

Key words: Log-likelihood, Classification Log-likelihood, Multilevel Latent Class

1 Introduction

Latent Class Models can be extended to deal with multilevel cross-classified (CC) data structures in which units are simultaneously grouped within multiple higher level units (for example, children nested within both schools and neighborhoods). A similar extension is that of Hierarchical Multilevel Latent Class (MLC) (see [3]) where it is considered the nesting of each observation within a single group. Un-

Silvia Columbu,
Department of Mathematics and Computer Science, University of Cagliari, e-mail: silvia.columbu@unica.it

Nicola Piras,
Department of Mathematics and Computer Science, University of Cagliari, e-mail: nicola.piras97@unica.it

Jeroen K. Vermunt,
Department of Methodology and Statistic, Tilburg University, e-mail: J.K.Vermunt@tilburguniversity.edu

like the MLC, in the cross-classified context the presence of a double missing data structure at the higher level causes the intractability of the derived log-likelihood that cannot be factorized as a product of the mixing probabilities. To overcome this issue, in [1] we have proposed the implementation of a Stochastic version of the EM algorithm including a Gibbs sampler between the E and the M step. The Gibbs step consists in the consecutive sampling from marginal posterior distributions of higher and lower level latent variables. This sampling procedure generates an irreducible Markov chain with a unique stationary distribution concentrated around the maximum likelihood parameter estimates. Therefore, our estimation procedure, unlike what happens in the MLC (see [3]), does not provide a finite estimate of the maximum log-likelihood.
The availability of such a finite approximation would be useful in the assessment of convergence of the estimation algorithm as well as in the definition of global model selection procedures based on the BIC. In this work we propose an approach to obtain a good approximation of the maximum of log-likelihood moving from the Markov chains obtained in the stochastic estimation algorithm. Given the finite estimate of maximum log-likelihood in the MLC setting, we propose to assess the best approximation procedure in stochastic versions of EM algorithms by performing comparative simulation studies in that frame. These results can then be extended to the Multilevel cross-classified latent class (MCCLC) formulation.

2 Log-likelihood approximation

The estimation of parameters in latent class models usually involves the introduction of the complete data log-likelihood, known in a classification context also as classification log-likelihood (CL). Its expression can be directly linked to the log-likelihood itself and therefore used to derive an approximation once that an estimation of model parameters is plugged in. In fact, following [2], the log-likelihood can be expressed as

$$\log L(\hat{\theta}) = CL(\hat{\theta}) + EN(\hat{\theta}) \quad (1)$$

where $CL(\hat{\theta})$ is the classification log-likelihood which is equivalent to the expected complete log-likelihood once the classification is given and $EN(\hat{\theta})$ is an entropy term which gives a measure of the separation of classes in the data. Their analytical expression depends on the specific model assumed. In particular, in this contribution we will consider the application of such a decomposition to estimate the maximum of the log-likelihood in two multilevel extensions (MLC and MCCLC) of latent class models. In what follows we will remind model formulations and their corresponding CL and EN expressions.

Log-likelihood approximation for MLC Let Y_{ijk} be the response on item i ($i = 1,...,I$) of individual or first level unit j ($j = 1,...,n_k$) belonging to the group level units k ($k = 1,...,K$). Following [3], we consider two sets of discrete latent variables X_{jk} and W_k with associated ℓ ($\ell = 1,...,L$) latent classes of level-1 and h

$(h = 1,...,H)$ of level-2. In this framework the model is described by two separate equations for the second level units and for the first level units, with resulting log-likelihood

$$\log L = \sum_{k=1}^{K} \log \sum_{h=1}^{H} P(W_k = h) \prod_{j=1}^{n_k} \left[\sum_{\ell=1}^{L} P(X_{jk} = \ell | W_k = h) P(\mathbf{Y}_{jk} | X_{jk} = \ell) \right]$$

The set of model parameters is $\theta = \{\pi_h = P(W_k = h), \pi_{\ell|h} = P(X_{jk} = \ell | W_k = h), \pi_{y_{i|\ell}} = P(\mathbf{Y}_{jk} | X_{jk} = \ell)\}$, where last ones are probability distribution parameters. Following [4] the Classification Log-likelihood for the estimated parameters, is computed as

$$\begin{aligned} CL(\hat{\theta}) = &\sum_{k=1}^{K} \sum_{h=1}^{H} P(W_k = h | \mathbf{y}_k) \log(P(W_k = h)) \\ &+ \sum_{h=1}^{H} \sum_{j=1}^{n} \sum_{\ell=1}^{L} P(W_k = h, X_{jk} = \ell | \mathbf{y}_k) log(P(X_{jk} = \ell | W_k = h) P(\mathbf{Y}_{jk} | X_{jk} = \ell)) \end{aligned}$$

where $P(W_k = h, X_{jk} = \ell | \mathbf{y}_k) = P(W_k = h | \mathbf{y}_k) P(X_{jk} = \ell | \mathbf{y}_{jk}, W_k = h)$, while the Entropy (EN) term is

$$\begin{aligned} EN(\hat{\theta}) = &\sum_{k=1}^{K} \sum_{h=1}^{H} -P(W_k = h | \mathbf{y}_k) \log(P(W_k = h | \mathbf{y}_k)) \\ &+ \sum_{h=1}^{H} \sum_{j=1}^{n} \sum_{\ell=1}^{L} -P(W_k = h, X_{jk} = \ell | \mathbf{y}_k) log(P(X_{jk} = \ell | W_k = h, \mathbf{y}_k)) \end{aligned}$$

Log-likelihood approximation for MCCLC The extension of MLC models to cross-classified structures requires the introduction of a third latent variable for level-2 memberships (Z_q) as well as a corresponding set of latent classes q ($q = 1,...,Q$) (see [1]). The expression of the log-likelihood becomes:

$$\begin{aligned} \log L(\theta) = \log &\sum_{h_1=1}^{H} \sum_{h_2=1}^{H} \cdots \sum_{h_K=1}^{H} \sum_{r_1=1}^{R} \sum_{r_2=1}^{R} \cdots \sum_{r_Q=1}^{R} \prod_{k=1}^{K} P(W_k = h_k) \prod_{q=1}^{Q} P(Z_q = r_q) \times \\ &\prod_{j=1}^{n_{kq}} \left[\sum_{\ell=1}^{L} P(X_{jkq} = \ell | W_k = h_k, Z_q = r_q) P(\mathbf{Y}_{jkq} | X_{jkq} = \ell) \right], \end{aligned}$$

so that the Classification Log-likelihood and the Entropy are computed as

$$\begin{aligned} CL(\hat{\theta}) = &\sum_{k=1}^{K} \sum_{h=1}^{H} P(W_k = h | \mathbf{y}_k) log(P(W_k = h)) + \sum_{q=1}^{Q} \sum_{r=1}^{R} P(Z_q = r | \mathbf{y}_q) log(P(Z_q = r)) \\ &+ \sum_{h=1}^{H} \sum_{r=1}^{R} \sum_{j=1}^{n} \sum_{\ell=1}^{L} P(W_k = h, Z_q = r, X_{jk} = \ell | \mathbf{y}_{jkq}) \log(P(X_{jkq} = \ell | W_k = h, Z_q = r) P(\mathbf{Y}_{jkq} | X_{jkq} = \ell)) \end{aligned}$$

with $P(W_k = h, Z_q = r, X_{jkq} = \ell|\mathbf{y}_{jkq}) = P(W_k = h, Z_q = r|\mathbf{y}_{jkq})P(X_{kq} = \ell|\mathbf{y}_{jkq}, W_k = h, Z_q = r)$, and

$$EN(\hat{\theta}) = \sum_{k=1}^{K}\sum_{h=1}^{H} -P(W_k = h|\mathbf{y}_k)log(P(W_k = h|\mathbf{y}_k)) + \sum_{q=1}^{Q}\sum_{r=1}^{R} -P(Z_q = r|\mathbf{Y}_q)log(P(Z_q = r|\mathbf{y}_q))$$
$$+ \sum_{h=1}^{H}\sum_{r=1}^{R}\sum_{j=1}^{n}\sum_{\ell=1}^{L} -P(W_k = h, Z_q = r, X_{jk} = \ell|\mathbf{y}_{jkq})\log(P(X_{jkq} = \ell|W_k = h, Z_q = r, \mathbf{y}_{jkq}))$$

3 Estimation and assessment of log-likelihood approximation

With the aim of assessing the performances of the estimation obtained through the decomposition in (1), in the MLC we use two estimation approaches, the upward-downward EM proposed in [3], and a Stochastic version of the EM (SMLC). The finite estimation obtained with the first version is compared with the approximated one in two separate simulation studies for binary and categorical observations.
The SMLC algorithm scheme is the following, after initialization of $\pi_h, \pi_{\ell|h}, \pi_{y_{i|\ell}}$, iterate the following sampling steps:

SE step:

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

$$P(W_k = h|\mathbf{y}_k) = \frac{\pi_h P(\mathbf{Y}_k|W_k = h)}{P(\mathbf{Y}_k)}, \text{ where } \quad P(\mathbf{Y}_k|W_k = h) = \prod_{j=1}^{n_k} P(\mathbf{Y}_{jk}|W_k = h)$$

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

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

M step:

$$\pi_h = \frac{\sum_{k=1}^{K} w_k^{h(t)}}{K}, \quad \pi_{\ell|h} = \frac{\sum_{j=1}^{n} w_{jk}^{h(t)} x_{jk}^{\ell(t)}}{\sum_{j=1}^{n} w_{jk}^{h(t)}}, \quad \pi_{i|\ell} = \frac{\sum_{j=1}^{n} x_{jk}^{\ell(t)} y_{ijk}}{\sum_{j=1}^{n} x_{jk}^{\ell(t)}}.$$

with w_k^h binary indicator of units' membership at higher level, x_{jk}^{ℓ} binary indicator of units' membership at lower level and w_{jk}^h the expansion of higher level latent class indicators over the first level units j. Final estimates are then calculated as the mean over the total number of iterations after a burn-in period.
Following the Iterative scheme of the stochastic EM the final computation of CL and EN can be performed following two different approaches:

(a) taking $\hat{\theta}$ as the mean over the iterations of the stochastic algorithm
(b) plugging-in the current $\theta^{(t)}$ parameter values at each iteration and computing the final $CL(\hat{\theta})$ and $EN(\hat{\theta})$ as average over the iterations.

To evaluate the goodness of these different approaches we have performed two simulation studies each including 50 synthetic datasets with $K = 50$, $n_k = 30$ (number of units per group) and a fixed number of classes $L = 4$, $H = 3$. A total of 500 iterations were considered in the SMLC implementation, including 150 as burn-in period. In simulation 1 six binary indicators were generated, in simulation 2 we generated six categorical variables: two binary, two with three modalities and two with four. In Table 1 the results of both simulations scenarios are summarised through the average absolute error resulted in the comparison of the approximated values of $\log L$, as decomposition in the sum of CL and EN, obtained by applying the two approaches proposed, and the finite values of the three components in (1) obtained through the EM, computed with LatentGOLD 6.0 software (see Table 2).
Results from simulations show that the approximation of log-likelihood values computed through approach (a) are preferable, in terms of the error produced, than those given from approach (b). More specifically, we observe that the estimates though approach (b) are to be preferred in the approximation of the CL component, whereas the other approach works better for the approximation of EN.

Table 1 Average absolute errors in the comparison between Stochastic EM approximation and EM finite estimate

	Log-likelihood	CL	EN
Simulation 1			
Approach (a)	1.88	11.00	9.30
Approach (b)	11.14	7.15	14.42
Simulation 2			
Approach (a)	1.08	10.35	9.69
Approach (b)	14.79	8.76	17.67

Table 2 Average values for Log-likelihood, CL and EN computed with LatentGOLD 6.0

	Log-likelihood	CL	EN
Simulation 1	-5552.33	-6221.07	668.74
Simulation 2	-8062.26	-8655.73	593.47

4 The case of MCCLC models

In the extension to cross-classified data structures the estimation algorithm takes into consideration the simultaneous belonging of observations to W_k and Z_q level-2 latent variables, see [1]. In the stochastic algorithm the first point of SE step becomes

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

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

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

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

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

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

In this case, the computation of CL and EN terms requires the knowledge of the double joint posterior probability $P(W_k = h, Z_q = r|\mathbf{y}_{kq})$ which cannot be factorised as product of the conditional probability of each higher level latent class. The estimation through the approach (b) proposed in Section 3 can be performed running an additional Gibbs sampler at each iteration of the stochastic EM, with π_h and π_r set to their current values $\pi_h^{(t)}$ and $\pi_r^{(t)}$. In this case steps 1.1) and 1.2) are repeated several times and that probability is given as mean of the samplings over these iterations.

5 Conclusions

We have provided preliminary results on the study of likelihood approximations in stochastic EM for multilevel latent classes with the aim of extending these insights to situations in which no finite likelihood estimate is available. The simulations in the hierarchical model have suggested that the best likelihood approximation is obtained taking approach (a). We assume that this behaviour is also valid in the cross-classified extension where a double sampling is performed at level-2. In this case the choice of approach (a) over (b) is particularly convenient in terms of computational complexity, indeed it is sufficient a single Gibbs sampling at each E step.

References

1. Columbu, S., Piras, N., Vermunt, J. K. (2003) Multilevel Cross-Classified Latent Class Models, CLADAG Book of Abstracts and Short Papers, ISBN: 9788891935632, 390–393
2. McLachlan, G.J., Peel, D. (2000) Finite Mixture Models. New York: John Wiley & Sons, Inc.
3. Vermunt, J. K. (2008) Latent class and finite mixture models for multilevel data sets. Statistical Methods in Medical Research, 17, 33–51
4. Vermunt, J. K., Magidson, J. (2016) Technical Guide for Latent GOLD 5.1: Basic, Advanced, and Syntax. Belmont, MA: Statistical Innovations Inc