Supplementary material for the paper
Maximum likelihood estimation for discrete latent variable models via evolutionary algorithms

Luca Brusa[1], Fulvia Pennoni[1], Francesco Bartolucci[2]

[1]Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy
Email: luca.brusa@unimib.it, fulvia.pennoni@unimib.it

[2]Department of Economics, University of Perugia, Via Alessandro Pascoli 20, 06123 Perugia, Italy
Email: francesco.bartolucci@unipg.it

In the following sections we provide additional results concerning the simulation study and the applications presented in the paper. The standard expectation-maximization (EM) and variational EM (VEM) algorithms are compared to the proposed evolutionary EM (EEM) and evolutionary VEM (EVEM) algorithms for the following discrete latent variable models (DLV, Bartolucci et al., 2022): latent class (LC), hidden Markov with discrete (HMcat) and continuous (HMcont) responses, and stochastic block (SB) models. Section S1 reports the performance of the EEM and EVEM algorithms in terms of computational time. Section S2 shows a comparison between the EEM and the tempered EM algorithms proposed in Brusa et al. (2023). Section S3 reports additional results for the applicative examples.

S1 Computation time

Simulations presented in Section 4 of the paper are performed by employing a *Standard_HC44rs* virtual machine with 44 vCPUs and 352 GB of RAM. In this section we provide the computational time required for convergence by the proposed EEM and EVEM algorithms under the simulation settings illustrated in Section 4 of the paper. Results are summarized in Table 1. The EEM algorithm is usually slower than the EM; the major difference is observed under each scenario of the LC and HMcat models estimated with the right number of latent components. Considering the SB model, there are instances where the EVEM algorithm demonstrates faster convergence than the VEM. It is worth noting, as shown in the paper, that the advantage of the EEM algorithm with respect to the EM and VEM algorithms in locating the global optimum is relevant, thus a higher computing time is reasonable.

S2 Comparison with the tempered EM algorithm

In this section we provide a comparison of the EEM algorithm with the TEM algorithm for the LC and HM models, and outlined in Section 2 of the paper. Table 2 shows the frequency of convergence to the optimum for each of the mentioned DLV models. Each value is computed as the average over 50 samples and 100 starting values as presented in Section 4 of the paper. Although both TEM and EEM algorithms show a superior performance with respect to the standard EM algorithm, the evolutionary approach always provides the best results, clearly outperforming also the tempered version of the algorithm. A relevant difference between EEM and TEM algorithms concerns the selection of the required constants, which is crucial for both algorithms. As stated in Section 2 of the paper, for the TEM algorithm the interpretation of the tempering constants is not straightforward, and it is not possible to state in which way a change of these values influences the behavior of the TEM algorithm, thus requiring a suitable grid where to search. On the contrary, the EEM algorithm ensures a much simpler interpretation; for example, increasing the number of offspring (N_O) or the probability of mutation (p_m), encourage a broader parameter space exploration.

S3 Additional results of the applications

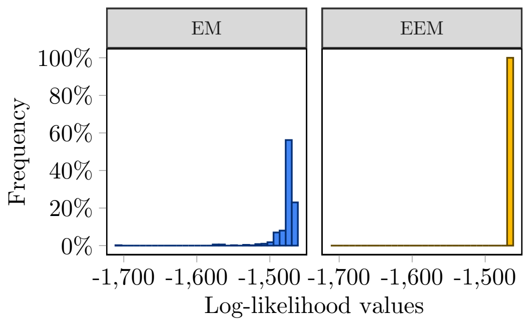
In the following we report some additional results of the applications illustrated in Section 5 of the paper. Figure 1 depicts the frequency of the log-likelihood values at convergence in order to compare the proposed and standard algorithms for each model analyzed. Results reveal that values are significantly different and those obtained with the proposed algorithms are always superior.

Table 1: *Computational time in seconds using EM or VEM algorithm (in blue) and EEM or EVEM algorithm (in yellow) under the simulated scenarios presented in Table 4 of Appendix A of the paper for the LC, HMcat, HMcont, and SB models with correctly specified (top panel) and misspecified (bottom panel) latent structure. Each value is computed as the average over 50 samples and 100 starting values, as presented in Section 4 of the paper. The colored bars show the value obtained with the EEM or EVEM algorithm as a proportion of the corresponding ones computed with the standard counterparts*

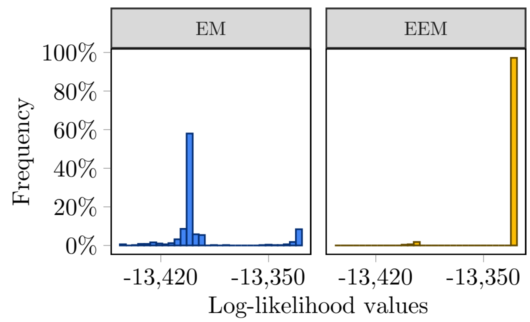
	LC	HMcat	HMcont	SB
Correctly specified latent structure				
A	0.064 0.176	0.505 4.643	1.325 2.995	28.004 22.391
B	0.960 24.940	0.458 5.164	1.688 4.81	20.708 19.348
C	0.571 23.287	0.754 4.078	2.515 4.504	62.346 53.109
D	0.550 22.103	1.014 4.771	2.202 5.879	70.685 57.866
E	0.339 0.771	0.884 8.226	8.345 11.933	137.060 214.009
F	-	38.953 116.228	-	157.011 234.64
Misspecified latent structure				
A	0.209 0.483	25.200 72.236	11.050 32.053	59.874 117.427
B	12.814 44.219	31.295 96.272	19.623 60.128	58.222 121.965
C	3.848 33.133	12.392 38.149	15.010 40.474	210.676 288.333
D	5.924 35.764	13.258 38.707	24.644 81.217	264.795 326.226
E	0.374 0.888	19.863 66.393	16.762 44.995	152.538 361.739
F	-	20.547 67.828	-	166.975 397.955

Table 2: *Percentage of convergence to the global maximum using EM (in blue), TEM (in red), and EEM (in yellow) algorithms under the simulated scenarios presented in Table 4 in Appendix A of the paper for the LC, HMcat, and HMcont models with misspecified latent structure. Each value is computed as the average over 50 samples and 100 starting values as presented in Section 4 of the paper*

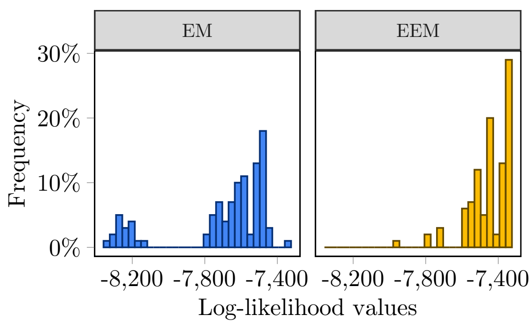
	LC	HMcat	HMcont
A	29% 31% 93%	14% 18% 51%	33% 36% 83%
B	41% 48% 98%	27% 34% 68%	50% 53% 85%
C	18% 21% 92%	12% 23% 50%	43% 49% 88%
D	16% 28% 96%	16% 24% 52%	24% 26% 65%
E	3% 7% 54%	11% 16% 44%	8% 10% 15%
F	-	2% 6% 9%	-



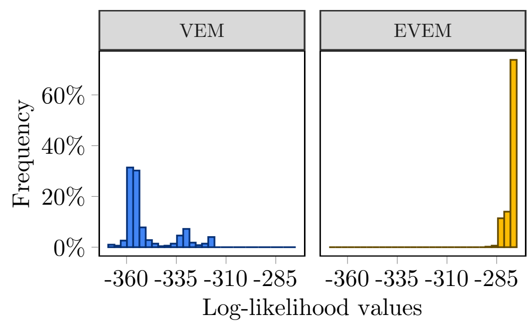
(a) *LC*



(b) *HMcat*



(c) *HMcont*



(d) *SB*

Figure 1: *Comparison between log-likelihood values at convergence obtained estimating the LC, HMcat, HMcont, and SB models using EM or VEM algorithms (in blue) and EEM or EVEM algorithms (in yellow) with data illustrated in Section 5 of the paper for each DLV model*

References

BARTOLUCCI, F., PANDOLFI, S., AND PENNONI, F. (2022). Discrete latent variable models. *Annu. Rev. Stat. Appl.*, **6**, 1–31.

BRUSA, L., BARTOLUCCI, F., AND PENNONI, F. (2023). Tempered expectation-maximization algorithm for the estimation of discrete latent variable models. *Comput. Stat.*, **38**, 1391–1424.