Misspecification tests for hidden Markov models based on a new class of finite mixture models

Fulvia Pennoni
Department of Statistics and Quantitative Methods
University of Milano-Bicocca
Email: fulvia.pennoni@unimib.it

Joint work with

Francesco Bartolucci
&
Silvia Pandolfi

University of Perugia

Outline

- A class of finite mixture models that includes hidden Markov models
- Maximum likelihood estimation and misspecification tests for hidden Markov models
- Simulation studies and empirical illustration
- Conclusions

Introduction

- In the context of longitudinal data, we introduce a new class of finite mixture (FM) models (McLachlan et al., 2019) that generalizes that of hidden Markov (HM) models (Bartolucci et al., 2013)
- This new class of models, denoted as FM2 models, is based on an augmented set of components and suitable constraints on the conditional response probabilities
- We derive conditions under which an FM2 model coincides with an HM model
- We develop a likelihood ratio (LR) misspecification test for the latent structure of an HM model and a multiple version of this test that may be used in the presence of many latent states or time occasions

Notation

- $\boldsymbol{Y}_{it}$ vectors of response variables collected in vector $\boldsymbol{Y}_i$ for every sample unit i and time occasion t, with $i = 1, \ldots, n$ and $t = 1, \ldots, T$
- The FM formulation with k components:

$$f_{FM}(\boldsymbol{y}_i) = \sum_{u=1}^{k} \pi_u g_u(\boldsymbol{y}_i),$$

  - U_i: discrete latent variables with probabilities $\pi_u = p(U_i = u)$, $u = 1, \ldots, k$
  - $g_u(\boldsymbol{y}_i)$ conditional density for component u
  - Conditional independence assumption: $g_u(\boldsymbol{y}_i) = \prod_{t=1}^{T} g_{tu}(\boldsymbol{y}_{it})$

Notation

- The HM formulation with l latent states:

$$f_{HM}(\boldsymbol{y}_i) = \sum_{v_1=1}^{l} \lambda_{v_1} h_{v_1}(\boldsymbol{y}_{i1}) \sum_{v_2=1}^{l} \rho_{v_1 v_2} h_{v_2}(\boldsymbol{y}_{i2}) \cdots \cdots \sum_{v_T=1}^{l} \rho_{v_{T-1} v_T} h_{v_T}(\boldsymbol{y}_{iT}),$$

  - $(V_{i1}, \ldots, V_{iT})$: discrete latent variables following a first-order Markov process
  - $\lambda_v = P(V_{i1} = v)$: initial probabilities
  - $\rho_{\bar{v}v} = P(V_{it} = v | V_{i,t-1} = \bar{v})$: transition probabilities
  - $h_v(\boldsymbol{y}_{it})$: conditional response density

A new class of finite mixture models

- FM2 is defined as a model with a number of mixture components equal to the number of possible latent state configurations $k = l^T$ and suitable constraints on the conditional distribution

- The requirement about the conditional probabilities is:

$$g_u(\boldsymbol{y}_i) = \prod_{t=1}^{T} h_{v_t(u)}(\boldsymbol{y}_{it}), \tag{1}$$

  - $v_t(u)$: value of the t-th latent state corresponding to the u-th mixture component

FM2 model assumptions

- Manifest distribution

$$f_{FM2}(\boldsymbol{y}_i) = \sum_{u=1}^{l^T} \left[\pi_u \prod_{t=1}^{T} h_{v_t(u)}(\boldsymbol{y}_{it}) \right]$$

- Condition under which the FM2 and the HM formulations become equivalent

$$\pi_u = \lambda_{v_1(u)} \prod_{t=2}^{T} \rho_{v_{t-1}(u)v_t(u)},$$

- The number of independent constraints on the parameters π_u of an FM2 model to be equivalent to an HM model corresponds to the difference between the number of free parameters of the two models

$$\#\mathrm{df} = l^T - l^2 = l^2(l^{T-2} - 1)$$

Equivalence conditions

- Example of equivalence between FM and HM models when $T = 3$ time occasions and $l = 2$ latent states for a number of mixture components ranging from 1 to 8

mixture component (u)	sequence of latent states ($\boldsymbol{v}$)	FM conditional distribution	HM conditional distribution	FM probability	HM probability
1	(1,1,1)	$g_1(\boldsymbol{y}_i)$	$h_1(\boldsymbol{y}_{i1})h_1(\boldsymbol{y}_{i2})h_1(\boldsymbol{y}_{i3})$	π_1	$\lambda_1\rho_{11}\rho_{11}$
2	(1,1,2)	$g_2(\boldsymbol{y}_i)$	$h_1(\boldsymbol{y}_{i1})h_1(\boldsymbol{y}_{i2})h_2(\boldsymbol{y}_{i3})$	π_2	$\lambda_1\rho_{11}\rho_{12}$
3	(1,2,1)	$g_3(\boldsymbol{y}_i)$	$h_1(\boldsymbol{y}_{i1})h_2(\boldsymbol{y}_{i2})h_1(\boldsymbol{y}_{i3})$	π_3	$\lambda_1\rho_{12}\rho_{21}$
4	(1,2,2)	$g_4(\boldsymbol{y}_i)$	$h_1(\boldsymbol{y}_{i1})h_2(\boldsymbol{y}_{i2})h_2(\boldsymbol{y}_{i3})$	π_4	$\lambda_1\rho_{12}\rho_{22}$
5	(2,1,1)	$g_5(\boldsymbol{y}_i)$	$h_2(\boldsymbol{y}_{i1})h_1(\boldsymbol{y}_{i2})h_1(\boldsymbol{y}_{i3})$	π_5	$\lambda_2\rho_{21}\rho_{11}$
6	(2,1,2)	$g_6(\boldsymbol{y}_i)$	$h_2(\boldsymbol{y}_{i1})h_1(\boldsymbol{y}_{i2})h_2(\boldsymbol{y}_{i3})$	π_6	$\lambda_2\rho_{21}\rho_{12}$
7	(2,2,1)	$g_7(\boldsymbol{y}_i)$	$h_2(\boldsymbol{y}_{i1})h_2(\boldsymbol{y}_{i2})h_1(\boldsymbol{y}_{i3})$	π_7	$\lambda_2\rho_{22}\rho_{21}$
8	(2,2,2)	$g_8(\boldsymbol{y}_i)$	$h_2(\boldsymbol{y}_{i1})h_2(\boldsymbol{y}_{i2})h_2(\boldsymbol{y}_{i3})$	π_8	$\lambda_2\rho_{22}\rho_{22}$

- Third and fourth columns illustrate condition (1) and the last two columns illustrate condition (2)

Testing procedure

- Proposed LR misspecification test:

$$LR = -2(\hat{\ell}_{HM} - \hat{\ell}_{FM2}),$$

  - $\hat{\ell}_{HM}$ and $\hat{\ell}_{FM2}$: maximum log-likelihood of the HM and FM2 models, respectively

- Assuming standard regularity conditions and locally identified models the asymptotic $p-$value for LR is computed according to a χ^2 distribution with $\#\mathrm{df} = I^T - I^2$ degrees of freedom

- The null hypothesis is rejected when the latent structure of the HM model is not correctly specified

- Maximum likelihood estimation of the model parameters for both models is based on the complete data log-likelihood function and it is performed through the Expectation-Maximization (EM) algorithm

- Suitable recursions (forward and backward) are able to reduce the computational burden of estimation strongly

LR test

- The number of possible sequences of latent states (l^T) may become easily huge
  - the estimation of the FM2 models requires high computational times
  - very reduced power of the test in rejecting an alternative hypothesis

- We suggest an alternative approach based on performing the test for each possible subsequence of three consecutive time periods and including the Bonferroni correction for multiple tests

- The overall p-value is taken as

$$\min\{1, (T-2)\min(p_1, \ldots, p_{T-2})\}$$

  where $p_1, \ldots, p_{T-2}$ are the p-values of each test

- Model selection is performed with the Bayesian Information Criterion (BIC, Schwarz, 1978); however, the index may constantly decrease for large data: the proposed test is valid also to choose the suitable number of hidden states for the HM model

Simulation study

- We considered a large scale Monte Carlo simulation study based on three distinct experimental scenarios under different for model specifications

- We simulated random samples:
  - under the null hypothesis of a multivariate time homogeneous Gaussian HM model
  - we considered two settings where the latent structure of the HM model is not correctly specified: by simulating from both a second-order HM model and from an HM model with time heterogeneous transition probabilities

- We estimate FM2 model having a number of components equal to l^T assuming a homoscedastic multivariate Gaussian distribution

$$f_{FM2}(\boldsymbol{y}_i) = \sum_{u=1}^{l^T} \left[\pi_u \prod_{t=1}^{T} h_{v_t(u)}(\boldsymbol{y}_{it}) \right] = \sum_{u=1}^{l^T} \left[\pi_u \prod_{t=1}^{T} \phi(\boldsymbol{y}_{it}; \boldsymbol{\mu}_{v_t(u)}, \boldsymbol{\Sigma}) \right]$$

Simulation study

- In order to evaluate the performance of the testing procedure, we considered the distribution of the LR test statistic used to compare the HM model against the FM2
- The simulation results show that when the number of latent states l is moderate, the asymptotic null distribution of the LR statistic is a reasonable approximation of the finite sample distribution, especially as the sample size increases
- Under both alternative experimental settings the power of the test increases with the sample size n, with relatively large probabilities of detecting deviations from the null hypothesis

Empirical illustration

- Data from the National Longitudinal Survey of Youth frequently employed in microeconometrics and labor economics applications
- We selected all the respondents who had completed the first five waves of the survey since 1979: 729 individuals up to 1983
- We model the natural log of the hourly rate of pay (log-wage) at the respondent's current or most recent job and a variable indicating the temporal man's labor force history associated with each observed value of the wage
- We estimated a first-order multivariate HM model with time homogeneous transition probabilities assuming a conditional distribution of Gaussian type for the response variables

Results: model selection

- HM and FM2 models with a number of states from 1 to 6

k	AIC	BIC	LR	#df	p-value	p-value$_m$
1	17332.743	17355.711	-	-	-	-
2	15022.925	15068.842	398.669	28	0.0000	0.0000
3	13550.814	13628.872	528.810	234	0.0000	0.0517
4	12362.549	12481.932	367.242	1008	1.0000	0.7785
5	11232.539	11402.431	164.628	3100	1.0000	1.0000
6	10205.395	10434.979	94.957	7740	1.0000	1.0000

  - The estimated p-values show that the HM model is misspecified when it is estimated with a number of hidden states less than 4
  - The realized p-values of the multiple LR misspecification tests based on consecutive triplets of observations lead us to select the HM model with 4 hidden states
  - The proposed LR test statistic permits choosing the most parsimonious HM model when indices like Akaike Information Criterion and BIC constantly decrease

Results: conditional means

- The estimated conditional means of the HM model with 4 hidden states show that:

	$v=1$	$v=2$	$v=3$	$v=4$
Log wage	1.734	1.797	1.905	2.033
Experience	0.830	2.126	3.485	5.273

  - Individuals for the first group, about 97% in the first year of the survey, have the lowest wages and relatively low working experience
  - Individuals in the second group, about 2.7% of males, earn the same wage as individuals in the first group but holds more working experience
  - Males in the third and especially in the fourth group earn more and have much more working experience than those of the first two groups

Results: transition probabilities

- The estimated transition probabilities show that:

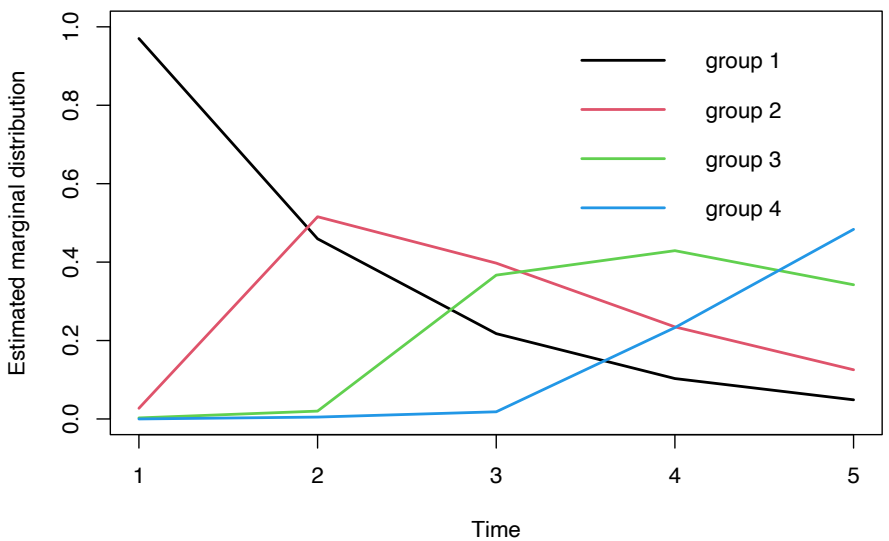
	$v = 1$	$v = 2$	$v = 3$	$v = 4$
$\hat{\lambda}_v$	0.970	0.027	0.003	0.000
$\hat{\rho}_{v\|1}$	0.474	0.523	0.000	0.003
$\hat{\rho}_{v\|2}$	0.000	0.305	0.695	0.000
$\hat{\rho}_{v\|3}$	0.000	0.000	0.417	0.583
$\hat{\rho}_{v\|4}$	0.000	0.000	0.000	1.000

  - There are progressive career advancements for some individuals
  - High persistence in the same group over the years is observed for those with the highest wages (group 4)
  - Around 70% transit from group 2 to group 3

Results: marginal probabilities

- The temporal trend of each group estimated according to the marginal distribution of the latent variable

- Probabilities of groups 1 and 2 decrease over time while the probability of group 3 increases up to the fourth wave, and that of group 4 increases in the last three waves



Conclusions

- We compared two broad and important classes of models used to analyze longitudinal data: finite mixture (FM) models and hidden Markov (HM) models
- We showed that it is possible to formulate an FM model, named FM2 that generalizes an assumed HM model
- We implemented a likelihood ratio (LR) misspecification test on the latent structure of an HM model
- We proposed a multiple version of this test based on each possible subsequence of three consecutive observations considering the Bonferroni correction in order to apply the proposed testing procedure for many time occasions and several hidden states

Main References

- Bartolucci, F., S. Pandolfi, and F. Pennoni (2022). Discrete latent variable models. *Annual Review of Statistics and its Application* 9, 1-31.
- Bartolucci, F, Farcomeni, A, and Pennoni, F. (2013) *Latent Markov Models for Longitudinal Data*. Boca Raton, FL: Chapman and Hall/CRC.
- Bartolucci, F., Pandolfi, S., and Pennoni, F. (2017). LMest: An R package for latent Markov models for longitudinal categorical data. *Journal of Statistical Software*, **81**, 1-38.
- McLachlan, G. and D. Peel (2000). *Finite Mixture Models*. New York: Wiley.
- McLachlan, G. J., S. X. Lee, and S. I. Rathnayake (2019). Finite mixture models. *Annual Review of Statistics and its Application* 10, 355–378.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6, 461-464.