Latent Markov models for longitudinal data in R by LMest package

Francesco Bartolucci

Department of Economics

University of Perugia (IT)

https://sites.google.com/site/bartstatistics/

francesco.bartolucci@unipg.it

Presentation at the European R Users Meeting - Milan

June 18th, 2020

Outline

1. Elements of theory:
   - Preliminary definitions
   - Latent Markov models
   - Estimation & model selection
   - Prediction of latent states
2. `LMest` package:
   - Data processing
   - Specification of formulas
   - Example with continuous responses
   - Examples with categorical responses

Credits

1. `LMest` package by:
   - F.Bartolucci - University of Perugia, IT
   - S.Pandolfi - University of Perugia, IT
   - F.Pennoni - University of Milano-Bicocca, IT
   - A.Farcomeni - University of Rome-Tor Vergata, IT
   - A.Serafini - Sose Rome, IT

2. Illustration of the `LMest` package by:
   - F.Pennoni - University of Milano-Bicocca, IT
   - S.Pandolfi - University of Perugia, IT
   - F.Bartolucci - University of Perugia, IT

Preliminary definitions

- Given a set of response variables $\boldsymbol{Y}$ and a set of covariates $\boldsymbol{X}$, *latent variables* ($\boldsymbol{U}$) are unobservable variables supposed to exist and to affect $\boldsymbol{Y}$; these may be correlated with $\boldsymbol{X}$

- A *latent variable model* (LVM) formulates assumptions on:
  - the conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{U}$ and $\boldsymbol{X}$, $f(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{x})$ (*measurement model*)
  - the conditional distribution of $\boldsymbol{U}$ given $\boldsymbol{X}$, $f(\boldsymbol{u}|\boldsymbol{x})$ (*structural model*)

- By marginalizing out the latent variables we obtain the *manifest distribution* $f(\boldsymbol{y}|\boldsymbol{x})$; by the Bayes theorem we obtain the *posterior distribution* $f(\boldsymbol{u}|\boldsymbol{x},\boldsymbol{y})$

- A common assumption of LVMs is that of *local independence*, according to which the response variables are conditionally independent given the latent variables and the covariates
- Latent variables are typically *included in a LVM* with different aims, such as:
  - accounting for the *unobserved heterogeneity* among subjects
  - accounting for *measurement errors*
  - *summarizing different measurements* of the same (directly) unobservable characteristics
- We focus in particular on *Latent (Hidden) Markov* (LM) models for longitudinal data

Latent Markov models (basic version)

- These are models for the analysis of *longitudinal categorical (or continuous)* data; they are used in many contexts, e.g., economics, education, psychology, sociology
- *Main references*: Wiggins (1973), Bartolucci *et al.* (2013), Zucchini *et al.* (2016)
- For individual sequences of response variables at T time occasions, $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{iT})'$, $i = 1, \dots, n$, the *basic version of the LM model* assumes that:
  - (*local independence*, LI) the response variables in $\boldsymbol{Y}_i$ are conditionally independent given a latent process $\boldsymbol{U}_i = (U_{i1}, \dots, U_{iT})'$
  - every latent process $\boldsymbol{U}_i$ follows a *first-order Markov chain* with state space $\{1, \dots, k\}$, initial probabilities π_u, and transition probabilities $\pi_{v|u}$, $u, v = 1, \dots, k$

Possible interpretation

- The LM model may be seen as a *generalization of the LC model* in which subjects are allowed to move between latent classes

LC:

$$
\begin{tikzcd}
Y_{i1} & Y_{i2} & \cdots & Y_{iT} \\
U_i \arrow[u] \arrow[ur] \arrow[urrr] & & &
\end{tikzcd}
$$

LM:

$$
\begin{tikzcd}
Y_{i1} & Y_{i2} & \cdots & Y_{iT} \\
U_{i1} \arrow[u] \arrow[r] & U_{i2} \arrow[u] \arrow[r] & \cdots \arrow[r] & U_{iT} \arrow[u]
\end{tikzcd}
$$

Model parameters

- Each latent state u ($u = 1, \dots, k$) corresponds to a *class of subjects* in the population, and is characterized by:
  - *initial probability*:
$$\pi_u = p(U_{i1} = u)$$
  - *transition probabilities* (which may also be time-specific in the non-homogenous case):
$$\pi_{v|u} = p(U_{it} = v | U_{i,t-1} = u), \quad t = 2, \dots, T, \; u, v = 1, \dots, k$$
  - *distribution of the response variables* (with categorical responses):
$$\phi_{y|u} = p(Y_{it} = y | U_{it} = u), \quad t = 1, \dots, T, \; u = 1, \dots, k, \; y = 0, \dots, l-1$$
- The transition probabilities are collected in the *transition matrix* $\mathbf{\Pi}$ of size $k \times k$

Manifest distribution

- LI implies that the *conditional distribution* of $\boldsymbol{Y}_i$ given $\boldsymbol{U}_i$ is:

$$p(\boldsymbol{y}_i|\boldsymbol{u}_i) \;=\; p(\boldsymbol{Y}_i=\boldsymbol{y}_i|\boldsymbol{U}_i=\boldsymbol{u}_i) = \prod_{t=1}^{T}\phi_{y_{it}|u_{it}}$$

- *Distribution* of $\boldsymbol{U}_i$: $$p(\boldsymbol{u}_i) = p(\boldsymbol{U}_i=\boldsymbol{u}_i) = \pi_{u_{i1}}\prod_{t>1}\pi_{u_{it}|u_{i,t-1}}$$

- *Manifest distribution* of $\boldsymbol{Y}_i$: $$p(\boldsymbol{y}_i) = p(\boldsymbol{Y}_i=\boldsymbol{y}_i) = \sum_{\boldsymbol{u}} p(\boldsymbol{y}_i|\boldsymbol{u})p(\boldsymbol{u})$$

- This may be *efficiently computed* through suitable recursions known in the hidden Markov literature (Baum *et al.*, 1970, Welch 2003)

Multivariate extension of the basic LM model

- We consider a *vector of J response variables* $\boldsymbol{Y}_{it} = (Y_{i1t}, \ldots, Y_{iJt})'$ for each individual i and time occasion t, $i = 1, \ldots, n$, $t = 1, \ldots, T$
- With categorical responses, the elements of $\boldsymbol{Y}_{it}$ are assumed to be *conditionally independent* given U_{it} (LI), so that

$$\begin{aligned} p(\boldsymbol{y}_{it}|u_{it}) &= p(\boldsymbol{Y}_{it} = \boldsymbol{y}_{it}|U_{it} = u_{it}) = \prod_{j=1}^{J} \phi_{jy_{ijt}|u_{it}} \\ \phi_{jy|u} &= p(Y_{ijt} = y|U_{it} = u) \end{aligned}$$

- With *continuous responses*, it is typically assumed that

$$\boldsymbol{Y}_{it}|u_{it} \sim N(\boldsymbol{\mu}_{u_{it}}, \boldsymbol{\Sigma})$$

- The latent variables $U_{i1}, \ldots, U_{iT}$ are assumed to follow a *first-order Markov chain* (possibly non-homogeneous)

Inclusion of covariates in the basic LM model

- Two possible *choices to include individual covariates* collected in $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})$

- The first is in the *measurement model* so that we have random intercepts; for binary variables we could assume:

$$\lambda_{itu} = p(Y_{it} = 1 | U_{it} = u, \boldsymbol{X}_i),$$

$$\log \frac{\lambda_{itu}}{1 - \lambda_{itu}} = \alpha_u + \boldsymbol{x}'_{it}\boldsymbol{\beta}, \quad i = 1, \ldots, n, \;\; t = 1, \ldots, T, \; u = 1, \ldots, k$$

- The latent variables are used to account for the *unobserved heterogeneity* and then the model may be seen as a "discrete version" of the logistic model for longitudinal data with random effects

- The second is in the *structural model* governing the distribution of the latent variables (via a multinomial logit parametrization):
  - initial probabilities:

$$\pi_{iu} = p(U_{i1} = u|\boldsymbol{x}_{i1}), \qquad \log\frac{\pi_{iu}}{\pi_{i1}} = \boldsymbol{x}'_{i1}\boldsymbol{\beta}_u, \quad u = 2,\dots,k$$

  - transition probabilities:

$$\begin{aligned} \pi_{itv|u} &= p(U_{it} = v|U_{i,t-1} = u, \boldsymbol{x}_{it}), \\ \log\frac{\pi_{itv|u}}{\pi_{itu|u}} &= \boldsymbol{x}'_{it}\boldsymbol{\gamma}_{uv}, \quad u, v = 1,\dots,k,\ u \neq v \end{aligned}$$

- The *main interest is on the latent variable* which is measured through the observable response variables (e.g., health status) and on how this latent variable evolves depending on the covariates

Mixed latent Markov models

- *Additional random effects/latent variables* may be included in an LM model to account for further sources of unobserved heterogeneity (van de Pol and Langeheine, 1990; Altman, 2007; Maroutti, 2011)

- Among the mixed LM models, we focus on that based on initial and transition probabilities of the individual latent processes defined *conditional on a discrete latent variables* V_i, $i = 1, \dots, n$

- The model assumes that individuals are divided in *latent clusters*, with individuals in the same cluster following the same LM model, while the measurement model is common to all individuals

- Mixed LM models may be also used for *multilevel longitudinal data* with individuals collected in observable groups (Bartolucci et al., 2011)

Maximum likelihood estimation of the basic LM model

- Model *log-likelihood*: $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log p(\boldsymbol{y}_i) = \sum_{\boldsymbol{y}} n(\boldsymbol{y}) \log p(\boldsymbol{y})$
  - $\boldsymbol{\theta}$: vector of all model parameters (π_u, $\pi_{v|u}$, $\phi_{y|u}$ for categorical data)
  - $n(\boldsymbol{y})$: frequency of the response configuration $\boldsymbol{y}$
- $\ell(\boldsymbol{\theta})$ may be maximized by an *EM algorithm* (Baum *et al.*, 1970, Dempster *et al.*, 1977)
- The EM algorithms alternates *two steps* until convergence:
  - *E-step*: compute the *posterior distribution* of the latent states given the observed data and the current value of the parameters
  - *M-step*: maximize the posterior expected value of the *complete data log-likelihood* with respect to the model parameters

- Suitable *recursions* are required to compute the $\ell(\boldsymbol{\theta})$ and to perform the E-step (Baum *et al.*, 1970; Welch, 2003)

- Being $\ell(\boldsymbol{\theta})$ *multimodal*, different initializations (deterministic and random) of the algorithm are necessary to increase the chance to get its global maximum

- *Extended models* (multivariate, with covariates, and mixed) are still fitted by an EM algorithm in which the main adjustments are in the M-step

- If necessary, *selection of k* may be based on suitable statistical criteria (Akaike, 1973, Schwarz, 1978):

$$
\begin{aligned}
AIC &= -2\ell(\hat{\boldsymbol{\theta}}) + 2\#\text{param.} \\
BIC &= -2\ell(\hat{\boldsymbol{\theta}}) + \log(n)\#\text{param.}
\end{aligned}
$$

Prediction of latent states (dynamic pattern recognition)

- The posterior probabilities

$$p(u_t|\boldsymbol{y}_i) = p(U_{it} = u_t|\boldsymbol{Y}_i = \boldsymbol{y}_i)$$

may be used to *assign a subject* to a latent state at a given time occasion (*local decoding*); state assigned to subject i at occasion t:

$$\hat{u}_{it} : p(\hat{u}_{it}|\boldsymbol{y}_i) = \max_{u_t} p(u_t|\boldsymbol{y}_i)$$

- More sophisticated is the problem of *path prediction*, i.e., finding the most likely sequence $\tilde{\boldsymbol{u}}_i = (\tilde{u}_{i1}, \dots, \tilde{u}_{iT})'$ for subject i (*global decoding*):

$$\tilde{\boldsymbol{u}}_i : p(U_{i1} = \tilde{u}_{i1}, \dots, U_{iT} = \tilde{u}_{iT}|\boldsymbol{y}_i) = \max_{\boldsymbol{u}} p(\boldsymbol{u}|\boldsymbol{y}_i)$$

- For this aim we need an *iterative algorithm* due to Viterbi (1967) and further elaborated by Juan & Rabiner (1991)

References

- Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In: B. N. Petrov and F. Csaki, eds., *Second International Symposium on Information Theory*, 267–281. Akademiai Kiado, Budapest.
- Altman, R. M. (2007). Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal data setting. *Journal of the American Statistical Association*, **102**:201–210.
- Bartolucci, F., Farcomeni, A., and Pennoni, F. (2013). *Latent Markov Models for Longitudinal Data*. Chapman and Hall/CRC press, Boca Raton, FL.
- 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, doi:10.18637/jss.v081.i04.
- Bartolucci, F., Pennoni, F., and Vittadini, G. (2016). Causal latent Markov model for the comparison of multiple treatments in observational longitudinal studies. *Journal of Educational and Behavioral Statistics*, **41**:146–179.
- Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, **41**:164–171.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**:1–38.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, **61**:215–231.

- Juang, B. H. and Rabiner, L. R. (1991). Hidden Markov models for speech recognition. *Technometrics*, **33**:251–272.
- Lazarsfeld, P. F. and Henry, N. W. (1968). *Latent Structure Analysis*. Houghton Mifflin, Boston.
- Maruotti, A. (2011). Mixed hidden Markov models for longitudinal data: An overview. *International Statistical Review*, **79**, 427–454.
- McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley, New York.
- Montanari, G. E., Doretti, M., & Bartolucci, F. (2018). A multilevel latent Markov model for the evaluation of nursing homes' performance. *Biometrical Journal*, **60**:962-978.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**:461–464.
- van de Pol, F. and Langeheine, R. (1990). Mixed Markov latent class models. *Sociological Methodology*, 20, 213–247.
- Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, **13**:260–269.
- Welch, L. R. (2003). Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter*, **53**:1–13.
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Addison-Wesley Pub. Co.
- Zucchini, W. and MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: an Introduction Using R*. Chapman and Hall/CRC.