Basic Latent Markov model

Silvia Bacci

silvia.bacci@unipg.it
Dipartimento di Economia - Università degli Studi di Perugia (IT)

Outline

Introduction

- Background:

  Latent Markov (LM) models (Wiggins, 1973; Bartolucci et al., 2012) are successfully applied in the analysis of longitudinal data: they allow to take into account several aspects, such as serial dependence between observations, measurement errors, unobservable heterogeneity

  LM models assume that one or more occasion-specific response variables depends only on a discrete latent variable characterized by a given number of latent states which in turn depends on the latent variables corresponding to the previous occasions according to a first-order Markov chain

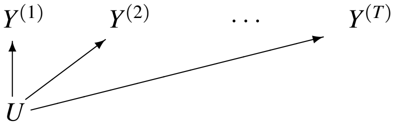
  LM models are characterized by several parameters: the initial probabilities to belong to a given latent state, the transition probabilities from a latent state to another one, the conditional response probabilities given the discrete latent variable

The basic LM model may be seen as

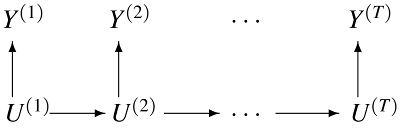
1. a generalization of a discrete-time Markov chain model to account for measurement errors in the observed variables of interest
2. a generalization of a latent class (LC) model for longitudinal data, in which each subject may move between latent classes

E.g., in the univariate case

LC:



LM:



Notation

- Repeated measurements of the same response variable on the same subjects at different occasions
- $\tilde{\boldsymbol{Y}} = (Y^{(1)}, \ldots, Y^{(T)})$: vector of values assumed by the categorical response variable Y at time t ($t = 1, \ldots, T$), having c categories
- $U^{(t)}$: latent state at time t with state space $\{1, \ldots, k\}$
- $\boldsymbol{U} = (U^{(1)}, \ldots, U^{(T)})$: vector describing the latent process

Main assumptions

- local independence: response variables in $\tilde{\boldsymbol{Y}}$ are conditionally independent given the latent process $\boldsymbol{U}$, i.e.,

  each occasion-specific observed variable $Y^{(t)}$ is independent of all its previous values $Y^{(t-1)}, \dots, Y^{(1)}$, given $U^{(t)}$

- latent process $\boldsymbol{U}$ follows a first-order Markov chain with k latent states, i.e.,

  each latent variable $U^{(t)}$ is independent of $U^{(t-2)}, \dots, U^{(1)}$, given $U^{(t-1)}$

Parameters

- $k(c-1)$ conditional response probabilities

  $\phi^{(t)}_{y|u} = p(Y^{(t)} = y|U^{(t)} = u) \quad t = 1,\dots,T;\ u = 1,\dots,k;\ y = 0,\dots,c-1$

- $(k-1)$ initial probabilities

  $\pi_u = p(U^{(1)} = u) \quad u = 1,\dots,k$

- $(T-1)k(k-1)$ transition probabilities

  $\pi^{(t)}_{u|v} = p(U^{(t)} = u|U^{(t-1)} = v) \quad t = 2,\dots,T;\ u,v = 1,\dots,k$

- $\#\text{par} = k(c-1) + (k-1) + (T-1)k(k-1)$

Probability distributions

- $p(\boldsymbol{U}=\boldsymbol{u}) = \pi_u \prod_{t=2}^{T} \pi_{u|v}^{(t)} = \pi_u \cdot \pi_{u_2|u}^{(2)} \dots \pi_{u_T|u_{T-1}}^{(T)}$
- $p(\tilde{\boldsymbol{Y}}=\boldsymbol{y}|\boldsymbol{U}=\boldsymbol{u}) = \prod_{t=1}^{T} \phi_{y|u}^{(t)} = \phi_{y|u}^{(1)} \cdot \phi_{y|u}^{(2)} \dots \phi_{y|u}^{(T)}$
- manifest distribution of $\tilde{\boldsymbol{Y}}$

$$
\begin{aligned}
p(\tilde{\boldsymbol{Y}}=\boldsymbol{y}) &= \sum_{\boldsymbol{u}} p(\tilde{\boldsymbol{Y}}=\boldsymbol{y}, \boldsymbol{U}=\boldsymbol{u}) = \sum_{\boldsymbol{u}} p(\boldsymbol{U}=\boldsymbol{u}) \cdot p(\tilde{\boldsymbol{Y}}=\boldsymbol{y}|\boldsymbol{U}=\boldsymbol{u}) \\
&= \sum_{u} \pi_u \phi_{y|u}^{(1)} \cdot \sum_{u_2} \pi_{u_2|u}^{(2)} \phi_{y|u}^{(2)} \dots \sum_{u_T} \pi_{u_T|u_{T-1}}^{(T)} \phi_{y|u}^{(T)} \\
&= \sum_{u} \sum_{u_2} \dots \sum_{u_T} \pi_u \prod_{t=2}^{T} \pi_{u|v}^{(t)} \prod_{t=1}^{T} \phi_{y|u}^{(t)}
\end{aligned}
$$

Note that computing $p(\tilde{\boldsymbol{Y}}=\boldsymbol{y})$ involves all the possible k^T configurations of vector $\boldsymbol{u}$

Computing of the manifest distribution: example

- We assume three occasions ($T = 3$) and three latent states ($k = 3$)
- We have $3^3 = 27$ possible configurations of vector $\boldsymbol{u}$
- The manifest distribution of $\tilde{\boldsymbol{Y}}$ is given by:

$$\begin{aligned} p(\tilde{\boldsymbol{Y}} = \boldsymbol{y}) = & \, \pi_1 \pi^{(2)}_{1|1} \pi^{(3)}_{1|1} \phi^{(1)}_{y|1} \phi^{(2)}_{y|1} \phi^{(3)}_{y|1} + \\ & + \pi_1 \pi^{(2)}_{1|1} \pi^{(3)}_{2|1} \phi^{(1)}_{y|1} \phi^{(2)}_{y|1} \phi^{(3)}_{y|2} + \\ & + \ldots + \\ & + \pi_3 \pi^{(2)}_{3|3} \pi^{(3)}_{3|3} \phi^{(1)}_{y|3} \phi^{(2)}_{y|3} \phi^{(3)}_{y|3} \end{aligned}$$

Notation

- $\boldsymbol{Y}^{(t)} = (Y_1^{(t)}, \ldots, Y_r^{(t)})$: vector of categorical response variables Y_j ($j = 1, \ldots, r$) observed at time t ($t = 1, \ldots, T$), having c_j categories
- $\boldsymbol{Y} = (\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(T)})$: vector of observed responses made of the union of vectors $\boldsymbol{Y}^{(t)}$; usually, it is referred to repeated measurements of the same variables Y_j ($j = 1, \ldots, r$) on the same individuals at different time points
- $U^{(t)}$: latent state at time t with state space $\{1, \ldots, k\}$
- $\boldsymbol{U} = (U^{(1)}, \ldots, U^{(T)})$: vector describing the latent process

Main assumptions

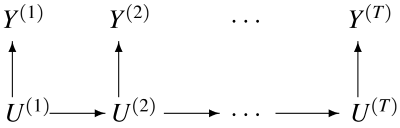
- Local independence: vectors $\boldsymbol{Y}^{(t)}$ $(t = 1, \ldots, T)$ are conditionally independent given the latent process $\boldsymbol{U}$ and the response variables in each $\boldsymbol{Y}^{(t)}$ are conditionally independent given $U^{(t)}$, i.e.,

  each occasion-specific observed variable $Y_j^{(t)}$ is independent of $Y_j^{(t-1)}, \ldots, Y_j^{(1)}$ and of each $Y_h^{(t)}$, for all $h \neq j = 1, \ldots, r$, given $U^{(t)}$

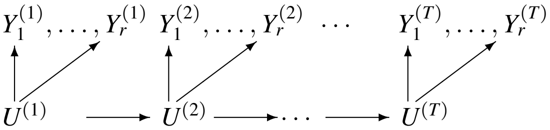
- latent process $\boldsymbol{U}$ follows a first-order Markov chain with k latent states, i.e.,

  each latent variable $U^{(t)}$ is independent of $U^{(t-2)}, \ldots, U^{(1)}$, given $U^{(t-1)}$

Univariate LM:



Multivariate LM:



Parameters

- $k\sum_{j=1}^{r}(c_j-1)$ conditional response probabilities

  $\phi_{jy|u}^{(t)} = p(Y_j^{(t)} = y|U^{(t)} = u) \quad j=1,\ldots,r;\ t=1,\ldots,T;\ u=1,\ldots,k;\ y=0,\ldots,c_j-1$

  $\phi_{\mathbf{y}|u}^{(t)} = \prod_{j=1}^{r}\phi_{jy|u}^{(t)} = p(Y_1^{(t)} = y_1,\ldots,Y_r^{(t)} = y_r|U^{(t)} = u)$

- $(k-1)$ initial probabilities

  $\pi_u = p(U^{(1)} = u) \quad u=1,\ldots,k$

- $(T-1)k(k-1)$ transition probabilities

  $\pi_{u|v}^{(t)} = p(U^{(t)} = u|U^{(t-1)} = v) \quad t=2,\ldots,T;\ u,v=1,\ldots,k$

- $\#\text{par} = k\sum_{j=1}^{r}(c_j-1)+(k-1)+(T-1)k(k-1)$

Probability distributions

- $p(\boldsymbol{U} = \boldsymbol{u}) = \pi_u \prod_{t=2}^{T} \pi_{u|v}^{(t)} = \pi_u \cdot \pi_{u_2|u}^{(2)} \dots \pi_{u_T|u_{T-1}}^{(T)}$
- $p(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{U} = \boldsymbol{u}) = \prod_{t=1}^{T} \phi_{\boldsymbol{y}|u}^{(t)} = \phi_{\boldsymbol{y}|u}^{(1)} \cdot \phi_{\boldsymbol{y}|u}^{(2)} \dots \phi_{\boldsymbol{y}|u}^{(T)}$
- Manifest distribution of $\boldsymbol{Y}$

$$
\begin{aligned}
p(\boldsymbol{Y} = \boldsymbol{y}) &= \sum_{\boldsymbol{u}} p(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{U} = \boldsymbol{u}) = \sum_{\boldsymbol{u}} p(\boldsymbol{U} = \boldsymbol{u}) \cdot p(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{U} = \boldsymbol{u}) \\
&= \sum_{u} \pi_u \phi_{\boldsymbol{y}|u}^{(1)} \cdot \sum_{u_2} \pi_{u_2|u}^{(2)} \phi_{\boldsymbol{y}|u}^{(2)} \dots \sum_{u_T} \pi_{u_T|u_{T-1}}^{(T)} \phi_{\boldsymbol{y}|u}^{(T)} \\
&= \sum_{u} \sum_{u_2} \dots \sum_{u_T} \pi_u \prod_{t=2}^{T} \pi_{u|v}^{(t)} \prod_{t=1}^{T} \phi_{\boldsymbol{y}|u}^{(t)}
\end{aligned}
$$

Maximum likelihood (ML) estimation

- Log-likelihood of the model

$$\ell(\boldsymbol{\theta}) = \sum_{\boldsymbol{y}} n_{(\boldsymbol{y})} \log[p(\boldsymbol{Y} = \boldsymbol{y})]$$

  - $\boldsymbol{\theta}$: vector of all model parameters (π_u, $\pi^{(t)}_{u|v}$, $\phi^{(t)}_{jy|u}$)
  - $n_{(\boldsymbol{y})}$: frequency of the response configuration $\boldsymbol{y}$ in the sample

- $\ell(\boldsymbol{\theta})$ may be maximized with respect to $\boldsymbol{\theta}$ by an *Expectation-Maximization (EM) algorithm* (Dempster *et al.*, 1977)

EM algorithm

Complete data log-likelihood of the model

$$\ell^*(\boldsymbol{\theta}) = \sum_{j=1}^{r}\sum_{t=1}^{T}\sum_{u=1}^{k}\sum_{y=0}^{c-1} a_{juy}^{(t)} \log \phi_{jy|u}^{(t)} + \\ + \sum_{u=1}^{k} b_u^{(1)} \log \pi_u + \sum_{t=2}^{T}\sum_{v=1}^{k}\sum_{u=1}^{k} b_{vu}^{(t)} \log \pi_{u|v}^{(t)}$$

- $a_{juy}^{(t)}$: frequency of subjects responding by y for the j-th response variable and belonging to latent state u, at time t
- $b_u^{(1)}$: frequency of subjects in latent state u at time 1
- $b_{vu}^{(t)}$: frequency of subjects which move from latent state v to u at time t

EM algorithm

- The algorithm *alternates two steps* until convergence in $\ell(\boldsymbol{\theta})$:

  E: compute the expected values of frequencies $a^{(t)}_{juy}$, $b^{(1)}_{u}$, and $b^{(t)}_{vu}$, given the observed data and the current value of $\boldsymbol{\theta}$, so as to obtain the expected value of $\ell^*(\boldsymbol{\theta})$

  M: update $\boldsymbol{\theta}$ by maximizing the expected value of $\ell^*(\boldsymbol{\theta})$ obtained above; explicit solutions for $\boldsymbol{\theta}$ estimations are available

- The E-step is performed by means of certain recursions

Forward and backward recursions

To efficiently compute the probability $p(\boldsymbol{Y} = \boldsymbol{y})$ and, then, the posterior probabilities $f_{u|\boldsymbol{y}}^{(t)}$ and $f_{u|v,\boldsymbol{y}}^{(t)}$ we can use forward and backward recursions for obtaining the following intermediate quantities

- Forward recursions

$$q_{u,\boldsymbol{y}}^{(t)} = p(U^{(t)} = u, \boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(t)}) = \sum_{v=1}^{k} q_{v,\boldsymbol{y}}^{(t-1)} \pi_{u|v}^{(t)} \phi_{\boldsymbol{y}|u}^{(t)} \quad u = 1, \ldots, k$$

starting with $q_{u,\boldsymbol{y}}^{(1)} = \pi_u \phi_{\boldsymbol{y}|u}^{(1)}$

- Backward recursions

$$\bar{q}_{v,\boldsymbol{y}}^{(t)} = p(\boldsymbol{Y}^{(t+1)}, \ldots, \boldsymbol{Y}^{(T)} | U^{(t)} = v) = \sum_{u=1}^{k} \bar{q}_{u,\boldsymbol{y}}^{(t+1)} \pi_{u|v}^{(t+1)} \phi_{\boldsymbol{y}|u}^{(t+1)} \quad v = 1, \ldots, k$$

starting with $\bar{q}_{v,\boldsymbol{y}}^{(T)} = 1$

Variants of basic LM model

- LM model with covariates: An LM model may be generalized in a similar way to the basic LC model introducing individual covariates
- Constrained LM model: Several interesting constraints may be introduced in an LM model to reduce the number of parameters and to make easier the interpretation of results
    - constraints on the conditional distribution
    - constraints on the transition probabilities
- In what follows we describe some of the most interesting constraints on the transition probabilities

Constraints on $\mathbf{\Pi}^{(t)}$

- We denote by $\mathbf{\Pi}^{(t)} = \{\pi_{u|v}^{(t)}\}$ the matrix of transition probabilities
- Linear constraint: $\boldsymbol{\rho}_v^{(t)} = \boldsymbol{Z}_v^{(t)}\boldsymbol{\delta}$, with $\boldsymbol{\rho}_v^{(t)}$ denoting a column vector containing the off-diagonal elements of the v-th row of $\mathbf{\Pi}^{(t)}$
- More in general, a GLM may be imposed on the transition probabilities $\boldsymbol{\lambda}_v^{(t)} = \boldsymbol{Z}_v^{(t)}\boldsymbol{\delta}$, with $\boldsymbol{\lambda}_v^{(t)} = g(\boldsymbol{\pi}_v^{(t)})$; e.g., $g(\cdot)$ may be a logit link function, so that the generic element of $\boldsymbol{\lambda}_v^{(t)}$ is $\lambda_{u|v}^{(t)} = \log \frac{\pi_{u|v}^{(t)}}{\pi_{v|v}^{(t)}};\ u = 1, \ldots, k,\ u \neq v$

Examples of constraints on $\mathbf{\Pi}^{(t)}$

C1 Time-homogeneous Markov-chain: $\mathbf{\Pi}^{(t)} = \mathbf{\Pi}$

$\rightarrow$ Transition probability from state v to state u is independent of the occasion t

C2 All the off-diagonal transition probabilities are equal to each other

$$\mathbf{\Pi}^{(t)} = \begin{pmatrix} 1-2\pi^{(t)} & \pi^{(t)} & \pi^{(t)} \\ \pi^{(t)} & 1-2\pi^{(t)} & \pi^{(t)} \\ \pi^{(t)} & \pi^{(t)} & 1-2\pi^{(t)} \end{pmatrix}, \quad t = 2, \ldots, T$$

C3 Symmetric transition matrix: transition probability from state v to state u is the same as the reverse transition

$$\mathbf{\Pi}^{(t)} = \begin{pmatrix} 1-(\pi^{(t)}_{2|1}+\pi^{(t)}_{3|1}) & \pi^{(t)}_{2|1} & \pi^{(t)}_{3|1} \\ \pi^{(t)}_{2|1} & 1-(\pi^{(t)}_{2|1}+\pi^{(t)}_{3|2}) & \pi^{(t)}_{3|2} \\ \pi^{(t)}_{3|1} & \pi^{(t)}_{3|2} & 1-(\pi^{(t)}_{3|1}+\pi^{(t)}_{3|2}) \end{pmatrix},$$

C4 Upper-triangular transition matrix: a subject in state v may move only in state $u = v+1,\dots,k$

$$\mathbf{\Pi}^{(t)} = \begin{pmatrix} \pi^{(t)}_{1|1} & \pi^{(t)}_{2|1} & \pi^{(t)}_{3|1} \\ 0 & \pi^{(t)}_{2|2} & \pi^{(t)}_{3|2} \\ 0 & 0 & 1 \end{pmatrix}, \quad t = 2,\dots,T$$

C5 Tridiagonal transition matrix: transition from state v is only allowed to state $u = v - 1, v + 1$

$$\mathbf{\Pi}^{(t)} = \begin{pmatrix} \pi^{(t)}_{1|1} & \pi^{(t)}_{2|1} & 0 & 0 \\ \pi^{(t)}_{1|2} & \pi^{(t)}_{2|2} & \pi^{(t)}_{3|2} & 0 \\ 0 & \pi^{(t)}_{2|3} & \pi^{(t)}_{3|3} & \pi^{(t)}_{4|3} \\ 0 & 0 & \pi^{(t)}_{3|4} & \pi^{(t)}_{4|4} \end{pmatrix}, \quad t = 2, \ldots, T$$

C6 Basic LC model: transition from state v to state u equals 0, for all $v \neq u$

$$\mathbf{\Pi}^{(t)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad t = 2, \ldots, T$$

Model selection

- Given the number k of latent states, it is convenient to sequentially introduce the constraints on $\mathbf{\Pi}^{(t)}$ and retain the constraint that, at each attempt, leads to a reduction of the BIC index
- Also an LR test is possible, based on the test statistic

$$LR = -2(\hat{\ell}_0 - \hat{\ell}_1)$$

  - $\hat{\ell}_1$: maximum log-likelihood value of the unconstrained model
  - $\hat{\ell}_0$: maximum log-likelihood value of the constrained model
- What's about the distribution of LR statistic?
  - when the usual regularity conditions hold (e.g., case C1), LR statistic has a chi-square distribution with a number of degrees of freedom equal to the number of free parameters
  - in presence of constraints on the boundary space (e.g., cases C4 and C5), LR statistic has a chi-bar-squared distribution (i.e., a mixture of chi-squared distributions)

Analysis of marijuana consumption

- We again consider the dataset about marijuana consumption, assuming that individuals may move from one state to another one during the time
- We first estimate an unconstrained basic LM model, characterised by a completely general transition matrix $\mathbf{\Pi}^{(t)}$ (model LM1)
- Then, we estimate some more realistic constrained LM models, characterised by
  - a time-homogeneous transition matrix $\mathbf{\Pi}$ (constraint C1; model LM2)
  - a time-homogeneous and tridiagonal transition matrix $\mathbf{\Pi}$ (constraints C1 and C5; model LM3)
  - a time-homogeneous and upper-triangular transition matrix $\mathbf{\Pi}$ (constraints C1 and C4; model LM4)
- Note that LM2 is nested in LM1, LM3 is nested in LM2, and LM4 is nested in LM2; LM3 and LM4 are not nested

Unconstrained basic LM model (LM1)

Table : Estimates of conditional response probabilities, $\hat{\phi}_{y|u}$

	$y = 0$	$y = 1$	$y = 2$
$u = 1$	0.996	0.000	0.004
$u = 2$	0.305	0.687	0.008
$u = 3$	0.012	0.083	0.905

- Results are coherent with those obtained by the LC model (Lecture 1, Example 1): state 1 corresponds to the lowest tendency of marijuana consumption, state 2 to an intermediate tendency, and state 3 to the highest tendency
- Note that the results outline the presence of measurement errors, as some off-diagonal values differ from 0

Table : Estimates of initial, $\hat{\pi}_u$, and transition probabilities, $\hat{\pi}^{(t)}_{u|v}$

		$u = 1$	$u = 2$	$u = 3$
$t = 1$		0.898	0.083	0.019
$t = 2$	$v = 1$	0.831	0.154	0.015
	$v = 2$	0.318	0.228	0.454
	$v = 3$	0.057	0.000	0.943
$t = 3$	$v = 1$	0.810	0.190	0.000
	$v = 2$	0.056	0.482	0.461
	$v = 3$	0.000	0.147	0.853
$t = 4$	$v = 1$	0.908	0.064	0.028
	$v = 2$	0.059	0.718	0.224
	$v = 3$	0.000	0.186	0.814
$t = 5$	$v = 1$	0.789	0.163	0.048
	$v = 2$	0.099	0.821	0.080
	$v = 3$	0.020	0.035	0.945

- Individuals tend to be in state 1 (low tendency to marijuana consumption) at the beginning of the study ($t = 1$)
- However, the completely general transition matrices $\mathbf{\Pi}^{(t)}$ are not easy to be interpreted
- In order to have information about the time trend of the tendency of marijuana consumption, we may
  - calculating the marginal distribution of latent states for each time occasion
  - constraining the transition matrices in order to reduce the number of parameters

Table : Estimated marginal probabilities of latent states, $\hat{p(U_i^{(t)} = u)}$, $u = 1, \ldots, 5$

	$u = 1$	$u = 2$	$u = 3$
$t = 1$	0.8980	0.0835	0.0185
$t = 2$	0.7736	0.1576	0.0688
$t = 3$	0.6355	0.2331	0.1314
$t = 4$	0.5905	0.2325	0.1770
$t = 5$	0.4924	0.2933	0.2143

- We observe that the tendency to stay in state 1 (low marijuana consumption) decreases with the age
- The tendency to consume marijuana (states 2 and 3) increases with the age

Constrained basic LM model (LM1)

Table : Estimates of conditional response probabilities, $\hat{\phi}_{y|u}$

	$y=0$	$y=1$	$y=2$
$u=1$	0.996	0.000	0.004
$u=2$	0.305	0.687	0.008
$u=3$	0.012	0.083	0.905

- Results are coherent with those obtained by the LC model (Lecture 1, Example 1): state 1 corresponds to the lowest tendency of marijuana consumption, state 2 to an intermediate tendency, and state 3 to the highest tendency
- Note that the results outline the presence of measurement errors, as some off-diagonal values differ from 0

Time-homogeneous LM model (LM2)

Table : Estimates of initial, $\hat{\pi}_u$, and transition probabilities, $\hat{\pi}^{(t)}_{u|v}$

		$u=1$	$u=2$	$u=3$
$t=1$		0.912	0.071	0.017
$t=2,\dots,5$	$v=1$	0.842	0.141	0.017
	$v=2$	0.080	0.670	0.250
	$v=3$	0.000	0.132	0.868

- High persistency in each latent state, but also a given tendency to move to adjacent states

Time-homogeneous LM model with tridiagonal transition matrix (LM3)

Table : Estimates of initial, $\hat{\pi}_u$, and transition probabilities, $\hat{\pi}^{(t)}_{u|v}$

		$u = 1$	$u = 2$	$u = 3$
$t = 1$		0.896	0.089	0.015
$t = 2, \ldots, 5$	$v = 1$	0.835	0.165	0.000
	$v = 2$	0.070	0.686	0.244
	$v = 3$	0.000	0.082	0.918

- High persistency in each latent state, but also a given tendency to move to higher adjacent states

Model selection

Table : Model selection for $k = 3$: maximum log-likelihood value, number of parameters, and BIC index

Model	$\hat{\ell}$	# par	BIC
LC	-658.238	32	1491.454
LM1	-646.895	32	1468.768
LM2	-658.593	14	1393.738
LM3	-660.600	12	1375.890
LM4	-661.930	11	1373.070

R package LMest

- This package includes a set of functions to fit LM models in the basic version and in the extended version with individual covariates
- The main function for the model estimation is `est_lm_basic`

- Data structure

```
> data(data_drug)
> data_drug = as.matrix(data_drug)
> head(data_drug)

> S=data_drug[,1:5]-1 # matrix of item responses
> yv=data_drug[,6] # vector of weights

> k=3 # number of latent states
```

- Model estimation

```
> # Basic unconstrained LM model
> LM1 = est_lm_basic(S,yv,k,mod=0)
> # Time-homogeneous LM model
> LM2 = est_lm_basic(S,yv,k,mod=1)
```

- Output

```
> LM1$piv #  latent states initial probabilities
> LM1$Pi # transition probabilities
> LM1$Psi #  conditional response probabilities

> # Marginal probabilities
> marg_prob_2 = colSums(LM1$Pi[,,2]*LM1$piv)
> marg_prob_3 = colSums(LM1$Pi[,,3]*marg_prob_2)
> marg_prob_4 = colSums(LM1$Pi[,,4]*marg_prob_3)
> marg_prob_5 = colSums(LM1$Pi[,,5]*marg_prob_4)
> round(rbind(LM1$piv, marg_prob_2, marg_prob_3,
marg_prob_4, marg_prob_5), 4)
```

Main references

- Bartolucci F., Farcomeni A., Pennoni F. (2012), Latent Markov Models for Longitudinal Data, Chapman & Hall/CRC
- Wiggins, L. (1973), Panel Analysis: Latent probability models for attitude and behaviours processes, Elsevier, Amsterdam
- Zucchini, W., MacDonald, I. L. (2009) Hidden Markov Models for Time Series: an Introduction using R. New York: Springer