

Likelihood inference for latent Markov models under linear hypotheses on the transition probabilities

Inferenza per modelli latent Markov sotto vincoli lineari sulle probabilità di transizione

Francesco Bartolucci¹

Dipartimento di Economia, Finanza e Statistica, Università di Perugia
e-mail: bart@stat.unipg.it

Riassunto: Si illustra un approccio per la formulazione e la verifica di ipotesi lineari sulle probabilità di transizione di un modello latent Markov (LM) per dati discreti longitudinali. Per la stima di massima verosimiglianza sotto ipotesi di questo tipo si fa uso di un algoritmo EM basato su recursioni note nella letteratura sui modelli hidden Markov. Si deriva inoltre la distribuzione asintotica del rapporto di verosimiglianza tra due modelli LM specificati tramite diverse ipotesi lineari sulle probabilità di transizione. Come caso particolare si considera quello in cui l'ipotesi di interesse è l'assenza di transizione tra gli stati latenti, ipotesi sotto la quale un modello LM si riduce a un modello a classi latenti.

Keywords: boundary problem, chi-bar-squared distribution, latent class model.

1. Introduction

The latent Markov (LM) model was introduced by Wiggins (1973) for the analysis of longitudinal discrete data and has been successfully applied in several fields, such as psychological and educational measurement, criminology and the analysis of customer behavior (for a review see Langeheine and van de Pol, 2002). This model may be seen as an extension of both a Markov chain model and a latent class (LC) model (Lazarsfeld and Henry, 1968). It is in fact based on the assumption that any occasion-specific response variable depends only on a discrete latent variable, which in turn depends on the latent variables corresponding to the previous occasions according to a first-order Markov chain.

In this paper we summarize, and partially extend, the main results of Bartolucci (2005) about a class of LM models in which: (i) the conditional distribution of the response variables given the latent process is formulated as in a generalized linear model; (ii) the latent process is time-homogeneous with transition probabilities formulated as in a linear probability model, so that it is possible to constrain certain of these probabilities to be equal to 0. Bartolucci (2005) dealt in particular with maximum likelihood (ML) estimation, based on an EM algorithm (Dempster *et al.*, 1977), and the asymptotic distribution of the likelihood ratio (LR) statistic between two LM models specified through two nested linear hypotheses on the transition probabilities. One of the most interesting hypotheses of this type is that of absence of transition between latent states, hypothesis under which a LM model specializes into a LC model. Note that under hypotheses of this type the parameters may be on the boundary of the parameter space and therefore an inferential problem under non-standard conditions (Self and Liang, 1987, Silvapulle and Sen, 2004) arises. It is shown, however, that the LR statistic has an asymptotic chi-bar-squared distribution under the null hypothesis.

The remainder of the paper is organized as follows. The class of LM models is illustrated in more detail in Section 2, while their ML estimation is illustrated in Section 3. LR testing of linear hypotheses on the transition probabilities is dealt with in Section 4. For reasons of space, the results of a simulation study on the finite sample accuracy of the proposed testing approach are not reported here. For these results we refer to Bartolucci (2005), where an empirical illustration based on the analysis two datasets, about educational assessment and behavior of young people, is also available.

2. A class of homogeneous latent Markov models

Let $\mathbf{Y} = \{Y_t, t = 1, \dots, s\}$ denote a sequence of s response variables with support $\{1, \dots, d\}$. The basic assumption of the LM model is that these variables are conditionally independent given a latent process $\mathbf{X} = \{X_t, t = 1, \dots, s\}$ which follows a first-order Markov chain with state space $\{1, \dots, c\}$. We also assume that the latent process is time-homogeneous, so that the transition probabilities $\pi_{x|w}$, $w, x = 1, \dots, c$, do not depend on t , whereas the initial probabilities λ_x , $x = 1, \dots, c$, are completely unconstrained.

As will be shown below, we also allow for the presence of constraints on the conditional distribution of the response variables given the latent process and on the transition probabilities of the latter. Regardless of the specific constraint, the distribution of $\mathbf{X}$ and the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$ may be expressed as $p(\mathbf{x}) = \lambda_{x_1} \prod_{t=2}^s \pi_{x_t|x_{t-1}}$ and $p(\mathbf{y}|\mathbf{x}) = \prod_t \phi_{t,y_t|x_t}$, respectively, with $\phi_{t,y|x} = p(Y_t = y|X_t = x)$. Consequently, the *manifest distribution* of $\mathbf{Y}$ has the following expression

$$p(\mathbf{y}) = \sum_{x_1} \phi_{1,y_1|x_1} \lambda_{x_1} \left(\sum_{x_2} \phi_{2,y_2|x_2} \pi_{x_2|x_1} \cdots \left(\sum_{x_s} \phi_{s,y_s|x_s} \pi_{x_s|x_{s-1}} \right) \right).$$

This distribution may be easily computed, even for very large values of s , through recursions known in the hidden Markov literature (MacDonald and Zucchini, 1997).

2.1. Modelling conditional distribution of response variables

Let ϕ be a column vector with elements $\phi_{t,y|x}$, $t = 1, \dots, s$, $x = 1, \dots, c$, $y = 1, \dots, d$, arranged in a suitable order and let $\boldsymbol{\eta} = \boldsymbol{\eta}(\phi)$ be a link function that maps ϕ onto $\mathbb{R}^{sc(d-1)}$. We assume that $\boldsymbol{\eta} = \mathbf{Z}\boldsymbol{\gamma}$, where $\mathbf{Z}$ is a full rank design matrix and $\boldsymbol{\gamma}$ is a vector of parameters. An interesting case with binary response variables is when we parameterize the conditional distribution of any Y_t given $X_t = x$ through a logit that, by assumption, has an additive form in a parameter indexed by t and another indexed by x . In this way we can formulate a LM version of the Rasch (1961) model.

In the present approach, we also allow for inequality constraints of the type $\mathbf{K}\boldsymbol{\gamma} \geq \mathbf{0}_k$ on the parameters $\boldsymbol{\gamma}$, where $\mathbf{K}$ is a full rank matrix with k rows and $\mathbf{0}_j$ is a vector of j zeros. The main use of these constraints is for making the latent states uniquely identifiable by ordering them in a suitable way.

2.2. Modelling transition probabilities

Let $\boldsymbol{\rho}$ denote the column vector with elements $\pi_{x|w}$, $x, w = 1, \dots, c$, $x \neq w$. We assume that $\boldsymbol{\rho} = \mathbf{W}\boldsymbol{\beta}$, where $\mathbf{W}$ is a full rank matrix of size $c(c-1) \times b$ with at most one positive element in any row and all the other elements equal to 0. Note that we are formulating a

linear model directly on the transition probabilities, so that we can deal with hypotheses under which one or more of these probabilities are equal to 0. However, in order to ensure that all the transition probabilities are non-negative, we have to impose the constraints $\beta \geq \mathbf{0}_b$ and $TW\beta \leq \mathbf{1}_c$, with $T = \mathbf{I}_c \otimes \mathbf{1}'_{c-1}$. Consequently, we are in a non-standard inferential problem in order to derive the asymptotic distribution of the LR statistic for testing hypotheses on β .

Let Π denote the transition matrix of the latent process. An interesting hypothesis that may be formulated as above is that all the off-diagonal elements of Π are equal to each other. A less stringent hypothesis is that Π is symmetric. Finally, when the latent states are ordered in a suitable way, we can formulate the hypothesis that Π is upper triangular, so that a subject in latent state w may move only to latent state $x = w + 1, \dots, c$.

3. Maximum likelihood estimation

Let n_y be the frequency of the response configuration y in a sample of n subjects. By assuming that these subjects are independent of each other, the log-likelihood of any model in the class of LM models outlined in the previous section may be expressed as $\ell(\theta) = \sum_y n_y \log[p(y)]$, with θ denoting the vector of non-redundant parameters, i.e. $\theta = (\alpha', \beta', \gamma')'$, with $\alpha = \{\log(\lambda_{x+1}/\lambda_1), x = 1, \dots, c-1\}$.

In order to estimate θ , we can maximize $\ell(\theta)$ by means of an EM algorithm (Dempster *et al.*, 1977) which is based on the concept of complete data log-likelihood, i.e. the log-likelihood that we could compute if we knew the latent configuration x of any subject in the sample. In our case, it may be expressed as

$$\ell^\dagger(\theta) = \sum_x f_x \log(\lambda_x) + \sum_w \sum_x g_{w,x} \log(\pi_{x|w}) + \sum_t \sum_x \sum_y h_{t,x,y} \log(\phi_{t,y|x}),$$

where f_x is the number of subjects which at the first occasion are in the latent state x , $g_{w,x}$ is the number of transitions from latent state w to latent state x and $h_{t,x,y}$ is the number of subjects which at occasion t are in latent state x and provide response y .

Note that the components of $\ell^\dagger(\theta)$ may be maximized separately through simple iterative algorithms. However, since the frequencies f_x , $g_{w,x}$ and $h_{t,x,y}$ are unknown, the EM algorithm maximizes $\ell^\dagger(\theta)$ (M step), once these frequencies have been substituted with the corresponding conditional expected values given the observed data and the current value of the parameters (E step). This process is iterated until convergence in $\ell(\theta)$. The E step, in particular, is again based on certain recursions known in the hidden-Markov literature (see MacDonald and Zucchini, 1997, Sec. 2.2).

4. Testing linear hypotheses on the transition probabilities

Consider a hypothesis of the type $H_0 : M\beta = \mathbf{0}_m$, with M denoting a full rank matrix of dimension $m \times b$, on the latent process parameters. This hypothesis may be tested through the LR statistic $D = -2[\ell(\hat{\theta}_0) - \ell(\hat{\theta})]$, where $\hat{\theta}_0$ is the constrained ML estimate of θ under H_0 and $\hat{\theta}$ is the unconstrained estimate. Both estimates may be computed through the EM algorithm above.

Let $\theta_0 = (\alpha'_0, \beta'_0, \gamma'_0)'$ be the true value of θ under H_0 . Since certain elements of β_0 may be equal to 0, we are in an inferential problem under non-standard conditions.

Situations of this type have been studied by many authors (see Self and Liang, 1987, and Silvapulle and Sen, 2004, and the references therein). In particular, from Silvapulle and Sen (2004, Proposition 4.8.2) we have the following Theorem, where $\mathbf{H}(\boldsymbol{\theta})$ is the average information matrix of the assumed LM model, $\mathbf{G}$ is the block of g rows of $\mathbf{I}_b$ such that $\mathbf{G}\boldsymbol{\beta}$ contains the elements of $\boldsymbol{\beta}$ which are constrained to 0 under H_0 and $\mathbf{J}$ is the block of the remaining rows of $\mathbf{I}_b$.

Theorem 1 *Provided that $\mathbf{H}(\boldsymbol{\theta}_0)$ is of full rank and that H_0 holds with $\mathbf{J}\boldsymbol{\beta}_0 > \mathbf{0}_{b-g}$, $\mathbf{TW}\boldsymbol{\beta}_0 < \mathbf{1}_c$ and $\mathbf{K}\boldsymbol{\gamma}_0 > \mathbf{0}_k$, the asymptotic distribution of the LR statistic D is $\chi^2_{m-g} + \bar{\chi}^2(\boldsymbol{\Sigma}_0, \mathcal{O}^g)$, where $\bar{\chi}^2(\boldsymbol{\Sigma}_0, \mathcal{O}^g)$ denotes the chi-bar-squared distribution with parameters $\boldsymbol{\Sigma}_0 = \mathbf{G}\mathbf{H}(\boldsymbol{\theta}_0)^{-1}\mathbf{G}'$ and $\mathcal{O}^g$, the non-negative orthant of dimension g .*

Note that the distribution of $\chi^2_{m-g} + \bar{\chi}^2(\boldsymbol{\Sigma}_0, \mathcal{O}^g)$ may be expressed as a mixture of conventional chi-squared distributions with weights that may be computed through a simple Monte Carlo simulation procedure (e.g. Silvapulle and Sen 2004, Sec. 3.5). On the basis of these weights we can easily compute an asymptotic p -value for D , once the unknown parameters have been substituted with their consistent estimate. This procedure is much faster than a bootstrap procedure.

As a particular case of Theorem 1, we have that of no elements of $\boldsymbol{\beta}$ constrained to 0 under H_0 . In this case, the LR statistic has χ^2_m asymptotic null distribution. Another interesting case is when we test a LC model against its LM version based on a transition matrix with all the off-diagonal elements equal to each other. The asymptotic null distribution is in this case $0.5\chi^2_0 + 0.5\chi^2_1$, and therefore the weights of the mixture of chi-squared distributions mentioned above are explicitly given.

References

- Bartolucci F. (2005) Likelihood inference for a class of latent Markov models under linear hypotheses on the transition probabilities, *Journal of the Royal Statistical Society, series B*, in press (available at <http://www.stat.unipg.it/~bart>).
- Dempster A. P., Laird N. M., Rubin D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B*, 39, 1–22.
- Langeheine R., van de Pol F. (2002) Latent Markov chains, in: *Applied latent Class Analysis*, Cambridge University Press, 304–344.
- Lazarsfeld P. F., Henry N. W. (1968) *Latent Structure Analysis*, Houghton Mifflin, Boston.
- MacDonald I. L., Zucchini W. (1997) *Hidden Markov and other Models for Discrete-Valued Time Series*, Chapman and Hall, New York.
- Rasch G. (1961) On general laws and the meaning of measurement in psychology, *Proceedings of the IV Berkeley Symposium on Mathematical Statistics and Probability*, 4, 321–333.
- Self S. G., Liang K.-Y. (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *Journal of the American Statistical Association*, 82, 605–610.
- Silvapulle M. J., Sen P. K. (2004) *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*, Wiley, New York.
- Wiggins L. M. (1973) *Panel Analysis*, Elsevier, Amsterdam.