Bayesian multivariate latent Markov models with an unknown number of regimes

Alessio Farcomeni and Silvia Pandolfi

Abstract We develop latent Markov models (LM) with non informative drop out in the Bayesian framework. We describe Metropolis algorithms for efficiently sampling parameters, and a reversible jump to infer the unknown number of latent regimes. We extend the basic model to models with constraints on the manifest and latent distribution. We illustrate the approach with an application to a real dataset on Marjuana consumption.

Key words: Bayesian statistics, latent Markov model, longitudinal data, reversible jump

1 Introduction

Latent Markov (LM) models are commonly used for longitudinal categorical data to describe the evolution of a latent characteristic of a group of individuals. They can also be used to model overdispersion, approximating time-varying unobserved effects without normality assumptions on them (Bartolucci and Farcomeni, 2009). A general overview of LM models can be found in Bartolucci *et al.* (2010).

LM models can be seen as a generalization of hidden Markov models (HMM) for time series (MacDonald and Zucchini, 1997), and many inferential tools are derived as a direct extension. Bayesian inference has so far been considered only for HMM.

Besides the usual advantages, there are at least three reasons why the Bayesian framework we propose in this paper could be more suitable in many situations: (i) when the sample size is small, the likelihood is often multimodal and has near

Alessio Farcomeni
Sapienza - University of Rome e-mail: alessio.farcomeni@uniroma1.it

Silvia Pandolfi
University of Perugia e-mail: pandolfi@stat.unipg.it

flat regions, and the prior acts as a regularizer (ii) the iterative algorithms used to maximize the likelihood may be slower to convergence than MCMC, (iii) reversible jump allows formal assessment of the number of regimes and model averaging is straightforward if desired. In this paper we also remove the common balanced panel assumption that all subjects are observed at exactly the same number of occasions, describing the model and the algorithms in the more general case of non informative drop out. This simple extension allows to apply LM models in many new contexts.

We outline a simple sampling strategy to approximate the posterior with an unknown number of regimes. Even this simple sampling strategy will prove efficient when the number of outcomes is small. In a forthcoming paper we will outline more complex models, also with the presence of covariates, and a more suitable sampling strategy for situations with a large number of parameters.

In next section we set up the basic LM model, in Section 3 we outline the proposed sampling strategy to approximate the posterior. We discuss extensions of the basic LM model in Section 4 and conclude with a real data example in Section 5.

2 Preliminaries

We introduce the preliminary concepts about the basic LM model for categorical data, based on a first-order Markov chain, time homogeneous transition probabilities and no covariates.

In the univariate case, $\mathbf{Y}_i = (Y_i^{(1)}, \dots, Y_i^{(T_i)})$, $i = 1, \dots, n$, denotes a sequence of T_i categorical response variables with l levels or categories, coded from 0 to $l-1$, independently observed over n subjects. In the multivariate case, $\mathbf{Y}_i = (\mathbf{Y}_i^{(1)}, \dots, \mathbf{Y}_i^{(T_i)})$, $i = 1, \dots, n$, and $\mathbf{Y}_i^{(t)}$ is a vector of r outcomes, each with l_h categories, $h = 1, \dots, r$. Let $T = \max_i T_i$, and assume drop-out is non informative.

The main assumption behind LM models is that of *local independence*, i.e. for every subject the response variables are conditionally independent given a latent process $\mathbf{U}_i = (U_i^{(1)}, \dots, U_i^{(T_i)})$. This latent process is assumed to follow a first-order Markov chain with state space $\{1, \dots, k\}$. Then, for all $t > 2$, the latent variable $U_i^{(t)}$ is conditionally independent of $U_i^{(1)}, \dots, U_i^{(t-2)}$ given $U_i^{(t-1)}$.

Parameters of the univariate model are the conditional response probabilities $\phi_{y|u}^{(t)} = p(Y_i^{(t)} = y | U_i^{(t)} = u)$, $t = 1, \dots, T_i$, $u = 1, \dots, k$, $y = 0, \dots, l-1$, the initial probabilities $\pi_u = p(U_i^{(1)} = u)$, $u = 1, \dots, k$, and the transition probabilities $\pi_{v|u} = p(U_i^{(t)} = v | U_i^{(t-1)} = u)$, $t = 2, \dots, T_i$, $u, v = 1, \dots, k$. For the multivariate model, we let $\phi_{\mathbf{y}|u}^{(t)} = p(\mathbf{Y}_i^{(t)} = \mathbf{y} | U_i^{(t)} = u)$, where $\mathbf{y}$ evolves over all possible $\prod l_h$ combinations of levels of the r responses, arranged in lexicographical order. When r is large, the corresponding contingency table could be sparse. In that case, a constrained version based on marginal logits and log-odds ratios as the one outlined in Section 4 should be preferred. Note that the latent process is time homogeneous, so that the transition probabilities do not depend on t, moreover the initial probabilities are completely unconstrained. Furthermore all these probabilities do not depend on i since, in its basic version, the model does not account for individual covariates.

The assumptions above imply that the distribution of $\mathbf{U}_i$ may be expressed as $p(\mathbf{U}_i = \mathbf{u}) = \pi_{u^{(1)}} \prod_{t>1} \pi_{u^{(t)}|u^{(t-1)}}$ where $\mathbf{u} = (u^{(1)}, \dots, u^{(T_i)})$. Moreover, the conditional distribution of $\mathbf{Y}_i$ given $\mathbf{U}_i$ may be expressed as $p(\mathbf{Y}_i = \mathbf{y}|\mathbf{U}_i = \mathbf{u}) = \prod_t \phi_{y^{(t)}|u^{(t)}}$ and, consequently, for the *manifest distribution* of $\mathbf{Y}_i$ we have

$$f(\mathbf{y}) = p(\mathbf{Y}_i = \mathbf{y}) = \sum_{u^{(1)}} \phi^{(1)}_{y^{(1)}|u^{(1)}} \pi_{u^{(1)}} \sum_{u^{(2)}} \phi^{(2)}_{y^{(2)}|u^{(2)}} \pi_{u^{(2)}|u^{(1)}} \cdots \sum_{u^{(T_i)}} \phi^{(T_i)}_{y^{(T_i)}|u^{(T_i)}} \pi_{u^{(T_i)}|u^{(T_i-1)}}, \tag{1}$$

where $\mathbf{y} = (y^{(1)}, \dots, y^{(T_i)})$. In order to efficiently compute $f(\mathbf{y})$ we must use a forward recursion (Baum *et al.*, 1970), obtaining $q^{(t)}(u, \mathbf{y}) = p(U_i^{(t)} = u, Y_i^{(1)} = y^{(1)}, \dots, Y_i^{(t)} = y^{(t)})$ for $t = 1, \dots, T_i$. The recursion is as follows: let $q^{(1)}(u, \mathbf{y}) = \phi^{(1)}_{y|u} \pi_u$. If $T_i > 1$, for $t = 2, \dots, T_i$ let $q^{(t)}(v, \mathbf{y}) = \phi^{(t)}_{y|v} \sum_u q^{(t-1)}(u, \mathbf{y}) \pi_{v|u}$. Finally, $f(\mathbf{y}) = \sum_u q^{(T_i)}(u, \mathbf{y})$. The likelihood is obviously given by $\prod_i \sum_u q^{(T_i)}(u, \mathbf{y})$.

We can more efficiently use matrix operations and let $f(\mathbf{y}) = \mathbf{q}^{(T)}(\mathbf{y})'\mathbf{1}$ where $\mathbf{1}$ is a column vector of ones of suitable dimension and $\mathbf{q}^{(t)}(\mathbf{y})$ is a column vector with elements $q^{(t)}(u, \mathbf{y})$. The recursion is then expressed as:

$$\mathbf{q}^{(t)}(\mathbf{y}) = \begin{cases} \mathrm{diag}[\phi^{(1)}_{y^{(1)}}]\pi, & \text{if} \quad t = 1, \\ \mathrm{diag}[\phi^{(t)}_{y^{(t)}}]\Pi' \mathbf{q}^{(t-1)}(\mathbf{y}), & \text{otherwise.} \end{cases} \tag{2}$$

with $\pi = \{\pi_u, u = 1, \dots, k\}$ denoting the initial probability vector, $\phi^{(t)}_y = \{\phi^{(t)}_{y|u}, u = 1, \dots, k\}$ denoting the conditional probability vector and $\Pi = \{\pi_{v|u}, u, v = 1, \dots, k\}$ denoting the transition probability matrix.

The basic LM model is considered here in the Bayesian setting in order to estimate both the dimension and the unknown parameters of the model. At this aim, we choose to parameterize the transition probability matrix Π by $\Omega = \{\omega_{uv}, u, v = 1, \dots, k\}$, as

$$\pi_{v|u} = \frac{\omega_{uv}}{\sum_{h=1}^k \omega_{uh}},$$

where $\omega_{uv} > 0$ for any u, v in the state space. The ω_{uv} are not identified, but this transformation facilitates the MCMC moves, since it relaxes the constraints on the initial and transition probabilities (Cappé *et al.*, 2003). In particular this transformation makes the random walk Metropolis-Hastings (MH) moves more efficient and often allows to improve the mixing of the corresponding MCMC algorithms.

Following Cappé *et al.* (2005) we choose a Gamma prior $\omega_{uv} \sim Ga(\delta_v, 1)$ for any $u, v = 1, \dots, k$, which leads to a Dirichlet prior with parameters $\delta_1, \dots, \delta_k$ on the corresponding rows of the transition matrix. We consider the same reparametrization also for the initial probabilities, using the vector $\lambda = \{\lambda_u, u = 1, \dots, k\}$, and for the conditional probabilities, through the vectors $\psi^{(t)}_y = \{\psi^{(t)}_{yu}, u = 1, \dots, k\}$. We also assume a Gamma prior distribution for $\lambda_u \sim Ga(\delta'_u, 1)$ and for $\psi^{(t)}_{yu} \sim Ga(\delta''_u, 1)$.

Finally, for the parameter k we define a discrete uniform prior distribution between 1 and $k_{\max}$, where $k_{\max}$ is the maximum number of regimes we admit a priori. Usually $k_{\max}$ is greater than the most complex model that could be visited by the algorithm.

As in Stephens (2000) and in Cappé *et al.* (2003) we do not use *completation* to run the algorithm, i.e. the latent Markov chain $\mathbf{U}_i$ is not simulated. This can be avoided because of the forward recursive representation of the likelihood we have outlined.

3 MCMC Algorithm

In this section we outline the MCMC algorithm for parameter and model dimension estimation. Every iteration of this algorithm is based on two different types of move: one aimed at updating the unknown parameters of the current model conditionally on the number of regimes, and a transdimensional one that allows updating the number of states.

In the first type of move, the simulation from the posterior density is accomplished through a MH step. We update the parameters by random walk MH moves instead of implementing Gibbs steps because the resulting algorithm is easier to implement within the reversible jump.

In the dimension-changing move we implement a reversible jump (RJ) MCMC algorithm based on split-combine and birth-death moves, as in Richardson and Green (1997), but without requiring the moment matching in split and combine moves and without limiting birth and death moves to the empty components (see Cappé *et al.* (2005), Ch. 13, and Spezia (2010)).

We tackle label switching by post-processing the output of the algorithm: we use relabelling-invariant priors and run the MCMC unconstrained. Then, as in Marin *et al.* (2005), label switching is managed by selecting one of the modal regions and relabelling according to proximity to the selected modal region.

3.1 Metropolis-Hastings Algorithm

We sample parameters from the posterior, when k is held fixed, through three separate MH steps. At each step, the elements of Ω, λ and $\psi_y^{(t)}$ are generated by multiplicative random walk moves. We consider a logarithmic transformation of the positive quantities ω_{uv}, λ_u, and $\psi_{yu}^{(t)}$ and the proposed moves are as follows:

$$\begin{aligned}
\log \omega'_{uv} &= \log \omega_{uv} + \varepsilon_{uv}, \\
\log \lambda'_u &= \log \lambda_u + \varepsilon_u, \\
\log \psi_{yu}^{(t)\prime} &= \log \psi_{yu}^{(t)} + \varepsilon_{yu}^{(t)},
\end{aligned}$$

where $\varepsilon_{uv} \sim N(0, \tau_\omega)$, $\varepsilon_u \sim N(0, \tau_\lambda)$ and $\varepsilon_{yu}^{(t)} \sim N(0, \tau_\psi)$.

The proposed values are accepted with probabilities equal, respectively, to

$$\frac{p(\theta_k')L(\mathbf{y}|\theta_k')}{p(\theta_k)L(\mathbf{y}|\theta_k)} \prod_{u,v} \frac{\omega_{uv}'}{\omega_{uv}}, \quad \frac{p(\theta_k')L(\mathbf{y}|\theta_k')}{p(\theta_k)L(\mathbf{y}|\theta_k)} \prod_{u} \frac{\lambda_u'}{\lambda_u}, \quad \frac{p(\theta_k')L(\mathbf{y}|\theta_k')}{p(\theta_k)L(\mathbf{y}|\theta_k)} \prod_{y,u,t} \frac{\psi_{yu}^{(t)\prime}}{\psi_{yu}^{(t)}}, \tag{3}$$

where $\theta_k = (\Omega, \lambda, \psi_y^{(t)})$ is a short-hand notation for the current parameter vector, $p(\theta_k)$ is the prior distribution and $L(\mathbf{y}|\theta_k)$ is the likelihood computed via the forward algorithm. The last term in all equations is derived from the Jacobian of the log-scale transformation.

3.2 Reversible Jump Algorithm

When the number of regimes is unknown, transdimensional MCMC is performed alternating the Metropolis strategy of the previous section with a RJ algorithm. The RJ algorithm is made by two steps. First, if $k > 1$ and $k < k_{\max}$, we choose between a combine and a split move with probability 0.5. When $k = 1$, we always propose a split move; when $k = k_{\max}$, we propose a combine move. Then, at a second step we propose a birth or a death move along the same lines.

We now describe split-combine moves. Suppose that the current state of the chain is (k, θ_k). The split move randomly selects a state u_0 and splits it into two new ones, u_1 and u_2. The corresponding parameters are split as follows:

1. Split λ_{u_0} as
$$\lambda_{u_1} = \lambda_{u_0}\rho, \quad \lambda_{u_2} = \lambda_{u_0}(1-\rho) \quad \text{with } \rho \sim U(0,1).$$
2. Split column u_0 of Ω as
$$\omega_{u\,u_1} = \omega_{u\,u_0}\rho_u, \quad \omega_{u\,u_2} = \omega_{u\,u_0}(1-\rho_u) \quad \text{with } \rho_u \sim U(0,1) \text{ for } u \neq u_0.$$
3. Split row u_0 of Ω as
$$\omega_{u_1\,v} = \omega_{u_0\,v}\xi_v, \quad \omega_{u_2\,v} = \omega_{u_0\,v}/\xi_v \quad \text{with } \xi_v \sim Ga(a_\omega, b_\omega) \text{ for } v \neq u_0.$$
4. Split $\omega_{u_0\,u_0}$ as
$$\begin{aligned} \omega_{u_1\,u_1} &= \omega_{u_0\,u_0}\rho_{u_0}\xi_{u_1}, \quad \omega_{u_1\,u_2} = \omega_{u_0\,u_0}(1-\rho_{u_0})\xi_{u_2}, \\ \omega_{u_2\,u_1} &= \omega_{u_0\,u_0}\rho_{u_0}/\xi_{u_1}, \quad \omega_{u_2\,u_2} = \omega_{u_0\,u_0}(1-\rho_{u_0})/\xi_{u_2}, \end{aligned}$$
with $\rho_{u_0} \sim U(0,1)$ and $\xi_{u_1}, \xi_{u_2} \sim Ga(a_\xi, b_\xi)$.
5. Split $\psi_{yu_0}^{(t)}$ as
$$\psi_{yu_1}^{(t)} = \psi_{yu_0}^{(t)}\xi_y^{(t)}, \quad \psi_{yu_2}^{(t)} = \psi_{yu_0}^{(t)}/\xi_y^{(t)}$$
with $\xi_y^{(t)} \sim Ga(a_\psi, b_\psi)$.

It is worth noting that the densities of the proposals are identical, as ρ and $1-\rho$ have the same distribution and likewise ξ and $1/\xi$ (here subscripts on these variables are omitted) so the symmetry constraints are satisfied.

In the reverse combine move two distinct states, u_1 and u_2, are picked at random and combined into a single state u_0 as follows:

1. $\lambda_{u_0} = \lambda_{u_1} + \lambda_{u_2}$;
2. $\omega_{u\,u_0} = \omega_{u\,u_1} + \omega_{u\,u_2}$ for $u \neq u_0$;
3. $\omega_{u_0\,v} = (\omega_{u_1\,v}\,\omega_{u_2\,v})^{1/2}$ for $v \neq u_0$;
4. $\omega_{u_0\,u_0} = (\omega_{u_1\,u_1}\,\omega_{u_2\,u_1})^{1/2} + (\omega_{u_1\,u_2}\,\omega_{u_2\,u_2})^{1/2}$;
5. $\psi^{(t)}_{y u_0} = (\psi^{(t)}_{y u_1}\,\psi^{(t)}_{y u_2})^{1/2}$.

The split move is accepted with probability $\min\{1, A\}$ whereas the combine move is accepted with probability $\min\{1, A^{-1}\}$ where A can be computed as

$$\begin{aligned}
&\frac{L(\mathbf{y}|\theta_{k+1})p(\theta_{k+1})p(k+1)}{L(\mathbf{y}|\theta_k)p(\theta_k)p(k)} \times \frac{(k+1)!}{k!} \times \frac{P_c(k+1)/[(k+1)k/2]}{P_s(k)/k} \\
&\times \frac{|J_{split}|}{2\,p(\xi_{u_1})p(\xi_{u_2})\prod_{v\neq u_0} p(\xi_v)\prod_y \prod_t p(\xi_y^{(t)})} \\
&= \frac{L(\mathbf{y}|\theta_{k+1})p(\theta_{k+1})}{L(\mathbf{y}|\theta_k)p(\theta_k)} \times \frac{P_c(k+1)}{P_s(k)} \times \frac{|J_{split}|}{p(\xi_{u_1})p(\xi_{u_2})\prod_{v\neq u_0} p(\xi_v)\prod_y \prod_t p(\xi_y^{(t)})},
\end{aligned} \tag{4}$$

where $P_s(k)/k$ and $P_c(k+1)/[(k+1)k/2]$ are respectively the probabilities to split a specific component out of k available ones and to combine one of $(k+1)k/2$ possible pairs of components. The factorials and the coefficient 2 arise from combinatorial reasoning related to label switching. $|J_{split}|$ is the Jacobian of the transformation from θ_k to θ_{k+1}, which is the product of five determinants $|J_1| = \lambda_{u_0}$, $|J_2| = \prod_{u\neq u_0} \omega_{u\,u_0}$, $|J_3| = 2^{k-1}\prod_{v\neq u_0} \omega_{u_0\,v}/\xi_v$, $|J_4| = 4\omega^3_{u_0\,u_0}\rho_{u_0}(1-\rho_{u_0})/\xi_{u_1}\xi_{u_2}$, $|J_5| = \prod_y \prod_t 2\psi^{(t)}_{y u_0}/\xi^{(t)}_y$. Birth and death moves are derived along similar lines. A birth is accomplished by generating a new regime, denoted by u_0. The new parameters are drawn from their respective priors; the position of the new regime is generated at random. The remaining parameters are simply copied to the proposed new state θ_{k+1}. In the death move a regime u_0 is selected at random and then deleted along with the corresponding parameters. The acceptance probability of the birth move is $\min\{1, A\}$, where A is given by

$$\begin{aligned}
&\frac{L(\mathbf{y}|\theta_{k+1})p(\theta_{k+1})p(k+1)}{L(\mathbf{y}|\theta_k)p(\theta_k)p(k)} \times \frac{(k+1)!}{k!} \times \frac{P_d(k+1)/(k+1)}{P_b(k)/(k+1)} \\
&\times \frac{|J_{birth}|}{\prod_{u\neq u_0} p(\omega_{u\,u_0})\prod_{v\neq u_0} p(\omega_{u_0\,u})p(\omega_{u_0\,u_0})}.
\end{aligned} \tag{5}$$

The death move is accepted with probability $\min\{1, A^{-1}\}$. Since the proposal densities are equal to the prior of the corresponding parameters, and because the components in θ_k remain the same in θ_{k+1}, many terms cancel out in the expression above. Note that $|J_{birth}| = 1$.

4 More General Versions of the LM model

In the basic LM model outlined in the previous sections, all the probabilities are completely unconstrained. There are two generalizations which may be of interest and commonly arise in applications. First, we may have observed covariates together with the outcomes. Secondly, we may desire to put restrictions on the parameter space. Constraints can be adopted in order to give a more parsimonious and easily interpretable LM model. Both generalizations may concern either the distribution of the response variables $\phi_{y|u}^{(t)}$ (i.e., the *measurement model*), the distribution of the latent process summarized in π_u and $\pi_{v|u}$ (i.e., the *latent model*), or even both. For a more detailed description see Bartolucci *et al.* (2010). We outline here for reasons of space only the constrained models for the measurement model. In order to fit the constrained models we use a very similar MCMC strategy (see Section 5). In the univariate case a sensible constraint may be

$$\phi_{y|u}^{(t)} = \phi_{y|u}, \quad t = 1, \dots, T,\ u = 1, \dots, k,\ y = 0, \dots, l-1. \tag{6}$$

This constraint corresponds to the hypothesis that the distribution of the response variables only depends on the corresponding latent variable and there is no dependence of this distribution on time.

Other constraints may be expressed by

$$\eta_u^{(t)} = \mathbf{Z}_u^{(t)} \beta \tag{7}$$

where $\eta_u^{(t)} = g(\phi_{0|u}^{(t)}, \dots, \phi_{l-1|u}^{(t)})$, with $g(\cdot)$ being a suitable link function, $\mathbf{Z}_u^{(t)}$ being a design matrix and β being a vector of parameters. A Gaussian prior is used for β. With binary response variables, we can parameterize the conditional probabilities through the logit link function $\eta_u^{(t)} = \log(\phi_{1|u}^{(t)}/\phi_{0|u}^{(t)})$. With response variables having more than two categories, a natural choice is that of multinomial logit link function, so that $\eta_u^{(t)} = \log(\phi_{y|u}^{(t)}/\phi_{0|u}^{(t)})$, $y = 1, \dots, l-1$. However when the response variables have an ordinal nature, global logits are more suitable, i.e. $\eta_u^{(t)} = \log(\phi_{y|u}^{(t)} + \dots + \phi_{l-1|u}^{(t)})/(\phi_{0|u}^{(t)} + \dots + \phi_{y-1|u}^{(t)})$, $y = 1, \dots, l-1$.

Example 1 (LM Rasch model). In the case of binary variables assuming that

$$\eta_u^{(t)} = \log(\phi_{1|u}^{(t)}/\phi_{0|u}^{(t)}) = \zeta_u - \gamma^{(t)}, \quad t = 1, \dots, T,\ u = 1, \dots, k,$$

we can formulate a LM version of the Rasch model (Rasch, 1961). This model find its natural application in psychological and educational assessment, for data derived from the responses that a group of n subjects provide to a set of T test items. In this settings, the parameter ζ_u may be interpreted as the ability level of the subjects in latent state u and $\gamma^{(t)}$ as the difficulty level of item t (see Bartolucci *et al.* (2010) for more details).

In the multivariate case, we use a multivariate link function of the kind

$$\eta_u^{(t)} = C \log[M\phi_u^{(t)}] = \mathbf{Z}_u^{(t)}\beta, \tag{8}$$

where C and M are matrices of simple construction. In particular, the vector $\eta_u^{(t)}$ contains marginal logits and marginal log-odds ratios of different types (e.g. local, global, continuation). Refer to Bartolucci and Farcomeni (2009) for more details.

Example 2 (Marginal model for two response variables). Consider the case of $r = 2$ variables with two and three levels ($l_1 = 2$, $l_2 = 3$, $l_3 = 3$), which are treated with logits of type local and global, respectively. Overall, there are 3 logits and 2 log-odds ratios. The logits may be parametrized as

$$\log \frac{\phi_{1|u}^{(t)}}{\phi_{0|u}^{(t)}} = \mathbf{Z}_{u1}^{(t)}\beta_1, \qquad \log \frac{\phi_{1|u}^{(t)} + \phi_{2|u}^{(t)}}{\phi_{0|u}^{(t)}} = \mathbf{Z}_{u2}^{(t)}\beta_2, \qquad \log \frac{\phi_{2|u}^{(t)}}{\phi_{0|u}^{(t)} + \phi_{1|u}^{(t)}} = \mathbf{Z}_{u3}^{(t)}\beta_3,$$

whereas for the log-odds ratios we have

$$\log \frac{p(Y_{i1}^{(t)} = 1, Y_{i2} \geq y | U_i^{(t)} = u, x_i^{(t)}) p(Y_{i1}^{(t)} = 0, Y_{i2} < y | U_i^{(t)} = u, x_i^{(t)})}{p(Y_{i1}^{(t)} = 0, Y_{i2} \geq y | U_i^{(t)} = u, x_i^{(t)}) p(Y_{i1}^{(t)} = 1, Y_{i2} \geq y | U_i^{(t)} = u, x_i^{(t)})} = \beta_{2+y},$$

with $y = 1,2$. In this case, the parameter vector β is made of the subvectors β_1, β_2 and β_3, and the elements β_4 and β_5 and $\mathbf{Z}_u^{(t)}$ is obtained similarly.

5 Empirical illustration: Marijuana Consumption Data Set

To illustrate the Bayesian inference for basic LM model proposed in this paper, we describe the analysis of a dataset which concerns the use of marijuana among young people. The marijuana consumption data set has been taken from five annual waves (1976-1980) of the National Youth Survey (Elliott *et al.*, 1989) and is based on $n = 237$ respondents who were aged 13 years in 1976. The use of marijuana is measured through $T = 5$ ordinal variables, one for each annual wave, with $l = 3$ levels coded as 0 for never in the past year, 1 for no more than once a month in the past year and 2 for more than once a month in the past year. We want to explore whether there is an increase of marijuana use with age. As illustrated by Vermunt and Hagenaars (2004), a variety of models may be used for the analysis of this data set but a LM approach is desirable for its flexibility and easy interpretation (see Bartolucci (2006) for an application of the basic LM model, in the frequentist framework, to this dataset). In our implementation we used a uniform prior on k with $k_{\max} = 15$, and took $\delta_v = 1$, $\delta'_u = 1$ and $\delta''_u = 1$ for all u and v. The parameters were updated for fixed k through a $N(0,0.1)$, $N(0,0.5)$ and $N(0,0.2)$ increment random walk MH proposal on each $\log\omega_{u,v}$, $\log\lambda_u$ and $\log\psi^{(t)}_{yu}$ respectively. The sampler parameters were tuned so as to achieve acceptance rates in the range $0.1-0.25$ for all values $k \leq 15$. In the split move ve used $a_\omega = a_\xi = a_\psi = 1$ and $b_\omega = b_\xi = b_\psi = 1$ as parameters of the Gamma distributions. We also decided to simplify the model using constraint (6), i.e. assuming that the conditional probabilities are time homogeneous and thus $\psi^{(t)}_{yu} = \psi_{yu}$. Finally, the algorithm was run for $2,000,000$ iterations with a burn-in of $400,000$ iterations. Each iteration includes a MH move, in order to update the parameters of the current model with k fixed, a split-combine move and a birth-death move, that allow changes in the dimension of the model. The acceptance rates are showed in Table 1. The estimated posterior probabilities are

Table 1 Acceptance rate for the MH update, the split-combine move and the birth-death move under the basic LM model

	Performed	Accepted	Proportion accepted (%)
MH with fixed k	2,000,000		
Initial probabilities		425,674	21.284
Transition probabilities		234,137	11.707
Conditional probabilities		319,634	15.982
Birth	999,536	676	0.068
Death	1,000,464	659	0.066
Split	1,000,519	2,110	0.211
Combine	999,481	2,125	0.213

0.9624, 0.0372 and 0.0003 for $k = 3,4,5$ and below 0.0001 for smaller and larger values of k. Hence, the most probable model is that with three latent states, the same as selected by Bartolucci (2006) using BIC. In order to face the label switching problem, at the end of all the iterations, the MCMC output was sorted on the basis of the conditional probabilities of the last category of the response variable. Doing this, the last class of the LM model may be interpreted as that of subjects with high tendency to use marijuana. Once the output was post processed, we estimated the parameters of the model by the averages based on the last 1,600,000 iterations. The parameter estimates are shown in Table 2. We now put constraints on the distribution of the response variable. Following Bartolucci (2006) we assumed a parametrization for the conditional local logits of any response variable given the latent state: $\eta^{(t)}_u = \log\frac{\phi_{y|u}}{\phi_{y-1|u}} = \zeta_u + \gamma_y, \quad u = 1,\ldots,k,\ y = 1,\ldots,l-1$, where ζ_u may be interpreted as the tendency to use marijuana for a subject in the state u and γ_y, for $y = 1,\ldots,l-1$, are the cutpoints common to all the response variables. This parametrization requires the choice of the prior distributions on ζ_u and γ_y that we

Table 2 Estimated initial probabilities and transition probabilities under the basic LM model

u	$\hat{\pi}_u$	v	$\hat{\pi}_{v\|u}$		
			$u=1$	$u=2$	$u=3$
1	0.8535	1	0.8273	0.1336	0.0391
2	0.1218	2	0.1033	0.6550	0.2418
3	0.0247	3	0.0237	0.0897	0.8816

assumed to be $N(0,1)$ for all $u=1,\dots,k$ and $y=1,\dots,l-1$. In the MH step, the elements of both the parameters were updated through a normal random walk proposal $N(0,0.5)$; moreover in the split move the parameter ζ_{u_0} was split into $\zeta_{u_1}=\zeta_{u_0}-\varphi_u$ and $\zeta_{u_1}=\zeta_{u_0}+\varphi_u$ where $\varphi_u \sim N(0,0.2)$ with $u=1,\dots,k$, while in the reverse combine move the selected states were combined into $\zeta_{u_0}=(\zeta_{u_1}+\zeta_{u_2})/2$. The results under the final model are illustrated in Table 3, which shows the acceptance rates for the random walk MH move and for the dimension changing moves. As in the basic LM model, the results are based on 2,000,000 iterations of the MCMC algorithm after a burn-in of 400,000 sweeps; $k_{\max}$ is set equal to 15. The posterior distribution of the number of states k leads to choose again a model with three latent states, with a probability equal to 0.9699. The posterior probabilities of $k=4,5$ are respectively 0.0296 and 0.0004. The plot of the first 40,000 values of k after the burn-in are shown in Figure 1.

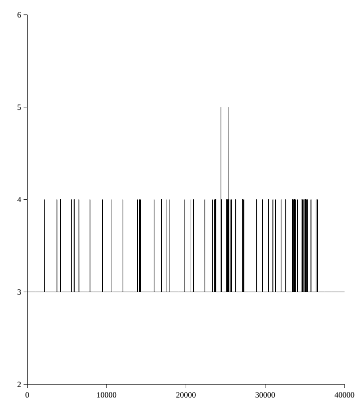

Fig. 1 The number of latent states in the first 40,000 iterations after the burn-in for the final LM model

From Tables 1 and 3 we note that the acceptance rates for the RJ moves are a bit lower than desired. A higher rate could be achieved by modifying the proposal distributions; we note anyway that the posterior distribution of k is highly concentrated on $k=3$, and this obviously affects the acceptance rate.

Tables 4 and 5 show the parameter estimates, computed after post-processing of the MCMC output. Even in this case, the latent states may be ordered, representing subjects with “no tendency to use marijuana”, “incidental users of marijuana”

Table 3 Acceptance rate for the MH update, the split-combine move and the birth-death move under the final LM model

	Performed	Accepted	Proportion accepted (%)
MH with fixed k	2,000,000		
Initial probabilities		468,355	23.418
Transition probabilities		303,666	15.183
ζ_u		202,448	10.122
γ_y		261,086	13.054
Birth	1,000,258	939	0.094
Death	999,742	942	0.094
Split	1,000,241	2,710	0.271
Combine	999,759	2,715	0.272

and “high tendency to use marijuana”. From the estimated transition matrices we

Table 4 Estimates of the parameters ζ_u and γ_y for the final LM model

u	$\hat{\zeta}_u$	y	$\hat{\gamma}_y$
1	-4.2054	1	0.4830
2	-0.0588	2	-1.6057
3	3.1381		

Table 5 Estimated initial probabilities and transition probabilities for the final LM model

u	$\hat{\pi}_u$	v	$\hat{\pi}_{v\|u}$		
			$u=1$	$u=2$	$u=3$
1	0.9079	1	0.8404	0.1442	0.0154
2	0.0612	2	0.0488	0.7113	0.2400
3	0.0309	3	0.0328	0.0443	0.9228

can see that although many subjects remain in the same latent class, moves towards higher frequencies are more likely than moves towards lower frequencies. For instance, around 24% of incidental users moves to the class of high frequency users, and only 4.9% towards the no tendency class: use of marjuana has increased over time.

Acknowledgements The first author was supported by EIEF research grant “Advances in non-linear panel models with socio-economic applications”.

References

F. Bartolucci (2006). Likelihood inference for a class of latent markov models under linear hypotheses on the transition probabilities. *Journal of the Royal Statistical Society, series B*, **68**, 155–178.

F. Bartolucci and A. Farcomeni (2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association*, **104**, 816–831.

F. Bartolucci, A. Farcomeni, and F. Pennoni (2010). An overview of latent Markov models for longitudinal categorical data. *arXiv:1003.2804*.

L. E. Baum, T. Petrie, G. Soules, and N. Weiss (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics*, **41**, 164–171.

O. Cappé, E. Moulines, and T. Rydén (2005). *Inference in Hidden Markov Models*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

O. Cappé, C.P. Robert, and T. Rydén (2003). Reversible jump, birth-and-death and more general continuous time markov chain monte carlo samplers. *Journal Of The Royal Statistical Society Series B*, **65**, 679–700.

D. S. Elliott, D. Huizinga, and S. W. Menard (1989). *Multiple Problem Youth: Delinquency, Substance Use, and Mental Health Problems*. Springer-Verlag New York, Inc.

I. L. MacDonald and W. Zucchini (1997). *Hidden Markov and other Models for Discrete-Valued Time Series*. Chapman and Hall, London.

J. M. Marin, K. L. Mengersen, and C. P. Robert (2005). Bayesian modelling and inference on mixture of distributions. In: D. Dey and C.R. Rao, eds., *Handbooks of Statistics*, vol. 25, 459–507. Elsevier Science, Amsterdam.

G. Rasch (1961). On general laws and the meaning of measurement in psychology. Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability 4, 321-333.

S. Richardson and P.J. Green (1997). On bayesian analysis of mixture with an unknown number of components. *Journal fo the Royal Statistical Society, Series B*, **59**, 731–792.

L. Spezia (2010). Bayesian analysis of multivariate Gaussian hidden Markov models with an unknown number of regimes. *Journal of Time Series Analysis*, **31**, 1–11.

M. Stephens (2000). Bayesian analysis of mixture models with an unknown number of components - an alternative to reversible jump methods. *The Annals of Statistics*, **28**, 40–74.

J.K. Vermunt and J.A. Hagenaars (2004). Ordinal longitudinal data analysis. in *Methods in Human Growth Research* (eds R.C. Hauspie, N. Cameron and L. Molinari), pp. 374-393. Cambridge: Cambridge University Press.