Estimating Markov and semi-Markov switching linear mixed models with individual-wise random effects

Florence Chaubert-Pereira, Yann Guédon, Christian Lavergne, Catherine Trottier

Estimating Markov and Semi-Markov Switching Linear Mixed Models with Individual-Wise Random Effects

Florence Chaubert-Pereira[1], Yann Guédon[1], Christian Lavergne[2], and Catherine Trottier[2]

[1] CIRAD, UMR DAP & INRIA, Virtual Plants,
Avenue Agropolis, TA A-96/02, 34398 Montpellier, France,
chaubert@cirad.fr, guedon@cirad.fr

[2] UM3, UMR I3M,
Place Eugène Bataillon, 34095 Montpellier Cedex 5, France,
Christian.Lavergne@math.univ-montp2.fr, trottier@math.univ-montp2.fr

Abstract. We address the estimation of Markov (and semi-Markov) switching linear mixed models i.e. models that combine linear mixed models with individual-wise random effects in a (semi-)Markovian manner. A MCEM-like algorithm whose iterations decompose into three steps (sampling of state sequences given random effects, prediction of random effects given the state sequence and maximization) is proposed. This statistical modeling approach is illustrated by the analysis of successive annual shoots along Corsican pine trunks.

Keywords: Markov switching model, linear mixed model, MCEM algorithm, plant growth.

1 Introduction

Lindgren (1978) introduced Markov switching linear models, i.e hidden Markov models (Cappé et al. (2005)) with linear models as output process; see Frühwirth-Schnatter (2006) for an overview of Markov switching models. In the literature, hidden Markov models with random effects in the output process have been used in a limited way. Chaubert et al. (2007) applied to forest tree growth data Markov switching linear mixed models (MS-LMM), i.e models that combine linear mixed models in a Markovian manner. These models broaden the class of Markov switching linear models by incorporating individual-wise random effects in the output process. Altman (2007) introduced Markov switching generalized linear mixed models (MS-GLMM) where the output process is supposed to belong to the exponential family, and applied these models to brain lesion counts observed on multiple sclerosis patients. Since covariates and individual-wise random effects are incorporated in the output process, the generalization of MS-LMM to hidden semi-Markov model (Guédon (2007)) is straightforward. The resulting models are called semi-Markov switching linear mixed models (SMS-LMM).

The remainder of this paper is organized as follows. MS-LMM are formally defined in Section 2. A Monte Carlo EM-like (MCEM) algorithm (McLachlan and Krishnan (2008)) whose iterations decompose into three steps (sampling of state sequences given random effects, prediction of random effects given state sequence and maximization) is presented in Section 3. This statistical modeling approach is illustrated in Section 4 by the analysis of successive annual shoots along Corsican pine trunks using SMS-LMM. Section 5 consists of concluding remarks.

2 Markov switching linear mixed models (MS-LMM)

Let $\{S_t\}$ be a Markov chain with finite-state space $\{1, \ldots, J\}$. This J-state Markov chain is defined by the following parameters:

- initial probabilities $\pi_j = P(S_1 = j),\ j = 1, \ldots, J$, with $\sum_j \pi_j = 1$,
- transition probabilities $p_{ij} = P(S_t = j | S_{t-1} = i),\ i, j = 1, \ldots, J$, with $\sum_j p_{ij} = 1$.

Let Y_{at} be the observation and S_{at} the non-observable state for individual a, $a = 1, \ldots, N$, at time t, $t = 1, \ldots, T_a$. Let $\sum_{a=1}^{N} T_a = T$. We denote by $Y_{a1}^{T_a}$ the T_a-dimensional vector of observations for individual a, and by Y_1^T the T-dimensional vector of all the observations; i.e. the concatenation of $Y_{a1}^{T_a}$; $a = 1, \ldots, N$. The vectors of non-observable states, $S_{a1}^{T_a}$ and S_1^T, are defined analogously.

A **Markov switching linear mixed model** can be viewed as a pair of stochastic processes $\{S_{at}, Y_{at}\}$ where the output process $\{Y_{at}\}$ is related to the state process $\{S_{at}\}$, which is a finite-state Markov chain, by the following linear mixed model:

$$Y_{at}|_{S_{at}=s_{at}} = X_{at}\beta_{s_{at}} + \tau_{s_{at}}\xi_{as_{at}} + \epsilon_{at}, \tag{1}$$

$$\xi_{as_{at}} \sim \mathcal{N}(0, 1), \qquad \epsilon_{at}|_{S_{at}=s_{at}} \sim \mathcal{N}(0, \sigma^2_{s_{at}}),$$

where X_{at} is the Q-dimensional row vector of covariates. Given the state $S_{at} = s_{at}$, $\beta_{s_{at}}$ is the Q-dimensional fixed effect parameter vector, $\xi_{as_{at}}$ is the individual a effect, $\tau_{s_{at}}$ is the standard deviation for the random effect and $\sigma^2_{s_{at}}$ is the residual variance. For convenience, random effects are supposed to follow the standard Gaussian distribution. Including random effects in the output process relaxes the assumption that the observations are conditionally independent given the non-observable states. The observations are here assumed to be conditionally independent given the non-observable states and the random effects.

3 Maximum likelihood estimation

Altman (2007) proposed to estimate the MS-GLMM parameters by maximizing directly the observed-data likelihood. Her approach based on Gaussian quadrature and quasi-Newton methods is strongly sensitive to starting values and to the number of quadrature points. Since both the states of the underlying Markov chain and the random effects are non observable, the EM algorithm (McLachlan and Krishnan (2008)) is a natural candidate to estimate MS-LMM. Let us consider the complete-data log-likelihood where both the outputs y_1^T, the random effects $\xi_1^J = \{\xi_{a1}^J = (\xi_{aj})_{j=1,\ldots,J}; a = 1, \ldots, N\}$ and the states s_1^T of the underlying Markov chain are observed

$$\begin{aligned}\log f(y_1^T, s_1^T, \xi_1^J; \theta) &= \log f(s_1^T) + \log f(\xi_1^J) + \log f(y_1^T | s_1^T, \xi_1^J)\\ &= \sum_{a=1}^{N} \log \pi_{s_{a1}} + \sum_{a=1}^{N} \sum_{t=2}^{T_a} \log p_{s_{a,t-1},s_{a,t}} + \sum_{a=1}^{N} \sum_{j=1}^{J} \log \phi(\xi_{aj}; 0, 1)\\ &+ \sum_{a=1}^{N} \sum_{t=1}^{T_a} \log \phi(y_{at}; X_{at}\beta_{s_{at}} + \tau_{s_{at}}\xi_{as_{at}}, \sigma_{s_{at}}^2). \end{aligned} \qquad (2)$$

where $\theta = (\pi, P, \beta, \tau, \sigma^2)$ is the set of parameters to be estimated and $\phi(y; \mu, \sigma^2)$ is the density of the Gaussian distribution with mean μ and variance σ^2.

The EM algorithm for hidden Markov chains cannot be transposed because the observations are not conditionally independent given the non-observable states; see Section 2. The EM algorithm for finite mixture of linear mixed models (Celeux et al. (2005)) cannot be adapted because the distribution of $\xi_1^J | Y_1^T = y_1^T$ cannot be analytically derived. Altman (2007) proposed a MCEM algorithm to estimate MS-GLMM where the random effects are sampled by Monte Carlo methods like Gibbs sampling. In the M-step, numerical methods like quasi-Newton routines are necessary to obtain updates for the parameter estimates. Altman (2007) noted the prohibitive computation burden due to the Monte Carlo and quasi-Newton methods, the slowness to converge and the sensitivity to starting values. Since sampling both a state sequence and random effects $\{s_1^T, \xi_1^J\}$ from their conditional distribution $S_1^T, \xi_1^J | Y_1^T = y_1^T$ is rather complicated, we propose here a MCEM-like algorithm where the Monte Carlo E-step is decomposed into two conditional steps:

- Monte Carlo Conditional E-step : given the random effects, state sequences are sampled for each individual a using a "forward-backward" algorithm (Chib (1996)).
- Conditional E-step : given the state sequence, the random effects are predicted.

In the M-step, the quantities $\sum_a \sum_t \log \phi(y_{at}; X_{at}\beta_{s_{at}} + \tau_{s_{at}}\xi_{as_{at}}, \sigma_{s_{at}}^2)$, $\sum_a \log \pi_{s_{a1}}$ and $\sum_a \sum_{t=2}^{T_a} \log p_{s_{a,t-1},s_{a,t}}$ in Equation (2) can be maximized separately.

3.1 Forward-backward algorithm for sampling state sequences given the random effects

For each individual a, the state sequences are sampled from the conditional distribution $P(S_{a1}^{T_a} = s_{a1}^{T_a} | Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J)$.
For a Markov switching linear mixed model, since

$$P\big(S_{a1}^{T_a} = s_{a1}^{T_a} | Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big) = \Big\{ \prod_{t=1}^{T_a - 1} P\big(S_{at} = s_{at} | S_{a,t+1}^{T_a} = s_{a,t+1}^{T_a}, Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big) \Big\} \times P\big(S_{aT_a} = s_{aT_a} | Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big),$$

the following conditional distributions should be used for sampling state sequences:

- final state (initialization) $P\big(S_{aT_a} = s_{aT_a} | Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big)$,
- previous state $P\big(S_{at} = s_{at} | S_{a,t+1}^{T_a} = s_{a,t+1}^{T_a}, Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big)$.

The forward-backward algorithm for sampling state sequences given the random effects can be decomposed into two passes, a forward recursion which is similar to the forward recursion of the forward-backward algorithm for hidden Markov chains, and a backward pass for sampling state sequences.

Forward recursion

The forward recursion is initialized for $t = 1$ by:

$$F_{aj}(1) = P\big(S_{a1} = j | Y_{a1} = y_{a1}, \xi_{a1}^J\big) = \frac{\phi(y_{a1}; X_{a1}\beta_j + \tau_j \xi_{aj}, \sigma_j^2)\pi_j}{N_{a1}} = \frac{G_{aj}(1)}{N_{a1}},$$

where $N_{a1} = P(Y_{a1} = y_{a1} | \xi_{a1}^J) = \sum_{j=1}^J G_{aj}(1)$ is a normalizing factor.

For $t = 2, \ldots, T_a$, the forward recursion is given by:

$$F_{aj}(t) = P\big(S_{at} = j | Y_{a1}^t = y_{a1}^t, \xi_{a1}^J\big) = \frac{\phi(y_{at}; X_{at}\beta_j + \tau_j \xi_{aj}, \sigma_j^2) \sum_{i=1}^J p_{ij} F_{ai}(t-1)}{N_{at}} = \frac{G_{aj}(t)}{N_{at}}.$$

The normalizing factor $N_{at} = P\big(Y_{at} = y_{at} | Y_{a1}^{t-1} = y_{a1}^{t-1}, \xi_{a1}^J\big) = \sum_{j=1}^J G_{aj}(t)$ is obtained directly during the forward recursion. The forward recursion can be used to compute the observed-data log-likelihood given the random effects for the parameter θ, as $\log P(Y_1^T = y_1^T | \xi_1^J; \theta) = \sum_a \sum_t \log N_{at}$.

Backward pass

The backward pass can be seen as a stochastic backtracking procedure. The final state s_{aT_a} is drawn from the smoothed probabilities

$$\Big(P\big(S_{aT_a} = j | Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big) = F_{aj}(T_a); j = 1, \ldots, J\Big).$$

For $t = T_a - 1, \ldots, 1$, the state s_{at} is drawn from the conditional distribution

$$\Big(P\big(S_{at} = j | S_{a,t+1}^{T_a} = s_{a,t+1}^{T_a}, Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big) = \frac{p_{js_{a,t+1}} F_{aj}(t)}{\sum_{i=1}^J p_{is_{a,t+1}} F_{ai}(t)}; j = 1, \ldots, J\Big).$$

3.2 Random effect prediction given the state sequence

The predicted random effects ξ_{a1}^J for each individual a is:

$$\xi_{a1}^J = \mathrm{E}[\xi_{a1}^J|y_{a1}^{T_a}] = \mathrm{E}\Big[\mathrm{E}[\xi_{a1}^J|y_{a1}^{T_a}, s_{a1}^{T_a}]|y_{a1}^{T_a}\Big] \approx \frac{1}{M}\sum_{m=1}^{M}\mathrm{E}[\xi_{a1}^J|y_{a1}^{T_a}, s_{a1}^{T_a}(m)], \quad (3)$$

with,

$$\mathrm{E}[\xi_{a1}^J|y_{a1}^{T_a}, s_{a1}^{T_a}(m)] = \Omega U_a^{(m)\prime}\Big(U_a^{(m)}\Omega^2 U_a^{(m)\prime} + \mathrm{Diag}\{U_a^{(m)}\sigma^2\}\Big)^{-1}\Big(Y_{a1}^{T_a} - \sum_{j=1}^{J} I_{aj}(m)X_a\beta_j\Big),$$

where:

- $s_{a1}^{T_a}(m)$ is the mth state sequence sampled for individual a,
- $\Omega = \mathrm{Diag}\{\tau_j; j = 1, \ldots, J\}$ is the $J \times J$ random standard deviation matrix,
- $U_a^{(m)}$ is the $T_a \times J$ design matrix associated with state sequence $s_{a1}^{T_a}(m)$, composed of 1 and 0 with $\sum_j U_a^{(m)}(t,j) = 1$ and $\sum_t \sum_j U_a^{(m)}(t,j) = T_a$,
- $\mathrm{Diag}\{U_a^{(m)}\sigma^2\}$ is the $T_a \times T_a$ diagonal matrix with $\{u_{at}^{(m)}\sigma^2; t = 1, \ldots, T_a\}$ on its diagonal,
- $u_{at}^{(m)} = \Big(\mathrm{I}(s_{at}(m) = 1) \cdots \mathrm{I}(s_{at}(m) = J)\Big)$ is the tth row of the design matrix $U_a^{(m)}$, I() is the indicator function,
- $\sigma^2 = (\sigma_1^2 \cdots \sigma_J^2)'$ is the J-dimensional residual variance vector,
- $I_{aj}(m) = \mathrm{Diag}\{I(s_{at}(m) = j); t = 1, \ldots, T_a\}$ is a $T_a \times T_a$ diagonal matrix,
- X_a is the $T_a \times Q$ matrix of covariates.

3.3 Extension to hidden semi-Markov models

Semi-Markov chains generalize Markov chains with the distinctive property of explicitly modeling the sojourn time in each state. Let $\{S_t\}$ be a semi-Markov chain defined by the following parameters:

- initial probabilities $\pi_j = P(S_1 = j)$, with $\sum_j \pi_j = 1$,
- transition probabilities
  - nonabsorbing state i: for each $j \neq i$, $\tilde{p}_{ij} = P(S_t = j|S_t \neq i, S_{t-1} = i)$, with $\sum_{j \neq i} \tilde{p}_{ij} = 1$ and $\tilde{p}_{ii} = 0$,
  - absorbing state i: $p_{ii} = P(S_t = i|S_{t-1} = i) = 1$ and for each $j \neq i$, $p_{ij} = 0$.

An occupancy distribution is attached to each nonabsorbing states:

$$d_j(u) = P(S_{t+u+1} \neq j, S_{t+u-v} = j, v = 0, \ldots, u-2|S_{t+1} = j, S_t \neq j), u = 1, 2, \ldots$$

As for the MS-LMM, the output process $\{Y_{at}\}$ of the semi-Markov switching linear mixed model (SMS-LMM) for individual a is related to the underlying semi-Markov chain $\{S_{at}\}$ by the linear mixed model (1). Since covariates and individual-wise random effects are incorporated in the output process, the observations are assumed to be conditionally independent given the non-observable states and the random effects. The proposed MCEM-like algorithm can therefore be directly transposed to SMS-LMM. Given the random effects, the state sequences are sampled using the "forward-backward" algorithm proposed by Guédon (2007). Given a state sequence, the random effects are predicted as previously described. The underlying semi-Markov chain parameters and the linear mixed model parameters are obtained by maximizing the Monte Carlo approximation of the complete-data log-likelihood.

4 Application to forest trees

The use of SMS-LMM is illustrated by the analysis of forest tree growth. The data set comprised four sub-samples of Corsican pines: 31 6-year-old trees, 29 12-year-old trees, 30 18-year-old trees and 13 23-year-old trees. Tree trunks were described by annual shoot from the base to the top where the length (in cm) was recorded for each annual shoot. The annual shoot is defined as the segment of stem established within a year. The observed growth is mainly the result of the modulation of the endogenous growth component by climatic factors. The endogenous growth component is assumed to be structured as a succession of roughly stationary phases separated by marked change points (Guédon et al.(2007)). The length of successive annual shoots along tree trunks was previously analyzed using a hidden semi-Markov chain (Guédon et al.(2007)) and a MS-LMM (Chaubert et al. (2007)). In the first case, the influence of climatic factors and the inter-individual heterogeneity were not explicitly modeled while in the second case, the length of the successive growth phases was not explicitly modeled.

A "left-right" three-state SMS-LMM composed of two successive transient states followed by a final absorbing state was estimated. We chose to use an intercept and the centered cumulated rainfall during a period recovering one organogenesis period and one elongation period as fixed effects for each linear mixed model. The linear mixed model attached to the growth phase j is:

$$Y_{at}|_{S_{at}=j} = \beta_{j1}+\beta_{j2}X_t+\tau_j\xi_{aj}+\epsilon_{at}, \quad \xi_{aj} \sim \mathcal{N}(0,1), \quad \epsilon_{at}|_{S_{at}=j} \sim \mathcal{N}(0,\sigma_j^2),$$

where Y_{at} is the length of the annual shoot for individual a at time t, β_{j1} is the intercept, X_t is the centered cumulated rainfall at time t ($\mathrm{E}(X_t) = 0$) and β_{j2} is the cumulated rainfall parameter. As the cumulated rainfall is centered, the intercept represents the average length of successive annual shoots in each growth phase. The estimation algorithm was initialized with the parameter values π, P, β and σ^2 estimated without taking into account random effects (hence, $\xi_1^J = 0$). The algorithm converged in 62 iterations with $m = 100$ state sequences sampled for each tree at each iteration. The convergence of the algorithm was monitored using the log-likelihood of the observed sequences given the random effects, which is directly obtained as a byproduct of the forward recursion; see Section 3.1.

The marginal distribution of the linear mixed model attached to growth phase j is $\mathcal{N}(\mu_j, \Gamma_j^2)$ with $\mu_j = \beta_{j1} + \beta_{j2}\mathrm{E}_j(X)$ and $\Gamma_j^2 = \tau_j^2 + \sigma_j^2$ where $\mathrm{E}_j(X)$ is the mean of the cumulated rainfalls X in growth phase j. The marginal distributions of the linear mixed models attached to each growth phase are well separated (few overlapping between marginal distributions corresponding to two successive states); compare the mean difference $\mu_{j+1} - \mu_j$ between consecutive states and the standard deviations Γ_j and Γ_{j+1} in Table 1. The standard deviation of the cumulated rainfall effect was computed as $\beta_{j2} \times sd(X)$ for each state j where $sd(X)$ is the standard deviation of the cumulated rainfalls X. The standard deviation of the cumulated rainfall effect represents the average amplitude of the climatic fluctuations in each growth phase. The influence of the cumulated rainfall is weak in the first growth phase (of slowest growth) while it is strong in the last two growth phases (a little less in the second phase than in the third phase); see Table 1.

	State		
	1	2	3
Intercept β_{j1}	7.19	26.08	50.48
Cumulated rainfall parameter β_{j2}	0.0042	0.0171	0.0304
Cumulated rainfall effect standard deviation	0.56	2.23	3.97
Random variance τ_j^2	6.81	52.34	72.83
Residual variance σ_j^2	5.13	39.75	76.54
Part of inter-individual heterogeneity	57.04%	56.84%	48.76%
Marginal distribution (μ_j, Γ_j)	$7.05, 3.46$	$26.17, 9.60$	$50.55, 12.22$

Table 1. Intercepts, regression parameters, centered cumulated rainfall effect, variability decomposition and marginal distributions of the estimated SMS-LMM.

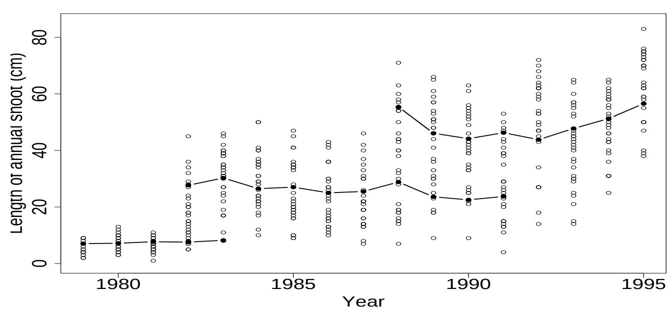

Fig. 1. 18-year-old Corsican pines: observed annual shoot lengths (points) and fixed part of the three observation linear mixed models (point lines).

The part of inter-individual heterogeneity, defined by the ratio between the random variance τ_j^2 and the total variance Γ_j^2 , is greater at the beginning of the plant life (first two growth phases with more than 56%) and decreases slightly in the last growth phase (near 49%). The most probable state sequence given the predicted random effects was computed for each observed sequence using a Viterbi-like algorithm. The fixed part of the three observation linear mixed models (i.e. $\beta_{j1} + \beta_{j2} X_t$ for each growth phase j) for 18-year-old trees is represented in Figure 1. The growth phases are well separated with few overlapping.

5 Concluding remarks

SMS-LMM enables to separate and to characterize the different growth components (endogenous, environmental and individual components) of forest trees. The behavior of each tree within the population can be investigated on the basis of the random effects predicted for each growth phase.
An interesting direction for further research would be to develop the statistical methodology for semi-Markov switching generalized linear mixed models. Since the hidden semi-Markov chain likelihood cannot be written as a simple product of matrices, the MCEM algorithm proposed by Altman (2007) for the MS-GLMM cannot be directly extended to the semi-Markovian case. In our MCEM-like algorithm proposed for MS-LMM and SMS-LMM, the difficulty lies mainly in the prediction of the random effects.

References

ALTMAN, R.M. (2007) : Mixed hidden Markov models : An extension of the hidden Markov model to the longitudinal data setting. *J. Amer. Stat. Assoc., 102, 201-210.*

CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005): *Inference in hidden Markov models.* Springer Series in Statistics. New York, NY: Springer. xvii, 652 p.

CELEUX, G., MARTIN, O. and LAVERGNE, C. (2005): Mixture of linear mixed models for clustering gene expression profiles from repeated microarray experiments. *Stat. Model., 5, 243-267.*

CHAUBERT, F., CARAGLIO, Y., LAVERGNE, C., TROTTIER, C. and GUÉDON, Y. (2007): A statistical model for analyzing jointly growth phases, the influence of environmental factors and inter-individual heterogeneity. Applications to forest trees. *Proceedings of the 5th International Workshop on Functional-Structural Plant Models, P43, 1-3.*

CHIB, S. (1996): Calculating posterior distributions and modal estimates in Markov mixture models. *J. Econometrics, 75, 79-97.*

FRÜHWIRTH-SCHNATTER, S. (2006): *Finite Mixture and Markov Switching Models.* Springer Series in Statistics. New York, NY: Springer. xix, 492 p.

GUÉDON, Y. (2007): Exploring the state sequence space for hidden Markov and semi-Markov chains. *Comput. Stat. and Data Anal., 51 (5), 2379-2409.*

GUÉDON, Y., CARAGLIO, Y., HEURET, P., LEBARBIER, E. and MEREDIEU, C. (2007): Analyzing growth components in trees. *J. of Theor. Biology, 248 (3), 418-447.*

LINDGREN, G. (1978): Markov regime models for mixed distributions and switching regressions. *Scand. J. Stat., Theory Appl., 5, 81-91.*

McLACHLAN, G.J. and KRISHNAN, T. (2008): *The EM algorithm and extensions. 2nd Edition.* Wiley Series in Probability and Statistics. New York, NY: John Wiley and Sons. xviii, 360 p.