

A semiparametric approach to mixed non-homogeneous hidden Markov models

Antonello Maruotti and Roberto Rocci

Abstract Longitudinal data, where the same set of subjects is repeatedly observed over time, are frequently analyzed by using Hidden Markov models (HMMs). In this situation, heterogeneity may arise at individual and/or time level affecting the hidden process, i.e. the transition probabilities among hidden states. Non-homogeneous (NH) HMMs face with this problem by expressing the transition probabilities as a function of a set of observed covariates, time and/or individual dependent, able to explain the non-homogeneity of the hidden processes. In this paper we introduce a finite mixture of NH-HMMs where each component is a NH-HMM differing only in the hidden process. The aim is twofold: classify subjects according to their behavior dynamic; estimate by non parametric maximum likelihood a mixed NH-HMM, where a random term is inserted accounting for omitted covariates at the individual level. The effectiveness of the proposal is tested through a simulation study.

Key words: Hidden Markov Models, Random effects models, NPML, Longitudinal data.

1 Introduction

Many time series data analyses focus on the evolution of a latent individual characteristic which is indirectly measured by one or more response variables. A suitable model for this type of analysis is the hidden Markov model (HMM) proposed by Baum and Petrie (1966), where the evolution of a latent characteristic of interest is represented by an unobservable Markov chain with a few number of states and

Antonello Maruotti
Dept. DIPES, University of Roma Tre. e-mail: antonello.maruotti@uniroma3.it

Roberto Rocci
Dept. SEFeMeQ, University of Tor Vergata. e-mail: roberto.rocchi@uniroma2.it

the response variables are assumed to be conditionally independent given the latent process. Under a longitudinal scheme, where several individuals are observed at periodic times, homogeneous HMMs (H-HMMs) are often used to describe disease or behavior progression (see e.g. Cooper and Lipsitch, 2004; Altman and Petkau, 2005). However, the assumption of homogeneity in the Markov model, i.e. that the transition probabilities are constant over time and among individuals, may represent a limitation on the use of these models. To overcome this problem, the non-homogeneous HMM (NH-HMM) has been introduced. It allows the transition matrix of the hidden chain to be time and/or individual varying on the basis of a set of exogenous covariates, time and/or individual specific (Hughes et al., 1999). However, the available observed covariates are not always able to capture all the individual heterogeneity sources. In this case, a random effect is included in the model specification to account for such unobserved individuals differences, leading to the so-called class of mixed HMMs (Altman, 2007; Shirley et al., 2009). The unobserved individual heterogeneity can also be modeled by using a latent class approach. Individuals are grouped into latent classes in an unsupervised manner on the basis of their dynamic behavior. The resulting model is a finite mixture of H-HMMs (Dias et al., 2010; Alfò and Maruotti, 2010), i.e. a mixed HMM where the unobserved heterogeneity sources, due to e.g. omitted covariates, are represented by means of time-constant discrete latent variables. The afore mentioned modeling framework, in theory, allows us to analyze data in a wide range of situations. However, such range could be in practice greatly reduced by computational complexities arising in the estimation step. The estimation of a HMM is carried out through several different methods such as maximum likelihood (Baum et al, 1970), Monte Carlo (Rydén, 2008) or via subspace estimation methods (Andersson and Rydén, 2009), whose application to HMMs is at the beginning. All methods pose challenging problems. In the context of mixed HMMs an interesting and useful method is the non-parametric maximum likelihood (NPML, Laird, 1978), where the distribution of the random effects is estimated by a discrete multivariate random variable with a finite number of support points. The resulting NPML estimator leads to a finite mixture of HMMs as in the case of a latent class HMM.

In this paper, a finite mixture of NH-HMM is proposed by extending the work of Dias et al. (2010) to the non-homogeneous case. The unobserved heterogeneity affects only the hidden process, while the distribution of the individual, time-constant, random term is assumed to be discrete. In this way, we specify a flexible model to describe a wide variety of behaviors providing both an unsupervised classification of the individuals and a semi-parametric estimator of a mixed HMM. The model parameters are estimated via maximum likelihood by using an EM-based algorithm. This can be seen as a possible solution to computational issues arising with mixed NH-HMMs. The proposal is illustrated by a large-scale simulation study in order to investigate the empirical behavior of the proposed approach with respect to several factors, such as the number of observed units and times, the distributional form of the random term (focusing on its variability) and the inclusion or not of an exogenous covariate in the logistic regression for the transition probabilities. The plan of the paper is as follows. In Section 2 we specify the proposed model, while in sec-

tion 3 some comments and remarks on the maximum likelihood estimation of model parameters are provided. A comparison of the performance of several model specifications under different data generation schemes is given in Section 4 by means of a simulation study. Some remarks are pointed out along with drawbacks that may arise by adopting the proposed methodology in Section 5.

2 Model specification

In statistical literature, HMMs form a wide and computationally tractable class of stochastic process models, where the dynamics of the stochastic process are governed by a latent Markov chain. In a basic HMM for longitudinal data, two processes are present: an unobservable finite-state first-order Markov chain, S_{it} , $i = 1, \dots, n$, $t = 0, \dots, T$, with state space $S = \{1, \dots, m\}$, and an observable process, Y_{it} , where Y_{it} denotes the response variable for individual i at time t (similarly for S_{it}). The distribution of Y_{it} depends only on S_{it} specifically, the Y_{it} are conditionally independent given the S_{it} . Typically, it is assumed that the state-dependent distributions, i.e. the distributions of Y_{it} given $S_{it} = j$, $j \in S$, come from a parametric family of continuous or discrete distributions. Thus, the unknown parameters in an HMM involve both the Markov chain and the state-dependent distributions of the random variables Y_{it} . In particular, the Markov chain parameters are the transition probabilities q_{itjk} , i.e. the probability that individual i visits state k at time t given that at time $t - 1$ he/she was in state j , and the initial probabilities δ_{ij} , i.e. the probability of being in state j at time 0.

The most simple model in this framework is the homogeneous HMM (H-HMM), which assumes common transition and initial probabilities. By assuming a logit type specification, we can write

$$q_{itjk} = \frac{\exp(c_{jk})}{1 + \sum_{h:h \neq j}^m \exp(c_{jh})}, \quad \forall j, k \in S, \quad (1)$$

with $c_{jk} = 0$ when $j = k$, and

$$\delta_{ij} = \delta_j.$$

Under the assumption of homogeneity of the Markov model, the transition probabilities are constant over time and among individuals. Since, in a longitudinal setting, multiple subjects are present at each period and individual covariates are often available, it may be desirable to estimate a different transition matrix for each individual. The previous specification of the transition probability matrix lacks in taking into account individual and time heterogeneity and, thus, the transition probabilities of the Markov chains can be parameterized as functions of p exogenous covariates $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})'$ as

$$q_{itjk} = \frac{\exp(\mathbf{x}_{it}' \boldsymbol{\gamma}_{jk} + c_{jk})}{1 + \sum_{h:h \neq j}^m \exp(\mathbf{x}_{it}' \boldsymbol{\gamma}_{jh} + c_{jh})} \quad \forall j, k \in S, \quad (2)$$

where $\gamma_{jk} = (\gamma_{jk1}, \dots, \gamma_{jkp})$ is a vector of fixed regression parameters. Of course, $\gamma_{jk} = c_{jk} = 0$ if $j = k$. This model specification, also known as non-homogeneous HMM (NH-HMM), is usually adopted when the interest is on the evolution of a hidden individual characteristic and the aim is to understand how the evolution of this characteristic is affected by the covariates. Unobserved individual heterogeneity, due to, for example, omitted covariates, could make the previous model inadequate. This can be solved by including a set of subject-specific random effects with zero mean, c_{ijk} , able to capture unobserved individual heterogeneity. The transition probabilities become

$$q_{ijk} = \frac{\exp(\mathbf{x}'_{it}\gamma_{jk} + c_{ijk})}{1 + \sum_{h:h \neq j}^m \exp(\mathbf{x}'_{it}\gamma_{jh} + c_{ijh})} \quad \forall j, k \in S, \quad (3)$$

while

$$\delta_{ij} = \frac{\exp(c_{ij0})}{1 + \sum_{h=1}^{m-1} \exp(c_{ih0})} \quad \forall j \in S. \quad (4)$$

Even in (3) and (4) we impose the identification constraints $\gamma_{jk} = c_{ijk} = 0$ if $j = k$ for $k = 0, 1, \dots, m$.

In our model the distribution of the random effects is assumed to be discrete $\mathbf{c}_i = (c_{ijk}), (j = 1, \dots, m; k = 0, 1, \dots, m; j \neq k)$ being a multivariate discrete random variable that takes values on a finite set $\{\mathbf{c}_1, \dots, \mathbf{c}_g, \dots, \mathbf{c}_G\} \subset \mathcal{R}^{m^2-m}$ with probabilities $\pi_1, \dots, \pi_g, \dots, \pi_G$. This model specification is the most general and can be very useful. It is a finite mixture of NH-HMM allowing the Markov chain parameters to vary across time and individuals. Further, it contains H-HMM, latent class and NH-HMM as special cases.

3 Maximum likelihood estimation of a mixed NH-HMM

The likelihood of models (3)-(4), for n independent units (individuals, subjects, etc.), can be expressed as

$$\begin{aligned} L(\cdot) &= \prod_{i=1}^n \sum_{g=1}^G \pi_g \sum_{s_{i0} \in S} \cdots \sum_{s_{iT} \in S} \{f(y_{i0} | s_{i0}) Pr(s_{i0} | \mathbf{c}_g) \\ &\quad \times \prod_{t=1}^T f(y_{it} | s_{it}) Pr(s_{it} | s_{it-1}, \mathbf{c}_g)\} \end{aligned} \quad (5)$$

Here, the initial and transition probabilities, $Pr(s_{i0} | \mathbf{c}_g)$ and $Pr(s_{it} | s_{it-1}, \mathbf{c}_g)$, denote the Markov chain parameters for the g -th component of the finite mixture; $f(y_{it} | s_{it})$ is the state-dependent distribution for the observed process. The basic idea of this approach is to assume that among the n hidden chains (one for each unit in the longitudinal study) G hidden groups are present, whereby all the units within each group are characterized by a non-homogeneous Markov chain with both fixed and

group specific parameters. An important feature of this approach is to assume that group membership of a certain unit is unknown a priori and is estimated along with all other model parameters. In our approach, model parameters are estimated by using an EM-type algorithm (not shown here for sake of brevity).

In the continuous case the likelihood for n independent units could be written as

$$L(\cdot) = \prod_{i=1}^n \int_B \sum_{s_{i0} \in S} \cdots \sum_{s_{iT} \in S} \{f(y_{i0} | s_{i0}) Pr(s_{i0} | \mathbf{c}_i) \\ \times \prod_{t=1}^T f(y_{it} | s_{it}) Pr(s_{it} | s_{it-1}, \mathbf{c}_i)\} dH(\mathbf{c}_i) \quad (6)$$

where H is the distribution of $\mathbf{c}_i$ and B is the corresponding support.

It is clear that the likelihood in (6) is more complex than the one in (5). The integration over the random effects may not be available in closed form, and therefore evaluation of (6) would require numerical integrations slowing the computation down considerably. In the case of normal distributed random effects, the previous algorithm could be easily adapted by using Gaussian quadrature. The main difference would be that the mass points $\mathbf{c}_g$ s and their probabilities π_g s would not be estimated but fixed a priori. A disadvantage of this method is the possible inaccuracy of the quadrature even with a large number of mass points, and its computational complexity due to the multivariate nature of the random effects distribution. A more appealing approach relies on the non parametric maximum likelihood (NPML) theory (Laird, 1978), where the random effects distribution is left unspecified and estimated from the data. The NPML estimator is simply the finite mixture in (5), which clearly simplifies the computational burden and provides a flexible specification that may be preferred to parametric alternatives. Finite mixture models have several significant advantages over parametric models; first, the discrete nature of the estimator helps to classify subjects in homogeneous clusters. This is particularly appealing in health sciences, where components can be interpreted as groups with a similar behavior with respect to the event of interest. Second, the locations and the corresponding probabilities are completely free to vary over the corresponding support and, therefore, extreme departures from the basic (homogeneous) model can be accommodated.

4 Simulation study

A large-scale simulation study was carried out to investigate the performance of the proposed approach focusing on the potential loss in efficiency resulting from model misspecification. Several samples have been generated from HMM models with the observed process given by a nominal variable with three categories. The simulation study has been planned to cover longitudinal schemes with different underlying *null models* by considering two possible distributions for the random ef-

fect (two-point discrete and normal) and its inclusion or not in the linear predictor of the logistic regression for the transition probabilities. In this way the following three different underlying models are obtained: a finite mixture of NH-HMM (fm-NH-HMM); a Normal mixture of NH-HMM (nm-NH-HMM); a NH-HMM, where no random effects are included in model specification. For each null model, we simulated 250 samples considering two experimental factors: the number of analyzed subjects, $n = 100, 300$, and the number of repeated measurements for each subject, $T = 10, 30, 90$. The parameter values of four different models have been estimated on each sample: a finite mixture of NH-HMMs, which takes into account non-homogeneity in the hidden process and among units; a NH-HMM, ignoring the presence of the random effects; a finite mixture of H-HMMs (fm-H-HMM), ignoring the influence of the covariate; a simple H-HMM, ignoring the presence of a non-homogeneous model in the hidden process. To evaluate model performance in guessing the true model structure, we compared the models on the basis of their ability to accurately estimate the vector of state-dependent probabilities (for the observed process) and the parameters γ associated with covariates included in the model. The accuracy have been measured by using the mean and the median root squared error. Each run has been started from a randomly chosen set of parameters. The algorithms were stopped when the increase in the log-likelihood was less than 10^{-4} ; for each of the four models, whenever possible, we have used the same starting points. The experiments provide evidence to draw some conclusions. The mean root squared error (MRSE) and the median root squared error (MedRSE) of parameter estimates are shown in Tables 1-4. When the data are generated according to a two-component finite mixture of NH-HMMs, the *true* model performs better in terms of MRSE than the other (more parsimonious) models; even if a simple NH-HMM works satisfactorily in situations where the sample size is small in terms of n and T . In these cases, the use of a more complex model does not cause any significant improvement (see Table 1).

As described in Section 3, the finite mixture can be used as a semiparametric estimator when the random effects distribution is continuous. Although our aim is not to estimate the random effects distribution (i.e. our goal is to avoid possibly misleading inferences from inappropriate distribution assumptions), we would investigate finite mixture behavior in such a context, where the distribution for the random effects is continuous. We simulated data from a normal mixture of NH-HMM and fit a 2-component finite mixture of NH-HMMs, as well as the other simpler models (see Table 2).

When the generating model is a Normal mixture of NH-HMM, the gain obtained by the proposed model with respect to the simple NH-HMM is sensible. Results show a clear and consistent pattern in that disregarding the random term seems to lead to less efficient estimates both of the regression parameters and the probabilities of the observed process. To give further insights in the finite mixture behavior, we analyzed the sensitivity of the results by varying dispersion in the random effects distribution and simulating data from a normal mixture of NH-HMM decreasing the standard deviation from 2 to 1 (see Table 3).

Table 1 Simulation results: MRSE and MedRSE for the probabilities of the observed process (p) and for the fixed effect in the logistic regression for the transition probabilities (γ) - Data generated from a two-component fm-NH-HMM

n	T	Parameter	H-HMM		fm-H-HMM		NH-HMM		fm-NH-HMM	
			MRSE	MedRSE	MRSE	MedRSE	MSE	MedRSE	MRSE	MedRSE
100	10	p	0.115	0.102	0.125	0.105	0.069	0.050	0.066	0.047
		γ	n.a.	n.a.	n.a.	n.a.	1.458	1.113	1.615	1.302
300	10	p	0.089	0.077	0.089	0.069	0.043	0.028	0.040	0.026
		γ	n.a.	n.a.	n.a.	n.a.	1.047	0.729	1.236	0.879
100	30	p	0.098	0.086	0.090	0.073	0.045	0.028	0.036	0.025
		γ	n.a.	n.a.	n.a.	n.a.	1.072	0.724	1.070	0.691
300	30	p	0.082	0.059	0.055	0.040	0.028	0.018	0.020	0.013
		γ	n.a.	n.a.	n.a.	n.a.	0.938	0.535	0.963	0.432
100	90	p	0.092	0.078	0.056	0.037	0.027	0.017	0.019	0.012
		γ	n.a.	n.a.	n.a.	n.a.	0.888	0.523	0.890	0.355

Table 2 Simulation results: MRSE and MedRSE for the probabilities of the observed process (p) and for the fixed effect in the logistic regression for the transtion probabilities (γ) - Data generated from a nm-NH-HMM with std. dev. equal to 2

n	T	Parameter	H-HMM		fm-H-HMM		NH-HMM		fm-NH-HMM	
			MRSE	MedRSE	MRSE	MedRSE	MSE	MedRSE	MRSE	MedRSE
100	10	p	0.114	0.104	0.116	0.098	0.083	0.068	0.078	0.063
		γ	n.a.	n.a.	n.a.	n.a.	1.462	1.006	1.456	1.067
300	10	p	0.099	0.093	0.083	0.070	0.055	0.043	0.048	0.035
		γ	n.a.	n.a.	n.a.	n.a.	1.097	0.834	1.069	0.805
100	30	p	0.107	0.098	0.071	0.059	0.055	0.045	0.039	0.032
		γ	n.a.	n.a.	n.a.	n.a.	1.036	0.822	0.922	0.691
300	30	p	0.096	0.089	0.056	0.050	0.046	0.033	0.026	0.020
		γ	n.a.	n.a.	n.a.	n.a.	0.918	0.775	0.766	0.554
100	90	p	0.107	0.097	0.059	0.051	0.048	0.035	0.022	0.017
		γ	n.a.	n.a.	n.a.	n.a.	0.986	0.796	0.769	0.531

Table 3 Simulation results: MRSE and MedRSE for the probabilities of the observed process (p) and for the fixed effect in the logistic regression for the transition probabilities (γ) - Data generated from a nm-NH-HMM with std. dev. equal to 1

n	T	Parameter	H-HMM		fm-H-HMM		NH-HMM		fm-NH-HMM	
			MRSE	MedRSE	MRSE	MedRSE	MSE	MedRSE	MRSE	MedRSE
100	10	p	0.120	0.110	0.138	0.125	0.072	0.058	0.079	0.058
		γ	n.a.	n.a.	n.a.	n.a.	1.434	1.029	1.670	1.382
300	10	p	0.099	0.091	0.109	0.093	0.048	0.033	0.047	0.033
		γ	n.a.	n.a.	n.a.	n.a.	0.980	0.676	1.188	0.860
100	30	p	0.115	0.109	0.105	0.091	0.045	0.033	0.043	0.031
		γ	n.a.	n.a.	n.a.	n.a.	0.944	0.607	1.020	0.671
300	30	p	0.105	0.098	0.076	0.063	0.029	0.018	0.026	0.017
		γ	n.a.	n.a.	n.a.	n.a.	0.764	0.435	0.770	0.443
100	90	p	0.120	0.111	0.073	0.064	0.029	0.017	0.023	0.014
		γ	n.a.	n.a.	n.a.	n.a.	0.780	0.452	0.748	0.375

It is plain to notice that for small n and small standard deviation for the random term distribution, standard models can be used for parameter estimation instead of the proposed mixed model, which requires a more intensive computational procedure. The four models have been also estimated on samples generated according to a NH-HMM without individual heterogeneity. In this case the best estimates are obtained with the NH-HMM, with the finite mixture of NH-HMMs giving comparable results when the sample size is large (see Table 4).

In conclusion, the simulation study provides some expected results, showing that the use of more complex models is not efficient only when the sample size is small. Furthermore, an important result concerning the sensitivity of probability estimates in the observed process with respect to the hidden parameter estimates is shown: ignoring the effects of observed and unobserved covariates on transition probabilities leads to inefficient estimates of the observed process, denoting that the distribution of a hidden Markov model is sensitive to perturbations in the transition probability matrix and the initial distribution of the underlying Markov chain, i.e. altering these probabilities produces high relative changes in the final values of the HMM parameters.

5 Discussion

In this paper we propose a flexible approach for modeling longitudinal data by using NH-HMMs in a finite mixture context. The model is able to deal with several

Table 4 Simulation results: MRSE and MedRSE for the probabilities of the observed process (p) and for the fixed effect in the logistic regression for the transition probabilities (γ) - Data generated from a NH-HMM

n	T	Parameter	H-HMM		fm-H-HMM		NH-HMM		fm-NH-HMM	
			MRSE	MedRSE	MRSE	MedRSE	MSE	MedRSE	MRSE	MedRSE
100	10	p	0.099	0.088	0.121	0.106	0.066	0.043	0.080	0.050
		γ	n.a.	n.a.	n.a.	n.a.	1.339	1.041	1.503	1.236
300	10	p	0.076	0.062	0.089	0.070	0.037	0.024	0.039	0.026
		γ	n.a.	n.a.	n.a.	n.a.	0.996	0.689	1.128	0.886
100	30	p	0.084	0.079	0.093	0.079	0.045	0.023	0.045	0.025
		γ	n.a.	n.a.	n.a.	n.a.	1.003	0.668	1.091	0.797
300	30	p	0.066	0.052	0.070	0.053	0.025	0.014	0.025	0.014
		γ	n.a.	n.a.	n.a.	n.a.	0.790	0.379	0.859	0.494
100	90	p	0.077	0.061	0.081	0.064	0.030	0.015	0.030	0.016
		γ	n.a.	n.a.	n.a.	n.a.	0.825	0.445	0.824	0.462

sources of heterogeneity that may affects the hidden process. In particular, it is a finite mixture where the components are non homogeneous HMMs differing only in the initial and transition probabilities. Through the paper the model has been considered from different perspectives of view: a latent class model able to classify the individual on the basis of their dynamic behavior; a mixed NH-HMM where the random term has a discrete distribution; a model based on a semi-parametric estimator of a mixed NH-HMM following the NPML approach.

This model can be extended in several directions. It has been discussed using only a set of random intercepts, but it can be straightforwardly generalized to the case where even the regression coefficients of the logits are random, taking into account that the computational effort is only linear in the dimension of the random coefficients vector. Another possible extension could be to include covariates and/or random effects in the conditional model for the observed process (Rijmen et al., 2008; Zucchini et al., 2008; Maruotti and Rydén, 2009) in a generalized linear mixed model framework, taking into account individual-specific effects due to unobserved heterogeneity at the observed level. Of course, this extension could also deal with counts or continuous response data (it requires the use of a different kind of distribution from the multinomial used here).

In the paper we addressed some issues regarding the proposed model specification, but more detailed studies are still needed on some aspects. In particular, further investigation should be devoted to the choice of the number of components and states. In the simulation study, on each set of 250 samples we computed the average $BIC = -2 \log(\max L) + \#parameters \times \log(nT)$ for each model. The results show that the BIC does not always suggests the true number of components when

the sample size is small. For small sample sizes, BIC fails in indicating the most efficient number of components when the true distribution of the random effects is normal (in this case does not exist a *true* number of components but still remains the problem of choosing the right number of components able to efficiently estimate model parameters). Although these results are limited because the comparison have been made only between one or two components, they are encouraging because show that BIC works well for large samples. However, the problem regarding which criterion is able to identify the model with the best trade-off between fit and complexity, remains still open, especially for small sample sizes. A further analysis should also examine the starting points in quantity and quality. Increasing the number of replications improve the estimation results for the proposed finite mixture of NH-HMMs. Nevertheless, the starting points can not be completely at random. Non reasonable starting values for the parameters may lead to local maxima or degenerate solutions where the weight π_g of one component tends to be zero.

References

1. Alfò, M., Maruotti, A.: A hierarchical Model for Time Dependent Multivariate Longitudinal Data. In Palumbo, F., Lauro, C.N., Greenacre, M. (Eds.), *Data Analysis and Classification - Studies in Classification, Data Analysis, and Knowledge Organization*, Springer Verlag, pp. 271–279 (2010)
2. Altman, R.J.: Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal setting. *Journal of the American Statistical Association* **102**, 201–210 (2007)
3. Altman, R.J., Petkau, A.J.: Application of hidden Markov models to multiple sclerosis lesion count data. *Statistics in Medicine* **24**, 2335–2344 (2005)
4. Andersson, S., Rydén, T.: Subspace estimation and prediction methods for hidden Markov models. *Annals of Statistics* **37**, 4131–4152 (2009)
5. Baum, L.E., Petrie, T.: Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics* **37**, 1559–1563 (1966)
6. Baum, L.E., Petrie, T., Soules, G., Weiss, N.: A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, **41**:164–171 (1970)
7. Cooper, B., Lipsitch, M.: The analysis of hospital infection data using hidden Markov models. *Biostatistics* **5**, 223–237 (2004)
8. Dias, J.G., Vermunt, J.K., Ramos, S.: Mixture hidden Markov models in finance research. In: Fink, A., Berthold, L., Seidel, W. and Ultsch, A. (eds.) *Advances in data analysis, data handling and business intelligence*, Springer: Berlin-Heidelberg, pp. 451–459 (2010)
9. Hughes, J.P., Guttorp, P., Charles, S.P.: A non-homogeneous hidden Markov model for precipitation occurrence. *Journal of the Royal Statistical Society - Series C* **48**, 15–30 (1999)
10. Laird, N.: Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* **73**, 805–811 (1978)
11. Maruotti, A., Rydén, T.: A semiparametric approach to hidden Markov models under longitudinal observations. *Statistics and Computing* **19**, 381–393 (2009)
12. Rijmen, F., Ip, E.H., Rapp, S., Shaw, E.G.: Qualitative longitudinal analysis of symptoms in patients with primary and metastatic brain tumors. *Journal of the Royal Statistical Society - Series A* **171**, 739–753 (2008)
13. Rydén, T.: EM versus Markov chain Monte Carlo for Estimation of Hidden Markov Models: A Computational Perspective. *Bayesian Analysis* **3**, 659–688 (2008)

14. Shirley, K.E., Small, D.S., Lynch, K.G., Maisto, S.A., Oslin, D.W.: Hidden Markov models for alcoholism treatment trial data. *Annals of Applied Statistics* in press. (2010)
15. Zucchini, W., Raubenheimer, D., MacDonald, I.L.: Modeling Time Series of Animal Behavior by Means of a Latent-State Model with Feedback. *Biometrics* **64**, 807–815 (2008)