Coupled Markov-switching regression: inference and a case study using electronic health record data

Jennifer Pohle[1], Ruth King[2], Mihaela van der Schaar[3], Roland Langrock[1]

[1] Bielefeld University, Germany
[2] University of Edinburgh, United Kingdom
[3] University of Oxford, United Kingdom

E-mail for correspondence: jennifer.pohle@uni-bielefeld.de

Abstract: Coupled hidden Markov models (HMMs) are designed to capture the structure of multivariate time series whose underlying latent state variables interact, but do not evolve synchronously. Here we extend coupled HMMs to allow for covariates in the observed time series, which leads to the class of coupled Markov-switching regression models. The method is applied to electronic health record data of 702 patients from an intensive care unit at the UCLA, where the aim is to gain a better understanding of the course of a disease as well as early-warning signs of potentially critical developments.

Keywords: hidden Markov model; Markov chain; medical statistics; time series

1 Introduction

Hidden Markov models (HMMs) are time series models which assume the observations to depend on an underlying unobserved Markov chain with finitely many states. They have been applied in many different areas, e.g. speech recognition, finance, medicine, and ecology (for an overview, see Zucchini *et al.*, 2016). In the case of multivariate time series, within a basic HMM formulation, the variables would be expected to evolve synchronously in the sense that they are driven by the same underlying state sequence. However, in some applications, e.g. in medicine, the observed variables do not necessarily evolve in lockstep, although they may be correlated. For example, a substantial change in a patient's rate of breathing may or may not be accompanied by immediate visible changes in other vital signs. Coupled

This paper was published as a part of the proceedings of the 33rd International Workshop on Statistical Modelling (IWSM), University of Bristol, UK, 16-20 July 2018.

hidden Markov models (CHMMs) overcome this limitation by assuming separate but correlated state sequences to underlie the different variables observed, hence "coupling" the state processes of multiple HMMs (Brand, 1997).
However, the observations often depend not only on the underlying state, but also on external factors, e.g. the blood pressure of a patient might depend on its general health state as well as certain medications. Therefore, we extend CHMMs to allow for covariates in the observation processes, which leads to the flexible class of coupled Markov-switching regression models (CMSR). We apply this method to electronic health record data collected for 702 patients within the medical intensive care unit at the University of California in Los Angeles (UCLA).

2 Coupled Markov-switching regression models

2.1 Basic formulation of coupled hidden Markov models

We consider an M-dimensional observed time series of length T, denoted by $\{\mathbf{Y}_t\}_{t=1}^T$, with $\mathbf{Y}_t = (Y_{1,t}, \ldots, Y_{M,t})$. A CHMM for $\{\mathbf{Y}_t\}$ involves M state sequences, summarized in the vector $\{\mathbf{S}_t\}_{t=1}^T$, with $\mathbf{S}_t = (S_{1,t}, \ldots, S_{M,t})$, where $S_{m,t} \in \{1, \ldots, N_m\}$. The dependence structure of a CHMM with two underlying state processes is displayed in Figure 1.

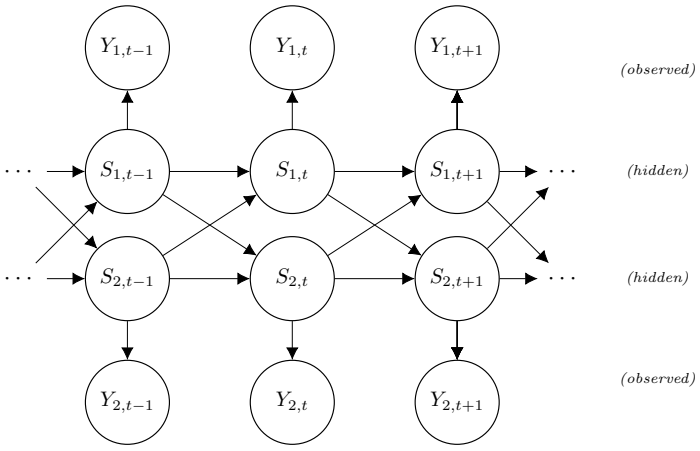

FIGURE 1. Basic structure of a CHMM with two underlying Markov chains.

Similar to HMMs, given the state sequences, the distribution of $Y_{m,t}$ is fully determined by the current state $S_{m,t}$ of its associated state sequence. However, for each state sequence, the future state $S_{m,t+1}$ depends not only on its current state $S_{m,t}$ — as would be the case if we were to consider M

separate HMMs — but on the current states of all M state sequences:

$$\Pr(S_{m,t+1}|\mathbf{S}_1,\ldots,\mathbf{S}_t) = \Pr(S_{m,t+1}|\mathbf{S}_t) \neq \Pr(S_{m,t+1}|S_{m,t}).$$

For simplicity of notation, from now on we assume that the number of states is the same for each state sequence, i.e. $N_m = N$ for all m. A CHMM can be written as an HMM with an $N^M \times N^M$ transition probability matrix $\Gamma = (\gamma_{ij})$, $i,j = 1,\ldots,N^M$, with each state representing an M–tuple corresponding to the possible states of $\{\mathbf{S}_t\}$. The state space thus is simply the Cartesian product of the M individual state spaces. Using the formulation with extended state space, the parameters can be estimated using standard HMM machinery, in particular by conducting a numerical maximization of the likelihood, which is evaluated using the forward algorithm. For model checking, pseudo residuals can be used, and the states can be decoded using the standard Viterbi algorithm.

2.2 Coupled Markov-switching regression models

To allow for covariates in the observation processes, we assume the expectations $\mathbb{E}(Y_{m,t})$ of the state-dependent distributions to be state-specific functions of (variable-specific) covariate vectors $\mathbf{X}_t^m = (X_{1,t}^m,\ldots,X_{p_m,t}^m)$:

$$\mathbb{E}(Y_{m,t}|s_{m,t}) = \beta_0^{(m,s_{m,t})} + \beta_1^{(m,s_{m,t})} X_{1,t}^m + \ldots + \beta_{p_m}^{(m,s_{m,t})} X_{p_m,t}^m.$$

Thus, the CMSR model comprises state-specific regression functions for each variable Y_m while also taking into account possible interactions in the state processes. Again, the forward algorithm and numerical optimization can be used to find the maximum likelihood estimates. Further extensions, for instance to generalized CMSR models with response distributions from the exponential family, are straightforward.

3 Case study on UCLA electronic health record data

Our dataset contains hourly information on vital signs, treatments and also personal data of 702 intensive care unit (ICU) patients at the UCLA, with a total number of 114927 observed time points. All of the patients considered underwent dialysis, which reduces the heterogeneity between the patients in the sense that individuals in a comparable situation are considered. Furthermore, these patients stayed in the ICU for at least 24 hours. Our aim is to model the observed heart rate, respiratory rate, systolic and diastolic blood pressure. The time series of observed vital signs all exhibit substantial changes over time, however, these changes do not always occur synchronously, which motivates the use of CMSR models rather than multivariate HMMs.

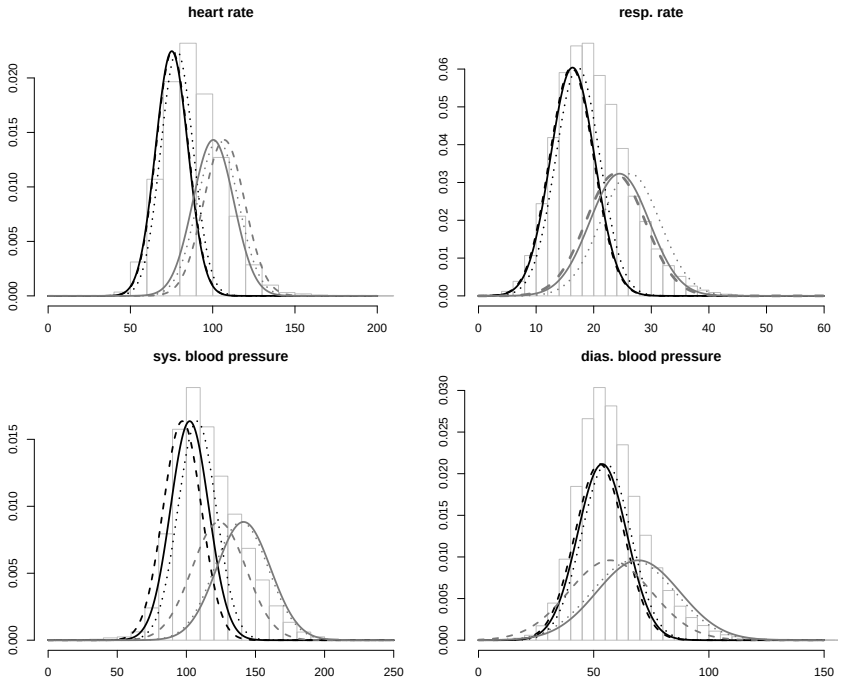

FIGURE 2. Estimated (marginal) state-dependent densities of the fitted CMSR model. The black lines correspond to the first state of the model, and the gray lines to the second state. Dashed lines indicate the presence of ventilation, dotted lines indicate the presence of vasopressors. The estimated correlations between systolic and diastolic blood pressure are $\rho^{(1)} = 0.48$ and $\rho^{(2)} = 0.43$ for state 1 and state 2, respectively.

We fitted a CMSR model with state-dependent normal distributions and $N = 2$ possible states per variable, considering the presence of ventilation and vasopressors as dummy covariates for all four vital signs. Since the blood pressure variables are highly correlated, we assume these two variables to depend on the same state sequence, with state-dependent bivariate normal distributions. The resulting marginal state-dependent distributions of our CMSR model are displayed in Figure 2. The estimated distributions are distinct, the model is thus able to capture changes in all four vital signs. However, only the presence of vasopressors is estimated to highly affect the state-dependent distributions, and only for the blood pressure variables (see Table 1).

TABLE 1. Estimated parameters of the state-dependent distributions for the four observed vital signs and both states, respectively.

param.	heart rate	resp. rate	sys. b. p.	dias. b. p.
$\beta_0^{(1)}$	75.15	16.33	102.41	53.71
$\beta_{ven}^{(1)}$	3.16	0.93	5.38	1.79
$\beta_{vaso}^{(1)}$	0.28	-0.17	-5.35	-1.14
$\sigma^{(1)}$	9.83	3.83	13.81	10.68
$\beta_0^{(2)}$	100.39	24.49	141.39	69.36
$\beta_{ven}^{(2)}$	2.80	1.86	-1.64	-2.16
$\beta_{vaso}^{(2)}$	6.59	-0.70	-17.74	-12.28
$\sigma^{(2)}$	12.44	5.18	19.60	18.04

4 Conclusion

Our preliminary results suggest that CMSR models are promising tools for analyzing the interplay of latent variables which do not evolve in lockstep, in particular such that are related to the health state of a patient. To decrease the computational costs, sparser parametrization of the transition probabilities could be used, as proposed for example by Saul and Jordan (1999). Current work focuses on model selection, on refinements of the particular model applied to the UCLA data, and on drawing meaningful medical inference.

Acknowledgments: We would like to thank Dr. Scott Hu (Division of Pulmonary and Critical Care Medicine, David Geffen School of Medicine, UCLA) for providing us with the clinical data.

References

Brand, M. (1997). Coupled hidden Markov models for modeling interacting processes. *Technical Report 405*, Massachusetts Institute of Technology Media Laboratory, Cambridge.

Saul, L. K. and Jordan, M. I. (1999). Mixed memory Markov models: Decomposing complex stochastic processes as mixtures of simpler ones. *Machine Learning*, **37**, 75–87.

Zucchini, W., MacDonald, I. L. and Langrock, R. (2016). *Hidden Markov models for time series: An introduction using R*, 2nd Edition. Boca Raton: Chapman & Hall/CRC.