Bolano, D., A. Berchtold, & G. Ritschard (2016)

A discussion on Hidden Markov Models for Life Course Data

in G. Ritschard & M. Studer (eds), Proceedings of the International Conference on Sequence Analysis and Related Methods, Lausanne, June 8-10, 2016, pp 241-260.

A discussion on Hidden Markov Models for Life Course Data

Danilo Bolano, André Berchtold, Gilbert Ritschard

Abstract This is an introduction on discrete-time Hidden Markov models (HMM) for longitudinal data analysis in population and life course studies. In the Markovian perspective, life trajectories are considered as the result of a stochastic process in which the probability of occurrence of a particular state or event depends on the sequence of states observed so far. Markovian models are used to analyze the transition process between successive states. Starting from the traditional formulation of a first-order discrete-time Markov chain where each state is liked to the next one, we present the hidden Markov models where the current response is driven by a latent variable that follows a Markov process. The paper presents also a simple way of handling categorical covariates to capture the effect of external factors on the transition probabilities and existing software are briefly overviewed. Empirical illustrations using data on self reported health demonstrate the relevance of the different extensions for life course analysis.

Key words: : Sequence Analysis, Life course approach, Hidden Markov Model.

Danilo Bolano
Institute for Social Science Research, Life Course Center, The University of Queensland, e-mail: `D.Bolano@uq.edu.au`

André Berchtold
Institute for Social Sciences, NCCR LIVES, University of Lausanne e-mail: `Andre.Berchtold@unil.ch`

Gilbert Ritschard
Institute of Demography and Socioeconomics, NCCR LIVES University of Geneva e-mail: `Gilbert.Ritschard@unige.ch`

Danilo Bolano, André Berchtold, Gilbert Ritschard

1 Introduction

Markovian models are stochastic models dedicated to the analysis of the transitions between successive states in sequences. More specifically, a Markovian model aims to describe how the current distribution of the possible values of a characteristic of interest depends on the previously observed values, and possibly how this dependence may be moderated by observable covariates or some unobservable latent factor.

Markov models have been extensively applied in many research areas such as speech recognition, behavior analysis, climatology, or finance. In demographic and population studies, they have been used for multistate analysis (e.g. Rogers, 1975; Land and Rogers, 1982; Willekens, 2014), for modelling population processes and life cycles (e.g., Feichtinger, 1973; Caswell, 2009), for analysing social mobility (since e.g., Hodge, 1966; McFarland, 1970), and so on. However, most of these studies consider only a simple Markov process, a first-order homogeneous Markov chain. Extensions such as higher order Markov chains, latent-based and non-homogeneous Markov models are instead rarely used. The aim of this paper is to underline the potential of latent-based Markov models, the so-called Hidden Markov Models, for life course studies.

In the Markovian perspective, life trajectories are considered as the result of a stochastic process in which the probability of occurrence of a particular state or event depends on the sequence of states observed so far. In other words, considering life trajectories as sequences of mutually exclusive states—e.g., sequences of employment statuses or of health conditions—a Markovian process focuses on successive transitions and attempts to depict the life history of an individual looking at the probabilities to switch to the different states of interest given the state history lived so far.

Markovian models form an extremely flexible class of models. The most immediate application is the study of transition mechanisms between successive states (e.g., studies on working career, social mobility, evolution of health conditions). A basic Markov chain models directly the transitions between visible states, while hidden Markov models prove useful to study how the succession of observed states may be governed by an underlying latent process. The latter approach is particularly interesting in life course studies where many not- or hardly observable aspects such as motivations, beliefs, levels of frailty, may influence the observed behavior of an individual. The salient aspect of hidden Markov models is that, unlike other latent models, it allows for a time-varying latent characteristic. Such models prove useful, for example, to study the transitions between unobserved vulnerability statuses revealed through some observable variable such as the health condition or the general satisfaction level.

The paper focuses on several important aspects of using Markov based models for life course studies addressing the analysis of transition mechanisms (Section 3), the modelling of latent processes (Section 4), probabilistic clustering, and how to evaluate the impact of covariates on the transition process (Section 6). Finally, existing software are briefly commented in Section 7 and the concluding discussion in Section 8 recaps the scope and applications of Markov modelling for life course analysis. The concepts presented all over the article are illustrated with longitudinal data on self-rated health represented in Section 2.

2 Longitudinal data on self-rated health condition

For illustration we use data from 14 waves of the Swiss Household Panel (Voorpostel et al., 2013). It is a yearly panel study started in 1999. We focus here on an unbalanced sub-sample of 1,331 individuals aged 50 years and more at the first interview and with at least three measurement occasions.

We intend to study the change over time in self-rated health conditions (SRH). We will analyse the transitions between SRH conditions (Section 3) and, by the means of a hidden Markov model, we will test whether there is some underlying hidden process that drives the observed changes (Section 4). Finally we will investigate the effects of the educational level on the process (Section 6).

The SRH condition is defined from the question "How do you feel right now?". Five possible answers were proposed: "not well at all", "not very well", "so-so", "well", "very well" that we shall denote respectively as P (poor), B (bad), M (medium), W (well) and E (excellent) health condition. The distribution on the overall dataset shows a general condition of good health. Near 80% of the respondents feel well (W) or very well (E) and only 2% bad (B) or very bad (P).

3 Markov chains

A discrete-time Markov chain is a stochastic process that describes how individuals transit between a finite number of pre-defined categorical states.

A Markov process models the probabilities of being in a state given the states visited in the past. In its basic formulation, the next value of the variable of interest depends only on the current state that is assumed to summarize the whole history of the individual. This is known as the *Markov property* and it defines a first-order Markov chain (Figure 1). For instance, with the two state alphabet corresponding to being in good (G) and bad health (B), we would consider, for people in a bad health condition, the probability to stay in the same condition the next period, $p(B|B)$, versus the probability to improve their health condition, $p(G|B)$. And, for those who are in a good health, the probability to have a deterioration in health condition, $p(B|G)$, versus the probability to stay in good health, $p(G|G)$.

The probability of switching from a given state to another is often assumed to remain unchanged over time. This defines a *time-homogeneous* Markov process. This assumption is tenable in fields as machine learning, biology but is often violated in applications in the social sciences. For example the probability of recovering from a bad health condition is likely to change over time with the age of the respondent.

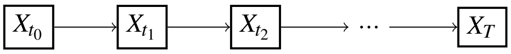


Fig. 1: A graphical representation of a first-order Markov chain.

For simplicity, this paper focuses on time-homogenous Markov processes but extensions to relax the homogeneity assumption have been proposed in the literature (see for instance the Double Chain Markov models proposed by Berchtold, 1999).

Table 1: First order Markov chain for SRH trajectories. The transition matrix.

$$\mathbf{A} = \begin{array}{cccccc} & & & X_t & & \\ X_{t-1} & \mathrm{P} & \mathrm{B} & \mathrm{M} & \mathrm{W} & \mathrm{E} \\ \mathrm{P} & 0.261 & 0.217 & 0.348 & 0.13 & 0.043 \\ \mathrm{B} & 0.024 & 0.159 & 0.53 & 0.271 & 0.018 \\ \mathrm{M} & 0.006 & 0.049 & 0.473 & 0.435 & 0.038 \\ \mathrm{W} & 0.002 & 0.009 & 0.141 & 0.726 & 0.123 \\ \mathrm{E} & 0.000 & 0.003 & 0.043 & 0.521 & 0.433 \end{array}$$

The transition probabilities are generally represented in matrix form in the so-called *transition matrix* (see for example Table 1). It is a square matrix of order m, with m the number of states.

Considering our illustration on self-rated health conditions, Table 1 reports the transition probabilities between health conditions estimated for a first order Markov chain. The probabilities to stay in the current state, reported on the main diagonal, are, with one exception, smaller than 50% meaning that there are frequent changes within the health trajectories. The probability to change the health condition is particularly high for people in bad condition. The probabilities of changing from a poor and a bad condition are respectively 73.9% and 84.1%.

4 Hidden Markov models

4.1 Including a latent process in life course data

Instead of modelling the stochastic process of the variable of interest—the health status in our illustration—it is often more realistic to assume that the successive values of this variable are governed by the underlying process of a latent variable such as motivation, belief, or vulnerability. Assuming that such categorical latent variable can change over time (i.e., is time-varying) following a Markov process, we get the so-called hidden Markov model (HMM) (see, e.g., Rabiner, 1989). The modalities assumed by the latent variable are called hidden states.

HMMs are widely used in biosciences and genetics (e.g., Le Strat and Carrat, 1999; Shirley et al., 2010) to study sequences of DNA and protein. An extensive literature exists in speech recognition since Baum and Petrie (1966). HMMs are also used in behavioral and criminal studies (Bijleveld and Mooijaart, 2003; Bartolucci et al., 2007), psychology (e.g., Visser et al., 2002) and in economics and finance where they are known as regime switching models (e.g., Elliott et al., 1998; Hayashi, 2004; Netzer et al., 2008).

There are several alternative ways of interpreting the latent variable in HMM providing multiple potential usages of this approach in life course studies. First, the HMM latent variable can be seen as a related unobserved characteristic of interest. For instance, assuming the reported health condition depends on the unobserved frailty level of the individual, an HMM would allow to study the stochastic process of this unobserved frailty.

Second, the hidden states may serve to capture the process heterogeneity or more specifically the 'person-position' heterogeneity, i.e., differences in the individual outcome probabiility distribution at the successive positions (e.g., McLachlan and Peel, 2000; Zucchini and MacDonald, 2009). In that case, the levels of the latent variable do not receive any specific interpretation but are just supposed to render the diversity of the person-period behaviors.

Third, HMM can be used for probabilistic clustering (see Section 5). This is similar to the capture of the process heterogeneity except that here a higher focus is put on each level of the latent variable being interpreted as a distinct latent class.

4.2 The HMM framework

When modelling life course data with an HMM, the sequence of observed events/states are supposed to be stochastically generated from a hidden Markov process. For each hidden state we have a different distribution of the visible state. So, it is the hidden process that selects at each position the distribution of the visible state.[1] Figure 2 shows a path diagram of a first-order hidden Markov process.

While a basic discrete first order Markov model is characterized by a response variable $X(t)$ with m modalities and a $m \times m$ matrix of transition probabilities $\mathbf{A}$, a first-order discrete HMM consists of five elements: i) a response variable $X(t)$ with m modalities;[2] ii) a categorical latent variable $S(t)$ with k modalities; iii) a $k \times k$ matrix $\mathbf{Q}$ of transition probabilities between two successive hidden states; iv) the emission—or outcome—probabilities, i.e., the probabilities $p_i(x_t)$, of observing

---

[1] For the sake of simplicity, in the rest of the paper, we shall omit the adjective "hidden" or "visible" when the nature of the state is unambiguous from the context.

[2] Even though the outcome variable $X(t)$ could also be numeric, we consider here only the case of a categorical response variable for simplicity.

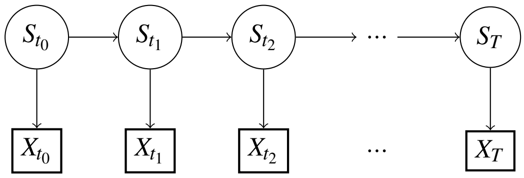


Fig. 2: A graphical representation of a first-order hidden Markov model. S_t is the latent variable at time t, X_t the observed response variable.

$X_t = x_t$ when being in the hidden state i; v) the initial probability distribution π of the k hidden states

The simplest HMM—a homogeneous hidden Markov model of order one—can be summarized using the following equations:

$$q_{ij} = p(S_t = i | S_{t-1} = j) \quad t = 1, ..., T \tag{1a}$$

$$\pi_i = p(S_{t_0} = i) \quad i = 1, ..., k \tag{1b}$$

$$p_i(x_t) = p(X_t = x_t | S_t = i) \quad i = 1, ..., k \tag{1c}$$

The first two equations represent the unobservable part of the model. Equation (1a) states that the latent variable S_t follows a first-order Markov process. So the current hidden state depends only on the previous one. As for visible Markov chains, a higher order dependence can be introduced. Equation (1b) gives the initial probability of the hidden states, i.e., the probability at the first time point $t = 1$.

The third equation (Eq 1c) refers to the measurement part of the model. It states that the visible state is determined by a hidden-state-dependent process—the emission probabilities (vertical arrows in Figure 2). Such emission probabilities are also known in the literature as response probabilities. The probability distribution of X_t depends only on the current hidden state and does not depend on previous observations nor on previous hidden states. In other words, Equation (1c) assumes that the observations are conditionally independent given the latent process. This is known as the *local independence* assumption.

4.3 Model comparison and selection of the number of hidden states

An important aspect with HMM is the choice of the number of hidden states. In some settings, the relevant number of states can be determined on theoretical grounds. This would be the case, for instance, when the latent variable explicitly stands for an a priori defined unobservable characteristic such as frail versus non-frail. Alternatively, we may want to let the number of states be determined from the data, i.e., choose the number of states on statistical grounds.

Since the HMM is an extension of mixture models, the issue of finding the number of hidden states is similar to find the number of mixture components (McLachlan and Peel, 2000). Information criteria such as the AIC and the BIC can be used. The model with the lowest BIC (or AIC) is chosen. Although other model selection criteria exists (e.g. cross-validated likelihood, Celeux and Durand, 2008), the BIC is the most commonly used.

For instance, in order to select the optimal number of hidden states in our empirical example, we compare several models in terms of likelihood and BIC increasing the number of hidden states up to 5 (Table 2). The lowest BIC (17971.8) is observed for the model with three hidden states.

Table 2: SRH trajectories. The choice of the number of hidden states.

Model	No. of hidden states	Free parameters	Log-Likelihood	BIC
HMM	2	11	-10532.6	21167.2
HMM	3	20	-8893.17	17971.82
HMM	4	31	-8887.315	18062.1
HMM	5	44	-8782.427	17972.87

Note: the number of parameters and the BIC do not include the null transition probabilities.

4.3.1 A 3-state HMM for SRH trajectories

As for other latent-based models, the order of the hidden states is not meaningful. The relationship between the data (i.e., the outcome variable) and the hidden states have to be analyzed using the emission probabilities (Eq. 1c, here reported in Table 3a) in order to give a "name" (labeling) each state. An alternative is to estimate for each individual the most likely sequence of hidden states (by the means of the Viterbi algorithm, Viterbi, 1967) and then to provide a cross tabulation between observations and the predicted hidden states (Table 4).

The first hidden state refers to individuals with perfect health condition with high chances to be in excellent (56.1%) or well (42.1%) condition. We will refer to this hidden state as a situation of "very good" condition (VG). Hidden State 2 is instead mainly associated with state M (65%) or with a worse health conditions (10% of probability of feeling B "not very well" or P "not well at all"). We will then label this state as "frail" health condition (F). This is the only state with individuals potentially at risk of frailty since in hidden state 1 and 3 we have almost no chance to be in a poor or very bad health condition. Finally, Hidden State 3 is an intermediate situation mainly associated with W (84%) or M (almost 10%). We will refer to State

Table 3: Three-state HMM

(a) Emission probability distribution by hidden states (columns).

SRH	Hidden State 1 *F*	Hidden State 2 *G*	Hidden State 3 *VG*
P	0.013	0.000	0.000
B	0.082	0.002	0.002
M	0.649	0.098	0.016
W	0.245	0.841	0.421
E	0.011	0.059	0.561

(b) Initial hidden state distribution π

F	*G*	*VG*
0.199	0.528	0.273

(c) Transition matrix between hidden states.

$$\mathbf{Q}=\begin{array}{cccc} & & S_t & \\ S_{t-1} & \text{F} & \text{G} & \text{VG} \\ \text{F} & 0.943 & 0.057 & 0.000 \\ \text{G} & 0.034 & 0.957 & 0.009 \\ \text{VG} & 0.000 & 0.084 & 0.916 \end{array}$$

Table 4: Cross tabulation between observed states (by rows) and predicted hidden states (1,331 individual sequences.)

SRH	*F*	*G*	*VG*
P	29	0	0
B	181	10	3
M	1383	655	29
W	426	5471	798
E	22	414	1228

3 as a state of "good" health (G)[3]. In the rest of the paper we will then (re)label and (re)order the hidden states as F, G and VG.

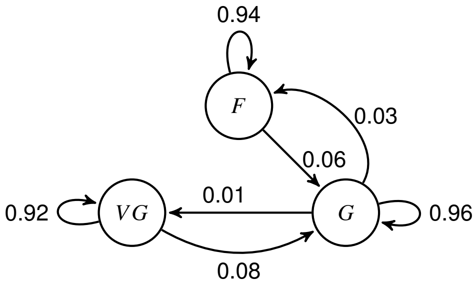
The transition probabilities between hidden states can be represented in matrix form (Table 3c) or since we have only three hidden states (i.e. the latent variable has three categories) as a path diagram (Table5). In the diagram, the arrows correspond-

---

[3] Here, the labels of the hidden states will be printed in italics since such states are not observed but inferred from the data.

ing to probabilities estimated as zero are not shown and for readability purposes transition probabilities have been rounded to two decimals.

Table 5: Transition probabilities as path diagram.



Despite the overall healthy aging of the Swiss population, we identify a relevant risk of vulnerability since the first observation. The initial distribution of hidden state (π in Table 3b) even if dominated by the hidden state G (52.7%), reports a 19.9% of chance to start the trajectory in a potentially at risk situation of frailty (hidden state F).

According to the transition probabilities (Table 3c), the states are very persistent meaning that a stability in health patterns is observed. There is more than 90% of probability to stay in the same state for two consecutive periods and three transitions, $(F - VG)$, $(G - VG)$, $(VG - F)$, are extremely rare or impossible. The transition probabilities for individuals with a good health condition, hidden state G, are particularly interesting. It is the most common hidden state and estimated to be used 59% of the time. Apart from those who stay in the same hidden state, they have more chance to fall down in the at frail condition rather than to improve their situation (3.4% vs 0.9%) confirming that a risk of a slight deterioration in health condition in the following period exists despite a general tendency of stability over time.

Once estimated the hidden states via the Viterbi algorithm (Viterbi, 1967) we can graphically represent their cross-sectional distributions as in the chronogram reported in Figure 3.

These results confirm the one reported for a visible Markov chain with a Swiss population in good health and with a pattern of stability or slightly deterioration over time. However, unlike before, a hidden Markov model allows to distinguish between three different conditions (the three hidden states) instead of analyzing the transition probabilities among the five observed health states.

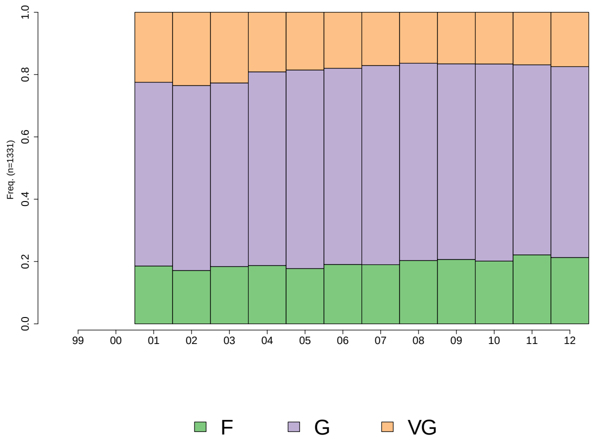


Fig. 3: Sequence of cross-sectional distributions of hidden states.

5 Using HMM for probabilistic clustering

As already mentioned, hidden states can be interpreted as latent classes and, therefore, HMM can serve for clustering. HMM belong to the class of mixture models so clustering with HMM is a model-based probabilistic clustering method. The class membership is given by the different emission probability distributions assigned to the hidden states.

HMM can perform two main types of clustering depending on whether the latent class is allowed to vary or not over time. We cluster person-position (person-period) states when the hidden state can vary over time, i.e., when using an unconstrained HMM, and trajectories when the latent variable is constrained to remain fixed over time. Other alternatives have been proposed in the literature as (e.g. Bicego et al., 2003) deriving pairwise dissimilarities between estimated sequences of hidden states and then to proceed with a dissimilarity-based clustering from those dissimilarities. Here, we focus on the first two alternatives.

5.1 *Person-position state clustering*

In an unconstrained conventional hidden Markov model, the latent variable is time-varying. Therefore, each individual may move from one cluster to another over time. In that case, we do not get clusters of individuals but clusters of states, actually of person-position states. Considering, for example, the three-hidden-state HMM fitted

Table 6: Emission probability distributions (columns). Constrained model.

SRH	$\overline{F}$	$\overline{G}$	$\overline{VG}$
P	0.011	0.001	0.000
B	0.069	0.003	0.003
M	0.55	0.114	0.021
W	0.346	0.801	0.458
E	0.024	0.081	0.518

before, we would have a cluster of frail states (F), one of good (G) and one of very good (VG) states of health conditions.

The membership probability at each time point is obtained multiplying the initial distribution π by the matrix of transition probabilities $\mathbf{Q}$ of the hidden process. For example, in our empirical example, the overall cluster membership probability distribution would be at position 3 given by $\pi * \mathbf{Q} * \mathbf{Q}$

$$\begin{array}{rccc} & F & G & VG \\ \pi\mathbf{Q}^2 = & \big(0.212 & 0.549 & 0.239\big) \end{array} ,$$

5.2 *Trajectory clustering*

To cluster the trajectories—the entire individual sequences—we have to constrain the transition matrix of the hidden process to be the diagonal identity matrix. This makes the latent variable time-invariant and individuals will belong to one and only one cluster during the whole period of observation. As in the period-position state case, we have to resort to the Viterbi algorithm to get cluster memberships.

5.2.1 HMM trajectory clustering of SRH conditions

We illustrate the clustering of trajectories with the SRH data by constraining the transition matrix $\mathbf{Q}$ of the hidden process to be the identity matrix $\mathbf{I}$. Re-estimating the model with this constraint, we get the emission probability distributions reported in Table 6. We use the notation $\overline{S}$ to designate the hidden states estimated using such a constrained transition matrix.

According to the initial probabilities,

$$\begin{array}{rccc} & \overline{F} & \overline{G} & \overline{VG} \\ \pi = & (0.245 & 0.554 & 0.201) \end{array}$$

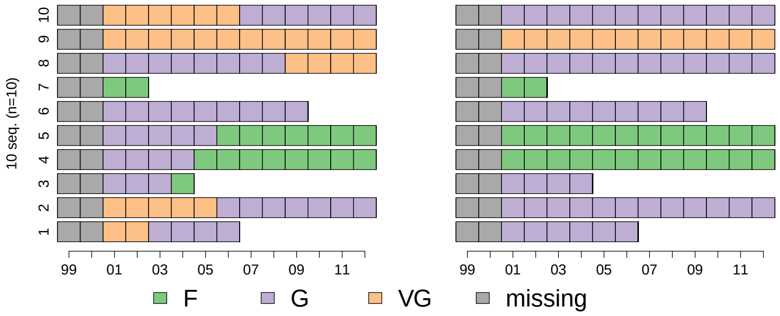


Fig. 4: Sequences of predicted hidden states. Unconstrained HMM (left-hand side) and with identity transition matrix (right-hand side).

we get a probability of 20.1% to belong to the hidden state $\overline{VG}$, 55.4% of being in hidden state $\overline{G}$ and 24.5% to be in “at risk” group $\overline{F}$

If we compare Table 6 with the emission probabilities for the unconstrained HMM (Table 3a), we observe only slight differences. Therefore, the interpretation of the hidden states remains the same and this justifies using a similar labelling of the clusters. Using the constrained transition matrix, the frail hidden state $\overline{F}$—actually a frail trajectory—seems to include a few more individuals with good health conditions than the frail state in the unconstrained case. People classified in this frail ‘trajectory’ have 37% chances to declare a well or excellent condition at one time point versus 26% for those in the frail state of the unconstraint model.

Figure 4 shows the differences in ten sequences of hidden states estimated using an unconstrained HMM (clusters of person-period states, left panel) and with the constrained transition matrix (clusters of trajectories, right panel). In the right panel where we cluster trajectories, each case remains in the same estimated hidden state for all periods where he/she responded. Looking at the first sequence for example, the individual is, according to the unconstrained HMM, in the hidden states VG for two periods and then, because of a worsening in its health status, he/she moves to hidden states G for the next four periods. According to the constrained model, he/she is estimated to have a good health $\overline{G}$ trajectory.

The appropriate approach should be chosen according to the research question. If the goal is to cluster individuals according to the entire trajectories and to study variation between groups, then a constrained matrix should be used. If we want to focus on the evolution of the situations lived by the individuals such as short periods at risk and recovery from positive or negative shocks, a person-period state clustering, i.e., an unconstraint HMM, is better suited.

6 Covariates

So far we have seen how Markov models can describe the probability to be in one or the other state of a response variable—the SRH in our illustrative example—in terms of the previously observed states of this same variable or of a latent Markov process. However, it is natural to make this relationship with previous values also depend on external factors so that we can test whether the fitted Markov process remains the same for different values of covariates. For example, we may want to know whether the evolution of the SRH is the same for men and women, or for people with different education levels. Covariates are a concern for both visible Markov models and HMMs.

In the literature, several methods have been considered to account for covariates. Berchtold and Raftery (1999) distinguish between two main approaches: By making the transition probabilities depend on the covariates, e.g., by means of a multinomial regression model, or through the interaction between the previous states and the values taken by the covariates. The first alternative is flexible. It can be used with multiple categorical as well as continuous covariates and also in case of multiple response variables (Bartolucci et al., 2015). However, parametrizing the transition probabilities dramatically increases the complexity of the model, which in turn involves many numerical computation difficulties. For a detail discussion on this approach refer to Bartolucci et al. (2012).

Here, we focus on the second alternative. The advantage of this second approach is its simplicity. It can be applied straightforwardly without modifying the estimation procedure both for Markov chains and Hidden Markov Models. However, this way of doing implies an increase of the number of probabilities to be estimated and works only with categorical covariates.

There are two main alternatives to make the current state depend on interaction between the modalities of categorical covariates and the previous states. We can either estimate directly one, possibly very large, transition matrix or approximate the matrix with a mixture model similar to the Mixture Transition Distribution model model proposed by Berchtold and Raftery (2002) for high-order Markov chains. We focus here on the first alternative where we estimate a single transition matrix with a row for each combination of the values taken by the covariates and the lag of the variable. The estimation of this matrix requires to simply count the number of observed transitions for each combination of modalities of the covariates. This approach is easy to implement however, increases the number of parameters to estimate. For the SRH data with two covariates—e.g. age and education—with each three modalities, the transition matrix would have $5 \times 3 \times 3 = 45$ rows and this number would increase exponentially with the number of variables and proportionally with the number of modalities of each variable. Below we include the covariates in the the three-state hidden model.

6.1 Effect of education level on the SRH process

We illustrate how we can account for covariates by studying the effect of educational level on the SRH trajectories More specifically, we include the covariate in the HMM with three hidden states considered before.

The level of education has been coded into three categories: *Low*, lower secondary level (22.29% of the 13,976 data points); *Medium*, secondary level and vocational school (44%); *High*, high educational level combining high level vocational school, maturity and university degrees (33.71%).[4]

The result of the direct estimation of the HMM with the education covariate is given in Table 7. The emission probabilities are very close from those of the model without covariate (Table 3). Therefore, we can maintain the same interpretation and labels for the hidden states.

From the transition matrix **D**, the probability of falling in the frail hidden state (F) decreases with the level of education. Moreover, less educated people have a probability to already be in a frail situation at the beginning of the sequence (Table 7b) twice bigger than the most educated ones (0.311 versus 0.139). The level of education has a slight positive impact also on chances to recover from a frail condition. For instance, people with high level of education have 2% more chances to move from a frail (F) to a good (G) situation (transition F-G) than those with a lower level of education (7% against 5%). Similarly, the probability of a worsening in the health condition (transition G-F) decreases with the educational attainment. All this indicates that education affects positively the evolution of the health condition.

7 Available software

According to some scholars (Ritschard and Oris, 2005; Scott et al., 2005), the limited use of Markovian models in life course and medical studies is also due to the limited software offer. Even the conventional homogeneous first-order Markov chains are not directly available in standard packages such as SPSS, Stata or SAS, and few dedicated software have been implemented. The March software[5] for categorical variables, for instance, was introduced in 2000 but, even though it offers advanced tools, March is not a free software, it runs only under the Microsoft Windows environment and is no longer maintained.

Fortunately, there are now several R packages that offer functions to model Markovian models. Most of these packages are designed for specific aims (e.g. for univariate time series or with specific type of processes) and not all propose the possibility to include the effect of external covariates. The more interesting R packages for fitting Markov models are `msm` (Jackson, 2011) and `Biograph` (Willekens,

---

[4] The percentages reported reflect the overall distribution of the educational level over the 14 waves considered.

[5] Software available at http://www.andreberchtold.com/march.html

Table 7: Direct estimation of the HMM with education as covariate.

(a) Emission probability distributions

	F	G	VG
P	0.013	0.000	0.000
B	0.082	0.002	0.002
M	0.647	0.097	0.016
G	0.247	0.844	0.423
V	0.011	0.057	0.559
H CI-Width	0.026	0.015	0.026

(b) Initial hidden state distributions (π)

	F	G	VG
Low	0.311	0.500	0.189
Medium	0.187	0.502	0.311
High	0.139	0.563	0.298

(c) Hidden transition distributions (**D**)

$t-1$	Education	F	G	VG
F	Low	0.950	0.050	0.000
F	Medium	0.949	0.051	0.000
F	High	0.922	0.074	0.005
G	Low	0.049	0.937	0.014
G	Medium	0.035	0.954	0.011
G	High	0.026	0.968	0.006
VG	Low	0.000	0.137	0.864
VG	Medium	0.000	0.091	0.909
VG	High	0.000	0.065	0.935

2014) for multi-state models in continuous time; `markovchain` (Spedicato and Signorelli, 2014) for discrete-time Markov chains; `HiddenMarkov` (Harte, 2010), `seqHMM` (Helske and Helske, 2016), and in particular `depmixS4` (Visser, 2010), `LMest` (Bartolucci et al., 2014) and `march` (the R port of the above mentioned March software, Berchtold, 2014) for discrete-time hidden Markov models. One of the main differences between the last three packages is the type of dependent variable used. In `depmixS4` the models can be fitted on data with distributions from the generalized linear model (glm) family, the (logistic) multinomial, or the multivariate normal distribution (i.e., continuous and discrete outcomes). The packages `LMest` and `march` are explicitly designed for discrete variables. These three packages differ also in the way of including covariates. `depmixS4` and `LMest` use a parametrization approach. The current version (version 1.0) of `march` does not support covariates, but will propose the method of transition probabilities by level of covariate values as presented in Section 6 in a next release. The empirical examples we provide in the paper have been computed using the Windows version of `march`.

8 Conclusion

The article illustrates the basic aspects and the flexibility of Markovian models with an illustration in life course studies to propose a general discussion on the relevance and possible applications of this class of models.

Starting from the traditional formulation of an homogeneous first order Markov chain, we presented the Hidden Markov Model and an intuitive and easy to understand way of including categorical covariates. Instead of using a parametrization of the transition probabilities, we consider directly the interaction between the states (observed or hidden) and the modalities assumed by the covariates. One of the main features that make Markovian models an interesting approach for life course studies is the specific role of time. The serial dependence between repeated measures is directly taken into account. In this transitional setting, the current measurement is described as a function of previous outcomes (Molenberghs and Verbeke, 2005) so that the distribution of the states depends on the own past of the subject. This aspect is particularly relevant for time-structured data such as individual life trajectories. For example, we have shown the significant relationship between previous self-reported health conditions and the current one.

In a life course perspective, modeling individual sequences at two levels, a visible and a latent one, as in the HMM, proves particularly interesting since many aspects of a life trajectory are not fully observable. For example, in the empirical example we have demonstrated that SRH trajectories are related to a latent variable representing frailty regimes. Moreover, unlike conventional latent variable approaches such as latent class or mixed effect models, HMM is a time-varying model. In many applications, the interest is not only to analyze the inter-individual differences in the response variable, but also the way in which individuals change their responses over time. In a HMM approach, the unobservable characteristic that drives the observed behavior has its own dynamics following a Markov process. The HMM then explore the dynamics in unobserved aspects which are measured by one or more response variables. The HMM can also be used as a clustering tool. HMM is a generalization of the mixture model (e.g., McLachlan and Peel, 2000) where each component is associated to one of the hidden states of the model. In particular, HMM can perform a probabilistic clustering in two ways. In its conventional formulation with a time-varying latent variable individuals can move among latent classes and we have a clustering of individual state observations. With a time-invariant latent variable (i.e., a HMM with each hidden state being fully absorbing), HMM performs a static probabilistic clustering of the individuals according to their entire observed sequences that can be seen as a clustering of the trajectories.

Although we focused here on the case of a single categorical outcome variable (the self-rated health condition), hidden Markov models can be applied to multivariate data (Bartolucci et al., 2012) and to numeric outcome variables. Bolano and Berchtold (2016) for example considered a double chain Markov model for numeric outcomes.

Interesting extensions of the HMM framework we did not discuss in the paper are the Double Chain Markov Models (see Berchtold, 1999)) that allow to relax the homogeneity assumption keeping the model parsimonious in terms of free parameters and the introduction of mixed effects (the so called Mixed Hidden Markov Model, Altman, 2007; Maruotti, 2011). By including (individual specific) random effects, the mixed HMM relaxes the assumption of conditional independence of the observations (Eq 1c). The resulting model is more flexible than a conventional HMM and it allows to distinguish between two sources of heterogeneity. The random effects capture the between-subject variations and the hidden states capture the heterogeneity in the individual trajectories.

Some aspects limit the diffusion of Markovian models in life course studies and related fields. For example, in social sciences and medical studies, a key aspect is to analyze the effect of external factors on the dynamics of the dependent variable. In the framework of hidden Markov models, although several alternatives to include covariates have been considered in the literature, an easily usable framework for estimating the effect of the covariates is still laking. Another limitation concerns the readability of the results in particular in presence of multiple states and/or multiple covariates. Comprehensive data visualization tools for representing empirical results are crucially missing too. However, given the numerous potential applications of Markovian models, and given the facilities offered by recently released software, we can expect Markovian modelling to overcome the above mentioned issues and this way to gain popularity also in life course research.

Acknowledgments

This publication benefited from the support of the Swiss National Centre of Competence in Research LIVES ”Overcoming vulnerability: life course perspectives”, which is financed by the Swiss National Science Foundation. The authors are grateful to the Swiss National Science Foundation for its financial assistance.

References

Altman, R. M. (2007). Mixed hidden markov models: An extension of the hidden markov model to the longitudinal data settings. *Journal of the American Statistical Association*, 102(477):201–210.

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2012). *Latent Markov Models for Longitudinal Data*. Chapman and Hall/CRC Press. Statistics in the Social and Behavioral Sciences, New York, NY.

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2014). The r package "lmest". comprehensive r archive network cran http://cran.r-project.org/web/packages/lmest/.

Bartolucci, F., Montanari, G. E., and Pandolfi, S. (2015). Three-step estimation of latent markov models with covariates. *Computational Statistics and Data Analysis*, 170(1):115–132.

Bartolucci, F., Pennoni, F., and Francis, B. (2007). A latent markov model for detecting patterns of criminal activity. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(1):115–132.

Baum, L. E. and Petrie (1966). Statistical inference for probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, 37 (6):1554–1563.

Berchtold, A. (1999). The double chain markov model. *Communications in Statistics: Theory and Methods*, 28(11):2569–2589.

Berchtold, A. (2014). The r package "march". comprehensive r archive network cran http://cran.r-project.org/web/packages/lmest/.

Berchtold, A. and Raftery, A. (1999). The mixture transition distribution (mtd) model for high-order markov chains and non-gaussian time series. Technical Report 360, Department of Statistics - University of Washington.

Berchtold, A. and Raftery, A. (2002). The mixture transition distribution model for high-order markov chains and non-gaussian time series. *Statistical Science*, 17(3):328–359.

Bicego, M., Murino, V., and Figueiredo, A. T. (2003). Similarity-based clustering of sequences using hidden markov models. *Machine Learning and Data Mining in Pattern Recognition. Lecture Notes in Computer Science*, 2734:pp 86–95.

Bijleveld, C. C. and Mooijaart, A. (2003). Latent markov modelling of recidivism data. *Statistica Neerlandica*, 57(3):305–320.

Bolano, D. and Berchtold, A. (2016). General framework and model building in the class of hidden mixture transition distribution models. *Computational Statistics & Data Analysis.*, 93:http://dx.doi.org/10.1016/j.csda.2014.09.011.

Caswell, H. (2009). Stage, age and individual stochasticity in demography. *Oikos*, 118(12):1763–1782.

Celeux, G. and Durand, J.-B. (2008). Selecting hidden markov model state number with cross-validated likelihood. *Computational Statistics*, 23(4):541–564.

Craig, B. A. and Sendi, P. P. (2002). Estimation of the transition matrix of a discrete-timemarkov chain. *Health economics*, 11(1):33–43.

Elliott, R. J., Hunterb, W. C., and Jamieson, B. M. (1998). Drift and volatility estimation in discrete time. *Journal of Economics Dynamics and Control*, 22:209–218.

Feichtinger, G. (1973). Markov models for some demographic processess. *Statistische Hefte*, 14(4).

Harte, D. S. (2010). R package "hiddenmarkov". comprehensive r archive network cran http://cran.r-project.org/web/packages/hiddenmarkov/.

Hayashi, T. (2004). A discrete-time model of high-frequency stock returns. *Quantitative finance*, 4:140–150.

Helske, S. and Helske, J. (2016). Mixture hidden markov models for sequence data:the seqhmm package in r.

Hodge, R. W. (1966). Occupational mobility as as probability process. *Demography*, 3(1):19–34.

Jackson, C. H. (2011). Multi-state models for panel data: The msm package for r. *Journal of Statistical Software*, 38(8):1–28.

Land, K. C. and Rogers, A. (1982). *Introduction to multiregional mathematical demography*. Academic Press, Waltham, MA.

Le Strat, Y. and Carrat, F. (1999). Monitoring epidemiologic surveillance data using hidden markov models. *Statistics in medicine*, 18(24):3463–78.

Maruotti, A. (2011). Mixed hidden markov models for longitudinal data: An overview. *International Statistical Review*, 79(3):427–454.

McFarland, D. D. (1970). Intragenerational social mobility as a markov process: Including a time-stationary markovian model that explains observed declines in mobility rates over time. *American Sociological Review*, 35(4):465–43.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. Wiley Series in Probability and Statistics, New York, NY.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer-Verlag. Series in Statistics, Berlin.

Netzer, O., Lattin, J. M., and Srinivasan, V. (2008). A hidden markov model of customer relationship dynamics. *Marketing Science*, 27(2):185–204.

Rabiner, L. I. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286.

Ritschard, G. and Oris, M. (2005). Life course data in demography and social sciences: Statistical and data-mining approaches. In Levy, R., Ghisletta, P., Le Goff, J.-M., Spini, D., and Widmer, E., editors, *Towards an Interdisciplinary Perspective on the Life Course. Advances in Life Course Research*, volume 10, pages 283–314.

Rogers, A. (1975). *Introduction to multiregional mathematical demography*. Wiley, New York, NY.

Scott, S. L., James, G. M., and Sugar, C. A. (2005). Hidden markov models for longitudinal comparisons. *Journal of the American Statistical Association*, 100(470):359–369.

Shirley, K. E., Small, D. S., Lynch, K. G., Maisto, S. A., and Oslin, D. W. (2010). Hidden markov models for alcoholism treatment trial data. *The Annals of Applied Statistics*, 4(1):366–395.

Spedicato, G. A. and Signorelli, M. (2014). The r package “markovchain”: Easily handling discrete markov chains. comprehensive r archive network cran http://cran.r-project.org/web/packages/markovchain/.

Visser, I. (2010). depmixs4 : An r package for hidden markov models. *Journal of Statistical Software*, 36(7):1–21.

Visser, I., Raijmakers, M. E. J., and Molenaar, P. C. M. (2002). Fitting hidden markov models to psychological data. *Scientific Programming*, 10:185–199.

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 16(2):260–269.

Voorpostel, M., Tillmann, R., Lebert, F., Kuhn, U., Lipps, O., Ryser, V.-A., Schmid, F., Rothenbühler, M., and Boris, W. (2013). Swiss household panel user guide. *Swiss Household Panel Userguide (1999-2012), Wave 14, November 2013. Lausanne FORS*.

Wiggins, L. M. (1973). *Panel Analysis: Latent Probability Models for Attitude and Behavior Processes*. Elsevier Scientific Pub, Amsterdam, Netherlands.

Willekens, F. (2014). *Multistate Analysis of Life Histories with R*. Springer-Verlag. Use R!, Berlin.

Zucchini, W. and MacDonald, I. L. (2009). *Hidden Markov models for time series. An introduction Using R*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, New York, NY.