A Comprehensive Exploration of Hidden Markov Models and Their Diverse Applications

Abstract:

Hidden Markov Models (HMMs) are statistical models to capture hidden information from observable sequential symbols. They have many applications in sequence analysis. A lot of Machine Learning techniques are based on HMMs and have been successfully applied to problems including speech recognition, optical character recognition, DNA sequencing, computational processing, progression of diseases into various stages, regime identification of financial time series, and they have become a fundamental tool in bioinformatics. In a HMM, the system being modelled is assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters from the observable parameters. The HMMs are described in brief along with examples. Some of its applications are chiefly discussed such as progression of diseases into various stages, and regime identification of financial time series.

Keywords*: Transition probabilities, Emission probabilities, Mean sojourn time, Stochastic process.*

UNDER PEER REVIEW

1. Introduction

1.1 Brief history of Markov chains

Andrey Andreyevich Markov (June 14, 1856 – July 20, 1922) was a Russian mathematician. He is well known for his work on the theory of stochastic Markov processes which include Markov chains. The Markov chains were introduced by Markov in 1906, when he produced the theoretical results for stochastic processes by using the term “chain” for the first time. In 1913, he determined the letter sequences of Russian language. But Markov appears to have pursued this out of a mathematical motivation. This approach led to the development of a general statistical instrument, the so-called stochastic Markov process. Generally in mathematics, particularly in probability theory and statistics, a Markov process can be considered as a stochastic process for which Markov property is achieved. The term Markov property refers to the memory-less property of a stochastic process i.e., the conditional probability distribution of future states of the process (conditional on both past and present states) depends only upon the present state, not on the sequence of events that preceded it. The use of the assumption that the Markov property holds for a certain random process in order to construct a stochastic model for that process is of more practical importance. In modelling terms, assuming that the Markov property holds is one among the limited number of simple ways of introducing statistical dependence into a model for a stochastic process in such a way that it allows the strength of dependence at different lags to decline as the lag increases.

1.2 Brief history of Hidden Markov Model:

Markov Models are too restrictive to be applicable to many problems of interest as each state corresponded to an observation (physical event). The concept of Markov models was extended to include the case where the observation is a probabilistic function of the state – i.e. the resulting model is a doubly embedded stochastic process with an underlying stochastic process that is not observable (it is hidden), but can only be observed through another set of stochastic process that produce the sequence of observations. The Hidden Markov Models (HMM) were later described in a series of statistical papers by Leonard E. Baum and other authors in the second half of the 1960s. One of the first applications of HMMs was speech recognition, starting in the mid-1970s. In the case of HMM, the current state is no longer directly visible to the observer, but in each state the model emits an observable output sequence. HMM involves two stochastic processes. Markov chain, being the first stochastic process, is characterised by the states and transition probabilities. The states of the chain are externally unobservable, therefore "hidden". Emissions observable at each moment are produced in the second stochastic process, depending on state-dependent probability distribution. It is conspicuous that the denomination "hidden" while defining a HMM refers to the states of the Markov chain, and not the parameters of the model.

UNDER PEER REVIEW

1.3. Brief history of algorithms needed to develop Hidden Markov Models:

In the middle of the 20th century, Claude Shannon (1916 – 2001), an American mathematician and electronic engineer, introduced in his paper "A mathematical theory of communication", published a very important historical step, that boosted the need of implementation and integration of the deterministic as well as stochastic automate in computer and electrical devices. Further there was also a need for important elements in the History of Algorithm Development in order to create, apply or understand Hidden Markov Models:

The expectation-maximization (EM) algorithm: An expectation-maximization (EM) algorithm is used for finding maximum likelihood estimates of parameters in probabilistic models, where the model depends on unobserved latent variables. EM alternates between performing an expectation (E) step and maximization (M) step. Where an expectation (E) step computes an expectation of the likelihood by including the latent variables as if they were observed, and maximization (M) step computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found on the E step. The parameters obtained in the M step are then used to begin another E step, and the process is repeated. EM is frequently used for data clustering in machine learning and computer vision. In natural language processing, two prominent instances of the algorithm are the Baum-Welch algorithm (also known as "forward-backward") and Viterbi algorithm.

The Baum-Welch algorithm: The Baum–Welch algorithm is a specific case of a generalized expectation-maximization (GEM) algorithm . The unknown parameters of a hidden Markov model (HMM) are found using the Baum–Welch algorithm. It makes use of the forward-backward algorithm hence also called as Forward-Backward algorithm.

The Viterbi Algorithm: It was introduced by Andrew Viterbi in 1967 as a decoding algorithm for convolution codes over noisy digital communication links. It is a dynamic programming algorithm. The Viterbi algorithm is used for finding the most likely sequence of hidden states, called the Viterbi path that results in a sequence of observed events. It is now generally used in speech recognition applications, keyword spotting, computational linguistics, and bioinformatics. For example, in certain speech-to-text recognition devices, the acoustic signal is treated as the observed sequence of events, and a string of text is considered to be the "hidden cause" of the acoustic signal. The Viterbi algorithm finds the most likely string of text given the acoustic signal.

2. Definitions and examples:

2.1. Markov chain: A stochastic process $\{ X_t, t = 0,1,2,\ldots\}$ is called a Markov chain, if for $j, k, j_1, \ldots, j_{t-1} \in N$,

$$P\{X_t = k \mid X_{t-1} = j, X_{n-2} = j_1, \ldots, X_0 = j_{n-1}\}$$

$$= P\{X_t = k \mid X_{t-1} = j\} = p_{jk} \text{ (say)},$$

whenever the first member $P\{ X_0 = j_{t-1}\}$ is defined.

UNDER PEER REVIEW

Example 1 (Medhi…): Consider a simple coin tossing experiment repeated a number of times. The possible outcomes at each trial are two: head with probability, say, p and tail with probability q, $p + q = 1$. Let us denote head by 1 and tail by 0 and the random variable denoting the result of the tth toss by X_t. Then for $t = 1, 2, 3, \dots$

$$P\{ X_t = 1 \} = p, P\{ X_t = 0 \} = q.$$

Thus we have a sequence of random variables $X_1, X_2, \dots$The trials are independent and the result of the tth trial does not depend in any way on the previous trials numbered 1,2, …, $t- 1$. The random variables are independent.

Consider now the random variable given by the partial sum $S_n = X_1 + X_2 + \dots + X_t$. The sum S_t gives the accumulated number of heads in the first t trials and its possible values are 0,1,2,…, t.

We have $S_{t+1} = S_t + X_{t+1}$. Given that $S_n = j$ ($j = 0,1,2,\dots, n$), the random variable S_{t+1} assume only two possible values : $S_{t+1} = j$ with probability q and $S_{t+1} = j + 1$ with probability p; these probabilities are not at all affected by the values of the variables $S_1 ,S_2 , \dots ,S_{t-1}$.

Thus

$$P\{ S_{t+1} = j+1 \mid S_t = j \} = p$$
$$P\{ S_{t+1} = j \mid S_t = j \} = q.$$

We have here an example of a Markov chain, a case of simple dependence that the outcome of (t +1)th trial depends directly on that of tth trial and *only* on it. The conditional probability of S_{t+1} given S_t depends on the value of S_t and the manner in which the value of S_t was reached is of no consequence.

The outcomes are called the states of the Markov chain; if X_t has the outcome j (i.e. $X_t = j$), the process is said to be at state j at tth trial. To a pair of states (j, k) at the two successive trials [tth and (t+1)th trials] there is an associated conditional probability p_{jk}. It is the probability of transition from the state j at tth trial to the state k at (t+1)th trial. The transition probabilities p_{jk} are basic to the study of the Markov chain.

2.2 Transition Probability Matrix (t.p.m.): The transition probability p_{jk} is the probability of moving from the state j at time $t-1$ to state k at time t. p_{jk} satisfy

$$p_{jk} \geq 0,\ \sum_{k=1}^{N} p_{jk} = 1 \text{ for all } j, \text{ where N is the number of states.}$$

These probabilities may be written in the matrix form

UNDER PEER REVIEW

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots \\ p_{21} & p_{22} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix}$$

This is called the transition probability matrix (t.p.m.) of the Markov chain. P is a matrix with non-negative elements and unit row sums.

The transition probability may or may not be independent of t. If the transition probability p_{jk} is independent of t i.e. if the transition probabilities are constant over time, the Markov chain is said to be *homogeneous* (or to have *stationary transition probabilities*). If it is dependent on t, the chain is said to be non-homogeneous.

Example 2 (Rabiner, 1989) : Consider a simple 3-state Markov model of the weather. Assume that in a day the weather is observed as being one of the following:

State 1: rainy (S_1)
State 2: snowy (S_2)
State 3: sunny (S_3)

Let the weather on day t is characterized by one of the three states above, and the weather on day t depends only on its previous day weather (weather at time t-1), then the matrix $\boldsymbol{A}$ of state transition probabilities is

$$\boldsymbol{A} = \{p_{ij}\} = \textbf{From} \begin{array}{c|ccc} & \multicolumn{3}{c}{\textbf{To}} \\ & S_1 & S_2 & S_3 \\ S_1 & 0.4 & 0.3 & 0.3 \\ S_2 & 0.2 & 0.6 & 0.2 \\ S_3 & 0.1 & 0.1 & 0.8 \end{array}$$

i.e. p_{31} is the probability of tomorrow being a *rainy* day given today is a sunny day.

Given that the weather on day 1 (t = 1) is sunny (state 3), we can find the probability that the weather for the next 7 days will be "sunny-sunny-rainy-rainy-sunny-snowy-sunny". Now let us define the observation sequence O as $O = \{S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3\}$ corresponding to $t = 1, 2, \ldots, 8$ and the required probability i.e. the probability of O, given the model can be expressed as

$$P(O \mid \text{Model}) = P\left[S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3 \mid \text{Model}\right]$$

$$= P\left[S_3\right] * P\left[S_3 \mid S_3\right] * P\left[S_3 \mid S_3\right] * P\left[S_1 \mid S_3\right] * P\left[S_1 \mid S_1\right] * P\left[S_3 \mid S_1\right] * P\left[S_2 \mid S_3\right] * P\left[S_3 \mid S_2\right].$$

$$= \pi_3 * p_{33} * p_{33} * p_{31} * p_{11} * p_{13} * p_{32} * p_{23}$$

$$= 1*(0.8)(0.8)(0.1)(0.4)(0.3)(0.1)(0.2)$$

UNDER PEER REVIEW

$= 1.536 * 10^{-4}$.

where we use the notation π_i, i = 1, 2, 3 to denote the initial state probabilities.
In situations where the states are unobserved / hidden, Hidden Markov Model is used.

2.3 Hidden Markov Model: A Hidden Markov Model (HMM) is a double stochastic process with the following two aspects:

- The first stochastic process is a finite set of states, where each of them is generally associated with a probability distribution. The transitions between the different states are statistically organized by a set of probabilities called transition probabilities.
- In the second stochastic process, the outputs which are dependent on the states are observed. Since the states are "hidden" to the observer, just what is observed is analysed, hence the name "Hidden Markov Model".

Elements of an HMM:

Each HMM is defined by states, observation symbols, initial probabilities, transition probabilities and emission probabilities.

1. N, the number of states in the Model; if $S_1, S_2, \dots, S_N$ are the possible hidden states of the HMM and X_t is the state at time *t*, then $X_t \in (S_1, S_2, \dots, S_N)$
2. M, the number of distinct observation symbols/values per state, if $v_1, v_2, \dots, v_M$ are the observed values and O_t is the value at time *t*, then $O_t \in (v_1, v_2, \dots, v_M)$.
3. $\pi = [\pi_j]_{1\times N}$, the initial state distribution vector, where

   $\pi_i = P(X_1 = S_j)$, $1 \le j \le N$.
4. A, the state transition probability matrix, A = $(p_{jk})_{N\times N}$, where

   $p_{jk} = P(X_{t+1} = S_k \mid X_t = S_j)$, $\qquad 1 \le j, k \le N$.

   the transition probabilities should satisfy the constraints, $p_{jk} \ge 0$, $1 \le j, k \le N$ *and* $\sum_{k=1}^{N} p_{jk} = 1$, $1 \le j \le N$.
5. B, the observation probability matrix, B = $(b_{jk})_{N\times M}$, where b_{jk} is the probability that symbol v_k is emitted in state S_j.

   $b_{jk} = P[O_t = v_k \mid X_t = S_j]$, $\qquad 1 \le i \le N$, $1 \le k \le M$.

   with the following constraints :

   $b_{jk} \ge 0$, $\qquad 1 \le j \le N$, $1 \le k \le M$ and $\sum_{k=1}^{M} b_{jk} = 1$, $1 \le j \le N$.

   Example 3: Consider example 2 above of weather Markov model. Consider the case of customer care support. Suppose one can only observe the customer's outlook i.e.

UNDER PEER REVIEW

customer may be wearing summer wear ($say\ o_1$), sweater (o_2) or have come with umbrella (o_3), but the weather outside cannot be seen. The states of weather are actually hidden, but one can only observe the emissions from the model. Hence it is a HMM. If the customer comes with an umbrella, then there is a greater probability that the weather outside is Rainy (say 65%), these probabilities are given by the emission / observation probability matrix.

$$B = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \end{array} \begin{array}{ccc} o_1 & o_2 & o_3 \\ \left[\begin{array}{c} 0.05 \\ 0 \\ 0.6 \end{array}\right. & \begin{array}{c} 0.3 \\ 0.5 \\ 0.1 \end{array} & \left.\begin{array}{c} 0.65 \\ 0.5 \\ 0.3 \end{array}\right] \end{array}$$

i.e. p_{33} is the probability of customer coming with an umbrella given it is a sunny day.

How do we calculate these probabilities?

Example 4 : Let us consider a two state model for the ease of calculation. Assume there are two friends A & B living far apart and B's mood changes according to the weather. i.e. B is mostly *Happy* (H) when it is *Sunny* (S) and he is mostly *Grumpy* (G) when it is *Rainy* (R). If B says A over telephone that he is *Happy* then she infers from this information that it must be *Sunny*. Similarly if B says A over telephone that he is *Grumpy* then she infers from this information that it must be *Rainy*.

Let q_t be the random variable denoting the state of weather at time *t* and o_t be the random variable denoting the observable (mood of Bob) at time *t*.

$$q_t=\{S,R\} \qquad , \qquad o_t=\{H,G\}$$

Consider the sequence of states and observables .

States :- S, S, S, S, R, R, R, S, S, S, S, R, R, S, S, S.

Observables :- G, H, H, H, G, G, H, G, H, H, H, G, H, H, H, H.

Initial probabilities:

P [$q_1 = S$] = 11/16 = 0.68

P [$q_1 = R$] = 5/16 = 0.32

Transition Probabilities:

P [$q_{t+1} = S \mid q_t = S$] = S,→ S,→S,→S,→ R, R,R, S,→S,→ S,→S,→R,R, S,→ S,→S.

8/10 =0.8

P [$q_{t+1} = R \mid q_t = S$] = S, →S,→S,→S,→R, R,R, S,→S,→S,→S,→R,R, S, →S,→S.

2/10 =0.2

P [$q_{t+1} = S \mid q_t = R$] = S, S, S, S, R,→R,→R,→S, S, S, S, R,→R,→S, S, S.

UNDER PEER REVIEW

2/5 =0.4

P [$q_{t+1} = R \mid q_t = R$] = S, S, S, S, R,➡R,➡ R,➡S, S, S, S,   R,➡R,➡S, S, S.

3/5 =0.6

Transition probability matrix:

$$A = \{p_{ij}\} = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$$

Emission probabilities:

P [$o_t = H \mid q_t = S$] = 9/11 = 0.82

P [$o_t = G \mid q_t = S$] = 2/11 = 0.18

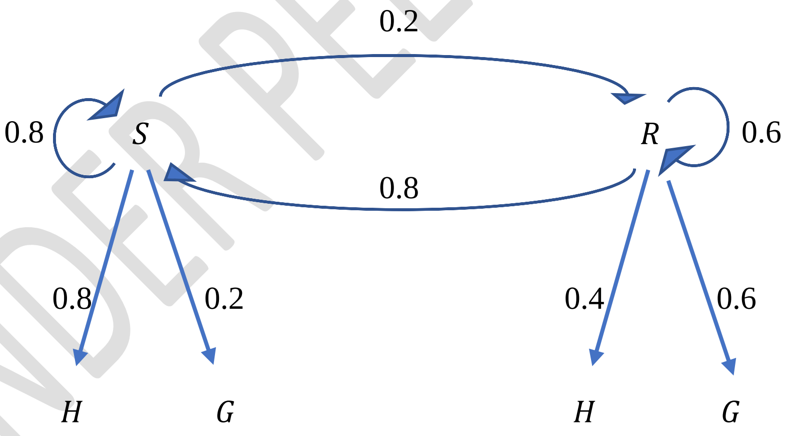
P [$o_t = H \mid q_t = R$] = 2/5 = 0.4

P [$o_t = G \mid q_t = S$] = 3/5 = 0.6

Emission probability matrix:

$$B = \{b_{jk}\} = \begin{array}{c} \\ S \\ R \end{array} \begin{array}{c} \begin{array}{cc} H & G \end{array} \\ \begin{bmatrix} 0.82 & 0.18 \\ 0.4 & 0.6 \end{bmatrix} \end{array}$$

Hidden Markov Model:



4. Uses of Hidden Markov Model:

- In many areas of probabilistic modelling.
- Regime identification of financial time series.
- Identification of the disease stage.
- Predicting the tourism demand.
- Prediction of Protein secondary structure.
- Modelling an oscillatory pattern in nucleosomes.

UNDER PEER REVIEW

- Modelling site dependence of evolutionary rates.
- Including evolutionary information in protein secondary structure prediction.

Some of the applications of HMM are chiefly discussed below:

Case 1) The estimation of misclassification probabilities of chronic kidney disease using Hidden Markov Models by Grover *et al.* (2019):-

The severity of Chronic Kidney Disease (CKD) is reflected in the form of stages of CKD and can be decided on the basis of estimated glomerular filtration rate (eGFR). The computation of eGFR is associated with computational and measurement errors which leads to misclassification of these stages of CKD. Objective is to estimate the transition intensities and mean sojourn times. For this purpose, the retrospective data of 117 patients suffering from CKD during the period March 2006 to October 2016 has been used. HMM has been developed to present the course of progression of CKD into various stages.

CKD is defined as the presence of kidney damage along with the subsequent decrease kidney function level. CKD is asymptomatic in its early stages and the determination of stage may be erroneous due to other prognostic factors which may lead to misclassification of stages. CKD can be classified into five categories of disease stages based on the eGFR as per the recommendations of National Kidney Foundation (NKF). The five stages are:

- stage 1 (GFR $\geq$ 90 ml/min/1.73 m^2)
- stage 2 (60 $\leq$ GFR $\leq$ 89 ml/min/1.73 m^2)
- stage 3 (30 $\leq$ GFR $\leq$ 59 ml/min/1.73 m^2)
- stage 4 (15 $\leq$ GFR $\leq$ 29 ml/min/1.73 m^2)
- stage 5 (GFR < 15 ml/min/1.73 m^2)

Stages 1, 2, 3 and 4 are transient states, i.e. movements from these states to other states in the forward direction are allowed. Stage 5 is an absorbing state and the movement from this state to any other state is not possible. It shows the loss of kidney functions and requires either dialysis or kidney transplantation. As the true states of disease are hidden. eGFR help in determining the actual state of disease. The progression of the disease is continuous in time and the time of transitions are random in nature. Hence the homogenous continuous time-HMM has been used as an appropriate model to describe the course of progression of CKD. The transition rates between various states and the misclassification probabilities between the true and observed stages of the disease have been estimated. These parameters help in computing

UNDER PEER REVIEW

the sojourn times of states and conditional probabilities. The mean sojourn time of a state is the length of time a CKD patient spends on an average in that stage before moving to the next stage.

The instantaneous risk of moving from stage ‘k’ to ‘l’ can be represented by the transition intensity(λ_{kl}). The transition intensity matrix Q takes the form as

$$Q = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{15} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{25} \\ . & . & & . \\ . & . & & . \\ . & . & \cdots & . \\ \lambda_{51} & \lambda_{52} & & \lambda_{55} \end{bmatrix}.$$

The matrix Q represents these transition intensities whose rows sum to zero, so that the diagonal entries are given by

$$\lambda_{kk} = -\sum_{l \neq k=1}^{5} \lambda_{kl}$$

The likelihood of transition intensities L(Q) is maximized in term of log(λ_{kl}) using optimization technique (Expectation maximisation) (Jackson ,2003). Therefore, the estimates λ_{kl} are obtained from log(λ_{kl}). Mean sojourn time for states can be computed from the estimated transition intensities as $-\frac{1}{\lambda_{kk}}$.

The transition states of CKD patients in their subsequent visits have been summarized in Table 1. Generally, number of visits depends on the severity of the disease and awareness about the disease. The table has been prepared by counting the number of transitions for each patient in their subsequent visits. The total number of times patients of stage 1 remains in stage 1 in their subsequent visits is 75. It is the frequency table counting the number of times each pair of states were observed in successive observation times.

Table 1. Number of state transitions

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Stage 1	75	8	4	3	0
Stage 2	0	116	6	3	1
Stage 3	0	0	293	29	2
Stage 4	0	0	0	145	51

UNDER PEER REVIEW

The number of transitions from higher stages to lower stages is zero as CKD is irreversible. Estimated transition intensities of simple homogeneous continuous-time multistate model based on Markov process have been shown in Table 2 (using optimisation technique).

Table 2. Estimated transition intensities

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Stage 1	-0.098	0.0599	0.0266	0.0115	0
Stage 2	0	-0.119	0.0725	0.0467	0
Stage 3	0	0	-0.115	0.115	0
Stage 4	0	0	0	-0.564	0.564

Mean sojourn time of each state is summarized in Table 3.

Table 3. Mean sojourn times at different stages

	Mean Sojourn Time
Stage 1	10.2030
Stage 2	8.3829
Stage 3	8.7051
Stage 4	1.7734

Estimated survival probability curves for stage 1, stage 2, stage 3 and stage 4 have been shown in Figure 1(msm package of R) . It is clear from the figure that there is a sharp decline in survivability of stage 4 patients nearing 1 year. The estimated transition intensities for HMM are summarized in Table 4.

UNDER PEER REVIEW

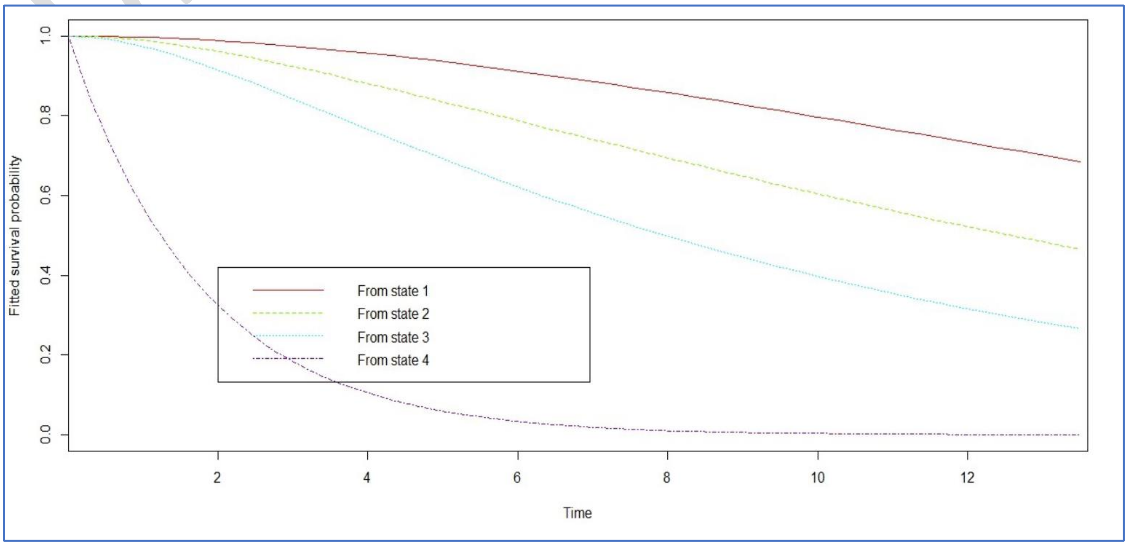


Figure1. Plot of fitted survival probability.

The number of transitions to next higher stage is maximum for stage 4 indicating the rapid progression of disease in advance stage 4. The diagonal entries of the transition matrix are negative which indicates the negation to the instantaneous risk of moving to other stage. The mean sojourn time for stage 1 is 10.2030 years. It means that a patient of stage 1 spends on an average 10.2030 years in stage 1 only before moving to stage 2. The mean sojourn time for stage 4 is 1.7734 years. This reveals the fact that the progression of disease is very slow in early stages as compared to higher stages. The knowledge of natural history of disease progression is necessary for devising a policy by policymakers. It helps in reducing the economic burden of the treatment.

Case 2) Hidden Markov Model for Financial Time Series and its Application to S&P 500 Index by *Lihn* (2017).

Lihn has studied the financial time series using HMM using R package “ldhmm” on S&P 500 index. S&P 500 index is the stock market index that measures stock performance of 500 large companies of US. He has identified different market regimes using the powerful tool of HMM. The Hidden Markov Models for two states, three states, four states, five states and six states have been fitted and analysed. He has observed that the models with more than five states capture more amount of auto-correlation, matching what has been observed in the data. Although the normal regime and crash regime are the two broad categories of stock market, he has gone beyond two-regime categories as the progression of HMM states allows the same. It is the fact that the stock market tends to rise when the volatility is low , while tends to fall when the volatility is high. Hence HMM is used to regime identification.

In general, kurtosis in each state decreases as number of states increases. Large outliers tend to get pushed to the high-volatility states. The ldhmm package is specifically designed for the analysis of SPX index. The package comes with the daily closing prices. There are two distinct features in the daily returns of SPX. First, its kurtosis is very large, around 30. Even after 10 largest outliers are dropped, the kurtosis is still as large as 8.5.

```
> sapply(0:10,function(drop) kurtosis(ldhmm.drop_outliers(ts$x,drop)))
 [1] 30.279540 12.479094 11.631622 10.983566 10.505473 10.041342
 [7]  9.595356  9.252240  8.903898  8.664750  8.518157
```

UNDER PEER REVIEW

Second, the auto-correlation of the returns is more than 20%. SPX's ACF moves around 23-28% for the first 6 lags. Dropping 10 largest outliers doesn't help reducing the ACF much.

```
> ldhmm.ts_abs_acf(ts$x,drop=0,lag.max=6)
[1] 0.2444656 0.2737506 0.2502788 0.2434303 0.2809230 0.2389925
> ldhmm.ts_abs_acf(ts$x,drop=10,lag.max=6)
[1] 0.2271080 0.2461340 0.2290556 0.2375244 0.2520413 0.2273132
```

Two states model:

The analysis starts with two states. The stock market can be classified to two regimes: the normal regime where the market is calm and rising, and the crash regime where the market is panic and plunges down. The former is also called "the bull market" while the latter "the bear market". Empirical evidence shows that the market spent "most" of time in the normal regime.

The theoretical statistics of each state is:

```
> ldhmm.ld_stats(hd)
              mean         sd kurtosis
[1,]  0.0005841558 0.00635735        3
[2,] -0.0006888484 0.01650986        3
```

Compare this to the statistics of the data classified in each state, as shown below:

```
> hd@states.local.stats
              mean          sd  kurtosis    skewness length
[1,]  0.0006007835 0.006333934  3.565389 -0.02491462  12868
[2,] -0.0007847466 0.016749016 14.828955 -0.71360850   3737
```

The kurtosis of the first state matches, but the kurtosis of the second state still differ quite a bit. This is because there are very large outliers in the data that cannot be captured by the two-state model. The package provides the following utility to drop those outliers then calculate statistics. It turns out that, after 11 largest moves are dropped, the kurtosis of the second state can be reduced. Not only the kurtosis is rectified, the large skewness of the second state is also reduced. The concept of dropping the few largest outliers is called "asymptotic statistics".

```
> ldhmm.calc_stats_from_obs (hd , drop =11)
              mean         sd kurtosis     skewness length
[1,]  0.0006027576 0.00629874 3.478128 -0.021274020  12857
[2,] -0.0006481724 0.01568384 4.275335  0.003376325   3726
```

UNDER REVIEW

We will compute the minus log-likelihood (MLLK) for model selection, as we gradually increase the number of states in the following sections.

Now shall make some interpretations based on the two-state HMM result:

(i) The first state represents the normal state, where the mean is positive. The second state represents the crash state, where the mean is negative.

(ii) The standard deviation of the crash state is more than twice of the normal state, depicting the panic and volatility of the bear markets.

(iii) According to the initial state probability vector delta, the market spent 76% of its time in normal state.

```
> hd@delta
[1] 0.7681356 0.2318644
```

The regime identification is to classify the stock market into a few regimes that human can comprehend. In terms of regime identification, we observe that the low volatility state corresponds to the bull markets, the high volatility state corresponds to the bear markets. The more states the better it is. We will explore the HMM with more states, from 3, 4, 5, then 6. By carefully crafting the parameter with different ranges of standard deviation, the MLE optimizer can converge reasonably well. The following are the results for 3, 4, 5, and 6 states. The main takeaway is that the MLE is optimized at five states based on MLLK. The MLLK of two state model is -56086.11.

```
> ldhmm.mllk(hd,ts$x)
[1] -56086.11
```

UNDER PEER REVIEW

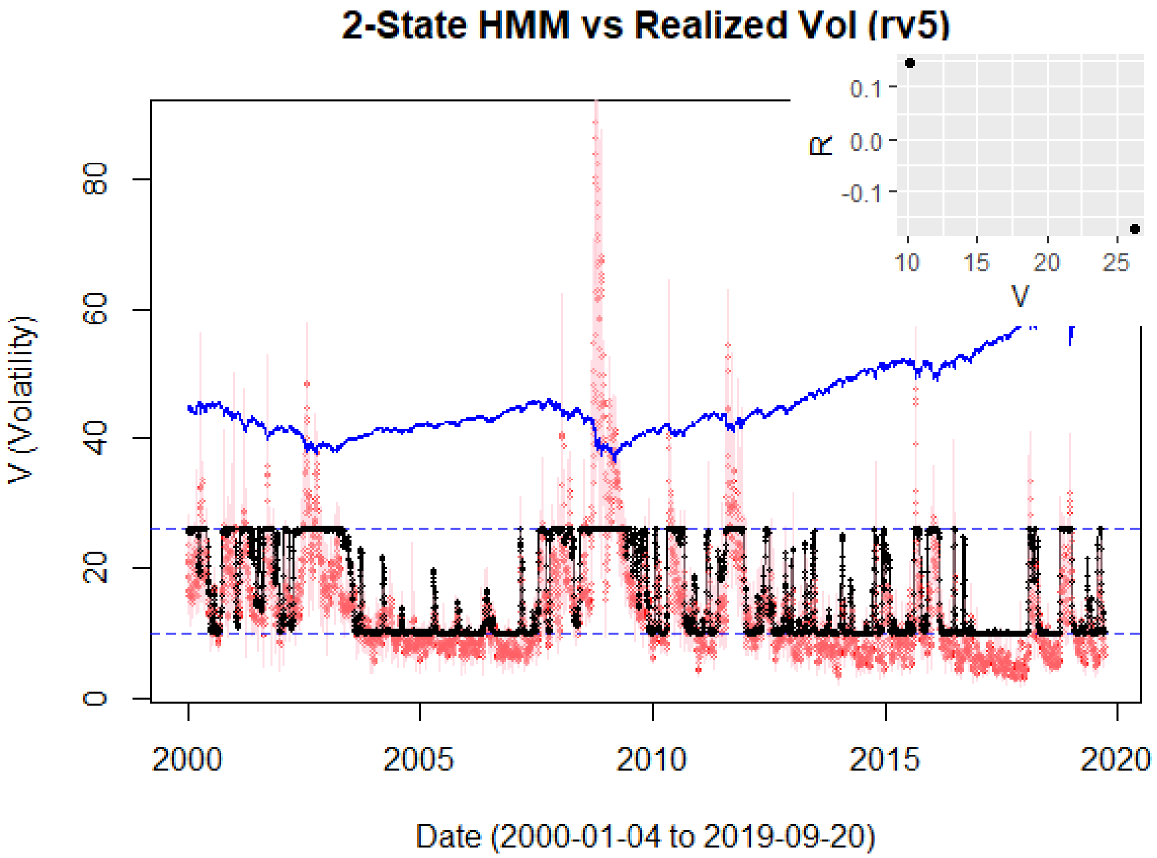


Figure 2 . Comparison of the expected volatility V (t) from two-state HMM (black) vs the realized volatility (red) from Oxford-Man realized variance data set. Two-state HMM is the simplest model to illustrate the normal and crash regimes. The red line is the daily realized volatility, and the red dots are the 5-day moving average. The dash blue lines are the volatilities of the HMM states. The solid blue line is the rescaled SPX price index.

UNDER PE

The best way to visualize the states is to compare the expected volatility V (t) to the realized volatility calculated from Oxford-Man data. The expected volatility is the measure of the amount by which a financial variable, such as share price, has fluctuated or is expected to fluctuate during a period, whereas the realized volatility is the measure of amount of actual fluctuation that occurs in a given underlying over a defined past period. Oxford man is an interdisciplinary research institute of the University of Oxford, England.

Three states model:

The three-state HMM produces better MLLK score than the two-state HMM. The MLLK is -56647.52.

```
> ldhmm.mllk(hd,ts$x)
```

```
[1] -56647.52
```

The visualization of 3 states model:

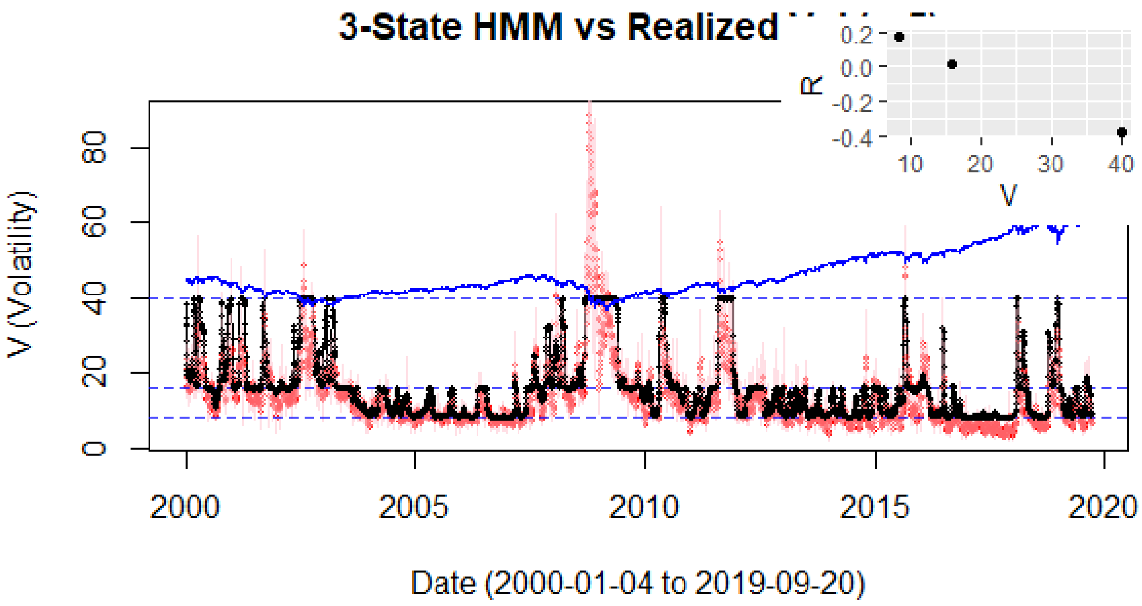


Figure 3. Comparison of the expected volatility V (t) from three-state HMM (black) vs the realized volatility (red). Three-state HMM reveals the intermediate state or the so-called transition state.

It is observed that the bull market is a condition of staying in state 1 and 2 persistently. On the other hand, the bear market was formed by oscillations between state 2 and 3. Maybe the definitive start of a bull market can be defined as the first touch of state 1, while the definitive start of a bear market as the first touch of state 3 (which ends previous bull market).

Four states model:

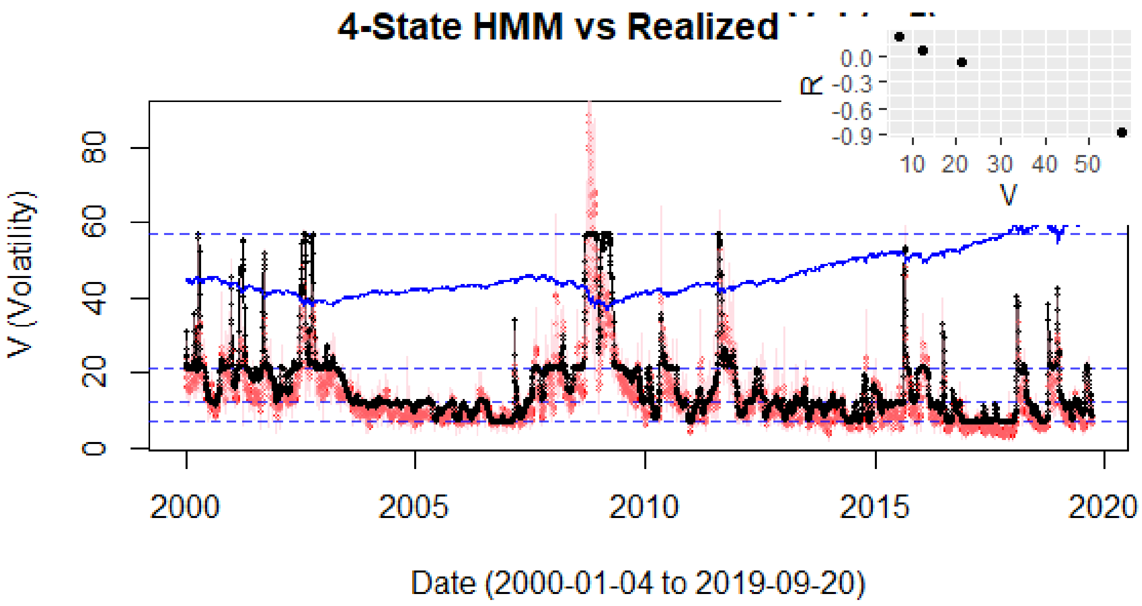
The MLLK scores continue to improve for four states. The MLLK is -56837.18.

```
> ldhmm.mllk(hd,ts$x)
[1] -56837.18
```

The visualization of the model:

UNDER PEER

Figure.4: Comparison of the expected volatility V (t) from four-state HMM (black) vs the realized volatility



(red).

Third state is the transition state. If it transitions back to state 1 or 2, then the bull market resumes. If it transitions to state 4, the crash state, then the market suffers violent downward moves.

Five states model:

The MLLK scores improves slightly. The MLLK is -56971.67. The five-state HMM has several attractive features:

(i) From MLLK perspective, five-state is the most optimized HMM in model selection.

(ii) The excess kurtosis of states, is most evenly distributed.

(iii) There are two low-volatility states (1, 2, bull market), one transition state (3), and two high-volatility states (4, 5, bear market).

The visualization of the model:

UNDER PEER RE

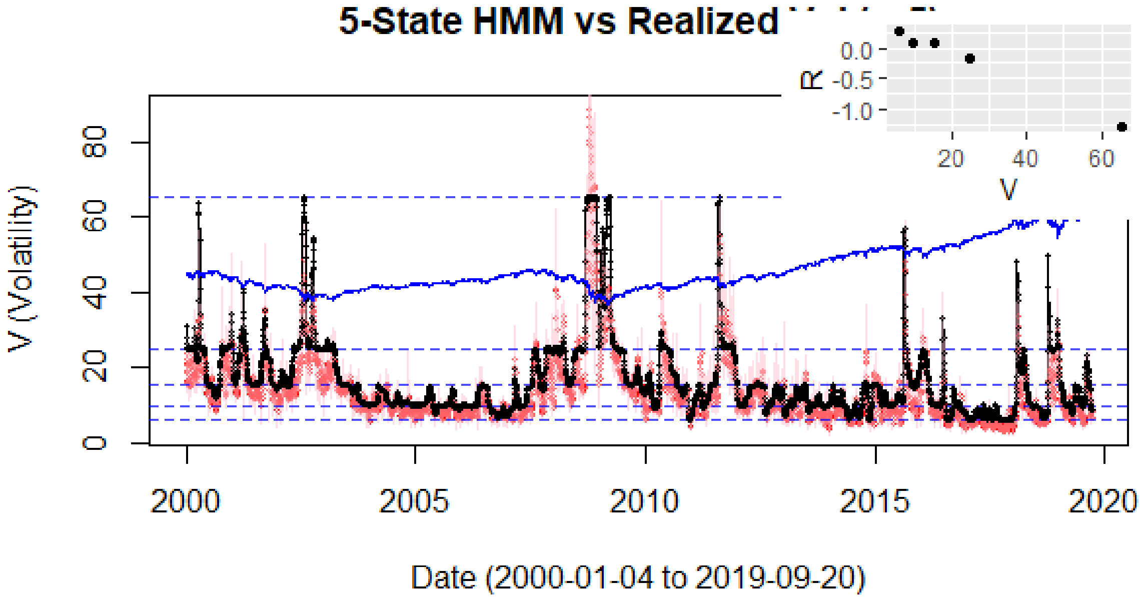


Figure 5: . Comparison of the expected volatility V (t) from five-state HMM (black) vs the realized volatility (red).

It is observed that the bull market is a condition to stay in the two low-volatility states, state 1 and 2, persistently. On the other hand, the bear market was formed by oscillations between state 3 and 4. And the fifth state is very destructive.

Six states model:

The AIC and BIC scores no longer improves. The MLLK is -56962.

Hence finally he says that the Five states model is the best fit model. This model divides the stock market into five regimes such that there are two low-volatility states (1, 2, bull market), one transition state (3), and two high-volatility states (4, 5, bear market). It can be further used for volatility forecasting, and also to know the behaviour of the stock market and planning accordingly.

Conclusions:

This study provides a comprehensive overview of Hidden Markov Models (HMMs) and their diverse applications in various fields such as bioinformatics, disease progression modeling, and statistical inference. The investigation delves into the fundamental concepts of HMMs, including transition probability matrices, state transitions, and the use of latent variables to model complex systems. Additionally, it highlights the significance of the Expectation-Maximization (EM) algorithm in parameter estimation for probabilistic models with unobserved variables. Through references to seminal works and research studies, the study underscores the importance of HMMs in capturing hidden information from observable data

UNDER PEER REV

sequences and their role in advancing statistical inference and modeling techniques. Overall, this study serves as a valuable resource for researchers, practitioners, and enthusiasts seeking to deepen their understanding of HMMs and their practical implications in real-world scenarios.

References:

Alizadeh, S. H., & Rezakhah, S. (2013). Hidden Markov mixture autoregressive models: stability and moments. *Communications in Statistics-Theory and Methods*, *42*(6), 1087-1104.

Bartolucci, F., & Farcomeni, A. (2010). A note on the mixture transition distribution and hidden Markov models. *Journal of Time Series Analysis*, *31*(2), 132-138.

Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The annals of mathematical statistics*, *37*(6), 1554-1563.

Cappé, O., Moulines, E., & Rydén, T. (2009, June). Inference in hidden markov models. In *Proceedings of EUSFLAT conference* (pp. 14-16).

Durbin, R., Eddy, S. R., Krogh, A., & Mitchison, G. (1998). *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge university press.

Elliott, R. J., Aggoun, L., & Moore, J. B. (2008). *Hidden Markov models: estimation and control* (Vol. 29). Springer Science & Business Media.

Gollery, M. (2008). *Handbook of hidden Markov models in bioinformatics*. Chapman and Hall/CRC.

Grover, G., Sabharwal, A., Kumar, S., & Thakur, A. K. (2019). A Multi-State Markov Model for the Progression of Chronic Kidney Disease. *Türkiye Klinikleri Biyoistatistik*, *11*(1), 1-14.

C. H. Jackson, L. D. Sharples, S. G. Thompson, S. W. Duffy, and E. Couto. Multistate Markov models for disease progression with classification error. *Journal of the Royal Statistical Society,Series D - The Statistician*, 52(2):193–209, July 2003

Jahromi, K. G., & Jahromi, V. G. (2018). Using discrete hidden Markov model for modelling and forecasting the tourism demand in Isfahan. *J Inf Syst Telecommun*, *6*, 112-118.

Kouemou, G. L., & Dymarski, D. P. (2011). History and theoretical basics of hidden Markov models. *Hidden Markov Models, Theory and Applications*, *1*.

Leite, P. B. C., Feitosa, R. Q., Formaggio, A. R., Da Costa, G. A. O. P., Pakzad, K., & IDA, S. (2008). Hidden Markov models applied in agricultural crops classification. *Proceeding of GEOBIA (GEOgraphic Object-Based Image Analysis for the 21St Century)*.

UNDER PEER REVIEW

Lihn, S. H. (2017). Hidden Markov Model for Financial Time Series and Its Application to S&P 500 Index. *Quantitative Finance, Forthcoming*.

Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suarez, C., & Andersen, P. K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical methods in medical research, 18*(2), 195-222.

Polisetti, H. (2016). Hidden Markov Chain Analysis: Impact of Misclassification on Effect of Covariates in Disease Progression and Regression.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257-286.

Rabiner, L. R., & Juang, B. H. (1986). An introduction to hidden Markov models. *ieee assp magazine*, *3*(1), 4-16.

Sipos, I. R., Ceffer, A., Horváth, G., & Levendovszky, J. Parallel MCMC sampling of AR-HMMs for prediction based option trading. *Algorithmic Finance*, (Preprint), 1-9.

Zhang, Y. (2004). *Prediction of financial time series with Hidden Markov Models* (Doctoral dissertation, Applied Sciences: School of Computing Science).

Zucchini, W., MacDonald, I. L., & Langrock, R. (2017). *Hidden Markov models for time series: an introduction using R*. Chapman and Hall/CRC. Medhi, J. (1994). *Stochastic processes*. New Age International.

UNDER PEER REVIEW