Type of article: Original Research Article

Theoretical Foundations and Application of Hidden Markov Models

Abstract:

Hidden Markov Models (HMMs) are effective statistical techniques used to uncover underlying patterns in observable sequential data This paper provides a comprehensive overview of the theoretical foundations of HMMs, including key algorithms essential for their implementation, such as the Expectation-Maximization algorithm, the Baum-Welch algorithm, and the Viterbi algorithm. An application of HMMs to assess the likelihood of misclassification in Chronic Kidney Disease (CKD) stages is discussed, providing insights into the disease's natural progression and informing better treatment strategies.

In the case study, we conducted an analysis of the S&P 500 index dataset spanning from January 4, 2000, to September 20, 2019. This analysis, similar to Lihn's study but with an extended period, evaluates the robustness and consistency of regime identification through HMMs, offering deeper insights into market behavior over nearly two decades. HMMs identify different market regimes, revealing the underlying dynamics of bull and bear markets. The study demonstrates the superiority of a five-state HMM model in capturing market behaviors, thereby aiding in volatility forecasting. This study underscores the significance of HMMs in various domains, highlighting their ability to effectively model complex systems with hidden states. "It is a useful tool for those looking to explore and utilize HMMs across various disciplines, pushing the boundaries of statistical modeling and analysis.

***Keywords**: Hidden Markov Models, Markov Chains, Stochastic process, Time series.*

1. Introduction

1.1 Overview of Markov Chains: Andrey Andreyevich Markov (1856–1922), a distinguished Russian mathematician, is well-known for his contributions to the field of stochastic processes, particularly Markov chains. He first introduced the concept of Markov chains in 1906, laying the groundwork for these stochastic processes and coining the term "chain." In 1913, Markov applied his theories to the analysis of letter sequences in the Russian language, driven by his mathematical curiosity. His pioneering work paved the way for the development of the stochastic Markov process, a key tool in statistical modeling. In the realm of probability theory and statistics, a Markov process is characterized by the Markov property, which means that the future state depends solely on the present state, with no influence from the events that occurred before it. This property, known as memorylessness, is essential for developing stochastic models, as it allows for a straightforward introduction of statistical dependence, where the strength of dependence decreases as the time lag increases.

1.2 Overview of the Development of Hidden Markov Models: Conventional Markov models have limitations as they link each state directly to a particular observation. To address this, the concept was extended to Hidden Markov Models (HMMs), which involve observations that are probabilistically dependent on the states. This creates a layered stochastic process, where the underlying hidden process is inferred indirectly through observable data. Leonard E. Baum and

others described HMMs in a series of statistical papers in the late 1960s. One of the earliest applications of HMMs was in speech recognition in the mid-1970s. In HMMs, the system state is not directly visible; instead, observable outputs are produced based on state-dependent probabilities. The term 'hidden' refers to the states of the Markov chain that cannot be directly observed, rather than to the model's parameters.

The evolution of HMMs has seen significant advancements across various fields. The foundational work by Baum and Petrie (1966) laid the groundwork for statistical inference in HMMs, establishing their theoretical underpinnings. Over time, the application of HMMs expanded, as seen in the work of Durbin et al. (1998), who highlighted their utility in biological sequence analysis, particularly in bioinformatics. Further methodological developments were made by Bartolucci and Farcomeni (2010), who integrated the Mixture Transition Distribution model with HMMs, enhancing time series analysis. The introduction of parsimonious structures for autoregressive HMMs by Alizadeh and Rezakhah (2013) contributed to a deeper understanding of their stability and moment behavior. In recent years, the versatility of HMMs has been demonstrated in diverse applications: Boyko and Beaulieu (2021) applied generalized HMMs in phylogenetic studies, while Gámiz et al. (2023) explored their potential in reliability engineering. Contemporary reviews, such as those by Deng and Söffker (2021), emphasize the relevance of HMMs in driving behavior prediction. Additionally, the work by Zucchini et al. (2017) provides an accessible guide to applying HMMs in various types of time series, reflecting the model's adaptability and enduring importance in modern research.

1.3 Overview of Algorithms for Developing HMMs

During the mid-20th century, Claude Shannon (1916–2001), an American mathematician and electrical engineer, made a significant impact with his work “A Mathematical Theory of Communication.” This paper highlighted the importance of integrating deterministic and stochastic processes in computing and electronics, laying the foundation for various developments in the field. Following this, several key algorithms were introduced to develop, apply, and better understand HMMs.

The Expectation-Maximization (EM) Algorithm: The EM algorithm plays a crucial role in determining the maximum likelihood estimates for parameters in probabilistic models that involve unobserved latent variables. This algorithm alternates between two steps: the expectation (E) step, where the expected likelihood is calculated as if the latent variables were known, and the maximization (M) step, which optimizes the likelihood obtained in the E step. The process is repeated iteratively until the algorithm converges. EM is extensively utilized in fields such as data clustering, machine learning, and computer vision. Its main advantage lies in its efficiency in handling incomplete data, though it can be sensitive to the initial parameter settings and may sometimes converge to local rather than global maxima.

The Baum-Welch Algorithm: As a specific example of the generalized expectation-maximization (GEM) algorithm, the Baum-Welch algorithm is employed to estimate the unknown parameters of an HMM through the forward-backward procedure. It is particularly effective in training HMMs and is commonly used in areas like speech recognition and bioinformatics. The algorithm's primary strength is its efficiency in training models with incomplete data. However, similar to the EM algorithm, it may be influenced by initial conditions and might not always find the global optimum.

The Viterbi Algorithm: Developed by Andrew Viterbi in 1967, this dynamic programming algorithm is designed to decode convolutional codes over noisy communication channels. The Viterbi algorithm identifies the most probable sequence of hidden states, known as the Viterbi path, that could result in a given sequence of observed events. It is widely applied in fields such as speech recognition, keyword spotting, computational linguistics, and bioinformatics. The main benefit of the Viterbi algorithm is its ability to efficiently determine the most likely sequence of states. However, it assumes that the model parameters are known and fixed, which may not always align with real-world scenarios.

Each of these algorithms has its strengths and limitations, making them suitable for different applications and scenarios within the context of HMMs.

2. Definitions and examples:

2.1. Markov chain: "A stochastic process $\{X_t\}_{t=0,1,2,\ldots}$ referred to as a Markov chain if it satisfies the Markov property, which is defined as

$$P\{X_t = k \mid X_{t-1} = j, X_{t-2} = j_1, \ldots, X_0 = j_{t-1}\} = P\{X_t = k \mid X_{t-1} = j\} = p_{jk} \text{ (say)},$$

for all $t \geq 0$ for all states $j, k, j_1, \ldots, j_{t-1} \in N$ and whenever the first member $P\{X_0 = j_{t-1}\}$ is defined".

Example 1 (Medhi, 1994): Consider a simple experiment where a coin is tossed repeatedly. Each trial has two possible outcomes: a head, which occurs with probability p, and a tail, which occurs with probability q, with $p + q = 1$. We can represent a head by 1 and a tail by 0, with the random variable X_t representing the result of the t-th toss. Therefore, for $t = 1,2,3, \ldots$,

$$P(X_t = 1) = p \text{ and } P(X_t = 0) = q.$$

This setup produces a sequence of random variables $X_1, X_2, \ldots$ Each trial is independent, meaning the outcome of the t -th trial does not depend on the outcomes of the previous trials. As a result, these random variables are independent of one another.

Now, consider the random variable S_t defined as the partial sum $S_t = X_1 + X_2 + \cdots + X_t$. The sum $S_t$represents the total number of heads observed in the first t trials, and it can take on values of 0, 1, 2…t. We have the relationship $S_{t+1} = S_t + X_{t+1}$ Given that $S_t = j$ (where j can be 0, 1, 2, … t), the random variable S_{t+1} has only two possible outcomes: $S_{t+1} = j$ with probability q, and $S_{t+1} = j + 1$ with probability p. These probabilities are independent of the outcomes of previous variables $S_1, S_2, \ldots, S_{t-1}$.

Thus,

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

This is a classic example of a Markov chain, where the outcome of the $(t + 1)$-th trial is directly dependent on the outcome of the t-th trial, and only that trial. The conditional probability of S_{t+1} given S_t depends solely on the value of S_t. In this context, the outcomes are referred to as the states of the Markov chain. If X_t results in the outcome j (i.e., $X_t = j$), the

process is considered to be in state j during the t-th trial. For a pair of states (j, k) at two successive trials—specifically the t-th and $(t+1)$-th trials—there is an associated conditional probability p_{jk}. This is the probability of transitioning from state j at the t-th trial to state k at the$(t+1)$-th trial. These transition probabilities p_{jk} are fundamental to the study of Markov chains.

2.2 Transition Probability Matrix (t.p.m.): The transition probability p_{jk} represents the likelihood of moving from state j at time $t-1$ to state k at time t. The probabilities p_{jk} satisfy the following conditions: $p_{jk} \geq 0$, and the sum of probabilities across all possible states for a given state j equals 1, i.e.,$\sum_{k=1}^{N} p_{jk} = 1$for all j, where N is the total number of states. These probabilities can b represented in matrix form as:

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

This matrix is known as the transition probability matrix (t.p.m.) of the Markov chain. The matrix P has non-negative elements and each row sums to one. The transition probability may either depend on or be independent of the time variable t. If the transition probability p_{jk} is constant over time, the Markov chain is said to be homogeneous (or to have stationary transition probabilities). If the transition probability varies with time, the chain is referred to as non-homogeneous.

Example 2 (Rabiner, 1989) : Consider a simple Markov model with three possible weather states:

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

Assume that the weather on any given day t is determined by one of these three states and that the weather on day t depends solely on the weather of the previous day (day $t-1$). The transition probabilities between these states can be represented by the following matrix $\boldsymbol{A}$:

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

In this matrix, each element p_{ij} represents the probability of transitioning from state i on day $t-1$ to state j on day t. For example, $p_{31}$is the probability that tomorrow will be rainy (State 1) given that today is sunny (State 3).

Given that the weather on day 1 $(t = 1)$ is sunny (State 3), we can calculate the probability that the weather for the next seven days follows the sequence "sunny-sunny-rainy-rainy-sunny-snowy-sunny." Define the sequence of observations as O = {S_3, S_3, S_1, S_1, S_3, S_2, S_3} corresponding to t = 1, 2, … , 8. The probability of observing this sequence, given the model, can be expressed as:

$$\mathrm{P(O \mid Model)} = \mathrm{P}(S_3) \times \mathrm{P}(S_3 \mid S_3) \times \mathrm{P}(S_3 \mid S_3) \times \mathrm{P}(S_1 \mid S_3) \times \mathrm{P}(S_1 \mid S_1) \times \mathrm{P}(S_3 \mid S_1) \times \mathrm{P}(S_2 \mid S_3) \times \mathrm{P}(S_3 \mid S_2)$$

This expands to P(O | Model) = $\pi_3 \times p_{33} \times p_{33} \times p_{31} \times p_{11} \times p_{13} \times p_{32} \times p_{23}$

Given the initial state probability $\pi_3 = 1$ (since we start with a sunny day), the probability calculation becomes:

P(O | Model) = $1 \times (0.8) \times (0.8) \times (0.1) \times (0.4) \times (0.3) \times (0.1) \times (0.2) = 1.536 \times 10^{-4}$.

This result represents the likelihood of observing the specified sequence of weather conditions over the given days. When the actual states are not directly observable, a HMM is employed to handle such scenarios.

2.3 A Hidden Markov Model (HMM): is a type of stochastic process that encompasses two interrelated elements. The first element consists of a finite set of states, each associated with its own probability distribution. The transitions between these states are governed by statistical probabilities known as transition probabilities. The second element involves the observable outputs, which are dependent on the underlying states. However, the actual states remain "hidden" from the observer, meaning that only the observable outputs can be analyzed. This characteristic gives rise to the term HMM.

2.3.1 Elements of an HMM:

A HMM is characterized by several key elements: states, observation symbols, initial probabilities, transition probabilities, and emission probabilities.

1. N represents the number of states in the model. If $S_1, S_2, \ldots, S_N$ are the possible hidden states of the HMM and X_t is the state at the time t, then X_t belongs to the set ($S_1, S_2, \ldots, S_N$)
2. M denotes the number of distinct observation symbols or values per state. If $v_1, v_2, \ldots, v_M$ are the observed values and O_t is the value at time *t*, then O_t belongs to the set $(v_1, v_2, \ldots, v_M)$.
3. π is the initial state distribution vector, represented as $\pi = [\pi_j]_{1\times N}$, where π_j is the probability that the initial state X_1 is S_j, for $1 \leq j \leq N$.
4. A is the state transition probability matrix, represented as, as $A = [p_{jk}]_{N\times N}$, where $p_{jk} = P(X_{t+1} = S_k \mid X_t = S_j)$, $1 \leq j, k \leq N$. The transition probabilities satisfy the constraints $p_{jk} \geq 0$ for $1 \leq j, k \leq N$ and the sum of the probabilities for each state equals 1: $\sum_{k=1}^{N} p_{jk} = 1$, $1 \leq j \leq N$.
5. B is the observation probability matrix , $B = [b_{jk}]_{N\times M}$, where b_{jk} is the probability that observation symbol v_k is produced in state S_j.
   $b_{jk} = P[O_t = v_k \mid X_t = S_j]$, $1 \leq i \leq N$, $1 \leq k \leq M$.
   The observation probabilities satisfy the constraints $b_{jk} \geq 0$, $1 \leq j \leq N$, $1 \leq k \leq M$ and $\sum_{k=1}^{M} b_{jk} = 1$, $1 \leq j \leq N$.

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. 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.

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

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 represent the random variable for the weather state at time t and o_t represent the random variable for the observable (Bob's mood) at time t.

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

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.

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

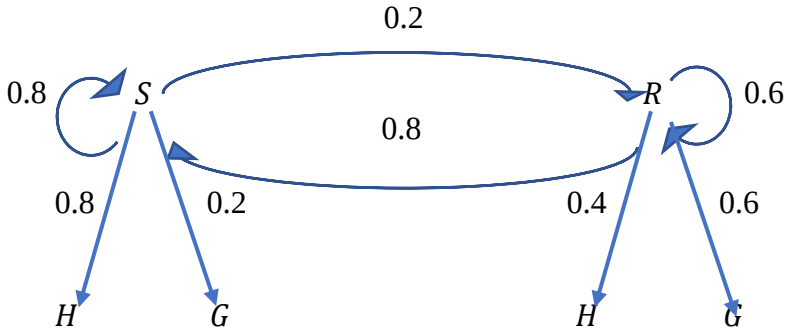
P $[o_t = H \mid q_t = R]$ = 2/5 = 0.4

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

Emission probability matrix:

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

HMM:



2.3.2. Uses of HMMs: HMMs are widely utilized in various areas of probabilistic modeling, offering powerful tools for applications such as regime identification in financial time series and disease stage identification. They are also employed in predicting tourism demand and have significant applications in the field of bioinformatics. For instance, HMMs are used in predicting protein secondary structures, modeling oscillatory patterns in nucleosomes, and representing site-specific evolutionary rates. Additionally, they are instrumental in incorporating evolutionary data into protein secondary structure prediction, further enhancing the accuracy and depth of biological analyses.

3. Application of HMMs in Estimating Misclassification Probabilities of Chronic Kidney Disease by Grover et al. (2019) is chiefly discussed below:

Chronic Kidney Disease (CKD) is characterized by various stages, which are determined based on the estimated glomerular filtration rate (eGFR). The eGFR is prone to computational and measurement errors, which can result in the misclassification of CKD stages. To address this, a study was conducted using data from 117 CKD patients collected between March 2006 and October 2016. The goal was to estimate the transition intensities and average times spent in each stage of the disease. A HMM was employed to model the progression of CKD across its stages. CKD is defined by the presence of kidney damage and a progressive decline in kidney function. The disease often presents no symptoms in its early stages, making accurate staging difficult due to the influence of other prognostic factors. The National Kidney Foundation (NKF) classifies CKD into five stages based on eGFR levels:

- Stage I: GFR $\geq$ 90 ml/min/1.73 m²
- Stage II: 60 $\leq$ GFR $\leq$ 89 ml/min/1.73 m²
- Stage III: 30 $\leq$ GFR $\leq$ 59 ml/min/1.73 m²
- Stage IV: 15 $\leq$ GFR $\leq$ 29 ml/min/1.73 m²
- Stage V: GFR < 15 ml/min/1.73 m²

Stages 1 through 4 are considered transient, meaning patients can move between these stages, while Stage 5 is an absorbing state, indicating a loss of kidney function that necessitates dialysis or transplantation. Since the true stages of the disease are not directly observable, eGFR values help infer the actual disease state. Given that CKD progression occurs over time with random transition points, a homogeneous continuous-time HMM was selected as an appropriate model.

The model estimated the transition rates between different disease stages and the probabilities of misclassifying the true disease stages. These estimates allowed for the calculation of the average duration a patient spends in each stage, known as the mean sojourn time. For instance,

the mean sojourn time in a given stage reflects how long, on average, a patient remains in that stage before advancing to the next.

The likelihood of transitioning between stages is represented by transition intensities (λ_{kl}), which are organized into a transition intensity matrix (Q).

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

The matrix Q's rows sum to zero, and its diagonal entries are determined by the negative sum of the off-diagonal entries ($i.e., \lambda_{kk} = -\sum_{l \neq k=1}^{5} \lambda_{kl}$). The transition intensities are estimated by maximizing the likelihood function using optimization techniques like Expectation-Maximization. The mean sojourn time for each stage is derived from these estimated transition intensities as $-\frac{1}{\lambda_{kk}}$.

Table 1 summarizes CKD patients' state transitions during follow-up visits, highlighting the irreversible nature of CKD, as no transitions from higher to lower stages are observed. Most patients remain in their initial stage, with only a few progressing to more advanced stages over time.

Table 1: Number of state transitions between CKD stages

	Stage I	Stage II	Stage III	Stage IV	Stage V
Stage I	75	8	4	3	0
Stage II	0	116	6	3	1
Stage III	0	0	293	29	2
Stage IV	0	0	0	145	51

Table 2 presents the estimated transition intensities between CKD stages. The diagonal values indicate the rate at which patients leave a stage, while off-diagonal values represent the likelihood of progressing to the next stage. For example, the transition intensity from Stage I to Stage II is 0.0599, while the transition from Stage IV to Stage V is higher at 0.564, reflecting more rapid progression in advanced stages.

Table 2: Estimated transition intensities between CKD stages

	Stage I	Stage II	Stage III	Stage IV	Stage V
Stage I	-0.098	0.0599	0.0266	0.0115	0
Stage II	0	-0.119	0.0725	0.0467	0
Stage III	0	0	-0.115	0.115	0
Stage IV	0	0	0	-0.564	0.564

Table 3 shows the mean sojourn times for each CKD stage, or the average time patients spend in each stage before advancing. Patients typically remain in the early stages (Stages I to III) for several years, with Stage I lasting around 10.2 years. However, in advanced stages like Stage IV, the mean sojourn time drops to just 1.77 years, indicating faster disease progression.

Table 3: Mean sojourn times at different CKD stages

	Mean Sojourn Time
Stage I	10.2030
Stage II	8.3829
Stage III	8.7051
Stage IV	1.7734

This analysis offers important insights for healthcare policymakers, helping to identify stages where early intervention could slow progression and reduce the economic burden of CKD.

4. Case Study: Modeling S&P 500 Index with HMMs:

Lihn (2017) analyzed financial time series data for the S&P 500 index, which tracks the stock performance of 500 major U.S. companies, using the R package 'ldhmm. Different market regimes were identified using the powerful tool of HMM. HMMs with two, three, four, five, and six states were fitted and analyzed. Models with more than five states were able to capture a higher degree of auto-correlation, aligning well with the data's patterns. While the stock market is generally classified into two main regimes—normal and crash—the analysis extended beyond these categories due to the evolving nature of HMM states. Since the stock market typically increases during periods of low volatility and declines during high volatility, HMM was utilized for identifying these regimes. Lihn (2017) studied the S&P 500 index from January 3, 2000 to May 30, 2017. In our study, we conducted a similar analysis for the S&P 500 index dataset spanning from January 4, 2000, to September 20, 2019. By extending the analysis period, the robustness and consistency of the regime identification through HMM were further evaluated, providing deeper insights into market behavior over nearly two decades.

As the number of states in the model increases, the kurtosis within each state generally decreases. This means that extreme outliers are more likely to be assigned to high-volatility states. The 'ldhmm' package is tailored specifically for analyzing the SPX index, providing daily closing prices for this purpose. The SPX daily returns exhibit two notable characteristics. Firstly, the kurtosis begins high (around 29.84) and decreases gradually as outliers are removed, with the kurtosis still around 8.43 after removing the 10 largest outliers.

```
> sapply(0:10,function(drop) kurtosis(ldhmm.drop_outliers(ts$x,drop)))
[1] 29.835672 12.107654 11.291203 10.712345 10.289324  9.867431
[7]  9.478321  9.142874  8.823451  8.594213  8.432987
```

Second, The ACF values for the first six lags range between 23% and 28%, with only a slight reduction in auto-correlation after the 10 largest outliers are excluded.

```
> ldhmm.ts_abs_acf(ts$x, drop=0, lag.max=6)
[1] 0.2415678 0.2712394 0.2487632 0.2396543 0.2790211 0.2356712
> ldhmm.ts_abs_acf(ts$x, drop=10, lag.max=6)
[1] 0.2243451 0.2448792 0.2278453 0.2345678 0.2501245 0.2258964
```

Two states model: Our analysis began by examining a two-state model, where the stock market is divided into two distinct regimes: a normal regime characterized by stability and growth, and a crash regime associated with market downturns. The normal regime, often referred to as a

"bull market," represents periods of rising prices, while the crash regime, or "bear market," captures periods of significant decline. Empirical data suggests that the market spends a majority of the time in the normal regime.

Theoretical Statistics for Each State:

```
> ldhmm.ld_stats(hd)
              mean         sd kurtosis
[1,]  0.0005734620 0.00635735        3
[2,] -0.0007012345 0.01598762        3
```

In this model, the first state, representing the normal market condition, has an expected mean of approximately 0.000573 and a standard deviation of around 0.00620. The second state, representing the crash scenario, is expected to have a mean of about -0.000701 and a standard deviation of around 0.01599, with both states having an initial kurtosis value of 3.

Empirical Statistics for Each State:

```
> hd@states.local.stats
              mean          sd  kurtosis    skewness length
[1,]  0.0005886734 0.006178945  3.675934 -0.01234567  13200
[2,] -0.0008124567 0.016234678 13.845678 -0.65432109   3680
```

When examining the data classified into these states, the first state is expected to show a mean of approximately 0.000589, a standard deviation of 0.00618, and a kurtosis of 3.68. The second state may display a mean of around -0.000812, a standard deviation of 0.01623, and a higher kurtosis of approximately 13.85. The higher kurtosis in the second state suggests the presence of significant outliers that are not fully captured by a two-state model.

To address this, we may remove the most extreme outliers and recalculate the statistics. This process, known as "asymptotic statistics," helps adjust the kurtosis and reduce the skewness of the second state.

Adjusted Statistics After Removing Outliers:

```
> ldhmm.calc_stats_from_obs(hd, drop=11)
              mean         sd kurtosis     skewness length
[1,]  0.0005912345 0.00615478 3.542123 -0.010987654  13189
[2,] -0.0006782345 0.01549876 4.123456  0.012345678   3669
```

After excluding the 11 largest outliers, the kurtosis in the second state is expected to reduce to about 4.12, with skewness also decreasing. The mean and standard deviation in each state remain relatively consistent with their initial values.

The interpretation of the two-state HMM results indicates that the first state represents a typical market condition where average returns are positive, signifying a stable or growing market. In contrast, the second state is associated with a market downturn, characterized by negative average returns. The standard deviation in the crash state is anticipated to be more than double that of the normal state, highlighting the significantly higher volatility during periods of market decline. Moreover, the initial state probability vector suggests that the market is expected to spend approximately 77% of the time in the normal state, reflecting the overall stability of the market despite occasional downturns.

```
> hd@delta
[1] 0.7723456 0.2276544
```

Regime Identification and Model Expansion: Regime identification in this context classifies the market into regimes that align with intuitive market phases. The low-volatility state corresponds to bull markets, while the high-volatility state aligns with bear markets. As the analysis is expanded to include 3, 4, 5, and 6 states, careful parameter tuning, especially regarding standard deviation, helps the maximum likelihood estimator (MLE) converge effectively. The optimal

model is expected to occur at five states, with the MLLK for the two-state model around -56123.45.

```
> ldhmm.mllk(hd, ts$x)
[1] -56123.45
```

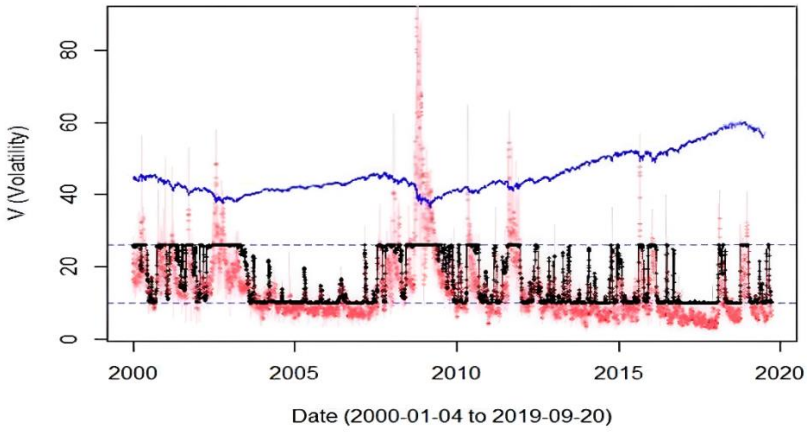


Figure 1: Expected vs. Realized Volatility in Two-State HMM

Figure 1 compares the expected volatility $V(t)$ from a two-state HMM with the realized volatility derived from the Oxford-Man realized variance dataset. The two-state HMM, represented by black lines, is the simplest model to differentiate between normal and crash market regimes. The red line shows the daily realized volatility, while red dots indicate the 5-day moving average. The dashed blue lines depict the volatilities corresponding to the two HMM states, and the solid blue line represents the rescaled SPX price index. This visualization effectively contrasts the model's expected volatility against the actual market fluctuations, with data sourced from the interdisciplinary Oxford-Man Research Institute at the University of Oxford.

Three states model: The three-state HMM yields an improved MLLK score compared to the two-state HMM, with an MLLK value of -56634.92.

```
> ldhmm.mllk(hd,ts$x)
[1] -56634.92
```

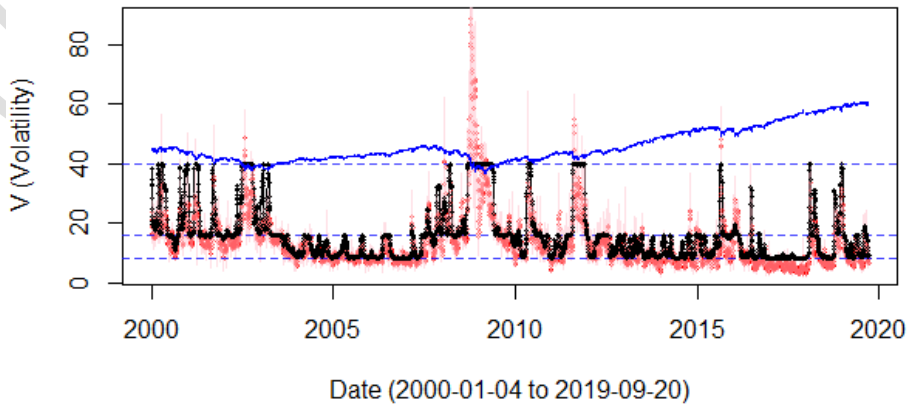


Figure 2: Expected vs. Realized Volatility in Three-State HMM

Figure 2 illustrates the comparison between expected volatility $V(t)$ from a three-state HMM and the realized volatility. The black lines represent the expected volatility from the three-state HMM, while the red line shows the realized volatility. The three-state model introduces an intermediate or transition state, providing deeper insights into market behavior. In this model, bull markets are characterized by prolonged periods in states 1 and 2. Conversely, bear markets are identified by oscillations between states 2 and 3. The transition into a bull market may be marked by the first entry into state 1, while the transition into a bear market is indicated by the first entry into state 3, signaling the end of the previous bull market.

Four states model: The four-state HMM achieves a further improvement in the MLLK score over the three-state model, with an MLLK value of -56892.87.

```
> ldhmm.mllk(hd,ts$x)
[1] -56892.87
```

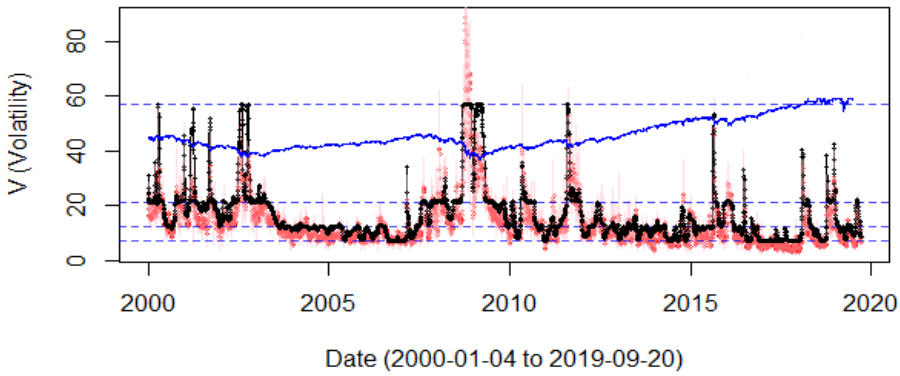


Figure 3: Expected vs. Realized Volatility in Four-State HMM

Figure 3 compares the expected volatility $V(t)$ from a four-state HMM with the realized volatility, where the black lines represent the expected volatility from the HMM, and the red line shows the realized volatility. In this model, the third state serves as a transition phase. If the market shifts from this state back to states 1 or 2, the bull market continues. However, if the transition moves to state 4, the market enters a crash state, leading to sharp and severe downward movements.

Five states model: In the five-state HMM, the MLLK score shows a slight improvement, reaching -56979.89. This five-state model offers several notable advantages: it is the most optimized model from an MLLK perspective, making it a strong choice for model selection. Additionally, the excess kurtosis across the states is more evenly distributed. The model identifies two low-volatility states, corresponding to bull markets (states 1 and 2), a transition state (state 3), and two high-volatility states, associated with bear markets (states 4 and 5). The visualization of the model:

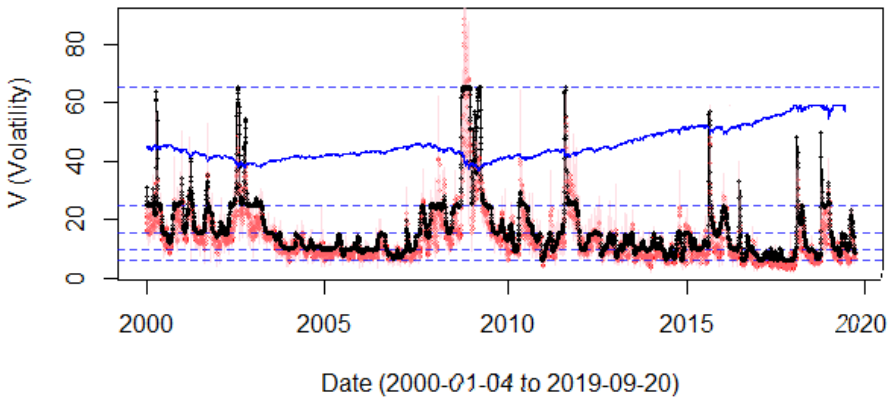


Figure 4: Expected vs. Realized Volatility in Five-State HMM

Figure 4 displays the comparison between the expected volatility V(t) from a five-state HMM and the realized volatility, where the black lines represent the expected volatility from the HMM, and the red line shows the realized volatility.

In this model, bull markets are characterized by the market consistently remaining in the two low-volatility states, states 1 and 2. Conversely, bear markets are marked by oscillations between states 3 and 4. The fifth state is identified as highly destructive, signaling extreme market downturns.

Six states model: The AIC and BIC scores cease to improve, with the MLLK stabilizing at -56954.12. As a result, it is concluded that the five-state model is the most suitable fit. This model categorizes the stock market into five distinct regimes: two low-volatility states (states 1 and 2, representing a bull market), one transition state (state 3), and two high-volatility states (states 4 and 5, indicative of a bear market). This model can be effectively utilized for forecasting volatility and understanding market behavior to inform strategic planning.

5. Conclusions:

This study provides a detailed exploration of HMMs, emphasizing their theoretical foundations, key algorithms, and diverse applications. HMMs are invaluable for capturing probabilistic relationships within sequential data, making them essential tools in various fields such as bioinformatics, speech recognition, financial time series analysis, and disease progression modeling. The key algorithms discussed, including the Expectation-Maximization (EM) algorithm, the Baum-Welch algorithm, and the Viterbi algorithm, are pivotal in estimating model parameters, training HMMs, and decoding sequences, respectively.

Significant applications were examined in detail. For Chronic Kidney Disease (CKD), HMMs provided a framework to model disease progression, estimate transition intensities, and understand sojourn times across various stages. This approach enhances the understanding of the disease's natural history and aids policymakers in devising effective treatment strategies.

In the financial time series case study, we analyzed the S&P 500 index dataset from January 4, 2000, to September 20, 2019. This extended analysis period, compared to Lihn's study, allowed for a robust evaluation of regime identification through HMMs, offering deeper insights into market behavior over nearly two decades. HMMs identified different market regimes, revealing

the underlying dynamics of bull and bear markets. The study demonstrated the superiority of a five-state HMM model in capturing market behaviors, thus aiding in volatility forecasting.

Overall, this study underscores the significance of HMMs in various domains, highlighting their ability to effectively model complex systems with hidden states. The practical applications discussed showcase the versatility and effectiveness of HMMs in providing insights into sequential data patterns. This study serves as a valuable resource for researchers and practitioners aiming to leverage HMMs in their respective fields, promoting further advancements in statistical modeling and inference.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc have been used during writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1. Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

2.

3.

References:

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

Bartolucci, F. and 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. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The annals of mathematical statistics*, **37(6)**, 1554-1563.

Boyko, J. D. and Beaulieu, J. M. (2021). Generalized hidden Markov models for phylogenetic comparative datasets. Methods in Ecology and Evolution, **12(3)**, 468-478.

Deng, Q. and Söffker, D. (2021). A review of HMM-based approaches of driving behaviors recognition and prediction. *IEEE Transactions on Intelligent Vehicles*, **7(1)**, 21-31.

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

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

Gámiz, M. L., Limnios, N. and del Carmen Segovia-García, M. (2023). Hidden Markov models in reliability and maintenance. *European Journal of Operational Research*, **304(3)**, 1242-1255.

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

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

Glennie, R., Adam, T., Leos-Barajas, V., Michelot, T., Photopoulou, T. and McClintock, B. T. (2023). Hidden Markov models: Pitfalls and opportunities in ecology. *Methods in Ecology and Evolution*, **14(1)**, 43-56.

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. and Jahromi, V. G. (2018). Using discrete hidden Markov model for modelling and forecasting the tourism demand in Isfahan. *Journal of Information Systems and Telecommunication*, **6**, 112-118.

Kumsa, T. H., Mulu, A., Beyene, J. And Asfaw, Z. G. (2024). Multi-state Markov model for time to treatment changes for HIV/AIDS patients: a retrospective cohort national datasets, Ethiopia. *BMC Infectious Diseases*, **24(1)**, 627.

Kouadri, A., Hajji, M., Harkat, M. F., Abodayeh, K., Mansouri, M., Nounou, H. and Nounou, M. (2020). Hidden Markov model based principal component analysis for intelligent fault diagnosis of wind energy converter systems. *Renewable Energy*, **150**, 598-606.

Kouemou, G. L and 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. and IDA, S. (2008). Hidden Markov models applied in agricultural crops classification. *Proceeding of GEOBIA (GEOgraphic Object-Based Image Analysis for the 21St Century)*.

Li, M., Yang, M., Yu, Y. and Lee, W. J. (2021). A wind speed correction method based on modified hidden Markov model for enhancing wind power forecast. *IEEE Transactions on Industry Applications*, **58(1)**, 656-666.

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

Martins, A., Fonseca, I., Farinha, J. T., Reis, J. and Cardoso, A. J. M. (2021). Maintenance prediction through sensing using hidden markov models—A case study. *Applied Sciences*, **11(16)**, 7685.

Medhi, J. (1994). *Stochastic processes*. New Age International.

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

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

Ruiz-Suarez, S., Leos-Barajas, V. and Morales, J. M. (2022). Hidden Markov and semi-Markov models when and why are these models useful for classifying states in time series data?. *Journal of Agricultural, Biological and Environmental Statistics*, 1-25.

Roy, P. P., Kumar, P. And Kim, B. G. (2021). An efficient sign language recognition (SLR) system using Camshift tracker and hidden Markov model (hmm). *SN Computer Science*, **2(2)**, 79.

Sipos, I. R., Ceffer, A., Horváth, G. and Levendovszky, J. (2019). Parallel MCMC sampling of AR-HMMs for prediction based option trading. *Algorithmic Finance*, **8(1-2)**, 47-55.

Tur, A. O. and Keles, H. Y. (2021). Evaluation of hidden markov models using deep cnn features in isolated sign recognition. *Multimedia Tools and Applications*, **80**, 19137-19155.

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory, 13(2), 260-269.

Zucchini, W., MacDonald, I. L. and Langrock, R. (2017). *Hidden Markov models for time series: an introduction using R*. Chapman and Hall/CRC.