Identifying Market Regimes Using Hidden Markov Models

Presentation for the Mathematics of Artificial Intelligence and Machine Learning at the University of Denver

Evans Hedges

March 6th 2023

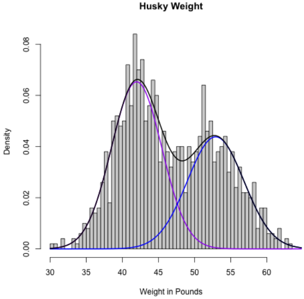
Outline

- Mixture Model Recap
- Markov Chain Theory
  - The Markov Property
  - Markov Chains
- Hidden Markov Models
  - Parameter Estimation of HMMs
  - Implementation Example: Identifying Market Regimes in the S&P 500

Mixture Model Recap

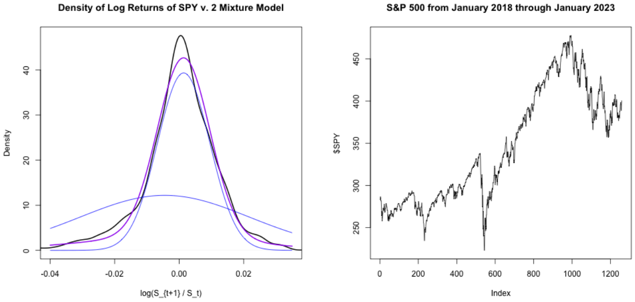
General Mixture Model:

$$f(x|\boldsymbol{\theta}) = \sum_{i=1}^{N} \pi_i f_i(x|\theta_i)$$



Mixture Model Recap

$$X = 0.811N(0.001387, 0.008114) + 0.189N(-0.004486, 0.02617)$$



Preliminary Theory: The Markov Property

A stochastic process is said to have the **Markov Property** if for any sequence of states $S_{1:t}$, the current state depends only on the previous state. i.e.:

$$P(S_t|S_{1:t-1}) = P(S_t|S_{t-1})$$

Example: A board game with dice. The next state of the board (S_t) depends only on the current state of the board (S_{t-1}) and the roll of the dice.

Preliminary Theory: Markov Chain

A (finite) Markov Chain can be described by:

- A finite collection of states $\mathcal{A}$ and
- A stochastic transition matrix A, and
- An initial distribution π.

Preliminary Theory: Markov Chain

- A finite collection of states $\mathcal{A}$

Our collection of states $\mathcal{A}$ denote the possible states our system can occupy (e.g. English language words or possible arrangement of pieces in a board game).

Preliminary Theory: Markov Chain

- A stochastic transition matrix A

The transition matrix $A = (a_{i,j})$ is an $n \times n$ stochastic matrix that denotes the probabilities of transitioning between states, defined such that

$$a_{i,j} = P(S_t = j | S_{t-1} = i)$$

By definition of a stochastic matrix we know $a_{i,j} \geq 0$ for all i, j, and $\sum_{j=1}^{n} a_{i,j} = 1$.

Preliminary Theory: Markov Chain

- An initial distribution π.

The initial distribution $\pi \in \mathbb{R}^n$ denotes the initial probability distribution of the states.
I.e., You will find yourself in state $i \in \mathcal{A}$ at time $t = 1$ with probability π_i.

Since π is a probability vector, we know $\pi_i \geq 0$ for all i and $\sum_{i=1}^{n} \pi_i = 1$.

Example: If we are using Markov Chains to model English sentences, π_i would denote the probability that a sentence would start with the ith word in our enumeration.

Preliminary Theory: Markov Chain

Given our initial distribution π at $t = 1$, the probability distribution of the states at $t = 2$ is exactly

$$A^T \pi$$

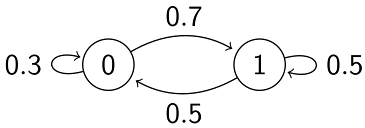
$$(A^T \pi)_i = \sum_{j=1}^{n} A^T_{i,j} \pi_j = \sum_{j=1}^{n} \pi_j A_{j,i}$$

$$= \sum_{j=1}^{n} P(S_1 = j) P(S_2 = i | S_1 = j) = P(S_2 = i)$$

Preliminary Theory: Markov Chain

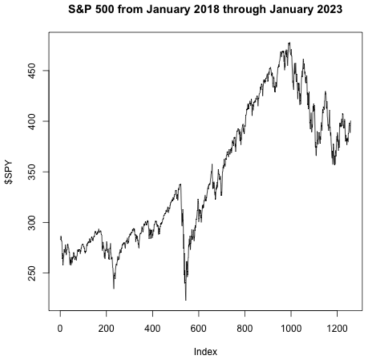
Example: 2 states with probability transition matrix:

$$A = \begin{bmatrix} 0.3 & 0.7 \\ 0.5 & 0.5 \end{bmatrix}$$



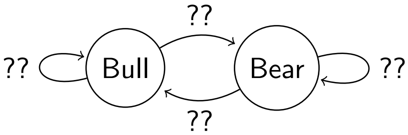
The Problem

In our example, we would like to determine whether or not the S&P 500 is currently in a "Bull" or "Bear" market, and what that means for our expected returns moving forward.



Framing as Markov Chain

We will suppose whether or not a given day is in a Bull or Bear market has the Markov Property with some probability transition matrix:



And for each state, we have some distribution of returns we can expect for the day.

$$f(x|Bull) = ?? \qquad f(x|Bear) = ??$$

The Model

Combining our Markov Model describing our states with a mixture model representing the probability distributions of each of those states, we have the joint probability of our dataset:

$$f(y_{1:T}, S_{1:T}) = P(S_1) f(y_1 | S_1) \times \prod_{t=2}^{T} P(S_t | S_{t-1}) f(y_t | S_t)$$

The Model

We can consider our hidden Markov Model consisting of three parts:

- Prior Model: $P(S_1|\theta_{pr})$ - the prior probabilities of each state given our model parameters.
- Transition Model: A model of $P(S_t = j|\theta_{tr}, \ S_{t-1} = i)$.
- Observation Model: A model of $P(y_t|\theta_{obs}, \ S_t = i)$.

Likelihood

The likelihood of some given parameters θ with observed data $y_{1:T}$ is defined as:

$$L(\theta|y_{1:T}) = \sum_{S_{1:T} \in \mathcal{S}^T} f(y_{1:T}, S_{1:T}|\theta)$$

$$= \sum_{S_{1:T} \in \mathcal{S}^T} \left(P(S_1) f(y_1|S_1) \times \prod_{t=2}^{T} P(S_t|S_{t-1}) f(y_t|S_t) \right)$$

$|\mathcal{S}^T| = 2^{1258} \approx 5 \times 10^{378}$.
The fastest supercomputer in the world today would take $\sim 10^{737}$ years to compute - comfortably into the black hole era of the universe.

Likelihood

Instead of looking at all possible paths through the state space, let's only consider the terminal state:

$$L(\theta|y_{1:T}) = \sum_{i=1}^{n} f(y_{1:T}, S_T = i|\theta)$$

Define the Forward Variables for all $1 \leq t \leq T$, $1 \leq i \leq n$ as:

$$\alpha_t(i) = f(y_{1:t}, S_t = i)$$

Computing Forward Variables

Initialize α_1 for $t = 1$ as follows:

$$\alpha_1(i) = f(y_1, S_t = i) = P(S_1 = i) f(y_1 | S_1 = i)$$

And remember: $P(S_1 = i), f(y_1 | S_1 = i) \in \theta$

Computing Forward Variables

For $t = 2$:

$$\alpha_2(i) = f(y_{1:2}, S_2 = i) \quad \text{Definition}$$

$$= \sum_{j=1}^{N} f(y_{1:1}, y_2, S_1 = j, S_2 = i) \quad \text{Sum over previous states}$$

$$= \sum_{j=1}^{N} f(y_{1:1}, S_1 = j) f(y_2, S_2 = i | S_1 = j) \quad \text{Bayes'}$$

$$= \sum_{j=1}^{N} \alpha_1(j) P(S_2 = i | S_1 = j) f(y_2 | S_2 = i) \quad \text{Definition and Bayes'}$$

Computing Forward Variables

In general for $t > 1$ compute:

$$\alpha_t(i) = \sum_{j=1}^{N} \alpha_{t-1}(j) P(S_t = i | S_{t-1} = j) f(y_t | S_t = i)$$

Likelihood Revisited

Key Equations

$$\alpha_1(i) = P(S_1 = i) f(y_1 | S_1 = i)$$

$$\alpha_t(i) = \sum_{j=1}^{N} \alpha_{t-1}(j) P(S_t = i | S_{t-1} = j) f(y_t | S_t = i)$$

$$L(\theta | y_{1:T}) = \sum_{i=1}^{N} \alpha_T(i)$$

$$P(S_t = i | S_{t-1} = j), P(S_t = i), f(y_t | S_t = i) \in \theta$$

Likelihood Revisited

In other notation:

$$\alpha_1(i) = \pi_i f_i(y_1)$$

$$\alpha_t(i) = \sum_{j=1}^{N} \alpha_{t-1}(j)(a_{j,i}) f_i(y_t)$$

$$L(\theta | y_{1:T}) = \sum_{i=1}^{N} \alpha_T(i)$$

$$\pi_i, f_i, (a_{j,i}) \in \theta$$

Likelihood Revisited

We can now use our Expectation Maximization algorithm to find θ!

$$L(\theta|y_{1:T}) = \sum_{i=1}^{N} \alpha_T(i)$$

Practical/Implementation Problem: Our forward variables α_t tend to zero relatively quickly when T is large. This can lead to precision difficulties in practical applications.

This can be addressed using scaled forward variables which we will not go into detail on today.

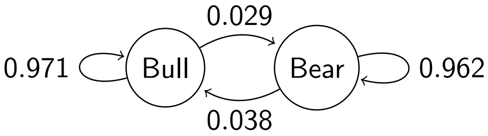
Code Implementation

```
spy_df = read.csv(file='SPY.csv', header=TRUE)
log_return = log(spy_df[['Close']][2:length(spy_df[['Close']])] /
                 spy_df[['Close']][1:length(spy_df[['Close']])-1])

model = hmm(data=log_return, nstates=2)
summary(model)
```

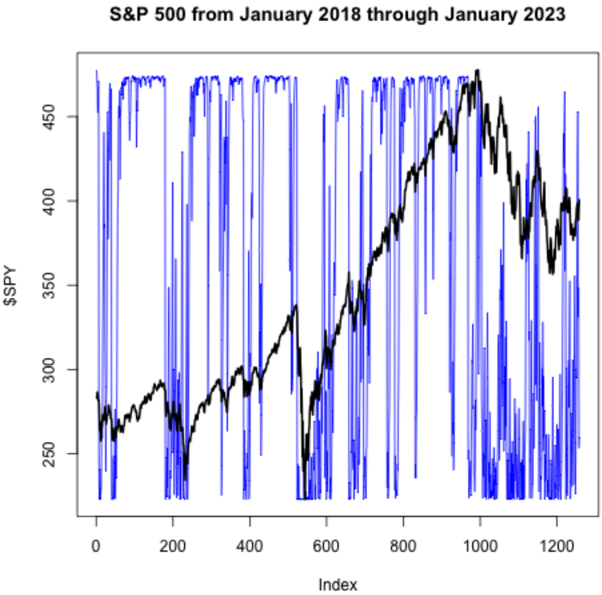
Results

$$\pi = (1, 0)$$



$$f(x|Bull) = N(0.01, 0.006) \qquad f(x|Bear) = N(-0.01, 0.02)$$

Historical Filtering Probabilities



Actionable Insights

On the last day of the dataset, the model said there was a 14.5% chance of being in a Bull market. So our state vector is:

$$\pi = \begin{bmatrix} .145 \\ .855 \end{bmatrix}$$

We can apply our model's transition matrix to get our estimate for the state we will be in the next day:

$$A^T\pi = \begin{bmatrix} 0.971 & 0.038 \\ 0.029 & 0.962 \end{bmatrix} \begin{bmatrix} .145 \\ .855 \end{bmatrix} = \begin{bmatrix} .173 \\ .826 \end{bmatrix}$$

Actionable Insights

Finally, applying our mixture model we have our estimated pdf for tomorrow's returns:

$$f(x) = 0.173N(0.001, 0.006) + 0.826N(-0.001, 0.02)$$

So, for tomorrow's returns we estimate:

$$E[\text{Return}] \approx -0.07\%$$

$$P(\text{Negative Returns}) \approx 50.5\%$$

$$P(\text{Returns } < -2\%) \approx 31\%$$

(Actual Return: -0.1%)

Some More Realistic Applications

- Given strategies A and B, identify market regimes of various performance profiles of the strategies. Use this information to allocate capital efficiently.
- Identify momentum vs. mean reverting regimes on various timescales.
- Identify different pricing model regimes (Black-Scholes Implied Volatility coefficient for options pricing)
- Increase the dimensionality of the return data to develop a more refined model of various "regimes." e.g.:
  - intra time period volatility
  - correlation levels between asset classes
  - correlation levels within asset classes, etc

Thank You

References:

- "Mixture and Hidden Markov Models with R" - Visser and Speekenbrink, 2022.
- https://github.com/depmix/hmmr

Appendix

In the slides below there is more information related to Markov Chains. In particular there are two statements of the Perron-Frobenius theorem as well as a number of statements related to the ergodicity and mixing properties of Markov Chains in general.

Perron-Frobenius 1

Theorem

Perron-Frobenius: If $A \geq 0$ is a real matrix with non-negative coefficients, then

- *A has a unique largest eigenvalue λ_A and*
- *the corresponding eigenvector v_A can be chosen with strictly positive components: $v_A > 0$.*

By applying Perron-Frobenius to A^T and defining $p = \frac{v_A}{\sum_{i=1}^{d}(v_A)_i}$, we get:

$$A^T p = p$$

We call p a stationary state for the Markov chain.

Perron-Frobenius 2

Definition

We call a non-negative matrix A **irreducible** if for each pair of indices i, j there exists some $k \geq 0$ such that $A^k_{i,j} > 0$.

Theorem

Perron-Frobenius: If $A \geq 0$ is an irreducible matrix then:

- *A has a positive eigenvector v_A with corresponding eigenvalue $\lambda_A > 0$*
- *λ_A is both geometrically and algebraically simple.*
- *If μ is another eigenvalue for A, then $|\mu| \leq \lambda_A$.*
- *Any positive eigenvector for A is a positive multiple of v_A.*

Stationary State and Perron-Frobenius

If we assume our stochastic transition matrix A is irreducible, then a stationary state p is unique due to the geometric simplicity of λ_A.

In other words, an irreducible Markov Chain has a unique stationary state distribution.

Some Facts (without proof)

Fact

A stochastic matrix A is irreducible if and only if for every i, j:

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} A^k_{i,j} = p_j$$

Note this means for any initial distribution π:

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} ((A^T)^k \pi)_j = \lim_n \frac{1}{n} \sum_{k=0}^{n-1} \sum_{i=1}^{d} (A^T)^k_{j,i} \pi_i$$

$$= \sum_{i=1}^{d} \pi_i \lim_n \frac{1}{n} \sum_{k=0}^{n-1} A^k_{i,j} = \sum_{i=1}^{d} \pi_i p_j = p_j$$

Some Facts (without proof)

Fact

Using a Markov measure μ, the Markov Shift (σ, μ) is ergodic if and only if the transition matrix A is irreducible.

Fact

A Markov Shift (σ, μ) is strong mixing if and only if A is aperiodic.

Fact

A is aperiodic if and only if

$$\lim_n A^n_{i,j} = p_j$$

For strong mixing Markov Shifts, for any initial distribution π:

$$\lim_n (A^T)^n \pi = p$$

Stationary State and Perron-Frobenius

For our example earlier where

$$A^T = \begin{bmatrix} 0.3 & 0.5 \\ 0.7 & 0.5 \end{bmatrix}$$

A is irreducible and aperodic, so we know the stationary state p is unique and the limit of any initial distribution π.

$$p = \begin{bmatrix} 5/12 \\ 7/12 \end{bmatrix} \approx \begin{bmatrix} 0.42 \\ 0.58 \end{bmatrix}$$