HIDDEN MARKOV MODELS IN INDEX RETURNS

JUSVIN DHILLON

ABSTRACT. In this paper, we consider discrete Markov Chains and hidden Markov models. After examining a common algorithm for estimating matrices of probability, we conclude by constructing a hidden Markov model of S&P 500 index returns.

CONTENTS

1. MARKOV CHAINS: BASIC PRINCIPLES

Definition 1.1. Let $(E, \mathcal{E})$ be a measurable space and Ω be a sample space with σ-algebra $\mathcal{F}$ and probability measure $P : \mathcal{F} \to [0,1]$. A *random variable* is an $(\mathcal{F}, \mathcal{E})$-measurable function $X : \Omega \to E$.

We consider in this paper a useful and intuitive collection of these functions.

Definition 1.2. A *discrete-time Markov chain* on a state space Ω is an Ω-valued sequence of random variables $X_1, X_2, X_3, ...$ such that for all times k in $\mathbb{N}$ and states $x_1, x_2, x_3, ...$

$$P(X_k = x_k | X_{k-1} = x_{k-1}, X_{k-2} = x_{k-2}, ...X_1 = x_1) = P(X_k = x_k | X_{k-1} = x_{k-1})$$

Thus the chain's next move depends only on the most recent state.

Example 1.3. The simple random walk on integer lattice $\mathbb{Z}^n$ with standard vectors $e_1, e_2, ...e_n$ is a Markov chain satisfying

$$P(X_k = \alpha | X_{k-1} = \alpha \pm e_i) = \frac{1}{2n} \qquad \forall \alpha \in \mathbb{Z}^n$$

Definition 1.4. If Ω is finite with $|\Omega| = n$, the chain can be represented by an $n \times n$ *transition matrix* $\boldsymbol{P}$ where $p(i,j) = P(X_k = j | X_{k-1} = i)$.

Date: August 30, 2010.

$$\begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,j} & \cdots \\ p_{2,1} & p_{2,2} & \cdots & p_{2,j} & \vdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ p_{i,1} & p_{i,2} & \cdots & p_{i,j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{pmatrix}$$

If the chain starts at state $X_1 = j$ and e_j denotes the standard unit vector, then $e_j\mathbf{P}$ indicates naturally the probability distribution of X_2 given that $X_1 = j$. In general, if $|\Omega| = n$ and $\mu \in \mathbb{R}^n$ with $\sum_{i=1}^n \mu_i = 1$ and $\mu_i \geq 0$ for all i, then $\mu\mathbf{P}^k$ is the probability distribution of the process after k transitions starting at μ. We denote $P(X_{i+k} = y|X_i = x)$ by $p^k(x, y)$, the $(x, y)th$ entry of $\mathbf{P}^k$.

Definition 1.5. A probability distribution π is *stationary* if $\pi = \pi\mathbf{P}$.

A stationary distribution thus represents an equilibrium for the random process. We prove the existence of this equilibrium in the next section.

2. Krylov-Bogoliubov Argument

Definition 2.1. A chain P is *irreducible* if for any two states $x, y \in \Omega$ there exists $\alpha \in \mathbb{Z}$ such that $p^\alpha(x, y) > 0$.

Thus states in an irreducible chain are always accessible from any given state. In our $\mathbb{Z}^n$ random walk, a point in the lattice can be reached from any other, so that the walk is irreducible.

Definition 2.2. The *period* of state x is gcd $\{n \in \mathbb{N}|p^n(x, x) > 0\}$. If every state in a chain has period 1 the chain is *aperiodic.* Otherwise, it is *periodic.*

Proposition 2.3. *If chain P is irreducible, then for all $x, y \in \Omega$*

$$gcd\,\{n \in \mathbb{N}|p^n(x, x) > 0\} = gcd\,\{n \in \mathbb{N}|p^n(y, y) > 0\}$$

Proof. Given states x and y, there exist i and $j \in \mathbb{Z}$ such that $p^i(x, y) > 0$ and $p^j(y, x) > 0$. Let $\{n \in \mathbb{N}|p^n(\alpha, \alpha) > 0\} = \mathcal{G}(\alpha)$ and let $k = i + j$.

Then $k \in \mathcal{G}(x) \cap \mathcal{G}(y)$, implying $\mathcal{G}(x) \subset \mathcal{G}(y)$ - $\{k\}$. Thus $gcd\mathcal{G}(y)$ divides all elements of $\mathcal{G}(x)$ and

$$gcd\mathcal{G}(y) \leq gcd\mathcal{G}(x)$$

But $k \in \mathcal{G}(x) \cap \mathcal{G}(y)$ also implies that $\mathcal{G}(y) \subset \mathcal{G}(x)$ - $\{k\}$ so that

$$gcd\mathcal{G}(x) \leq gcd\mathcal{G}(y)$$

□

All possible probability vectors μ axiomatically satsify $\sum_{i=1}^n \mu_i = 1$ and $\mu_i \geq 0$ for all i. Thus the set $\mathcal{P}$ of probablity distributions on Ω with $|\Omega| = n$ is the standard N-simplex in $\mathbb{R}^n$. Containing $\mathcal{P}$ in the $(n-1)$-sphere with radius 2 and observing that convexity connects any 2 points by a (closed) line, we see that $\mathcal{P}$ is bounded, closed, and hence compact by the *Heine-Borel Theorem.* In the following proof, we will use the *Bolzano-Weierstrass* property of compact spaces to extract our convergent subsequence.

Theorem 2.4. *Krylov-Bogoliubov Argument: If Markov chain P is finite-state and discrete, then there exists π in $\mathcal{P}$ with $\pi = \pi\boldsymbol{P}$.*

Proof. Fix φ in $\mathcal{P}$ and consider *Cesàro average* $\frac{1}{n}\sum_{i=1}^{n}\varphi\mathbf{P}^i$.

As the Cesàro average is an arimethic average of probability vectors, its components remain positive and sum to $\frac{n}{n}$. Thus Cesàro averages remain in $\mathcal{P}$.

Consider $\left\{\frac{1}{n}\sum_{i=1}^{n}\varphi\mathbf{P}^i\right\}_{n=1}^{\infty}$ with convergent subsequence $\left\{\frac{1}{n_k}\sum_{i=1}^{n_k}\varphi\mathbf{P}^i\right\}_{n_k=1}^{\infty}$ and denote, albeit prophetically, π as the limit of the subsequence. Observing the continuity of linear map $\mathbf{P}$,

$$\begin{aligned}
\pi\mathbf{P} &= \lim_{k\to\infty}\frac{1}{n_k}\sum_{i=1}^{n_k}\varphi\mathbf{P}^i\cdot\mathbf{P} \\
&= \lim_{k\to\infty}\frac{1}{n_k}\sum_{i=2}^{n_k+1}\varphi\mathbf{P}^i \\
&= \lim_{k\to\infty}\frac{1}{n_k}\left(\sum_{i=1}^{n_k}\varphi\mathbf{P}^i + \varphi\mathbf{P}^{n_k+1} - \varphi\mathbf{P}\right) \\
&= \lim_{k\to\infty}\frac{1}{n_k}\left(\sum_{i=1}^{n_k}\varphi\mathbf{P}^i\right) + \lim_{k\to\infty}\frac{1}{n_k}\left(\varphi\mathbf{P}^{n_k+1} - \varphi\mathbf{P}\right) \\
&= \lim_{k\to\infty}\frac{1}{n_k}\left(\sum_{i=1}^{n_k}\varphi\mathbf{P}^i\right) \\
&= \pi
\end{aligned}$$

□

3. Hidden Markov Models

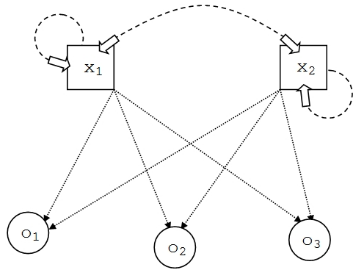
Given sufficient observations of a Markov chain, we can estimate transition matrix $\mathbf{P}$ by counting or bootstrapping the sample probabilities. A more interesting problem arises when the states are unobservable, but act as random variables on *observable output* set $\{o_1, o_2, o_3...\}$.

Definition 3.1. A *hidden Markov model* $(\mathbf{P}, \mathbf{Q}, \mu)$ is a 3-tuple of transition matrix $\mathbf{P}$, matrix of observation probabilities $\mathbf{Q}$ where $q(i,j) = P(O_k = o_i|X_k = x_j)$ for all times k in $\mathbb{N}$, and initial probability vector μ.

Note that the model is hidden irrespective of whether $\mathbf{P}$ is explicitly known. Instead, the chain states are difficult to observe or are defined by their output.

Example 3.2. Consider a speech recognition machine which gathers observable sound waves $\{o_1, o_2, o_3...\}$ as a proxy for spoken words $\{x_1, x_2, x_3...\}$. Given (statistical) noise and speech variation, o_i may indicate more than one x_j, but since the machine knows $\mathbf{P}$ it weights its estimation in favor of likelier word orderings.

Fig. A



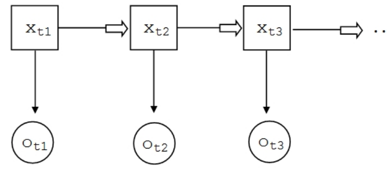


Fig. B

The above diagrams summarize hidden Markov models. In figure A, dashed arrows represent possible state changes, while each state can emit all of the three possible observable outputs. In figure B, known as a *trellis diagram,* arrows represent conditional dependence and we consider a possible sequence of states and their outputs. Continuing our example above, the sentence $\{x_{t1}, x_{t2}, x_{t3}...\}$ would be occurring while the speech recognition machine sees only sound waves $\{o_{t1}, o_{t2}, o_{t3}...\}$.

4. Baum-Welch Inference Algorithm

Suppose we have K observations $O_1 = o_{t_1}, O_2 = o_{t_2}, ...O_K = o_{t_K}$, where o_{t_k} is any element from $\{o_1, o_2, o_3...\}$. Denoting $\mathcal{H} = (\mathbf{P}, \mathbf{Q}, \mu)$, we will consider two simple ways to find $P(o_{t_1}, o_{t_2}, ...o_{t_K}|\mathcal{H})$.

Definition 4.1. For each time k and state x_i, *forward variable* α is defined by

$$\alpha_k(x_i) = P(X_k = x_i \cap o_{t_1}, o_{t_2}, ...o_{t_k}|\mathcal{H}) \tag{4.2}$$

This is the probability of the kth state being x_i and seeing the observations through time k, given our model. From the definition of $\mathbf{Q}$ and μ, we see that

$$\alpha_1(x_i) = P(X_1 = x_i \cap O_1 = o_{t_1}|\mathcal{H}) = q(t_1, i)\mu_i$$

and

$$\alpha_2(x_i) = P(X_2 = x_i \cap O_1 = o_{t_1}, O_2 = o_{t_2}|\mathcal{H}) = q(t_2, i)\sum_{j=1}^{n}\alpha_2(x_j)p_{ji}$$

so that recursively

$$\alpha_{k+1}(x_i) = q(t_k + 1, i)\sum_{j=1}^{n}\alpha_k(x_j)p_{ji}.$$

In particular, we can find $\alpha_K(x_i)$. Since this is the conditional probability given state x_i,

$$\sum_{i=1}^{n}\alpha_K(x_i) = P(o_{t_1}, o_{t_2}, ...o_{t_K}|\mathcal{H})$$

Definition 4.3. For each time k and state x_i, *backward variable* β is defined by

$$\beta_k(x_i) = P(o_{t_{k+1}}, o_{t_{k+2}}, ...o_{t_K}|X_k = x_i \cap \mathcal{H}) \tag{4.4}$$

This is the probability of seeing the observations after time k, given our model and that the kth state is x_i. Since given o_{t_K} and $\mathcal{H}$

$$\beta_K(x_i) = P(O_K = o_{t_K} | X_K = x_i \cap \mathcal{H}) = 1$$

and

$$\beta_{K-1}(x_i) = P(O_{K-1} = o_{t_{K-1}}, O_K = o_{t_K} | X_{K-1} = x_i \cap \mathcal{H}) = \sum_{j=1}^{n} \beta_K(x_j) p_{ij} q(t_K, j)$$

recursively we have

$$\beta_k(x_i) = \sum_{j=1}^{n} \beta_{k+1}(x_j) p_{ij} q(t_{k+1}, j).$$

From the definitions of α and β we can write

$$\alpha_k(x_i)\beta_k(x_i) = P(o_{t_1}, o_{t_2}, ...o_{t_K} | X_k = x_i, \mathcal{H})$$

Since this is the conditional probability given state x_i,

$$\sum_{i=1}^{n} \alpha_k(x_i)\beta_k(x_i) = P(o_{t_1}, o_{t_2}, ...o_{t_K} | \mathcal{H})$$

Thus we have two ways to judge the fit of hidden Markov model $\mathcal{H}$. Many useful probabilities can be expressed as a combination of forward and backward probabilities. For example, the probability of two states at two adjacent times given our observations and model is

$$\begin{aligned}
\xi_k(i,j) &:= P(X_k = x_i \cap X_{k+1} = x_j | o_{t_1}, o_{t_2}, ...o_{t_K}, \mathcal{H}) \\
&= \frac{P(X_k = x_i \cap X_{k+1} = x_j \cap o_{t_1}, o_{t_2}, ...o_{t_K} | \mathcal{H})}{P(o_{t_1}, o_{t_2}, ...o_{t_K} | \mathcal{H})} \\
&= \frac{\alpha_k(x_i) P(X_{k+1} = x_j \cap o_{t_{k+1}}, o_{t_{k+2}}, ...o_{t_K} | \mathcal{H})}{\sum_{i=1}^{n} \alpha_k(x_i)\beta_k(x_i)} \\
&= \frac{\alpha_k(x_i) p_{ij} \beta_{k+1}(x_j) q(t_{k+1}, j)}{\sum_{i=1}^{n} \alpha_k(x_i)\beta_k(x_i)}
\end{aligned}$$

The probability of state x_i at time k given our observations and model is

$$\gamma_k(x_i) := P(X_k = x_i | o_{t_1}, o_{t_2}, ...o_{t_K}, \mathcal{H}) = \frac{\alpha_k(x_i)\beta_k(x_i)}{\sum_{i=1}^{n} \alpha_k(x_i)\beta_k(x_i)}$$

Observe that since $\beta_k(x_i) = \sum_{j=1}^{n} \beta_{k+1}(x_j) p_{ij} q(t_{k+1}, j)$,

$$\gamma_k(x_i) = \sum_{j=1}^{n} \xi_k(i,j)$$

Definition 4.5. The Baum-Welch inference algorithm reestimates $\mathbf{P}, \mathbf{Q}$, and μ as

$$\bar{\mu}_i = \gamma_1(x_i) \tag{4.6}$$

$$\bar{p}_{ij} = \frac{\sum_{k=1}^{K-1} \xi_k(i,j)}{\sum_{k=1}^{K-1} \gamma_k(x_i)} \tag{4.7}$$

$$\bar{q}(i,j) = \frac{\sum_{k=1}^{K} \gamma_k(x_j) 1_{o_{t_k}=o_i}}{\sum_{k=1}^{K} \gamma_k(x_j)} \tag{4.8}$$

The origin of these values will be introduced in the next section.

5. Kullback-Leibler Divergence Derivation

Definition 5.1. For probability distributions P and P' on discrete random variable X, their **Kullback-Leibler divergence** is

$$D_{KL}(P, P') = \sum_x P(x) \log\left(\frac{P(x)}{P'(x)}\right)$$

wherever $\log\left(\frac{P(x)}{P'(x)}\right)$ is defined.

As this is a (weighted) average of the log-differences between distributions P and P', it is often used to judge the fit of an estimated distribution P' to a theoretical distribution P. We consider a basic attribute of K-L divergence.

Lemma 5.2. *Jensen's inequality: If X is a discrete random variable and φ a convex function,*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$$

Proof. Since $\mathbb{E}[X] = \sum_i P(x_i)x_i$ with $\sum_i P(x_i) = 1$, by convexity

$$\varphi(P(x_1)x_1 + P(x_2)x_2) \leq P(x_1)\varphi(x_1) + P(x_2)\varphi(x_2)$$

Suppose as an inductive hypothesis that for all $(k-1)$-valued X with P

$$\varphi\left(\sum_{i=1}^{k-1} P(x_i)x_i\right) \leq \sum_{i=1}^{k-1} P(x_i)\varphi(x_i)$$

Then

$$\varphi\left(\sum_{i=1}^{k} P(x_i)x_i\right) = \varphi\left(P(x_1)x_1 + (1 - P(x_1))\sum_{i=2}^{k} \frac{P(x_i)}{1 - P(x_1)} x_i\right)$$

But since

$$\sum_{i=2}^{k} \frac{P(x_i)}{1-P(x_1)} = \frac{\left(\sum_{i=1}^{k} P(x_i)\right) - P(x_1)}{1-P(x_1)} = 1$$

our hypothesis gives

$$\varphi(\mathbb{E}[X]) = \varphi\left(\sum_{i=1}^{k} P(x_i)x_i\right) \leq \sum_{i=1}^{k} P(x_i)\varphi(x_i) = \mathbb{E}[\varphi(X)]$$

□

The continuous case is similar (but not needed for our purposes), and it is clear from above that Jensen's inequality holds for any weighted sum, not just discrete expectations.

Proposition 5.3. *Gibb's inequality*

$$D_{KL}(P, P') \geq 0$$

Proof. Use Jensen's inequality with $X = \frac{P'}{P}$ and $\varphi = \log\left(\frac{1}{x}\right)$. Then

$$D_{KL}(P, P') = \sum_x P(x) \log\left(\frac{P(x)}{P'(x)}\right) = \mathbb{E}[\varphi(X)] \geq \log\left(\frac{1}{\mathbb{E}[P'/P]}\right)$$

where

$$\log\left(\frac{1}{\mathbb{E}[P'/P]}\right) = \log\left(\frac{1}{\sum_i P(x_i)P'(x_i)/P(x_i)}\right) = \log\left(\frac{1}{1}\right) = 0$$

□

Definition 5.4. For observations $o_{t_1}, ...o_{t_K}$ and $\mathcal{H} = (\mathbf{P}, \mathbf{Q}, \mu)$ *proxy probability* Q is defined by

$$Q(\mathcal{H}') = \sum_{x_{t_1},...x_{t_K}} P(x_{t_1}, ...x_{t_K} \cap o_{t_1}, ...o_{t_K}|\mathcal{H}) \cdot \log P(x_{t_1}, ...x_{t_K} \cap o_{t_1}, ...o_{t_K}|\mathcal{H}')$$

This probability mimics the probability of observations given a model, as shown below.

Proposition 5.5. *If* $Q(\mathcal{H}) \leq Q(\bar{\mathcal{H}})$*, then* $P(o_{t_1}, ...o_{t_K}|\mathcal{H}) \leq P(o_{t_1}, ...o_{t_K}|\bar{\mathcal{H}})$.

Proof. Let $\mathcal{O}$ = (given) observations $o_{t_1}, ...o_{t_K}$. Using Gibb's Inequality with

$$P = \frac{P(x_{t_1}, ...x_{t_K}, \mathcal{O}|\mathcal{H})}{P(\mathcal{O}|\mathcal{H})}, P' = \frac{P(x_{t_1}, ...x_{t_K}, \mathcal{O}|\bar{\mathcal{H}})}{P(\mathcal{O}|\bar{\mathcal{H}})}$$

we have

$$0 \leq \sum_{x_{t_1},...x_{t_K}} \frac{P(x_{t_1},...x_{t_K},\mathcal{O}|\mathcal{H})}{P(\mathcal{O}|\mathcal{H})} \cdot \log\left(\frac{P(x_{t_1},...x_{t_K},\mathcal{O}|\mathcal{H})P(\mathcal{O}|\bar{\mathcal{H}})}{P(\mathcal{O}|\mathcal{H})P(x_{t_1},...x_{t_K},\mathcal{O}|\bar{\mathcal{H}})}\right)$$

$$= \log\left(\frac{P(\mathcal{O}|\bar{\mathcal{H}})}{P(\mathcal{O}|\mathcal{H})}\right) + \sum_{x_{t_1},...x_{t_K}} \frac{P(x_{t_1},...x_{t_K},\mathcal{O}|\mathcal{H})}{P(\mathcal{O}|\mathcal{H})} \cdot \log\left(\frac{P(x_{t_1},...x_{t_K},\mathcal{O}|\mathcal{H})}{P(x_{t_1},...x_{t_K},\mathcal{O}|\bar{\mathcal{H}})}\right)$$

$$= \log\left(\frac{P(\mathcal{O}|\bar{\mathcal{H}})}{P(\mathcal{O}|\mathcal{H})}\right) + \frac{Q(\mathcal{H}) - Q(\bar{\mathcal{H}})}{P(\mathcal{O}|\mathcal{H})}$$

Rearranging gives

$$\frac{Q(\bar{\mathcal{H}}) - Q(\mathcal{H})}{P(\mathcal{O}|\mathcal{H})} \leq \log\left(\frac{P(\mathcal{O}|\bar{\mathcal{H}})}{P(\mathcal{O}|\mathcal{H})}\right)$$

Since

$$0 \leq Q(\bar{\mathcal{H}}) - Q(\mathcal{H})$$

we conclude

$$P(o_{t_1},...o_{t_K}|\mathcal{H}) \leq P(o_{t_1},...o_{t_K}|\bar{\mathcal{H}})$$

□

Thus proxy probability Q allows us to compare the fit of different $\mathcal{H}$ to the observed data. Why examine Q over $P(o_{t_1},...o_{t_K}|\mathcal{H})$? Maximizing real-valued $Q(\mathcal{H})$ produces a critical point of $\bar{\mathcal{H}}$ equal precisely to $\bar{\mathcal{H}}$ in the Baum-Welch algorithm. As space prohibits, the reader may check this lengthy exercise in taking derivatives (Baum, 1972).

Since $\bar{\mathcal{H}} = (\bar{\mathbf{P}}, \bar{\mathbf{Q}}, \bar{\mu})$ is therefore a *maximum log-likelihood estimation*, we have

$$P(o_{t_1}, o_{t_2}, ...o_{t_K}|\bar{\mathcal{H}}) \geq P(o_{t_1}, o_{t_2}, ...o_{t_K}|\mathcal{H}),$$

We can repeat the process iteratively to obtain $\left\{\mathcal{H}, \bar{\mathcal{H}}, \bar{\bar{\mathcal{H}}}...\right\}$ with probabilities given $\{o_{t_1}, o_{t_2}, ...o_{t_K}\}$ converging by the *monotone convergence theorem*. Repeated reestimation is the algorithmic component of Baum-Welch inference.

6. Hidden Markov Chains in Index Returns

We conclude with an example that applies and optimizes a hidden Markov model using Baum-Welch inference.

Recall that hidden Markov models allow for phenomena where states of the world $\{x_{t_1}, x_{t_2}, ...x_{t_K}\}$ are difficult to observe, but rather can be understood by the observable output $\{o_{t_1}, o_{t_2}, ...o_{t_K}\}$ they produce. Following the global liquidity and financial crisis of 2007-2009, general interest in market performance has reached new heights, accompanied as always by speculation over arrivals of "Bull" and "Bear" markets.

Despite a vague sense that "Bull" equates with "Good" while "Bear" embodies the reverse (for the majority of equity investors), when to apply these two terms is not well-defined. As they are characterized by the difficult to observe (and often conflicting) phenomenon of investor sentiment, determination of the states

$$\{x_{BullMarket}, x_{BearMarket}\}$$

lend themselves agreeably to hidden Markov structure.

We will use as observable output monthly log-returns of the S&P 500, a capitalization-weighted index of 500 large companies based in the United States. To simplify matters and produce accessible probability matrices, we divide returns into three classes

$$\{o_{negative}, o_{flat}, o_{positive}\}$$

where returns become negative or positive by exceeding - or + $\frac{15}{12}$%, respectively. No sound methodology was used to determine the cutoff, though annual returns exceeding ± (10 to 20 %) are generally accepted markers of bullish or bearish markets. We examine monthly returns from July 1970 through July 2010.

Operatively, to model bull and bear markets as a hidden Markov chain we input hypothesized probabilities of each return given each state, and each state given a previous state. The true probabilities are then optimized from the observations using Baum-Welch. Programming and iterating for this paper were done in R and Excel.

Below are results after 10 iterations of the Baum-Welch inference algorithm for three different starting parameters.

Initial I

	$P(x_{Bear}\|...)$	$P(x_{Bull}\|...)$	$P(o_n\|...)$	$P(o_f\|...)$	$P(o_p\|...)$
$P(...\|x_{Bear})$	0.7	0.3	0.6	0.3	0.1
$P(...\|x_{Bull})$	0.3	0.7	0.1	0.3	0.6

Baum-Welch Optimization I

	$P(x_{Bear}\|...)$	$P(x_{Bull}\|...)$	$P(o_n\|...)$	$P(o_f\|...)$	$P(o_p\|...)$
$P(...\|x_{Bear})$	0.661	0.339	0.452	0.286	0.262
$P(...\|x_{Bull})$	0.209	0.781	0.241	0.195	0.564

Initial II

	$P(x_{Bear}\|...)$	$P(x_{Bull}\|...)$	$P(o_n\|...)$	$P(o_f\|...)$	$P(o_p\|...)$
$P(...\|x_{Bear})$	0.8	0.2	0.7	0.2	0.1
$P(...\|x_{Bull})$	0.2	0.8	0.1	0.2	0.7

Baum-Welch Optimization II

	$P(x_{Bear}\|...)$	$P(x_{Bull}\|...)$	$P(o_n\|...)$	$P(o_f\|...)$	$P(o_p\|...)$
$P(...\|x_{Bear})$	0.774	0.226	0.462	0.311	0.227
$P(...\|x_{Bull})$	0.118	0.872	0.243	0.184	0.57

Initial III

	$P(x_{Bear}\|...)$	$P(x_{Bull}\|...)$	$P(o_n\|...)$	$P(o_f\|...)$	$P(o_p\|...)$
$P(...\|x_{Bear})$	0.6	0.4	0.8	0.1	0.1
$P(...\|x_{Bull})$	0.4	0.6	0.1	0.1	0.8

Baum-Welch Optimization III

	$P(x_{Bear}\|...)$	$P(x_{Bull}\|...)$	$P(o_n\|...)$	$P(o_f\|...)$	$P(o_p\|...)$
$P(...\|x_{Bear})$	0.607	0.393	0.550	0.301	0.148
$P(...\|x_{Bull})$	0.279	0.711	0.154	0.178	0.668

The Baum-Welch inference algorithm adjusts probabilities away from starting conditions, which as proved increases the likelihood of our model given the observations. Here, the observation probabilities appear most consistently altered. Although it remains sensitive to initial conditions, some relativistic patterns emerged which were confirmed in further trials. For instance,

$$P(x_{BullMarket}|x_{BullMarket}) > P(x_{BearMarket}|x_{BearMarket})$$

Thus the model adapts to account for more frequent (and longer) Bull Markets. Somewhat more surprisingly,

$$P(o_{negative}|x_{BullMarket}) \approx P(o_{flat}|x_{BullMarket})$$

After some research into whether this was possible, I read of *market corrections*, which are short-term drops in asset prices that realign them with the asset's economic value. During bull markets, positive investor sentiment and speculation lead to stock prices continually inflating, until checked by these short drops. Thus the model optimized to account for market features unknown (to me) a priori.

Acknowledgments. It is a pleasure to thank my mentor, Zachary Madden, for keeping this project on track and buoyed with helpful ideas. Thanks are of course due for Peter May and the whole REU staff for allowing us this opportunity to learn and research.

References

[1] Baum, Petrie, Soules, and Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann. Math. Statist., vol. 41, no. 1, pp. 164171, 1970.

[2] Levin, Peres, and Wilmer. Markov Chains and Mixing Times. http://www.uoregon.edu/ dlevin/MARKOV/markovmixing.pdf

[3] Lloyd Welch. Hidden Markov Models and the Baum-Welch Algorithim. http://www-bcf.usc.edu/ lototsky/MATH508/Baum-Welch.pdf

[4] Narada Warakagoda. The Evaluation Problem and the Forward Algorithm. http://jedlik.phy.bme.hu/ gerjanos/HMM/node7.htmleqm10dash.