Hidden Markov Models

Garrett M. Dancik, PhD

The **conditional** probability of an event A, given that B has occurred, is denoted by $P(A|B)$ and has the formula

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Example

Find the probability that a die lands on 4, given that the die roll is even. We want to find $P(A|B)$, where $A =$ a 4 is rolled and B equals an even number is rolled.

$$\begin{aligned} P(4|\text{even}) &= \frac{P(\text{even}|4)P(4)}{P(\text{even})} \\ &= \frac{(1)\times\frac{1}{6}}{1/2} \\ &= \frac{1}{3} \end{aligned}$$

If the event A can be broken down into disjoint events A_1, A_2, etc, then **Bayes' Theorem** states that

$$P(A|B) = \frac{P(B|A)P(A)}{\sum P(B|A_i)P(A_i)}$$

Example

	M	F	Total
Soph	4	2	6
Junior	3	2	5
Senior	3	1	4
Total	10	5	15

$$
\begin{aligned}
P(\text{Soph}) &= P(\text{Soph}|\text{M})P(\text{M}) + P(\text{Soph}|\text{F})P(F) \\
&= (\tfrac{4}{10})(\tfrac{10}{15}) + (\tfrac{2}{5})(\tfrac{5}{15}) \\
&= 0.4
\end{aligned}
$$

For some calculations, we do not need to worry about the denominator from Bayes' Theorem, and instead can work with the formula

$$P(A|B) \propto P(B|A)P(A)$$

where $\propto$ means "is proportional to."

Example

A sophomore student is selected. Is the student more likely to be a male or a female? Using the above version of Bayes Theorem:

$$\begin{aligned} P(M|Soph) &\propto P(Soph|M)P(M) \\ &\propto (\tfrac{4}{10})(\tfrac{10}{15}) \\ &\propto \approx 0.267 \end{aligned}$$

and

$$\begin{aligned} P(F|Soph) &\propto P(Soph|F)P(F) \\ &\propto (\tfrac{2}{5})(\tfrac{5}{15}) \\ &\propto \approx 0.133 \end{aligned}$$

Since $P(M|Soph) \propto 0.267$ and $P(F|Soph) \propto 0.133$, the selected individual is $\frac{0.267}{0.133} = 2$ times as likely to be a male than a female.

A **Markov chain** is a sequence of random variables (or states) X_1, X_2, X_3, with the property that the next state X_{n+1} depends on the m previous states (including the current one). Usually, m is taken to be 1 in which case the next state depends only on the current one, and the Markov chain is said to have the Markov property and is a first order Markov model. Formally, first order Markov models have the property

$$P(X_{n+1}|X_1, X_2, \ldots X_n) = P(X_{n+1}|X_n)$$

Example

An individual has two coins, a fair (F) coin and a biased (B) coin. Before each coin toss, there is a 10% chance that the individual will switch coins. Initially, there is a 50% chance the individual selects the fair (or biased) coin. Find the probability that the selected coins are FFB. Note that this is a 1st order Markov Chain. The subscript i will be used for the i^{th} selected coin.

$$
\begin{aligned}
P(F_1F_2B_3) &= P(F_1)P(F_2|F_1)P(B_3|F_2, F_1) \\
&= P(F_1)P(F_2|F_1)P(B_3|F_2) \\
&= 0.50(0.90)(0.10) \\
&= 0.045
\end{aligned}
$$

A **Hidden Markov model** (HMM) is a Markov chain where the states are hidden (unobserved), but hidden states emit observed values with certain probabilities. The hidden states can then be deduced based on the observed values.
For a first order HMM with observed states O_1, $O_2, \ldots, O_n$ and hidden states $H_1, H_2, \ldots, H_n$, the HMM is characterized by the following probabilities:

Initial state probabilities:	$P(H_1)$
Transition state probabilities:	$P(H_{i+1} \vert H_i)$ for all i
Emission probabilities:	$P(O_i \vert H_i)$ for all i

The goal of a HMM is to predict the set of hidden states (such as the unknown gene structure). The most likely sequence of hidden states is the state sequence $H_1, H_2, \ldots, H_n$ that maximizes

$$P(H_1, \ldots, H_n | O_1, \ldots, O_n) \propto \\ P(O_1|H_1)P(H_1) \times P(O_2|H_2)P(H_2|H_1) \times \cdots \times P(O_n|H_n)P(H_n|H_{n-1})$$

Because these probabilities are often extremely small, it is often useful to work on the log scale and find the state sequence that maximizes

$$\log P(H_1, \ldots, H_n | O_1, \ldots, O_n) = \\ \log P(O_1|H_1)P(H_1) + \log P(O_2|H_2)P(H_2|H_1) + \cdots + \log P(O_n|H_n)P(H_n|H_{n-1}) + k,$$

where k is a constant which is not needed for our purposes.

Coin flipping example

We will use the following probabilities:

$$
\begin{aligned}
&P(F_1) = P(B_1) = 0.50 \\
&P(F_i|F_{i-1}) = P(B_i|B_{i-1}) = 0.90, i > 1 \\
&P(F_i|B_{i-1}) = P(B_i|F_{i-1}) = 0.10, i > 1 \\
&P(h|F) = P(\text{tail}|F) = 0.50 \\
&P(h|B) = 0.80 \\
&P(t|B) = 0.20
\end{aligned}
$$

Suppose we flip 3 coins and observe (h, t, h). Which is more likely, BFB or BBB?

For two different hidden state sequences, we can determine which is more likely. But how do we find the optimal sequence of hidden states?

Continuing this example, suppose we knew that the two optimal states for the first two observations, ending with F or ending with B, were the following: FF and BB. There are 4 possibilities for the last hidden state:

- $F \rightarrow F$
- $F \rightarrow B$
- $B \rightarrow F$
- $B \rightarrow B$

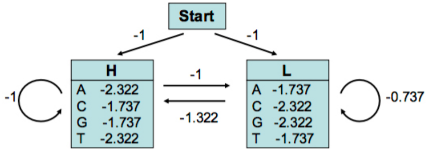
The optimal hidden state sequence is the above possibility that maximizes the overall probability.
This lends itself to a dynamic programming solution (known as the **Viterbi algorithm**).

Example

Using the HMM described previously, what sequence of coins most likely generated *h,t,h?*

High (H) vs. Low (L) GC content example

HMM : Viterbi algorithm - a toy example



back-tracking
(= finding the path which corresponds to the highest probability, -24.49)

GGCACTGAA

	G	G	C	A	C	T	G	A	A
H	**-2.73**	**-5.47**	**-8.21**	-11.53	-14.01	...			-25.65
L	-3.32	-6.06	-8.79	**-10.94**	**-14.01**	...			**-24.49**

The most probable path is: **HHHLLLLLL**

Its probability is $2^{-24.49} = 4.25E{-8}$
(remember that we used $\log_2(p)$)

Note: probabilities are on the $\log_2$ scale
Source: `http://homepages.ulb.ac.be/~dgonze/TEACHING/viterbi.pdf`