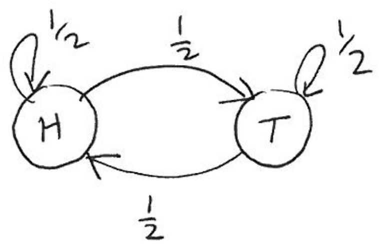
MARKOV, BAYES, AND VITERBI

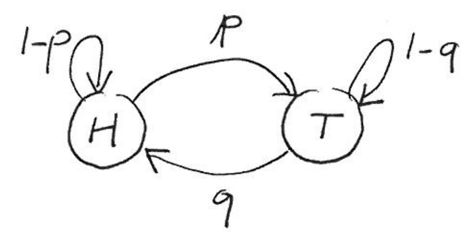
ONE POSSIBLE (AND SIMPLE) WAY TO USE BAYESIAN ANALYSIS IS TO DECIDE WHICH MARKOV CHAIN IS MORE LIKELY TO HAVE PRODUCED A GIVEN OBSERVATION.

FOR INSTANCE, CONSIDER THE FOLLOWING MODELS FOR A FAIR COIN AND A "WEIRD" COIN.



FAIR

$P(H) = P(T) = \frac{1}{2}$



WEIRD

$$P(H) = \frac{q}{p+q} \quad \text{(STATIONARY)}$$

GIVEN AN OBSERVATION OF HEADS AND TAILS, SAY $X_1 X_2 \ldots X_n$, WHICH COIN (MARKOV CHAIN) IS MORE LIKELY TO HAVE GENERATED THE SEQUENCE?

IN OTHER WORDS, COMPARE:

$$P(\text{FAIR} \mid X_1 \ldots X_n) \quad \text{AND} \quad P(\text{WEIRD} \mid X_1 \ldots X_n)$$

$$P(\text{FAIR} \mid X_1 \ldots X_n) \propto P(X_1 \ldots X_n \mid \text{FAIR})\, P(\text{FAIR})$$

USING THE MARKOV PROPERTY

$$P(X_1 \ldots X_n \mid \text{FAIR}) = P(X_1) \cdot P_{X_1 X_2}\, P_{X_2 X_3} \cdots P_{X_{n-1} X_n}$$

WHERE $P_{X_i X_{i+1}}$ IS THE TRANSITION PROBABILITY FROM X_i TO X_{i+1} IN THE "FAIR" MARKOV CHAIN.

BUT WHAT IF THE SEQUENCE $X_1 \dots X_n$ WAS GENERATED BY DIFFERENT COINS AT DIFFERENT TIMES? FOR THIS KIND OF QUESTION, A HIDDEN MARKOV MODEL IS MORE APPROPRIATE.

HIDDEN MARKOV MODEL

A HIDDEN MARKOV MODEL IS GIVEN BY

- A MARKOV CHAIN WITH A SET OF STATES AND TRANSITION PROBABILITIES
- AN ALPHABET Σ
- EMISSION PROBABILITIES.

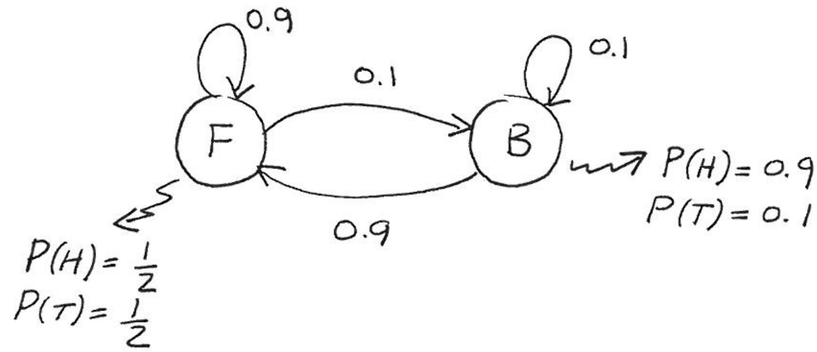
  $e_K(b)$ IS THE PROBABILITY OF EMITTING SYMBOL $b \in \Sigma$ IN STATE K, $\sum_b e_K(b) = 1$.

THE OBSERVATIONS $X_1 \dots X_n$ ARE NOW SYMBOLS IN Σ. THE STATES EMITTING THESE SYMBOLS ARE UNKNOWN (HIDDEN). WE ALSO ASSUME THE MARKOV PROPERTY

$$P(\pi_n = j \mid X_0, \dots, X_m, \pi_0, \dots, \pi_m = i) = P(\pi_n = j \mid \pi_m = i)$$

FOR $m < n$, WHERE π_n STANDS FOR THE STATE AT TIME n.

AN EXAMPLE: (FAIR V.S. BIASED COIN)



THE BAYESIAN QUESTION IS: GIVEN A SEQUENCE OF COIN TOSSES (H & T), WHAT IS THE MOST LIKELY PATH THAT GENERATED IT? IN GENERAL,

FIND $\pi = \pi_1 \ldots \pi_n$ THAT MAXIMIZES

$$P(\pi \mid X_1 \ldots X_n) \propto P(X_1 \ldots X_n \mid \pi)\, P(\pi)$$

$$= P(\pi_1, X_1, \pi_2, X_2, \ldots, \pi_n, X_n)$$

$$= P(\pi_0)\, p_{\pi_0 \pi_1}\, e_{\pi_1}(X_1)\, p_{\pi_1 \pi_2}\, e_{\pi_2}(X_2) \ldots p_{\pi_{n-1} \pi_n}\, e_{\pi_n}(X_n)$$

WHERE π_0 IS A "FICTICIOUS" <u>STARTING</u> STATE WITH EQUAL TRASITION PROBABILITIES TO ALL OTHER STATES.

IN PRINCIPLE, WE CAN TRY ALL POSSIBLE PATHS AND COMPUTE THE ABOVE PROBABILITY FOR EACH PATH AND IDENTIFY THE ONE WITH THE LARGEST PROBABILITY. BUT THIS IS NOT EFFICIENT. THERE ARE EXPONENTIALLY MANY PATHS.

THE VITERBI ALGORITHM

Let $v_k(i)$ BE THE PROBABILITY OF THE OPTIMAL PATH $\pi_1 \dots \pi_i$ THAT GENERATES $x_1 \dots x_i$. AND ENDS IN STATE k. THEN,

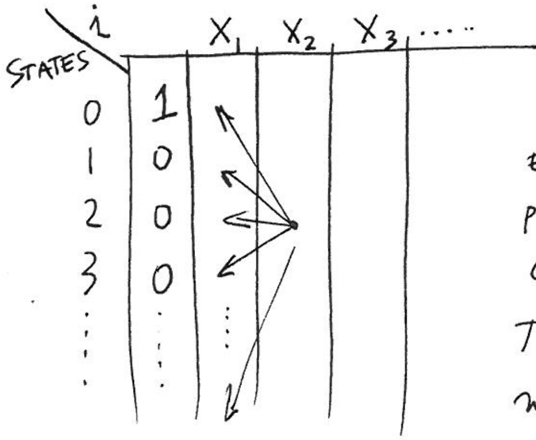
$$v_k(i) = e_k(x_i) \cdot \max_{\ell} \left[v_\ell(i-1) \cdot p_{\ell k} \right]$$

IN OTHER WORDS, SUCH PATH MUST GENERATE SYMBOL x_i IN STATE k, MUST HAVE MADE A TRANSITION TO STATE k FROM SOME STATE ℓ, AND MUST HAVE GENERATED $x_1 \dots x_{i-1}$ OPTIMALLY WHILE ENDING IN STATE ℓ FOR x_{i-1}.

THIS RECURSIVE DEFINITION FOR $v_k(i)$ WILL ALLOW US TO COMPUTE $v_k(n)$ FOR EVERY k, THEN

$\max_k v_k(n)$ IS THE PROBABILITY OF THE OPTIMAL PATH.

TO COMPUTE $v_k(i)$ WE ONLY NEED VALUES OF $v(i-1)$. WE START WITH $v_0(0) = 1$ AND $v_k(0) = 0$ FOR $k \neq 0$.



EVERY $v_k(i)$ NEEDS THE PREVIOUS COLUMN TO BE COMPUTED. TO FILL THE TABLE TIME IS PROPORTIONAL TO nK^2 WHERE K IS # STATES.

EXAMPLE : THE COIN PREVIOUSLY DESCRIBED.
ASSUME WE OBSERVE HH.

		H	H
O	1	0	0
F	0	0.25	0.2025
B	0	0.45	0.0405

KEEP TRACK WHERE YOU COME FROM.

SO THE OPTIMAL PATH IS OBF WITH PROB. 0.2025.
IF WE DROP THE INITIAL TRANSITION FROM O TO B
THE OPTIMAL PATH IS BF WITH PROBABILITY 0.405.

PRACTICAL CONSIDERATION

PROBABILITIES OF PATHS GET TOO SMALL FOR THE MACHINE PRECISION. IT MIGHT BE BETTER TO WORK WITH THE LOG OF PROBABILITIES. LET

$$V_k(i) = \log v_k(i)$$
$$E_k(b) = \log e_k(b)$$
$$P_{\ell k} = \log p_{\ell k}$$

THEN

$$V_k(i) = E_k(b) + \max_{\ell}\left[V_\ell(i-1) + P_{\ell k} \right]$$

AND

$$V_0(0) = 0 \,, \quad V_k(0) = -\infty \quad \text{FOR } k \neq 0$$