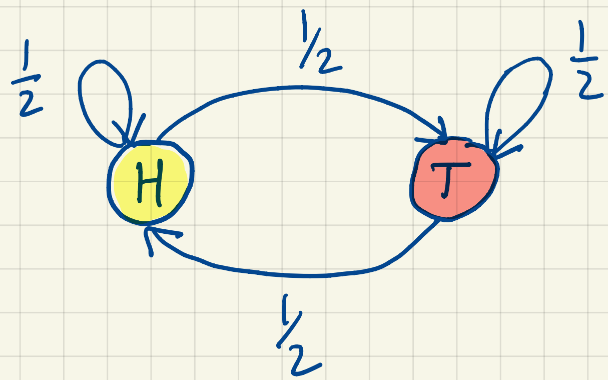
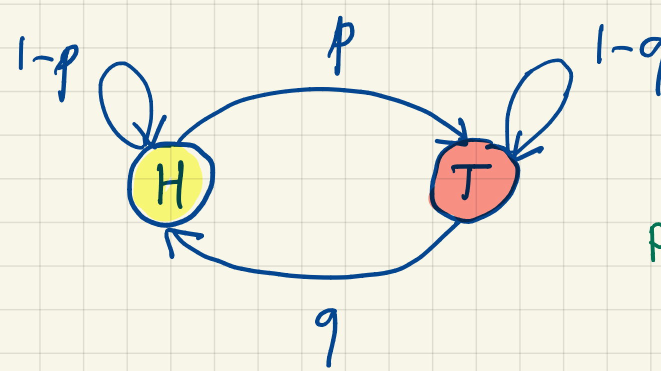
Markov, Bayes, and Viterbi

- One might be interested in deciding, among multiple Markov chains, which one is more likely to have produced a given observation.

- For instance, consider two Markov models for a coin.





$P(H|T) = q$

$P(H|H) = 1-p$

$1-p = q \Rightarrow$ independent

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

(stationary dist.)

Biased: $P(H) = \frac{q}{p+q}$

(stationary dist.)

(Tosses also dependent if $p+q \neq 1$)

- Given an observation $x_1 x_2 \dots x_n$, which coin (Markov chain) is more likely to have generated the sequence.

$$P(\text{Fair} \mid x_1 \dots x_n) \text{ v.s. } P(\text{Biased} \mid x_1 \dots x_n)$$

- $P(\text{Fair} \mid x_1 \dots x_n) \propto P(x_1 \dots x_n \mid \text{Fair})\, P(\text{Fair})$

- Using Markov Property

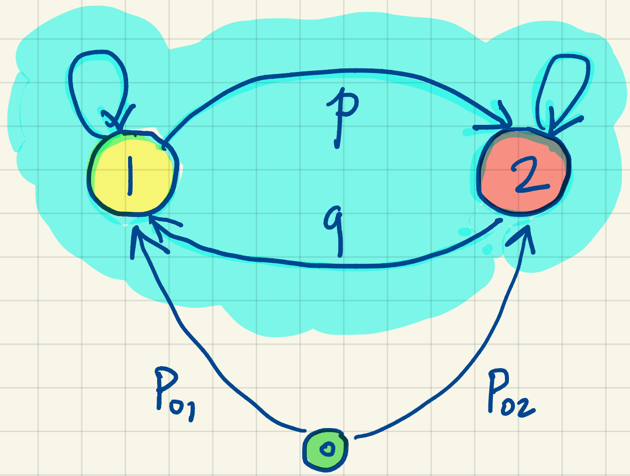
$P(x_1)\, P(x_2 \mid x_1)\, P(x_3 \mid x_1 x_2)\, P(x_4 \mid x_1, x_2, x_3) \dots$

$$P(x_1 \dots x_n \mid \text{Fair}) = P(x_1)\, P_{x_1 x_2}\, P_{x_2 x_3} \dots P_{x_{n-1} x_n}$$

where $P_{x_i x_{i+1}}$ is the transition prob. from x_i to x_{i+1} in the "Fair" Markov chain.

- What is $P(x_1)$? It's part of the model.

Modeling a starting state



We always start here

Note: (but not relevant)
Not irreducible!

But equivalent to an irreducible chain with initial probabilities

$$[p_{01} \quad p_{02}]$$

Stationary dist. $\left[0, \frac{q}{p+q}, \frac{p}{p+q}\right]$

Question: What if the sequence $x_1\, x_2 \ldots x_n$ was generated by different coins at different times?

Hidden Markov Model

- A Hidden Markov Model (HMM) is given by:
  - A Markov chain with a set of states & transition probabilities
  - An alphabet Σ (set of symbols)
  - Emission probabilities:
    - $e_k(b)$ = prob. of emitting symbol $b \in \Sigma$ in state k
    - $\sum_b e_k(b) = 1$

- The observations $x_1\, x_2 \ldots x_n$ are now symbols in Σ.
- The states emitting these symbols are unknown (hidden)
- Markov Property: $(m < n)$

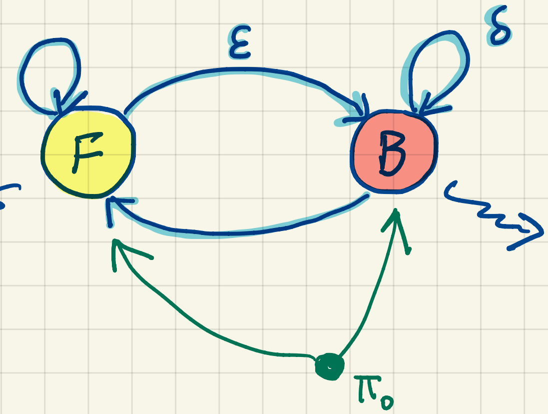
$$P(\pi_n = j \mid x_0, \ldots, x_m, \pi_0, \ldots, \pi_m = i) = P(\pi_n = j \mid \pi_m = i)$$

where π_n = state at time n.

Example HMM "Dishonest casino"

$e_F(H) = \frac{1}{2}$

$e_F(T) = \frac{1}{2}$



$e_B(H) = 0.9$

$e_B(T) = 0.1$

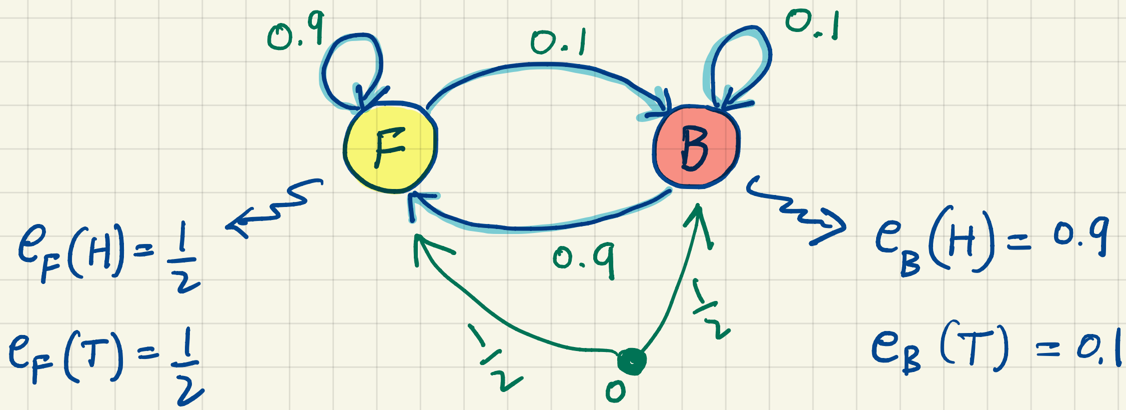
Given $x_1 \ldots x_n$ where $x_i \in \{H, T\}$,

find $\pi = \pi_1 \ldots \pi_n$ that maximizes

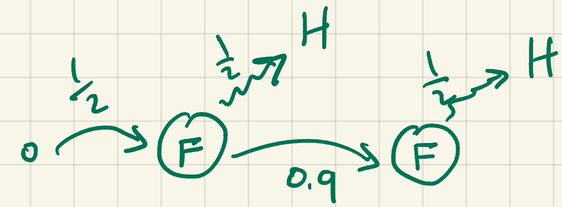
$$P(\pi \mid x_1 \ldots x_n) \propto P(x_1 \ldots x_n \mid \pi) P(\pi) = P(x_1 \ldots x_n \wedge \pi_1 \ldots \pi_n)$$

$$= P_{\pi_0 \pi_1} e_{\pi_1}(x_1) P_{\pi_1 \pi_2} e_{\pi_2}(x_2) \ldots P_{\pi_{n-1} \pi_n} e_{\pi_n}(x_n)$$

Example:



Let's say $X_1 X_2 = HH$



$0FF$: $\frac{1}{2} \times \frac{1}{2} \times 0.9 \times \frac{1}{2}$

$0FB$: $\frac{1}{2} \times \frac{1}{2} \times 0.1 \times 0.9$

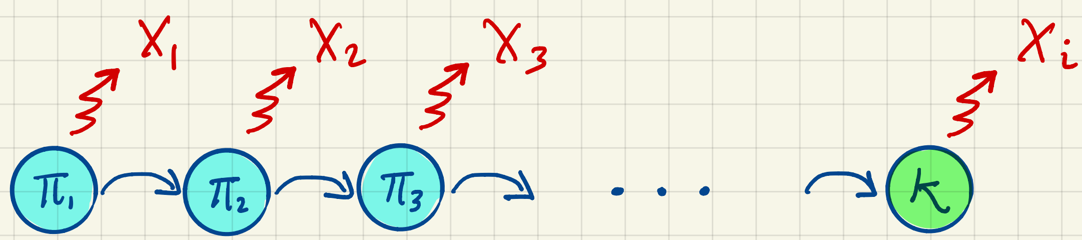
$0BF$: $\frac{1}{2} \times 0.9 \times 0.9 \times \frac{1}{2}$ ✓

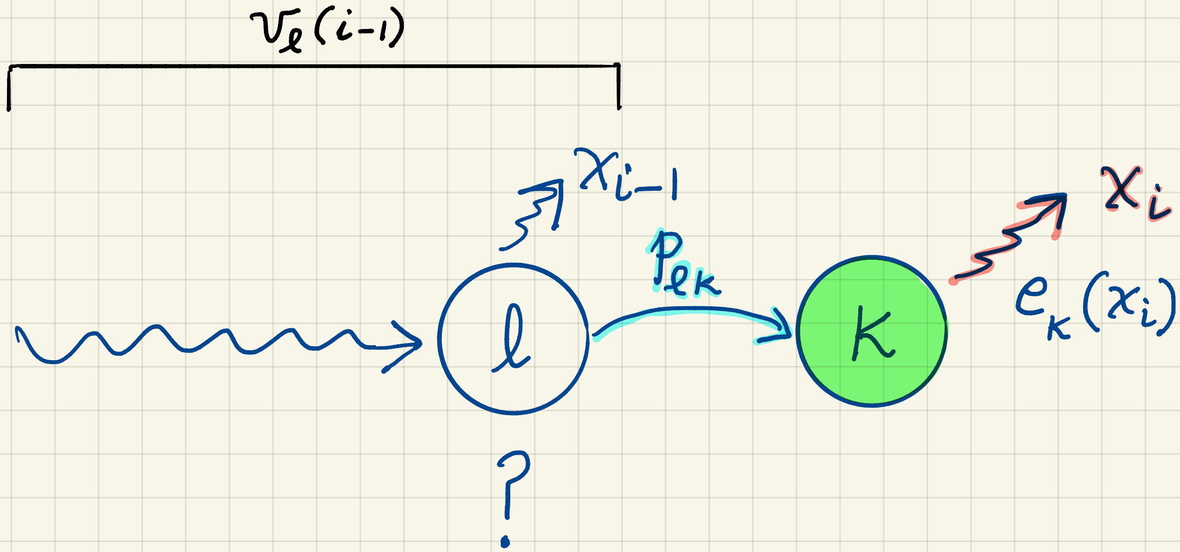
$0BB$: $\frac{1}{2} \times 0.9 \times 0.1 \times 0.9$

- Consider all possible paths
- Compute above probability for each path
- Identify the one with largest probability
- Correct, but NOT efficient.
- There are exponentially many path, e.g. 2^n for our example.

Solution: Viterbi algorithm.

Define $v_k(i)$ = prob. of <u>OPTIMAL</u> path $\pi_1 \dots \pi_i$ that generates $x_1 \dots x_i$ and ends in state k $(\pi_i = k)$





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

We can compute $v_k(i)$ if we have

$v_0(i-1), v_1(i-1), \dots v_K(i-1)$ where K = # states

Finally, we need $\max_k v_k(n)$

Viterbi

$$V_0(0) = 1 \qquad V_k(0) = 0 \text{ for all } k \neq 0$$

$$V_k(i) = e_k(x_i) \max_{\ell} \left[V_\ell(i-1)\, P_{\ell k} \right]$$

$$\text{return } \max_{k} V_k(n)$$

Init

for $k = 0 \ldots K$

do $V[k, 0] = 0$

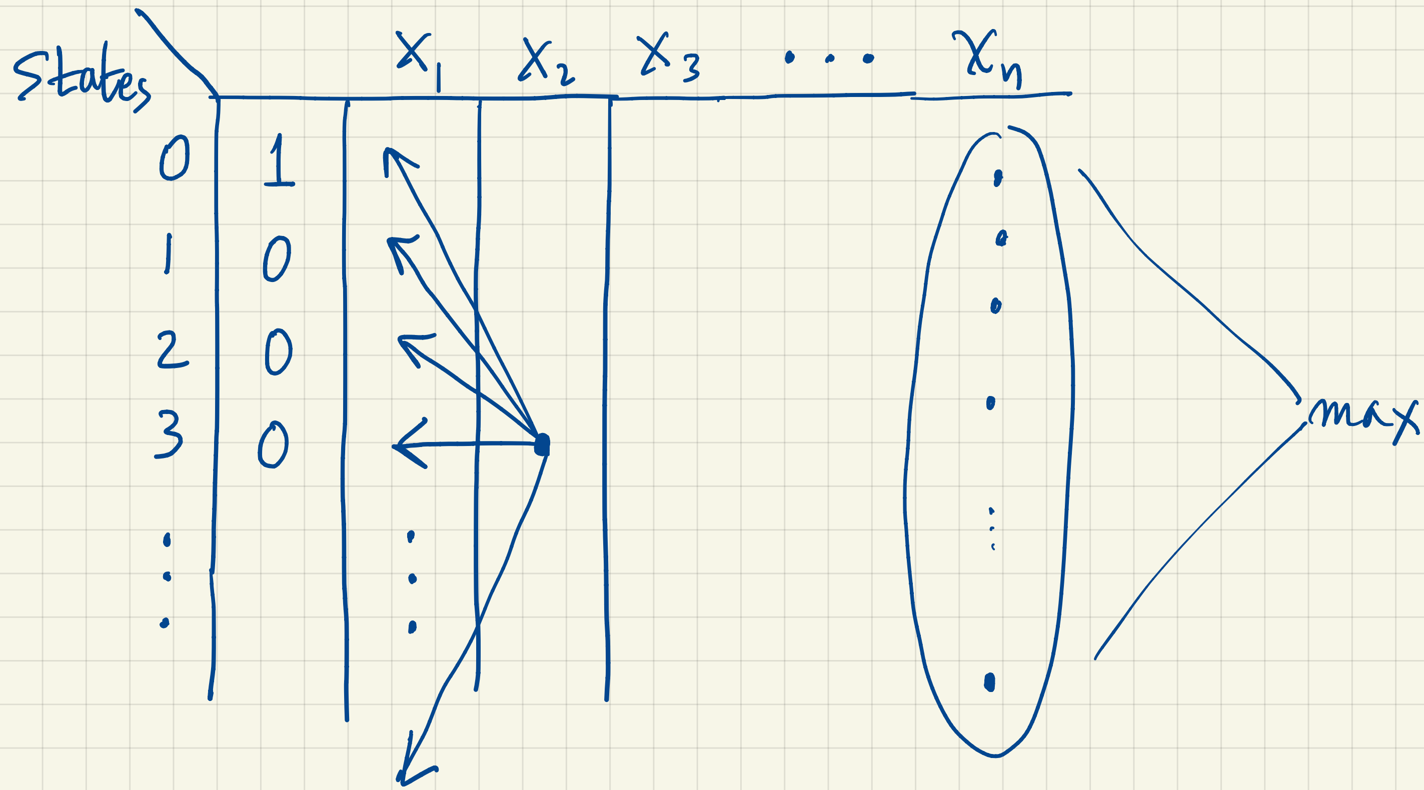
$V[0,0] = 1$

for $k = 0 \ldots K$

for $i = 1 \ldots n$

$$V[k, i] = \underbrace{e_k(x_i) \cdot \max_{\ell} \left[V[\ell, i-1] \cdot P[\ell, k] \right]}_{\text{for loop to find max}}$$

$V_0(0) = 1 \qquad V_k(0) = 0$ for $k \neq 0$

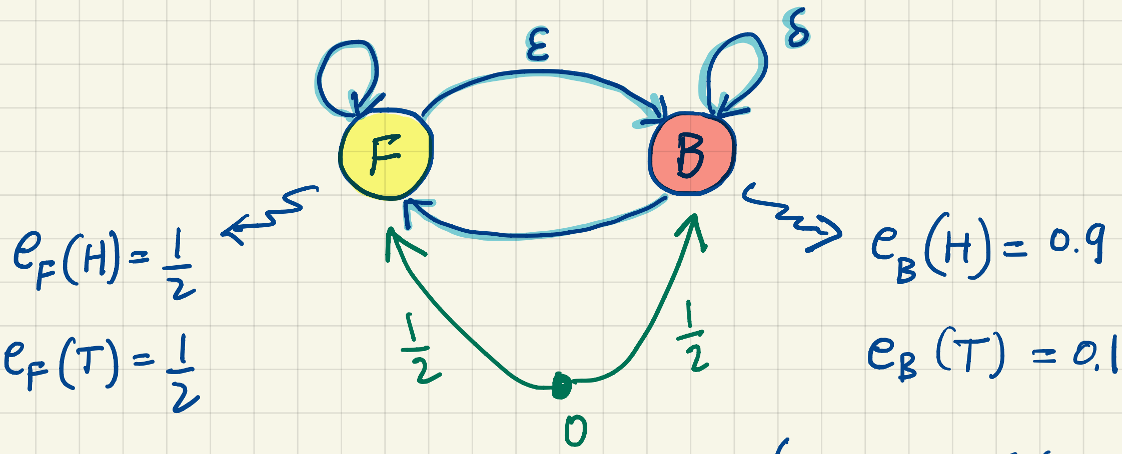


$V_k(i)$ needs the previous column to be computed.

Time needed proportional to nK^2, K = # states.

Example:

Assume $P_{0F} = P_{0B} = \frac{1}{2}$, $\varepsilon = \delta = 0.1$



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

$$v_B(1) = \max \begin{cases} e_B(H)\, v_0(0)\, P_{0B} = 0.9 \times 1 \times \frac{1}{2} \\ e_B(H)\, v_F(0)\, P_{FB} = 0.9 \times 0 \times 0.1 \\ e_B(H)\, v_B(0)\, P_{BB} = 0.9 \times 0 \times 0.1 \end{cases}$$

$$v_F(2) = \max \begin{cases} e_F(H)\, v_0(1)\, P_{0F} = \frac{1}{2} \times 0 \times \frac{1}{2} \\ e_F(H)\, v_F(1)\, P_{FF} = \frac{1}{2} \times 0.25 \times 0.9 \\ e_F(H)\, v_B(1)\, P_{BF} = \frac{1}{2} \times 0.45 \times 0.9 \end{cases}$$