MA3259 Lecture 17

Hidden Markov Models and Applications

LX Zhang
Department of Mathematics
National University of Singapore
matzlx@nus.edu.sg

Outline

- Markov chain
- Hidden Markov Models

Reference: Biological sequence analysis

by R. Durbin, S. Eddy, A. Krogh, G. Mitchison

Finite Markov Chain (Processes)

A discrete-time finite Markov chain has

-- a finite set of states , $\mathcal{E}=\{E_1, E_2, \ldots, E_s\}$

-- a transition probability matrix $P_{s\times s} = (p_{ij})$

- p_{ij} is the probability from state E_i to state E_j

-- an initial state distribution $\pi=\{\pi_i \mid 1\leq i\leq s\}$

π_i is the probability that the process is at state E_i at time 0.

At each time point t=0, 1, 2, …, the process occupies a state E in $\mathcal{E}$;

if at time point i, the process is at state E_i, then, in step from i to $i+1$, it either stays at E_i with probability p_{ii} or move to some other states E_j with probability p_{ij}

A finite Markov chain has the following distinguishing properties:

- **The memoryless property.** If at some time i the process is in the state E_i, the probability that one time unit later it is in state E_j depends only one E_j, and not on the past history of the states it was in before time t.
- **The time homogeneity property.** Given that at time i, the process is in state E, the probability that one time unit later, it is in state E' is independent of time i.

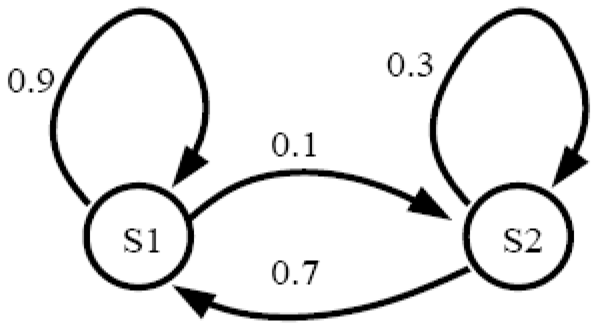
Give the present state, the future and past states are independent.

Formally, a Markov chain is a sequence of random variables (having state as their values) with the following Markov property:

$$\Pr[X_{n+1} = x \mid X_n = x_n, X_{n-1} = x_{n-1}, \cdots, X_0 = x_0] = \Pr[X_1 = x \mid X_0 = x_n].$$

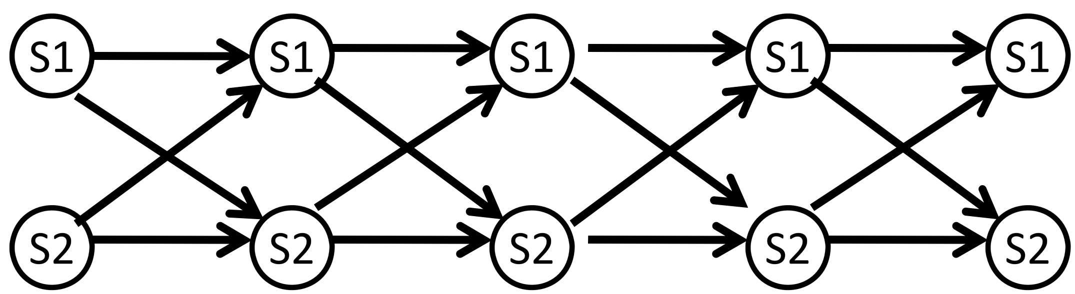
Remark: Since p_{ij} is the probability that the process moves to E_j from E_i, therefore, the sum of entries in each row is 1 in P.

It is often to present a finite Markov chain as a directed graph with labeled edges. For example,



$$\begin{array}{cc} & \begin{array}{cc} S1 & S2 \end{array} \\ \begin{array}{c} S1 \\ S2 \end{array} & \begin{bmatrix} 0.9 & 0.1 \\ 0.7 & 0.3 \end{bmatrix} \end{array}$$

represents a finite Markov chain with states S1 and S2 and the transition matrix given above.



Time:	0	1	2	3	4	…
State variables:	X1	X2	X3	X4	X5	…
Observed states	S1	S2	S2	S1	S1	…

Markov chain process is a sequence of random variables X_i, $i=0, 1, 2, ...$

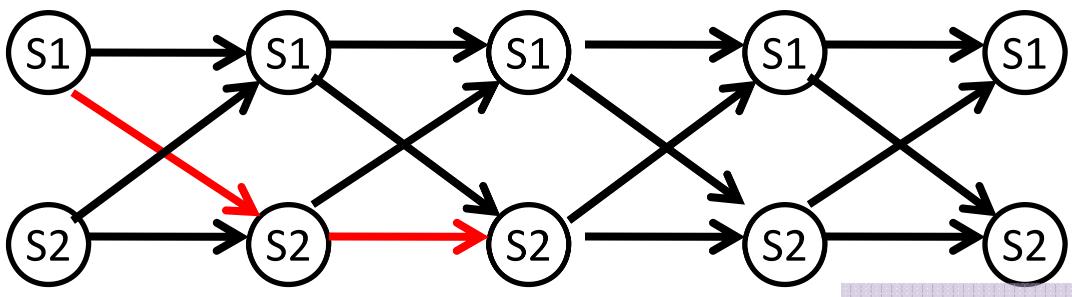
Hence, we are interested in the following probabilities.

$$\Pr[X_0 = S1, X_1 = S2, X_2 = S2 \mid X_0 = S1]$$

(Prob that the process moves from S1 to S2 and Stay at S2 in two steps.)

$$= \Pr[X_1 = S2 \mid X_0 = S1] \times \Pr[X_2 = S2 \mid X_1 = S2]$$

$$= 0.1 \times 0.3$$



(Prob that the process starts at S1, moves from S1 to S2 and stay at S2 in tw Steps.)

$$\Pr[X_0 = S1, X_1 = S2, X_2 = S2]$$

$$= \Pr[X_0 = S1] \times \Pr[X_0 = S1, X_1 = S2, X_2 = S2 \mid X_0 = S1]$$

$$= \Pr[X_0 = S1] \times 0.1 \times 0.3$$

Markov chain process is a sequence of random variables X_i, $i=0, 1, 2, \ldots$

$$p^{(3)}_{12} = \Pr[X_3 = S2 \mid X_0 = S1]$$

Probability that it moves from S1 to S2 in three steps.

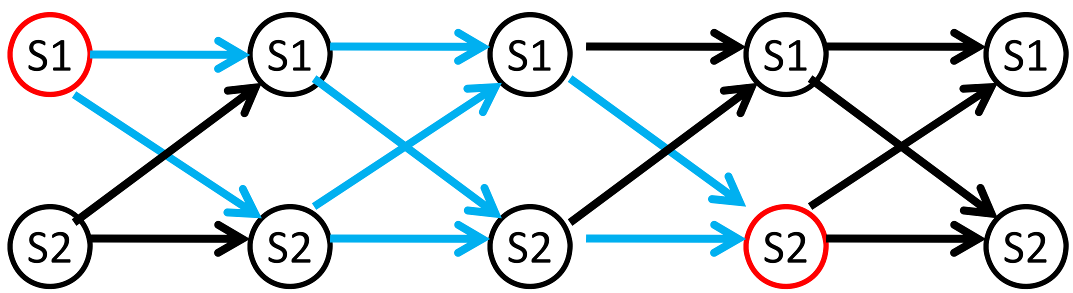
$$= \sum_{x1}\sum_{x2} \Pr[X_0 = S1, X_1 = x_1, X_2 = x_2, X_3 = S2 \mid X_0 = S1]$$

$$= \Pr[X_0 = x_0, X_1 = S1, X_2 = S1, X_3 = S2 \mid X_0 = S1]$$

$$+ \Pr[X_0 = x_0, X_1 = S1, X_2 = S2, X_3 = S2 \mid X_0 = S1]$$

$$+ \Pr[X_0 = x_0, X_1 = S2, X_2 = S1, X_3 = S2 \mid X_0 = S1]$$

$$+ \Pr[X_0 = x_0, X_1 = S2, X_2 = S2, X_3 = S2 \mid X_0 = S1]$$



By induction, we can prove that $p^{(3)}_{ij}$ is the (i, j)-entry of the power matrix P^3.

Example:

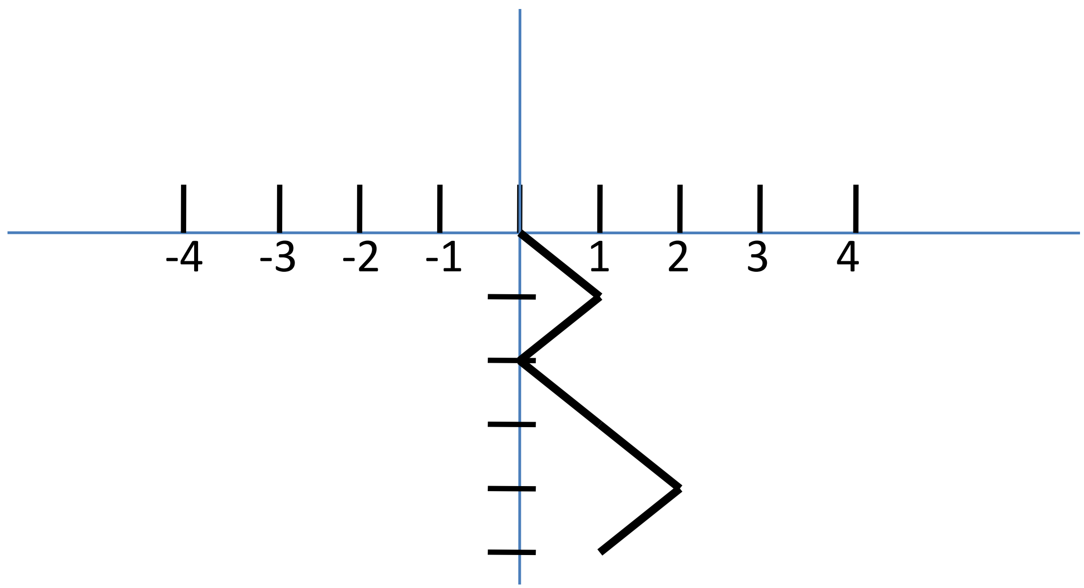
One dimensional random walk with absorb states.

-- States: integers from $-k$ to k.

-- Move from state j ($|j|<k$) to $j-1$ and $j+1$ with equal probability. At state $j=-k$, k, the process stays there forever.

-- The process starts at 0.

Time:	Outcome
0	
1	R
2	L
3	R
4	R
5	L



Hidden Markov Models(HMMs)

Definition: A HMM is a finite Markov chain with an extra function: At each state, it emits a letter with time-independent, but state-dependent probability.

It has

-- a set of states $\mathcal{E}=\{E_1, E_2, \ldots, E_s\}$;

-- an alphabet A;

-- a transition probabilities $P=(p_{ij})$;

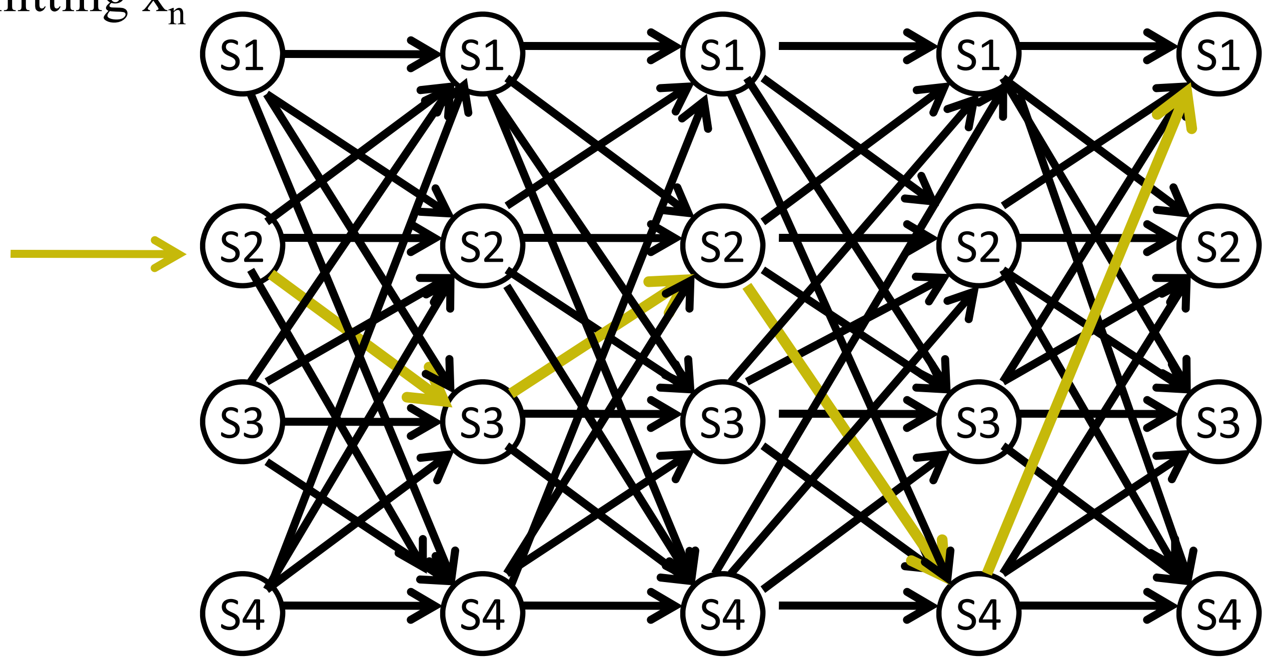
-- an initial state distribution $\pi=\{\pi_i \mid 1 \le i \le s\}$

-- emission probabilities:

$b_i(o)$ is the probability that letter o (in A) is emitted at state i.

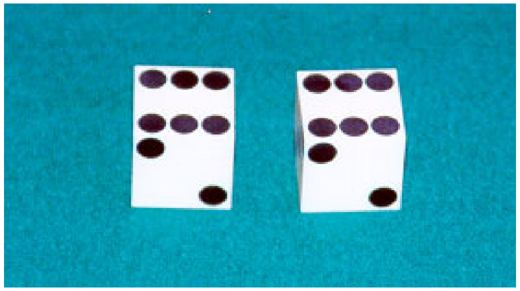
HMM is a sequence generator!

1. Start at state Si according to prob π_i
2. Emit letter x_1 according to prob $e_1(x_1)$
3. Move from Si to state Sj according to probability p_{ij}
4. … until emitting x_n



Time	0	1	2	3	4	….
States	S2	S4	S2	S4	S1	….
Emitted letters	x1	x2	x3	x4	x5	…..

Example: The dishonest casino (Batzoglou's slide)



A casino has two dice:

Fair die

$P(1) = P(2) = P(3) = P(5) = P(6) = 1/6$

Loaded die

$P(1) = P(2) = P(3) = P(5) = 1/10$

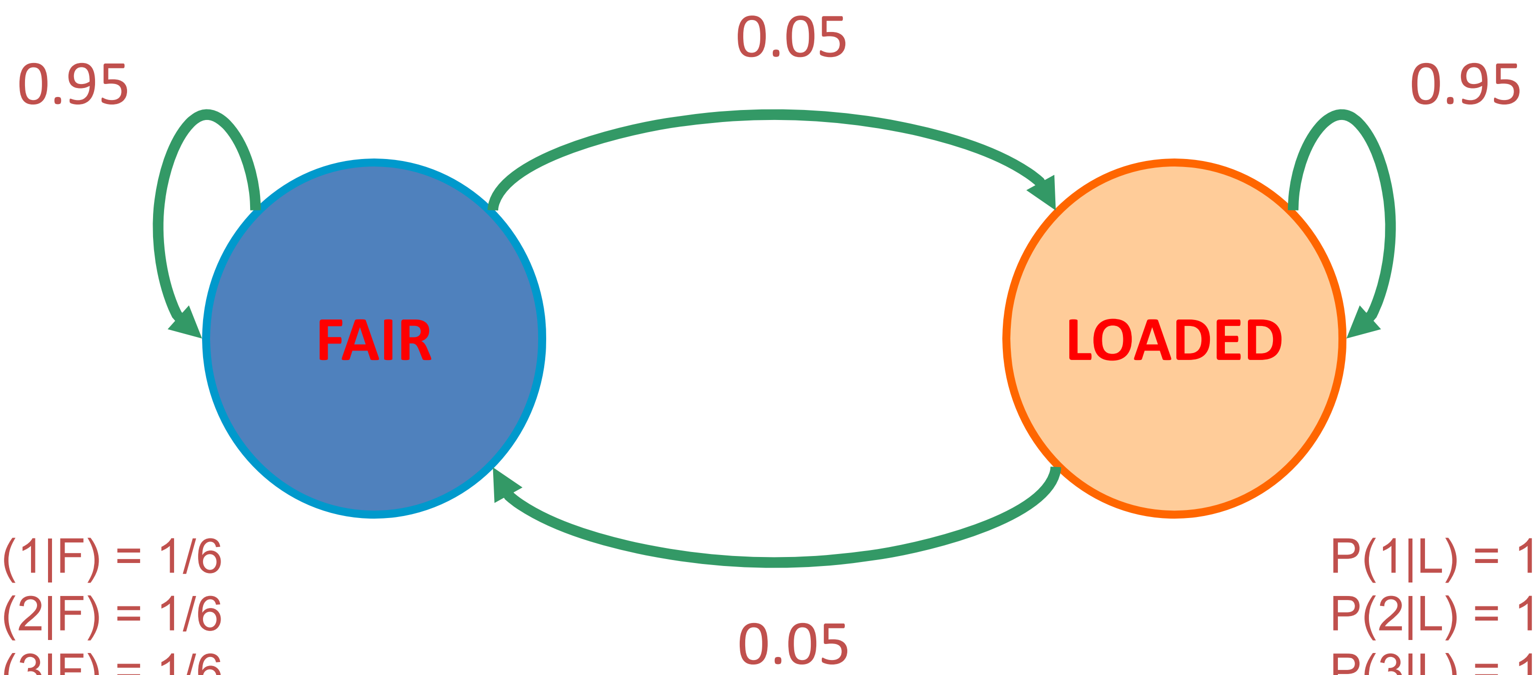
$P(6) = 1/2$

Casino player switches back-&-forth between fair and loaded die once every 20 turns

Game:

1. You bet $1
2. You roll (always with a fair die)
3. Casino player rolls (maybe with fair die, maybe with loaded die)
4. Highest number wins $2

The dishonest casino model (Batzoglou’s slide)



P(1|F) = 1/6
P(2|F) = 1/6
P(3|F) = 1/6
P(4|F) = 1/6
P(5|F) = 1/6
P(6|F) = 1/6

P(1|L) = 1/10
P(2|L) = 1/10
P(3|L) = 1/10
P(4|L) = 1/10
P(5|L) = 1/10
P(6|L) = 1/2

Problem 1: Evaluation

Input: a sequence of rolls by the casino player

1255526461146146136136661664661636616366163616515615116146123562344

Question: How likely is this observed sequence given our model of how the casino plays?

Problem 2: Decoding

Input: a sequence of rolls by the casino player

1255526461146146136136661664661636616366163616515615116146123562344

Fair die Fair die

Question: What is the most likely way of producing such a sequence of rolls given our model? In other words, which portion of the sequence was generated with the fair die, and which portion with the loaded die?

Problem 3: Learning

Input: a sequence of rolls by the casino player

1255526461146146136166661664666636616366163616515615116146123562344

Question: How unfair is the loaded die?
How often the dice are exchanged back and forth?

In summary, the following three problems arise in HMM applications.

(1). Given the parameters $\lambda = (P, B, \pi)$, what is $P[\mathcal{O}|\lambda]$, the probability of $\mathcal{O}$ as an observed output?

(2). What is the hidden sequence $\mathcal{Q} : q_1, q_2, \cdots, q_t$ of states that is most likely to have occured, given $\mathcal{O}$? That is, we need to find $argmax_{\mathcal{Q}} P[\mathcal{Q}|\mathcal{O}]$.

(3). Assuming a fixed topology of the model, what are the parameters $\lambda = (P, B, \pi)$ that maximize $P[\mathcal{O}|\lambda]$?

Remarks:

(a) The model M includes (1) Architecture (States and links)
(2) parameters θ= (P, π, b()).

P[x | M] is the same with P[x | θ], and P[x], when the architecture, and the parameters, respectively, are implied.
It is the probability that the sequence x is generated by the model.

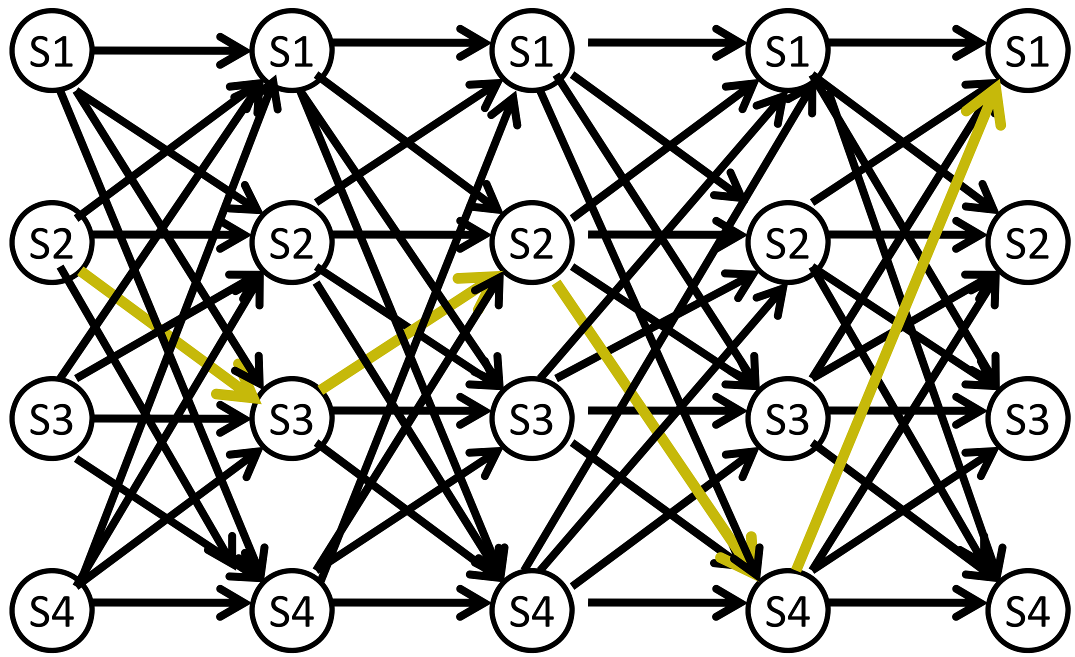
(b) For a sequence $\mathcal{O}$ over alphabet A,

$$\Pr[\mathcal{O}]\ \Pr[\mathcal{Q} | \mathcal{O}] = \Pr[\mathcal{O}, \mathcal{Q}].$$

Hence, find a $\mathcal{Q}$ that has the largest conditional probabilities is equivalent to find $\mathcal{Q}$ that maximizes Pr[$\mathcal{O}$, $\mathcal{Q}$]

Pr[$\mathcal{O}$, $\mathcal{Q}$ | M], Pr[$\mathcal{O}$, $\mathcal{Q}$ | θ] and Pr[$\mathcal{O}$, $\mathcal{Q}$] are the same when the architecture, and the parameters, are implied

A parse is a sequence of states visited at times 0, 1, 2, …, N.



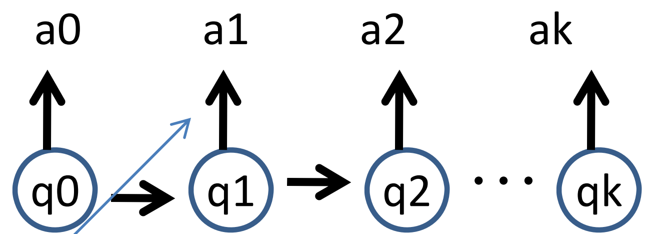
Time	0	1	2	3	4
Parse	S2	S4	S2	S4	S1

Likelihood of a parse

Give a sequence $a_0, a_1, \cdots a_k$ over Σ,

and a parse $q_0, q_1, \cdots q_k$

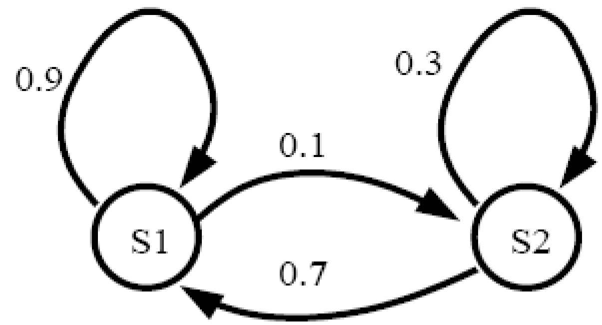
To find how likely the parse produces the sequence (given a HMM model)



$$\Pr[a_0, \ldots, a_k, q_0, \ldots\ldots, q_k]$$
$$= \Pr[q_0]\Pr[a_0 \mid q_0] \Pr[q_1 \mid q_0] \Pr[x_1 \mid q_1] \ldots P[q_k \mid q_{k-1}] \Pr[a_k \mid q_k]$$
$$= \pi_{q0} p_{q0q1} \ldots\ldots p_{qk-1qk}\, b_{q0}(a_0) \ldots\ldots b_{qk}(a_k)$$

Example 1:

For sequence $\mathcal{O} : 2, 2, 2$

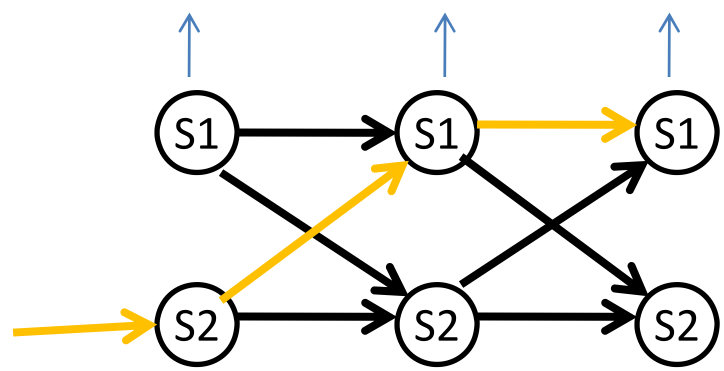


Pr[1]=1/2　　Pr[1]=1/4
Pr[2]=1/2　　Pr[2]=3/4

The state sequence S_2, S_1, S_1 has the following likelihood:

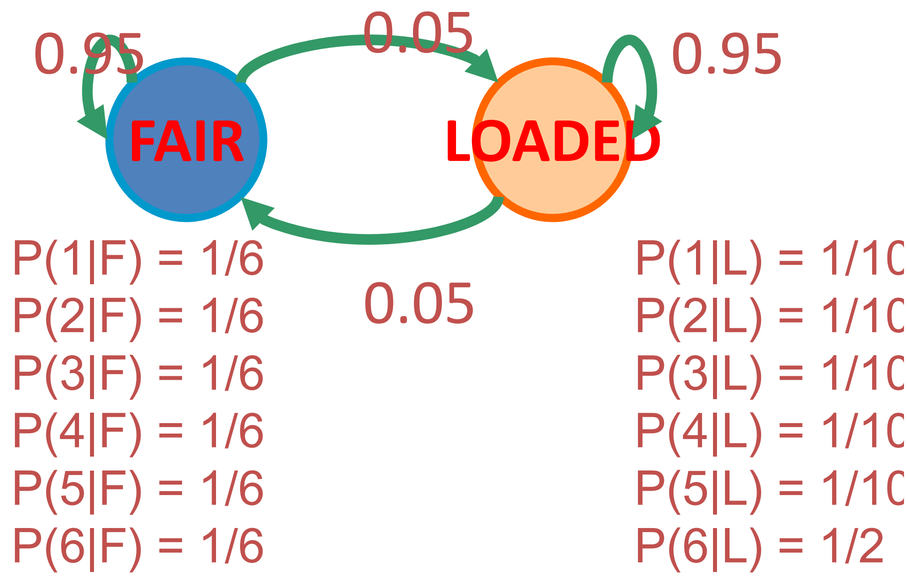
Pr[{2, 2, 2}, {S2, S1, S1}]
= Pr[Initial state S2] Pr[Emit 2 at S2]
Pr[S1|S2] Pr[Emits 2 at S1] Pr[S1|S1] Pr[Emit 2 at S1]
= (1/2) x (3/4) x 0.7 x (1/2) x 0.9 x (1/2)
= 0.0591.

(Assume initial state distribution is uniform)



Sum the likelihood over all the state sequences, we have
Pr[{2, 2, 2}]= 0.1867.

Example: the dishonest casino (Batzoglou's slide)



Let the sequence of rolls be:

O = 1, 6, 6, 5, 6, 2, 6, 6, 3, 6

Now, what is the likelihood Q = Fair, Fair, …, Fair?

$\frac{1}{2} \times (1/6)^{10} \times (0.95)^9$ ~= 0.5×10^{-9}, same as before

What is the likelihood

π = L, L, …, L?

$\frac{1}{2} \times (1/10)^4 \times (1/2)^6 (0.95)^9 = .00000049238235134735$ ~= 0.5×10^{-7}

So, it is 100 times more likely the die is loaded