Hidden Markov Model

Hidden Markov Model

- Hidden Markov model (HMM) is a statistical **Markov model** in which the system being modeled is assumed to be a **Markov process** with **unobserved (hidden) states**.

- In probability theory and statistics, a **Markov process** is a stochastic process that satisfies the **Markov property**.

- **Markov property**: A Markov process is called to satisfy the Markov property if one can make prediction of the future for the process based solely on its present state. In other words, the Markov process is "**memory less**".

- For a discrete-time Markov chains

$$P\left[{}^{X_n = x_n}\!/_{X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_0 = x_0}\right] = P\left[{}^{X_n = x_n}\!/_{X_{n-1} = x_{n-1}}\right]$$

Markov Model

Example

Let us say that you have \$10 and participate in a betting game. You bet \$1 each time and continue until you loose all your money. So, we want to know if this process of betting follows a Markov property.

Let us take that X_n: the number of dollars you have after n bets. It means that $X_0=10$. We want to prove that the sequence $\{X_n, n \in N)$ is a Markov process.

Let us say that it is given that $X_1=11$, $X_2=10$, $X_3=11$ and $X_4=12$, then

$P[X_5=x_5 / X_0=10, X_1=11, X_2=10, X_3=11, X_4=12] = P[X_5=x_5 / X_4=12] = ½$ (if $x_5=11$ or 13)

Thus, it is a Markov process.

Exercise

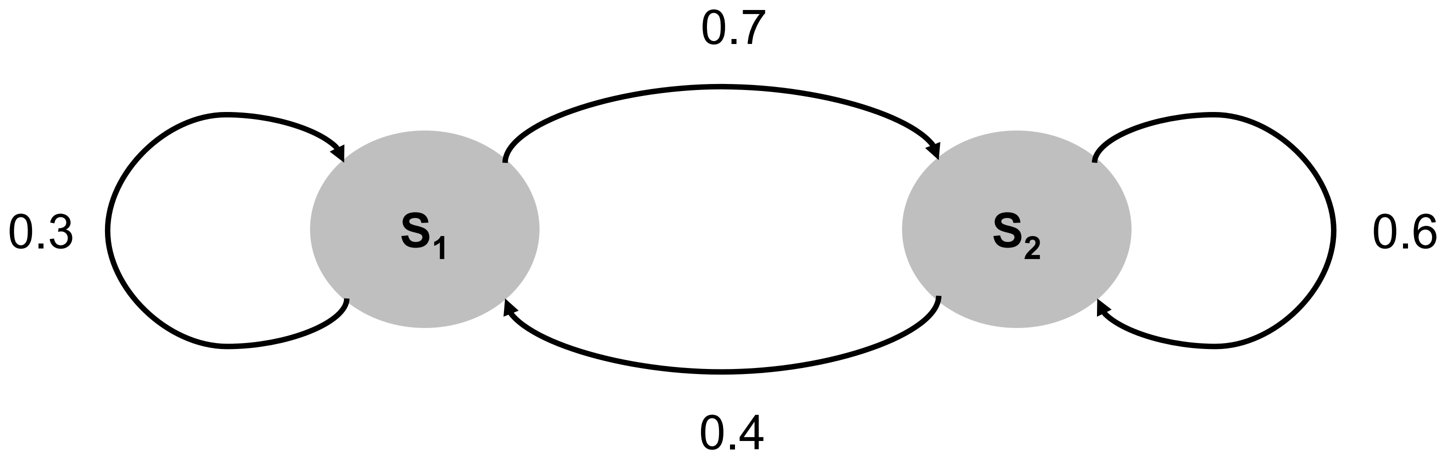
Suppose you have a wallet with five quarters, nickels and dimers each worth 25c, 5c and 10c, respectively.

Now, you draw coins one-by-one from the wallet and keep them on a table. If X_n represents the total value of the coins kept on the table after n draws with $X_0=0$, then find out if the sequence $\{X_n, n \in N)$ is a Markov process or not.

Markov Model

Markov process / chain: is a random process that undergoes transitions from one state to another on a state space in such a way that it follows Markov property.

$$P\left[{}^{X_{t+1}}/_{X_t, X_{t-1}, \ldots, X_0}\right] = P\left[{}^{X_{t+1}}/_{X_t}\right]$$



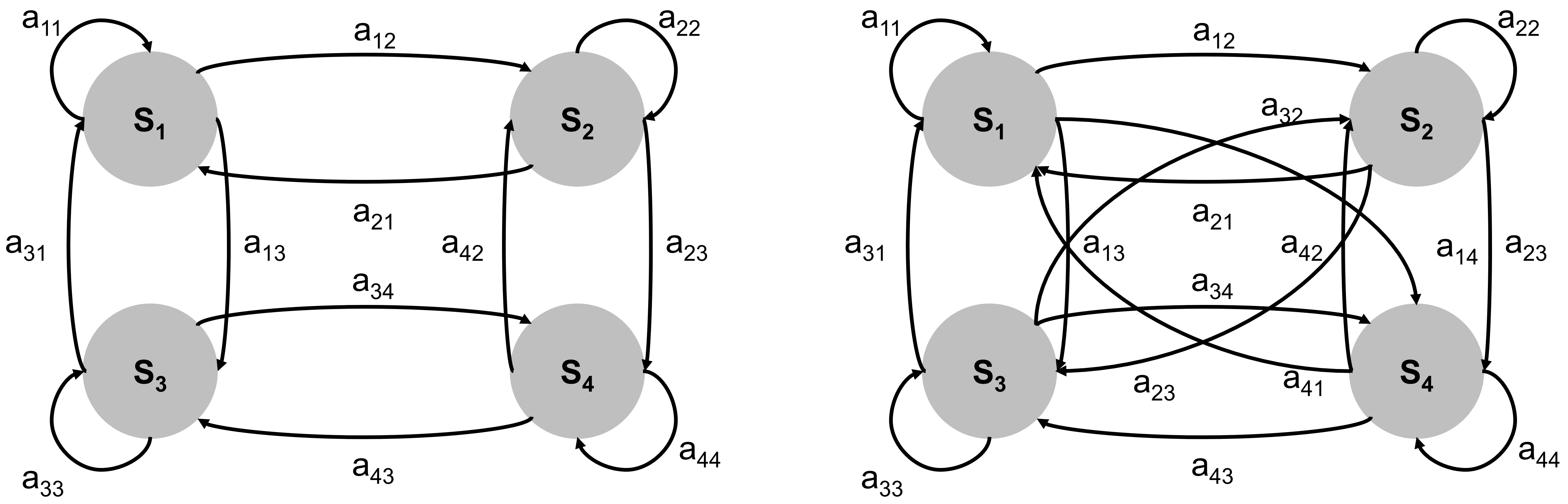
Markov chain: A process with discrete-time Markov events (DTMC).

Markov process: A process with continuous-time Markov events (CTMC).

Discrete Markov process (chain)

A system example

Consider a system which can be described at any time as being in one of a set of N states (S_1, S_2, …, S_N). Let us take N=4, then system looks like as follows:



Ergodic model

Discrete Markov process (chain)

Transition probability

$$P[S] = P[S_1\, S_2\, S_3 \ldots S_{N-1}\, S_N] = P[S_1] \times P\left[{}^{S_2}/_{S_1}\right] \times P\left[{}^{S_3}/_{S_2\, S_1}\right] \times \ldots \times P\left[{}^{S_N}/_{S_{N-1}\, S_{N-2} \ldots S_2\, S_1}\right]$$

For a Markov chain,

$$P[S] = P[S_1\, S_2\, S_3 \ldots S_{N-1}\, S_N] = P[S_1] \times P\left[{}^{S_2}/_{S_1}\right] \times P\left[{}^{S_3}/_{S_2}\right] \times \ldots \times P\left[{}^{S_N}/_{S_{N-1}}\right]$$

Let us define a variable q_t denoting a state S_i (i = 1, 2, …, N) at time t.

If $a_{ij} = P\left[{}^{q_{t+1} = S_j}/_{q_t = S_i}\right] = P\left[{}^{S_j}/_{S_i}\right]$ where $aij \geq 0$ and $\sum_{j=1}^{N} aij = 1$,

then

$$P[S] = P[S_1] \times a_{12} \times a_{23} \times \ldots \times a_{N-1\,N}$$

a_{ij} are called transition probabilities.

Discrete Markov process (chain)

Initial state probability

Let us define

$$\pi_i = P[q_1 = S_i] \quad where\ 1 \leq i \leq N$$

This is known as initial state probability.

Thus,

$$\mathrm{P[S]} = \pi_1 \times a_{12} \times a_{23} \times \ldots \times a_{N-2\ N-1} \times a_{N-1\ N}$$

Discrete Markov process (chain)

So, as of now, we know the following:

(1) States

$$\{S_i\} \qquad 1 \leq i \leq N$$

(2) Transition probabilities

$$a_{ij} \quad 1 \leq i \leq N \text{ such that } a_{ij} \geq 0 \quad \text{and} \quad \sum_{j=1}^{N} a_{ij} = 1$$

(3) Initial state probability

$$\pi_i = P[q_1 = S_i] \quad where\ 1 \leq i \leq N$$

Discrete Markov process (chain)

Weather forecasting example

Let us consider that once a day (e.g. at noon), the weather is observed as being one of the following: Rain, Cloudy and Sunny. Let us call S_1, S_2 and S_3 as the states denoting these weathers.

Suppose that the transition probabilities are given as follows:

$$A = [a_{ij}] = \begin{array}{c} S_1 \\ S_2 \\ S_3 \end{array} \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

Given that the weather on day 1 is Sunny , what is the probability of observing the following weathers for the next 7 days: O= {Sunny, Sunny, Rain, Rain, Sunny, Cloudy, Sunny}.

Discrete Markov process (chain)

Weather forecasting example

We can represent the weather observation as $O = [S_3\ S_3\ S_3\ S_1\ S_1\ S_3\ S_2\ S_3]$, then the probability of observing 'O' can be written as

$$
\begin{aligned}
P[O] = P[S_3\ S_3\ S_3\ S_1\ S_1\ S_3\ S_2\ S_3] &= P[S_3] \times P[S_3/S_3] \times P[S_3/S_3] \times P[S_1/S_3] \times P[S_1/S_1] \times P[S_3/S_1] \times P[S_2/S_3] \times P[S_3/S_2] \\
&= \pi_3 \times a_{33} \times a_{33} \times a_{31} \times a_{11} \times a_{13} \times a_{32} \times a_{23} \\
&= 1 \times 0.8 \times 0.8 \times 0.1 \times 0.4 \times 0.3 \times 0.1 \times 0.2 \\
&= 1.536 \times 10^{-4}
\end{aligned}
$$

Discrete Markov process (chain)

Forecasting a state

Given that the model is in a known state, what is the probability that it will stay in that state for exactly ‘d’ days?

$O=\{{}^{1}S_i\ {}^{2}S_i\ {}^{3}S_i \dots {}^{d}S_i,\ {}^{d+1}S_j\}$ such that $S_j \neq Si$ and $1 \leq i,j \leq N$

$P[O] = \pi_i \times (a_{ii})^{d-1} \times (1-a_{ii}) = p_i(d)$

$p_i(d)$ represents a discrete probability density function of duration ‘d’ in state S_i.

Based on the function $p_i(d)$, we can calculate the expected number of observation (duration) in a state, given that the model is in that state.

$$\bar{d}_i = \sum_{d=1}^{\infty} d.\, p_i(d) = \sum_{d=1}^{\infty} d\, (a_{ii})^{d-1}\, (1-a_{ii}) = \frac{1}{1-a_{ii}}$$

Discrete Markov process (chain)

Forecasting a state

Thus, the expected number of consecutive days of sunny weather, according to the model is

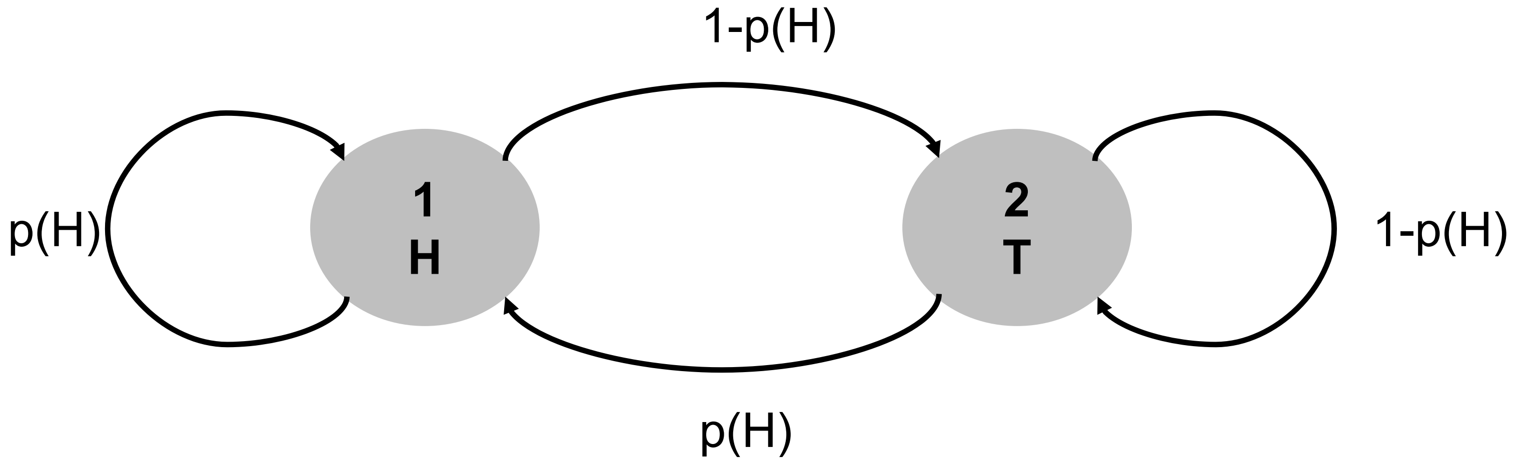
$$\bar{d}_{S1} = \frac{1}{1-0.4} = \frac{1}{0.6} = 1.67\ days$$

$$\bar{d}_{S2} = \frac{1}{1-0.6} = \frac{1}{0.4} = 2.5\ days$$

$$\bar{d}_{S3} = \frac{1}{1-0.8} = \frac{1}{0.2} = 5\ days$$

Hidden Markov model

One coin toss model



Let us say that we have the following observations of heads and tails: O = H H T T H T H H T T H
Then

S = 1 1 2 2 1 2 1 1 2 2 1

Hidden Markov model

One coin toss model

$$P[S] = P[1] \times P[1/1] \times P[2/1] \times P[2/2] \times P[1/2] \times P[2/1] \times P[1/2] \times P[1/1] \times P[2/1] \times P[2/2] \times P[1/2]$$

$$= \pi_1 \times a_{11} \times a_{12} \times a_{22} \times a_{21} \times a_{12} \times a_{21} \times a_{11} \times a_{12} \times a_{22} \times a_{21}$$

$$= 1 \times p(H) \times (1-p(H)) \times (1-p(H)) \times p(H) \times (1-p(H)) \times p(H) \times p(H) \times (1-p(H)) \times (1-p(H)) \times p(H)$$

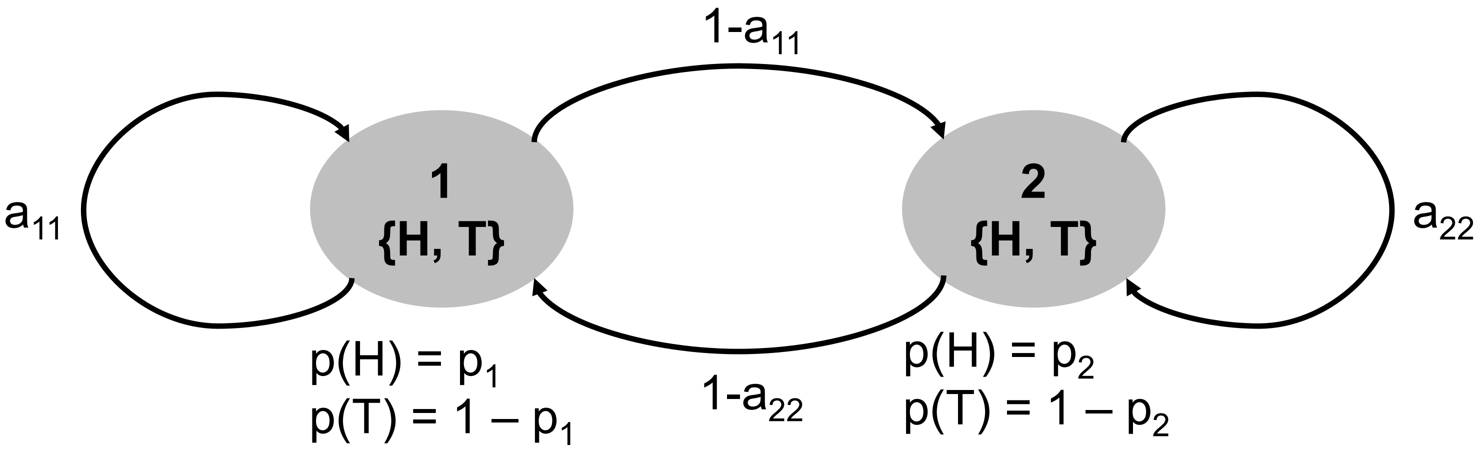
If $p(H) = 0.5$, then

$$P[HHTTHTHHTTH] = P[S] = 1 \times 0.5 \times 0.5 \times 0.5 \times 0.5 \times 0.5 \times 0.5 \times 0.5 \times 0.5 \times 0.5 \times 0.5$$

$$= 9.765625 \times 10^{-4}$$

Hidden Markov model

Two coin toss model



Let us say that we have the same following observations of heads and tails: O = H H T T H T H H T T H

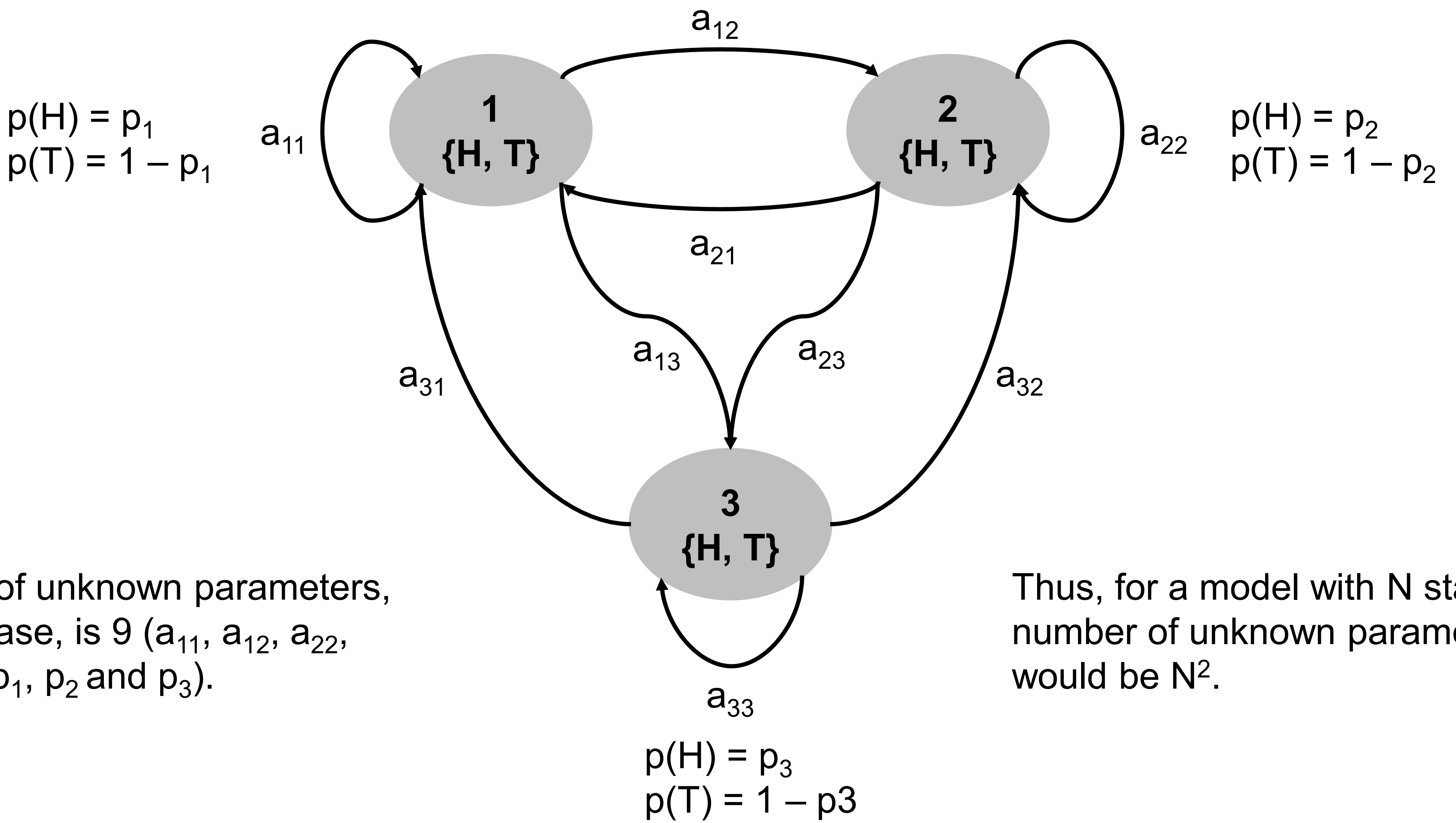
Then can we write S = 1 1 2 2 1 2 1 1 2 2 1 for sure?

It means that the states are ***hidden***. This type of modeling is called ***Hidden Markov Model***.

The number of unknown parameters, here in this case, is 4 (a_{11}, a_{22}, p_1 and p_2).

Hidden Markov model

Three coin toss model



The number of unknown parameters, here in this case, is 9 (a_{11}, a_{12}, a_{22}, a_{23}, a_{33}, a_{31}, p_1, p_2 and p_3).

Thus, for a model with N states, the number of unknown parameters would be N^2.

Elements of Hidden Markov Model

An HMM is characterized by the following elements:

(1) N: ***number of states*** in the model. States are represented as S={S_1, S_2, …, S_N) and a state at time t is denoted as q_t. In addition to observed states, there are two ***end states*** called as ***start*** and ***end***.

(2) M: ***number of distinct observation symbols per state*** i.e., discrete alphabet size. The symbols correspond to the physical output of the system. It is represented as V = {v_1, v_2, …, v_m).

(3) A: the ***state transition probability distribution***. It is represented as A = {a_{ij}} where $a_{ij} = P\left[\frac{q_{t+1} = S_j}{q_t = S_i}\right]$ and 1 ≤ i,j ≤ N.

(4) B: the ***emission probability distribution*** of observation symbols. It is represented as $B = \{b_j(k)\}$, where

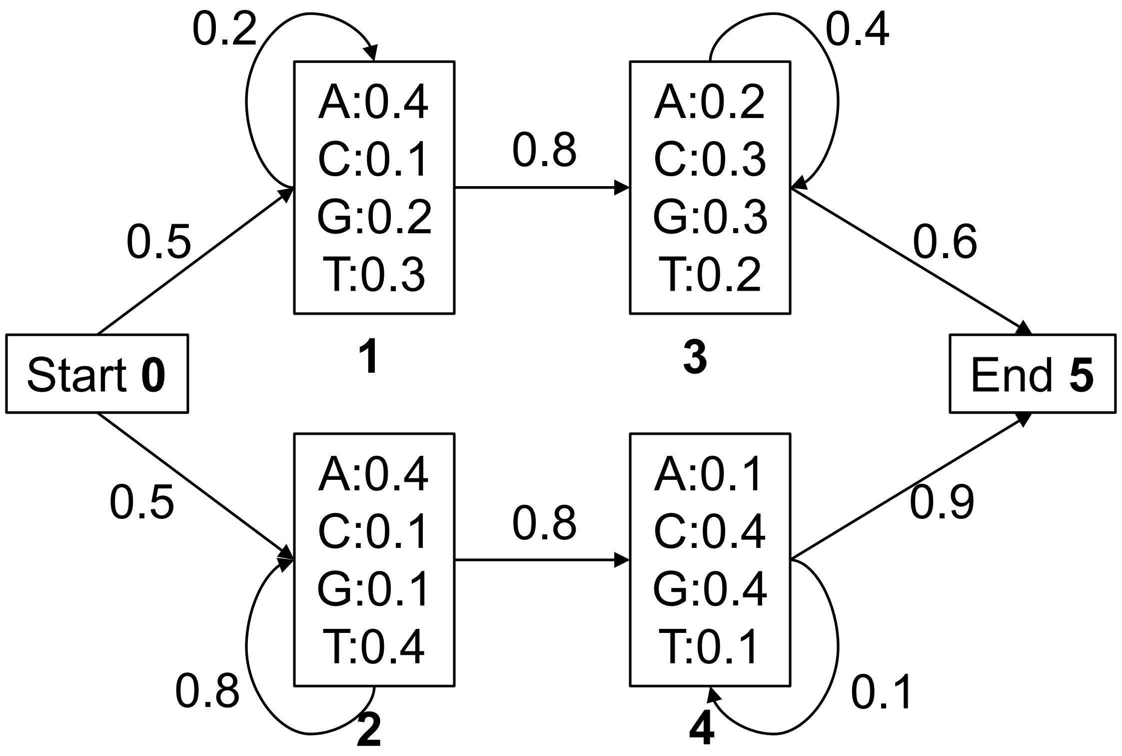
$$b_j(k) = P\left[{}^{v_k\ at\ time\ t}/_{q_t = S_j}\right]$$ and 1 ≤ j ≤ N, 1 ≤ k ≤ M

(5) π: the initial state probability distribution. It is represented as π = {π_i} where

$$\pi_i = P[q_1 = S_i] \quad where\ i \le i \le N$$

Calculating the probability of an observation given an Hidden Markov Model

Find the probability of observing a DNA sequence 'AAC', given the following HMM and path = {0, 1, 1, 3, 5}.



The probability of a given observation (sequence) is given by

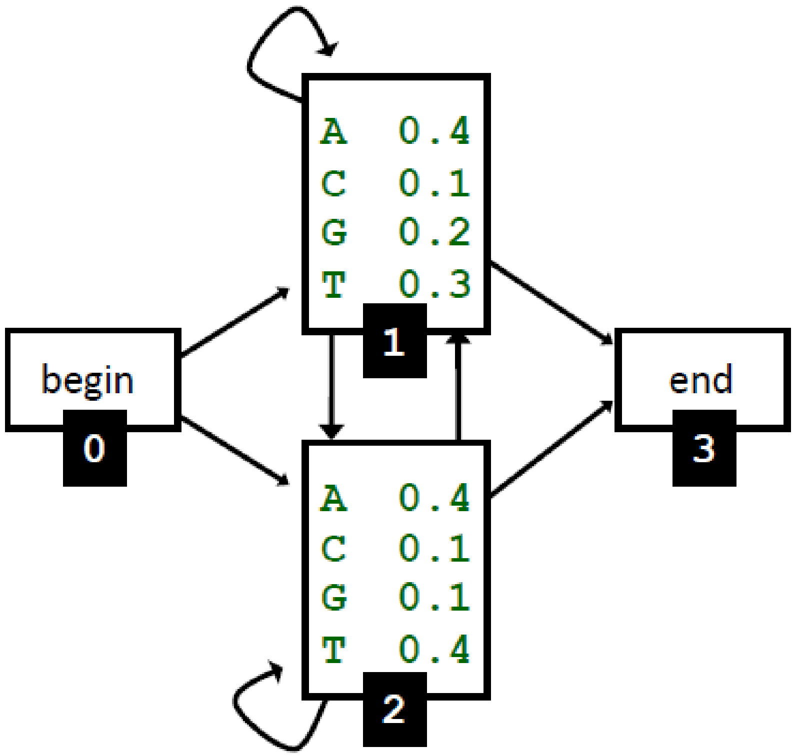
$$P[{}^{O}/_{\pi}] = a_{01} \prod_{i=1}^{L} e_{\pi_i}(O_i)\, a_{\pi_i \pi_{i+1}} \qquad where\ \pi_{L+1} = 0$$

Thus,

$P[AAC / \pi] = a_{01} \times e_1(A) \times a_{11} \times e_1(A) \times a_{13} \times e_3(C) \times a_{35}$
$= 0.5 \times 0.4 \times 0.2 \times 0.4 \times 0.8 \times 0.3 \times 0.6$
$= 2.304 \times 10^{-3}$

However, the probability of an observation 'O' is the sum of probabilities of all the possible paths.

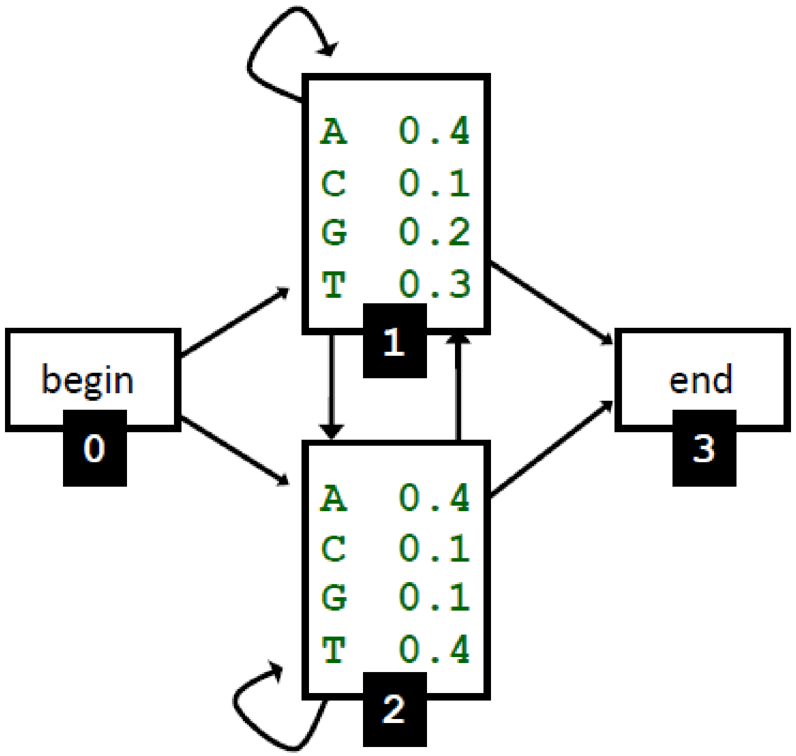
Number of possible paths



For an observation of length (L) =1, the possible paths are {0, 1, 3} and {0, 2, 3}.

For an observation of length (L) =2, the possible paths are {0, 1, 1, 3}, {0, 1, 2, 3}, {0, 2, 2, 3} and {0, 2, 1, 3}.

Number of possible paths



For an observation of length (L) =3, the possible paths are {0, 1, 1, 1, 3}, {0, 1, 1, 2, 3}, {0, 1, 2, 1, 3}, {0, 1, 2, 2, 3}, {0, 2, 2, 2, 3}, {0, 2, 2, 1, 3}, {0, 2, 1, 1, 3} and {0, 2, 1, 2, 3}.

Thus, for an observation of length (L), number of possible paths for an ergodic model is $\mathbf{2^L}$.

Three basic problems for HMMs

Problem 1: How likely is a given sequence of observations?
Given the observation sequence $O = O_1, O_2, \ldots, O_L$ and a model $\lambda=(N, M, A, B, \pi)$, how do we efficiently compute $P(O/\lambda)$, the probability of the observation sequence given the model.
[The Forward algorithm]

Problem 2: What is the most probable "path" for generating a given sequence?
Given the observation sequence $O = O_1, O_2, \ldots, O_L$ and a model $\lambda=(N, M, A, B, \pi)$, how do we choose a corresponding state sequence $Q=q_1, q_2, \ldots, q_L$ which is optimal in some meaningful sense (i.e. best 'explains' the observation).
[The Viterbi algorithm]

Problem 3: How can we learn the HMM parameters given a set of sequences?
How do we adjust the model parameters $\lambda = (N, M, A, B, \pi)$ to maximize $P(O/ \lambda)$.
[The Forward-Backward (Baum-Welch) algorithm]

How likely is a given sequence: the Forward algorithm

Let us define $f_k(i)$ denote the probability of being in state *k* having observed the first *i* characters of *O* (length *L*). It means that we want to find the value of $f_N(L)$, the probability of being in the end state having observed all the characters of the given observation *O*.

(1) Initialization

$f_0(0) = 1$, probability that we are in start '0' state and have observed '0' characters from the sequence '*O*'.

$f_k(0) = 0$ $\forall$ k that are not silent states (k = 1, 2, …, N)

(2) Recursion for emitting states (i=1, 2, …, L)

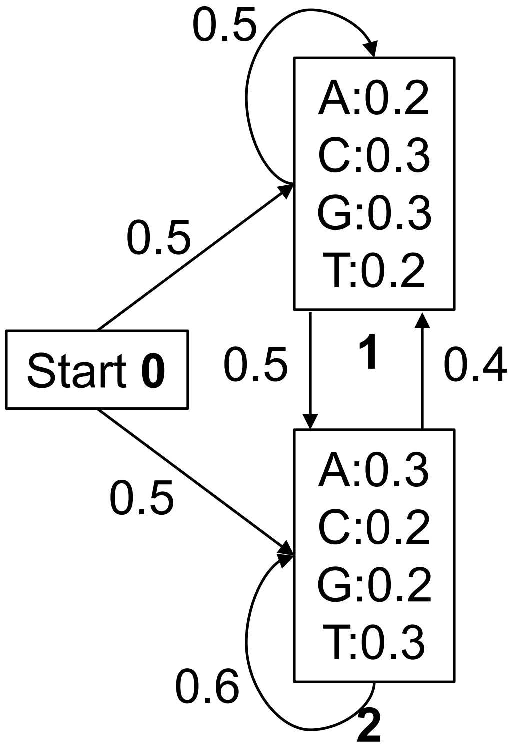
$$f_l(i) = e_l(i) \sum_k f_k(i-1)\, a_{kl} \qquad 1 \le i \le L$$

(3) Termination

$$P[O] = P[O_1\, O_2 \ldots O_L] = f_N(L) = \sum_k f_k(L) a_{kN}$$ where N denotes the last state.

How likely is a given sequence: the Forward algorithm

Given the following HMM, find the probability of O=GGCA using the Forward algorithm



States	Start	G	G
1	0	0.5 x 0.3 = 0.15	0.15 x 0.5 x 0.3 + 0.10 x 0.4 x 0.3 = 0.0345
2	0	0.5 x 0.2 = 0.10	0.10 x 0.6 x 0.2 + 0.15 x 0.5 x 0.2 = 0.027

C
0.0345 x 0.5 x 0.3 + 0.027 x 0.4 x 0.3 = 0.008415
0.027 x 0.6 x 0.2 + 0.0345 x 0.5 x 0.2 = 0.00669

A
0.008415 x 0.5 x 0.2 + 0.00669 x 0.4 x 0.3 = 0.00206505
0.00669 x 0.6 x 0.3 + 0.008415 x 0.5 x 0.3 = 0.00246645

Thus, the probability of the observation O=GGCA given the model = 0.00206505 + 0.00246645 = 0.00453150.
The probability of the observation O=GGCA with a random model = $(0.25)^4$ = 0.00390625
Thus, the given HMM performs better than a random model.

What is the most probable “path” for generating a given sequence?

Let us define $v_k(i)$ denote the probability of the most probable path accounting for the first *i* characters of *O* (length *L*) and has ended in state *k*.

It means that we want to find the value of $v_N(L)$, the probability of the most probable path accounting for all the sequences and ending in the ‘end’ state.

(1) Initialization

$v_0(0) = 1$

$v_k(0) = 0$ $\forall$ k that are not silent states (k = 1, 2, …, N)

(2) Recursion for emitting states (i=1, 2, …, L)

$$v_l(i) = e_l(i)\, max_k\, [vk(O_{i\ \_\ 1})a_{kl} \qquad 1 \le i \le L,\ 1 \le k, l \le \text{N}$$
$$ptr_l(i) = \arg max_k\, [vk(O_{i\ \_\ 1})a_{kl}]$$

What is the most probable "path" for generating a given sequence?

(3) Termination

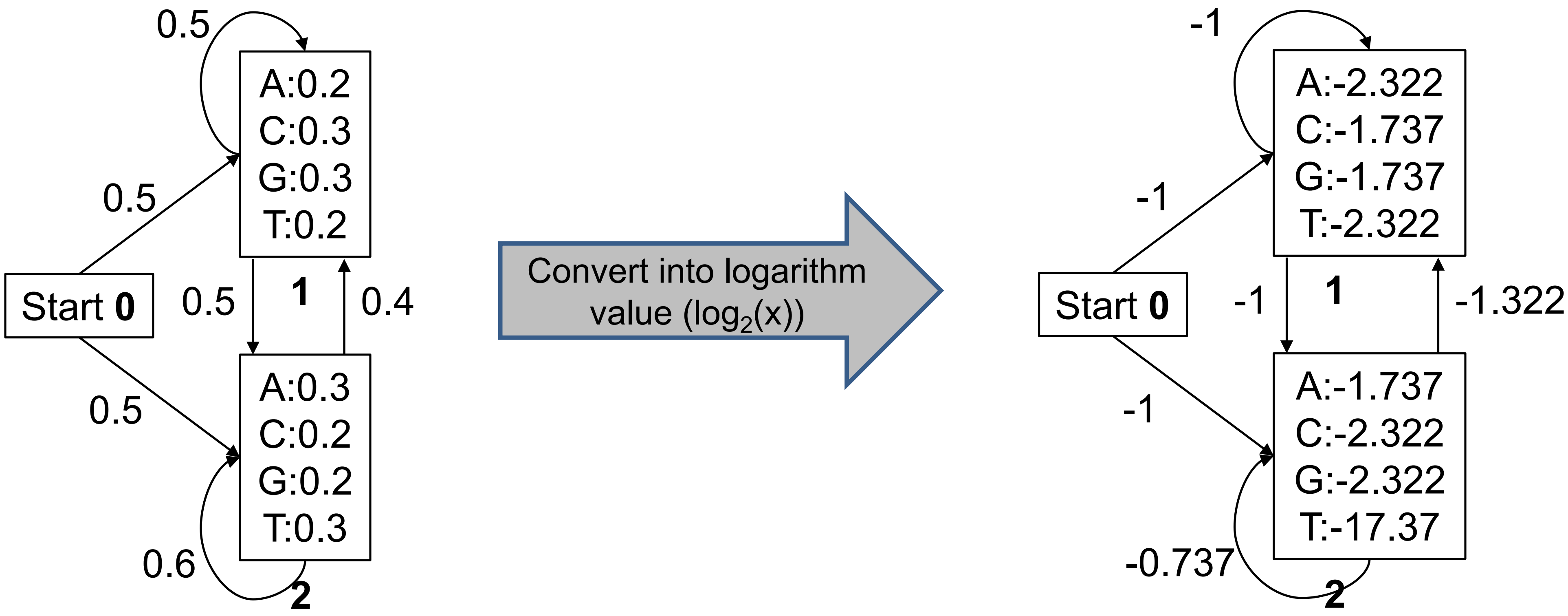
$$\mathrm{P}(\mathrm{O}) \ = \ max_k \ [vk(L)a_{kN}$$
$$\pi_L = \arg max_k \ [vk(L)a_{kN}]$$

(4) Trace back

Follow pointers back starting at π_L.

What is the best possible path to generate a given sequence: the Viterbi algorithm

Given the following HMM, find the best possible path to generate O=GGCACTGAA using the Viterbi algorithm



What is the best possible path to generate a given sequence: the Viterbi algorithm

Probability that G at the first position was emitted by state 1
$v_1(G,1) = -1 + (-1.737) = -2.737$

Similarly, probability that G at the first position was emitted by state 2
$v_2(G,1) = -1 + (-2.322) = -3.322$

Probability that G at the second position was emitted by state 1

$$\begin{aligned} v_1(G,2) &= -1.737 + \max [\, v_1(G,1) + a_{11},\ v_2(G,1) + a_{21}] \\ &= -1.737 + \max [\, -2.737 -1,\ -3.322 - 1.322] \\ &= -1.737 + \max [\, -3.737,\ -4.644] \\ &= -1.737 -3.737 = -5.474 \end{aligned}$$

Similarly, probability that G at the second position was emitted by state 2

$$\begin{aligned} v_2(G,2) &= -2.322 + \max [\, v_1(G,1) + a_{12},\ v_2(G,1) + a_{22}] \\ &= -2.322 + \max [\, -2.737 -1,\ -3.322 - 0.737] \\ &= -2.322 + \max [\, -3.737,\ -4.059] \\ &= -2.322 -3.737 = -6.059 \end{aligned}$$

Similarly, all the probabilities of finding different characters at different positions can be calculated.

What is the best possible path to generate a given sequence: the Viterbi algorithm

The final probability and path (underlined) are follows. Note: (1) kindly check if the values are correct and (2) in case of equal probability, any state randomly can be chosen as the best possible state, producing that character.

States	Start	G	G	C	A	C	T	G	A	A
1	0	<u>-2.737</u>	<u>-5.47</u>	<u>-8.21</u>	-11.53	-14.01	-17.332	<u>-19.543</u>	-22.865	-25.65
2	0	-3.322	-6.06	<u>-8.79</u>	<u>-10.94</u>	<u>-14.01</u>	<u>-16.484</u>	-19.543	<u>-22.017</u>	<u>-24.49</u>

Thus, the most probable path to generate the sequence GGCACTGAA is 111222122 with a probability of $2^{-24.49}$ = approx. 4.25×10^{-8}.

Thank You