CENG 465

Introduction to Bioinformatics

Spring '2019-2020

Assignment 2

Deadline: April 5, 23:59

Submission: via ODTUCLASS

Student Information

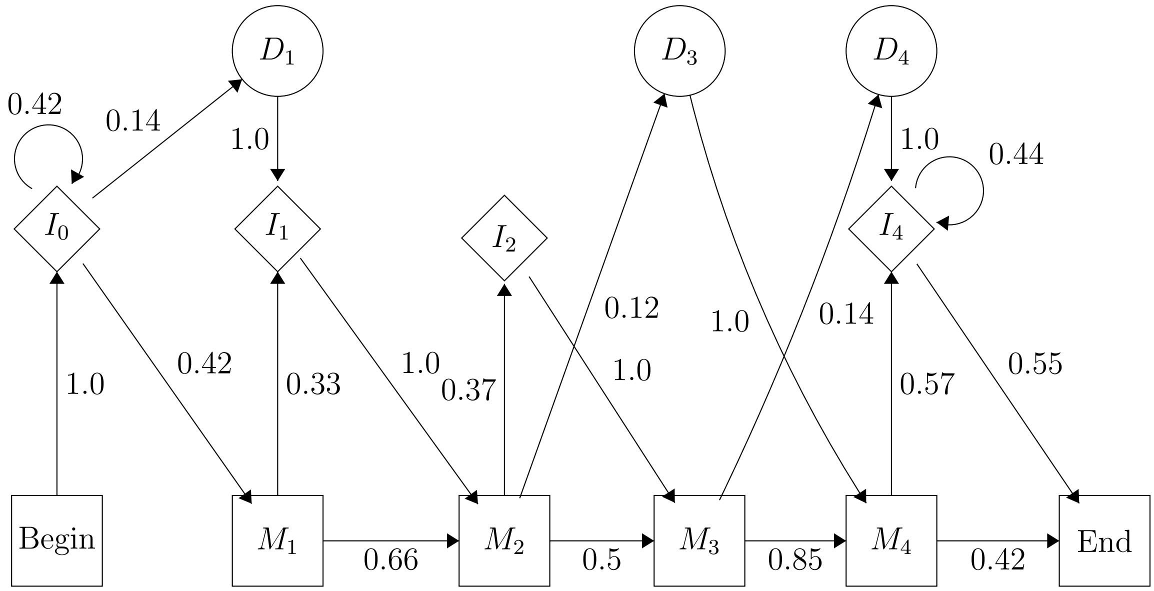
Full Name : Yasin Fatih ALPUL
Id Number : 2098739

Part A

Below are the state transitions for the 8 sequences.

$$1 : \text{Begin} \rightarrow I_0 \rightarrow I_0 \rightarrow D_1 \rightarrow I_1 \rightarrow M_2 \rightarrow M_3 \rightarrow M_4 \rightarrow I_4 \rightarrow \text{End}$$
$$2 : \text{Begin} \rightarrow I_0 \rightarrow I_0 \rightarrow M_1 \rightarrow I_1 \rightarrow M_2 \rightarrow M_3 \rightarrow M_4 \rightarrow I_4 \rightarrow I_4 \rightarrow \text{End}$$
$$3 : \text{Begin} \rightarrow I_0 \rightarrow M_1 \rightarrow I_1 \rightarrow M_2 \rightarrow I_2 \rightarrow M_3 \rightarrow M_4 \rightarrow \text{End}$$
$$4 : \text{Begin} \rightarrow I_0 \rightarrow I_0 \rightarrow M_1 \rightarrow M_2 \rightarrow D_3 \rightarrow M_4 \rightarrow I_4 \rightarrow I_4 \rightarrow I_4 \rightarrow \text{End}$$
$$5 : \text{Begin} \rightarrow I_0 \rightarrow D_1 \rightarrow I_1 \rightarrow M_2 \rightarrow M_3 \rightarrow M_4 \rightarrow \text{End}$$
$$6 : \text{Begin} \rightarrow I_0 \rightarrow I_0 \rightarrow M_1 \rightarrow M_2 \rightarrow I_2 \rightarrow M_3 \rightarrow M_4 \rightarrow I_4 \rightarrow I_4 \rightarrow \text{End}$$
$$7 : \text{Begin} \rightarrow I_0 \rightarrow I_0 \rightarrow M_1 \rightarrow M_2 \rightarrow I_2 \rightarrow M_3 \rightarrow M_4 \rightarrow \text{End}$$
$$8 : \text{Begin} \rightarrow I_0 \rightarrow I_0 \rightarrow M_1 \rightarrow M_2 \rightarrow M_3 \rightarrow D_4 \rightarrow I_4 \rightarrow \text{End}$$

By computing the frequencies of state transitions, we can create the following profile Hidden Markov Model. Note that the computations were conducted as finding the number of transitions from a state and finding the number of transitions to each of other states. For example, we have 14 transitions originating from I_0 and 6 of them goes to M_1. Thus $I_0 \rightarrow M_1$ transitions should have $6/14 = 0.42$ probability.



$\pi_{Begin} = 1.0$ and π for all other states is 0.0.
The following table shows the emission probabilities.

	ω_A	ω_C	ω_G	ω_T
M_1	0.16	0.83	0.0	0.0
M_2	0.0	0.0	0.13	0.88
M_3	0.0	0.86	0.14	0.0
M_4	0.71	0.0	0.0	0.29
I_0	0.25	0.25	0.25	0.25
I_1	0.25	0.25	0.25	0.25
I_2	0.25	0.25	0.25	0.25
I_3	0.25	0.25	0.25	0.25
I_4	0.25	0.25	0.25	0.25

Table 1: Emission Probabilities

Delete states and Begin-End states do not emit any symbols.

The table below also shows the transition probabilities as a table. Note that each row sums up to 1 except for the End state.

	Begin	M_1	M_2	M_3	M_4	I_0	I_1	I_2	I_4	D_1	D_3	D_4	End
Begin	0.0	0.0	0.0	0.0	0.0	**1.0**	0.0	0.0	0.0	0.0	0.0	0.0	0.0
M_1	0.0	0.0	**0.66**	0.0	0.0	0.0	**0.33**	0.0	0.0	0.0	0.0	0.0	0.0
M_2	0.0	0.0	0.0	**0.5**	0.0	0.0	0.0	**0.37**	0.0	0.0	**0.12**	0.0	0.0
M_3	0.0	0.0	0.0	0.0	**0.85**	0.0	0.0	0.0	0.0	0.0	0.0	**0.14**	0.0
M_4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**0.57**	0.0	0.0	0.0	**0.42**
I_0	0.0	**0.42**	0.0	0.0	0.0	**0.42**	0.0	0.0	0.0	**0.14**	0.0	0.0	0.0
I_1	0.0	0.0	**1.0**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
I_2	0.0	0.0	0.0	**1.0**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
I_4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**0.44**	0.0	0.0	0.0	**0.55**
D_1	0.0	0.0	0.0	0.0	0.0	0.0	**1.0**	0.0	0.0	0.0	0.0	0.0	0.0
D_3	0.0	0.0	0.0	0.0	**1.0**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
D_4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**1.0**	0.0	0.0	0.0	0.0
End	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 2: Transition Probabilities

Part B

	Begin	I_0	M_1	I_1	D_1	M_2	I_2	M_3	D_3	M_4	I_4	D_4	End
""	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
C	0.0	0.25	0.0	0.0	3.57e-2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T	0.0	2.62e-2	0.0	8.93e-3	3.67e-3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
C	0.0	2.75e-3	9.13e-3	9.18e-4	3.85e-4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T	0.0	2.89e-4	0.0	7.53e-4	4.05e-5	5.3e-3	0.0	0.0	6.36e-4	0.0	0.0	0.0	0.0
G	0.0	3.03e-5	0.0	1.01e-5	4.24e-6	9.79e-5	4.9e-4	3.71e-4	1.17e-5	0.0	0.0	5.19e-5	0.0
A	0.0	3.18e-6	2.04e-6	1.06e-6	4.45e-7	0.0	9.06e-6	0.0	0.0	2.24e-4	1.3e-5	0.0	9.4e-5

Table 3: Partial Probability Table

If we trace back the table that we filled using *Viterbi Algorithm*, we get the following sequence of states:

$$\text{Begin} \rightarrow I_0 \rightarrow I_0 \rightarrow M_1 \rightarrow M_2 \rightarrow M_3 \rightarrow M_4 \rightarrow \text{End}$$

According to the table, we end up in M_4 state with $2.24e-4$ probability and since we go to *End* state with 0.42 probability, we get $2.24e-4 \times 0.42 = 9.4e-5$.