

HMM for DNA Sequence

In this problem, you will use HMM to decode a simple DNA sequence. It is well known that a DNA sequence is a series of components from $\{A, C, G, T\}$. Now let's assume there is one hidden variable S that controls the generation of DNA sequence. S takes 2 possible states $\{S_1, S_2\}$. Assume the following transition probabilities for HMM M

$$P(S_1|S_1) = 0.8, P(S_2|S_1) = 0.2, P(S_1|S_2) = 0.2, P(S_2|S_2) = 0.8$$

emission probabilities as following

$$\begin{aligned} P(A|S_1) &= 0.4, P(C|S_1) = 0.1, P(G|S_1) = 0.4, P(T|S_1) = 0.1 \\ P(A|S_2) &= 0.1, P(C|S_2) = 0.4, P(G|S_2) = 0.1, P(T|S_2) = 0.4 \end{aligned}$$

and start probabilities as following

$$P(S_1) = 0.5, P(S_2) = 0.5$$

Assume the observed sequence is $x = CGTCAG$, calculate:

- $P(x|M)$