MATH 448 — Homework 7

Exercise 11.1. Three sequences are aligned:

$$\begin{aligned} S_0&: \ \texttt{ACGTTAGCAT} \\ S_1&: \ \texttt{ACTTCAGCAT} \\ S_2&: \ \texttt{ACATTAGTAT} \end{aligned}$$

S_1 descends from S_0, and S_2 descends from S_1. (a) What fraction of sites differ between S_0 and S_2? (b) List every substitution from S_0 to S_1 and from S_1 to S_2, and label each as a transition or a transversion. How many substitutions occurred in all, per site? (c) Explain why (a) and (b) differ, pointing to the sites responsible.

Exercise 11.2. The sequence

$$\texttt{AGGATTTAATCACGTATCGT}$$

has length 20. Treat each site as a trial. (a) Estimate $P(\text{A})$, $P(\text{G})$, $P(\text{C})$, $P(\text{T})$ and check that they add to 1. (b) Estimate $P(\text{purine})$. (c) Why can you add the probabilities in (b)?

Exercise 11.3. Use the model of Problem 11.4 (one site, change probability 0.02 per generation, independently of the other generations) but over three generations. Find the probability that the site is in its original class (purine or pyrimidine) after three generations. *Hint:* List the sequences of change and no change that end in the original class.

Exercise 11.4. Two aligned sequences, an ancestor S_0 and a descendant S_1:

$$\begin{aligned} S_0&: \ \texttt{GACCTATCCTCATATC} \\ S_1&: \ \texttt{GACCCATTTTCATGTC} \end{aligned}$$

(a) Find $P(S_1 = \text{T} \mid S_0 = \text{T})$ and $P(S_1 = \text{C} \mid S_0 = \text{T})$, and check that they add to 1 with the other possibilities. (b) Find $P(S_0 = \text{T} \text{ and } S_1 = \text{T})$. (c) Verify that your answers to (a) and (b) satisfy $P(F \mid E) = P(F \text{ and } E)/P(E)$, with E the event $S_0 = \text{T}$ and F the event $S_1 = \text{T}$.

Exercise 12.1. Use the aligned sequences of Exercise 11.4. (a) Write the matrix M with entries $P(S_1 = i \mid S_0 = j)$ and the vector $\mathbf{p}_0$ of base probabilities in S_0, with the bases in the order A, G, C, T. (b) Compute $\mathbf{p}_1 = M\mathbf{p}_0$. (c) Count the bases in S_1 directly and compare with $\mathbf{p}_1$.

Project proposal. Write half a page naming the model you intend to study, the question you want it to answer, and where your data will come from if you are using any. You may work alone or in pairs; if you work in a pair, hand in one proposal with both names. You are allowed to change your topic later if you find one that interests you more. The purpose of this exercise is that you have at least one topic to start from.