MATH 448 — Homework 7 — Solutions

Exercise 11.1. (a) Sites 3 and 8 differ between S_0 and S_2 (G vs. A and C vs. T), so the fraction is $2/10 = 0.2$. (b) From S_0 to S_1: site 3, G→T (transversion); site 5, T→C (transition). From S_1 to S_2: site 3, T→A (transversion); site 5, C→T (transition); site 8, C→T (transition). That is 5 substitutions in 10 sites, 0.5 per site. (c) Site 3 changed twice (G→T→A), so S_0 and S_2 still differ there, but the two substitutions show up as one difference. Site 5 changed and then changed back (T→C→T), so S_0 and S_2 agree there although two substitutions occurred. Only site 8 is a single visible substitution. Comparing S_0 and S_2 directly counts 2 where 5 occurred.

Exercise 11.2. (a) Counting gives 6 A's, 4 G's, 3 C's, 7 T's, so

$$P(\text{A}) \approx \tfrac{6}{20} = .30, \quad P(\text{G}) \approx \tfrac{4}{20} = .20, \quad P(\text{C}) \approx \tfrac{3}{20} = .15, \quad P(\text{T}) \approx \tfrac{7}{20} = .35,$$

and $.30 + .20 + .15 + .35 = 1$. (b) The purines are A and G, so $P(\text{purine}) \approx .30 + .20 = .50$, which agrees with 10/20. (c) A site cannot be both A and G, so the events "A" and "G" are mutually exclusive and the addition rule applies.

Exercise 11.3. Over three generations there are 8 equally structured sequences of change (c) and no change (n). The site is in its original class after an even number of changes: none, or exactly two. Zero changes: nnn, with probability $(.98)^3 = .941192$. Exactly two changes: ccn, cnc, ncc, each with probability $(.02)^2(.98) = .000392$, so $3(.000392) = .001176$. These outcomes are mutually exclusive, so the probability is

$$.941192 + .001176 = .942368.$$

Exercise 11.4. The sites with T in S_0 are 5, 7, 10, 13 and 15, which is 5 of the 16 sites. At those sites S_1 has C, T, T, T, T. (a) $P(S_1 = \text{T} \mid S_0 = \text{T}) = 4/5 = .8$ and $P(S_1 = \text{C} \mid S_0 = \text{T}) = 1/5 = .2$. The probabilities of A and G given $S_0 = \text{T}$ are 0, and $.8 + .2 = 1$. (b) Four of the 16 sites have T in both sequences, so $P(S_0 = \text{T and } S_1 = \text{T}) = 4/16 = .25$. (c) $P(E) = 5/16$, so $P(F \mid E) = \dfrac{4/16}{5/16} = \dfrac{4}{5}$, which matches (a).

Exercise 12.1. (a) The count table (rows S_1, columns S_0, order A, G, C, T) is

$S_1 \backslash S_0$	A	G	C	T
A	3	0	0	0
G	1	1	0	0
C	0	0	4	1
T	0	0	2	4
column sum	4	1	6	5

Dividing each column by its sum,

$$M = \begin{pmatrix} 3/4 & 0 & 0 & 0 \\ 1/4 & 1 & 0 & 0 \\ 0 & 0 & 2/3 & 1/5 \\ 0 & 0 & 1/3 & 4/5 \end{pmatrix}, \qquad \mathbf{p}_0 = \frac{1}{16}\begin{pmatrix} 4 \\ 1 \\ 6 \\ 5 \end{pmatrix}.$$

(b) $\mathbf{p}_1 = M\mathbf{p}_0 = \dfrac{1}{16}(3,\ 1+1,\ 4+1,\ 2+4) = \dfrac{1}{16}(3,2,5,6) = (.1875,\ .125,\ .3125,\ .375)$. (c) Counting S_1 directly gives 3 A's, 2 G's, 5 C's, and 6 T's, so the base probabilities are $(3,2,5,6)/16$, which agrees with $\mathbf{p}_1$. Each product $P(S_1 = i \mid S_0 = j)P(S_0 = j)$ is $P(S_1 = i \text{ and } S_0 = j)$, and the sum over j is $P(S_1 = i)$.