Mathematical Models in Biology *An Introduction*
by Elizebeth Allman and John A. Rhodes

4.2.4. a. $P(A) \approx .05$, $P(G) \approx .4$, $P(C) \approx .3$, $P(T) \approx .25$

b. $P(purine) \approx .45$, $P(pyrimidine) \approx .55$

c. G, a purine, is the most likely base. This may, at first, appear to contradict part (b) which shows the base is most likely to be a pyrimidine. However, there is no real contradiction: While G is the most likely base, the probability of either a C or T is higher than that of an A or G.

4.2.7. a. not mutually exclusive, independent

b. mutually exclusive, dependent

c. not mutually exclusive, dependent

4.2.8. Two mutually exclusive events E_1 and E_2 with positive probabilities can not be independent since $0 = P(E_1 \cap E_2) \neq P(E_1)P(E_2) > 0$. More informally, if

the events cannot occur together, then knowing whether one has occurred does give us information as to whether the other has.

4.3.1. a. $\{(F, F), (F, M), (M, F), (M, M)\}$, all with probability .25

b. 3/4

c. 1/2

d. 2/3

e. 1

f. No. Knowledge that one child is female effects the likelihood that the youngest child is female, since $1/2 \neq 2/3$. Alternately, $3/4 \neq 1$ shows that knowledge that the youngest child is female affects the likelihood that one child is female.

4.3.8. a. $P(S_0 = A) = .225$, $P(S_0 = G) = .275$, $P(S_0 = C) = .275$, $P(S_0 = T) = .225$, $P(S_1 = A) = .225$, $P(S_1 = G) = .3$, $P(S_1 = C) = .275$, $P(S_1 = T) = .2$.

b. No, since $P(S_1 = i$ and $S_0 = j) \neq P(S_0 = i)P(S_1 = j)$. For instance, since $P(S_1 = i$ and $S_0 = j) = (1/40)$(the (j, i) entry of the table), we find $P(S_1 = A$ and $S_0 = A) = 7/40 = .175 \neq (.225)(.225) = .050625$.

c. Since the sequences are related and mutations are rare, the appearance of a particular base at a site in S_0 means it is highly probable that the same base would appear at the same site in S_1, i.e. the events $\{S_0 = i\}$ and $\{S_1 = j\}$ are not independent.

4.3.9. a. Since there is no relationship between the two sequences, knowing information about one should convey nothing about the other.

b. All the columns would be the same.

4.3.10. a. The formula calculates the conditional probability of a purine occurring in S_2 given a purine occurred in S_0 by accounting for either a purine or a pyrimidine occurring in the intermediate sequence S_1.
$P(S_2 = pyr \mid S_0 = pur) = P(S_2 = pyr \mid S_1 = pur) \cdot P(S_1 = pur \mid S_0 = pur) + P(S_2 = pyr \mid S_1 = pyr) \cdot P(S_1 = pyr \mid S_0 = pur)$, etc.

b. $P(S_2 = pur \mid S_0 = pur) = .9606$; $P(S_2 = pyr \mid S_0 = pur) = .0394$; $P(S_2 = pur \mid S_0 = pyr) = .0197$; $P(S_2 = pyr \mid S_0 = pyr) = .9803$

c. Note that with the given assumptions

$$P(S_2 = pur \mid S_1 = pur) \cdot P(S_1 = pur \mid S_0 = pur)$$
$$= P(S_2 = pur \mid S_1 = pur \text{ and } S_0 = pur) \cdot P(S_1 = pur \mid S_0 = pur)$$
$$= \frac{P(S_2 = pur \text{ and } S_1 = pur \text{ and } S_0 = pur)}{P(S_1 = pur \text{ and } S_0 = pur)} \cdot \frac{P(S_1 = pur \text{ and } S_0 = pur)}{P(S_0 = pur)}$$
$$= \frac{P(S_2 = pur \text{ and } S_1 = pur \text{ and } S_0 = pur)}{P(S_0 = pur)}.$$

Similarly,

$$P(S_2 = pur \mid S_1 = pyr) \cdot P(S_1 = pyr \mid S_0 = pur)$$
$$= \frac{P(S_2 = pur \text{ and } S_1 = pyr \text{ and } S_0 = pur)}{P(S_0 = pur)}.$$

Therefore

$$P(S_2 = pur \mid S_1 = pur) \cdot P(S_1 = pur \mid S_0 = pur)$$
$$+ P(S_2 = pur \mid S_1 = pyr) \cdot P(S_1 = pyr \mid S_0 = pur)$$
$$= \frac{P(S_2 = pur, S_1 = pur, S_0 = pur) + P(S_2 = pur, S_1 = pyr, S_0 = pur)}{P(S_0 = pur)}$$
$$= \frac{P(S_2 = pur, S_0 = pur)}{P(S_0 = pur)} = P(S_2 = pur \mid S_0 = pur).$$

4.4.1. a. A plot in the forest can be in the state "occupied by an A tree" or the state "occupied by a B tree."

b. All the entries are non-negative and the column sums are one.

c. $(1,1)$ entry: the conditional probability that a spot which is occupied by an A tree in one year remains occupied by an A tree the next year; $(1,2)$ entry: the conditional probability that a spot which is occupied by a B tree in one year is occupied by an A tree the next year; $(2,1)$ entry: the conditional probability that a spot which is occupied by an A tree in one year is occupied by an B tree the next year; $(2,2)$ entry: the conditional probability that a spot which is occupied by a B tree in one year remains occupied by an B tree the next year

d. $(.01, .99)$

4.4.4. a. $\alpha = .06$ is faster.

b. Yes.

c. The larger the value of α, the more mutation occurs and the quicker any initial vector $\mathbf{p_0}$ will move towards equilibrium.

4.4.8. a. The first theorem applies to M, but the second does not since M has some zero entries. (However, since M^2 has all non-zero entries, you can apply the second theorem to it.)

b. $(.1849, .3946, .2819, .1386)$

4.4.9. a. $\mathbf{p_0} = (.3, .225, .25, .225)$, $M = \begin{pmatrix} .833 & 0 & 0 & .111 \\ .083 & .889 & 0 & 0 \\ 0 & .111 & 1 & .111 \\ .083 & 0 & 0 & .778 \end{pmatrix}$

b. $\mathbf{p_0}$ is reasonable close to $(.25, .25, .25, .25)$. M may seem less close to a Jukes-Cantor matrix than you might expect, because of the variation in the off-diagonal entries. One way to estimate α is to average the off-diagonal entries to estimate $\alpha/3$. This gives $\alpha/3 = .0416$, so $\alpha = .1248$.