Allman Ch 4.4 Prob 3,5,8,9,10, 15 Allman Ch 4.5 Ex 1,3,5,7,13

4.4.3. a. About 27 steps to be within .05; about 67 steps to be within .01.

4.4.5. Because mutation is rare, the conditional probabilities describing no change should be largest.

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 $\alpha = .1248$.

4.4.10. a. The Jukes-Cantor model is more appropriate for the pair S_0', S_1', since a particular base seems to mutate to any of the other three bases with roughly the same frequency. Note also that the bases in S_0' are in roughly equal numbers.
b. The Kimura 2-parameter model is more appropriate for the pair S_0, S_1, since the data shows that transitions are more likely than transversions. Note also that the bases in S_0 are in roughly equal numbers.

4.4.15. a. $\mathbf{p}_0 = (.15, .25, .35, .25)$ is not an equilibrium base distribution for the Jukes-Cantor matrix $M = \begin{pmatrix} .7 & .1 & .1 & .1 \\ .1 & .7 & .1 & .1 \\ .1 & .1 & .7 & .1 \\ .1 & .1 & .1 & .7 \end{pmatrix}$

b. $\mathbf{p}_0 = (.19, .25, .31, .25)$ is not an equilibrium base for the transition matrix
$M = \begin{pmatrix} .5526 & .06 & .0484 & .06 \\ .1316 & .7 & .0806 & .1 \\ .1842 & .14 & .7903 & .14 \\ .1316 & .1 & .0806 & .7 \end{pmatrix}$, which is not Jukes-Cantor.

c. $\mathbf{p}_0$ is unchanged by multiplying by M; it is an equilibrium. Notice that $\mathbf{p}_0$ is an eigenvector of M with eigenvalue 1.
d. The initial vector is drawn towards the stable equilibrium $(.25, .25, .25, .25)$. This $\mathbf{p}_0$ corresponds to an initial sequence comprised entirely of G's.

4.5.1. .1367

4.5.3. a. .2224580274

b. .2308224444

c. The Kimura 2-parameter distance is probably a better choice (assuming we did not already know that the sequences were created with the Kimura 2-parameter model). The frequency table shows a definite pattern of more transitions than transversions. Notice too that the distances differ in the second decimal position.

4.5.5. Graph the Jukes-Cantor distance on a graphing calculator or computer.

a. If the sequences are identical, then $p = 0$. This means the Jukes-Cantor distance is $-.75\log(1) = 0$.

b., c. Mathematically, if two sequences differ in more than 3/4 of the sites, then $p > 3/4$. Then the Jukes-Cantor distance formula requires taking the logarithm of a negative number, which is impossible. This is not a limitation with real data. If we took two sequences that were in no way related, we would expect that about 1/4 of the sites agree and about 3/4 of the sites disagree, since with a uniform distribution of bases about 25% of the time the two sequences should agree if everything is chosen at random. For related sequences the formulas for the Jukes-Cantor model derived in the last section show p is at most 3/4, and in practice p is usually much less than 3/4. Notice that the Jukes-Cantor distance gets huge as the values of p get close to .75. This is desirable, since distances should be large when comparing sequences that appear almost unrelated.

4.5.7. Substituting $(1-q)$ for p yields $d_{JC} = -\frac{3}{4}\ln\left(1-\frac{4}{3}p\right) = -\frac{3}{4}\ln\left(1-\frac{4}{3}(1-q)\right) = -\frac{3}{4}\ln\left(\frac{4}{3}q-\frac{1}{3}\right) = -\frac{3}{4}\ln\left(\frac{4q-1}{3}\right)$.

4.5.13. The Kimura 3-parameter distance is given by $d_{K3} = -\frac{1}{4}\left(\ln(1-2\beta-2\gamma) - \ln(1-2\beta-2\delta) - \ln(1-2\beta-2\gamma)\right)$. Substituting $\alpha/3$ for β, γ, and δ gives

$$d = -\frac{1}{4}\left(\ln(1-2\alpha/3-2\alpha/3) - \ln(1-2\alpha/3-2\alpha/3) - \ln(1-2\alpha/3-2\alpha/3)\right)$$
$$= \frac{1}{4}\left(3\ln(1-4\alpha/3)\right) = d_{JC}.$$