Allman Ch 4.3 Prob 3,5,6,7,8,9

Ch 4.3

4.3.3. a. Sensitivity is $\mathcal{P}(+ \text{ result} \mid \text{disease})$; Specificity is $\mathcal{P}(- \text{ result} \mid \text{no disease})$.
b. False positive: $\mathcal{P}(+ \text{ result} \mid \text{no disease})$; False negative: $\mathcal{P}(- \text{ result} \mid \text{disease})$.
c. Sensitivity $= 22/30 = .7333$; Specificity $= 1739/1790 = .9715$.

4.3.5

$S_1 \backslash S_0$	A	G	C	T
A	.778	0	.111	.111
G	.083	.75	.167	0
C	0	.182	.636	.182
T	.125	0	.125	.75

b. $\mathcal{P}(S_1 = i \mid S_0 = j)$ is the conditional probability that given a j in S_0 it mutates to become an i in S_1. However, $\mathcal{P}(S_0 = i \mid S_1 = j)$ is the conditional probability that given a j in the descendent, it came from an i in the ancestor. The first is found by dividing an entry in the table by its column sum, the second by dividing by its row sum.

4.3.6. a. The diagonal entries correspond to no mutation occurring. These are likely to be the largest, since point mutations are rare.
b. Transitions: entries $(1,2)$, $(2,1)$, $(3,4)$, $(4,3)$; Transversions: entries $(1,3)$, $(1,4)$, $(2,3)$, $(2,4)$, $(3,1)$, $(3,2)$, $(4,1)$, $(4,2)$. This table does not support the hypothesis that transitions are more common than transversions.

4.3.7. a. The distribution of bases in S_0 is estimated by $p_A = .225$, $p_G = .275$, $p_C = .275$, $p_T = .225$.
b. The distribution of bases in S_1 is estimated by $p_A = .225$, $p_G = .3$, $p_C = .275$, $p_T = .2$.

4.3.8. a. $\mathcal{P}(S_0 = A) = .225$, $\mathcal{P}(S_0 = G) = .275$, $\mathcal{P}(S_0 = C) = .275$, $\mathcal{P}(S_0 = T) = .225$, $\mathcal{P}(S_1 = A) = .225$, $\mathcal{P}(S_1 = G) = .3$, $\mathcal{P}(S_1 = C) = .275$, $\mathcal{P}(S_1 = T) = .2$.
b. No, since $\mathcal{P}(S_1 = i \text{ and } S_0 = j) \neq \mathcal{P}(S_0 = i)\mathcal{P}(S_1 = j)$. For instance, since $\mathcal{P}(S_1 = i \text{ and } S_0 = j) = (1/40)(\text{the } (j,i) \text{ entry of the table})$, we find $\mathcal{P}(S_1 = A \text{ and } S_0 = A) = 7/40 = .175 \neq (.225)(.225) = .050625$.
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.