M469 Spring 2020, Assignment 7, due Fri. Mar. 20

Suggested Readings. *Snowshoe hare populations: squeezed from below and above*, by Nils Christian Stenseth, and *Impact of food and predation on the snowshoe hare cycle*, by Charles J. Krebs et al, both in Science **269** (1995) respective pages 1061-1062 and 1112-1115. The first of these articles is a review and commentary on the second. In the second article, the authors discuss an eight-year study carried out near Kluane Lake, Yukon (near the Canada-Alaska border in the southwest corner of Yukon) regarding snowshoe hare populations. They are trying to understand the most important factors in the long-observed cyclic patterns of snowshoe hare populations, and in particular they focus on the effects of predation and food abundance. By manipulating both predation patterns and food patterns, they observe that each of these contributes substantially to the cyclic behavior.

The discussion by Stenseth includes an intriguing method for determining the number of significant factors required for modeling an observed process. For example, we would consider competition between individuals in a single species as one factor (leading to a carrying capacity), and predation by a second species as a second factor. A changing environment could be a third factor. Stenseth suggests that we can get a sense of the number of expected effects by determining the number of delay terms required in modeling the population. That is, a single-effect process would be modeled by

$$y_{t+1} = f(y_t),$$

while a two-effect process would be modeled by

$$y_{t+1} = f(y_t, y_{t-1}).$$

See particularly Stenseth's figure and his endnote 7 for his comments on this.

1. [10 pts] Suppose we have a 40-base strand of ancestral DNA, S_0, and an aligned descendent sequence as follows:

$$S_0 : ACTTGTCGGA|TGATCAGCGG|TCCATGCACC|TGACAACGGT$$
$$S_1 : ACATGTTGCT|TGACGACAGG|TCCATGCGCC|TGAGAACGGC$$

(The vertical lines divide the sequence into sub-strands of 10.)

a. Compute the transition matrix

$$M = \begin{pmatrix} P_{A|A} & P_{A|G} & P_{A|C} & P_{A|T} \\ P_{G|A} & P_{G|G} & P_{G|C} & P_{G|T} \\ P_{C|A} & P_{C|G} & P_{G|G} & P_{G|T} \\ P_{T|A} & P_{T|G} & P_{T|C} & P_{T|T} \end{pmatrix}.$$

b. Compute the Jukes-Cantor distance between these strands.

2. [10 pts] Answer the following.

a. Suppose J is the Jukes-Cantor matrix

$$J = \begin{pmatrix} 1-\alpha & \frac{\alpha}{3} & \frac{\alpha}{3} & \frac{\alpha}{3} \\ \frac{\alpha}{3} & 1-\alpha & \frac{\alpha}{3} & \frac{\alpha}{3} \\ \frac{\alpha}{3} & \frac{\alpha}{3} & 1-\alpha & \frac{\alpha}{3} \\ \frac{\alpha}{3} & \frac{\alpha}{3} & \frac{\alpha}{3} & 1-\alpha \end{pmatrix},$$

and use our calculations from class to show that J^t is a Jukes-Cantor method for some appropriate parameter $\tilde{\alpha}$ that will depend on t.

b. Show that if J_1 and J_2 are Jukes-Cantor matrices

$$J_1 = \begin{pmatrix} 1-\alpha_1 & \frac{\alpha_1}{3} & \frac{\alpha_1}{3} & \frac{\alpha_1}{3} \\ \frac{\alpha_1}{3} & 1-\alpha_1 & \frac{\alpha_1}{3} & \frac{\alpha_1}{3} \\ \frac{\alpha_1}{3} & \frac{\alpha_1}{3} & 1-\alpha_1 & \frac{\alpha_1}{3} \\ \frac{\alpha_1}{3} & \frac{\alpha_1}{3} & \frac{\alpha_1}{3} & 1-\alpha_1 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 1-\alpha_2 & \frac{\alpha_2}{3} & \frac{\alpha_2}{3} & \frac{\alpha_2}{3} \\ \frac{\alpha_2}{3} & 1-\alpha_2 & \frac{\alpha_2}{3} & \frac{\alpha_2}{3} \\ \frac{\alpha_2}{3} & \frac{\alpha_2}{3} & 1-\alpha_2 & \frac{\alpha_2}{3} \\ \frac{\alpha_2}{3} & \frac{\alpha_2}{3} & \frac{\alpha_2}{3} & 1-\alpha_2 \end{pmatrix},$$

then the matrix product $J_1 J_2$ is a Jukes-Cantor matrix.

3. [10 pts] Show that the Jukes-Cantor distance is additive in the following sense: for any three aligned genetic sequences S_0, S_1, and S_2,

$$d_{JC}(S_0, S_2) = d_{JC}(S_0, S_1) + d_{JC}(S_1, S_2).$$

Notes. This is the notation of Allman-Rhodes, but it's a bit misleading. Here, we are taking the step from S_0 to S_1 as a single-step process with Jukes-Cantor matrix J_{01} (with parameter α_{01}), and we are taking the step from S_1 to S_2 as a single step process with a *different* Jukes-Cantor matrix J_{12} (with parameter α_{12}). You need to compute the Jukes-Cantor matrix associated with the step from S_0 to S_2, and to find an appropriate expression for α_{02} in terms of α_{01} and α_{12}.

4. [10 pts] As discussed in class, the nucleotides A and G are called purines, while the nucleotides C and T are called pyrimidines. (See our reference Saltzman, Section 3.3.1, for details; it's easy to remember which are which because the ones with y— cytosine and thymine—are pyrimidines.) If a purine is substituted for a purine or a pyrimidine is substituted for a pyrimidine, we refer to the substitution as a transition; if a purine is substituted for a pyrimidine or vice versa, we refer to the substitution as a transversion. Generally, transitions are more common than transversions, because the molecular rearrangement involved is less substantial. The Kimura 2-parameter model assumes transitions occur with a different probability than transversions: in particular, transitions occur with probability β and transversions occur with probability γ. This gives rise to the transition matrix

$$K = \begin{pmatrix} 1-\beta-2\gamma & \beta & \gamma & \gamma \\ \beta & 1-\beta-2\gamma & \gamma & \gamma \\ \gamma & \gamma & 1-\beta-2\gamma & \beta \\ \gamma & \gamma & \beta & 1-\beta-2\gamma \end{pmatrix}.$$

Show that K has the same eigenvectors as the Jukes-Cantor matrix J, and use this observation to identify the eigenvalues of K. Find an expression for the matrix K^t, $t = 1, 2, \ldots$.