

Homework #5 due Monday April 11.

Use an LRT to test for the constancy of the rate of substitutions (the “molecular clock” hypothesis) assuming:

- the simplest model of sequence evolution, the Jukes-Cantor model,
- that each site in a sequence evolves independently from other sites, and
- that a change (or lack of change) on one branch, is independent of what happens on any other branch of the tree.
- the process of sequence evolution has come to equilibrium before the root of the tree for which we have data.

A sequence alignment for 3 species is posted on the website. Omit sites that have - or ? for any of the three sequences.

Treat the first sequence as the “outgroup”. The next 2 sequences are named “in1” and “in2”

The transition probabilities for the Jukes-Cantor model (where  $\nu$  is the expected number of changes per site on a single branch) are:

$$\begin{aligned}\Pr(\text{end} = i \mid \nu, \text{start} = i) &= \frac{1}{4} + \frac{3}{4}e^{-4\nu/3} \\ \text{for } i \neq j: \Pr(\text{end} = j \mid \nu, \text{start} = i) &= \frac{1}{4} - \frac{1}{4}e^{-4\nu/3}\end{aligned}$$

The equilibrium frequencies are simply  $1/4$  for each state.

**#1** What are the parameters in the null and alternative models?

**#2** What is the  $\ln L$  under the null model?

**#3** What is the  $\ln L$  under the alternative model?

**#4** What is the approximate  $P$ -value for the test?

**#5** If  $\rho$  is ratio of lengths for the branches that connect “in2” and “in1” to the rest of the tree, construct an approximate 95% confidence interval for  $\rho$ .