

Assignment 6

1. Use your previous code to read in a data matrix with DNA characters and a tree with branch lengths. Then write new code to calculate the likelihood of the data on the tree using the Jukes Cantor model of DNA evolution. You can use the exact formula for calculating the transition probabilities ($P(v)$) under the Jukes Cantor model given the branch length (v):

$$P(v) = \begin{bmatrix} \frac{1}{4} + \frac{3}{4}e^{-4v/3} & \frac{1}{4} - \frac{1}{4}e^{-4v/3} & \frac{1}{4} - \frac{1}{4}e^{-4v/3} & \frac{1}{4} - \frac{1}{4}e^{-4v/3} \\ \frac{1}{4} - \frac{1}{4}e^{-4v/3} & \frac{1}{4} + \frac{3}{4}e^{-4v/3} & \frac{1}{4} - \frac{1}{4}e^{-4v/3} & \frac{1}{4} - \frac{1}{4}e^{-4v/3} \\ \frac{1}{4} - \frac{1}{4}e^{-4v/3} & \frac{1}{4} - \frac{1}{4}e^{-4v/3} & \frac{1}{4} + \frac{3}{4}e^{-4v/3} & \frac{1}{4} - \frac{1}{4}e^{-4v/3} \\ \frac{1}{4} - \frac{1}{4}e^{-4v/3} & \frac{1}{4} - \frac{1}{4}e^{-4v/3} & \frac{1}{4} - \frac{1}{4}e^{-4v/3} & \frac{1}{4} + \frac{3}{4}e^{-4v/3} \end{bmatrix}$$

Remember to weight the root conditional likelihoods for the different states by the stationary state frequency ($1/4$) of each state when calculating the overall likelihood of the data on the tree.

2. (Optional) Allow the user to choose another model of DNA evolution, such as the Felsenstein 1981 model, and then calculate the likelihood of the tree given the model chosen by the user. Note that the user will also have to specify the model parameters, such as the stationary state frequencies. You may consult Felsenstein's book for the exact formulas of the transition probabilities, or you can use standard linear algebra algorithms to exponentiate the instantaneous rate matrix.