Exercise 1: With

$$P(\ell_i) = \left(P_{x\to y}(\ell_i)\right)_{x,y\in\{A,C,G,T\}}, \qquad W_k = \begin{pmatrix} w_{k,1}(A) & w_{k,2}(A) & \dots & w_{k,n}(A) \\ w_{k,1}(C) & w_{k,2}(C) & \dots & w_{k,n}(C) \\ w_{k,1}(G) & w_{k,2}(G) & \dots & w_{k,n}(G) \\ w_{k,1}(T) & w_{k,2}(T) & \dots & w_{k,n}(T) \end{pmatrix},$$

can the Felsenstein pruning recursion really be written as $W_k = (P(\ell_i) \cdot W_i) \circ (P(\ell_j) \cdot W_j)$? Or is it maybe $W_k = (W_i \cdot P(\ell_i)) \circ (W_j \cdot P(\ell_j))$? Or perhaps $W_k = (W_i \circ P(\ell_i)) \cdot (W_j \circ P(\ell_j))$? Check which of these equations is/are correct.

Exercise 2: A Markov process $X_1, X_2, X_3, \dots$ on three states a, b, c has the transition matrix

$$M = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

for one time step.

(a) Calculate the following probabilities with paper, pencil and a pocket calculator (or with R, using only functions that you have also on your pocket calculator): $\Pr(X_2 = b \mid X_1 = a)$, $\Pr(X_5 = b \mid X_1 = a)$, $\Pr(X_5 = b \mid X_1 = b)$.

(b) Use R to calculate the for the transition matrix for 20 time steps.

(c) Calculate the stationary distribution of the Markov chain.

Exercise 3: Calculate the equilibrium distributions of the Markov process with rate matrix Q and check whether the processes are reversible:

$$Q = \begin{pmatrix} -1.0 & 0.4 & 0.6 \\ 0.6 & -0.9 & 0.3 \\ 1.5 & 0.5 & -2.0 \end{pmatrix}$$

Exercise 4:

$$M = \begin{pmatrix} 0.5 & 0.5 \\ 0.2 & 0.8 \end{pmatrix}$$

(a) Calculate with paper and pencil the eigenvalues and the left and the right eigenvectors of the matrix M.

(b) To what distribution will a Markov process converge if M is its transition matrix?

(c) Calculate M^{100} with paper, pencil and pocket calculator (or R, but only for calculations that are also possible with a common pocket calculator).

Exercise 5: Calculate the transition matrix for $t = 0.4$ for the nucleotide-substitution model rate-matrix

$$Q = \begin{pmatrix} -0.10 & 0.03 & 0.040 & 0.030 \\ 0.02 & -0.13 & 0.020 & 0.090 \\ 0.04 & 0.03 & -0.085 & 0.015 \\ 0.02 & 0.09 & 0.010 & -0.120 \end{pmatrix}.$$

(a) Calculate this in `R` using the command `expm` of the `Matrix` package.

(b) Compare this to the solution that you get when using the approximation $(1 + (t/n) \cdot Q)^n$ of the matrix exponential when the matrix power is calculated in `R` by iterating the usual matrix multiplication (`R` operation: `%*%`). That is, avoid calculating eigenvectors and eigenvalues.

(c) Compare both solutions to what you get when you calculate the matrix exponential with the eigenvectors calculated by the R command `eigen`. You can also use R matrix operations and funtions like `t`, `diag` and `solve`, but no functions from special packages for computing matrix exponentials.

(d) Compare the different ways of calculating matrix exponentials also for Felsenstein-84 model rate matrix, for which you can calculate the transition probabilities also without matrix operations.

Exercise 6: Assume an F84 substitution model with nucleotide frequencies $(\pi_A, \pi_C, \pi_G, \pi_T) = (0.2, 0.3, 0.3, 0.2)$, a rate $\lambda = 0.1$ of "crosses" and a rate $\mu = 0.2$ of "bullets" (see lecture).

(a) Assume that the nucleotide at some site is A at time t. Calculate the probabilities that the nucleotide is A, C or G at time $t + 0.2$.

(b) Assume the nucleotide distribution in a genomic region is $(0.1, 0.2, 0.3, 0.4)$ at time t, but from this time on the genomic region evolves according to the model above. Calculate the expectation values for the nucleotide distributions at time points $t + 0.2$ and $t + 2$.

(c) Do (b) also for other time points $t + s$ and visualize how the nucleotide frequencies change in time.

Exercise 7: Let nucleotide frequencies $(\pi_A, \pi_C, \pi_G, \pi_T)$ and mutation rates α and β of the HKY model be given. Under which conditions can you find rates λ and μ of the Felsenstein 84 model, such that the transition matrices of the two models are the same, and which are the appropriate values for λ and μ? Discuss also the opposite direction. The F84 rate matrix is:

$$\begin{pmatrix} -\lambda(1-\pi_A) - \frac{\mu\pi_G}{\pi_A+\pi_G} & \lambda\pi_C & \lambda\pi_G + \frac{\mu\pi_G}{\pi_A+\pi_G} & \lambda\pi_T \\ \lambda\pi_A & -\lambda(1-\pi_C) - \frac{\mu\pi_T}{\pi_C+\pi_T} & \lambda\pi_G & \lambda\pi_T + \frac{\mu\pi_T}{\pi_C+\pi_T} \\ \lambda\pi_A + \frac{\mu\pi_A}{\pi_A+\pi_G} & \lambda\pi_C & -\lambda(1-\pi_G) - \frac{\mu\pi_A}{\pi_A+\pi_G} & \lambda\pi_T \\ \lambda\pi_A & \lambda\pi_C + \frac{\mu\pi_C}{\pi_C+\pi_T} & \lambda\pi_G & -\lambda(1-\pi_T) - \frac{\mu\pi_C}{\pi_C+\pi_T} \end{pmatrix}$$

You can insert positive values for the F84 parameters.

Exercise 8: Calculate the rate matrix for the nucleotide substution process for which the substitution matrix for time t is

$$S(t) = \begin{pmatrix} P_{A\to A}(t) & \frac{1-e^{-t/10}}{10} & \frac{21+9e^{-t/10}-30e^{-t/5}}{70} & \frac{1-e^{-t/10}}{5} \\ \frac{2-2e^{-t/10}}{5} & P_{C\to C}(t) & \frac{3-3e^{-t/10}}{10} & \frac{3+7e^{-t/10}-10e^{-t/5}}{15} \\ \frac{14+6e^{-t/10}-20e^{-t/5}}{35} & \frac{1-e^{-t/10}}{10} & P_{G\to G}(t) & \frac{1-e^{-t/10}}{5} \\ \frac{2-2e^{-t/10}}{5} & \frac{3+7e^{-t/10}-10e^{-t/5}}{30} & \frac{3-3e^{-t/10}}{10} & P_{T\to T}(t) \end{pmatrix},$$

where the diagonal entries $P_{A\to A}(t)$, $P_{C\to C}(t)$, $P_{G\to G}(t)$ and $P_{T\to T}(t)$ are the values that fulfill that each row sum is 1 in the matrix $S(t)$.