Exercise 1: The two values 1.0 and 1.1 are independently drawn from the same normal distribution $\mathcal{N}(\mu, \sigma^2)$ with unknown μ and unknown σ^2. We would like to know whether $\mu = 0$.

(a) Apply a t-test to address this question.

(b) Compute Bayes-factors to compare the following two hypotheses

H_0 : $\mu = 0$ and σ^2 has a uniform prior on $[0, s]$

H_A : μ has a uniform prior on $[-m, m]$ and σ^2 has a uniform prior on $[0, s]$

How does the decision for one or the other model depend on m and s?

Exercise 2: It is known that the balls in an urn are numbered from 1 to n, but there are two hypotheses about n. The null hypothesis H_0 says that $n = 100$, and the alternative hypothesis H_A says that n could be any positive number. Ball number 99 is drawn from the urn.

(a) Discuss how the classical concept of hypothesis testing could be applied to decide whether H_0 can be rejected on the 5% level.

(b) Apply a Bayes factor analysis to decide between the two hypotheses. For H_A assume a uniform prior on $\{1, 2, \ldots, N\}$ for n. How does the result depend on N?

(c) For $N \in \{100, 300, 1000\}$ apply a Bayes factor analysis also for the case that the prior for n on $\{1, 2, \ldots, N\}$ is of the form $(n+1) \cdot (N+1-n) \cdot c_N$ (where c_N is the constant that makes sure that the sum of the probabilities is 1).

Exercise 3: Simulate sequence datasets for various trees and various substitution models. Explore which models are preferred for these datasets by jmodeltest and by the Bayes-factor analysis in MrBayes. How does the accuracy of phylogeny reconstruction depend on which of the proposed models is used?

Exercise 4: Calculate transition probabilities between the states $D = \begin{smallmatrix}\texttt{B}\\\texttt{B}\end{smallmatrix}$, $H = \begin{smallmatrix}\texttt{-}\\\texttt{B}\end{smallmatrix}$ and $V = \begin{smallmatrix}\texttt{B}\\\texttt{-}\end{smallmatrix}$ for the simplified TKF91 model with $\lambda = \mu = 0.2$ and $t = 0.5$.

Exercise 5: With the insertion–deletion model from exercise 4 with $\lambda = \mu = 0.2$ and $t = 0.5$, and the Jukes-Cantor substitution model with the rate matrix

$$\begin{pmatrix} -0.3 & 0.1 & 0.1 & 0.1 \\ 0.1 & -0.3 & 0.1 & 0.1 \\ 0.1 & 0.1 & -0.3 & 0.1 \\ 0.1 & 0.1 & 0.1 & -0.3 \end{pmatrix}$$

carry out the following calculations for the pair of (very short) sequences `AC` and `AGC`. For this, assume that the sequences were taken an alignment of longer sequences and have homologous neighbors. That is, the left and the right neighbors in the alignment are in state $D = \begin{smallmatrix}\texttt{B}\\\texttt{B}\end{smallmatrix}$.

(a) Calculate the probability of the sequence pair by dynamic programming (but without programming) implicitly accounting for all possible alignments.

(b) Sample an alignment according to the posterior probabilities for the given sequences.

(c) Find the most probable alignment for the sequences (according to the posterior probability). For this, modify the dynamic programming algorithm such that each edge in the graph is labeled with the probability of the most probable alignment (including nucleotide types) of the sequence beginnings ending in that edge.

Exercise 6: Calculate the transition probabilities between the states $D = \begin{smallmatrix}\mathtt{B}\\ \mathtt{B}\end{smallmatrix}$, $H = \begin{smallmatrix}\mathtt{-}\\ \mathtt{B}\end{smallmatrix}$ and $V = \begin{smallmatrix}\mathtt{B}\\ \mathtt{-}\end{smallmatrix}$

(a) ... for the FID model (that is, the simplified TKF92 model with $\lambda = \mu$ and fragments of geometrically distributed length) for $\lambda = \mu = 0.2$, $t = 0.5$ and an average fragment length of $\gamma = 3$.

(b) ... for the FID model for general values of $\lambda = \mu$, t and the average fragment length γ.