

Exercise 1: Two alternative phylogenies are discussed for the four taxa A, B, C, D: $((A:1, B:1):1, (C:1, D:1):1)$ and $((A:1, D:1):1, (B:1, C:1):1)$, assuming a Jukes-Cantor sequence evolution model with $P_{X \rightarrow Y}(1) = 0.1$ for $X \neq Y$. The following alignment of DNA sequences of the four taxa is given:

A: ACGT
 B: ACGA
 C: GTGA
 D: GGGT

Apply the Kishino-Hasegawa-Test (stepwisely, without specific software for this test) to check whether the first tree explains the data significantly better than the second.

Exercise 2: Create an unrooted tree with at least 8 leaves in which one of the inner branches has length 0 (or if this is not possible, use a positive value that is as small as possible but still accepted as input for the programs that you use in this exercise). Simulate 100 sequence datasets for this tree with the program `seq-gen`, using a model that is also available in RAxML. Then calculate the log likelihood of the true topology for each of the datasets by using the option “-f e” in RAxML, which involves that RAxML optimizes the branch lengths and the model parameters. Then repeat the RAxML analyses with the two trees that you obtain from the true tree by NNI steps that eliminate the branch that has length 0 (or nearly 0). For each of the 100 dataset calculate the difference of the log likelihood of the most likely tree and that of the most likely that has a different topology than the most likely tree (that is, the second most likely tree among the three trees that are optimal for their respective topology). How well does the distribution of the 100 log likelihood-ratios fit to the chi-square distribution approximation that is applied in Anisimova and Gascuel’s aLRT?

Exercise 3: 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.

- Discuss how the classical concept of hypothesis testing could be applied to decide whether H_0 can be rejected on the 5% level.
- Apply a Bayes factor analysis to decide between the two hypotheses. For H_A assume a uniform prior on $\{1, 2, \dots, N\}$ for n . How does the result depend on N ?
- For $N \in \{100, 300, 1000\}$ apply a Bayes factor analysis also for the case that the prior for n on $\{1, 2, \dots, N\}$ 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 4: Simulate data for some example trees and explore how the type-I-errors and the type-II-errors of the KH test and the SH test (as implemented in the `ape` package in R) depend on the amount of available data and on the branch lengths of the trees.