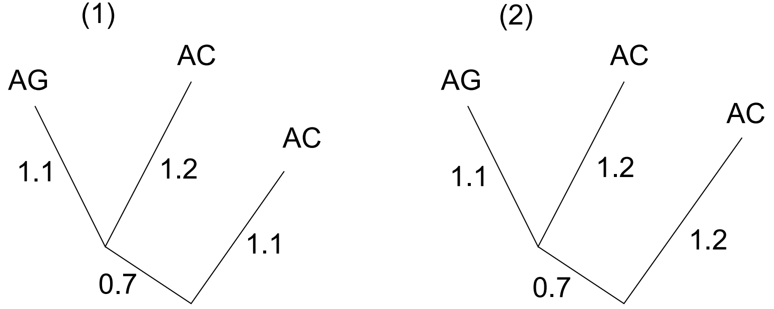
Exercise 1: An MCMC method is applied to a sequence alignment of length 2 to infer the phylogeny of the species from which they were sampled. Following a Gibbs-sampling approach, the method combines Metropolis–Hastings (MH) steps to update branch lengths, tree topologies and mutation parameter values.

To analyse the sequences you apply a Jukes–Cantor model, with the rate λ of events in which nucleotides forget their type and are replaced by a nucleotide that is randomly drawn from $\{$A, C, G, T$\}$ according to the unifor distribution. Your prior for λ is the exponential distribution with rate 10, that is expectation value 0.1, and the same λ applies to both sites. Your prior for each branch length is the exponential distribution with rate 1. Your method uses rooted binary trees and the two branches at the root count separately. A priori, you assume that all tree topologies are equally probable.

(a) The proposal chain in the MH step to update branch lenghths first picks a random branch and then proposes to add to the branch a number that is drawn from the interval [-0.15, 0.15] according to the uniform distribution on this interval. If this leads to a negative branch length $-\ell$, then $+\ell$ is used as proposed branch length.

The current value of λ is 0.08, the current tree is tree (1) and the proposed tree is (2):



Calculate the acceptance probability for this proposed change.

(b) If λ_1 is the current "event" rate for the substitution model, the proposed new "event" rate λ_2 is drawn from an exponential distribution with expectation value λ_1. Assume that the current state value of λ is 0.08 and the current tree is tree (2). The proposed value for λ is 0.04. Calculate the acceptance probability for this change.

Exercise 2:

(a) The log-likelihoods of a series of MCMC samples are

$$-5342.2 \qquad -5341.7 \qquad -5342.1 \qquad -5343.8 \qquad -5345.5.$$

Calculate the effective sample size (ESS) of this sample.

(b) In another MCMC run the log likelihoods were

$$-5341.7 \qquad -5342.1 \qquad -5345.5 \qquad -5343.8 \qquad -5342.2.$$

Calculate the ESS of that sample.

Exercise 3: We flip a coin 1000 times to test whether it is fair. If we apply Bayesian inference with a uniform prior on $[0, 1]$ for the probability p that the coin shows "head", how probable is it that we estimate the probability of $\{p > 0.5\}$ to be higher than 90% or lower than 10%? For this, assume that the coin is actually fair.

Exercise 4:

(a) Analyze the primates data data with Beast and with RAxML with bootstrapping and compare the results, that is the tree topologies, the branch lengths, and the posterior probabilities vs. bootstrap values.

(b) Do the same also for the lizards data. Explore the role of priors in the Bayesian analyses.

(c) How much do the results of Beast and RAxML depend on the choice of DNA substitution models? Explore this with the primates data and for the lizards data. For the latter, also explore the effect of partioning the data.

Exercise 5:

(a) Implement Felsenstein's pruning algorithm in R as a function that reads a tree in newick format, a sequence alignment and parameters of the F84 model and returns the log-likelihood of the tree. You can use the R script `https://evol.bio.lmu.de/_statgen/compevol/pruning.R` as a starting point.

(b) Discuss how your implementation can be made more efficient.

Exercise 6: Find publications about at least two different genera, where fossils have been used to time-calibrate phylogenetic trees. Find out in detail which traits of the fossils were used for their taxonomic classification—you may have to trace this back in other publications—and how this was used in the phylogeny analysis.