

18. January 2012

Exercise 1: Simulate several datasets for trees with five taxa $\{A,B,C,D,E\}$ that is nearly, but not exactly ultrametric. Assume that the taxa $\{A,B,C\}$ are known to form a monophyletic group because they have some morphological traits that $\{D,E\}$ do not have. A fossil is found and dated with high precision that has the same morphological traits as $\{A,B,C\}$. We neglect the possibility that D or E stem from species that once had these morphological traits but lost them. Feed your simulated data and the fossil information into BEAST and check how precisely BEAST can date the inner nodes in the tree. Repeat this analysis with TreeTime instead of Beast:

http://evol.bio.lmu.de/_statgen/software/treetime/

Exercise 2: Assume that 10000 trees have been MCMC-sampled from the posterior distribution for a given data set. To summarize these trees we would like to show a single tree that has all branches that appear in more than 5000 of the sampled trees (as splits of the taxa sets). Proof that this is always possible or give a counter-example.

Exercise 4: Perform simulation studies to examine how robust posterior probabilities given by BEAST are against misspecification of priors, substitution models and relaxed-clock models.

Exercise 5: It is known that the balls in an urn are numbered from 1 to n , but there are two Hypotheses about n . H_0 says that $n = 100$, and H_A says that n could be any positive number. Ball number 99 is drawn from the urn.

- Apply hypothesis testing 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 ?

Exercise 6: 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 $\mu = 0$.

- Apply a t -test to address this question.
- Compute Bayes-factors to compare the following two hypotheses

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

$H_A : \mu$ 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 7: We flip a coin n 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%?

Exercise 8: 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?