

HW4 – Probability and Bayesian Stats – Due Tuesday 13 February

Suppose you are studying a pair of cryptic species. In your area 5% of individuals are species A and 95% of individuals are species B. There is currently no genetic assay capable of telling them apart. They differ however in the frequency of a rare color pattern. Species A has the rare color pattern 50% of the time while species B has the rare color pattern only 2% of the time. Assume these numbers are known with certainty, from many years of field research.

- 1) Now suppose you find one of these species with the rare color pattern. Use Bayes theorem to compute the probability that it is from species A.
- 2) A new paper comes out that states that species B has a morphological feature 50% time that is only present 10% of the time in species A. Your sample has this feature. What is the probability that you have a sample of species A now?
- 3) Suppose now that a genetic test is developed that can identify the species of our sample. But the test, like all tests, is imperfect. This is the information you have about the test:
 - a. The probability it correctly identifies species A is 0.7
 - b. The probability it correctly identifies species B is 0.98You run the test and it is positive for species A. Compute the posterior probability that your sample is species A using all the information available.

Analyzing MCMC log files

For this question download the file `mcmc.log` from the website. Typical software that uses a Bayesian approach produces log files that document the parameter values sampled during the MCMC (Markov chain Monte Carlo) run. You have just completed running *Beast* to reconstruct the gene tree for a collection of odorant receptor genes in your species of interest.

- 4) You know that 3rd codon positions should evolve more quickly than 1st or 2nd codons. You can use this knowledge to make sure that you set up the analysis correctly. Use your R skills to plot a comparison of the rate of evolution for all 3 codons.
- 5) In Bayesian stats, we compare highest posterior densities rather than confidence intervals. Highest posterior densities are the narrowest part of a distribution that contains 95% of our sample. Install the package `coda`. Calculate and report the 95% highest posterior density for the rate of evolution for each codon position (hint: you may need the `as.mcmc` function). Do you believe that the rate of evolution for the three codons is different?

Homework

Prepare a document with your answers and R code you used to solve all five problems.