**HW4 – Probability and Bayesian Stats – Due Tuesday 13 February**

Suppose there are two species of panda bear. Both are equally common in the wild and live in the same places. There is no genetic assay capable of telling them apart. They differ however in their family sizes. Species A gives birth to twins 10% of the time, otherwise birthing a single infant. Species B births twins 20% of the time, otherwise birthing singleton infants. Assume these numbers are known with certainty, from many years of field research. Now suppose you are managing a captive panda breeding program.

1. You have a new female panda of unknown species, and she has just given birth to twins. Use Bayes theorem to compute the probability that the panda you have is from species A.
2. What is the probability that the next birth will also be twins?
3. Suppose the same panda mother has a second birth and that it is not twins, but a singleton infant. Compute the posterior probability that this panda is species A.

A common boast of Bayesian statisticians is that Bayesian inference makes it easy to use all of the data, even if the data are of different types. So, suppose now that a veterinarian comes along who has a new genetic test that she claims can identify the species of our mother panda. But the test, like all tests, is imperfect. This is the information you have about the test:

The probability it correctly identifies a species A panda is 0.8.

The probability it correctly identifies a species B panda is 0.65.

1. The vet administers the test to your panda and tells you that the test is positive for species A. Compute the posterior probability that your panda is species A. Take into account that the panda has had one birth that was twins.

**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 MrBayes to reconstruct the gene tree for a collection of odorant receptor genes in your species of interest.

1. 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.
2. 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. 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 six problems.