

Homework 6

Trevor Caughlin

March 5th, 2019

Question 1: For Bayesian analysis, we will be using the software package Stan. Stan works using a physics-based algorithm to sample from the posterior distributions.

A) Install the following R packages on your computer:

```
install.packages(c("rstan","rstanarm","shinystan"))
```

B) Search the Stan forums (<http://discourse.mc-stan.org/>) for an example of Stan code that is relevant to your project. Below, include a web link to the forum discussion and write a brief description of how you might adapt the discussed model for your own purposes.

Question 2: Choose one of the following texts to read (both texts are available in github) and answer the associated question with approximately one paragraph of text:

Choice 1: Dall et al. 2005: "Information and its use by animals in evolutionary ecology"

Answer the following: **Are foraging animals Bayesians? Why or why not?**

Choice 2: "The Garden of Forking Paths" (Original Spanish title: "El jardín de senderos que se bifurcan") by Jorge Luis Borges, 1941.

Answer the following: **Why has the phrase "Garden of forking paths" become so important and widespread in modern statistics?**

The following Questions are adapted from Richard McElreath's Rethinking Statistics book, Chapter 2:

Question 3: Earth is 70% covered in water, while Mars is 100% land. A randomly selected satellite pixel from one of the planets was classified as "land." Assuming the satellite pixel was equally likely to be from Earth or Mars, show that the posterior probability that the pixel was from Earth, conditional on seeing land $[P(\text{Earth}|\text{Land})]$, is 0.23.

Question 4: Suppose there are two species of panda bear. Both are equally common in the wild and live in the same places. They look exactly alike and eat the same food, and there is

yet 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. You have a new female panda of unknown species, and she has just given birth to twins. What is the probability that her next birth will also be twins?

- (A) Compute the probability that the panda we have is from species A, assuming we have only observed the first birth and that it was twins.
- (B) 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.
- (C) 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:
 - a. The probability it correctly identifies a species A panda is 0.8
 - b. The probability it correctly identifies a species B panda is 0.65.

The vet administers the test to your panda and tells you that the test is positive for species A. First ignore your previous information from the births and compute the posterior probability that your panda is species A. Then redo your calculation, now using the birth data as well.