Homework 6

Katie Callery

2/25/2018

## Question 1

# A.)

Is this supposed to take an extremely long time?

I tried it in R Markdown first and it got weird. A message came up in the console about compiling and offered a y/n option. I hit y and it helped but still took a weird amount of time.

I installed these using a separate R script, which I can upload to GitHub if you want.

install.packages("rstan")  
install.packages("rstanarm")  
install.packages("shinystan")

# B.)

<http://discourse.mc-stan.org/t/a-more-efficient-implementation-of-the-sum-of-independent-binomials/3012>

Title: **A more efficient implementation of the sum of independent binomials?**

*Write a brief description of how you might adapt the discussed model for your own purposes.*

The model is about estimating the abundance of hatchery fish that pass a point with capture data. Some of the fish are coming down the stream from a hatchery and some are wild fish, but they know how many fish are being released from the hatchery at any given time. When there are no hatchery fish being released, they can get baseline levels of fish abundance representing just the wild fish. The abundance of hatchery fish is dependent on the survival rate and arrival rate of the fish between the hatchery and the capture point. This information feeds into a modified Cormack-Jolly-Seber mark-recapture model, which is a model often used for measuring survival of mobile animals over time.

I could adapt this model for my data on adult kestrels. The survival rate of hatchery fish would be substituted for the survival rate of the kestrels, and would be based on prior knowledge of annual apparent survival, based on return rates to the breeding grounds. The arrival rate component would be substituted for the rate of dispersal from the breeding site instead, since both are basically the chance that an organism will arrive at the capture point.

## Question 2

# 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?** *(Answer the associated question with approximately one paragraph of text)*

Foraging animals are Bayesians. In Box 1, it shows how previously collected information informs an animal’s decision-making process. This is kind of like the posterior-probability. It is an iterative process in which the more information an animal collects in previous events reduces the uncertainty in future, similar events. Uncertainty cannot be eliminated, just as - in other instances - prior information or sampling cannot fully inform about current conditions. Some of the information foraging animals consider are time spent foraging per food items collected and the quality of those food items. Gathering and retaining information like this would be advantageous for to maximize resource acquisition efficiency, which could improve survival odds.

## Question 3

*Earth is 70% covered in water (so 30% land), 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 the posterior probability that the pixel was from Earth, conditional on seeing land [P(Earth|Land)], is 0.23.**

[P(Earth|Land)]=[P(Land|Earth)]\*[P(Earth)])/[P(Land)]

P(Earth) = 0.5 P(Land) = (0.3+1.0)/2.0 = 0.65 [P(Land|Earth)]= 0.3

[P(Earth|Land)]= 0.3 \* 0.5 /(0.65)

PEarthLand<-(0.3 \* 0.5)/(0.65)  
PEarthLand

## [1] 0.2307692

The posterior probability that the pixel was from Earth is indeed 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?**

1. 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.

P(twins | A) = 0.1 P(twins | B) = 0.2 P(twins) = (0.1 + 0.2)/ 2 = 0.15 P(A) = 0.5 P(B) = 0.5 P(A | twins) = ?

[P(A|twins)]=([P(twins|A)]*[P(A)])/[P(twins)][P(A|twins)]=0.1*0.5/0.15

PAtwins<-0.1\*0.5/0.15  
PAtwins

## [1] 0.3333333

The probability of the panda being species A given that it gave birth to twins (once) is 0.33.

1. 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.

P(twins | A) = 0.1 P(single | A) = 0.9 P(offspring | A) = 0.1 \* 0.9 = 0.09 P(offspring | B) = 0.2 \* 0.8 = 0.16 P(offspring) = (0.09 + 0.16) / 2 = 0.125

P(A | offspring) = ?

P(A | offspring) = P(offspring | A) \* P(A) / P(offspring)

P(A | offspring) = 0.09 \* 0.5 / 0.125

PAoffspring<-0.09 \* 0.5 / 0.125  
PAoffspring

## [1] 0.36

P(A | offspring) = 0.36

The probability of the panda being species A given the current offspring produced is 0.36. The probability went up, since species A have higher probability of producing singletons than species B.

1. 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:
2. The probability it correctly identifies a species A panda is 0.8
3. 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.

P(A) = 0.5 P(B) = 0.5 P(identify|A) = 0.8 P(identify|B) = 0.65 P(identify) = (0.8 + 0.65) / 2 = 0.725

[P(A|identify)]=([P(identify|A)]\*[P(A)])/[P(identify)]

P(A|identify) = 0.8 \* 0.5 / 0.725

P(A | identify) = 0.55

Then redo your calculation, now using the birth data as well.

In this case, the probability that the bear is A is 0.55, and the probability that the bear is B is 0.45. So P(A)=0.55.

P(A) = 0.55 P(B) = 0.45 P(twins | A) = 0.1 P(single | A) = 0.9 P(offspring | A) = 0.1 \* 0.9 = 0.09 P(offspring | B) = 0.2 \* 0.8 = 0.16 P(offspring) = (0.09 + 0.16)/ 2 = 0.125

P(A | offspring) = ?

P(A | offspring) = P(offspring | A) \* P(A) / P(offspring)

P(A | offspring) = 0.09 \* 0.55 / 0.125

P(A | offspring) = 0.396

The probability of the bear being species A given its offspring and that it tested as bear A is 0.396. It went up since the test makes P(A) go up from 0.5 to 0.55.