Chapter 2

Hard question 1

Suppose there are two species of Species 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 Species 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?

The important information about the two species.

Species A:

- 10% Twins, P(Twin | Species A)
- 90% Single infant

Species B:

- 20% Twins, P(Twin | Species B)
- 80 % Single infant

The problem is divided up using total probaility:

$$\text{P(Twin} \mid \text{Previous twin)} = \text{P(Twin} \mid \text{Species A)} \times \text{P(Species A} \mid \text{Previous twin)} + \text{P(Twin} \mid \text{Species B)} \times \text{P(Species A} \mid \text{Previous}$$

Two of the probabilites are given directly in the excersise. Inserting these into the expression gives

$$\text{P(Twin} \mid \text{Previous twin)} = 0.1 \times \text{P(Species A} \mid \text{Previous twin)} + 0.2 \times \text{P(Species A} \mid \text{Previous twin)}$$

Calculating the missing probabilites is done using bayes rule. First for Species A

$$\text{P(Species A} \mid \text{Previous twin)} = \frac{\text{P(Previous twin} \mid \text{Species A)} \times \text{P(Species A)}}{\text{P(Previous twin)}} = \frac{0.1 \times 0.5}{\text{P(Previous twin)}}$$

To complete the calculation the probability of a twin is needed. Total probaility can be used

$$\text{P(Previous twin)} = \text{P(Previous twin} \mid \text{Species A)} \times \text{P(Species A)} + \text{P(Previous twin} \mid \text{Species B)} \times \text{P(Species B)} = 0.1 \times 0.5 + 0.2$$

Inserting this back into the bayes rule applied on the probability the panda being from Species A completes that calculation

$$\text{P(Species A} \mid \text{Previous twin)} = \frac{0.1 \times 0.5}{0.15} = 0.33$$

The same calculation is repeated for Species B

$$\text{P(Species B} \mid \text{Previous twin)} = \frac{\text{P(Previous twin} \mid \text{Species B)} \times \text{P(Species B)}}{\text{P(Previous twin)}} = \frac{0.2 \times 0.5}{0.15} = 0.66$$

Finally the two probabilites calcutated using Bayes rule can go back into the original expression

$$\text{P(Twin} \mid \text{Previous twin)} = 0.1 \times \text{P(Species A} \mid \text{Previous twin)} + 0.2 \times \text{P(Species A} \mid \text{Previous twin)} = 0.1 \times 0.33 + 0.2 \times 0.66 = 0.1$$

16.5% makes sence. The probability is more than just taking the average probability of having twins from the two species. SInce the previous litter was twins, it increased the probability of the panda being from species B and therefore the probability of the panda giving birth to twins.

Hard Question 2

Recall all the facts from the problem above. Now compute the probability that the panda we have is from species A, assuming we have observed only the first birth and that it was twins.

This was calculated in question 1

$$\text{P(Species A} \mid \text{Previous twin)} = \frac{\text{P(Previous twin} \mid \text{Species A)} \times \text{P(Species A)}}{\text{P(Previous twin)}} = \frac{0.1 \times 0.5}{0.15} = 0.33 = 33\%$$

Hard Question 3

Continuing on from the previous problem, 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.

The question expressed in symbols becomes

$$\text{P(Species A} \mid \text{(Second singelton} \cap \text{First twin))}$$

Aproach 1

This can be solved using bayes

$$\text{P(Species A} \mid \text{(Second singelton} \cap \text{First twin))} = \frac{\text{P((Second singelton} \cap \text{First twin)} \mid \text{Species A)} \times \text{P(Species A)}}{\text{P(Second singelton} \cap \text{First twin)}}$$

$$\frac{\text{P((Second singelton} \cap \text{First twin)} \mid \text{Species A)} \times \text{P(Species A)}}{\text{P(Second singelton} \cap \text{First twin)}} = \frac{\text{P(Second singelton} \mid \text{Species A)} \times \text{P(First twin} \mid \text{Species A}}{\text{P(Second singelton} \cap \text{First twin)}}$$

Using that the two events: the second being a singelton given the species is A i and the first being a twin given the species is A, are independent. We can calulate the probaiblity useing multiplication.

$$\frac{\text{P(Second singelton} \mid \text{Species A)} \times \text{P(First twin} \mid \text{Species A))} \times \text{P(Species A)}}{\text{P(Second singelton} \cap \text{First twin)}} = \frac{0.9 \times 0.1 \times 0.5}{\text{P(Second singelton} \cap \text{First twin)}}$$

To adress the denominator we use total pobability.

$$\text{P(Second singelton} \cap \text{First twin)} = \text{P((Second singelton} \cap \text{First twin)} \mid \text{Species A)} \times \text{P(Species A)} + \text{P((Second singelton} \cap \text{Firs}$$

Using that the conditional probabilites are independent the probabilites can be calulated using multiplication.

$$\text{P(Second singelton} \cap \text{First twin)} = 0.9 \times 0.1 \times 0.5 + 0.8 \times 0.2 \times 0.5 = 0.125$$

Inserting this

$$\text{P(Species A} \mid \text{(Second singelton} \cap \text{First twin))} = \frac{0.9 \times 0.1 \times 0.5}{0.125} = 0.36 = 36\%$$

Aproach 2 Appraoch 1 was allot of work. Stating with what we know from question 2 it can be done faseter.

The prior is now updated and with it so are the probabilites. Keeping the notation simple P(Species A now takes a new value without an updated notation.

$$\text{P(Species A} \mid \text{Second singelton)} = \frac{\text{P(Second singelton} \mid \text{Species A)P(Species A)}}{\text{P(Second singeltion)}} = \frac{0.9 \times 0.33}{\text{P(Second singeltion)}}$$

$$\text{P(Second singeltion)} = \text{P(Second singeltion} \mid \text{Species A)} \times \text{P(Species A)} + \text{P(Second singeltion} \mid \text{Species B)} \times \text{P(Species A)} = 0.$$

$$\text{P(Species A} \mid \text{Second singelton)} = \frac{0.9 \times 0.33}{\text{P(Second singelton)}} = \frac{0.9 \times 0.33}{0.825} = 0.36 = 36\%$$

Hard Question 4

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. The vet administers the test to your Species 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.

Ignoring the adjustemnts to the priors we use bayes

$$\text{P(Species A} \mid \text{Tested as A)} = \frac{\text{P(Tested as A} \mid \text{Species A)} \times \text{P(Species A)}}{\text{P(Tested as A)}} = \frac{0.8 \times 0.5}{\text{P(Tested as A)}}$$

$$\text{P(Tested as A)} = \text{P(Tested as A} \mid \text{Species A)} \times \text{P(Species A)} + \text{P(Tested as A} \mid \text{Species B)} \times \text{P(Species B)} = 0.8 \times 0.5 + 0.35 \times 0.5$$

$$\text{P(Species A} \mid \text{Tested as A)} = \frac{0.8 \times 0.5}{\text{P(Tested as A)}} = \frac{0.8 \times 0.5}{0.575} = 0.696 = 69.6\%$$

Including the updated priors due to the testing gives anoter result. As in question 3 approach 2 this means the probabilites are updated wihout updating the notation.

$$\text{P(Species A} \mid \text{Tested as A)} = \frac{\text{P(Tested as A} \mid \text{Species A)} \times \text{P(Species A)}}{\text{P(Tested as A)}} = \frac{0.8 \times 0.36}{\text{P(Tested as A)}}$$

$$\text{P(Tested as A)} = \text{P(Tested as A} \mid \text{Species A)} \times \text{P(Species A)} + \text{P(Tested as A} \mid \text{Species B)} \times \text{P(Species B)} = 0.8 \times 0.36 + 0.35 \times 0.$$

$$\text{P(Species A} \mid \text{Tested as A)} = \frac{0.8 \times 0.36}{\text{P(Tested as A)}} = \frac{0.8 \times 0.36}{0.484} = 0.595 = 59.5\%$$