Chapter 2. Small Worlds and Large Worlds

29/10/2021

Here we'll solve the medium to hard exercises for the chapter.

```
library(rethinking)
```

2M1

Recall the globe tossing model from the chapter. Compute and plot the grid approximate posterior distribution for each of the following set of observations. In each case, assume a uniform prior for p.

(1) W, W, W
(2) W, W, W, L
(3) L, W, W, L, W, W, W

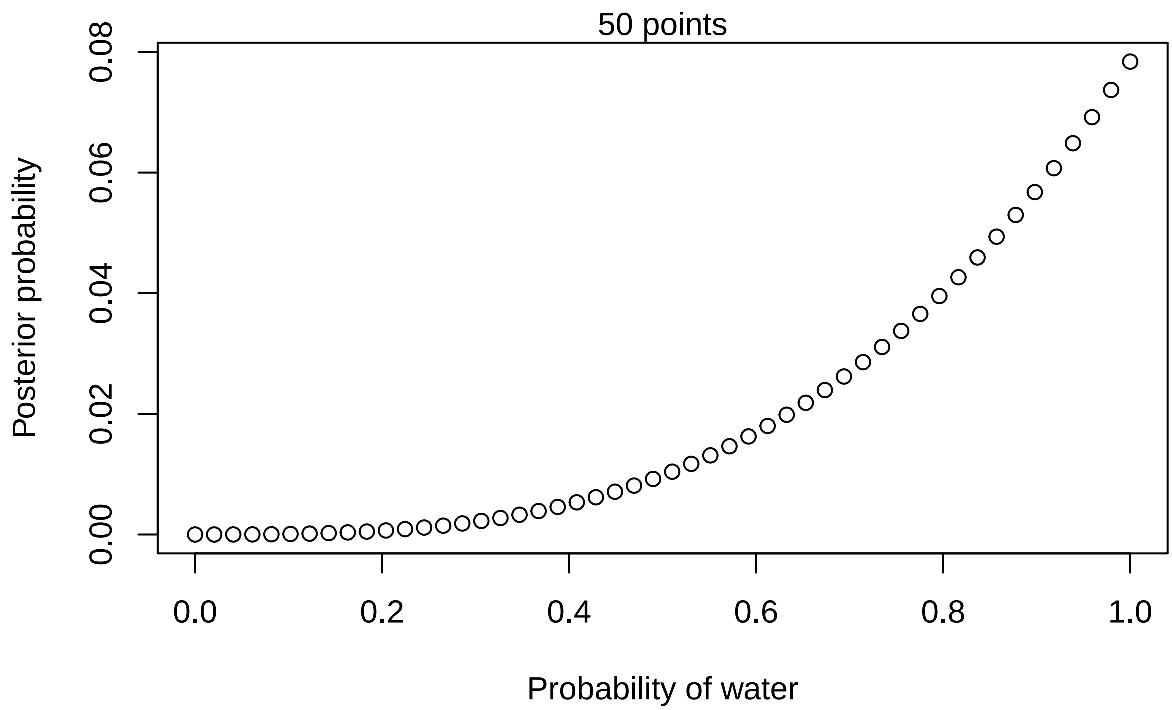
Answer

Define a function to do the work for us

```
# Where w is the amount of water observed
# and s is the total size
gripap <- function(w, s){

        # define grid
        p_grid <- seq(from = 0, to=1, length.out = 50)

        # define prior
        prior <- rep(1, 50)

        # compute likelihood at each value of the grid
        likelihood <- dbinom(w, size =s, prob = p_grid)

        # compute product of likelikood and prior
        unstd.posterior <- likelihood * prior

        # standardise the posterior, so it sums 1
        posterior <- unstd.posterior/sum(unstd.posterior)

        # And plot the results plot
        plot(p_grid, posterior, type="b", xlab="Probability of water", ylab = "Posterior probability")
        mtext("50 points")

}
```

So for (1)

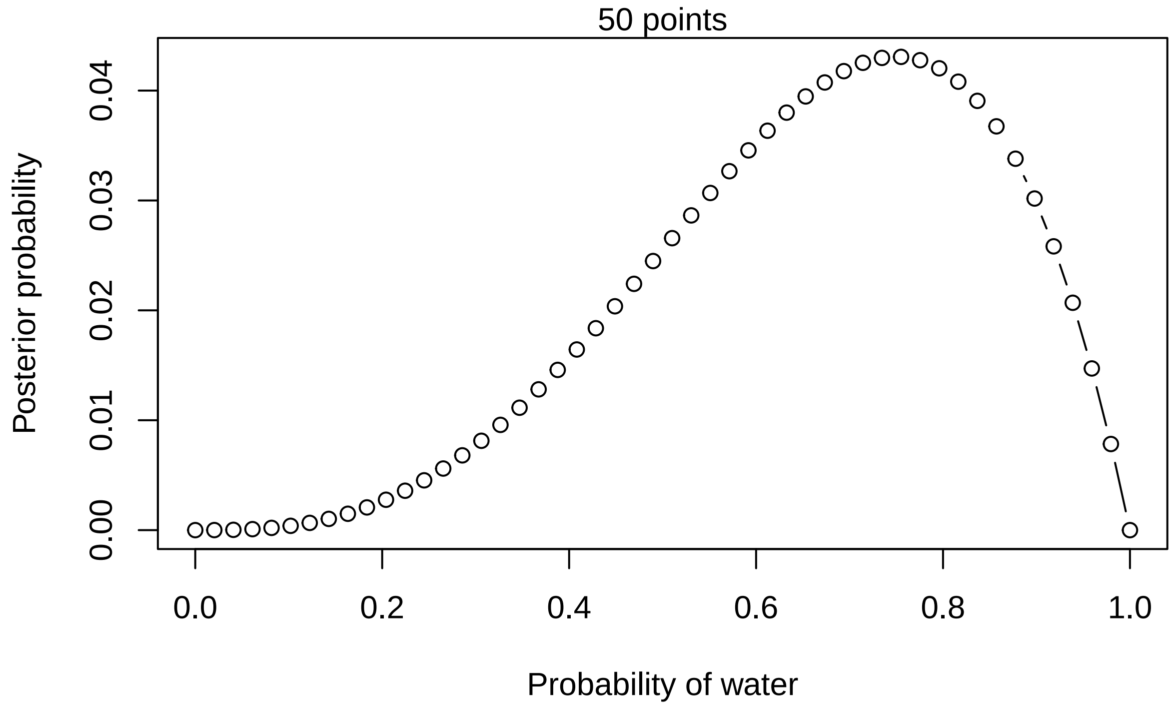
```
gripap(w = 3, s = 3)
```



We've seen water only, so it makes sense that observing land is unlikely for the model, although not impossible.

Let's try (2)

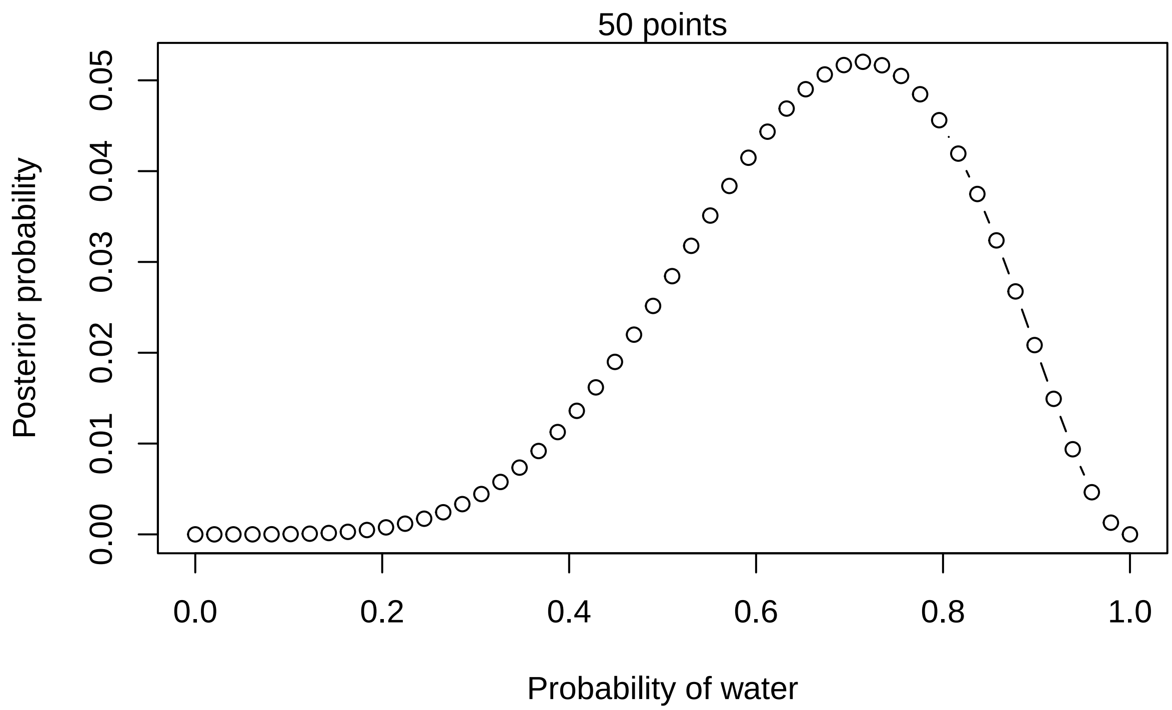
```
gripap(w = 3, s = 4)
```



We saw one land! Probabilities are still mostly for water, but observing water only is now off the table.

And lastly (3)

```
gripap(w = 5, s = 7)
```



Now we have a more balanced distribution, although still biased to water.

2M2

Now assume a prior for p that is equal to zero when $p < 0.5$ and is a positive constant when p >= 0.5. Again compute the grid approximate posterior distribution for each of the sets of observations in the problem just above.

Answer

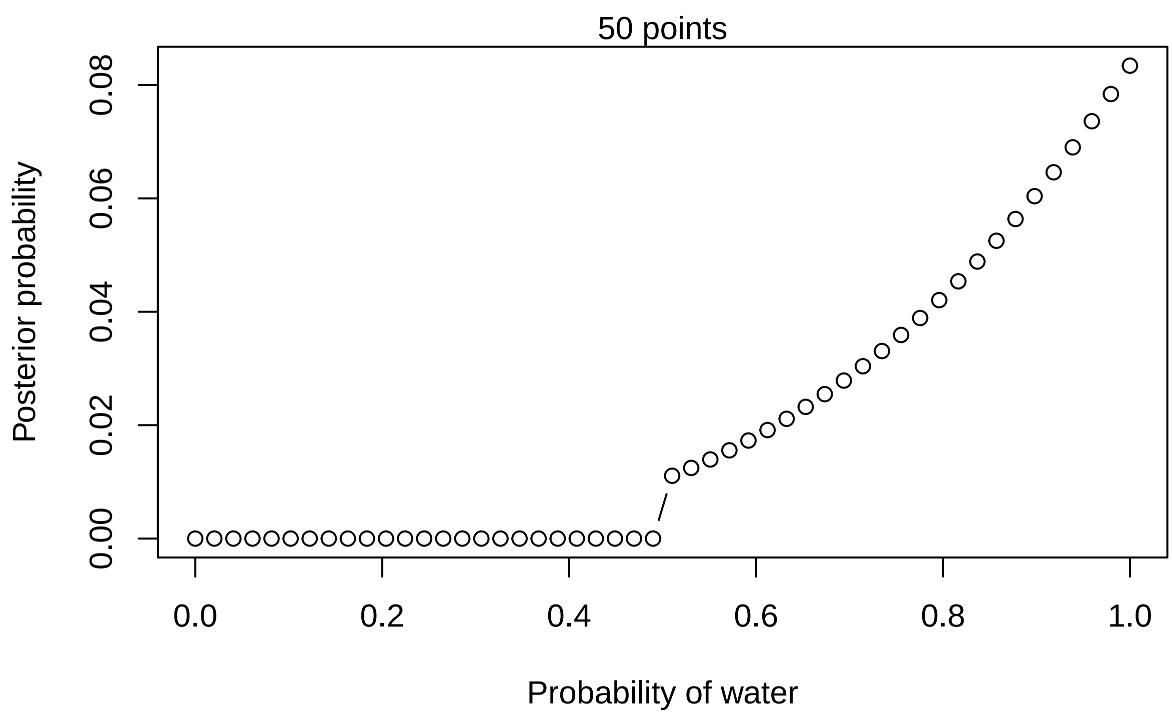
Again, let's define a function to have this flexiprior

```
gripap.cprior <- function(w, s){

        # define grid
        p_grid <- seq(from = 0, to=1, length.out = 50)

        # define prior
        prior <- ifelse(p_grid < 0.5, 0, 1)

        # compute likelihood at each value of the grid
        likelihood <- dbinom(w, size =s, prob = p_grid)

        # compute product of likelikood and prior
        unstd.posterior <- likelihood * prior

        # standardise the posterior, so it sums 1
        posterior <- unstd.posterior/sum(unstd.posterior)
```

```r
        # And plot the results plot
        plot(p_grid, posterior, type="b", xlab="Probability of water", ylab = "Posterior probability")
        mtext("50 points")

}
```

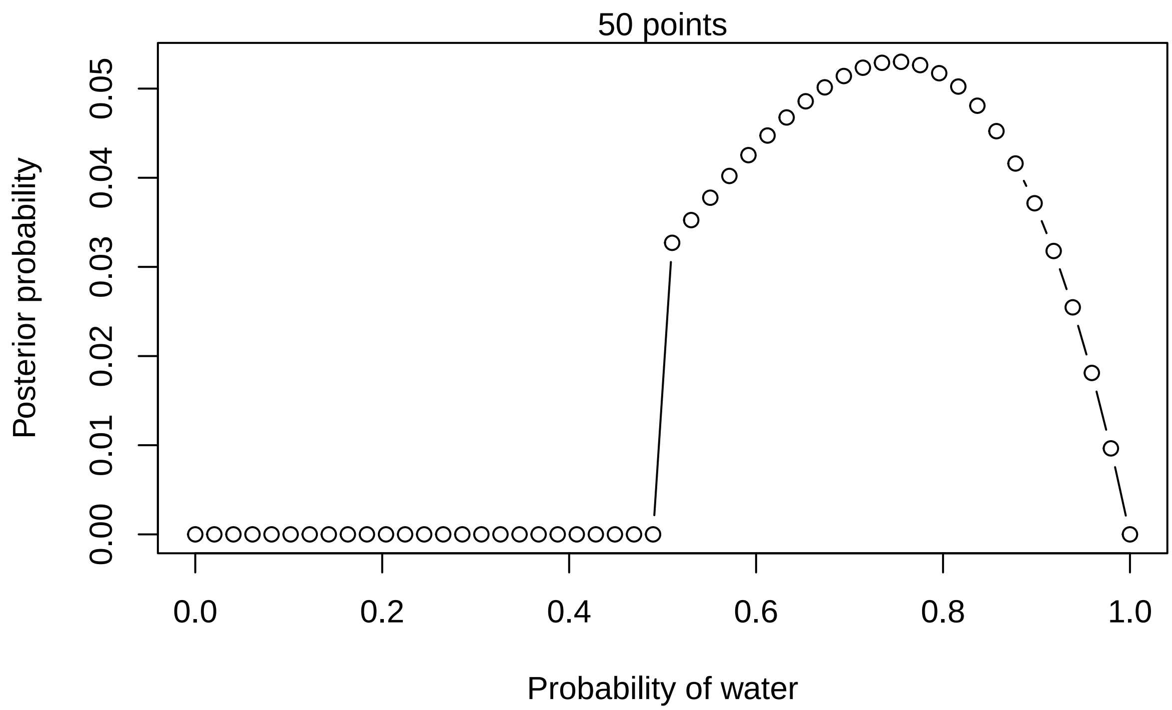
So for (1)

```r
gripap.cprior(w = 3, s = 3)
```



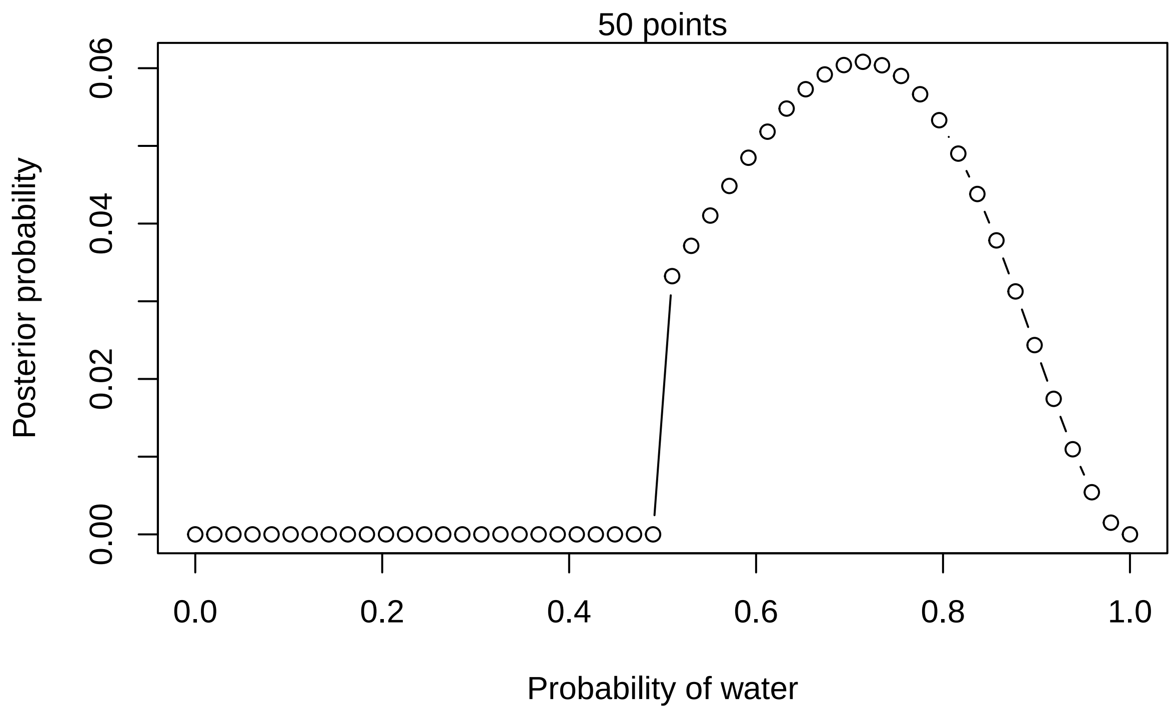
For (2)

```r
gripap.cprior(w = 3, s = 4)
```



and for (3)

```
gripap.cprior(w = 5, s = 7)
```



With this prior, we assume that there must be more water than land.

2M3

Suppose there are two globes, one for Earth and one for Mars. The Earth globe is 70% covered in water. The Mars globe is 100% land. Further suppose that one of these globes – you don't know which – was tossed and produced a "land" observation. Assume that each globe was equally likely to be tossed. Show that the posterior probability that the globe was the Earth, conditional on seeing "land", is 0.23.

Answer

Let's take a look at what we have.

$Pr(Earth|land) = 0.23$

$Pr(Earth) = 0.5$

$Pr(Mars) = 0.5$

$Pr(land|Earth) = 0.3$

$Pr(land|Mars) = 1$

Thus,

$Pr(Earth|land) \propto P(land|Earth) * P(Earth)$

So it follows that

$P(land|Earth) * P(Earth) = 0.7 * 0.5 = 0.15$

$P(land) = P(land|Earth)P(Earth) + P(land|Mars) * P(Mars) = 0.15 + 0.5 = 0.65$

$Pr(Earth|land) = \frac{P(land|Earth)*P(Earth)}{P(land)} = \frac{0.15}{0.65} = 0.23$

2M4

Suppose you have a deck with only 3 cards. Each card has two sides, and each side is each black or white.

One card has two black sides. The second card has one black and one black side. The third card has two white sides.

Supposed all three cards are placed in a bag and shuffled. Some reaches into the bag and pulls out a card and places it flat on a table. A black side is shown facing up, but you don't know the colour of the side facing down. Show that the probability of the other side is also black is 2/3. Use the counting method to approach this problem. This means counting up the ways that each card could produce the observed data (a black side facing up).

Answer

Let's count!

1st card (2 black sides) = 2 ways

2nd card (one side each) = 1 way

3rd card (2 white sides) = 0 ways

Since we have 3 possible total ways we could observe a black side, and the 1st card has 2 of those ways, it follows that the probability for that card to be the 1st one is 2/3.

2M5

Now suppose there are four cards: B/B, B/W, W/W and another B/B. Again suppose a black side appears face up. Again calculate the probability that the other side is black.

Answer

2 B/B cards = 4 ways

1 B/W card = 1 way

1 W/W card = 0 ways

Following the same logic, the probability that this is a B/B card is **4/5**.

2M6

Imagine that black inc is heavy, and so cards with black sides are heavier than white sides. As a result, it's less likely that a card with black sides is pulled from the bag.

Again assume there are three cards: B/B, B/W, W/W. After experimenting a number of times you conclude that for every way to pull the B/B card, there are two ways to pull the B/W card, and 3 ways to pull the W/W card. Again, suppose that a card is pulled and a black side appears face up. Show that the probability the other side is black is now 0.5.

Answer

B/B = 2 ways (sides) * 1 way (weight) = 2 ways

B/W = 1 * 2 = 2 ways

W/W = 0 * 3 = 0 ways

Since both B/B and B/W have 2 ways each to show, the probability for the card to be a B/B is **0.5**.

2M7

Assume again the original card problem. Before looking at the other side, we draw another card, the face of the new card is white. Show that the probability that the first card, the one showing a black side has black on it's other side is now 0.75. Use the counting method, if you can.

Answer

We need to compute the ways a B + W can be produced

Let's see the ways that could possibly yield this result

B/B + B/W = 2 ways (since B/B has two sides that could possibly pop up).

B/B + W/W = 4 ways (sinnce both B/B and W/W have two sides).

B/W + W/W = 2 ways (since W/W has to sides).

In 6 out of 8 ways we could have a B/B card as the first draw. 6/8 = 3/4 = **0.75**.

2H1

Suppose there are two species of panda. 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 in their family sizes.

Species A gives birth to twins 10% of the time. Species B gives birth to 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're managing a captive panda breeding programme. You have a female panda of unknown species, and she's just given birth to twins.

What's the probability that her next birth will also be twins?

Answer

$Pr(twins|SpA) = 0.1$

$Pr(twins|SpB) = 0.2$

$Pr(SpA) = Pr(SpB) = 0.5$

$Pr(SpA|twins) = (Pr(twins|spA)) * Pr(spA)/Pr(twins)$

$Pr(SpB|twins) = (Pr(twins|spB)) * Pr(spB)/Pr(twins)$

$Pr(twins) = Pr(spA)Pr(twins|spA) + Pr(spB)Pr(twins|spB)$
$= 0.5 * 0.1 + 0.5 * 0.2$
$= 0.15$

$Pr(SpA|twins) = 0.1 * 0.5/0.15 = 0.3333333$

$Pr(SpB|twins) = 0.2 * 0.5/0.15 = 0.6666667$

$Pr(twins) = Pr(spA) * (Pr(twins|spA)) + Pr(spB) * (Pr(twins|spB))$
$= 0.3333333 * 0.1 + 0.6666667 * 0.2$
$= 0.1666667$

2H2.

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.

Answer

$Pr(SpA|twins) = 0.1 * 0.5/0.15$
$= 0.3333333$

2H3.

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.

Answer

$Pr(single|SpA) = 0.9$

$Pr(single|SpB) = 0.8$

$Pr(SpA) = 0.3333333$

$Pr(SpB) = 0.6666667$

$Pr(SpA|single) = (Pr(single|spA)) * Pr(spA)/Pr(single)$

$Pr(SpB|single) = (Pr(single|spB)) * Pr(spB)/Pr(single)$

$Pr(single) = Pr(spA) * (Pr(single|spA)) + Pr(spB) * (Pr(single|spB))$
$= 0.3333333 * 0.9 + 0.6666667 * 0.8$
$= 0.8333333$

$Pr(SpA|single) = 0.9 * 0.3333333/0.8333333$
$= 0.36$

$Pr(SpB|single) = 0.8 * 0.6666667/0.8333333$
$= 0.6400001$

2H4.

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 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.

Answer

- **Ignoring birth information**

$Pr(test|spA) = 0.8$

$Pr(test|spB) = 0.65$

$Pr(SpA) = Pr(SpB) = 0.5$

$Pr(spA|test) = (Pr(test|spA)) * Pr(spA)/Pr(test)$

$Pr(test) = Pr(spA) * (Pr(test|spA)) + Pr(spB) * (Pr(test|spB))$
$= 0.5 * 0.8 + 0.5 * 0.65$
$= 0.725$

$Pr(spA|test) = 0.8 * 0.5/0.725 = 0.5517241$

The probability for the individual to be from species A is then **0.55**.

- **1 twin is born**

$Pr(SpA) = 0.3333333$

$Pr(SpB) = 0.6666667$

$Pr(test) = Pr(spA) * (Pr(test|spA)) + Pr(spB) * (Pr(test|spB))$
$= 0.3333333 * 0.8 + 0.6666667 * 0.65$
$= 0.7$

$Pr(spA|test) = 0.8 * 0.3333333/0.7$
$= 0.3809523$

The probability for the individual to be from species A is then **0.38**.

- **Singleton born after previous twin**

$Pr(SpA) = 0.36$

$Pr(SpB) = 0.6400001$

$Pr(test) = Pr(spA) * (Pr(test|spA)) + Pr(spB) * (Pr(test|spB))$
$= 0.36 * 0.8 + 0.6400001 * 0.65$
$= 0.7040001$

$Pr(spA|test) = 0.8 * 0.36/0.7040001$
$= 0.4090909$

The probability for the individual to be from species A is then ~**0.41**.