Chapter 2 Questions

Melissa Van Bussel

June 1, 2018

Chapter 2 points: 2 + 14 + 4 = 20 points total.

Easy (2 points total)

2E1

The probability of rain on Monday can be represented by P(rain | Monday).

2E2

The expression Pr(Monday | rain) corresponds to "The probability that it is Monday, given that it is raining."

2E3

"The probability that it is raining, given that it is Monday" can be represented either by Pr(rain|Monday) or, equivalently by Baye's Theorem, $\frac{\text{Pr(Monday|rain) Pr(rain)}}{\text{Pr(Monday)}}$

2E4

What Finetti means by this sentence in context of the globe tossing example is that there is no "objective reality" when it comes to the "probability of water". There is no actual statistical distribution of water on Earth, but rather, saying that the probability of water is 0.7 is saying that if you were to randomly pick coordinates anywhere on Earth, there would be a 0.7 probability that those coordinates would be on water, not land. This is because $\approx$ 70% of the Earth is covered in water.

Medium (14 points total)

2M1 (2 points)

Recall the globe tossing model from the chapter. Compute and plot the grid approximate posterior distribution for each of the following sets 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

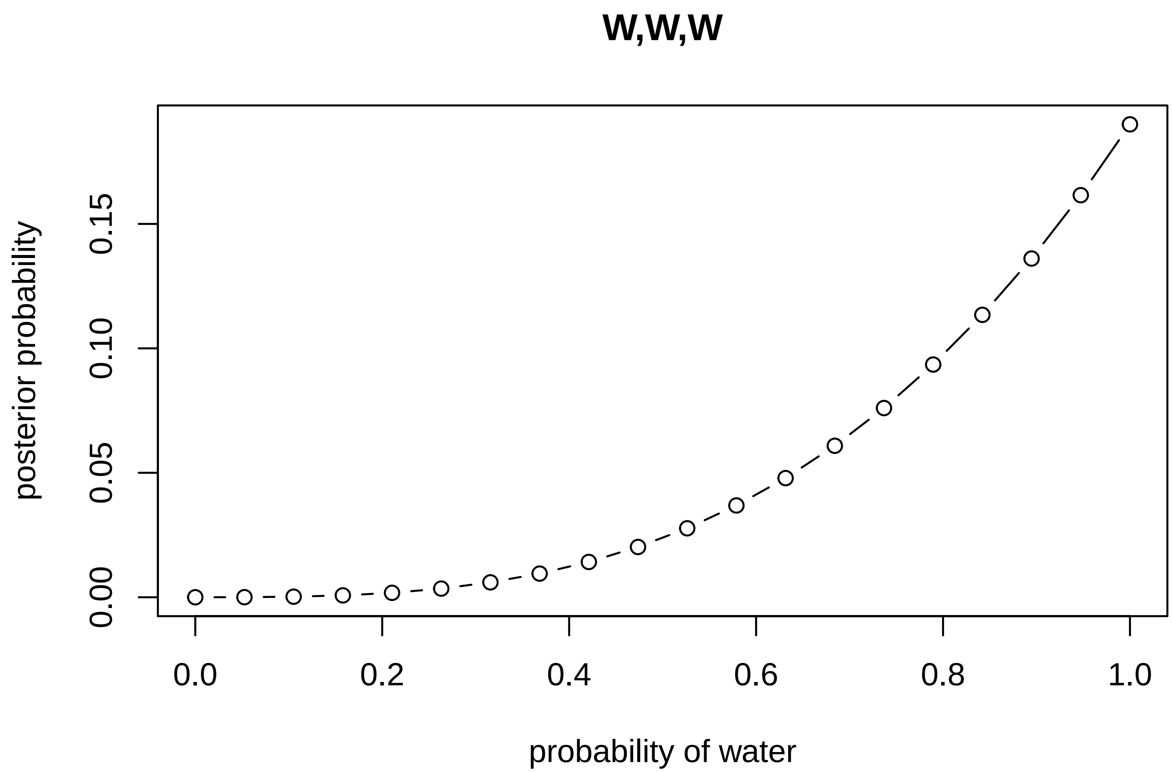
1)

```
library(rethinking)
```

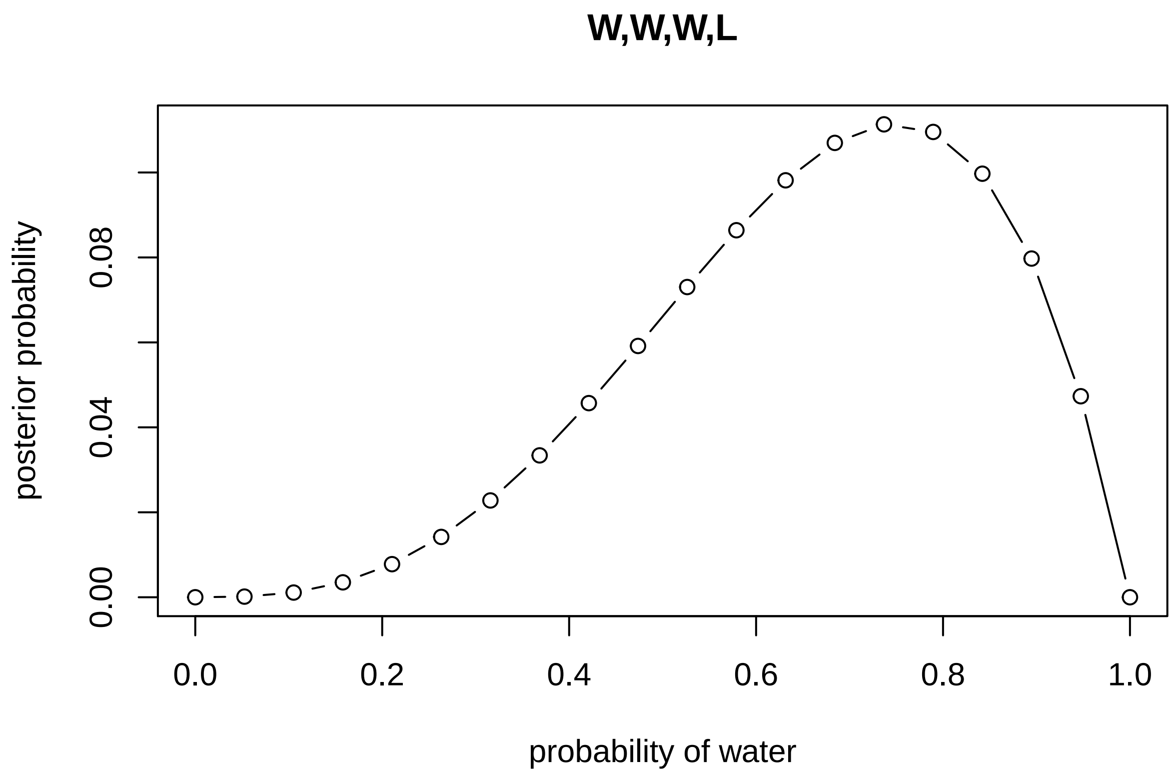
```
## Loading required package: rstan
## Warning: package 'rstan' was built under R version 3.3.3
## Loading required package: ggplot2
## Warning: package 'ggplot2' was built under R version 3.3.3
```

```
## Loading required package: StanHeaders

## Warning: package 'StanHeaders' was built under R version 3.3.3

## rstan (Version 2.17.3, GitRev: 2e1f913d3ca3)

## For execution on a local, multicore CPU with excess RAM we recommend calling
## options(mc.cores = parallel::detectCores()).
## To avoid recompilation of unchanged Stan programs, we recommend calling
## rstan_options(auto_write = TRUE)

## Loading required package: parallel

## rethinking (Version 1.59)
```

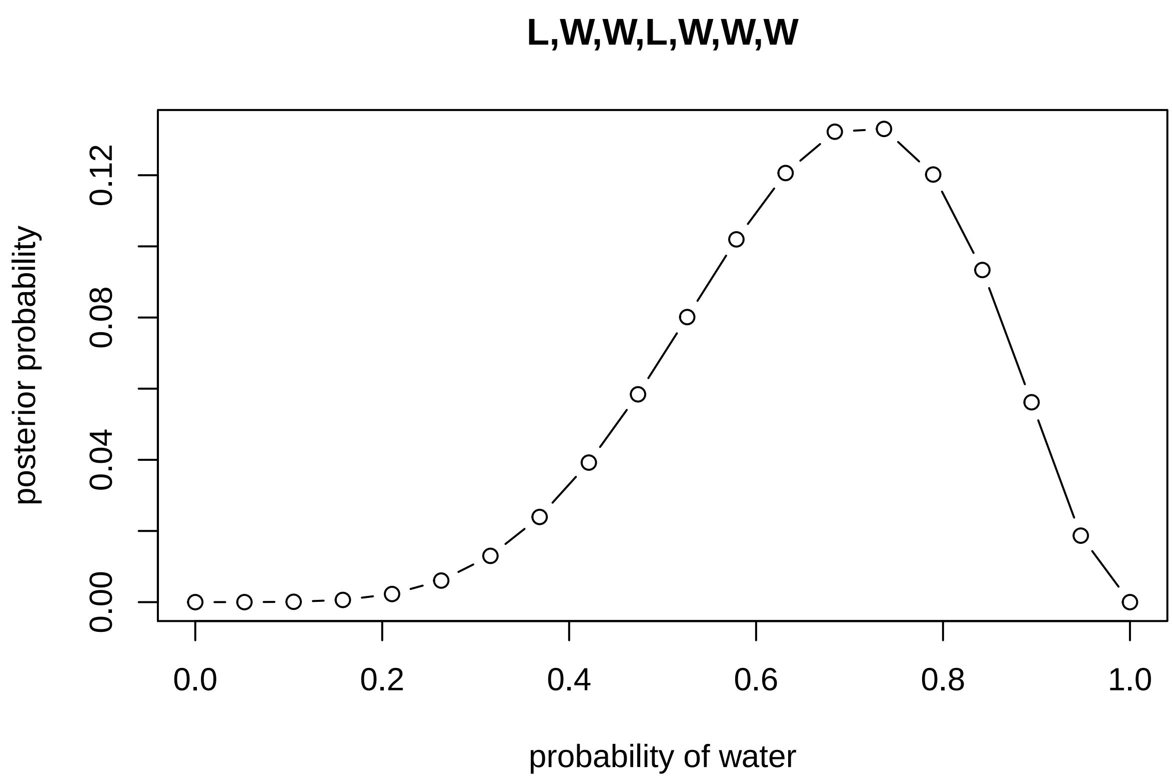
```r
# define grid
p_grid <- seq(from = 0, to = 1, length.out = 20)
# define prior
prior <- rep(1, 20)
# compute likelihood at each value in grid
likelihood1 <- dbinom(3, size = 3 , prob = p_grid)
likelihood2 <- dbinom(3, size = 4, prob = p_grid)
likelihood3 <- dbinom(5, size = 7, prob = p_grid)
# compute product of likelihood and prior
unstd.posterior1 <- likelihood1 * prior
unstd.posterior2 <- likelihood2 * prior
unstd.posterior3 <- likelihood3 * prior
# standardize the posterior, so it sums to 1
posterior1 <- unstd.posterior1 / sum(unstd.posterior1)
posterior2 <- unstd.posterior2 / sum(unstd.posterior2)
posterior3 <- unstd.posterior3 / sum(unstd.posterior3)
# plot the 3 posteriors
plot(p_grid, posterior1, type="b",
     xlab = "probability of water",
     ylab = "posterior probability",
     main = "W,W,W")
```




```r
plot(p_grid, posterior2, type="b",
     xlab = "probability of water",
     ylab = "posterior probability",
     main = "W,W,W,L")
```



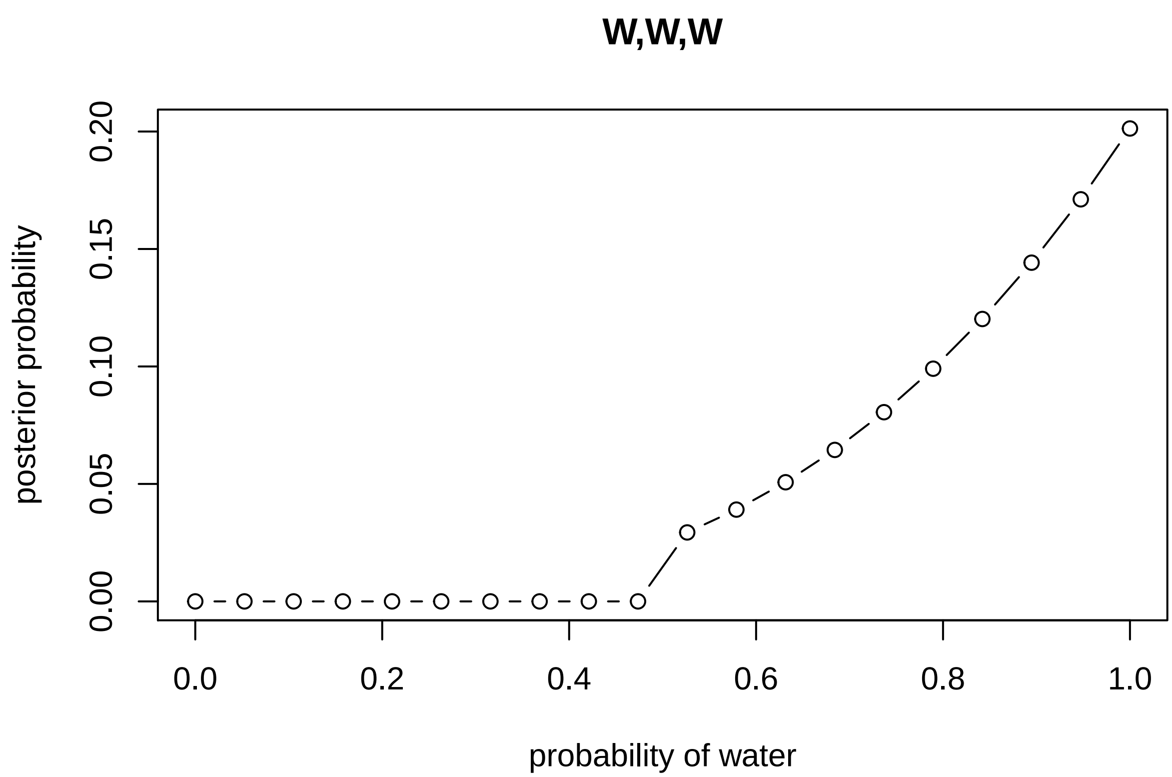

```r
plot(p_grid, posterior3, type="b",
     xlab = "probability of water",
     ylab = "posterior probability",
     main = "L,W,W,L,W,W,W")
```



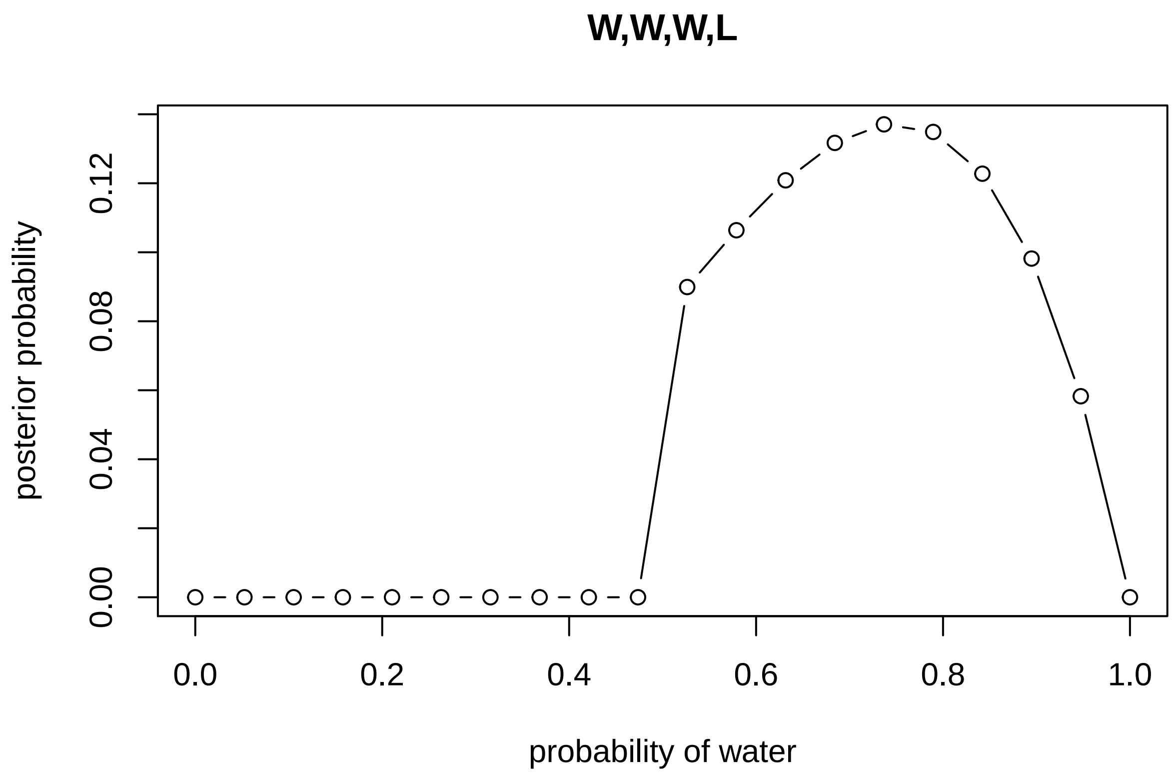
2M2 (2 points)

```
prior <- p_grid < 0.5
for (i in 1:length(prior)) {
    if (prior[i] == FALSE) {
      prior[i] <- 1
    } else {
      prior[i] <- 0
  }
}
# compute likelihood at each value in grid
likelihood1 <- dbinom(3, size = 3 , prob = p_grid)
likelihood2 <- dbinom(3, size = 4, prob = p_grid)
likelihood3 <- dbinom(5, size = 7, prob = p_grid)
# compute product of likelihood and prior
unstd.posterior1 <- likelihood1 * prior
unstd.posterior2 <- likelihood2 * prior
unstd.posterior3 <- likelihood3 * prior
# standardize the posterior, so it sums to 1
posterior1 <- unstd.posterior1 / sum(unstd.posterior1)
posterior2 <- unstd.posterior2 / sum(unstd.posterior2)
posterior3 <- unstd.posterior3 / sum(unstd.posterior3)
# plot the 3 posteriors
plot(p_grid, posterior1, type="b",
```

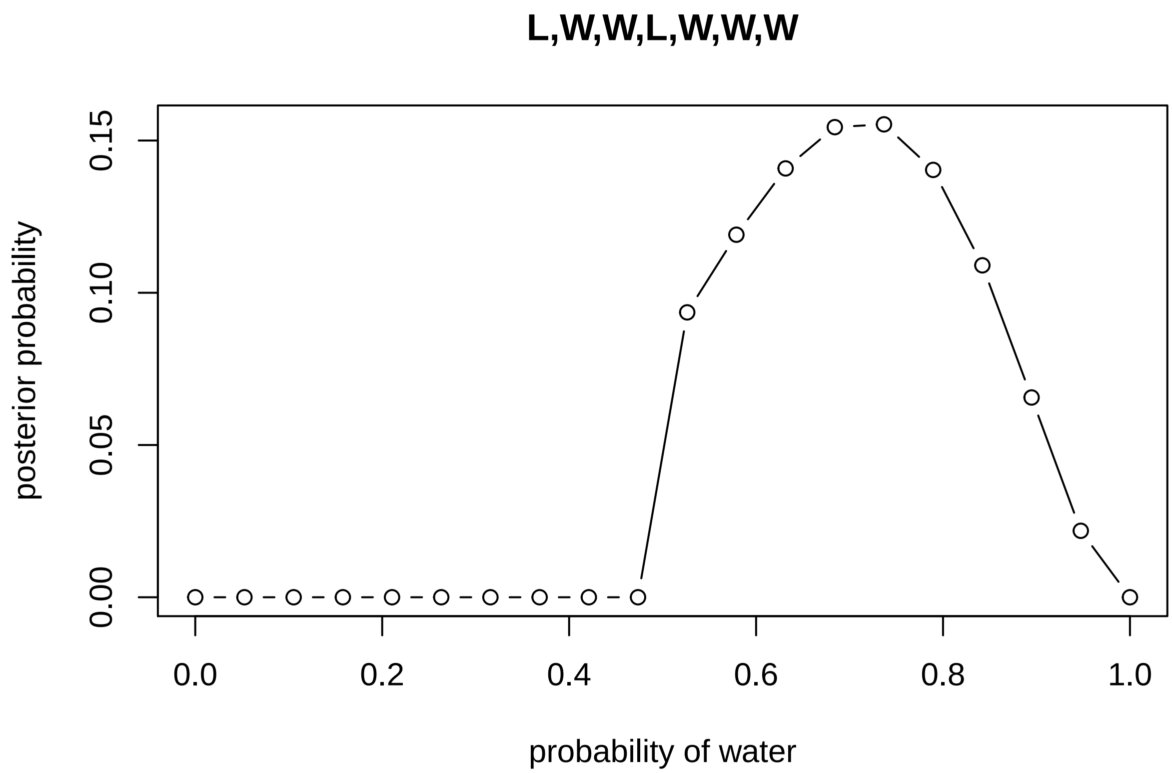
```r
     xlab = "probability of water",
     ylab = "posterior probability",
     main = "W,W,W")
```




```r
plot(p_grid, posterior2, type="b",
     xlab = "probability of water",
     ylab = "posterior probability",
     main = "W,W,W,L")
```




```
plot(p_grid, posterior3, type="b",
     xlab = "probability of water",
     ylab = "posterior probability",
     main = "L,W,W,L,W,W,W")
```



2M3 (2 points)

The "prior" in this case is just the probability of it being the Earth or Mars that you tossed. Since the question says each one is equally likely to be tossed, the prior is just 0.5 for both.

The "likelihood" is the likelihood of seeing land for each of the two globes. This will be 0.3 and 1 for Earth and Mars respectively.

```
prior <- rep(0.5, 2)
likelihood <- c(0.3, 1)
unstd.posterior <- prior * likelihood
posterior <- unstd.posterior / sum(unstd.posterior)
posterior
```

```
## [1] 0.2307692 0.7692308
```

Therefore the posterior probability that the globe was Earth, conditional on having seen land, is 0.2307692.

2M4 (2 points)

Let the B/B card be "card A". Let the "B/W" card be "card B". Let the "W/W" card be "card C".

Let B_2 be side 2 of card B (so this side is white), and so on. Then we have the following possibilities for seeing BLACK:

A_1, A_2, B_1.

Note that there are 3 ways to get the observed data (a black card facing up on the table), and only two of those ways are from cards which have the other side of the card as being black. Therefore, the probability that the other side is black conditional on seeing black is 2/3.

2M5 (2 points)

This time the possibilities for seeing a black card are:

A_1, A_2, B_1, D_1, D_2, and 4 out of these 5 possibilities have a black face on the other side of the card. Therefore, the probability is 4/5.

2M6 (2 points)

This is essentially the same thing as saying that there is now one B/B card, 2 B/W cards, and 3 W/W cards.

This time the possibilities for the black card are:

A_1, A_2, B_1, B_2, and 2 out of these 4 possibilites have a black face on the other side of the card, so the probability is 2/4 or 0.5

2M7 (2 points)

We now have a B/B card, a B/W card, and a W/W card. We reveal the first card to see black, which means the possibilities are A_1, A_2, B_1. We reveal another card and see white, so the possibilities are B_2, C_1, C_2. The possible combinations of cards we could have are:

$(A_1, B_2), (A_1, C_1), (A_1, C_2), (A_2, B_2), (A_2, C_1), (A_2, C_2), (B_1, C_1), (B_1, C_2)$

Note that (B_1, B_2) cannot possibly happen, because we know that the two cards drawn are different cards.

We've just listed 8 possible different combinations, and only 6 of them correspond to having a black side on the other side of the first card drawn. Thus the probability that the first card has two black sides is 6/8 or 0.75.

Hard (4 points total)

2H1 (1 point)

We can solve this using Baye's Theorem.

```
likelihood1 <- 0.1
likelihood2 <- 0.2
likelihood <- c(likelihood1, likelihood2)
prior <- rep(1, 1)
unstd.posterior <- prior * likelihood
posterior <- unstd.posterior / sum(unstd.posterior)
posterior
```

```
## [1] 0.3333333 0.6666667
```

```
# probability that second birth is twins
posterior[1] * 0.1 + posterior[2] * 0.2
```

```
## [1] 0.1666667
```

2H2 (1 point)

The probability that the panda is from species 1 is just the first element of our posterior vector. Therefore, the probability is 0.3333333.

2H3 (1 point)

By the rule of multiplication, P(A) followed by P(B) is just P(A)*P(B). Therefore, the likelihood of having twins followed by a single infant is (in the case of species A) 0.1×0.9. Then,

```
likelihood1 <- 0.1*0.9
likelihood2 <- 0.2*0.8
likelihood <- c(likelihood1, likelihood2)
prior <- rep(1, 2)
unstd.posterior <- prior * likelihood
posterior <- unstd.posterior / sum(unstd.posterior)
posterior[1]
```

```
## [1] 0.36
```

Therefore, the probability that the panda is from species A is 0.36.

2H4 (1 point)

Again, we can use Baye's Theorem. First we'll do it without the birth information.

```
# Without birth information
vettestlikelihood1 <- 0.8
vettestlikelihood2 <- 1 - 0.65 # probability that test doesn't identify it as species B even though it i
vettestlikelihood <- c(vettestlikelihood1, vettestlikelihood2)
prior <- rep(0.5, 2)
vettestunstd.posterior <- vettestlikelihood * prior
vettestposterior <- vettestunstd.posterior / sum(vettestunstd.posterior)
vettestposterior[1]
```

```
## [1] 0.6956522
```

Therefore the probability that the panda is actually from species A is 0.6956522.

Now we compute it using the birth information:

```
birthinfolikelihood1 <- 0.1 * 0.9 # twins followed by single for species A
birthinfolikelihood2 <- 0.2 * 0.8 # twins followed by single for species B
birthinfolikelihood <- c(birthinfolikelihood1, birthinfolikelihood2)
birthinfounstd.posterior <- birthinfolikelihood * prior
birthinfoposterior <- birthinfounstd.posterior / sum(birthinfounstd.posterior)

# probability of species A
composite.unstd.posterior <- vettestposterior * birthinfoposterior
compositeposterior <- composite.unstd.posterior / sum(composite.unstd.posterior)
compositeposterior[1]
```

```
## [1] 0.5625
```