Notes on and Solutions for Statistical Rethinking

Alexander Pastukhov

2020-11-04

Contents

Chapter 1

Precis

This is a collection of solutions for exercises but also of notes that attempt to provide further details and intuition for some topics, such as information theory, information criteria, MCMC algorithms, etc.

Chapter 2

Small Worlds and Large Worlds

```r
library(tidyverse)
library(rethinking)
```

2.1 Solutions to Exercises

2E1

Which of the expressions below correspond to the statement: the probability of rain on Monday?

1. Pr(rain)
2. **Pr(rain | Monday)**
3. Pr(Monday | rain)
4. Pr(rain, Monday) / Pr(Monday)

2E2

Which of the following statements corresponds to the expression: Pr(Monday|rain)?

1. The probability of rain on Monday.
2. The probability of rain, given that it is Monday.
3. **The probability that it is Monday, given that it is raining.**
4. The probability that it is Monday and that it is raining.

2E3

Which of the expressions below correspond to the statement: the probability that it is Monday, given that it is raining?

1. **Pr (Monday | rain)**
2. Pr(rain | Monday)
3. Pr(rain | Monday) Pr(Monday)
4. **Pr(rain | Monday) Pr(Monday) / Pr(rain)**
5. Pr(Monday | rain) Pr(rain) / Pr(Monday)

2E4

The Bayesian statistician Bruno de Finetti (1906– 1985) began his 1973 book on probability theory with the declaration: "PROBABILITY DOES NOT EXIST." The capitals appeared in the original, so I imagine de Finetti wanted us to shout this statement. What he meant is that probability is a device for describing uncertainty from the perspective of an observer with limited knowledge; it has no objective reality. Discuss the globe tossing example from the chapter, in light of this statement. What does it mean to say "the probability of water is 0.7"?

––––––––––––––––––––––––––––––

He meant that, at least at macro level, there are no truly random events. At that scale, all events, such as motion of a globe throughout its flight, are described by deterministic laws of physics. Therefore, the outcome of globe tossing is deterministic and predictable if we have a complete knowledge about its initial momentum and about other forces at play (e.g., air movement). In most cases, we do not have the full knowledge and, therefore, our predictions are likely to diverge from the observed outcome and be mostly but not always correct. Hence, the concept of probability. At the end, noise is never truly random, noise is information that we did not include into our model explicitly.

2M1

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.

```
#' Computes posterior for water probability assuming binomial likelihood
#'
#' @param observations vector of "W" and "L"
#' @param prior numeric vector, its length determines grid resolution
```

```
#'
#' @return tibble with Pwater (from 0 to 1), Posterior (probability of water for given Pwater), c
#'
#' @examples
#' probability_of_water(c("W", "W", "W"), rep(1, 100))
probability_of_water <- function(observations, prior){
  p_grid <- seq(from=0, to=1, length.out=length(prior))
  likelihood <- dbinom(sum(observations == "W"), size=length(observations), prob=p_grid)
  unstandardized_posterior <- likelihood * prior
  tibble(Pwater = p_grid,
         Posterior = unstandardized_posterior / sum(unstandardized_posterior),
         Prior = prior)
}
```
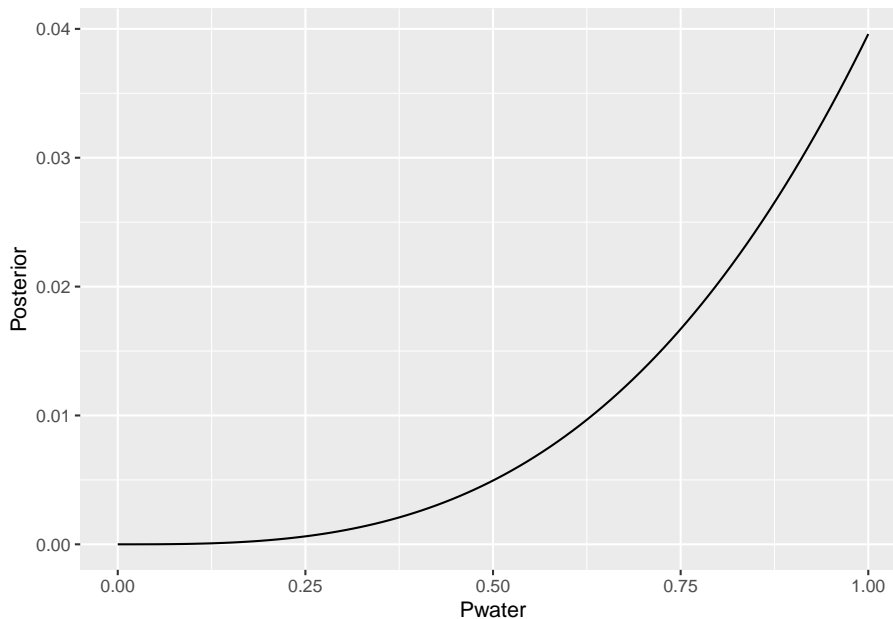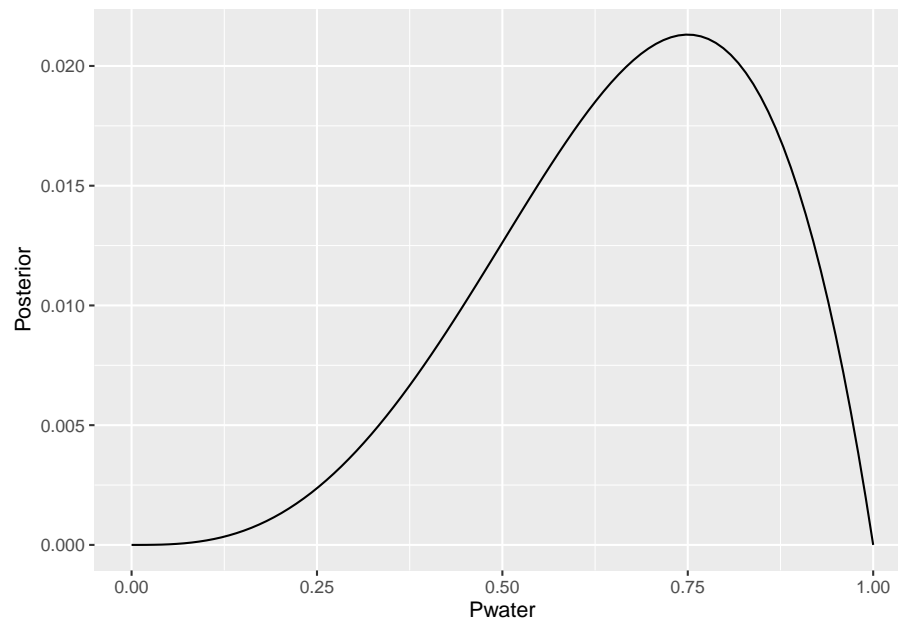
1. W, W, W

```
posterior2M1 <- probability_of_water(c("W", "W", "W"), rep(1, 100))
ggplot(posterior2M1, aes(x=Pwater, y=Posterior)) +
  geom_line()
```
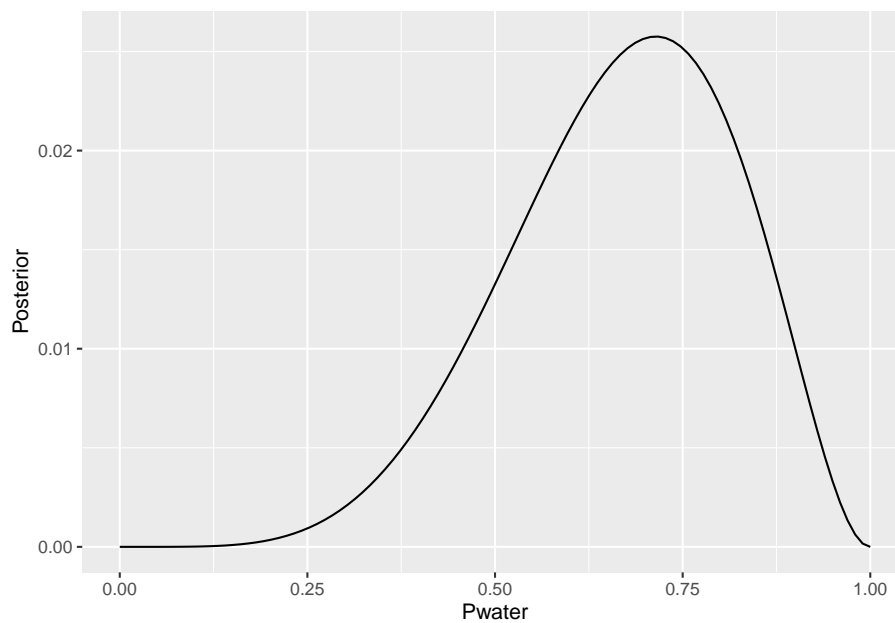

2. W, W, W, L

```r
posterior2M2 <- probability_of_water(c("W", "W", "W", "L"), rep(1, 100))
ggplot(posterior2M2, aes(x=Pwater, y=Posterior)) +
  geom_line()
```


3. L, W, W, L, W, W, W

```r
posterior2M3 <- probability_of_water(c("L", "W", "W", "L", "W", "W", "W"), rep(1, 100))
ggplot(posterior2M3, aes(x=Pwater, y=Posterior)) +
  geom_line()
```

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 and plot the grid approximate posterior distribution for each of the sets of observations in the problem just above.

```r
p_grid <- seq(from=0, to=1, length.out = 100)
priorM2 <- as.numeric(p_grid >= 0.5)

posterior2M3 <- probability_of_water(c("L", "W", "W", "L", "W", "W", "W"), priorM2)
ggplot(posterior2M3, aes(x=Pwater, y=Posterior)) +
  geom_line()
```
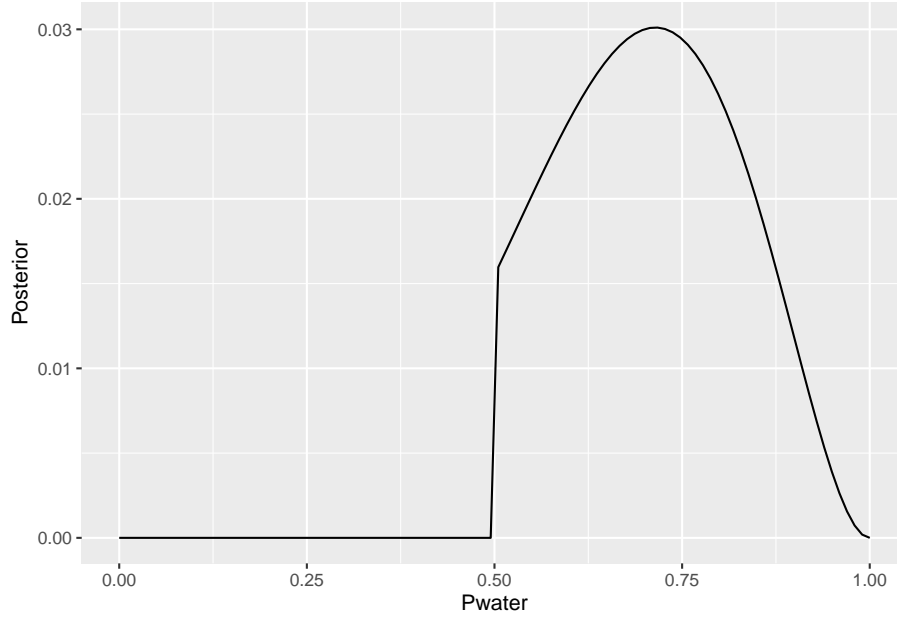
2M3

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 in the air 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" (Pr (Earth | land)), is 0.23.

Bayes Formula tells us that

$$Pr(Planet|land) = \frac{Pr(land|Planet)Pr(Planet)}{Pr(land)}$$

As $Pr(land)$ is a normalization constant, we can ignore it for a moment. Accordingly,

$$uPr(Earth|land) = 0.3 \cdot 0.5 = 0.15 \quad uPr(Mars|land) = 1 \cdot 0.5 = 0.5$$

where $uPr()$ is unstandardized plausibility. Normalizing it, we get

$$Pr(Earth|land) = \frac{0.15}{0.15 + 0.5} = 0.2308$$

2M4

Suppose you have a deck with only three cards. Each card has two sides, and each side is either black or white. One card has two black sides. The second card has one black and one white side. The third card has two white sides. Now suppose all three cards are placed in a bag and shuffled. Someone 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 color of the side facing down. Show that the probability that the other side is also black is 2/3. Use the counting method (Section 2 of the chapter) to approach this problem. This means counting up the ways that each card could produce the observed data (a black side facing up on the table).

- B|B: B(1) → **B(2) : 1**
- B|B: B(2) → **B(1) : 1**
- B|W : B → W : 1
- W|B : 0
- W|W : 0
- W|W : 0

There are three possible outcomes, given the visible side is black and two out of three lead to a black back side: $\frac{2}{3}$.

2M5

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

- B|B: B(1) → **B(2): 1**
- B|B: B(2) → **B(1) : 1**
- B|W : B → W : 1
- W|B : 0
- W|W : 0
- W|W : 0
- B|B: B(3) → **B(4): 1**
- B|B: B(4) → **B(3) : 1**

There are three possible outcomes, given the visible side is black and four out of five lead to a black back side: $\frac{4}{5}$

2M6

Imagine that black ink is heavy, and so cards with black sides are heavier than cards with white sides. As a result, it's less likely that a card with black sides is pulled from the bag. So again assume there are three cards: B/B, B/W, and W/W. After experimenting a number of times, you conclude that for every way to pull the B/B card from the bag, there are 2 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. Use the counting method, as before.

- B|B: B(1) → **B(2): 1 × 1 (prior) = 1**
- B|B: B(2) → **B(1) : 1 × 1 (prior) = 1**
- B|W : B → W : 1 × 2 (prior) = 2
- W|B : 0 × 2 (prior) = 0
- W|W : 0 × 3 (prior) = 0
- W|W : 0 × 3 (prior) = 0

Now the counts are two out of four that other side is black, i.e. 0.5.

2M7

Assume again the original card problem, with a single card showing a black side face up. Before looking at the other side, we draw another card from the bag and lay it face up on the table. The face that is shown on the new card is white. Show that the probability that the first card, the one showing a black side, has black on its other side is now 0.75. Use the counting method, if you can. Hint: Treat this like the sequence of globe tosses, counting all the ways to see each observation, for each possible first card.

Possible card sequences, bold means that back side of the first card is black:

- **B(1)|B(2) → W|B : 1**

- **B(2)|B(1) → W|B : 1**

- B(1)|B(2) → B|W : 0

- B(2)|B(1) → B|W : 0

- **B(1)|B(2) → W(1)|W(2) : 1**

- **B(2)|B(1) → W(1)|W(2) : 1**

- **B(1)|B(2) → W(2)|W(1) : 1**

- **B(2)|B(1) → W(2)|W(1) : 1**

- B|W → B(1)|B(2) : 0

- W|B → B(1)|B(2) : 0

- B|W → B(2)|B(1) : 0

- W|B → B(2)|B(1) : 0

- B|W → W(1)|W(2) : 1

- W|B → W(1)|W(2) : 0

- B|W → W(2)|W(1) : 1

- W|B → W(2)|W(1) : 0

- W(1)|W(2) → W|B : 0

- W(2)|W(1) → W|B : 0

- W(1)|W(2) → B|W : 0

- W(2)|W(1) → B|W : 0

- W(1)|W(2) → B(1)|B(2) : 0

- W(2)|W(1) → B(1)|B(2) : 0

- W(1)|W(2) → B(2)|B(1) : 0

- W(2)|W(1) → B(2)|B(1) : 0

Total of eight possible path, six of them have black back for the first card: 0.75.

2H1

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?

$$uPr(A|twins) = Pr(twins|A)*Pr(A) = 0.1*0.5 = 0.05uPr(B|twins) = Pr(twins|B)*Pr(B) = 0.2*0.5 = 0.1$$

After normalization $Pr(A|twins) = 1/3$ and $Pr(B|twins) = 2/3$.

Probability that you will see twins again per species is

$Pr(twins|A, twins) = Pr(A|twings) * Pr(twins|A) = 1/3*0.1 = 0.1/3 Pr(twins|B, twins) = Pr(B|twing$

The total probability is $\frac{0.1}{3} + \frac{0.4}{3} = \frac{0.5}{3} \approx 0.167$

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.

$uPr(A|twins) = Pr(twins|A)*Pr(A) = 0.1*0.5 = 0.05 uPr(B|twins) = Pr(twins|B)*Pr(B) = 0.2*0.5 =$

After normalization $Pr(A|twins) = 1/3$

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.

$uPr(A|twins) = Pr(twins|A)*Pr(A) = 0.1*0.5 = 0.05 uPr(B|twins) = Pr(twins|B)*Pr(B) = 0.2*0.5 =$

After normalization $Pr(A|twins) = 1/3$ and $Pr(B|twins) = 2/3$.

Given that next birth is a singleton:

$uPr(A|twins, singleton) = Pr(A|twins)*(1-Pr(twins|A)) = 1/3·0.9 = 0.9/3 uPr(B|twins, singleton) =$

After normalization $Pr(A|twins, singleton) = 0.36$ (and $Pr(B|twins, singleton) = 0.64$).

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.

$uPr(A|test) = Pr(test|A)*Pr(A) = 0.8*0.5 uPr(B|test) = (1-Pr(test|B))*Pr(B) = (1-0.65)*0.5 = 0.35*0.5$

After normalization $Pr(A|test) \approx 0.7$ and $Pr(B|test) \approx 0.3$.

Given the twins, this becomes

$uPr(A|test, twins) = Pr(A|test)*Pr(twins|A) = 0.7*0.1 = 0.07 uPr(B|test, twins) = Pr(B|test)*Pr(twins|B) = 0.3$

After normalization $Pr(A|test, twins) \approx 0.533$ and $Pr(B|test, twins) \approx 0.467$.

Given the singleton, this becomes

$uPr(A|test, twins, singleton) = Pr(A|test, twins)*(1-Pr(twins|A)) = 0.533*0.9 = 0.48 uPr(B|test, twins, singleton)$

After normalization $Pr(A|test, twins, singleton) \approx 0.5625$ and $Pr(B|test, twins, singleton) \approx 0.4375$.