

Chapter 2

Easy.

2E1. Which of the expressions below correspond to the statement: the probability of rain on Monday? (1) $\Pr(\text{rain})$ (2) $\Pr(\text{rain}|\text{Monday})$ (3) $\Pr(\text{Monday}|\text{rain})$ (4) $\Pr(\text{rain}, \text{Monday}) / \Pr(\text{Monday})$

#(2) $\Pr(\text{rain}|\text{Monday})$

2E2. Which of the following statements corresponds to the expression: $\Pr(\text{Monday}|\text{rain})$? (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.

(3) The probability that it is Monday, given 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) $\Pr(\text{Monday}|\text{rain})$ (2) $\Pr(\text{rain}|\text{Monday})$ (3) $\Pr(\text{rain}|\text{Monday}) \Pr(\text{Monday})$ (4) $\Pr(\text{rain}|\text{Monday}) \Pr(\text{Monday}) / \Pr(\text{rain})$ (5) $\Pr(\text{Monday}|\text{rain}) \Pr(\text{rain}) / \Pr(\text{Monday})$

(4) $\Pr(\text{rain}|\text{Monday}) \Pr(\text{Monday}) / \Pr(\text{rain})$

2E4. The Bayesian statistician Bruno de Finetti (1906–1985) began his 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’?

Essentially, it means that of the experiments we have conducted, on average, water is observed 70% of

Medium.

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

skip

2M2. Now assume a prior for p that is equal to zero when $p < 0.5$ and is a positive constant when $p \geq 0.5$. Again compute and plot the grid approximate posterior distribution for each of the sets of observations in the problem just above.

skip

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’ ($\Pr(\text{Earth}|\text{land})$), is 0.23.

```
mars_p_land = 1
earth_p_land = .3
```

```
#  $P(E \mid \text{land}) = P(\text{land} \mid E) * P(E) / P(\text{land})$ 

earth_p_land * .5 / (.5*mars_p_land + .5*earth_p_land)

## [1] 0.2307692
```

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).

```
# c1 = BB = 2 ways
# c2 = BW = 1 way
# c3 = WW = 0 ways

# there are 3 ways to get the observed data. of those 3 ways, 2 out of 3 of them have a black side on t
```

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.

```
# BB = 2 ways
# BW = 1 way
# WW = 0 ways
# BB = 2 ways

# 5 total ways to realize the observed data. Of those 5 ways, 4 have other side being black, so 80%
```

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.

```
#BB = 2 ways
#BW = 1 way (2x)
#WW = 0 ways (3x)

# normalize all "ways" to the same scale:
#BB = 2*1 = 2 ways
#BW = 1*2 = 2 way
#WW = 0*3 = 0 ways

# 4 ways to get observed data, 2 of them result in a BB results, hence 50%
```

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.

```

# BB:
# BW:
# WW:

# Bb, bW
# Bb, Ww
# Bb, ww

# bB, bW
# bB, Ww
# bB, ww

# Bw: Ww
# Bw: ww

# 6 out of 8 ways have the first card being the BB card

```

Hard.

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?

```

p_a = .5
p_b = .5

p_t_g_a = .1
p_t_g_b = .2

p_twins = .5*.1 + .5*.2

# if panda is species A, p = .1
# if panda is species B, p = .2

p_a_g_twins = p_t_g_a * p_a / p_twins
p_b_g_twins = p_t_g_b * p_b / p_twins

p_b_g_twins * .2 + p_a_g_twins*.1

```

```
## [1] 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.

```
p_a_g_twins
```

```
## [1] 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.

```

p_a = .33
p_s_g_a = .9
p_s = .9*.33 + .8*.66

p_a_g_s = p_s_g_a * p_a / p_s
p_a_g_s

```

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

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.

```

p_a = .5
p_b = .5

p_test_a_given_a = .8

p_test_a = .5*.8 + .5*.35 # my answer
#p_test_a = .5*.8 + .5*.65 # right answer ..?

p_a_g_positive_test = p_test_a_given_a * p_a / p_test_a
p_a_g_positive_test

```

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

```

p_a = .36
p_b = .64

p_test_a_given_a = .8

p_test_a = .36*.8 + .35*.64 # my answer
#p_test_a = .36*.8 + .65*.64 # right answer ..?

p_a_g_positive_test = p_test_a_given_a * p_a / p_test_a
p_a_g_positive_test

```

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