Appendix C

Answers to Exercises

Part I, Introduction to Probability

Chapter 1

Q1. Rewrite the following statements as equations using the mathematical notation you learned in this chapter:

The probability of rain is low

The probability of rain given that it is cloudy is high

The probability of you having an umbrella given it is raining is much greater than the probability of you having an umbrella in general.

A1.

$$P(\text{rain}) = \text{low}$$

$$P(\text{rain} \mid \text{cloudy}) = \text{high}$$

$$P(\text{umbrella} \mid \text{rain}) >> P(\text{umbrella})$$

Q2. Organize the data you observe in the following scenario into a mathematical notation, using the techniques we've covered in this chapter. Then come up with a hypothesis to explain this data:

> You come home from work and notice that your front door is open and the side window is broken. As you walk inside you immediately notice that your laptop is missing.

A2. We first want to describe our data with a variable:

$$D = \text{door open, window broken, laptop missing}$$

Our data represents three facts you observed upon arriving home. An immediate explanation for this data is that you've been robbed! We would express this mathematically as:

$$H_1 = \text{you've been robbed!}$$

Now we can express this as "The probability of seeing all these things, given that you've been robbed " as:

$$P(D \mid H_1)$$

Q3. The following scenario adds data to the previous one. Demonstrate how this new information changes your beliefs and come up with a second hypothesis to explain the data, using the notation you've learned in this chapter.

> A neighborhood child runs up to you and apologizes profusely for accidentally throwing a rock through your window. They claim that they saw the laptop and didn't want it stolen so they opened the front door to grab it, and your laptop is safe at their house.

A3. Now we have another hypothesis for the things you observed:

$$H_2 = \text{child accidentally broke your window} \quad \text{and took the laptop for safekeeping}$$

We can express this as:

$$P(D \mid H_2) >> P(D \mid H_1)$$

And we would expect:

$$\frac{P(D \mid H_2)}{P(D \mid H_1)} = \text{a large number}$$

Of course, you might think that this child is untrustworthy and notorious for causing trouble, which might change your mind about how likely their explanation is and lead you to hypothesize that they have robbed you! As you move through this book, you'll learn more about how you can reflect that mathematically.

Chapter 2

Q1. What is the probability of rolling two six-sided dice and getting a value greater than 7?

A1. There are 36 possible ways that we could roll the two dice (if we consider 1 and 6 different from 6 and 1). You can list this all out on paper (or find a way to do it in code, which will be faster). Fifteen of these 36 pairs are *greater* than 7. So the probability that you'll get a value greater than 7 is $\frac{15}{36}$.

Q2. What is the probability of rolling three six-sided dice and getting a value greater than 7?

A2. With three rolls there are 216 different possible outcomes. You can write these out on a sheet of paper, which is fine but will take you quite a while. You can see why learning the basics of coding is helpful, as there are various programs (even messy ones) you can write to solve this problem. For example, we can find the answer with this simple set of `for` loops in R:

```
count <- 0
for(roll1 in c(1:6)){
  for(roll2 in c(1:6)){
    for(roll3 in c(1:6)){
      count <- count + ifelse(roll1+roll2+roll3 > 7,1,0)
    }
  }
}
```

Here you can see the count is 181, so the probability of the rolls totaling more than 7 is $\frac{181}{216}$. As noted, however, there are many ways to compute this. One alternative is this single (difficult to read!) line of R, which does the same thing as the `for` loops:

```
sum(apply(expand.grid(c(1:6),c(1:6),c(1:6)),1,sum) > 7)
```

When learning to code, you should focus on getting the correct answer over using a particular approach to arrive at it.

Q3. The Yankees are playing the Red Sox. You're a diehard Sox fan and bet your friend they'll win the game. You'll pay your friend $30 if the Sox lose and your friend will have to pay you only $5 if the Sox win. What is the probability you have intuitively assigned to the belief that the Red Sox will win?

A3. We can see that the odds you've given for the Red Sox to win is:

$$O(\text{Red Sox win}) = \frac{30}{5} = 6$$

Recalling our formula for converting odds to probabilities, we can translate the odds into a probability that the Red Sox will win:

$$P(\text{Red Sox win}) = \frac{O(\text{Red Sox win})}{1 + O(\text{Red Sox win})} = \frac{6}{7}$$

So, based on the bet you take, you would say there's about an 86 percent chance that the Red Sox will win!

Chapter 3

Q1. What is the probability of rolling a 20 three times in a row on a 20-sided die?

A1. The probability of rolling a 20 is $\frac{1}{20}$, and to determine the probability of rolling three in a row, we must use our product rule:

$$P(\text{three 20s}) = \frac{1}{20} \times \frac{1}{20} \times \frac{1}{20} = \frac{1}{8{,}000}$$

Q2. The weather report says there's a 10 percent chance of rain tomorrow, and you forget your umbrella half the time you go out. What is the probability that you'll be caught in the rain without an umbrella tomorrow?

A2. Again, we can use the product rule to solve this problem. We know that *P*(rain) = 0.1 and *P*(forgetting umbrella) = 0.5, so:

$$P(\text{rain, forget umbrella}) = P(\text{rain}) \times P(\text{forget umbrella}) = 0.05$$

As you can see, there's only a 5 percent chance that you'll find yourself caught in the rain without an umbrella.

Q3. Raw eggs have a 1/20,000 probability of having salmonella. If you eat two raw eggs, what is the probability you ate a raw egg with salmonella?

A3. For this question, we need to use the sum rule because if *either* egg has salmonella, you'll get sick:

$$P(\text{egg}_1) + P(\text{egg}_2) - P(\text{egg}_1) \times P(\text{egg}_2) = \frac{1}{20{,}000} + \frac{1}{20{,}000} - \frac{1}{20{,}000} \times \frac{1}{20{,}000}$$

$$= \frac{39{,}999}{400{,}000{,}000}$$

…which is just a hair under $\frac{1}{10{,}000}$.

Q4. What is the probability of either flipping two heads in two coin tosses or rolling three 6s in three six-sided dice rolls?

A4. For this exercise, we need to combine our product rule and our sum rule. First let's calculate *P*(two heads) and *P*(three 6s) separately. Each probability uses the product rule:

$$P(\text{two heads}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

$$P(\text{three 6s}) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216}$$

Now we need to use the sum rule to figure out the probability of either of these happening, *P*(two heads or three 6s):

$$P(\text{two heads}) + P(\text{three 6s}) - P(\text{two heads}) \times P(\text{three 6s})$$

$$= \frac{1}{4} + \frac{1}{216} - \frac{1}{4} \times \frac{1}{216} = \frac{73}{288}$$

…which is just a little bit more than a 25 percent chance.

Chapter 4

Q1. What are the parameters of the binomial distribution for the probability of rolling either a 1 or a 20 on a 20-sided die, if we roll the die 12 times?

A1. We're looking for an event to happen 1 time out of 12 trials, so n = 12, and k = 1. We have 20 sides and care about 2 of them, so $p = \frac{2}{20} = \frac{1}{10}$.

Q2. There are four aces in a deck of 52 cards. If you pull a card, return the card, then reshuffle and pull a card again, how many ways can you pull just one ace in five pulls?

A2. We don't even need combinatorics for this one. There are five possible cases, if we imagine A stands for "ace" and x for anything else:

Axxxx

xAxxx

xxAxx

xxxAx

xxxxA

We could just call this $\binom{5}{1}$ or, in R, `choose(5,1)`. Either way, the answer is 5.

Q3. For the example in question 2, what is the probability of pulling five aces in 10 pulls (remember the card is shuffled back in the deck when it is pulled)?

A3. This is the same as $B(5; 10, \frac{1}{23})$

As expected, the probability of this is extremely low: about $\frac{1}{32{,}000}$.

Q4. When you're searching for a new job, it's always helpful to have more than one offer on the table so you can use it in negotiations. If you have a 1/5 probability of receiving a job offer when you interview, and you interview with seven companies in a month, what is the probability you'll have at least two competing offers by the end of that month?

A4. We can use the following R code to compute this answer:

```
> pbinom(1,7,1/5,lower.tail = FALSE)
 0.4232832
```

As you can see, there's about a 42 percent chance of receiving two *or more* job offers if you interview at seven companies.

Q5. You get a bunch of recruiter emails and find out you have 25 interviews lined up in the next month. Unfortunately, you know this will leave you exhausted, and the probability of getting an offer will drop to 1/10 if you're tired. You really don't want to go on this many interviews unless you are at least twice as likely to get at least two competing offers. Are you more likely to get at least two offers if you go for 25 interviews, or stick to just 7?

A5. Let's write a bit more R code to sort this out:

```
p.two.or.more.7 <- pbinom(1,7,1/5,lower.tail = FALSE)
p.two.or.more.25 <- pbinom(1,25,1/10,lower.tail = FALSE)
```

Even with the reduced probability of an offer, your probability of getting at least two offers in 25 interviews is 73 percent. However, you'll go this route only if you are *twice* as likely. As we can see in R:

```
> p.two.or.more.25/p.two.or.more.7
[1] 1.721765
```

you're only 1.72 times more likely to get two or more offers, so all the hassle isn't worth it.

Chapter 5

Q1. You want to use the beta distribution to determine whether or not a coin you have is a fair coin—meaning that the coin gives you heads and tails equally. You flip the coin 10 times and get 4 heads and 6 tails. Using the beta distribution, what is the probability that the coin will land on heads more than 60 percent of the time?

A1. We would model this as Beta(4,6). We want to calculate the integral from 0.6 to 1, which we can do in R like so:

```
integrate(function(x) dbeta(x,4,6),0.6,1)
```

This tells us there is about a 10 percent chance that the true probability of getting heads is 60 percent or greater.

Q2. You flip the coin 10 more times and now have 9 heads and 11 tails total. What is the probability that the coin is fair, using our definition of fair, give or take 5 percent?

A2. Our beta distribution is now Beta(9,11). But we want to know the probability that the coin is fair, meaning the chance of getting heads is 0.5, within 0.05 probability either way. This means we need to integrate our new distribution between 0.45 and 0.55. We can do so with this line of R:

```
integrate(function(x) dbeta(x,9,11),0.45,0.55)
```

Now we find that there's a 30 percent chance that our coin is fair, given the new data we have.

Q3. Data is the best way to become more confident in your assertions. You flip the coin 200 more times and end up with 109 heads and 111 tails. Now what is the probability that the coin is fair, give or take 5 percent?

A3. Given the previous question, this answer is pretty straightforward:

```
integrate(function(x) dbeta(x,109,111),0.45,0.55)
```

Now we're 86 percent certain that the coin is reasonably fair. Notice that the key to becoming more certain was to include more data.

Part II, Bayesian Probability and Prior Probabilities

Chapter 6

Q1. What piece of information would we need in order to use Bayes' theorem to determine the probability that someone in 2010 who had GBS *also* had the flu vaccine that year?

A1. We want to figure out *P*(flu vaccines | GBS). We can solve this using Bayes' theorem, provided we have all these pieces of information:

$$P(\text{flu vaccine} \mid \text{GBS}) = \frac{P(\text{flu vaccine}) \times P(\text{GBS} \mid \text{flu vaccine})}{P(\text{GBS})}$$

Of these pieces of information, the only one we don't know is the probability of getting the flu vaccine in the first place. We could probably get this information from the Centers for Disease Control and Prevention or another national data collection service.

Q2. What is the probability that a random person picked from the population is female and is *not* color blind?

We know that *P*(female) = 0.5 and that *P*(color blind | female) = 0.005, but we want to know the probability that someone is female and *not* color blind, which is 1 – *P*(color blind | female) = 0.995. So:

$$P(\text{female, not color blind}) = P(\text{female}) \times P(\text{not color blind} \mid \text{female}) = 0.5 \times 0.995 = 0.4975$$

Q3. What is the probability that a male who received the flu vaccine in 2010 is either color blind or has GBS?

A3. This problem may initially seem complex, but we can simplify it a bit. Let's start by just working on the probability of being color blind given someone is male, and the probability of having GBS given they've received the flu vaccine. Notice that we're taking a bit of a shortcut, since being male is independent from GBS (as far as we're concerned here) and having a flu vaccine has no impact on being color blind. We'll make each of these into a separate probability:

$$P(A) = P(\text{color blind} \mid \text{male})$$

$$P(B) = P(\text{GBS} \mid \text{flu vaccine})$$

Luckily we already did all this work earlier in the chapter, so we know that $P(A) = \frac{4}{1000}$ and $P(B) = \frac{3}{100{,}000}$.

Now we can just use our sum rule to solve this:

$$P(\text{A or B}) = P(A) + P(B) - P(A) \times P(B \mid A)$$

And because the probability of being color blind, as far as we know, has nothing to do with the probability of GBS, we know that *P*(B | A) = *P*(B). Plugging in our numbers, we get an answer of $\frac{100,747}{25,000,000}$ or 0.00403. This is just bit larger than the chance of being color blind given someone is male, because the probability of GBS is so small.

Chapter 7

Q1. Kansas City, despite its name, sits on the border of two US states: Missouri and Kansas. The Kansas City metropolitan area consists of 15 counties, 9 in Missouri and 6 in Kansas. The entire state of Kansas has 105 counties and Missouri has 114. Use Bayes' theorem to calculate the probability that a relative who just moved to a county in the Kansas City metropolitan area also lives in a county in Kansas. Make sure to show *P*(Kansas) (assuming your relative lives either in Kansas or Missouri), *P*(Kansas City metropolitan area), and *P*(Kansas City metropolitan area | Kansas).

A1. Hopefully it is pretty clear that there are 15 counties in the Kansas City metro area, and 6 of them are in Kansas, so the probability of being in Kansas, given you know someone lives in the Kansas City metro area, should be $\frac{6}{15}$, which is equivalent to $\frac{2}{5}$. The purpose of this question, however, is not just to get an answer but to show that Bayes' theorem provides the tools to solve it. When we work on harder problems, it will be very helpful to have established trust in Bayes' theorem.

So, to solve *P*(Kansas | Kansas City), we can use Bayes' theorem as follows:

$$P(\text{Kansas} \mid \text{Kansas City}) = \frac{P(\text{Kansas City} \mid \text{Kansas}) \times P(\text{Kansas})}{P(\text{Kansas City})}$$

From our data we know that of the 105 counties in Kansas, 6 are in the Kansas City metro area:

$$P(\text{Kansas City} \mid \text{Kansas}) = \frac{6}{105}$$

And between Missouri and Kansas there are 219 counties, 105 of which are in Kansas:

$$P(\text{Kansas}) = \frac{105}{219}$$

And of this total of 219 counties, 15 are in the Kansas City metro area:

$$P(\text{Kansas City}) = \frac{15}{219}$$

Filling in all of the parts of Bayes' theorem, then, gives us:

$$P(\text{Kansas} \mid \text{Kansas City}) = \frac{\frac{6}{105} \times \frac{105}{219}}{\frac{15}{219}} = \frac{2}{5}$$

Q2. A deck of cards has 52 cards with suits that are either red or black. There are four aces in a deck of cards: two red and two black. You remove a red ace from the deck and shuffle the cards. Your friend pulls a black card. What is the probability that it is an ace?

A2. As with the previous question, we can easily see there are 26 black cards and 2 of them are aces, so there is a $\frac{2}{26}$ or $\frac{1}{13}$ probability of getting an ace if we have a black card. But, again, we want to establish some trust in Bayes' theorem and not take so many mathematical mental shortcuts. Using Bayes' theorem we get:

$$P(\text{ace} \mid \text{black card}) = \frac{P(\text{black card} \mid \text{ace}) \times P(\text{ace})}{P(\text{black card})}$$

There are 26 black cards in the deck, out of what is now 51 cards since we removed 1 red ace. If we know that we have an ace, the probability it is black is:

$$P(\text{black card} \mid \text{ace}) = \frac{2}{3}$$

In this deck there are now 51 cards, only 3 of which are aces, so we have:

$$P(\text{ace}) = \frac{3}{51}$$

Finally, we know that of the remaining 51 cards, 26 of them are black, so:

$$P(\text{black card} = \frac{26}{51})$$

Now we have enough information to solve our problem:

$$P(\text{ace} \mid \text{black card}) = \frac{\frac{2}{3} \times \frac{3}{51}}{\frac{26}{51}} = \frac{1}{13}$$

Chapter 8

Q1. As mentioned, you might disagree with the our original probability assigned to the likelihood:

***P*(broken window, open front door, missing laptop | robbed) = 3/10**

How much does this change our strength in believing H_1 over H_2?

A1. To start, remember that:

$$P(\text{broken window, open front door, missing laptop} \mid \text{robbed}) = P(D \mid H_1)$$

To see how this changes our beliefs, all we have to do now is replace this part in our ratio:

$$\frac{P(H_1) \times P(\mathrm{D} \mid \mathrm{H}_1)}{P(H_2) \times P(D \mid H_2)}$$

We already know that the denominator of our formula is $\frac{1}{21{,}900{,}000}$ and that $P(H_1) = \frac{1}{1000}$, so to get our answer we just have to add our changed belief in $P(\mathrm{D} \mid H_1)$:

$$\frac{\frac{1}{1{,}000} \cdot \frac{3}{100}}{\frac{1}{21{,}900{,}000}} = 657$$

So when we believe $(\mathrm{D} \mid H_1)$ is 10 times less likely, our ratio is 10 times smaller (though still very much in favor of H_1).

Q2. How unlikely would you have to believe being robbed is—our prior for H_1—in order for the ratio of H_1 to H_2 to be even?

A2. In the previous answer, decreasing our probability in $P(\mathrm{D} \mid H_1)$ by 10 times reduced our ratio 10 times. This time, we want to change $P(H_1)$ so that our ratio is 1, which means we need to make it 657 times smaller:

$$\frac{\frac{1}{1{,}000 \times 657} \times \frac{3}{100}}{\frac{1}{21{,}900{,}000}} = 1$$

So our new $P(H_1)$ needs to be $\frac{1}{657{,}000}$, which is a pretty extreme belief in the unlikeliness of getting robbed!

Chapter 9

Q1. A friend finds a coin on the ground, flips it, and gets six heads in a row and then one tails. Give the beta distribution that describes this. Use integration to determine the

probability that the true rate of flipping heads is between 0.4 and 0.6, reflecting that the coin is reasonably fair.

A1. We can represent this as a beta distribution with $\alpha = 6$ and $\beta = 1$, since we have six heads and one tail. In R we can integrate this as follows:

```
> integrate(function(x) dbeta(x,6,1),0.4,0.6)
0.04256 with absolute error < 4.7e-16
```

With about a 4 percent chance this coin is fair, based on likelihood alone we would consider it unfair.

Q2. Come up with a prior probability that the coin *is* fair. Use a beta distribution such that there is at least a 95 percent chance that the true rate of flipping heads is between 0.4 and 0.6.

A2. Any $\alpha_{\text{prior}} = \beta_{\text{prior}}$ will give us a "fair" prior; and the larger those values are, the stronger that prior is. For example, if we use 10 we get:

```
> prior.val <- 10
> integrate(function(x) dbeta(x,6+prior.val,1+prior.val),0.4,0.6)
0.4996537 with absolute error < 5.5e-15
```

But of course that's only a 50 percent chance that the coin is fair. Using a bit of trial and error, we can find a number that works for us. Using $\alpha_{\text{prior}} = \beta_{\text{prior}} = 55$, we find that this gives a prior that achieves our goal:

```
> prior.val <- 55
> integrate(function(x) dbeta(x,6+prior.val,1+prior.val),0.4,0.6)
0.9527469 with absolute error < 1.5e-11
```

Q3. Now see how many more heads (with no more tails) it would take to convince you that there is a reasonable chance that the coin is *not* fair. In this case, let's say that this means that our belief in the rate of the coin being between 0.4 and 0.6 drops below 0.5.

A3. Again, we can solve this problem simply through trial and error until we get an answer that works. Remember that we're still using Beta(55,55) as our prior. This time, we want to see how much we can add to our α in order to change the probability of a fair coin to around 50 percent. We can see that with five more heads, our posterior drops to 90 percent:

```
> more.heads <- 5
> integrate(function(x) dbeta(x,6+prior.val+more.heads,1+prior.val),0.4,0.6)
0.9046876 with absolute error < 3.2e-11
```

And if we got 23 more heads, we'd find that the probability of the coin being fair now would be about 50 percent. This shows that even a strong prior belief can be overcome with more data.

Part III, Parameter Estimation

Chapter 10

Q1. It's possible to get errors that don't quite cancel out the way we want. In the Fahrenheit temperature scale, 98.6 degrees is the normal body temperature and 100.4 degrees is the typical threshold for a fever. Say you are taking care of a child that feels warm and seems sick, but you take repeated readings from the thermometer and they all read between 99.5 and 100.0 degrees: warm, but not quite a fever. You try the thermometer yourself and get several readings between 97.5 and 98. What could be wrong with the thermometer?

A1. It looks like the thermometer might be giving *biased* measurements that tend to be off by 1 degree F. If you added 1 degree to your results, you'd see that they were between 98.5 and 99, which seems correct for someone that normally has a 98.6 degree body temperature.

Q2. Given that you feel healthy and have traditionally had a very consistently normal temperature, how could you alter the measurements 100, 99.5, 99.6, and 100.2 to estimate if the child has a fever?

A2. If measurements are biased, it means that they are systematically wrong, so no amount of sampling will correct this on its own. To correct our original measurements, we could just add 1 degree to each.

Chapter 11

Q1. One of the benefits of variance is that squaring the differences makes the penalties exponential. Give some examples of when this would be a useful property.

A1. Exponential penalties are very useful for many everyday situations. One of the most obvious is physical distance. Suppose someone invents a teleporter that can transport you to another location. If you miss the mark by 3 feet, that's fine; 3 miles might be okay; but 30 miles could be incredibly dangerous. In this case, you want the penalty for being far away from your target to get much more severe as it grows.

Q2. Calculate the mean, variance, and standard deviation for the following values: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

A1. Mean = 5.5, variance = 8.25, standard deviation = 2.87.

Chapter 12

A Note on Standard Deviation

R has a built-in function, `sd`, that computes the *sample* standard deviation, rather than the standard deviation we've discussed in the book. The idea of sample standard deviation is that you average by $n - 1$ instead of n. Sample standard deviation is used in classical statistics to make estimates about population means given data. Here, the function `my.sd` computes the standard deviation used in this book:

```
my.sd <- function(val){
  val.mean <- mean(val)
  sqrt(mean((val.mean-val)^2))
}
```

As your data set grows in size, the difference between sample standard deviation and the true standard deviation will become irrelevant. But for the small data sizes in these examples, it will make a small difference. For all the examples in Chapter 12 I've used `my.sd`, but sometimes for convenience I'll just use the default, `sd`.

Q1. What is the probability of observing a value five sigma greater than the mean or more?

A1. We can use `integrate()` on a normal distribution with a mean of 0 and standard deviation of 1. Then we just integrate from 5 to some reasonably large number like 100:

```
> integrate(function(x) dnorm(x,mean=0,sd=1),5,100)
2.88167e-07 with absolute error < 5.6e-07
```

Q2. A fever is any temperature greater than 100.4 degrees Fahrenheit. Given the following measurements, what is the probability that the patient has a fever?

100.0, 99.8, 101.0, 100.5, 99.7

A2. We'll start by figuring out the mean and standard deviation of our data:

```
temp.data <- c(100.0, 99.8, 101.0, 100.5, 99.7)
temp.mean <- mean(temp.data)
temp.sd <- my.sd(temp.data)
```

Then we just use `integrate()` to find out the probability that the temperature is over 100.4:

```
> integrate(function(x) dnorm(x,mean=temp.mean,sd=temp.sd),100.4,200)
0.3402821 with absolute error < 1.1e-08
```

Given these measurements, there's about a 34 percent chance of fever.

Q3. Suppose in Chapter 11 we tried to measure the depth of a well by timing coin drops and got the following values:

2.5, 3, 3.5, 4, 2

The distance an object falls can be calculated (in meters) with the following formula:

$$\text{distance} = 1/2 \times G \times \text{time}^2$$

where G is 9.8 m/s/s. What is the probability that the well is over 500 meters deep?

A3. Let's start by putting our time data in R:

```
time.data <- c(2.5,3,3.5,4,2)
time.data.mean <- mean(time.data)
time.data.sd <- my.sd(time.data)
```

Next we need to figure out how much time it takes to reach 500 meters. We need to solve:

$$\frac{1}{2} \times G \times t^2 = 500$$

If G is 9.8, we can work out that time (t) is about 10.10 seconds (you can also solve this by making a function in R and just manually iterating, or look up the solution on something like Wolfram Alpha). Now we just have to integrate our normal distribution to beyond 10.1:

```
> integrate(function(x)
dnorm(x,mean=time.data.mean,sd=time.data.sd),10.1,200)
2.056582e-24 with absolute error < 4.1e-24
```

This is basically 0 probability, so we can be pretty certain that the well is *not* 500 meters deep.

Q2. What is the probability there is no well (i.e., the well is really 0 meters deep)? You'll notice that probability is higher than you might expect, given your observation that there *is* a well. There are two good explanations for this probability being higher than it should be. The first is that the normal distribution is a poor model for our measurements; the second is that, when making up numbers for an example, I chose values that you likely wouldn't see in real life. Which is more likely to you?

A2. If we do the same integration but with –1 to 0, we get:

```
integrate(function(x)
dnorm(x,mean=time.data.mean,sd=time.data.sd),-1,0)
1.103754e-05 with absolute error < 1.2e-19
```

It's small, but the probability that *there is no well* is greater than 1 in 100,000. But you can see a well! It's right in front of you! So, even if the probability is small, it's not really that close to zero. Now should we question the model, or should we question the data? As a Bayesian, generally you should favor questioning the model over the data. For example, movement in stock prices will typically have very high σ events during financial crises. This means that the normal distribution is a bad model for stock movements. However, in this example, there's no reason to question the assumptions of the normal distribution, and in fact these are the original numbers that I picked for the previous chapter until my editor pointed out that the values seemed too spread out.

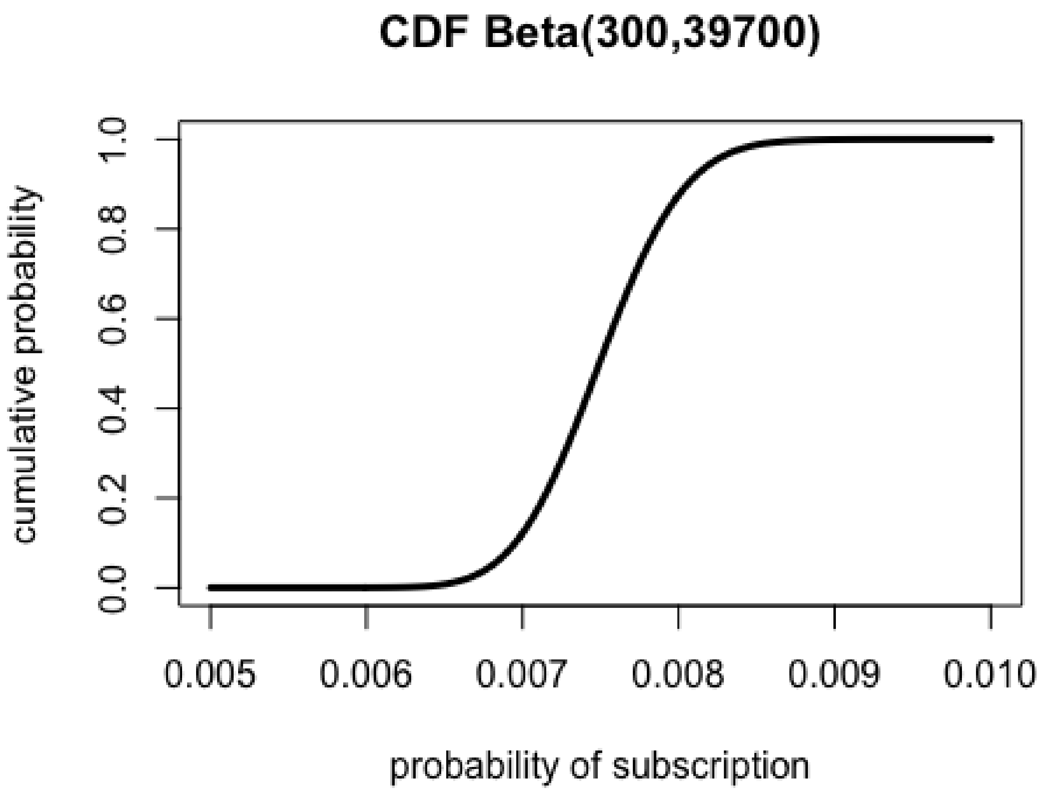
One of the greatest virtues in statistical analysis is skepticism. In practice I have been given bad data to work with on a few occasions. Even though models are always imperfect, it's very important to make sure that you can trust your data as well. See if the assumptions you have about the world hold up and, if they don't, see if you can be convinced that you still trust your model and your data.

Chapter 13

Q1. Using the code example for plotting the PDF on page 127, plot the CDF and quantile functions.

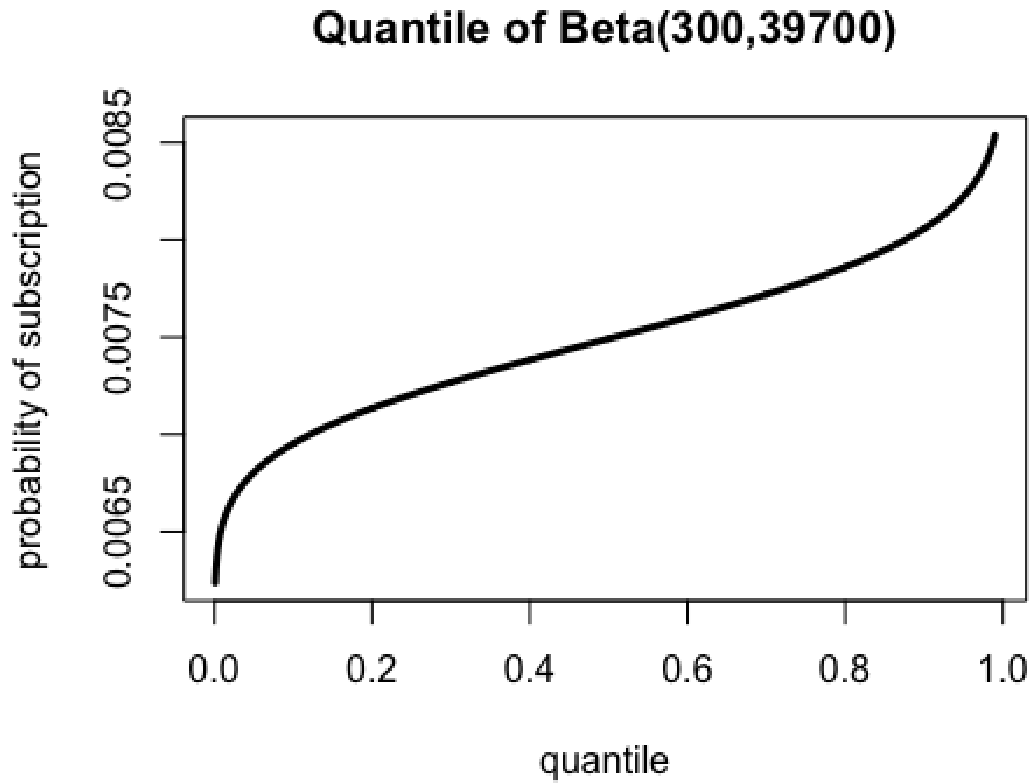
A1. Taking the code from the chapter, you just need to substitute `dbeta()` with `pbeta()` for the CDF like so:

```
xs <- seq(0.005,0.01,by=0.00001)
plot(xs,pbeta(xs,300,40000-300),type='l',lwd=3,
     ylab="cumulative probability",
     xlab="probability of subscription",
     main="CDF Beta(300,39700)")
```



And for quantile we need to change the *x*s to the actual quantiles:

```
xs <- seq(0.001,0.99,by=0.001)
plot(xs,qbeta(xs,300,40000-300),type='l',lwd=3,
     ylab="probability of subscription",
     xlab="quantile",
     main="Quantile of Beta(300,39700)")
```



Q2. Returning to the task of measuring snowfall from Chapter 10, say you have the following measurements (in inches) of snowfall:

7.8, 9.4, 10.0, 7.9, 9.4, 7.0, 7.0, 7.1, 8.9, 7.4

What is your 99.9 percent confidence interval for the true value of snowfall?

A2. We'll calculate the mean and standard deviation for this data first:

```
snow.data <- c(7.8, 9.4, 10.0, 7.9, 9.4, 7.0, 7.0, 7.1, 8.9, 7.4)
snow.mean <- mean(snow.data)
snow.sd <- sd(snow.data)
```

Then we use `qnorm()` to calculate the 99.9 percent confidence interval upper and lower bounds.

lower is `qnorm(0.0005,mean=snow.mean,sd=snow.sd)` = 4.46

upper is `qnorm(0.9995,mean=snow.mean,sd=snow.sd)` = 11.92

This means that we're very confident that there's no less than 4.46 inches of snowfall and no more than 11.92.

Q3. A child is going door to door selling candy bars. So far she has visited 30 houses and sold 10 candy bars. She will visit 40 more houses today. What is the 95 percent confidence interval for how many candy bars she will sell the rest of the day?

A3. First we have to calculate the 95 percent confidence interval for the probability of selling a candy bar. We can model this as Beta(10,20) and then use `qbeta()` to figure out these values:

lower is `qbeta(0.025,10,20)` = 0.18

upper is `qbeta(0.975,10,20)` = 0.51

Given there is 40 houses left, we can expect she'll sell between 40 × 0.18 = 7.2 and 40 × 0.51 = 20.4 candy bars. Of course, she can sell only whole bars, so we'll say we're pretty confident she'll sell between 7 and 20 candy bars.

If you really want to be particular, we could actually calculate the quantile for the binomial distribution at each extreme of her selling rates using `qbinom()`! I'll leave *that* as an exercise for you to explore on your own.

Chapter 14

Q1. Suppose you're playing air hockey with some friends and flip a coin to see who starts with the puck. After playing 12 times, you realize that the friend who brings the coin almost always seems to go first: 9 out of 12 times. Some of your other friends start to get suspicious. Define prior probability distributions for the following beliefs:

One person who weakly believes that the friend is cheating and the true rate of coming up heads is closer to 70 percent.

One person who very strongly trusts that the coin is fair and provided a 50 percent chance of coming up heads.

One person who strongly believes the coin is biased to come up heads 70 percent of the time.

A1. Picking these priors is a bit subjective, but here are some examples that correspond to each of the beliefs:

Beta(7,3) is a reasonably weak prior representing the belief that the rate is closer to 70 percent.

Beta(1000,1000) is a very strong belief that the coin is fair.

Beta(70,30) is a much stronger belief that the coin is biased to 70 percent heads.

Q2. To test the coin, you flip it 20 more times and get 9 heads and 11 tails. Using the priors you calculated in the previous question, what are the updated posterior beliefs in the true rate of flipping a heads in terms of the 95 percent confidence interval?

A2. Now we have an updated data set with a total of 32 observations, which includes 18 heads and 14 tails. Using R's `qbeta()` and the priors from the preceding questions, we can come up with the 95% confidence intervals for these different beliefs:

We'll just show the code for Beta(7,3) since the other examples are identical.

The lower bound for the 95 percent interval is `qbeta(0.025,18+7,14+3)` = 0.445 and the upper bound is `qbeta(0.975,18+7,14+3)` = 0.737.

For Beta(1000,1000) we have: 0.479 – 0.523.

And for Beta(70,30) we have: 0.5843 – 0.744.

So, as you can see, the weak prior provides the widest range of possibility, the very strong fair prior remains quite certain that the coin is fair, and the strong 70 percent prior still leans toward a higher range of possible values for the true rate of the coin.

Part IV, Hypothesis Testing: The Heart of Statistics

Chapter 15

Q1. Suppose a director of marketing with many years of experience tells you he believes very strongly that the variant without images (B) won't perform any differently than the original variant. How could you account for this in our model? Implement this change and see how your final conclusions change as well.

A1. You can account for this by increasing the strength of the prior. For example:

```
prior.alpha <- 300
prior.beta <- 700
```

This will require much more evidence to change our beliefs. To see how this changes our conclusions, we can rerun our code:

```
a.samples <- rbeta(n.trials,36+prior.alpha,114+prior.beta)
b.samples <- rbeta(n.trials,50+prior.alpha,100+prior.beta)
p.b_superior <- sum(b.samples > a.samples)/n.trials
```

And our new `p.b_superior` is 0.74, which is much lower than our original 0.96.

Q2. The lead designer sees your results and insists that there's no way that variant B should perform better with no images. She feels that you should assume the conversion rate for variant B is closer to 20 percent than 30 percent. Implement a solution for this and again review the results of our analysis.

Rather than using one prior to change our beliefs, we want to use two—one that reflects the original prior we had for A and one that reflects the lead designer's belief in B. Rather than use the weak prior, we'll use a slightly stronger one:

```
a.prior.alpha <- 30
a.prior.beta <- 70

b.prior.alpha <- 20
b.prior.beta <- 80
```

And when we run this simulation, we need to use two separate priors:

```
a.samples <- rbeta(n.trials,36+a.prior.alpha,114+a.prior.beta)
b.samples <- rbeta(n.trials,50+b.prior.alpha,100+b.prior.beta)
p.b_superior <- sum(b.samples > a.samples)/n.trials
```

The `p.b_superior` this time is 0.66, which is lower than before, but still slightly suggests that B might be the superior variant.

Q3. Assume that being 95 percent certain means that you're more or less "convinced" of a hypothesis. Also assume that there's no longer any limit to the number of emails you can send in your test. If the true conversion for A is 0.25 and for B is 0.3, explore how many samples it would take to convince the director of marketing that B was in fact superior. Explore the same for the lead designer.

A3. Here's the basic code to figure out this problem for the case of the director of marketing (for the lead designer, you'll need to add the separate priors). You can use a `while` loop in R to iterate through the examples (or just manually try new values).

```
a.true.rate <- 0.25
b.true.rate <- 0.3



prior.alpha <- 300
prior.beta <- 700

number.of.samples <- 0
#using this as an initial value so that the loop starts
p.b_superior <- -1
while(p.b_superior < 0.95){
  number.of.samples <- number.of.samples + 100
  a.results <- runif(number.of.samples/2) <= a.true.rate
  b.results <- runif(number.of.samples/2) <= b.true.rate
  a.samples <- rbeta(n.trials,
                     sum(a.results==TRUE)+prior.alpha,
                     sum(a.results==FALSE)+prior.beta)
  b.samples <-  rbeta(n.trials,
                     sum(b.results==TRUE)+prior.alpha,
                     sum(b.results==FALSE)+prior.beta)
  p.b_superior <- sum(b.samples > a.samples)/n.trials
}
```

Note that because this code itself is a simulation, you'll get different results each time you run it, so run it a few times (or build a more complicated example that runs itself a few more times!).

For the director of marketing it should take about 1,200 samples to be convinced. The lead designer should take about 1,000 samples. Notice that even though the lead designer believes that B is worse, she also has weaker priors in our example, so it takes less evidence to change her mind.

Chapter 16

Q1. Returning to the dice problem, assume that your friend made a mistake and suddenly realized that there were, in fact, two loaded dice and only one fair die. How does this

change the prior, and therefore the posterior odds, for our problem? Are you more willing to believe that the die being rolled is the loaded die?

A1. The original prior odds were $\frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$, and the Bayes factor was 3.77, giving us posterior odds of 1.89. Our new prior odds are $\frac{\frac{2}{3}}{\frac{1}{3}} = 2$, so our posterior odds are 2 × 3.77 = 7.54. We're certainly more willing now to believe that the die being rolled is loaded, but our posterior odds are still not very strong either way. We'd want to collect more evidence before completely giving up.

Q2. Returning to the rare diseases example, suppose you go to the doctor, and after having your ears cleaned you notice that your symptoms persist. Even worse, you have a new symptom: vertigo. The doctor proposes another possible explanation, labyrinthitis, which is a viral infection of the inner ear in which 98 percent of cases involve vertigo. However, hearing loss and tinnitus are less common in this disease; hearing loss occurs only 30 percent of the time, and tinnitus occurs only 28 percent of the time. Vertigo is also a possible symptom of vestibular schwannoma, but occurs in only 49 percent of cases. In the general population, 35 people per million contract labyrinthitis annually. What is the posterior odds when you compare the hypothesis that you have labyrinthitis against the hypothesis that you have vestibular schwannoma?

A2. We'll mix things up a bit and make H_1 "has labryinthitis" and H_2 "has vestibular schwannoma," since we already saw how unlikely vestibular schwannoma is. We need to recalculate every piece of our posterior odds because we're looking at a new piece of data, "has vertigo," and an entirely new hypothesis as well.

Let's start with the Bayes factor. For H_1 we have:

$$P(D \mid H_1) = 0.98 \times 0.30 \times 0.28 = 0.082$$

And the new likelihood for H_2 is:

$$P(D \mid H_2) = 0.63 \times 0.55 \times 0.49 = 0.170$$

So the Bayes factor for the new hypothesis is:

$$\frac{P(D \mid H_1)}{P(D \mid H_2)} = 0.48$$

This means that given the Bayes factor alone, vestibular schwannoma is a roughly two times better explanation than labyrinthitis. Now we have to look at the odds ratio:

$$O(H_1) = \frac{P(H_1)}{P(H_2)} = \frac{\frac{35}{1{,}000{,}000}}{\frac{11}{1{,}000{,}000}} = 3.18$$

Labyrinthitis is much less common than impacted earwax, and only about three times more common than vestibular schwannoma. When we put posterior odds together, we can see:

$$O(H_1) \cdot \frac{P(D|H_1)}{P(D|H_2)} = 3.18 \cdot 0.48 = 1.53$$

The end result is that labyrinthititis is only a slightly better explanation than vestibular schwannoma.

Chapter 17

Q1. Every time you and your friend get together to watch movies, you flip a coin to determine who gets to choose the movie. Your friend always picks heads, and every Friday for 10 weeks, the coin lands on heads. You develop a hypothesis that the coin has two heads sides, rather than both a heads side and a tails side. Set up a Bayes factor for the hypothesis that the coin is a trick coin over the hypothesis that the coin is fair. What does this ratio alone suggest about whether or not your friend is cheating you?

A1. Let's say H_1 is the hypothesis that the coin is in fact a trick coin, and H_2 is the hypothesis that it is fair. If the coin is indeed a trick coin, the probability of getting 10 heads in a row is 1, so we know that:

$$P(D \mid H_1) = 1$$

And if the coin is fair, then the probability of observing 10 heads is $0.5^{10} = \frac{1}{1024}$. So we know that:

$$P(D \mid H_2) = \frac{1}{1024}$$

The Bayes factor for this tells us that:

$$\frac{P(D \mid H_1)}{P(D \mid H_2)} = \frac{1}{\frac{1}{1024}} = 1024$$

This means that, given the Bayes factor alone, it is 1,024 times more likely that the coin is a trick coin.

Q2. Now imagine three cases: that your friend is a bit of a prankster, that your friend is honest most of the time but can occasionally be sneaky, and that your friend is very trustworthy. In each case, estimate some prior odds ratios for your hypothesis and compute the posterior odds.

A1. This is a bit subjective, but let's make some estimates. We need to come up with three different prior odds ratios. For each case we just multiply the prior odds by the Bayes factor from the previous question to get our posterior.

Being a prankster means our friend is more likely than not to trick us, so we'll set $O(H_1) = 10$. Then our posterior odds becomes 10 × 1,024 = 10,240.

If your friend is mostly honest but can be sneaky you wouldn't be that surprised if he was tricking you, but don't expect it, so we'll make the prior odds $O(H_1) = ¼$, which means that our posterior odds become 240.

If you really trust your friend, you might want to put the prior odds very low for cheating. Prior odds here might be $O(H_1) = \frac{1}{10{,}000}$, which gives you a posterior odds of roughly $\frac{1}{10}$, meaning you still think it's 10 times more likely that the coin is fair than that your friend is cheating.

Q3. Suppose you trust this friend deeply. Make the prior odds of them cheating 1/10,000. How many times would the coin have to land on heads before you feel unsure about their innocence—say, a posterior odds of 1?

A3. At 14 coin tosses the Bayes factor would be $\frac{1}{\frac{1}{0.5^{14}}} = 16{,}384$. Your posterior odds would be $\frac{16{,}384}{10{,}000} = 1.64$. At this point, you start to feel unsure about your friend's innocence. But with fewer than 14 coin tosses, you might still favor the idea that the coin is fair.

Q4. Another friend of yours also hangs out with this same friend and, after only four weeks of the coin landing on heads, feels certain you're both being cheated. This confidence

implies a posterior odds of about 100. What value would you assign to this other friend's prior belief that the first friend is a cheater?

A4. We can solve this by filling in the blanks. We know that $P(D \mid H_2) = 0.5^4 = \frac{1}{16}$, meaning that our Bayes factor would be 16. We just need to find a value to multiply by 16 that equals 100.

$$100 = O(H_1) \times 16$$

$$O(H_1) = \frac{100}{16} = 6\frac{1}{4}$$

And now we've assigned an exact value to the prior odds in your suspicious friend's mind!

Chapter 18

Q1. When two hypotheses explain the data equally well, one way to change our minds is to see if we can attack the prior probability. What are some factors that might increase your prior belief in your friend's psychic powers?

A1. Since we're talking about prior beliefs, the answers to this are likely to be a little bit different for everyone. For me, merely predicting the outcome of the roll of a die seems particularly easy to fake. I'd like to see this friend demonstrate psychic powers in an experiment of my choosing—for example, asking the friend to predict the last digit on the serial number of the dollar bills in my wallet—so that it would be much more difficult to trick me.

Q2. An experiment claims that when people hear the word *Florida*, they think of the elderly and this has an impact on their walking speed. To test this, we have two groups of 15 students walk across a room; one group hears the word *Florida* and one does not. Assume H_1 = the groups don't move at different speeds, and H_2 = the Florida group is slower because of hearing the word *Florida*. Also assume:

$$\boldsymbol{BF} = (\boldsymbol{P}(\boldsymbol{D}|\boldsymbol{H2}))/(\boldsymbol{P}(\boldsymbol{D}|\boldsymbol{H1}))$$

SET AS EQUATION AS IN CHAPTER 18

The experiment shows that H_2 has a Bayes factor of 19. Suppose someone is unconvinced by this experiment because H_2 had a lower prior odds. What prior odds would explain

someone being unconvinced and what would the BF need to be to bring the posterior odds to 50 for this unconvinced person?

A2. This question comes from an actual paper, "Automaticity of Social Behavior."[1] If the experiment seems questionable, you're not alone. The results of the study have been notoriously difficult to reproduce. If you were unconvinced, we'll say that means prior odds must be about $\frac{1}{19}$ to negate the results. In order to have a posterior odds of 50, you would need:

$$50 = \frac{1}{19} \times 950$$

So you'd need a Bayes factor of 950 to get your posterior odds into the "strong belief" range, given your initial skepticism.

Now suppose the prior odds do not change the skeptic's mind. Think of an alternate H_3 that explains the observation that the Florida group is slower. Remember if H_2 and H_3 both explain the data equally well, only prior odds in favor of H_3 would lead someone to claim H_3 is true over H_2, so we need to rethink the experiment so that these odds are decreased. Come up with an experiment that could change the prior odds in H_3 over H_2.

A3. It is entirely possible that the second group was on average slower. With only 15 participants, it's not hard to imagine that the group hearing the word *Florida* just happened to include a higher number of shorter people who might walk a short distance in a longer time. To be convinced I would need to, at minimum, see this experiment reproduced many times with many different groups of people to ensure that it wasn't just chance that led the group hearing the word *Florida* to be slower.

Chapter 19

Q1. Our Bayes factor assumed that we were looking at H_1: P(prize) = 0.5. This allowed us to derive a version of the beta distribution with an alpha of 1 and a beta of 1. Would it matter if we chose a different probability for H_1 ? Assume H_1 is P(prize) = 0.24 , then see if

---

[1] John A. Bargh, Mark Chen, and Lara Burrows, "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action," *Journal of Personality and Social Psychology* 71, no. 2 (1996).

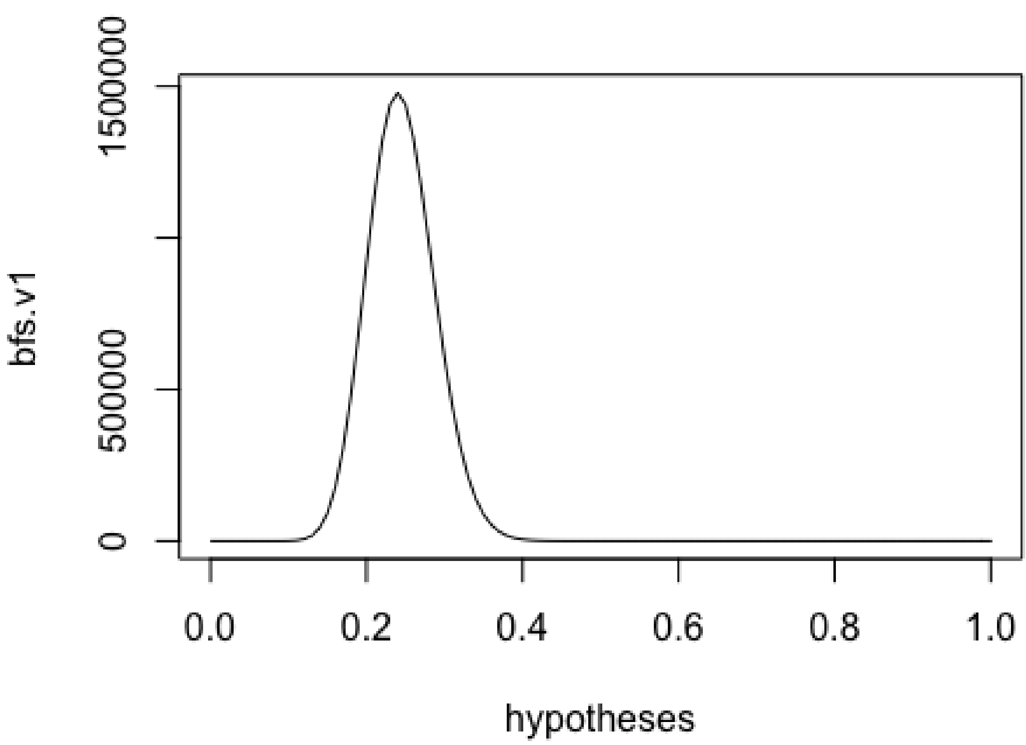
the resulting distribution, once normalized to sum to 1, is any different than the original hypothesis.

A1. We can rerun all of our code but this time make one group of `bfs` for the 0.5 version, and another for the 0.24 version:

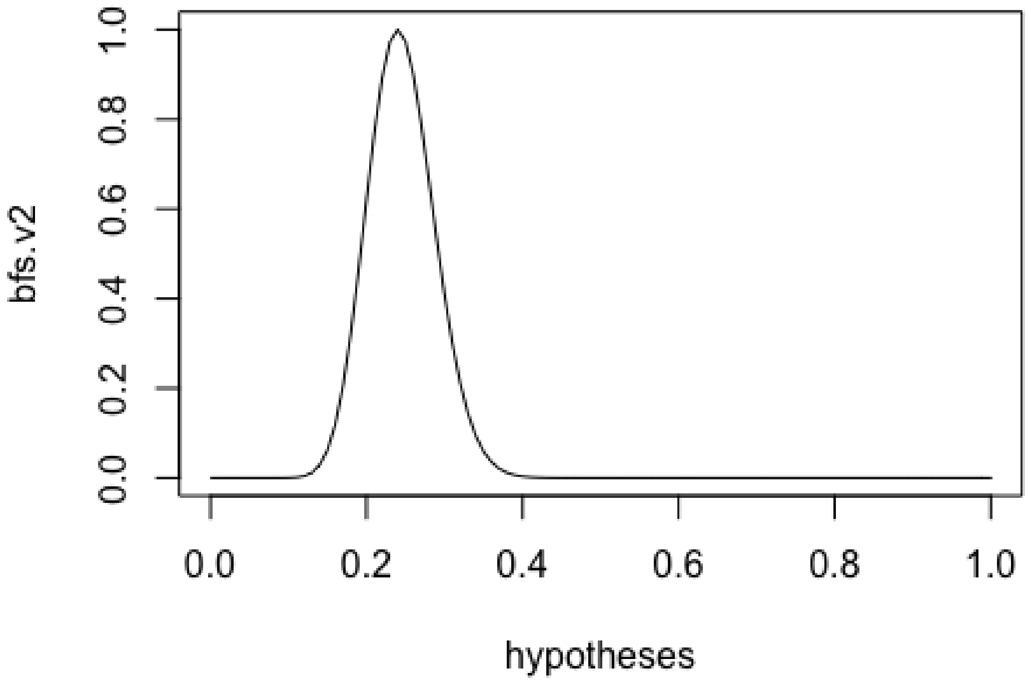
```
dx <- 0.01
hypotheses <- seq(0,1,by=0.01)
bayes.factor <- function(h_top,h_bottom){
  ((h_top)^24*(1-h_top)^76)/((h_bottom)^24*(1-h_bottom)^76)
}
bfs.v1 <- bayes.factor(hypotheses,0.5)
bfs.v2 <- bayes.factor(hypotheses,0.24)
```

Next we'll plot these each out separately:

```
plot(hypotheses,bfs.v1,type='l')
```

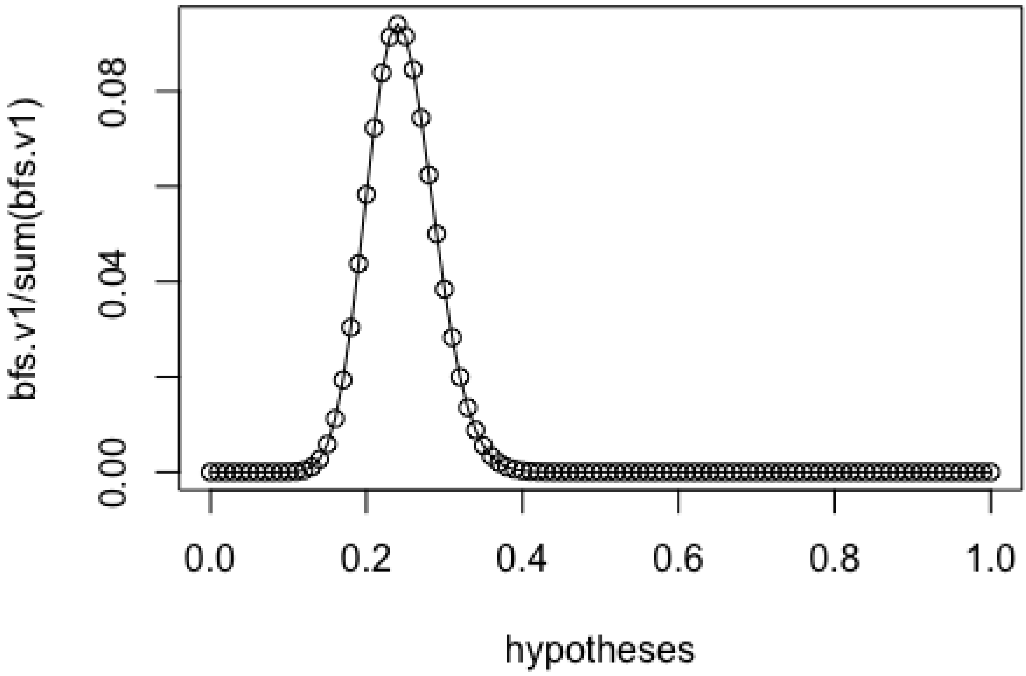



```
plot(hypotheses,bfs.v2,type='l')
```



Here we see the only difference is the y-axis. Choosing a weaker or stronger hypothesis changes only the scale of the distribution, not the shape of it. If we normalize and plot these two together, we see they are identical:

```
plot(hypotheses,bfs.v1/sum(bfs.v1),type='l')
points(hypotheses,bfs.v2/sum(bfs.v2))
```



Q2.Write a prior for the distribution in which each hypothesis is 1.05 times more likely than the previous hypothesis (assume our `dx` remains the same).

A1. Let's re-create our `bfs` from the original (see the code in the previous answer for the first part of this):

```
bfs <- bayes.factor(hypotheses,0.5)
```

Next our new priors are going to start with 1 (since there is no previous hypothesis), then `1.05, 1.05*1.05, 1.05*1.05*1.05`, and so on. There's a few ways to do this, but we'll just start with a vector of 1.05s one less than the length our hypotheses (since the first one is 1), using R's `replicate()` function:

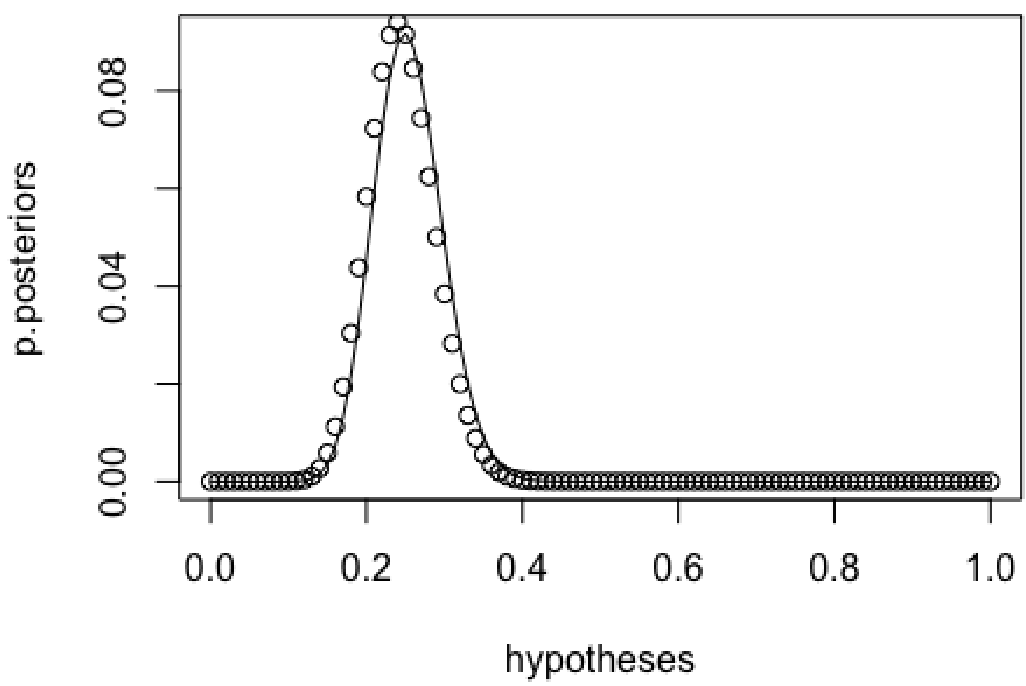
```
vals <- replicate(length(hypotheses)-1,1.05)
```

Then we add 1 to this list, and we can use the `cumprod()` function (which is just like `cumsum()` but for multiplying) to create our priors:

```
vals <- c(1,vals)
priors <- cumprod(vals)
```

Finally, we just compute our posteriors and normalize them, and then we can visualize our new distribution:

```
posteriors <- bfs*priors
p.posteriors <- posteriors/sum(posteriors)
plot(hypotheses,p.posteriors,type='l')
#add the bfs alone for comparision
points(hypotheses,bfs/sum(bfs))
```



Note that this doesn't change our final distribution all that much. Even though it gives much stronger prior odds to the last hypothesis—around 125 times more likely—the Bayes factor is so low that it doesn't make much of a difference in the end.

Q3. Suppose you observed another duck game that included 34 ducks with prizes and 66 ducks without prizes. How would you set up a test to determine "What is the probability that you have a better chance of winning a prize in this game than in the game we used in our example?" Implementing this requires a bit more sophistication than the R used in this

book, but see if you can learn this on your own to kick off your adventures in more advanced Bayesian statistics!

A3. Clearly what we need to do to solve this problem is to set up an A/B test like in Chapter 15. We can easily come up with our two distributions for the "34 prizes, 66 no prizes" example just by repeating the process we used in this chapter. The tricky part is sampling from our posterior that we created ourselves. In the past, to sample from a known distribution, we used built-in functions like `rbeta()`, but we have no equivalent function for this case. To solve this problem you'll need to use an advanced sampling technique like rejection sampling or even Metropolis–Hastings. If you're eager to solve this problem, it's a good time to start looking into a more advanced book on Bayesian analysis. But you should be proud, as it means you have a solid understanding of the basics!