Take Home Quiz 1

Solutions

Due Tuesday Sep. 17 at 11:00am

There is no time limit, but this quiz should take you about 30-40 minutes. Place your answers into this markdown document, knit it, and hand in the result as a PDF or Word document. You may use R, any reference material, and information already available on the internet. Do not work together, do not use AI, and do not get help from anyone but Dr. Clair.

Problem 1

Rabies is a virus that is almost always fatal to humans. People get rabies when bitten by a wild animal with the disease. About 4% of all bats carry rabies.

Bat bites are not very common in the U.S., but when they do occur 10% of those bats have rabies.

Let B be the event "the bat bit a person", and R be the event "the bat has rabies".

a. What is $P(R)$?
b. What is $P(R|B)$?
c. Dr. Speegle may have been bitten by a bat, but there's no way to tell. The bat was never tested for rabies. Suppose Dr. Speegle believes there was a 20% chance he was bitten. What is $P(B|R)$? (This is the probability he was bitten, assuming the bat had rabies).

Solution $P(R) = 0.04$, $P(R|B) = 0.1$. The prior is $P(B) = 0.2$. Using Bayes' Rule,

$$P(B|R) = \frac{P(B)P(R|B)}{P(R)} = \frac{0.2 \cdot 0.1}{0.04} = 0.5$$

If the bat actually had rabies, Dr. Speegle was a lot more likely to be bitten than he thought. He'd better get the rabies shots.

Problem 2

Klaus watches a lot of movies, and believes that most Americans keep their car keys in the car, tucked above the sun visor. If the probability of finding keys in the sun visor is π, Klaus thinks the value of π is probably around 60%, but could range anywhere from 25%-90%.

a. Choose an appropriate Beta distribution to model Klaus' prior.
b. Klaus gets a job at a parking lot and observes that only 2 of the 300 cars he's seen have their keys in the sun visor. What is his posterior distribution for π?
c. What is the mean for Klaus' posterior on π?
d. Did the prior or the data have more influence on the posterior?

Solution

Beta(7,5) has a mode of 60% and about the right spread for the prior, though other answers are also fine. Posterior is Beta(9,303), which has mean $9/312 \approx 0.029$. The data had more influence, Klaus is now pretty sure the keys in the visor thing only happens in movies.

Problem 3

There's a soda machine you use all the time, and it's never broken.

a. Explain why Beta(10,0.5) would be a reasonable prior for π, the probability the machine is working.

b. One day, you try the machine and it is broken. Money gone, no soda. What is the posterior probability distribution for π after this sad day? Is this an accurate reflection of how you would feel about the soda machine?

Solution a. A Beta(10,0.5) prior has a vertical asymptote at 1, indicating that you have a strong believe the machine will always work.

b. The posterior is Beta(10,1.5). The distribution is now zero at $pi = 1$, indicating you've lost your faith the machine will always work.

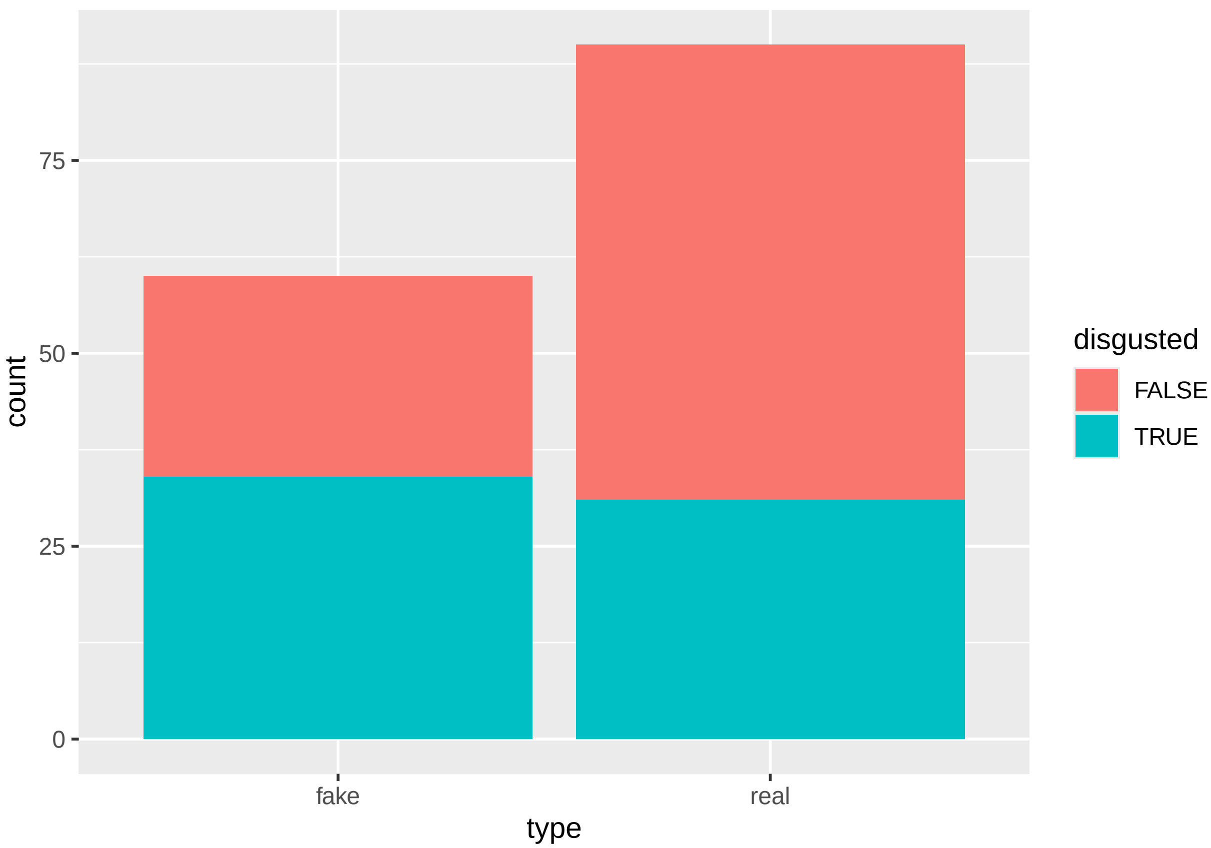
The last two problems use the `fake_news` data from the `bayesrules` library (also available on our course page at https://turtlegraphics.org/bayes/data/fake_news.csv).

Problem 4

Make a new variable `disgusted` which is TRUE when the `disgust` variable is bigger than 1.

Make a graphic to compare the difference in disgusted articles between real and fake news.

```
library(dplyr)
library(ggplot2)
fake_news <- read.csv("https://turtlegraphics.org/bayes/data/fake_news.csv")
fake_news <- fake_news |> mutate(disgusted = (disgust > 1))
fake_news |> ggplot(aes(x = type, fill = disgusted)) + geom_bar()
```



Problem 5

Use the `fake_news` data to estimate the prior probabilities that an article is real and fake. Suppose you observe an article with a disgust score larger than 1. What is the posterior probability that the article is fake news?

```
fake_news |> count(type) |> mutate(prior = n/nrow(fake_news))
```

```
##   type  n prior
## 1 fake 60   0.4
## 2 real 90   0.6
```

Compute the likelihood function $L(\text{type}|\text{disgusted})$ and multiply by the prior.

```
fake_news |> group_by(type) |>
  summarize(prior = n()/nrow(fake_news),
            likelihood = mean(disgusted)) |>
  mutate(posterior = prior*likelihood / sum(prior*likelihood))
```

```
## # A tibble: 2 x 4
##   type  prior likelihood posterior
##   <chr> <dbl>      <dbl>     <dbl>
## 1 fake    0.4      0.567     0.523
## 2 real    0.6      0.344     0.477
```

The probability of fake news given disgusted is 0.523.

Note you could compute this directly from the data:

```
fake_news |> filter(disgusted) |> summarize(mean(type == "fake"))
```

```
##   mean(type == "fake")
## 1            0.5230769
```