Homework 1

Caitlin Nordheim

1/16/2023

Bayesian Statistical Modeling Winter 2023

Homework, Week 1

When is homework due? Homework is due on Mondays (Tuesday if Monday is a holiday). This assignment is due on Monday, 1/16/2023. Submit homework to gradescope.

Before submitting your work You should use this file as a template for completing the assignment. You should change the author to your name. You should use the features of Rmarkdown to write your text answers below and include R code and equations where necessary. You should turn in both the .Rmd file as well as the pdf file that you generate by using the `Knit` button.

Running the shiny app

This assignment is a walkthough of using the virutal machine to run a shiny app. Once you have your VM up and running you can navigate the filesystem and find the file app.R under "Distributed Class Material". This app performs the Bayesian calculation we discussed in class, namely what is the probability of having covid given cold-like symptoms.

Question 1: Baye's rule

The calculation uses Bayes rule and expresses relationships between conditional probabilities.

1A: What is the conditional probability that we are trying to calculate? Write this in words and symbols, as best you can.

We are trying to calculate the conditional probability of having covid given that you have symptoms (here, symptoms being a runny nose). In "normal terms" we are trying to calculate the probability that your runny nose is due to covid.

in symbols: Pr(covid | symptoms)

1B: I gave two versions of the calculation in class. The simpler version assumed that you could not have both a cold and covid, while the more complicated one assumed you could have both. Write out one of these as best you can.

The probability of covid given symptoms is equal to the probability of symptoms given covid (aka how common is it to have a runny nose if you have covid) times the probability of covid (aka the prevalence of covid in the population) divided by the probability of symptoms given covid times the probability of covid plus the probability of symptoms given the flu (aka how common is it to have a runny nose if you have the flu) times the probability of the flu (aka the prevalence of flu in the population).

In symbols: Pr(covid|symptoms) = Pr(symptoms|covid) * Pr(covid) / (Pr(symptoms|covid) * Pr(covid)) + (Pr(symptoms|flu) * Pr(flu))

Question 2: Bayesian updated covid odds

2A: Low covid prevalence

Set the slider to a low value for covid prevalence. Report the value you chose, and describe the graph and explain why you beleive it has the shape and values.

I chose a covid prevalence of 0.05 (aka 5%). The probability of covid given symptoms (the chances your runny nose is due to covid) is low while the prevalence of flu is high. However, where the prevalence of flu is low and comparable to the chosen prevalence chosen for covid (between 0 - 0.05 on the x axis), then the chance of your runny nose being due to covid is higher, up to the inevitable odds of your runny nose being due to covid 100% if there is no flu in the population. The overall shape of the graph is exponential decay, with the odds of your runny nose being covid dropping rapidly as the prevalence of the flu increases from 0-25%, and the odds of your runny nose being due to covid continuing to fall and asymptote to around 0.05 as the flu prevalence increases to 100%.

2B: High covid prevalence

Set the slider to a high value for covid prevalence. Report the value you chose, and describe the graph and explain why you beleive it has the shape and values. How is it different from the low value?

I set the covid prevalence to 0.99 (99%).The graph takes on a shape of essentially a horizontal line close to a probability of 1 that your runny nose is due to covid, because nearly everyone has covid, meaning, there's a really good chance your runny nose is due to covid. In other words, the numerator is large in the equation since the Pr(covid) is so large.