Lab Probably

Abhi Thanvi, Paul Holaway

June 29th, 2022

Contents

Lab Probably

Welcome

Just like learning a new spoken language, you will not learn the language without practice. Labs are an important part of this course. Collaboration on labs is **extremely encouraged**. If you find yourself stuck for more than a few minutes, ask a neighbor or course staff for help. When you are giving help to your neighbor, explain the **idea and approach** to the problem without sharing the answer itself so they can figure it out on their own. This will be better for them and for you. For them because it will stick more and they will have a better understanding of the concept. For you because if you can explain it to other students, that means you understand it better too.

The Idea of this Lab

The idea behind this lab is to get familiar with probability. Probability provides us the information about the likelihood that something will happen or not. For example, Meteorologists use weather patterns to predict the probability of rain. In epidemiology, probability theory is used to understand the relationship between exposures and the risk of health effects. This is considered to be an abstract topic by many statisticians as it is more of computation than conceptual, but we will try to include some foundation conceptual ideas here.

"According to CDC, the odds of being injured by a toilet is 1 in 10,000" - Curious Abhi

Problem 1: It All Adds Up

Before we start computing probabilities in R, it is a good practice to know how to do probability questions by ourselves. Having a good conceptual foundation of what probability is and why the calculations change based on the problem statement is super crucial. Once we grasp that, doing probability in R should be light work. In this case, think of R as an calculator, it can sometimes help us do messy calculations, but in the end, we should know what we are doing. Let's start!

Question 1: If you take out a single card from a regular pack of cards, what is probability that the card is either an ace or a spade?

Answer: $P(Ace \cup Spade) = P(Ace) + P(Spade) - P(Ace \cap Spade) = \frac{1}{13} + \frac{1}{4} - (\frac{1}{13} * \frac{1}{4}) = \frac{1}{13} + \frac{1}{4} - \frac{1}{52} = \frac{16}{52}$

Question 2: If you take out a single card from a regular pack of cards, what is probability that the card is both an ace and a spade?

Hint: Pay attention to what the question is asking, you do not need to do "formal calculation using multiplication rule".

Answer: $P(Ace(number) \cap Spade(suit)) = P(Ace) * P(Spade) = \frac{1}{13} * \frac{1}{4} = \frac{1}{52}$

Question 3: If you take out a single card from a regular pack of cards, what is probability that the card is both an Club and Hearts?

Hint: In other words what is the probability of getting a club of hearts?

Answer: $P(Club(number) \cap Hearts(suit)) = P(Club) * P(Hearts) = \frac{0}{13} * \frac{1}{4} = 0$
Note: $P(Club(number)) = 0$ because there is no "number" called Club, because it's a suit!

Question 4: Oh wow... You are a genius with probability! Let me throw some curve balls at you now! I have tossed a regular coin 5 times already, and I have gotten the sequence of $\{Head, Tail, Tail, Tail, Head\}$ for my 5 tosses? Based on the previous 5 tosses, what is the probability that my 6th toss will be a Tail? Why do you think your answer is correct?

Hint: Think about the dependencies of our outcome during coin tosses

Answer: The probability that the 6th toss is a tail is $\frac{1}{2}$ (just like before), because the coin tosses are independent each time; in other words, the probabilities of the previous tosses don't influence the probability of the upcoming ones.

Question 5: Well, this is my last shot to prove myself that I can come up with the hardest probability questions! If you get this one too, I think we will need to celebrate... probably. ;) See what I did there!

Anyways, I am playing monopoly with my friends and I am stuck in jail. Apparently, my friends are making it hard for me (probably because I own Boardwalk) and told me I can only get out of jail if I roll a sum of six with two die. They don't know I am like Dr. Strange when it comes to probability, except that I did not mess up the multiverse. But again, I am on a sabbatical when it comes to using my probability powers, kind of like Hulk. So you tell me, what is the probability of me rolling a sum of six?

Hint: This is a question where you might want to list down all the possibilities of getting a sum of six with two die

Answer: Possibilities: (1,5), (2,4), (3,3), (4,2), (5,1). Therefore, the answer is $\frac{5}{36}$

Problem 2: What "R" the odds?

Paul has challenged me for a Baseball match, so I need a coach! I am pretty good at pitching, but need to change my Tennis swing into a baseball swing while batting. I am choosing a Baseball Batter as my coach, and since I am from Texas I am choosing some batter from Texas team! I am not sure what the probability of me finding a Batting Coach from a Texas team would be from the list I have. Help me before Paul wins!

Question 1: Import `MLB_Batters` csv and figure out the probability of choosing a batting coach who is from Texas team.

Answer:

```
# Import the CSV.
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
MLB_Batters <- read.csv("~/Desktop/DPI 2022/MLB_Batters.csv", stringsAsFactors=TRUE)
# Filter Texas Team Players.
texas = MLB_Batters %>% filter(Team == "TEX")
# Probability of choosing a person from Texas Team from the list.
nrow(texas)/nrow(MLB_Batters)
```

```
## [1] 0.02741703
```

Question 2: Based on your probability calculation, do you think it is likely for me to find a coach from Texas team?

Answer: It's pretty unlikely, because 0.02 is a pretty small number.

This lab was short, because we wanted you to get familiar with probability conceptually and using data frames. Next lesson, we will be learning more rules and tricks! Next lab will be a bit more involved and will ask you to switch back and forth between different rules and tricks. So make sure, you feel comfortable with today's lesson and lecture. Feel free to ask questions with your group members, your instructors, or join the office hours! We love seeing people outside our assigned lectures/labs and encourage everyone to show up when they can!

Submission

Once you have finished your lab...

1. Go to the top left and click `File` and `Save`.
2. Click on the `Knit` button to convert this file to a PDF.
3. Submit **BOTH** the `.Rmd` file and `.pdf` file to Blackboard by 11:59 PM tonight.