Lab Probably II

Abhi Thanvi, Paul Holaway

June 30th, 2022

Contents

Lab Probably II

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. This is the sequel to the Lab Probably.

"In a room of just 23 people there is a 50% chance of any two sharing a birthday. In a room of 75 people this becomes 99.9%" - Curious Abhi

Problem 1: Rules Forever!

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 two cards from a regular pack of cards (without replacing), what is probability that you get an ace (first card) and then a king (second card)?

Hint: Are the events independent? **Answer:** (Student Response Here) $P(K_2 \cap A_1) = P(A_1)P(K_2|A_1) = \frac{4}{52} * \frac{4}{51} = \frac{16}{2652} \approx 0.0060$

Question 2: Billy Bob Joe (one person) observed what customers ordered at his ice cream shop and found the following probabilities:

1. P(Vanilla) = 0.3
2. P(Sundae) = 0.2
3. P(Vanilla and Sundae) = 0.15

Find the probability that a customer ordered vanilla ice cream given they ordered a sundae

Answer: (Student Response Here) $P(Vanilla|Sundae) = \frac{P(V \cap S)}{P(S)} = \frac{0.15}{0.2} = 0.75$

Question 3: Paul is interested in elections so he made the following two-way table categorizing the US senators in 2018 by their political party and whether or not it was their first term in the senate.

	Democrat	Republican	Independent	**Total**
First Term	11	28	1	40
Second Term	33	26	1	60
Total	44	54	2	100

Find the probability that the senator was in the Democratic party, given that the senator was returning to office.

Hint: There are two ways to do it! Intuitive and Calculation Method! Make sure to show the method you choose and it's appropriate work!

Answer: (Student Response Here) Intuitive -> $\frac{33}{60} = 0.55$
Calculate -> $P(Democratic|Returning) = \frac{P(D \cap Ret)}{P(Ret)} = \frac{33/100}{60/100} = \frac{33}{60} = 0.55$

Question 4: Let me throw some curve balls again 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 and 7th Toss will look like $\{Tail, Tail\}$? Why do you think your answer is correct?

Answer: (Student Response Here) $P(Tail_6 \cap Tail_7) = 0.25$ since each toss is independent.

Question 5: Well, this is my last shot (again!) 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... probability. ;) See what I did there (again)!

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 twelve two times in a row. 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 twelve two times in a row?

Answer: (Student Response Here) P(Rolling a Twelve) = P(Dice 1 = 6) and P(Dice 2 = 6) = P(Rolling a Six) and P(Rolling a Six) = 1/36 Rolling two twelves in a row would be (1/36)*(1/36) = 1/1296

Problem 2

Part 1: What "R" the odds again!?

Unfortunately I lost the last baseball game because my Coach was from Texas but not experienced enough to work with "people like me". Paul has challenged me for another Baseball match, so I need a new 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 my last coach didn't work out I am not choosing some batter from a Texas team!

Question 1: Import MLB_Batters csv and figure out the probability of choosing a batting coach who is not 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("~/Classes/DPISu22/Data Sets/MLB_Batters.csv", stringsAsFactors=TRUE)
# Filter Non Texas Team Players.
non_texans <- MLB_Batters %>% filter(Team != "TEX" & Team != "HOU")

# Probability of choosing a person from Non Texas Team from the list.
prob = length(non_texans$Team) / length(MLB_Batters$Team)
prob
```

```
## [1] 0.9422799
```

Question 2: Based on your probability calculation, do you think it is likely for me to find a coach from a non Texas team? Does this make sense? Justify your response with the response from your previous lab!

Hint: We are looking for something with sample space total equaling "something" **Answer:** (Student Response Here) Yes, the probability is 94.23% which is quite high. This makes sense because last time finding a Texan coach was 5.77% so P(Total) - P(Texan) = P(Non Texan) => 100% - 5.77% = 94.23%.

Question 3: Well, okay so I can definitely find a coach, but I want to make sure I win this time! I think finding a player who is above the age 30. Uhm, let's skip people who are 30... I am desperate for that win! Let's figure out the probability of picking a coach who is not from Texas **AND** is older than 30.

Hint: Use the "Non-Texan" data frame to shorten your code **Answer:**

```
old_non_texan <- non_texans %>% filter(Age > 30)
prob_2 = length(old_non_texan$Team) / length(MLB_Batters$Team)
prob_2
```

```
## [1] 0.2597403
```

Part 2: Reflection

Question 1: Of course, the probability of finding a Coach who is not a Texas team player and above 30 is quite low. Do you think this was an Addition Rule or Multiplication Rule?

Answer: (Student Response Here) Multiplication Rule

Question 2: Do you think that finding the probability of Non Texan player above the age of 30 be easier to do by hand or using `R` (given the table or CSV)? Why do you think so?

Answer: (Student Response Here) Most likely using R is better, because it filters the people for us!

Big Brain Bonus

This question should be filled out, but it is a higher level of thinking question. It involves something that will help you become a better statistician, data analyst and/or data scientist!

Question: So I am in a middle of test, and Chegg is not working (I do not endorse it) so I need your help! The question is mysterious. . . it has mystery numbers! All I know the relation among the probabilities, but not the actual probabilities. I am stuck on what rule to use addition or multiplication. Is the Mystery asking for using addition or multiplication rule? Help me out and give me a reason why you think you are right. . . just so I can make sure I don't lose points if I use your answer!

–> Question on my Test :

P(A) = x

P(B) = y

P(Mystery) = z

z < x = y

```
The mystery is some event or condition relation to A and B.
```

Hint: Think about if the probabilities are higher or lower when using addition rule and when using multiplication rule Feel free to discuss among your group! This is a fun question and promotes exploring!

Answer: (Student Response Here) This is a multiplication rule problem because our P(Mystery) = z is lower than x and y, which means that it should be lower probability. x = y tells us that the probability of these two events are equal. z < x = y tells us that P(Mystery) is lower than both P(A) and P(B). Addition Rule allows for more possibilities which increases probability, and Multiplication Rule does the opposite. Therefore, the mystery is most likely a multiplication rule relation and that's what I should probably use. Student Responses will vary, but should have an answer and reasoning along this logical path.

Feedback

Hey this is Abhi! As this second week comes to an end, I would like to know (again) whether you are liking the course or you hate your summer because of us (hopefully not!).

Please give some feedback of what you like about the course and what you would like to change about this course! We will try our best to make this the best course and have the best time as much we can! Have a great weekend and Paul and I will see you on the other side :)

Feedback: (Student Feedback)

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.