

Homework 03: Probability

Name: Julia Troni

This assignment is due on Canvas by **6:00PM on Friday September 16**. Your solutions to theoretical questions should be done in Markdown directly below the associated question. Your solutions to computational questions should include any specified Python code and results as well as written commentary on your conclusions. Remember that you are encouraged to discuss the problems with your classmates, but **you must write all code and solutions on your own**.

NOTES:

- Any relevant data sets should be available in the Homework 01 assignment write-up on Canvas. To make life easier on the grader if they need to run your code, do not change the relative path names here. Instead, move the files around on your computer.
- If you're not familiar with typesetting math directly into Markdown then by all means, do your work on paper first and then typeset it later. Remember that there is a [reference guide](#) linked on Canvas on writing math in Markdown. **All** of your written commentary, justifications and mathematical work should be in Markdown.
- Because you can technically evaluate notebook cells in a non-linear order, it's a good idea to do **Kernel → Restart & Run All** as a check before submitting your solutions. That way if we need to run your code you will know that it will work as expected.
- It is **bad form** to make your reader interpret numerical output from your code. If a question asks you to compute some value from the data you should show your code output **AND write a summary of the results** in Markdown directly below your code.
- This probably goes without saying, but... For any question that asks you to calculate something, **you must show all work and justify your answers to receive credit**. Sparse or nonexistent work will receive sparse or nonexistent credit.

Import Pandas and NumPy anytime you are doing data analysis.

In [1]:

```
# Per the standard import pandas as 'pd' and numpy as 'np'
import pandas as pd
import numpy as np
```

You may or may not need this depending on how you create your own homework, but if you need to create a graph, you should load Matplotlib's Pylab library to set up Jupyter so that it will plot directly in the notebook.

In [2]:

```
import matplotlib.pyplot as plt
%matplotlib inline
# Recall 'inline' puts your graph in the cell versus a new popup window
```

Problem 1

Suppose you are asked to manage/interview the employees at a new Boulder start-up called **Programmers of Large Opulent Problems**, or PLOP.

PLOP employs 100 programmers. Forty-five of the programmers are proficient in Java, 30 in C#, 20 in Python, six in C# and Java, one in Java and Python, five in C# and Python, and just one programmer is proficient in all three languages.

(7 points) Suppose an employee is randomly chosen to interview with you concerning their coding proficiency and general productivity. What is the probability that the chosen employee is not proficient in any of the three languages mentioned?



Write up your solution to Problem 1 here:

Solution:

$$|\Omega| = 100$$

$$P(OnlyJava) = \frac{39}{100}$$

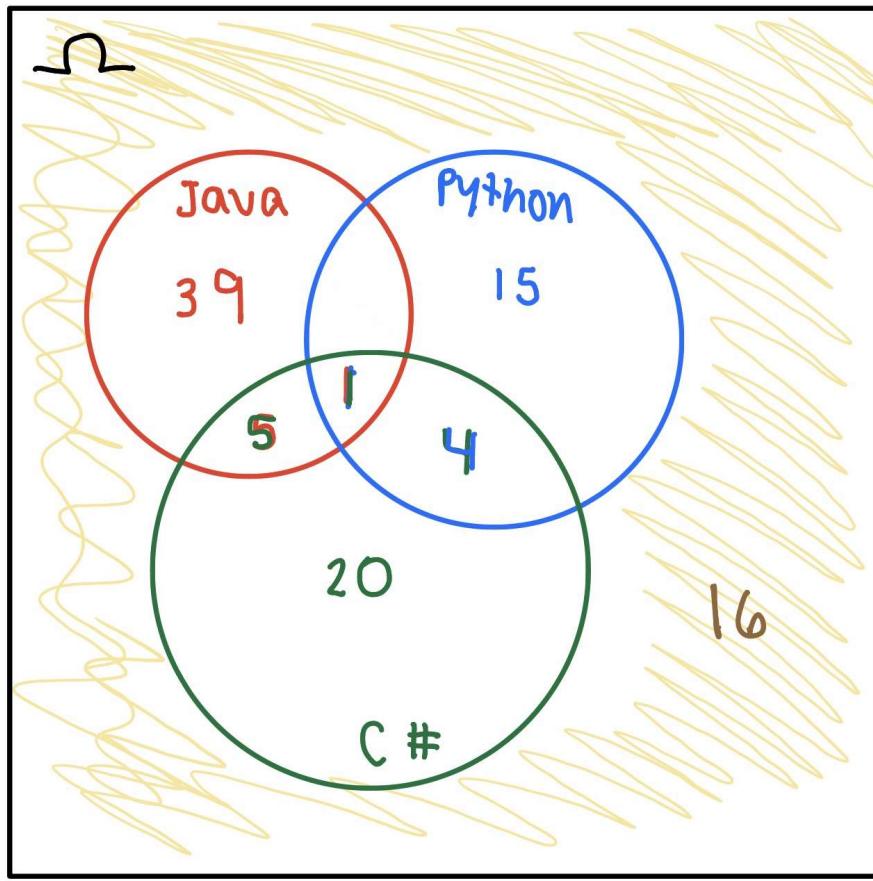
$$P(OnlyPython) = \frac{15}{100}$$

$$P(OnlyC\#) = \frac{20}{100}$$

$$P(Java^C \cap (Python)^C \cap (C\#)^C) = P(Java \cup Python \cup C\#)^C = 1 - P(Java \cup Python \cup C\#)$$

- See the visual below, with the yellow shading showing

$$(Java \cup Python \cup C\#)^C$$



Problem 2

For a certain period of time (months and months) you decide that sometimes wearing a colorful hat should be your latest fashion statement. Looking back over past time you note that there were some days you wore a hat and some days that you did not wear a hat.

'A' is the event that you wore a red hat. 'B' is the event that you wore a blue hat. 'C' is the event that it rained.

You do some data analysis on this past period of time.

You discover that the probability that you wore a red hat or that it rained is $\frac{2}{3}$.

The probability that you wore a blue hat or that it rained is $\frac{3}{4}$.

The probability that you wore a red hat, or that you wore a blue hat, or that it rained is $\frac{11}{12}$.

solution:

Event A and B are disjoint since you cannot wear a red hat and a blue hat on the same day so

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap C) - P(B \cap C) - P(A \cap B) + P(A \cap B \cap C)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap C) - P(B \cap C) - 0 + 0$$

$$P(B \cup C) = P(B) + P(C) - P(B \cap C)$$

$$P(A \cup C) = P(A) + P(C) - P(A \cap C)$$

Events A and C, and events B and C are independent so

$$P(B \cap C) = P(B) \cdot P(C)$$

and

$$P(A \cap C) = P(A) \cdot P(C)$$

So,

$$P(B \cup C) = P(B) + P(C) - P(B) \cdot P(C)$$

$$P(A \cup C) = P(A) + P(C) - P(A) \cdot P(C)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A) \cdot P(C) - P(B) \cdot P(C)$$

Now we have a system of 3 equations with knowns

$$P(B \cup C) = \frac{9}{12}$$

$$P(A \cup C) = \frac{8}{12}$$

$$P(A \cup B \cup C) = \frac{11}{12}$$

So

$$P(A \cup B \cup C) = P(A) + P(B \cup C) - P(A) \cdot P(C)$$

We can reduce to

$$\frac{11}{12} = P(A) + \frac{9}{12} - P(A) \cdot P(C)$$

$$\frac{2}{12} = P(A) - P(A) \cdot P(C)$$

$$P(A) \cdot P(C) = P(A) - \frac{2}{12}$$

Likewise

$$\frac{11}{12} = P(B) + \frac{8}{12} - P(B) \cdot P(C)$$

$$\frac{3}{12} = P(B) - P(B) \cdot P(C)$$

$$P(B) \cdot P(C) = P(B) - \frac{3}{12}$$

So plug that into

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A) + \frac{2}{12} - P(B) + \frac{3}{12}$$

After cancelations

$$P(A \cup B \cup C) = P(C) + \frac{2}{12} + \frac{3}{12}$$

Thus

$$P(C) = \frac{6}{12} = \frac{1}{2}$$

Then

$$\frac{2}{12} = P(A) - P(A) \cdot \frac{6}{12}$$

$$P(A) = \frac{2}{6} = \frac{1}{3}$$

And lastly

$$\frac{3}{12} = P(B) - P(B) \frac{6}{12}$$

$$P(B) = \frac{3}{6} = \frac{1}{2}$$

Part A

(4 points)

What is the probability that you wore a red hat?

solution:

$$P(A) = \frac{1}{3}$$

Part B

(4 points)

What is the probability that you wore a blue hat?

solution:

$$P(B) = \frac{1}{2}$$

Part C

(4 points)

What is the probability that it rained?

solution:

$$P(C) = \frac{1}{2}$$

Problem 3

It is free donut day on campus! There are three locations offering each participant one free donut; either a glazed donut or a cake donut.

The C4C is offering donuts, 75% of their donuts are glazed and 25% are cake.

The Alfred Packer Grill is offering donuts. 60% of their donuts are glazed and 40% are cake.

The SEEC Cafe (East campus) is offering donuts, 45% are glazed and 55% are cake.

Part A

(4 points)

If you randomly choose a donut location (each location is equally likely) and then your chosen location randomly grabs a donut and hands it to you, then what is the probability that your newly acquired donut is glazed?

The picture below is merely here to help visualize this as a law of total probability type problem.

You need not use the pictured variables.



solution:

$$P(\text{Glazed}) = \frac{P(\text{GlazedDonut})}{P(\text{TotalDonuts})}$$

$$P(\text{Glazed}) = \frac{P(\text{GlazedC4C}) + P(\text{GlazedAlfredPacker}) + P(\text{GlazedSEEC})}{P(\text{TotalDonuts})} = \frac{75 + 60 + 45}{300}$$



Part B

(5 points)

For the same situation mentioned above, create some code that will run a simulation to estimate the probability of getting a glazed donut. (Use $n=1000$)

Run the code and verify that it agrees with the by-hand computation you arrived at above.

```
In [3]: C4C = {'donuts' : np.array(["glazed", "cake"]), 'probs' : np.array([75/100, 25/100])}
AlfredPacker = {'donuts' : np.array(["glazed", "cake"]), 'probs' : np.array([60/100, 40/100])}
SEEC = {'donuts' : np.array(["glazed", "cake"]), 'probs' : np.array([45/100, 55/100])}

loc_choices = {'locations' : np.array([C4C, AlfredPacker, SEEC]), 'probs' : np.array([1, 1, 1])}

def sample_donut(loc_choices):
    # randomly choose a donut location
    location = np.random.choice(loc_choices['locations'], p = loc_choices['probs'])
    # randomly choose a donut from that location
    return np.random.choice(location['donuts'], p = location['probs'])

def probability_of_donut(variety, loc_choices, num_samples=1000):
    # get a bunch of donuts
    donuts = np.array([sample_donut(loc_choices) for ii in range(num_samples)])
    # compute fraction of donuts of desired type
    return np.sum(donuts == variety) / num_samples
```

```
In [4]: probability_of_donut("glazed", loc_choices, num_samples=50000)
```

```
Out[4]: 0.6012
```

The code and the by-hand computation agree. Both show the probability that your newly acquired donut is glazed is 60%

Problem 4

Suppose you decide to join a game club to meet other folks on campus. You become fascinated with the tetrahedron, or 4-sided die.

This particular die has its sides marked with '1', '2', '3', and '4'.

You roll the tetrahedron *three* times:



Part A

(3 points)

What is the probability that you roll a '4' three times in a row?

Put your answer to number 4, part A here:

solution:

These are independent events, thus

$$P(4, 4, 4) = \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{64}$$

Part B

(3 points)

What is the probability that you roll only a single '4' in the three rolls?

Put your answer to number 4, part B, here:

solution:

Let event A be denoted by the sample space

$$\Omega = \{\{YES4, Not4, Not4\}, \{Not4, YES4, Not4\}, \{Not4, Not4, YES4\}\}$$

$$P(A) = \binom{3}{1} \cdot (1 - P(4))^3 \cdot P(4)$$

$$P(A) = 3 \cdot \left(\frac{3}{4}\right)^3 \cdot \frac{1}{4} = \frac{81}{256} = 0.31640625$$

Part C

(4 points)

Given that you have observed at least one '4', what is the probability that you observe at least two 4's?

solution: Translation ==> If you have already rolled the three dice but you can't see them, and someone else tells you whether there is at least one that has a 4. Then given that they told you at least one of your rolls was a 4, what is the chance that out of the remaining 2 dies at least one of them is a 4?

let event C represent "at least one 4". Let event D represent "you observe at least 2 4's"

$$P(D|C) = \frac{P(C \cap D)}{P(C)}$$

$$P(D|C) = \frac{P(D)}{P(C)}$$

$$P(C) = \frac{1}{4} + \left(\frac{1}{4} \cdot \frac{1}{4}\right) + \left(\frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4}\right)$$

$$P(D) = \left(\frac{1}{4} \cdot \frac{1}{4}\right) + \left(\frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4}\right)$$

$$P(D|C) = 0.2380$$

Problem 5

In an attempt to avoid COVID, a group of 100 students separate themselves from CU society and choose to live in connected housing apart from all other people.

Unfortunately, and unbeknown to anyone, one of the students had COVID when the group went into isolation.

It is known that 90% of COVID tests will detect the virus (i.e. true positives) while 9.6% of the tests are false positives.

(7 points)

If, on move-in day, a randomly chosen person, within this group of 100, gets a positive test result, what is the probability that they actually have COVID?

solution: Let the events:

- V= virus is actually in the person
- T= test is positive (test says the person has the virus)

Then

$$P(\text{truepositive}) = P(T|V) = \frac{90}{100}$$

$$P(\text{falsepositive}) = P(T|V^C) = \frac{96}{1000}$$

Then since

$$P(T|V) + P(T^C|V) = 1$$

and

$$P(T|V^C) + P(T^C|V^C) = 1$$

$$P(\text{falsenegative}) = P(T^C|V) = \frac{10}{100}$$

$$P(\text{truenegative}) = P(T^C|V^C) = \frac{904}{1000}$$

We know that

$$P(V) = \frac{1}{100}$$

so

$$P(V^C) = \frac{99}{100}$$

We are being asked to find

$$P(V|T)$$

$$P(V|T) = \frac{P(V \cap T)}{P(T)} = \frac{P(T|V) \cdot P(V)}{P(T|V) \cdot P(V) + P(T|V^C) \cdot P(V^C)}$$

$$P(V|T) = \frac{\frac{90}{100} \cdot \frac{1}{100}}{\frac{90}{100} \cdot \frac{1}{100} + \frac{96}{1000} \cdot \frac{99}{100}}$$

$$P(V|T) = 0.0865051903114187$$

Therefore there is a 8.65% chance that the randomly chosen person with the positive test result actually has COVID.

```
In [5]:  
num=90/100/100  
denom= num+ (96/1000*99/100)  
num/denom*100
```

```
Out[5]: 8.65051903114187
```

Problem 6

Formatting (**5 points total**)

- a. Problems not linked to Gradescope (**2 points**)
- b. Poor comments, or no commenting (**1 point**)
- c. Markdown not rendered (**1 point**)
- d. Explanations are not formatted correctly, or are not printed (**1 point**)