# Homework 03: Probability

---

**Name**: CJ Kennedy

---

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 is 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.

```python
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.
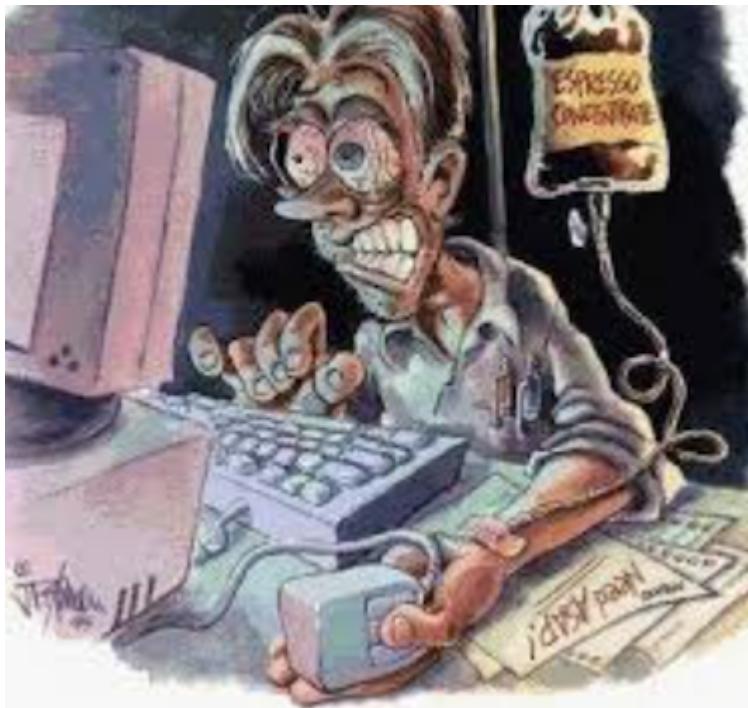
```python
In [2]:  import matplotlib.pylab as plt
         %matplotlib inline
             # Recall 'inline' puts your graph in the cell versus a new popup window
```

# 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.

**(8 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:**

First, assign variables to the information we have.

$N(a) = 45$ (number of people proficient in Java)

$N(b) = 30$ (number of people proficient in C#)

$N(c) = 20$ (number of people proficient in Python)

$N(d) = 6$ (number of people proficient in a and b)

$N(e) = 1$ (number of people proficient in a and c)

$N(f) = 5$ (number of people proficient in b and c)

$N(g) = 1$ (number of people proficient in a,b,c)

Now, total the number of people proficient in two languages:

$N_2 = N(d) + N(e) + N(f) = 6 + 1 + 5 = 12$

Now, the total number of people proficient in one language:

$N_1 = N(a) + N(b) + N(c) = 45 + 30 + 20 = 95$

So, the number of people not proficient in at least one language is:

$N = N_1 - N_2 + N(g) = 95 - 12 + 1 = 16$

Therefore, the probability that any chosen employee is not proficient in any of the languages is:

$P(N) = 16/100 = .16$

# 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 $\frac{11}{12}$.

## Part A

### *(4 points)*

What is the probability that you wore a red hat?

First, write the given probabilities below:

$$P_{red} \lor P_{rain} = \frac{2}{3}$$

$$P_{blue} \lor P_{rain} = \frac{3}{4}$$

$$P_{red} \lor P_{blue} \lor P_{rain} = \frac{11}{12}$$

Now, substitute the second equation into the third:

$$P_{red} \lor \left(\frac{3}{4}\right) = \frac{11}{12}$$

$$\Rightarrow P_{red} + \left(\frac{3}{4}\right) = \frac{11}{12}$$

$$\therefore P_{red} = \frac{11}{12} - \frac{3}{4} = \frac{2}{12}$$

So, the probability that a red hat was worn is $\frac{1}{6}$.

# Part B

*(4 points)*

What is the probability that you wore a blue hat?

Similiar to part a, substitute the first equation into the third:

$$P_{blue} \lor \left(\frac{2}{3}\right) = \frac{11}{12}$$

$$\Rightarrow P_{blue} + \left(\frac{2}{3}\right) = \frac{11}{12}$$

$$\therefore P_{blue} = \frac{11}{12} - \frac{2}{3} = \frac{4}{12}$$

So, the probability that a blue hat was worn is $\frac{1}{3}$.

# Part C

*(4 points)*

What is the probability that it rained?

The probability that it rained can be solved via the solutions to the previous problems and the third equation:

$$P_{red} \lor P_{blue} \lor P_{rain} = \frac{11}{12}$$

$$P_{red} + P_{blue} + P_{rain} = \frac{11}{12}$$

$$\frac{1}{6} + \frac{1}{3} + P_{rain} = \frac{11}{12}$$

$$\Rightarrow P_{rain} = \frac{11}{12} - \frac{1}{6} - \frac{1}{3}$$

So, the probability that it rained is $\frac{1}{2}$.

# 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.
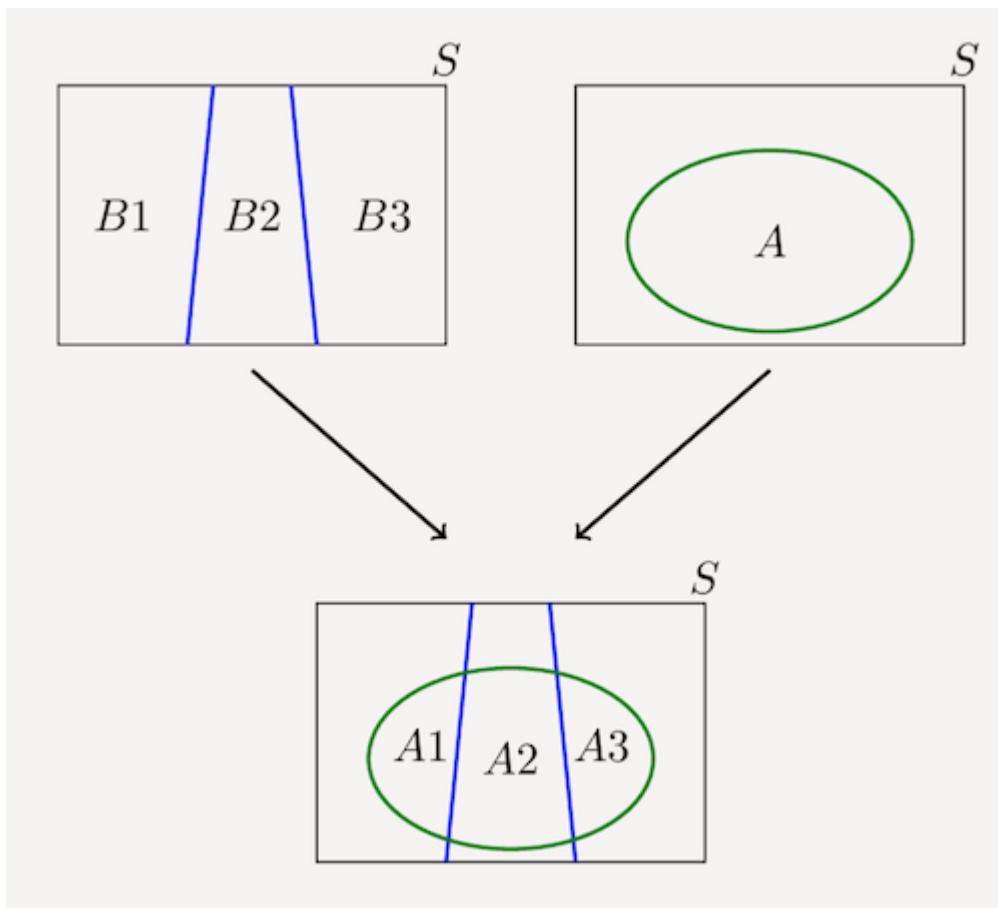
# Part A

*(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.

Note the glazed offering percentages for each location:

$$x_1 = C4C = .75$$

$$x_2 = APG = .60$$

$$x_3 = APG = .45$$

Since each location is just as likely to be choosen, the mean of the three probabilties represents the probability that given donut is glazed:

$$\bar{x} = \frac{x_1 + x_2 + x_3}{3} = \frac{.75 + .60 + .45}{3} = 0.6$$

Therefore, the probability the donut is glazed is $\frac{6}{10}$.

# Part B

**(6 points)**

For the same situation mentioned above, create some code that will run a simulation to estimate the probability of getting a glazed donut.

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

```
In [3]:   # adapted from nb05
          c4c = {'donuts' : np.array(["glazed", "cake"]), 'probs' : np.array([3/4, 1/4])}
          apg = {'donuts' : np.array(["glazed", "cake"]), 'probs' : np.array([6/10, 4/10])}
          seec = {'donuts' : np.array(["glazed", "cake"]), 'probs' : np.array([.45, .55])}
          location_choices = {'locations' : np.array([c4c, apg, seec]), 'probs' : np.array([1/3,

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

          def probability_of_donutType(donutType, location_choices, num_samples=1000):
              # get a bunch of locations
              locs = np.array([sample_location(location_choices) for ii in range(num_samples)])
              # compute fraction of donut type
              return np.sum(locs == donutType) / num_samples

          answer = probability_of_donutType("glazed", location_choices, num_samples=50000);
          print("Probability of getting a glazed donut is: ",answer);
```

Probability of getting a glazed donut is:   0.60056

# Problem 4

---

Suppose you decide to join a game club to meet other folks on campus. You become facinated 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?

First, write the probability that a '4' is rolled:

$$P(4) = \tfrac{1}{4}$$

Next, the probability that a '4' is rolled again:

$$P_2(4) = P(4) \wedge P(4)$$

$$P_2(4) = \tfrac{1}{4} \cdot \tfrac{1}{4} = \tfrac{1}{16}$$

Again:

$$P_3(4) = P_2(4) \cdot \tfrac{1}{4} = \tfrac{1}{64}$$

Thus, the probability that a '4' is rolled three times in a row is $\tfrac{1}{64}$.

# Part B

*(3 points)*

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

The probability that a '4' is rolled:

$$P(A) = \tfrac{1}{4}$$

The probability that a '4' is not rolled:

$$P(B) = \tfrac{3}{4}$$

The probability that a '4' is rolled once in two rolls:

$$P_{two} = P(A) \wedge P(B)$$

The probability that a '4' is rolled once in three rolls:

$$P_{three} = P(A) \wedge P(B) \wedge P(B)$$

$$= P(A) \cdot P(B) \cdot P(B)$$

$$= \tfrac{1}{4} \cdot \tfrac{3}{4} \cdot \tfrac{3}{4}$$

$$P_{three} = \tfrac{3^2}{4^3} = \tfrac{9}{64}$$

Therefore, the probability that you roll only a single '4' in the three rolls is $\frac{9}{64}$.

# Part C

*(5 points)*

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

Let the event of observing at least one '4' as B and the probability of observing at least two 4's as B. Then:

$$P(A|B) = \frac{P(B \cap A)}{P(B)}$$

Now, $P(B) = \frac{1}{4}$ and $P(A) = \frac{1}{4} \cdot \frac{1}{4}$. So, simplify the above equation:

$$P(A|B) = \frac{P(B) \cdot P(A)}{P(B)}$$

$$P(A|B) = \frac{\frac{1}{4} \cdot \frac{1}{16}}{\frac{1}{4}} = \frac{1}{16}$$

So, the probability of observing at least two 4's given at least one has been observed is $\frac{1}{16}$.

# Problem 5

---

In an attempt to avoid COVID, a group of 100 students seperate 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.

*(9 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?

Put your answer to problem 5 here:

*solution:*

Let V be the event that the virus is in the person and let T be the event the test result is positive.

$$P(T|V) = .90$$

$$P(T|V^c) = .096$$

We know the person is either $V$ or $V^c$

$$P(T) = P(T \cap V) \cup P(T \cap V^C)$$

$$P(T) = (P(V) \cdot P(T|V)) \cup (P(V^c) \cdot P(T|V^c))$$

$$P(T) = P(T|V) \cdot P(V) + P(V^c) \cdot P(T|V^c)$$

Now, substitute the previously established probabilities of events:

$$P(T) = .90 \cdot P(V) + P(V^c) \cdot .096$$

Next, recall there are 100 students and only 1 of them is infected.

Thus, the likelihood of any selected person having the virus is $\frac{1}{100}$:

$$P(V) = .01$$

$$P(V^c) = 1 - P(V) = .99$$

$$P(T) = .90 \cdot .01 + .99 \cdot .096 = 0.10404$$

Now, the probability that a selected person tests positive and they actually are infected:

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

$$P(V|T) = \frac{.90 \cdot .01}{0.10404} = 0.0865$$

Therefore, $P(V|T)$ (the chance the randomly chosen person's test result is accurate) is about 8.65%.

In [ ]: