CSCI3022 S23

Homework 4: Discrete Random Variables

Due Monday, February 13th at 11:59pm to Gradescope

Name: _______Rocky Erdenebat____

Collaboration Policy

While completing the assignment you are not allowed to consult any source other than the course textbooks/online reference links provided on Canvas, your own class notes, and/or the posted lecture slides/in-class Jupyter notebooks. You may discuss questions you have with your classmates or on Piazza or in office hours, but all work you submit must be your own, which means when writing up your solutions or code, you MUST do it entirely by yourself.

You should be able to easily reproduce from scratch and explain a solution that was your own when asked in office hours by a TA/Instructor or on a quiz/exam without referencing your notes/book/HW.

Do not search/ask for a solution online: You may not actively search for a solution to the problems below from the internet. This includes posting to or using sources like ChatGPT, StackOverflow, StackExchange, Reddit, Chegg, CourseHero, etc.

We are here to help! Visit HW Hours and/or post questions on Piazza!

Copying/consulting from the solution of another classmate or an online solution (or providing a classmate your solution) constitutes a **violation of the course's collaboration policy and the honor code and will result in an F in the course and a trip to the honor council**.

Instructions for Submitting in Correct Format

You must submit a PDF of this Juptyer notebook to Gradescope by the deadline listed above. Submissions that are not a PDF or that are not submitted to Gradescope will not be counted for credit.

Before submitting your PDF, make sure that your LaTeX has rendered correctly in you
Any of your solutions with incorrectly rendered or incompletely rendered LaTeX will be

- There are several ways to quickly make a .pdf out of this notebook for Gradescope submission.
- If you are running Juptyer locally on your computer:
  - Option1 : Select Kernel->Restart & Run All. Then select File -> Print Preview, and then Right-Click -> Print using your default browser and "Print to PDF"
  - Option 2: Select Kernel->Restart & Run All. Then select File -> Download as PDF via LaTeX. This will require your system path find a working install of a TeX compiler
- If you are running using CSEL:
  - Option1 : Go to File ->Save & Export Notebook As-> HTML. Then open the HTML, and then Right-Click -> Print and select "Print to PDF".
  - Option2 : Go to File ->Download. Then use this converter https://htmtopdf.herokuapp.com/ipynbviewer/ to convert ipynb to pdf.

Notes

- For full points you must correctly match your questions to the respective Gradescope problem, and include clear comments in your code. Please note that any LaTeX that is not correctly rendered in your submitted PDF will result in a 0 on the entire problem(s) that involves the unrendered LaTeX.
- 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.
- Any relevant data sets are available on Canvas.
- LaTeX Tips: Here is a [reference guide] (https://math.meta.stackexchange.com/questions/5020/mathjax-basic-tutorial-and-quick-reference). **All** of your written commentary, justifications and mathematical work should be in Markdown. I also recommend the wikibook for LaTex.
- Because you can technically evaluate notebook cells is a non-linear order, it's a good idea to do **Kernel** $\to$ **Restart & Run All** as a check before submitting your solutions.
- 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.
- There is *not a prescribed API* for these problems. You may answer coding questions with whatever syntax or object typing you deem fit. Your evaluation will primarily live in the

> clarity of how well you present your final results, so don't skip over any interpretations! Your code should still be commented and readable to ensure you followed the given course algorithm.

Import Pandas and NumPy.

In [1]:
```
# Per the standard import pandas as 'pd' and numpy as 'np'
import pandas as pd
import numpy as np

# You may need binom to calculate a binomial coefficient
from scipy.special import binom
```

In case you create a graph, load Matplotlib's Pylab library to set up Jupyter so that it will plot directly in the notebook.

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

Problem 1 (14 pts)

Dana and Brennan have created a new variation of baseball. It's hard to play with only 2 of them, so they instead have a simplifed variant of a game with the following properties:

- The game consists of rounds:
  - In each round, Dana gets one attempt to hit a pitch thrown by Brennan. Her probability of hitting any one pitch is p and is independent of what occurred in previous rounds.
  - Then Brennan gets one attempt to hit a pitch thrown by Dana. His probability of hitting any one pitch is q and is also independent from Dana's result and what occurred in previous rounds.
  - If one of them succeeds and the other doesn't, the one that succeeded is the winner and the game ends!
  - If they both fail **OR** they both succeed, they proceed to the next round.
- The game will continue until one of them wins.

Denote the following:

- D is the event that Dana wins the **game** at the end.
- E is the event that the game ends in the first round.

Part A: The game will continue until one of them wins. Represent the outcomes of individual hits by H for a hit and M for a miss, and assume that Dana has the first attempt in each round.

Write out the first few events (i.e. all the events that include one round and all the events that include 2 rounds) in the sample space Ω. Notice that since the sample space Ω is infinite, so you won't be able to actually write out the entire set.

Solution:

Round 1: {HM, MH} Round 2: {HHHM, HHMH, MMHM, MMMH}

B: Determine $P(D \mid E)$ in terms of p and q.

Solution: $\frac{p(1-q)}{p+q-2pq}$

Using Bayes' Theorum $P(D \mid E)$ = $\frac{P(D|E)(P(D)}{P(E)}$

The probability that Dana wins the game on the first round is $P(D) = (1-p) * q^0 \implies q$
This is the probability bre

C: Explain in words why the events D and E^C are independent (for this problem you can do this using words/intuition; you do not need to prove this mathematically)

Solution:

This is because the probability of D occuring such that E^C is still $P(D)$. In other words, the probability of D happening is unchanged, or not dependent on the probability that the game does not end on the first round.

D: Use the Law of Total Probability to show that
$P(D) = p(1-q) + (2pq - p - q + 1)P(D \mid E^C)$.

Solution:

E: Use the fact from **Part C** that $P(D \mid E^C) = P(D)$ and your work in **Part D** to determine $P(D)$ in terms of p and q.

Solution:

Part F (2 Points): Explain why the answers to **Part B** and **Part E** are the same.

It is because they are both independent of each other

Solution:

Problem 2 (7 pts)

You drink 3 cups of coffee each day, and in a continuation of your plans to change your diet you decide to stop using sugar in your coffee.

However, you find it too difficult to completely stop using sugar. Therefore, 20% of the time you end up using sugar in your coffee anyway.

Let X be the random variable that indicates the number of 'sugared' coffees **out of a total of three** that you drink per day.

Part A ***(3 points)***

What is the PMF (i.e. probability mass function) for X? Represent your answer in 3 different ways:

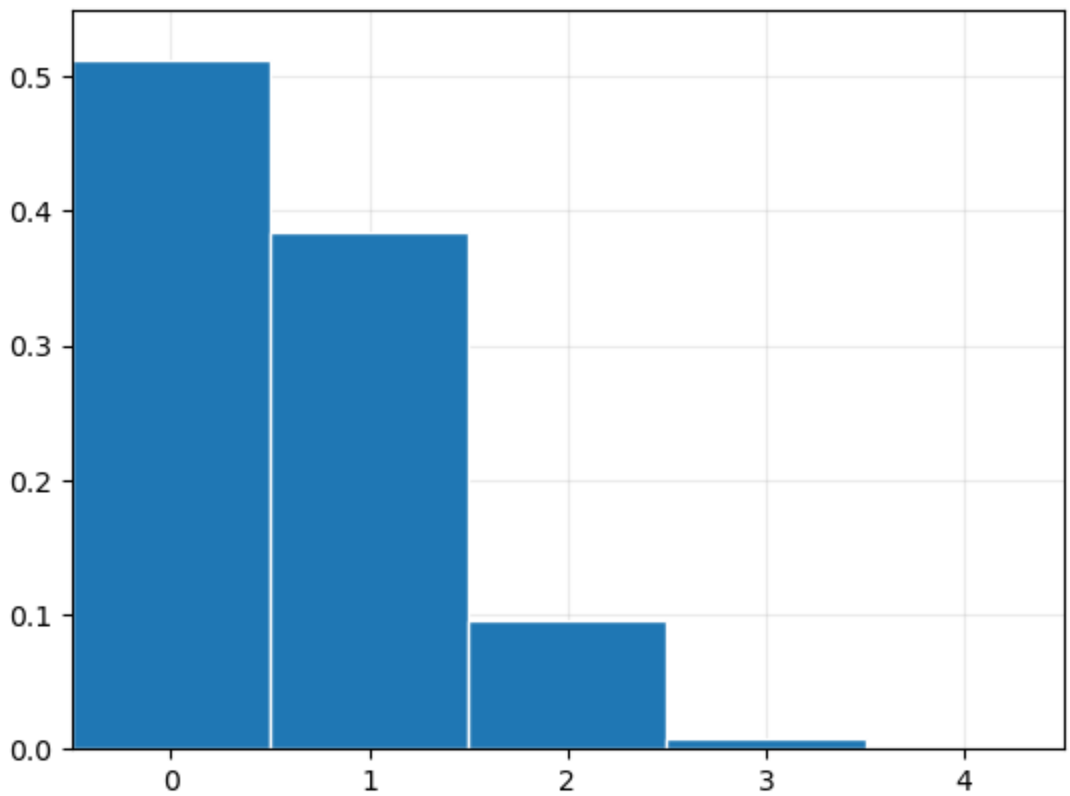
- As a closed form function
- In a table
- As a density histogram (with bin width = 1)

Solution:

P(X)= $\{ 0.2^k * 0.8^{(3-k)}$ for $k = 0, 1, 2, 3$

a	0	1	2	3
$P(x)$	$.512$	$.384$	$.096$	$.008$

In [5]:
```
def binomial_pmf(n, p):
    pmf = np.array([binom(n,k)*(p**k)*((1-p)**(n-k)) for k in range(n+1)])
    return pmf

n = 3
p = 0.2

bins = np.arange(1)

X = np.array(range(n+1))
#Use the function we wrote above:
pmf = binomial_pmf(n, p)

fig, ax = plt.subplots()

#Plot pmf distribution
ax.set_ylim([0, 0.55])
ax.set_xlim([-0.5, 4.5])
ax.bar(X, pmf,width=1, edgecolor='white')
ax.set_axisbelow(True)
ax.grid(alpha=0.25);
```



Part B *(3 points)*

What is the CDF (i.e. cumulative distribution function) for X? Represent your answer in 3 different ways:

- A piecewise function.
- A table.
- A piecewise graph (write code to create this graph).

$$P(X) = \begin{cases} .512 & \text{if } 0 \leq x < 1 \\ .896 & \text{if } 1 \leq x < 2 \\ .998 & \text{if } 2 \leq x < 3 \\ 1 & \text{if } x3 \end{cases}$$

a	0	1	2	3
$P(x)$	$.512$	$.896$	$.992$	1

In [ ]:
```
data = np.arange(0, 6)
y = np.array([.512, .896, .998,1])

#Create a 2nd array for formatting piecewise endpoints
yn = np.insert(y, 0, 0)

fig, ax = plt.subplots()
ax.set_facecolor('white')

# https://matplotlib.org/api/_as_gen/matplotlib.axes.Axes.hlines.html
ax.hlines(y=yn, xmin=data[:-1], xmax=data[1:],
          color='red', zorder=1)

# https://matplotlib.org/api/_as_gen/matplotlib.axes.Axes.vlines.html
#ax.vlines(x=data[1:-1], ymin=yn[:-1], ymax=yn[1:], color='red',linestyle='dashed',

ax.scatter(data[1:-1], y, color='red', s=18, zorder=2)
ax.scatter(data[1:-1], yn[:-1], color='white', s=18, zorder=2,
           edgecolor='red')
ax.grid(False)
ax.set_xlim(data[0], data[-1])
ax.set_ylim([-0.01, 1.01])
```

Part C

(1 points) What is the probability that you drank 2 or fewer sugared coffees in one day?

$$P(X \leq 2) = \frac{12+48+64}{125} = .998$$

Problem 3 (8 pts)

Consider the probability mass function $P(X=x)=2nx^2-2nx$ with the discrete random variable taking on values $1,2,3,4$.

Part A

*(**2 points**)* What is the value of n?

$P(X=1)=2n(1)^2-2n(1)$ = 0

$P(X=2)=2n(2)^2-2n(2)$ = 4n

$P(X=3)=2n(3)^2-2n(3)$ = 12n

$P(X=4)=2n(4)^2-2n(4)$ = 24n

$P(X)=1 \implies 0+4n+12n+24n=1$

$\implies 40n=1$

$\implies n=\frac{1}{40}$

*(**solution:**)*

Part B

*(**2 points**)* What is $P(X=4)$?

*(**solution:**)*

$P(X=4)=24n,\ 24*\frac{1}{40}=\frac{3}{5}$

Part C

*(**2 points**)* What is $P(X\leq 3)=$?

*(**solution:**)*

$$P(X \leq 3) = P(X=3) + P(X=2) + P(X=1) \implies 12n + 4n + 0n$$

$$\implies P(X \leq 3) = 4 * \frac{1}{40} + 12 * \frac{1}{40}$$

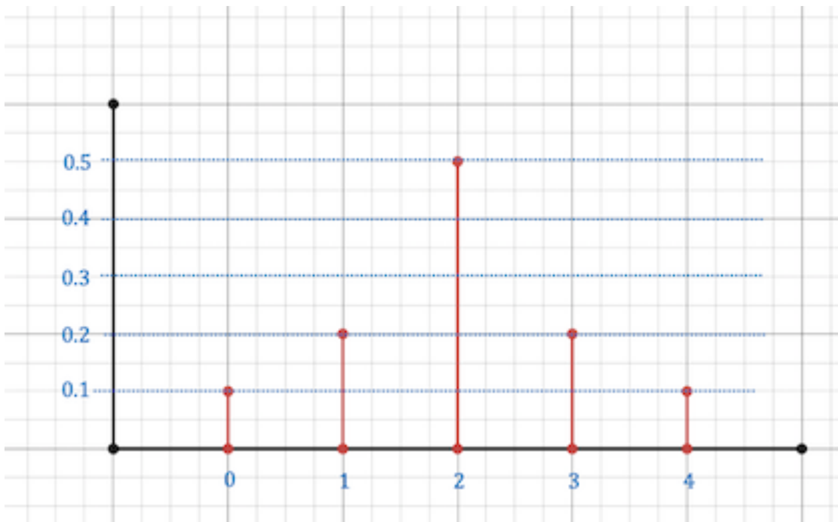
$$\implies P(X \leq 3) = \frac{2}{5}$$

Part D

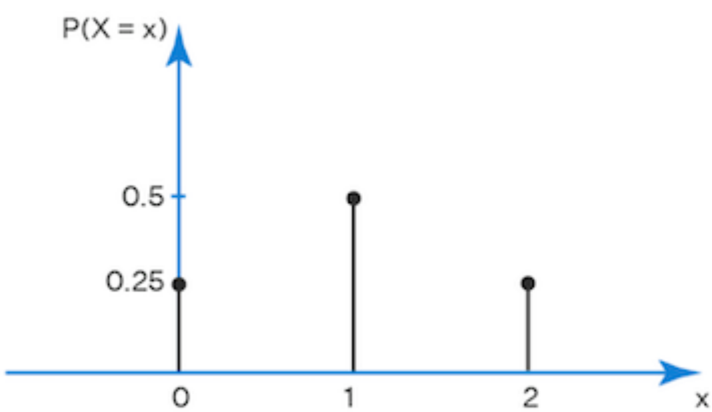
The following part is NOT related to parts a, b, and c above.

One of the following graphs (either graph 1 or graph 2) is a PMF and one is not.

*(**2 points)*** Write code to draw the CDF for the one valid PMF.



GRAPH 1



GRAPH 2:

solution:

In [ ]:

```

```

Problem 4 (12 pts)

You are the proud new owner of a three sided die!

The sides are marked '1', '2', and '3'.

Let X be the amount of 3's obtained after rolling the die twice.

If S is the sample space, then $X : S \to \mathbb{R}$, and $p(x_i) = P(X = x_i)$

Part A

1] *(1 point)* What is $p(2)$?

2] *(1 point)* What is $p(1)$?

1. $p(2)$ is $\frac{1}{3} * \frac{1}{3} \implies \frac{1}{9}$
2. $p(1)$ is $(1,3), (2,3), (3,2), (3,1) \implies \frac{4}{9}$

Part B

1] *(1.5 points)* What is $X((3,3))$?

2] *(1.5 points)* What is $X((2,3))$?

solution:

$X((3,3)) = \frac{1}{3} * \frac{1}{3} \implies \frac{1}{9}$

$X((2,3)) = \frac{1}{3} * \frac{1}{3} \implies \frac{1}{9}$

Part C

1] *(2 points)* Create a table of values for the PMF.

2] *(1 points)* Create (code) a density histogram for the PMF.

solution:

In [ ]:
```
# code your solution to Part C #2 here:
```

Part D

1] *(2 points)* Create a table of values for the CDF.

2] *(1 points)* Create a piecewise function for the CDF.

3] *(1 points)* Create (code) a graph for the CDF.

solution:

In [ ]:
```
# Code your solution to part D, #3,  here:
```

Problem 5 (9 pts)

Part A

We are still playing with our new three sided die and we are still considering rolling a '3' a success. Only now we are rolling the die 10 times!

Suppose you actually rolled the 3-sided die ten times and counted how many times you rolled a '3'.

You could get zero amount of 3's.

You could roll a '3' only once.

You could roll a '3' two out of ten times.

You might even roll a '3' ten out of ten times!

Write a function that takes in the parameters $n = 10$ (for ten rolls of the 3-sided die) and $p = \frac{1}{3}$ (for the probability of rolling a '3'). The function should return the PMF as a Numpy array.

(4 points) Use the function to print out the PMF as a table of values after rolling the 3-sided die 10 times.

i.e. the table should show the probability of rolling zero 3's, one 3, two 3's,..., ten 3's.

solution:

In [ ]:
```
# Put your code to Part A here:
```

Part B

Suppose you rolled the die ten times and wrote down how many 3's resulted.

Then, you again rolled the die ten times and again wrote down how many 3's resulted.

And again you roll ten times and record. And again. And again.

In totality, lets say you recorded results 20 times. That is, twenty times in a row you rolled the 3-sided die 10 times and recorded the amount of 3' that appeared out of the 10 rolls.

You might get 20 results like [2 2 4 4 2 4 5 2 5 2 2 4 2 1 3 1 3 2 3 3] representing 2 out of 10, 2 out of ten, 4 out of ten, etc.

In order to determine how many successes (amount of 3's) TYPICALLY result when you roll this die ten times, you could look at a histogram (a distribution) of your 20 recordings. Better yet, a more accurate picture results from looking at a distribution of 100000 recordings.

(4 points) Create (code) a density histogram of 100000 results to get an estimation of the distribution (aka PMF).

In [ ]:
```
# code your answer to Part B here:
```

Part C

(1 point) From the PMF just created, what appears to be the most common result? In other words, how many times will '3' most commonly appear after rolling a 3-sided die ten times?

solution: Put your solution to Part C here:

In [ ]:
```
```

Final Answers To Selected Problems

These are final answers for selected problems so you can check your work and determine if you are on the right track. To receive credit on each problem you must show all steps leading to these answers, fully answer the problem, and justify your answer using correct mathematical notation.

1b). $\frac{p(1-q)}{p+q-2pq}$

2c). $\frac{124}{125}$

3b). $\frac{3}{5}$

4a Part 2). $\frac{4}{9}$

4b Part 2). $X((2,3)) = 1$

5c). Three

```
In [ ]:
```