Assignment 3
CMPT 210

Due: In class on Friday, 15 July

(1) **[35 marks]** Suppose we have a bowl containing N balls where w of the balls are white.

- If we draw n balls simultaneously (where $n \leq \min\{N - w, w\}$), calculate the probability that we draw k white balls (where $k \leq w$)? [10 marks]
- Let us define the random variable X equal to the number of white balls drawn among the n total balls. Assuming $n \leq \min\{N - w, w\}$ and $k \leq w$, what is the domain of X? [2 marks]
- Prove that $\mathbb{E}[X] = \frac{n\,w}{N}$ [13 marks]

  Instead of drawing the n balls simultaneously, suppose we draw the n balls one at a time *with replacement*. This means that after each draw, we put the ball back into the bowl. Let Y be the random variable equal to the number of white balls drawn among the n total balls.
- Fully specify PDF_Y and compute $\mathbb{E}[Y]$? [10 marks]

(2) **[20 marks]** The CDF of a random variable X is given as:

$$
\begin{aligned}
F(x) &= 0 && \text{(if } x < 0) \\
&= \frac{1}{2} && \text{(if } 0 \leq x < 1) \\
&= \frac{2}{3} && \text{(if } 1 \leq x < 2) \\
&= \frac{11}{12} && \text{(if } 2 \leq x < 3) \\
&= 1 && \text{(if } x \geq 3)
\end{aligned}
$$

- Fully specify the PDF $f(x)$ for X. [5 marks]
- Calculate $\Pr[X < 3]$. [3 marks]
- Calculate $\Pr[2 < X \leq 4]$. [2 marks]
- Calculate $\mathbb{E}[X]$ [3 marks]
- Calculate $\Pr[X = 2 | X \geq 2]$ [3 marks]
- Calculate $\mathbb{E}[X | X \geq 2]$ [4 marks]

(3) **[15 marks]** If $X \sim \text{Bin}(n, p)$, then $\mathbb{E}[X] = np$. In Lecture 13, we calculated this using the linearity of expectation. Using the definition of expectation of a Binomial random variable, prove that,

$$\mathbb{E}[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np$$

For this, first prove that for $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$,

$$\sum_{k=0}^{n} k \binom{n}{k} x^k y^{n-k} = x\, n(x+y)^{n-1}\,, \text{[13 marks]}$$

and use it to compute $\mathbb{E}[X]$ [2 marks]

(4) **[40 marks]** Suppose we have an election between 3 candidates – A, B and C. Since all the candidates are quite bad, the voters vote randomly between the 3 candidates with probabilities p_A, p_B and p_C ($p_A + p_B + p_C = 1$). Assuming that there are n independent voters and let Y_A be the random variable equal to the number of votes candidate A receives.

- What is the distribution of Y_A? Calculate $\mathbb{E}[Y_A]$. [5 marks]
- Let us define the joint distribution PDF_{Y_A,Y_B,Y_C}. For i, j, k such that $i, j, k \in [0, n]$ and $i+j+k = n$, calculate $\text{PDF}_{Y_A,Y_B,Y_C}[i, j, k]$. [15 marks].
- If $p_A = 0.2$, $p_B = 0.3$ and $p_C = 0.5$ and $n = 8$, calculate the probability that A receives half the votes and candidates B and C receive a quarter of the votes each. [5 marks].
- In this same scenario, compute the probability that A receives 3 votes. [5 marks].
- In this same scenario, if A receives 3 votes, compute the probability that B and C receive 3 and 2 votes respectively. [5 marks].
- If A receives 3 votes, calculate the expected number of votes that B receives? [5 marks].

(5) **[30 marks]** We have a set of n keys, one of which fits the door to the apartment. We try the keys randomly until we find the key that fits. Let T be the random variable equal to the number of times we try the keys until we find the right key.

- Suppose each time we try a key that does not fit the door, we simply put it back. This means we might try the same incorrect key several times before finding the right key. On average, how many tries do we need to find the right key? [5 marks]
- Now suppose we throw away each incorrect key that we try. This means that we choose a key randomly from amongst the keys we have not yet tried. What is the maximum number of tries we need to find the right key? [2 marks]
- Given that we have not found the right key on the first $k-1$ tries, prove that the probability that we do not find it on the k^{th} trial is given by:

$$\Pr[T > k | T > k-1] = \frac{n-k}{n-k+1} \quad \text{[3 marks]}$$

- Using the previous result and induction on k prove that $\Pr[T > k] = \frac{n-k}{n}$. [10 marks]
- Using the previous results, prove that the expected number of tries needed to find the right key is $\frac{n+1}{2}$. [10 marks]

(6) **[20 marks]** For a random variable X with range $\mathbb{N}$, prove the following statement:

$$\Pr[X = i] = \Pr[X \geq i] - \Pr[X \geq i+1] \text{ [2 marks]}$$

Use the above result to prove that

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \Pr[X \geq i] \text{ [9 marks]}$$

Use the above result to prove that if $R \sim \text{Geo}(p)$, then, $\mathbb{E}[R] = \frac{1}{p}$. [9 marks]

(7) **[15 marks]** Bruce Lee is practicing by breaking 5 boards with his fists. He is able to break a board with probability 0.8 and he breaks each board independently.

- What is the probability that he breaks exactly 2 out of the 5 boards that are placed before him? [2 marks]
- What is the probability that he breaks at most 3 out of the 5 boards that are placed before him? [3 marks]
- What is the expected number of boards that he will break? [2 marks]
- If X is the r.v. equal to the number of boards that he will break, plot the PDF_X and CDF_X. [8 marks]

(8) **[20 marks]** Suppose we throw a 6-sided dice that has probability p of getting the number 6. We roll the dice independently and multiple times until we get a pair of two consecutive 6's. Define R to be the random variable equal to the number of times we need to throw the dice to get a pair of two consecutive 6's. Using the law of total expectation, prove that

$$\mathbb{E}[R] = \frac{1+p}{p^2}$$

(9) **[30 marks]** Implementing Max Cut

- For a given partition of the nodes $\mathcal{U}$, write a function in the language of your choice (preferably C or Python) to compute $|\delta(\mathcal{U})|$. [5 marks]
- Write a function to compute the exact maximum cut for a given graph. [10 marks]
- Write a function that implements Erdos' randomized algorithm (from Lecture 14) to approximate the Max Cut for a given graph. [5 marks]

Print out your code and submit it with the assignment.

Use the following graph of 5 nodes and 6 edges (given in the form of an adjacency matrix) in order to test the code. $A = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}$. We can compute the expected size of the cut for Erdos' algorithm by repeating the experiment for 1000 independent runs (each run of the algorithm involves selecting a random set of nodes to form $\mathcal{U}$).

- Report the expected size of the cut ($\mathbb{E}[|\delta(\mathcal{U})|]$) and the approximation ratio equal to $\frac{\mathbb{E}[|\delta(\mathcal{U})|]}{OPT}$ where OPT is the size of the exact maximum cut calculated using the above function. [10 marks]