

CS 215

Assignment 1

Due Date: 25 August 2024

Instructions

1. You should type out a report containing all the answers to the written problems in Word (with the equation editor) or using \LaTeX , or write it neatly on paper and scan it. In either case, prepare a single pdf file.
2. The report should contain names and roll numbers of all group members on the first page as a header.
3. Put the pdf file and the code for the programming parts all in one zip file. The pdf should contain the names and ID numbers of all students in the group within the header. The pdf file should also contain instructions for running your code. Name the zip file as follows: A1-RollNumberOfFirstStudent-RollNumberOfSecondStudent-RollNumberOfThirdStudent.zip. (If you are doing the assignment alone, the name of the zip file is A1-RollNumber.zip, if there are two students it should be A1-RollNumberOfFirstStudent-RollNumberOfSecondStudent.zip).
4. Upload the file on moodle BEFORE 11:59 pm on the due date. No assignments will be accepted thereafter.
5. Note that only one student per group should upload their work on moodle, though all group members will receive grades.
6. Please preserve a copy of all your work until the end of the semester.

1 Let's Gamble

There are two friends playing a dice-roll game. Friend A has $(n + 1)$ fair dice and Friend B has n fair dice (a fair die has equal probability of every face). On every roll, a win is achieved if we get a prime number on the top. What is the probability that A will have more wins than B if both roll all of their dice? [5 marks]

2 Two Trading Teams

You are playing a trading game against two teams A and B (will happen in reality soon). The game is played in the form of a three-set series with A and B alternately. Also, Team B is better at trading than Team A. To encourage your trading career, the exchange (an organization responsible for managing the trades) gives you two options A-B-A (which means you play a game with Team

A, then Team B and at last Team A again) or B-A-B. You will win if you win two sets in a row. Which of the two options should you choose? Justify your choice with proper calculations.

[5 marks]

3 Random Variables

3.1 Let Q_1, Q_2 be non-negative random variables. Let $P(Q_1 < q_1) \geq 1 - p_1$ and $P(Q_2 < q_2) \geq 1 - p_2$, where q_1, q_2 are non-negative. Then show that $P(Q_1 Q_2 < q_1 q_2) \geq 1 - (p_1 + p_2)$ [3 marks]

3.2 Given n distinct values $\{x_i\}_{i=1}^n$ with mean μ and standard deviation σ , prove that for all i , we have $|x_i - \mu| \leq \sigma\sqrt{n-1}$. How does this inequality compare with Chebyshev's inequality as n increases? (give an informal answer) [3+2 marks]

4 Staff Assistant

You need a new staff assistant, and you have n people to interview. You want to hire the best candidate for the position. When you interview a candidate, you can give them a score, with the highest score being the best and no ties being possible.

You interview the candidates one by one. Because of your company's hiring practices, after you interview the k^{th} candidate, you either offer the candidate the job before the next interview or you forever lose the chance to hire that candidate. We suppose the candidates are interviewed in a random order, chosen uniformly at random from all $n!$ possible orderings.

We consider the following strategy. First, interview m candidates but reject them all: these candidates give you an idea of how strong the field is. After the m th candidate, hire the first candidate you interview who is better than all of the previous candidates you have interviewed.

(a) Let E be the event that we hire the best assistant, and let E_i be the event that i^{th} candidate is the best and we hire him. Determine $Pr(E_i)$, and show that

$$Pr(E) = \frac{m}{n} \sum_{j=m+1}^n \frac{1}{j-1}$$

[4 marks]

(b) Bound $\sum_{j=m+1}^n \frac{1}{j-1}$ to obtain:

$$\frac{m}{n}(\ln(n) - \ln(m)) \leq Pr(E) \leq \frac{m}{n}(\ln(n-1) - \ln(m-1))$$

[3 marks]

(c) Show that $\frac{m}{n}(\ln(n) - \ln(m))$ is maximized when $m = \frac{n}{e}$, and explain why this means $Pr(E) \geq \frac{1}{e}$ for this choice of m . [3 marks]

5 Free Trade

Imagine an infinitely long line of traders waiting outside a brokerage firm to place their trades. Each trader is assigned an ID number from 1 to 200 (both inclusive, obviously these IDs are not unique). The firm's director announces a special offer: the first trader in the queue whose ID number matches the ID of any trader who has already placed a trade will receive a free trade (i.e., a trade without any margins). You have the option to choose your position in this queue. However, you don't know the ID numbers of the traders ahead of you or behind you. Your goal is to maximize

your chances of being the first trader whose ID matches someone who has already placed a trade. Given this situation, what position in the queue should you choose to maximize your chances of receiving the free trade?

[6 marks]

6 Update Functions

Suppose that you have computed the mean, median and standard deviation of a set of n numbers stored in array A where n is very large. Now, you decide to add another number to A . Write a python function to update the previously computed mean, another python function to update the previously computed median, and yet another python function to update the previously computed standard deviation. Note that you are not allowed to simply recompute the mean, median or standard deviation by looping through all the data. You may need to derive formulae for this. Include the formulae and their derivation in your report. Note that your python functions should be of the following form:

```
function newMean = UpdateMean(OldMean, NewDataValue, n, A),  
function newMedian = UpdateMedian(OldMedian, NewDataValue, n, A),  
function newStd = UpdateStd(OldMean, OldStd, NewMean, NewDataValue, n, A).
```

Also explain, how would you update the histogram of A , if you received a new value to be added to A ? (Only explain, no need to write code.) Please specify clearly if you are making any assumptions.

[10 marks]

7 Plots

Read about the following plots:

- Violin Plot
- Pareto Chart
- Coxcomb Chart
- Waterfall Plot

Describe the uses of these plots. Take some sample data and generate one example plot for each of them.

[8 marks]

8 Monalisa

Download the image of Monalisa from [here](#). Read the image using matplotlib ([example](#)). Write a piece of python code to shift the image along the X direction by t_x pixels where t_x is an integer ranging from -10 to +10 (so, in total you need to do this for 20 values). While doing so, assign a value of 0 to unoccupied pixels. For each shift, [compute the correlation coefficient between the original image and its shifted version](#). Make a plot of correlation coefficients across the shift values. Also, generate a normalized histogram for the original image. You might need to refer to section 3.3 from this [book](#). You are not allowed to use any inbuilt function for generating the histogram. If you are using any other libraries, then please mention about them in the pdf.

[8 marks]