CS 215

Assignment 1 Solutions

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]

Solution:

Credits: Aayush Borkar, Sandeep Reddy Nallamilli, Yash Sabale

A can either win after both of them throwing n throws each, or A can win on the last throw by drawing in the first n throws and getting one more win in the last throw. In the first case, the last throw from A does not matter, and in the latter, A needs to win the last throw.

We define the following events:

E_A : A has more wins after n throws

E_B : B has more wins after n throws

E_D : A and B have the same number of wins after n throws

The probability that A is leading after n throws must be equal to the probability that B is leading after n throws:

$$P(E_A) = P(E_B)$$

Also, we have:

$$P(E_A) + P(E_B) + P(E_D) = 1$$

Thus,

$$P(E_A) = \frac{1 - P(E_D)}{2}$$

To find the probability that A will have more wins, we consider both cases where A could win:

1. A wins after n throws: This is already covered by $P(E_A)$.

2. A wins on the last throw given that both A and B have the same number of wins after n throws: The probability of A winning the last throw is $\frac{1}{2}$, and the probability of ending in a draw after n throws is $P(E_D)$.

Combining these,

$$P(\text{A having more wins}) = P(E_A) + P(\text{A winning on the last throw} \mid E_D) \times P(E_D)$$

$$P(\text{A having more wins}) = \frac{1 - P(E_D)}{2} + \frac{1}{2} \times P(E_D)$$

$$P(\text{A having more wins}) = \frac{1 - P(E_D) + P(E_D)}{2}$$

$$P(\text{A having more wins}) = \frac{1}{2}$$

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]

Solution:
Credits: Malay Kedia, Aakash Gupta, Yash Singh

Solution:
Let us assign the probabilities a and b to both teams A and B winning while playing against us and E be the event of our team winning.

Since Team B is better at trading than Team A, we have:

$$P(A) = a, P(B) = b$$
$$0 \leq a, b \leq 1$$
$$\text{a} \leq b$$

For the 2 scenarios, lets start with A-B-A first. Where the 2 possibilities of our team winning is when we either win the first 2 match first with A then with B, or we lose the first match with A and then go on to win the other two with B and A consecutively.

We can write it as:

$$\begin{aligned} P(E_1) =& (1-a) \cdot (1-b) + a \cdot (1-b) \cdot (1-a) \\ =& (1+a) \cdot (1-a) \cdot (1-b) \\ =& (1-a^2) \cdot (1-b) \end{aligned}$$

For the Second scenario, we have B-A-B. Here also the 2 possibilities of us winning is when we defeat B and A consecutively, or we lose the first match with B and go on to defeat A and B in the next 2 rounds.

This will look like:

$$\begin{aligned} P(E_2) =& (1-b) \cdot (1-a) + b \cdot (1-a) \cdot (1-b) \\ =& (1+b) \cdot (1-b) \cdot (1-a) \\ =& (1-a) \cdot (1-b^2) \end{aligned}$$

Comparing the 2 probabilities for both scenarios we'll have,

$$\frac{P(E_1)}{P(E_2)} = \frac{1+a}{1+b} \leq 1, \qquad \{\because \ b \geq a\}$$

Showing that the second scenario is more favourable for our team.

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 $P(Q_1Q_2 < q_1q_2) \geq 1-(p_1+p_2)$ [3 marks]

Solution:

Credits: Aditya Neeraje, Balaji Karedla, Moulik Jindal

Adding p_i and subtracting $P(Q_i < q_i)$ on both sides for $i \in \{1, 2\}$, we get

$$P(Q_1 \geq q_1) \leq p_1$$

$$P(Q_2 \geq q_2) \leq p_2$$

The event $(Q_1Q_2 \geq q_1q_2)$ is a subset of the union of the events $(Q_1 \geq q_1)$ and $(Q_2 \geq q_2)$. This is because if $Q_1Q_2 \geq q_1q_2$, at least one of Q_1, Q_2 is $\geq$ the corresponding q_i. Thus,

$$P(Q_1Q_2 \geq q_1q_2) \leq p_1 + p_2$$

Since $p_1 + p_2$ itself is $\geq$ the probability of the union of events which occur with probabilities p_1 and p_2. Subtracting from 1 on both sides, we get

$$P(Q_1Q_2 < q_1q_2) \geq 1 - (p_1 + p_2)$$

as required.

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]

Solution: By definition, we have $\sum_{i=1}^n (x_i - \mu)^2 = \sigma^2(n-1)$ and hence for any i, we have $(x_i - \mu)^2 \leq \sigma^2(n-1)$, i.e. $x_i - \mu \leq \sigma\sqrt{n-1}$.
Comparison: Chebyshev's inequality provides a bound which is satisfied by some fraction of all the sample points in the dataset, not all of them, whereas this inequality is for all points. However, this inequality is much looser than Chebyshev's inequality, especially for large n. Note that Chebyshev's inequality is unaffected by the number of points. For small n, the bound in this question is better than Chebyshev's bound. For example, when $n = 10$, this inequality tell us that all values lie within $\pm 3\sigma$ of μ, whereas Chebyshev's inequality states that 99.4% of the values lie within $\pm 13\sigma$ of μ. When $n = 100$, this inequality tell us that all values lie within $\pm 9.95\sigma$ of μ, whereas Chebyshev's inequality states that 99.4% of the values lie within $\pm 13\sigma$ of μ. When $n = 1000$, this inequality tell us that all values lie within $\pm 31.6\sigma$ of μ, whereas Chebyshev's inequality states that 99.4% of the values lie within $\pm 13\sigma$ of μ. As n increases, you see Chebyshev's bound becoming better.

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 mth 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]

Credits: Aditi Singh, Abhilasha Sharma Suman, Navya Garg

Solution

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]

Credits: Vavilala Chidvilas Reddy, Rishi Kalra, Mehul Borad

Solution

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]

Credits: Jigyasa Chouhan, Gowri K Balakrishnan, Aryan Mahajan

Code

Histogram Update

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]

Credits: Abhi Jain, Anushka Singhal, Sabil Ahmad

Solution

Code

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]

Credits: Niral Charan, N.K. Vishwaajith, Shravya Karna
Solution
Code