CS215 Assignment 1 Solutions

Satyam Sinoliya, 23B0958
Vaibhav Singh, 23B1068
Shaik Awez Mehtab, 23B1080

August 2024

Question 1. 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 1. So A or B wins when they get either 2, 3 or 5, whose probability is $\frac{1}{2}$. So the probability that A wins i number of times when all of the $(n+1)$ dice are rolled is

$$p(A_i) = \binom{n+1}{i}\left(\frac{1}{2}\right)^i\left(\frac{1}{2}\right)^{n+1-i} \tag{1}$$

$$= \binom{n+1}{i}\left(\frac{1}{2}\right)^{n+1} \tag{2}$$

Similarly, probability that B gets i wins when all of the n are rolled is

$$p(B_i) = \binom{n}{i}\left(\frac{1}{2}\right)^i\left(\frac{1}{2}\right)^{n-i} \tag{3}$$

$$= \binom{n}{i}\left(\frac{1}{2}\right)^{n} \tag{4}$$

Now the probability that A has more wins than B is

$$p(i > j) = \sum_{i=0}^{n+1}\sum_{\substack{j=0\\ i>j}}^{n} p(A_i)p(B_j) \tag{5}$$

Since number of wins of A is independent from that of B. By writing each possible term $\binom{n+1}{i}$ and $\binom{n}{j}$ in a grid manner. We observe that our required terms cover exactly half of the grid. Thus equation (5) simplifies to

$$p(i > j) = \frac{\sum_{i=0}^{n+1}\binom{n+1}{i}\left(\frac{1}{2}\right)^{n+1}\cdot\sum_{j=0}^{n}\binom{n}{j}\left(\frac{1}{2}\right)^{n}}{2} = \frac{1}{2} \tag{6}$$

Thus, the probability of A winning more times than B is $\frac{1}{2}$.

Question 2. **3.1** Let Q1, Q2 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 $P(Q_1Q_2 < q_1q_2) \geq 1 - (p_1 + p_2)$ [5 marks]

Solution 2. Define two events, E_1 and E_2:

1. $E_1 = \{Q_1 < q_1\}$
2. $E_2 = \{Q_2 < q_2\}$

So,

$$P(E_1) \geq 1 - p_1$$

$$P(E_2) \geq 1 - p_2$$

We need to prove that,

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

Let's define another event E_3, where $E_3 = \{Q_1Q_2 < q_1q_2\}$
If we consider the complement of E_3, which is

$$\{Q_1Q_2 < q_1q_2\}^{\mathsf{C}} = \{Q_1Q_2 \geq q_1q_2\}$$

$$E_3^{\mathsf{C}} = \{Q_1Q_2 \geq q_1q_2\}$$

If $Q_1 Q_2 \geq q_1 q_2$, and Q_1, Q_2 are non-negative integers, it is very clear that, atleast one of the following has to be true:

$$Q_1 \geq q_1 \text{ or } Q_2 \geq q2$$

This means

$$E_3^{\complement} \subseteq \{Q_1 \geq q_1\} \cup \{Q_2 \geq q_2\}$$

This implies,

$$P(E_3^{\complement}) \leq P(\{Q_1 \geq q_1\} \cup \{Q_2 \geq q_2\})$$
$$P(E_3^{\complement}) \leq P(\{Q_1 \geq q_1\}) + P(\{Q_2 \geq q_2\})$$

It is clear that:

$$\{Q_1 \geq q_1\} = E_1^{\complement} \text{ and } \{Q_2 \geq q_2\} = E_2^{\complement}$$

So,

$$1 - P(E_3) \leq P(E_1^{\complement}) + P(E_2^{\complement})$$
$$1 - P(E_3) \leq 1 - P(E_1) + 1 - P(E_2)$$
$$1 - P(E_3) \leq 1 - (1 - p_1) + 1 - (1 - p_2)$$
$$1 - P(E_3) \leq p_1 + p_2$$

This implies,

$$P(E_3) \geq 1 - (p_1 + p2)$$

Question 3. Read about the following plots:

1. Violin Plot
2. Pareto Chart
3. Coxcomb Chart
4. Waterfall Plot

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

[8 marks]

Solution 3. Here are the descriptions and usages of the given plots along with their examples:

1. **Violin plot:** It's a hybrid of box plot and kernel density plot. A box plot represents data in a linear fashion. It's made of a straight line from lowest value to highest value along with a box from first to third quartiles, marking all the quartiles of the dataset. Here's an example:

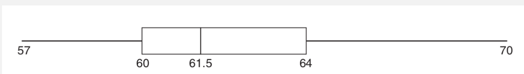

Figure 1: Box plot

A kernel density plot represents the density/frequency of data points. It's similar to a histogram, but smooth. In a violin graph, this is kept vertical, with two mirror images of it reflected along y-axis. This is useful in the sense that we can look into both centrel tendencies of the data (like mean, median etc.) but also how the data is distributed. Both at once. We can visualise the following data which I've taking from (*) containing the average number of hours a person studies given the number of courses taken

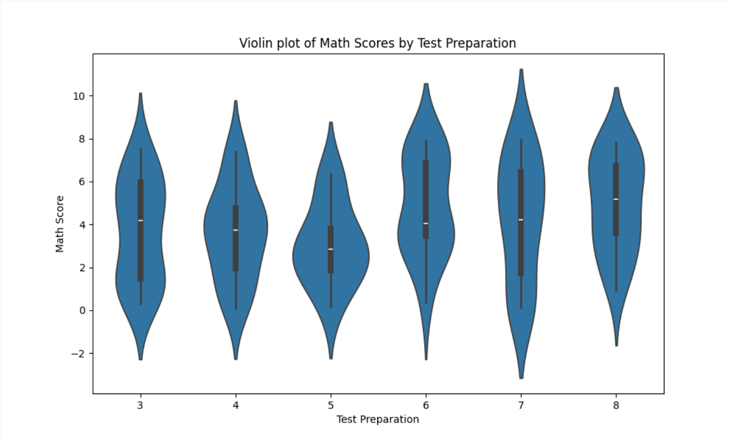

Figure 2: Violin Plot

As mentioned before, a violing plot can show both statistical summary along with distribution, which a normal plot can't. Here, the gray line represents the box plot component of it. And the plot you get by rotating it by 90° is the distribution plot.

2. **Pareto Chart:**

Question 4. 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 tx pixels where tx 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]

Solution 4. The image of Mona Lisa was read using the `matplotlib` library. The image was then shifted horizontally by `tx` pixels for each value of `tx` in the range of -10 to +10. The shifting operation was implemented manually, ensuring that unoccupied pixels were assigned a value of 0.
A custom function `shiftimg` was created to handle the shifting process:

- If `tx > 0`, pixels were shifted rightwards by `tx` units.
- If `tx < 0`, pixels were shifted leftwards by `tx` units.
- If `tx = 0`, the function returned the original image.

For each shifted image, the correlation coefficient between the original and shifted image was calculated. This coefficient quantifies the linear relationship between the two images, with values ranging from -1 to 1, where 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation. A normalized histogram of the original image was generated by calculating the frequency of each pixel intensity and then normalizing the values. This was done manually without using any inbuilt histogram function.

The correlation coefficients for the different shift values were plotted to observe the relationship between the magnitude of the shift and the correlation with the original image.

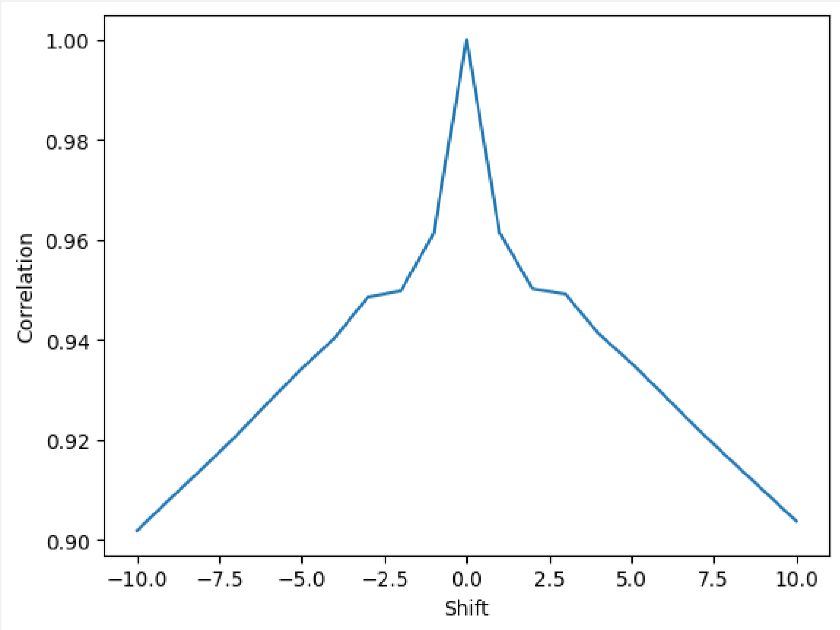

Figure 3: Correlation Coefficient vs Shift Values

As observed in Figure 3, the correlation decreases as the shift increases in either direction. This is expected as the more the image is shifted, the less it resembles the original, resulting in lower correlation values.
The normalized histogram of the original image was also plotted to visualize the distribution of pixel intensities.

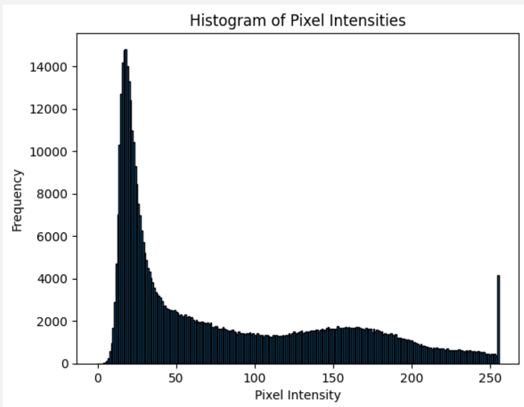

Figure 4: Normalized Histogram of the Original Image

The histogram in Figure 4 shows the frequency of occurrence of each pixel intensity, normalized over the total number of pixels.