CS 215 Assignment 1 Report

Nimish Manware(22b0944)
Prasanna Nage (22b0953)
Jayesh Jadhav (22b1056)

August 2023

Contents

Q 1

(a)

Here the first person who picks the book has to select his own from the n books. So the probability that this happens is $P(p_1b_1) = \frac{1}{n}$. Similarly, the probability that the second person chooses his own book from the remaining $n-1$ books is $P(p_2b_2) = \frac{1}{n-1}$, and for the i^{th} to choose his book is $P(p_ib_i) = \frac{1}{n-p+1}$ So assuming that these events are independent, the probability that everyone chooses their own book is

$$\begin{aligned} P(E_1) &= P(p_1b_1) \cdot P(p_2b_2) \cdots P(p_nb_n) \\ &= \frac{1}{n} \cdot \frac{1}{n-1} \cdot \frac{1}{n-2} \cdots \frac{1}{1} \\ &= \frac{1}{n!} \end{aligned}$$

(b)

Here we give the same argument as before, that the first person picks the first book and the second person picks the second book, but the $(m+1)^{th}$ person onwards can pick any book. So the probability after the $(m)^{th}$ term is just 1.

$$\begin{aligned} P(E_2) &= P(p_1b_1) \cdot P(p_2b_2) \cdots P(p_mb_m) \cdot 1 \\ &= \frac{1}{n} \cdot \frac{1}{n-1} \cdot \frac{1}{n-2} \cdots \frac{1}{n-m+1} \\ &= \frac{(n-m)!}{n!} \end{aligned}$$

(c)

Now the first person has to pick amongst the m books of the m persons, so the probability that this happens is $P_1 = \frac{m}{n}$. Similarly for the i^{th} person, $i \leq m$ is $P_i = \frac{m-i+1}{n-i+1}$. Rest of them can now pick any book so that probablity is just 1.

$$\begin{aligned} P(E_3) &= P_1 \cdot P_2 \cdots P_m \cdot 1 \\ &= \frac{m}{n} \cdot \frac{m-1}{n-1} \cdot \frac{m-2}{n-2} \cdots \frac{1}{n-m+1} \\ &= \frac{(n-m)! \times (m)!}{n!} \end{aligned}$$

(d)

Probability that the first person gets clean book is $P_1 = (1-p)$. Same is for the next person. After the m^{th} person the rest of them may get any book, clean or unclean.

$$\begin{aligned} P(E_4) &= P_1 \cdot P_2 \cdots P_m \cdot 1 \\ &= (1-p) \cdot (1-p) \cdots (1-p) \\ &= (1-p)^m \end{aligned}$$

(e)

We first select the m persons who gets a clean books by ${}^nC_m$. The rest have to get a unclean books. Now the probability the the selected persons all get clean books is $P_c = (1-p)^m$ and the rest getting unclean books is $P_{uc} = p^{n-m}$ thus the total prob is

$$\begin{aligned} P(E_5) &=^n C_m \cdot P_1 \cdot P_2 \\ &=^n C_m (1-p)^m p^{n-m} \end{aligned}$$

Q 2

We know that the formula for standard deviation is

$$\begin{aligned} \sigma^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2 \\ \Rightarrow \sigma^2(n-1) &= \sum_{i=1}^{n}(x_i-\mu)^2 \\ \Rightarrow \sigma^2(n-1) &\geq \sum_{i=1}^{n}(x_i-\mu)^2 \\ \therefore \sigma\sqrt{(n-1)} &\geq |x_i-\mu| \quad \forall i \leq n \end{aligned} \tag{2.1}$$

The equality will hold only when all the elements are equal.As then all the data values will be equal to μ.
Now the Chebyshev's inequality is

$$S_k = \{x_i : |x_i-\mu| \geq k\sigma\} \Rightarrow |S_k| < \frac{n}{k^2}$$

$$\text{Here } k = \sqrt{n-1},$$

$$|S_k| < \frac{n}{n-1}$$

For higher values of n

$$|S_k| \leq 1 \tag{2.2}$$

We know from equation (2.1) that the number of element in the set S_k is zero and the equation (2.2) is saying that this set will either contain 0 or 1 elements. So the two inequalities are consistent.

Q 3

$$F = \{|Q_1+Q_2| > \epsilon\} \quad \text{and} \quad E = \{|Q_1|+|Q_2| > \epsilon\}$$

Here We know that,

$$\begin{aligned} &|Q_1+Q_2| \leq |Q_1|+|Q_2| \quad \forall Q_1, Q_2 \in \mathbb{R} \\ &\therefore |Q_1+Q_2| > \epsilon \quad \Rightarrow \quad |Q_1|+|Q_2| > \epsilon \end{aligned}$$

Thus the event F is the subevent of E. Therefore $P(F) < P(E)$. (3.1)
Now,

$$|Q_1| + |Q_2| > \epsilon \;\Rightarrow\; |Q_1| > \epsilon \;\vee\; |Q_2| > \epsilon$$

Thus $E = E_1 \cup E_2 \Rightarrow$

$$\begin{aligned} \Rightarrow P(E) &= P(E_1) + P(E_2) - P(E_1 \cap E_2) \\ &\leq P(E_1) + P(E_2) \end{aligned} \tag{3.2}$$

From (3.1) and (3.2)

$$P(F) \leq P(E_1) + P(E_2)$$

Q 4

Consider the inequalities

$$Q_1 < q_1 \quad \text{and} \quad Q_2 < q_2$$

As q_1 and q_2 are non-negative this implies

$$\begin{aligned} Q_1Q_2 &< q_1q_2 \\ \therefore P(Q_1Q_2 < q_1q_2) &= P((Q_1 < q_1) \cap ((Q_2 < q_2)) \\ &= P(Q_1 < q_1) + P(Q_2 < q_2) - P((Q_1 < q_1) \cup ((Q_2 < q_2)) \\ &\geq 1 - p_1 + 1 - p_2 - P((Q_1 < q_1) \cup ((Q_2 < q_2)) \\ &\geq 1 - p_1 + 1 - p_2 - 1 \\ \therefore P(Q_1Q_2 < q_1q_2) &\geq 1 - (p_1 + p_2) \end{aligned}$$

Q 5

(a)

As the car and stones are already placed behind the doors the event of choosing a door will not alter it's probability i.e. these events are independent.

$$P(C_i|Z_i) = P(C_i) = 1/3 \quad \forall \quad i \in \{1, 2, 3\}$$

(b)

Now $P(H_3|C_1, Z_1)$, means that car is behind door1 and contestant has chosen 1 is given so,

$$P(H_3|C_1, Z_1) = 1/2 \quad \text{as the guest will only open door with stones}$$

for i=2

$$P(H_3|C_2, Z_1) = 1 \quad \text{as he has only one option}$$

for i=3

$$P(x_3|c_1, z_1) = 0$$

(c)

$$\begin{aligned}
\frac{P(H_3|C_1, Z_1)P(C_1, Z_1)}{P(H_3, Z_1)} &= \frac{P(H_3|C_1, Z_1)P(C_1, Z_1)}{P(H_3|Z_1)P(Z_1)} \\
&= \frac{(1)(1/3)(1/3)}{(1/2)(1/3)} \\
&= \frac{2}{3}
\end{aligned}$$

Here we used $P(H_3, Z_1) = P(H_3|Z_1)P(Z_1)$ and also the fact that $P(H_3|Z_1) = \frac{1}{2}$

(d)

$$\begin{aligned}
P(C_1|H_3, Z_1) &= \frac{P(H_3|C_1, Z_1)P(C_1, Z_1)}{P(H_3, Z_1)} \\
&= \frac{P(H_3|C_1, Z_1)P(C_1, Z_1)}{P(H_3|Z_1)P(Z_1)} \\
&= \frac{(1/2)(1/3)(1/3)}{(1/2)(1/3)} \\
&= \frac{1}{3}
\end{aligned}$$

(e)

In this case the host will open the door 3 with equal probability as door 2. Thus

$$P(H_3|C_1, Z_1) = \frac{1}{2}$$

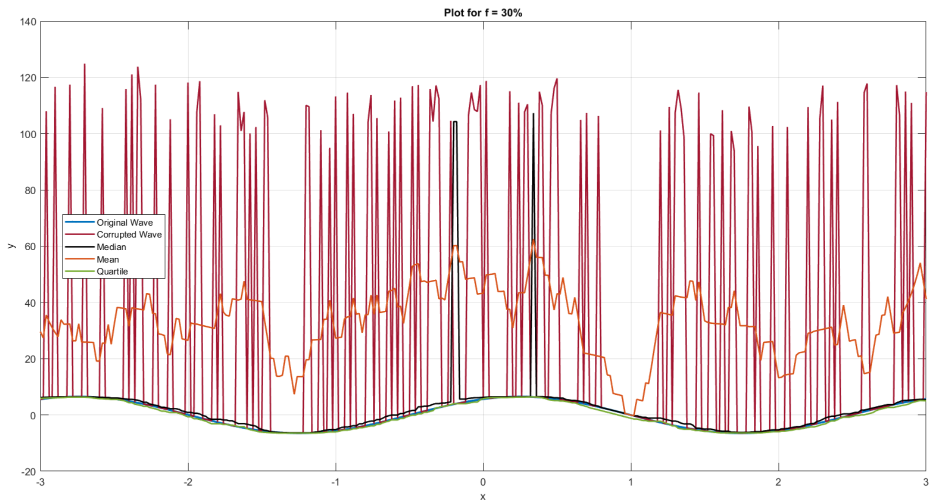
$$P(H_3|C_2, Z_1) = \frac{1}{2}$$

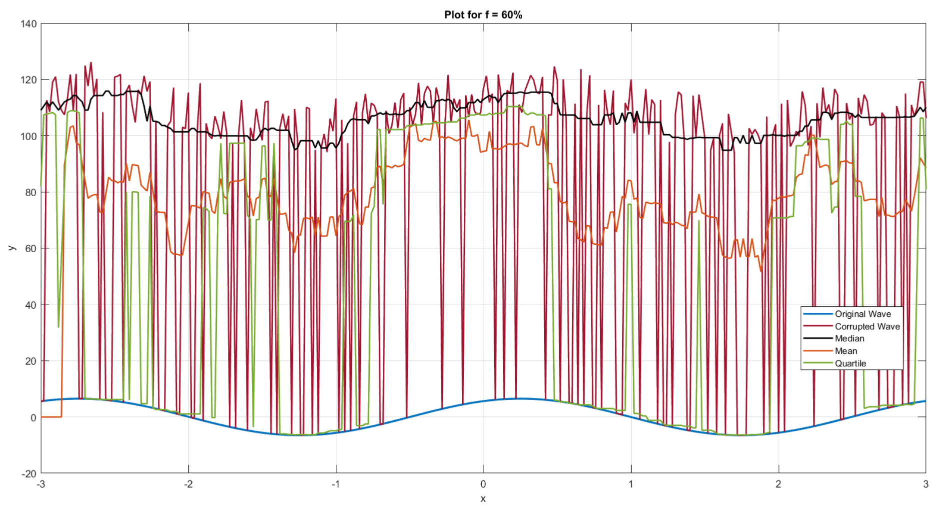
$$P(H_3|C_3, Z_1) = \frac{1}{2}$$

$$
\begin{aligned}
P(C_2|H_3,Z_1) &= \frac{P(H_3|C_2,Z_1)P(C_2,Z_1)}{P(H_3|Z_1)} \\
&= \frac{(1/2)(1/3)(1/3)}{((1/2)(1/3))} \\
&= \frac{1}{3} \\
P(C_1|H_3,Z_1) &= \frac{P(H_3|C_1,Z_1)P(C_1,Z_1)}{P(H_3|Z_1)} \\
&= \frac{(1/2)(1/3)(1/3)}{((1/2)(1/3))} \\
&= \frac{1}{3}
\end{aligned}
$$

Q 6

Plots





Analyzing the filters

We Find mean squared values for each filtering for each case
For f = 30
Mean squared error for median filtering: 4.7856
Mean squared error for mean filtering: 57.1198

Mean squared error for quartile filtering: 0.013507
For f = 60
Mean squared error for median filtering: 520.787
Mean squared error for mean filtering: 288.8463
Mean squared error for quartile filtering: 169.6121

Explanation

For f = 30 percent.

We see that mean squared error for quartile filtering produced better relative mean squared error than mean and median filtering. This is because it returns the 25 percentile of each which is less affected by larger values which in this case is outliers. So it produces better results than median which takes the lower 50 percent which will be more affected by these outliers. Mean just takes the average of these so it is more affected. Here the percentage of corruption is 30 percent so median affects less than mean.

For f = 30 percent.

Here the percentage of corruption is more than 50 percent so it is more likely that the median of windows will be the corrupted values. So it performs worse than mean because that mean will take the average value of the corrupted values and the original rather than taking the corrupted values directky like median does.

Q 7

Let the newly added element be A_{n+1}. Following are the derivations for change in mean, median and standard deviation.

i) Change in Mean

Let,

$$Old\,Mean = \overline{A} = \frac{\sum_{i=1}^{n} A_i}{n}$$

and

$$New\,Mean = \overline{A'}$$

Then, we have

$$
\begin{aligned}
\overline{A'} &= \frac{\sum_{i=1}^{n+1} A_i}{n+1} \\
&= \frac{\sum_{i=1}^{n+1} A_i}{n} && \text{(As } n \text{ is large so } n+1 \approx n) \\
&= \frac{\sum_{i=1}^{n} A_i + A_{n+1}}{n} \\
&= \frac{\sum_{i=1}^{n} A_i}{n} + \frac{A_{n+1}}{n} \\
&= \overline{A} + \frac{A_{n+1}}{n}
\end{aligned}
$$

So,

$$\boxed{New\,Mean = Old\,Mean + \frac{A_{n+1}}{n}}$$

We will use this to derive equation for new standard deviation.

ii) Change in Median

We divide the problem in cases: Case 1: n is even
Further there can be 3 possibilities:
1. $A_{n+1} \leq A_{n/2}$, new median will be $A_{n/2}$ itself as it comes in the middle of the new array
2.$An+1 \geq An/2+1$, new median will be $A_{n/2+1}$ as it will be in the middle of the new array
3. similarly if $A_{n/2} \leq A_{n+1} \leq A_{n/2+1}$, then new median will be A_{n+1}. Case 2: n is odd
1. if $A_{n+1} < A_{\frac{n-1}{2}}$, new median will be $(Old\,Median + A_{\frac{n-1}{2}})/2$
where $Old\,Median = (A_{n/2} + A_{n/2+1})/2$
2. if $A_{n+1} > A_{\frac{n+3}{2}}$, new median will be $(Old\,Median + A_{\frac{n+3}{2}})/2$
3. if $A_{\frac{n+3}{2}} < A_{n+1} < A_{\frac{n-1}{2}}$, new median will be $(Old\,Median + A_{n+1})/2$.

iii) Change in Standard Deviation

Let,

$$Old\,Standard\,Deviation = S$$

and

$$New\,Standard\,Deviation = S'$$

According to given formula,

$$
\begin{aligned}
S &= \sqrt{\frac{\sum_{i=1}^{n}(A_i - \overline{A})^2}{n-1}} \\
&= \sqrt{\frac{\sum_{i=1}^{n}(A_i^2 + \overline{A}^2 - 2 \cdot A_i \cdot \overline{A})}{n-1}} \\
&= \sqrt{\frac{\sum_{i=1}^{n}(A_i)^2 + \sum_{i=1}^{n}(\overline{A})^2 - \sum_{i=1}^{n}(2 \cdot A_i \cdot \overline{A})}{n-1}} \\
&= \sqrt{\frac{\sum_{i=1}^{n}(A_i)^2 + n \cdot \overline{A}^2 - 2 \cdot n \cdot \overline{A}^2}{n-1}} && \left(\text{since } \sum_{i=1}^{n} A_i = n \cdot \overline{A}\right) \\
&= \sqrt{\frac{\sum_{i=1}^{n}(A_i)^2 - n \cdot \overline{A}^2}{n-1}} && \text{...(1)}
\end{aligned}
$$

And so,

$$
\begin{aligned}
S' &= \sqrt{\frac{\sum_{i=1}^{n+1}(A_i)^2 - (n+1) \cdot \overline{A'}^2}{n+1-1}} \\
&= \sqrt{\frac{\sum_{i=1}^{n}(A_i)^2 + A_{n+1}^2 - n \cdot \overline{A}^2 + n \cdot \overline{A}^2 - (n+1) \cdot \overline{A'}^2}{n}} \\
&= \sqrt{\frac{\sum_{i=1}^{n}(A_i)^2 - n \cdot \overline{A}^2 + n \cdot \overline{A}^2 + A_{n+1}^2 - (n+1) \cdot \overline{A'}^2}{n}} \\
&= \sqrt{\frac{(n-1) \cdot S^2 + n \cdot \overline{A}^2 + A_{n+1}^2 - (n+1) \cdot \overline{A'}^2}{n}} && (\text{using (1)}) \\
&= \sqrt{\frac{(n-1) \cdot S^2}{n} + \overline{A}^2 + \frac{A_{n+1}^2}{n} - \frac{(n+1) \cdot \overline{A'^2}}{n}} \\
&= \sqrt{S^2 + \overline{A}^2 - \overline{A'}^2 + \frac{A_{n+1}^2}{n}} && \left(\text{given } n \text{ is large } \frac{n-1}{n} \approx \frac{n+1}{n} \approx 1\right)
\end{aligned}
$$

So,

$$
\boxed{S' = \sqrt{S^2 + \overline{A}^2 - \overline{A'}^2 + \frac{A_{n+1}^2}{n}}}
$$

Updating the Histogram of A

For updating histogram of A after addition of the new element we find the bin corresponding to the value of the new element and then we increment the frequency of this particular bin by 1. If the bin is not existing, then we will create a new bin and assign it a frequency value of 1.