CS215 Assignment 1

Aadish Jain - 190050001
Deepanshu - 190050032
Gaurang Dev - 19D070024

September 3, 2020

Contents

1 Q1

Anyone can pick any book
So, here the Sample Space is, $S(n) = n!$

(a)

First person has to choose his book among the n books, **only one way is possible**.
Simlilarly for second person,
and so for the n^{th}
hence,

$$E(x) = 1 * 1 * 1 * ... * 1 \quad (n\ times)$$

$$E(x) = 1 \qquad\qquad \text{where x is the event}$$

As we know that,

$$P(x) = \frac{E(x)}{S(x)}$$

So, here probability will be,

$$\boxed{P(x) = \frac{1}{n!}}$$

(b)

For first m persons it reduce to part(a) itself.
For the rest they can pick any book among the remaining ones. i.e.,

$$({}^{n-m}C_1 * {}^{n-m-1}C_1 * ... * {}^{1}C_1) \tag{1}$$

this reduce to

$$= (n - m)!$$

E(first m receive their book back)

$$= 1 * (n - m)!$$

Hence,

$$\boxed{P(x) = \frac{(n - m)!}{n!}}$$

(c)

Assuming

$$2m < n$$

So first m persons have to choose a book from the book which belong to last m persons
So, first m persons can choose any book from last m
Total possibilities for first m

$$= ({}^{m}C_1 * {}^{m-1}C_1 * ... * {}^{1}C_1)$$

i.e.,

$$m!$$

For the rest, total possibilities will be,

$${}^{n-m}C_1 * {}^{n-m-1}C_1 * ... * {}^{1}C_1$$

$$(n-m)!$$

Totally,

$$E(x) = m! * (n-m)!$$

So finally,

$$\boxed{P(x) = \frac{1}{{}^{n}C_m}}$$

(d)

$$\text{probability of book being clean} = (10 - p)$$

Total possibilities for first m person,

$$E(x) = ((1-p) * (1-p) * ... * (1-p))$$

$$E(x) == (1-p)^m$$

So, total probability is,

$$\boxed{P(x) = (1-p)^m}$$

(e)

Any m persons could have clean book, so
Taking binomial distribution into account, total possibilities will be,

$$E(x) = {}^{n}C_m * (1-p)^m * p^{n-m}$$

Hence the probability will be,

$$\boxed{P(x) = {}^{n}C_m * (1-p)^m * p^{n-m}}$$

2 Q2

For n distinct values x_i where i=1..n with variance σ^2 and mean μ, we know that

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2/(n-1)$$

$$\implies (n-1)\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2$$

Since every term of the summation on the RHS is positive, this means that the sum will be greater than any individual term,
Therefore, we may write

$$\sum_{i=1}^{n} (x_i - \mu)^2 \geq (x_i - \mu)^2, \forall i$$

$$\implies (n-1)\sigma^2 \geq (x_i - \mu)^2, \forall i$$

Taking square root both the sides,

$$\boxed{\sigma\sqrt{n-1} \geq \mid x_i - \mu \mid, \forall i}$$

Comparing with Chebyshev's inequality :
We know that Chebyshev's inequality says :

$$P(\mid X - \mu \mid \leq k\sigma) \geq 1 - \frac{1}{k^2}$$

In this equation, let $k = \sqrt{n-1}$, and it becomes :

$$P(\mid X - \mu \mid \leq \sigma\sqrt{n-1}) \geq 1 - \frac{1}{n-1}$$

As n approaches infinity, the RHS approaches 1, but probability cannot be greater than 1, so we may write

$$\lim_{n\to\infty} P(\mid X - \mu \mid \leq \sigma\sqrt{n-1}) = 1$$

which means that $P(\mid x_i - \mu \mid \leq \sigma\sqrt{n-1}) = 1, \forall i$ which we proved earlier for every n. So we may say that Chebyshev's inequality is loose as compared to the given inequality because in case of Chebyshev's inequality, it holds only when n approaches infinity whereas it is true for all n.

3 Q3

By Chebyshev-Cantelli inequality we have:

$$S_k = \{x_i : x_i - \mu \geq k\}$$

Then:

$$\frac{|S_k|}{N} \leq \frac{1}{1 + \frac{k^2}{\sigma^2}}$$

Given N is even we can say that exactly half the number of total elements will be less than median and exactly half will be greater than that. (Property of median when N is even).
Therefore putting $k = t - \mu$ in the above inequality, exactly half of the terms $(x_i - \mu)$ will be greater than k; Therefore, For $k = t - /mu$:

$$\frac{|S_k|}{N} = \frac{1}{2}$$

$$\frac{1}{2} \leq \frac{1}{1 + \frac{(t-\mu)^2}{\sigma^2}}$$

$$2 \geq 1 + \frac{(t-\mu)^2}{\sigma^2}$$

$$1 \geq \frac{(t-\mu)^2}{\sigma^2}$$

$$\sigma^2 \geq (t-\mu)^2$$

Taking positive Square roots both sides we get:

$$\boxed{|\mu - t| \leq \sigma \quad \square}$$

4 Q4

Given:

$$\text{total rickshaws} = 100$$

$$\text{red rickshaws} = 1$$

$$\text{blue rickshaws} = 99$$

P(red is observed given that its actually red),

$$P(r_o|r_a) = \frac{99}{100}$$

P(red is observed given that its actually blue),

$$P(r_o|b_a) = \frac{2}{100}$$

Solution:
We have to find probability such that rickshaw is actually red given that it was observed red.
Applying Baye's theorem,

$$P(r_a|r_o) = \frac{P(r_o|r_a) * P(r_a)}{P(r_o|r_a) * P(r_a) + P(r_o|\bar{r_a}) * P(\bar{r_a})}$$

or in other words,

$$P(r_a|r_o) = \frac{P(r_o|r_a) * P(r_a)}{P(r_o|r_a) * P(r_a) + P(r_o|b_a) * P(b_a)}$$

putting all the values from the given values,

$$P(r_a|r_o) = \frac{\frac{99}{100} * \frac{1}{100}}{\frac{99}{100} * \frac{1}{100} + \frac{2}{100} * \frac{99}{100}}$$

So the final probability will be,

$$\boxed{P(r_a|r_o) = \frac{1}{3}}$$

5 Q5

(a)

By definition:

$$P(C_i|Z_1) = \frac{P(C_i \cap Z_1)}{P(Z_1)} \qquad \forall i \in \{1,2,3\}$$

Since events C_i and Z_1 are independent events, we can write:

$$P(C_i|Z_1) = P(C_i) \qquad \forall i \in \{1,2,3\}$$

Since occurrence of a car behind any door is equally likely. Therefore:

$$\boxed{P(C_i|Z_1) = P(C_i) = \frac{1}{3}} \qquad \forall i \in \{1,2,3\}$$

(b)

we can write:

$$P(H_3|Z_1, C_i) = \frac{P(C_i \cap (Z_1 \cap H_3))}{P(Z_1 \cap C_i)} \qquad \forall i \in \{1,2,3\}$$

Define a set U which contains all ordered pairs (x,y) where x denotes the door no. where car is present and y denotes the door that host opens. **Note** that Set U assumes that event Z_1 is given. Therefore:

$$U = \{(1,3),(1,2),(2,3),(3,2)\}$$

For $i = 1$

i=1 denotes that Car is present behind door 1.
Given that car is present behind door 1. Sample Space (S) and favourable set (E) becomes:

$$S = \{(1,3),(1,2)\}$$

and

$$E = \{(1,3)\}$$

Therefore;

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

For $i = 2$

i=2 denotes that Car is present behind door 2.
Given that car is present behind door 2. Sample Space (S) and favourable set (E) becomes:

$$S = \{(2, 3)\}$$

and

$$E = \{(2, 3)\}$$

Therefore;

$$\boxed{P(H_3|Z_1, C_2) = \frac{|E|}{|S|} = 1}$$

For $i = 3$

i=3 denotes that Car is present behind door 3.
Given that car is present behind door 3. Sample Space (S) and favourable set (E) becomes:

$$S = \{(3, 2)\}$$

and

$$E = \phi$$

Therefore;

$$\boxed{P(H_3|Z_1, C_3) = \frac{|E|}{|S|} = 0}$$

(c)

Using (a) and (b):

$$P(H_3|Z_1, C_2) = 1$$

$$P(C_2, Z_1) = P(C_2|Z_1)P(Z_1) = \frac{1 \times 1}{3 \times 3} = \frac{1}{9}$$

$$P(H_3|Z_1, C_i) = \frac{P(C_i \cap (Z_1 \cap H_3))}{P(Z_1 \cap C_i)} \qquad \forall i \in \{1, 2, 3\}$$

Now the event $(H_3 \cap Z_1)$ can be partitioned into 3 disjoint exhaustive events $(C_i \cap H_3 \cap Z_1) \forall i \in \{1, 2, 3\}$. Therefore,

$$P(H_3, Z_1) = \sum_{i=1}^{3} P(C_i \cap H_3 \cap Z_1)$$

$$P(H_3, Z_1) = \sum_{i=1}^{3} P(H_3|Z_1, C_i)P(Z_1 \cap C_i)$$

Since Z_1 and C_i are independent events. Therefore:

$$P(H_3, Z_1) = \sum_{i=1}^{3} P(H_3|Z_1, C_i)P(Z_1)P(C_i)$$

$$P(H_3, Z_1) = \frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} + 1 \times \frac{1}{3} \times \frac{1}{3} + 0 \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{6}$$

Therefore:

$$\boxed{P(C_2|H_3, Z_1) = \frac{1 \times \frac{1}{9}}{\frac{1}{6}} = \frac{2}{3}}$$

(d)

Using (a), (b) and (c).

$$P(H_3, Z_1) = \frac{1}{6}$$

$$P(C_1, Z_1) = P(C_1|Z_1)P(Z_1) = \frac{1 \times 1}{3 \times 3} = \frac{1}{9}$$

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

Therefore:

$$\boxed{P(C_1|H_3, Z_1) = \frac{\frac{1}{2} \times \frac{1}{9}}{\frac{1}{6}} = \frac{1}{3}}$$

(e)

Since

$$P(C_2|H_3, Z_1) > P(C_1|H_3, Z_1)$$

Hence we can conclude that switching the door is **beneficial**.

(f)

If the host were whimsical,
For part (a), it does not matter as the sample space and events we are interested in are independent of the actions of the host. Therefore:

$$\boxed{P(C_i|Z_1) = \frac{1}{3}} \qquad \forall i \in \{1, 2, 3\}$$

For part (b), Our sample space (S) and favourable event (E) changes: For i=1,

$$S = \{(1,3),(1,2)\}$$

$$E = \{(1,3)\}$$

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

For i=2,

$$S = \{(2,2),(2,3)\}$$

$$E = \{(2,3)\}$$

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

For i=3,

$$S = \{(3,2),(3,3)\}$$

$$E = \{(3,3)\}$$

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

For part (c),

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

$$P(C_2, Z_1) = P(C_2|Z_1)P(Z_1) = \frac{1 \times 1}{3 \times 3} = \frac{1}{9}$$

$$P(H_3, Z_1) = \frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{6}$$

therefore

$$\boxed{P(C_2|H_3, Z_1) = \frac{\frac{1}{2} \times \frac{1}{9}}{\frac{1}{6}} = \frac{1}{3}}$$

For part (d),

$$P(H_3, Z_1) = \frac{1}{6}$$

$$P(C_1, Z_1) = P(C_1|Z_1)P(Z_1) = \frac{1 \times 1}{3 \times 3} = \frac{1}{9}$$

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

Therefore:

$$\boxed{P(C_1|H_3, Z_1) = \frac{\frac{1}{2} \times \frac{1}{9}}{\frac{1}{6}} = \frac{1}{3}}$$

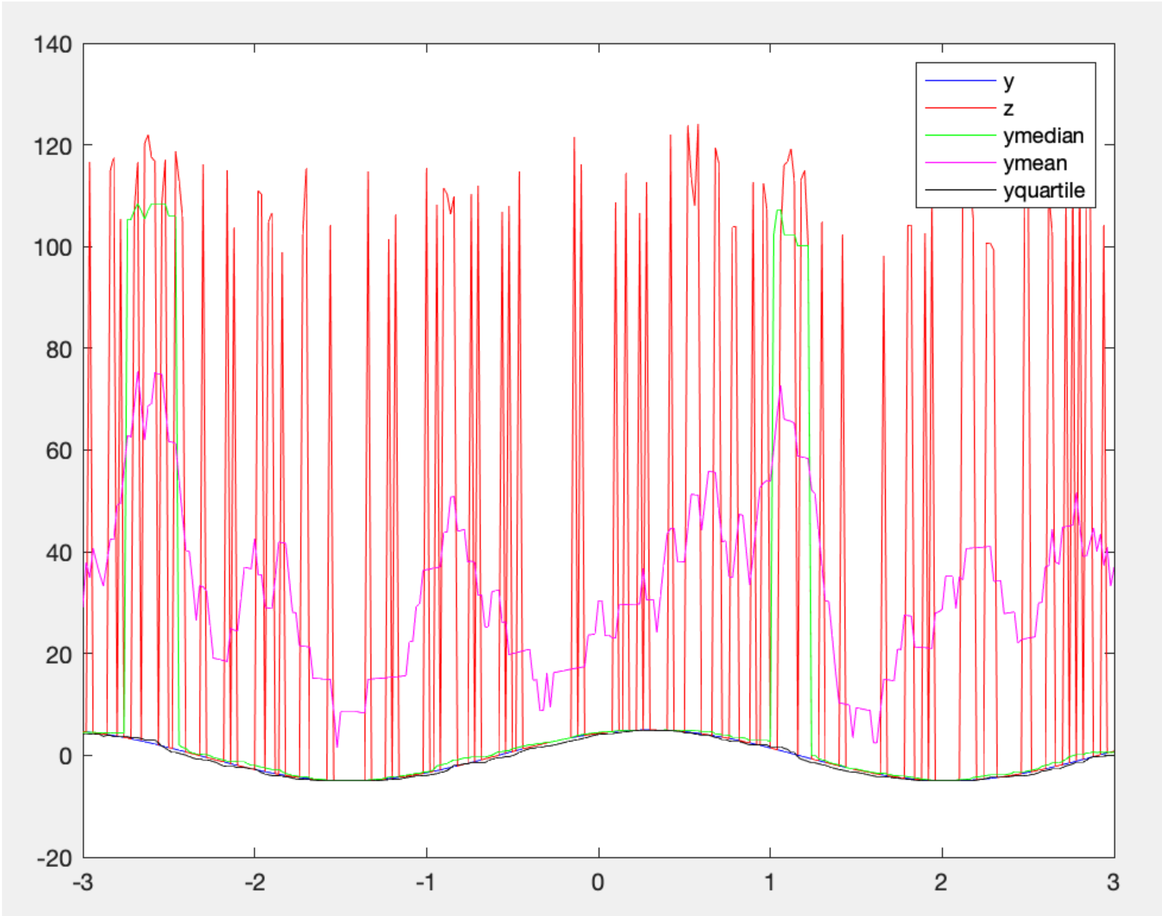
And finally for part(e),
Since

$$P(C_2|H_3, Z_1) = P(C_1|H_3, Z_1)$$

Hence, we can conclude that switching door is **not** beneficial if the host were whimsical.

6 Q6

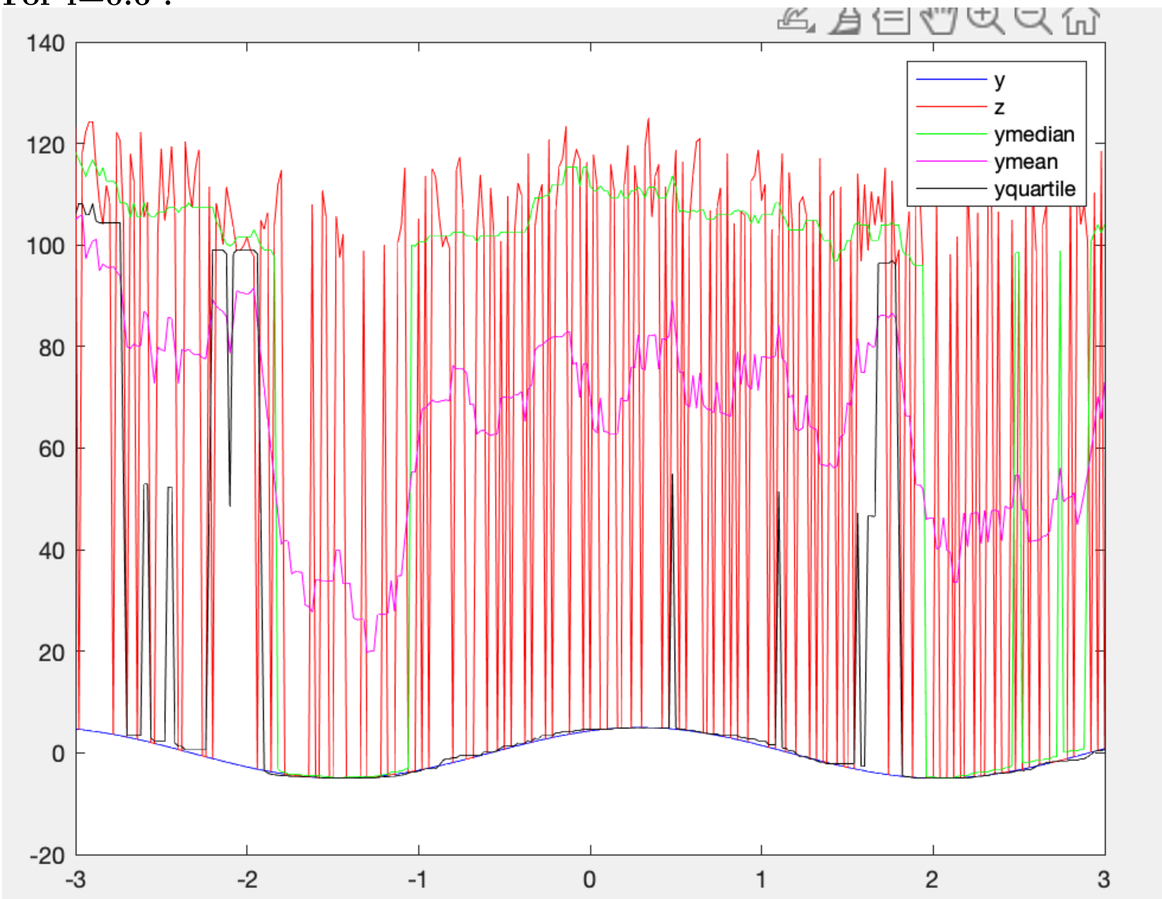
For f=0.3 :



Type of filtering	Mean Squared Error
median	73.46
mean	106.39
first quartile	0.01

Table 1: Mean squared error values for f=0.3

For f=0.6 :



Type of filtering	Mean Squared Error
median	753.79
mean	376.18
first quartile	47.57

Table 2: Mean squared error values for f=0.6

Median is more robust to outliers as compared to mean upto some threshold, but as noise crosses threshold, the relative error for median filter increases more rapidly as compared to mean which is evident from the data.
Moving first quartile filtering produced the best relative mean squared error. This is because in quartile filtering, we essentially take median of the values less than the median and since median is quite robust, the first quartile is even more robust to noise. But it's error also increases as noise increases.
The source code file is attached with the pdf as q6.m, to compile it type q6 on command window and enter.

7 Q7

(a)newMean

$$newMean = \frac{OldSum + NewDataValue}{n+1}$$
$$newMean = \frac{OldMean \times n + NewDataValue}{n+1}$$

(b)newMedian

Let

$$a = NewDataValue$$

(i) n is even

$$OldMedian = \frac{A[\frac{n}{2}] + A[\frac{n}{2}+1]}{2}$$

Case (1):

$$A[\frac{n}{2}] > a$$

then

$$NewMedian = A[\frac{n}{2}]$$

Case (2):

$$A[\frac{n}{2}+1] < a$$

then

$$NewMedian = A[\frac{n}{2}+1]$$

Case (3):

$$A[\frac{n}{2}] < a \;\; \& \;\; A[\frac{n}{2}+1] > a$$

then

$$NewMedian = a$$

(ii) n is odd

$$OldMedian = A[\frac{n+1}{2}]$$

Case (1):

$$A[\frac{n-1}{2}] > a$$

then

$$NewMedian = \frac{A[\frac{n-1}{2}] + A[\frac{n+1}{2}]}{2}$$

Case (2):

$$A[\frac{n+3}{2}] < a$$

then

$$NewMedian = \frac{A[\frac{n+1}{2}] + A[\frac{n+3}{2}]}{2}$$

Case (3):

$$A[\frac{n-1}{2}] < a \ \ \& \ \ A[\frac{n+3}{2}] > a$$

$$NewMedian = \frac{A[\frac{n+1}{2}] + a}{2}$$

(c)newStd

Let

$$\mu_0 = OldMean$$

$$\mu = newMean$$

$$\sigma_0 = OldStd$$

$$\sigma = newStd$$

$$x_{n+1} = NewDataValue = a$$

Now,

$$\sum_{i=1}^{n}(x_i - \mu_0)^2 = \sigma_0^2(n-1)$$

$$\sum_{i=1}^{n+1}(x_i - \mu)^2 = \sigma^2 n$$

On subtracting the above two eqautions we get,

$$(\mu_0 - \mu)\sum_{i=1}^{n}(2x_i - \mu - \mu_0) + (a-\mu)^2 = \sigma^2 n - \sigma_0^2(n-1)$$

$$(\mu_0 - \mu)(2n\mu_0 - n\mu - n\mu_0) + (a-\mu)^2 = \sigma^2 n - \sigma_0^2(n-1)$$

$$n(\mu_0 - \mu)(\mu_0 - \mu) + (a-\mu)^2 = \sigma^2 n - \sigma_0^2(n-1)$$

$$n(\mu_0 - \mu)^2 + (a - \mu)^2 = \sigma^2 n - \sigma_0^2(n-1)$$

$$n(\mu_0 - \mu)^2 + (a - \mu)^2 + \sigma_0^2(n-1) = \sigma^2 n$$

$$\boxed{\sqrt{(\mu_0 - \mu)^2 + \frac{(a-\mu)^2}{n} + \frac{\sigma_0^2(n-1)}{n}} = \sigma}$$

(d) updating histogram

First we will find the bin which contains our NewDataValue and then we will increase the frequency of that bin by 1.

3 files UpdateMean.m, UpdateMedian.m and UpdateStd.m are attached along with this pdf. We can implement these functions with the arguments given in question.