Assignment 1: CS 215

190050113 Shivam Raj **190050080** Pawan Kumar
190020010 Aman Singh

September 3, 2020

Contents

Question 1

(a) The situation is equivalent to distributing n books to n people. The total number of ways of doing that is $n!$. There is only 1 way in which everyone gets his book back. So the probability of it happening is $\boxed{1/n!}$

(b) There is only 1 way of distributing m books to their respective m owners. And for this way there are $(n-m)!$ ways of distributing the left $n-m$ books among left $n-m$ people for a total of $1 \,\mathrm{x}\, (n-m)! = (n-m)!$ ways. So the probability of it happening is $\boxed{(n-m)!/n!}$

(c) There are $m!$ way of distributing the m books belonging to the last m people to the first m people. And for each such way there are $(n-m)!$ ways of distributing the left $n-m$ books among left $n-m$ people for a total of $m! \,\mathrm{x}\, (n-m)! = m!(n-m)!$ ways. So the probability of it happening is $\boxed{m!(n-m)!/n!}$

(d) Every book has a probability p of getting unclean which is independent of who picked up which book and independent of whether other books became unclean. So the probability of first m people picking up books that are unclean will be $\prod_m p = \boxed{p^m}$

(e) The probability of some *particular* m people picking up books that are unclean and the rest clean books is $p^m(1-p)^{n-m}$. But there are ${}^nC_m$ different ways of choosing a particular set of m people from a group of n. So the probability of it happening is $\boxed{{}^nC_m p^m(1-p)^{n-m}}$

Question 2

2\. For n distinct values $\{x_i\}_{i=1}^{n}$

we have

$$\sigma = \frac{1}{\sqrt{n-1}} \sqrt{\sum (x_i-\mu)^2}$$

where μ is mean and σ is standard deviation

consider the inequality

$$\sqrt{a_1^2+a_2^2+\dots+a_n^2} \geq |a_i| \quad \forall\, i=1,2,\dots n$$

because

$$|a_i|^2 \leq \sum_{i=1}^{n} a_i^2 \Rightarrow |a_i| \leq \sqrt{\sum_{i=1}^{n} a_i^2}$$

$$\therefore \quad \sigma = \frac{1}{\sqrt{n-1}} \sqrt{\sum (x_i-\mu)^2} \geq \frac{1}{\sqrt{n-1}} |x_i-\mu| \quad \forall\, i=1 \text{ to } n$$

$$\therefore \quad (\sqrt{n-1})\,\sigma \geq |x_i-\mu|$$

consider the above proof for $n>1$ and for $n=1$

$\mu = x_1$ and $\sigma = 0$, therefore inequality $|x_i-\mu| \leq \sigma\sqrt{n-1}$ holds.

~~The~~ Thus the inequality is proved for all $n \geq 1$.

Now let us consider chebyshev's inequality for calculating fraction (estimitated) of x_i values satisfying this

$$|x_i-\mu| \geq \sqrt{n-1}\,\sigma$$

By using chebyshev's inequality,

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

Therefore frection satisfying

$|x_i - \mu| \leq \sigma\sqrt{n-1}$ will be greater than

or equal to $1 - \frac{1}{n-1}$

As when n increase $\frac{1}{n-1} \to 0$ and the above fraction $\to 1$

Thus we can say that as n increases chebyshev's inequality starts stating that $|x_i - \mu| \leq \sigma\sqrt{n-1}$ is true $\forall\ i = 1$ to n which is actually true.

As n increases chebyshev's inequality gives a very good approximination.

Question 3

3. We will try to prove it by contradiction.

suppose $|\mu - \tau| > \sigma$

Case ①: when $\tau > \mu$;

we have $\tau - \mu > \sigma \Rightarrow \tau > \sigma + \mu$

As median is a ~~elem~~ number will almost divides the data in such a manner that median is a central element.

Let say $n = 2k$ and $\{y_i\}_{i=1}^{n}$ is in ascending order and y_i are elements of x_i then μ lies between y_k and y_{k+1}

$\therefore \quad y_k \leq \mu \leq y_{k+1}$

Thus we can say that fraction of element such that $y_i \geq \tau$ must be $\geq \frac{1}{2}$

As $\tau > \sigma + \mu$,

$\therefore$ we can say that fraction of element such that $y_i > \sigma + \mu$ must be $\geq \frac{1}{2}$

But from one sided chebyshev's inequality fraction of element such that $y_i - \mu > \sigma$ must be $< \frac{1}{1+(1)^2} = \frac{1}{2}$

$\therefore$ We got contradiction

this means $|\mu - \tau| > \sigma$ cannot be true when $\tau > \mu$

Case ②: when $\tau < \mu$;

By similar argument above and using the $(k<0)$ form of Chebyshev-Cantelli inequality.

$|\mu - \tau| > \sigma$ cannot be true when $\tau < \mu$

Teacher's Signature ....

when $\tau > \mu$ or $\tau < \mu$.

Therefore $|\mu - \tau| \leq \sigma$ will always be true as we showed that $|\mu - \tau| > \sigma$ is always false.

For $\tau = \mu$

$0 \leq \sigma$ which is obviously true

$\therefore$ $|\mu - \tau| \leq \sigma$ will be always true

Question 4

4. $P(A)$ = probability of rickshaw to be red = 0.01

$P(B)$ = probability of rickshaw to be blue = 0.99

Let C be an event in which the rickshaw is ~~actually~~ observed red.

Then, $P(C|A) = 0.99$

$P(C|B) = 0.02$

Therefore by Bayes theorem,

$$P(A|C) = \frac{P(C|A)\,P(A)}{P(C|A)\,P(A) + P(C|B) * P(B)}$$

$$= \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.99 \times 0.02}$$

$$= \frac{0.01}{0.01 + 0.02} = \frac{1}{3}$$

Therefore the probability that the rickshaw is actually a red one when XYZ observed it to be red $= \frac{1}{3}$

Question 5

(a).

$$P(C_1|Z_1) = \tfrac{1}{3}$$
$$P(C_2|Z_1) = \tfrac{1}{3}$$
$$P(C_3|Z_1) = \tfrac{1}{3}$$

(b) $P(H_3|C_1, Z_1) = \frac{1}{2}$ (if car is present behind first door so rest two doors have equal probability)

$P(H_3|C_2, Z_1) = 1$ (if car is present behind door 2, so the host will always pick door 3)

$P(H_3|C_3, Z_1) = 0$ (if car is present behind door 3 , them host will never select it)

(c)

$$P(H_3|C_2, Z_1) = 1$$
$$P(C_2, Z_1) = P(C_2|Z_1) \times P(Z_1) = \tfrac{1}{3} \cdot \tfrac{1}{3} = \tfrac{1}{9}$$
$$P(H_3|Z_1) = \tfrac{1}{3} \cdot \tfrac{1}{2} + \tfrac{1}{3} \cdot 1 + \tfrac{1}{3} \cdot 0 = \tfrac{1}{2}$$

$$\begin{aligned} P(H_3, Z_1) &= P(H_3|Z_1) \times P(Z_1) \\ &= \frac{1}{2} \cdot \frac{1}{3} \\ &= \frac{1}{6} \end{aligned}$$

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

(d)

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

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

So, Switching will be beneficial

(f)

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

$$\forall_i P(C_i, Z_1) = \frac{1}{3}$$
$$P(H_3|Z_1) = \frac{1}{2}$$
$$P(H_3, Z_1) = P(H_3|Z_1) \times P(Z_1) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$

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

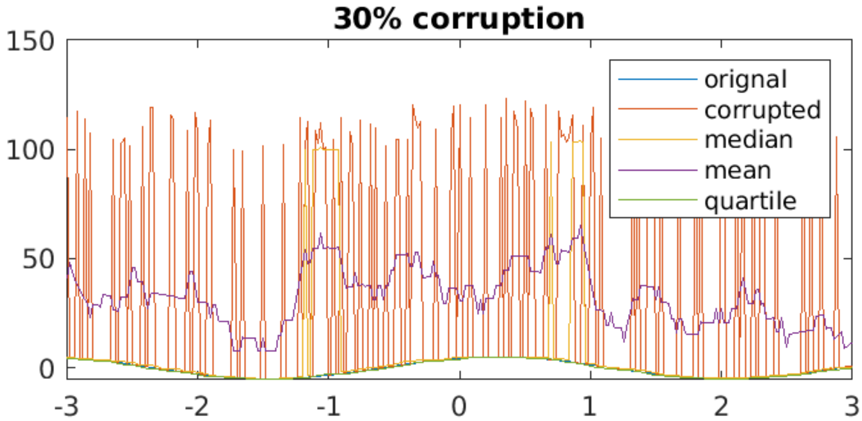
$$\begin{aligned} P(C_1|H_3, Z_1) &= \frac{P(H_3|C_1, Z_1)P(C_1, Z_1)}{P(H_3, Z_1)} \\ &= \frac{\frac{1}{2} \times \frac{1}{9}}{\frac{1}{6}} \\ &= \frac{1}{3} \end{aligned}$$

$P(C_2|H_3, Z_1) = P(C_1|H_3, Z_1)$. So switching won't create any difference.

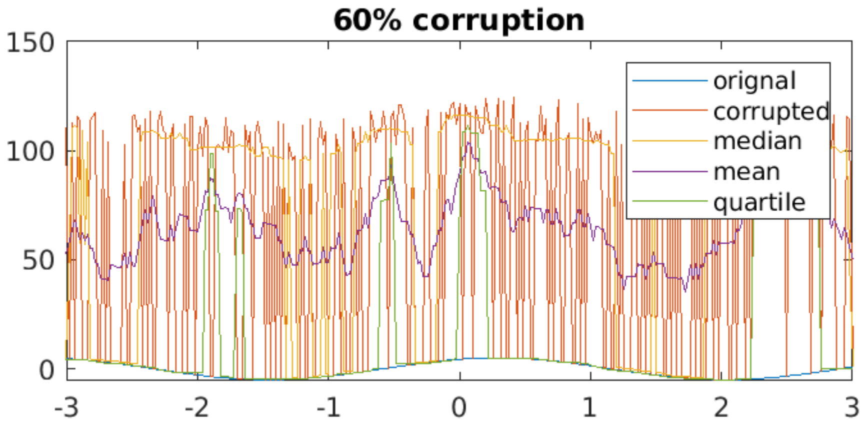
Question 6

The script for this question is under the file named `q6.m`. Please note that the relative mean squared error values listed below are averaged from multiple runs of `q6.m` but the plots are representative of only one run. The relative mean squared error values in the case of 30% and 60% corruption are given below respectively :

Moving Median Filtering 42.10
Moving Mean Filtering 103.81
Moving Quartile Filtering 0.01



Moving Median Filtering 744.02
Moving Average Filtering 376.55
Moving Quartile Filtering 76.17



In both the cases **moving quartile filtering** gives the least relative mean squared error. This can be explained as follows:

Lets choose a random x point from the set (-3 : 0.02 : 3). Let the set to be considered for filtering be $S = \{x_{-8}, x_{-7} \ldots x_0 \ldots x_7, x_8\}$ where $x_0 = x$ and $x_i > x_j$ for $i > j$. On an average if the probability of corruption is p we can expect $p\,|S|$ values to be corrupted. All the corrupted values are necessarily greater than the uncorrupted ones. And as we consider p values, here, to be 0.3 or 0.6 which is sufficiently greater than 0.75 we can expect the 25th percentile to be in S and hence close to x.
The average will be in between the the corrupted and uncorrupted (hence not close to x) and as the median is just the 50th percentile, chance of it being in S is lower in the case of $p = 0.3$ and much lower when $p = 0.6$.

Question 7

The script of this question is under the file named `q7.m`

$$\begin{aligned}
\mu &= Old\ Mean \\
\mu_n &= New\ Mean \\
\sigma &= Old\ Standard\ Deviation \\
\sigma_n &= New\ Standard\ Deviation \\
X &= Old\ Median \\
X_n &= New\ Median \\
A_i &= Elemnts\ of\ array \\
b &= New\ element \\
n &= Size\ of\ old\ array
\end{aligned}$$

1.

$$\sum_i A_i = \mu \times n$$
$$\mu_n = \frac{\sum_i A_i + b}{n+1}$$
$$\mu_n = \frac{(\mu \times n) + b}{n+1}$$

2. - **n is odd.**

     if b is less than the element left to median then the new median will lie betwee old median and number left to it.

     if b lies betwee number left and right to old median then new median will lie between old median and added number.

     if b is greater than the number right to meadian the new median will lie between old median and number right to it.

   - **n is even**

     lets say the old median lies between i and j.

     if $b < i$, $X_n = i$

     else if $b > j, X_n = j$

     else $X_n = b$

3.

$$
\begin{aligned}
\sigma_n^2 &= \frac{\sum_i^n (A_i-\mu)^2 + (b-\mu)^2}{n} \\
n \cdot \sigma_n^2 &= \sum_i^n (A_i-\mu)^2 + (b-\mu)^2 \\
Now, \\
\sum_i^n A_i^2 &= (n-1)\sigma^2 + 2n\mu^2 + -nx^2 \\
Using\ above\ result \\
n \cdot \sigma_n^2 &= (n-1)\sigma^2 + n\mu^2 + n\mu_n^2 - 2n\mu\mu_n + (b-\mu_n)^2 \\
&= (n-1)\sigma^2 + b^2 + n\mu^2 + (n+1)\mu_n^2 - 2n\mu\mu_n - 2b\mu_n \\
Using\ result\ of\ first\ qsn \\
n \cdot \sigma_n^2 &= (n-1)\sigma^2 + b^2 + n\mu^2 - (n+1)\mu_n^2
\end{aligned}
$$

4. We will increase the frequency of the bin which will contain the new element by 1.