

CS 215 - Assignment 1

Niraj Mahajan
180050069

Raaghav Raaj
180050082

August 16, 2019

- 1 Question 1
- 2 Question 2
- 3 Question 3
- 4 Question 4

5 Question 5

A sine wave of the form $y = 5\sin(2.2x + \pi/3)$ is plotted, corrupted and filtered by three methods - moving median filtering, moving average filtering and moving quartile filtering and the **root mean squared errors** of the outputs of each method are :

- 30 percent data corrupted
 1. **Moving Median Filtering** = 28.121810
 2. **Moving Mean Filtering** = 99.607732
 3. **Moving Quartile Filtering** = 0.017173
- 60 percent data corrupted
 1. **Moving Median Filtering** = 692.400109
 2. **Moving Mean Filtering** = 351.490789
 3. **Moving Quartile Filtering** = 59.808615

The clean sine wave, corrupted wave, and the filtered waves using all three methods are plotted in Figure 1 and Figure 2.

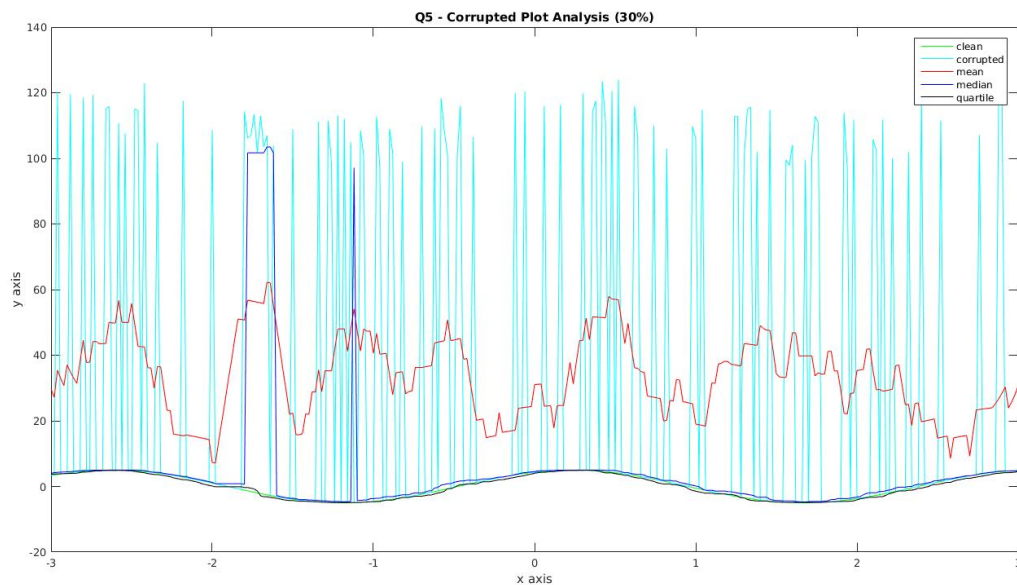


Figure 1: 30% data corrupted

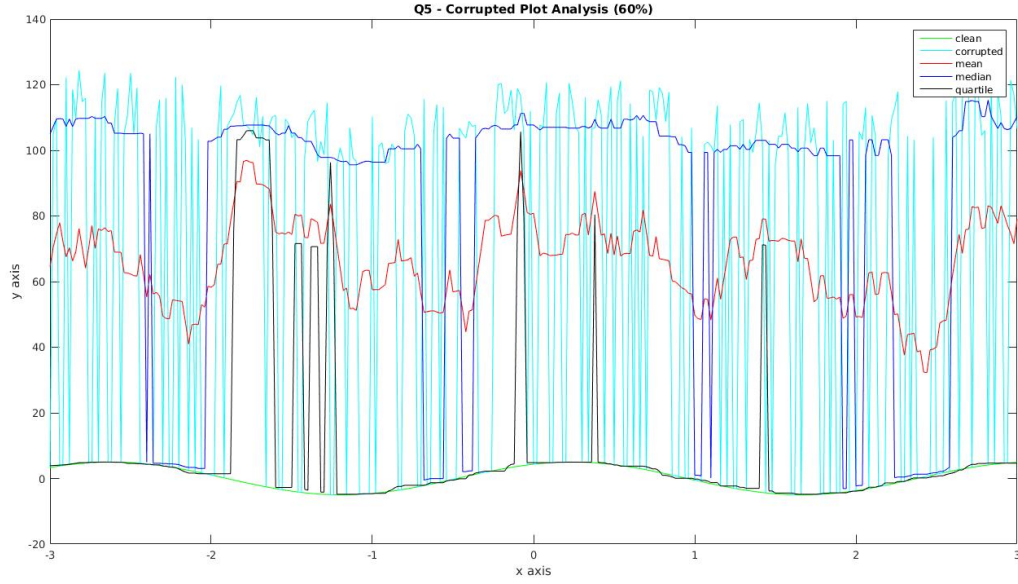


Figure 2: 60% data corrupted

From the plots and the root mean square error data, it is clearly evident that the **moving quartile filtering** method has the **least root mean square error** and, hence, is the **best** method of filtering corrupted data.

This is because since the data is increased by around 100 when it is corrupted, we need to consider the smallest values in any interval while filtering.

Mean of a data considers even the **corrupted values** and is heavily **influenced** by them. Hence we can safely rule out 'mean' as an optimum method for filtering.

Median of a data considers the middle term and is not influenced as heavily as mean by corruption but still, if **more than 50 %** of the data is corrupted then the median will also be a **corrupted value**, and so we can rule out median as well.

Quartile of a dataset focusses on the lower 25% data by value and we can say that it will **prioritize the clean (non corrupted) data** in any interval and hence will not get influenced by the corrupted data until more than 75% data in an interval has been corrupted.

Hence moving quartile filtering is the best way to filter our data.

6 Question 6

7 Question 7

Formula

The number of people 'n' such that any two of them have the same birthday with atleast 'p' probability is given by the inequality

$$\left(1 - \frac{365C_n}{365^n}\right) \leq p$$

The situation can be visualized as the negation of the event where all n people have unique birthdays.

The plot of values of n vs p is given in Figure 3

$$\forall p \in \{5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 95, 99, 99.99, 99.9999, 100\}$$

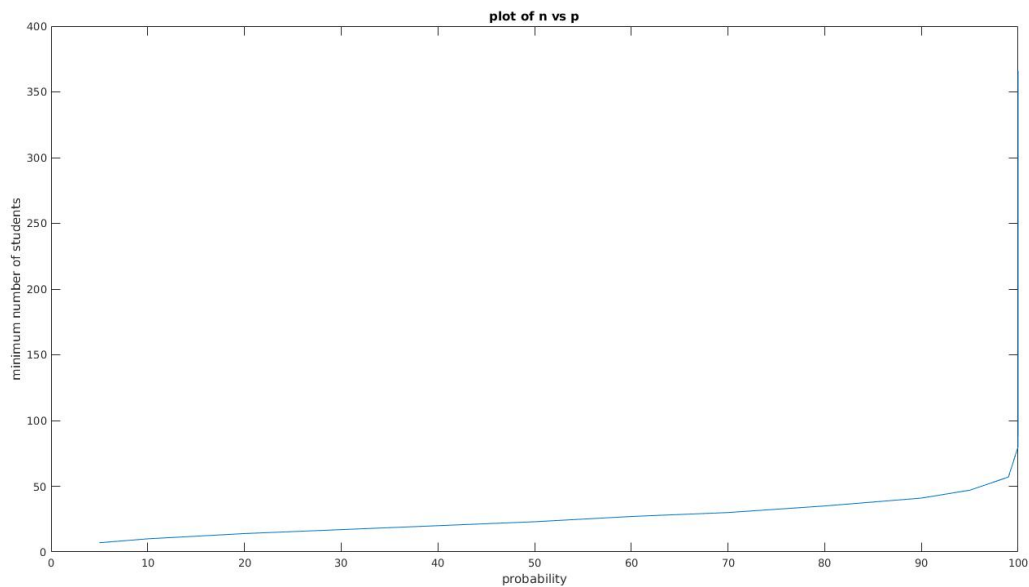


Figure 3: plot of n vs p

Usage of the MATLAB Code

- Load code in the following path 'matlab_code/q5/q5.m'
- Run the code
- This should output all values of n for corresponding values of p
- This should also plot the required plot of n vs p