DATA ANALYSIS PROJECT

RAVURU CHIDAKSH (200010046)
200010046@iitdh.ac.in

IIT Dharwad

September 20, 2021

Contents

1 Approach towards Problem 1

1.1 Calculating mean and varience:

From the given dataset we know that first 1000 data were mixed with the value of random variables which we wantedly assigned them to 0. So, by calculating the mean of the first 1000 values in the dataset we get the mean of the noise data. We can also calculate varience for the first 1000 samples in our dataset.

So now we know the mean and varience of noise , we can calculate mean for the remaining data say corrupted_toss_data and also the varience of the same data.

1.2 Code explained briefly:(explanation is written as comments)

```python
import numpy as np
import pandas as pd
#we are reading data from excel file using read_excel function into a dataframe or a series object.
series =  pd.read_excel("comp_1_200010046.xls",header=None)
#we know first 1000 data are from noise combined with random variable (whose value is 0).
#hence we can charecterize the noise using first 1000 values of the series object.
noise = series.values[:1000]
# from the above noise we can find the mean and all required to charecterize the noise.
currupted_toss_data = series.values[1001:]
#which contains noise with the value of random variable
# so we are basically finding expected value of the random variable by ,
#E[X] = E[currupted_toss_data] - E[noise]
#we can find bias by dividing by,
#E[X] = (0* num_tails + 5*num_heads)/10000
#bias = num_heads/10000
bias = (np.mean(currupted_toss_data) - np.mean(noise))/5
# same as E[X] = 0*(1-p) + 5*p; (where p = bias)
#please check the .ipynb file for the clear and complete code
```

Listing 1: Calculating bias

2 Calculating Error from the calculated bias:

We discuss about the error because sometimes even when the get a toss and transmit 0, the receiver might receive that signal as head due to noise.

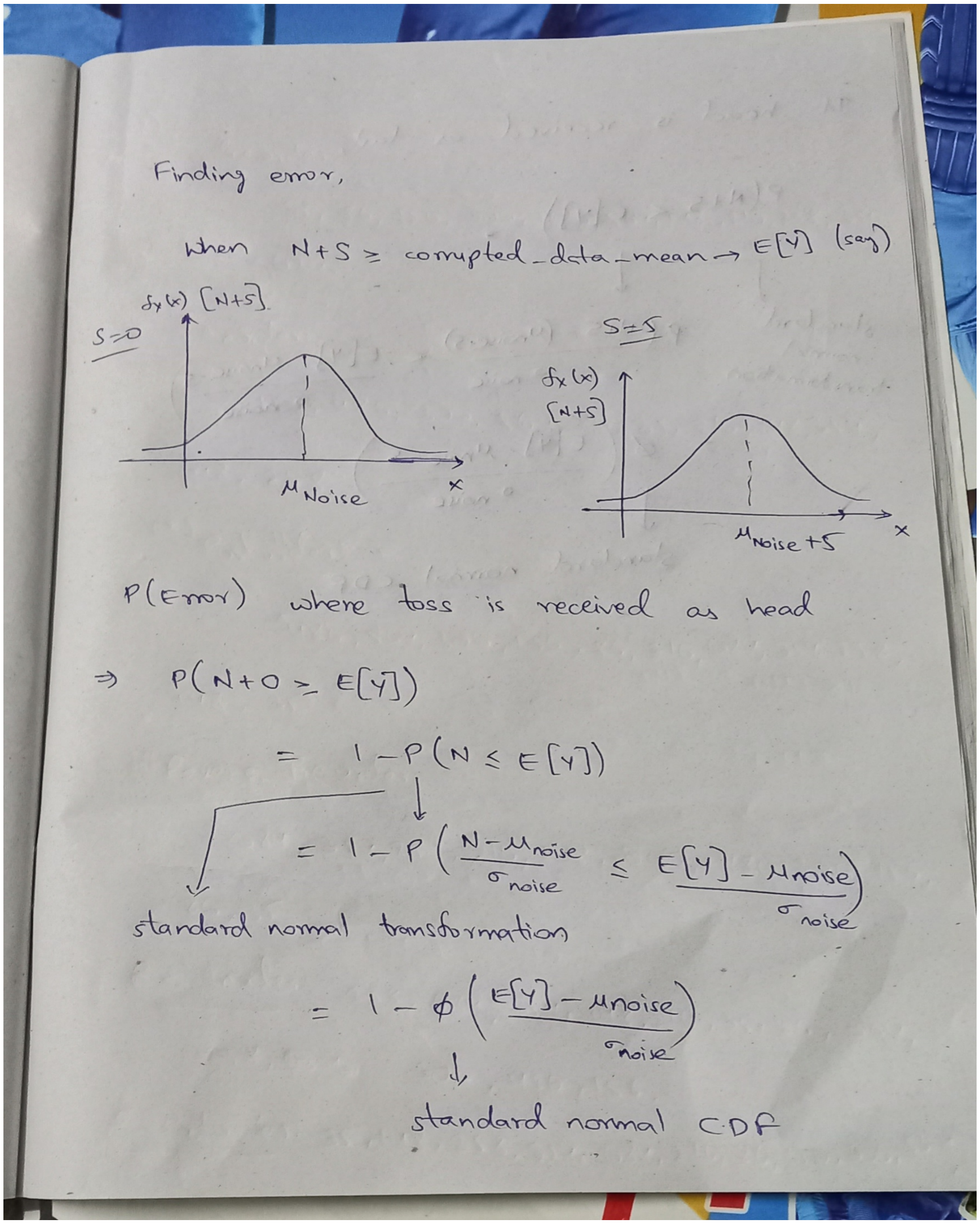
We get error in two cases when we toss a head but receiver receives as a tail and vice-versa. So, we need to calculate both the errors and we can do that with the given information that thy noise is guassian.

Since when we send a tail we get N+S (where S is signal) as the same guassian distribution with the expected value μ (say μ is the expected value of noise) and when $S = 5$ the expected value will be $\mu + 5$ but still remains guassian.

Hence, we can treat the recieving signal as head if $N+S >= E[currupted\_toss\_data]$ as a head and the data where $N+S < E[currupted\_toss\_data]$ as tail.(corrupted_toss_data is defined in section1.2)

Since we know the guassian graph , it's like finding area after the certain point and adding it up with the help of $\phi - table$.

3 Images explaining calculation of error



Finding error,

When $N+S \geq$ corrupted_data_mean $\rightarrow E[Y]$ (say)

S=0
$f_Y(x)$ [N+S]
μ_{Noise}
x
S=5
$f_X(x)$
[N+S]
$\mu_{Noise}+5$
x

P(Error) where toss is received as head

$$\Rightarrow P(N+0 \geq E[Y])$$

$$= 1 - P(N \leq E[Y])$$

$$\downarrow$$

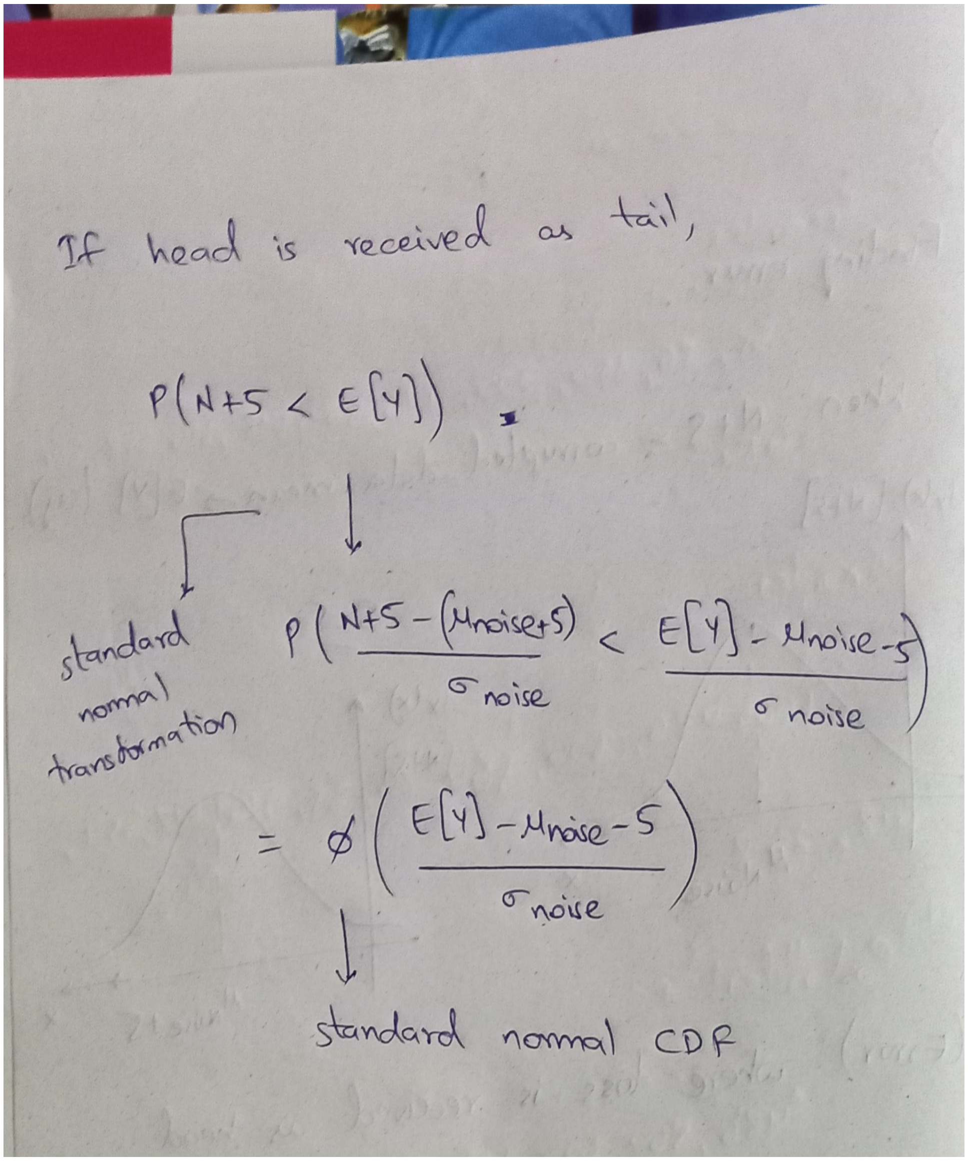
$$= 1 - P\left(\frac{N-\mu_{noise}}{\sigma_{noise}} \leq \frac{E[Y]-\mu_{noise}}{\sigma_{noise}}\right)$$

standard normal transformation

$$= 1 - \phi\left(\frac{E[Y]-\mu_{noise}}{\sigma_{noise}}\right)$$

$$\downarrow$$

standard normal CDF



If head is received as tail,

$$P(N+5 < E[Y])$$

standard normal transformation

$$P\left(\frac{N+5-(\mu_{noise}+5)}{\sigma_{noise}} < \frac{E[Y]-\mu_{noise}-5}{\sigma_{noise}}\right)$$

$$= \phi\left(\frac{E[Y]-\mu_{noise}-5}{\sigma_{noise}}\right)$$

standard normal CDF

And the total error will be sum of both the errors , and with the values calculated from the dataset we can substitute the values and get the total error and error percentage. In my case the net error was around 0.0018. Around 0.18%.

4 Note:

- I wrote the code in python in Google Colab , which i'm submitting with extension ”ipynb”. So, to run the entire file , please upload the file on drive and open it with google colabaratory.
- I worked with .xls file in the entire process which i'm going to zip it with this pdf file . Please upload that .xls file on Google colab before running the cells.