Data Analytics Assignment - 4

G Ramesh Babu
SR No. 20950

Visual Neuroscience

Abstract

This Report shows the analysis done on Visual Neuroscience as part of Data Analytics Assignment - 4.

1 Implementation Summary

1.1 Importing Data

- Two .csv files are given one of which has the search times and the other has the firing rates.
- Both the files are loaded using pandas dataframes and all the values are converted to float values
- In case of firing rate file, columns [**19**, **21**, **23**, **25**, **27**, **29**] are removed inorder to match the **12** pairs in search times file

1.2 Finding Behavioural Indices, L1 Distances & Relative Entropies

- The **behIndex**() function calculates the behavioural indices by taking the search times data. It takes the mean of each column, subtracts **328ms** from it and returns a list of behavioural indices which are the inverses of the mean search times
- The **l1distance**() function calculates the L1 Distance of each group pair by taking the L1 Norm of both the columns of the group(L & R) and dividing it by the no. of columns. The columns are of different length and the function takes care of it
- The **relentropy**() function calulates the Relative Entropy of every pair of columns using the formula and returns a list of Relative Entropies

1.3 Finding KS statistic

- **AMGM**() function calculates the AM / GM by taking the search times and L1 Distances or Relative Entropies
- α and β of the gamma function are found by using the **findAlpha**() and **findRate**() function using the search times. α is calculated by finding the mean and standard deviation of half of the columns (first 12). β is calculated by using the half of the samples of remaining columns(13 to 24)
- KS Statistic is calculated by using the rest of the samples and by using Empirical CDF and Gamma CDF (by using α and β). **findKSstat**() function takes the Rate parameters, alpha and search times and returns a list of KS statistics found using **scipy**.**stats**()

2 Results

2.1 Question 1

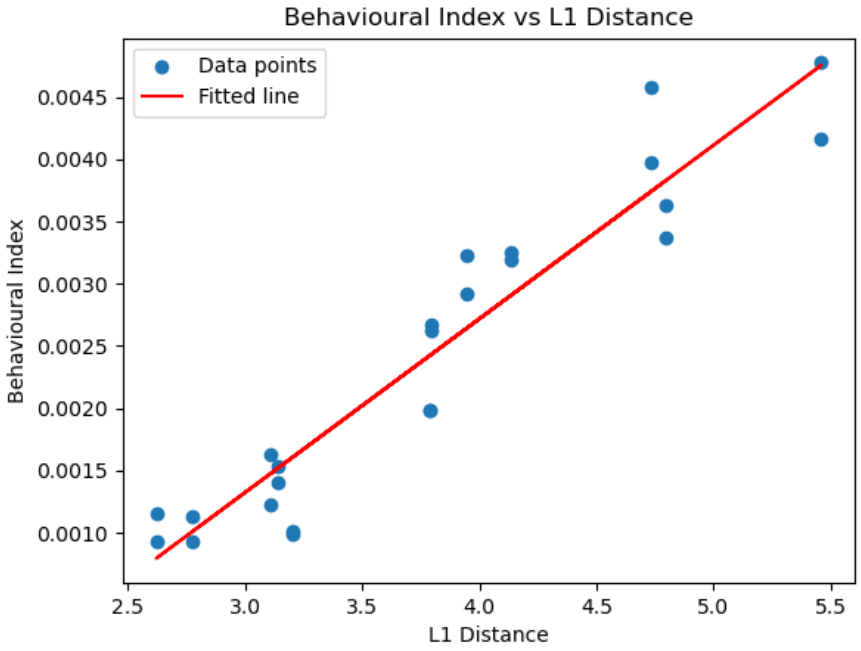


Figure 1: Behavioural Indices vs L1 Distances

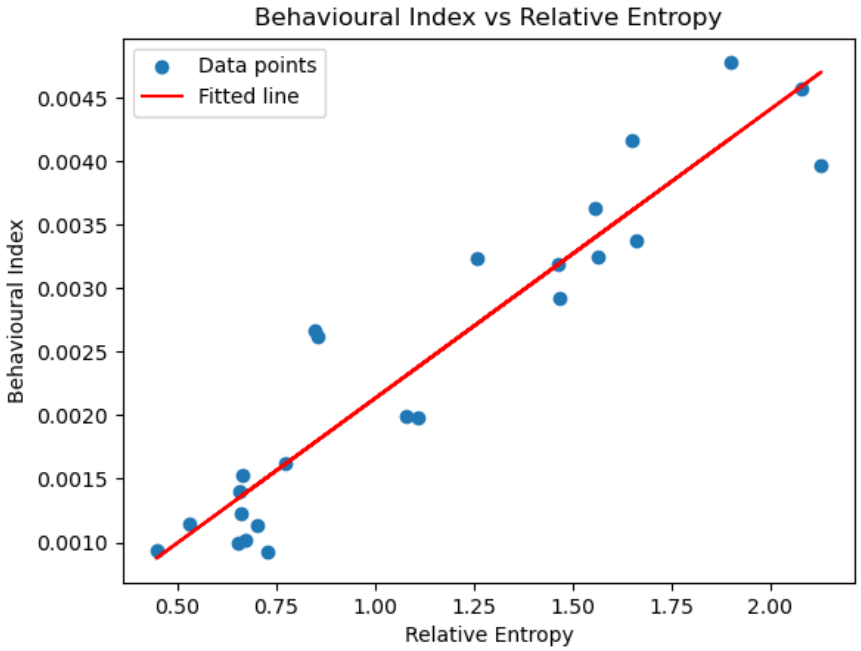


Figure 2: Behavioural Indices vs Relative Entropies

Comparison	R-squared	p-value	Standard Deviation
Behavioural Index vs L1 Distance	0.9049	1.0177×10^{-12}	9.6518×10^{-5}
Behavioural Index vs Relative Entropy	0.8845	8.6434×10^{-12}	0.0002

Table 1: Summary of Regression Results

It can be seen that the R-square value for the regression done using L1 Distance is better than that of using Relative Entropy. The L1 distance is often considered better than relative entropy when a straightforward and intuitive measure of difference between two probability distributions is needed, especially in cases with outliers or non-standard distributions

2.2 Question 2

Comparison	AM / GM Ratio
Search Delay vs L1 Distance	1.065
Search Delay vs Relative Entropy	1.022

Table 2: AM/GM Ratios

The AM/GM ratio for "search delay x L1 distance" can help identify the presence of outliers or extreme differences in L1 distances and provide a measure of spread that is more robust to such variations. This makes the AM/GM ratio a suitable choice for assessing the spread of "search delay x L1 distance" when compared to the AM/GM ratio for "search delay x relative entropy."

2.3 Question 3

2.4 Estimation

- We estimate the parameters for the Gamma Distribution from samples.
- To estimate the shape parameter (α), we randomly selected half of the groups and plotted the standard deviation against means, resulting in an estimated $\alpha \approx \mathbf{2.5871}$.
- The chosen method relies on the well-established principle that the standard deviation-to-mean ratio is linked to the shape parameter of the Gamma distribution. This provides a meaningful and interpretable estimate of α. The approach is statistically robust and provides an unbiased estimate of the shape parameter.
- To estimate the rate parameter (β) of the Gamma distribution, half of the samples are randomly selected from the unutilized groups in the shape parameter estimation and we calculate the means of these samples and estimate the rate parameters by using $\beta = \alpha/\mathbf{mean}$.
- we employed the Maximum Likelihood Estimation (MLE) method. MLE is a well-established technique for parameter estimation in various statistical distributions, including the Gamma distribution.

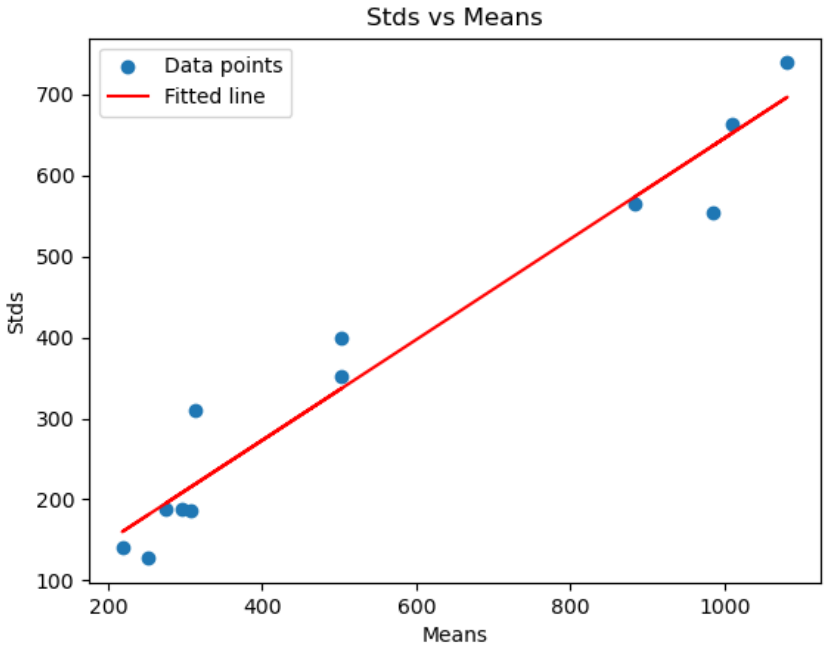


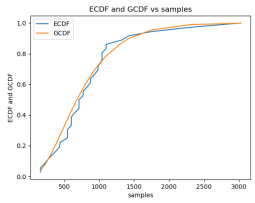
Figure 3: Standard Deviation vs Mean

Parameter	Value
Shape Parameter (α)	2.5871
Rate Parameters	[0.0033, 0.0042, 0.0084, 0.0067, 0.0102, 0.0119, 0.0030, 0.0024, 0.0041, 0.0038, 0.0068, 0.0065]
KS Statistics	[0.1389, 0.0833, 0.1944, 0.3611, 0.2500, 0.1944, 0.2222, 0.1111, 0.0694, 0.1528, 0.1944, 0.1944]
KS p-values	[0.8849, 0.9998, 0.5100, 0.0176, 0.2123, 0.5100, 0.0569, 0.7695, 0.9956, 0.3722, 0.1317, 0.1317]

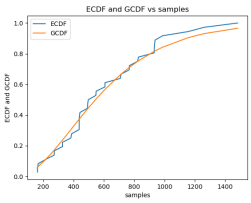
Table 3: Summary of Parameters, KS Statistics, and KS p-values

2.5 Inference

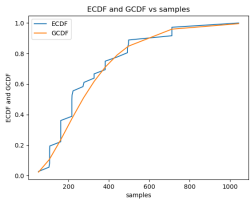
- We can see from the KS pvalues that some of the p-values are > 0.05 and some of them are < 0.05. If the p-value is less than 0.05, it suggests that you have sufficient evidence to reject the null hypothesis. If the p-value is greater than 0.05, it suggests that you do not have sufficient evidence to reject the null hypothesis.
- Hence, in some cases the search delays follow a gamma distribution and in some other cases, they do not follow the gamma distribution



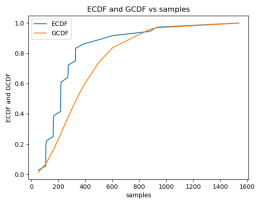
(a) Figure 1



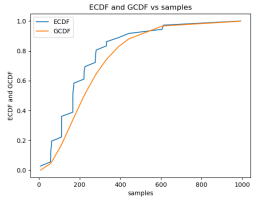
(b) Figure 2



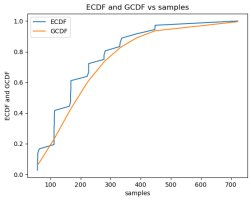
(c) Figure 3



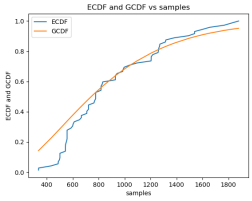
(d) Figure 4



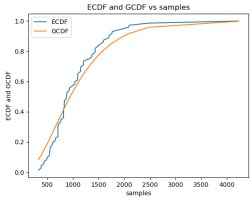
(e) Figure 5



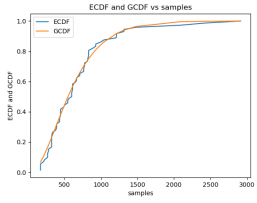
(f) Figure 6



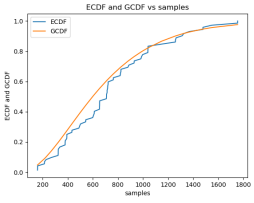
(g) Figure 7



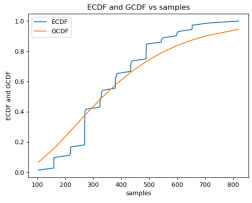
(h) Figure 8



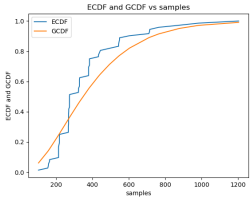
(i) Figure 9



(j) Figure 10



(k) Figure 11



(l) Figure 12

Figure 4: ECDF and GCDF vs samples