Project 2. Necker Cube – Correlation, Recurrence measure of dependence, Statistical analysis

Leyla Aminova, Maxim Rassabin

I. Introduction

In this work we look into correlation, recurrence measure of dependence and statistical analysis methods using the measures electroencephalography of observe the Necker cube with different values of brightness.

A. Pearson correlation coefficient

In statistics, correlation is any statistical relationship, whether causal or not, between two random variables or bivariate data. In the broadest sense correlation is any statistical association, though it commonly refers to the degree to which a pair of variables are linearly related. [3] The Pearson correlation coefficient (PCC) is a statistic that measures linear correlation between two variables X and Y. [4] Pearson correlation coefficient for a pair of random variables (X, Y) is:

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

where:

- μ_X, μ_Y are mean of X, Y;
- σ_X, σ_Y are standard deviation of X, Y;
- E is the expectation.

B. Recurrence measure of dependence

Recurrence-based measure of dependence (RMD) is a connectivity measure. This measure indicates "non-independence" and it's direction obtained from given time series introducing a lag in a possible driving process. RMD determines the presence or absence of a causal relation in a pair of processes in terms of establishment of a functional relationship (linear or nonlinear) between them [1]. And defined as:

$$RMD|_{XY} = log_2 \left(\frac{1}{N} \sum_{i=1}^{N} RMD_i|_{XY} \right)$$

$$RMD_i|_{XY} = \frac{P(E_X(t_i), E_Y(t_i))}{P(E_X(t_i))P(E_Y(t_i))}$$

where N is a length of the time series, $P(E_X(t_i)$ is probability, that calculated as:

$$P(E_X(t_i)) = \frac{1}{N} \sum_{j=1}^{N} R_X(i, j)$$

$$P(E_X(t_i))P(E_Y(t_i)) = \frac{1}{N} \sum_{j=1}^{N} JR(i, j)$$

$$JR = R_X(i, j)R_Y(i, j)$$

where $R_X(i, j)$ are recurrence matrix of $E_X(t)$ time series and JR is joint recurrence matrices of X, Y.

Recurrence matrix can be calculated to each process [5]:

$$R(i, j) = \begin{cases} 1, x_i \approx x_j \\ 0, other \end{cases} \quad i, j = 1, ..., N,$$

where $x_i \approx x_j$ means equality up to an error (or distance) ϵ.

C. Non-parametric test

Non-parametric (or distribution-free) inferential statistical methods are mathematical procedures for statistical hypothesis testing which, unlike parametric statistics, make no assumptions about the probability distributions of the variables being assessed. The non-parametric statistical test is performed in the following way:

1) Collect the trials of the two experimental conditions in a single set.
2) Randomly draw as many trials from this combined data set as there were trials in condition 1 and place those trials into subset 1. Place the remaining trials in subset 2. The result of this procedure is called a random partition.
3) Calculate the test statistic on this random partition.
4) Repeat steps 2 and 3 a large number of times and construct a histogram of the test statistics.
5) From the test statistic that was actually observed and the histogram in step 4, calculate the proportion of random partitions that resulted in a larger test statistic than the observed one. This proportion is called the p-value.
6) If the p-value is smaller than the critical alpha-level (typically, 0.05), then conclude that the data in the two experimental conditions are significantly different.

Non-parametric tests have some distinct advantages over parametric. Outcomes that are ordinal, ranked, subject to outliers or measured imprecisely are difficult to analyze with parametric methods without making major assumptions about their distributions as well as decisions about coding some values (e.g., ”not detected”). Non-parametric tests can also be relatively simple to conduct.

II. Analysis

First of all, before starting analysis was performed filtration to all data in order to extract feature information corresponding to α (8-12 Hz) and β (15-30 Hz) bands. For this propose was used Python script based on ”scipy.signal” module. So the filtrating resut you can see on the fig 1.

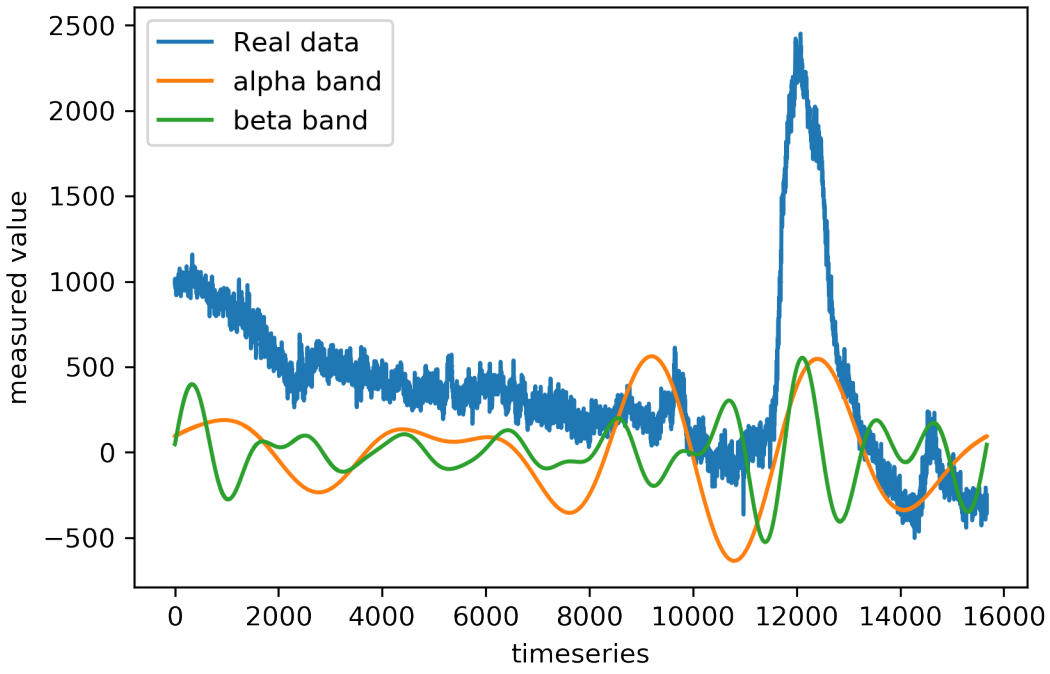
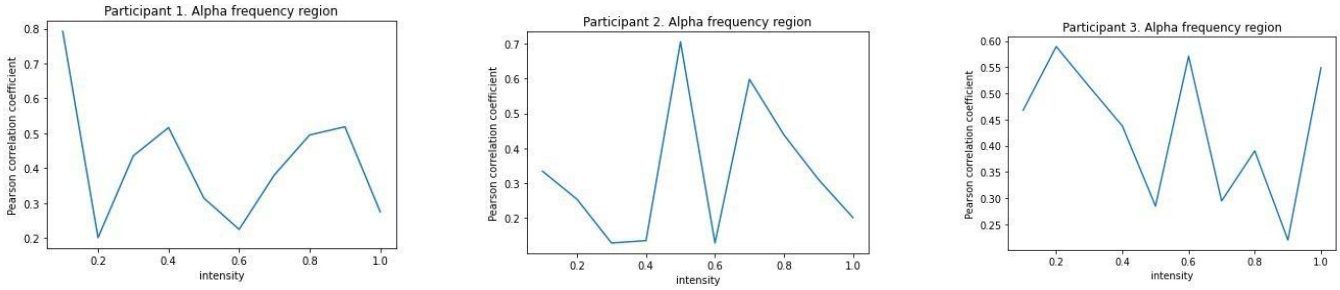


Fig. 1: Bands filtering of data

Then using the scipy.stats.pearsonr module was calculated Pearson coefficient for each pair of feature (from 1 to 31), then was calculated mean value of coefficient corresponding to particular participant and intensity value. Based on it was plot the dependency of coefficient from intensity. Examples of such dependencies on fig 2.



(a) (b) (c)

Fig. 2: Dependencies of Pearson coefficient from intensity value

The next is to calculate Recurrence measure of dependence (RMD). During calculations was considered to take only first 10 sec from measured datasets in terms of computational complexity. The first step of RMD calculations is calculation recurrence matrix. For this propose was used the pyhton pacakage ”pyunicorn” with function ”RecurrencePlot”. So, this opertaion is complex for computation and we have big amount of data, threrfore we create temp files each contains recurrence matrix for each column from each data file. Visualisation of recurrence matrix on the fig. 3.

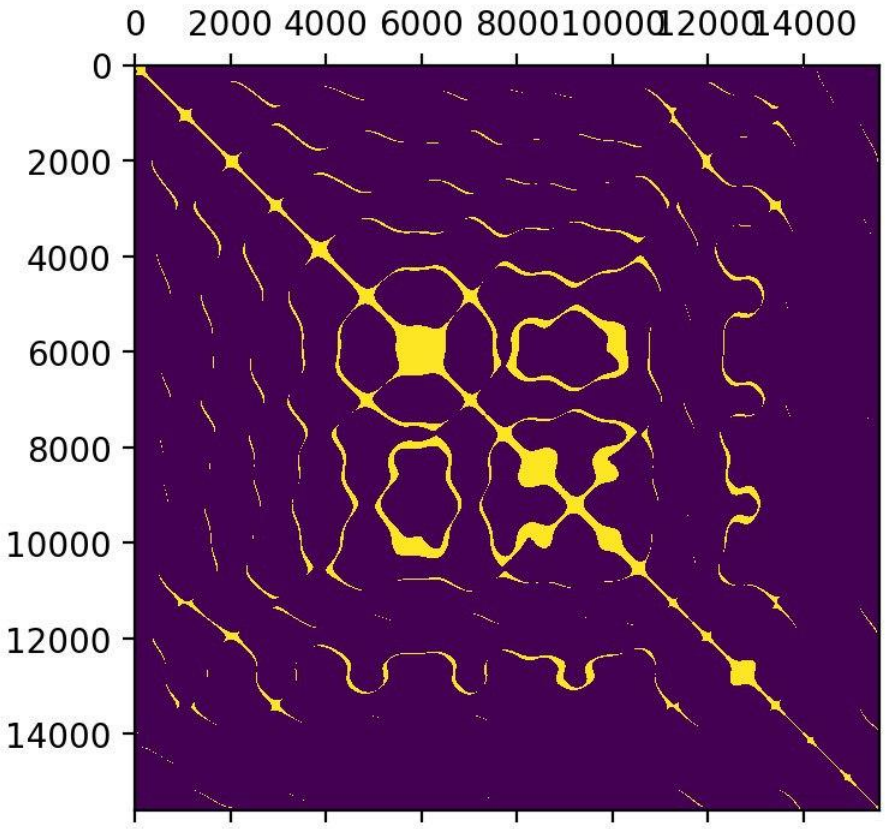
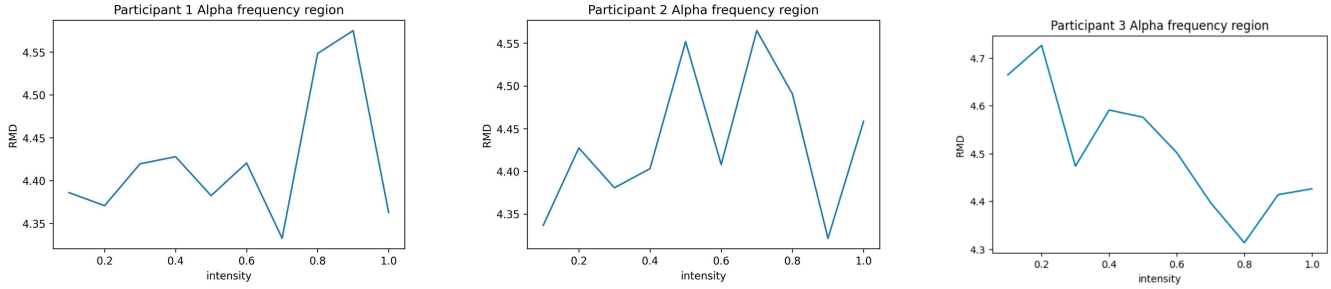


Fig. 3: Recurrence matrix visualization

After it used the temporary files contains recurrence matrix was calculated mean RMD value for each

participant and each band as was done before. The examples of dependency of RMD from intensity value for alpha region on fig 4.



(a) (b) (c)

Fig. 4: Dependencies of RMD coefficient from intensity value

Based on all plots all dependencies and on biggest Pearson coefficient and RMD values was chosen the optimal values of intensity for each coefficient and each Participant. For Pearson coefficient:

- **Participant 1 - 0.1** ;
- **Participant 2 - 0.5** ;
- **Participant 3 - 0.2** ;
- **Participant 4 - 0.2** ;
- **Participant 5 - 0.2** ;

For RMD:

- **Participant 1 - 0.9** ;
- **Participant 2 - 0.5** ;
- **Participant 3 - 0.2** ;
- **Participant 4 - 0.7** ;
- **Participant 5 - 0.1** ;

III. Estimating efficiency of Pearson correlation and RMD

As was mentioned in [1] the high image ambiguity and high task complexity accounted for intensity range 0.4-0.6 for Necker cube task. So, therefore was assumed that the maximum mean connectivity should accounted for this range. As was described in previous section the maximum values for Pearson correlation coefficient comes on this range only for Participant 2.Overall for α frequency region the major mean connectivity does not have a pronounced severity for the average of participants (only for Participant 2). For β region this region have a pronounced severity for Participant 2 and 3 (not so high).

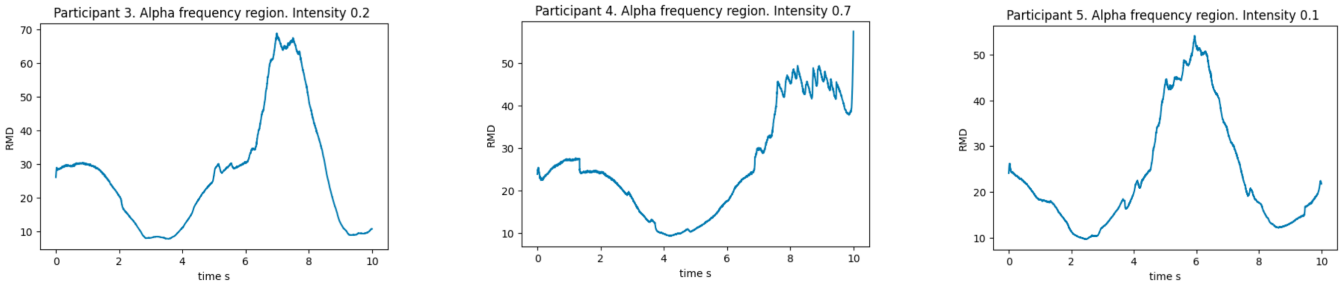
For RMD's characteristics the maximum values comes from high task complexity range only for Participant 2. Overall for α region mean best connectivity come to this range for 2, 3 and 4 Participant. And for β region only for Participant 2.

Based on the result that was described above, Pearson correlation and RMD correlation shows same result for Participant 2 for each frequency region, this shows that they are similar in highlighting the biggest connectivity between brain regions for the experiment under consideration. Because the most obvious brain connectivity in the most difficult task is shown precisely by participant 2. But in general,

for all participants each methods do not show such connectivity for 0.4-0.6 region intensity. However, RMD metric shown more valid on alpha frequency (the majority number of participants).

IV. Coupling strength

Investigation the time-dependence of the coupling strength based on Person correlation coefficient can not be held, because it is taken over the entire timeline. But RMD can help in this investigation. To find the optimal value of the time window we use optimal value of intensity for each Participant found in section II.



(a) (b) (c)

Fig. 5: Dependencies of RMD from time

We can see in fig. 5 the optimal time window is from 6s to 8s. And as RMD represents a relevant measure of the coupling strength between two process [1], we can say, that the value of coupling strength is in range from 50 to 70.

V. Non-parametric Test

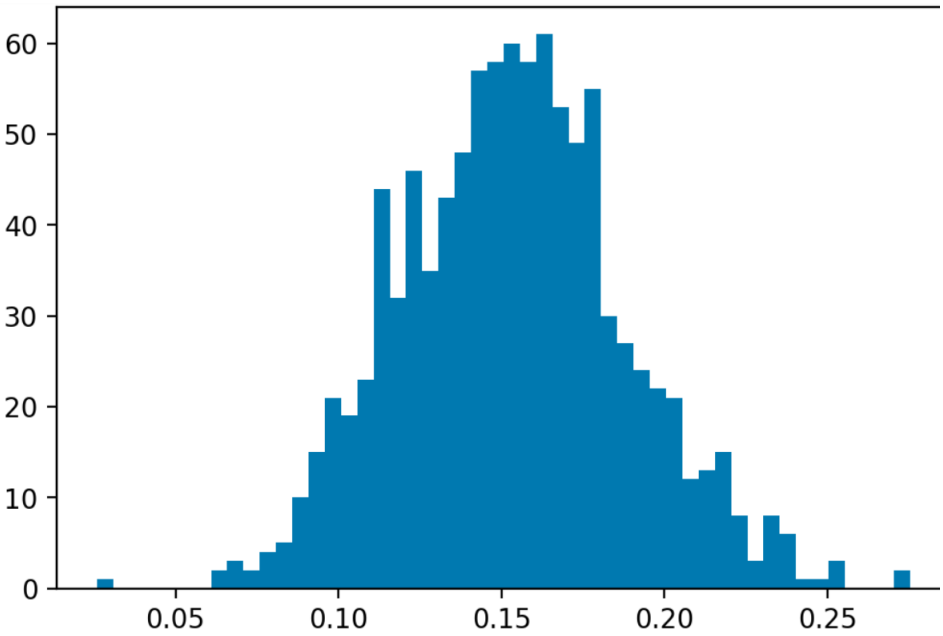


Fig. 6: Histogram of RMD

As was was described in previous most pronounced demonstration of connectivity in more complex intensity range shown in Participant 2, for intensity value 0.5. Also, in [1] described that during Necker cube experiment one of the biggest value of connectivity shows pair of Brain zone O1 - FP2. Sp, for non-parametric test was chosen Participant 2 intensity 0.5 and 02-09 feature data. Based on this data was calculated the RMD's histogram, fig. 6.

For histogram plotting was performed random partitions RMD calculations with 1000 repeating. So, for MVC problem [2] propose the threshold α level calculate as:

$$\alpha = 0.05/N$$

where N is equal number of time samples. For our case $\alpha = 0.000025$

Calculated ρ metric equal 0.0321 that significantly more than α level. Therefore, non-parametric test proved that there is a connectivity level is big enough between the parameters under consideration.

REFERENCES

[1] Neural Interactions in a Spatially-Distributed Cortical Network During Perceptual Decision-Making Maksimenko, Vladimir A. and Frolov, Nikita S. and Hramov, Alexander E. and Runnova, Anastasia E. and Grubov, Vadim V. and Kurths, Jürgen and Pisarchik, Alexander N. 2019

[2] Nonparametric statistical testing of EEG- and MEG-data Eric Maris , Robert Oostenveld 2017

[3] en.wikipedia.org/wiki/Correlation_and_dependence

[4] en.wikipedia.org/wiki/Pearson_correlation_coefficient

[5] Recurrence plots for the analysis of complex systems. Marwan N., Carmenromano M., Thiel M., Kurths J. 2007