IL 2223 Embedded Intelligence
Lab 1
Time Series Visualization and Feature Extraction

Rui Shi
srui@kth.se

April 5, 2022

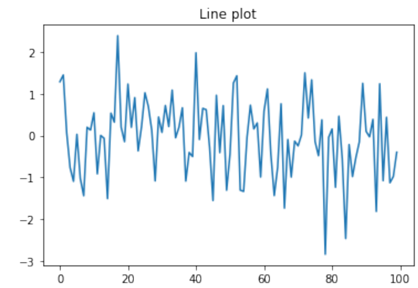
1 Task 1. Exploratory Data Analysis

In this task, I learnt to generate white noise series, random-walk series, find actual mean, standard deviation and draw line plot, histogram, density plot, box plot, lag-1 plot, ACF and PACF graphs of them. The series are generated from **pandas.series**, and the graphs are plotted by **pandas.series.plot** and **statsmodels**.

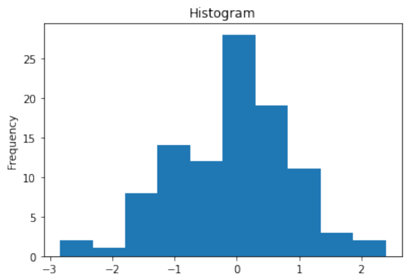
1.1 Task 1.1 White noise series

1.1.1 Features and plots of white noise series

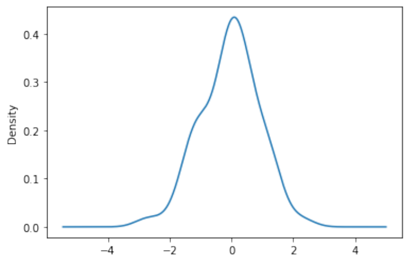
In this section, I generate a white noise series with N data points. Here are the features and plots of the series.



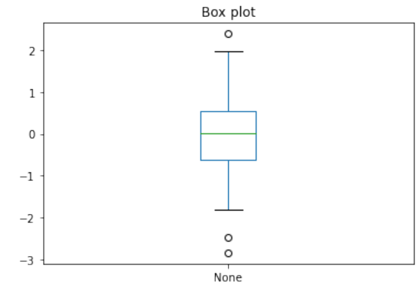
(a) Line plot for white noise



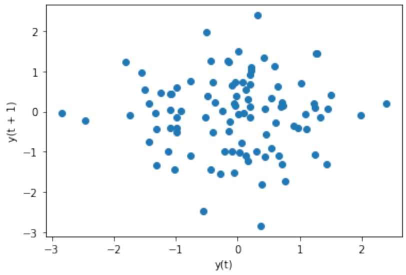
(b) Histogram for white noise



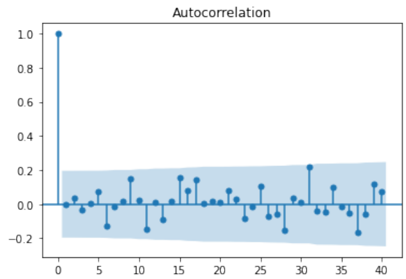
(c) Density plot for white noise



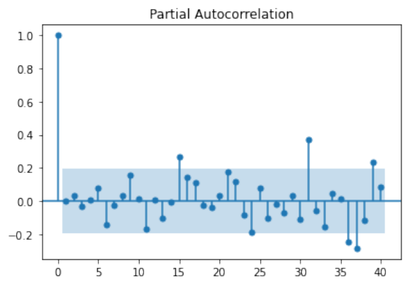
(d) Box plot for white noise



(e) Lag-1 plot for white noise



(f) ACF for white noise



(g) PACF for white noise

Figure 1: Graphs for white noise

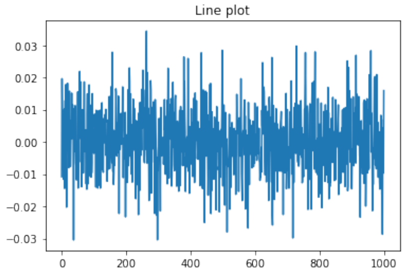
Table 1: Features of white noise

Data	Mean	Standard Deviation
White noise	-0.050576379210951254	0.9440133469488994

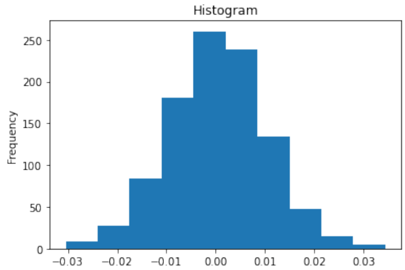
It can be observed that the standard deviation is close to 1, as I set before.

1.1.2 Features and plots of mean value of white noise series

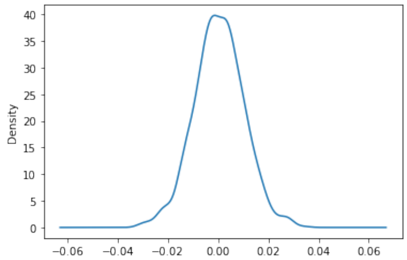
In this section, I generate 100 random series with length 1000 data points, then use the average values at each time to produce an average value series. Then repeat the same process above.



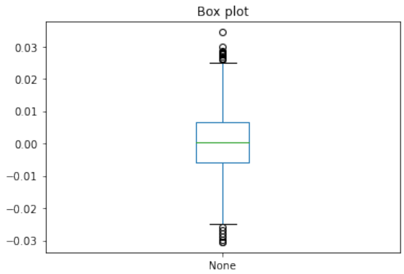
(a) Line plot for mean value of white noise



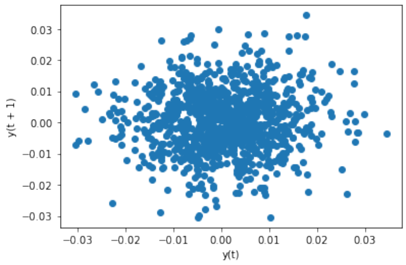
(b) Histogram for mean value of white noise



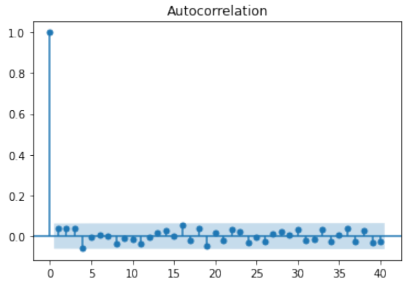
(c) Density plot for mean value of white noise



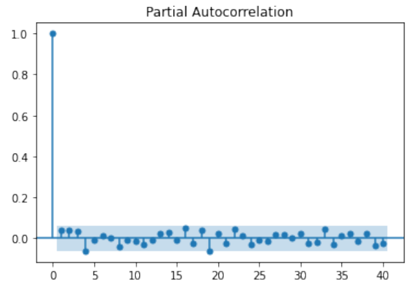
(d) Box plot for mean value of white noise



(e) Lag-1 plot for mean value of white noise



(f) ACF for mean value of white noise



(g) PACF for mean value of white noise

Figure 2: Graphs for mean value of white noise

Table 2: Features of mean value of white noise series

Data	Mean	Standard Deviation
White noise	0.00041705181698558112	0.009882040818504639

For this new time series, the standard deviation is close to 0.01.

1.1.3 Perform randomness test and stationarity test on the white noise series using the Ljung-Box test and ADF test

Table 3: Randomness test and stationarity test on the white noise

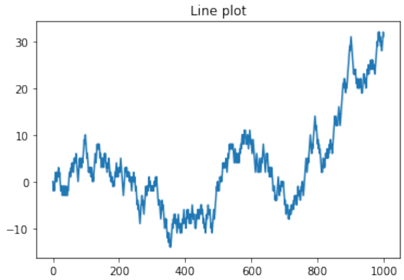
Data	lb_pvaluet	p_value
Random-walk series	0.837929	0.000000

As the lb_pvalue ¿ 0.05, then accept the null hypothesis, meaning that the series are random. p_value ¡¡ 0, then reject the null hypothesis and judge that the series is stationary.

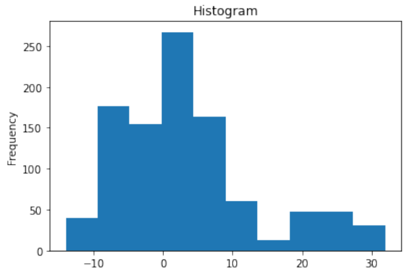
1.2 Task 1.2 Random-walk series

In this section, I define a random walk series. The current value depends on the previous value.

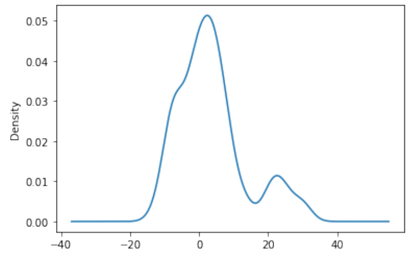
1.2.1 Find its actual mean, standard deviation, and draw its line plot, histogram, density plot, box plot, lag-1 plot, ACF and PACF graphs (lags up to 40).



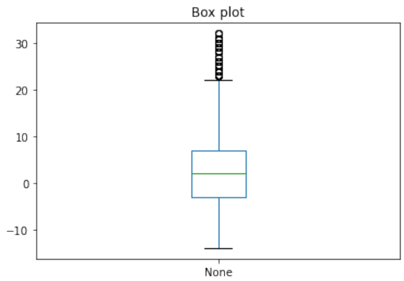
(a) Line plot of random walk series



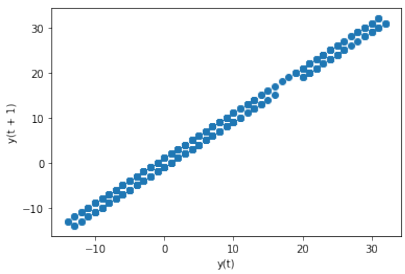
(b) Histogram of random walk series



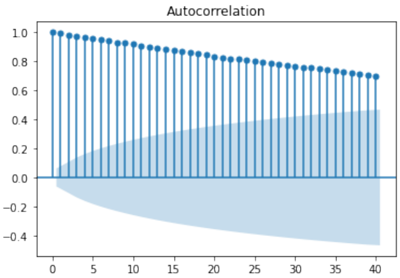
(c) Density plot for of random walk series



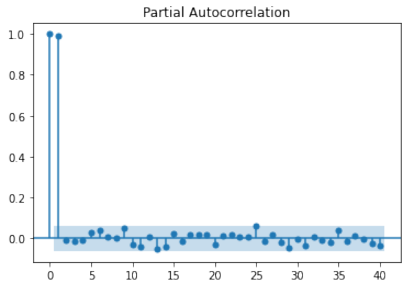
(d) Box plot of random walk series



(e) Lag-1 plot of random walk series



(f) ACF of random walk series



(g) PACF of random walk series

Figure 3: Graphs for random walk series

Table 4: Features of random-walk series

Data	Mean	Standard Deviation
White noise	3.495	9.866066859456973

1.2.2 Perform randomness test and stationarity test on the random-walk noise series using the Ljung-Box test and ADF test

Table 5: randomness test and stationarity test on the random-walk noise series

Data	lb_pvaluet	p_value
Random-walk series	3.524787e-216	0.908966

As the lb_pvalue << 0.05, then reject the null hypothesis, meaning that the series is not random. p_value > 0.5, then reject the null hypothesis and judge that the series is not stationary.

By generating a one order difference, the new series will be stationary.

1.2.3 Answer for the questions

1. What methods can be used to check if a series is random? Describe both visualization and statistic test methods.

   Answer: For visualization methods, we can check the trend in line plot, data distribution in histogram and density plot. A useful method is to check log plot, ACF graph and PACF graph.

For statistic test methods, we could use Ljung-Box to test if the residual series after prediction is random noise. If lb_pvalue > 0.05, than accept the Null hypothesis that the series is random, vice versa.

2. What methods can be used to check if a series is stationary? Describe both visualization and statistic test methods. Answer: For visualization test methods, we can check the trend in line plot. We can also use Augmented Dickey Fuller test. If p_value is lower than the critical value which equals 0.05, reject the null hypothesis that the series is not stationary, and consider that the time serious is stationary.

3. Why is white noise important for time-series prediction?

   Answer: Fist of all, white noise cannot be predicted. Secondly, the remainder series (forecast errors) from a prediction model should ideally be white noise.

4. What is the difference between a white noise series and a random walk series?

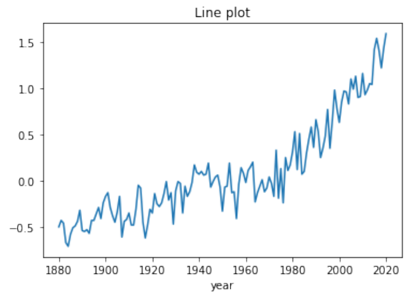
   Answer: White noise series is stationary, while random walk series is not. It depends on the previous value.

5. Is it possible to change a random walk series into a series without correlation across its values ? If so, how? Explain also why it can.

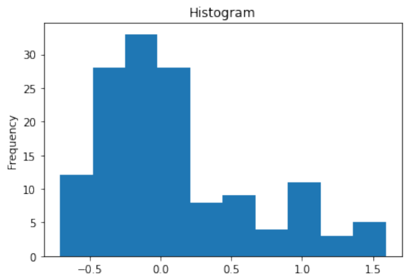
   Answer: I think it is possible. We can generate a new series that is the first order difference of random walk series. Than the new series is totally random. Notice that two or more order differencing also works.

1.3 Task 1.3 Global land temperature anomalies series

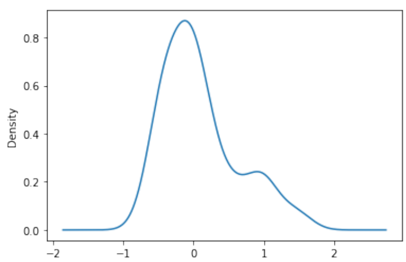
1.3.1 Plots of Global Land Temperature Anomalies Series



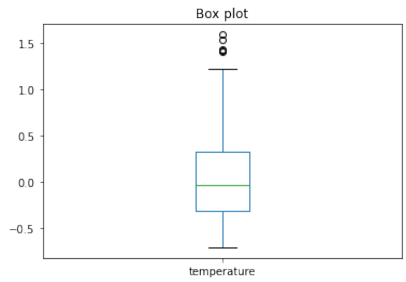
(a) Line plot for global land temperature anomalies series



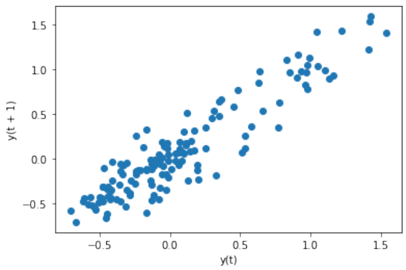
(b) Histogram for global land temperature anomalies series



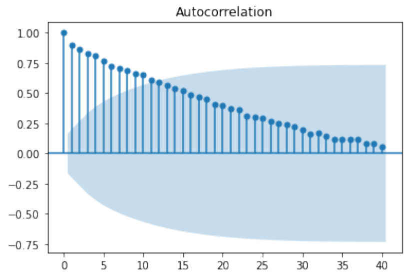
(c) Density plot for global land temperature anomalies series



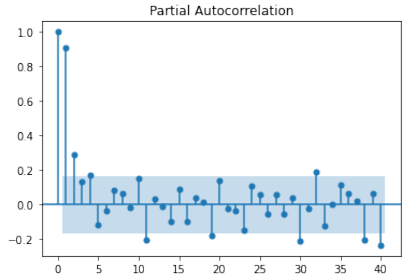
(d) Box plot for global land temperature anomalies series



(e) Lag-1 plot for global land temperature anomalies series



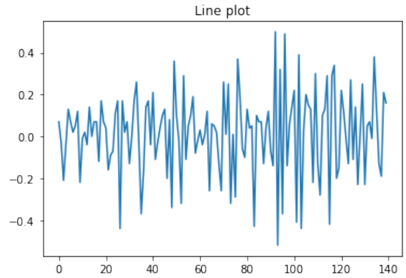
(f) ACF for global land temperature anomalies series



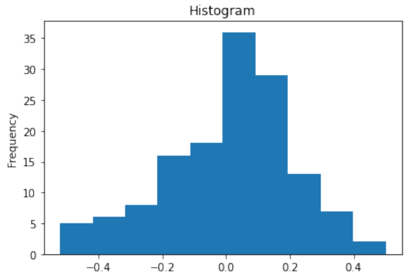
(g) PACF for global land temperature anomalies series

Figure 4: Graphs for global land temperature anomalies series

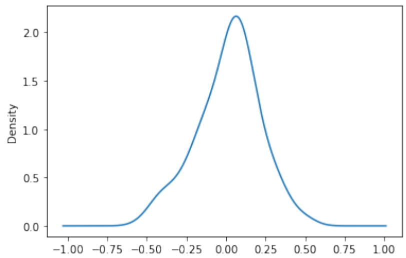
1.3.2 Plots of the First Order Difference of the Dataset



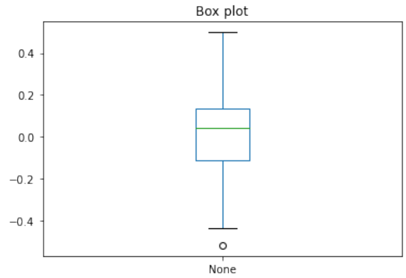
(a) Line plot for first order difference of the dataset



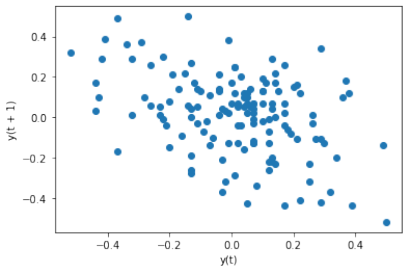
(b) Histogram for first order difference of the dataset



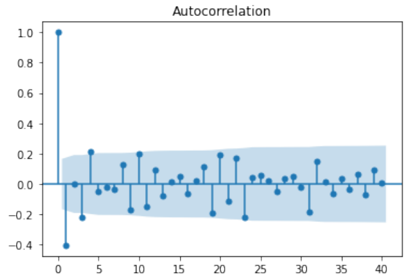
(c) Density plot for first order difference of the dataset



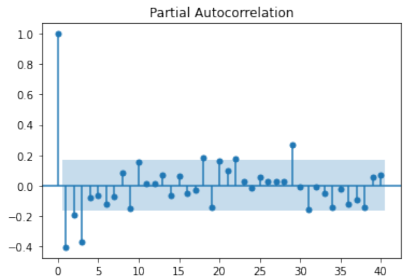
(d) Box plot for first order difference of the dataset



(e) Lag-1 plot for first order difference of the dataset



(f) ACF for first order difference of the dataset



(g) PACF for first order difference of the dataset

Figure 5: Graphs for first order difference of the dataset

1.3.3 Perform randomness test and stationarity test on the temperature anomaly dataset

Table 6: randomness test and stationarity test on the temperature anomaly dataset

Data	lb_pvalue	p_value	Random	Stationary
Original	4.169678e-27	0.938293	not random	not stationary
Differenced	0.000001	0.000000	not random	stationary

Both the original and differenced series are not random. But the original series is not stationary, the differenced series is stationary.

1.3.4 Performance the classical decomposition and STL decomposition on the dataset

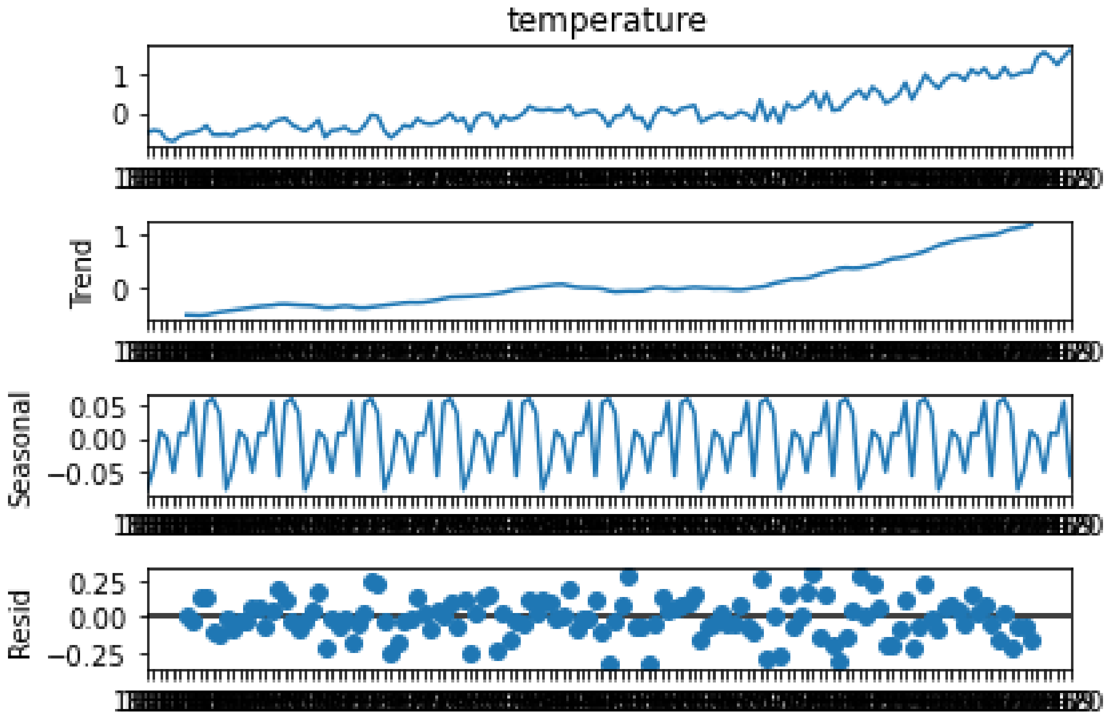


Figure 6: classical decomposition

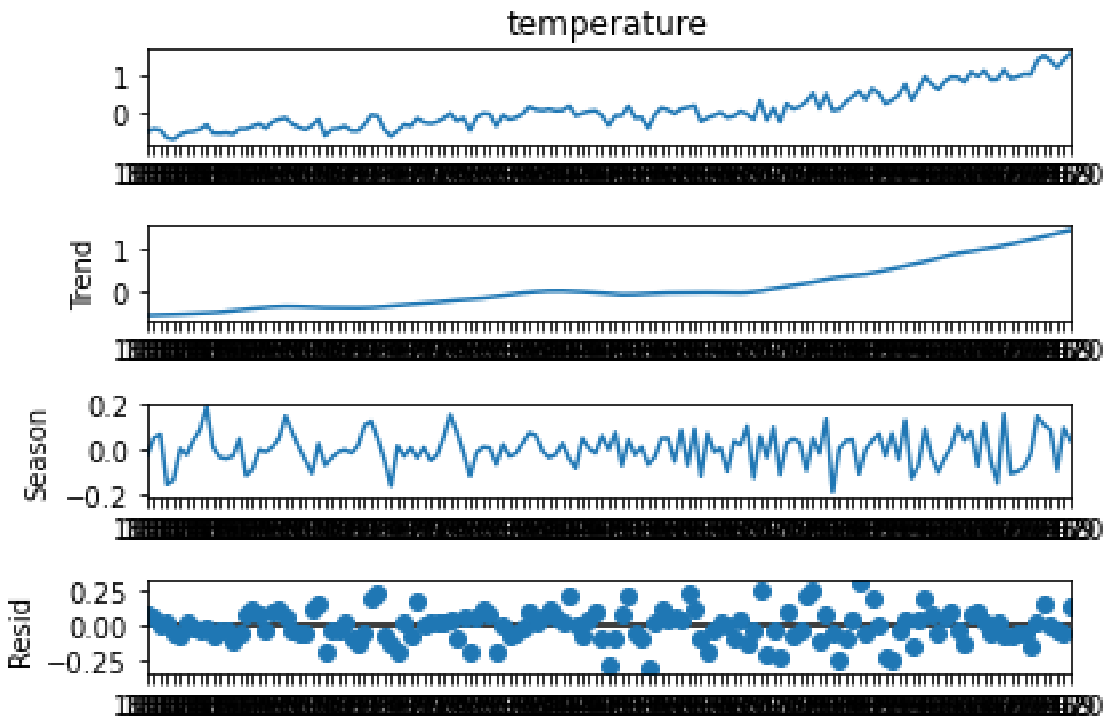


Figure 7: STL decomposition

1.3.5 Answers to the questions

1. What is a stationary time series?

Answer: A stationary time series is one whose properties do not depend on the time at which the series is observed.

2. If a series is not stationary, is it possible to transform it into a stationary one? If so, give one technique to do it?

   Answer: Yes, by differencing. Also it could be decomposed, and the seasonal component could be stationary.

3. Is the global land temperature anomaly series stationary? Why or why not?

   Answer: No, because p_value is larger than 0.05, then we can accept the null hypothesis and judge that the series is not stationary.

4. Is the data set after the first-order difference stationary?

   Answer: Yes, because p_value is less than 0.05, then we can reject the null hypothesis and judge that the series is stationary.

5. Why is it useful to decompose a time series into a few components? What are the typical components in a time-series decomposition?

   Answer: Decomposition provides a useful abstract model for thinking about time series generally and for better understanding problems during time series analysis and forecasting. Trend-cycle component, seasonal component and remainder component are the typical components in a time-series decomposition,

2 Task 2. Feature Extraction

2.1 Task 2.1. Frequency components of a synthetic time-series signal

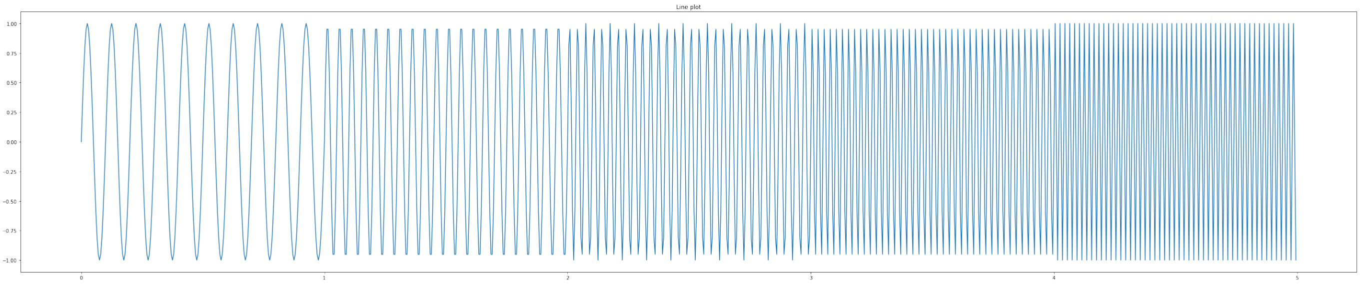


Figure 8: Line plot

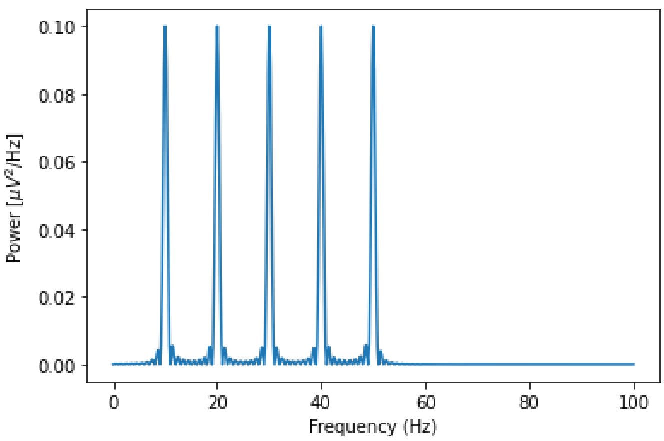


Figure 9: power spectrum

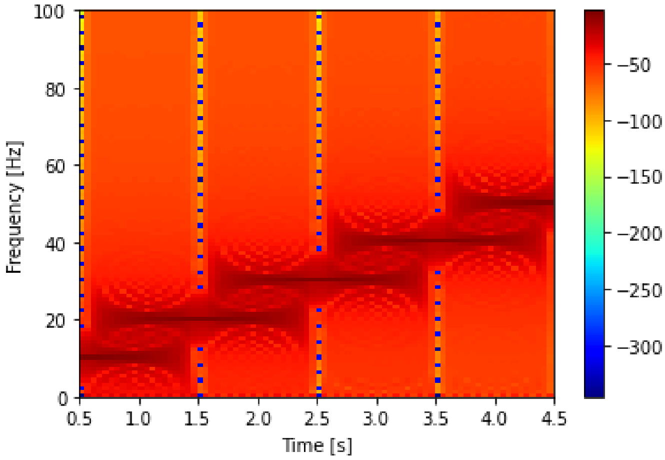
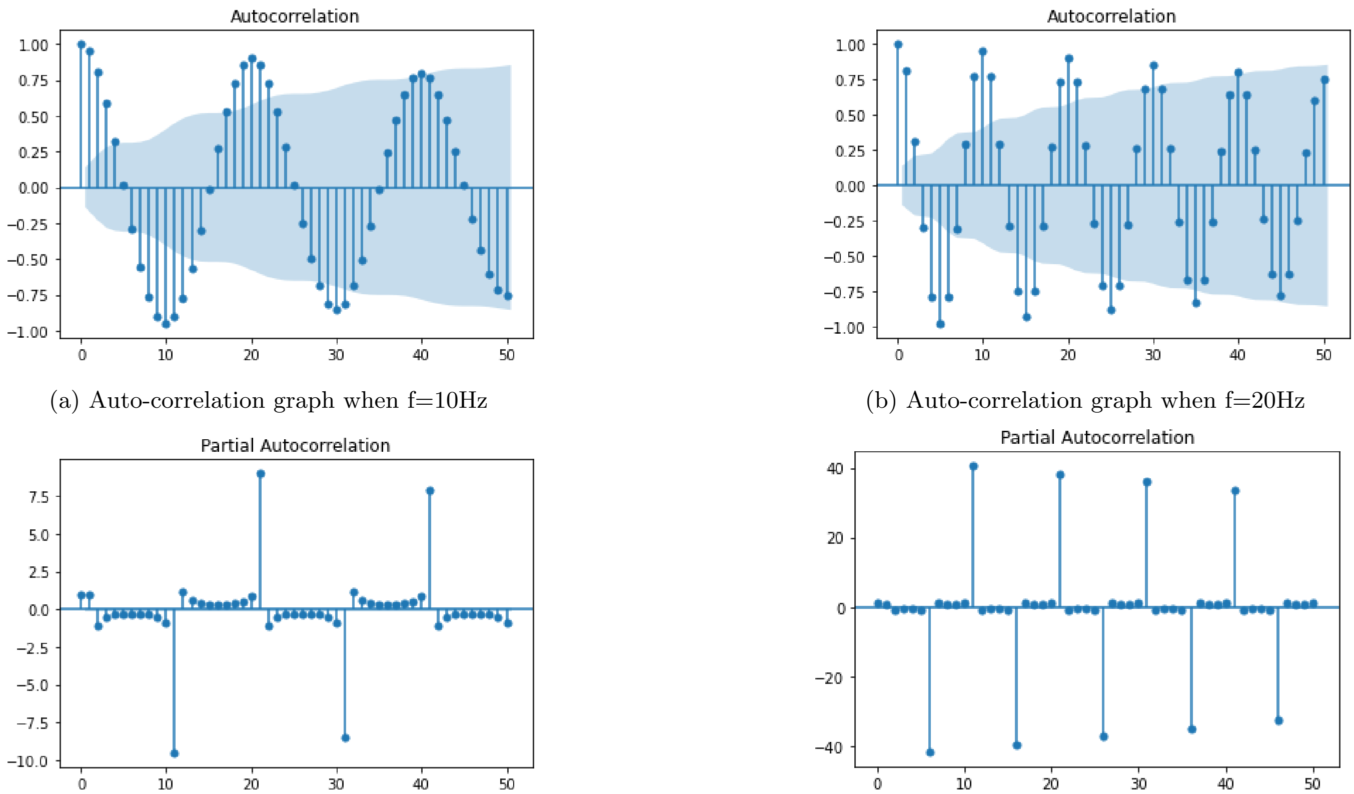


Figure 10: spectrum



(a) Auto-correlation graph when f=10Hz

(b) Auto-correlation graph when f=20Hz

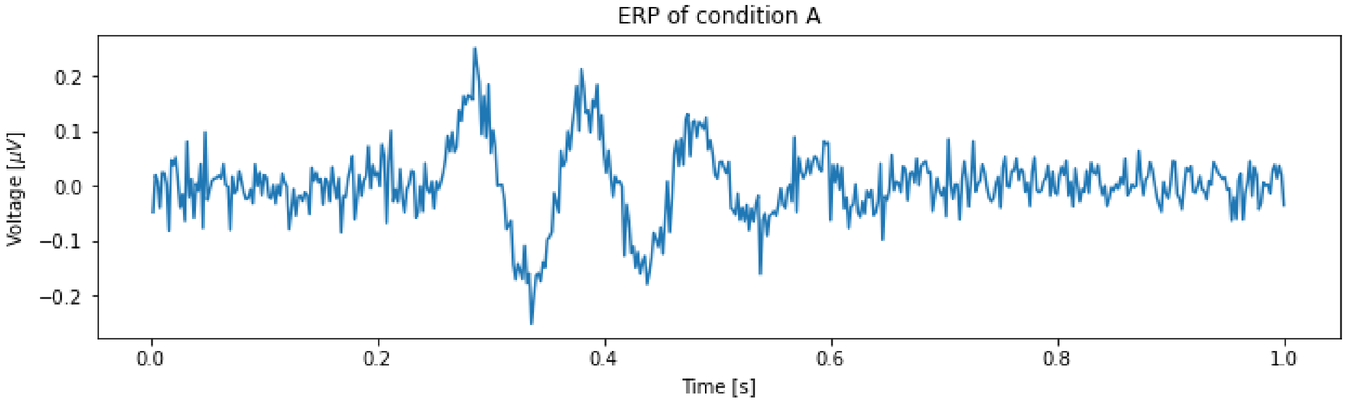
(c) PACF when f = 10Hz

(d) PACF graphs when f = 20Hz

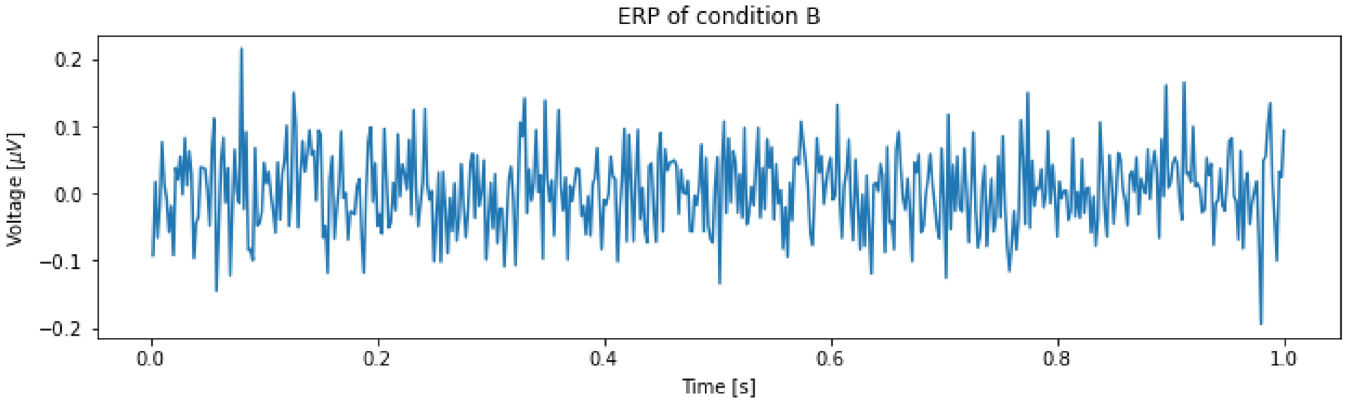
Figure 11: ACF and PACF graphs of the two series

We can observe that the periods of the first one-second and the second one-second series are different. In case of 10 Hz, peaks repeat every 20 units, while peaks repeat every 10 units in case of 20 Hz.

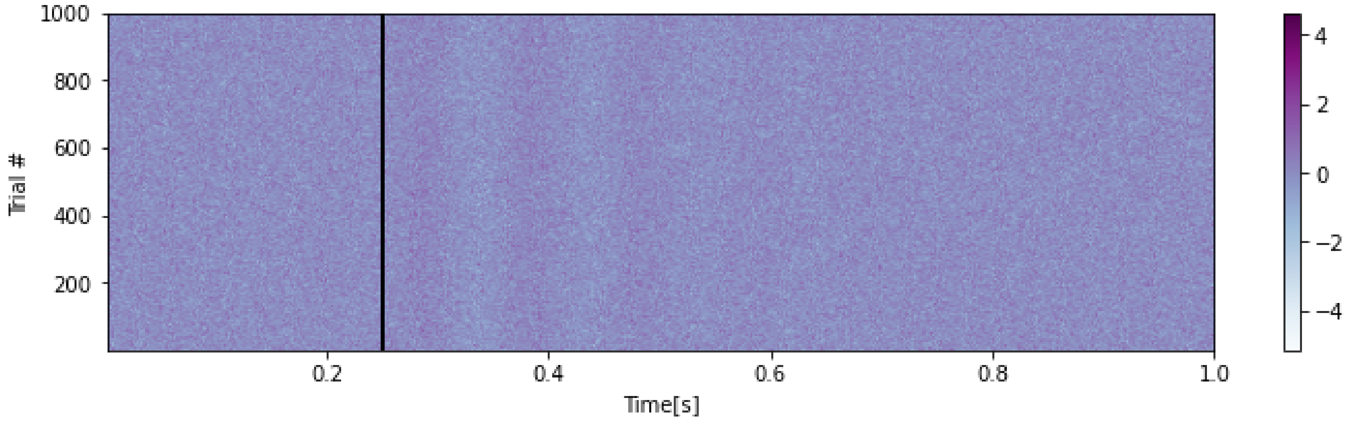
2.2 Task 2.2. Statistical features and discovery of event-related potential



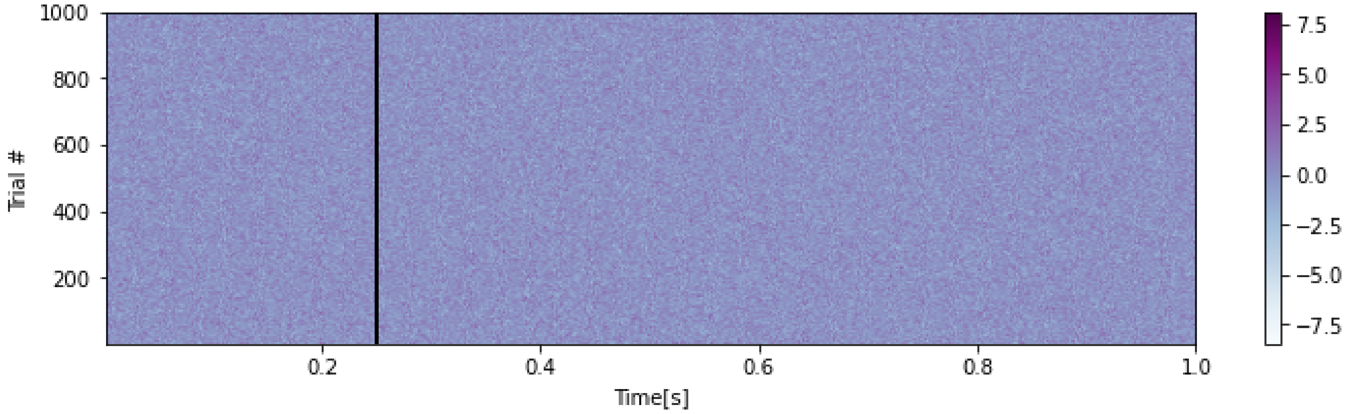
(a) Line plot for mean value of ERP of condition A



(b) Line plot for mean value of ERP of condition B
Visualization of the ERP in two conditions



(a) Condition A



(b) Condition B

Visualization at all trails in two conditions

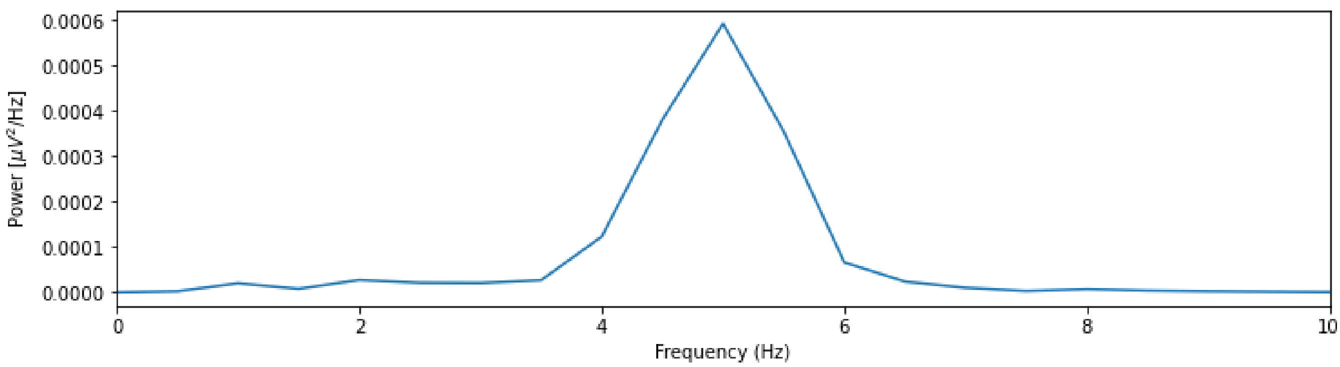
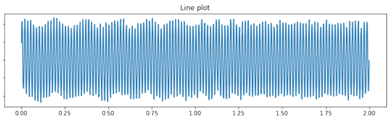


Figure 14: Brain activity frequency

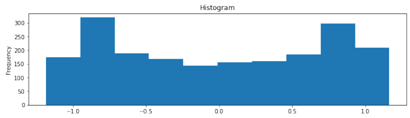
From the figure, we can judge that the frequency of brain activity is near 5Hz.

2.3 Features of observed rhythms in EEG

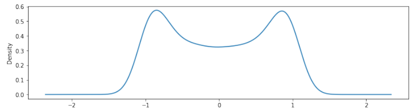
2.3.1 Plots of EEG dataset



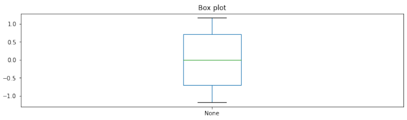
(a) Line plot of EEG signals



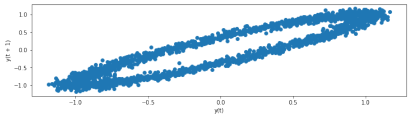
(b) Histogram of EEG signals



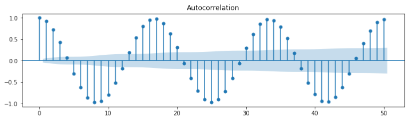
(c) Density plot of EEG signals



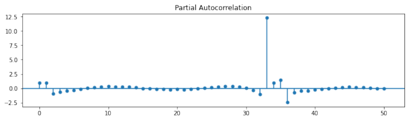
(d) Box plot of EEG signals



(e) Lag-1 plot of EEG signals



(f) ACF of EEG signals



(g) PACF of EEG signals

Figure 15: Graphs of EEG signals

Table 7: Features of EEG signals

Data	Mean	Variance	Standard Deviation
EEG	2.481868877080018e-17	0.5049697256484698	0.7106122188989363

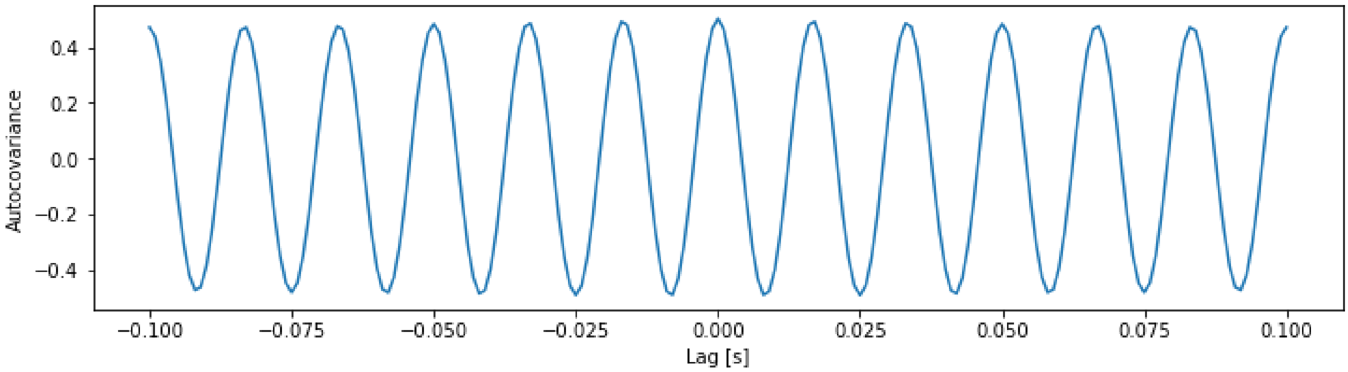
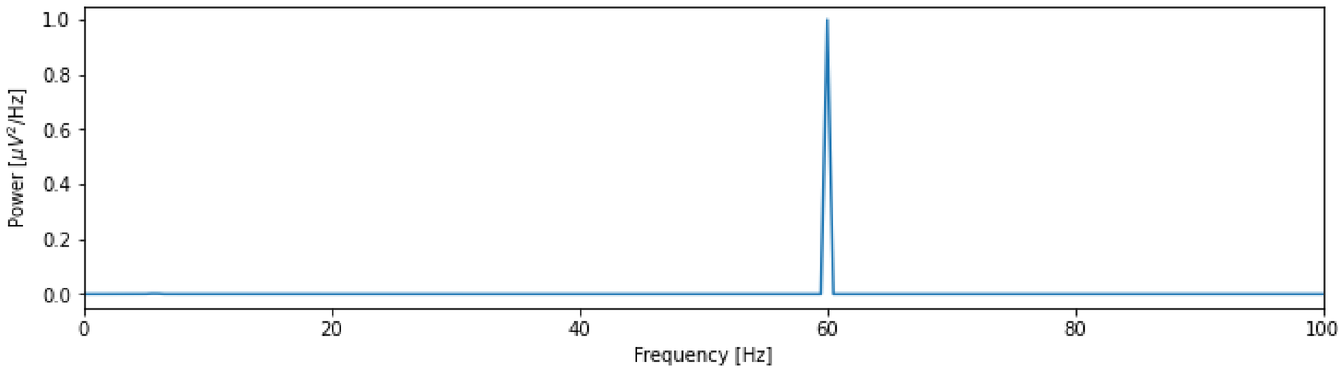
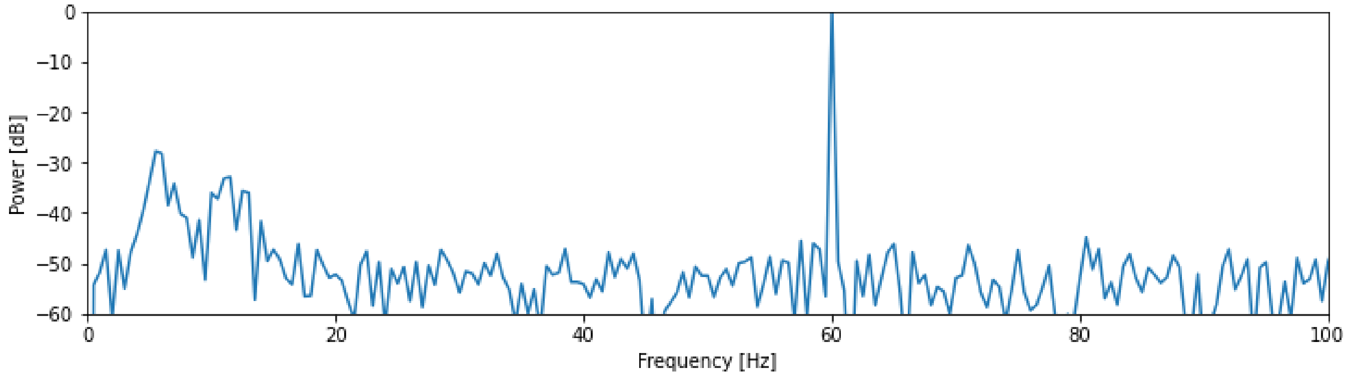


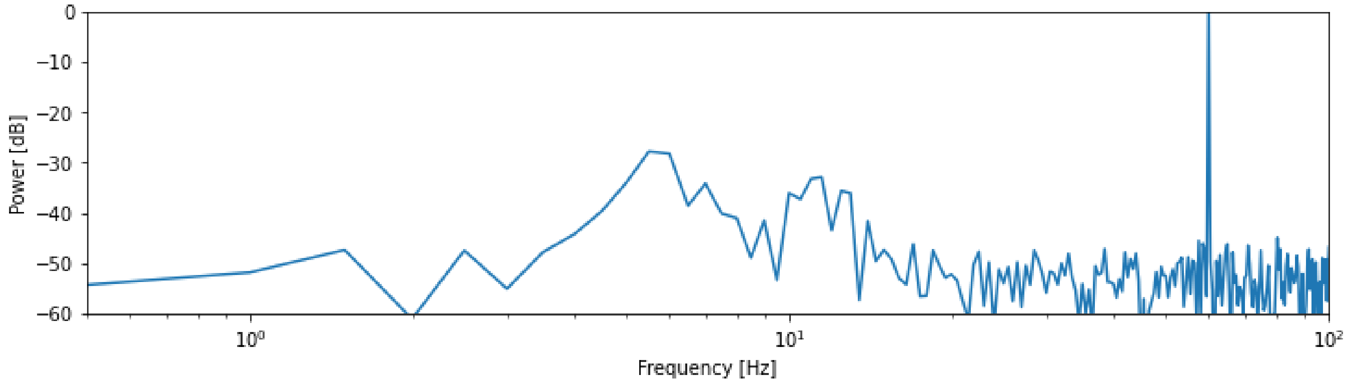
Figure 16: Auto variance of EEG signals



(a) Linear scale of the power-spectrum



(b) Log power scale of the power-spectrum



(c) Log frequency scale of the power-spectrum

Figure 17: power-spectrum of the EEG signals

The auto-covariance exhibit repeated peaks and troughs approximately every 0.166s, because by checking linear scale of the power-spectrum, we could find that the frequency is about 60 Hz and $\frac{1}{60} \approx 0.01667$.

2.3.2 Answers to the questions

1. What features do you typically consider useful for analyzing and modeling time series data?

   Answer: If the series is random, stationary. The mean value, standard deviation and so on.

2. What features are specific for time-series, and what are general for both time-series and non-time-series data? Answer: Usually, time series data has features like trend, seasonal and nonseasonal cycles, stationarity and so on. In general, they all have like mean value, standard deviation and so on.

3. How are auto-covariance and auto-correlation are defined for a time series? Give mathematical formulas for the definitions.

$$c_k = \frac{1}{T} \sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y}) \tag{1}$$

Notice that auto-correlation function is auto-covariance normalized with $r_{xx}[0]$.

$$r_k = c_k \; c_0 \tag{2}$$

Biased estimate of the auto-covariance:

$$r_{xx}[L] = \frac{1}{N} \sum_{n=1}^{N-L} (x_{n+L} - \bar{x})(x_{n+} - \bar{x}) \tag{3}$$

Unbiased estimate of the auto-covariance:

$$r_{xx}[L] = \frac{1}{N-L} \sum_{n=1}^{N-L} (x_{n+L} - \bar{x})(x_{n+} - \bar{x}) \tag{4}$$

4. Assume a short time-series 1, 2, 3, 4, 5, 6, 7, 8, 7, 6, 5, 4, 3, 2, 1. (1) Calculate the auto-covariance and auto-correlations for all valid lags. Do the calculations manually. (2) Write a Python program to validate your calculations. (3) Draw the ACF graph for the time series.

   Answer: In this case, N = 15 and $\bar{x} = \frac{64}{15}$. I calculate the first two lags as examples.

$$r_{xx}[0] = \frac{1}{N} \sum_{n=1}^{N} (x_n - \bar{x})(x_n - \bar{x}) = 4.72889 \tag{5}$$

$$r_{xx}[1] = \frac{1}{N-1} \sum_{n=1}^{N} (x_{n+1} - \bar{x})(x_n - \bar{x}) = 3.5508148148148146 \tag{6}$$

   Here are the results of auto-covariance and auto-correlations for all valid lags.

   Auto-covariance: [4.72888889, 3.55081481, 2.07496296, 0.50133333, -0.97007407, -2.13925926, -2.80622222, -2.77096296, -1.83348148, -0.93155556, -0.13185185, 0.49896296, 0.89422222, 0.98725926, 0.71140741]

   Auto-correlation: [1. , 0.75087719, 0.43878446, 0.10601504, -0.20513784, -0.45238095, -0.59342105, -0.58596491, -0.3877193 , -0.19699248, -0.02788221, 0.10551378, 0.18909774, 0.20877193, 0.1504386]

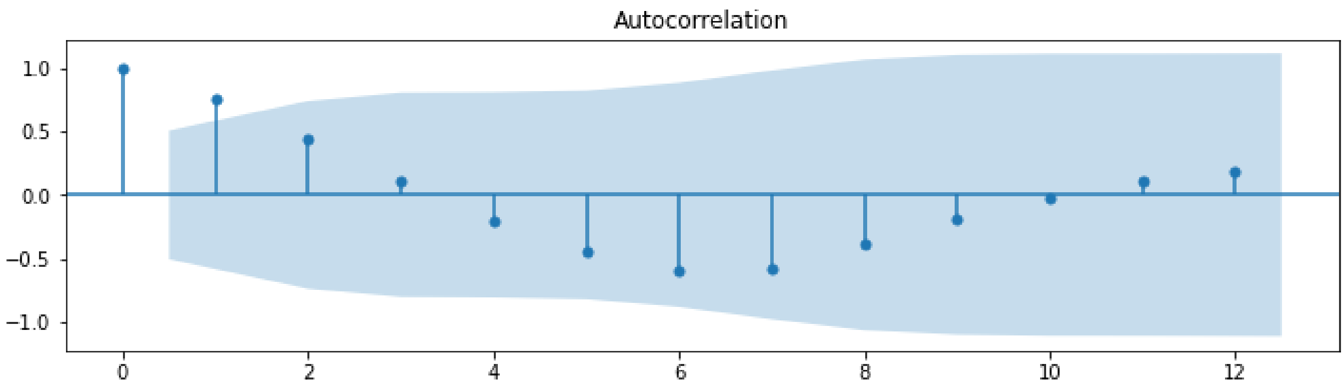


Figure 18: ACF graph for the time series