Lecture 3. Basics on Time-Series Analysis

Feature extraction, Stationarity and Decomposition

Prof. Zhonghai Lu

KTH Royal Institute of Technology

March 25, 2024

Outline

A time-series signal

- As an ordered data sequence, a time series may be treated as a digitalized signal.
- Basic digital signal processing methods such as filtering, frequency-domain analysis etc. are applicable.

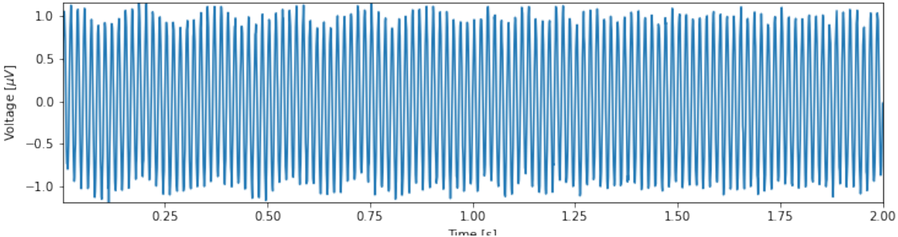


Figure: A 2s scalp EEG signal sampled at 1000 Hz from one electrode for an unconscious patient at the Massachusetts General Hospital (MGH) emergency room, USA.

Electroencephalography (EEG)

- EEG is a medical monitoring technique that records an electrogram of the electrical activity on the scalp representing the macroscopic activity of the brain.
- EEG measures the brain voltage activity with high temporal resolution (typically on the order of milliseconds) but poor spatial resolution (on the order of 10 cm^2 of cortex).

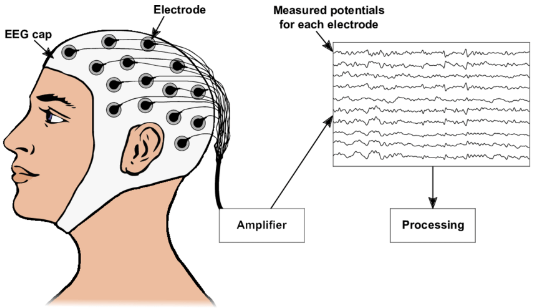


Figure: Monitoring brain activities via EEG

Electroencephalography (EEG)

- The waveforms of the brain activities are usually sinusoidal.
- The signals are very weak, measured from peak to peak with amplitude that ranges from 0.5 to 100 μV.
- A widely used convention of frequency bands in EEG is given below.

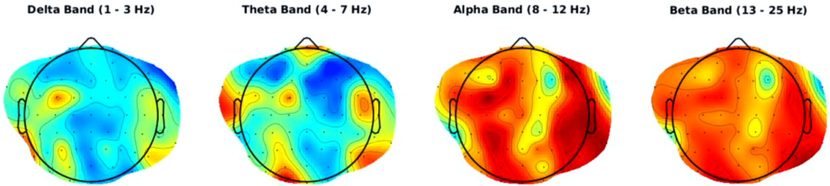


Figure: Frequency bands in EEG

https://towardsdatascience.com/time-series-signals-the-fourier-transform-f68e8a97c1c2

Questions to answer?

- What features characterize the collected measurement data?
    - Statistical features
    - Temporal features
    - Spectral features
- For the EEG data, what rhythmic activities can we find there? In particular, anything in the common frequency bands, $\delta, \theta, \alpha, \beta$?

Signal sampling

Let's define

- Δ: the sampling interval, i.e., time interval between samples, in this case, $\Delta = 1$ ms.
- N: the total number of points observed, and
- T: the total time of the recording or sampling.
  These three terms are related: $T = N\Delta$.
  For the T=2 s of EEG data, there are $N = T/\Delta = 2/0.001 = 2000$ points.
- The sampling frequency $f_0 = 1/\Delta$ which is 1000 Hz.

We define a symbol for the data, x, and use x_n to explicitly indicate the index $n \in \{1, 2, 3, \ldots, N\}$ corresponding to the sample number.

Dominant rhythmic activity?

Is there a dominant rhythmic activity in the data? If so, what is its frequency?

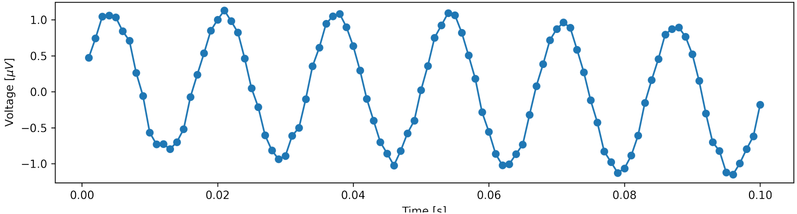


Figure: The first 0.1 second of data

We count about 6 peaks/troughs in the first 0.1 s, corresponding to approximately 60 peaks/troughs in 1 s, i.e, 60 cycles per second or 60 Hertz (Hz).

Data points in one cycle

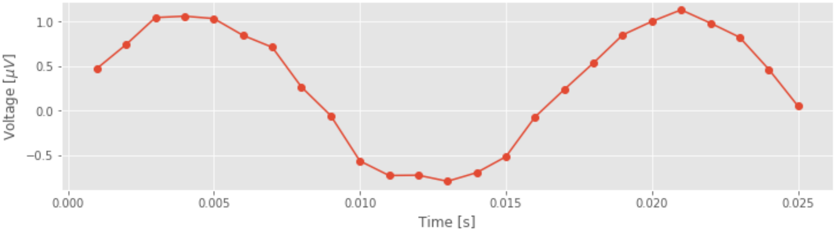


Figure: The first 25 points of data

The signal is approximately 60 cycles per second or 60 Hertz (Hz).
Since the sampling rate is 1000 Hz (1000 data points per second), one cycle has 1000/60 = 16.66667 data points spanning 1/60 = 0.01667 second.

Is anything lurking in the signal background?

- The dominant rhythmic activity is remarkably regular for the EEG data.
- Important fact: The alternating current in North American electrical grids alternates at 60 Hz.
- We conclude that the data are dominated by electrical noise.

The visual inspection suggests a dominant 60 Hz signal, which is noise!
But perhaps some other rhythmic activities are there, lurking in the signal background?

Statistical features

- Let x be a sequence of data, x_n being the value of x at index n, $n \in \{1, 2, 3, \ldots, N\}$. If n is a time instant, then x is a time series.
- To estimate its mean $\bar{x}$:

$$\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

- To estimate its variance (σ^2), standard deviation (σ):

$$\sigma^2 = \frac{1}{N} \sum_{n=1}^{N} (x_n - \bar{x})^2$$

Temporal feature: Auto-covariance

- Autocovariance is often used to assess the dependent structure in the data. The autocovariance, $r_{xx}[L]$, evaluated at lag L, is defined as

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

- Note that this is a *biased* estimate of the true auto-covariance. To compute an *unbiased* estmiate of the auto-covariance, we replace the $1/N$ term with $1/(N-L)$.

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

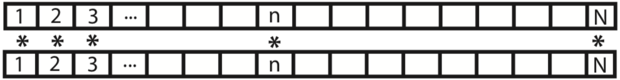
Auto-covariance

Let's represent *x* graphically as a one-dimensional row vector.

x

- When $L = 0$, the auto-covariance is simply the element-by-element product of x with itself (after subtracting mean), summed over all indices.

L=0



sum $r_{xx}[0]$

Figure: Auto-covariance when L=0

Auto-covariance

- When $L = 1$, we shift x by one index, multiply element-by-element the original x by the shifted version (after subtracting mean $\bar{x}$), and sum over all indices.

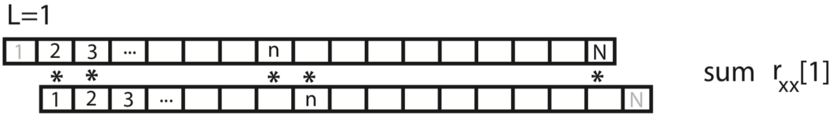


Figure: Auto-covariance when L=1

- When $L = 2$, we shift x by two indices, multiply element-by-element the original x by the shifted version (after subtracting mean $\bar{x}$), and sum over all indices.

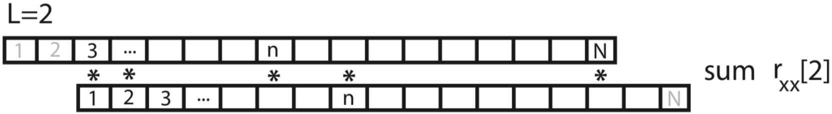
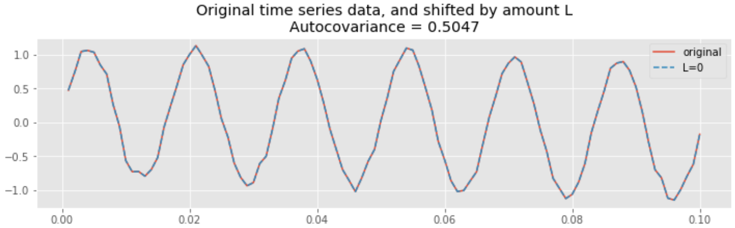


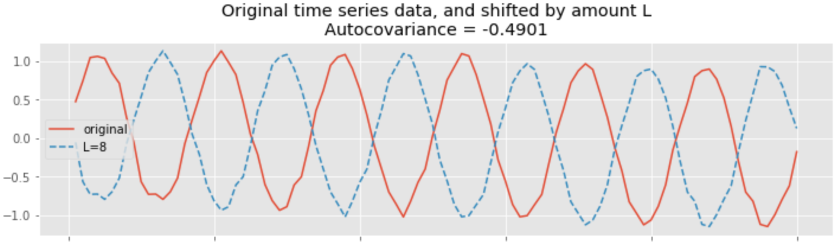
Figure: Auto-covariance when L=2

Auto-covariance of the EEG data

- When $L = 0$, compute the co-variance of itself.



- When $L = 8$, shift x by half cycle.



Co-variance of the EEG signal

- The EEG signal is periodic, and so does its auto-covariance.

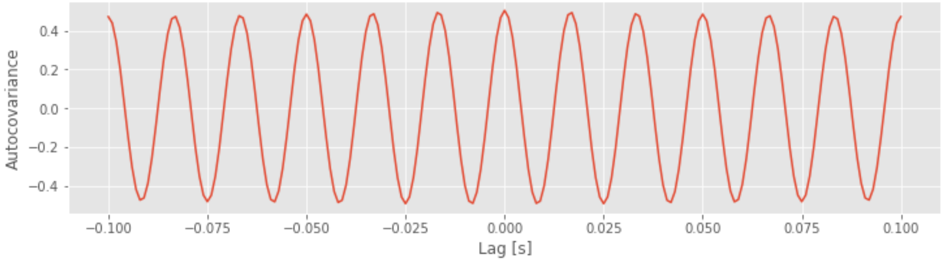


Figure: Auto-covariance with lags in time

ACF of the EEG signal

- Autocorrelation Function (ACF): Autocovariance as a function of Lag.

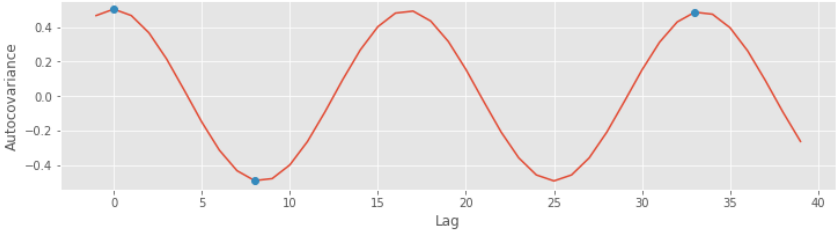


Figure: Auto-covariance with lags in index

- The dominant 60 Hz activity manifests as periodic peaks and troughs in the autocovariance function.

ACF of the EEG signal

- Auto-Correlation Function (ACF) is indeed auto-covariance normalized with the autocovaraince $r_{xx}[0]$ (Lag=0).

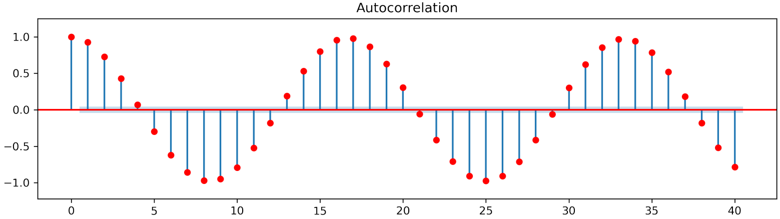


Figure: ACF of the EEG signal

- The dominant 60 Hz activity manifests as periodic peaks and troughs in the autocovariance function.

Power spectral density or spectrum

Are there other rhythmic activities in the signal at all?

- The power spectrum indicates the amplitude of rhythmic activity in x as a function of frequency.
- The power spectrum of signal x is the magnitude squared of the DFT of x.
- The power spectral density (psd) or spectrum describes the extent to which sinusoids of a single frequency capture the structure of the data.
  To compute the power over any range of frequencies, we would integrate (or for discrete frequencies, sum) the spectrum over that frequency range.

Discrete Fourier Transform (DFT)

- The Discrete-time Fourier Transform (DFT) of signal x with N data points for duration T seconds

$$X_j = \frac{1}{N}\sum_{n=1}^{N} x_n \cdot e^{-2\pi i \cdot f_j \cdot t_n}$$

  - DFT sums over all time indices $t_n = \Delta\{1, 2, 3, ..., N\}$ of the data x_n multiplied by sinusoids oscillating at a given frequency $f_j = j/T$, where $j = \{-\frac{N}{2}+1, -\frac{N}{2}+2, ..., \frac{N}{2}-1, \frac{N}{2}\}$.
  - X_j is a new signal as a function of frequency f_j rather than time t_n.

- By Euler's formula

$$e^{(-2\pi i f_j t_n)} = cos(-2\pi f_j t_n) + i \cdot sin(-2\pi f_j t_n)$$

Power spectrum

- The power spectrum is then

$$S_{xx,j} = \frac{2\Delta^2}{T} X_j \cdot X_j^*$$

  which is the product of the DFT of x, X_j, with its complex conjugate X_j^* (indicated by superscript *), scaled by the sampling interval Δ and the total duration of the recording T.
- The term $2\Delta^2/T$ is simply a numerical scaling.
- The unit of the power spectrum is, in this case, $(\mu V)^2/\mathrm{Hz}$.

Power spectrum

In this plot, the horizontal axis corresponds to the indices of x, beginning at index 0 and ending at index $N = 1000$. [Useful fact: When a signal is real, the negative frequencies in the spectrum are redundant.]

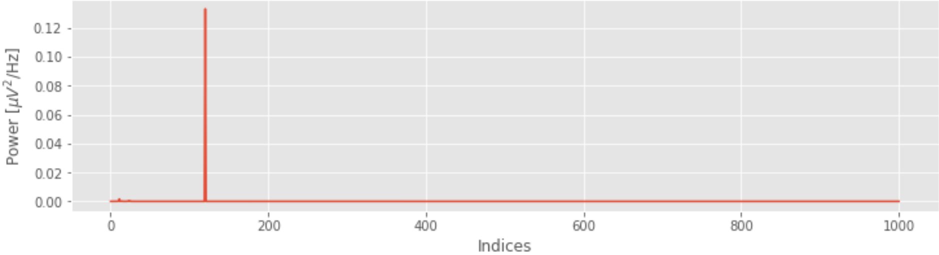


Figure: Power spectrum of the EEG data

To convert the x-axis from indices to frequencies, we define two new quantities:

- the frequency resolution, $df = 1/T$, or the reciprocal of the total recording duration;
- the Nyquist frequency (maximum observable frequency), $f_{NQ} = f_0/2 = 1/(2\Delta)$, or half of the sampling frequency $f_0 = 1/\Delta$.

Power spectrum

- For the EEG data, the total recording duration is T = 2 seconds, so the frequency resolution $df = 1/(2s) = 0.5$ Hz.
- The sampling frequency f_0 is 1000 Hz, so $f_{NQ} = 1000/2$ Hz$=$ 500 Hz.

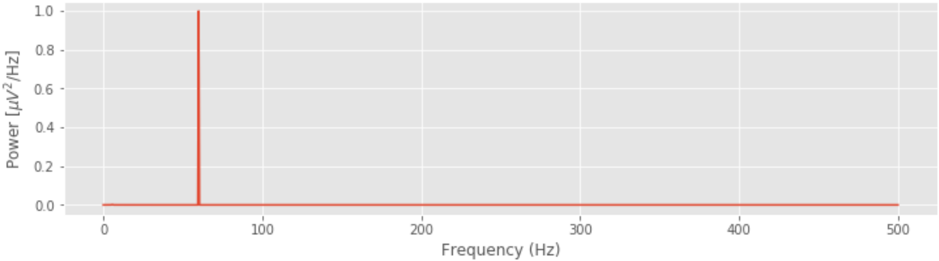


Figure: Power spectrum

Decibel scaling on the Y-axis

- The spectrum is dominated by a single peak at 60 Hz. Other weaker rhythmic activity might be present but invisible, because the large 60 Hz peak saturates the vertical scale.
- To emphasize lower-amplitude rhythms hidden by larger-amplitude oscillations, we can change the scale of the spectrum to decibels.
- The decibel is a logarithmic scale. Different conventions exist to define the decibel scale.

Decibel scaling on the Y-axis

- To change to the decibel scale, we first divide the spectrum by the maximum value observed (at the dominant 60 Hz) and then take the logarithm base 10 of this ratio and multiply the result by 10.
- The decibel scale reveals new structure in the spectrum. In particular, two peaks have emerged at frequencies 5–15 Hz.
- These peaks are much weaker than the 60 Hz signal, approximately 30 dB below the maximum at 60 Hz, or equivalently, three orders of magnitude weaker.

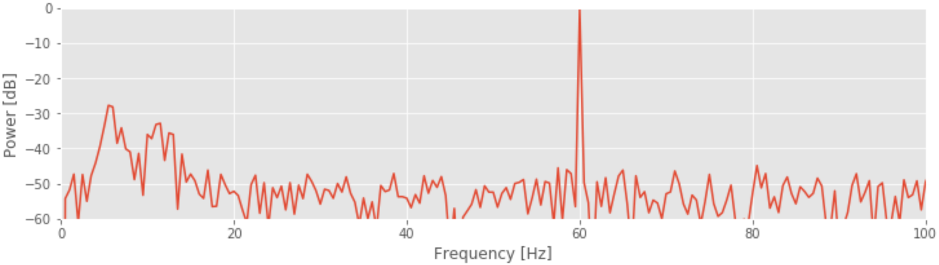


Figure: Power spectrum (power in dB)

Decibel scaling on the X-axis

- Similarly, we can use the logarithmic scale to stretch the low-frequency part of the horizontal axis, the two low-frequency peaks become more apparent.
- By using the logarithmic scale to stretch the low-frequency part of the horizontal axis, the two low-frequency peaks (6 Hz and 11 Hz) become more apparent.

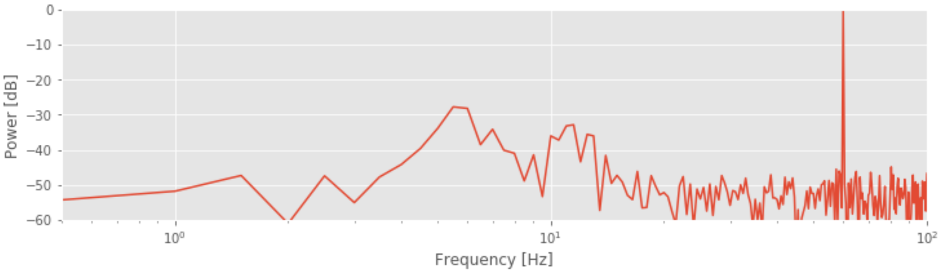


Figure: Power spectrum (frequency in log10 scale)

Spectrogram

Do the two weaker signals (6 Hz and 11 Hz) last the entire measurement duration for 2 seconds?

- The idea of the spectrogram is to break up the time series into smaller intervals of data and then compute the spectrum in each interval.
- These intervals can be quite small and can even overlap.
- The result is the spectrum as a function of frequency and time. It provides insight into spectral features that change in time.

Spectrogram

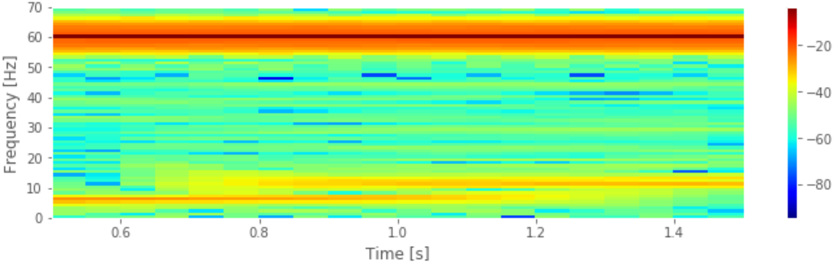


Figure: Power spectrogram of the EEG signal

- The EEG data are dominated by 60 Hz noise throughout the 2 s interval.
- However, weaker low-frequency activity emerges during two intervals: a 6 Hz rhythm in the θ band from 0 s to 1 s, and an 11 Hz rhythm in the α from 1 s to 2 s.

Summary

- Time series has a rich features such as statistical, temporal, and spectral features which may be interesting for different purposes.
  - Statistical: mean, variance, etc.
  - Temporal: covariance, ACF, Partial ACF (PACF) etc.
  - Spectral: power spectrum, spectrogram etc.
- Through the case study, we can characterize different features using respective function calls in Python libraries.

Ipython demo

Now we will see how the entire investigation is coded in Python.

`https://mark-kramer.github.io/Case-Studies-Python/03.html#supplement-acv`

- Visual inspection
- Extract statistical features
- Extract temporal features
- Extract spectral features

`http://localhost:8888/notebooks/IL2233VT22/Lec3_feature_extraction/03-onramp.ipynb`

Stationarity

- Classical time series analysis method, ARIMA (Auto-Regression Integrated Moving-Average), assumes that the series is stationary.
- If y_t is a stationary series, then for all s, the joint distribution of $(y_t, y_{t+1}, \cdots, y_{t+s})$ is dependent on s but independent of t.
- A stationary series has
  - Constant mean μ
  - Constant variance σ^2

Checking stationarity

Stationarity can be checked by different means.

- Visual inspection of line plot
- Statistic quantity check, e.g. Augmented Dickey-Fuller (ADF) test

Is the series stationary?

Judge if the series is stationary from its line plot. Explain why or why not?

Are they stationary?

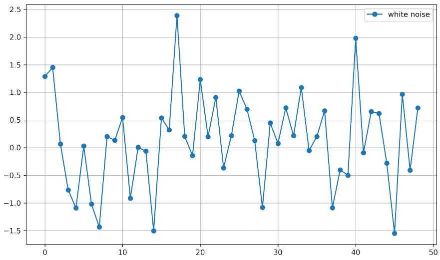


Figure: A random number series

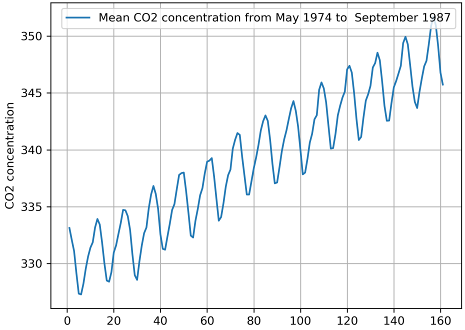


Figure: A CO_2 concentration data set

ADF test

Augmented Dickey-Fuller (ADF) Test is one of *hypothesis testing* methods used to test the stationarity of a series.

- Null hypothesis (H_0): The series is not stationary.
- Alternative hypothesis (H_a): The series is stationary.

At first a test statistic is computed and p-value gets reported. From the test statistic and the p-value, you can judge whether a given series is stationary or not.

- If the p-value $\leq$ the significance level (default value 0.05), then you reject the null hypothesis, and consider that the series is stationary.
- If the p-value $>$ the significance level, you accept the null hypothesis, and consider that the series is not stationary.

ADF test and Lag plot of white noise

- ADF test result for the random series:

Test Statistic	p-value
-6.630003e+00	5.746791e-09

  Since p-value < 0.05, we reject the Null hypothesis. Thus the series is stationary.

- To show better the result, the lag plot is generated with 500 random points.

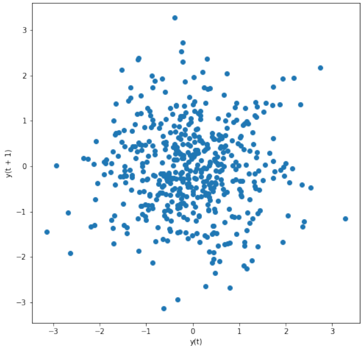


Figure: Lag plot of a white noise

White noise vs. Random walk

- White noise is a stationary series.

$$y_t = \epsilon_t$$

- A random-walk series is a non-stationary series.

$$y_t = c + y_{t-1} + \epsilon_t$$

  where c is a drift constant.

- It can be turned into a stationary series by first-order differencing.

$$y'_t = y_t - y_{t-1} = \epsilon'_t$$

Time-series differencing

- Differencing is a commonly used transformation to turn a non-stationary series into a stationary one.
- First-order and second-order differencing

$$y : y_1, y_2, y_3, \cdots, y_n, \cdots y_N$$

$$y' : y_2 - y_1, y_3 - y_2, \cdots, y_n - y_{n-1}, \cdots, y_N - y_{N-1}$$

$$y'' : y'_2 - y'_1, y'_3 - y'_2, \cdots, y'_n - y'_{n-1}, \cdots, y'_{N-1} - y'_{N-2}$$

- How many orders are needed?
  - Statistical tests can be used to determine the required order of differencing. Need to avoid under-differencing and over-differencing.
  - Lower-order differencing is more common than higher-order differencing.

Time-series decomposition

A time series is often structured, patterned.

- Trend: a pattern showing there is a long-term increase or decrease in the data.
- Seasonal: a pattern exists when a series is influenced by a seasonal factor (e.g. a quarter, a month, or each day of a week.), or more generally, a fixed period.
- Cyclic: a pattern exists when data exhibit rise and fall at not-a-fixed period (duration usually at least 2 years for chronological data)

Seasonal or cyclic?

Differences between season and cycle can be checked from:

- Period
  Seasonal pattern: constant length
  Cyclic pattern: variable length
- Short-Long term
  Average length of a cycle is longer than that of a season.
- Magnitude
  The magnitude of a cycle is more variable than that of a season.
- Predictability: Generally, the timing of peaks and troughs is predictable with seasonal data, but unpredictable in the long term with cyclic data.

Time-series decomposition

General decomposition

$$y_t = f(S_t + T_t + R_t)$$

where
y_t is data at time t.
T_t = trend-cycle component at time t
S_t = seasonal component at time t
R_t = remainder component at time t

Common decomposition means

- Additive decomposition: $y_t = S_t + T_t + R_t$
- Multiplicative decomposition: $y_t = S_t \times T_t \times R_t$

- The log() operation turns a multiplicative relationship into an additive relationship. $log(y_t) = log(S_t \times T_t \times R_t) = log(S_t) + log(T_t) + log(R_t)$
- For negative data, one can add a suitable constant to make all the data positive before applying the transformation. This constant can then be subtracted from the model to obtain predicted (i.e., the fitted) values and forecasts for future data points.

Decomposition algorithms

- Classical simple decomposition: seasonal_decompose
  - additive
  - multiplicative
- STL: "Seasonal and Trend decomposition using Loess" (Loess: Local regression)
  - Very versatile and robust to outliers.
  - STL allows seasonal component to change over time, and rate of change controlled by user.
  - Smoothness of trend-cycle also controlled by user.
  - Only additive. Take logs to get multiplicative decomposition.
  - Use Box-Cox transformations to get other decompositions.

The CO_2 Example

- The CO_2 data set contains monthly mean CO_2 concentrations at the Mauna Loa Observatory as measured by the continuous infrared analyser of the Geophysical Monitoring for Climatic Change division of NOAA's Air Resources Laboratory.
  In total, it has 161 data.
- Is the series stationary?

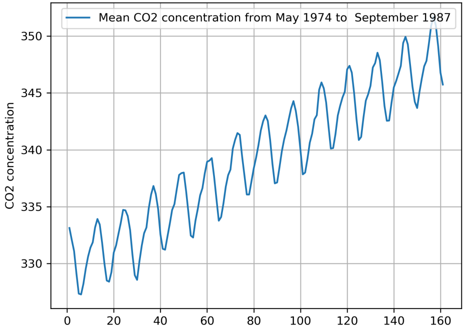


Figure: The CO_2 concentration data set

CO2 ADF test and Lag plot

- Results of ADF Test:

Test Statistic	p-value
-0.355668	0.917242

  Since p-value > 0.05, we accept the Null hypothesis.
- Thus the series is Not stationary.

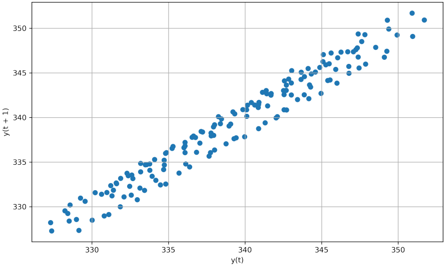


Figure: Lag-1 plot of the CO_2 data set

Trend and seasonality in ACF plots

- When data have a trend, the autocorrelations for small lags tend to be large and positive.
- When data are seasonal, the autocorrelations will be larger at the seasonal lags (i.e., at multiples of the seasonal frequency)
- When data are trended and seasonal, you see a combination of these effects.

ACF of the CO_2 data

For this data which are trended and seasonal, we see a combination of the trend and season effects.

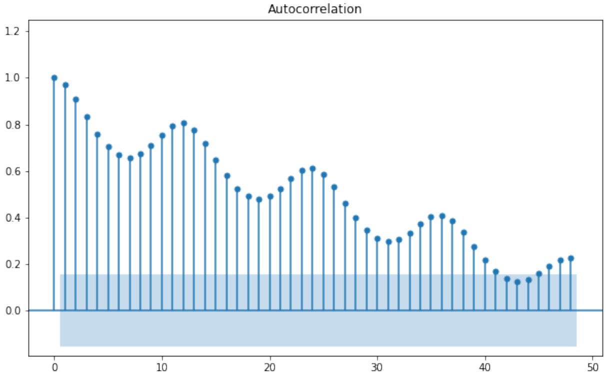


Figure: ACF of the CO_2 data

Decomposition of the CO_2 data

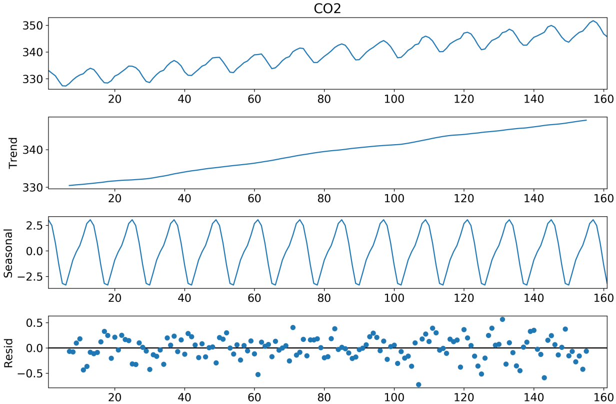


Figure: Decomposition of the CO_2 data set

```
from statsmodels.tsa.seasonal import seasonal_decompose
result = seasonal_decompose(series.dropna(),model='additive', period=12)
fig = result.plot()
```

IPython demo

This lecture uses the following iPython demos under directory IL2233VT22 Lecture_3[1]:

- co2.ipynb using the CO2 data.

[1] The ipython code in my lecture slides is for my in-classroom demonstration only, not available to students.

Acknowledgements

- Case Studies in Neural Data Analysis, by Mark Kramer and Uri Eden. The book uses Matlab for analysis. The online version uses Python. `https://mark-kramer.github.io/Case-Studies-Python/intro.html`
- Nagel, Sebastian. (2019). Towards a home-use BCI: fast asynchronous control and robust non-control state detection. 10.15496/publikation-37739.