Chapter 1 - Introduction to Time Series

Basic Definitions

- **Definition: A time series is a sequence of information which attaches a time period to each value.**
- **Definition: A time period is the difference between the start and end time.**
- **Definition: Frequency refers to how often values of the data set are recorded.**
- To analyze time series, all time periods must be equal and clearly defined. This results in a constant frequency.
- Patterns observed in time series are expected to persist in the future.
- Example of time series data: Stock market, weather forecast.
- Time series data suffer from seasonality. (eg. Spring, autumn, summer and winter happen every year)

Notation for Time Series

- For the time period, let t denote a single period and let T denote the entire period.
- Let X denote the value variable. (eg. stock price of S&P 500)
- To denote the value of the value variable at single time period t, write X_t. Then, the value of the value variable at the next time period is X_{t+1}.

Peculiarities of Time Series Data

- Time series intervals need to be constant. Inconsistent intervals may be due to missing data. Missing data needs to be approximated.
- Time series data require chronological data. This means that machine learning model cannot shuffle the data when splitting it. It requires splitting the data based on a specific time period

(ie. Values before the time period are training data while values after the time period are test data).

- Time series graphs do not follow any standard model because they do not follow Gauss-Markov assumption.

Concepts in Time Series

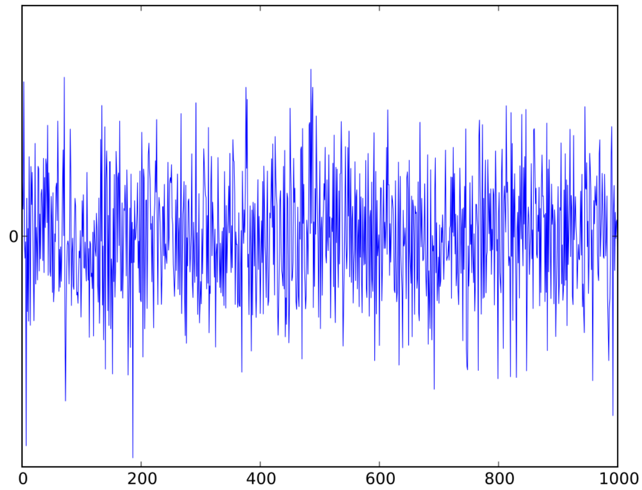
White Noise

- **Definition: White noise is a special type of time series, where the data does not follow any pattern (ie. is random).**
- Since it does not follow any pattern, it is unpredictable (ie. Future time step values cannot be predicted using past time steps).
- **Definition: Autocorrelation denotes how correlated the time series is with past versions of itself.**

$$\rho = corr(x_t, x_{t-1})$$

▼ Conditions for white noise:

- Zero mean and constant variance.
- No autocorrelation (ie. No clear relationship between past and present data).



Random Walk

- **Definition: A random walk is a special type of time series where values tend to persist over time and the differences between periods are simply white noise.**
- The best estimator for today's value is yesterday's value. The best estimator for tomorrow's value is today's value.
- Suppose that X_t denotes the price at time t, then $X_t = X_{t-1} + \epsilon_t$ where $\epsilon_t - WN(\mu, \sigma^2)$ is a random walk.

Proof:

$$X_t - X_{t-1} = \epsilon_t - \epsilon_{t-1}$$

which is simply white noise.

- If stock market prices are modelled by random walks, then prices cannot be predicted efficiently.

Stationarity

- **Definition: A weak form stationarity is satisfied if $cov(S_1) = cov(S_2)$ for two series S_1 and S_2 with the same period, regardless of the starting point.**

▼ Conditions for weak form stationarity:

  - Constant mean and constant variance.
  - $cov(X_n, x_{n+k}) = cov(x_m, x_{m+k})$

- An example of a stationary time series is white noise.
- Random walks are non-stationary time series.
- Dickey-Fuller test is used to determine if a time series has stationarity.

Seasonality

▼ Seasonality in data can be detected by decomposing the data into 3 components:

  - Trend detects pattern in data.
  - Seasonal detects cyclical effects.

- Residual detects error of prediction.

- ▼ Types of naive decomposition:
  - ▼ Additive
    - Observed = Trend + Seasonal + Residual
  - ▼ Multiplicative
    - Observed = Trend * Seasonal * Residual

Autocorrelation Function

Autocorrelation Function (ACF)

- **Definition: Autocorrelation function computes the autocorrelation between different lags (ie. current time period and past time period).**
- Computes indirect lag.
- Eg. Prices from two days ago affect prices one day ago, which then affect prices today.

Partial Autocorrelation Function (PACF)

- Computes direct lag.
- **Definition: PACF at lag k is the autocorrelation between X_t and X_{t-k} that is not accounted for by lags 1 through $k-1$.**
- The ACF measures the accumulated effects past lags have on the current value while PACF measures the direct effect of each lag on the current value.