Go Back to Practical Data Science

Course Content

Stationary and Non-Stationary Time Series

In the previous pre-read, we read about the various components of a time series.

In the real world, time series data has specific properties that help to determine which type of time series it represents.

There are 2 broad categories of time series:

1. Non-Stationary Time Series

2. Stationary Time Series

Non-Stationary Time Series

In non-stationary time series, the statistical properties of the time series change over time. A time series containing a trend or seasonality is therefore non-stationary. This is due to the fact that the presence of a trend or seasonality affects the mean, variance, and other features of the data at any given period.

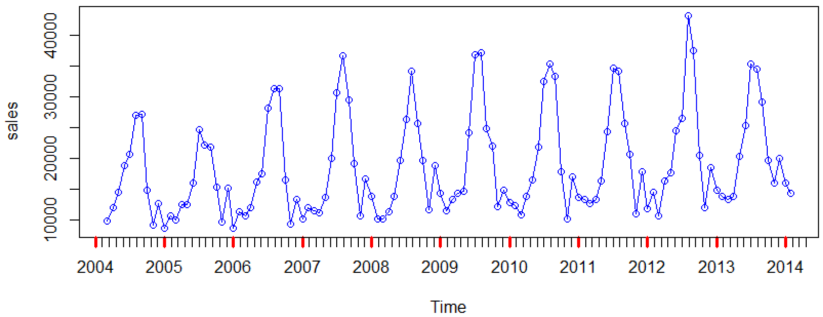

Image Source

In the above plot, we observe the presence of a certain amount of seasonality over the time series, which is an example of non-stationarity.

Stationary Time Series

The formal definition of a stationary process is a random process whose unconditional joint probability distribution does not change when shifted in time. Many time series forecasting algorithms rely on statistical consistency to create predictable distributions. In basic terms, we can slice up the time series data into equally sized parts and yet have the same probability distribution. if and only if, the time series is stationary.

A stationary time series has a constant variance and it always returns to the long-run mean, i.e, the mean is almost constant in the time series up to **n** time stamps.

An example of a stationarity time series is observed from the below plot.

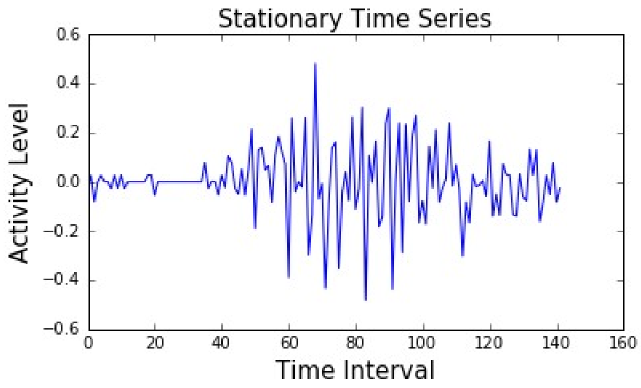

Image source

The above graph does not show any trend or seasonality in the series and is stationary. There are some spikes in the data, but the series has an almost consistent long-run mean.

Why do we convert a Non-Stationary Time Series to Stationary?

Forecasting a stationary time series is relatively easy, and the forecasts are more reliable. An important reason is that forecasting methods by default assume that the time series is stationary. And some forecasting models are essentially linear regression models that utilize the lag(s) of the series itself as predictor variables. We know that linear regression works best if the predictors are not correlated with each other. So, making the series stationary solves the problem of correlation since it removes any persistent trend thereby making the predictors (the lags of the series) in the forecasting models nearly independent of each other, and hence enabling linear regression-based methods.

Detecting the Stationarity of a Time Series

The most common method to check whether a given data comes from a stationary series or not is to simply plot the data. Both stationary and non-stationary time series have some properties that can be detected very easily from the plot of their data. For example, in a stationary series, the data points would always return towards the long-run mean with a constant variance, whereas, in a non-stationary time series, the data points might show some trend or seasonality.

Alternatively, we can also use hypothesis tests like Augmented Dickey–Fuller (ADF) test, to check whether the time series is stationary or not.

How to Make a Time Series Stationary

The most common and convenient method to make a time series stationary is by **differencing** the series (at least once) until it becomes approximately stationary.

Differencing is a technique that subtracts the next value from the current value of the time series. Mathematically, if Y_t is the value at a time ‘t’, then the first difference is $Y_t - Y_{t-1}$. If the first difference doesn’t make a series stationary, you can go for the second differencing, and so on. The number of times that differencing is performed is called the difference order.

For example, consider the given series: **[2, 3, 4, 10, 20]**

First differencing gives: [3-2, 4-3, 10-4, 20-10] = [1, 1, 6, 10]

Second differencing gives: [1-1, 6-1, 10-6] = [0, 5, 4]

and so on.

Note that the number of values in the time series is decreasing with each differencing because the first value would not have any previous value, hence its output would be NaN and will be removed from the time series.

Happy Learning!

Previous Next