Time Series Analysis Theory Document

What the tutorial covers?

The tutorial will cover using time series to forecast how many passengers are flying per month on an airline for several years.

Key Terms

Time series data: A set of observations on the values that a variable takes at different times

Stationarity: Shows the mean value of the series that remains constant over time; if past effects accumulate and the values increase toward infinity, then stationarity is not met

Differencing: Used to make the series stationary, to De-trend, and to control the auto-correlations; however, some time series analyses do not require differencing and over-differenced series can produce inaccurate estimates.

ARIMA: stands for autoregressive integrated moving average. This method is also known as the Box-Jenkins method.

Statistical stationarity

A stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time. Most statistical forecasting methods are based on the assumption that the time series can be rendered approximately stationary through the use of mathematical transformations. A stationary series is relatively easy to predict: you simply predict that its statistical properties will be the same in the future as they have been in the past. The predictions for the stationary series can then be "untransformed," by reversing whatever mathematical transformations were previously used, to obtain predictions for the original series. Thus, finding the sequence of transformations needed to make a time series stationary often provides important clues in the search for an appropriate forecasting model. Stationarizing a time series through differencing is an important part of the process of fitting an ARIMA.

Dickey-Fuller Test

This is one of the statistical tests for checking stationarity. Here the null hypothesis is that the TS is non-stationary. The test results comprise of a Test Statistic and some Critical Values for difference confidence levels. If the 'Test Statistic' is less than the 'Critical Value', we can reject the null hypothesis and say that the series is stationary. The different Critical values represent the difference confidence intervals so that if the Test statistic is below the 1% critical value then we can be 99% sure that the time series is stationary

ARIMA -Auto-Regressive Integrated Moving Averages.

The ARIMA forecasting for a stationary time series is nothing but a linear (like a linear regression) equation. The predictors depend on the parameters (p,d,q) of the ARIMA model:

Number of AR (Auto-Regressive) terms (p): AR terms are just lags of dependent variable. For instance if p is 5, the predictors for x(t) will be x(t-1)….x(t-5).

Number of MA (Moving Average) terms (q): MA terms are lagged forecast errors in prediction equation. For instance if q is 5, the predictors for x(t) will be e(t-1) ….e(t-5) where e(i) is the difference between the moving average at ith instant and actual value.

Number of Differences (d): These are the number of nonseasonal differences, i.e. in this case we took the first order difference. So, either we can pass that variable and put d=0 or pass the original variable and put d=1. Both will generate same results.

An important concern here is how to determine the value of ‘p’ and ‘q’. We use two plots to determine these numbers. Lets discuss them first.

Autocorrelation Function (ACF): It is a measure of the correlation between the TS with a lagged version of itself. For instance, at lag 5, ACF would compare series at time instant ‘t1’…’t2’ with series at instant ‘t1-5’…’t2-5’ (t1-5 and t2 being end points).

Partial Autocorrelation Function (PACF): This measures the correlation between the TS with a lagged version of itself but after eliminating the variations already explained by the intervening comparisons. E.g. at lag 5, it will check the correlation but remove the effects already explained by lags 1 to 4.