Analysis of Time Series

Prasanta K. Jana
IEEE Senior Member

Department of Computer Science and Engineering
Indian Institute of Technology (ISM), Dhanbad

September 21, 2018

Introduction to Time Series

- A sequence of observations of collected at regular time intervals.
- The values are typically measured at equal time intervals (e.g. hourly, daily, weekly).
- Establishes relation between "Cause" and "Effects".
- One variable is "Time" which is independent and the second variable is "Data" which is dependent.
- Mathematically, a Time Series is defined as $Y = F(t)$

Importance of Time Series Analysis

Time Series Analysis finds applications in:

- Stock Market Analysis
- Economic and Sales Forecasting
- Budgetary Analysis
- Utility Studies
- Inventory Studies
- Yield Projections
- Work load Projections
- Process and Quality control
- Medical treatments
- Scientific and Engineering experiments
- Observation of natural phenomena (such as atmosphere, temperature, wind, earthquake)

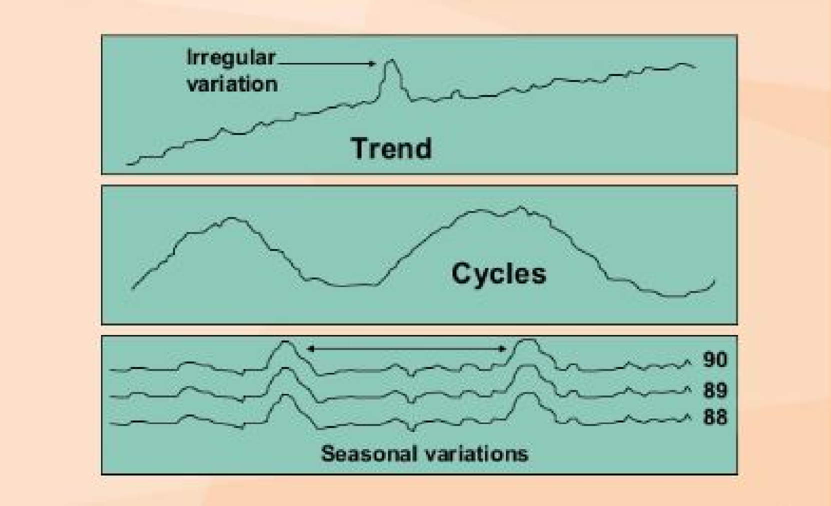
Components of Time Series

The changes in Time Series due to Economic, Social, Natural, Industrial and Political reasons are called components of Time Series.

They are broadly classified into four categories:

- Secular Trend or Trend
- Seasonal Variation
- Cyclic Variation
- Irregular or Random Variation

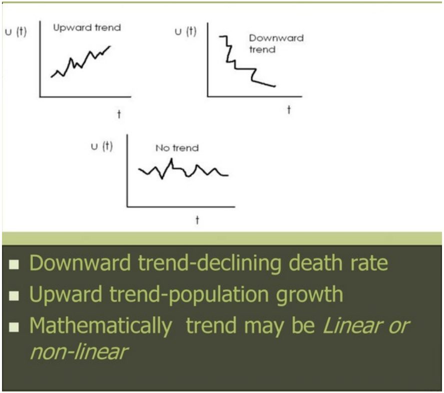
Components of Time Series



Secular Trend

- The valus of the variable tends to increase or decrease over a long period of time is called Secular Trend.

- A time series data may show upward trend or downward trend for a period of years and this may be due to :
  - increase in population
  - change in technological progress
  - large scale shift in consumer demands



Seasonal Variation

Seasonal Variations are short term fluctuations in time series which occur periodically in a year.

The factors that cause seasonal variations are :

- Climate and weather conditions
- Customs, traditions and habits

For example :

- Crops are sown and harvested at certain times every year and demand for labour goes up during that time.
- Demands for woolen clothes goes up in winter.
- Price increases during festivals
- Withdraws increase during first week of month

Cyclic Variations

- Long term movements that represent consistently recurring rises and declines in activity.
- Usually lasts longer than a year.
- The variations are not as regular as Seasonal Variations.
- Consists of four phases :

- Prosperity
- Recession
- Depression
- Recovery

Irregular Variations

- Also called Erratic, Random or Accidental variations.
- Does not repeat in a definite pattern.
- Short in duration and unpredictable
- Referred to as Residual Variation since they represent what is left out in time series after Trend, Seasonal and Cyclic Varitions.
- Results due to occurence of unforeseen events like :
    - Flood
    - Earthquake
    - War
    - Famine

Mathematical Model of Time Series

Time Series modelling is referred to as the decomposition of Time Series into four basic movements namely,

- Trends
- Seasonal Movements
- Cyclic Movements
- Irregular Movements

There are two types of models :

- Additive Model
- Multiplicative Model

Additive Model

- Data is the sum of the Time Series components.
  $Y = T + S + C + I$
  where,
  - $Y =$ Original Data (Observed vale o tim series)
  - $T =$ Trend
  - $S =$ Seasonal Variation
  - $C =$ Cyclic Variation
  - $I =$ Irregular Variation
- If the data do not contain one of the components, the value for that missing component is zero.
- The Seasonal component is independent of Trend and thus the magnitude of seasonal swing is constant over time.
- This model works best when time series has roughly same variability through the length of the series i.e. values of series falls within a band with constant width centered on the Trend.

Multiplicative Model

- Data is the product of the Time Series components.
  Y = T x S x C x I
  where,
  - Y = Original Data
  - T = Trend
  - S = Seasonal Variation
  - C = Cyclic Variation
  - I = Irregular Variation
- If the data do not contain one of the components, the value for that missing component is 1.
- The Seasonal component is proportional to Trend and thus the magnitude of seasonal swing increases or decreases according to the behaviour of Trend.
- This model works best when variability of time series increase with level i.e. value of series becomes larger as Trend increases.

The secular variation of a time series is its long-term non-periodic variation.

The following methods are used for calculation of Trend.

- Free Hand Curve Method
- Semi Average Method
- Moving Average Method
- Least Square Method

Out of these four methods, the Moving Average Method and the Least Square Method are widely which are explained in the following slides.

Semi-Average Method

- Step 1 : Divide the time series into the two equal segments.
- Step 2: Compute the Arithmetic Mean (AM) for each segment.
- Step 3: Draw a straight line passing through the AM.
  showing the trend of the series.

Note : For odd number of periods.Drop the value of the middle period.

Semi-Average Method

Trend Equation:

$Y = a + bX$

Where : $a = \frac{S_1+S_2}{t_1+t_2}$ & $b = \frac{S_2-S_1}{t_1(n-t_2)}$

Where : S_i = Sum of Part i

t_i = No. of Periods of Part i

n = Total Observed values

Semi-Average Method

Illustration :

Fit a trend line using the method of Semi-Average for the following data.

Period:	Apr.	May.	June	July	Aug.	Sep.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.
Sales:	28.0	30.0	28.0	28.0	27.0	24.0	23.0	23.0	22.0	20.0	21.0	20.0

Solution: Average of the first half series $= \frac{28+30+28+28+27+24}{6} = 27.5$

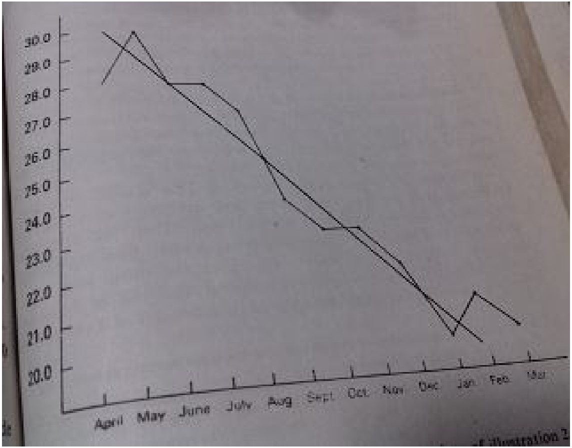
Average of the second half series $= \frac{23+23+22+20+21+20}{6} = 21.5$

The straight line trend:

$a = \frac{S_1+S_2}{t_1+t_2} = \frac{165+129}{12} = 24.5$ $b = \frac{S_2-S_1}{t_1(n-t_2)} = \frac{129-165}{6x6} = -1.0$

The Trend equation is $Y = 24.5 - 1.0X$

Semi-Average Method



Trend line by semi-average method

Moving Average Method

- Used for calculating long term Trend. This method is also used for Seasonal Fluctuations, Cyclic Fluctuations and Irregular Fluctuations. We calculate the Moving Average for several years.
- Tends to reduce the amount of variation present in the data set.
- The process of replacing Time Series by its Moving Average eliminates unwanted fluctuations and is therefore also referred to as "Smoothing of Time Series".
- For example, if we calculate "Three years Moving Average" then according to this method, it will be :

$$= \frac{(1)+(2)+(3)}{3}, \quad \frac{(2)+(3)+(4)}{3}, \quad \frac{(3)+(4)+(5)}{3}, \quad \ldots\ldots\ldots\ldots$$

where, (1), (2), (3), ... are the various years of Time Series.

❑ **Example:** Find out the five year's moving Average:

Year	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Price	20	25	33	33	27	35	40	43	35	32	37	48	50	37	45

Year (1)	Price of sugar (Rs.) (2)	Five year's moving Total (3)	Five year's moving Average (Col 3/5) (4)
1982	20	-	-
1983	25	-	-
1984	33	135	27
1985	30	150	30
1986	27	165	33
1987	35	175	35
1988	40	180	36
1989	43	185	27
1990	35	187	37.4
1991	32	195	39
1992	37	202	40.4
1993	48	204	40.8
1994	50	217	43.4
1995	37	-	-
1996	45	-	-

Least Square Method

- A Trend line is fitted to data in such a manner that the following two conditions are met :
  - The sum of deviations of the actual values of Y and the computed values of Y is zero.
    $\sum Y - Y_c = 0$
  - The sum of squares of the deviations of the actual values of Y and the computed values of Y is least from this line. That is why this method is called as method of Least Squares.
    $\sum (Y - Y_c)^2$ is least
- The line obtained by this method is called line of "Best Fit".

Least Square Method

- The method of Least Square can be used either to fit a straight line Trend or a parabolic Trend.
- The straight line trend is represented by the equation :
  $Y = a + bX$
  where,
  - $Y =$ Trend value to be computed
  - $X =$ Unit of Time (independent variable)
  - $a,b =$ Constants to be calculated

For example,

Draw a straight line trend and estimate trend value for 1996:

Year	1991	1992	1993	1994	1995
Production	8	9	8	9	16

Least Square Method

Year (1)	Deviation From 1990 X (2)	Y (3)	XY (4)	X^2 (5)	Trend $Y_c = a + bx$ (6)
1991	1	8	8	1	$5.2 + 1.6(1) = 6.8$
1992	2	9	18	4	$5.2 + 1.6(2) = 8.4$
1993	3	8	24	9	$5.2 + 1.6(3) = 10.0$
1994	4	9	36	16	$5.2 + 1.6(4) = 11.6$
1995	5	16	80	25	$5.2 + 1.6(5) = 13.2$
N= 5	$\sum x = 15$	$\sum y = 50$	$\sum xy = 166$	$\sum x^2 = 55$	

Least Square Method

$\sum Y = Na + b\sum X$
$\sum XY = a\sum X + b\sum X^2$
Now we put the values of these :
50 = 5a + 15b
166 = 15a + 55b
Multiplying first eq. by 3 and subtracting it from second eq. we get
-10b = -16
b = 1.6
Now we put the value of b in first equation to get value of a
5a + 15(1.6) = 50
a = 5.2
As according the value of a and b the Trend line :
Y = a + bX
Y = 5.2 + 1.6X
Now we calculate the Trend line for 1996 :
Y = 5.2 + 1.6(6) = 14.8

Time Series Forecasting

- The most common application of time series analysis is forecasting future values of a numeric value using the temporal structure of data.

- The available observations are used to predict values from the future.

- The most widely used model for Time Series Analysis is called Autoregressive Moving Average (ARMA).
    - The model consists of two parts, an autoregressive (AR) part and a moving average (MA) part.
    - The model is usually then referred to as the $ARMA_{p,q}$ model where p is the order of the autoregressive part and q is the order of the moving average part.

Autoregressive Moving Average

Autoregressive Model

The AR_p is read as an Autoregressive Model of order p. Mathematically it is written as :

$$X_t = c + \varepsilon_t + \sum_{i=1}^{P} \phi_i X_{t-1} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

where ϕ_i are parameters to be estimated,
c is a constant
ϵ_t represents the white noise.

Moving Average Model

The notation MA_q refers to the moving average model of order q. Mathematically it is written as :

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

where θ_i are parameters of the model,
μ is the expectation of X_t
ϵ_{t-i} are white noise error terms.

Autoregressive Moving Average

The ARMA model combines p autoregressive terms and q moving-average terms.
Mathematically the model is expressed with the following formula

$$X_t = c + \varepsilon_t + \sum_{i=1}^{P} \phi_i X_{t-1} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

We see that $ARMA_{p,q}$ model is a combination of AR_p and MA_q models.

The AR part of the equation seeks to estimate parameters for X_{t-i} observations of in order to predict the value of the variable in X_t.
It is in the end a weighted average of the past values.

The MA section uses the same approach but with the error of previous observations, ϵ_{t-i}
So in the end, the result of the model is a weighted average.