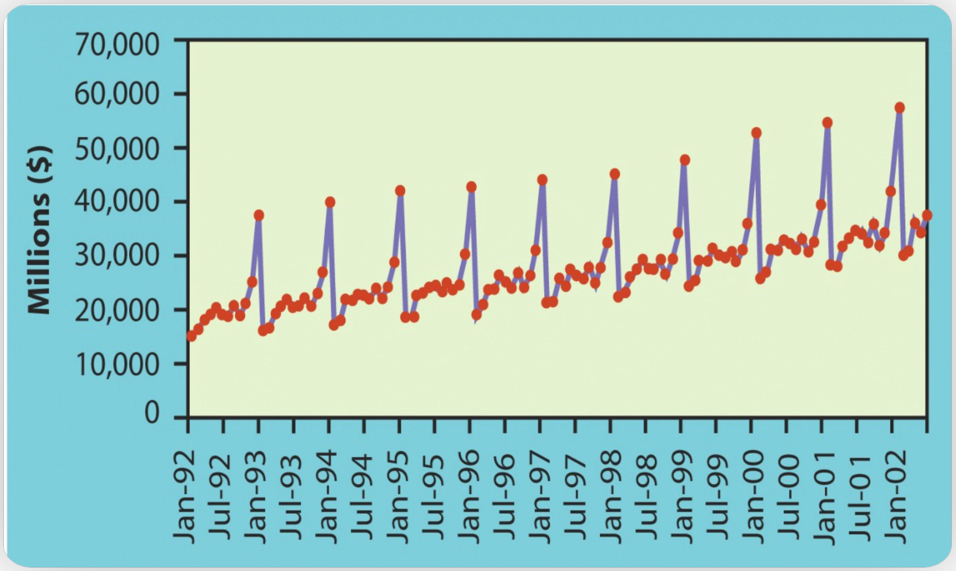
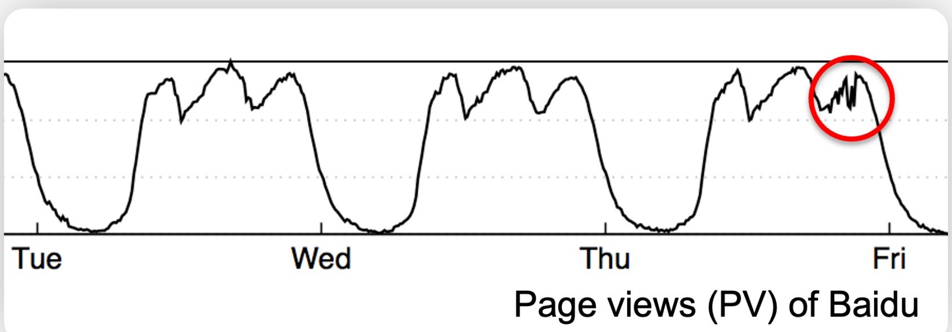
Time Series Forecasting and Anomaly Detection





Introduction to Time Series (I)

ZHANG RONG

Department of Social Networking Operations
Social Networking Group
Tencent Company

November 20, 2017

https://zhuanlan.zhihu.com/p/32584136

Outline

Outline

Time Series

Definition and Methods

Definition of Time Series

A time series is a series of data points indexed in time order.
Methods for time series analysis may be divided into two classes:

- Frequency-domain methods: spectral analysis and wavelet analysis;
- Time-domain methods: auto-correlation and cross-correlation analysis.

Methods of Time Series

Methods for time series analysis may be divided into another two classes:

- Parametetic methods
- Non-parametetic methods

Moving Average

Moving Average

Let $\{x_i : i \geq 1\}$ be an observed data sequence. A simple moving average (SMA) is the unweighted mean of the previous w data. If the w-days' values are $x_i, x_{i-1}, \ldots, x_{i-(w-1)}$, then the formula is

$$M_i = \frac{1}{w}\sum_{j=0}^{w-1} x_{i-j} = \frac{x_i + x_{i-1} + \cdots + x_{i-(w-1)}}{w}.$$

When calculating successive values, a new value comes into the sum and an old value drops out, that means

$$M_i = M_{i-1} + \frac{x_i}{w} - \frac{x_{i-w}}{w}.$$

Moving Average

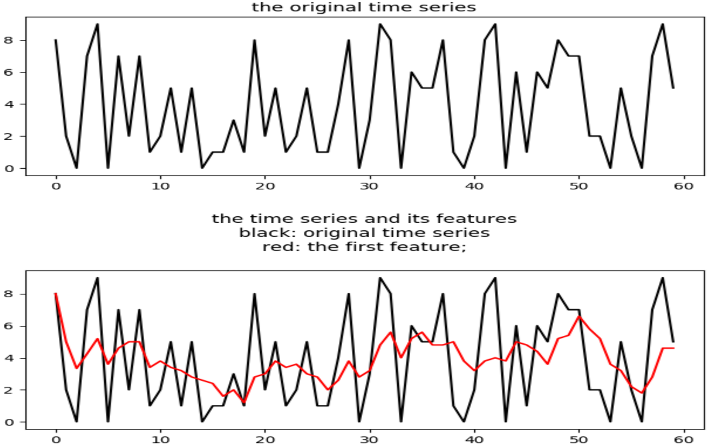


Figure: Moving Average Method for $w = 5$

Cumulative Moving Average

Cumulative Moving Average

Let $\{x_i : i \geq 1\}$ be an observed data sequence. A cumulative moving average is the unweighted mean of all datas. If the w-days values are $x_1, \cdots, x_i$, then

$$CMA_i = \frac{x_1 + \cdots + x_i}{i}.$$

If we have a new value x_{i+1}, then the cumulative moving average is

$$\begin{aligned} CMA_{i+1} &= \frac{x_1 + \cdots + x_i + x_{i+1}}{i+1} \\ &= \frac{x_{i+1} + i\,\cancel{n} \cdot CMA_i}{i+1} \\ &= CMA_i + \frac{x_{i+1} - CMA_i}{i+1}. \end{aligned}$$

Weighted Moving Average

Weighted Moving Average

A weighted moving average is the weighted mean of the previous w-datas. Suppose $\sum_{j=0}^{w-1} weight_j = 1$ with all $weight_j \geq 0$, then the weighted moving average is

$$WMA_i = \sum_{j=0}^{w-1} weight_j \cdot x_{i-j}.$$

Weighted Moving Average

A Special Case

In particular, let $\{weight_j : 0 \le j \le w-1\}$ be a weight with

$$weight_j = \frac{w-j}{w+(w-1)+\cdots+1} \text{ for } 0 \le j \le w-1.$$

In this situation,

$$WMA_i = \frac{wx_i + (w-1)x_{i-1} + \cdots + 2x_{i-w+2} + x_{i-w+1}}{w+(w-1)+\cdots+1}.$$

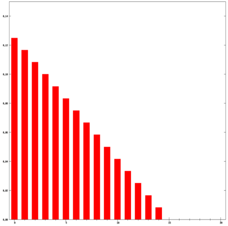


Figure: WMA weights $w = 15$

Weighted Moving Average

A Special Case

Weighted Moving Average

Suppose

$$
\begin{aligned}
Total_i &= x_i + \cdots + x_{i-w+1}, \\
Numerator_i &= wx_i + (w-1)x_{i-1} + \cdots + x_{i-w+1},
\end{aligned}
$$

then the update formulas are

$$
\begin{aligned}
Total_{i+1} &= Total_i + x_{i+1} - x_{i-w+1}, \\
Numerator_{i+1} &= Numerator_i + wx_{i+1} - Total_i, \\
WMA_{i+1} &= \frac{Numerator_{i+1}}{w + (w-1) + \cdots + 1}.
\end{aligned}
$$

Exponential Weighted Moving Average

Exponential Weighted Moving Average

Suppose $\{Y_t : t \geq 1\}$ is an observed data sequence, the exponential weighted moving average series $\{S_t : t \geq 1\}$ is defined as

$$S_t = \begin{cases} Y_1, & t = 1 \\ \alpha \cdot Y_t + (1 - \alpha) \cdot S_{t-1}, & t \geq 2 \end{cases}$$

- $\alpha \in [0, 1]$ is a constant smoothing factor.
- Y_t is the observed value at a time period t.
- S_t is the value of the EMWA at any time period t.

Exponential Weighted Moving Average

Moreover, from above definition,

$$S_t = \alpha[Y_t + (1-\alpha)Y_{t-1} + \cdots + (1-\alpha)^k Y_{t-k}] + (1-\alpha)^{k+1} S_{t-(k+1)}$$

for any suitable $k \in \{0, 1, 2, \cdots\}$. The weight of the point Y_{t-i} is $\alpha(1-\alpha)^{i-1}$.

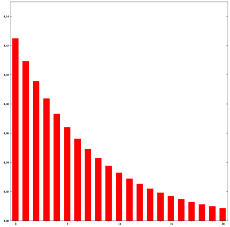


Figure: EMA weights $k = 20$

Exponential Weighted Moving Average

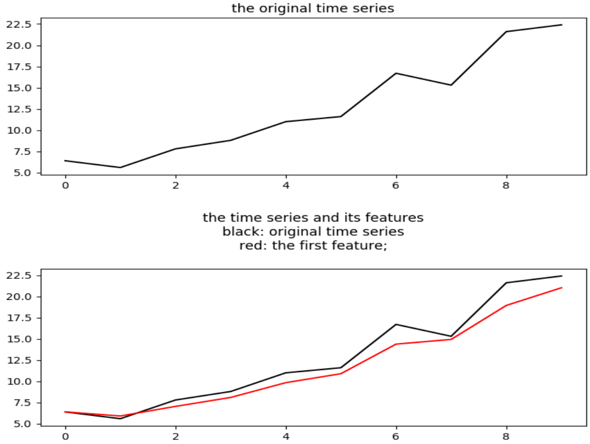


Figure: Exponential Weighted Moving Average Method for $\alpha = 0.6$

Double Exponential Smoothing

Double Exponential Smoothing

Suppose $\{Y_t : t \geq 1\}$ is an observed data sequence, there are two equations associated with double exponential smoothing:

$$
\begin{aligned}
S_t &= \alpha Y_t + (1-\alpha)(S_{t-1} + b_{t-1}), \\
b_t &= \beta(S_t - S_{t-1}) + (1-\beta) b_{t-1},
\end{aligned}
$$

where $\alpha \in [0,1]$ is the data smoothing factor and $\beta \in [0,1]$ is the trend smoothing factor.

Double Exponential Smoothing

Double Exponential Smoothing

Here, the initial values are $S_1 = Y_1$ and b_1 has three possibilities:

$$\begin{aligned} b_1 &= Y_2 - Y_1, \\ b_1 &= \frac{(Y_2 - Y_1) + (Y_3 - Y_2) + (Y_4 - Y_3)}{3} = \frac{Y_4 - Y_1}{3}, \\ b_1 &= \frac{Y_n - Y_1}{n - 1}. \end{aligned}$$

Forecast

- The one-period-ahead forecast is given by $F_{t+1} = S_t + b_t$.
- The *m*-period-ahead forecast is given by $F_{t+m} = S_t + mb_t$.

Double Exponential Smoothing

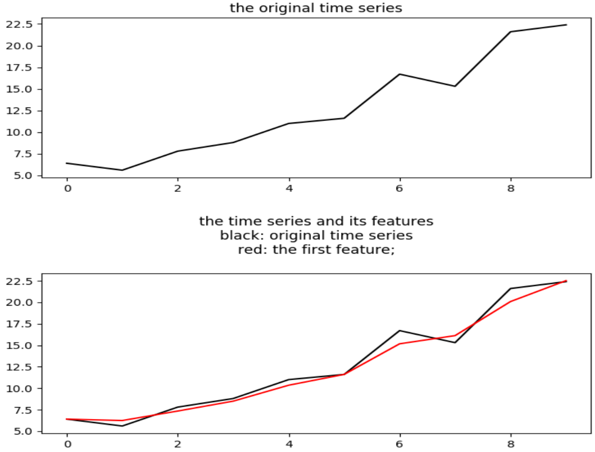


Figure: Double Exponential Smoothing for $\alpha = 0.6$ and $\beta = 0.4$

Triple Exponential Smoothing

Additive Seasonality

a.k.a Holt-Winters

Triple Exponential Smoothing (Additive Seasonality)

Suppose $\{Y_t : t \geq 1\}$ is an observed data sequence, then the triple exponential smoothing is

$$S_t = \alpha(Y_t - c_{t-L}) + (1-\alpha)(S_{t-1} + b_{t-1}), \quad \text{Overall Smoothing}$$
$$b_t = \beta(S_t - S_{t-1}) + (1-\beta)b_{t-1}, \quad \text{Trend Smoothing}$$
$$c_t = \gamma(Y_t - S_{t-1} - b_{t-1}) + (1-\gamma)c_{t-L}, \quad \text{Seasonal Smoothing}$$

We wish to estimate C_t at every time t mod L in the cycle that the observations take on

where $\alpha \in [0,1]$ is the data smoothing factor, $\beta \in [0,1]$ is the trend smoothing factor, $\gamma \in [0,1]$ is the seasonal change smoothing factor.

The *m*-period-ahead forecast is given by

$$F_{t+m} = S_t + mb_t + c_{(t-L+m) \bmod L}.$$

Triple Exponential Smoothing

Multiplicative Seasonality

Triple Exponential Smoothing (Multiplicative Seasonality)

Suppose $\{Y_t : t \geq 1\}$ is an observed data sequence, then the triple exponential smoothing is

$$
\begin{aligned}
S_t &= \alpha \frac{Y_t}{c_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}), && \text{Overall Smoothing} \\
b_t &= \beta(S_t - S_{t-1}) + (1-\beta)b_{t-1}, && \text{Trend Smoothing} \\
c_t &= \gamma \frac{Y_t}{S_t} + (1-\gamma)c_{t-L}, && \text{Seasonal Smoothing}
\end{aligned}
$$

where $\alpha \in [0,1]$ is the data smoothing factor, $\beta \in [0,1]$ is the trend smoothing factor, $\gamma \in [0,1]$ is the seasonal change smoothing factor.

Triple Exponential Smoothing

Multiplicative Seasonality

Forcast

The *m-period-ahead forecast* is given by
$F_{t+m} = (S_t + mb_t)c_{(t-L+m) \mod L}$.

Triple Exponential Smoothing

Initial values are

$$
\begin{aligned}
S_1 &= Y_1, \\
b_0 &= \frac{(Y_{L+1} - Y_1) + (Y_{L+2} - Y_2) + \cdots + (Y_{L+L} - Y_L)}{L}, \\
c_i &= \frac{1}{N} \sum_{j=1}^{N} \frac{Y_{L(j-1)+i}}{A_j}, \forall i \in \{1, \cdots, L\}, \\
A_j &= \frac{\sum_{i=1}^{L} Y_{L(j-1)+i}}{L}, \forall j \in \{1, \cdots, N\}.
\end{aligned}
$$

Time series Decomposition

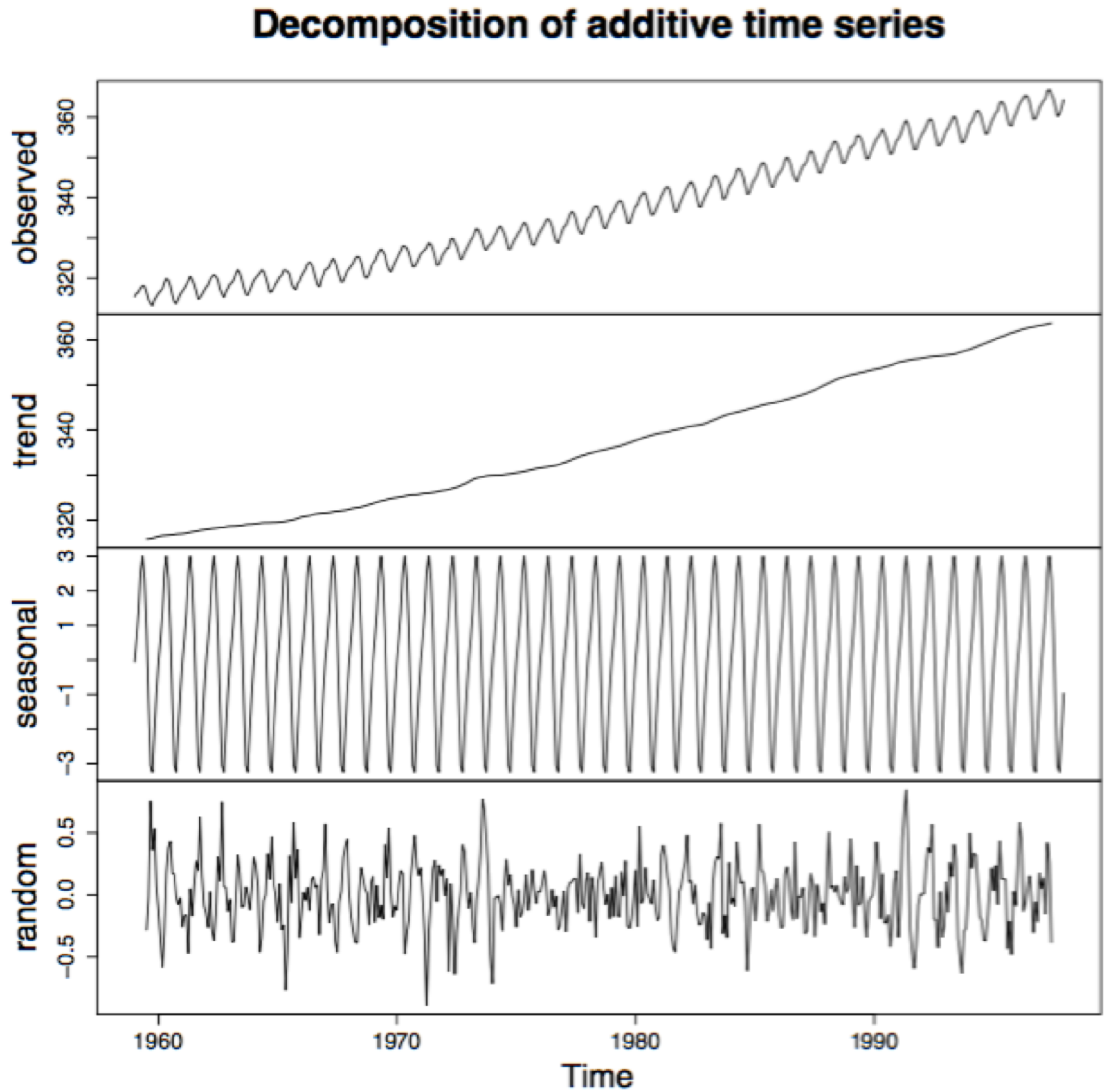
Farideh Dehkordi-Vakil

Introduction

- One approach to the analysis of time series data is based on smoothing past data in order to separate the underlying pattern in the data series from randomness.
- The underlying pattern then can be projected into the future and used as the forecast.

Introduction

- The underlying pattern can also be broken down into sub patterns to identify the component factors that influence each of the values in a series.
- This procedure is called **decomposition**.
- Decomposition methods usually try to identify two separate components of the basic underlying pattern that tend to characterize economics and business series.
  - Trend Cycle
  - Seasonal Factors



Decomposition returned by the R package `forecast`.

Introduction

- The Trend Cycle represents long term changes in the level of series.
- The Seasonal factor is the periodic fluctuations of constant length that is usually caused by known factors such as rainfall, month of the year, temperature, timing of the Holidays, etc.
- The decomposition model assumes that the data has the following form:

$$\text{Data} = \text{Pattern} + \text{Error}$$

$$= f\,(\text{trend cycle, Seasonality , error})$$

Decomposition Model

- Mathematical representation of the decomposition approach is:

$$Y_t = f(S_t, T_t, E_t)$$

  - Y_t is the time series value (actual data) at period t.
  - S_t is the seasonal component (index) at period t.
  - T_t is the trend cycle component at period t.
  - E_t is the irregular (remainder) component at period t.

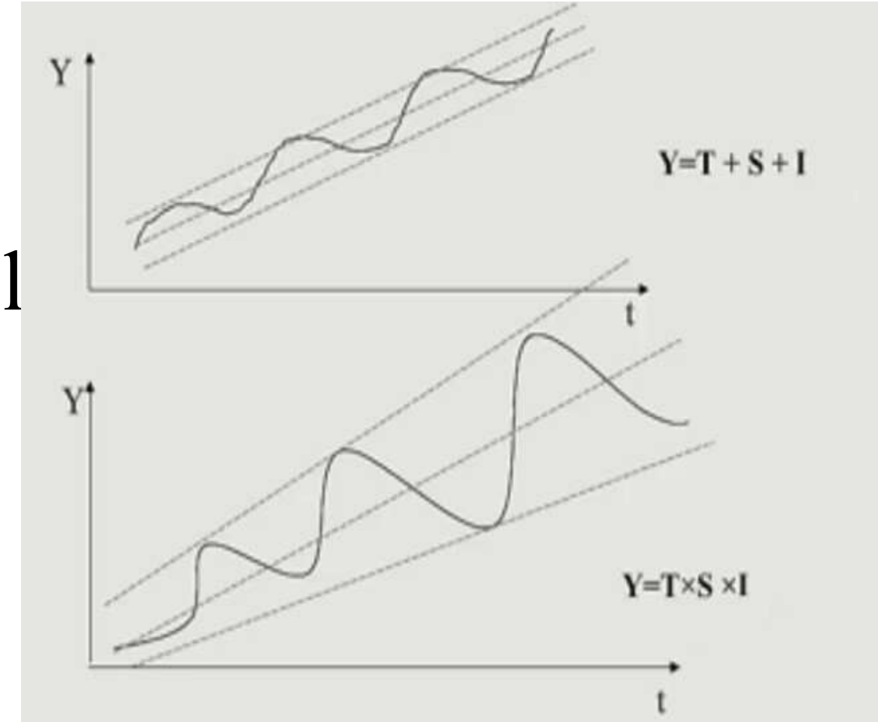
Decomposition Model

- The exact functional form depends on the decomposition model actually used. Two common approaches are:
- Additive Model

$$Y_t = S_t + T_t + E_t$$

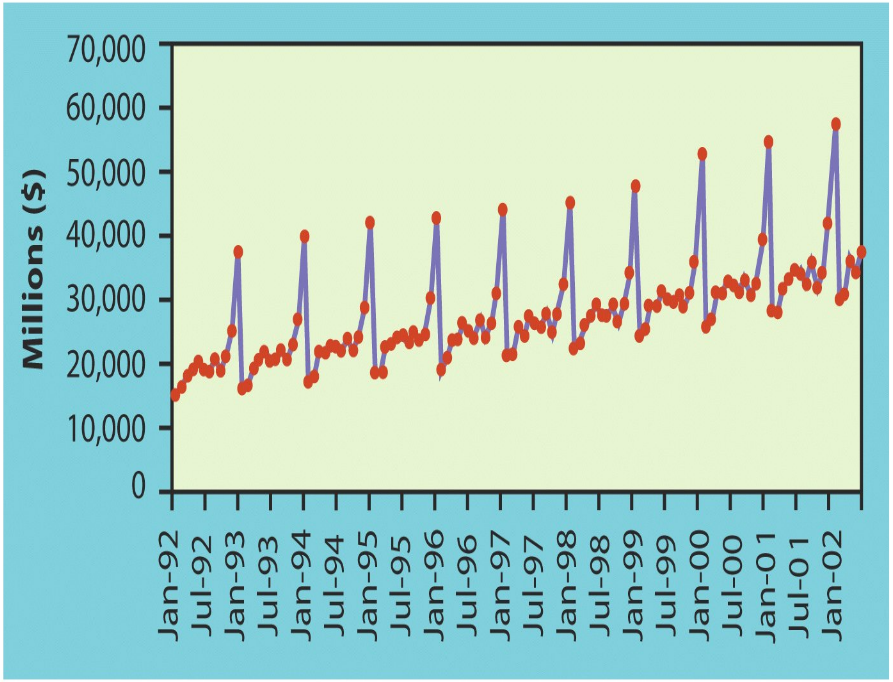
- Multiplicative Model

$$Y_t = S_t \times T_t \times E_t$$



Decomposition Model

- An additive model is appropriate if the magnitude of the seasonal fluctuation does not vary with the level of the series.
- Time plot of U.S. retail Sales of general merchandise stores for each month from Jan. 1992 to May 2002.



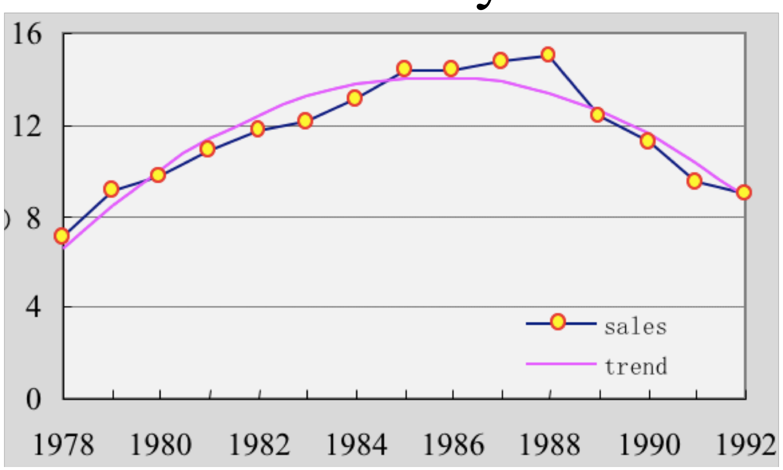
$$Y_t = T_t + S_t + E_t$$
$$Y_t = T_t \times S_t \times E_t$$

Trend-Cycle Estimation

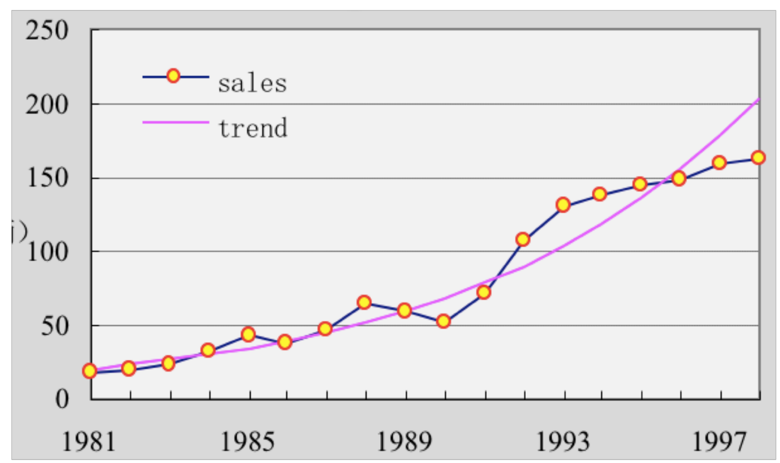
- Step 1. Estimate the Trend-Cycle
  - Moving Average
    - Simple moving average
  - Local Regression Smoothing
    - Least squares estimates

Trend-Cycle Estimation

- Instead of fitting one straight line to the entire dataset, a series of straight lines will be fitted to sections of the data.
- A straight trend line is not always appropriate, there are many time series where some curved trend is better. Then the trend maybe like these:



$$T_t = a + bt + ct^2$$



$$T_t = at^b$$

$$Y_t = T_t + S_t + E_t \qquad Y_t - T_t = S_t + E_t$$
$$Y_t = T_t \times S_t \times E_t \qquad Y_t / T_t = S_t \times E_t$$

Step 2: Estimating Seasonal factors

A. Seasonal factors are estimated using the de-trended series. Season *L* can be day, week, month,…

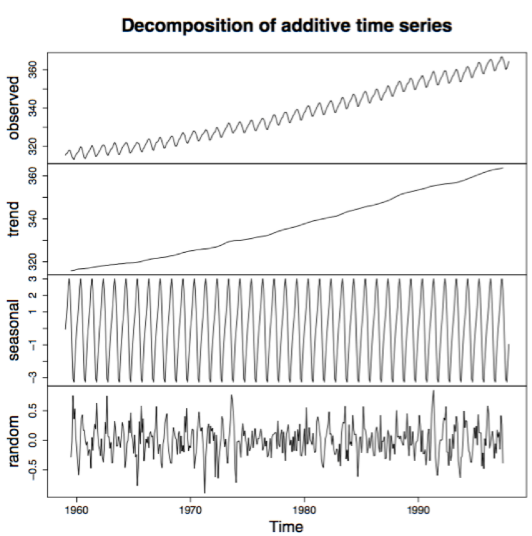
B. L=day, compute the average of S_t= Y_t – T_t (t=0h,1h,2h,…23h) for all days in the dataset.

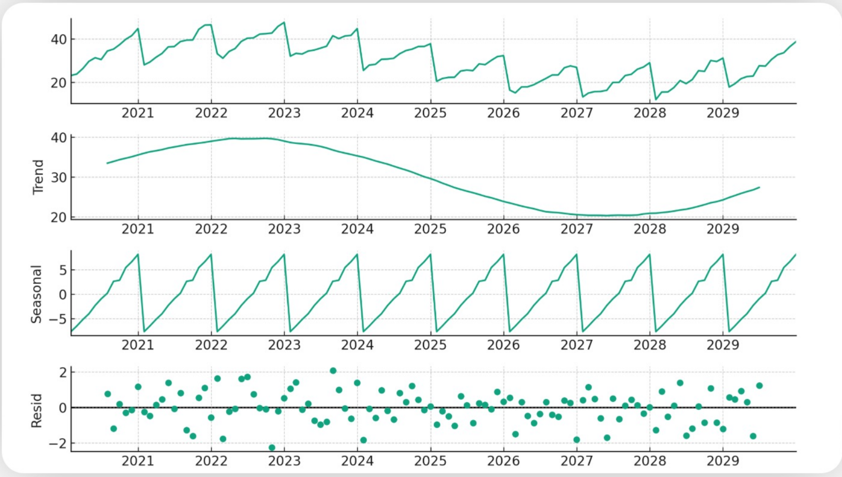
C. E_t = Y_t –T_t – S_t (random = series – trend – seasonal)

$$Y_t = T_t + S_t + E_t$$
$$Y_t = T_t \times S_t \times E_t$$

conclusion

- Till now, we finish decomposition of the original time series.
  - We can use the estimation of T_t and S_t for forecasting.
  - Use the E_t for anomaly detection.





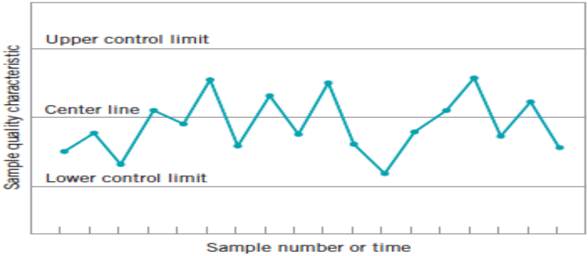
Outline

Definition of Control Chart Theory

Control Chart

The control chart is a graphical display of a quality characteristic that has been measured from a sample versus the sample number or time.

- Center Line: the average value of the quality characteristic
- Upper Control Limit (UCL) and Lower Control Limit (LCL): two horizontal lines.



3σ Control Chart

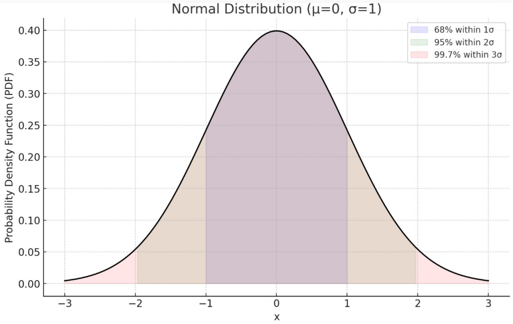
Simplest Control Chart

3σ Control Chart

Suppose that w is a sample statistic that measures some quality characteristic, the mean of w is μ_w and the standard deviation of w is σ_w. Then the center line, the upper control limit and the lower control limit becomes:

$$\begin{aligned} \text{UCL} &= \mu_w + L\sigma_w \\ \text{Center line} &= \mu_w \\ \text{LCL} &= \mu_w - L\sigma_w \end{aligned}$$

where L is the "distance" of the control limits from the center line, expressed in standard deviation units. In particular, if $L = 3$, then it is the 3σ control chart.



Here's the replot of the normal distribution. This graph visualizes the probability density function with the mean (μ) at 0 and standard deviation (σ) of 1. It shows the areas under the curve corresponding to within one, two, and three standard deviations from the mean, covering approximately 68%, 95%, and 99.7% of the data respectively.

The Cumulative Sum Control Chart

CUSUM Control Chart

Let x_i be the i-th observation on the process $\{x_i : 1 \leq i \leq n\}$, $\{x_i : 1 \leq i \leq n\}$ has a normal distribution with mean μ and standard deviation σ. The cumulative sum control chart is calculated by, for all $1 \leq i \leq n$,

$$C_i = \sum_{j=1}^{i}(x_j - \mu_0) = C_{i-1} + (x_i - \mu_0),$$

where $C_0 = 0$ and μ_0 is the target for the process mean.

- If $|C_i|$ exceed the decision interval H, then the process is considered to be out of control.
- The decision interval H is 3σ or 5σ.

The Cumulative Sum Control Chart

Data for the Cusum Example

Data for the Cusum Example

Sample, i	(a) x_i	(b) $x_i - 10$	(c) $C_i = (x_i - 10) + C_{i-1}$
1	9.45	−0.55	−0.55
2	7.99	−2.01	−2.56
3	9.29	−0.71	−3.27
4	11.66	1.66	−1.61
5	12.16	2.16	0.55
6	10.18	0.18	0.73
7	8.04	−1.96	−1.23
8	11.46	1.46	0.23
9	9.20	−0.80	−0.57
10	10.34	0.34	−0.23
11	9.03	−0.97	−1.20
12	11.47	1.47	0.27
13	10.51	0.51	0.78
14	9.40	−0.60	0.18
15	10.08	0.08	0.26
16	9.37	−0.63	−0.37
17	10.62	0.62	0.25
18	10.31	0.31	0.56
19	8.52	−1.48	−0.92
20	10.84	0.84	−0.08
21	10.90	0.90	0.82
22	9.33	−0.67	0.15
23	12.29	2.29	2.44
24	11.50	1.50	3.94
25	10.60	0.60	4.54
26	11.08	1.08	5.62
27	10.38	0.38	6.00
28	11.62	1.62	7.62
29	11.31	1.31	8.93
30	10.52	0.52	9.45

The Cumulative Sum Control Chart

The first 20 of these observations were drawn at random from a normal distribution with $\mu = 10$ and standard deviation $\sigma = 1$. They are plotted on a Shewhart control chart.

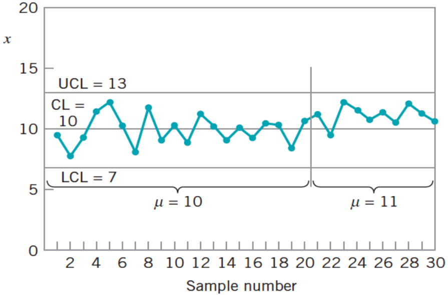


Figure: A Shewhart control chart for the data

The Cumulative Sum Control Chart

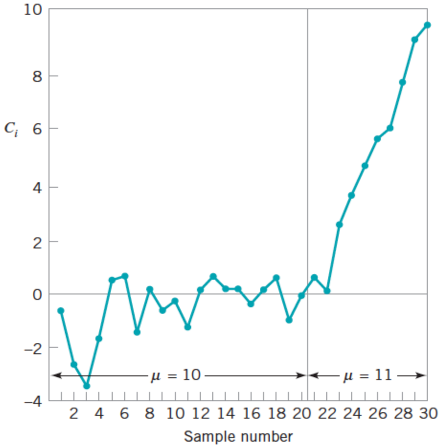


Figure: Plot of the cumulative sum from column (c) in above table

The Cumulative Sum Control Chart

Comparison to three-sigma control limit

Difference

- Three-sigma control limit: one or more points beyond a three-sigma control limit
- CUSUM control limit: it is a good choice when small shifts are important.

TSFRESH python package

TSFRESH python package

- tsfresh is used to to extract characteristics from time series.
- Paper: Time Series Feature extraction based on scalable hypothesis tests
- Spend less time on feature engineering
- Automatic extraction of 100s of features

TSFRESH python package

TSFRESH python package

Let $\{x_1, \cdots, x_n\}$ be a time series, some features are

- max, min, median, mean μ, variance σ^2, standard deviation σ,
- range is maximum minus minimum
- skewness is the third standardized moment:

$$skewness = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^3,$$

- kurtosis is the fourth standardized moment:

$$kurtosis = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^4.$$

TSFRESH python package

TSFRESH python package

Let $\{x_1, \cdots, x_n\}$ be a time series, some features are

- absolute energy: $E = \sum_{i=1}^{n} x_i^2$,
- absolute sum of changes: $E = \sum_{i=1}^{n-1} |x_{i+1} - x_i|$,
- aggregate autocorrelation:

$$\frac{1}{n-1} \sum_{\ell=1}^{n} \frac{1}{(n-\ell)\sigma^2} \sum_{t=1}^{n-\ell} (x_t - \mu)(x_{t+\ell} - \mu),$$

- autocorrelation: parameter is lag ℓ,

$$\frac{1}{(n-\ell)\sigma^2} \sum_{t=1}^{n-\ell} (x_t - \mu)(x_{t+\ell} - \mu).$$

TSFRESH python package

TSFRESH python package

Let $\{x_1, \cdots, x_n\}$ be a time series, some features are

- count above mean, count below mean
- variance larger than standard deviation
- first location of maximum, first location of minimum
- last location of maximum, last location of minimum
- has duplicate, has duplicate max, has duplicate min
- longest strike above mean, longest strike below mean

TSFRESH python package

TSFRESH python package

Let $\{x_1, \cdots, x_n\}$ be a time series, some features are

- mean change: $\sum_{i=1}^{n-1}(x_{i+1} - x_i)/n = (x_n - x_1)/n$
- mean second derivative central:

$$\frac{1}{n}\sum_{i=1}^{n-2}\frac{1}{2}(x_{i+2} - 2 \cdot x_{i+1} + x_i)$$

- percentage of reoccurring data points to all data points
- percentage of reoccurring values to all values
- ratio value number to time series length
- sum of reoccurring data points
- sum of reoccurring values

TSFRESH python package

Initialization of Time Series

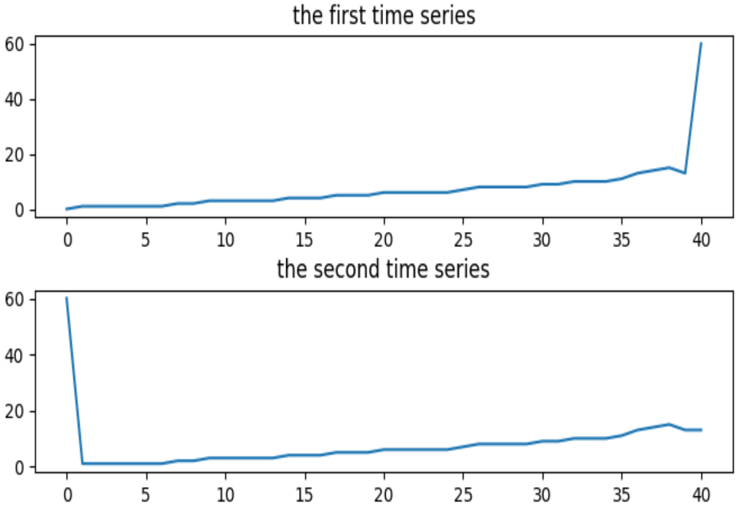
Let $\{x_1, \cdots, x_n\}$ be a time series, some initialization methods are, for $1 \leq i \leq n$,

$$
\begin{aligned}
y_i &= \frac{x_i}{mean(\{x_i : 1 \leq i \leq n\})}, \\
y_i &= \frac{x_i}{median(\{x_i : 1 \leq i \leq n\})}, \\
y_i &= \frac{x_i}{\max - \min}, \\
y_i &= \frac{x_i}{(\max - \min)/10},
\end{aligned}
$$

where max and min denotes the maximum and minimum value of the time series, respectively.

TSFRESH python package

Example of Two Lists



TSFRESH python package

Features of the Above Two Lists

nonParametersFeatures	th	value_list1	value_list2
feature	0	60	60
feature	1	0	1
feature	2	7.19512195122	7.51219512195
feature	3	85.0350981559	84.493753718
feature	4	9.22144772559	9.1920483962
feature	5	4.71450748799	4.67091571882
feature	6	26.5796091617	26.2452662595
feature	7	6.0	6.0
feature	8	5609	5778
feature	9	64	75
feature	10	1	1
feature	11	15	16
feature	12	26	25
feature	13	0.975609756098	0.0
feature	14	0.0	0.0243902439024
feature	15	1.0	0.0243902439024
feature	16	0.0243902439024	0.170731707317
feature	17	True	True
feature	18	False	False
feature	19	False	True
feature	20	15	15
feature	21	26	25
feature	22	1.6	1.875
feature	23	1.5	-1.175
feature	24	0.589743589744	0.75641025641
feature	25	0.625	0.666666666667
feature	26	0.853658536585	0.878048780488
feature	27	0.390243902439024 4	0.365853658536585 36
feature	28	188	201
feature	29	61	61
feature	30	295	308
feature	31	60	59

Thank you for watching!

https://zhuanlan.zhihu.com/p/32584136

ZHANG RONG
zr9558@gmail.com
zr9558.wordpress.com