VIVEKANANDA COLLEGE
THAKURPUKUR
KOLKATA-700063

NAAC ACCREDITED 'A' GRADE

Topic: Time_Series_Analysis

Course Title: Stationary time series

Paper: CC10

Unit: 3 & 4

Semester: 4

Name of the Teacher: Sutapa Biswas

Name of the Department: Statistics

Stationary Time Series

A time series is said to be stationary if there is no systematic change in mean (no trend), if there is no systematic change in ~~variation~~ variance, and if strictly periodic variations have been removed.

Most of the probability theory of time series is concerned with stationary time series, and for this reason time series analysis often requires one to turn a non-stationary series into a stationary one so as to use this theory. For example it may be of interest to remove the trend and seasonal variation from a set of data and then try to model the variation in the residuals by means of a stationary stochastic process.

[Stochastic Process is a family of random variables $\{X_t, t \in T\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$, where T is an index set. For each fixed $t \in T$, X_t is a function on the set Ω. For each fixed $\omega \in \Omega$, $X_t(\omega)$ is a function on T. Here $\Omega = \{\omega\}$ is the set of all elementary events ω, $\mathcal{F}$ collection of all subsets A of Ω and P, a probability measure.

In a stochastic time series, Y_t; $t \in \mathbb{Z} = \{0, \pm 1, \pm 2, \cdots\}$ is a family of random variables, Y_t denoting the value of characteristic of interest at time t. Thus, $y = (y_1, \cdots, y_n)'$ is seen as a realised value of of a random vector $Y = (Y_1, Y_2, \cdots, Y_n)'$ with joint probability density function $f_Y(y)$. The joint distribution function of a finite set of random variables

$$\{Y_{t_1}, Y_{t_2}, \cdots, Y_{t_n}\}, \quad t_1 < t_2 < \cdots < t_n$$

from the collection $\{Y_t, t \in T\}$ is

$$F_{Y_{t_1}, \cdots, Y_{t_n}}(y_{t_1}, \cdots, y_{t_n}) = P[Y_{t_1} \le y_{t_1}, \cdots, Y_{t_n} \le y_{t_n}],$$
$$(y_{t_1}, \cdots, y_{t_n}) \in \mathbb{R}^n. \ ]$$

Mathematical Definition of a Stationary Time Series

Denote by $(\Omega, \mathcal{F}, P)$, the probability space and T be an index set. Then a real valued time series (stochastic process) is a real valued function $x(t, \omega)$ belonging to the product space $T \times \Omega$.

For a fixed $t \in T$, $x(t, \omega)$ is a random variable defined over Ω and is denoted by X_t. The time series $\{X_t, t \in T\}$ is therefore a collection of all such random variables.

The series $\{X_t\}$ is said to be <u>strictly stationary</u> if

$$F_{X_{t_1}, \cdots, X_{t_n}}(x_{t_1}, \cdots, x_{t_n}) = F_{X_{t_1+h}, \cdots, X_{t_n+h}}(x_{t_1}, \cdots, x_{t_n})$$

for $(t_1, \cdots, t_n)$ and $(t_1+h, \cdots, t_n+h) \in T$ and

$(X_{t_1}, \ldots, X_{t_n})$ lies in the range of the random variables X_t.

Note: i) The distribution of any point in the index set remains the same.

ii) From (i) it is clear that the joint distribution of a finite number of points in the index set do not involve the points themselves, instead it is a function of the distance between points.

iii) $t_1, \ldots, t_n$ are not necessarily consecutive.

iv) If the second order moment $E(X_t^2) < \infty$, then

$$\left.\begin{array}{l} E(X_t) = \text{constant for all } t \\ Var(X_t) = \text{constant for all } t \end{array}\right\} \text{ for a such stationary time series (process).}$$

Recall that $\{X_t, t \in T\}$ is covariance stationary if $E(X_t) = \text{constant}$ and $cov(X_t, X_{t+h}) = \gamma(h)$ for all $t \in T$.

strictly stationary $\not\Rightarrow$ covariance stationary, since the variance may not exist.

Also covariance stationary $\not\Rightarrow$ strictly stationary.

$\gamma(h)$ = auto covariance for covariance stationary (or weakly or second order stationary).

$\rho(h) = \dfrac{\gamma(h)}{\gamma(0)}$ auto correlation function for cov. stationary process.

[A less restrictive requirement, called weak stationarity of order f, is that moments of the joint distribution $f(x_{t_1}, \ldots, x_{t_n})$ upto order f depend only on the time differences.

Thus a time series $\{X_t, t \in T\}$ is weak stationary of order 2 or is covariance stationary or is simply stationary if:

(i) $E(X_t) = \mu \quad \forall t$

(ii) Covariance matrix of $(X_{t_1}, \ldots, X_{t_n})$ = Covariance matrix of $(X_{t_1+h}, \ldots, X_{t_n+h})$ for all non-empty finite set of indices $(t_1, \ldots, t_n)$ and all such 'h' such that $(t_1+h, \ldots, t_n+h)$ are contained in the index set T.

Since a multivariate normal distribution is characterised by its two moments moments, if a Gaussian process is stationary of order 2, then it must be strictly stationary.

A strictly stationary process with the first two moments finite is also covariance stationary. However, a strictly stationary time series may not posses finite (first two) moments and hence may not be covariance stationary].

Some stationary process useful in representing different time series models:

(1) <u>Purely random process</u>: A time series $\{X_t, t \in T\}$ will be called a purely random process if the random variables, X_t's are, a collection of iid random variables with,

$E(X_t) = \mu$, $Var(X_t) = \sigma^2$ and $Cov(X_t, X_{t+h}) = \gamma(h) = 0$

A purely random process $\{X_t, t \in T\}$ is also called White Noise (W.N.) when the process is a discrete parameter-time series. e.g. - autoregressive process.

$$Z_t = \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + \cdots + \alpha_p Z_{t-p} + \epsilon_t \text{ ; where } \epsilon_t \text{ is a White Noise.}$$

Purely random processes are used as constituent of more complicate stationary processes viz, moving average process, ARMA (Auto Regressive Moving Average) Process etc.

(2) <u>Moving Average process</u> [MA(q)]:

A time series $\{X_t, t \in T\}$ is called a moving average process of order q if it can be expressed in the form

$$X_t = \beta_0 \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q} \text{ ; } \beta\text{'s are constants (or parameter) } \beta_0 = 1$$

and $\{\epsilon_t\}$ is a White Noise.

This is simply a moving of a random series generating a stationary process with $E(X_t) = 0 \ \forall t$

& $\quad var(X_t) = \sigma^2 \sum_{i=0}^{q} \beta_i^2 \quad \forall t$

$$\gamma(k) = Cov(X_t, X_{t+k}) = Cov(\beta_0 \epsilon_t + \cdots + \beta_q \epsilon_{t-q}, \beta_0 \epsilon_{t+k} + \cdots + \beta_q \epsilon_{t+k-q})$$

$$= \begin{cases} 0 & \text{if } k > q \\ \sigma^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k} & \text{if } k \le q \ \forall t \\ \gamma(-k) & \text{if } k < 0 \end{cases}$$

Hence this MA(q) is covariance stationary (or weakly stationary)

also
$$\rho(k) = \begin{cases} 0 & \text{if } k > q \\ 1 & \text{if } k = 0 \\ \sum_{i=0}^{q-k} \beta_i \beta_{i+k} \Big/ \sum_{i=0}^{q} \beta_i^2 & \text{if } k = 1, 2, \ldots, q \end{cases}$$

and $\rho(k) = \rho(-k)$ if $k < 0$

(3) Autoregressive process [AR(p)]

A time series $\{x_t\}$ is called an autoregressive process of order p if

$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + \epsilon_t$, where ϵ_t is a White Noise, with $E(\epsilon_t)=0$, $Var(\epsilon_t)=\sigma^2$, $cov(\epsilon_t, \epsilon_{t'})=0$ if $t \neq t'$.

Here, the process arises from oscillatory series giving rise to an autoregressive scheme.

(i) when $p=1$, this is simply a Markov process.

(ii) " $p=2$, this is a Yule process

For $p=1$,
$$\begin{aligned} x_t &= \alpha x_{t-1} + \epsilon_t \\ &= \alpha(\alpha x_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \alpha^2 x_{t-2} + \alpha \epsilon_{t-1} + \epsilon_t \\ &\;\vdots \\ &= \epsilon_t + \alpha\epsilon_{t-1} + \alpha^2 \epsilon_{t-2} + \cdots \quad \text{if } |\alpha|<1 \\ &= \beta_0 \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \cdots \quad \text{where } \beta_i = \alpha^i;\ i \geq 0 \end{aligned}$$

is a MA(∞) process if $|\alpha|<1$.

In general an AR(p) is an MA(∞) process (A duality property between AR(p) and MA(∞) exists), if all the roots of the characteristic equation

$x^p - \alpha_1 x^{p-1} - \alpha_2 x^{p-2} - \cdots - \alpha_p = 0$ are less than one in absolute value.

Back to AR(1), $E(x_t) = \mu \sum_{i\geq 0} \beta_i = 0$ if $\mu = 0$ & $\sum_{i\geq 0} \beta_i^2 < \infty$ i.e. $\sum_{i \geq 0} |\beta_i| < \infty$

$V(x_t) = \sigma^2 \sum_{i \geq 0} \beta_i^2$

and satisfied if $|\alpha|<1$

Then $\gamma(k) = Cov(x_t, x_{t+k}) = \sigma^2 \sum_{i\geq 0} \beta_i \beta_{i+k} < \infty$ if $|\alpha|<1$ (by Cauchy-Schwartz)

$\therefore$ AR(1) is covariance stationary.

[An informal proof of "AR(p) is an MA(∞) process".

$$\begin{aligned} x_t &= \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \epsilon_t \\ &= \alpha_1 B x_t + \cdots + \alpha_p B^p x_t + \epsilon_t \quad \text{where} \end{aligned}$$

B is the back shift operator such that $x_{t-p} = B^p x_t$

Therefore, white noise,

$$\epsilon_t = \underbrace{(1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p)}_{f(B)} x_t = f(B)\, x_t \text{ say.}$$

$\therefore$ If $f(B)$ is invertible, we have, $x_t = f^{-1}(B)\, \epsilon_t$

$\therefore\ X_t = (1-\alpha_1 B - \cdots - \alpha_p B^p)^{-1} \epsilon_t$

$= \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \cdots$ where $\beta_0 = 1$

AR(p) can be expressed as MA(∞)

i.e. $X_t = \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \cdots$; $\beta_0 = 1$

Like AR(2) and AR(1) ;

$E(X_t) = 0 \quad \forall\ t = 0, \pm 1, \pm 2, \cdots$

$Var(X_t) = \sigma^2 \sum_{i=0}^{\infty} \beta_i^2 \quad \forall\ t = 0, \pm 1, \pm 2, \cdots$

and $\gamma(k) = \sigma^2 \sum_{i=0}^{\infty} \beta_i \beta_{i+k}$.

Autocorrelation Function

The autocorrelation function expresses how the correlation between any two values of the series changes as their extent of seperation in time changes.

Since $\rho_k = \rho_{-k}$, the autocorrelation function is symmetric about $k=0$.

The graph of autocorrelation function ρ_k against k is called autocorrelogram.

If we have N observations $x_1, \cdots, x_N$ in a time series, a satisfactory estimate of ρ_k of k-th lag correlation, is

$$\hat{\rho}_k = r_k = \frac{c_k}{c_0}$$

where $c_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$, $\bar{x} = \frac{1}{N} \sum_{t=1}^{N} x_t$

In c_k we have used in the denominator N in place of $(N-k)$, This ensure that the estimated autocovariance matrix Γ_N is positive semi-definite.

To obtain useful estimates of autocorrelation function, we would need atleast 50 observations and r_k should be calculated for $k = 0, 1, \cdots, K$ where $K \le \frac{N}{4}$.

The sample autocorrelation function is a listing or graph of the sample autocorrelations r_k at lags $k = 1, 2, \cdots$.

Correlogram Analysis

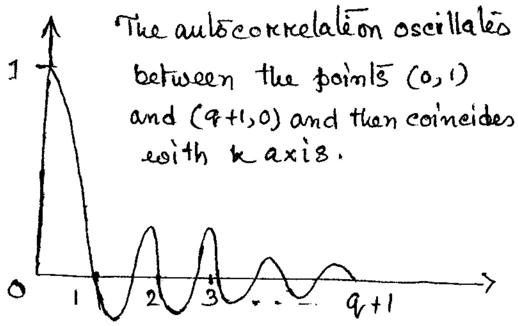
Consider a covariance stationary time series $\{X_t\}$. Let ρ_k be the autocorrelation of lag k. Different covariance stationary time series possesses different forms of correlograms, ρ_k, when plotted against k is called the correlogram. By analysing the shape of the correlogram we can infer about the interval structure of $\{X_t\}$. For different well known stochastic models we present the correlograms.

1) MA (q)

$x_t = \beta_0 \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}$; $\beta_0 = 1$

where ϵ_t is white noise with $E(\epsilon_t) = 0$; $var(\epsilon_t) = \sigma^2$ and

$$\rho_k = \begin{cases} 1 & \text{if } k=0 \\ \dfrac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^{q} \beta_i^2} & \text{for } k=1,2,\cdots q \\ 0 & \text{if } k>q \end{cases}$$



The autocorrelation oscillates between the points (0,1) and (q+1,0) and then coincides with k axis.

In particular, $\beta_i = \frac{1}{q+1}$ $\forall i$, indicates Simple MA process

Then $$\rho_k = \begin{cases} 1 - \dfrac{k}{q+1} & \text{if } k = 0,1,2,\cdots,q \\ 0 & \text{if } k>q \end{cases}$$

If it is a simple MA process then autocorrelation will decline smoothly, o.w., it may have some oscillatory movement depending on the coefficients β_i, and it will ultimately coincide with k-axis for $k \geq q+1$.

2) 1st order autoregressive (or Markov) process:

$x_t = \alpha x_{t-1} + \epsilon_t$; $|\alpha| < 1$ [this assumption is needed to make time series stationary]

where ϵ_t's are white noise with $E(\epsilon_t) = 0$ and $var(\epsilon_t) = \sigma^2$.

$$x_t = \alpha(\alpha x_{t-2} + \epsilon_{t-1}) + \epsilon_t$$
$$\vdots$$
$$= \alpha^n x_{t-n} + \sum_{i=0}^{n-1} \alpha^i \epsilon_{t-i}$$

will be stationary and convergent if $|\alpha| < 1$

$x_t \sim \sum_{i=0}^{\infty} \alpha^i \epsilon_{t-i}$ is MA (∞) process if $|\alpha|<1$

Now $E(x_t) = 0$ and $Var(x_t) = \sigma^2 \sum_{i=0}^{\infty} \alpha^{2i} = \frac{\sigma^2}{1-\alpha^2}$ if $|\alpha| < 1$

and $$\gamma_k = E(x_t x_{t+k}) = E\left[\left(\sum_{i=0}^{\infty} \alpha^i \epsilon_{t-i}\right)\left(\sum_{j=0}^{\infty} \alpha^j \epsilon_{t+k-j}\right)\right]$$

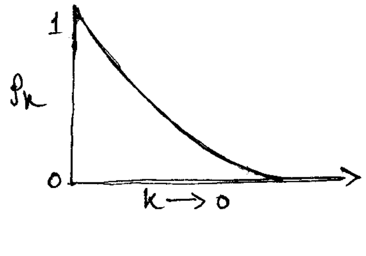
$$= E\left[\left(\sum_{l=-\infty}^{t} \alpha^{t-l} \epsilon_l\right)\left(\sum_{m=-\infty}^{t+k} \alpha^{t+k-m} \epsilon_m\right)\right] \quad \text{where } l = t-i \ ; \ m = t+k-j \ ;$$

$$= \sigma^2 \alpha^k \sum_{i=0}^{\infty} \alpha^{2i} = \frac{\sigma^2 \alpha^k}{1-\alpha^2}.$$

$\therefore$ $\rho_k = \alpha^k$

To get an even function defined for all integer k, $\rho_k = \alpha^{|k|}$; $k = 0, \pm 1, \pm 2, \cdots$

As $k \to \infty$, $\rho_k \to 0$ if $|\alpha| < 1$



3) Correlogram of AR (2) [Yule Process]

Consider $X_t = a X_{t-1} + b X_{t-2} + \epsilon_t$; where ϵ_t is white noise with $E(\epsilon_t)=0$, $var(\epsilon_t) = \sigma^2$.

Multiplying bothsides by X_{t-k}, taking expectation and dividing by $var(X_t)$ we get

$\rho_k = a\rho_{k-1} + b\rho_{k-2}$ (Yule-Walker equation)

homogeneous difference equation of order k in ρ_k's.

The general solution is given by

$\rho_k = A_1 \pi_1^k + A_2 \pi_2^k$; $|\pi_1|<1$, $|\pi_2|<1$

where A_1 & A_2 has to be found from initial conditions and π_1 & π_2 are roots of equation $\pi^2 - a\pi - b = 0$ (Characteristic or auxiliary equation of the difference equation)

Case (i) : π_1 and π_2 are real and unequal :

i.e. $\frac{a \pm \sqrt{a^2+4b}}{2}$ are real number i.e. if $a^2+4b > 0$

A_1 and A_2 are found as follows:

Since $\rho_0 = 1 = A_1 + A_2$ from first Yule-Walker equation.

we have $\rho_1 = a\rho_0 + b\rho_{-1} = a + b\rho_1$

$\therefore \rho_1 = \frac{a}{1-b}$

Also $\rho_1 = A_1\pi_1 + A_2\pi_2$

$= A_1\pi_1 + (1-A_1)\pi_2 = A_1(\pi_1 - \pi_2) + \pi_2$

$\therefore A_1 = \left(\frac{a}{1-b} - \pi_2\right) / (\pi_1 - \pi_2)$ and $A_2 = 1 - A_1$

(ii) For real equal roots of the characteristic equation $\pi_1 = \pi_2 = \pi$.

The general solution is given by $\rho_k = (C_1 + C_2 k)\pi^k$.

$\therefore \rho_k = \left\{1 + k\left(\frac{1-\pi^2}{1+\pi^2}\right)\right\}\pi^k \to 0$ as $k \to \infty$

In both the cases $\rho_k \to 0$ as $k \to \infty$ and the correlogram is oscillatory and decaying in character.

(iii) π_1, π_2 are complex conjugates if $a^2+4b<0$

write $\pi_1 = p(\cos\theta + i\sin\theta)$

and $\pi_2 = p(\cos\theta - i\sin\theta)$

$\therefore \rho_k = p^k [A_1(\cos\theta + i\sin\theta)^k + A_2(\cos\theta - i\sin\theta)^k]$

$= p^k [A_1(\cos k\theta + i\sin k\theta) + A_2(\cos k\theta - i\sin k\theta)]$

$= p^k [A^* \cos k\theta + B^* \sin k\theta]$; where $A^* = A_1 + A_2$ and $B^* = i(A_1 - A_2)$

If $k = 0$; $\rho_0 = 1 = A^*$

$k = 1$; $\rho_1 = p\,[\cos\theta + B^*\sin\theta]$

$k = -1$; $\rho_{-1} = p^{-1}\,[\cos\theta - B^*\sin\theta]$

Since $\rho_1 = \rho_{-1}$; equating the above two equations

$$B^* = \frac{1-p^2}{1+p^2}\,\frac{\cos\theta}{\sin\theta} = \frac{1-p^2}{1+p^2}\cot\theta = \cot\psi \text{ (say)}.$$

Now $\rho_k = p^k\,[\cos\theta k + \cot\psi\sin\theta k] = p^k\,\dfrac{\sin(\theta k+\psi)}{\sin\psi} \longrightarrow 0$

$\left[\text{since } \left|\dfrac{\sin(\theta k+\psi)}{\sin\psi}\right| \le c\right]$

Now $p^2 = \pi_1\pi_2$ and also from theory of quadratic equations $\pi_1\pi_2 = -b$

Since $a^2 + 4b < 0 \Rightarrow b < 0$

$\therefore\ p = \sqrt{-b} < 1$ if $|\pi_1| < 1$, $|\pi_2| < 1$

$\therefore\ \rho_k \to 0$ as $k \to \infty$ and ρ_k is damped sine curve.

Exponential Smoothing

Given a stationary, non-renewal time series $x_1, x_2, \dots, x_n$, it is to take as an estimate of x_{N+1}. The weighted sum of past observations

$$\hat{x}(N,1) = c_0 x_N + c_1 x_{N-1} + \cdots \quad (1)$$

where $\{c_j\}$ are weights, geometric weights

$$c_j = \alpha(1-\alpha)^j \; ; \; j = 0, 1, 2, \dots$$

where $0 < \alpha < 1$ are chosen to give more weights to recent observations and less weights to the observations further in the past.

Since, in reality we have only finite number of observations, let us make some modification of observations and then (1) can be written as

$$\hat{x}_{N+1} = \hat{x}(N,1) = \alpha x_N + (1-\alpha)\left[\alpha x_{N-1} + \alpha(1-\alpha) x_{N-2} + \cdots\right]$$

$$= \alpha x_N + (1-\alpha)\hat{x}(N-1,1) \quad \text{---} \quad (2)$$

If we set $\hat{x}(1,1) = x_1$ then equation (2) can be used recursively to compute forecasts.

Also it reduces the amount of arithmatic involved, since forecasts can easily be updated using the latest observations and previous forecast. The procedure defined by (2) is called exponential (forecasting) smoothing.

The adjective 'exponential' arises from the fact that the geometric weights lie on a exponential curve.

Equation (2) can be written as

$$\hat{x}(N,1) = \alpha\left[x_N - \hat{x}(N-1,1)\right] + \hat{x}(N-1,1)$$

$$= \alpha e_N + \hat{x}(N-1,1) \; ; \quad \text{where } e_N = x_N - \hat{x}(N-1,1)$$

$= \text{prediction error at time } N.$

It can be shown that exponential smoothing is optimal if the underlying model for the time series is given by

$$x_t = \mu + \alpha \sum_{j<t} \epsilon_j + \epsilon_t$$

This infinite $MA(\infty)$ is non-stationary, but the first difference $W_t = x_t - x_{t-1}$ from $MA(1)$ process so that x_t is $ARIMA(0,1,1)$.

The value of the smoothing constant α depends on the properties of the given time series.

Values between 0·1 and 0·3 are often used. The value of α may be estimated by past data by a similar procedure as used in estimating the parameter of MA process. The sum of squared prediction errors is computed for different values of α and the values is chose which minimizes the

~~the~~ Sum of Squares i.e.

$\hat{x}(1,1) = x_1$, $e_2 = x_2 - \hat{x}(1,1)$.

$\hat{x}(2,1) = \alpha e_2 + \hat{x}(1,1)$, $e_3 = x_3 - \hat{x}(2,1)$ and so on.

In general $e_N = x_N - \hat{x}(N-1,1)$ $N = 2, 3, ---, N$.

Compute $\sum_{i=2}^{N} e_i^2$.

Repeat this procedure for different values of α between 0 & 1, say in steps of 0.1 and select the value which minimizes $\sum_{i=2}^{N} e_i^2$.

Usually the Sum of Squares surface is quite flat near the minimum and so the choice of α is not critical.