Lecture Notes #3

Time series analysis (STAT 5140/4140)

1 Examples of Stationary process

Example Consider the process $X_t = Sin(2\pi ut)$, with $u \sim Uniform(0, 1)$

$$\begin{aligned} E[X_t] &= \int_0^1 Sin(2\pi ut)du \\ &= -\tfrac{Cos(2\pi ut)}{2\pi t}\Big|_0^1 \\ &= \tfrac{1}{2\pi t}(Cos0 - Cos2\pi t) \\ &= 0 \end{aligned}$$

The auto-covariance function of the process is given by:

$$\begin{aligned} Cov(X_t, X_{t+h}) &= Cov\{Sin(2\pi ut), Sin(2\pi u(t+h))\} \\ &= E[Sin(2\pi ut)Sin(2\pi ut + 2\pi uh)] \\ &= \frac{1}{2}\int_0^1 2Sin(2\pi ut)Sin(2\pi ut + 2\pi uh)du \\ &= \frac{1}{2}\int_0^1 Cos(2\pi ut + 2\pi ut + 2\pi uh)du + \frac{1}{2}\int_0^1 Cos(2\pi ut - 2\pi ut + 2\pi uh)du \\ &= \frac{1}{2}\int_0^1 Cos(4\pi ut + 2\pi uh)du + \frac{1}{2}\int_0^1 Cos(2\pi uh)du \\ &= \frac{1}{2}\{\frac{Cos(4\pi ut + 2\pi uh)}{4\pi ut + 2\pi uh}\Big|_0^1\} + \frac{1}{2}\{\frac{Sin(2\pi uh)}{2\pi uh}\Big|_0^1\} \\ &= \frac{1}{2}\{\frac{Sin(2\pi(2t + uh)}{4\pi t + 2\pi h} + \frac{Sin(2\pi t}{2\pi t}\} \\ &= 0. \end{aligned} \tag{1}$$

Example Let $\{W_t\}$ is a Uncorrelated Normal process with $\mu = 0$ and $\sigma = 1$. Define $X_t = W_tW_{t-1}$, show that $\{X_t\}$ is stationary (weak). To prove this is a stationary process first compute its mean:

$$E[X_t] = E[W_tW_{t-1}] = E[W_t]E[W_{t-1}] = 0$$

Joint distribution of W_t, and W_{t-1} is bivariate normal, implies $rho(W_tW_{t-1}) = 0$. So the auto covariance function $\gamma(t, t+h) = Cov(W_tW_{t-1})$ is given by: When $h \neq 0$, $\gamma(t, t+h) = 0$. But $when h = 0$:

$$\begin{aligned}
\gamma(t, t+h) &= E[W_tW_{t-1}W_tW_{t-1}] \\
&= E[W_t^2]E[W_{t-1}^2] \quad , as \quad W_t^2 \sim \chi^2_{(1)} \\
&= 1
\end{aligned}$$

According to the definition, the process is stationary (weak).

Example Let $\{X_t\}$ be a normal process with mean μ_X and autocovariance function $\sigma(h)$. Define a new process $\{Y_t\}$ as $Y_t = exp(X_t)$. Using the moment generating function of a Normal distribution: $M_x(t) = exp(\mu_X t + \frac{1}{2}t^2\sigma^2)$. The mean of the process is given by:

$$E[Y_t] = E[exp(X_t)] = exp(\mu_X + \frac{1}{2}\sigma^2) = exp\{\mu_X + \frac{1}{2}\gamma(0)\}$$

Now the autocovariance function is:

$$\begin{aligned}
\gamma(t, t+h) &= Cov(Y_t, Y_{t+h}) \\
&= E[exp(X_t)exp(X_{t+h})] - E[exp(X_t)]E[exp(X_{t+h})] \\
&= E[exp(X_t + X_{t+h})] - E[exp(X_t)]E[exp(X_{t+h})] \\
&= E[exp(W_t)] - E[exp(X_t)]E[exp(X_{t+h})] \\
&= exp(\mu_W + \frac{1}{2}\sigma_W^2) - exp\{\mu_X + \frac{1}{2}\gamma(0)\}exp\{\mu_X + \frac{1}{2}\gamma(0)\} \\
&= exp\{2\mu_X + \frac{1}{2}2^2\gamma(0)\} - exp\{2\mu_X + \gamma(0)\} \\
&= exp\{2\mu_X + 2\gamma(0)\} - exp\{2\mu_X + \gamma(0)\}
\end{aligned}$$

Where $W_t = X_t + X_{t+h} \sim N(\mu_W = 2\mu_X, \sigma_W^2 = 2^2\gamma(0))$. This process is also a stationary process (weak).

Example Consider a sequence of uncorrelated random variables, $\{e_t\}$ with mean 0 and variance σ^2. This process is a stationary. As $E[e_t] = 0$, $Var(X_t) = \sigma^2$ and $\gamma(h) = 0$ for $h \neq 0$.

Definition *The process $\{e_t\}$ described above is called White noise process, and we will write as $e_t \sim WN(0, \sigma^2)$.*

Remark Every *iid* process is a WN process.

Example Consider the process $\{X_t\}$ defined as, $X_t = e_t + \theta e_{t-1}$, where $e_t \sim WN(0, \sigma^2)$ and $\theta \in \mathbb{R}$. For this process:

$$\begin{aligned} E[X_t] &= 0 \\ Var(X_t) &= E[X_t^2] = \sigma^2(1+\theta^2) \end{aligned} \tag{2}$$

and the autocovariance function is given by:

$$\gamma(h) = \begin{cases} \sigma^2(1+\theta^2), & \text{for } h = 0 \\ \theta\sigma^2 & \text{for } |h| = 1 \\ 0 & \text{for } |h| > 1 \end{cases}$$

This is called a MA(1) process. We will discuss the properties of this process later.

Example Consider a process $\{X_t\}$, where $X_t = \phi X_{t-1} + e_t$. It can be proved that the auto covariance function is given by $\gamma(h) = \phi^{|h|}\gamma(0)$. This process is called AR(1), we will discuss this process in detail later.

Example Consider a zero mean random process $\{X_t\}$. For prediction at time t based on the values X_{t-1} and X_{t-2} consider a linear predictor as:

$$\hat{X}_t = -d_1 X_{t-1} - d_2 X_{t-2}$$

. Goal is to find the coefficient d_1 and d_2 such that them mean square error $E[X_t - \hat{X}_t]$ is minimized. In later notes we will show that the optimal values of d_1 and d_2 can be obtain from the equation below:

$$\begin{bmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(2) \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} -\gamma(1) \\ -\gamma(2) \end{bmatrix}$$

This set of equations is call Yule Walker equations. Note to obtain a linear predictor we only need first and second order moments of the process. Also the estimated parameters are not dependent on time, this means this model can be used for future prediction.

2 Non-Stationary processes

If a process is not stationary (weak) then its called non-stationary process. There could be several factors that can cause a non-stationary process. Generally we will discuss two types of non-stationary processes: (1) Deterministic type and (2) Stochastic type.

For example look at the plot the air passenger data available in R.

As you can see from the plot there is a upward (linear/quadratic ?) trend in the data. Also note the variance is changing over time.

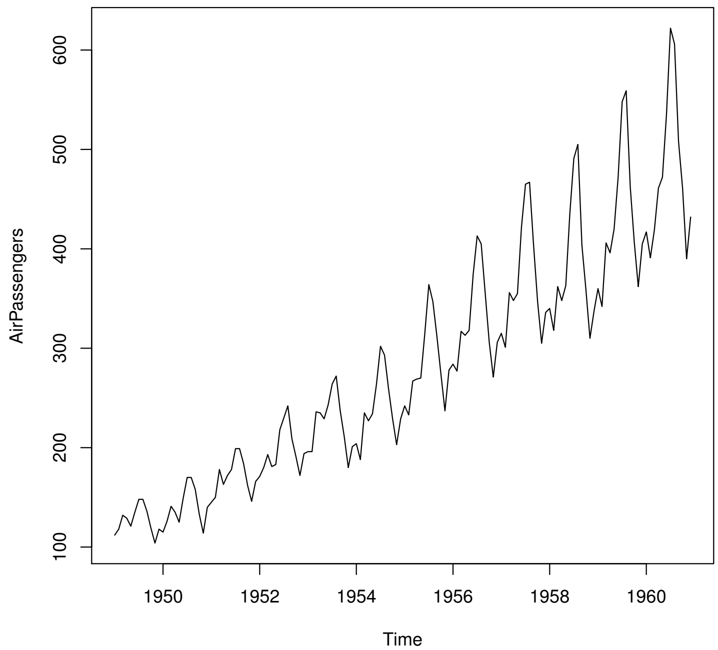


Figure 1: Air passenger data

2.1 Trend type models

A simple way of modeling trend type non-stationary model is to use the flowing model:

$$X_t = m(t) + Y_t$$

Where $\{Y_t\}$ is a stationary process. A simple version of of this model is:

$$X_t = m(t) + e_t \text{ where } e_t \sim WN(0, \sigma^2)$$

Where $m(t)$ is a deterministic p^{th} order polynomial function. If a linear trend is present choose

$$m(t) = a + bt$$

For Air Passenger data we fit a linear trend and plot the residuals. The plots are given below:

It is obvious that the residuals are not stationary but they show variance type non stationary (variance is a function of time).

2.2 Stochastic type non-stationary

There are random processes where the trend is not deterministic. Consider a white noise process $\{e_t\} \sim WN(0, \sigma^2)$. Define a random process as:

$$X_t = \sum_{k=0}^{t} e_k$$

. The process $\{X_t\}$ is a non-stationary process. Another way of writing the process is:

$$X_t = X_{t-1} + e_t$$

This process is called integrated type / random walk process. Some properties of the process are:

1. Given the history of the process up to $t-1$, the best prediction of $\hat{X}_t$ is X_{t-1}.
2. $E(X_t) = 0$.
3. $V(X_t) = t\sigma^2$, is a function of time.
4. From the construction of the model $X_t - X_{t-1}$ is white noise. But it is not necessary that one should obtain white noise after differencing the series once, multiple differencing may require.
5. This type of non-stationarity can be easily detected by visual inspection. Later we will discuss statistical test to detect integrated effect.
6. Run the codes in R to simulate such a process. Try to guess the trend for each simulation.

3 Modeling time series data

The main goal of time series data analysis is prediction. But is prediction always possible? Predictability is a generic trem and can largely depend on what type of model is being used. Most of the theories revolves around linear model. We will study linear models. As covariance is a measure of linear dependence between two random variables, we will study covariance first.

Definition *Auto covariance function (ACVF) of a random process is defined by:*

$$\gamma(h) = Cov(X_{t+h}, X_t)$$

Auto correlation function (ACF) is defined as:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

Properties of ACVF:

1. $\gamma(0) \geq 0$ [Which is the variance of the process]
2. $|\gamma(h)| \leq |\gamma(0)|$
3. $\gamma(h) = \gamma(-h)$

Definition *A sequence $\gamma(.)$ is called non-negative definite if:*

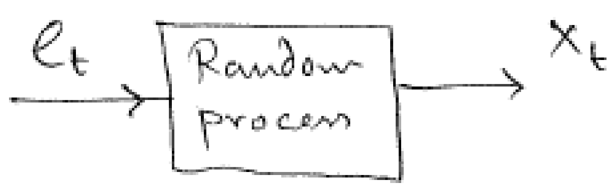
$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\gamma(|i-j|)a_j \geq 0$$

; for all $n \in \{0, 1, 2, \dots\}$ and any real number a_i.

Theorem *ACVF of a stationary process, X_t, is non-negative definite.*

3.0.1 White Noise process

One of the most important use of ACVF or ACF is to define a pure random process, which is back bone of linear random process theory. Consider e_t be a sequence of un-correlated random variables such that $E[e_t] = 0$ and $V[e_t] = \sigma^2$. The process is called White noise. As there is no correlation in the process a linear model is not useful to make prediction. For a good model the residuals must behave like White noise. As those are the part of the process which can't be predicted using a liner model. Unlike a deterministic process we do not know the input of a random process or if there is any input at all. For modeling purpose one can think a White noise, e_t, process as an input of the random process.



Note: White noise process is not an IID (Identical and Independent Distribution) process, but all IID processes are White noise process.

4 Linear random process

To come up with a model we will assume the process is linear and the input of the process is a White noise (This is an mathematical assumption only). Using the two assumptions we can write our model as:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j e_{t-j}$$

Where $\{e_t\} \sim WN(0, \sigma^2)$, and $\{\psi_j\}$ is a sequence of constants such that $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. But this model can't applied in real life for two reasons.

1. The model has infinite parameters. It is not possible to estimate infinite parameters.
2. The model depends on future terms. That is if X_t is your present state the model requires terms like $\{X_{t+1}, X_{t+2} \dots\}$. This type of model is not causal.

To overcome those problems we consider a simple linear model:

$$X_t = \sum_{j=0}^{q} \psi_j e_{t-j}$$

For uniqueness of the model we also assume $\psi_0 = 1$

Definition *A random process is called Moving Average of order q, $MA(q)$, if the model can be written as:*

$$X_t = e_t + \sum_{j=1}^{q} \theta_j e_{t-j}$$

Where $\{\theta_1 \dots \theta_{t-q}\}$ are real numbers and $\{e_t\} \sim WN(0, \sigma^2)$.

Note: X_t is a linear combination of $q+1$ white noise variables, that is X_t and X_{t+l} are uncorrelated for any $l > q$

Example Consider MA(1) process:

$$X_t = e_t + \theta_1 e_{t-1}$$

Easy to see that $E[X_t] = 0$, but in general the model can also be written as:

$$X_t = \mu + e_t + \theta_1 e_{t-1}$$

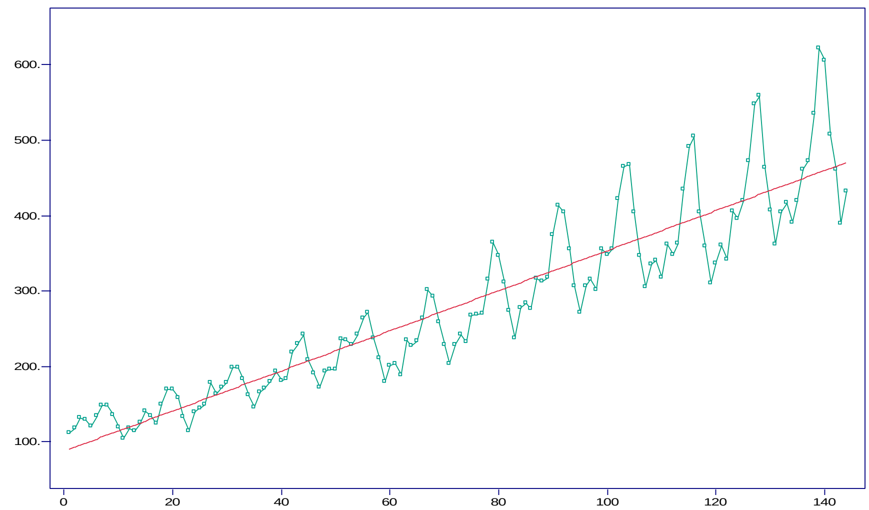
In this case the $E[X_t] = \mu$. To compute ACVF of this process:

$$\begin{aligned} Cov(X_t, X_{t+h}) &= Cov(e_t + \theta_1 e_{t-1}, e_{t+h} + \theta_1 e_{t-1+h}) \\ &= Cov(e_t, e_{t+h}) + \theta_1 Cov(e_t, e_{t+h-1}) + {\theta_1}^2 Cov(e_{t-1}, e_{t+h-1}) + \theta_1 Cov(e_{t-1}, e_{t+h}) \end{aligned} \tag{3}$$

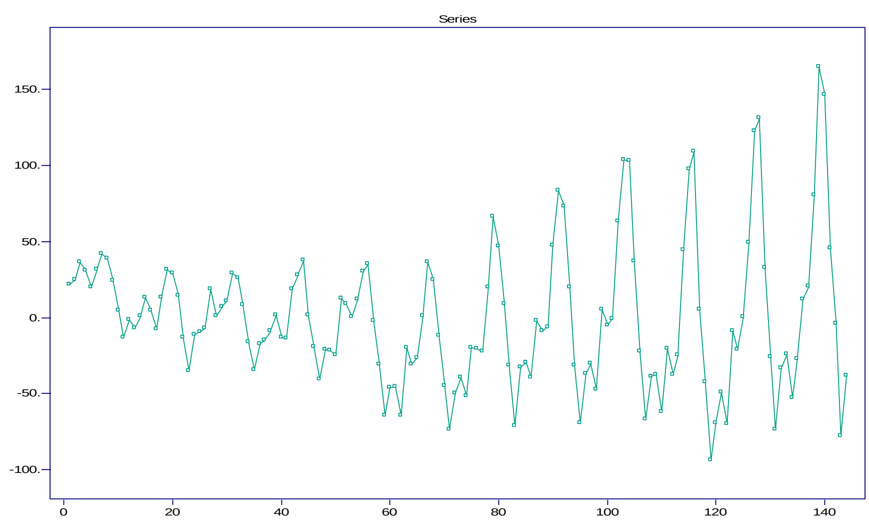
So the ACVF function is:

$$\gamma(h) = \begin{cases} \sigma^2(1 + {\theta_1}^2), & \text{for } h = 0 \\ \theta_1 \sigma^2 & \text{for } |h| = 1 \\ 0 & \text{for } |h| > 1 \end{cases}$$

Note the Auto covariance function is zero after lag=1 ($h = +1$ or -1). This property can be used to identify the $MA(1)$ process.



Linear Trend fit



Residual plot after the linear fit