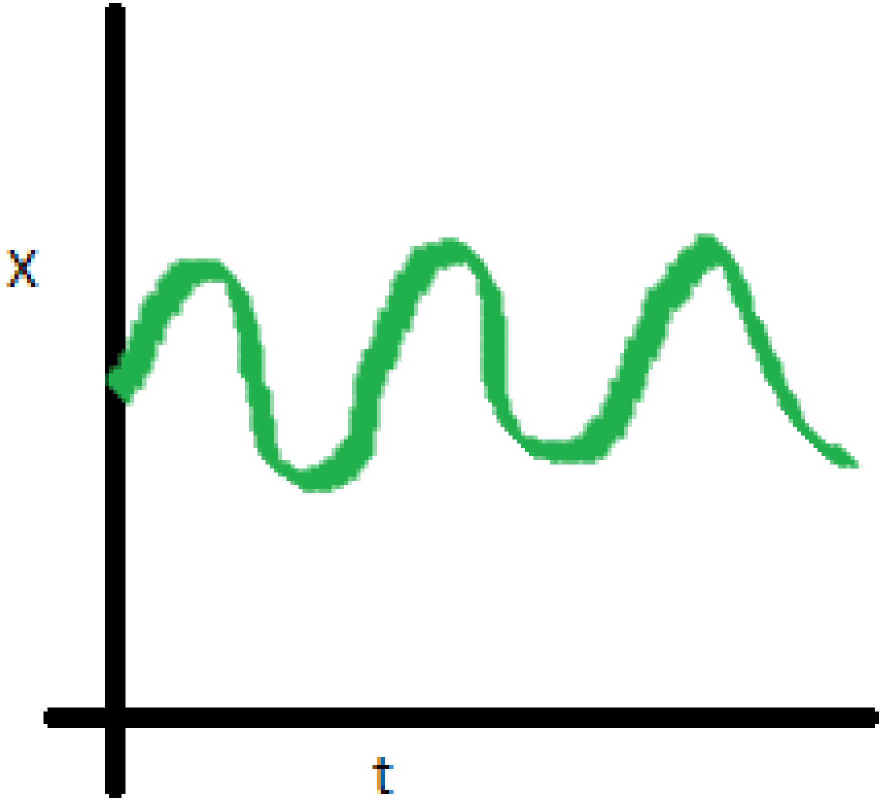
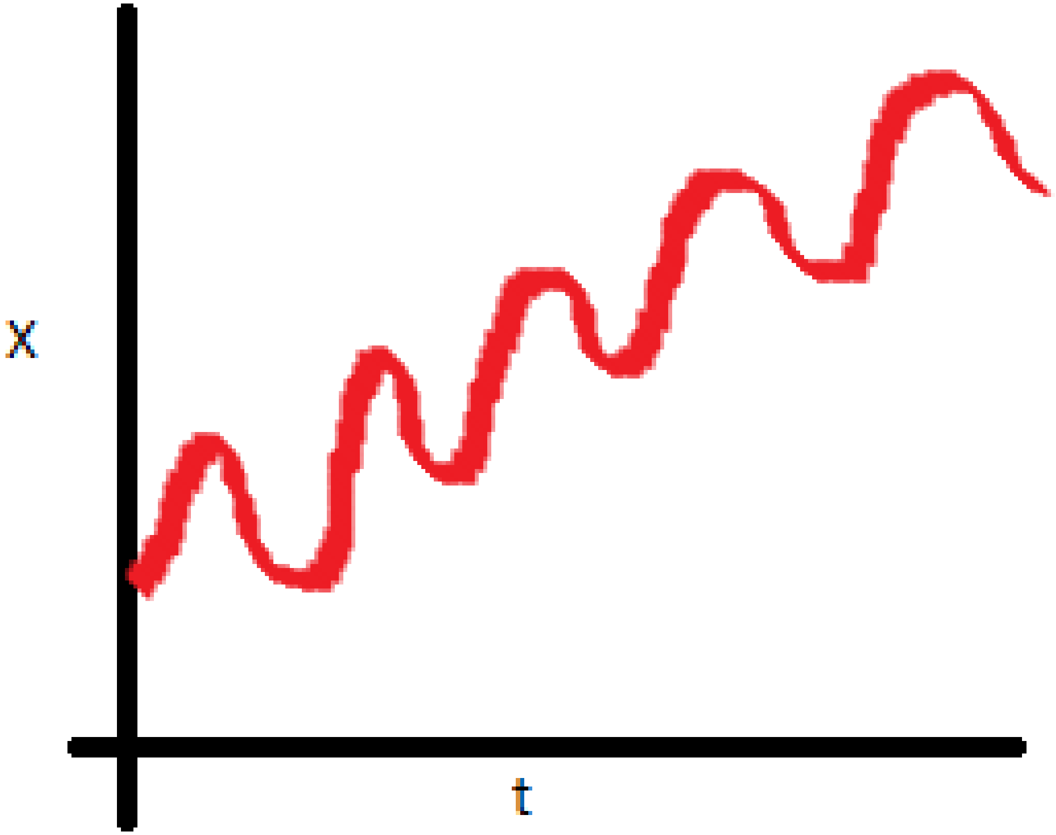
MS4S09

Data Mining and Statistical Modelling

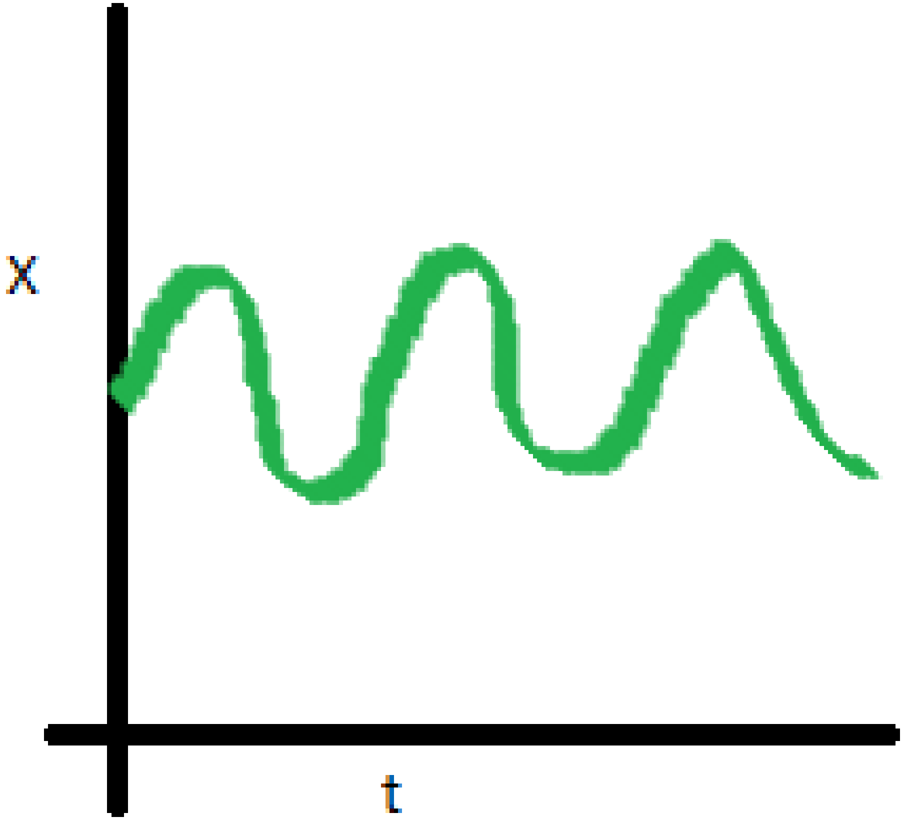
Stationarity



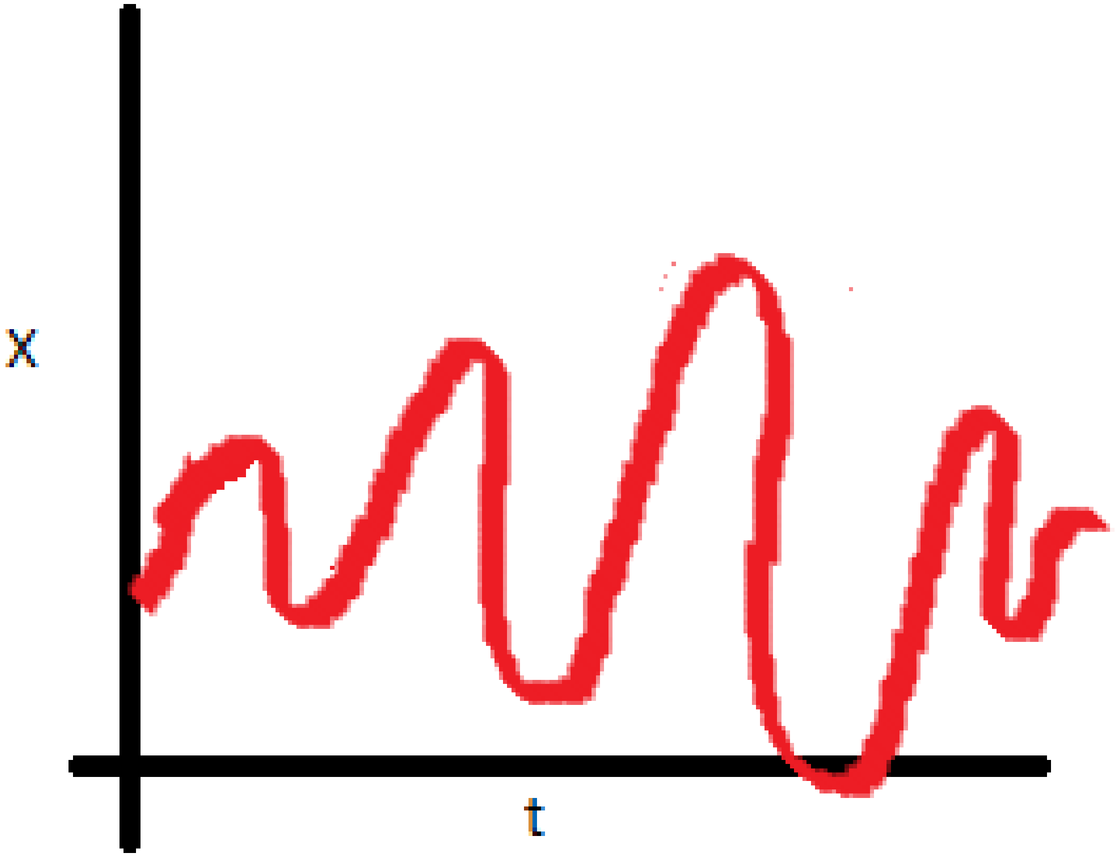
Stationary series



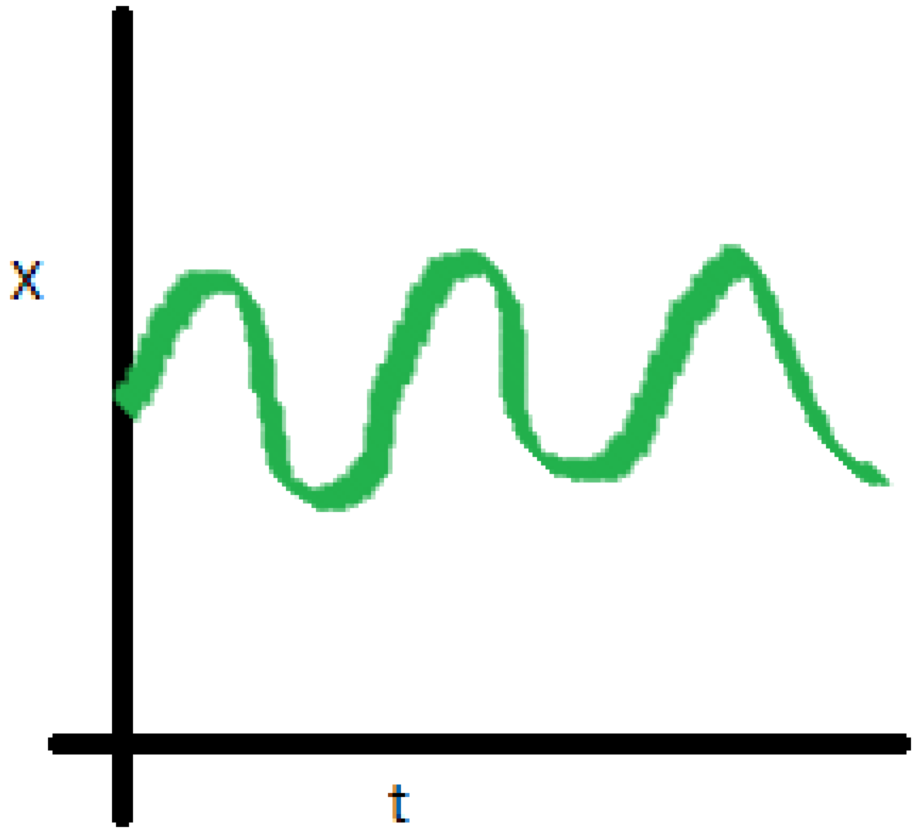
Non-Stationary series



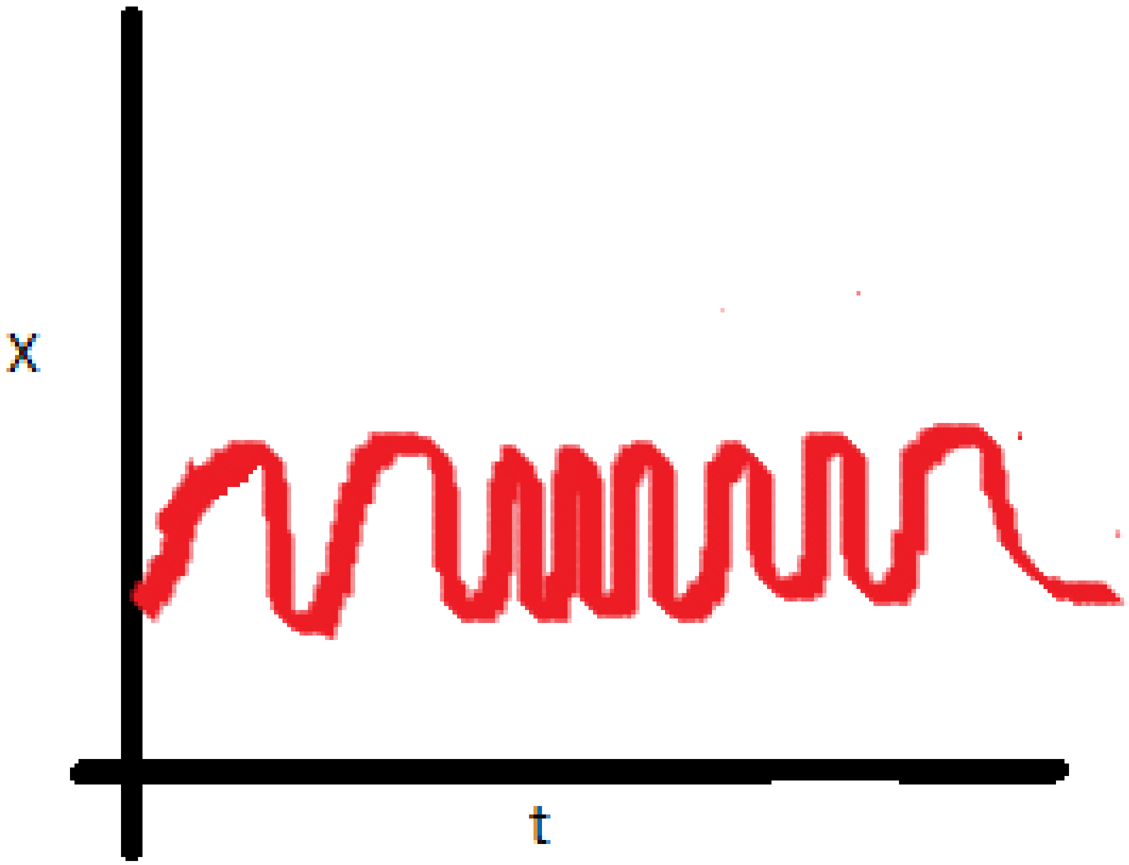
Stationary series



Non-Stationary series



Stationary series



Non-Stationary series

Time Series: Stationarity

The *auto-covariance function* of a time series $\{X_t, t \in \mathbb{Z}\}$:

$$\gamma_X(r,s) = \mathrm{E}[(X_r - \mathrm{E}[X_r]) \cdot (X_s - \mathrm{E}[X_s])].$$

$\{X_t\}$ is *(weakly) stationary* if

1. $\mathrm{E}[X_t] = m$ for all $t \in \mathbb{Z}$
2. $\mathrm{E}\left[{X_t}^2\right] < \infty$ for all $t \in \mathbb{Z}$, and
3. $\gamma_X(r,s) = \gamma_X(r+t, s+t)$ for all $\mathrm{r}, \mathrm{s}, t \in \mathbb{Z}$.

Examples of Stationary Time Series

1. If $\{X_t\}$ is a sequence of random variables with

$$\gamma_X(r,s) = \begin{cases} \sigma^2, & r = s \\ 0, & otherwise \end{cases}$$

with $\sigma^2 < \infty$ and $\mathrm{E}[X_t] = 0$, then $\{X_t\}$ is called *white noise* and we write

$$X_t \sim \mathrm{WN}(0, \sigma^2).$$

2. IID noise with finite second moment:

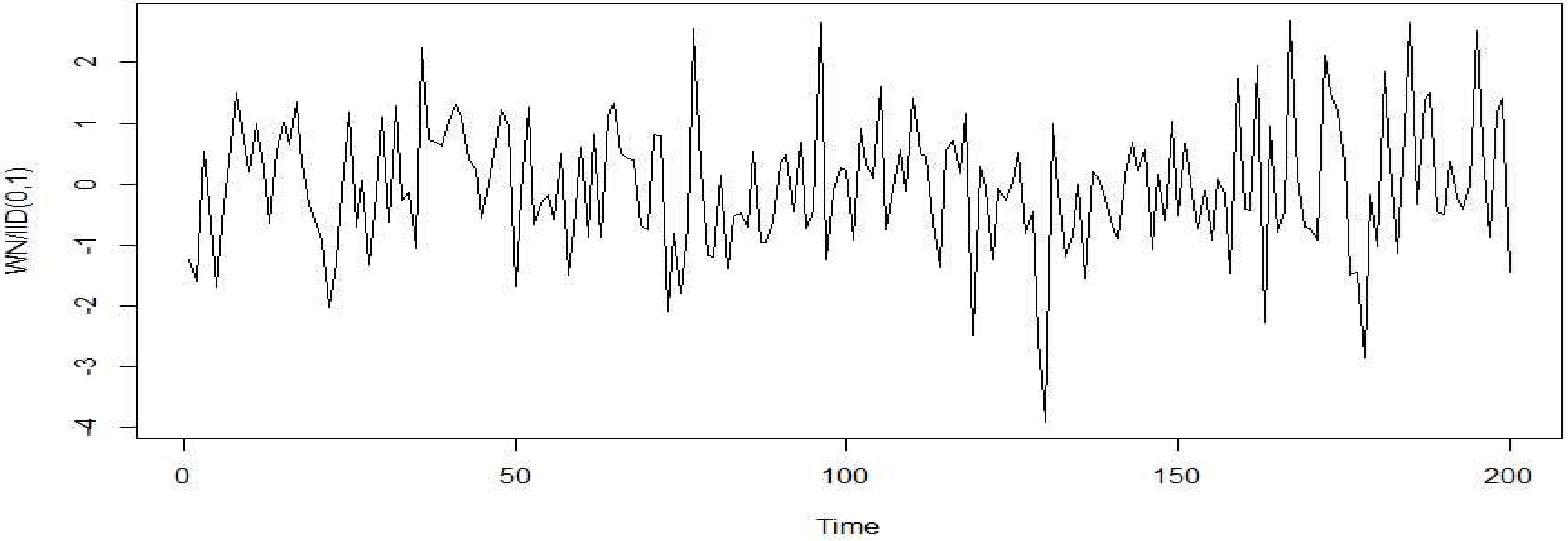
If $\{X_t\}$ is a sequence of independent identically distributed random variables with mean zero and second moment equal to $\sigma^2 < \infty$, we write

$$X_t \sim \mathrm{IID}(0, \sigma^2).$$

Examples of Stationary Time Series

How to generate in the R statistical software?

Use `rnorm` for normal (Gaussian) random variables

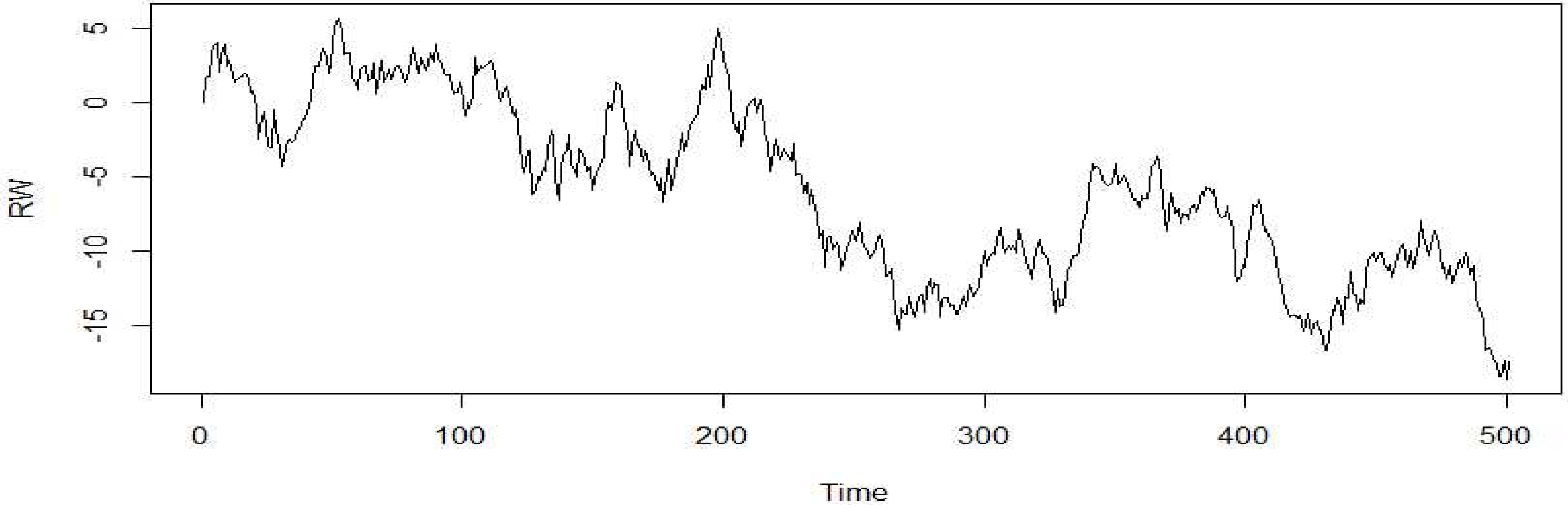


Examples of Non-Stationary Time Series

1. Random walk: Suppose $X_t \sim \mathrm{IID}(0, \sigma^2)$,

$$S_t = \sum_{j=1}^{t} X_j \, .$$

$\{S_t, t = 1, 2, \ldots\}$ is called a *random walk*.



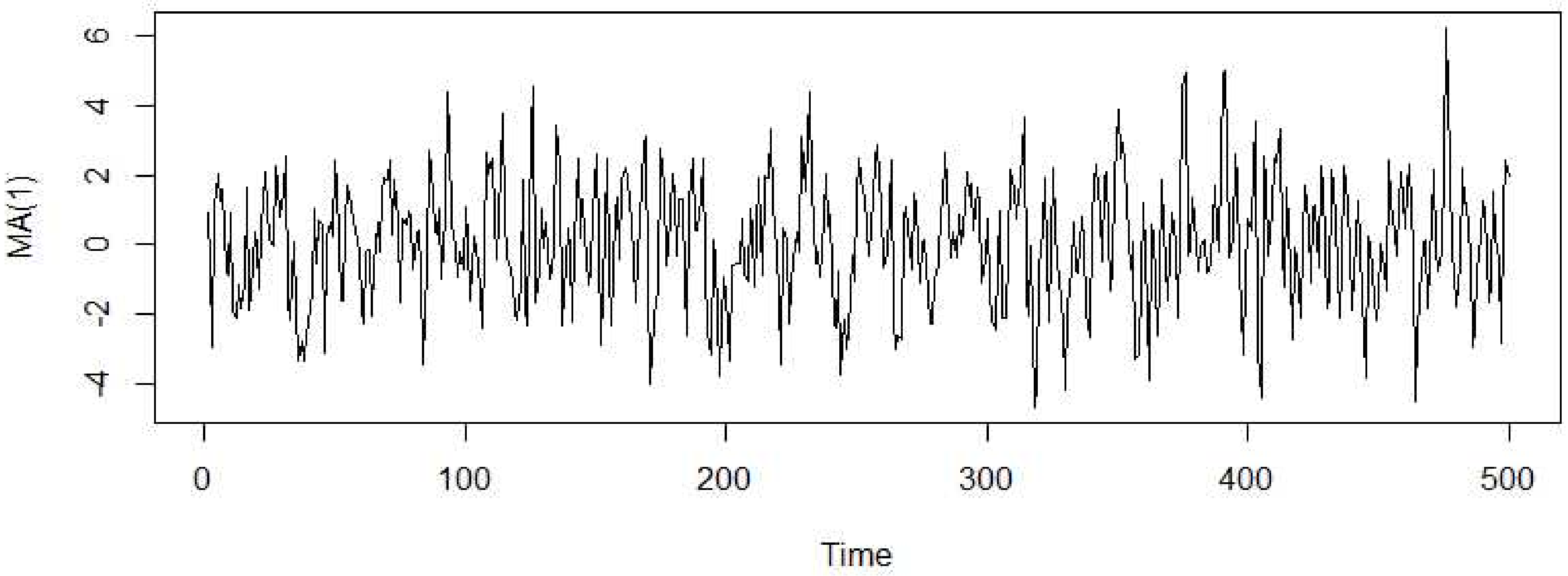
Examples of Stationary Time Series

2. A Moving Average Process of Order 1 (MA(1)):

$$X_t = Z_t + \theta Z_{t+1}, \qquad t \in \mathbb{Z}$$

where

$Z_t \sim \mathrm{WN}(0, \sigma^2)$.



ARMA Model: Definition

The process $\{X_t, t \in \mathbb{Z}\}$ is said to be an $\mathrm{ARM}A(p, q)$ process if $\{X_t\}$ is stationary and if for every t,

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p}$$
$$= Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$$

Auto Regression

Moving Average

where $Z_t \sim \mathrm{WN}(0, \sigma^2)$.

- AR order *p* and MA order *q*
- $\{X_t\}$ is said to be an $\mathrm{ARM}A(p, q)$ process with mean μ if $\{X_t - \mu\}$ is an $\mathrm{ARM}A(p, q)$ process.

ARMA Model: Notation

Write the ARMA model in the more compact form

$$\phi(B)X_t = \theta(B)Z_t,$$

where

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$$

and

$$\theta(z) = 1 - \theta_1 z - \ldots - \theta_p z^p$$

The polynomials are called the autoregressive and moving average polynomials, respectively.

ARMA Model: General

The class of "autoregressive moving average" or ARMA models forms an important family of stationary time series, for a number of reasons, including the following two.

For *any* autocovariance function $\gamma(\cdot)$ such that $\lim_{h\to\infty} \gamma(h) = 0$, and any integer $k > 0$, it is possible to find an ARMA process with autocovariance $\gamma_X(\cdot)$ such that $\gamma_X(h) = \gamma(h)$ for $h = 0, 1, 2, \ldots, k$.

The linear structure of ARMA models makes prediction "easy" to carry out.

ARMA Model: Stationarity

Not all formulations $\phi(B)X_t = \theta(B)Z_t$ model a stationary time series:

A stationary solution to the ARMA equation exists and is unique if and only if

$$\phi(z) \neq 0 \ \text{ for all } \ z \in \mathbb{C} \ \text{ such that } \ |z| \leq 1$$

(That is, it exists and is unique if and only if no zeroes of $\phi(z)$ lie inside the unit circle.)

GTx

Autocovariance Function: Stationarity

For a stationary time series $\{X_t\}$, the *autocovariance* function is

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

with the following properties:

1. $\gamma(0) \geq 0$,
2. $|\gamma(h)| \leq \gamma(0)$, and
3. $\gamma(h) = \gamma(-h)$.

The *autocorrelation* function of a stationary time series $\{X_t\}$ is

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} \ ,$$

and has all the properties of the autocovariance function, except that $\rho_X(0) = 1$.

Autocovariance Function: Estimation

Objective: Given $\{x_1, \ldots, x_n\}$ observations of a stationary time series $\{X_t\}$, estimate the autocovariance function $\gamma_X(\cdot)$ of $\{X_t\}$

- The sample *autocovariance function* is

$$\hat{\gamma}_X(h) = \frac{1}{n}\sum_{j=1}^{n-h}\left(x_{j+h} - \bar{x}\right)\left(x_j - \bar{x}\right), \qquad 0 \le h < n,$$

with $\hat{\gamma}_X(h) = \hat{\gamma}_X(-h), \ \ -n < h \le 0$, where $\bar{x} = \frac{1}{\mathrm{n}}\sum_{j=1}^{n} x_j$.

- The *sample autocorrelation function* is defined by

$$\hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}, \qquad |h| < n.$$

ACF and MA(q) Process

Now suppose that $\{X_t\}$ is the stationary solution of

$$X_t = \theta(B)Z_t,$$

where $\theta(z) = 1 - \theta_1 z - \ldots - \theta_p z^p$ and $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$.

- It can be shown that $\gamma_X(h) = 0$ for $|h| > q$.

Partial Autocorrelation Function

Suppose $\{X_t\}$ is a stationary time series with mean zero, for which $\gamma_X(h) \to 0$ as $h \to \infty$. The *partial autocorrelation function (PACF)* $\alpha_X(h)$ is defined by

$$\alpha_X(0) = 1,$$
$$\alpha_X(h) = \alpha_{hh},$$

where α_{hh} is the last component of

$$\alpha_h = \Gamma_h^{-1}\gamma_h(1),$$

with

$$\Gamma_h = [\gamma_X(i-j)]_{i,j=1,\ldots,h} \quad \text{and} \quad \gamma_h(1) = \left(\gamma_X(1), \gamma_X(2), \ldots, \gamma_X(h)\right)^T.$$

PACF and Prediction

$\{X_t\}$ is a stationary time series with mean zero, for which $\gamma_X(h) \to 0$ as $h \to \infty$:

$$P_h X_{h+1} = a_1 X_h + a_2 X_{h-1} + \ldots + a_h X_1$$

the *one-lag linear prediction* given $X_1, \ldots, X_h$. If $a_1, \ldots, a_h$ are selected such that we minimize

$$S(a_1, \ldots, a_h) = \mathrm{E}[(X_{h+1} - a_1 X_h - \ldots + a_h X_1)^2].$$

then $P_h X_{h+1}$ is called the *Best Linear Unbiased Predictor (BLUP)* for X_{h+1}. We define the <u>partial autocorrelation function</u> as

$$\alpha(h) = a_h$$

PACF and AR(p) Process

Now suppose that $\{X_t\}$ is the stationary solution of

$$\phi(B)X_t = Z_t,$$

where $\phi(z) = 1 + \phi_1 z + \ldots + \phi_q z^q$ and $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$.

- It can be shown that $\alpha_X(h) = 0$ for $|h| > p$.

MA(q) and AR(p) Processes

- Summarizing:

- An AR(p) process has PACF $\alpha(h) = 0$ for $|h| > p$.
- An MA(q) process has ACF $\rho(h) = 0$ for $|h| > q$.

- Unfortunately, there are no such simple rules for ARMA(p, q) processes in general.