Time Series Analysis

B. Banerjee

October 28, 2019

BOOKS

- Text Book

1. Time Series Analysis and Its Applications:
   With R Examples By Robert H. Shumway , David S. Stoffer
2. Introduction to Time Series and Forecasting
   By Brockwell, Peter J., Davis, Richard A.

- Reference book

1. Forecasting: principles and practice
   By Rob J Hyndman, George Athanasopoulos
2. Time Series: Theory and Method
   By Brockwell, Peter J., Davis, Richard A.

Introduction

Definition

Time series is a collection of random variables $\{X_t| \ \ t \in T\}$ over a time index set T, which might be a finite, countably infinite or uncountable set.

- Realized values: What we observe are the realized values of the time series i.e. the data set is $\{X_1 = x_1, X_2 = x_2, \cdots X_n = x_n\}$, where the x_is are some numeric or categorical values.

Categories of Time series

- Discerte time series: If T is a countable set then it is a discerte time series.
- Continuous time series: If T is an interval then it is a continuous time series.
- Note: Discerte or continuous are the adjectives of time but NOT of random variable X_t

Operators

- Back-shift operator B, such that $B^h X_t = X_{t-h}$
- Difference operator: $I - B = \nabla$ which gives

$$\nabla X_t = X_t - X_{t-1} = (I - B)X_t$$

which implies

$$\nabla^h X_t = (I - B)^h X_t$$

- Seasonal difference

$$\nabla_s X_t = (1 - B^s)X_t$$

Polynomial $y = 1 + 2x + 3x^2$

```
x<-seq(1,1.5, by=0.1)
y<-1+2*x+3*x^2
l<-length(x)
dtable<-array(0,dim = c(l,l+2))
dtable[,1]<-x
dtable[,2]<-y
for(i in 1: 3){
  dtable[1:(l-i),(i+2)]<-diff(y,1,i)
}
round(dtable, 2)
```

Difference Table

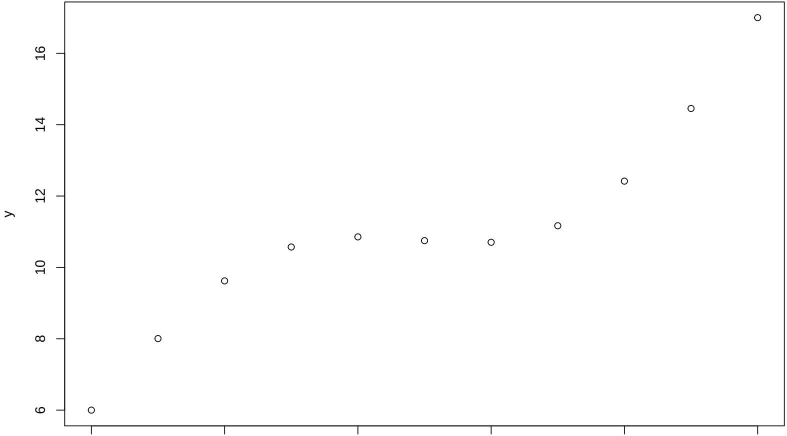
Polynomial $y = 1 + 2x + 3x^2$

```
##       [,1]  [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,]   1.0  6.00 0.83 0.06    0    0    0    0
## [2,]   1.1  6.83 0.89 0.06    0    0    0    0
## [3,]   1.2  7.72 0.95 0.06    0    0    0    0
## [4,]   1.3  8.67 1.01 0.06    0    0    0    0
## [5,]   1.4  9.68 1.07 0.00    0    0    0    0
## [6,]   1.5 10.75 0.00 0.00    0    0    0    0
```

Difference Table

Polynomial & periodic $y = 1 + 2x + 3x^2 + \sin(2\pi x)$

```
x<-seq(1,2, by=0.1)
y<-1+2*x+3*x^2+2*sin(2*pi*x)
plot(y~x)
```



Difference Table

Polynomial & periodic $y = 1 + 2x + 3x^2 + \sin(2\pi x)$

```
##        [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7] [,8]  [,9] [,10]
##  [1,]  6.00  2.01 -0.39 -0.28  0.28  0.00 -0.11 0.04  0.03 -0.03
##  [2,]  8.01  1.62 -0.67  0.00  0.28 -0.11 -0.07 0.07  0.00 -0.03
##  [3,]  9.62  0.95 -0.67  0.28  0.17 -0.17  0.00 0.07 -0.03  0.00
##  [4,] 10.57  0.28 -0.39  0.45  0.00 -0.17  0.07 0.04  0.00  0.00
##  [5,] 10.86 -0.11  0.06  0.45 -0.17 -0.11  0.11 0.00  0.00  0.00
##  [6,] 10.75 -0.05  0.51  0.28 -0.28  0.00  0.00 0.00  0.00  0.00
##  [7,] 10.70  0.46  0.79  0.00 -0.28  0.00  0.00 0.00  0.00  0.00
##  [8,] 11.17  1.25  0.79 -0.28  0.00  0.00  0.00 0.00  0.00  0.00
##  [9,] 12.42  2.04  0.51  0.00  0.00  0.00  0.00 0.00  0.00  0.00
## [10,] 14.45  2.55  0.00  0.00  0.00  0.00  0.00 0.00  0.00  0.00
## [11,] 17.00  0.00  0.00  0.00  0.00  0.00  0.00 0.00  0.00  0.00
```

Difference Table

Polynomial & periodic & random error $y = 1 + 2x + 3x^2 + \sin(2\pi x) + \epsilon$

```
x<-seq(1,2, by=0.1)
l<-length(x)
y<-1+2*x+3*x^2+ 2*sin(2*pi*x)+ rnorm(l,0,0.1)

dtable<-array(0,dim = c(l,l+2))
dtable[,1]<-x
dtable[,2]<-y
for(i in 1: (l-1)){
  dtable[1:(l-i),(i+2)]<-diff(y,1,i)
}
round(dtable[,2:11], 2)
```

Difference Table

Polynomial & periodic & random error $y = 1 + 2x + 3x^2 + \sin(2\pi x) + \epsilon$

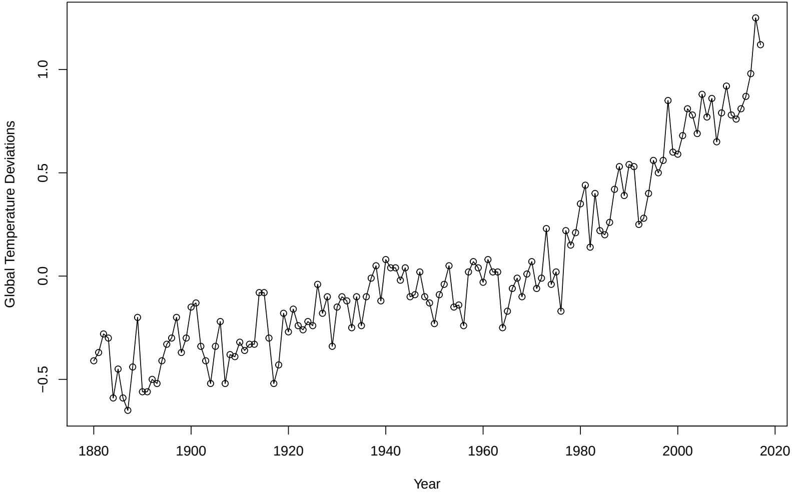
```
##         [,1] [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9] [,10]
##  [1,]  6.04 1.88 -0.12 -0.72  0.82 -0.33 -0.59  2.11 -3.89  3.51
##  [2,]  7.92 1.76 -0.84  0.09  0.49 -0.92  1.52 -1.77 -0.37  9.56
##  [3,]  9.68 0.92 -0.74  0.58 -0.43  0.60 -0.25 -2.14  9.19  0.00
##  [4,] 10.60 0.18 -0.16  0.15  0.17  0.34 -2.40  7.05  0.00  0.00
##  [5,] 10.78 0.02 -0.01  0.32  0.51 -2.05  4.65  0.00  0.00  0.00
##  [6,] 10.80 0.01  0.31  0.83 -1.54  2.60  0.00  0.00  0.00  0.00
##  [7,] 10.80 0.31  1.13 -0.72  1.06  0.00  0.00  0.00  0.00  0.00
##  [8,] 11.11 1.44  0.42  0.34  0.00  0.00  0.00  0.00  0.00  0.00
##  [9,] 12.56 1.86  0.76  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## [10,] 14.42 2.62  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## [11,] 17.04 0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
```

Global Temperature Deviations

Meteorological station data were used to estimate the global annual-mean surface air temperature deviation from 1880 to 2018.
source : https://data.giss.nasa.gov/gistemp/graphs/

```
gtmp<-read.csv("graph.csv",header = T,sep =",")
plot(gtmp$No_Smoothing~gtmp$Year, type="o",
     ylab="Global Temperature Deviations", xlab='Year')
```

Global Temperature Deviations



White Noise

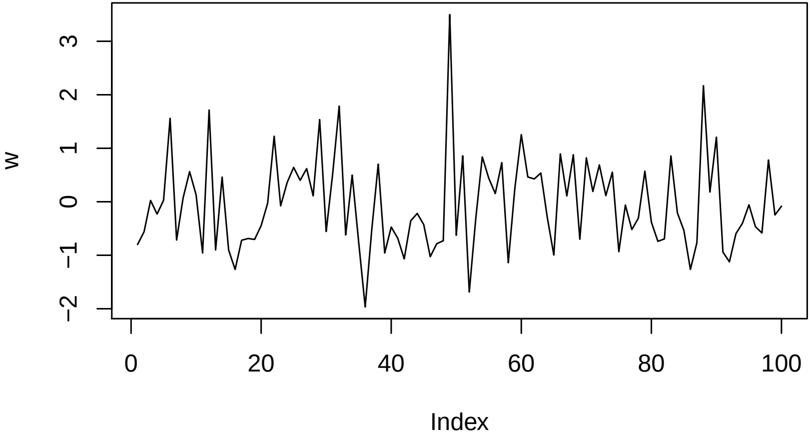
Definition

The time series generated from uncorrelated variables with zero mean and fixed finite variance is called white noise [*Notation* $W_t \sim WN(0, \sigma_w^2)$]

- Example1: $X_t = W_t = \begin{cases} N(0,1) \text{ if t is even }, \\ Exp(1) - 1 \text{ if t is odd }. \end{cases}$
- Example2 : $X_t = W_t \quad i.i.d. \quad N(0, \sigma^2)$

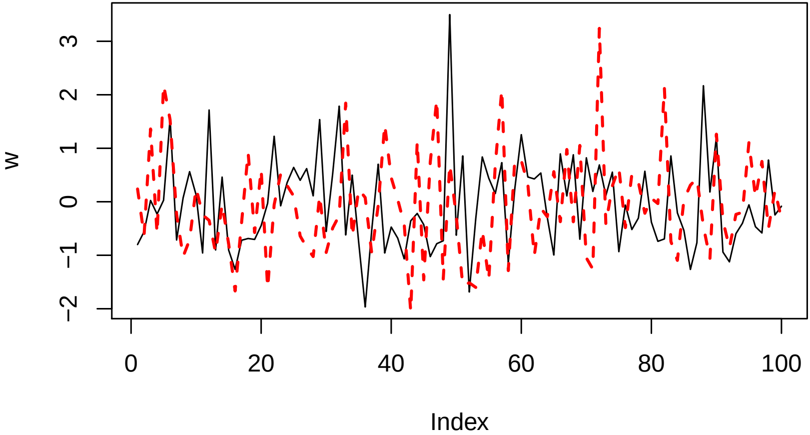
White Noise (Example)

```
set.seed(123); n<-100; w<-array(0,dim = c(n))
e<-seq(from = 2,to = n, by=2) ; o<-(e-1);
w[e]<-rnorm((n/2),0,1) ; w[o]<-rexp((n/2),1)-1;
plot(w, type = "l")
```



White Noise (Example)

```
set.seed(123); n<-100; w<-array(0,dim = c(n))
e<-seq(from = 2,to = n, by=2) ; o<-(e-1);
w[e]<-rnorm((n/2),0,1) ; w[o]<-rexp((n/2),1)-1;
plot(w, type = "l"); lines(rnorm(n),col=2,lty=2,lwd=2)
```



Trend

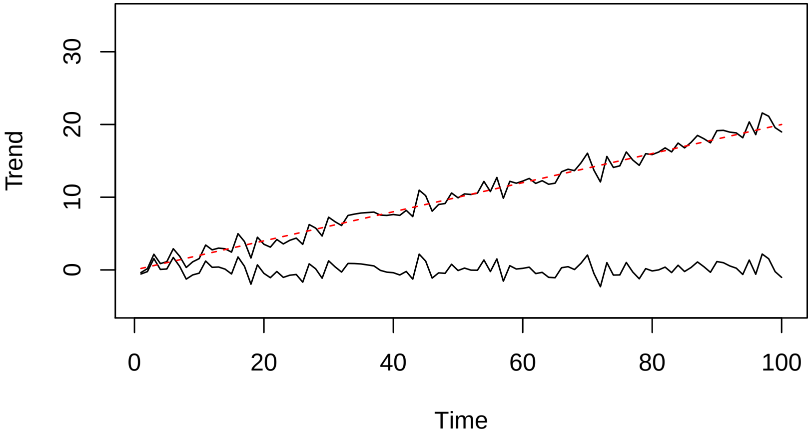
Definition

A deterministic pattern (T_t) which persist through out in the time sires. For example $X_t = T_t + W_t$

- Example: $X_t = T_t + W_t = 0.2t + W_t$
- Example: $X_t = T_t + W_t = t^{(1/3)} + W_t$

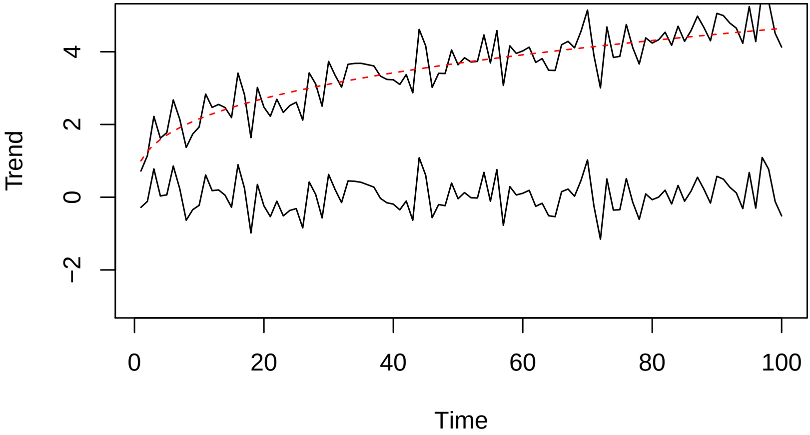
Trend (Example)

```
set.seed(123); n<-100;  w <- rnorm(n,0,1);
xw<-w; xz<-0.2*(1:n)+w
plot.ts(xw, ylim=c(-5,35), ylab="Trend")
lines(xz); lines(0.2*(1:n), col=2, lty="dashed")
```



Trend (Example)

```
set.seed(123); n<-100;  w <- rnorm(n,0,.5);
xw<-w; xz<-(1:n)^(1/3)+w
plot.ts(xw, ylim=c(-3,5), ylab="Trend")
lines(xz); lines((1:n)^(1/3), col=2, lty="dashed")
```



Seasonality

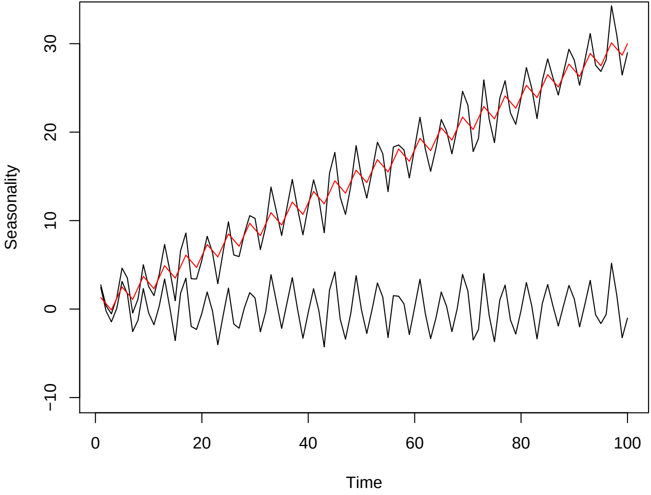
Definition

A deterministic pattern (S_t) which returns in the time sires after a fixed interval. For example $X_t = T_t + S_t + W_t$

- Example: $X_t = T_t + S_t + W_t = 0.3t + 3\sin(4.5t\pi) + W_t$
- Example: $X_t = S_t + W_t = 3\sin(4.5t\pi) + W_t$

```r
set.seed(123); n<-100;
w <- rnorm(n,0,1);
xw<-0.3*(1:n) + 3*sin(4.5*(1:n)*pi)+w;
xz<-  3*sin(4.5*(1:n)*pi)+w;
plot.ts(xw, ylim=c(-10,33), ylab="Seasonality");
lines(xz);
lines(0.3*(1:n)+ sin(4.5*(1:n)*pi), col=2)
```

Seasonality (Example)



Sunspot

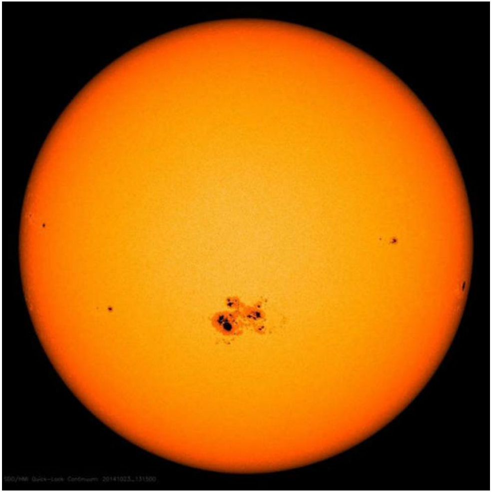
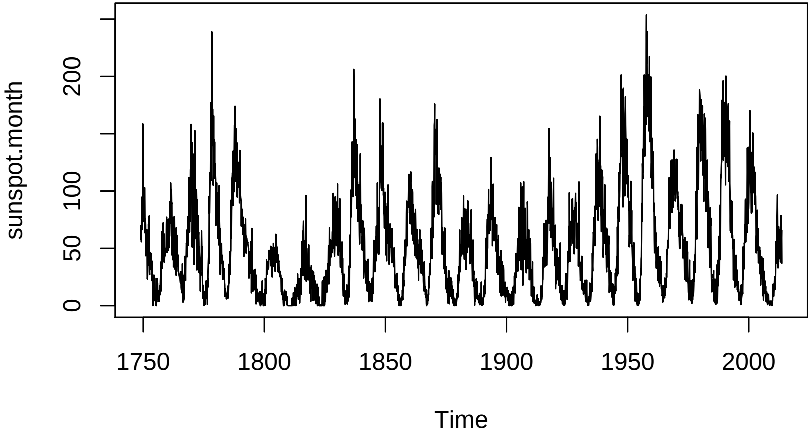


Figure 1: Sunspots

Sunspot data

Monthly numbers of sunspots, as from the World Data Center, aka SIDC. The univariate time series “sunspot.month" contain 2988 observations.

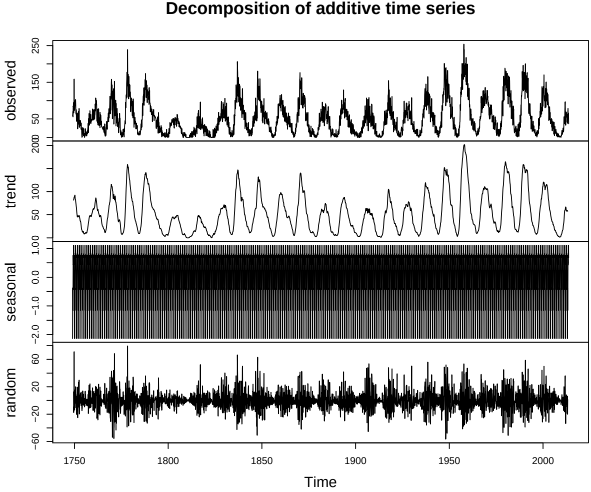
```
plot(sunspot.month)
```



Additive model : $X_t = T_t + S_t + W_t$

```
plot(decompose(sunspot.month,type = "additive"))
```

Sunspot data (Additive model)



Stationariry

Mean and Autocovariance

Definition

Suppose that $\{X_t\}$ is a time series with $E|X_t| < \infty$.
Its mean function is

$$\mu_t = E[X_t]$$

.

Definition

Suppose that $\{X_t\}$ is a time series with $E[X_t^2] < \infty$, then its autocovariance function (ACVF) is

$$\gamma_X(s, t) = Cov(X_s, X_t) = E[(X_s - \mu_s)(X_t - \mu_t)]$$

.

Definition

A time series $\{X_t\}$ is said to be weakly stationary if
1. μ_t is independent of t, and
2. For each $h \in \mathbb{Z}$, ACVF $\gamma(t+h, t)$ is independent of t

- Example: $X_t \sim WN(0, \sigma_w^2)$
- NOT an Example: $X_t = \sum_{i=0}^{t} W_i$, where $W \sim WN(0, \sigma_w^2)$

Strongly Stationary Time seris

Definition

$\{X_t\}$ is strongly stationary if for all k, $h, t_1, \ldots, t_k$, $x_1, \ldots, x_k$shifting the time axis does not affect the distribution i.e.

$$P(X_{t_1} \leq x_1, \ldots, X_{t_k} \leq x_k) = P(X_{t_1+h} \leq x_1, ..., X_{t_k+h} \leq x_k)$$

- Example: $X_t \sim WN = N(0, \sigma^2)$
- NOT an Example: Random walk defined as

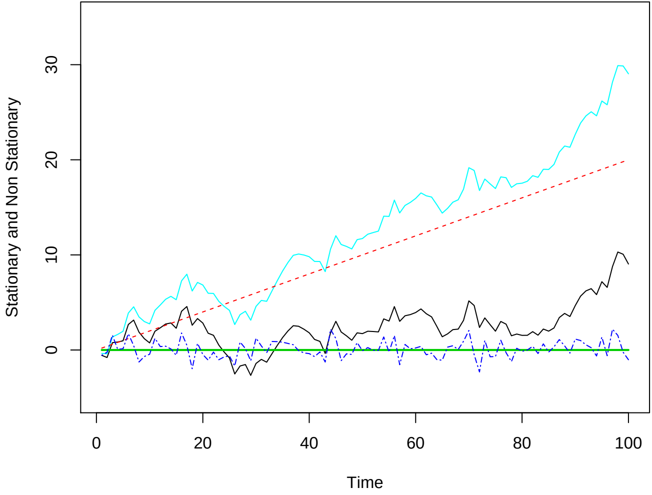
$$S_t = \sum_{i=0}^{t} X_i, \quad \text{where} \;\; X_i \quad i.i.d. \quad N(0, \sigma^2)$$

- Strong Stationarity $\implies$ Weak Stationarity

Stationary and Non Stationary Time seris (Example)

```
set.seed(123);
n<-100;
w <- rnorm(n,0,1);
z<-w+0.2;
xw<-cumsum(w);
xz<-cumsum(z) ;
plot.ts(xw, ylim=c(-5,35),
        ylab="Stationary and Non Stationary")
lines(xz, col=5);
lines(0.2*(1:n), col=2, lty="dashed");
lines(0.0*(1:n), col=3,lwd=2)
lines(w,lty=6, col=4)
```

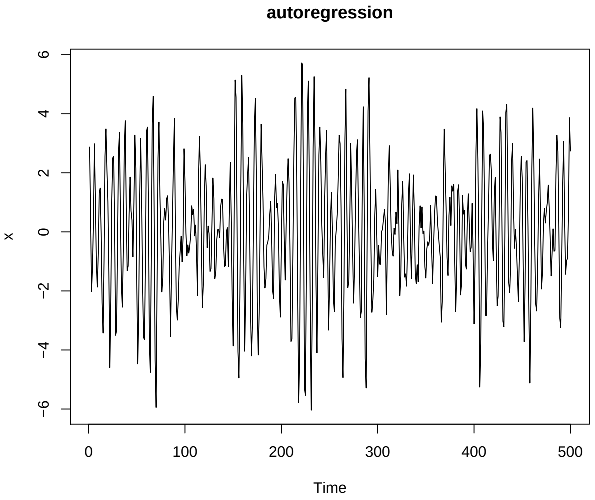
Stationary and Non Stationary Time seris (Example)



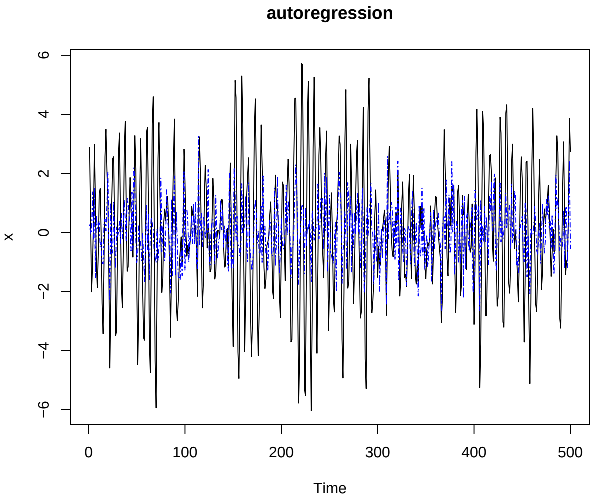
Weakly Stationary Time seris (Example)

```
set.seed(123);
n<-550;
w <- rnorm(n,0,1);
x = filter(w, filter=c(1,-.9),
           method="recursive")[-(1:50)]
plot.ts(x, main="autoregression")
lines(w[-(1:50)],lty=6, col=4)
```

Weakly Stationary Time seris (Example)



Weakly Stationary Time seris (Example)



Properties of a Strongly Stationary Time Series X_t

1. The random variables X_t are identically distributed.
2. $(X_t, X_{t+h}) \stackrel{d}{=} (X_1, X_{1+h})$ for all integers t and h.
3. X_t is weakly stationary if $E(X_t^2) < \infty$ for all t.
4. Weak stationarity does not imply strongly stationarity.
5. An i.i.d. sequence is strongly stationary.

An important class of weakly stationary time series:

Definition

$$X_t = \mu + \sum_{j=-\infty}^{j=+\infty} \psi_j W_{t-j}$$

where $\mu \in \mathbb{R}$ and $\sum_{j=-\infty}^{j=+\infty} |\psi_j| < \infty$ and $W_j \sim WN(0, \sigma_w^2)$

- $E(X_t) = \mu$
- $\gamma_X(h) = \sigma_w^2 \sum_{j=-\infty}^{j=+\infty} \psi_j \psi_{j-h} < \infty$

WN as a linear proccess

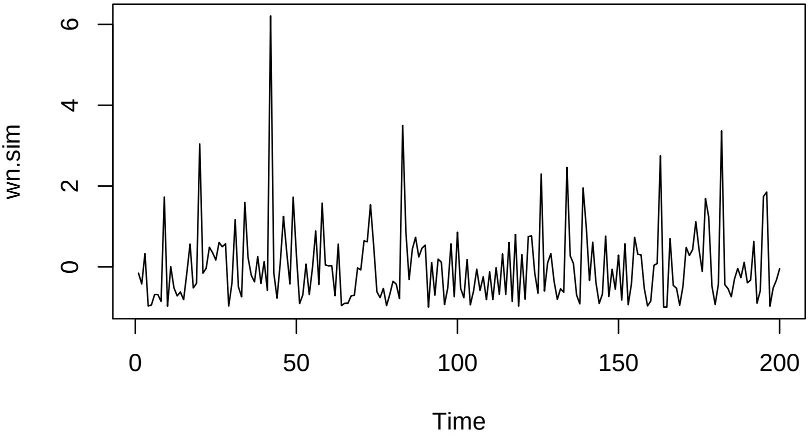
- Time series $X_t = W_t$ where $W_t \sim WN(0, \sigma_w^2)$
- WN process is a linear process with $\mu = 0$ and

$$\psi_j = \begin{cases} 1 & \text{if} \quad j = 0 \\ 0 & \text{otherwise} \end{cases}$$

- WN had fixed and finite mean and variance with correlation zero
- WN need not be normally distribute
- WN need not be iid
- WN is always weakly stationary
- Normally distribute WN is strongly stationary
- iid sequence is always white noise and strongly stationary

WN Example

```
set.seed(123); n<-200;
wn.sim<-rexp(n,1)-1
plot.ts(wn.sim)
```



Genaral result

Theorem

Consider a (weakly) stationary time series $\{X_t\}$ with mean zero and define

$$Y_t = \mu + \sum_{j=-\infty}^{j=+\infty} \psi_j X_{t-j}$$

where $\mu \in \mathbb{R}$ and $\sum_{j=-\infty}^{j=+\infty} |\psi_j| < \infty$ then

$$E(Y_t) = \mu$$

$$\gamma_Y(h) = \sum_{k=-\infty}^{k=+\infty} \sum_{j=-\infty}^{j=+\infty} \psi_k \psi_j \gamma_X(h - j + k) \quad \textit{if exists.}$$

Auto correlation function (ACF)

Suppose that X_t is at least weakly stationary time series with

$$E(X_t) = \mu \quad and \quad \gamma_X(h) = E[(X_t - \mu)(X_{t+h} - \mu)]$$

Definition

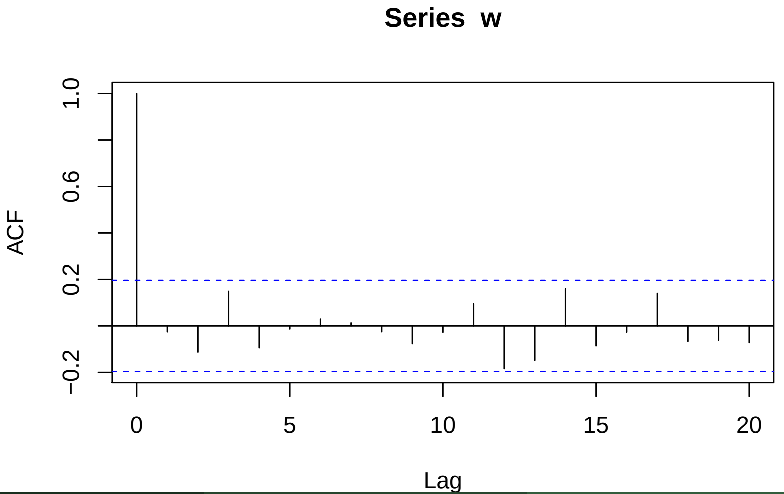
Auto correlation function (ACF) of X_t and X_{t+h} is defined as

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

- $\rho(h) = \rho(-h)$
- Let $1 \leq i, j \leq n$ and define a matrix $R = ((\rho(|i-j|)))_{i,j}$ then $\mathbf{a}^T R\mathbf{a} \geq 0$ for all $\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^n$.
- R is positive semidefinite matrix and $\gamma()$ and $\rho()$ are positive semidefinite functions.

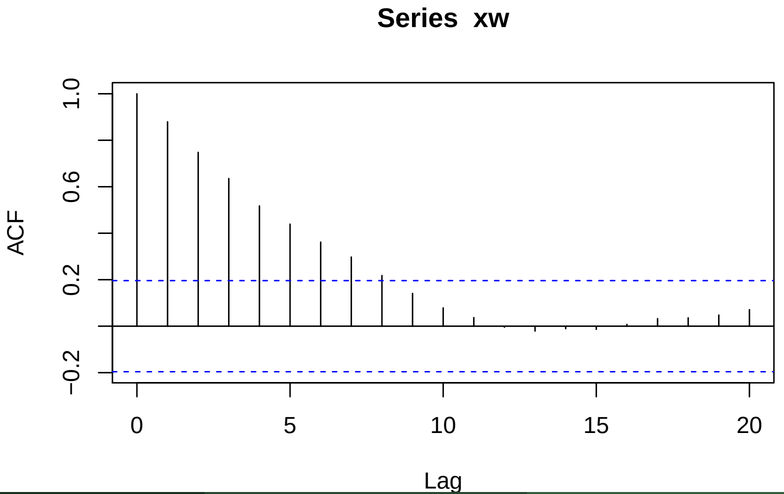
ACF (Example): WN

```
set.seed(123); n<-100;  w <- rnorm(n,0,1); z<-w+0.2;
xw<-cumsum(w); xz<-cumsum(z) ;
acf_xw<-acf(w,type = "correlation",plot = T)
```



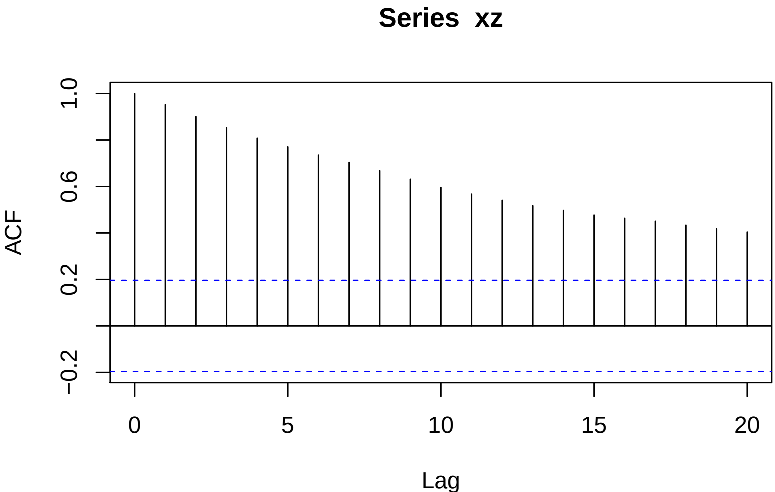
ACF (Example): Random walk

```
set.seed(123); n<-100;  w <- rnorm(n,0,1); z<-w+0.2;
xw<-cumsum(w); xz<-cumsum(z) ;
acf_xw<-acf(xw,type = "correlation",plot = T)
```



ACF (Example): Trend

```
set.seed(123); n<-100;  w <- rnorm(n,0,1); z<-w+0.2;
xw<-cumsum(w); xz<-cumsum(z) ;
acf_xz<-acf(xz,type = "correlation",plot = T)
```



MA(q) processes

Definition

MA(q),a moving average process of order q is defined as

$$X_t = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$$

Moving average process (MA)

Definition

A time series $\{X_t\}$is said to be an moving average process (of order one) if $X_t = W_t + \theta W_{t-1}$ where $W_t \sim WN(0, \sigma_w^2)$

- MA(1) process is a linear process with $\mu = 0$ and

$$\psi_j = \begin{cases} 1 & \text{if} \quad j = 0 \\ \theta & \text{if} \quad j = 1 \\ 0 & \text{otherwise} \end{cases}$$

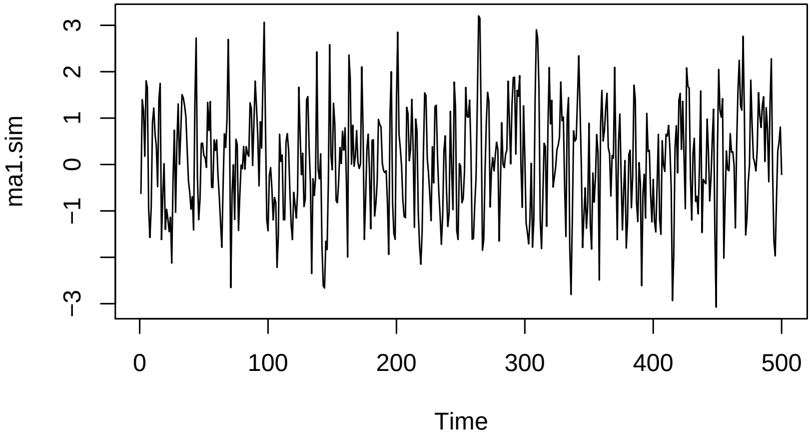
- $E(X_t) = 0$
-

$$\gamma_X(h) = \begin{cases} \sigma_w^2(1+\theta^2) & \text{if} \quad h = 0 \\ \sigma_w^2\theta & \text{if} \quad h = +1, -1 \\ 0 & \text{otherwise} \end{cases}$$

- MA(1) is a at least weakly stationary process

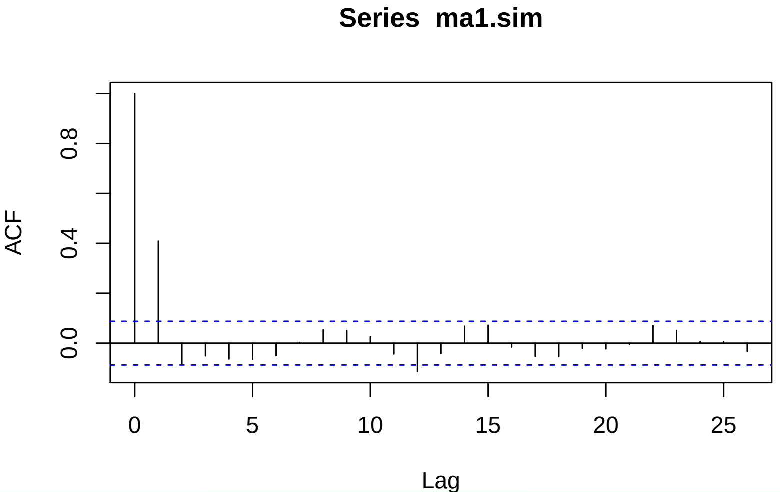
MA(1) Example

```
set.seed(123); n<-500; p<-0; d<-0;q<-1;
ma1.sim<-arima.sim(list(order=c(p,d,q), ma=0.7), n)
ts.plot(ma1.sim)
```



MA(1) Example

```r
set.seed(123); n<-500; p<-0; d<-0;q<-1;
ma1.sim<-arima.sim(list(order=c(p,d,q), ma=0.7), n)
acf(ma1.sim,type = "correlation",plot = T)
```



AR(p) processes

Definition

AR(p), an auto-regressive process of order p is defined as

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + W_t$$

Auto regressive process (AR)

Definition

A time series $\{X_t\}$is said to be an auto regressive process (of order one) if $X_t = \phi X_{t-1} + W_t$ where $W_t \sim WN(0, \sigma_w^2)$

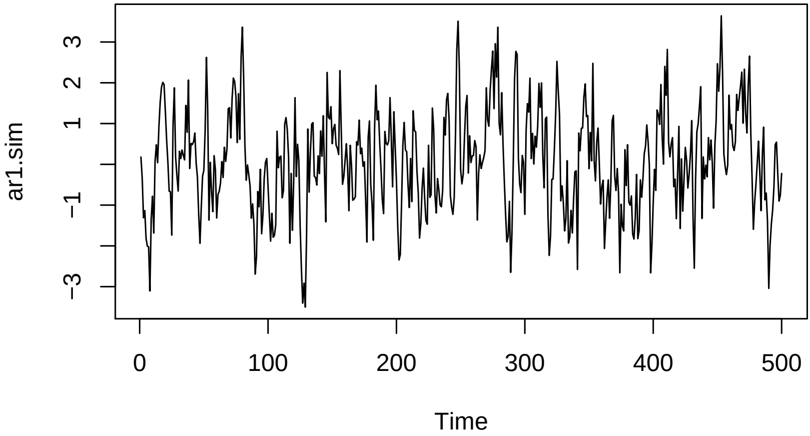
- AR(1) process is a linear process with $\mu = 0$ and

$$\psi_j = \begin{cases} 1 & \text{if} \quad j = 0 \\ \phi^j & \text{if} \quad j \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

- $E(X_t) = 0$
- $\gamma_X(h) = \frac{\sigma_w^2 \phi^{|h|}}{1-\phi^2}$ if $|\phi| < 1$, $\phi \neq 0$ [exponential decay]
- If $\phi = 0$ then AR(1) process is a WN process
- AR(1) is at least a weakly stationary process

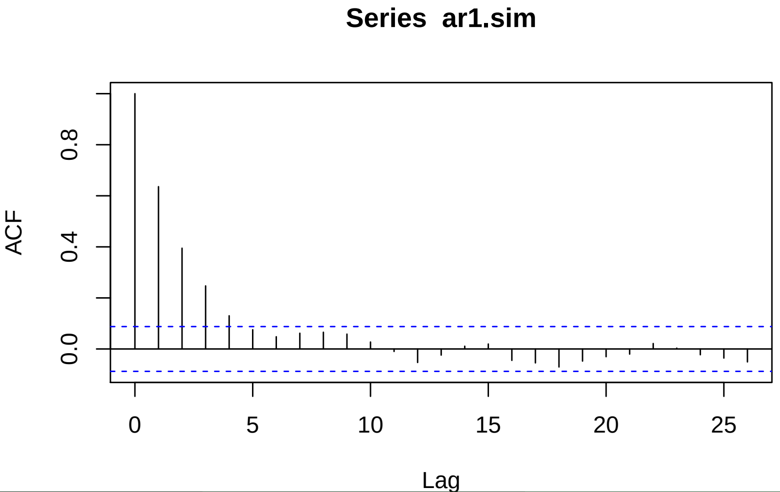
AR(1) Example

```
set.seed(123); n<-500; p<-1; d<-0;q<-0;
ar1.sim<-arima.sim(list(order=c(p,d,q), ar=0.7), n)
ts.plot(ar1.sim)
```



AR(1) Example

```
set.seed(123); n<-500; p<-1; d<-0;q<-0;
ar1.sim<-arima.sim(list(order=c(p,d,q), ar=0.7), n)
acf(ar1.sim,type = "correlation",plot = T)
```



Converges in mean square

Definition

A sequence of random variables $Y_1, Y_2, \cdots$ converges in mean square to Z if for which

$$\lim_{n\to\infty} E(Y_n - Z)^2 = 0$$

- Consider AR(1) process $X_t = \phi X_{t-1} + W_t$ where, $W_t \sim WN(0, \sigma_w^2)$ and $|\phi| < 1$.
- AR(1) is an $MA(\infty)$ process i.e.

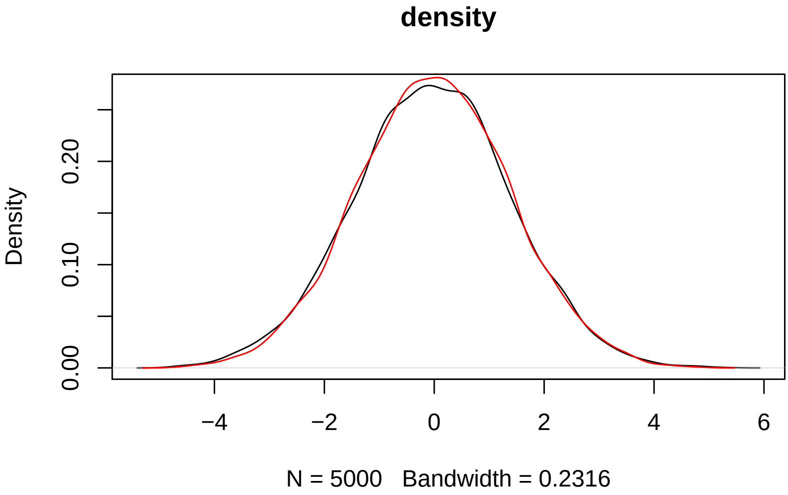
$$\lim_{k\to\infty} E\left(X_t - \sum_{j=0}^{k} \phi^j W_{t-j}\right)^2 = 0$$

- Convergence in mean square $\implies$ Convergence in probability $\implies$ Convergence in distribution, BUT NOT THE OTHER WAY ROUND IN GENERAL.

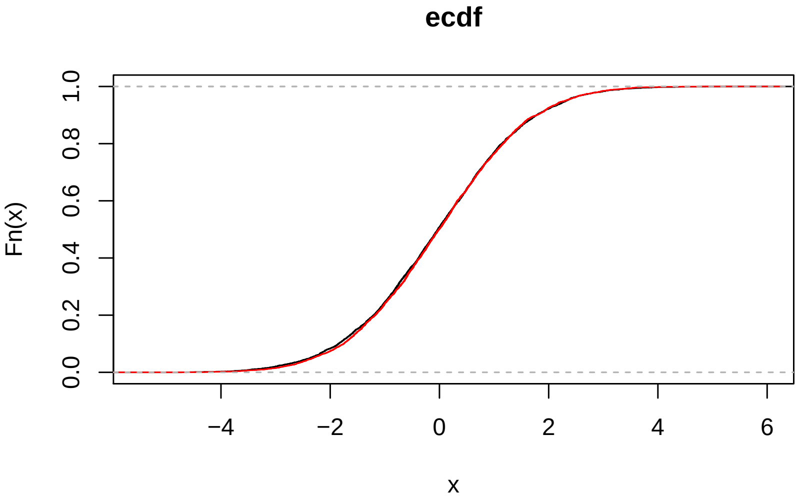
AR(1) and MA(∞) Example

```r
set.seed(123);
t<-100
d<-0;
arsim<-numeric(0)
masim<-numeric(0)
for(i in 1 : 5000){
  p<-1; q<-0;
  ar1<-arima.sim(list(order=c(p,d,q), ar=0.7), t)
  p<-0; q<-500;
  mainf<-arima.sim(list(order=c(p,d,q),
                        ma=(0.7)^(seq(1:q))), t)
  arsim[i]<-ar1[t] ;masim[i]<-mainf[t]}
plot(density(arsim), main="density");
lines(density(masim),col="red")
#plot(ecdf(arsim), main="ecdf");
#lines(ecdf(masim),col="red")
```

$AR(1)$ and $MA(\infty)$ Example



AR(1) and MA(∞) Example



Use of sampe ACF

Time series	ACF
White Noise	Zero
MA(q)	Zero for $\|h\| > q$
AR(p)	Decays to zero exponentially

Casulity

Definition

A linear process X_t is causal function of W_t if

$$X_t = \left(1 + \sum_{i=1}^{\infty} \phi_i B^i\right) W_t$$

where $\sum_{i=1}^{\infty} |\phi_i| < \infty$

- AR(1) i.e. $X_t = \phi X_{t-1} + W_t$ is casual iff $|\phi| < 1$ i.e. $1 - \phi z$ has a solution out of the unit circle on complex plane $\mathbb{C}$.

Invertibility

Definition

A linear process X_t is invertible function of W_t if there

$$W_t = \left(1 + \sum_{i=1}^{\infty} \theta_i B^i\right) X_t$$

where $\sum_{i=1}^{\infty} |\theta_i| < \infty$

- MA(1) i.e. $X_t = W_t + \theta W_{t-1}$ is invertable iff $|\theta| < 1$ i.e. $1 + \theta z$ has a solution out of the unit circle on complex plane $\mathbb{C}$.

Auto regressive-Moving average process (ARMA (p,q))

Definition

An **ARMA(p,q)** process X_t is a stationary process that satisfies

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$$

where $W_i \sim WN(0, \sigma_w^2)$.

- AR(p)= ARMA(p,0)
- MA(q)= ARMA(0,q)

$ARMA(p,q)$

- Back-shift operator B, such that $BX_t = X_{t-1}$
- $\Phi_p(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ if $\phi_p \neq 0$
- $\Theta_q(z) = 1 + \sum_{i=1}^{q} \theta_i z^i$ if $\theta_q \neq 0$
- Then ARMA(p,q) model can be written as

$$\Phi_p(B)X_t = \Theta_q(B)W_t$$

if $\Phi_p(z)$ and $\Theta_q(z)$ have no common factor i.e. not a lower order ARMA model is possible.

ARMA (1,1)

$$X_t - \phi X_{t-1} = W_t + \theta W_{t-1} \quad with \quad |\phi| < 1$$

- $E(X_t) = 0$
-

$$\gamma(h) = \begin{cases} \sigma^2 \left[1 + \frac{(\theta+\phi)^2}{1-\phi^2}\right] & if \quad h = 0 \\ \sigma^2 \left[\theta + \phi + \frac{\phi(\theta+\phi)^2}{1-\phi^2}\right] & if \quad h = 1 \\ \gamma(1)\phi^{h-1} & if \quad h \geq 2 \end{cases}$$

ARMA(p,q): Stationarity, causality, and invertibility

Theorem

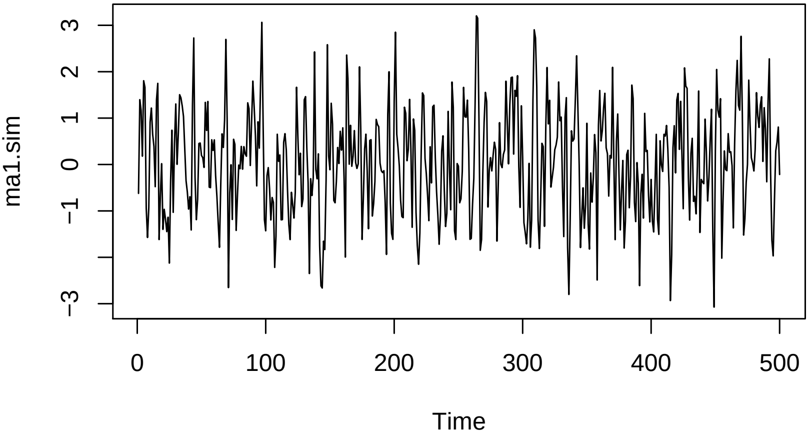
If Φ_p and Θ_q have no common factors, a (unique) stationary solution to $\Phi_p(B)X_t = \Theta_q(B)W_t$ exists iff the roots of $\Phi_p(z)$ avoid the unit circle i.e.

$$|z| = 1 \implies \Phi_p(z) \neq 0.$$

- This ARMA(p,q) process is causal iff the roots of $\Phi_p(z)$ are out- side the unit circle i.e. $|z| \leq 1 \implies \Phi_p(z) \neq 0$.
- It is invertible iff the roots of $\Theta_q(z)$ are outside the unit circle i.e. $|z| \leq 1 \implies \Theta_q(z) \neq 0$.

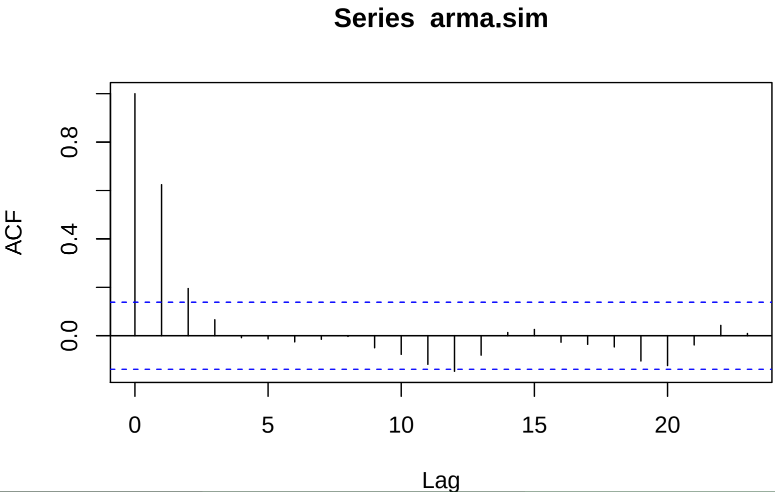
ARMA(1,1) Example

```
set.seed(123); n<-200; p<-1; d<-0;q<-1;
arma.sim<-arima.sim(list(order=c(p,d,q), ma=0.7, ar=.4), n)
ts.plot(ma1.sim)
```



ARMA(1,1) Example

```
set.seed(123); n<-200; p<-1; d<-0;q<-1;
arma.sim<-arima.sim(list(order=c(p,d,q), ma=0.7,ar=c(0.4)), n)
acf(arma.sim,type = "correlation",plot = T)
```



Partial Auto correlation function (PACF)

Definition

Partial Auto correlation function $\alpha(h)$ between X_h and X_0 for given $X_1, \cdots, X_{h-1}$ the correlation between the linear prediction errors $X_h - Ln(X_h|X_1, \ldots, X_{h-1})$ and $X_0 - Ln(X_0|X_1, \ldots, X_{h-1})$

$$\alpha(h) = \frac{\gamma(h) - \widetilde{\boldsymbol{\gamma}}^T_{h-1}(1)\Gamma^{-1}_{h-1}\boldsymbol{\gamma}_{h-1}(1)}{\gamma(0) - \boldsymbol{\gamma}^T_{h-1}(1)\Gamma^{-1}_{h-1}\boldsymbol{\gamma}_{h-1}(1)}$$

where

$$\widetilde{\boldsymbol{\gamma}}^T_{h-1}(1) = (\gamma(h-1)^T, \ldots, \gamma(1))$$

and

$$\boldsymbol{\gamma}^T_{h-1}(1) = (\gamma(1), \ldots, \gamma(h-1))^T$$

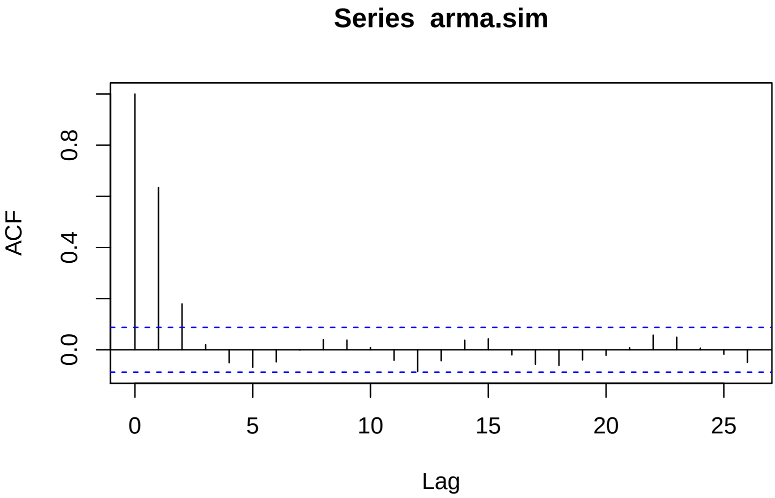
and

$$\Gamma_{h-1} = ((\gamma(|i-j|)))_{ij}$$

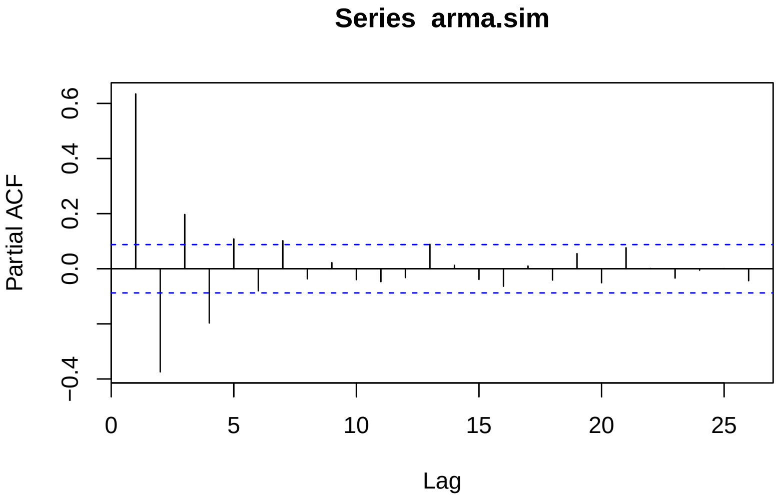
Classification

Model:	ACF:	PACF:
AR(p)	decays	zero for $h > p$
MA(q)	zero for $h > q$	decays
ARMA(p,q)	decays	decays

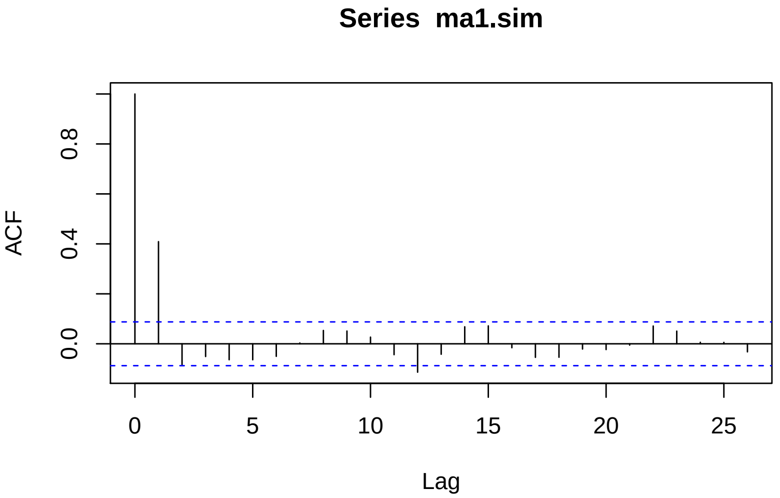
ARMA(1,1) Example



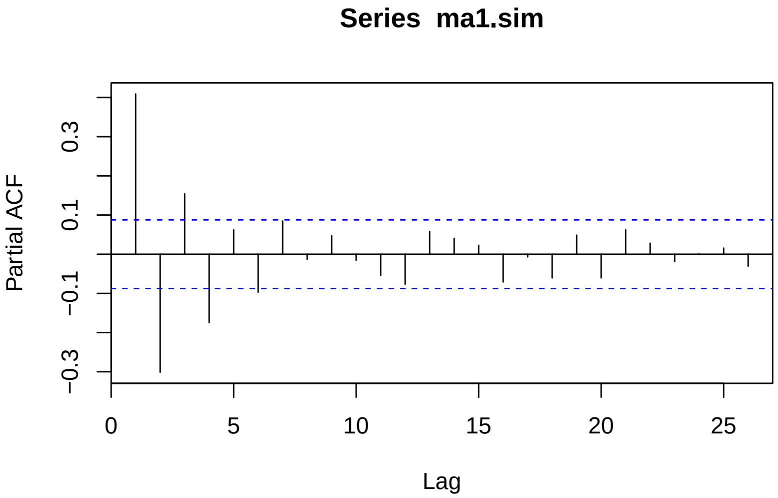
ARMA(1,1) Example



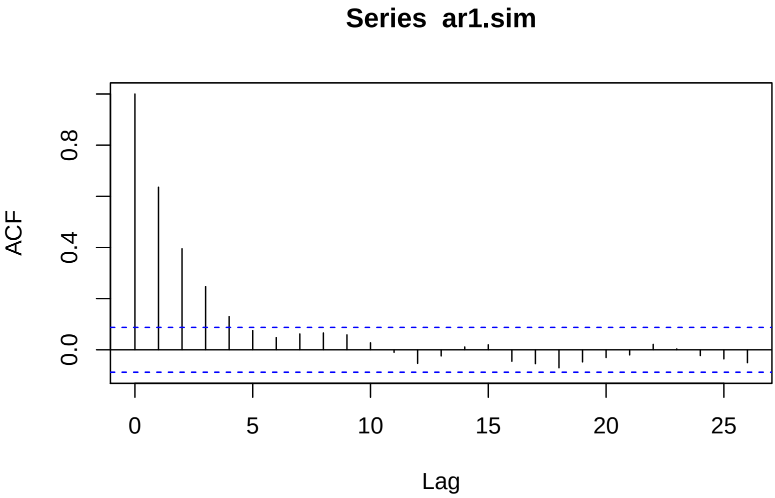
MA(1) Example



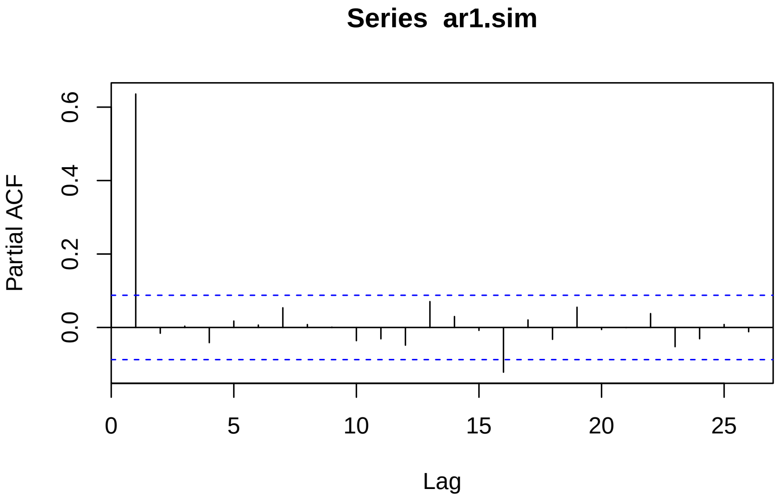
MA(1) Example



AR(1) Example



AR(1) Example



Esimation

For observations $x_1, \ldots, x_n$ of a time series,

$$\textit{Sample mean} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

The sample autocovariance function is

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_t - \bar{x})(x_{t+|h|} - \bar{x}) \quad \textit{if} \quad -n < h < n$$

The sample auto-correlation function is

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

About sample mean $\bar{x}$

- $E(\bar{x}) = \mu$ [unbiased estimator]
- $Var(\bar{x}) = \frac{1}{n} \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right) \gamma(h) \to 0 \;\; as \;\; n \to \infty$
- $\widehat{Var(\bar{x})} = \frac{1}{n} \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right) \widehat{\gamma(h)}$

Definition

Long run variance

$$\lim_{n\to\infty} nVar(\bar{x}) = \lim_{n\to\infty} \sum_{h=-\infty}^{\infty} \left(1 - \frac{|h|}{n}\right) \gamma(h) = \sigma_w^2 \left(\sum_{j=-\infty}^{\infty} \psi_j\right)^2$$

- $nVar(\bar{x}) \to \sum_{-\infty}^{\infty} \gamma(h)$ if $\sum_{-\infty}^{\infty} |\gamma(h)| < \infty$ as $n \uparrow \infty$.

About sample mean $\bar{x}$

If $\{X_t\}$ is a stationary Gaussian sequence then,

$$\sqrt{n}(\bar{x} - \mu) \sim N\left(0, \sum_{|h|<n}(1 - \frac{|h|}{n})\gamma(h)\right)$$

- 95% CI of μ is

$$\left(\bar{x} - 1.96\frac{\sigma_n}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma_n}{\sqrt{n}}\right)$$

where $\sigma_n^2 = \sum_{|h|<n}\gamma(h) \approx \sum_{|h|<\sqrt{n}}\hat{\gamma}(h)$

Bartlett's formula about $\hat{\rho}(h)$

If $X_t = \mu + \sum_{j=-\infty}^{j=+\infty} \psi_j W_{t-j}$ with $E(X_t^4) < \infty$ then the approximate joint density of $(\hat{\rho}(i), \hat{\rho}(j))$ is

$$\begin{pmatrix} \hat{\rho}(i) \\ \hat{\rho}(j) \end{pmatrix} \sim N\left[\begin{pmatrix} \rho(i) \\ \rho(j) \end{pmatrix}, \frac{1}{n}\begin{pmatrix} v_{ii} & v_{ij} \\ v_{ji} & v_{jj} \end{pmatrix}\right]$$

where,

$$v_{ij} = \sum_{h=1}^{\infty} (\rho(h+i) + \rho(h-i) - 2\rho(i)\rho(h)) \times (\rho(h+j) + \rho(h-j) - 2\rho(j)\rho(h))$$

Forecasting

- Given $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ the best linear predictor

$$\hat{X}^n_{m+n} = \sum_{i=1}^{n} \alpha_i X_i$$

satisfying the following conditions

1. $E(\hat{X}^n_{m+n} - X_{m+n}) = 0$ [Unbiased prediction]
2. $E[(\hat{X}^n_{m+n} - X_{m+n})X_i] = 0$ [Error is orthogonal to predictors]

- Durbin-Levinson estimate : Coefficient for 1 step prediction

$$\hat{\boldsymbol{\alpha}} = \Gamma_n^{-1} \boldsymbol{\gamma}_n(1)$$

- Prediction error: $E(X_{n+1} - \hat{\boldsymbol{\alpha}}^T \mathbf{X})^2 = \gamma(0) - \boldsymbol{\gamma}_n^T(1) \Gamma_n^{-1} \boldsymbol{\gamma}_n(1)$

- Smoothing with a finite moving average filter: Let q be a nonnegative integer and consider the two-sided moving average

$$m_t = \frac{1}{2q+1}\sum_{i=t-q}^{t+q} X_i \approx \frac{1}{2q+1}\sum_{i=t-q}^{t+q} T_i \approx T_t$$

  if T_t is linear in $[t-q, t+q]$

- Linea filter :

$$m_t = \sum_{i=-q}^{q} a_i X_{t-i}$$

Estimation of Trend in the Absence of Seasonality

- Exponential smoothing: For any fixed $\alpha \in [0, 1]$, the one-sided moving $m_t = X_1$ and

$$m_t = \alpha X_t + (1 - \alpha) m_{t-1} \quad for \quad t = 2, \ldots, n$$

- Polynomial fitting: For example fit $m_t = a_0 + a_1 t + a_2 t^2$ which minimize the suare distance $\sum_{t=1}^{n} (x_t - m_t)^2$.
- Trend Elimination by Differencing: Estimate k such that $\nabla^k X_t \approx constant$. Then fit k-degree polynomial

- Additive model : $X_t = T_t + S_t + W_t$

with $E(W_t) = 0$, $S_{t+d} = S_t$ and $\sum_{t=1}^{d} S_t = 0$ then

$$\nabla_d X_t = T_t - T_{t-d} + W_t - W_{t-d}$$

- Now the trend $T_t - T_{t-d}$ can be removed by using any of the above methods.

Testing the Estimated Noise Sequence

- If there is no dependence among these residuals, then we can regard them as observations of independent random variables, and there is no further modeling to be done except to estimate their mean and variance.

1. The sample autocorrelation function, 95% CI$=(-1.96/\sqrt{n}, +1.96/\sqrt{n})$
2. Ljung and Box (1978) test : $Q = n(n+2)\sum_{i=1}^{h} \hat{\rho}^2(j)/(n-j)$ whose distribution is the chi-squared distribution with h degrees of freedom.
3. Non-parametric test: Rank test, run testst, sign test etc..

- Given $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ the best linear predictor

$$\hat{X}^n_{m+n} = \sum_{i=1}^{n} \alpha_i X_i$$

satisfying the following conditions

1. $E(\hat{X}^n_{m+n} - X_{m+n}) = 0$ [Unbiased presdiction]
2. $E[(\hat{X}^n_{m+n} - X_{m+n})X_i] = 0$ [Error is orthogonal to predictors]

- Durbin-Levinson estimate : Coefficient for 1 step prediction

$$\hat{\boldsymbol{\alpha}} = \Gamma_n^{-1} \boldsymbol{\gamma}_n(1)$$

- Prediction error: $E(X_{n+1} - \hat{\boldsymbol{\alpha}}^T \mathbf{X})^2 = \gamma(0) - \boldsymbol{\gamma}_n^T(1) \Gamma_n^{-1} \boldsymbol{\gamma}_n(1)$

$$\hat{X}_{n+1}^{n} = \begin{cases} 0 \quad if \quad n = 1 \\ \sum_{j=1}^{n} \theta_{nj}(X_{n-j+1} - \hat{X}_{n-j+1}^{n-j}) \quad otherwise \end{cases}$$

- The innovations $(X_{n-j+1} - \hat{X}_{n-j+1}^{n-j})$ are uncorrelated.

Estimation of Model parameters (Yule-Walker)

- Remove the trend
- Remove the seasonal effect
- Now for weakly stationary process use

1. Yule-Walker estimator : This eventually a method of moment estimation process. So equate the theoretical moments with the corresponding sample moments of ARMA(p,q) and solve them.

- $$\hat{\Gamma}_n \hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\gamma}}_n(1)$$
- $$\hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\boldsymbol{\alpha}}^T \hat{\boldsymbol{\gamma}}_n(1)$$

- Remove the trend
- Remove the seasonal effect
- Now for weakly stationary process use

2. MLE: Suppose that $X_1, X_2, ..., X_n$ is drawn from a zero mean Gaussian ARMA(p,q) process. The likelihood of parameters $\boldsymbol{\phi} \in \mathbb{R}^p$, $\boldsymbol{\theta} \in \mathbb{R}^q$ and $\sigma_w^2 > 0$ is defined as the density of $\mathbf{X} = (X_1, X_2, ..., X_n)^T$ under the multivariate Gaussian model with those parameters.

$$f_{\mathbf{X}}(x_1, \ldots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x})^{\mathrm{T}}\boldsymbol{\Gamma_n}^{-1}(\mathbf{x})\right)}{\sqrt{(2\pi)^k|\boldsymbol{\Gamma_n}|}}$$

ARIMA

Definition

For $p, d, q \geq 0$, we say that a time series X_t is an $ARIMA(p, d, q)$ process if $Y_t = (1 - B)^d X_t$ is $ARMA(p, q)$. We can write

$$\Phi_p(B)\nabla^d X_t = \Theta_q(B)W_t.$$

- Example: [ARIMA(0,1,0)] Random walk with drift

$$X_t = \mu t + \sum_{i=0}^{t} W_i \quad where \quad W_i \sim N(0, \sigma^2) \quad i.i.d.$$

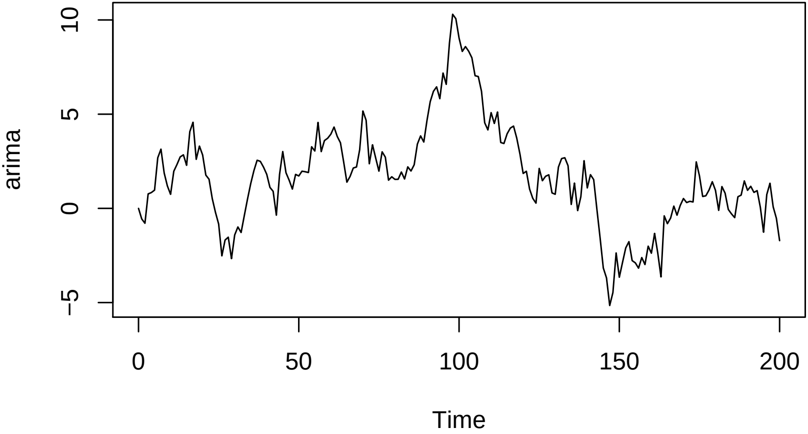
This implies

$$\nabla X_t - \mu \sim N(0, \sigma^2) \quad which \ is \ a \ White \ noise$$

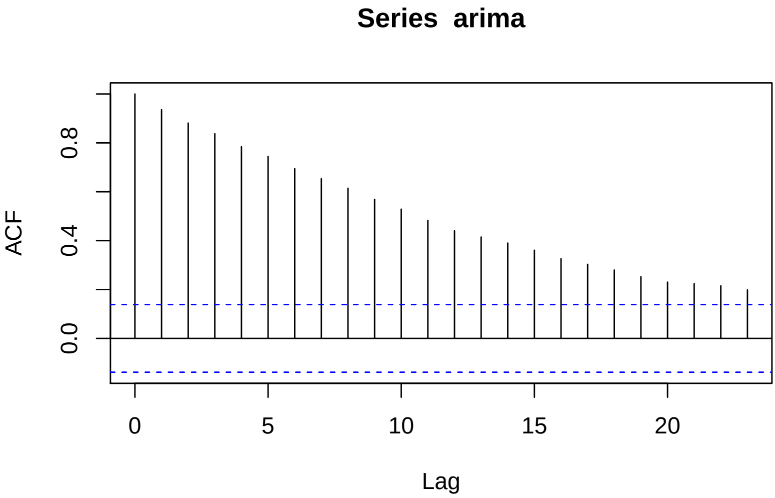
ARMA(0,1,0) Example


```
set.seed(123); n<-200; p<-0; d<-1;q<-0;
arima<-arima.sim(list(order=c(p,d,q)), n)
par(mfrow=c(1,1))
ts.plot(arima)
acf(arima.sim,type = "correlation",plot = T)
pacf(arma.sim,plot = T)
```

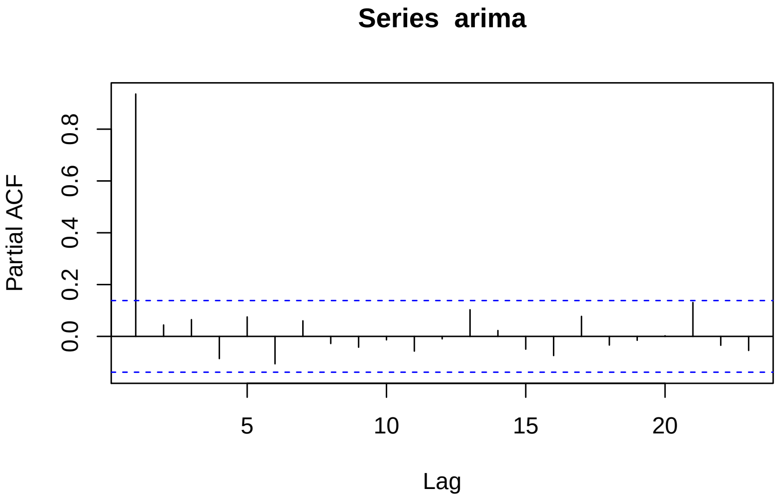
ARMA(0,1,0) Example



ARMA(0,1,0) Example



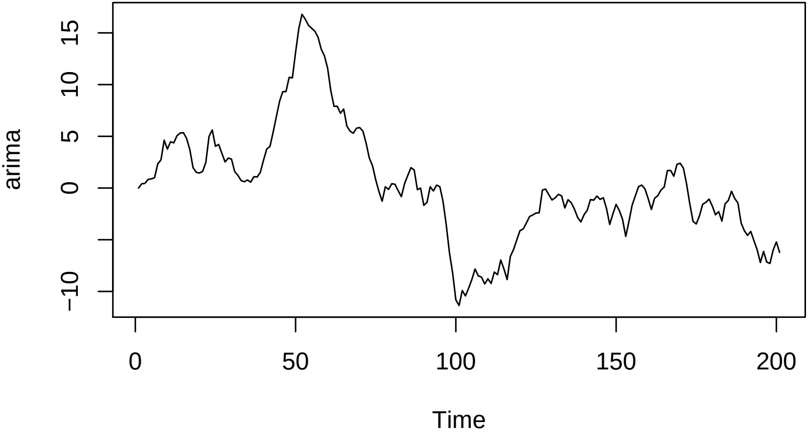
ARMA(0,1,0) Example

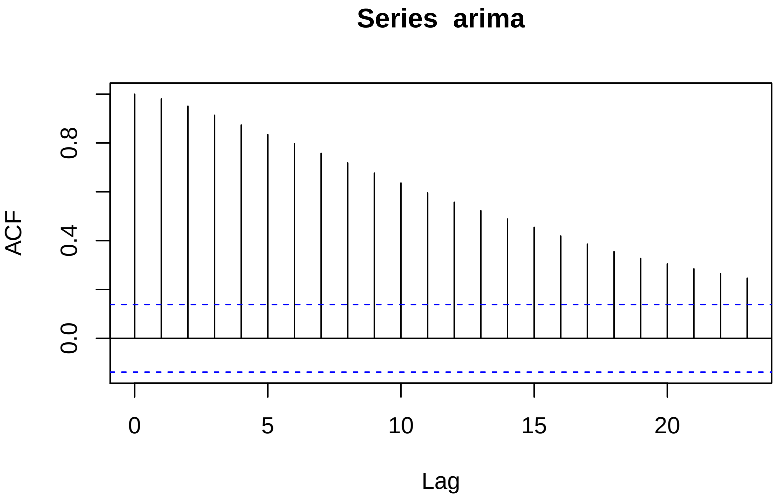


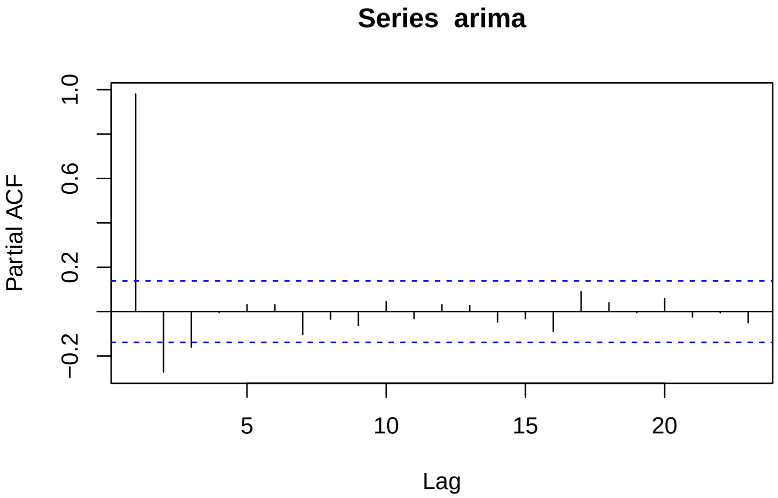
ARMA(2,1,1) Example

```r
set.seed(123); n<-200; p<-2; d<-1;q<-1;
arima<-arima.sim(list(order=c(p,d,q), ar=c(-0.3, 0.5),
                                                ma=c(0.7)), n)
#arima1<-diff(arima, lag = 1)
par(mfrow=c(1,1))
ts.plot(arima1)
acf(arima,type = "correlation",plot = T)
pacf(arima, plot = T)
```

ARMA(2,1,1) Example



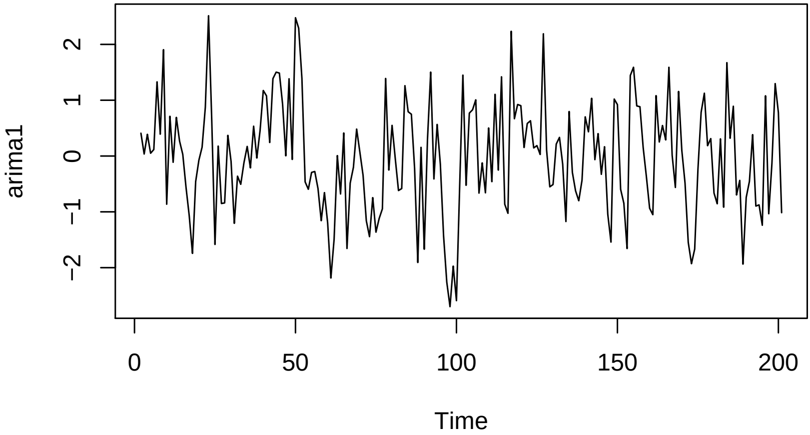

Series arima
ACF
0.0
0.4
0.8
0
5
10
15
20
Lag


Series arima
Partial ACF
1.0
0.6
0.2
−0.2
5
10
15
20
Lag

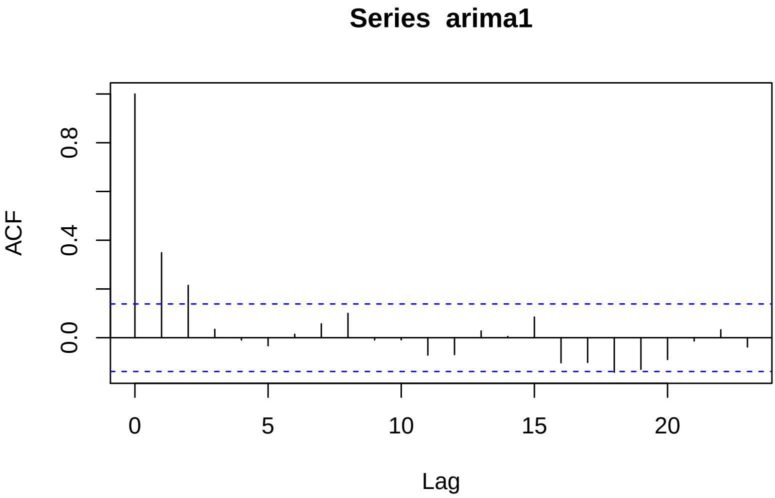
ARMA(2,1,1) Example

```
set.seed(123); n<-200; p<-2; d<-1;q<-1;
arima<-arima.sim(list(order=c(p,d,q), ar=c(-0.3, 0.5),
                       ma=c(0.7)), n)
arima1<-diff(arima, lag = 1)
par(mfrow=c(1,1))
ts.plot(arima1)
acf(arima1,type = "correlation",plot = T)
pacf(arima1, plot = T)
```

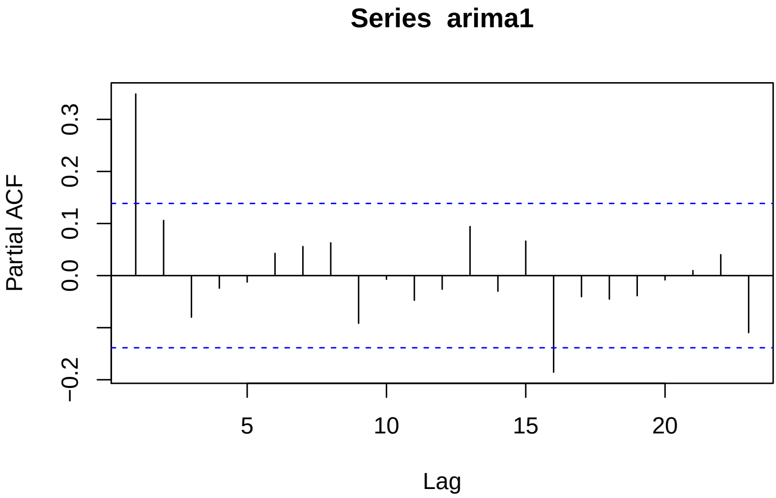
ARMA(2,1,1) Example



ARMA(2,1,1) Example



ARMA(2,1,1) Example



ARMA(2,1,1) Example

```
set.seed(123); n<-200; p<-2; d<-1;q<-1;
arima<-arima.sim(list(order=c(p,d,q), ar=c(-0.3, 0.5),
                      ma=c(0.7)), n)
library('forecast')
auto.arima(arima)
```

ARMA(2,1,1) Example

```
## Warning: package 'forecast' was built under R version 3.4.4

## Warning in as.POSIXlt.POSIXct(Sys.time()): unknown timezone 'zone
## 1.0/zoneinfo/Asia/Kolkata'

## Series: arima
## ARIMA(1,1,0)
##
## Coefficients:
##          ar1
##       0.3497
## s.e.  0.0662
##
## sigma^2 estimated as 0.8799:  log likelihood=-270.55
## AIC=545.11   AICc=545.17   BIC=551.7
```

ARMA(2,1,1) Example

```
set.seed(123); n<-1000; p<-2; d<-1;q<-1;
arima<-arima.sim(list(order=c(p,d,q), ar=c(-0.3, 0.5),
                      ma=c(0.7)), n)
auto.arima(arima)
```

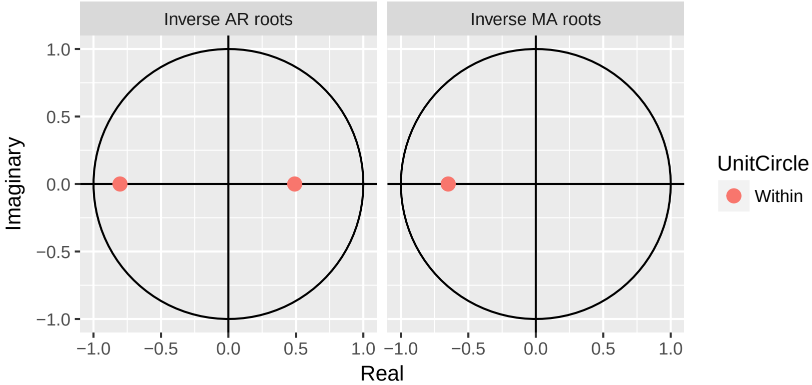
```
## Series: arima
## ARIMA(2,1,1)
##
## Coefficients:
##          ar1     ar2     ma1
##      -0.3081  0.4640  0.6713
## s.e.  0.0514  0.0288  0.0522
##
## sigma^2 estimated as 1.005:  log likelihood=-1419.96
## AIC=2847.92   AICc=2847.96   BIC=2867.55
```

ARMA(2,2,1) Example

```r
set.seed(123); n<-1000; p<-2; d<-2;q<-1;
arima<-arima.sim(list(order=c(p,d,q), ar=c(-0.3, 0.5),
                      ma=c(0.7)), n)
auto.arima(arima)
```

```
## Series: arima
## ARIMA(2,2,1)
##
## Coefficients:
##           ar1     ar2     ma1
##       -0.3081  0.4640  0.6713
## s.e.   0.0514  0.0288  0.0522
##
## sigma^2 estimated as 1.005:  log likelihood=-1419.96
## AIC=2847.92   AICc=2847.96   BIC=2867.55
```

ARMA(2,2,1) Example

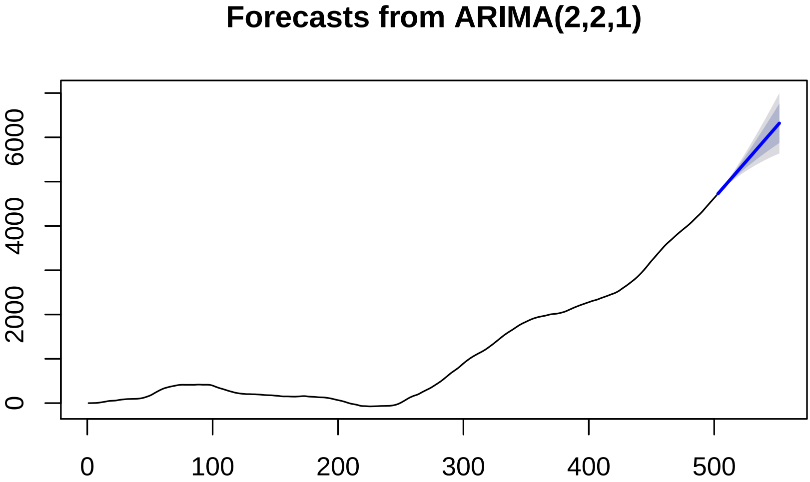


ARMA(2,2,1) Example

```r
set.seed(123); n<-500; p<-2; d<-2;q<-1;
arima<-arima.sim(list(order=c(p,d,q), ar=c(-0.3, 0.5),
                      ma=c(0.7)), n)
aarima<-auto.arima(arima)
fc<-forecast(aarima, h=5)
print(fc)
```

```
##     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 503       4726.819 4725.588 4728.050 4724.936 4728.702
## 504       4759.264 4756.135 4762.393 4754.479 4764.050
## 505       4791.706 4785.911 4797.502 4782.843 4800.570
## 506       4824.187 4815.150 4833.223 4810.367 4838.007
## 507       4856.654 4843.822 4869.487 4837.028 4876.280
```

ARMA(2,2,1) Example



Multiplicative seasonal ARIMA Models

Definition

For $p, q, P, Q \geq 0$, $s > 0$, $d, D \geq 0$, we say that a time series X_t is a multiplicative seasonal ARIMA model $ARIMA(p, d, q) \times (P, D, Q)_s$

$$\Phi_P(B^s)\phi_p(B)\nabla_s^D \nabla^d X_t = \Theta_Q(B^s)\Theta_q(B)W_t,$$

where the seasonal difference operator of order D is defined by

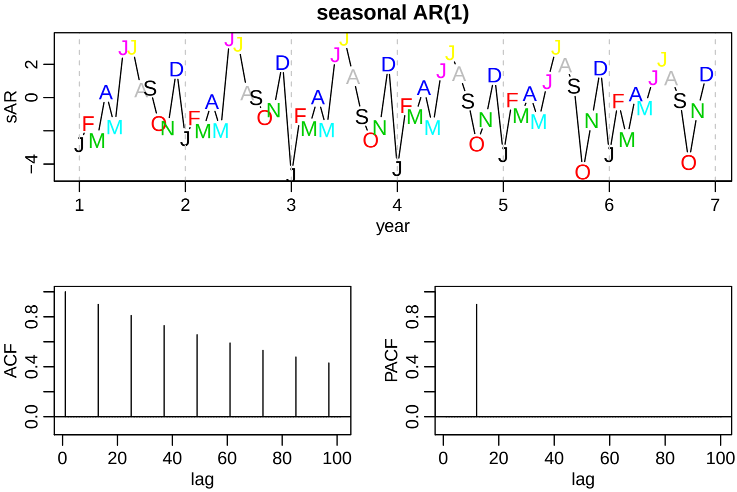
$$\nabla_s^D X_t = (1 - B^s)^D X_t$$

.

- Example: Auto regressive conditional heteroskedasticity (ARCH) model

```
set.seed(666)
phi = c(rep(0,11),.9)
sAR = arima.sim(list(order=c(12,0,0), ar=phi), n=72)
sAR = ts(sAR, freq=12)
layout(matrix(c(1,2, 1,3), nc=2))
par(mar=c(3,3,2,1), mgp=c(1.6,.6,0))
plot(sAR, axes=FALSE, main='seasonal AR(1)', xlab="year", type='c')
Months = c("J","F","M","A","M","J","J","A","S","O","N","D")
points(sAR, pch=Months, cex=1.25, font=12, col=1:12)
axis(1, 1:12); abline(v=1:12, lty=2, col='#cccccc')
axis(2); box()
ACF = ARMAacf(ar=phi, ma=0, 100)
PACF = ARMAacf(ar=phi, ma=0, 100, pacf=TRUE)
plot(ACF,type="h", xlab="lag", ylim=c(-.1,1)); abline(h=0)
plot(PACF, type="h", xlab="lag", ylim=c(-.1,1));
abline(h=0)
```

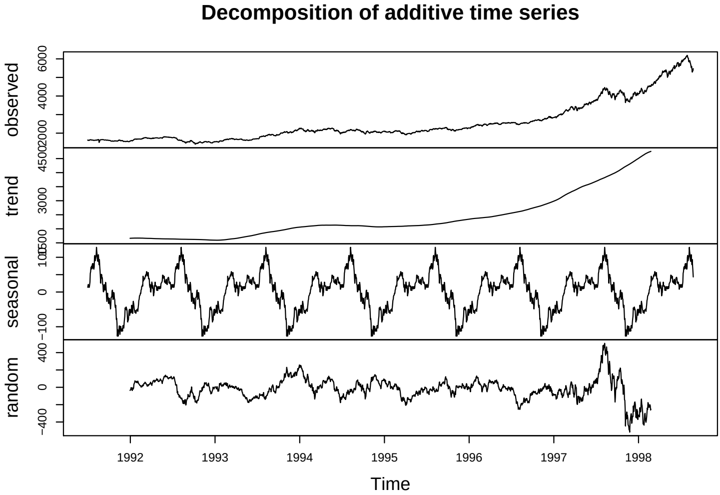
Multiplicative seasonal ARIMA Models



- Daily Closing Prices of Major European Stock Indices, 1991-1998
- Description: Contains the daily closing prices of major European stock indices: Germany DAX (Ibis), Switzerland SMI, France CAC, and UK FTSE. The data are sampled in business time, i.e., weekends and holidays are omitted.
- Format: A multivariate time series with 1860 observations on 4 variables. The object is of class "mts".
- Source: The data were kindly provided by Erste Bank AG, Vienna, Austria.

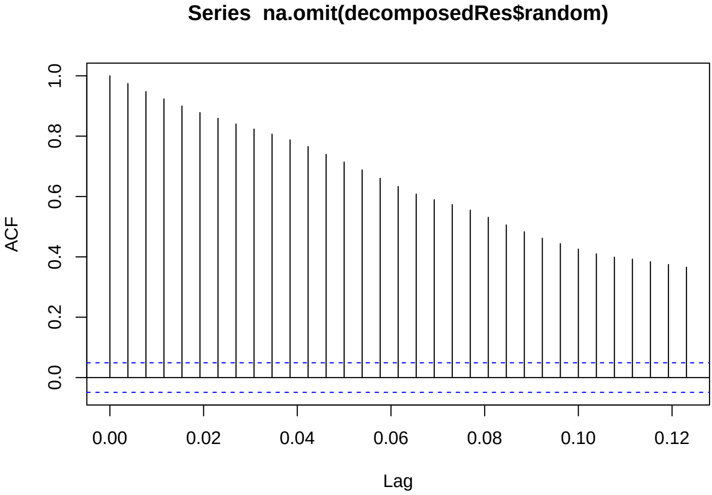
European stock indices

```r
tsData <- EuStockMarkets[, 1] # ts data
decomposedRes <- decompose(tsData, type="additive")
plot (decomposedRes) # see plot below
```


Decomposition of additive time series
observed
2000
4000
6000
trend
1500
3000
4500
seasonal
−100
0
100
random
−400
0
400
1992
1993
1994
1995
1996
1997
1998
Time

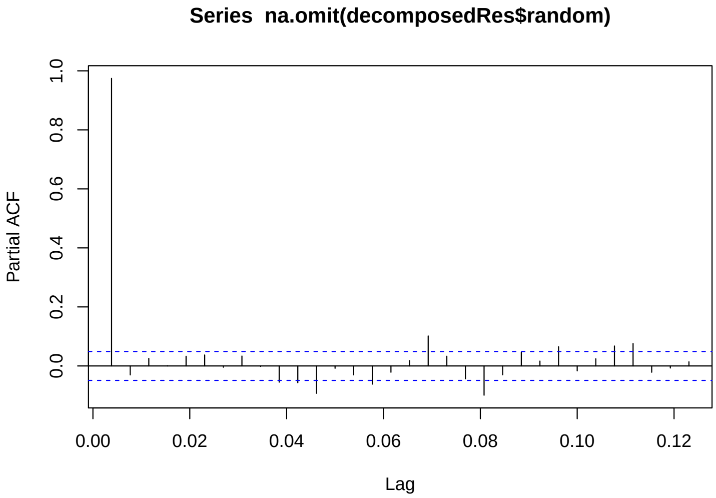
European stock indices


```r
tsData <- EuStockMarkets[, 1] # ts data
decomposedRes <- decompose(tsData, type="additive")
acf(na.omit(decomposedRes$random))
```


Series na.omit(decomposedRes$random)
ACF
Lag
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.02
0.04
0.06
0.08
0.10
0.12

European stock indices

```
tsData <- EuStockMarkets[, 1] # ts data
decomposedRes <- decompose(tsData, type="additive")
pacf(na.omit(decomposedRes$random))
```


Series na.omit(decomposedRes$random)
Partial ACF
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Lag

Augmented Dickey-Fuller Test of stationary

```
tsData <- EuStockMarkets[, 1] # ts data
decomposedRes <- decompose(tsData, type="additive")
library('tseries')
# p-value < 0.05 indicates the TS is stationary
adf.test(na.omit(decomposedRes$random))
```

```
## Warning: package 'tseries' was built under R version 3.4.4

## Warning in adf.test(na.omit(decomposedRes$random)): p-value smal
## printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  na.omit(decomposedRes$random)
## Dickey-Fuller = -4.6053, Lag order = 11, p-value = 0.01
## alternative hypothesis: stationary
```

The Akaike information criterion (AIC) is an estimator of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection.

Suppose that we have a statistical model of some data. Let k be the number of estimated parameters in the model. Let $\hat{L}$ be the maximum value of the likelihood function for the model. Then the AIC value of the model is the following.

$$\mathrm{AIC} = 2k - 2\ln(\hat{L})$$

BIC

In statistics, the Bayesian information criterion (BIC) is a criterion for model selection among a finite set of models; the model with the lowest BIC is preferred.

The BIC is formally defined as

$$\mathrm{BIC} = \ln(n)k - 2\ln(\hat{L}).$$