TIME SERIES

Buddhananda Banerjee
Department of Mathematics
&
Centre for Artificial Intelligence
Indian Institute of Technology Kharagpur
e-mail:bbanerjee@maths.iitkgp.ac.in

Last updated on November 3, 2021

Contents

1. Estimation

Let $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ be the observed/ realized values of a set of i.i.d. random variables $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ where$X_i \overset{iid}{\sim} f_\theta$ for some $\theta \in \Theta$. Here a family of distributions is denoted by

$$\mathcal{F} = \{f(x|\theta)|\theta \in \Theta\} \quad \text{or} \quad \{F(x|\theta)|\theta \in \Theta\}$$

Parametric Estimation: In a parametric inference problem it is assumed that the family of the distribution is known but the particular value of the parameter is unknown. We estimate the value of the parameter θ as a function of the observations $\mathbf{x}$. The ultimate goal is to approximate the p.d.f f_θ or F_θ through the estimation of θ itself. Parametric estimation has two aspects, namely, (a) **Point estimation** and (b) **Interval estimation** .

Definition 1. Statistic: A statistic is a function of random variables and it is free from any unknown parameter. Being a (measurable) function,$T(\mathbf{X})$ say , of random variables it is also a random variable.

Definition 2. Estimator: If the statistic $T(\mathbf{X})$ is used to estimate a parametric function $g(\theta)$ then $T(X)$ is said to be {an estimator of $g(\theta)$. And a realized value of it for $\mathbf{X} = \mathbf{x}$ i.e. $T(\mathbf{x})$ is know as **an estimate** of θ. We often abuse the notation as $g(\hat{\theta}) = T(\mathbf{x})$ and $g(\hat{\theta}) = T(\mathbf{X})$ which are understood from the context.

Definition 3. Unbiased estimator: An estimator $T(\mathbf{X})$ is said to be an unbiased estimator of a parametric function $g(\theta)$ if $E(T(\mathbf{X}) - g(\theta)) = 0 \ \forall \ \theta \in \Theta$.

Remark 4. It does not require$T(\mathbf{x}) = g(\theta)$ to be hold or it may hold with probability zero.

Definition 5. Bias: The bias of an estimator $T(\mathbf{X})$ while estimating a parametric function $g(\theta)$ is $B_{g(\theta)}(T(\mathbf{X})) = E(T(\mathbf{X}) - g(\theta)) \ \forall \ \theta \in \Theta$.

Definition 6. Asymptotically unbiased estimator: Denoting $T_n = T(X_1, X_2, \cdots, X_n)$ an estimator T_n is said to be asymptotically unbiased of $g(\theta)$ if

$$\lim_{n\to\infty} B_{g(\theta)}(T_n) = \lim_{n\to\infty} E(T_n - g(\theta)) = 0$$

Definition 7. Consistent estimator: An estimator T_n is said to be consistent estimator $g(\theta)$ if $T_n \xrightarrow{P} g(\theta)$ i.e.

$$\lim_{n\to\infty} P(|T_n - g(\theta)| < \epsilon) = 1 \ \forall \ \theta \in \Theta, \epsilon > 0$$

Definition 8. Mean squared error (MSE): The MSE of an estimator $T(\mathbf{X})$ while estimating a parametric function $g(\theta)$ is

$$MSE_{g(\theta)}(T(\mathbf{X})) = E[(T(\mathbf{X}) - g(\theta))^2] \ \forall \ \theta \in \Theta.$$

Remark 9. $MSE_{g(\theta)}(T(\mathbf{X})) = Var(T(\mathbf{X})) + B^2_{g(\theta)}(T(\mathbf{X}))$

Remark 10. If $MSE_{g(\theta)}(T_n(\mathbf{X})) \downarrow 0$ as $n \uparrow \infty$ then $(T_n(\mathbf{X}))$ is a consistent estimator.

Remark 11. Asymptotic unbiasedness and consistency are large sample properties and both are based on L_1 norm. . MSE is defined based on L_2 norm.

Remark 12. Let $(X_1, X_2, \cdots, X_n)$ be i.i.d random variables with $E(X) = \mu$ and $Var(X) = \sigma^2$. and define $T_n(\mathbf{X}) = \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, $S_1^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ and $S_2^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$. Then
(a) $T_n(\mathbf{X})$ is an unbiased estimator of μ.
(b) S_1^2 is a biased estimator of σ^2
(c) S_2^2 is an asymptotically unbiased estimator of σ^2

Remark 13. Let $(X_1, X_2, \cdots, X_n) \overset{iid}{\sim} N(\mu, \sigma^2)$. Show that $MSE(S_2^2) < MSE(S_1^2)$. Note: Unbiased estimator need not have minimum MSE.

Definition 14. Method of Moment for Estimation (MME): Consider $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ be the observed/ realized values of a set of i.i.d. random variables $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ where $X_i \overset{iid}{\sim} f_\theta$ for some $\theta \in \Theta$. Then
Step 1: Computer theoretical moments from the p.d.f.
Step 2: Computer empirical moments from the data.
Step 3: Construct k equations if you have k unknown parameters.
Step 4: Solve the equations for the parameters.

Remark 15. We can not use MME to estimate the parameters of $C(\mu, \sigma)$, because the moments does not exists for Cauchy distribution.

Definition 16. Maximum Likelihood Estimator: Consider $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ be the observed/ realized values of a set of i.i.d. random variables $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ where $X_i \overset{iid}{\sim} f_\theta$ for some $\theta \in \Theta$. Then the joint p.d.f. of $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ is a function of $\mathbf{x}$ when the parameter value is fixed i.e.

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i, \theta)$$

and the likelihood of a function of parameter for a given set of data $\mathbf{X} = \mathbf{x}$ i.e.

$$\ell(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i, \theta).$$

Hence the maximum likelihood estimator of θ is

$$\hat{\theta}_{mle} = \arg\max_{\theta\in\Theta} \ell(\theta|\mathbf{x}) = \arg\max_{\theta\in\Theta} \log \ell(\theta|\mathbf{x})$$

NOTE: Finding the maxima through differentiation is possible **only of** ℓ is a smoothly differentiable function w.r.t θ. Otherwise it has to be maximized by some other methods. **Differentiation is not the only way of finding maxima or minima.**

Exercise 17. Let $(X_1, X_2, \cdots, X_n) \overset{iid}{\sim} N(\mu, \sigma^2)$.
(a) Find the MME and MLE of μ and σ^2. Are they same ?
(b) Are they unbiased estimators?

Exercise 18. Let $(X_1, X_2, \cdots, X_n) \overset{iid}{\sim} Gamma(\alpha, \lambda)$.
(a) Find the MME of (α, λ)?
(b) Find MLE of (α, λ) by by an iterative method of solution.
NOTE: You may use the MME as an initial value of iteration to obtain the MLE.

Propertied of MLE:
(a) MLE need not be unique.
(b) MLE need not be an unbiased estimator.
(c) MLE is always a consistent estimator.
(d) MLE is asymptotically normally distributed up to some location and scale when some regularity condition satisfied like
(1) Range of the random variable is free from parameter.
(2) Likelihood is smoothly differentiable for up to 3rd order and corresponding expectations exists.

Definition 19. Interval Estimation: Consider a pair of statistic $(L(\mathbf{X}), U(\mathbf{X}))$ such that for a parameter θ ,

$$P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})]) = 1 - \alpha$$

Then a $100(1-\alpha)\%$ confidence interval of θ is considered to be $[L(\mathbf{X}), U(\mathbf{X})]$.

Example 20. If $X_1, X_2, \ldots, X_n$ are i.i.d random variables with $N(\mu, \sigma^2)$ distribution with known value of σ^2. Then a $100(1-\alpha)\%$ CI of μ is

$$\left[L(\mathbf{X}) = \overline{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, U(\mathbf{X}) = \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right]$$

Example 21. If $X_1, X_2, \ldots, X_n$ are i.i.d random variables with $N(\mu, \sigma^2)$ distribution . Then a $100(1-\alpha)\%$ CI of μ is

$$\left[L(\mathbf{X}) = \overline{X} - \frac{\hat{\sigma_u}}{\sqrt{n}} \tau_{\alpha/2, n-1}, U(\mathbf{X}) = \overline{X} + \frac{\hat{\sigma_u}}{\sqrt{n}} \tau_{\alpha/2, n-1}\right]$$

$\hat{\sigma}_u^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ is an unbiased estimator of unknown variance and a $100(1-\alpha)\%$ CI of σ^2 is

$$\left[L(\mathbf{X}) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi^2_{\alpha/2,(n-1)}}, U(\mathbf{X}) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi^2_{1-\alpha/2,(n-1)}}\right]$$

2. Testing of Hypothesis

Definition 22. Hypothesis: A hypothesis in parametric inference is a statement about the population parameter. It has two categories. A **null hypothesis (H_0)** specifies a subset Θ_0 in the parameter space Θ. If Θ_a is a singleton set then it called a **simple null**, otherwise a **composite null.** On the other hand an **alternative hypothesis (H_1)** specifies another subset $\Theta_a \subset \Theta$ which is disjoint to Θ_0.

Definition 23. Test Rule: A test rule is a statistical procedure, based on the distribution of the test statistic, which will reject the null hypothesis in favour of the alternative hypothesis.

Definition 24. Rejection Region or Critical region: A rejection Region or critical region is a subset $C \subset \mathbb{R}^n$ such that $\mathbf{X} \in C \Leftrightarrow T(\mathbf{X})$ will reject the null hypothesis.

Definition 25. Level-α test: For any $\alpha \in (0, 1)$, a test is said to be level-α test if

$$\sup_{\theta \in \Theta_0} P_\theta(\mathbf{X} \in C) \leq \alpha.$$

Definition 26. Size-α test: For any $\alpha \in (0, 1)$, a test is said to be size-α test if

$$\sup_{\theta \in \Theta_0} P_\theta(\mathbf{X} \in C) = \alpha.$$

Definition 27. Power-function: A power function is a function

$$P_\theta(\mathbf{X} \in C) : \Theta_a \to [0, 1]$$

Remark 28. More than one tests with same level can be compared in terms of power functions. A test procedure with more power than the other with same level can be considered a better test.

Definition 29. Type-I error: The event $\mathbf{X} \in C$ when $\theta \in \Theta_0$ is known as Type-I error.

Definition 30. Type-II error: The event $\mathbf{X} \in C^c$ when $\theta \in \Theta_a$ is known as Type-II error. Power is 1-P(Type-II error).

Lemma 31. ***Neyman-Pearson Lemma (1933):*** *To test $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ reject H_0 in favour of H_1 at level/ size α if*

$$\Lambda(\mathbf{x}) = \frac{f(\mathbf{x}|\theta_0)}{f(\mathbf{x}|\theta_1)} \leq \xi \quad \textit{such that} \quad P_{\theta_0}(\Lambda(\mathbf{X}) \leq \xi) = \alpha$$

How to perform a test ??

Step1: Estimate the parameter for which the testing to be done.

Step2: Estimate the unknown parameters if any.

Step3: Construct the test statistic and obtain its value.

Step4: Obtain the exact or asymptotic distribution of the test statistic under the null hypothesis.

Step5: Depending on the alternative hypothesis (H_1) and level (α) decide the cut-off value or rejection condition.

Step6: Compare the observed value of test statistic (from Step 4) and the cut off value (from Step 5) to conclude the test. You may use **p-value** also.

Exercise 32. Let $X_1, ..., X_n \sim N(\mu_, \sigma^2)$ Perform a test at size 0.05 for

(a)$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$. when σ^2 is known

(b)$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$. when σ^2 is unknown

(a)$H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$ when μ is unknown

Exercise 33. Let $X_1, ..., X_n \sim N(\mu_1, \sigma^2)$ (iid) and $Y_1, ..., Y_m \sim N(\mu_2, \sigma^2)$ (iid) are independent. Perform a test at size 0.05 for $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$.

Exercise 34. Let $X_1, ..., X_n \sim N(\mu_1, \sigma_1^2)$ (iid) and $Y_1, ..., Y_m \sim N(\mu_2, \sigma_2^2)$ (iid) are independent. Perform a test at size 0.05 for $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$.

3. Introduction

Definition 35. 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. 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_i s are some numeric or categorical values.

Categories of Time series

(1) Discrete time series: If T is a countable set then it is a discrete time series.
(2) Continuous time series: If T is an interval then it is a continuous time series.

Note: Discrete or continuous are the adjectives of time but NOT of random variable X_t

- Johnson & Johnson Quarterly Earnings

```
library(astsa)
plot(jj, type="o", ylab="Quarterly Earnings per Share")
```

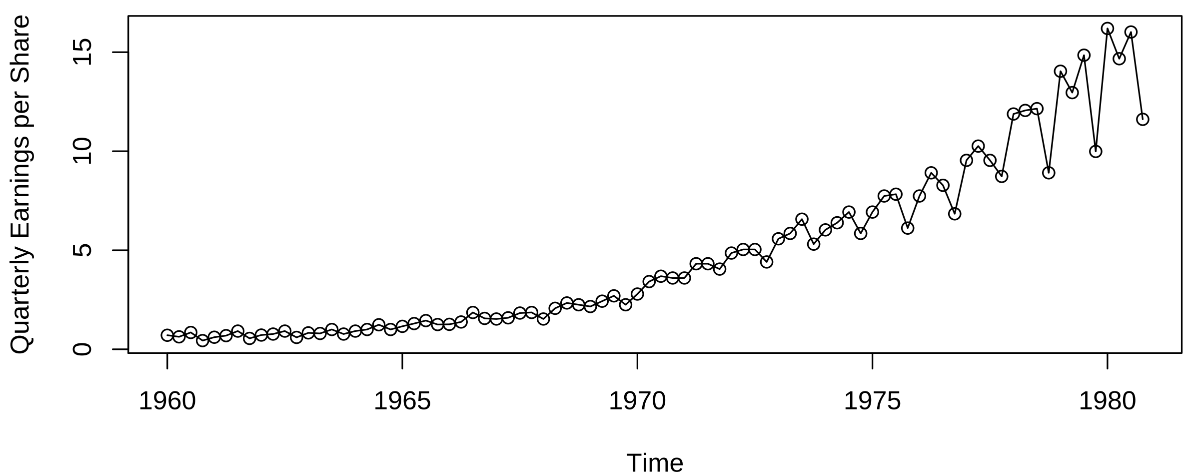
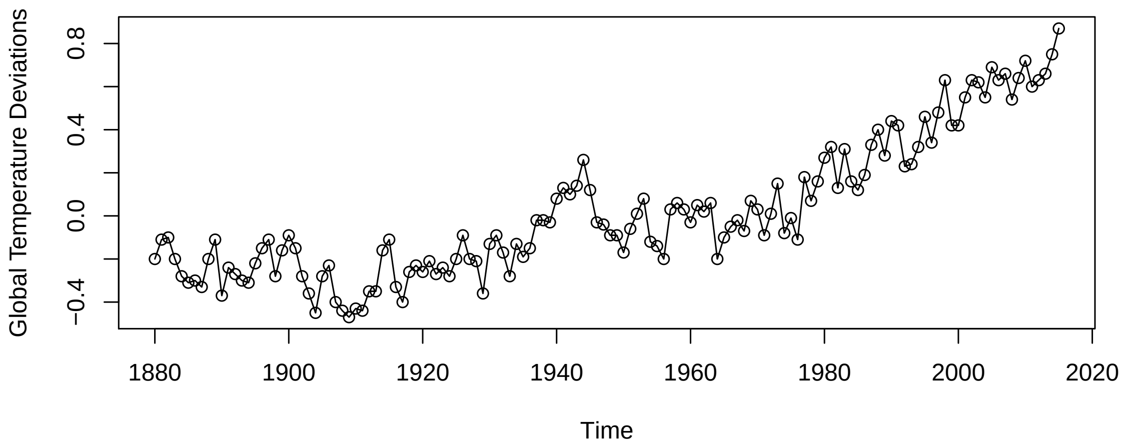


Figure shows quarterly earnings per share for the U.S. company Johnson & Johnson, furnished by Professor Paul Griffin (personal communication) of the Graduate School of Management, University of California, Davis. There are 84 quarters (21 years) measured from the first quarter of 1960 to the last quarter of 1980. Modeling such series begins by observing the primary patterns in the time history.

- Global Warming

```
library(astsa)
plot(globtemp, type="o", ylab="Global Temperature Deviations")
```



Consider the global temperature series record shown in Figure 1.2. The data are the global mean land–ocean temperature index from 1880 to 2015, with the base period 1951-1980. In particular, the data are deviations, measured in degrees centigrade, from the 1951-1980 average, and are an update of Hansen et al. (2006). We note an apparent upward trend in the series during the latter part of the twentieth century that has been used as an argument for the global warming hypothesis. Note also the leveling off at about 1935 and then another rather sharp upward trend at about 1970. The question of interest for global warming proponents and opponents is whether the overall trend is natural or whether it is caused by some human-induced interface.

- Speech Data

```r
library(astsa)
plot(speech, ylab=" speech ")
```

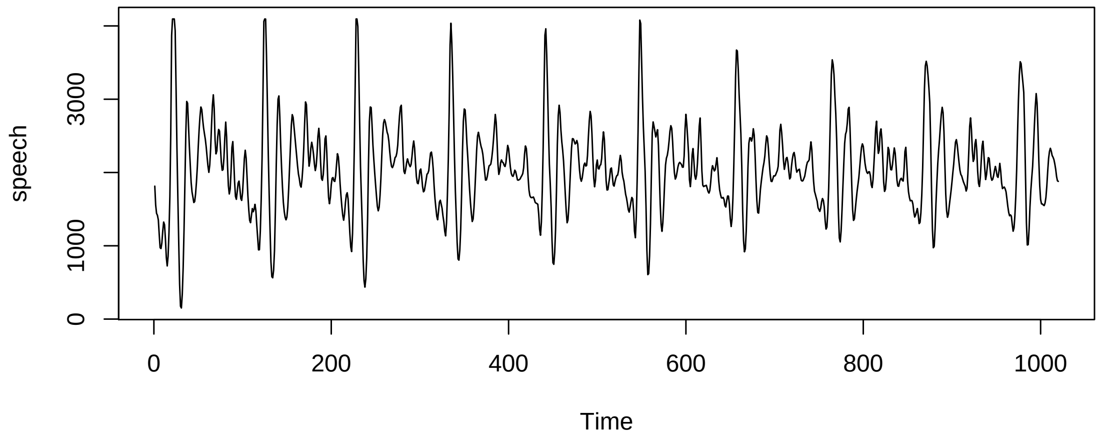
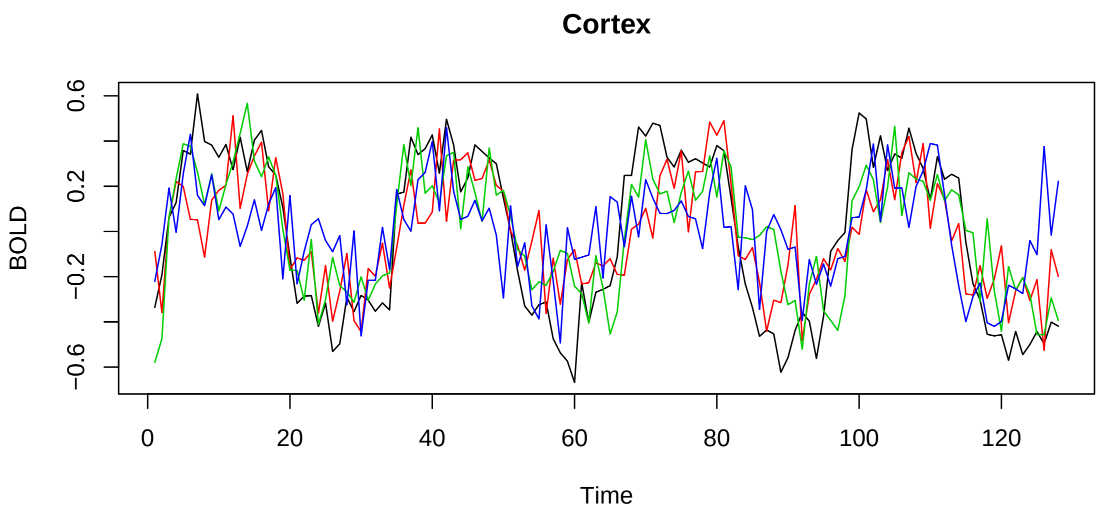
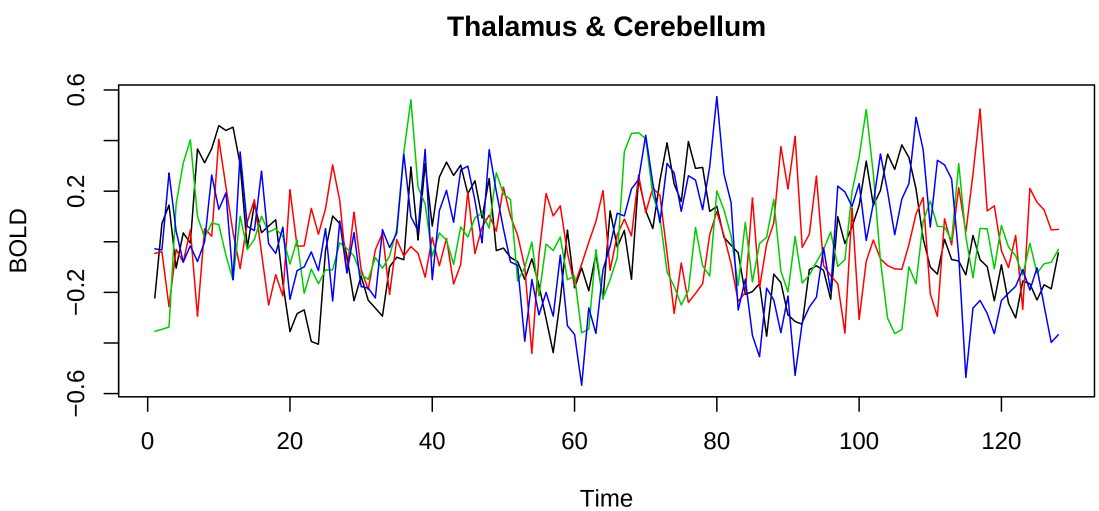


Figure shows a small .1 second (1000 point) sample of recorded speech for the phrase aaa $\cdots$ hhh, and we note the repetitive nature of the signal and the rather regular periodicities. One current problem of great interest is computer recognition of speech, which would require converting this particular signal into the recorded phrase aaa $\cdots$ hhh. Spectral analysis can be used in this context to produce a signature of this phrase that can be compared with signatures of various library syllables to look for a match. One can immediately notice the rather regular repetition of small wavelets. The separation between the packets is known as the pitch period and represents the response of the vocal tract filter to a periodic sequence of pulses stimulated by the opening and closing of the glottis.

- fMRI Imaging

```
library(astsa)
par(mfrow=c(2,1))
ts.plot(fmri1[,2:5], col=1:4, ylab="BOLD", main="Cortex")
ts.plot(fmri1[,6:9], col=1:4, ylab="BOLD", main="Thalamus & Cerebellum")
```

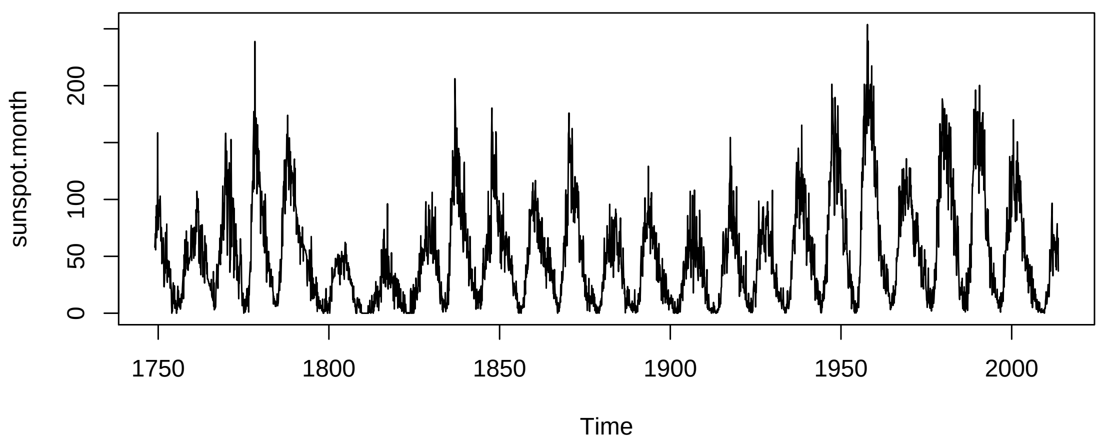




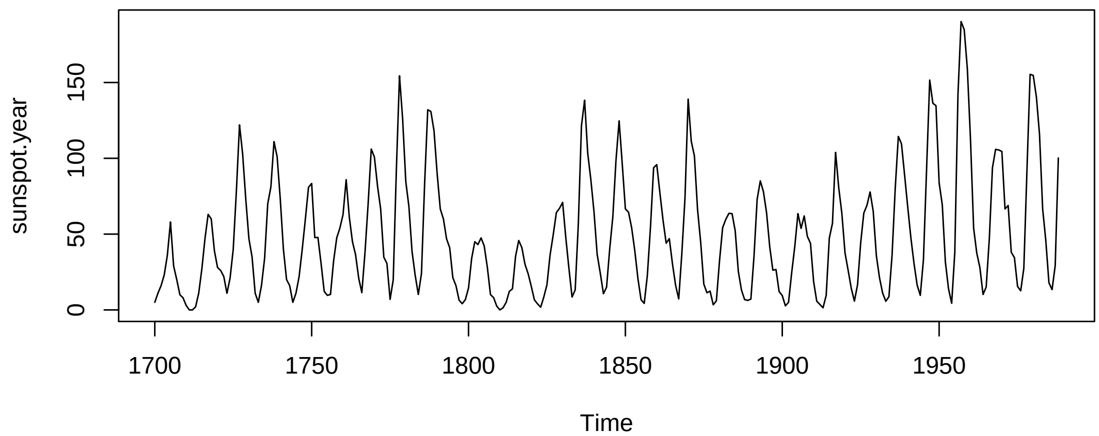
A fundamental problem in classical statistics occurs when we are given a collection of independent series or vectors of series, generated under varying experimental conditions or treatment configurations. Such a set of series is shown in Figure 1.6, where we observe data collected from various locations in the brain via functional magnetic resonance imaging (fMRI). In this example, five subjects were given pe- riodic brushing on the hand. The stimulus was applied for 32 seconds and then stopped for 32 seconds; thus, the signal period is 64 seconds. The sampling rate was one observation every 2 seconds for 256 seconds (n = 128). For this example, we averaged the results over subjects (these were evoked responses, and all subjects were in phase). The series shown in Figure 1.6 are consecutive measures of blood oxygenation-level dependent (bold) signal intensity, which measures areas of activation in the brain. Notice that the periodicities appear strongly in the motor cortex series and less strongly in the thalamus and cerebellum. The fact that one has series from different areas of the brain suggests testing whether the areas are responding differently to the brush stimulus.

- Sun Spot Data

```
library(astsa)
plot(sunspot.month)
```

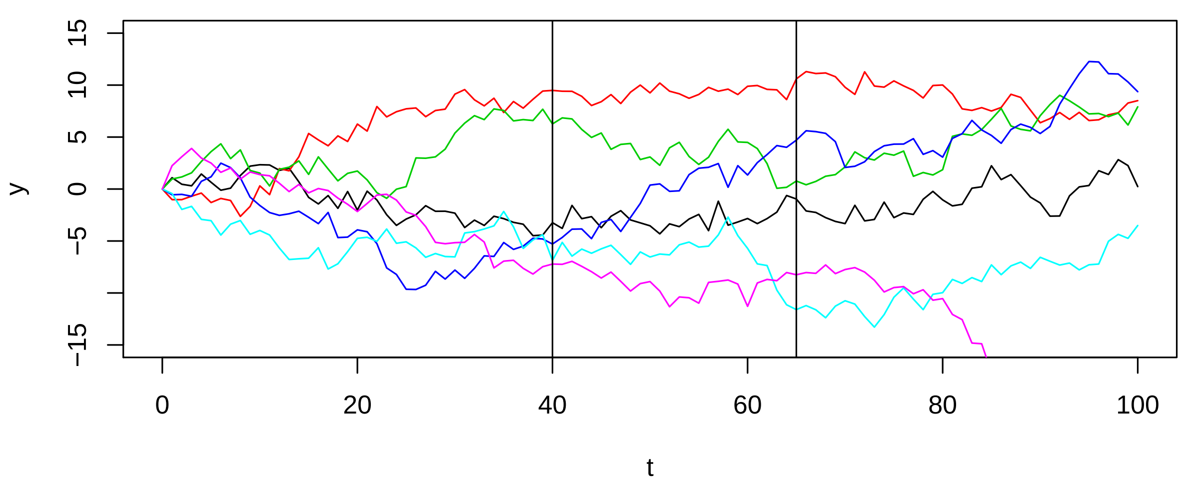



```
plot(sunspot.year)
```



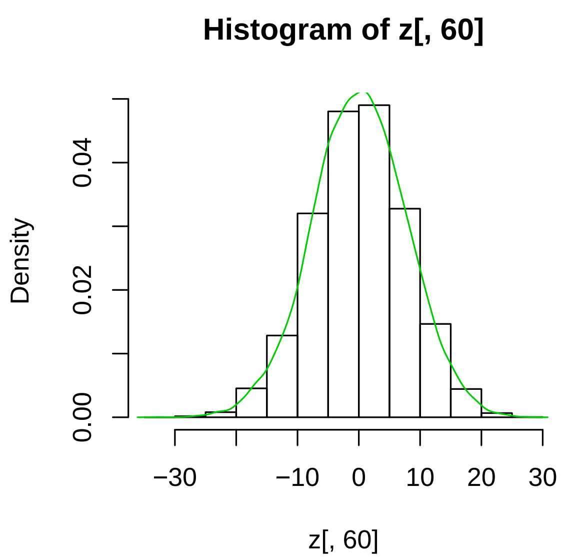
Monthly mean relative sunspot numbers from 1749 to 1983. Collected at Swiss Federal Observatory, Zurich until 1960, then Tokyo Astronomical Observatory

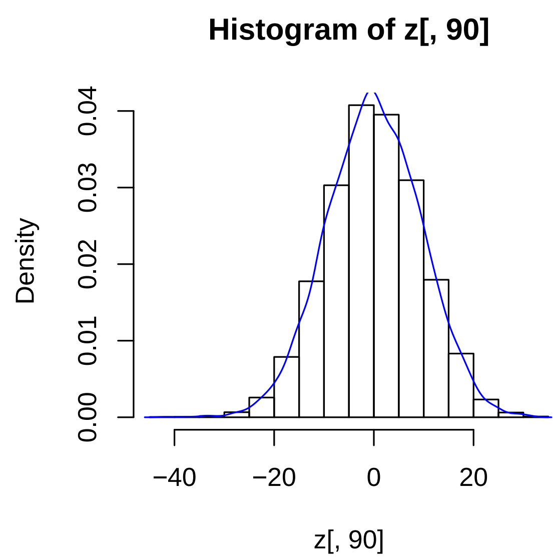
```
# Distribution of time series
n<-100
t<-0:n
x<-rnorm(n,0,1)
x<-c(0,x)
y<-cumsum(x)
plot(y~t, type='l', ylim=c(-15,15))

for (i in 1 : 5){
  x<-rnorm(n,0,1)
  x<-c(0,x)
  y<-cumsum(x)
  lines(y~t, col=(i+1) )
}
abline(v=40)
abline(v=65)
```




```
itrn<-10000
z<-array(0, dim=c(itrn,(n+1)))
#plot(t*0~t, col=2 ,ylim=c(-15,15) ,type='l', lwd=2)
for (i in 1 : itrn){
  x<-rnorm(n,0,1)
```

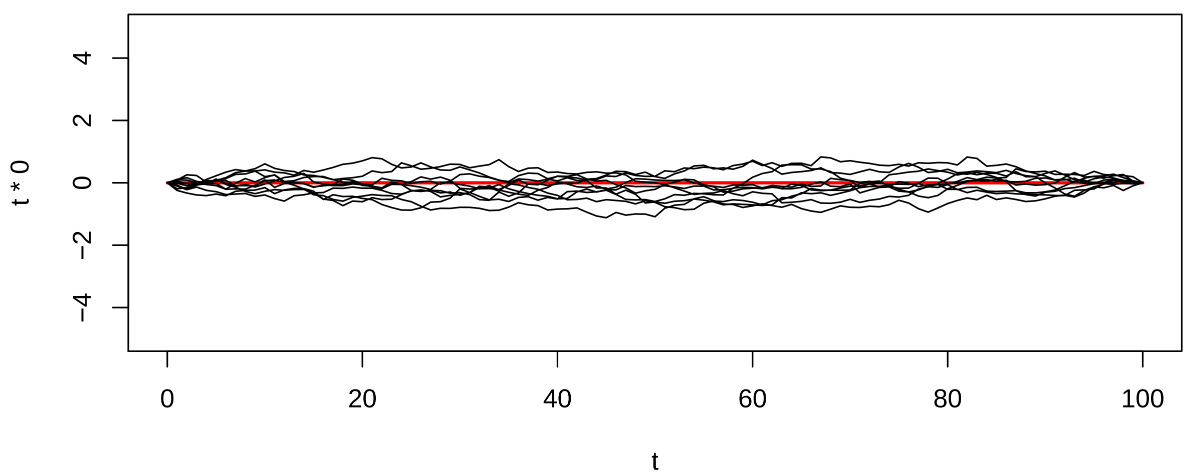
```
  z[i,2:(n+1)]<-cumsum(x)
 # lines(z[i,]~t )
}
par(mfrow=c(1,2))
hist(z[,60], probability = T)
lines(density((z[,60])), col=3)
hist(z[,90],probability = T)
lines(density((z[,90])), col=4)
```






```
print(var(z[,60]))
## [1] 59.32856
print(var(z[,90]))
## [1] 89.93363
print(cov(z[,60], z[,90]))
## [1] 59.43091
itrn<-10000
z1<-array(0, dim=c(itrn,(n+1)))
#par(mfrow=c(1,1))
#plot(t*0~t, col=2 ,ylim=c(-5,5) ,type='l', lwd=2)
for (i in 1 : itrn){
  x<-rnorm(n,0,1)
```

```
  z1[i,2:(n+1)]<-cumsum(x)/sqrt(n)
  #lines(z1[i,]~t )
}

print(var(z1[,60]))
## [1] 0.5914786
print(var(z1[,90]))
## [1] 0.8955987
print(cov(z1[,60], z1[,90]))
## [1] 0.5929031
itrn<-10
z2<-array(0, dim=c(itrn,(n+1)))
par(mfrow=c(1,1))
plot(t*0~t, col=2 ,ylim=c(-5,5) ,type='l', lwd=2)
for (i in 1 : itrn){
  x<-rnorm(n,0,1)
  z2[i,2:(n+1)]<-cumsum(x-mean(x))/sqrt(n)
   lines(z2[i,]~t )
}
```



```
print(var(z2[,60]))

## [1] 0.220481

print(var(z2[,90]))

## [1] 0.115939

print(cov(z2[,60], z2[,90]))

## [1] 0.09186109
```

4. Examples

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

(2) **Binary Process:** Consider the sequence of iid random variables $\{X_t\}$ with

$$P(X_t = 1) = p = 1 - P(X_t = -1)$$

(3) **Random Walk :** Let $\{W_t\}$ be a sequence of i.i.d. random variables with $E(W_t) = 0$ and $Var(W_t) = \sigma_w^2$. Define another sequence $X_0 = 0$ and $X_t = \sum_{i=1}^{t} W_i$. Then X_t is called a random walk.

(4) **Random Walk with Drift:** Let $\{W_t\}$ be a sequence of i.i.d. random variables with $E(W_t) = 0$ and $Var(W_t) = \sigma_w^2$. Define another sequence $X_0 = 0$ and $X_t = \delta t + \sum_{i=1}^{t} W_i$.

(5) **Signal in Noise:** Many realistic models for generating time series assume an underlying signal with some consistent periodic variation, contaminated by adding a random noise.

$$X_t = A \sin(2\pi f t + \phi) + W_t$$

(6) **Moving Average process:** Let $W_t \sim WN(0, \sigma_w^2)$ then $X_t = W_t + \theta W_{t-1}$ for is called a moving average process of order one.

(7) **Auto Regressive process of order one :** Let $W_t \sim WN(0, \sigma_w^2)$ then $X_t = \phi X_{t-1} + W_t$ for $|\phi| < 1$ is called an autoregressive process of order one.

(8) **Wiener process** : A process X_t is called Wiener process (Brownian motion) if it satisfies the following conditions
   (a) $X_0 = a$(for standard Brownian motion $a = 0$)
   (b) $X_{t+u} - X_t$ is independent of ant X_s for all $s \leq t$ and $u > 0$
   (c) $X_{t+u} - X_t$ is follows $N(0, u)$ for all $u > 0$
   (d) X_t is continuous over t

   Remark. Let $\{W_i\}$ be a sequence of i.i.d. random variables with $E(W_i) = 0$ and $Var(W_i) = 1$ then $Y_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} W_k$ converges to Wiener process (Brownian motion) X_t on $[0, 1]$ for large n.

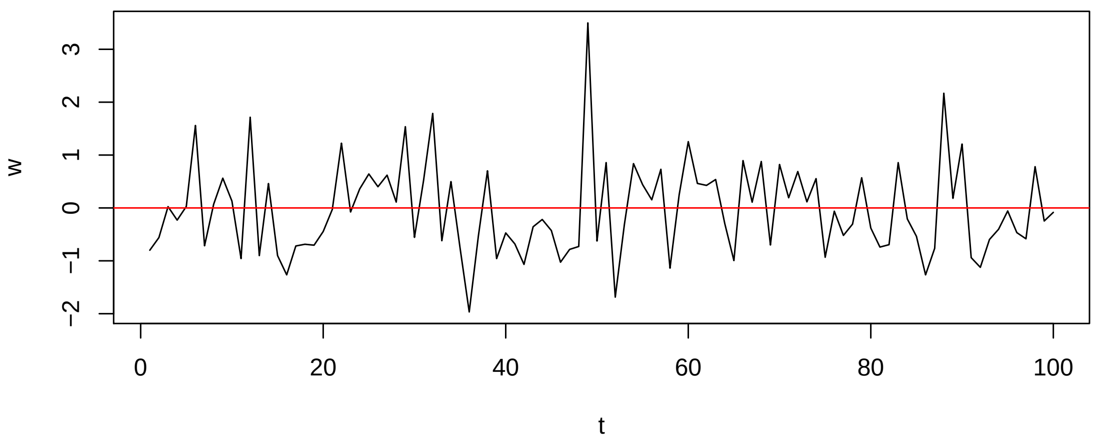
(9) **Brownian Bridge :** A Brownian bridge $B(t)$ is a conditional brownian motion X_t on $[0, T]$ such that $X_t = 0$ at $t = 0$ and $X_t = b$ at $t = T$. For a standard bridge $B_0(t)$, $T = 1$ and $b = 0$. The standard brownian bridge can be represented as

$$B_0(t) = X_t - tX_1 \;\forall t \in [0, 1]$$

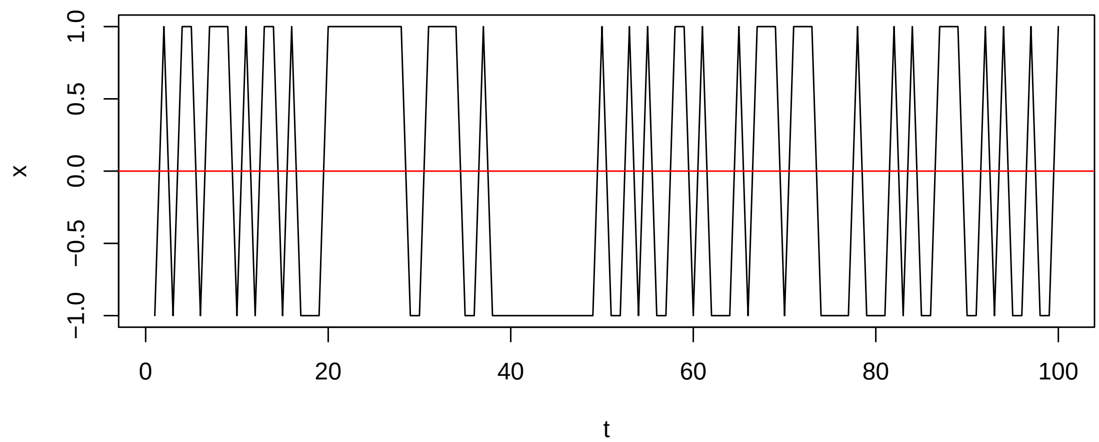
(10) **Process with trend and seasonality:**

$$X_t = \delta_0 + \delta_1 t + \delta_{2t} t^2 + A \sin(2\pi f t + \phi) + W_t$$

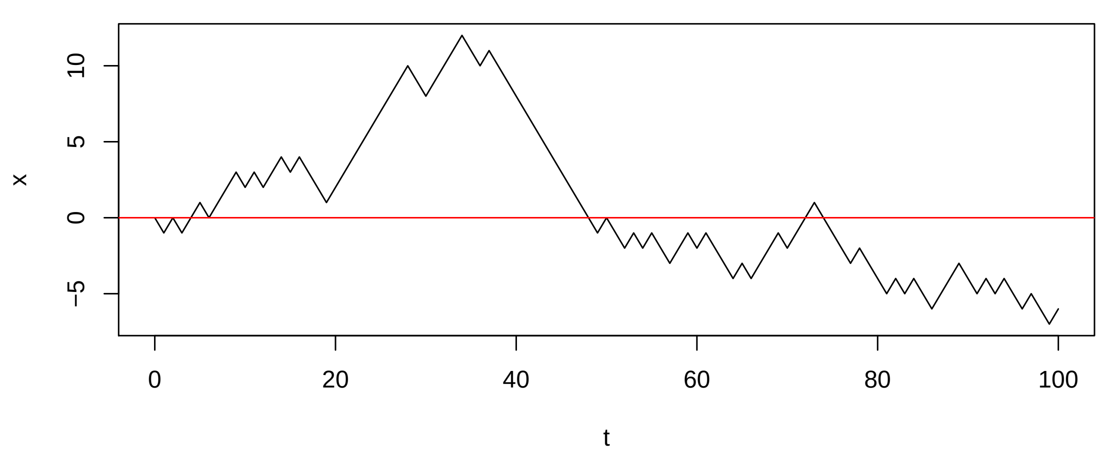
```
## White Noise
set.seed(123);
n<-100;
t<-1:n
w<-array(0,dim = c(n))
e<-seq(from = 2,to = n, by=2) ; o<-(e-1);
w[e]<-rnorm((n/2),0,1) ;
w[o]<-rexp((n/2),1)-1;
plot(w~t, type = "l")
abline(a=0,b = 0, col=2)
```



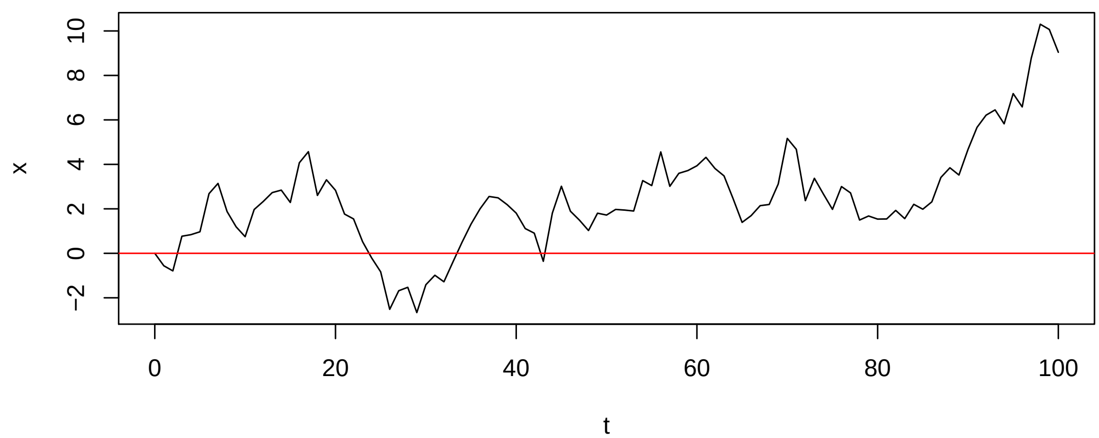

```
# Binary Process:
set.seed(123);
n<-100;
t<-1:n
w<-rbinom(n,size = 1,prob = 0.5)
x<-(w-0.5)/0.5
plot(x~t ,type="l")
abline(a=0,b = 0, col=2)
```



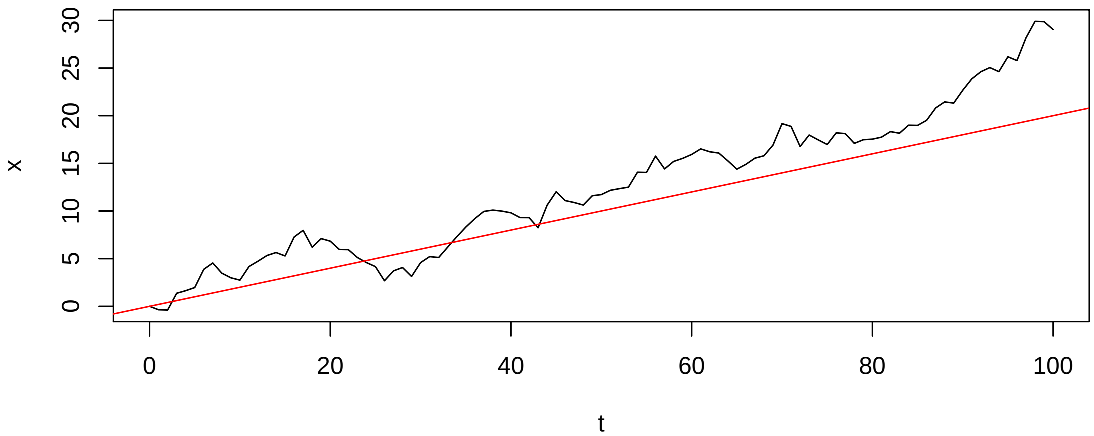

```
# Random Walk
set.seed(123);
n<-100;
t<-0:n
w<-rbinom(n,size = 1,prob = 0.5)
z<-(w-0.5)/0.5
x<-c(0,cumsum(z))
plot(x~t ,type="l")
abline(a=0,b = 0, col=2)
```



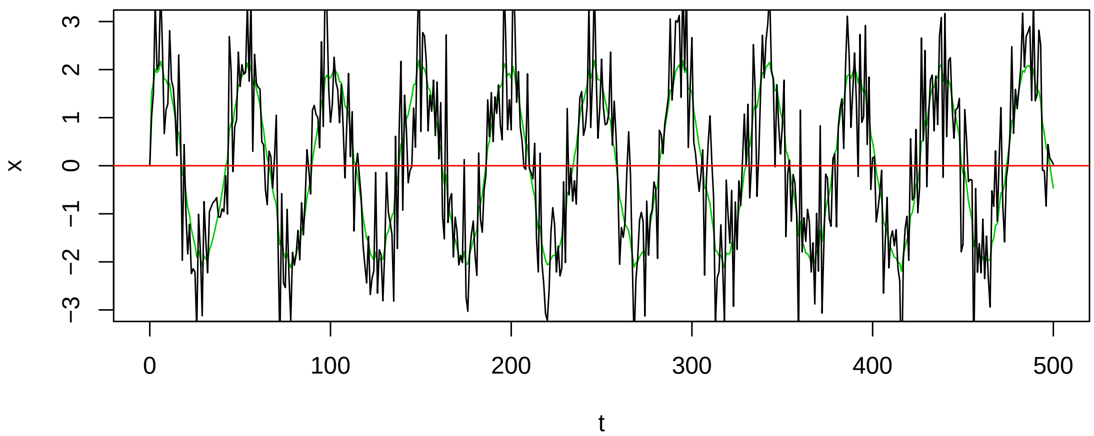
```
set.seed(123);
n<-100;
t<-0:n
w<- rnorm(n,0,1)
x<-c(0,cumsum(w))
plot(x~t ,type="l")
abline(a=0,b = 0, col=2)
```



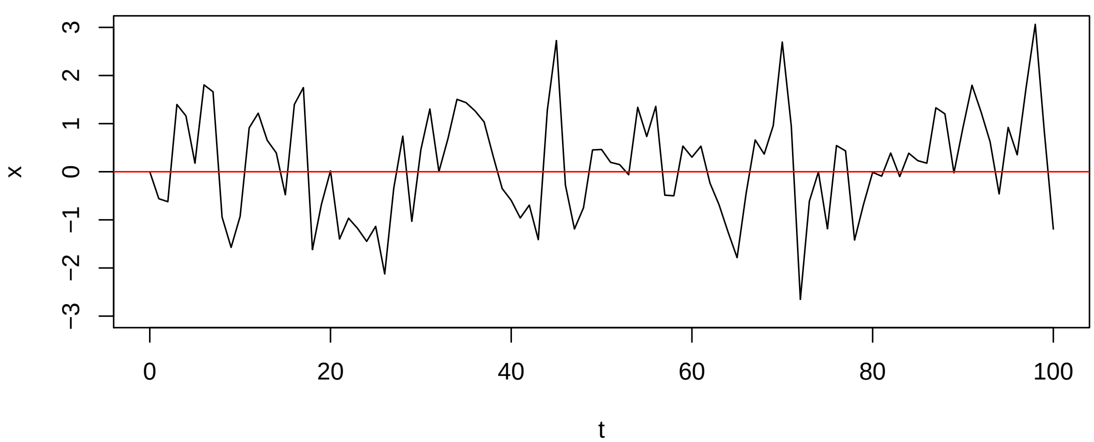

```
# Random Walk with drift
set.seed(123);
n<-100;
t<-0:n
w<- rnorm(n,0.2,1)
x<-c(0,cumsum(w))
plot(x~t ,type="l")
abline(a=0,b = 0.2, col=2)
```



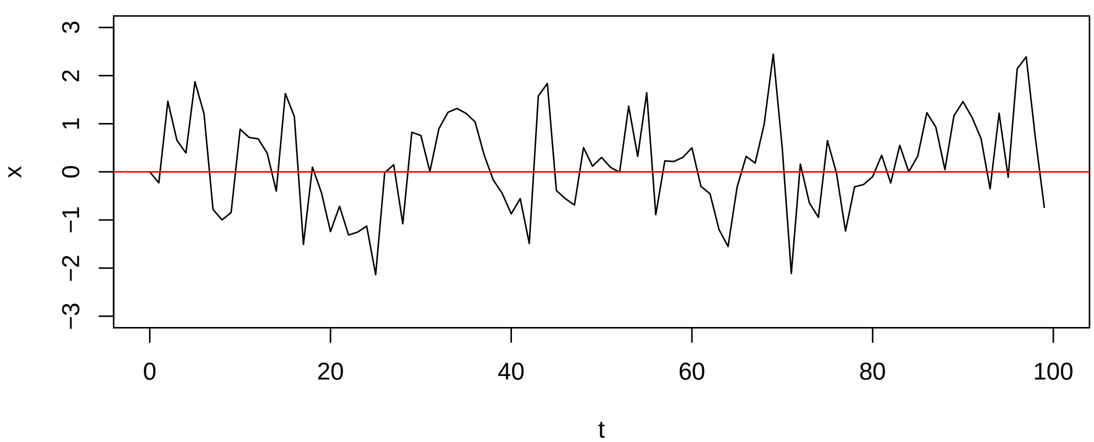

```
# Signal in Noise:
set.seed(123);
n<-500;
t<-0:n
w<- rnorm(n,0,1)
x<-c(0,2*sin(2*pi*(1/48)*t[-1]+pi/4)+w*0.1)
plot(x~t ,type="l", col=3, ylim=c(-3,3))
x<-c(0,2*sin(2*pi*(1/48)*t[-1]+pi/4)+w)
lines(x~t ,type="l")
abline(a=0,b = 0, col=2)
```




```
# Moving Average process:
set.seed(123);
n<-100;
t<-0:n
w<- rnorm(n,0,1)
x<-c(0,w)+0.7*c(0,0,w[-n])
plot(x~t ,type="l", ylim=c(-3,3))
abline(a=0,b = 0, col=2)
```

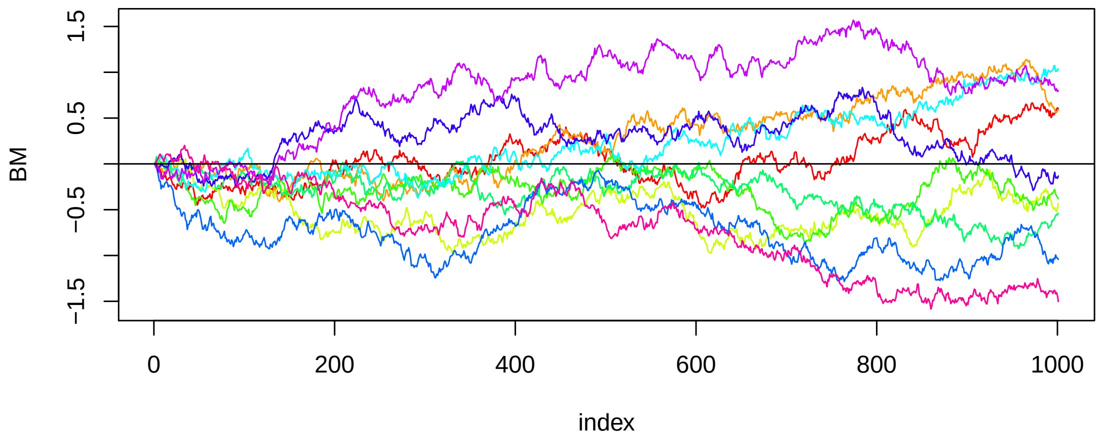


```
# Autoregressive process
set.seed(123);
n<-100;
t<-0:n
w<- rnorm(n,0,1)
x<-numeric(0)
x[1]<-0
for(i in 2 :(n+1))
  x[i]<-0.4*x[i-1]+w[i]
plot(x~t ,type="l", ylim=c(-3,3))
abline(a=0,b = 0, col=2)
```

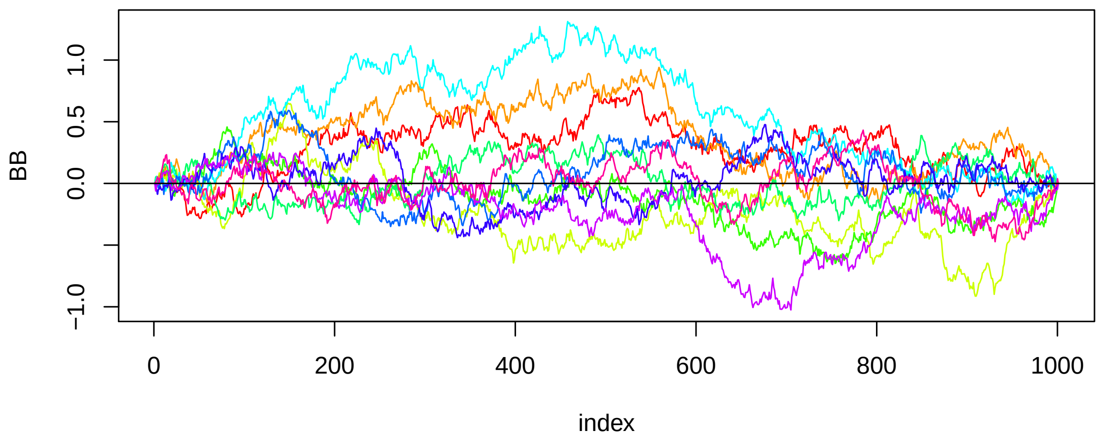



```
# Wiener process ( Brownian motion)
n<-1000;
t<-0:n
itrn=10
BM = replicate(itrn, {
  w<- rnorm(n,0,1)
  x<-c(0,cumsum(w)/sqrt(n))
})

cols = rainbow(itrn)
matplot(BM, type = "l", col = cols, lty = 1,xlab = "index")
```

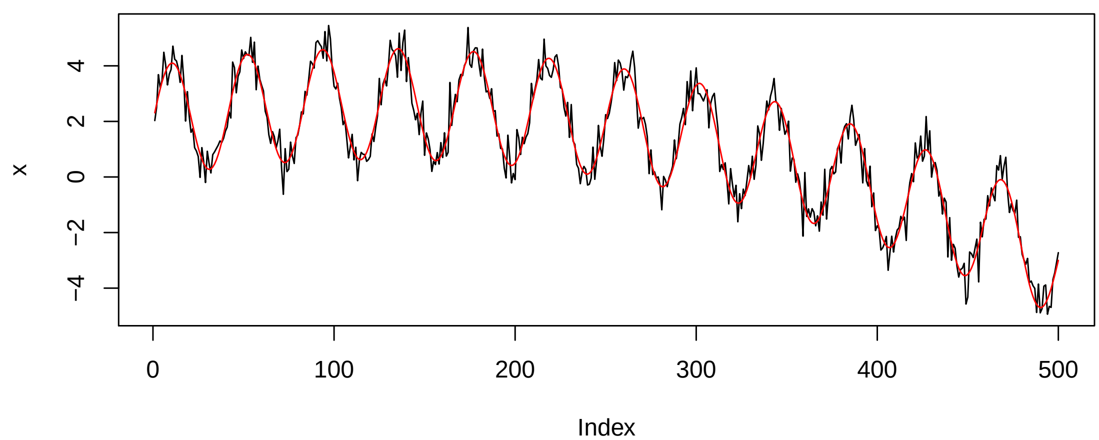
```
abline(a=0,b = 0, col=1)
```




```
# Brownian Bridge


n<-1000;
t<-0:n
itrn=10
BB = replicate(itrn, {
  w<- rnorm(n,0,1)
  bm<-c(0,cumsum(w)/sqrt(n))
  bb<-bm-(t/n)*bm[n+1]
})

cols = rainbow(itrn)
matplot(BB, type = "l", col = cols, lty = 1, xlab = "index")
abline(a=0,b = 0, col=1)
```




```
# Process with trend and seasonality:
set.seed(123);
n<-500;
t<-(1:n)/n
w<- rnorm(n,0,.5)
x<-2+5*t-10*t^2+ (2*sin(2*pi*(12)*t))+w
plot(x ,type="l")
lines(2+5*t-10*t^2+ (2*sin(2*pi*(12)*t)), col=2)
```



5. Operators

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

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

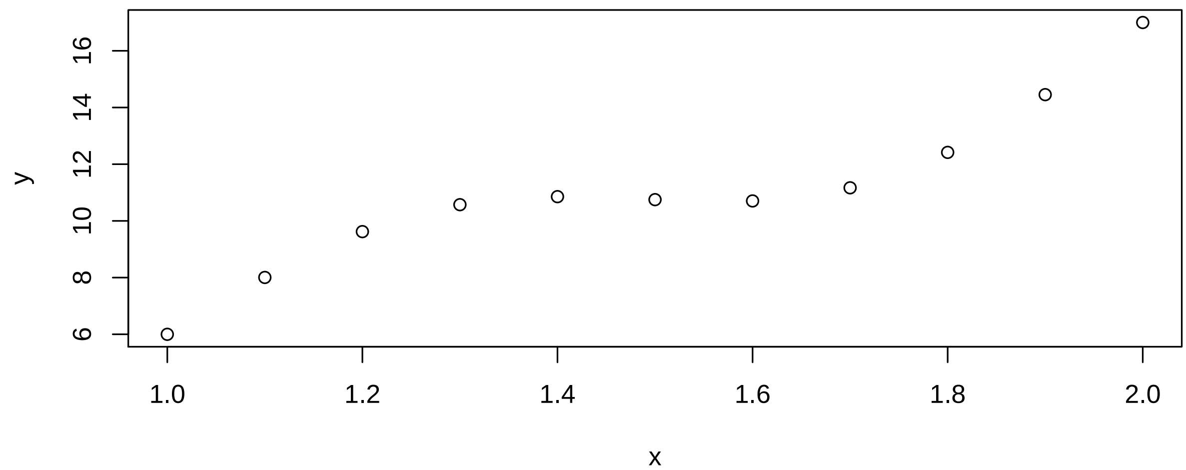
  which implies

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

- Seasonal difference

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

```
# Difference table


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)
}
print(round(dtable, 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
# 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)
```




```r
x<-seq(1,2, by=0.1)
y<-1+2*x+3*x^2+ 2*sin(2*pi*x)
l<-length(x)
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)
}
print(round(dtable, 2))
## 	    [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8] [,9] [,10] [,11]
##  [1,]  1.0  6.00  2.01 -0.39 -0.28  0.28  0.00 -0.11 0.04  0.03 -0.03
##  [2,]  1.1  8.01  1.62 -0.67  0.00  0.28 -0.11 -0.07 0.07  0.00 -0.03
##  [3,]  1.2  9.62  0.95 -0.67  0.28  0.17 -0.17  0.00 0.07 -0.03  0.00
##  [4,]  1.3 10.57  0.28 -0.39  0.45  0.00 -0.17  0.07 0.04  0.00  0.00
##  [5,]  1.4 10.86 -0.11  0.06  0.45 -0.17 -0.11  0.11 0.00  0.00  0.00
##  [6,]  1.5 10.75 -0.05  0.51  0.28 -0.28  0.00  0.00 0.00  0.00  0.00
##  [7,]  1.6 10.70  0.46  0.79  0.00 -0.28  0.00  0.00 0.00  0.00  0.00
##  [8,]  1.7 11.17  1.25  0.79 -0.28  0.00  0.00  0.00 0.00  0.00  0.00
##  [9,]  1.8 12.42  2.04  0.51  0.00  0.00  0.00  0.00 0.00  0.00  0.00
## [10,]  1.9 14.45  2.55  0.00  0.00  0.00  0.00  0.00 0.00  0.00  0.00
```

```
## [11,]  2.0 17.00  0.00  0.00  0.00  0.00  0.00  0.00 0.00  0.00  0.00
##       [,12] [,13]
##  [1,]     0     0
##  [2,]     0     0
##  [3,]     0     0
##  [4,]     0     0
##  [5,]     0     0
##  [6,]     0     0
##  [7,]     0     0
##  [8,]     0     0
##  [9,]     0     0
## [10,]     0     0
## [11,]     0     0
```

6. Linear process

Definition. An class of weakly stationary time series known as linear process defined by

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

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

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

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

$$E(Y_t) = \mu$$

and

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

- Suppose that X_t is at least weakly stationary time series with

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

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.

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

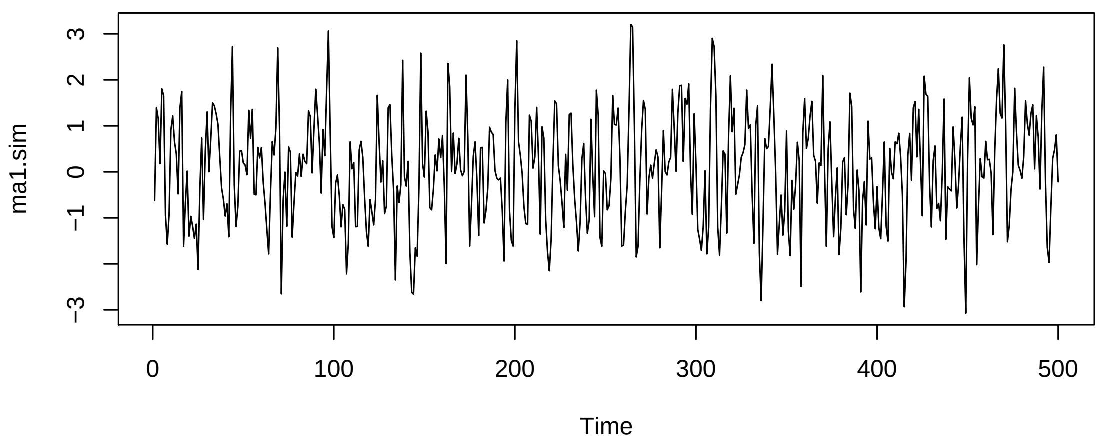
$$\lim_{n \to \infty} E(Y_n - Z)^2 = 0$$

- 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.

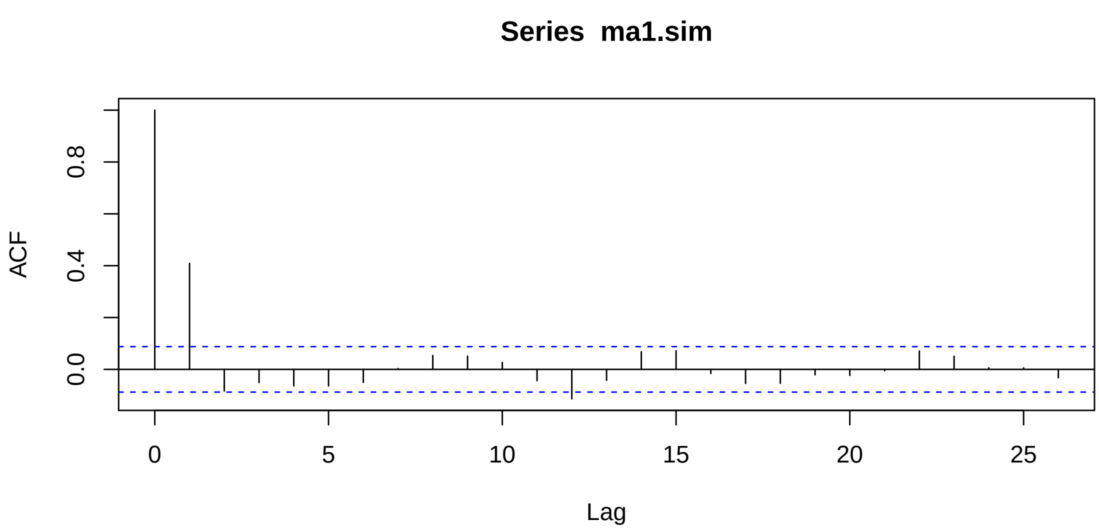
$$\lim_{k \to \infty} E\left(X_t - \sum_{j=0}^{k} \phi^j W_{t-j}\right)^2 = 0$$

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

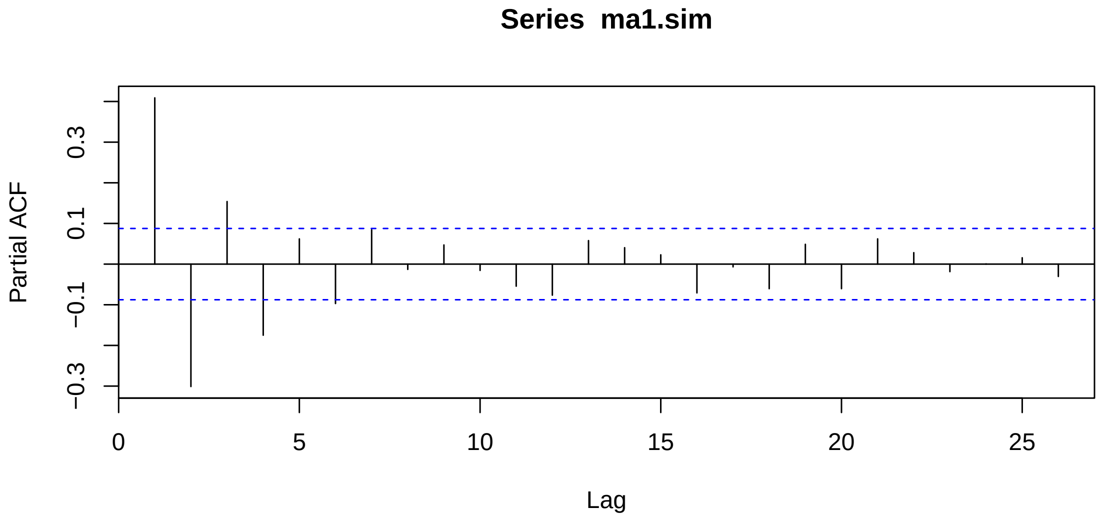
```
## MA(1) Example
set.seed(123);
n<-500; p<-0; d<-0;q<-1;
ma1.sim<-arima.sim(list(order=c(p,d,q), ma=0.7), n)
ts.plot(ma1.sim)
```



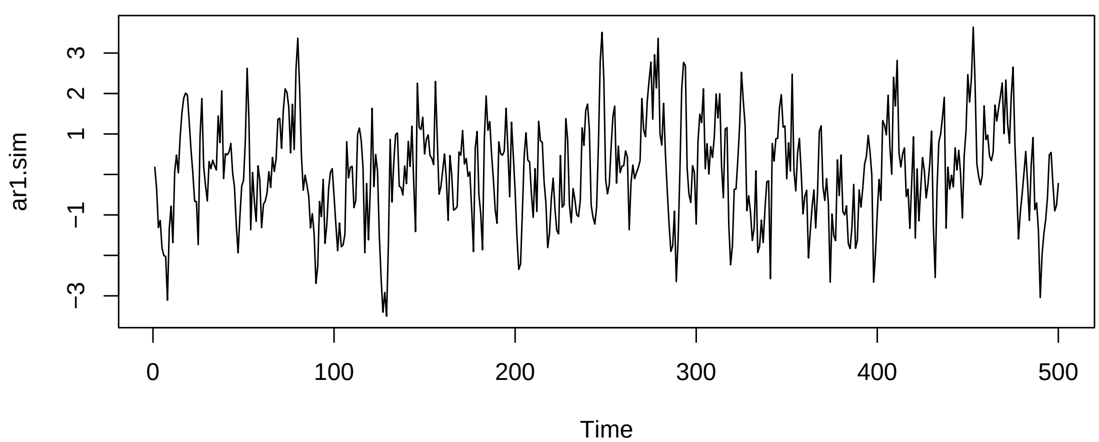

```
acf(ma1.sim,type = "correlation",plot = T)
```



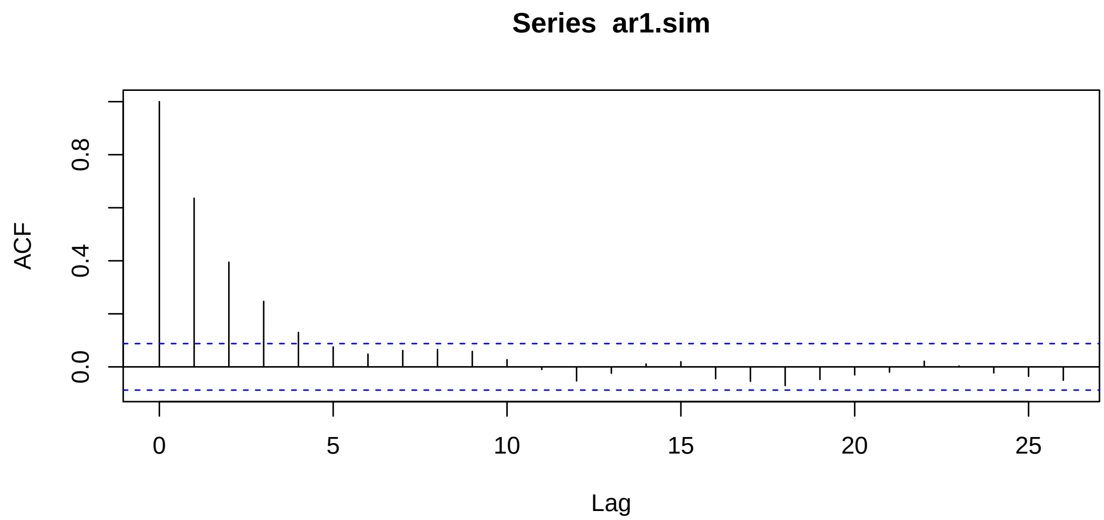
```
pacf(ma1.sim,plot = T)
```



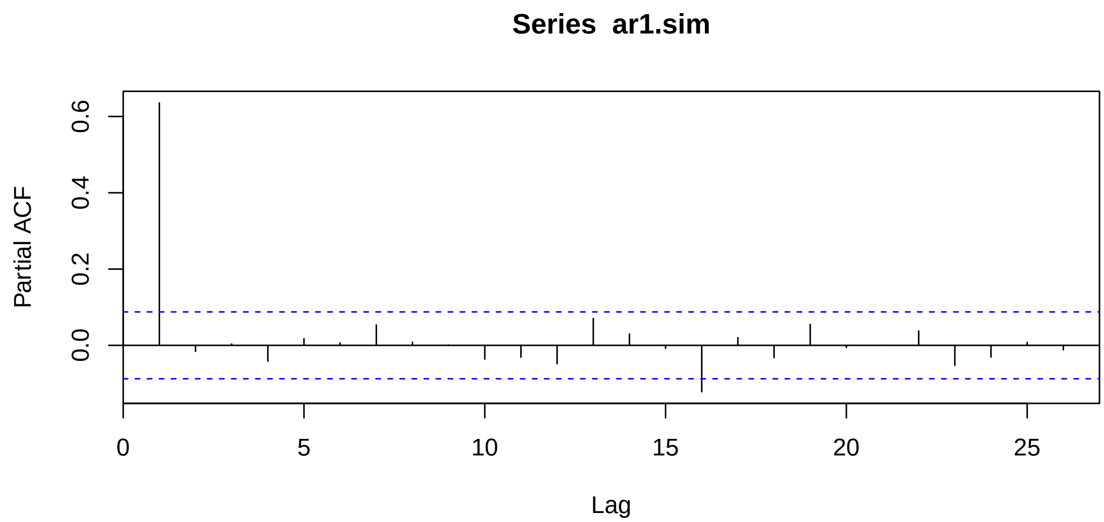

```
## AR(1) Example
set.seed(123);
n<-500;
p<-1; d<-0;q<-0;
ar1.sim<-arima.sim(list(order=c(p,d,q), ar=0.7), n)
ts.plot(ar1.sim)
```



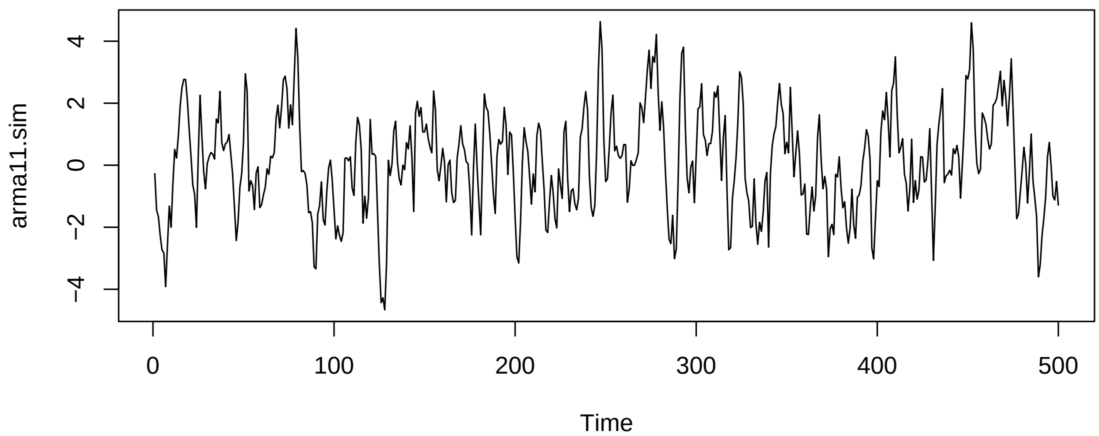
```
acf(ar1.sim,type = "correlation",plot = T)
```



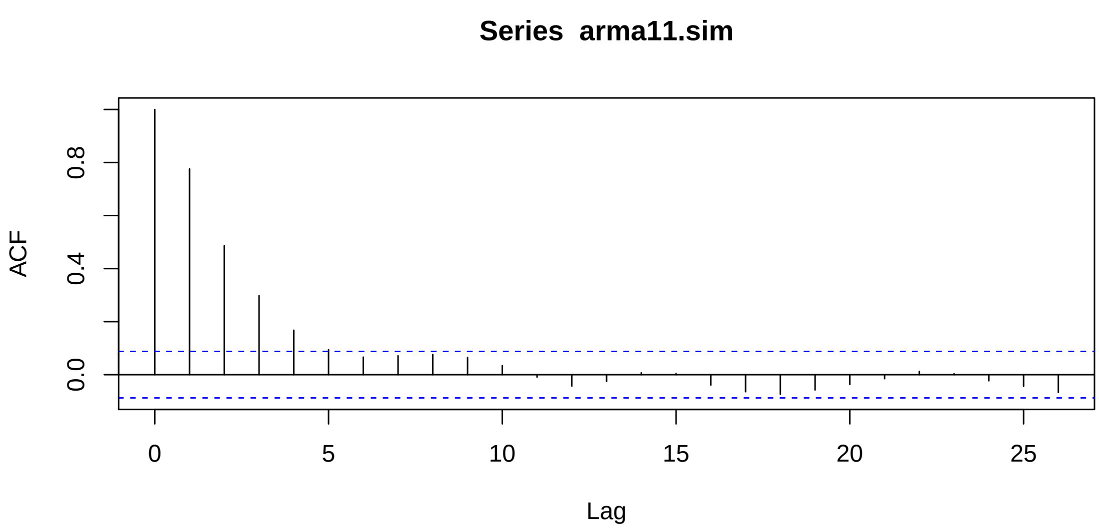

```
pacf(ar1.sim,plot = T)
```



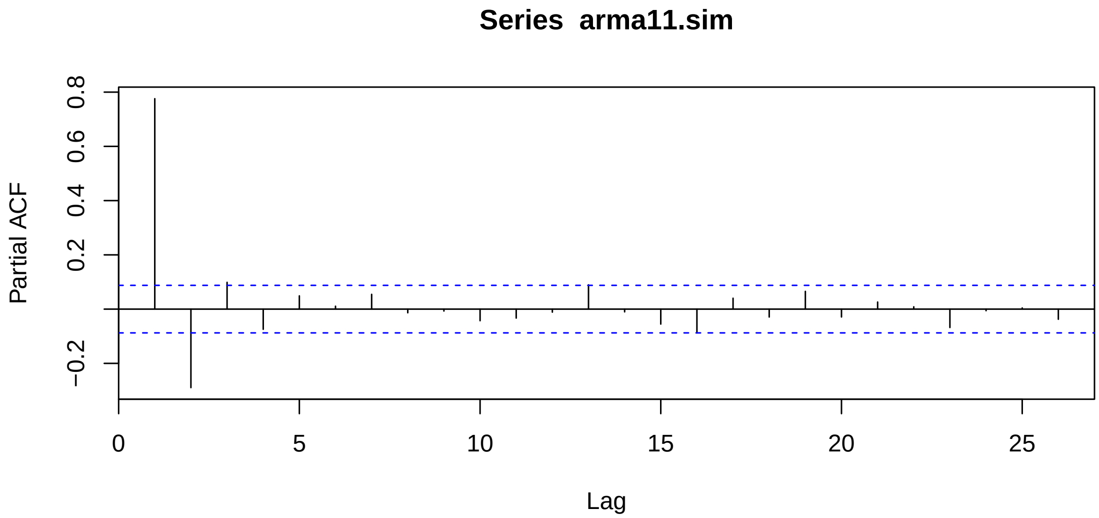

```
## ARMA(1,1) Example
set.seed(123);
n<-500;
p<-1; d<-0;q<-1;
arma11.sim<-arima.sim(list(order=c(p,d,q), ar=0.7, ma=0.4), n)
ts.plot(arma11.sim)
```



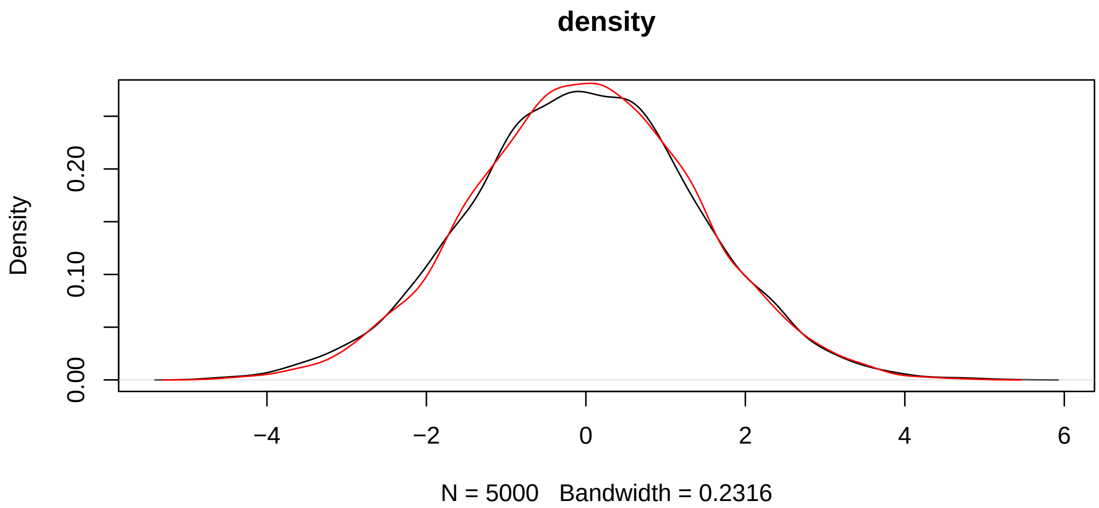

```r
acf(arma11.sim,type = "correlation",plot = T)
```



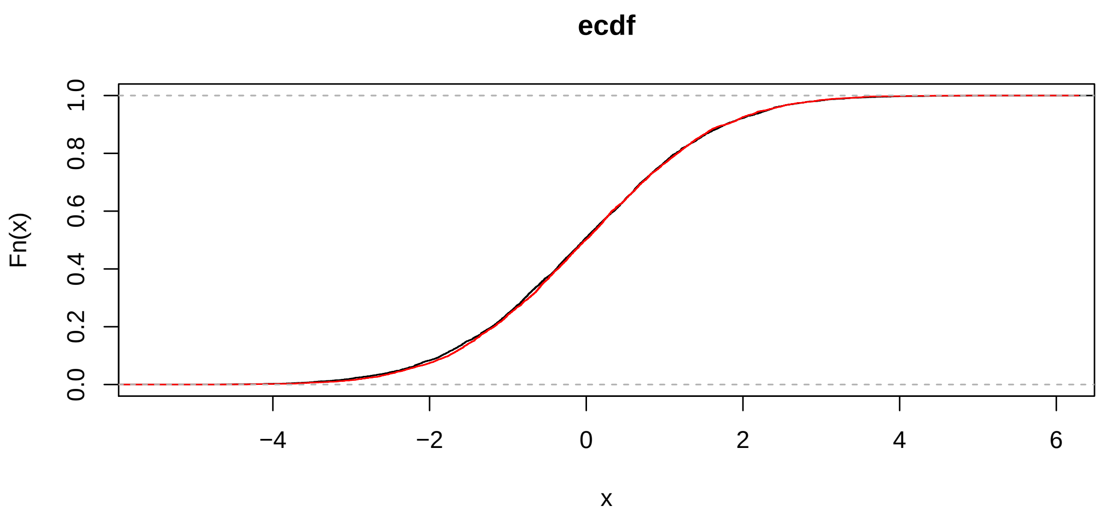

```r
pacf(arma11.sim,plot = T)
```




```r
## AR(1) and £MA(\infty)£ Example
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")
```




```r
plot(ecdf(arsim), main="ecdf");
lines(ecdf(masim),col="red")
```



- **Causality:** 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:** A linear process X_t is invertible function ofW_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 invertible iff $|\theta| < 1$ i.e. $1 + \theta z$ has a solution out of the unit circle on complex plane $\mathbb{C}$.
- In general ARMA (p, q) process $\Phi_p(B)X_t = \Theta_q(B)W_t$ causal if $\Phi_p(z) = 0$ has roots out of the unit circle and is invertible if of is invertible $\Theta_q(z) = 0$ has roots out of the unit circle in complex plane $\mathbb{C}$. Unique stationary solution exists iff $\Phi_p(z) = 0$ has no solution on $|z| = 1$.

7. Prediction

Estimation of Trend in the Absence of Seasonality:

- 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 $[t-q, t+q]$

- Linear filter :

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

- 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 square 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

Estimation of Trend and Seasonality.

- 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.

Linear Forecasting : Durbin-Levinson algorithm.

- 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

- $E(\hat{X}^n_{m+n} - X_{m+n}) = 0$ [Unbiased presdiction]
- $E[(\hat{X}^n_{m+n} - X_{m+n})X_i] = 0$ [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: $E(X_{n+1} - \hat{\boldsymbol{\alpha}}^T\mathbf{X})^2 = \gamma(0) - \boldsymbol{\gamma}_n^T(1)\Gamma_n^{-1}\boldsymbol{\gamma}_n(1)$
- 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.

$$\hat{\Gamma}_n \hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\gamma}}_n(1)$$

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

Integrated ARMA Models: ARIMA(p,d,q).

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

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

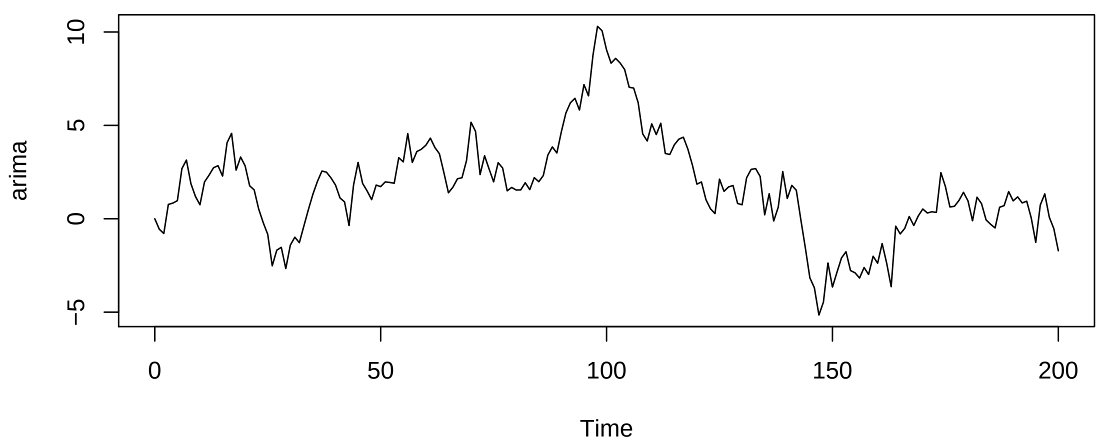
- [ARIMA(0,1,0)]: Random walk with drift

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

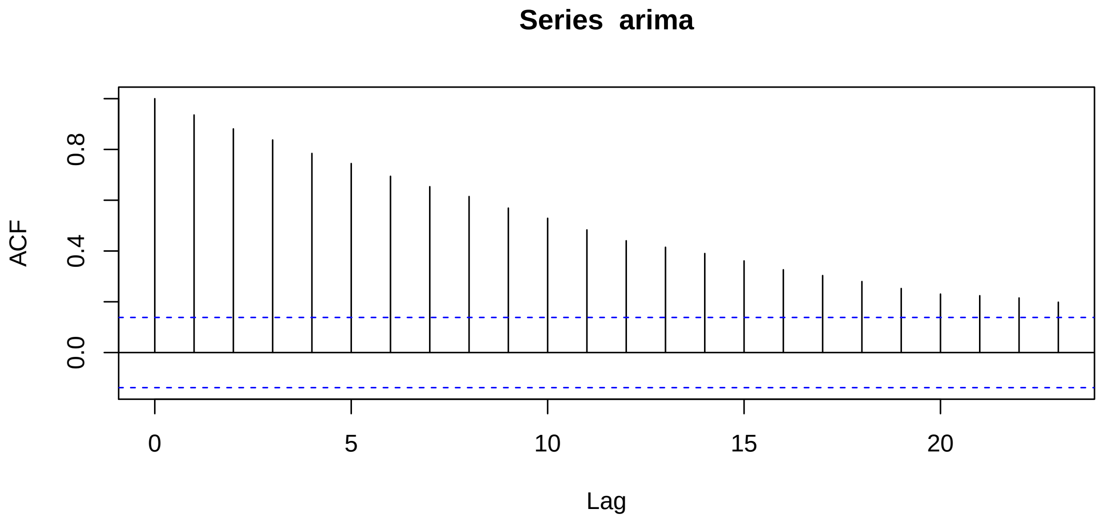
  This implies

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

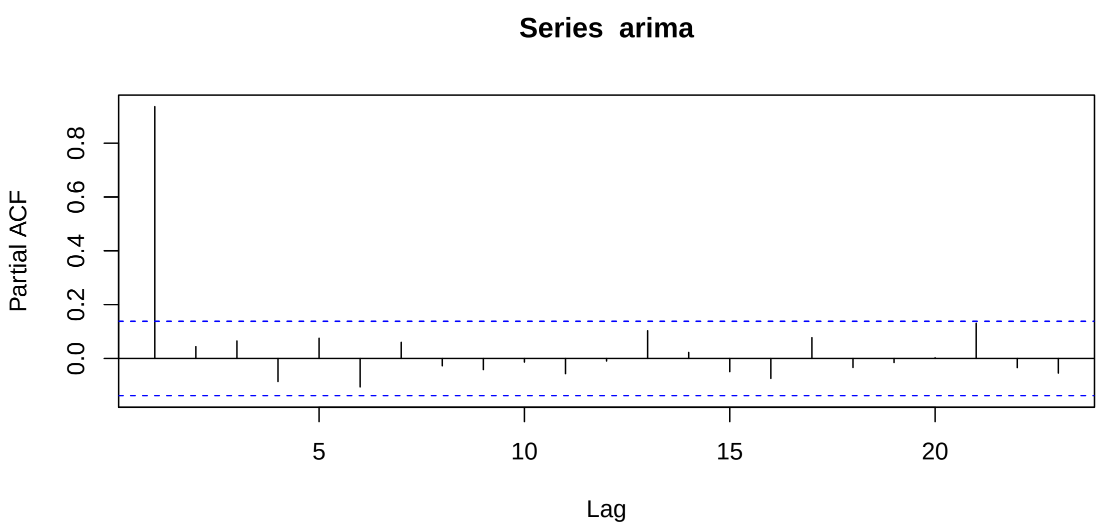
-
```
## ARIMA(0,1,0) Example
set.seed(123); n<-200; p<-0; d<-1;q<-0;
arima<-arima.sim(list(order=c(p,d,q)), n)
par(mfrow=c(1,1))
ts.plot(arima)
```



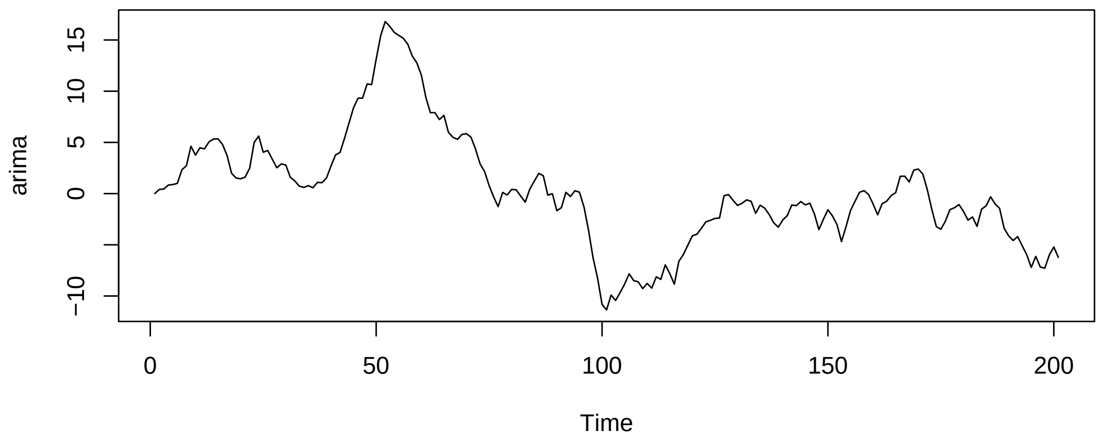

```
acf(arima,type = "correlation",plot = T)
```



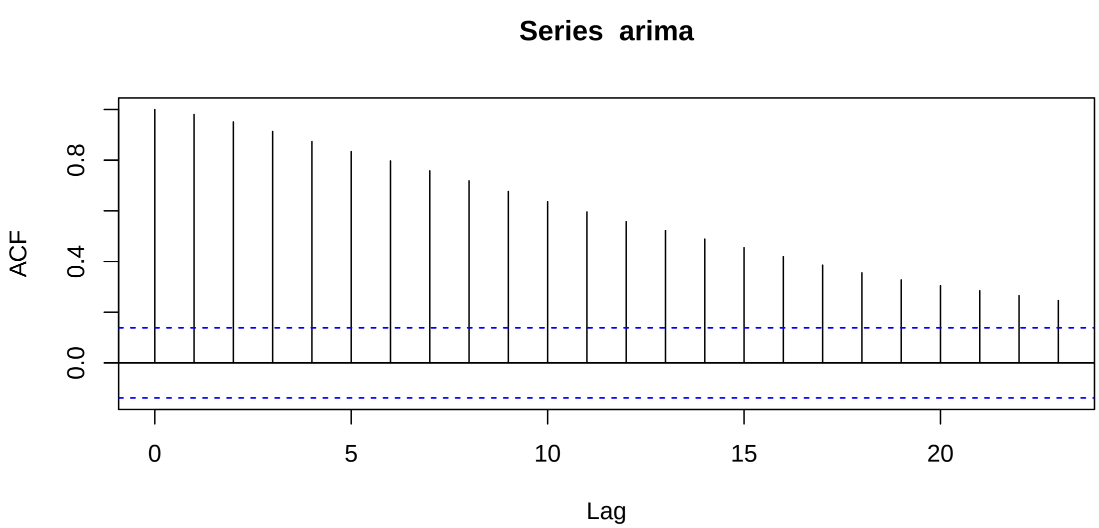

```r
pacf(arima,plot = T)
```



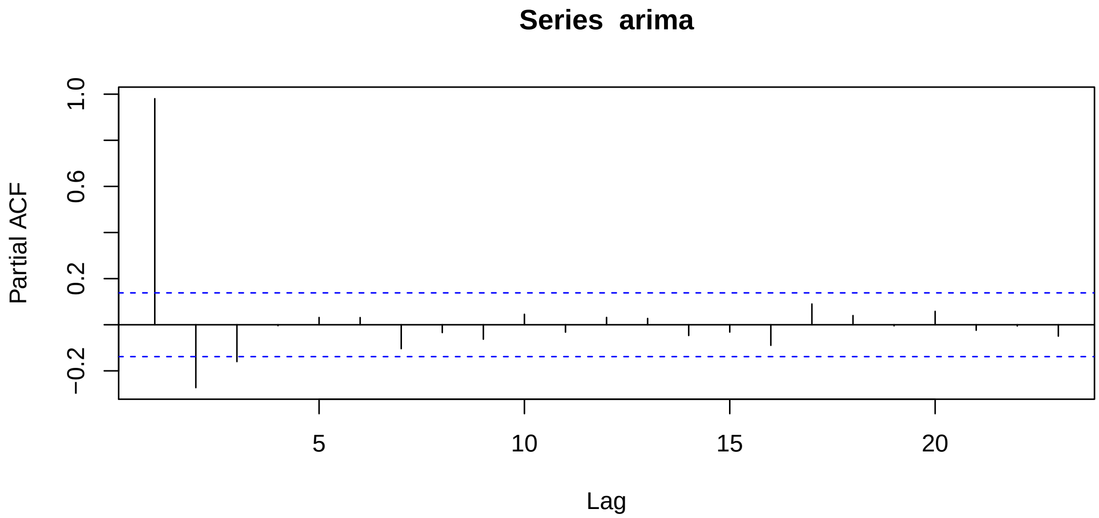

```r
## ARIMA(2,1,1) Example
set.seed(123); n<-200; p<-2; d<-1;q<-1;
arima<-arima.sim(list(order=c(p,d,q), ar=c(-0.3, 0.5),ma=c(0.7)), n)
par(mfrow=c(1,1))
ts.plot(arima)
```



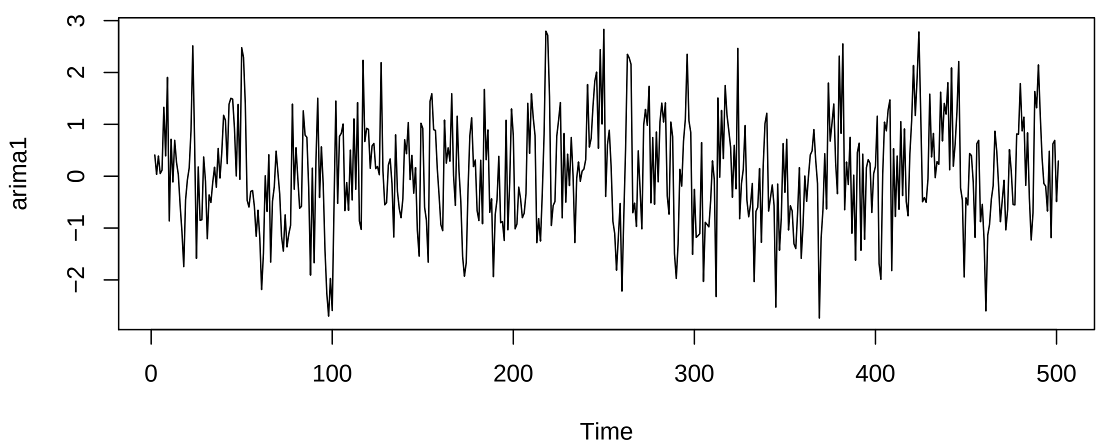

```
acf(arima,type = "correlation",plot = T)
```



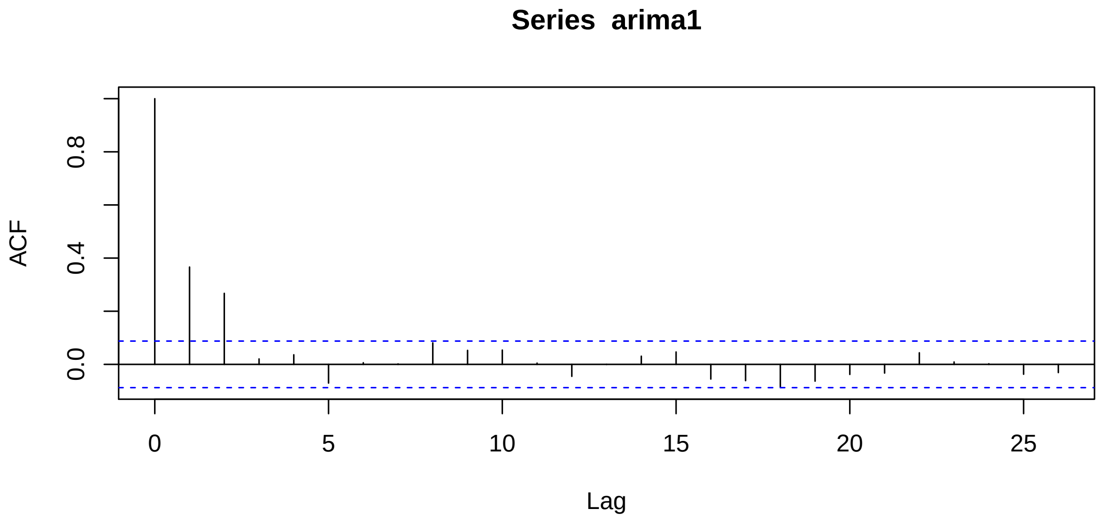

```
pacf(arima, plot = T)
```



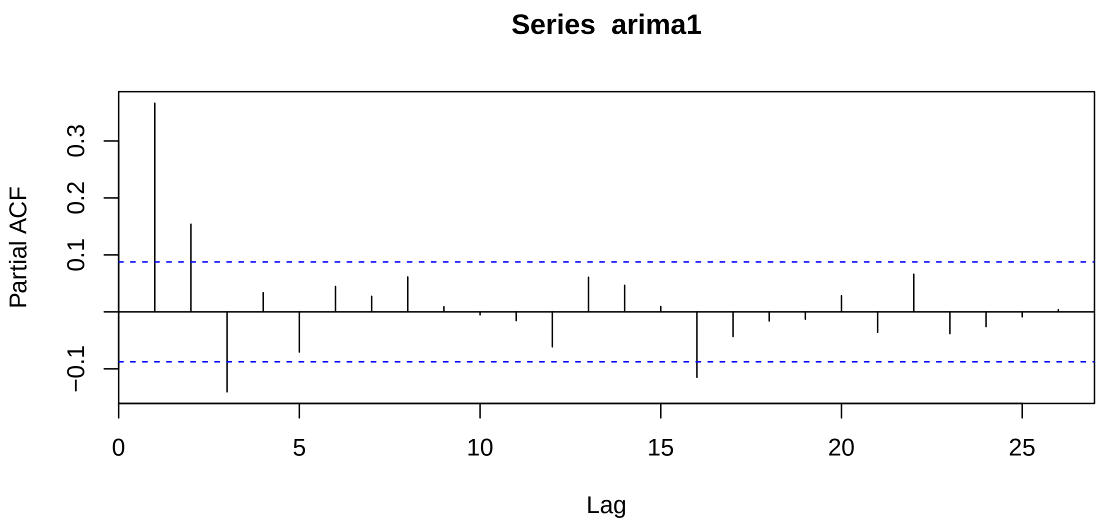

```
set.seed(123); n<-500; p<-2; d<-1;q<-1;
arima<-arima.sim(list(order=c(p,d,q), ar=c(-0.3, 0.5), ma=c(0.7)), n)
arima1<-diff(arima, lag = 1)
par(mfrow=c(1,1))
ts.plot(arima1)
```




```
acf(arima1,type = "correlation",plot = T)
```

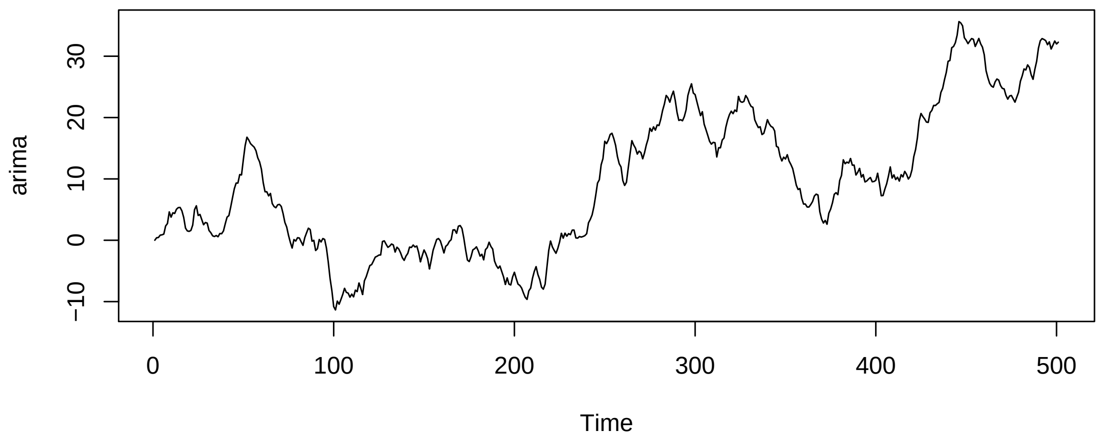



```
pacf(arima1, plot = T)
```

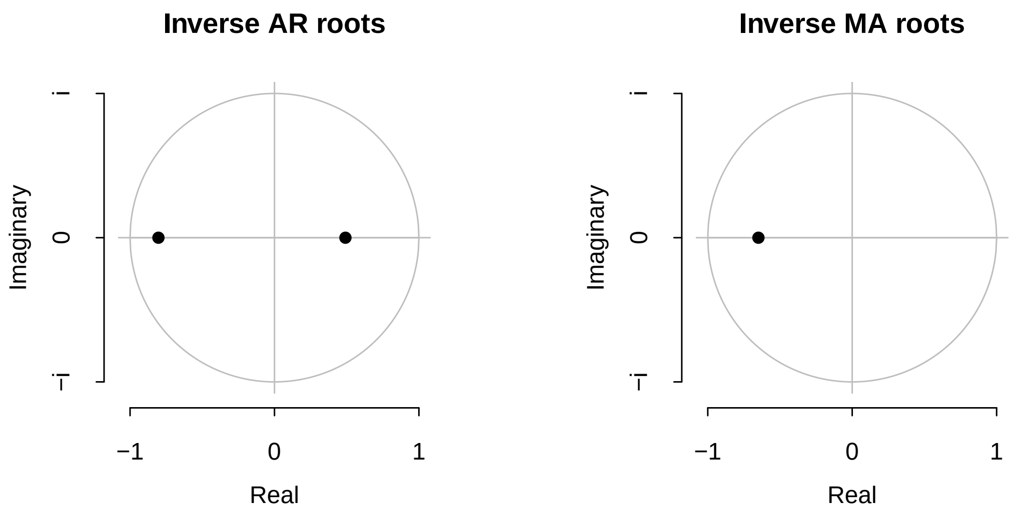



```
## ARMA(2,1,1) Example
set.seed(123); n<-500; p<-2; d<-1;q<-1;
arima<-arima.sim(list(order=c(p,d,q), ar=c(-0.3, 0.5),ma=c(0.7)), n)
par(mfrow=c(1,1))
ts.plot(arima)
library('forecast')
## Warning:  package 'forecast' was built under R version 3.4.4
## Warning in as.POSIXlt.POSIXct(Sys.time()):  unknown timezone 'zone/tz/2021a.3.0/zoneinfo/Asia/
```

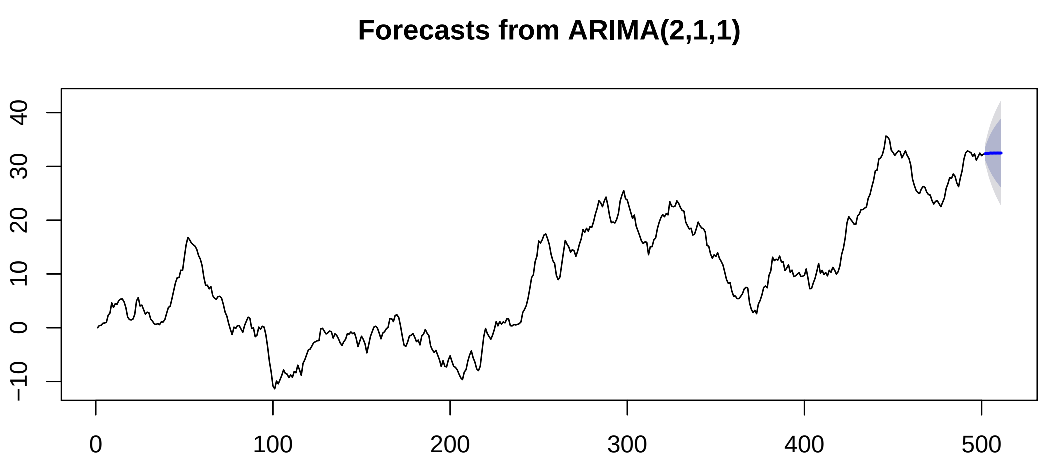
```
##
## Attaching package:  'forecast'
## The following object is masked from 'package:astsa':
##
##     gas
```




```
aarima<-auto.arima(arima)
print(auto.arima(arima))
## Series: arima
## ARIMA(2,1,1)
##
## Coefficients:
##           ar1     ar2     ma1
##       -0.3124  0.3947  0.6493
## s.e.   0.1026  0.0450  0.1059
##
## sigma^2 estimated as 0.9226:  log likelihood=-687.96
## AIC=1383.92   AICc=1384   BIC=1400.78
plot(aarima)
```




```
fc<-forecast(aarima, h=10)
plot(fc)
```



Multiplicative seasonal ARIMA Models: 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 $ARIMA(p,d,q) \times (P,D,Q)_s$

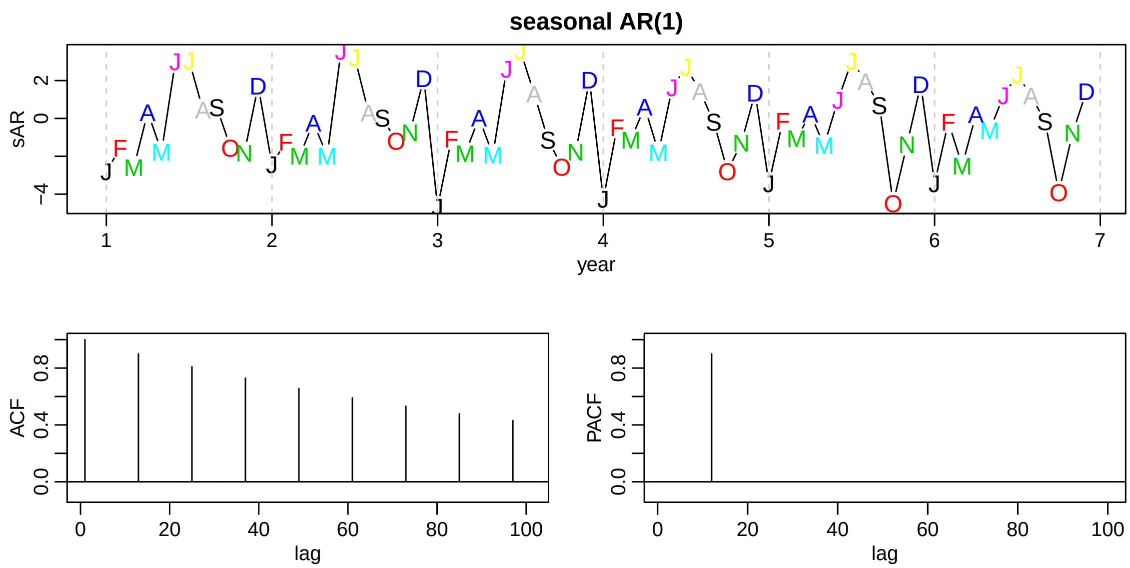
$$\Phi_P(B^s)\Phi_p(B)\nabla_s^D \nabla^d X_t = \Theta_Q(B^s)\Theta_q(B)W_t,$$

where the seasonal difference operator of order D is defined by

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

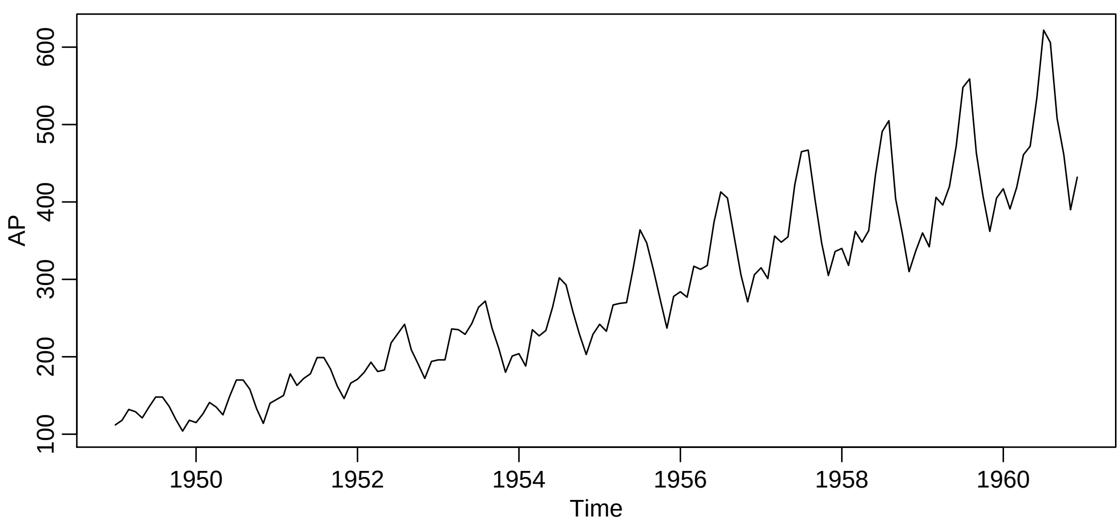
•

```
## Multiplicative seasonal ARIMA Models
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)
```

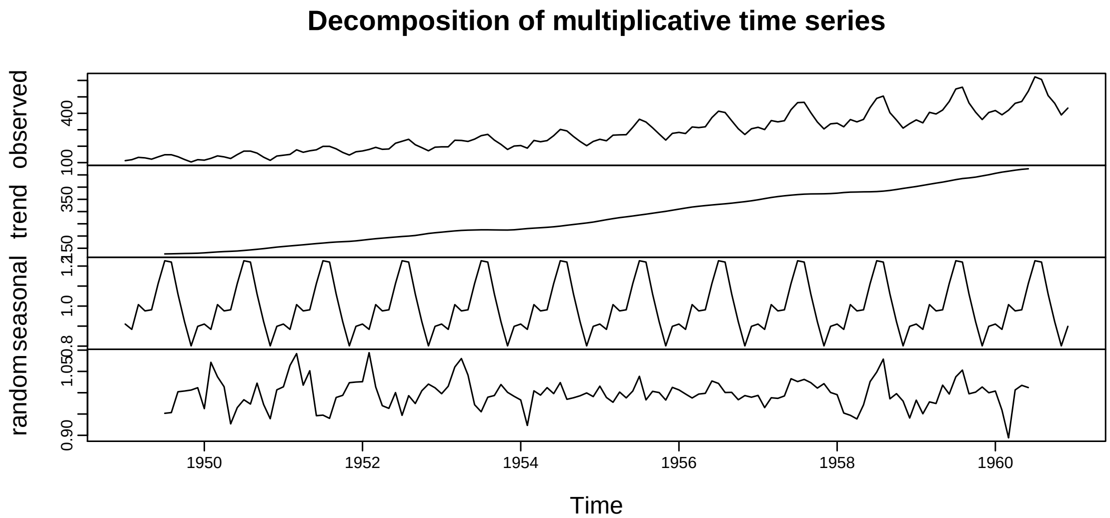



```
data(AirPassengers)
AP <- AirPassengers
print(AP)

##        Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
```

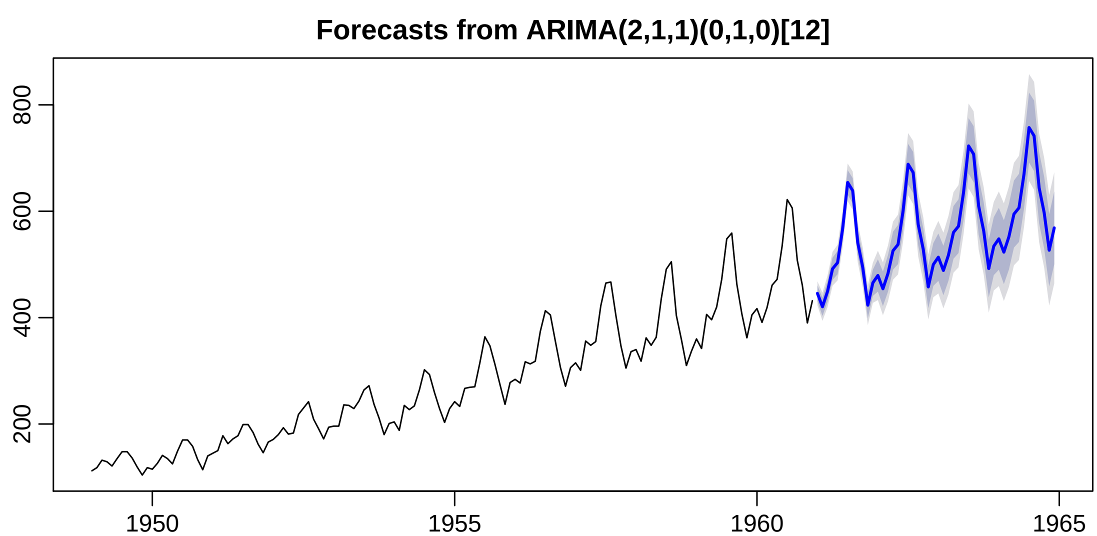
```
## 1949 112 118 132 129 121 135 148 148 136 119 104 118
## 1950 115 126 141 135 125 149 170 170 158 133 114 140
## 1951 145 150 178 163 172 178 199 199 184 162 146 166
## 1952 171 180 193 181 183 218 230 242 209 191 172 194
## 1953 196 196 236 235 229 243 264 272 237 211 180 201
## 1954 204 188 235 227 234 264 302 293 259 229 203 229
## 1955 242 233 267 269 270 315 364 347 312 274 237 278
## 1956 284 277 317 313 318 374 413 405 355 306 271 306
## 1957 315 301 356 348 355 422 465 467 404 347 305 336
## 1958 340 318 362 348 363 435 491 505 404 359 310 337
## 1959 360 342 406 396 420 472 548 559 463 407 362 405
## 1960 417 391 419 461 472 535 622 606 508 461 390 432
par(mfrow=c(1,1))
ts.plot(AP)
```




```
APD<-decompose(AP, type="mult")
par(mfrow=c(1,1))
plot(APD)
```




```r
par(mfrow=c(1,1))
library('forecast')
aarima<-auto.arima(AP)
fc<-forecast(aarima, h=48)
plot(fc)
```




```r
library('tseries')
## Warning:  package 'tseries' was built under R version 3.4.4
# p-value < 0.05 indicates the TS is stationary
adf.test(na.omit(APD$random))
```

```
## Warning in adf.test(na.omit(APD$random)):  p-value smaller than printed p-value
##
##  Augmented Dickey-Fuller Test
##
## data:  na.omit(APD$random)
## Dickey-Fuller = -6.4236, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
# AP.predict <- predict(AP.hw, n.ahead = 4 * 12)
# ts.plot(AP, AP.predict, lty = 1:2)
```

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