

# Time Series

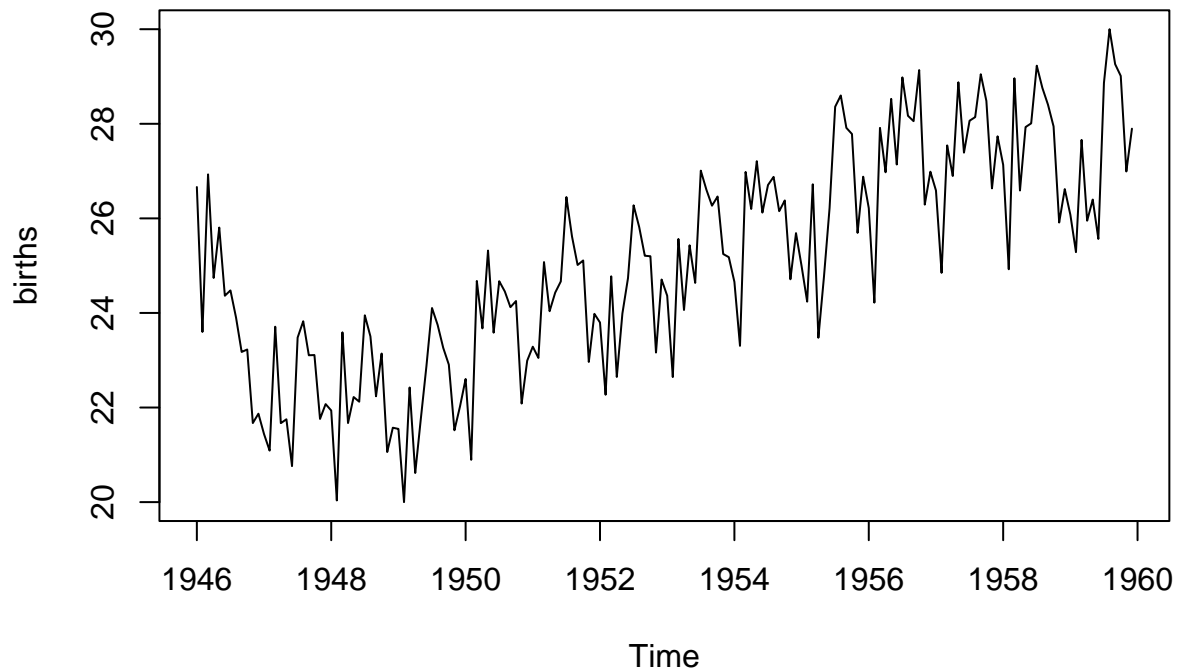
A time series is denoted by  $y_t$ , where  $t = 1, 2, 3, \dots$  are the time unit.

## Three Components of Time Series

To better analyze a time series, we often decompose it into three components: Time trend component (long term direction), Seasonal component (calendar related changes), and the random component.

For example, the time series of number of births per month in New York city, from January 1946 to December 1958, which is available at: <http://robjhyndman.com/tsdldata/data/nybirths.dat>

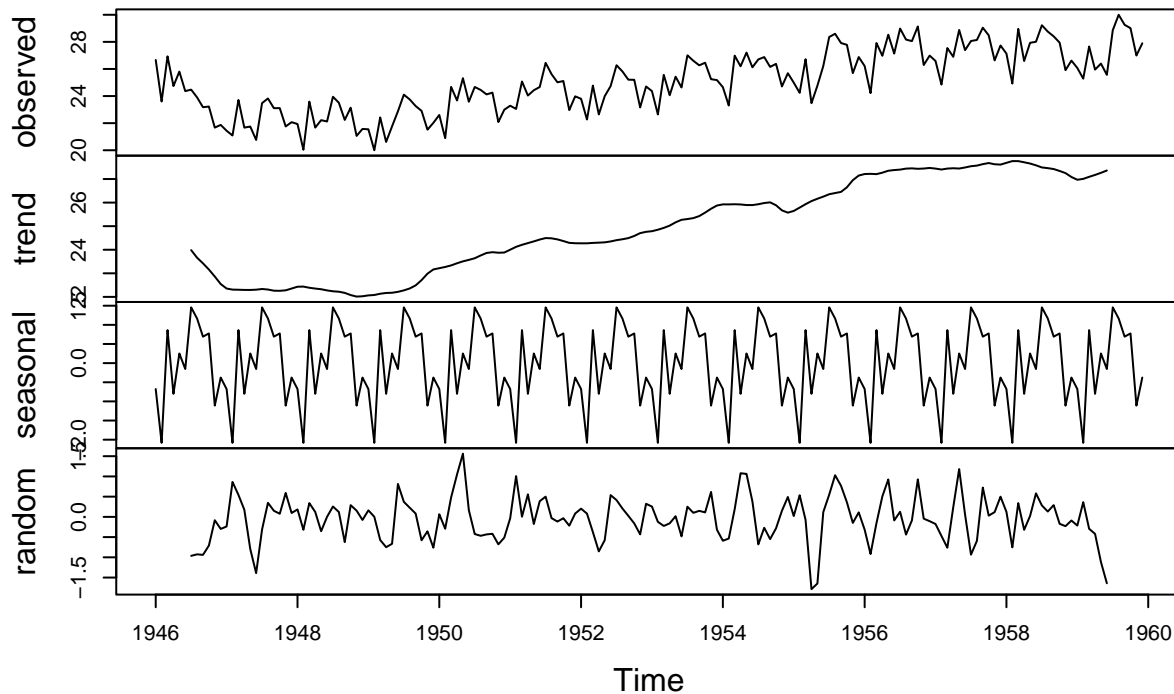
```
births <- scan("http://robjhyndman.com/tsdldata/data/nybirths.dat")
births <- ts(births, frequency = 12, start = c(1946, 1))
plot(births)
```



A plot shows that this time series has all three components, trend, seasonal, and random. We decompose this series to see each component individually.

```
birthsComp <- decompose(births)
plot(birthsComp)
```

## Decomposition of additive time series



## Stationarity

A time series  $y_t$  is called stationary if

- (1)  $E(y_t)$  does not depend on  $t$
- (2)  $Cov(y_t, y_s)$  is a function of  $|t - s|$

From (2),  $var(y_t) = cov(y_t, y_t)$  does not depend on  $t$ . In general, a stationary series has features that do not depend on the time  $t$ . Thus, series with seasonal or trend are not stationary series.

## Autocorrelation

We can calculate the correlation of the time series to itself with the one unit time delay. For example, we can calculate the correlation of  $(y_1, y_2, \dots, y_9)$  with  $(y_2, y_3, \dots, y_{10})$ . This is called a lag 1 autocorrelation  $r_1$ . More formally,

$$r_1 = \frac{\sum_{t=2}^T (y_{t-1} - \bar{y})(y_t - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

In general, the lag  $k$  autocorrelation is

$$r_k = \frac{\sum_{t=k+1}^T (y_{t-k} - \bar{y})(y_t - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2},$$

for  $k = 1, 2, 3, \dots$

We can consider lag- $k$  autocorrelation as a function of  $k$ . This function is called autocorrelation function (ACF).

We consider two series: the first one,  $a$ , with a trend and the second one,  $b$ , contains just a random component.

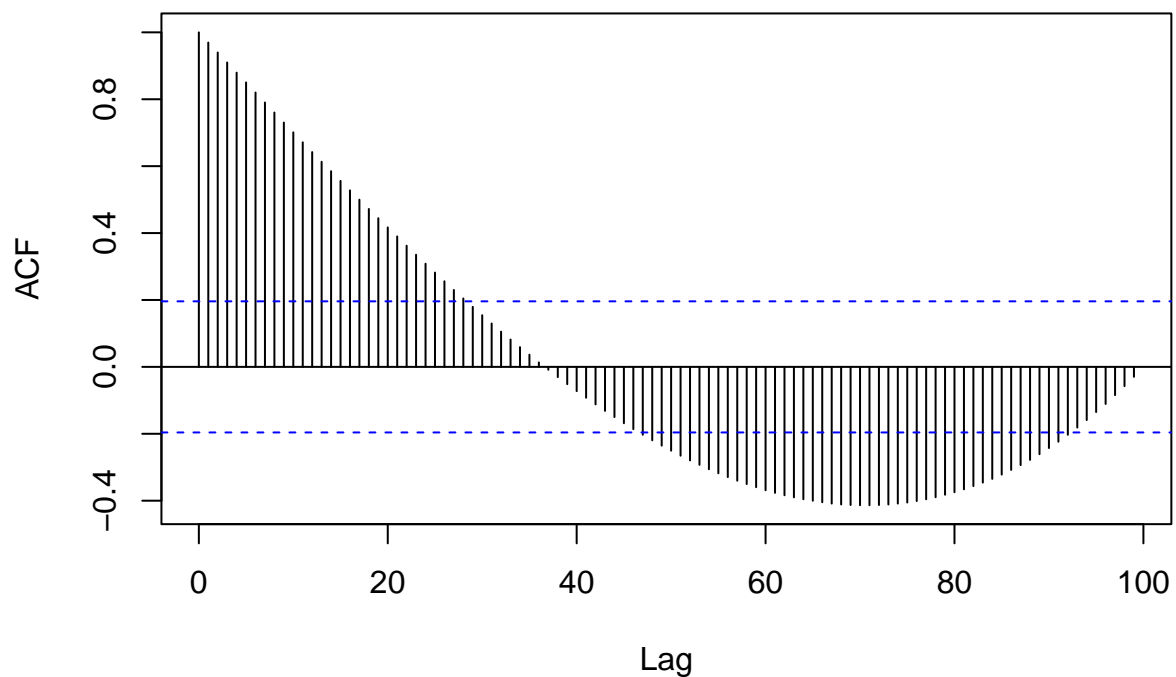
$$a_t = 3t + 4 + \epsilon, b_t = \epsilon$$

where  $\epsilon \sim N(0, 1)$ .

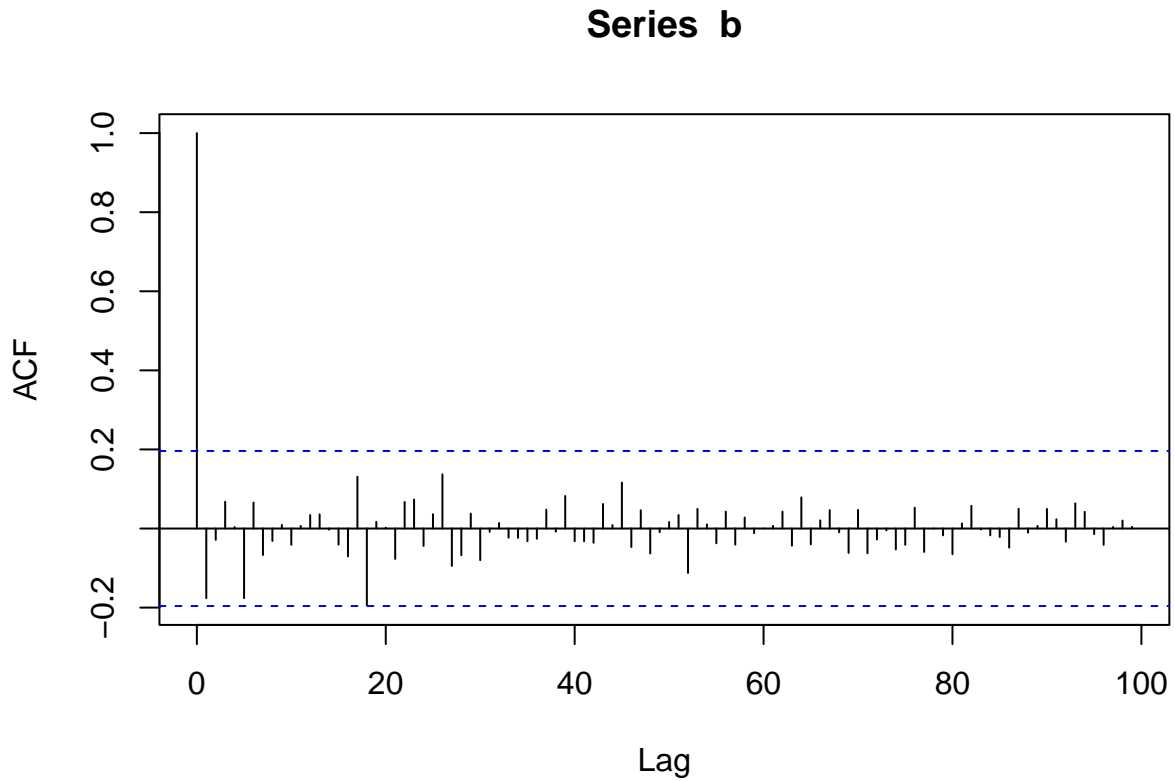
```
# acf of trend series
n=100
a = rep(0,n)
b = rep(0,n)

for (t in c(1:n))
{
  a[t] = 3*t + 4 + rnorm(0, 1, n=1)
  b[t] = rnorm(0, 1, n=1)
}
acf(a, lag.max = 100)
```

### Series a



```
acf(b, lag.max = 100)
```



The ACF of series a show a strong positive autocorrelations in the first range of lags and then reverse to negative correlations. The ACF of b does not show a very weak correlation (within the blue lines) and the correlations die out as the lag increased.

ACF is a good tool to check if a series is stationary, which is a series *b* in this case.

## Time Series Models

### White Noise

A white noise time series consists of terms that are independent and comes from a identical distribution (iid). The distribution is often assumed to be normal distribution. The mean of the distribution is usually assumed to be zero.

$$y_t \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2) \quad (1)$$

for  $t = 1, 2, \dots$ . A white noise is a stationary series.

### Random Walk

A random walk series is a series that has the following form

$$y_t = y_{t-1} + c_t$$

where  $c_1, c_2, \dots, c_T$  are observations of a white noise.

A random walk can be written as

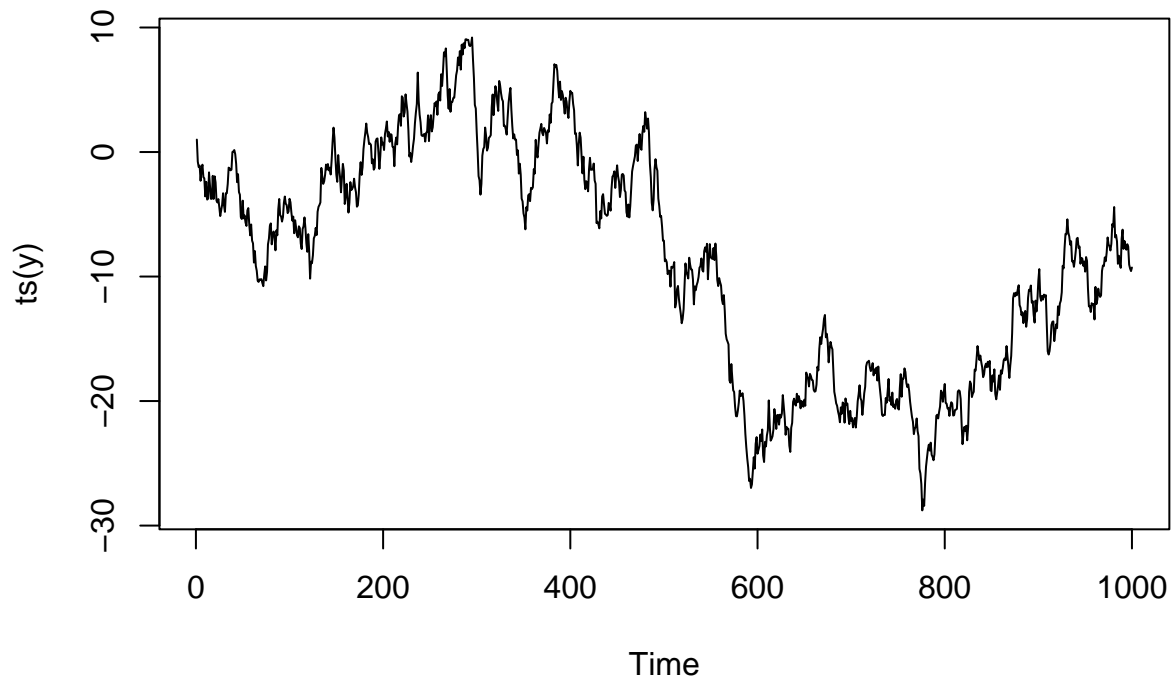
$$y_t = y_0 + c_0 + c_1 + c_2 + \dots + c_t$$

Let's plot a random walk  $y_t = y_{t-1} + \epsilon$  where  $y_1 = 1, \epsilon \sim N(0, 1)$  and its ACF.

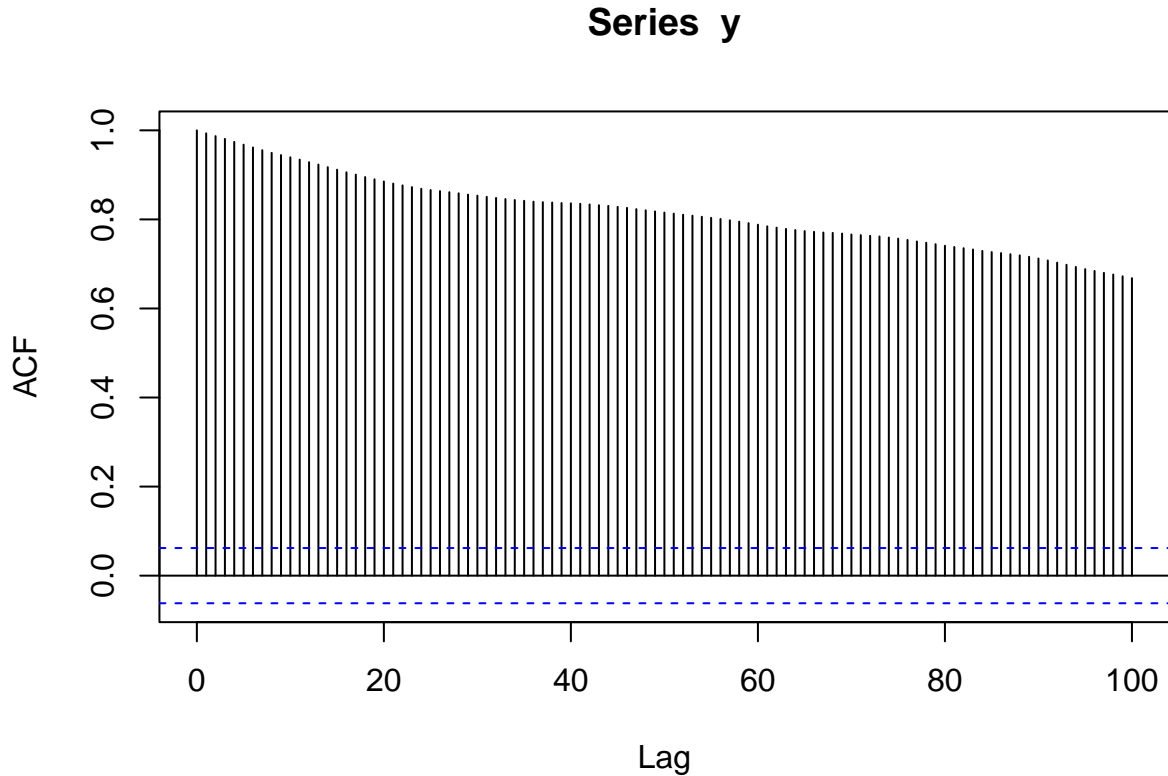
```
# acf of trend series
n=1000
y = rep(0,n)

y[1] = 1

for (t in c(2:n))
{
  y[t] = y[t-1]+ rnorm(0, 1, n=1)
}
plot(ts(y))
```



```
acf(y, lag.max = 100)
```



It can be seen from the ACF that the random walk is non-stationary. In general, random walks are not stationary as its variance depend on the time  $t$ .

If  $c_t \sim (\mu, \sigma^2)$ , we have  $var(y_t) = (t)\sigma^2$  and  $E(y_t) = y_1 + t\mu$ . Thus, both the mean and the variance of a random walks may depend on the time  $t$ .

### Autoregressive

The autoregressive model of order 1, denoted by AR(1), is defined by

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon,$$

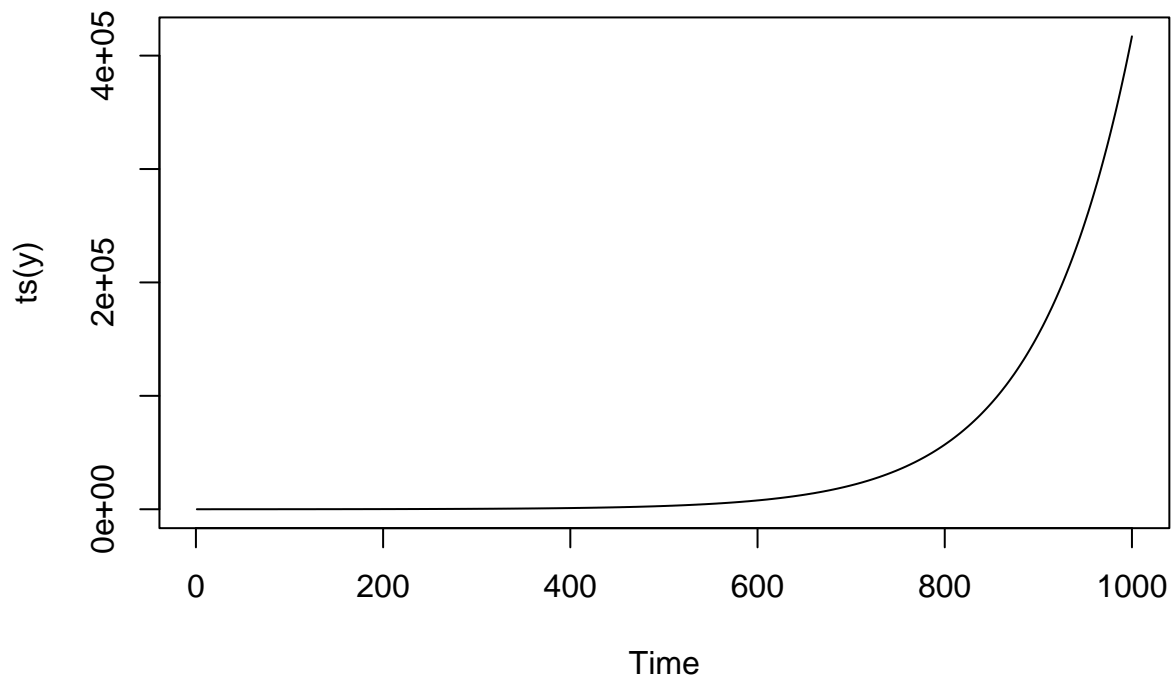
where  $\epsilon$  is a white noise process such that  $Cov(\epsilon_{t+k}, y_t) = 0$  for  $k > 0$ .  $\beta_0$  and  $\beta_1$  are parameters (non-random).  $-1 \leq \beta_1 \leq 1$ .

We notice that if  $\beta_1 = 1$ , an AR(1) becomes a random walk, and if  $\beta_1 = 0$  an AR(1) becomes a white noise.

In the below, we plot autoregressive models for different  $\beta_1$  values.

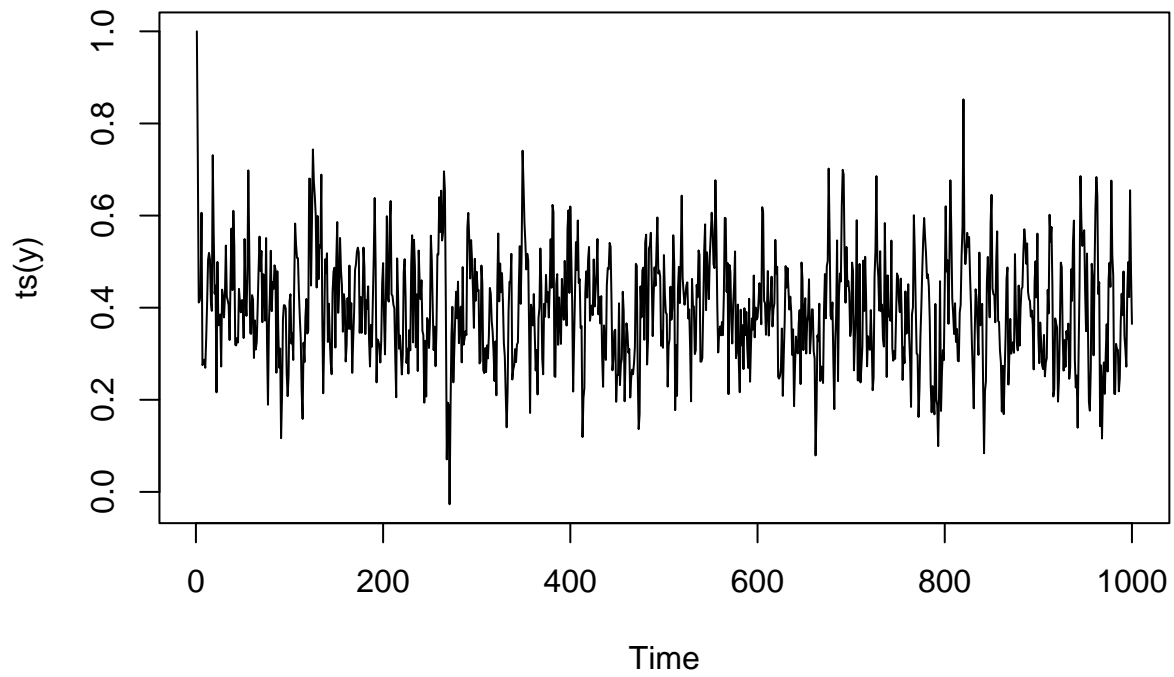
```
n=1000
y = rep(0,n)
y[1] = 1
for (t in c(2:n))
{
  y[t] = .2+ 1.01*y[t-1]+ rnorm(0, .1, n=1)
}
plot(ts(y))
title(main="Coefficient = 1.01")
```

**Coefficient = 1.01**



```
n=1000
y = rep(0,n)
y[1] = 1
for (t in c(2:n))
{
  y[t] = .2+ .5*y[t-1]+ rnorm(0, .1, n=1)
}
plot(ts(y))
title(main="Coefficient = .5")
```

### Coefficient = .5



The plot shows that the AR(1) with  $\beta_1 = .5$  is a stationary process. This is indeed true for any AR(1) with  $-1 < \beta_1 < 1$ . When  $\beta_1 = \pm 1$ , the AR(1) is non-stationary.

### Practice Problems