UNIT 15 STATIONARY PROCESSES

Structure

15.1 INTRODUCTION

In Units 13 and 14, you have learnt that a time series can be decomposed into four components, i.e., Trend (T), Cyclic (C), Seasonal (S) and Irregular (I) components. We have discussed methods for smoothing or filtering the time series and for estimating Trend, Seasonal and Cyclic components. We have also explained how to use them for forecasting.

In this unit, we describe a very important class of time series, called the **stationary time series**. In Sec. 15.2, we explain the concept of stationary process and define weak and strict stationary processes. We discuss autocovariance, autocorrelation function and correlogram of a stationary process in Secs. 15.3 and 15.4. If a time series is stationary, we can model it and draw further inferences and make forecasts. If a time series is not stationary, we cannot do any further analysis and hence cannot make reliable forecasts. If a time series shows a particular type of non-stationarity and some simple transformations can make it stationary, then we can model it.

In the next unit, we shall discuss certain stationary linear models such as Auto Regressive (AR), Moving Average (MA) and mixed Autoregressive Moving Average (ARMA) processes. We shall also discuss how to deal with models with trend by considering an integrated model called Autoregressive Integrated Moving Average (ARIMA) model.

Objectives

After studying this unit, you should be able to:

- describe stationary processes;
- define weak and strict stationary processes;
- define autocovariance and autocorrelation coefficients;
- estimate autocovariance and autocorrelation coefficients;
- plot the correlogram and interpret it; and
- make proper choice of probability models for further studies.

15.2 STATIONARY PROCESSES

In the course MST-004, you have studied random variables and their properties. Recall that a random variable Y is a function defined on a sample space. A family of random variables defined on the same sample space taking values over time is known as a **random process**. Most physical processes in real life situations involve random components or variables and a random process may be described as a statistical phenomenon that evolves in time. A random process may be defined as a family of random variables defined on a given probability space indexed by the parameter t. Here we denote a stochastic variable by the capital letter Y and assume that it is observable at discrete time points t_1, t_2, ...

A random process is a statistical phenomenon that evolves in time according to some laws of probability. The length of queue in a system, the number of accidents in a particular city in successive weeks, etc. are examples of a random process. Mathematically, a random process is defined as the family of random variables which are ordered in time, i.e., random process {Y(t); t belongs to T} is a collection of random variables, where T is a set for which all random variables Y_t are defined on the same sample space. If T takes continuous range of values, the random process is said to be a **continuous parameter process**. On the other hand, if T takes discrete set of values, the process is said to be a **discrete parameter process**. We use the notation Y_t for a random process when we deal with discrete parameter processes. When T represents time, the random process is referred to as a time series.

In Units 13 and 14, we have dealt with one set of observations recorded at different times. Thus, we had only a single outcome of the process and a single observation on the random variable at time t. This sample may be regarded as one time series out of the infinite set of time series, which might have been observed. This infinite set of time series is called an **Ensemble**.

Every member of the ensemble can be taken as a possible realisation of the stochastic process and the observed time series can be considered as one particular realisation.

15.2.1 Stationary Process

Broadly speaking, a time series is said to be **stationary** if there is no systematic change in mean, variance and covariance of the observations over a period of time. This means that the properties of one section of the time series are similar to the properties of the other sections of the time series. In other words, a process is said to be stationary if it is in a state of statistical equilibrium.

A random process is said to be stationary if the joint distribution of $Y_{t1}, Y_{t2}, Y_{t3}, \ldots, Y_{tk}$ is the same as the joint distribution of $Y_{t1+J}, Y_{t2+J}, Y_{t3+J}, \ldots, Y_{tk+J}$, for all $t_1, t_2, \ldots, t_k$ and J. In other words, shifting the origin of time by an amount J has no effect on the joint distribution. This means that it depends only on the interval between $t_1, t_2, \ldots, t_k$. This definition holds for any value of k.

For k =1,

$$E(Y_t) = \mu \qquad \text{and} \qquad V(Y_t) = \sigma^2$$

This implies that the mean and variance of Y_t are constant and do not depend on time.

For k = 2, the joint distribution of Y_{t1} and Y_{t2} depends only on the time difference $(t_2 - t_1) = J$, say, which is called **lag**.

Thus, the covariance term depends only on lag $J = t_2 - t_1$, i.e.,

$$\gamma(t_1, t_2) = E[(Y_t - \mu)(Y_{t+J} - \mu)]$$

$$= Cov(Y_t, Y_{t+J})$$

The variance function is a special case of the covariance function when $t_1 = t_2$, i.e., $J = 0$.

There are two types of stationary processes: Strict stationary processes and weak stationary processes. Let us discuss them one at a time.

15.2.2 Strict Stationary Process

Strict stationary process imposes the strong condition that the joint probability distribution remains the same and depends only on the time interval. If all the finite dimensional distributions of a random process are invariant under translation of time, then it is called a **Strict Sense Stationary process** or **SSS process**. In other words, if the joint distribution of $Y_{t1}, Y_{t2}, \ldots, Y_{tk}$ is the same as the joint distribution of $Y_{t1+J}, Y_{t2+J}, \ldots, Y_{tk+J}$, for all $t_1, t_2, \ldots, t_k$ and $J (> 0)$ for all $k \geq 1$, the random process Y_t is called a Strict Sense Stationary process.

The strict stationary process requires that for any t_1, t_2, the distributions of Y_{t1} and Y_{t2} for all i = 1, 2, …, n must be the same. Also, the bivariate distributions of pairs $[Y_{t1}, Y_{t1+J}]$ and $[Y_{t2}, Y_{t2+J}]$ are the same for all i = 1, 2, …, n and J.

Note that the requirement of strict stationarity is a severe one. Usually it is difficult to establish it mathematically and in most cases, the distributions of $[Y_{ti}, Y_{ti+J}]$ for all i = 1, 2, …, n and J are not known. That is why, the less restrictive notions of stationarity called **weak stationary processes** have been developed.

15.2.3 Weak Stationary Process

A stationary process is said to have weak stationarity of order m, if the moments up to order m depend only on time lag J. If m = 2, the stationarity (or weak stationarity) of order 2 implies that moments up to the second order depend only on time lag J. For a weak stationary process:

$$E[y_t] = \mu;$$

and $$Cov[Y_t, Y_{t+J}] = \gamma(J)$$

No requirements are placed on the moments of higher order. The definition also implies that the mean and variance remain constant and finite. Thus, a random process Y_t with finite first and second order moments is called a **weak stationary process**, if the means are constant and the covariance depends only on the time lag.

In the subsequent discussion in this unit, we shall assume weak stationarity as many properties of a stationary process depend only on the first and second order moments. One important class of processes is the **normal** process, where joint distributions of $Y_{t1}, Y_{t2}, \ldots, Y_{tk}$ are multivariate normal for all $t_1, t_1, \ldots, t_k$. This multivariate normal distribution is completely characterised by its first and second moments, i.e., μ_t and $\gamma_{(t1, t2)}$.

15.3 AUTOCOVARIANCE AND AUTOCORRELATION

In this section, we discuss the autocovariance and autocorrelation function for a stationary process. Of particular interest in the analysis of a time series are the covariance and correlation between Y_{t1} and Y_{t2}. Since these are the covariance and correlation within the same time series, they are called **autocovariance** and **autocorrelation**. Some important properties of time series can be studied with the help of autocovariance and autocorrelation. They measure the linear relationship between observations at different time lags. They provide useful descriptive properties of the time series being studied and are important tools for guessing a suitable model for the time series data. Let us define these parameters.

15.3.1 Autocovariance and Autocorrelation Coefficients

Suppose a weak stationary time series process is denoted by $Y_1, Y_2, \ldots, Y_t, Y_{t+1}, \ldots, Y_{t+k}, \ldots$. We are interested in finding the linear relationship between two consecutive observations, Y_t, Y_{t+1}. We are also interested in the relationship between observations that are apart by a time lag k, e.g., Y_t and Y_{t+k} . We shall study the linear relationship by studying covariances and correlations of observations at different time lags.

In the course MST-002, you have learnt how to calculate the covariance and correlation between two variables for given N pairs of observations on two variables X and Y, say $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$. Recall that the formulas for computation of covariance and correlation coefficient are given as:

$$\mathrm{Cov}(X,Y) = E\left[(X-\mu)(Y-\mu)\right]$$

$$\rho_{(X,Y)} = \frac{E\left[(X-\mu)(Y-\mu)\right]}{\sqrt{\left[E(X-\mu)^2 E(Y-\mu)^2\right]}} \quad = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Here we apply analogous formulas to the stationary time series data to measure whether successive observations are correlated.

The autocovariance between Y_t and Y_{t+k}, separated by the time interval k, for a stationary process must be the same for all t and is defined as:

$$\gamma_k = \gamma_{-k} = \mathrm{Cov}(Y_t, Y_{t+k}) = E\left[(Y_{t+k}-\mu)(Y_t-\mu)\right] \qquad \ldots(1)$$

Similarly, the autocorrelation at lag k is

$$\rho_k = \frac{E\left[(Y_{t+k}-\mu)(Y_t-\mu)\right]}{\sqrt{\left[E(Y_{t+k}-\mu)^2 E(Y_t-\mu)^2\right]}}$$

$$= \frac{\mathrm{Cov}(Y_t, Y_{t+k})}{\sigma_Y^2} \qquad \ldots (2)$$

From equation (1), we note that

$$\sigma_Y^2 = \gamma_0 \qquad \ldots (3)$$

Therefore, $\rho_k = \dfrac{\gamma_k}{\gamma_0}$ and $\rho_0 = 1$ … (4)

15.3.2 Estimation of Autocovariance and Autocorrelation Coefficients

So far, we have defined the autocovariance and autocorrelation coefficients for a random process. You would now like to estimate them for a finite time series for which N observations $y_1, y_2, ..., y_N$ are available. We shall denote a realisation of the random process $Y_1, Y_2, \ldots, Y_N$ by small letters $y_1, y_2, ..., y_N$. The mean μ can be estimated by

$$\bar{y} = \sum_{i=1}^{N} y_i / N \qquad \ldots (5)$$

and autocovariance γ_k at lag k can be estimated by the autocovariance coefficient c_k as follows:

$$c_k = \sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y}) / (N - k), \quad \text{for all } k \qquad \ldots (6)$$

The sample covariance is a special case of autocovariance when k = 0, i.e.,

$$c_0 = \sum_{t=1}^{N} (y_t - \bar{y})(y_t - \bar{y}) \Big/ N = \sum_{t=1}^{N} (y_t - \bar{y})^2 / N = \sigma_Y^2, \text{ for all } k$$

The autocorrelation coefficients (r_k) are usually calculated by computing the series of autocovariance coefficients (c_k) as follows:

$$r_k = \frac{c_k}{c_0} = \frac{c_k}{\sigma_Y^2} \qquad \ldots (7)$$

In practice, at least 50 observations are required for the estimation of correlations. It is also advisable that for calculations of r_k, the lag k should not exceed N/4.

Let us explain these concepts with the help of an example.

Example 1: A series of 10 consecutive yields from a batch chemical process are given as follows:

47, 64, 23, 71, 38, 64, 55, 41, 59, 48

Calculate the mean, autocovariance c_1 and autocorrelation coefficient r_1 for the given time series.

Solution: We first construct the following table:

S. No.	Y	Y^2	$(Y_t - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+1} - \bar{Y})$
1	47	2209	−4	
2	64	4096	13	−52
3	23	529	−28	−364
4	71	5041	20	−560
5	38	1444	−13	−260
6	64	4096	13	−169
7	55	3025	4	52
8	41	1681	−10	−40
9	59	3481	8	−80
10	48	2304	−3	−24
Total	$\sum y_i = 510$	$\sum y_i^2 = 27906$		$\sum (Y_t - \bar{Y})(Y_{t+1} - \bar{Y}) = -1497$

From equation (5), we get

$$\bar{y} = \frac{\sum_{i=1}^{N} y_i}{10} = \frac{510}{10} = 51.0$$

From equation (6), for k=0, the autocovariance coefficient is

$$c_0 = \frac{\sum_{t=1}^{N-k} \left(y_i - \bar{y}\right)^2}{N} = \frac{\left(\sum y_i^2 - N\bar{y}^2\right)}{N}$$

$$= \frac{(27906 - 26010)}{10} = 189.6$$

For k = 1,

$$c_1 = \frac{\sum_{t=1}^{9} \left(y_t - \bar{y}\right)\left(y_{t+1} - \bar{y}\right)}{9} = -166.33$$

From equation (7),

$$r_1 = \frac{c_1}{c_2} = \frac{-166.33}{189.6} = -0.88$$

You may now like to solve a problem to assess your understanding.

E1) Ten successive observations on a stationary time series are as follows:

1.6, 0.8, 1.2, 0.5, 0.6, 1.5, 0.8, 1.2, 0.5, 1.3.

Plot the observations and calculate r_1.

E2) Fifteen successive observations on a stationary time series are as follows:

34, 24, 23, 31, 38, 34, 35, 31, 29,
28, 25, 27, 32, 33, 30.

Plot the observations and calculate r_1.

15.4 CORRELOGRAM OF STATIONARY PROCESSES

A useful plot for interpreting a set of autocorrelations coefficient is called a **correlogram** in which the sample autocorrelation coefficients r_k are plotted versus the lag J where J=1, 2, 3, …, k. This helps us in examining the nature of time series. It is also a very important diagnostic tool for the selection of a suitable model for the process which generates the data. The correlogram is alternatively known as the **sample autocorrelation function** (**acf**).

The value of lag J is usually much less than N. For example, a time series of length N = 200 given in Fig 15.1a shows the plot of the time series for N=200 and Fig. 15.1b shows a plot of the correlogram for a lag up to order 17.

The relatively smooth nature of the time series plot indicates that observations which are close to each other (at smaller lags) are positively correlated. The correlogram suggests that observations with smaller lag are positively correlated and autocorrelation decreases as lag k increases.

In most time series, it is noticed that the absolute value of r_k, i.e., $| r_k|$ decreases as k increases. This is because observations which are located far away are not related to each other, whereas observations lying closer to each other may be positively (or negatively) correlated.

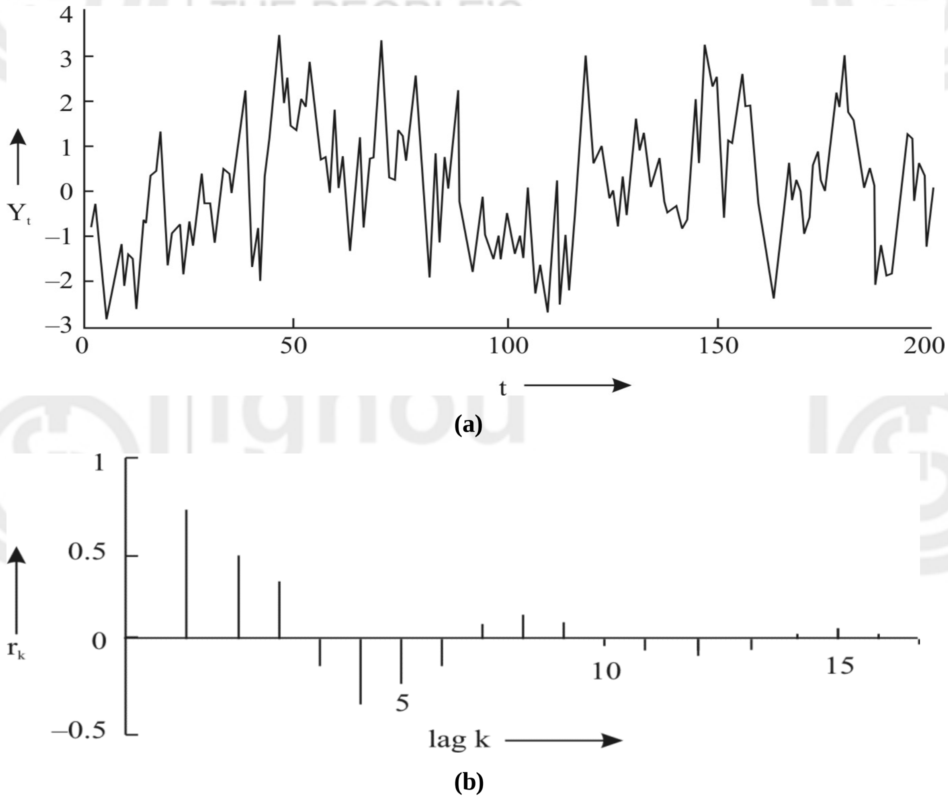


Fig. 15.1: a) Plot of a time series for N = 200; b) correlogram for lag k = 1, 2, .., 17.

15.4.1 Interpretation of Correlogram

Sometimes it is possible to recognise the nature of a time series from its correlogram, though it is not always easy. We shall describe certain types of time series processes and the nature of their correlograms. If the correlogram of a time series is one of these types, it is possible to get a good idea of the process which gave rise to that time series.

Random Series

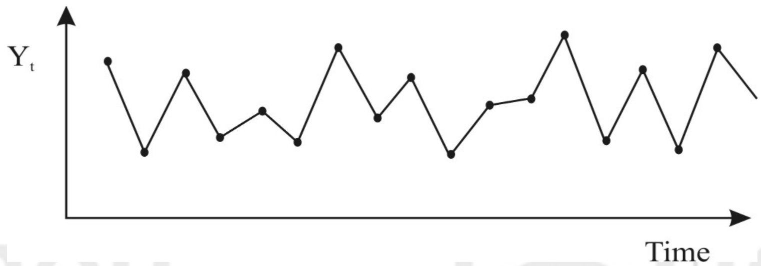
In case all observations are completely random, i.e., they contain only independent observations, the series is called a **random series**.
This means that $r_k \approx 0$ for all non zero k. The correlogam of such a random series will be oscillating around the axis (zero line). In fact, for such a series, for large N, the values of r_k approximately follow the N(0,1/N) distribution. Thus, in about 95% cases, the values of r_k lie in a range of $\pm 2/\sqrt{N}$. If the correlogram shows such a behaviour, it is a good indication that the time series is random. However, this behaviour may not always confirm that the time series is random and it may need further examination.

Short-Term Correlation

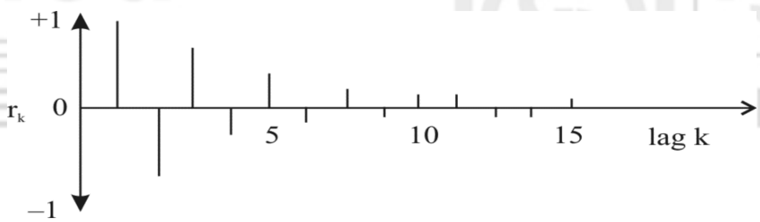
Stationary time series usually has a few large autocorrelations in absolute value for small lag k. They tend to zero very rapidly with increase in lag k (see Fig. 15.1b). When the first few autocorrelations are positive, the time series is smooth in nature, i.e., if an observation is above mean, it is likely to be followed by an observation above mean and if an observation is below mean, it is likely to be followed by an observation below mean. This gives an indication of stationary time series with most of the non-zero autocorrelations being either positive or negative.

Alternating Series

If a time series behaves in a very rough and zig-zag manner, alternating between values above and below mean, it is indicated by a negative r_1 and positive r_2. An alternating time series with its correlogram is shown in Fig.15.2.



(a)

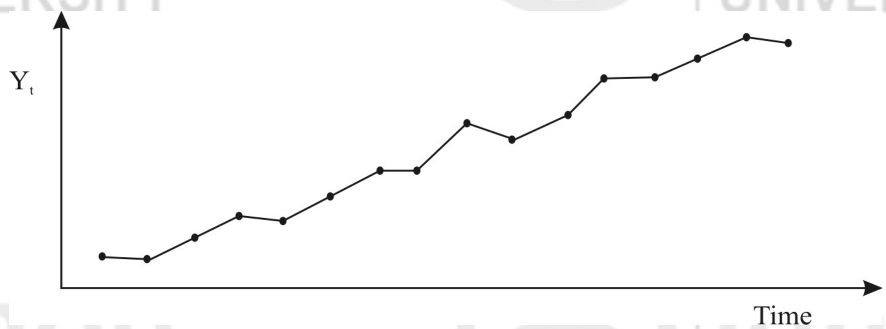


(b)

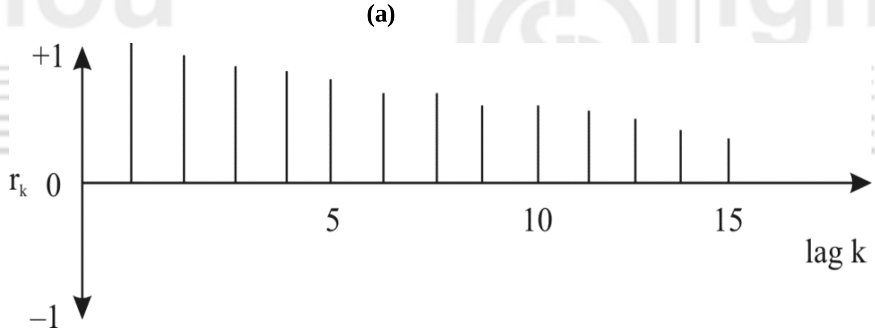
Fig. 15.2: a) Plot of alternating time series; b) correlogram for an alternating series with lag up to 15.

Non-Stationary Time Series

If a time series contains trend, it is said to be **non-stationary**. Such a series is usually very smooth in nature and its autocorrelations go to zero very slowly as the observations are dominated by trend. Due to the presence of trend, the autocorrelations move towards zero very slowly (see Fig. 15.3). One should remove trend from such a time series before doing any further analysis.



(a)



(b)

Seasonal Time Series

If a time series has a dominant seasonal pattern, the time plot will show a cyclical behaviour with a periodicity of the seasonal effect. If data have been recorded on monthly basis and the seasonal effect is of twelve months, i.e., s = 12, we would expect a highly negative autocorrelation at lag 6 (r_6) and highly positive correlation at lag 12 (r_{12}). In case of quarterly data, we expect to find a large negative r_2 and large positive r_4. This behaviour will be repeated at r_6, r_8 and so on. This pattern of cyclical behaviour of correlogram will be similar to the time plot of the data.

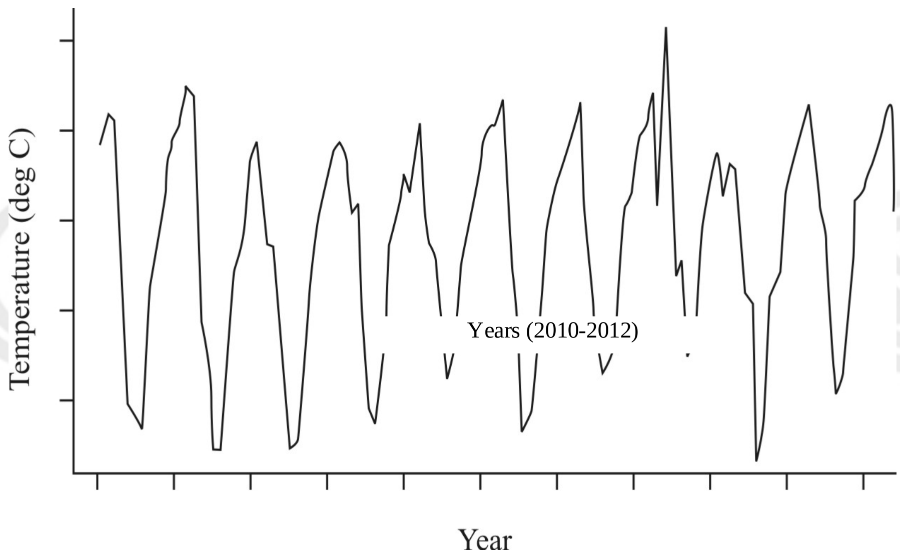
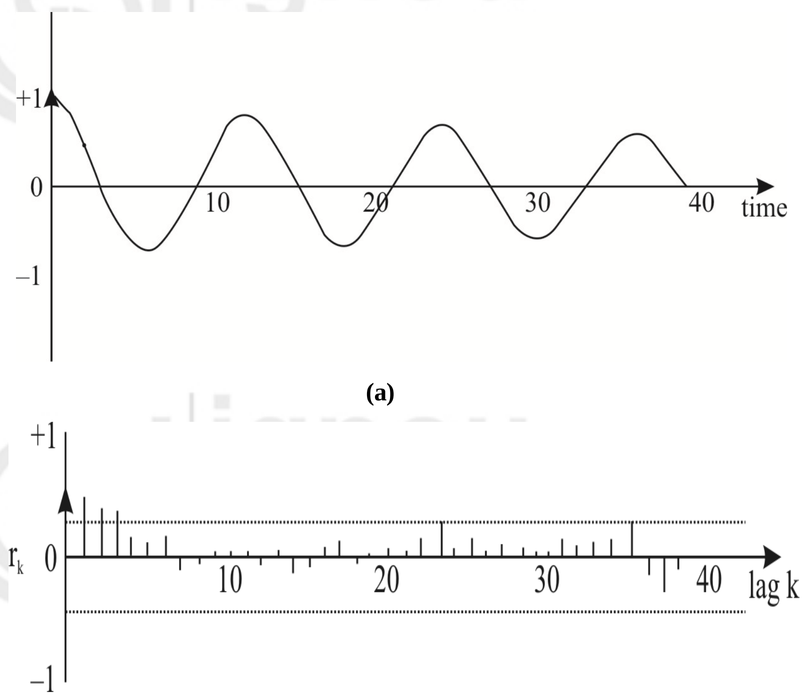


Fig. 15.4: Time plot of the average rainfall at a certain place, in successive months from 2010 to 2012.

Therefore, in this case the correlogram may not contain any more information than what is given by the time plot of the time series.



(b)

Fig. 15.5: a) Smoothed plot of the average rainfall at a certain place, in successive months from 2010 to 2012; b) correlogram of monthly observations of seasonal time series.

Fig. 15.5a shows a time plot of monthly rainfall and Fig. 15.5b shows the correlogram. Both show a cyclical pattern and the presence of a strong 12 monthly seasonal effect. However, it is doubtful that in such cases the correlogram gives any more information about the presence of seasonal effect as compared to the time plot shown in Fig 15.4.

In general, the interpretation of correlogram is not easy and requires a lot of experience and insight. Estimated autocorrelations (r_k) are subject to sampling fluctuations and if N is small, their variances are large. We shall discuss this in more detail when we consider a particular process. When all the population autocorrelations ρ_k ($k \neq 0$) are zero, as happens in a random series, then the values of r_k are approximately distributed as N(0,1/N). This is a very good guide for testing whether the population correlations are all zeros or not, i.e., the process is completely random or not.

Example 2: For the time series given in Example 1, calculate r_1, r_2, r_3, r_4 and r_5 and plot a correlogram.

Solution: From Example 1 and its results we have the following:

$$\bar{y} = 51.0, \quad c_0 = 189.6, \; c_1 = -166.33 \text{ and } r_1 = -0.88$$

Now we form the table for the calculations as follows:

S. No.	Y	Y^2	$(Y_t - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+2} - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+3} - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+4} - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+5} - \bar{Y})$
1	47	2209	−4				
2	64	4096	13				
3	23	529	−28	112			
4	71	5041	20	260	−80		
5	38	1444	−13	364	−169	52	
6	64	4096	13	260	−364	169	−52
7	55	3025	4	−52	80	−112	52
8	41	1681	−10	−130	130	−200	280
9	59	3481	8	32	104	−104	160
10	48	2304	−3	30	−12	−39	39
Total	**510**	**27906**		**876**	**−311**	**−234**	**479**

We now calculate the autocorrelation coefficients r_2, r_3, r_4 and r_5 as follows:

For k = 2, we get

$$c_2 = \sum_{t=1}^{8} \frac{(y_t - \bar{y})(y_{t+2} - \bar{y})}{8}$$

$$= 876/8 = 109.5$$

$$r_2 = \frac{c_2}{c_0} = 109.5/\,189.6 = 0.58$$

For k = 3, we get

$$c_3 = \frac{\sum_{t=1}^{7} (y_t - \bar{y})(y_{t+3} - \bar{y})}{7}$$

$= –311/7 = –44.43$

$$r_3 = \frac{c_3}{c_0} = -44.43/\ 189.6 = -0.2343$$

For k = 4, we get

$$c_4 = \frac{\sum_{t=1}^{6}(y_t - \bar{y})(y_{t+4} - \bar{y})}{6}$$

$$= -234/6 = -39$$

$$r_4 = \frac{c_4}{c_0} = -39/\ 189.6 = -0.2057$$

For k = 5, we get

$$c_5 = \frac{\sum_{t=1}^{5}(y_t - \bar{y})(y_{t+5} - \bar{y})}{5}$$

$$= 479/5 = 95.8$$

$$r_5 = \frac{c_5}{c_0} = 95.8/\ 189.6 = -0.5052$$

Thus, we have obtained the autocorrelation coefficients r_1, r_2, r_3, r_4 and r_5 as $r_1 = –0.88$, $r_2 = 0.58$, $r_3 = -0.2343$, $r_4 = -0.2057$, $r_5 = -0.5052$, respectively.

Now we plot the correlogram for the given time series by plotting the values of the autocorrelation coefficients versus the lag k for k = 1, 2, …, 5. The correlogram is shown in Fig. 15.6.

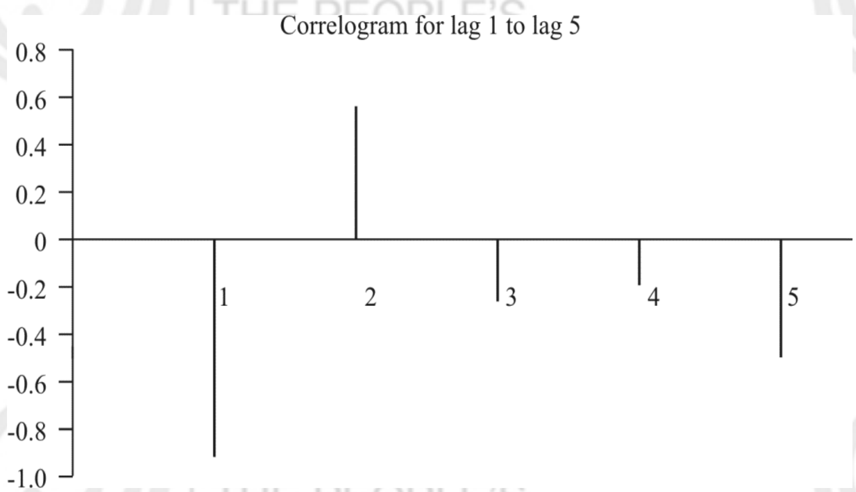


Fig. 15.6: Correlogram for the given time series.

Example 3: A computer generates a series of 200 observations that are supposed to be random. The first 10 sample autocorrelation coefficients of the series are:

$r_1 = 0.02$, $r_2 = 0.05$, $r_3 = –0.09$, $r_4 = 0.08$, $r_5 = –0.02$, $r_6 = –0.07$,
$r_7 = 0.12$, $r_8 = 0.06$, $r_9 = 0.02$, $r_{10} = –0.08$

Plot the correlogram.

Solution: The correlogram for the given values of autocorrelation coefficients is shown in Fig. 15.7.

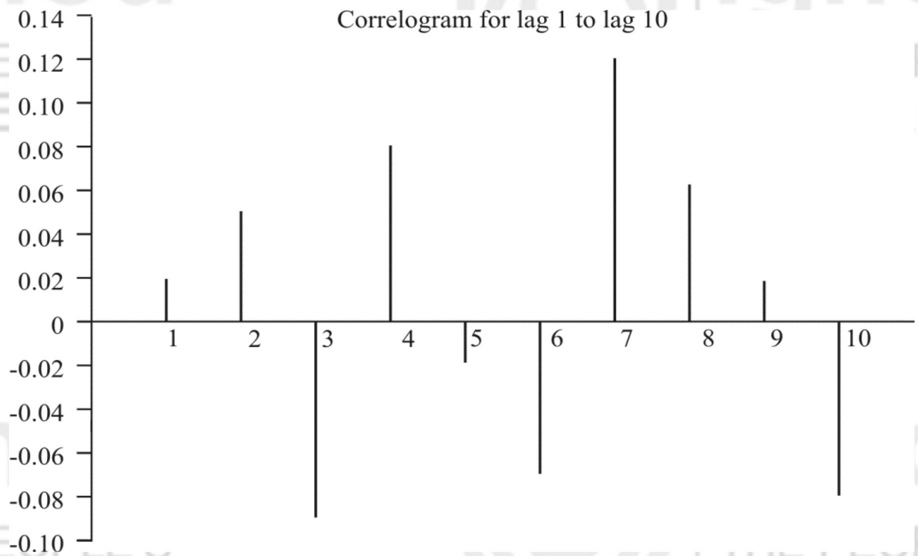


Fig. 15.7: Correlogram for 10 sample autocorrelation coefficients of the series of 200 observations.

Example 4: A random walk (S_t, t = 0, 1, 2, …) starting at zero is obtained by cumulative sum of independently and identically distributed (i.i.d) random variables. Check whether the series is stationary or non-stationary.

Solution: Since we have a random walk (S_t, t = 0, 1, 2, …) starting at zero obtained from cumulative sum of independently and identically distributed (i.i.d) random variables, a random walk with zero mean is obtained by defining $S_0 = 0$ and

$$S_t = Y_1 + Y_2 + \ldots + Y_t, \qquad \text{for } t = 1, 2, \ldots$$

where $\{Y_t\}$ is i.i.d. noise with mean zero and variance σ^2. Then we have

$$E(S_t) = 0, \quad E(S_t^2) = t\sigma^2 < \infty \quad \text{for all } t$$

$$\text{Cov}(S_t, S_{t+h}) = (S_t, S_t + Y_{t+1} + \ldots + Y_{t+h})$$

$$= \text{Cov}(S_t, S_t) = t\,\sigma^2$$

This depends upon t and hence the series S_t is non-stationary.

You may now like to solve the following exercises to assess your understanding about correlogram and stationary processes.

E3) Calculate r_2, r_3, r_4 and r_5 for the time series given in Exercise 1 and plot a correlogram.

E4) Calculate r_2, r_3, r_4 and r_5 for the time series given in Exercise 2 and plot a correlogram.

E5) A computer generates a series of 500 observations that are supposed to be random. The first 10 sample autocorrelation coefficients of the series are:

$r_1 = 0.09$, $r_2 = -0.08$, $r_3 = 0.07$, $r_4 = -0.06$, $r_5 = -0.05$, $r_6 = 0.04$, $r_7 = -0.3$, $r_8 = 0.02$, $r_9 = -0.02$, $r_{10} = -0.01$

Plot the correlogram.

Let us now summarise the concepts that we have discussed in this unit.

15.5 SUMMARY

1. A time series is said to be **stationary** if there is no systematic change in mean, variance and covariance of the observations over a period of time. If a time series is stationary, we can model it and draw further inferences and make forecasts. If a time series is not stationary, we cannot do any further analysis and make reliable forecasts. If a time series shows a particular type of non-stationarity and some simple transformations can make it stationary, then we can model it.

2. A **random process** is a statistical phenomenon that evolves in time according to some laws of probability. Mathematically, a random process is defined as the family of random variables which are ordered in time, i.e., a random process {Y(t); t belongs to T} is a collection of random variables, where T is a set for which all the random variables Y_t are defined on the same sample space.

3. A random process is said to be stationary if the joint distribution of Y_{t1}, Y_{t2}, Y_{t3},…, Y_{tk} is the same as the joint distribution of Y_{t1+J}, Y_{t2+J}, Y_{t3+J}, ..., Y_{tk+J}, for all $t_{1,}$ $t_{2, ...,}$ t_k and J. In other words, shifting the origin of time by an amount J has no effect on the joint distribution. This means that it depends only on the interval between t_1, t_2, …, t_k.

4. If all the finite dimensional distributions of a random process are invariant under the translation of time, it is called a **Strict Sense Stationary process** or **SSS process**. In other words, if the joint distribution of Y_{t1}, $Y_{t2,}$ …, Y_{tk} is the same as the joint distribution of Y_{t1+J}, Y_{t2+J},..., Y_{tk+J}, for all $t_{1,}$ $t_{2, ...,}$ t_k and J (> 0) for all k ≥ 1, the random process Y_t is called a Strict Sense Stationary process.

5. A **stationary process is said to have weak stationarity** of order m, if the moments up to order m depend only on the time lag J. If m = 2, the stationarity (or weak stationarity) of order 2 implies that moments up to the second order depend only on the time lag J.

6. Of particular interest in the analysis of a time series are the covariance and correlation between Y_{t1} and Y_{t2}. Since these are the covariance and correlation within the same time series, they are called **autocovariance** and **autocorrelation**. Some important properties of time series can be studied with the help of autocovariance and autocorrelation. They measure the linear relationship between observations at different time lags. They provide useful descriptive properties of the time series being studied and are important tools for guessing a suitable model for the time series data.

7. A useful plot for interpreting a set of autocorrelation coefficients is called a **correlogram** in which the sample autocorrelation coefficients r_k are plotted versus the lag J where J=1, 2, 3, …, k. This helps us in examining the nature of time series. It is also a very important diagnostic tool for selection of a suitable model for the process which generates the data. The correlogram is also known as the sample autocorrelation function (acf).

8. In most time series, it is noticed that the absolute value of r_k, i.e., | r_k| decreases as k increases. This is because observations which are located far away are not related to each other, whereas observations lying closer to each other may be positively (or negatively) correlated.

15.6 SOLUTIONS/ANSWERS

E1) We first plot the stationary time series values as shown in Fig. 15.8.

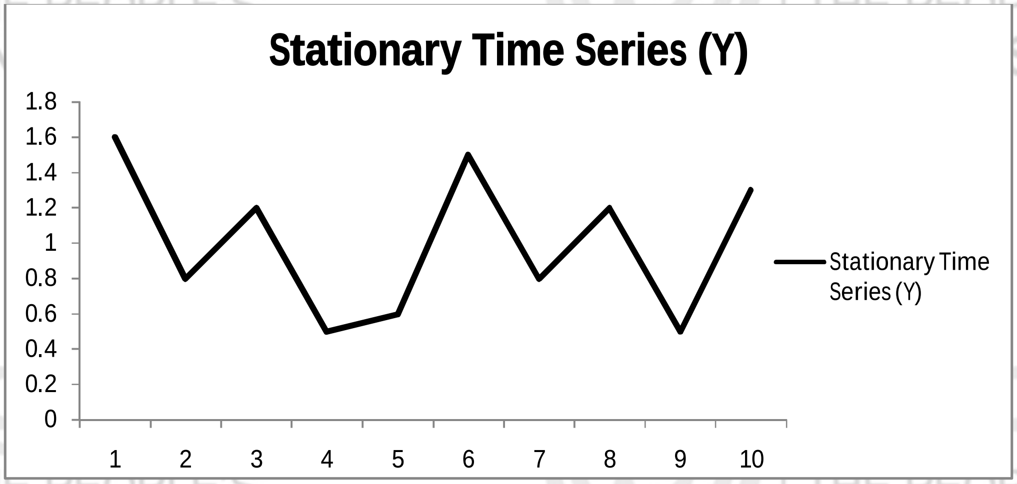


Fig. 15.8: Plot of the given stationary time series.

For calculating r_1 we first calculate the terms given in the table below:

S. No.	Y	Y^2	$(Y_t - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+1} - \bar{Y})$
1	1.6	2.56	0.6	
2	0.8	0.64	−0.2	−0.12
3	1.2	1.44	0.2	−0.04
4	0.5	0.25	−0.5	−0.1
5	0.6	0.36	−0.4	0.2
6	1.5	2.25	0.5	−0.2
7	0.8	0.64	−0.2	−0.1
8	1.2	1.44	0.2	−0.04
9	0.5	0.25	−0.5	−0.1
10	1.3	1.69	0.3	−0.15
Total	10	11.52		−0.65

$$\bar{y} = \sum_{i=1}^{N} y_i / 10 = 10/10 = 1.0$$

$$c_0 = \sum_{t=1}^{N-k} \frac{(y_i - \bar{y})^2}{N} = \frac{\left(\sum y_i^2 - N\bar{y}^2\right)}{N}$$

$$= (11.52 - 10)/10 = 0.152$$

$$c_1 = \frac{\sum_{t=1}^{N-1} (y_t - \bar{y})(y_{t+1} - \bar{y})}{N-1} = \frac{-0.65}{9} = -0.072$$

$$r_1 = \frac{c_1}{c_0} = -0.072/0.152 = -0.475$$

E2) We first plot the stationary time series values as shown in Fig. 15.9.

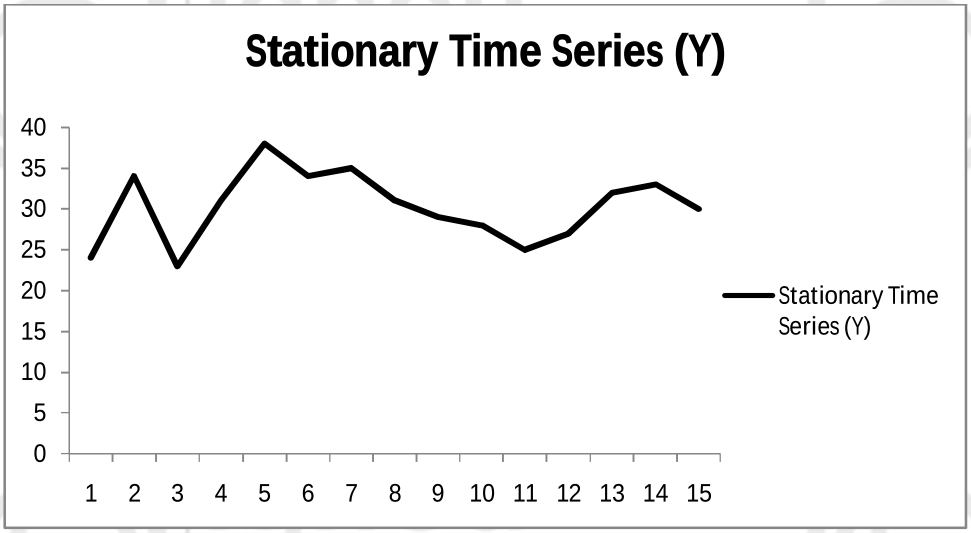


Fig. 15.9: Plot of the given stationary time series.

For the given time series we do the following calculations:

S. No.	Y	Y^2	$(Y_t - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+1} - \bar{Y})$
1	24	576	−6.267	
2	34	1156	3.733	−23.395
3	23	529	−7.267	−27.129
4	31	961	0.733	−5.329
5	38	1444	7.733	5.671
6	34	1156	3.733	28.871
7	35	1225	4.733	17.671
8	31	961	0.733	3.471
9	29	841	−1.267	−0.929
10	28	784	−2.267	2.871
11	25	625	−5.267	11.938
12	27	729	−3.267	17.204
13	32	1024	1.733	−5.662
14	33	1089	2.733	4.738
15	30	900	−0.267	−0.729
Total	**454**	**14000**		**1.773**

$$\bar{y} = \sum_{i=1}^{N} y_i / 15 = 454/15 = 30.267,$$

$$c_0 = \sum_{t=1}^{N-k} \frac{(y_i - \bar{y})^2}{N} = \frac{\left(\sum y_i^2 - N\bar{y}^2\right)}{N}$$

$$= (14000 - 13922.82)/15 = 5.145$$

$$c_1 = \frac{\sum_{t=1}^{N-1} (y_t - \bar{y})(y_{t+1} - \bar{y})}{N - 1}$$

$$= \frac{1.773}{14} = 0.1266$$

$$r_1 = \frac{c_1}{c_0} = 0.1266/5.145 = 0.0246$$

E3) In Exercise 1, we have obtained the following values:

$\bar{y} = 1.0$, c_0 = 0.152, $c_1 = -0.072$ and $r_1 = -0.475$

S. No.	Y	Y^2	$(Y_t - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+2} - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+3} - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+4} - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+5} - \bar{Y})$
1	1.6	2.56	0.60				
2	0.8	0.64	−0.2				
3	1.2	1.44	0.20	0.12			
4	0.5	0.25	−0.50	0.10	−0.30		
5	0.6	0.36	−0.40	−0.08	0.08	−0.24	
6	1.5	2.25	0.50	−0.25	0.10	−0.10	0.30
7	0.8	0.64	−0.20	0.08	0.10	−0.04	0.04
8	1.2	1.44	0.20	0.10	−0.08	−0.10	0.04
9	0.5	0.25	−0.50	0.10	−0.25	0.20	0.25
10	1.3	1.69	0.30	0.06	−0.06	0.15	−0.12
	10	11.52		0.23	−0.41	−0.13	0.51

We now use the same procedure to calculate the autocorrelation coefficients r_2, r_3, r_4 and r_5 as follows:

For k = 2,

$$c_2 = \sum_{t=1}^{8} \frac{(y_t - \bar{y})(y_{t+2} - \bar{y})}{8}$$

$$= 0.23/8 = 0.02875$$

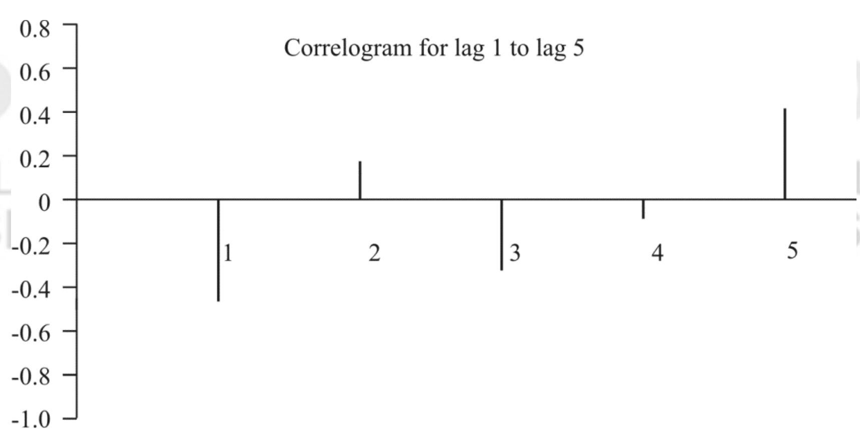
From equation (7),

$$r_2 = \frac{c_2}{c_0} = 0.02875/\ 0.152 = 0.1891$$

Similarly, r_3, r_4 and r_5 are obtained as

$$r_3 = –0.3371,\ \ r_4 = –0.1069,\ \ r_5 = 0.4194$$

Now to plot the correlogram for the given time series, we plot the values of the autocorrelation coefficients versus lag k for all k = 1, 2, …, 5. The correlogram is shown in Fig. 15.10.



E4) In Exercise 2, we have obtained the following values:

$$\bar{y} = 30.267,\ c_0 = 5.145,\ c_1 = 0.1266 \text{ and } r_1 = 0.0246$$

For the given time series, we do the following calculations:

S. No.	Y	Y^2	$(Y_t - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+2} - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+3} - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+4} - \bar{Y})$	$(Y_t - \bar{Y})(Y_{t+5} - \bar{Y})$
1	24	576	−6.267				
2	34	1156	3.733				
3	23	529	−7.267	45.538			
4	31	961	0.733	2.738	−4.595		
5	38	1444	7.733	−56.195	28.871	−48.462	
6	34	1156	3.733	2.738	−27.129	13.938	−23.395
7	35	1225	4.733	36.604	3.471	−34.395	17.671
8	31	961	0.733	2.738	5.671	0.538	−5.329
9	29	841	−1.267	−5.995	−4.729	−9.795	−0.929
10	28	784	−2.267	−1.662	−10.729	−8.462	−17.529
11	25	625	−5.267	6.671	−3.862	−24.929	−19.662
12	27	729	−3.267	7.404	4.138	−2.395	−15.462
13	32	1024	1.733	−9.129	−3.929	−2.195	1.271
14	33	1089	2.733	−8.929	−14.395	−6.195	−3.462
15	30	900	−0.267	−0.462	0.871	1.404	0.604
Total	454	14000		26.502	−9.169	−86.64	−29.511

We now calculate the autocorrelation coefficients r_2, r_3, r_4 and r_5 as follows:

For k=2,

$$c_2 = \sum_{t=1}^{14} \frac{(y_t - \bar{y})(y_{t+2} - \bar{y})}{13} = 26.502/13 = 2.0386$$

From equation (7),

$$r_2 = \frac{c_2}{c_0} = 2.0386/5.145 = 0.3962$$

Similarly, r_3, r_4 and r_5 may also be obtained as

$$r_3 = -0.1485,\ r_4 = -0.72,\qquad r_5 = -0.5736$$

Now to plot the correlogram for the given time series, we plot the values of the autocorrelation coefficients versus the lag k for all k = 1, 2, …, 5. The correlogram is shown in Fig. 15.11

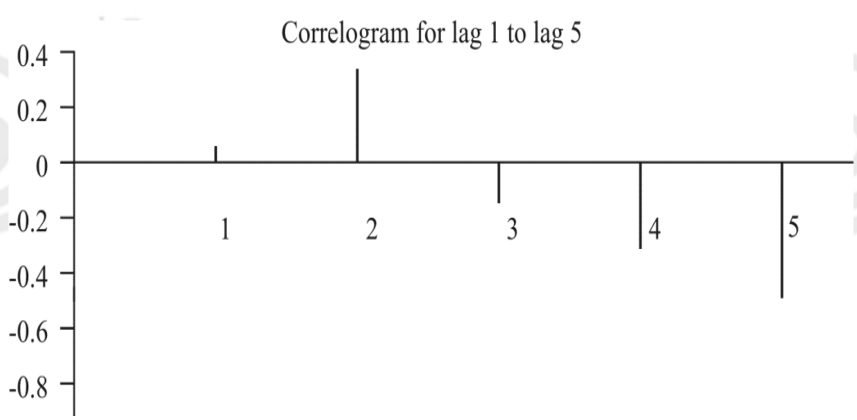


Fig. 15.11: Correlogram for the given time series

E5) The correlogram for the given values of autocorrelation coefficients of a stationary process is shown in Fig. 15.12.

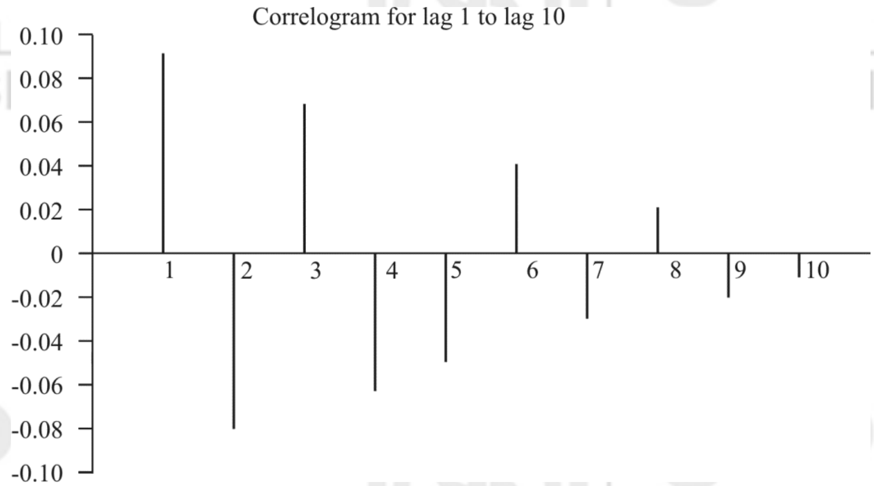


Fig. 15.12: Correlogram for 10 sample autocorrelation coefficients of the series of 400 observations.