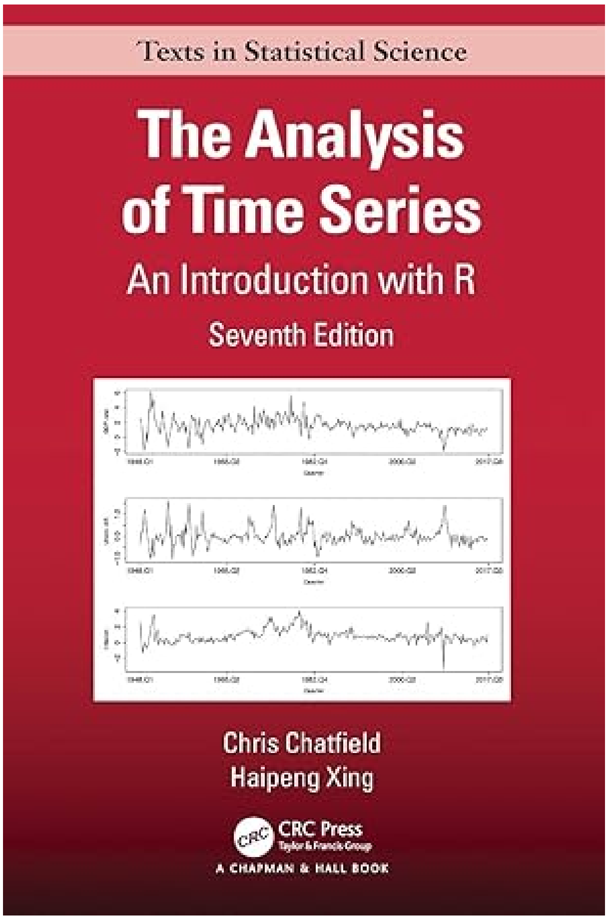
Notes for The Analysis of Time Series Seventh Edition by Chris Chatfield & Haipeng Xin

Gallant Tsao

August 8, 2025



Contents

1 Introduction

1.1 Some Representative Time Series

We begin with some examples of the sort of time series that arise in practice.

Economic and financial time series

Many time series are routinely recorded in economics and finance: Share prices on succesive days, export totals in successive months, average incomes in successive months, company profits in successive years, and so on.

The classic Beveridge wheat price index series consists of the average wheat price in nearly 50 places in various countries measured in successive years from 1500 to 1869. We can plot the series via

```
> library(tseries) # load the library
> data(bev) # load the dataset
> plot(bev, xlab="Year", ylab="Wheat price index", xaxt="n")
> x.pos = c(1500, 1560, 1620, 1680, 1740, 1800, 1860) # define x-axis labels
> axis(1, x.pos, x.pos)
```

Fig. 1.1 shows this series and some apparent cyclic behavior can be seen.

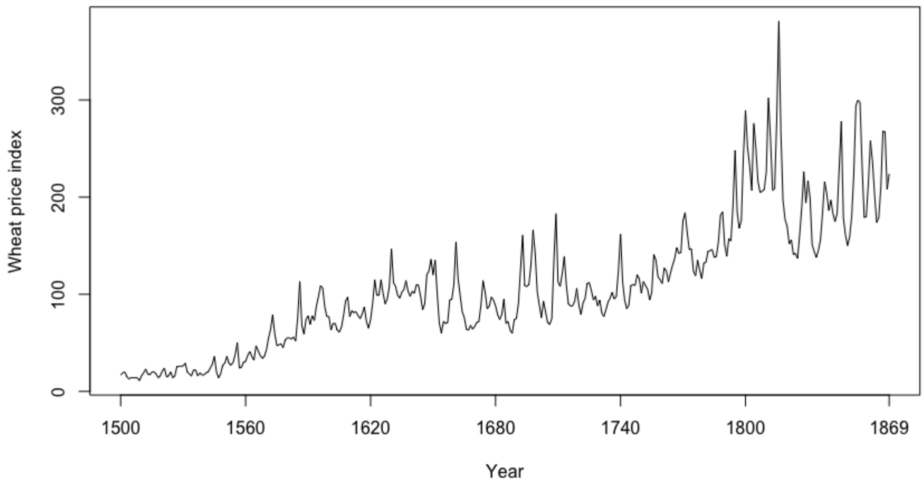


Figure 1.1: The Beverage wheat price annual index series from 1500 to 1869.

There is also an example of financial time series: Fig. 1.2 below hsows the daily returns (or percentage change) of the adjusted closing prices of the Standard & Poor's 500 (S&P) Index from January 4th, 1995 to December 30th, 2016.

To reproduce Fig. 1.2 in R, suppose we have the data as sp500_ret_1995–2016.csv in the directory mydata. Then we can load the data via the following piece of code:

```
> sp500<-read.csv("mydata/sp500\_ret\_1995-2016.csv")
> n<-nrow(sp500)
> x.pos<-c(seq(1,n,800),n)
> plot(sp500\$Return, type="l", xlab="Day", ylab="Daily return", xaxt="n")
> axis(1, x.pos, sp500\$Date[x.pos])
```

Physical time series Many types of times series occur in the physical sciences, particularly in meteorology, marine sciences, and geophysics. Examples are rainfall on succesive days, and air temperature measured in successive hours, days, or months. Fig. 1.3 shows the average air temperature in Anchorage, Alaska in the US in successive months over a 16 year time period. The series can be downloaded from the

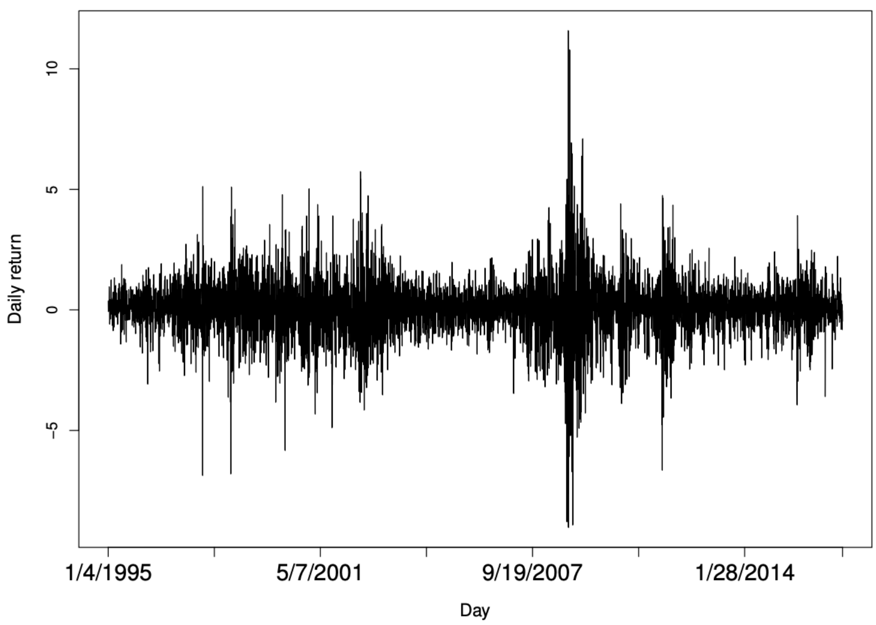


Figure 1.2: Daily returns of the adjusted closing proces of the S&P index from January 4th, 1995 to December 40th, 2016.

U.S. National Centers for Environmental Information. Seasonal fluctuations can be clearly seen in the series.

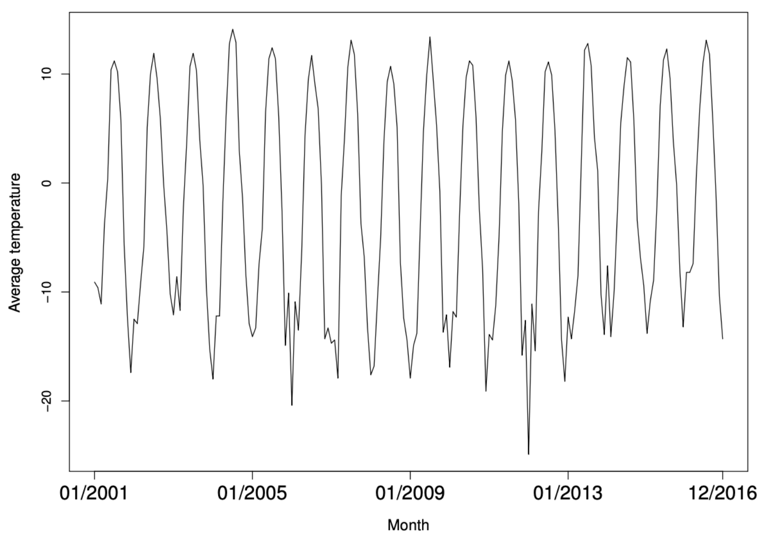


Figure 1.3: Monthly average air temperature (degrees Celcius) in Anchorage, Alaska, in successive months from 2001 to 2016.

Some mechanical recorders take measurements continuously and produce a continuous trace rather than observations at discrete intervals of time. However, for more detailed analysis, it is customary to convert the continuous trace to a series in discrete time by sampling the trace at appropriate intervals of time. The resulting analysis is more straightforward and can be handled by standard time series software.

Marketing time series

The analysis of time series arising in marketing is an important problem in commerce. Observed variables could include sales figures in successive weeks or months, monetary receipts, advertising costs and so on. As an example, Fig. 1.4 shows the domestic sales of Australian fortified wine by winemakers in successive quarters over a 30-year period, which are available at the Australian Bureau of Statistics.

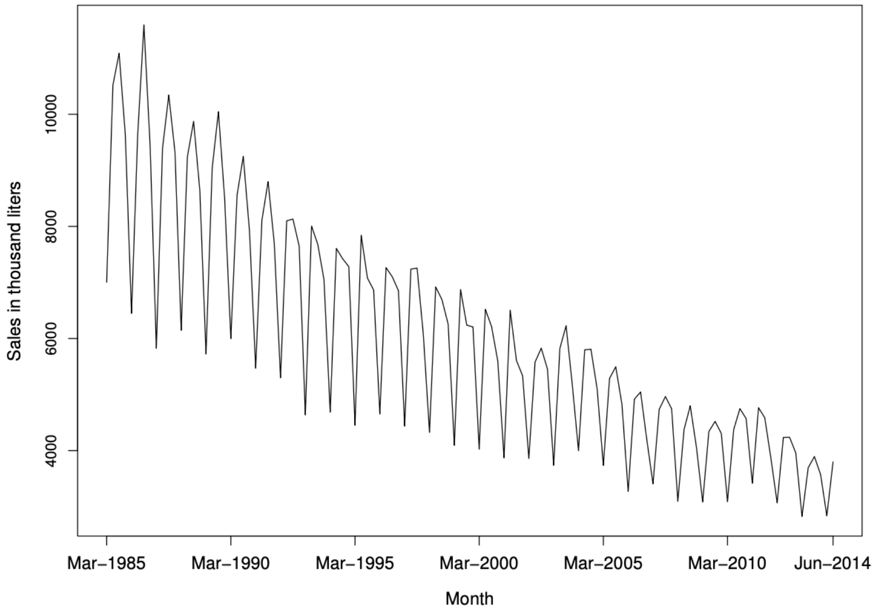


Figure 1.4: Domestic sales (unit: thousand liters) of Australian fortified wine by winemakers in successive quarters from March 1985 to June 2014.

Note the trend and seasonal variation above.

Demographic time series

Various time series occur in the study of population change. Examples include the total population of Canada measured annually, and monthly birth totals in England. Fig. 1.5 shows the total population and crude birth rate (per 1000 people) for the UNited States from 1965 to 2015. The data are available at the US Federal Reserve Bank of St. Louis.

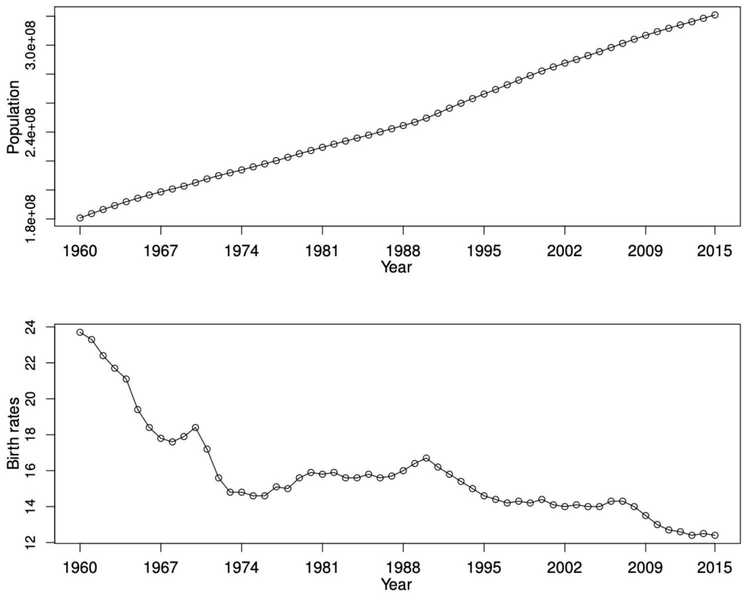


Figure 1.5: Total population and birth rate (per 1000 people) for the United States from 1965 to 2015.

We can reproduce Fig. 1.5 using the following piece of code:

```
> pop<-read.csv("mydata/US\_pop\_birthrate.csv", header=T)
> x.pos<-c(seq(1, 56, 7), 56)
> x.label<-c(seq(1960, 2009, by=7), 2015)
```

```
> par(mfrow=c(2,1), mar=c(3,4,3,4))
> plot(pop[,2], type="l", xlab="", ylab="", xaxt="n")
> points(pop[,2])
> axis(1, x.pos, x.label, cex.axis=1.2)
> title(xlab="Year", ylab="Population", line=2, cex.lab=1.2)

> plot(pop[,3], type="l", xlab="", ylab="", xaxt="n")
> points(pop[,3])
> axis(1, x.pos, x.label, cex.axis=1.2)
> title(xlab="Year", ylab="Birth rates", line=2, cex.lab=1.2)
```

Process control data

In process control, a problem is to detect changes in the performance of a manufacturing process by measuring a variable, which shows the quality of the process. These measurements can be plotted against time, as shown in Fig. 1.6 below.

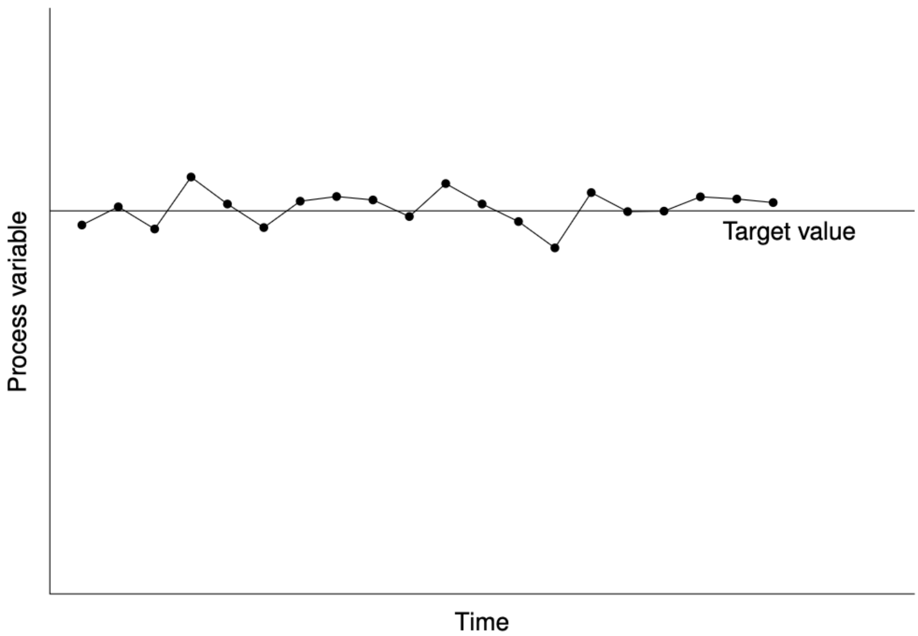


Figure 1.6: A process control chart.

Special techniques have been developed for this type of time series problems, namely statistical quality control (Montgomery, 1996).

Binary processes A special type of time series arisis when observations can only take one of only two values, usually denoted by 0 and 1 (see Fig. 1.7). For example, in computer science, the position of a switch could be recorded as 1 or 0 respectively.

Time series of this type occur in many situations, including the study of communication theory. A problem here is to predict when the process will take a different value.

Point processes We can consider a series of events occuring 'randomly' through time, such as the dates of major railway disasters. A series of events of this type is usualled called a point process. As an example, Fig. 1.8 shows the intraday transaction data of the International Business Machines (IBM) Corporation from 9:35:00 to 9:38:00 on January 4th, 2010. When a trade event occurs, the corresponding trading price and trading volume are observed.

Methods of analyzing point process data are generally very different from those used for analyzing standard time series data. The text by Cox and Isham (1980) is recommended.

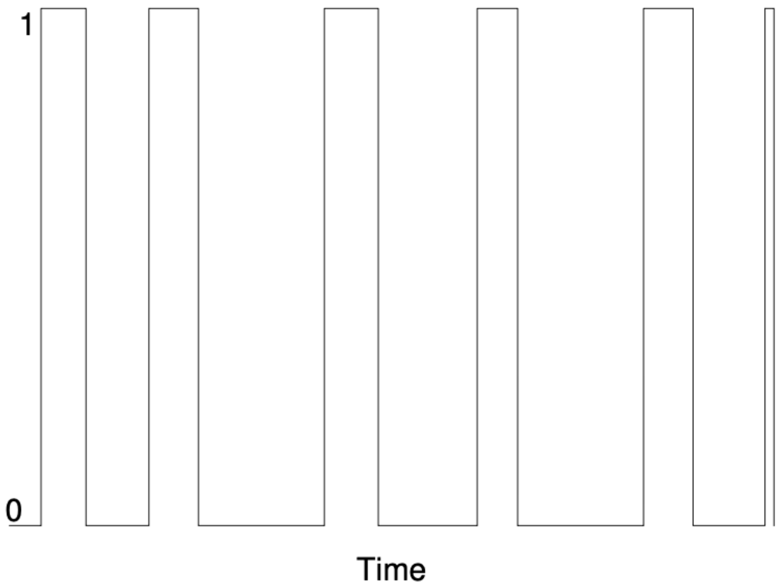


Figure 1.7: A realization of a binary process.

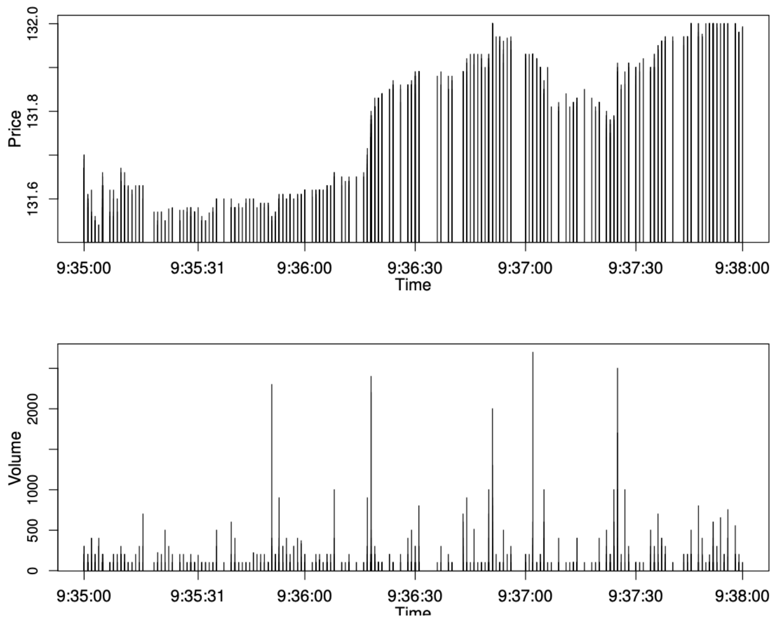


Figure 1.8: Transaction prices and volumes of the IBM stocks from 9:35:00 to 9:38:00 on January 4th, 2010.

To reproduce Fig. 1.8, we can use the following piece of code:

```
> ibm<-read.table("mydata/taq\_trade\_ibm\_100104.txt",
header=T, sep="\t")
> ibm.new<-ibm[,c(1,2,7)]
> ibm[,2]<-as.numeric(as.character(ibm[,2]))

> ### take 9:35:00-9:37:59am trading record
> data<-ibm.new[1458:2371,]
> newtime<-rep(0, nrow(data))
> for (i in 1:nrow(data)){
  min<-as.numeric(substr(as.character(data\$TIME[i]),3,4))
  sec<-as.numeric(substr(as.character(data\$TIME[i]),6,7))
  newtime[i]<- (min-30)*60+sec
}

> x.label<-c("9:35:00", "9:35:31", "9:36:00", "9:36:30",
"9:37:00", "9:37:30", "9:38:00")
> x.pos<-c(1, 139, 249, 485, 619, 776, 914)
> par(mfrow=c(2,1), mar=c(2,4,2,4))

> plot(newtime, data[,2],xlab="",ylab="",xaxt="n",type="h")
> axis(1, newtime[x.pos], x.label, cex.axis=1.2)
> title(xlab="Time", ylab="Price", line=2, cex.lab=1.2)
> plot(newtime, data[,3],xlab="",ylab="",xaxt="n",type="h")
> axis(1, newtime[x.pos], x.label, cex.axis=1.2)
> title(xlab="Time", ylab="Volume", line=2, cex.lab=1.2)
```

1.2 Terminology

Definition. A time series is said to be continuous when observations are made continuously through time.
A time series is said to be discrete when observations are taken only at specific times, usually equally spaced.

Note that the term 'discrete' is used for series of this type even when the measured variable is continuous. This book mostly concerns with discrete time series.

Discrete time series can arise in several ways. Given a continuous time series, we can read off/digitize the values at equal intervals of time to give a discrete time series, sometimes called sampled series. A different type arises when a variable does not have an instantaneous value but we can aggregate/accumulate the values over equal intervals of time (like monthly exports). Finally, some time series are inherently discrete (dividend paid by a company to shareholders in successive years).

Important: Most statistical theory relies on samples of independent observations. This is not true with time series! Successive observations are usually *not* independent hence we must take the *time order* into account when doing the analysis. If a time series can be predicted exactly, it is said to be deterministic. However, most time series are stochastic, hence future values have a probability distribution that depend on the past.

1.3 Objectives of Time Series Analysis

There are several objectives of analyzing a time series:

(i) Description

When presented with a time series, the firrst step is usually get the time plot from the data. For simpler time series, this is useful, like in Fig. 1.4, where we can see a regular seasonal effect. For more complex models, more sophisticated techniques are required.

The book devotes a greater amount of space to the more advanced techniques, but this does not mean the elementary descriptive techniques are not important. These are very important when dealing with outliers, which is a very complex subject in time series analysis.

Other features to look for in a time plot include sudden or gradual changes in the properties of the series, in which case we may or may not need to use piecewise model to fit parts of the time series one at a time.

(ii) Explanation

When observations are taken on two or more variables, it may be possible to use the variation in one time series to explain the variation in another series. This may lead to a deeper understanding of the mechanism that generated a given time series.

When dealing with many factors, regression is usually not too helpful in handling time series data. We'll use something called a linear system, which will be discussed in Chapter 9, which converts an input series to an output series via a linear operator. Given observations on the input and output to a linear system (see Fig. 1.9), the analyst wants to assess the properties of the linear system. For example, it is of interest to see how sea level is affected by temperature and pressure, and to see how sales are affected by price and economic conditions. A class of models, called transfer function models, enables us to model time series data in an appropriate way.

(iii) Prediction

Figure 1.9: Schematic representation of a linear system.

Given a time series, we may want to predict future values based on the current and past values, for example in sales forecasting. Here we are doing 'prediction' and 'forecasting'.

(iv) Control

Time series are sometimes collected to improve control over some physical or economic system. For example, if a time series is used to keep track of the quality of manufacturing, we would want it to always be at the 'high' level. Control problems are related to predictions, as we can use predictions for error correction.

This topic is only briefly touched on in Section 14.3.

1.4 Approaches to Time Series Analysis

This book describes various approaches to time series:

- Chapter 2: Simple descriptive techniques (plotting, trend and seasonality, etc.)
- Chapter 3: Probability models for time series
- Chapter 4: Fitting models to time series
- Chapter 5: Forecasting procedures
- Chapter 6: Spectral density function
- Chapter 7: Using spectral analysis to estimate the spectral density function
- Chapter 8: Analysis of bivariate time series
- Chapter 9: Using linear systems
- Chapter 10: State-space models + Kalman filter

- Chapter 11: Nonlinear time series analysis
- Chapter 12: Volatility time series model analysis
- Chapter 13: Multivariate time series model analysis
- Chapter 14: More advanced stuff!

1.5 Review of Books on Time Series

This subsections simply concerns other books on time series that may be helpful. The complete list is here:

Introductory texts

- *Introduction to Time Series and Forecasting Third Edition* by Brockwell & Davis
- *Time Series Models Second Edition* by Harvey
- *Time Series Third Edition* by Kendall & Ord
- *Time Series: A Biostatistical Introduction* by Diggle
- *Time Series Analysis: Univariate and Multivariate Models Second Edition* by Wei
- *Applied Econometric Time Series* by Enders
- *Applied Time Series Analysis: A Practical Guide to Modeling and Forecasting* by Mills
- *Applied Time Series Modelling and Forecasting* by Harris & Sollis

Advanced Texts

- *The Statistical Analysis of Time Series* by Anderson
- *Time Series Theory and Methods* by Brockwell & Davis
- *Introduction to Statistical Time Series Second Edition* by Fuller
- *Spectral Analysis and Time Series I & II* by Priestley
- *Time Series: Data Analysis and Theory* by Brillinger
- *Time Series Fourth Edition* by Kendall
- *Time Series Analysis* by Hamilton
- *Analysis of Financial Time Series Third Edition* by Tsay

Texts with statistical computing languages (R or S-Plus)

- *Introductory Time Series with R* by Cowpertwit & Metcalfe
- *Time Series Analysis and Its Applications Fourth Edition* by Shomway & Stoffer
- *Analysis of Financial Time Series Third Edition* by Tsay

Additional books may be referenced in later chapters.

2 Basic Descriptive Techniques

Statistical techniques for analyzing time series vary from relatively straightforward descriptive techniques to sophisticated inferential techniques. This chapters introduces the former. Descriptive techniques should be tried before attempting more complicated procedures, because they are important in 'cleaning' the data, and then getting a 'feel' for them, before trying to generate ideas as regards to a suitable model.

This chapter focuses on ways of understanding typical time-series effects, such as trend, seasonality, and correlations betyween successive observations.

2.1 Types of Variation

Traditional methods of time-series analysis are mainly concerned with decomposing the variation in a series into components representing trend, seasonal variation and other cyclic changes. Any remaining variation is attributed to 'irregular' fluctuations. This approach is not always the best but is particularly valuable when the variation is dominated by trend and seasonality.

Seasonal variation Many time series, such as sales figures and temperature readings, exhibit variation that is annual in period. For example, unemployment is typically 'high' in winter but low in summer. This yearly variation is easy to understand, and can readily be estimated if seasonality is of direct interest. Alternatively, seasonal variation can be removed from the data to give deseasonalized data.

Other cyclic variations Apart from seasonal effects, there are some other variations of time series at a fixed period due to some other physical cause. For example, daily variations in temperature depend on what time of the day it is.

Trend This may be loosely defined as 'long-term change in the mean level'. However, this largely depends how we define the term 'long-term'. It can be a year to even 50 years depending on the span of the time series.

Other irregular fluctutations After trend and cyclic variations have been removed from a set of data, we are left with residuals that may or may not be 'random'. We'll examine later whether this can be explained in terms of probability models, such as the moving average (MA) or autoregressive (AR) models.

2.2 Stationary Time Series

Broadly speaking a time series is said to be stationary if there is no systematic change in mean (no trend), if there is no systematic change in variance and if strictly periodic variations have been removed. In other words, the properties of one section of the data are much like those of any ohter section.

Much of the probability theory of time series is concerned with stationary time series, and for this reason time series analysis often requires one to transform a non-stationary time series to a stationary one so as to use the theory. For example, we can remove the trend and seasonality and model the resuals by a stationary stochastic process. However, sometimes we may be more interested in the non-stationary components too!

2.3 The Time Plot

Ths first, and most important, step in any time-series analysis is to plot the observations against time. This grap, called the time plot, will show up important features of the series such as trend, seasonality, outliers and discontinuities.

Plotting a time series is not as easy as is sounds. The choice of scales, siz eof intercept, way that points are plotted may substantially affect the way the plot looks. Not all computer software plots the time series well, so we'll have to adjust manually sometimes.

2.4 Transformations

Plotting the data may suggest that it is reasonable to transform them by, for example, taking logs or square roots. The three main reasons for making a transformation are as follows:

(i) To stabilize the variance

If there is a trend in the series and the variance appears to increase with the mean, they it may be advisable to transform the data. In particular, if the standard deviation is directly proportional to the mean, a logarithmic transformation is indicated. On the other hand, if the variance changes through time without a trend being present, then a transformation will not help.

(ii) To make the seasonal effect additive

If there is a trend in the series and the size of the seasonal effect appears to increase with the mean, then it may be advisable to transform the data so as to make the seasonal effect constant from year to year. The seasonal effect is then said to be additive. In particular, if the size of the seasonal effect is directly proportional to the mean, then the seasonal effect is said to be multiplicative and a logarithmic transformation is appropriate to make the effect additive. However, this transformation will only stabilize the variance if the error term is also thought to be multiplicative (see Section 2.6), a point that is sometimes overlooked.

(iii) To make the data normally distributed

Model building and forecasting are usually carried out on the assumption that the data are normally distributed. In practice this is not necessarily the case; there may, for example, be evidence of skewness in that there tend to be 'spikes' in the time plot that are all in the same direction (either up or down). This effect can be diffcult to eliminate with a transformation and it may be necessary to model the data using a different 'error' distribution.

The logarithmic and square root transformations above are special cases of a general class of transformations called the Box-Cox transformation. Given an observed time series $\{x_t\}$ and a transformation parameter λ, the transformed series is given by

$$y_t = \begin{cases} (x_t^\lambda - 1)/\lambda & \text{if } \lambda \neq 0, \\ \log(x_t) & \text{if } \lambda = 0. \end{cases}$$

The best value of λ is usually 'guesstimated'.

However, Nelson and Granger (1979) found that there is little improvement in forecast performance when a general Box–Cox transformation was tried on a number of series. There are cases where the transformation fails to stabilize the variance. Usually, transformations should be avoided wherever possible except where the transformed variable has a direct physical interpretation. For example, when percentage increases are of interest, then taking logarithms makes sense.

2.5 Analyzing Series that Contain a Trend and No Seasonal Variation

The simplest type of trend is the familiar 'linear trend + noise', for which the observation at time t is a random variable X_t given by

$$X_t = \alpha + \beta t + \varepsilon_t,$$

where α, β are constants and ε_t denotes a random error term with zero mean. The mean level at time t is given by $m_t = \alpha + \beta t$; which is some times called the 'trend term'. However, some authors denote β as the trend so it depends on the context.

The trend above is a deterministic function of time and is sometimes called a global linear trend. In practice, this generally provides an unrealistic model, and nowadays there is more emphasis on models that allow for local linear trends. One possibility is to fit a piecewise linear model where the trend line is locally linear but with change points where the slope and intercept change (abruptly). We can also assume that α, β evolve stochastically, giving rise to a stochastic trend.

Now we describe some methods to describing the trend.

2.5.1 Curve Fitting

A traditional method of dealing with non-seasonal data with a trend is to fit a simple function such as a polynomial, a Gompertz curve, or a logistic curve. The Gompertz curve can be written in the form

$$\log x_t = a + br^t$$

where a, b, r are parameters with $0 < r < 1$, or in the alternative form of

$$x_t = \alpha \exp\left[\beta \exp\left(-\gamma t\right)\right],$$

which is equivalent as long as $\gamma > 0$. The logistic curve is given by

$$x_t = a/(1 + be^{-ct}).$$

For curves of this type, the fitted function provides a measure of the trend, and the residuals provide an estimate of local fluctuations.

2.5.2 Filtering

A second procedure for dealing with a trend is by using a linear filter, which converts a time series $\{x_t\}$ into another time series $\{y_t\}$ by the linear operation

$$y_t = \sum_{r=-q}^{s} a_r x_{t+r},$$

where $\{a_r\}$ are a set of weights. In order to smooth out local fluctuations and estimate the local mean, we should clearly choose the weights so that $\Sigma a_r = 1$, and the operation is referred as the moving average.

Moving averages are often symmetric with $s = q$ and $a_j = a_{-j}$. The simplest example of a moving average is a symmetric moving average, defined by

$$\mathrm{Sm}(x_t) = \frac{1}{2q+1} \sum_{r=-q}^{+q} x_{t+r}.$$

The simple moving average is not generally recommended by itself for measuring trend, although it can be useful for removing seasonal variation.

Another example is to take $\{a_r\}$ to be successive terms in the expansion of $(\frac{1}{2} + \frac{1}{2})^{2q}$. As q gets large, the weights approximate a normal curve.

A third example is Spencer's 15-point moving average, which is used for smoothing mortality statistics to get life tables. This covers 15 consecutive points with $q = 7$, and the symmetric weights are

$$\frac{1}{320}[-3, -6, -5, 3, 21, 46, 67, 74, 67, 46, 21, \ldots]$$

A fourth example is the Henderson moving average. This moving average aims to follow a cubic polynomial trend without distortion, and the choice of q depends on the degree of irregularity. The symmetric nine-term moving average, for example, is given by

$$[-0.041, -0.010, 0.119, 0.267, 0.330, \ldots]$$

To demonstrate this effect of moving averages, we use the Beveridge wheat price annual index series from 1500 to 1869 as an example (Fig. 1.1).

The top panel of Fig. 2.1 shows the original time series, while the middle and bottom ones show the smoothed series via the simple moving average and Spencer's 15-point moving average.

To reproduce Fig. 2.1, we can use the following piece of code:

```
> smooth.sym<-function(my.ts, window.q){
    window.size<-2*window.q+1
    my.ts.sm<-rep(0, length(my.ts)-window.size)
    for (i in 1:length(my.ts.sm)) {
      my.ts.sm[i]<-mean(my.ts[i:(i+window.size-1)])
    }
    my.ts.sm
  } # moving average with equal and symmetric weights
```

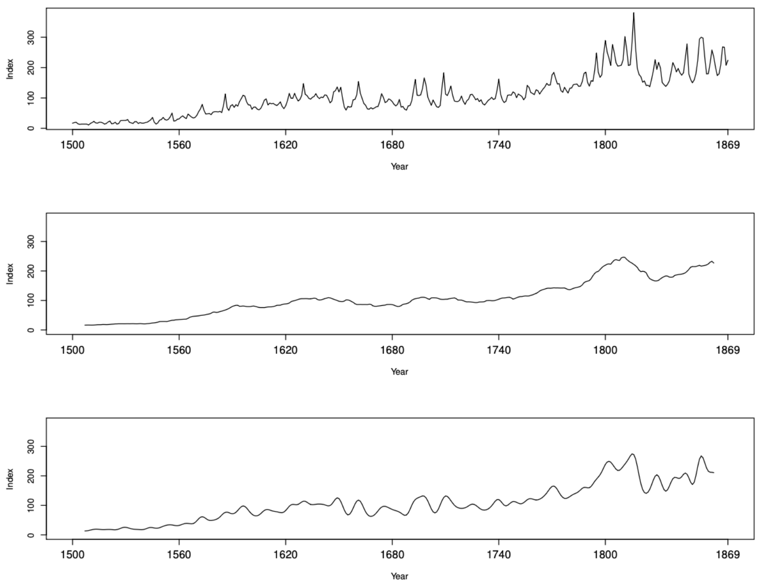


Figure 2.1: The original and smoother Beveridge wheat price annual index series from 1500 to 1869 (Top: The raw data; Middle: Smoothed series with simple moving average and $q = 7$; Bottom: Smoothed series with Spencer's 15-point moving average).

```
> smooth.spencer<-function(my.ts){
    weight<-c(-3,-6,-5,3,21,46,67,74,67,46,21,3,-5,-6,-3)/320
    my.ts.sm<-rep(0, length(my.ts)-15)
    for (i in 1:length(my.ts.sm)) {
          my.ts.sm[i]<-sum(my.ts[i:(i+14)]*weight)
    }
    my.ts.sm
  } # Spencer's 15-point moving average

> library(tseries)
> data(bev)
> bev.sm<-smooth.sym(bev, 7)
> bev.spencer<-smooth.spencer(bev)
> x.pos<-c(1500, 1560, 1620, 1680, 1740, 1800, 1869)
> par(mfrow=c(3,1), mar=c(4,4,4,4))
> plot(bev, type="l", xlab="Year", ylab="Index", xaxt="n")
> axis(1, x.pos, x.pos)
> plot(c(1, length(bev)), c(0, max(bev)), type="n", xlab="Year",
ylab="Index", xaxt="n")
> lines(seq(8, length(bev)-8), bev.sm)
> axis(1, x.pos-1500+1, x.pos)
> plot(c(1, length(bev)), c(0, max(bev)), type="n", xlab="Year",
ylab="Index", xaxt="n")
> lines(seq(8, length(bev)-8), bev.spencer)
> axis(1, x.pos-1500+1, x.pos)
```

Whenever a symmetric filter is chose, there is like to be an end-effects problem, since the symmetric filter does not take care of the end. Ideally we would want smoothened values all the way to the end of the time series, especially when doing forecasting. The analyst can project the smoothed values by eye or, alternatively, use an asymmetric filter that only involves present and past values, like exponential smoothing (Section 5.2.2):

$$\mathrm{Sm}(x_t) = \sum_{j=0}^{\infty} \alpha(1-\alpha)^j x_{t-j},$$

where α is a constant such that $0 < \alpha < 1$.

Having estimated the trend, we can look at the local fluctuations by examining

$$\begin{aligned}
\mathrm{Res}(x_t) &= \text{residual from smoothed value} \\
&= x_t - \mathrm{Sm}(X_t) \\
&= \sum_{r=-q}^{+s} b_r x_{t+r}.
\end{aligned}$$

This is also a linear filter with $b_0 = 1 - a_0$, and $b_r = -a_r$ for $r \neq 0$. If $\Sigma a_r = 1$, then $\Sigma b_r = 0$ and the filter is a trend remover.

How to choose the appropriate filter? The answer to this question requires considerable experience plus knowledge of the frequency aspects of time-series analysis. Some times we may want to get rid of local fluctuations to get smoothed values (low-pass filter), while other times, we want to investigate the residuals and remove the long-term fluctuations (high-pass filter).

Filters in series A smoothing prodedure may be carried out in two or more stages:

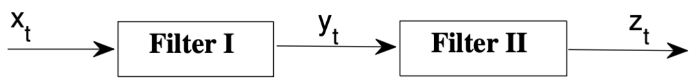


Figure 2.2: Two filters in series.

It is easy to show that a series of linear operations is still a linear filter overall. Suppose $\{a_r\}$ are the coefficients of the first filter which gets $\{y_t\}$ from $\{x_t\}$, and $\{b_j\}$ are the coefficients of the second filter which gets $\{z_t\}$ from $\{y_t\}$. Then

$$\begin{aligned}
z_t &= \sum_j b_j y_{t+j} \\
&= \sum_j b_j \sum_r a_r x_{t+j+r} \\
&= \sum_k c_k x_{t+k}
\end{aligned}$$

where

$$c_k = \sum_r a_r b_{k-r}$$

are the weights for the overall filter. In particular, the weights $\{c_k\}$ are obtained by a procedure called convolution, and write

$$\{c_k\} = \{a_r\} * \{b_j\}.$$

For example,

$$\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right) = \left(\frac{1}{2}, \frac{1}{2}\right) * \left(\frac{1}{2}, \frac{1}{2}\right)$$

The Spencer 15-point moving average is a convolution of four filters:

$$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) * \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) * \left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right) * \left(-\frac{3}{4}, \frac{3}{4}, 1, \frac{3}{4}, -\frac{3}{4}\right)$$

2.5.3 Differencing

A special type of filtering is to take the difference of a given time series until it becomes stationary. For instance, first differencing is defined as

$$\nabla x_t = x_t - x_{t-1} \text{ for } t = 2, 3, \ldots, N$$

while second differencing is defined as

$$\nabla^2 x_t = \nabla x_t - \nabla x_{t-1} = x_t - 2x_{t-1} + x_{t-2}.$$

Seasoning differences will be introduced in the next section.

2.5.4 Other approaches

More complicated approaches to handling trend will be mentiond later. In particular, state-space models involving trend terms will be mentioned in Chapter 10.

2.6 Analyzing Series that Contain a Trend and Seasonal Variation

Three seasonal models in common use are

$$\begin{array}{ll} A & X_t = m_t + S_t + \varepsilon_t \\ B & X_t = m_t S_t + \varepsilon_t \\ C & X_t = m_t S_t \varepsilon_t \end{array}$$

where m_t is the deseasonalized mean level at time t, S_t is the seasonal effect at time t, and ε_t is the random error.

Model A describes the additive case, while models B and C describe the multiplicative case. The analysis of time series, which exhibit seasonal variation, depends on whether one wants to (1) measure the seasonal effect and/or (2) eliminate seasonality.

For series that do contain a substantial trend, a more sophisticated approach is required. With monthly data, the most common way of eliminating the seasonal effect is to calculate

$$\mathrm{Sm}(x_t) = \frac{\frac{1}{2}x_{t-6} + x_{t-5} + \cdots + x_{t+5} + \frac{1}{2}x_{t+6}}{12}.$$

For quarterly data, the seasonal effect can be eliminated by calculating

$$\mathrm{Sm}(x_t) = \frac{\frac{1}{2}x_{t-2} + x_{t-1} + x_t + x_{t+1} + \frac{1}{2}x_{t+2}}{4}.$$

These smoothing procedures all effectively estimate the local (deseasonalized) level of the series. The seasonal effect itself can then be estimated by calculating $x_t - \mathrm{Sm}(x_t)$ or $x_t/\mathrm{Sm}(x_t)$ depending on whether the seasonal effect is additive or multiplicative.

As an example, Fig. 2.3 illustrates the decomposition of the domestic monthly sales series of Australian fortified wine by winemakers into trend, additive, seasonal, and remainder series. We can do this via the "decompose" command in R:

```
> wine<-read.csv("../data/aus\_wine\_sales.csv", header=F)
> wine.ts<-ts(wine[,2], frequency=4, start=c(1985,1))
# create a time series object
> wine.de<-decompose(wine.ts, type="additive")
> plot(wine.de)
```

Another technique of eliminating seasonal effects is seasonal filtering:

$$\nabla_{12} x_t = x_t - x_{t-12}.$$

Two general reviews of methods for seasonal adjustment are Butter and Fase (1991) and Hylleberg (1992). Currently, the most common package for removing trend and seasonal effects is called the X-12 method.

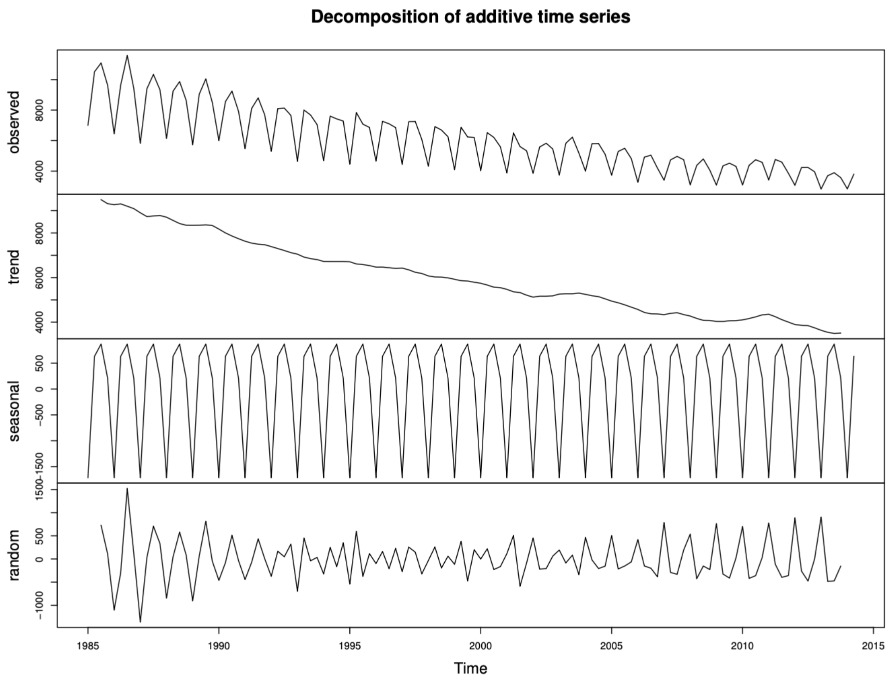


Figure 2.3: The decomposition of the domestic sales of Australian fortified wine.

It is a fairly complicated procedure that implements a series of linear filters and adopts a recursive approach. The new software for X-12 also gives more flexibility in handling outliers, and also allows the users to deal with the possible issue of calendar effects. The X-12 method is often combined with ARIMA modeling, which gives rise to the package X-12-ARIMA. Some other countries use SEATS (Signal Extraction in ARIMA Time Series) or TRAMO (Time-series Regression with ARIMA Noise).

2.7 Autocorrelation and the Correlogram

Let's first review on the regular correlation coefficient. Given N pairs of variables x and y, say $\{(x_1, y_1), \ldots, (x_N, y_N)\}$, the sample correlation coefficient is given by

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}.$$

This value lies in $[-1, 1]$ and measures the strength of linear association between the two variables.

Now, given N observations $x_1, \ldots, x_N$ of a time series, we can form $N-1$ pairs of observation by getting $\{(x_1, x_2), \ldots, (x_{N-1}, x_N)\}$, where each pair of observations is seperated by 1 time interval. Then we can get a sample correlation coefficient

$$r_1 = \frac{\sum_{t=1}^{N-1}(x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)})}{\sqrt{\sum_{t=1}^{N-1}(x_t - \bar{x}_{(1)})^2 \sum_{i=1}^{N-1}(x_{t+1} - \bar{x}_{(2)})^2}},$$

where

$$\bar{x}_{(1)} = \frac{1}{N-1}\sum_{t=1}^{N-1} x_t$$

is the mean of the first $N-1$ observations, while

$$\bar{x}_{(2)} = \frac{2}{N}\sum_{t=1}^{N-1} x_t$$

is the mean of the last $N-1$ observations. The quantity above is called the sample autocorrelation coefficient or a serial correlation coefficient at lag one.

However, the formula is rather complicated, so instead we can estimate using

$$r_1 = \frac{\sum_{t=1}^{N-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{(N-1)\sum_{t=1}^{N-1}(x_t - \bar{x})^2/N}$$

where $\bar{x}$ is just the overall mean. We can also simplify by dropping the factor $N/(N-1)$, which is close to one if N is large. This gives the even simpler formula

$$r_1 = \frac{\sum_{t=1}^{N-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} \quad (*)$$

which is the form that will be used in this book.

In a similar way, we can find the correlation between observations of lag k, meaning

$$r_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2}.$$

In practice, the autocorrelation coefficients are calculated from the autocovariance coefficients, $\{c_k\}$, which we define by analogy using the usual covariance formula:

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x}).$$

We then compute

$$r_k = c_k/c_0, \ \ k = 0, 1, \ldots, M, \ \text{where } M < N.$$

Some authors use this equation instead (NOT used in this book!):

$$c_k = \frac{1}{N-k}\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x}).$$

2.7.1 The correlogram

A useful aid in interpreting a set of autocorrelation coefficients it the correlogram. It contains the sample correlation coefficients r_k for $k = 0, 1, \ldots, M$ for some M which is usually much less than N. For example, if $N = 200$ then we can use $M = 30$. Examples are shown below, from Fig. 2.4 toFig. 2.8.

2.7.2 Interpreting the correlogram

Interpreting the coefficients from the correlogram is not easy. Here are some situations:

Random series

A time series is said to be completely random (i.i.d.) if it consists of a series of independent observations having the same distribution. For large N, we would expect that $r_k \approx 0$ for all $k \neq 0$. In fact, later we will see that, for a random time series, $r_k, k \geq 1$ is approximately $N(0, 1/N)$. Thus, if a time series, is random, we can expect 19 out of 20 of the values of r_k to lie between $\pm 1.96/\sqrt{N}$. As a result, it is common to regard any values of r_k outside of this range to be statistically significant. This also means that even if the time series is completely random, we can find statistically significant values of r_k, which is a type I error.

Fig. 2.4 can be reproduced using the following R code:

```
> set.seed(1)
> x<-rnorm(400)
> par(mfrow=c(2,1), mar=c(3,4,3,4))
> plot(x, type="l", xlab="", ylab="")
> title(xlab="Time", ylab="Series", line=2, cex.lab=1.2)
> acf(x, ylab="", main="")
> title(xlab="Lag", ylab="ACF", line=2)
```

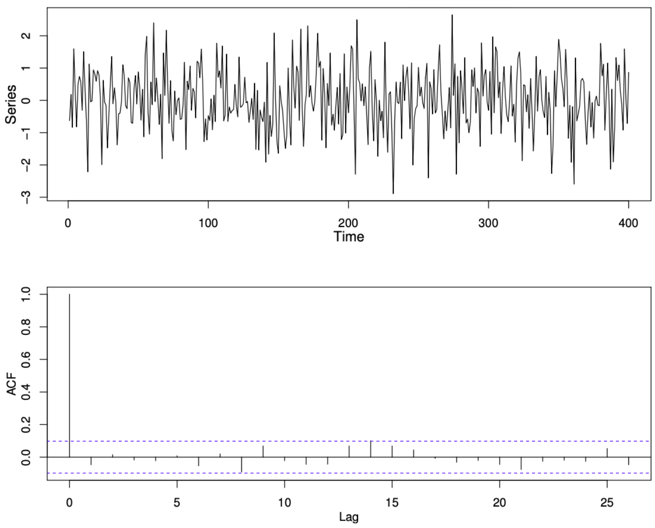


Figure 2.4: A completely random series together with its correlogram. The dotted lines in the correlogram are at $\pm 1.96/\sqrt{N}$. Values outside these lines are said to be significantly different from zero.

Short term correlation

Stationary series often exhibit short-term correlation characterized by a fairly large value of r_1 followed by one or two further coefficients that tend to decrease. Values of r_k for longer lags tend to be approximately zero. An example is shown in Fig. 2.5.

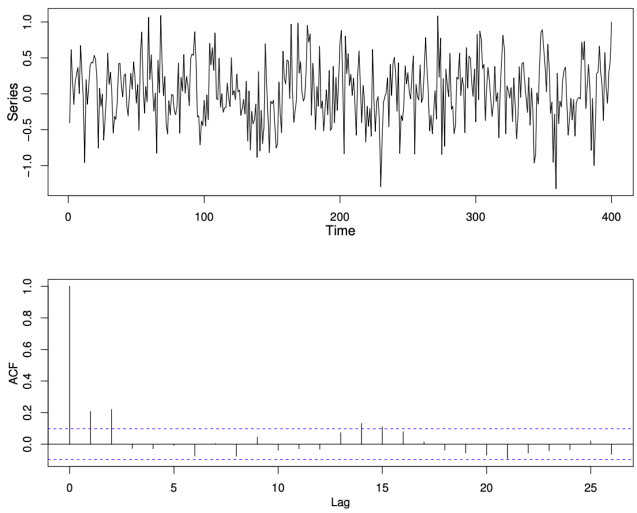


Figure 2.5: A time series showing short-term correlation together with its correlogram.

A time series that gives rise to such a correlogram is one for which an observation above the mean tends to be followed by one or more further observations above the mean, and similarly for observations below the mean.

Alternating series

If the time series tends to alternate, so will its corrologram. An example is shown in Fig. 2.6. The time series plots in Fig. 2.5 and Fig. 2.6 suggest that it is hard to distinguish between a time series with short-term correlation from that with alternating correlation.

Non-stationary series If a time series contains a trend, then the value of r_k will not come down to zero except for very large values of the lag. This is because an observation on one side of the overall mean tends to be followed by a large number of further observations on the same side of the mean because

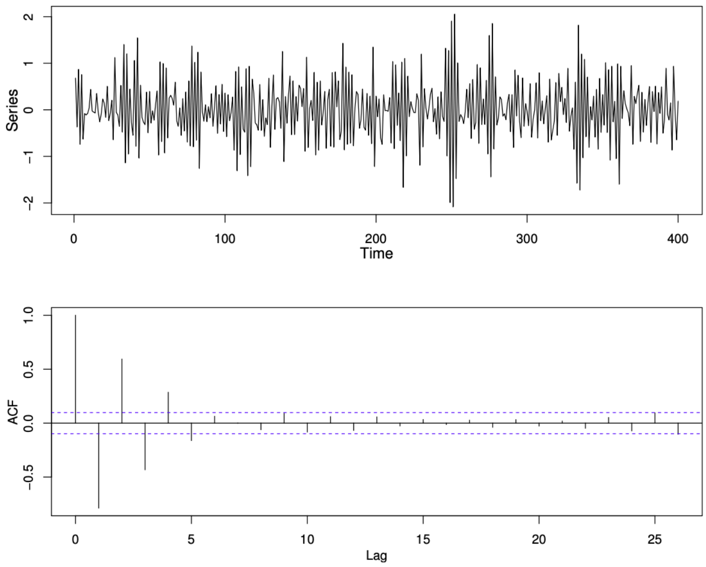


Figure 2.6: An alternating time series together with its correlogram.

of the trend. See Fig. 2.7 for an example. Little can be inferred from the corrologram, and any trend should be removed before calculating the set of autocorrelation coefficients $\{r_k\}$.

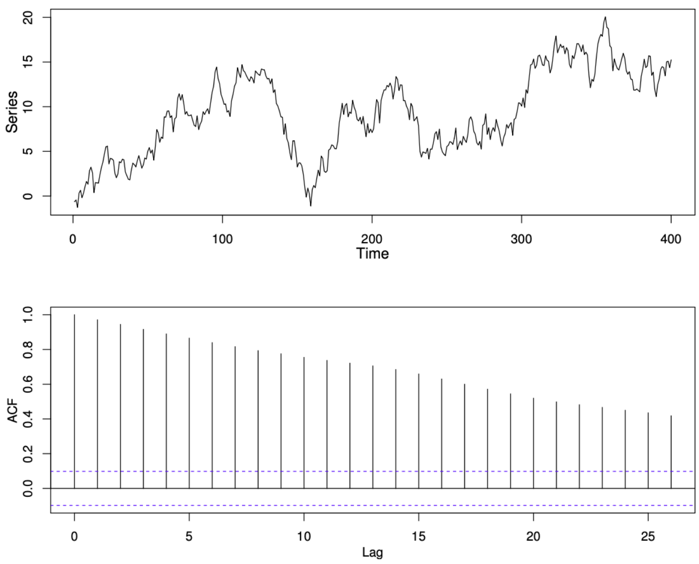


Figure 2.7: A non-stationary time series together with its correlogram.

Fig. 2.7 can be reproduced via the following R code:

```
> set.seed(1)
> ts.sim3<-cumsum(rnorm(400))
> par(mfrow=c(2,1), mar=c(3,4,3,4))
> plot(ts.sim3, type="l", xlab="", ylab="")
> title(xlab="Time", ylab="Series", line=2, cex.lab=1.2)
> acf(ts.sim3, ylab="",main="")
> title(xlab="Lag", ylab="ACF", line=2)
```

Seasonal series If a time series contains seeasonal variation, the corrologram will also exhibit oscillation at the same frequency. For example, with monthly observations, we expect r_6 to be 'large' and negative, while $r_1 2$ will be 'large' and positive. In particular, if x_t follows a sinusoidal pattern, so does r_k. For example, if

$$x_t = a \cos t\omega$$

where a is a constant and the frequency ω is such that $0 < \omega < \pi$, then it can be shown (Exercise 2.3) that

$$r_k \approxeq \cos k\omega \text{ for large } N.$$

See Fig. 2.8 for the corrologram of the monthly air temperature data shown in Fig. 1.3.

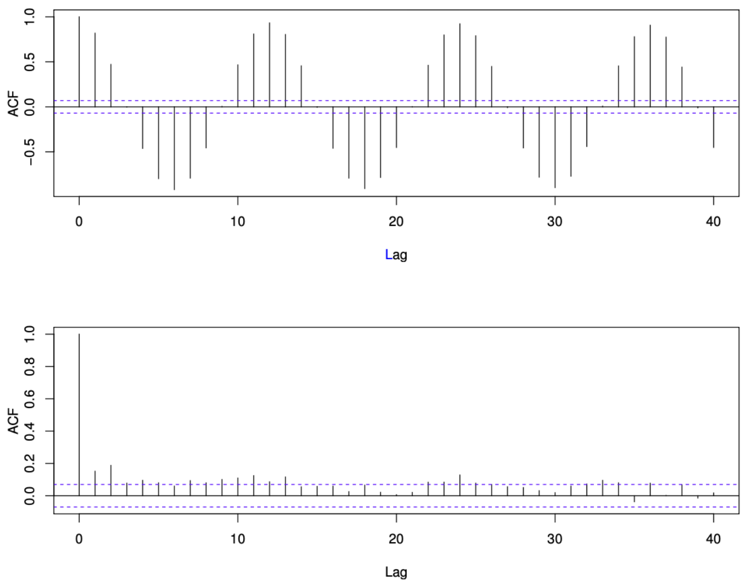


Figure 2.8: The correlograms of monthly observations on air temperature in Anchorage, Alaska for the raw data (top) and for the seasonally adjusted data (bottom).

The sinusoidal pattern of the correlogram is clearly evident, but for seasonal data of this type the correlogram provides little extra information, because the seasonal pattern is already apparent in the data.

If the seasonal variation is removed from seasonal data, then the correlogram may provide more useful information. For example, after the seasonal effect is removed from Fig. 1.3, the correlogram of the resulting series (bottom panel of Fig. 2.8) shows that the first three coefficients are significantly differnt from zero, meaning there is short-term correlation.

Outliers

If a a time series contains one or more outliers, the correlogram may be seriously affected and it is advisable to adjust these outliers before doing the formal analysis. For example, if there is one outlier in the time series at, say, time t_0, and if it is not adjusted, then the plot of x_t against x_{t+k} will contain two 'extreme' points, namely, (x_{t_0-k}, x_{t_o}) and (x_{t_0}, x_{t_0+k}).

General remarks Considerable experience is required to interpret sample autocorrelation coefficients. In addition, we need to study the probability theory of stationary series and learn about the classes of models that may be appropriate.

2.8 Other Tests of Randomness

In most cases, visual examination is enough to determine that the series is NOT random (there is trend/seasonality/short-term correlation). However, we can occasionally test whether the apparent stationary time series it 'random'. One type of approach is to carry out the test of randomness in which we test whether the observations $x_1, \ldots, x_N$ could have arisen in that order by chance. We'll mention a few such tests here.

One type of test is based on counting the number of *turning points*, meaning the number of times there is a local maximum or minimum in the time series. If the series is really random, one can work out the expected number of turning points and compare it with the observed value.

An alternative type of test is based on *runs* of observations. For example, the analyst can count the number of runs where successive observations are all greater than the median or all less than the median, which may indicate short-term correlation. Or they can cout the number of runs of monotonicity, which may incidate trend.

Tests above are not covered in the book, since examining the correlogram is sufficient and simple enough. If a test is required though, the pormanteau test is mentioned in Section 4.7, which is based on testing for residuals.

2.9 Handling Real Data

We'll close on how to handle real data. When receiving the data, we can't assume that it's structured well already, hence we will have to clean the data first, for example: Modifying outliers, correcting obvious errors, and filling in (imputing) missing observations. The analyst should also deal with any other known peculiarities, such as a change in the way that a variable is defined during the course of the data collection process.

After cleaning the data, the next step for the time-series analyst is to determine whether trend and seasonality are present. If so, how shold such eefects be modelled, measured, or removed? The treatment of such effects and missing values is often more important than the analysis and modeling of the time-series data.

The context of the problem is crucial in deciding how to modify data, if at all, and how to handle trend and seasonality. Therefore it is essential to get background knowledge about the problem.

We close by giving a non-exhaustive list of possible actions, and can be adapted to the particular problem:

- Do you understand the context? Have the 'right' variables been measured?
- Have all the time series been plotted?
- Are there any missing values? If so, what should be done about them?
- Are there any outliers? If so, what should be done about them?
- Are there any obvious discontinuities in the data? If so, what does this mean?
- Does it make sense to transform any of the variables?
- Is trend present? If so, what should be done about it?
- Is seasonality present? If so, what should be done about it?

3 Some Linear Time Series Models

This chapter introduces various probability models for time series. Some tools for describing the properties of such models are specified and the important notion of stationarity is formally defined.

3.1 Stochastic Processes and Their Properties

Definition. A stochastic process is a collection of random variables that are ordered in time and defined at a set of time points, which may be continuous or discrete. We denote it as $X(t)$ if time is continous, and X_t if time is discrete.

Definition. The mean function $\mu(t)$ is defined for all t by

$$\mu(t) = \mathbb{E}\left[X(t)\right].$$

The variance function $\sigma^2(t)$ is defined for all t by

$$\sigma^2(t) = \mathrm{Var}(X(t)) = \mathbb{E}\left[(X(t) - \mu(t))^2\right].$$

The autocovariance function (acv.f.) $\gamma(t_1, t_2)$ is simply the covariance of $X(t_1)$ and $X(t_2)$, namely

$$\gamma(t_1, t_2) = \mathbb{E}\left[(X(t_1) - \mu(t_1))(X(t_2) - \mu(t_2))\right].$$

Higher moments of a stochastic process may be defined in a similar way, but are rarely used in practice.

3.2 Stationary Processes

An important class of stochastic processes are those that are stationary.

Definition. A time series is strictly stationary if the joint distribution of $X(t_1), \ldots, X(t_k)$ is the same as the joint distribution of $X(t_1 + \tau), \ldots, X(t_k + \tau)$ for any $t_1, \ldots, t_k, \tau$.

The definition above for $k = 1$ implies that the distribution of $X(t)$ is the same for all t, hence

$$\mu(X(t)) = \mu, \sigma^2(X(t)) = \sigma^2$$

do not depend on the time t.

Furthermore, for $k = 2$, the joint distribution of $X(t_1)$ and $X(t_2)$ depends on only the time difference $t_2 - t_1 = \tau$, which is called the lag. Thus the acv.f. $\gamma(t_1, t_2)$ only depends on the lag τ and can be written as $\gamma(\tau)$, where

$$\begin{aligned}\gamma(\tau) &= \mathbb{E}\left[(X(t) - \mu)(X(t+\tau) - \mu)\right] \\ &= \mathrm{Cov}(X(t), X(t+\tau))\end{aligned}$$

is the autocovariance function at lag τ.

Definition. The autocorrelation function (ac.f.) is defined by

$$\rho(\tau) = \gamma(\tau)/\gamma(0).$$

This quantity measures the correlation between $X(t)$ and $X(t + \tau)$.

It may seem surprising to suggest that there are processes for which the distribution of $X(t)$ is the same for all t. However, from the theory of stochastic processes, there are many processes $\{X(t)\}$ for which have an equilibrium distribution as $t \to \infty$. Of course, the conditional distribution of $X(t_2)$ given $X(t_1)$ can be very different, but this does not conflict with the series being stationary.

In practice, it is often useful to define stationarity in a less restricted way:

Definition. A stochastic process is called second-order stationary or weakly stationary if its mean is constant and its acv.f. only depends on the lag, so that

$$\mathbb{E}[X(t)] = \mu \text{ and } \mathrm{Cov}(X(t), X(t+\tau)) = \gamma(\tau).$$

No requirements are placed on moments higher than the second order. By letting $\tau = 0$, the form of a stationary acv.f. implies that the variance, as well as the mean, is constant. The definition also implies that both the variance and the mean must be finite.

By stationary, we will mean 'weakly stationary' from now on.

3.3 Properties of the Autocorrelation Function

We'll introduce a few general properties of the ac.f. in this subsection.

Suppose a stationary process $X(t)$ has mean μ, variance σ^2, acv.f. $\gamma(t)$ and ac.f. $\rho(t)$. Then

$$\rho(t) = \gamma(t)/\gamma(0) = \gamma(t)/\sigma^2.$$

Note that $\rho(0) = 1$.

Proposition (Properties of the ac.f.)**.** For the autocorrelation function $\rho(\tau)$,
(a) The ac.f. is an even function of lag, so that $\rho(\tau) = \rho(-\tau)$;
(b) $|\rho(t)| \leq 1$;
(c) The ac.f. does not uniquely identify the underlying model.

Proof. (i)

$$\begin{aligned}\gamma(\tau) &= \mathrm{Cov}(X(t), X(t+\tau)) \\ &= \mathrm{Cov}(X(t-\tau), X(t)) \quad \text{(By stationarity)} \\ &= \gamma(-\tau).\end{aligned}$$

(ii) First of all,

$$\mathrm{Var}(\lambda_1 X(t) + \lambda_2 X(t+\tau)) \geq 0$$

for any constants λ_1, λ_2, since the variance is always nonnegative. We can expand the above:

$$\begin{aligned}\mathrm{Var}(\lambda_1 X(t) + \lambda_2 X(t+\tau)) &= \lambda_1^2 \mathrm{Var}(X(t)) + \lambda_2^2 \mathrm{Var}(X(t+\tau)) + 2\lambda_1\lambda_2 \mathrm{Cov}(X(t), X(t+\tau)) \\ &= (\lambda_1^2 + \lambda_2^2)\sigma^2 + 2\lambda_1\lambda_2\gamma(\tau).\end{aligned}$$

When $\lambda_1 = \lambda_2 = 1$, $\gamma(\tau) \geq -\sigma^2$ so that $\rho(\tau) \geq -1$.

When $\lambda_1 = 1, \lambda_2 = -1$, $\sigma^2 \geq \gamma(\tau)$, so that $\rho(t) \leq +1$.

(iii) An example is given in Jenkins and Watts (1968, p.170). The proof is complete.

□

For the ac.f. to guarantee uniqueness of the underlying model, we would also need the invertibility condition, which will be introduced in Section 3.6.

3.4 Purely Random Process

Definition. A discrete-time process is called a purely random process or white noise if it consists of a sequence of random variables, $\{Z_t\}$, which are mutually independent and identically distributed.

We normally further assume that the random variables are normally distributed with mean zero and variance σ_Z^2.

From definition, the process has constant mean and variance. Moreover, the independence assumption gives

$$\gamma(k) = \mathrm{Cov}(Z_t, Z_{t+k}) = \begin{cases} \sigma_Z^2 & \text{if } k = 0, \\ 0 & \text{if } k = \pm 1, \pm 2, \ldots \end{cases}$$

Therefore the ac.f. is given by

$$\rho(k) = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{if } k = \pm 1, \pm 2, \ldots \end{cases}$$

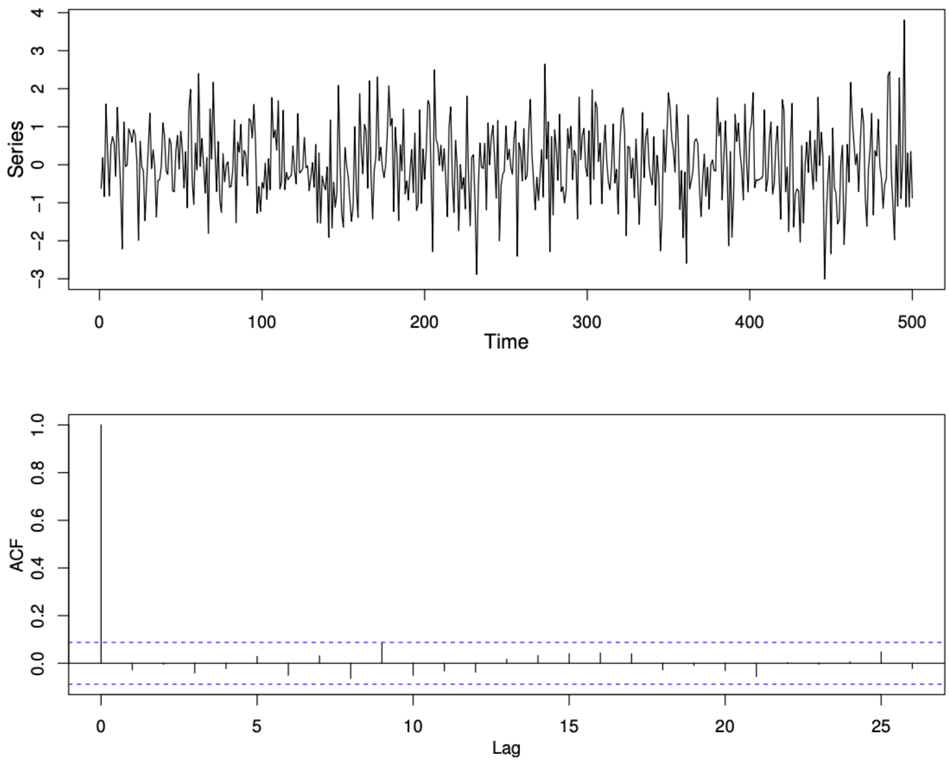


Figure 3.1: A purely random process iwith $\sigma_Z^2 = 1$ (top) and its correlogram (bottom).

As the mean and acv.f. do note depend on time, the process is weakly stationary. In fact, the independence assumption implies that it is strictly stationary.

Purely random processes are useful in many situations, particularly as building blocks for more complicated processes such as moving average processes (Section 3.6). In practice, if all sample ac.f.'s of a series are close to zero, then the series is considered as a realization of a purely random process. Fig. 3.1 shows an example where $Z_t \sim N(0, 1)$, $1 \leq t \leq 500$, and its correlogram, which can be reproduced via the following piece of code:

```
> z<-rnorm(500, 0, 1)
> par(mfrow=c(2,1), mar=c(3,4,3,4))
> plot(z, type="l", xlab="Time", ylab="Series")
> acf(z, xlab="Lag",ylab="ACF", main="")
```

Some authors prefer to make the weaker assumption that the Z_t's are mutually uncorrelated rather than independent. This is fine for linear, normal processes, but not ok for nonlinear models.

3.5 Random Walks

Definition. Suppose $\{Z_t\}$ is a discrete-time, purely random process with mean μ and variance σ_Z^2. A process $\{X_t\}$ is said to be a random walk if

$$X_t = X_{t-1} + Z_t = \sum_{i=1}^{t} Z_i, \text{ and } X_0 = 0.$$

We can easily see that by independence of Z_t,

$$\mathbb{E}[X_t] = t\mu \text{ and } \mathrm{Var}(X_t) = t\sigma_Z^2.$$

As the mean and variance changes with t, this process is not stationary. However, by taking the first differences, we get

$$\nabla X_t = X_t - X_{t-1} = Z_t,$$

which is stationary. Fig. 3.2 below shows a random walk series and its correlogram.

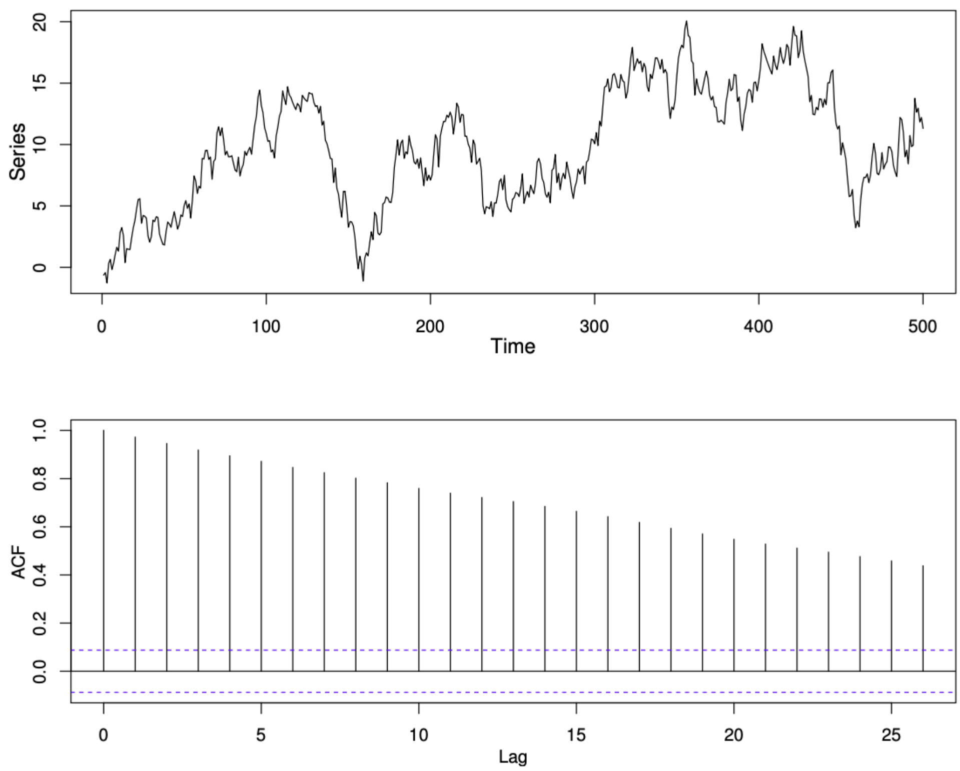


Figure 3.2: Simulated random walk (top) and its correlogram (bottom). The random walk series is generated from the white noise series in Fig. 3.1.

We can generate the random walk series via the cumulative sum of white noise series using the code below:

```
> n<-500
> z<-rnorm(n, 0, 1)
> x.rw<-cumsum(z)
> par(mfrow=c(2,1), mar=c(4,4,4,4))
> plot(x.rw, type="l", xlab="Time", ylab="Series")
> acf(x.rw, xlab="Lag", ylab="ACF", main="")
```

The best-known examples of time series which behave like random walks are share prices.

3.6 Moving Average Processes

Definition. Suppose $\{Z_t\}$ is a discrete-time, purely random process with mean zero and variance σ_Z^2. A process $\{X_t\}$ is said to be a moving average process of order q, denoted a MA(q) process, if

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q},$$

where $\{\beta_i\}$ are constants. The Zs are usually scaled so that $\beta_0 = 1$.

3.6.1 Stationarity and autocorrelation of an MA process

We can immediately get that

$$\mathbb{E}[X_t] = 0, \ \mathrm{Var}(X_t) = \sigma_Z^2 \sum_{i=0}^{q} \beta_i^2,$$

since the Z's are independent. We also have

$$\begin{aligned}
\gamma(k) &= \mathrm{Cov}(X_t, X_{t+k}) \\
&= \mathrm{Cov}(\beta_0 Z_t + \cdots + \beta_q Z_{t-q}, \beta_0 Z_{t+k} + \cdots + \beta_q Z_{t+k-q}) \\
&= \begin{cases} 0 & \text{if } k > q, \\ \sigma_Z^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k} & \text{if } k = 0, 1, \ldots, q \\ \gamma(-k) & \text{if} k < 0 \end{cases}
\end{aligned}$$

since

$$\mathrm{Cov}(Z_s, Z_t) = \begin{cases} \sigma_Z^2 & \text{if } s = t, \\ 0 & \text{if } s \neq t. \end{cases}$$

We can see from above that the process is weakly stationary for all values of $\{\beta_i\}$. Furthermore, if the Zs are normally distributed, then so are the Xs, and we have a strictly stationary normal process.

The ac.f. of the MA(q) process is

$$\rho(k) = \begin{cases} 1 & \text{if } k = 0, \\ \frac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^{q} \beta_i^2} & \text{if } k = 1, \ldots, q, \\ 0 & \text{if } k > q, \\ \rho(-k) & \text{if } k < 0. \end{cases}$$

Note that the ac.f. 'cuts off' at lag q, which is special of the MA process. In particular, the MA(1) process with $\beta_0 = 1$ has an ac.f. given by

$$\rho(k) = \begin{cases} 1 & \text{if } k = 0 \\ \beta_1/(1 + \beta_1^2) & \text{if } k = \pm 1 \\ 0 & \text{otherwise.} \end{cases}$$

Using the definition of the MA(q) process, we can construct a MA series. For example, Fig. 3.3 shows the series and their correlograms of these two MA processes:

$$X_t = Z_t - 0.8 Z_{t-1}, Z_t = Z_t + 0.7 Z_{t-1} - 0.2 Z_{t-2}.$$

We can reproduce Fig. 3.3 via the following piece of code:

```
> n<-500
> z<-rnorm(n)
> x.ma1<-z[2:n]-0.8*z[1:(n-1)]
> x.ma2<-z[3:n]+0.7*z[2:(n-1)]-0.2*z[1:(n-2)]

> par(mfrow=c(2,2), mar=c(4,4,4,4))
```

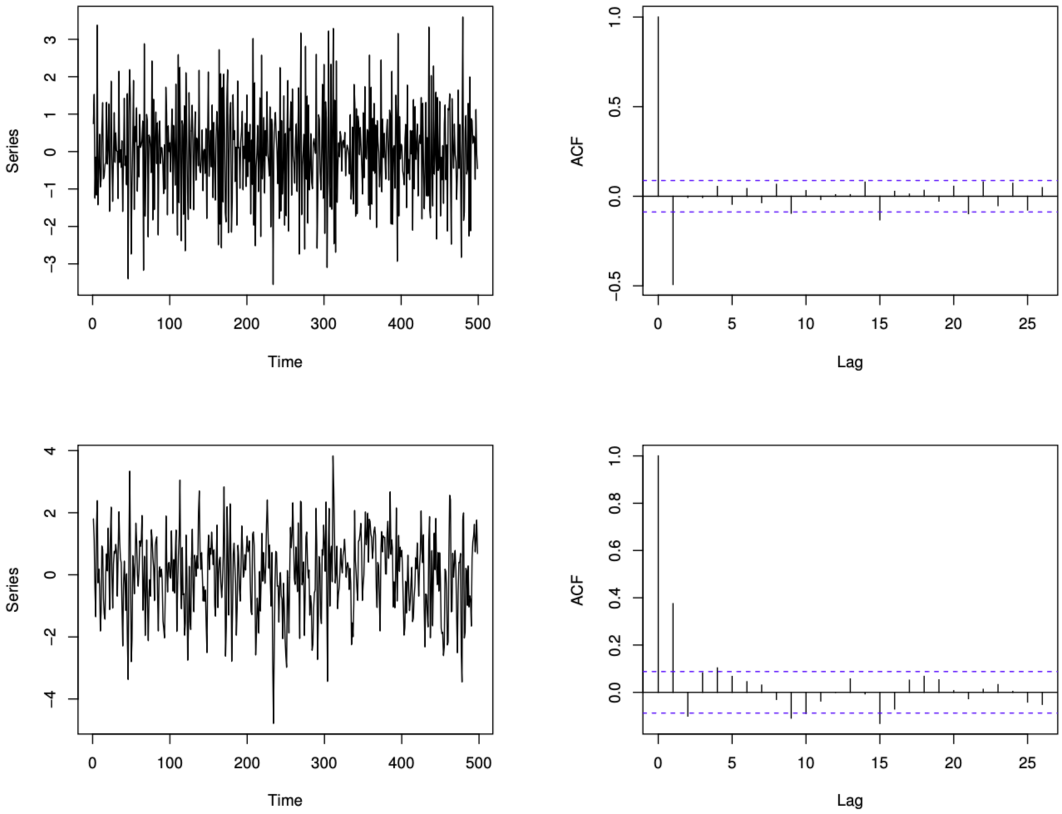


Figure 3.3: Simulated MA(1) (top left) and MA(2) (bottom left) processes and their corresponding correlograms (top right and bottom right).

```
> plot(x.ma1, type="l", xlab="Time", ylab="Series")
> acf(x.ma1, xlab="Lag", ylab="ACF", main="")
> plot(x.ma2, type="l", xlab="Time", ylab="Series")
> acf(x.ma2, xlab="Lag", ylab="ACF", main="")
```

Note that the r_k's are significant only for $k = 0, 1$ and $k = 0, 1, 2$ respectively.

3.6.2 Invertibility of an MA process

No restrictions on the $\{\beta_i\}$ are required for a (finite-order) MA process to be stationary, but it is generally advisable to impose restrictions on $\{\beta_i\}$ so that the process satisfies a condition called invertibility. This condition can be explained in the following way. Considering the following MA processes:

$$\begin{aligned} \text{A}: &\quad X_t = Z_t + \theta Z_{t-1}. \\ \text{B}: &\quad X_t = Z_t + \frac{1}{\theta} Z_{t-1}. \end{aligned}$$

Using the formula of the ac.f. for a MA(1) process, we can get that the two different processes have exactly the same ac.f. Therefore we cannot identify a MA process uniquely from a given ac.f. Now, if we 'invert' models A and B by expressing Z_t in terms of $X_t, X_{t-1}, \ldots$ via successive substitution, we get that

$$\begin{aligned} \text{A}: &\quad Z_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \cdots \\ \text{B}: &\quad Z_t = X_t - \frac{1}{\theta} X_{t-1} + \frac{1}{\theta^2} X_{t-2} - \cdots \end{aligned}$$

Of $|\theta| < 1$, the series of coefficients of X_{t-j} for model A converges whereas that of B does not. Thus model B cannot be 'inverted' this way.

Definition. A process $\{X(t)\}$ is said to be invertible if the random disturbance at time t, sometimes called the *innovation*, can be expressed as a convergent sum of present and past values of X_t in the form

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad \sum_j |\pi_j| < \infty.$$

Therefore this means a first order MA process can be rewritten in the form of an autoregressive process, possible of infinite order, whose coefficients form a convergent sum.

Definition. The backshift operator, denoted B, is defined by

$$B^j X_t = X_{t-j} \text{ for } j = 0, 1, \ldots$$

Then the MA process can be rewritten as

$$\begin{aligned} X_t &= (\beta_0 + \beta_1 B + \cdots + \beta_q B^q) Z_t \\ &= \theta(B) Z_t \end{aligned}$$

where $\theta(q)$ is a polynomial of order q in B.

Theorem (Invertibility criterion for MA process)**.** A MA(q) process is invertible if the roots of the equation

$$\theta(B) = \beta_0 + \beta_1 B + \cdots + \beta_q B^q = 0$$

all lie outside the unit circle, where we regard B as a complex variable.

Proof. First of all, from the MA equation we can directly get that

$$Z_t = \frac{1}{\theta(B)} X_t.$$

Suppose that $\theta(B)$ can be decomposed into the following form:

$$\theta(B) = (1 + \theta_1 B) \cdots (1 + \theta_q B),$$

where $\theta_1, \ldots, \theta_q$ could possibly take complex variables. Then the operator $1/\theta(B)$ can be written as

$$\frac{1}{\theta(B)} = \prod_{j=1}^{q} \frac{1}{1 + \theta_j B} = \prod_{j=1}^{q} \left(1 + \sum_{i=1}^{\infty} (-\theta_j)^i B^i \right).$$

When all the roots, $-1/\theta_1, \ldots, -1/\theta_q$, are outside the unit circle, the product on the right hand side above is convergent, hence we can indeed write $Z_t = 1/\theta(B) X_t$, and therefore $\{X_t\}$ is invertible. $\square$

MA processes have been used in many areas, particularly econometrics. For example, economic indicators are affected by a variety of 'random' events such as strikes, government decisions, shortages of key materials, and so on. Such events will not only have an immediate effect but may also affect economic indicators to a lesser extent in several subsequent periods, and so it is at least plausible that an MA process may be appropriate.

An arbitrary constant, say μ, can be added to the MA equation to give a process of mean μ. However, this does not affect the ac.f. (Exercise 3.5) hence has been omitted.

3.7 Autoregressive Processes

Autoregressive processes are slightly different from moving average processes, but they are closely related to each other.

Definition. Suppose $\{Z_t\}$ is a discrete-time, purely random process with mean zero and variance σ_Z^2. A process $\{X_t\}$ is said to be a autoregressive process of order p, denoted a AR(P) process, if

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t.$$

3.7.1 First-order process

We begin by examining the case where $p = 1$, i.e.

$$X_t = \alpha X_{t-1} + Z_t.$$

The AR(1) process is sometimes called the Markov process. By substitution to the equation above, we get

$$\begin{aligned} X_t &= \alpha(\alpha X_{t-2} + Z_{t-1}) + Z_t \\ &= \alpha^2(\alpha X_{t-3} + Z_{t-2}) + \alpha Z_{t-1} + Z_t \\ &= \cdots \end{aligned}$$

Therefore X_t can be expressed as an infinite-order MA process in the form

$$X_t = Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \cdots$$

provided that $|\alpha| < 1$ so that the sum converges.

This shows that there is a duality between AR and MA processes, which we can in fact see by writing in terms of the backshift operator:

$$(1 - \sigma B)X_t = Z_t,$$

so that

$$\begin{aligned} X_t &= Z_t/(1 - \alpha B) \\ &= (1 + \alpha B + \alpha^2 B^2 + \cdots)Z_t \\ &= Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \cdots \end{aligned}$$

When expressed in this form, it is clear that

$$\mathbb{E}\left[X_t\right] = 0,\ \mathrm{Var}(X_t) = \sigma_Z^2(1 + \alpha^2 + \alpha^4 + \cdots) = \frac{\sigma_Z^2}{1 - \alpha^2}$$

provided that $|\alpha| < 1$. The acv.f. is given by

$$\begin{aligned} \gamma(k) &= \mathbb{E}\left[X_t X_{t+k}\right] \\ &= \mathbb{E}\left[(\Sigma \alpha^i Z_{t-i})(\Sigma \alpha^j Z_{t+k-j})\right] \\ &= \sigma_Z^2 \sum_{i=0}^{\infty} \alpha^i \alpha^{k+i} \quad (\text{ for } k \geq 0) \\ &= \alpha^k \sigma_Z^2/(1 - \alpha^2) \quad (\text{provided } |\alpha| < 1). \end{aligned}$$

For $k < 0$, we find $\gamma(k) = \gamma(-k)$. Since $\gamma(k)$ does not depend on t, an AR(1) process is weakly stationary provided that $|\alpha| < 1$, and the ac.f. is then given by

$$\rho(k) = \alpha^k,\ k = 0, 1, 2, \ldots$$

To get an even functino defined for all integers k, we rewrite the above as

$$\rho(k) = \alpha^{|k|},\ k = 0, \pm 1, \pm 2, \ldots$$

Three examples of AR processes are shown in Fig. 3.4 for $\alpha = 0.8, -0.8$, and 0.3. All three series are constructed using the same noise series via the code below:

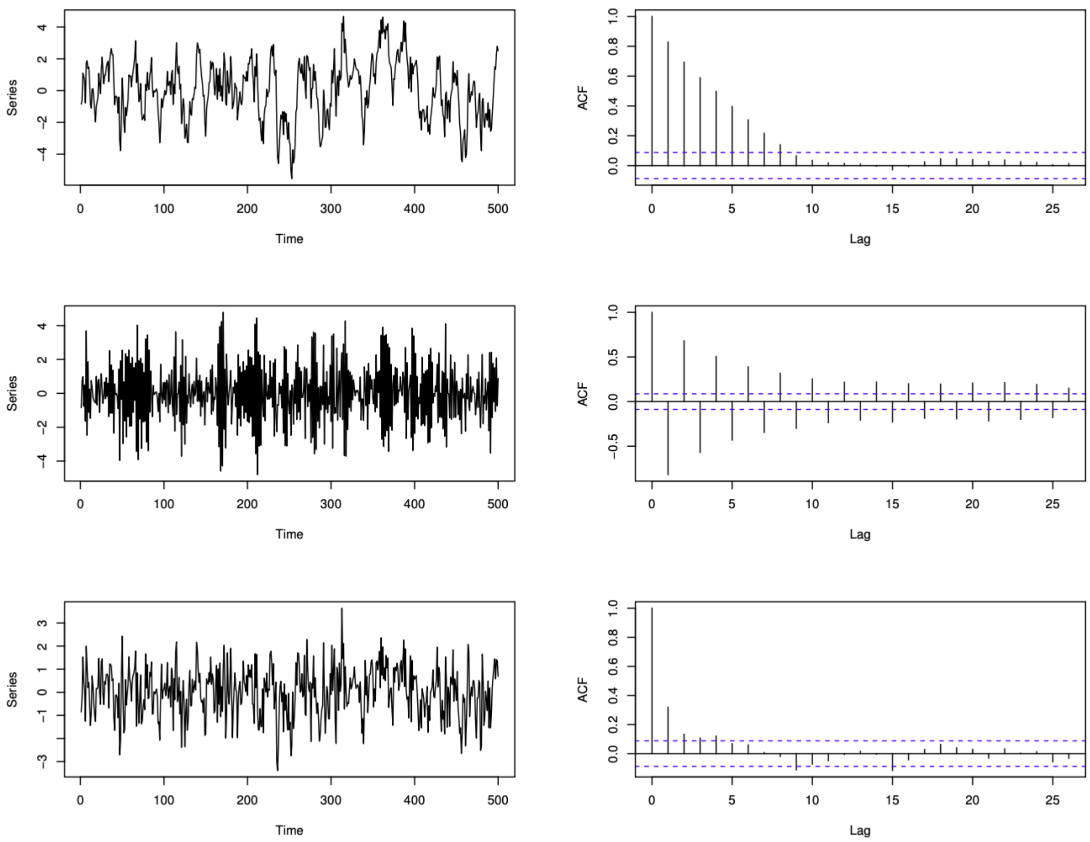


Figure 3.4: Three simulated AR(1) processes and their corrolograms. Top: $X_t = 0.8X_{t-1} + Z_t$; Middle: $X_t = -0.8X_{t-1} + Z_t$; Bottom: $X_t = 0.3X_{t-1} + Z_t$, $Z_t \sim N(0, 1)$.

```
> n<-500
> z<-rnorm(n, 0, 1)
> x.ar1.1<-x.ar1.2<-x.ar1.3<-rep(0,n)
> x.ar1.1[1]<-x.ar1.2[1]<-x.ar1.3[1]<-z[1]
> for (i in 2:n){
    x.ar1.1[i]<- 0.8*x.ar1.1[i-1]+z[i]
    x.ar1.2[i]<- -0.8*x.ar1.2[i-1]+z[i]
    x.ar1.3[i]<- 0.3*x.ar1.3[i-1]+z[i]
  }
```

Note how quick the ac.f. decays when $\alpha = 0.3$, and how it alternates when $\alpha = -0.8$.

3.7.2 General order process

Recall the AR equation:

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t \implies (1 - \alpha_1 B - \cdots - \alpha_p B^p)X_t = Z_t.$$

We can rewrite this as

$$X_t = Z_t/(1 - \alpha_1 B - \cdots - \alpha_p B^p) = f(B)Z_t,$$

where

$$\begin{aligned} f(B) &= (1 - \alpha_1 B - \cdots - \alpha_p B^p)^{-1} \\ &= (1 + \beta_1 B + \beta_2 B^2 + \cdots). \end{aligned}$$

The relationship between the a's and the β's may then be found. Since we just rewrote it as an MA(∞) process, it follows that $\mathbb{E}[X_t] = 0$. The variance is finite provided that $\sum \beta_i^2$ converges, and is a necessary condition for stationarity. We can rewrite the acv.f. by the β's:

$$\gamma(k) = \sigma_Z^2 \sum_{i=0}^{\infty} \beta_i \beta_{i+k}, \ \beta_0 = 1.$$

A sufficient condition for this to converge, and hence stationarity, is that $\sum |\beta_i|$ converges.

Yule-Walker equations We can in principle find the ac.f. of any AR(p) process using the above procedure, but the β_i may be algebraically hard to find. The simpler way is to *assume* the process is stationary, multiply through the AR equation by X_{t-k}, take expectations and divide by σ_X^2, assuming that the variance of X_t is finite. Then, using the fact that $\rho(k)$ is even, we find:

Definition. The Yule-Walker equations are the following set of linear equations:

$$\rho(k) = \alpha_1 \rho(k-1) + \cdots + \alpha_p \rho(k-p) \text{ for all } k = 1, 2, \ldots$$

It is a set of difference equations and has the general solution

$$\rho(k) = A_1 \pi_1^{|k|} + \cdots + A_p \pi_p^{|k|},$$

where $\{\pi_i\}$ are the roots of the auxiliary equation

$$y^p - \alpha_1 y^{p-1} - \cdots - \alpha_p = 0.$$

The constants $\{A_i\}$ are chosen to satisfy the initial conditions depending on $\rho(0) = 1$, meaning that $\sum A_i = 1$.

Stationarity conditions From the general form of $\rho(k)$, it is clear that $\rho(k)$ tends to zero as k increases provided that $|\pi_i| < 1$ for all i, and this is a necessary and sufficient condition for the AR(p) process to be stationary. Equivalently, this is equal to the roots of the equation

$$\phi(B) = 1 - \alpha_1 B - \cdots - \alpha_p B^p = 0$$

must lie outside the unit circle.

Of particular interest is the AR(2) process, when π_1, π_2 are the roots of the quadratic equation

$$y^2 - \alpha_1 y - \alpha_2 = 0.$$

Here $|\pi_i| < 1$ if

$$\left| \frac{\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}}{2} \right| < 1$$

from which we can show (Exercise 3.6) that the stationarity region is the triangular region satisfying

$$\alpha_1 + \alpha_2 < 1, \ \alpha_1 - \alpha_2 > -1, \alpha_2 > -1.$$

The roots are real if $\alpha_1 + 4\alpha_2 > 0$, in which case the ac.f. decreases exponentially with k, but the roots are complex if $\alpha_1^2 + 4\alpha_2 < 0$, in which case the ac.f. turns out to be a damped sinusoidal wave.

When the roots are real, $\rho(k) = A_1 \pi_1^{|k|} + A_2 \pi_2^{|k|}$ where the constants A_1, A_2 are also real and may be found as follows. Since $\rho(0) = 1$,

$$A_1 + A_2 = 1$$

while the first Yule-Walker equation gives

$$\begin{aligned} \rho(1) &= \alpha_1 \rho(0) + \alpha_2 \rho(-1) \\ &= \alpha_1 + \alpha_2 \rho(1). \end{aligned}$$

Solving the above gives $\rho(1) = \alpha_1 / (1 - \alpha_2)$, which in turn must equal

$$A_1 \pi_1 + A_2 \pi_2 = A_1 \pi_1 + (1 - A_1) \pi_2.$$

Therefore, we finally get

$$\begin{cases} A_1 & = \frac{\alpha_1/(1-\alpha_2)-\pi_2}{\pi_1-\pi_2} \\ A_2 & = 1 - A_1 \end{cases}$$

and we can write down the general form of the ac.f. of an AR(2) process with real roots.

For example, consider the AR(2) process defined by $X_t = 1/3X_{t-1} + 2/9X_{t-2} + Z_t$. Fig. 3.5 below shows a realization of this process and its correlogram. The roots are -3 and 3/2, hence we can use the formula above to find its ac.f. (Exercise 3.6).

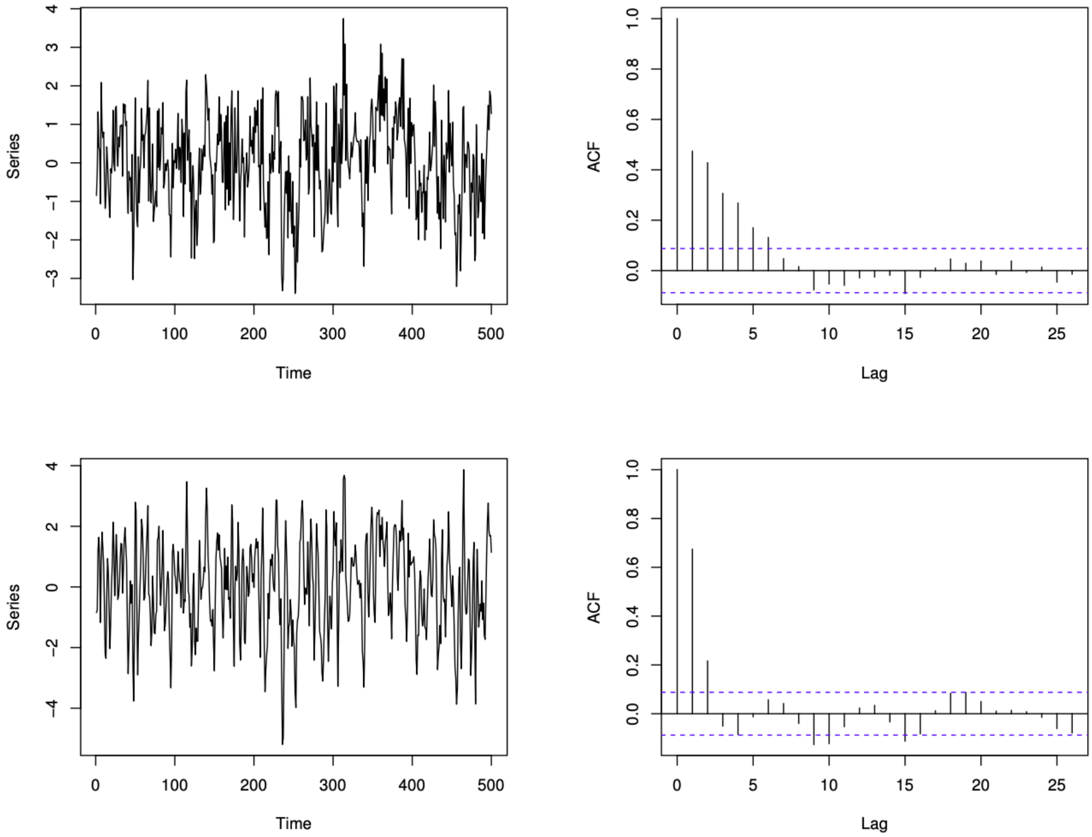


Figure 3.5: Two simulated AR(2) processes and their correlograms. Top: $X_t = \frac{1}{3}X_{t-1} + \frac{2}{9}X_{t-2} + Z_t$; Bottom: $X_t = X_{t-1} - \frac{1}{2}X_{t-2} + Z_t$; $Z_t \sim N(0, 1)$.

Now let's look at an example for complex roots:

Example 3.1. Consider the AR(2) process given by

$$X_t = X_{t-1} - \frac{1}{2}X_{t-2} + Z_t.$$

Is this process stationary? If so, what is its ac.f.?
Let's look at the roots of the polynomial, in this case, it's

$$\phi(B) = 1 - B + \frac{1}{2}B^2 = 0.$$

The roots are $1 \pm i$. The modulus of both roots exceed one, hence the process is stationary.
Now, to find the ac.f., we use the first Yule-Walker equation to give

$$\begin{aligned} \rho(1) = \rho(0) - \frac{1}{2}\rho(-1) \\ = 1 - \frac{1}{2}\rho(1), \end{aligned}$$

giving $\rho(1) = 2/3$.
We can use the Yule-Walker equations to solve for $\rho(2), \rho(3)$ and so on by successive substitution, but let's use a smarter approach by solving the set of Yule-Walker equations as a set of difference equations. The Yule-Walker equation above has the auxiliary equation

$$y^2 - y + \frac{1}{2} = 0$$

with roots $(1 \pm i)/2$. By Euler's formula, this is

$$\frac{1 \pm i}{2} = \frac{\cos(\pi/4) \pm i \sin(\pi/4)}{\sqrt{2}} = e^{\pm i\pi/4}/\sqrt{2}.$$

Since $\alpha_1^2 + 4\alpha_2 = 1 - 2 < 0$, and the roots are complex, the ac.f. is a dampled sinusoidal wave. Using $\rho(0) = 1$ and $\rho(1) = 2/3$, some messy trigonometry and algebra involving complex numbers gives

$$\rho(k) = \left(\frac{1}{\sqrt{2}}\right)^k \left(\cos\frac{\pi k}{4} + \frac{1}{3}\sin\frac{\pi k}{4}\right)$$

for $k = 0, 1, 2, \ldots$. Note that the values of the ac.f. are all real even if the roots of the auxiliary equation are complex. The bottom two panels of Fig. 3.5 show a realization of X_t and its correlogram.

Again, like MA processes, non-zero means may be dealth with by rewriting the AR equation as

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \cdots + \alpha_p(X_{t-p} - \mu) + Z_t.$$

This does not affect the ac.f. (Exercise 3.4).

3.8 Mixed ARMA Models

A useful class of models for time series is formed by combining AR and MA processes:

Definition. A mixed autoregressive moving average process containing p AR terms and q MA terms is said to be an ARMA process or order (p, q). It is given by

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q},$$

where $\{Z_t\}$ is a purely random process with mean zero and variance σ_Z^2.

Using the backshift operator, we can rewrite the above as

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

where $\phi(B), \theta(B)$ are polynomials of order p, q respectively, such that

$$\phi(B) = 1 - \alpha_1 B - \cdots - \alpha_p B^p$$

and

$$\theta(B) = 1 + \beta_1 B + \cdots + \beta_q B^q.$$

3.8.1 Stationarity and invertibility conditions

The conditions to make the process stationary and invertible are those combined for the pure AR and pure MA processes. This means the roots for the equation

$$\phi(B) = 0, \theta(B) = 0$$

all lie outside the unit circle. In is straightforward in principle, though algebraically tedious, to calculate the ac.f. of an ARMA process. This is not discussed in the text (Exercise 3.11).

The importance of ARMA processes lies in the fact that a stationary time series may often be adequately modelled by an ARMA model involving fewer parameters than a pure MA or AR process by itself. This is an early example of what is often called the Principle of Parsimony. This says that we want to find a model with as few parameters as possible, but which gives an adequate representation of the data at hand.

3.8.2 Yule-Walker equations and autocorrelations

The ac.f. of an ARMA process can be found using similar procedures as for AR processe. First, multiply through the ARMA equation by X_{t-k} and take expectations. Note that, for $k > q$, $Z_t, \ldots, Z_{t-q}$ are independent of X_{t-k}. Hence the expected values of $Z_t X_{t-k}, \ldots, Z_{t-q} X_{t-k}$ are all zero. If $k \geq p$, we can further divide both sides of the equation by $\gamma(0)$, then we get the Yule-Walker equations for the general ARMA(p, q) process

$$\rho(k) = \alpha_1 \rho(k-1) + \cdots + \alpha_p \rho(k-p),\ k \geq \max(p, q+1).$$

Note that the form is the exact same as that of a pure AR process, but the initial conditions are different. Here is an example of how to compute the ac.f.:

Example 3.2. Consider the ARMA$(1,1)$ process

$$X_t =_{t-1} + Z_t + \beta Z_{t-1},$$

where $|\alpha|, |\beta| < 1$. To derive the ac.f. of the process, the Yule-Walker equations are

$$\rho(k) = \alpha \rho(k-1),\ k \geq 2.$$

To obtain the initial conditions, we note that $\gamma(1)$ can be computed as follows:

$$\begin{aligned}
\gamma(1) &= \text{Cov}(X_t, X_{t-1}) \\
&= \text{Cov}(\alpha X_{t-1} + Z_t + \beta Z_{t-1}, X_{t-1}) \\
&= \alpha\gamma(0) + \beta \text{Cov}(Z_{t-1}, X_{t-1}) \\
&= \alpha\gamma(0) + \beta \text{Cov}(Z_{t-1}, \alpha X_{t-2} + Z_t + \beta Z_{t-2}) \\
&= \alpha\gamma(0) + \beta\sigma_Z^2.
\end{aligned}$$

The variance of X_t, or $\gamma(0)$, can be calculated as follows:

$$\begin{aligned}
\gamma(0) &= \text{Var}(\alpha X_{t-1} + Z_t + \beta Z_{t-1}) \\
&= \text{Cov}(\alpha X_{t-1} + Z_t + \beta Z_{t-1}, \alpha X_{t-1} + Z_t + \beta Z_{t-1}) \\
&= \alpha^2\gamma(0) + (1+\beta^2)\sigma_Z^2 + 2\alpha\beta \text{Cov}(X_{t-1}, Z_{t-1}) \\
&= \alpha^2\gamma(0) + (1 + 2\alpha\beta + \beta^2)\sigma_Z^2.
\end{aligned}$$

Solving for $\gamma(0)$ gives

$$\gamma(0) = \frac{1 + 2\alpha\beta + \beta^2}{1-\alpha^2}\sigma_Z^2.$$

Thewn plugging this into the equation for $\gamma(1)$ gives

$$\gamma(1) = \frac{(1+\alpha\beta)(\alpha+\beta)}{1-\alpha^2}\sigma_Z^2.$$

Therefore,

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{(1+\alpha\beta)(\alpha+\beta)}{1+2\alpha\beta+\beta^2}.$$

USing the Yule-Walker equations, we then have

$$\rho(k) = \frac{\gamma(1)}{\gamma(0)} = \frac{(1+\alpha\beta)(\alpha+\beta)}{1+2\alpha\beta+\beta^2}\alpha^{k-1},\ k \geq 1.$$

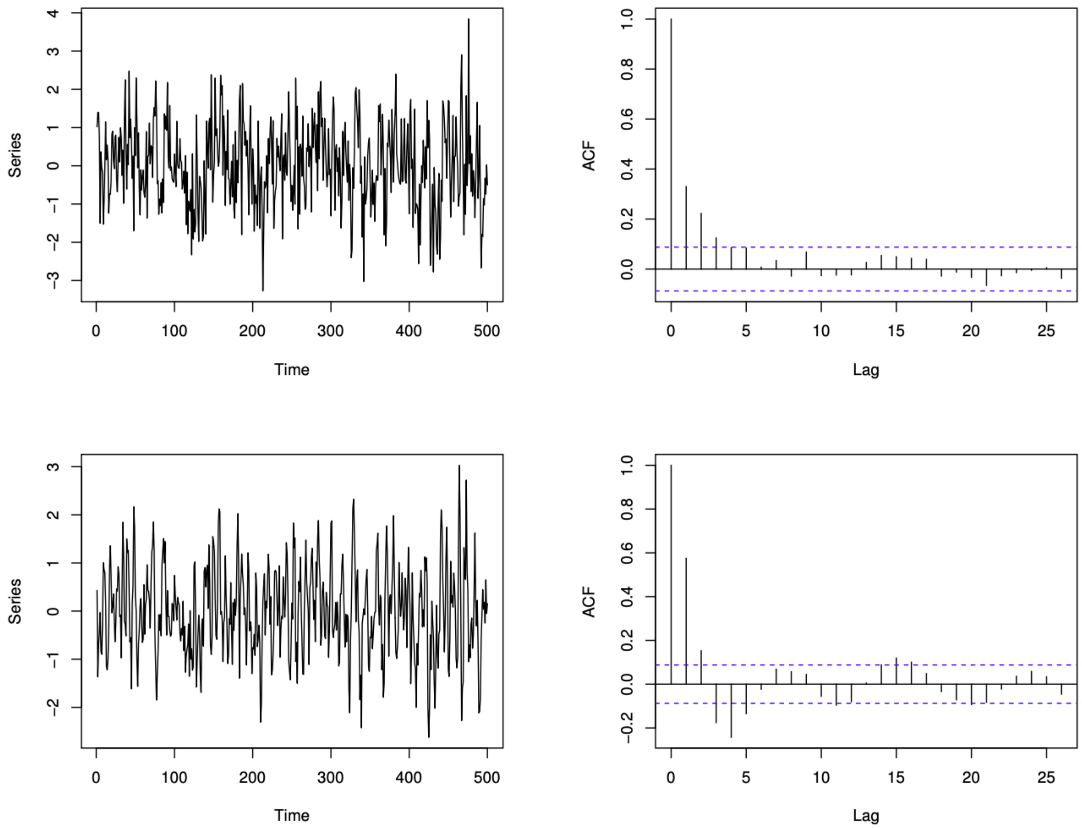


Figure 3.6: Two simulated ARMA processes and their correlograms. Top: $X_t = 0.7X_{t-1} + Z_t - 0.4Z_{t-1}$; Bottom: $X_t = 0.9X_{t-1} - 0.5X_{t-2} + Z_t - 0.2Z_{t-1} + 0.25Z_{t-2}$, $Z_t \sim N(0, 0.5)$.

In principle, the ac.f. for a general ARMA process can be computed, but are usually more complicated than for AR processes. Fig. 3.6 above shows two simulated AR processes and their correlograms using the following piece of code:

```
> x1<-arima.sim(n=500, list(ar=0.7, ma=-0.4))
> x2<-arima.sim(n=500, list(ar=c(0.9,-0.5), ma=c(-0.2,0.25)), sd=sqrt(0.5))
```

Note that, though the correlogram of the ARMA(2, 2) process in Fig. 3.6 looks like that from the AR(2) process in Fig. 3.5, these processes are completely different!

3.8.3 AR and MA representations

We can express ARMA processes as a pure MA process or a pure AR process:

$$X_t = \psi(B)Z_t \text{ or } \pi(B)X_t = Z_t,$$

where

$$\psi(B) = \sum_{i=0}^{\infty} \psi_i B^i \text{ and } \pi(B) = 1 - \sum_{i=1}^{\infty} \pi_i B^i.$$

We can immediately get that

$$\pi(B)\psi(B) = 1.$$

The weights of ψ or π can be obtained directly by division or by equating powers of B in an equation such as

$$\psi(B)\phi(B) = \theta(B).$$

Example 3.3. Find the ψ weights and π weights for the ARMA(1, 1) process given by

$$X_t = 0.5X_{t-1} + Z_t - 0.3Z_{t-1}.$$

Here $\phi(B) = 1 - 0.5B$ and $\theta(B) = 1 - 0.3B$. It follows that the process is stationary and invertible, because both equations have roots greter than one. Then

$$\begin{aligned}\psi(B) &= \theta(B)/\phi(B)\\ &= (1-0.3B)(1-0.5B)^{-1}\\ &= (1-0.3B))(1+0.5B+0.5^2B^2+\cdots)\\ &= 1+0.2B+0.1B^2+0.05B^2+\cdots\end{aligned}$$

Hence

$$\psi_0 = 1,\ \psi_i = 0.2 \times 0.5^{i-1} \text{ for } i = 1, 2, \ldots$$

Similarly, we find

$$\pi_0 = 1,\ \pi_i = 0.2 \times 0.3^{i-1} \text{ for } i = 1, 2, \ldots$$

Note that both the ψ and the π weights die away quickly, and this also indicates a stationary, invertible process.

3.9 Integrated ARMA (or ARIMA) Models

In practice most time series are non-stationary. If the time series is non-stationary in the mean, then we can difference the series, as suggested in Section 2.5.3.

Definition. Set

$$W_t = \nabla^d X_t = (1-b)^d X_t,\ d = 0, 1, 2, \ldots$$

The autoregressive integrated moving average (ARIMA) process of order (p, d, q) is of the form

$$W_t = \alpha_1 W_{t-1} + \cdots + \alpha_p W_{t-p} + Z_t + \beta_1 Z_{t-1} + \cdots + \beta_1 Z_{t-q}.$$

Of course, we can write the above using backshift operators:

$$\phi(B)W_t = \theta(B)Z_t$$

or

$$\phi(B)(1-B)^d X_t = \theta(B)Z_t.$$

The model for X_t is clearly non-stationary, as the AR operator has d roots on the unit circle. For example, the random walk can be regarded as an ARIMA(0, 1, 0) process, which is non-stationary. Applying the difference operator once makes the process stationary.

Fig. 3.7 shows a simulated ARIMA(1, 1, 1) and a simmmulated ARIMA(1, 1, 2) process via theh following piece of code:

```
> y1<-arima.sim(list(order=c(1,1,1), ar=-0.5, ma=-0.3), n=500)
> y2<-arima.sim(list(order=c(1,1,2), ar=0.3, ma=c(-0.3,0.5)), n=500)
```

ARIMA models can be generalized to include seasonal terms, which will be discussed in Section 4.8.

3.10 Fractional Differencing and Long-Memory Models

An interesting variant of ARIMA modeling arises with the use of fractional differencing, leading to a fractional integrated ARMA model:

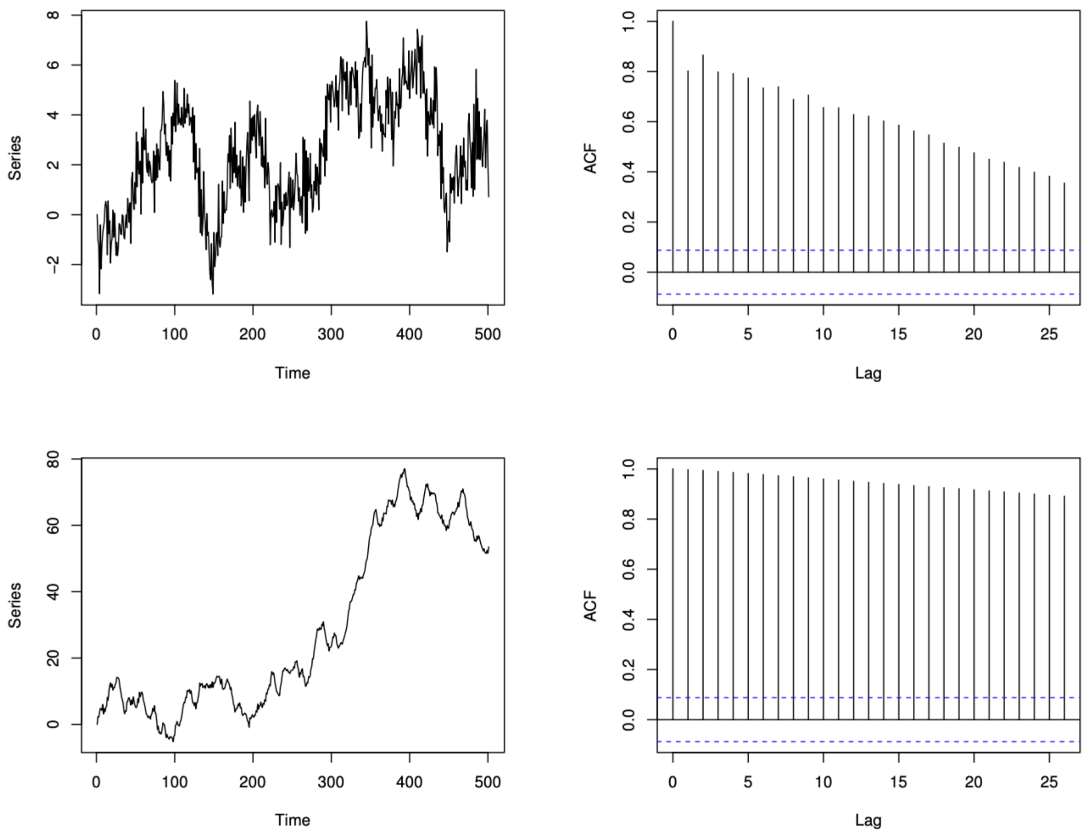


Figure 3.7: Two simulated ARIMA processes and their correlograms. Top: $(1+0.5B)(1-B)X_t = (1+0.3B)Z_t$; Bottom: $(1-0.6B)(1-B)X_t = (1-0.3B+0.5B^2)Z_t$.

Definition. A fractional integrated ARMA (ARFIMA) process is a process defined by

$$\phi(B)(1-B)^d X_t = \theta(B)Z_t,$$

where $\phi(B)$ and $\theta(B)$ are polynomials of oreder p, q respectively, and d does not need to be an integer.

When d is not an integer, then the dth difference $(1-B)^d X_t$ becomes a fractional difference, and may be represented by its binomial expansion, namely

$$(1-B)^d X_t = \left[1 - dB + \frac{d(d-1)}{2!}B^2 - \frac{d(d-1)(d-2)}{3!}B^3 + \cdots\right]$$

As such, it is an infinite weighted sum of past values.

It can be shown (Brockwell & Davis, 1991, Section 13.2) that an ARFIMA process is stationary provided that $-0.5 < d < 0.5$. For $d > 1/2$, the processes is not stationary in the usual sense, but further integer differencing can be used to give a stationary ARFIMA process. For example, first differencing of an ARFIMA$(p, d = 1.3, q)$ process gives a stationary ARFIMA$(p, d = 0.3, q)$ process.

A stationary ARFIMA model with $0 < d < 0.5$, is of particular interest as such a process is not only stationary, but is also an exmaple of a **long-memory model**. This means the correlations decay to zero very slowly.

Definition. A stationary process with ac.f. $\rho(k)$ is a long-memory process if

$$\sum_{k=0}^{\infty} |\rho(k)|$$

does not converge.

In particular, the latter condition applies when the ac.f. is of the form $\rho(k) \sim Ck^{2d-1}$ as $k \to \infty$, where X is a nonzero constant, and $0 < d < 0.5$. It can be shown that a stationary ARFIMA model with differencing parameter d in the range $0 < d < 0.5$, as an ac.f. ρ_k whose limiting form has the required structure.

As an example, Fig. 3.8 shows a simulated ARFIMA$(1, d = 0.4, 2)$ series,

$$(1 - 0.3B)(1 - B)^d X_t = (1 - 0.3B + 0.5B^2)Z_t, \; Z_t \sim N(0, 0.2^2),$$

and its correlogram in R. Note that a specific package, fracdiff, needs to be loaded before calling the function fracdiff.sim.

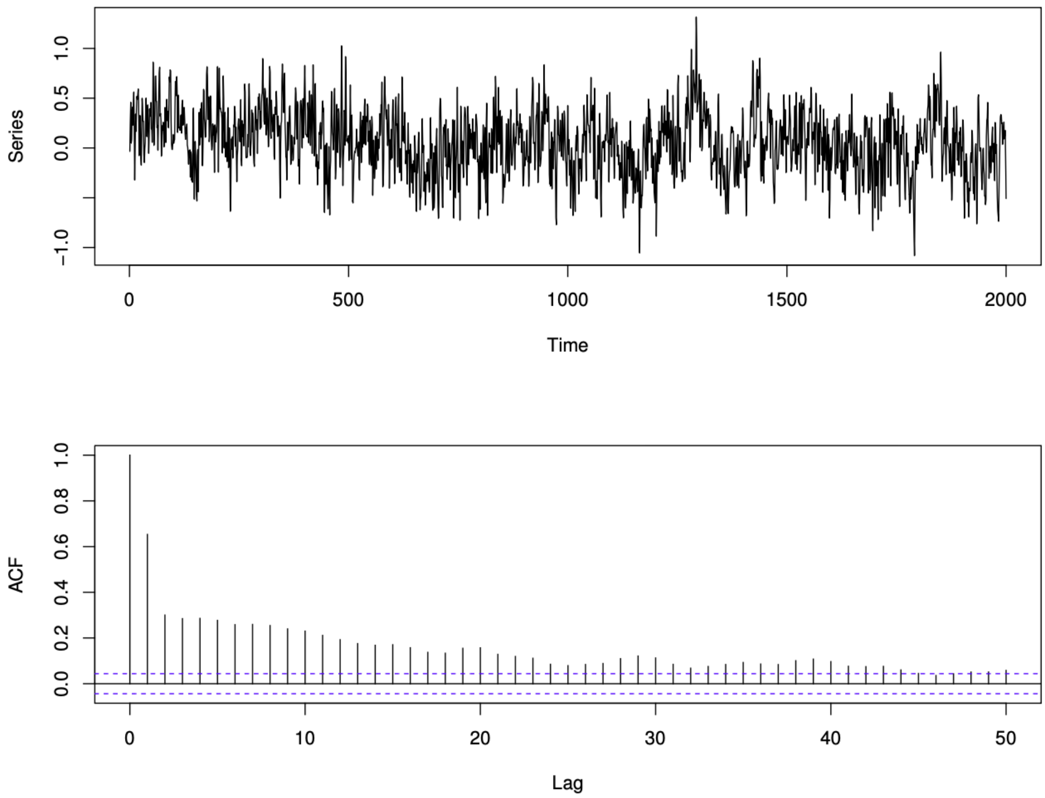


Figure 3.8: A simulated ARFIMA$(1, d = 0.4, 2)$ process, $(1 - 0.3B)(1 - B)^d X_t = (1 - 0.3B + 0.5B^2)Z_t$ with $Z_t \sim N(0, 0.04)$, and its correlogram.

In contrast, the ac.f. of a stationary ARMA process satisfies the condition that $|\rho(k)| < C\lambda^k$, where C is a constant and $0 < \lambda < 1$. Thus there correlations are absolutely summable and such processes may be called short-memory models.

Long-memory models have a number of interesting features, especially for forecasting. In time-series analysis, the variance of a sample mean can be expressed as

$$\frac{\sigma_Z^2}{N}[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{n}\right)\rho(k)];$$

see Section 4.1.2. When the correlationas are positive, as they usually are, the latter expression can be much larger then σ^2/N, especially for long-memory processes where the correlations die out slowly. In contrast to this result, it is intuitively clear that the larger and longer lasting the autocorrelations, the better will be the forecasts of the model. It can be shown both theoretically and practically (Bera, 1994, Section 8.7), but is not shown in this book.

A long-memory (stationary) process and a non-stationary process

It's hard to distinguish between a long-memory (stationary) process and a non-stationary process. A feature of both models is that the empirical ac.f. will die out slowly and the spectrum will large at zero frequency. Fig. 3.9 shows a demonstration.

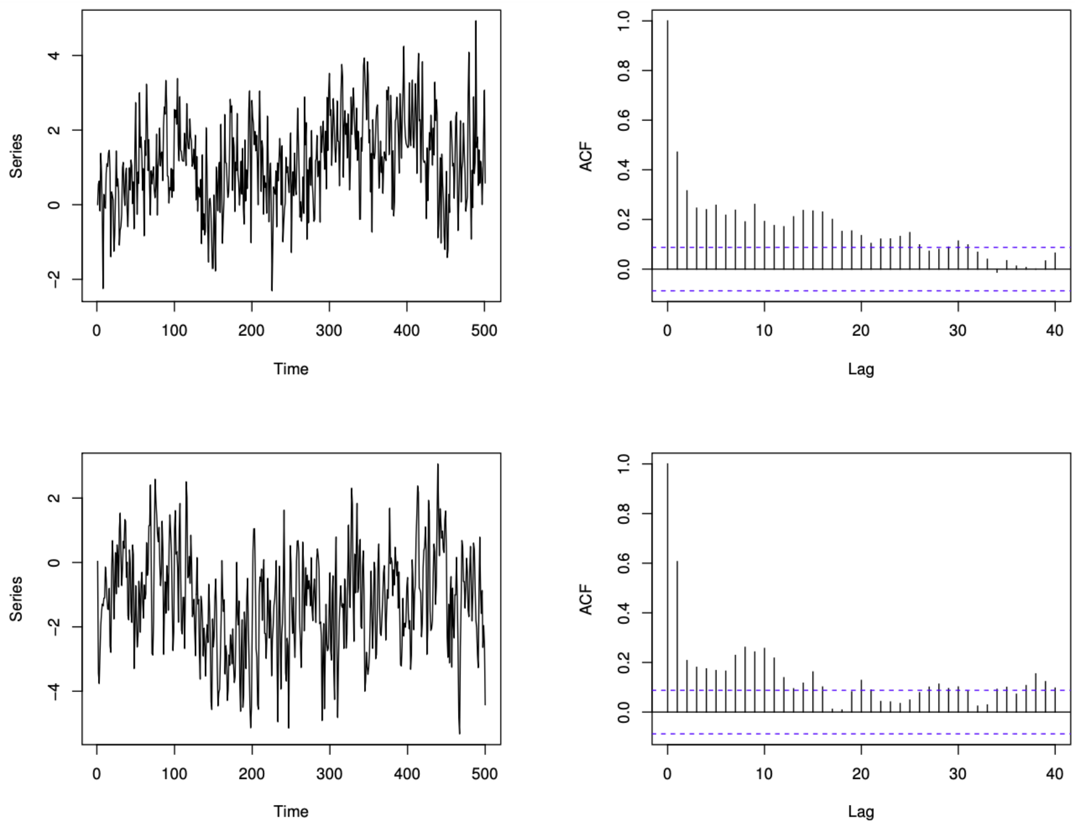


Figure 3.9: A simulated non-stationary process and a simulated ARFIMA$(1, d = 0.45, 2)$ process and their correlograms. Top: $(1 - 0.3B)(1 - B)X_t = (1 + 0.9B)Z_t$; Bottom: $(1 - 0.3B)(1 - B)^d X_t = (1 - 0.3B + 0.5B^2)Z_t$, $Z_t \sim N(0, 1)$.

Both the series themselves and their correlograms look quite similar. Therefore, given a set of data with these properties which seems to be non-stationary, it may be worth considering a fractional ARIMA model with $0 < d < 1$, as well as an ordinary model with $d = 1$. Now the question: will the resulting forecasts likely to be better than those from alternative models? Currently research is still being done on this subject.

3.11 The General Linear Process

Definition. A stationary process is called a general linear process is it can be written as an MA process, i.e.

$$X_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i}.$$

A sufficient condition for the sum to converge, and hence for the process to be stationary, is that the coefficients are absolutely summable. Of course, stationary AR and ARMA processes can also be expressed as a general linear process using the duality between AR and MA processes.

3.12 Continuous Processes

Let's give a brief introduction in the difficulties of continuous time series.

As an example, we consider a first-order AR process in continuous time. A first order AR process in discrete time can be rewritten as

$$X_t = \alpha_{t-1} + Z_t \implies (1-\alpha)X_t + \alpha\nabla X_t = Z_t,\ \nabla X_t = (1-B)X_t.$$

Differencing in discrete time corresponds to differentiation in continuous time, hence a natural way to define an AR(1) process in continuous time is via the equation

$$aX(t) + \frac{dX(t)}{dt} = Z(t),$$

where a is a constant, and $Z(t)$ denotes continuous white noise. However, $\{Z(t)\}$ cannot physically exist, hence let's rewrite the above:

$$dX(t) = -aX(t)dt + dU(t)$$

where $\{U(t)\}$ is a process with orthogonal increments such that the random variables $U(t_2) - U(t_1)$ and $U(t_4) - U(t_3)$ are uncorrelated for any two non-overlapping intervals (t_1, t_2) and (t_3, t_4). In the theory of Brownian motion, this is within the Ornstein-Uhlenbeck model and is sometimes called the Langevin equation. It can be shown that the process defined above has the ac.f.

$$\rho(\tau) = e^{-a|\tau|},$$

which is similar to the ac.f. of an AR(1) process in discrete time in that both decay exponentially.

Rigorous studies of continuous process require considerable mathematical theory, including a knowledge of stochastic integration. Hence this is not covered in this book.

3.13 The Wold Decomposition Theorem

The Wold Decomposition Theorem is mostly of theoretical interest.

Theorem (Wold's Decomposition)**.** Any discrete-time weakly stationary process can be expressed as the sum of a deterministic and a stochastic time series, i.e.

$$X_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} + \eta_t.$$

where

- X_t is the time series being considered,
- ε_t is an uncorrelated sequence which is the innovation process of X_t,
- b is the *possibly infinite* vector of moving average weights,
- η_t is a deterministic time series, in the sense that it is completely determined as a linear combination of its past values.

While the concept of a purely indeterministic process may sometimes be helpful, the Wold decomposition itself can be of little assistance. For a linear purely indeterministic process like AR or ARMA, trying an MA(∞) model is inappropriate, since there are too many parameters to estimate. For processes generated in a nonlinear way, the Wold decomposition is usually even of lest interest, as the best predictor is often nonlinear.

For example, consider

$$X_t = g\cos(\omega t + \theta),$$

where g is a constant, $\omega \in (0, \pi)$ is the frequency of the process, and $\theta \sim \text{Unif}(0, 2\pi)$ is a random variable, called the phase. Note that we need the term θ so that

$$\mathbb{E}[X_t] = 0 \text{ for all } t.$$

If this is not done, X_t will not be stationary.

As θ is fixed for a single realization, once enough value of X_t have been observed to evaluate θ, all subsequent values of X_t are determined. Then we can see that X_t defines a deterministic process. However, it is not 'purely deterministic' because it is nonlinear. In fact, it is 'purely stochastic' using a linear predictor!

4 Fitting Time Series Models in the Time Domain

5 Forecasting

6 Stationary Processes in the Frequency Domain

7 Spectral Analysis

8 Bivariate Processes

9 Linear Systems

10 State-Space Models and the Kalman Filter

11 Non-Linear Models

12 Volatility Models

13 Multivariate Time Series Modeling

14 Some More Advanced Topics