Chapter 1

Time Series: An Introduction

1.1 Introduction

A *time series* is a sequence of observations recorded in time order. Such data occur in numerous contexts, notably economics, meteorology, finance and marketing, and there is much theory regarding their study and interpretation. We will detail some of the most important issues in the chapters to follow, though the interested reader will find several books on the subject in the UBC library.

Example 1 *The* Beveridge Wheat Price Index *was made up of wheat prices in about fifty European countries taken yearly between 1500 and 1869. Many trade and agricultural policies were based on these data, and so the time series is of great interest to economic historians.*

Time series come in a range of varieties. For instance, two potentially key features characterizing a time series might be

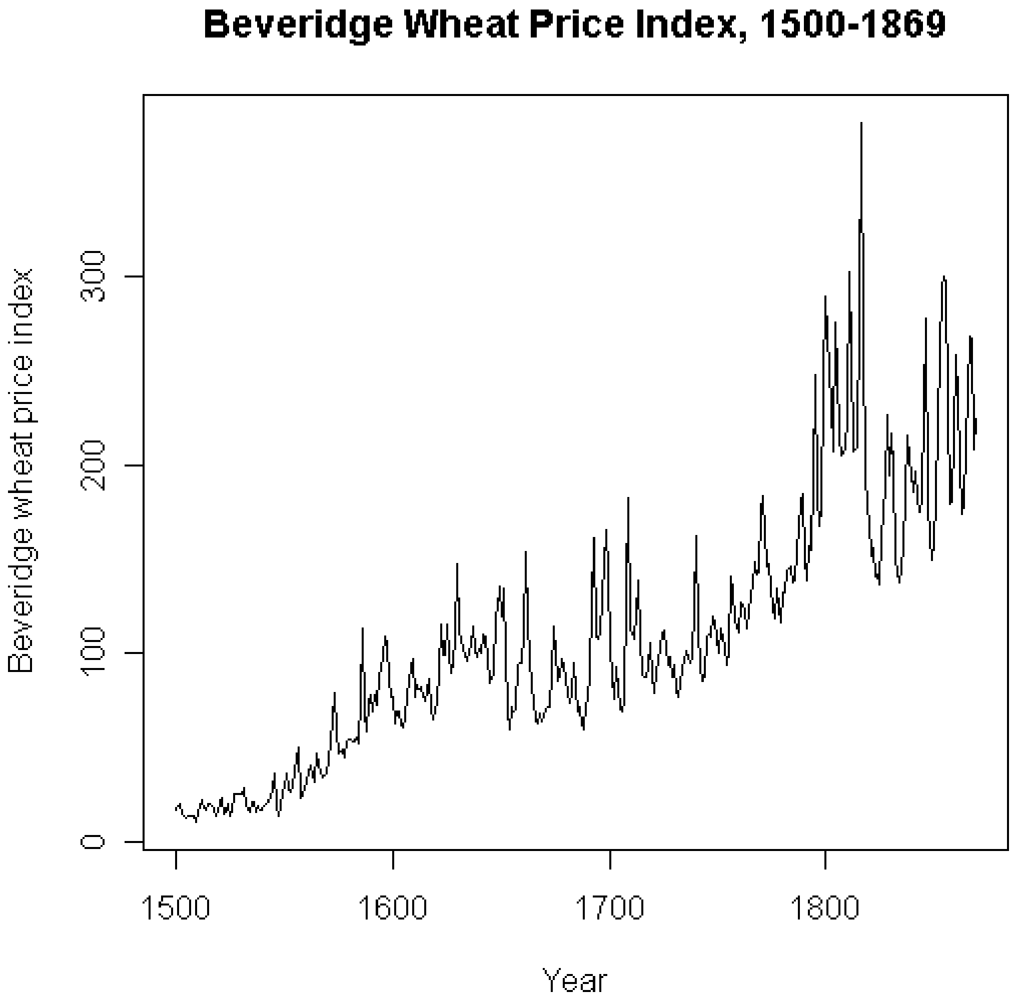
1. whether the *values* the time series can take are discrete or continuous;

2. whether the *time scale* on which the series was measured was discrete or continuous.

In fact the first feature above makes little difference in what is to follow, as although special methods might be adopted for series which are "very discrete" (such as those which are binary – only two values being possible), in general no special theory is required for the handling of discrete–valued series as opposed to continuous–valued ones. The second point is more important. In what follows we will assume that a time series under scrutiny has been recorded in *discrete time*. That is to say, we can label the time points, 1, 2, $\ldots, N$ say, at which we have observations. This makes life a little easier, though the extension to continuous time series is no major theoretical leap. In practice, discrete time series are more common.

Statistics is all about *modelling* things. It is about modelling processes that produce numbers. So in Statistics we tend to have two things under consideration at all times, namely the *model* and the *data*. It is important to have notation that distinguishes between the two entities, so that we are clear when we are referring to properties of the model and when we are referring to something to do with the data. In what follows we will denote the model for a time series by $X\left(t\right)$. This is consistent with notation that you have met in Stat 200 – upper case letters for random variables – but note that the model is a function of t, which is time. The data will be a realizations from some model, and we denote the data by $x\left(t\right).$ That is to say, the sequence $x\left(1\right),\ldots,x\left(N\right)$ is a sequence of numbers, and we are interested in the model $X\left(t\right)$, say, which underlies the process which produced the numbers.

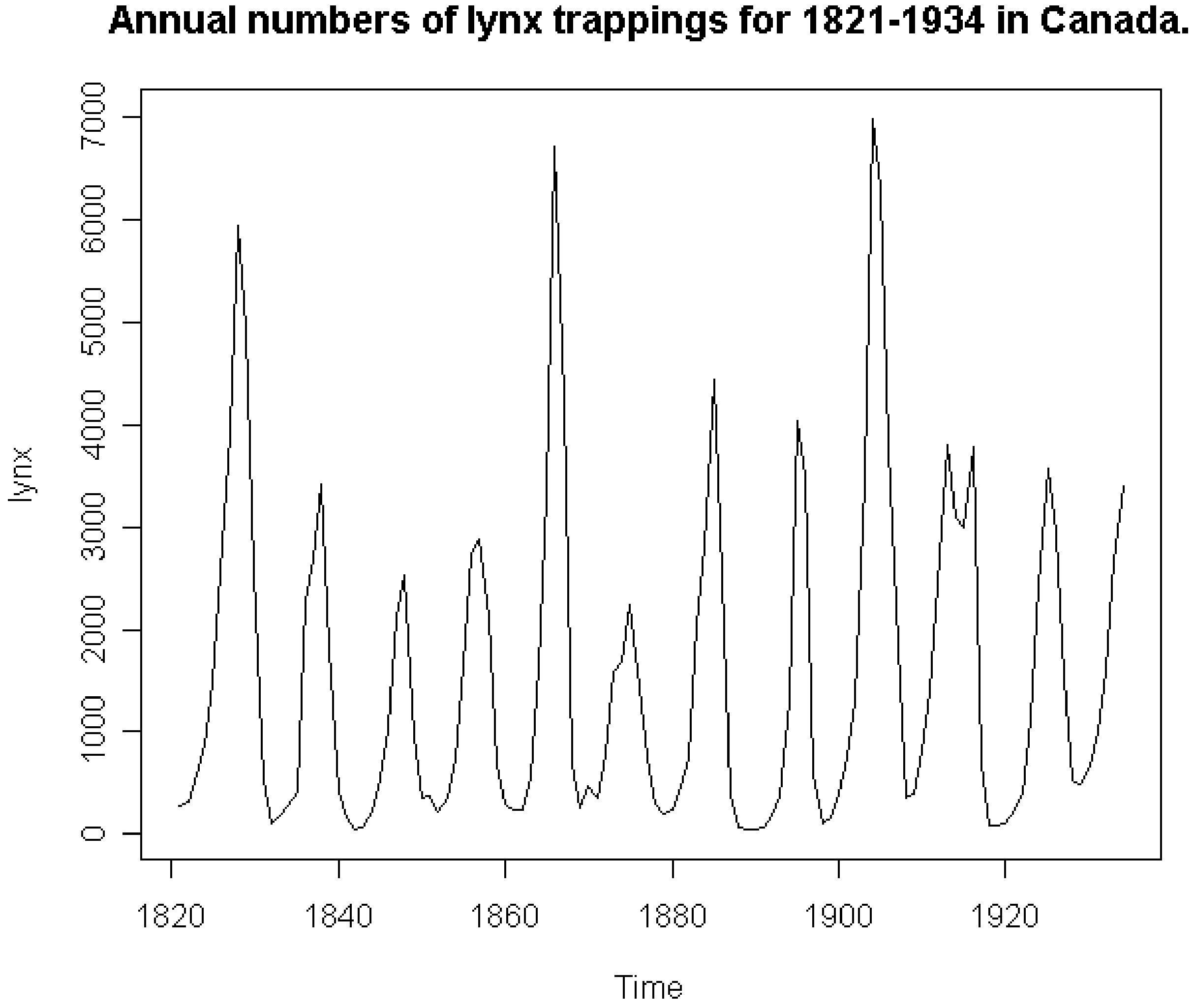
The key difference in the study of time series compared to data studied in earlier statistics courses is that the assumption of *independence* between the values $X\left(1\right),\ldots,X\left(N\right)$ will not usually be reasonable. It is the modelling of the *temporal* relationships in the series that sets the subject apart from other areas of statistics. Our motivation in studying a series may be to explain the variation present, possibly with a view to predicting future values.



1.2 Plotting time series

Given a time series $x(1), \ldots, x(N)$, it is a sensible idea to plot the data against t. This is almost certainly the first thing one should do when presented with a time series, as it will bring out the important features. Care must be taken with the choice of scale and labelling of the axes.

Example 2 *The number of lynx caught in traps in Canada between 1821 and 1934 comprise a much–studied time series.*



1.3 Terminology

There are some terms that will crop up frequently in what follows. We do not give formal definitions as such here, but list the words and give approximate meanings.

1. Roughly speaking, a *trend* is a long term change in the mean of the series, i.e., a tendency to go up or down.

2. A *seasonal effect* is a regular variation with “season”, where “season” might be month, day, year or whatever. Sales figures and temperature readings might be expected to exhibit such fluctuations. Such effects are presumed to be *periodic*, and in some sense predictable. Some series might exhibit other *cyclical effects*, which give oscillations of possibly unknown cause and variable period.

3. An *outlier* is a value which is seemingly inconsistent with the majority of the remainder of the series.

4. A time series is said to be *stationary* if, loosely, it has constant mean and variance and is without any periodic components. This property is often highly desirable, and there are operations we can perform on series to make them (look) stationary. We will give a formal definition of what this word means in the next chapter.

5. A *transformation* is an operation performed involving every element in a series, often used to stabilize the variance or to make certain components in the series easier to model. For example, it can be shown that a series whose variance appears to be increasing with time can usually be stabilized by either square rooting each value or taking logarithms.

6. Many time series models involve, as a building block, a *white noise* process, which we denote $Z\left(t\right).$ This is assumed to be a purely random series, of zero mean and constant variance. Often a white noise series might be thought of as an independent and identically distributed (i.i.d. for short) sequence of $N\left(0,\sigma^2\right)$ values.

1.4 Series with a trend

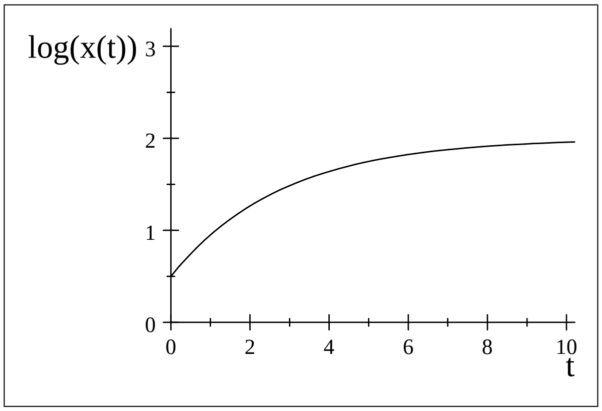
Confronted with a series with an obvious trend, we might want to (i) measure the trend, (ii) remove the trend or (iii) both (i) and (ii). One rather crude approach to this is *curve fitting*, where we attempt to fit a line or curve through the points. In doing this we would usually make use of a computer to estimate the unknown coefficients in the model we fit.

Some examples of families of curves are:

1. Linear: $x(t) = a + bt$.

2. Quadratic: $x(t) = a + bt + ct^2$.

3. Gompertz: $\log(x(t)) = a - br^t$, for $|r| < 1$.



$\log(x(t)) = a - br^t$, $a = 2$, $b = 1.5$, $r = 0.7$.

4. Logistic: $x(t) = a/(1 + be^{-ct})$.

The last two curves above both approach an asymptotic value as $t \to \infty$. They might be fitted after transforming the data.

Exercise 1.4.1 *Using a computer package of your choice, plot examples of Gompertz and Logistic curves.*

The main problem with curve fitting is that it is not very *dynamic*, inasmuch as it assumes that the same curve fits the data at all time points. In cases where it works well, however, there is often little need for further analysis.

We will later address other ways of extracting the trend from a series.

1.5 Series with seasonal variation

As with trend, we might want to estimate seasonal variation and/or remove it. For series with no (or very little) trend, we might model the process by

$$X(t) = \mu + S(t) + \varepsilon(t),$$

where μ is the underlying mean, $S(t)$ is the seasonal effect and $\varepsilon(t)$ is the random component. This says that the seasonal effect is *additive.*

Alternatively the seasonal effect may be *multiplicative* rather than additive. A model for this would be

$$X(t) = \mu(t)\, S(t)\, \varepsilon(t),$$

The above model might be used when the amplitude of the seasonal effect increases with the mean. Here, taking logarithms would be sufficient to convert the series to one with an additive seasonal term.

On the other hand, the random component may be additive, giving the model

$$X(t) = \mu(t)\, S(t) + \varepsilon(t).$$

To fix ideas, it is often convenient to make two assumptions about $S(t)$:

1. The function $S(t)$ has *period* p, that is, it repeats after p time units, so

   $$S(t+p) = S(t)$$

   for all t. As we often only have a few cycles in the data, this is usually the only possible assumption.

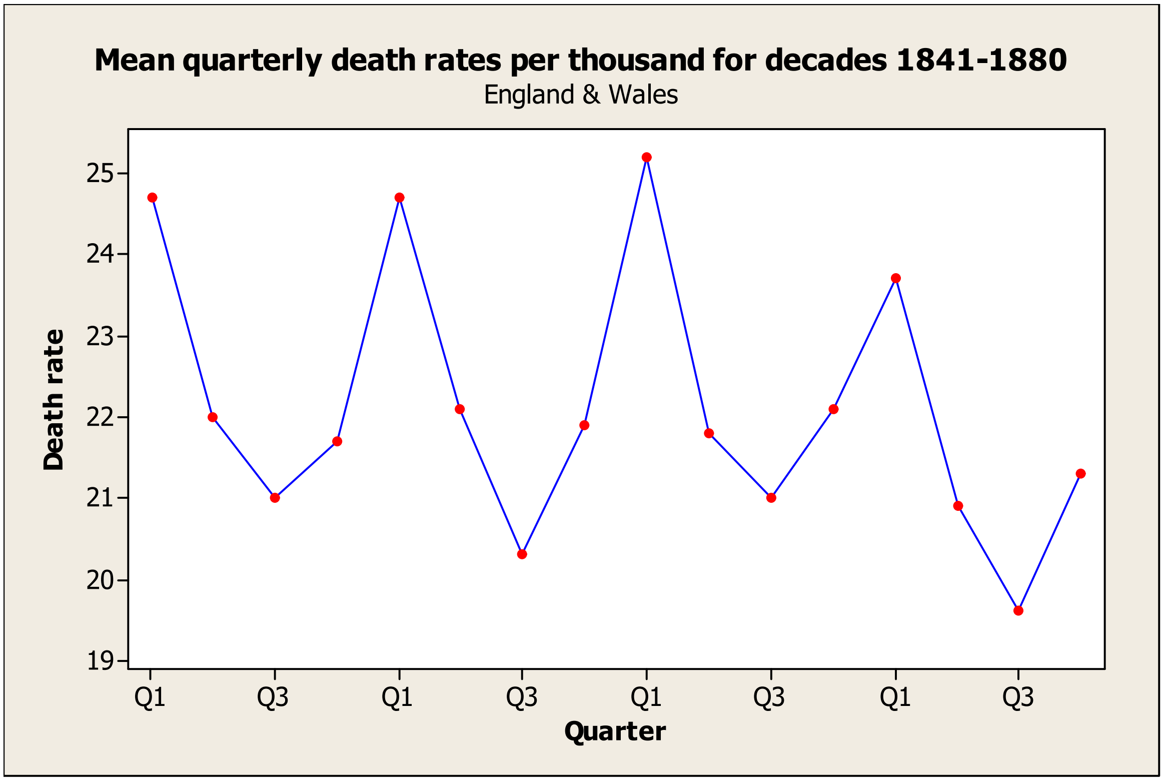
2. The sum of the seasonal effects over a complete cycle (or period) is zero, i.e.,

   $$\sum_{j=1}^{p} S(t+j) = 0.$$

Assuming a model of the above form, a simple approach to estimating the function $S(t)$ is just to average the values over each time in the periodic cycle and use these, minus the sample mean (our estimate of μ) as the seasonal estimates. So to estimate a seasonal effect find the average in that period (say, the first quarter) and subtract the yearly average (in the additive case, or in the multiplicative case *divide* by the yearly average).

Example 3 *The death rates, per thousand people, for England and Wales during year-quarters of four decades of the nineteenth century are given below.*

		Quarter			
		1	*2*	*3*	*4*
Decade	*1841-50*	*24.7*	*22.0*	*21.0*	*21.7*
	1851-60	*24.7*	*22.1*	*20.3*	*21.9*
	1861-70	*25.2*	*21.8*	*21.0*	*22.1*
	1871-80	*23.7*	*20.9*	*19.6*	*21.3*
		98.3	*86.8*	*81.9*	*87.0*



As might be expected, there is a seasonal effect apparent, with seemingly more people dying in the first quarter (after the Winter, ending in March). There is little evidence of a trend, however. The means in each of the quarters are 24.58, 21.70, 20.48 and 21.75 respectively, the overall mean being 22.13. So we could take estimates of our seasonal effects to be:

$$\begin{aligned} S(1) &= 24.58 - 22.13 = 2.45, & S(2) &= 21.70 - 22.13 = -0.43, \\ S(3) &= 20.48 - 22.13 = -1.65, & S(4) &= 21.75 - 22.13 = -0.38. \end{aligned}$$

Note that the values sum (approximately) to zero. We could now subtract the values above from the original in the appropriate quarter to leave the mean (assuming no trend is present) and the random component.

In cases where a trend in present, the above approach will not be adequate. We come shortly to methods for handling these cases.

1.6 Operations on time series

There are several classes of operations we can perform on a time series to bring out facets of interest. We give details of a few here.

1.6.1 Filtering

This involves creating a new series

$$\mathrm{Sm}(t) := \sum_{r=-q}^{s} a_r x(t+r)$$

where $\{a_r\}$ are weights such that (usually) $\sum_{r=-q}^{s} a_r = 1$. It is often the case that $s = q$ and $a_r = a_{-r}$ for all r, so that the weights are *symmetric.* For example, we might take

$$a_r = \frac{1}{2q+1},$$

for $r = -q, \ldots, q$. Another option is to take $\{a_r\}$ as the terms in the expansion of $\left(\frac{1}{2} + \frac{1}{2}\right)^{2q}$, which are Normal distribution like in shape, when plotted against r. Notice that in any case the new series has $2q$ fewer points than the original.

The aim of filtering, sometimes called *smoothing,* is to reduce the local variation, hence making a trend easier to estimate (and therefore remove). It can also be used to remove a seasonal effect, of period p say, by forming the *moving average* series

$$\frac{1}{p}\sum_{j=0}^{p-1} x(t+j), \frac{1}{p}\sum_{j=1}^{p} x(t+j), \frac{1}{p}\sum_{j=2}^{p+1} x(t+j), \ldots$$

As we have assumed that the sum of the seasonal effects is zero over a cycle, the above series will not have a seasonal effect since each term covers all the times in a cycle. One slight problem arises when the period of the seasonal effect is *even*, so the moving average involves averaging an even number of terms each time, for then the new series formed is for time points which do not correspond to those in the original series, being naturally positioned between two points in the original series. There are two ways around this:

1. Re-weight the moving average series to include an extra term, but give half-weight to the terms at each end of the cycle. For example, for quarterly data we could use
$$\mathrm{Sm}\left(x\left(t\right)\right)=\frac{\frac{1}{2}x\left(t-2\right)+x\left(t-1\right)+x\left(t\right)+x\left(t+1\right)+\frac{1}{2}x\left(t+2\right)}{4}.$$
2. Alternatively we could *centre* the moving average, by forming a new series that averages consecutive values of the smoothed one (i.e., takes a moving average of *order* 2 on the first moving average series). This takes us back to the right time points, but loses an observation from either end.

Given the (centred) moving average series, we could estimate the seasonal effects by looking at the differences between the moving average and the original series. To estimate the kth quarter seasonal, for example, we could just average the kth quarter differences, for $k=1,2,3,4$.

Example 4 *Returning to the death rate data, below we have the original series in the first column, the four point moving average in the second column, and the centred version in the third column. Using the centred MA, we can estimate the seasonal effects by simply averaging the differences between it and the original data for each of the four quarters. Hence*
$$\begin{aligned}S\left(1\right)&=2.392, & S\left(2\right)&=-0.467,\\ S\left(3\right)&=-1.567, & S\left(4\right)&=-0.346.\end{aligned}$$
For example,
$$S\left(1\right)=\frac{2.4125+2.815+1.950}{3}.$$

Since the sum of these is 0.0122 and we require the seasonal effects to sum to zero, we can deduct a correction factor of

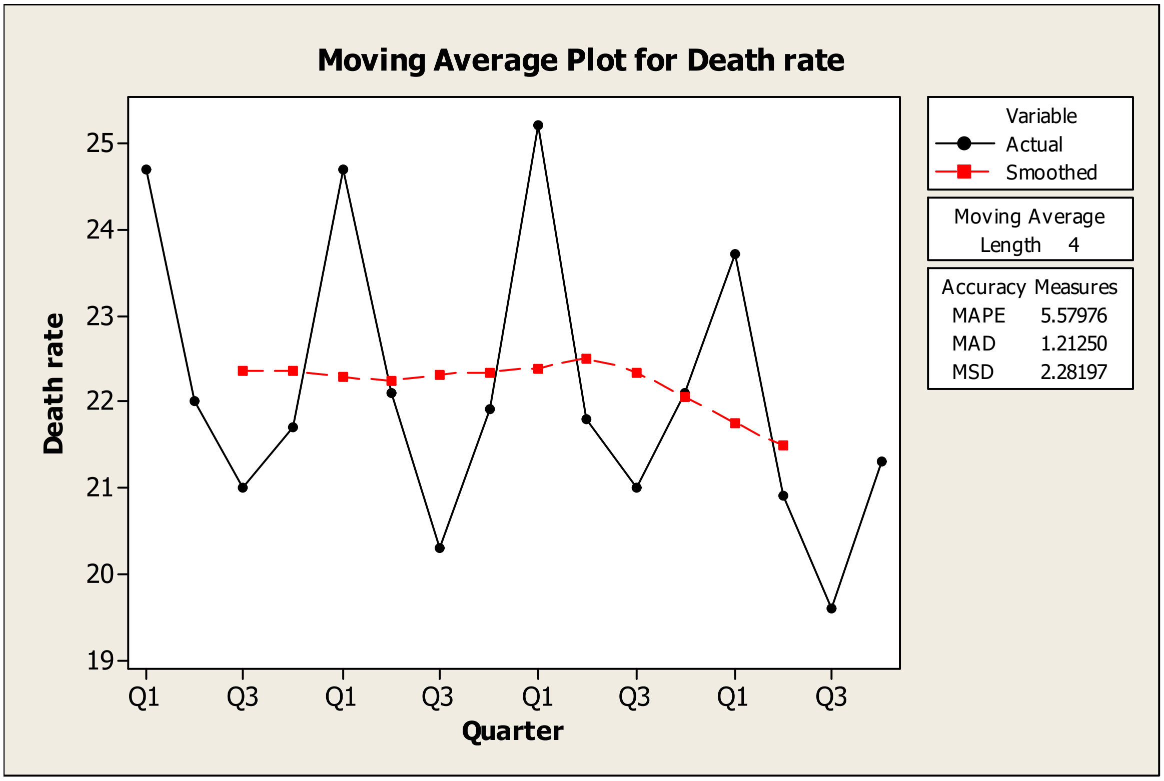
$$\frac{0.0122}{4} = 0.00304$$

from each. This leads to revised seasonal estimates of the seasonals as (to 3 d.p.)

$$\begin{array}{ll} S(1) = 2.389, & S(2) = -0.470, \\ S(3) = -1.570, & S(4) = -0.349. \end{array}$$

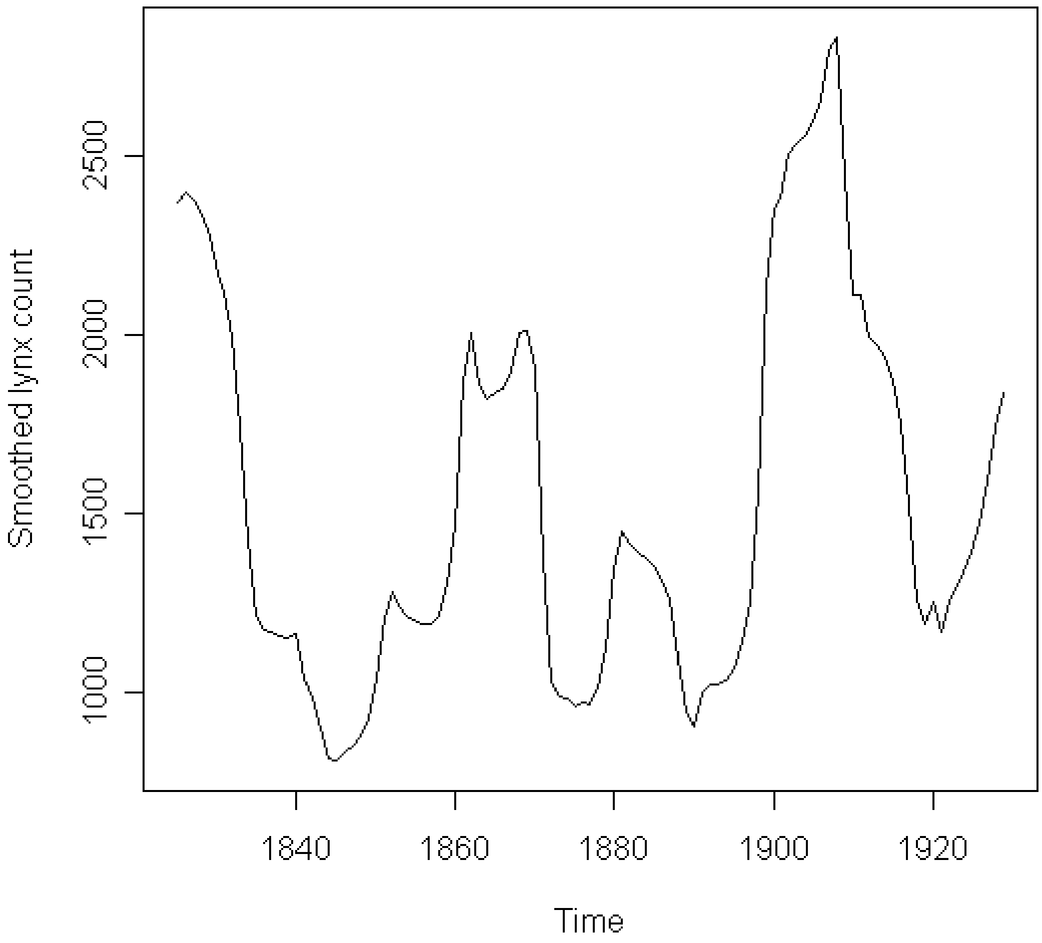
In the table below the columns are the raw data, the four–point moving averaged data and the centred version, $Cm(x(t))$ say, respectively.

$x(t)$	$Sm(x(t))$	$Cm(x(t))$
24.7		
22.0		
	22.35	
21.0		22.35
	22.35	
21.7		22.36
	22.37	
24.7		22.29
	22.20	
22.1		22.22
	22.25	
20.3		22.31
	22.37	
21.9		22.38
	22.30	
25.2		22.39
	22.47	
21.8		22.50
	22.52	
21.0		22.34
	22.15	
22.1		22.04
	21.92	
23.7		21.75
	21.57	
20.9		21.47
	21.37	
19.6		
21.3		



Example 5 *Below is a plot of the smoothed version of the Canadian lynx*

data, taking a ten point moving average.



Canadian lynx data, smoothed via a ten-point moving average

A special case of filtering that we will return to in the topic of prediction is *exponential smoothing.* This takes an *asymmetric* filter of the form

$$\mathrm{Sm}\left(x\left(t\right)\right)=\sum_{j=0}^{q}a_{j}x\left(t-j\right)$$

with $q=\infty$ and

$$a_{j}=\alpha\left(1-\alpha\right)^{j}$$

for some $\alpha\in\left[0,1\right]$.

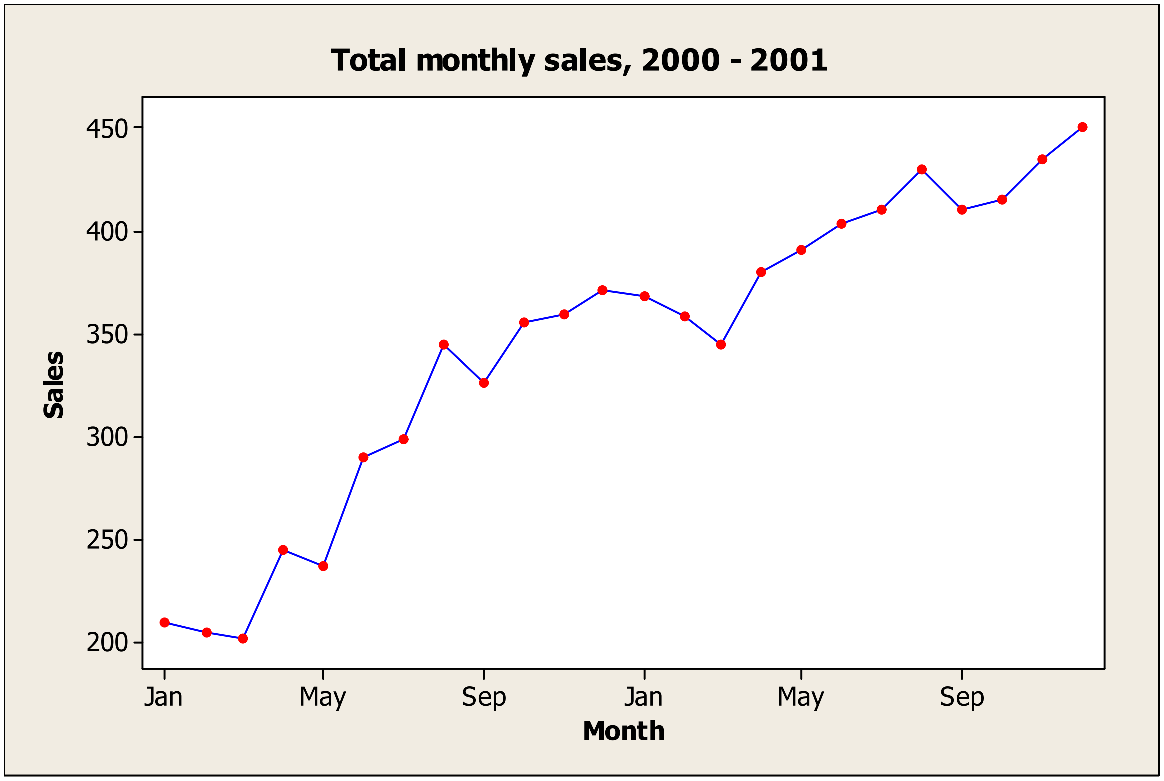
1.6.2 Differencing

A simple method which is often effective for the removal of trends and seasonal effects is the application of the *difference operator* ∇, defined by

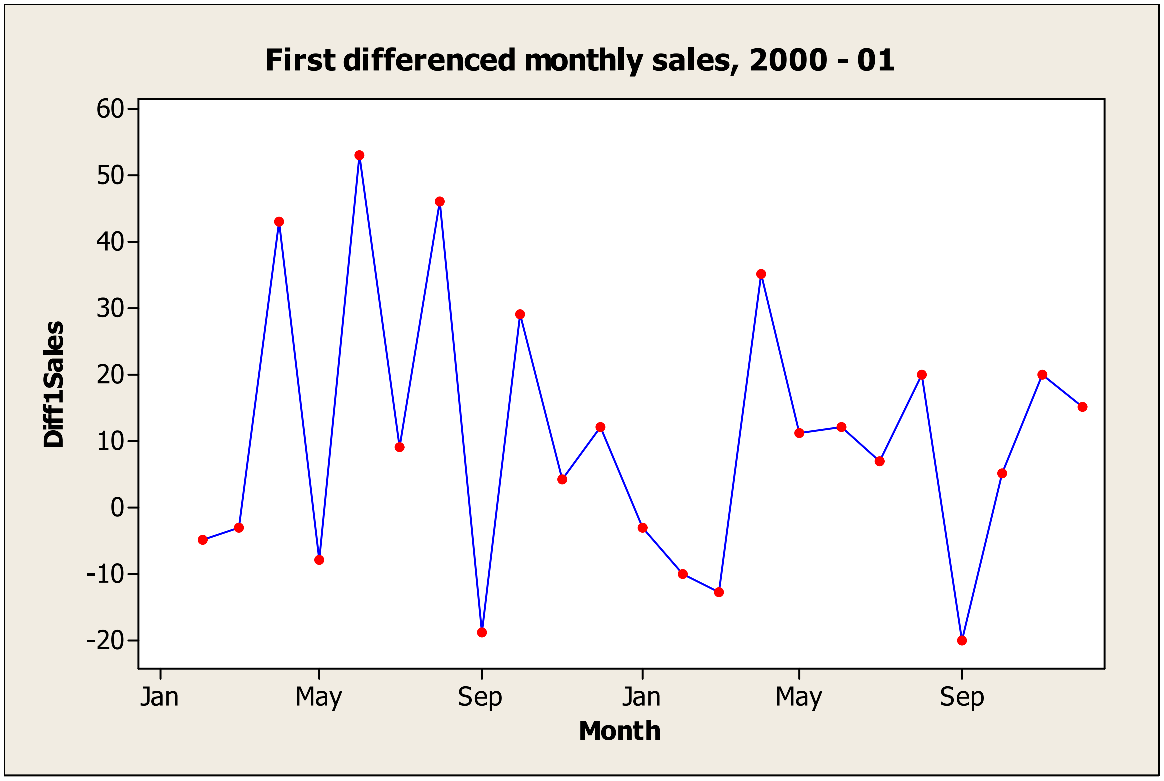
$$\nabla x(t):=x\left(t\right)-x\left(t-1\right).$$

This produces a new series of first differences $y\left(t\right):=\nabla x\left(t+1\right), t=$1,$\ldots,N-$ 1.

Example 6 *The series below has an apparent trend:*



Below is the first–differenced series:



Occasionally we may need to difference twice, and can apply the operator repeatedly in these cases, i.e.,

$$\begin{aligned}\nabla^2 x(t+2) &= \nabla x(t+2) - \nabla x(t+1) \\ &= x(t+2) - 2x(t+1) + x(t)\,.\end{aligned}$$

Seasonal effects can sometimes be removed by appropriate *seasonal differencing*; for example, monthly data might be de-seasonalised by taking

$$\nabla_{12} x(t) := x(t) - x(t-12)\,.$$

This is taking first differences at *lag* 12. The new series so defined *may* be non-seasonal.

1.6.3 Transformations

As we have already briefly mentioned, there are sometimes reasons why we might want to *transform* our original series. One typical example is when we suspect that the seasonal effect is *multiplicative* rather than additive. A model for such a situation would be

$$X\left(t\right)=\mu\left(t\right)S\left(t\right)\varepsilon\left(t\right),$$

where $\mu(t)$ is the mean at time t, $S\left(t\right)$ is the seasonal effect and $\varepsilon\left(t\right)$ the random term. The above model might be used when the amplitude of the seasonal effect increases with the mean. Here, simply taking logarithms would be sufficient to convert the series to one with an additive seasonal term, which may be easier to handle.

1.7 The autocorrelation function

Recall that for a set of N paired observations $\left(x_1,y_1\right),\ldots,\left(x_N,y_N\right)$ we define the *sample correlation coefficient* to be

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

where $\bar{x}$ is the mean of the x values, $\bar{y}$ the mean of the y's. Should you have forgotten the properties that r possesses, it is sensible to remind yourself forthwith, but suffice to say that r lies between -1 and $+1$, and is a measure of the linear relationship between the two variables in question.

Given a time series $x\left(1\right),\ldots,x\left(N\right)$ it is natural to look at the relation between *consecutive* values. To this end, consider the $N-1$ pairs of consecutive values $\left(x\left(1\right),x\left(2\right)\right),\left(x\left(2\right),x\left(3\right)\right),\ldots,\left(x\left(N-1\right),x\left(N\right)\right),$ and define the *autocorrelation* coefficient for the series at *lag* 1 as

$$r_1:=\frac{\sum_{t=1}^{N-1}\left(x\left(t\right)-\bar{x}_{(1)}\right)\left(x\left(t+1\right)-\bar{x}_{(2)}\right)}{\sqrt{\sum_{t=1}^{N-1}\left(x\left(t\right)-\bar{x}_{(1)}\right)^2\sum_{t=1}^{N-1}\left(x\left(t+1\right)-\bar{x}_{(2)}\right)^2}},$$

where

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

is the mean of the first $N-1$ observations and

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

is the mean of the second $N-1$ observations. Now the above formula for r_1 is a little unwieldy, but note that $\bar{x}_{(1)} \approx \bar{x}_{(2)} \approx \bar{x}$ and $(N-1)/N \approx 1$ (at least for N reasonably large) so

$$r_1 \approx \frac{\sum_{t=1}^{N-1} (x(t) - \bar{x})(x(t+1) - \bar{x})}{\sum_{t=1}^{N} (x(t) - \bar{x})^2}.$$

This is the formula we will use to define r_1 in what follows.

In a similar fashion, we define the autocorrelation function (acf for short) at lag k to be

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

If the *autocovariance function* at lag k is defined to be

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

then clearly

$$r_k = \frac{c_k}{c_0}.$$

Note that in practice, although it is possible to calculate r_k for values of k up to $N-1$, there is little point in looking at the acf for lags above about $N/4$, as such correlations are unlikely to provide useful information.

1.7.1 The correlogram

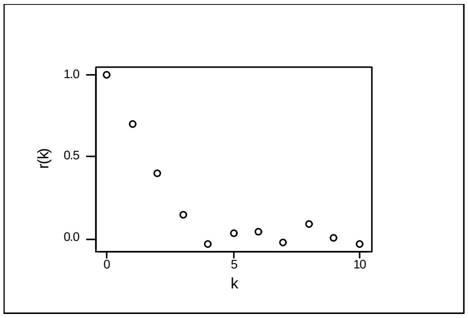
The correlogram is a plot of the acf r_k against the lag k. This is often a useful plot to examine for a time series.

Be aware for a completely random series, that approximately for reasonably large values of N we have

$$r_k \sim N\left(-\frac{1}{N}, \frac{1}{N}\right).$$

For such a series, roughly one in twenty values of r_k would be expected to lie outside the interval $\left(-\frac{1}{N} - \frac{2}{\sqrt{N}}, -\frac{1}{N} + \frac{2}{\sqrt{N}}\right)$.

More usually the acf will appear to decay with the lag, looking something like



The above indicates short–term correlation within the series, decaying off by about lag 3.

We briefly summarize what the acf will look like in some particular cases:

1. If the time series *alternates*, that is, consecutive values tend to be on opposite sides of the mean, then the acf will alternate. The value of r_1 will be negative, whilst the value of r_2 will most likely be positive, since every other value is on the same side of the mean.

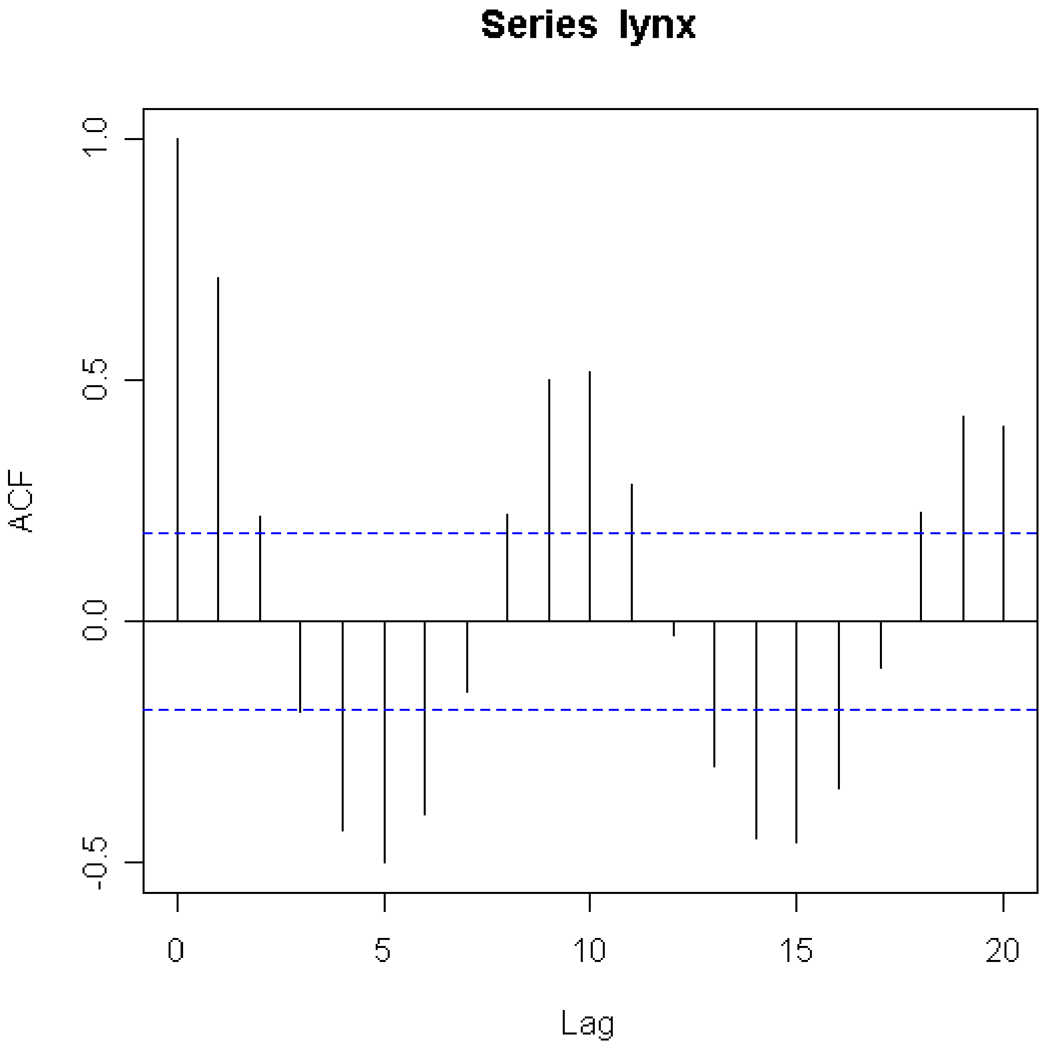
2. If the series has a trend, either up or down, the acf will not decay to values close to zero very quickly. This is basically because values in the time series close to one another in time will be tending to lie on the same side of the mean, and so be quite highly correlated.

3. A series with a seasonal fluctuation will have an acf with an oscillation at the same frequency. For example, for monthly data with a seasonal effect, r_{12} could be expected to be relatively large and positive.

It should be pointed out that in none of the above three cases does the acf give much more useful information about the time series than a simple plot of the series would yield. In general, experience is required to glean much from an acf plot, though we will return to acf properties for certain models in the next chapter.

Example 7 *Below is the acf for the original Canadian lynx data given ear-*

lier.



1.8 Time series in Minitab

The software package Minitab contains various features for processing and analyzing time series. All can be found by clicking Stat→Time Series on the menu bar. In truth, providing you are aware of what you are trying to do with the appropriate feature, the dialogue boxes are fairly self–explanatory, and pressing the “Help” button always provides a reasonably adequate description of what an option performs. Nevertheless we include here a brief summary of the commands that are useful to us so far.

1. Time series plot: As you would surmise, this plots the numbers in a column as a time series, and assumes by default that the observations are equally spaced in time. There is an option for times and dates to be stored in a separate column to appear as labels on the horizontal axis if required.

2. Trend analysis: This fits a trend model to a time series, of one of various forms, including linear and quadratic.

3. Decomposition: This is for seasonal data, and requires the period length to be entered (e.g., 12 for monthly data). This will store the trend line (if one is present), estimate the seasonals and give the seasonally adjusted data. Note that by default this command assumes a multiplicative model.

4. Moving average: This is principally for smoothing seasonal data. If the procedure requires a smoothing with an even number of weights, can Minitab centre the resulting series by smoothing with a two–point moving average.

5. Difference: This provides a new series, based on differences of lag k from a series stored in a column.

6. Lag: This moves a series forward in time by a given lag. The new series will have missing value(s) at its start.

7. Autocorrelation: Provides the acf for a series, up to a given number of lags, the default being $N/4$.

1.9 Time Series in R

The R statistical language has a comprehensive collection of tools for analyzing time series, and is probably the most powerful platform for serious numerical analysis in the subject. There are two useful packages to my knowledge, tseries and zoo. The latter allows for series with unequal time intervals between observations and missing values. The packages can be added in R via Packages → update packages. Note you should select a CRAN mirror for installing new packages, such as USA(CA).

With the above packages installed there are then two types of time series objects in R, namely ts and zoo. Check the help facilities on these objects, by entering

```
>library(tseries)
> help(zoo)
```

for example. Nearly always we will be using data taken at equally spaced time intervals, so the zoo facilities are not all required, but the package does increase functionality in other ways.

Certain R functions work differently on time series objects. For example, plot(x) returns a different output when x is a time series object compared to when it is a vector, whilst plot.ts is a similar function specifically for time series objects. The lynx data plot was obtained via

```
> plot(lynx, main = ''Annual numbers of lynx trappings for
1821-1934 in Canada.'')
```

R can compute the smoothed, moving average version of a time series object via the following zoo command:

```
> rollmean(x, k, na.pad = FALSE, align = c(''center'', ''left'',
''right''), ...)
```

in which x is the time series and k the integer width of the averaging process. The align command dictates where in time the smoothed series is located, and na.pad enables "missing" values to be replaced by NA's if required. For various reasons the term "rolling mean" is somewhat better than the "moving average" nomenclature, and commands rollmedian and rollmax also exist with similar syntax.

For computing the acf of a time series (or a vector) x the command

```
> acf(x, lag.max = NULL,...)
```

does the trick, where the "null" value of the maximum lag variable is taken as $10\log_{10}(N)$ unless otherwise defined, where N is the length of the series as usual. The `type` subcommand enables the autocovariance function to be determined if desired.

The command

```
> lag(x, k = 1, ...)
```

takes the time series `x` and creates a new time series lagged `k` units back in time, the default value being one time unit as indicated above. To difference a series, use

```
> diff(x, lag = 1, differences = 1, ...)
```

which creates the once–differenced version of the series in `x`, differences taken at lag 1. Differences can be taken more than once, and at lags other than one by changing the values of the arguments within the `diff` function.

1.10 Learning outcomes

On completion of this chapter, learners should be able to appreciate the important features that describe a time series, and perform simple analyses and computations on series. In particular:

1. (a) Informally define and explain terminology used to describe time series, including trend, seasonal effects, cyclical effects, outlier and white noise.

   (b) Recognize when curve–fitting may be an appropriate method for modelling a series, identifying linear, quadratic, Gompertz, and Logistic models where appropriate.

   (c) Describe models for seasonal variation, including additive and multiplicative models.

   (d) Apply a filter (that is, a smoother) to a time series, centring if necessary.

   (e) Use a filter to estimate the seasonal indices in a time series that has an additive seasonal component.

(f) Define and apply the difference operator, including the operator for seasonal differences.

(g) Recognize the role of transformations for time series, and identify possible transformations to address certain non-stationary features of a series, such as non-constant variance and multiplicative seasonal effects.

(h) Define the sample autocorrelation function and the correlogram.

(i) Describe the behaviour of the correlogram for series that alternate, have a trend, or show seasonal fluctuations.

(j) Use R to perform certain time series analyses including plots, smoothing, computation of the sample autocorrelation function, lagging and differencing.

1.11 Exercises 1

1. Plot the following time series:

0.8	0.2	1.1	1.4	0.9	1.8	1.0	2.1	0.5	2.0	1.1	2.9
2.6	3.2	2.7	2.5	3.4	1.5	2.8	3.3	3.7	2.9	3.1	2.7

   Apply two filters separately to the above series, with (i) five equal weights and (ii) weights that are successive terms in the expansion of $\left(\frac{1}{2}+\frac{1}{2}\right)^4$. In each case, use the smoothed process to estimate the gradient of the trend of the series, by least squares. Comment on your results.

2. The following series $x\left(t\right)$ gives the population size of a controlled population of fruit flies in an experiment lasting 39 days.

t (days)	0	9	12	15	18	21	24	27	30	33	36	39
$x\left(t\right)$	22	39	105	152	225	390	499	547	618	791	877	938

   Plot the data. A logistic model

$$X\left(t\right)=\frac{a}{1+ce^{-abt}}$$

has been suggested for the series. Assuming such a model is valid, use your plot to estimate the carrying capacity a. By transforming the data into linear form, estimate the values of b and c.

3. A filter with weights $\{a_j\}$ acts on a series $x(t)$ to produce a new series $y(t)$. A second filter with weights $\{b_j\}$ then acts on $y(t)$ to give a new series $z(t)$. Express $z(t)$ in terms of $x(t)$. What are the weights of the filter produced by applying the filter $\left\{\frac{1}{2}, \frac{1}{2}\right\}$ followed by $\left\{\frac{1}{3}, \frac{2}{3}\right\}$?

4. The following give the quarterly sales of a university textbook over the years 2003, 2004, and 2005.

		2003	2004	2005
Quarter	1	1690	1800	1850
	2	940	900	1100
	3	2625	2900	2930
	4	2500	2360	2615

   (a) Plot the time series. Describe the key features of the series.

   (b) Based on the plot, what model would you suggest for the data? Define clearly any notation you use.

   (c) After smoothing and de-trending, the series are as below:

		2003	2004	2005
Quarter	1	*	–190.62	–206.25
	2	*	–1107.50	–991.88
	3	672.50	903.75	*
	4	538.75	332.50	*

   Estimate the seasonal indices for this series.

   (d) Explain how you might calculate the residuals for the model fitted. Why might you use these residuals?

   (e) Suggest how you would predict the textbook sales for the next two quarters.

5. Plot the following stationary series:

$$2.7, 1.9, 2.3, 1.6, 3.0, 2.3, 2.2, 1.7, 2.6, 1.8, 2.0, 2.3, 1.6, 2.4, 1.8, 2.3.$$

   Plot $x(t)$ against $x(t+1)$ for the series, and make a guess of r_1. Use R to plot the correlogram for this series.

6. The first ten values $r_1, \ldots, r_{10}$ of the acf of a time series of length 400 are 0.02, 0.05, –0.09, 0.08, –0.02, 0.00, 0.12, 0.06, 0.02, –0.08. Is there any evidence to suggest the series is purely random?

7. Using the `rnorm` command in R, generate a sequence of 100 independent standard Normal values. Calculate the correlogram for this series. Comment on the appearance of this plot.

8. The R packages `tseries` and `zoo` contain various time series datasets. For example, the Beveridge wheat price index can be called via

```
> library(tseries)
> data(bev)
```

   Create a time series plot of the series in R, including a suitable title and sensible labels for the axes. A graphical parameter we may wish to tweak when creating a time series plot is the *aspect ratio,* `asp` in R. For `asp>0`, one unit in the x-axis direction will be equal in length on the graphic to `asp` times one unit in the y-axis. Insert different values for `asp` in your command for creating the plot of the wheat price index: try values 0.5, 0.8, 1, 1.2, and 1.5. Which, if any, aspect ratio do you prefer? Summarise how changing the aspect ratio can affect the visual impact of the plot.

9. Data sets can be read into R, usually without too much difficulty. The file `yearlysnow.txt` contains yearly snowfall and equivalent rainfall measurements in Boston, from 1892 to 2000. Read this file into R by copying it into the `bin` directory from the course web page, and then entering

```
> yearlysnow <- read.table(”yearlysnow.txt”, header=T, sep=”\t”)
```

   If you are unfamiliar with reading datasets into R, use the help facility to understand the command above. A suitable dataset can be coerced into a time series object via the `ts` command. Use this command to create a time series object of the snowfall data – in fact the snowfall figures have to be multiplied by ten to be in inches. Plot the series and comment on its features.

1.12 Revision topics

There are a couple of mathematics topics which will crop up in the next chapters. You will have met these before, perhaps briefly. It is a good idea to spend a short time re-familiarising yourself with them.

1.12.1 Geometric series

A series of the form

$$a + ar + ar^2 + ar^3 + \cdots$$

is called a *geometric series* (or *geometric progression,* sometimes therefore a "G.P." for short). Each term in the sum is the previous term multiplied by r.

The summation above is an infinite sum – there are an infinite number of terms being added together. One might think that such a sum must therefore be ∞ (or perhaps $-\infty$ if all the terms were negative), but in fact that is not necessarily the case. A geometric series of the form above *converges* to a finite limit provided r is suitably small: if $|r|< 1$ (i.e., $-1 < r < 1$) then the geometric series sums to

$$\frac{a}{(1-r)}.$$

For other values of r there is no finite limit possible in the infinite sum – in such cases we will think of the summation as not being properly defined.

The sum of the first n terms can be shown to be

$$\begin{aligned} a + ar + ar^2 + \cdots + ar^{n-1} &= \frac{a\,(r^n - 1)}{r - 1} \\ &= \frac{a\,(1 - r^n)}{1 - r}. \end{aligned}$$

1.12.2 Complex numbers

Recall that the roots of the quadratic equation

$$ax^2 + bx + c$$

are given by

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Exercise 1.12.1 *Solve the equation*

$$x^2 + x + 1 = 0.$$

What do you notice about the solutions?

Not every quadratic has *real* roots, that is, roots in the set of real numbers $\mathbb{R}$. Some have both roots in the set of *complex numbers,* denoted by $\mathbb{C}$. A complex number is of the form

$$a + b\mathbf{i}$$

where a and b are real numbers, and $\mathbf{i} = \sqrt{-1}$. Note that the number $\sqrt{-1}$ is *not* a real number.

Complex numbers are sometimes plotted on what are called *Argand diagrams.* For the complex number $a+b\mathbf{i}$, this is basically just a plot of the point (a, b) in two dimensions. The distance the point lies from the origin, which by Pythagoras' theorem is $\sqrt{a^2 + b^2}$, is sometimes called the *magnitude* (or *modulus*) of the number.

Not much knowledge is required about complex numbers in what follows, though you should be aware of their existence. Note that in general a polynomial *may* have complex roots (a cubic can have either two or no complex roots, for example). The statement that the roots of a polynomial, $f(x)$ say, are outside the unit circle simply means that any solution of the equation $f(x) = 0$ has magnitude (i.e., modulus, or *absolute value*) greater than unity (i.e., >1).

Any complex number z can be written in *exponential form,*

$$z = re^{\mathbf{i}\theta}$$

for some real numbers r and θ. Taking the series expansion of e^x when $x = \mathbf{i}\theta$ we deduce

$$e^{\mathbf{i}\theta} = \cos(\theta) + \mathbf{i}\sin(\theta).$$

Moreover,

$$\cos(\theta) = \frac{e^{\mathbf{i}\theta} + e^{-\mathbf{i}\theta}}{2}$$

and

$$\sin(\theta) = \frac{e^{\mathbf{i}\theta} - e^{-\mathbf{i}\theta}}{2\mathbf{i}}.$$