3 Stationary processes and autocorrelations

3.1 Introduction

A time series is a series of data, a sequence of observations recorded over time, usually at more or less equally spaced intervals of time. Many such series of data exhibit either a long-term trend, i.e., a systematic change of the mean level of the series over time, or seasonality (with period usually one year), or abrupt changes.

The following plot shows a series of monthly totals of international airline passengers in thousands, for the years 1949 - 1960, which in addition to an overall increasing trend has a period of length 12. (The data are taken from Box, Jenkins and Reinsel, *Time Series Analysis*):

```r
> air.rts <- rts(passengers, start = 1949, frequency = 12, units = "months")
> air.rts
> ts.plot(air.rts, xlab = "", ylab = "", main = "Monthly Totals of Airline Passengers in Thousands", las = 1)
```
However, some series appear to be *stationary*, i.e., their distributional properties do not change over time. In some approaches to the analysis of time series, when series with trend or seasonality are being analysed, they are first transformed into stationary series by some means or other. In any case, the study of stationary processes is fundamental to the analysis of time series.

In constructing models for time series, it turns out to be convenient to consider an underlying process that stretches back into the infinite past and forward into the infinite future. Thus we consider a doubly infinite sequence of random variables (r.v.s) \( \{ Y_t : t \in \mathbb{Z} \} \), a stochastic process in discrete time, where \( \mathbb{Z} \) is the set of integers \( \{\ldots,-2,-1,0,1,2,\ldots\} \). We will also need to refer to sequences of random variables in which the subscript indexing runs over just a subset of \( \mathbb{Z} \). Commonly used sets are listed below.

- \( \mathbb{Z}^+ = \{1,2,3,\ldots\} \): the positive integers;
- \( \mathbb{N} = \{0,1,2,3,\ldots\} \): the natural numbers;
- \( \{a, a+1, \ldots, b\} \): the set of integers running from \( a \) up to \( b \) (inclusive), where \( a < b \), and
  - \( i \geq a \): refers to the set of integers that are greater than or equal to \( a \), i.e. \( \{a, a+1, a+2, \ldots\} \).

The \( \{Y_t\} \) will usually be continuous r.v.s.

What we shall observe is a realization \( y_1, y_2, \ldots, y_T \) of a finite section of the process, where \( y_t \) is the value that the process takes at time \( t \). The observed data will be used to make inferences about the structure of the underlying process \( \{Y_t\} \).
### 3.2 The definition of stationarity

**Definition 3.2.1 (Strict Stationarity)**
The process \( \{Y_t\} \) is said to be **strictly stationary** if its probabilistic laws remain unchanged through shifts in time, i.e., if for every integer \( n \geq 1 \), every selection \( t_1, t_2, \ldots, t_n \) of distinct indices and every integer \( h \), the joint distribution of \( Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n} \) is the same as the joint distribution of \( Y_{t_1+h}, Y_{t_2+h}, \ldots, Y_{t_n+h} \).

**Remarks 3.2.2**
(i) In particular, taking \( n = 1 \), the marginal distribution of \( Y_t \) is the same for each \( t \). Assuming existence, the mean and variance of \( Y_t \) is the same for each \( t \).
(ii) Taking \( n = 2 \), for every pair \( t_1, t_2 \) and every \( h \), the joint distribution of \( Y_{t_1}, Y_{t_2} \) is the same as the joint distribution of \( Y_{t_1+h}, Y_{t_2+h} \). Another way of expressing this is that the joint distribution of \( Y_t, Y_{t-\tau} \) depends only on the lag \( \tau \) and not on \( t \). (The lag is the difference in the subscripts, the time difference.)

Assume from now on that all first and second order moments of the process exist and are finite.

**Definition 3.2.3 (Moments)**
(i) The first order moments of a stationary process are specified by the process mean \( \mu \), where

\[
\mu = E[Y_t] \quad \quad t \in \mathbb{Z}.
\]

(ii) The second order moments of a stationary process are specified by the autocovariances \( \{\gamma_{\tau}\} \) at lag \( \tau \), where

\[
\gamma_{\tau} = \text{cov}(Y_t, Y_{t-\tau}) = E[(Y_t - \mu)(Y_{t-\tau} - \mu)] \quad \quad \tau \in \mathbb{Z}.
\]

**Remarks 3.2.4**
Note that \( \gamma_0 = \text{var}(Y_t) \) and that \( \gamma_{\tau} = \gamma_{-\tau} \) for all \( \tau \), since, equivalently,

\[
\gamma_{\tau} = \text{cov}(Y_t, Y_{t+\tau}).
\]

The sequence \( \{\gamma_{\tau}\} \) is called the **autocovariance function**. The autocovariance (or ’self’ covariance) function represents the covariance of the process with previous values of itself.

**Definition 3.2.5 (Autocorrelation)**
The autocorrelation \( \rho_{\tau} \) at lag \( \tau \) is given by

\[
\rho_{\tau} = \text{cor}(Y_t, Y_{t-\tau}) = \frac{\gamma_{\tau}}{\gamma_0} \quad \quad \tau \in \mathbb{Z}.
\]

Note that \( \rho_0 = 1 \), \( \rho_{\tau} = \rho_{-\tau} \) for all \( \tau \) and \( |\rho_{\tau}| \leq 1 \) for all \( \tau \). \( \rho_{\tau} \) is a measure of the linear dependency between values of the process at lag \( \tau \) apart. The sequence \( \{\rho_{\tau}\} \) is called the autocorrelation function.
Because many of the properties of stationary processes follow from the properties of their first and second moments, a weaker definition of stationarity is often used, which imposes conditions on the means and covariances of the process random variables.

**Definition 3.2.6 (Weak Stationarity)**
The process is said to be *weakly stationary* (or *covariance stationary* or *wide-sense stationary*) if $E[Y_t]$ is constant over time and $\text{cov}(Y_t, Y_{t-\tau})$ depends only on the lag $\tau$ and not on $t$. \qed

The first and second order moments of the process are then again specified by the process mean $\mu$ and the autocovariance function $\{\gamma_\tau\}$.

Given the assumption that all first and second order moments exist and are finite, strict stationarity clearly implies weak stationarity. But not every weakly stationary process is also strictly stationary. However, any multivariate normal distribution is completely specified by its first and second order moments. Hence, if it is assumed that all the joint distributions of the $Y_t$ are multivariate normal then weak stationarity does imply strict stationarity.

From now on, we shall refer to weakly stationary processes simply as stationary processes.

If $\{Y_t\}$ is a stationary process with process mean $\mu$ then we may work instead with the r.v.s $Y_t - \mu$, which does not alter the autocovariance function $\{\gamma_\tau\}$ but sets the process mean to zero. So in dealing with much of the theory of stationary processes we may without any essential loss of generality assume that the process mean is zero, in which case

$$
\gamma_\tau = E[Y_t Y_{t-\tau}] \quad \tau \in \mathbb{Z}.
$$

### 3.3 White noise processes

A simple but important type of stationary process is one in which the $Y_t$ are independently and identically distributed (i.i.d.), with common variance $\sigma^2$, say, in which case

$$
\gamma_0 = \sigma^2 \quad \text{and} \quad \gamma_\tau = 0 \quad \tau \neq 0,
$$

$$
\rho_0 = 1 \quad \text{and} \quad \rho_\tau = 0 \quad \tau \neq 0.
$$

To use engineering/time series terminology, a stationary process whose autocovariances and autocorrelations satisfy the above conditions is known as a *white noise process*. A formal definition may be presented as follows.

**Definition 3.3.1 (White Noise)**
The sequence $\{Y_t\}$ is said to constitute a white noise process if it is stationary and the $\{Y_t\}$ are (pairwise) uncorrelated. \qed

This does not imply that the $Y_t$ are i.i.d.
Remarks 3.3.2
(i) We shall use the notation \( \{ \epsilon_t \} \) for a white noise process, which, unless otherwise stated, will be assumed to have mean zero and variance \( \sigma^2 \). For such a process, then,
\[
E[\epsilon_t^2] = \sigma^2 \quad (1)
\]
\[
E[\epsilon_t \epsilon_{t-\tau}] = 0 \quad \tau \neq 0. \quad (2)
\]

(ii) In the context of forecasting, we shall make the stronger assumption that the \( \epsilon_t \) are i.i.d. and, furthermore, that they are normally distributed.

3.4 Sample autocovariances and autocorrelations

Definition 3.4.1 (Sample Statistics)
Let \( y_1, y_2, \ldots, y_T \) be a realization of part of a stationary process.
(i) The sample mean, \( \bar{y} \), is given by
\[
\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t,
\]
(ii) The sample autocovariance, \( c_\tau \), at lag \( \tau \), is given by
\[
c_\tau = \frac{1}{T} \sum_{t=\tau+1}^{T} (y_t - \bar{y})(y_{t-\tau} - \bar{y}) \quad \tau = 0, 1, \ldots, T - 1.
\]
(iii) The sample autocorrelation, \( r_\tau \), at lag \( \tau \), is given by
\[
r_\tau = \frac{c_\tau}{c_0} \quad \tau = 0, 1, \ldots, T - 1.
\]

First, note that \( \bar{y} \) is an unbiased estimator of the process mean \( \mu \).
Also, the statistic \( c_\tau \) is used as an estimator of \( \gamma_\tau \) and \( r_\tau \) as an estimator of \( \rho_\tau \). We also note the following:

1. \( r_0 = 1 = \rho_0 \).
2. The range of \( t \)-values in the summation for \( c_\tau \) depends on the lag \( \tau \).
3. Although the divisor \( T - \tau \) might seem more natural in the definition of \( c_\tau \), the divisor \( T \) is generally preferred. This ensures that the sample autocorrelations satisfy \( |r_\tau| \leq 1 \) for all \( \tau \).
4. We could not expect to estimate autocovariances and autocorrelations at lag \( T \) or greater if the observed time series is of length \( T \), so the range of values of \( \tau \) for which \( c_\tau \) and \( r_\tau \) are defined is in accordance with common sense.

The following table provides a summary of our notation for autocovariances and autocorrelations.
The sequence \( \{ r_\tau : \tau = 0, 1, \ldots, T-1 \} \) is the sample autocorrelation function. The plot of \( r_\tau \) against \( \tau \) is sometimes called the correlogram.

The statistic \( r_\tau \) becomes less and less reliable as an estimator of \( \rho_\tau \) as \( \tau \) increases towards \( T \). In practice, values of \( r_\tau \) will be calculated and plotted over a limited range of \( \tau \) values such as, for example, \( 0 \leq \tau \leq T/4 \).

- The sample mean, sample autocovariances and sample autocorrelations may also be calculated for observed time series that are not necessarily assumed to come from an underlying stationary process. For example, if we have data with a highly regular seasonality of period \( s \), we will observe large autocorrelations at lags \( s, 2s, 3s, \ldots \).

### 3.5 Bread price example

The following S+ output gives a preliminary analysis of the average annual price in pennies of a 4 lb loaf of bread in London for the years 1634 to 1690 (original series runs until 1757). The bread price data are plausibly a realization of a stationary process, but successive values are positively correlated.

```r
> BP.rts <- rts(price, start = 1634, units = "years")
> summary(BP.rts)

Regular Time Series:
Observations: 124

    Min. 1st Qu.  Median    Mean  3rd Qu.   Max.
3.900  4.800   5.400   5.652   6.200   9.700

Time Parameters :
  start deltat frequency units
   1634    1     1     years

> ts.plot(BP.rts[1:(1 + (1690 - 1634))], xlab = "Year", ylab = "", main = + "Average Price in Pennies of 4lb Loaf of Bread in London", las = 1)
> BP.acf <- acf(BP.rts, 30)
> BP.acf
```

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>autocovariance</th>
<th>autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>process/population</td>
<td>( \mu )</td>
<td>( \gamma_{\tau} )</td>
<td>( \rho_{\tau} = \gamma_{\tau}/\gamma_{0} )</td>
</tr>
<tr>
<td>sample</td>
<td>( \bar{y} )</td>
<td>( c_{\tau} )</td>
<td>( r_{\tau} = c_{\tau}/c_{0} )</td>
</tr>
</tbody>
</table>
3.6 The model for a first order autoregressive process

Consider a simple model for the numbers of unemployed \(Y_t\) in successive months \(t\). Ignoring seasonal effects and trend, to a first approximation, suppose that the number unemployed in any month is a fixed proportion \(\phi\) of the unemployed in the previous month plus a number of newly unemployed workers seeking jobs. Assume that the newly unemployed form a white noise process with mean \(m\), so that

\[
Y_t = \phi Y_{t-1} + m + \epsilon_t \quad t \in \mathbb{Z},
\]

where \(\{\epsilon_t\}\) is a white noise process with mean zero. Assuming that the process \(\{Y_t\}\) as defined by Equation (3) is stationary, let \(\mu\) denote the process mean. Taking expectations in Equation (3) we find that

\[
\mu = \phi \mu + m,
\]

and hence, assuming that \(\phi \neq 1\),

\[
\mu = \frac{m}{1 - \phi}.
\]

The model of Equation (3) can be rewritten as

\[
Y_t - \mu = \phi (Y_{t-1} - \mu) + \epsilon_t, \quad t \in \mathbb{Z}.
\]

Replacing the r.v.s \(\{Y_t - \mu\}\) by \(\{Y_t\}\) in Equation (4), we obtain the simpler model

\[
Y_t = \phi Y_{t-1} + \epsilon_t \quad t \in \mathbb{Z},
\]

which has process mean zero.

In general, a stationary process \(\{Y_t\}\) which satisfies Equation (4) or Equation (5) is known as a first order autoregressive process, an AR(1) process, with autoregressive parameter \(\phi\). The term “autoregressive” is used, because \(Y_t\) is expressed as a regression on the previous value \(Y_{t-1}\) of the process itself. The autoregression is of “first order” because the regression is on only one previous process value.

This model, despite its simplicity, is a very useful one for modelling data, either on its own or as a component of more complex models; and in its simplicity it also has the attraction of being readily estimable.

For the present we shall use the version of the model specified by Equation (5), which is more convenient for the purposes of mathematical analysis. Both versions of the model have the same autocovariance and autocorrelation functions.

The process \(\{Y_t\}\) generally represents a sequence of observable variables, whereas \(\{\epsilon_t\}\) is a white noise process of unobservable random errors or innovations — each \(Y_t\) depends directly or indirectly on the previous values \(\{Y_{t-i} : i \geq 1\}\) and \(\{\epsilon_{t-i} : i \geq 1\}\), whereas each \(\epsilon_t\) is a new input into the model, uncorrelated with the previous history of the process. Thus \(\{\epsilon_t\}\) can be regarded as a process of mutually uncorrelated impulses, which drive the system.

If the white noise term \(\epsilon_t\) is removed from the model of Equation (5) then we are left with the deterministic model

\[
Y_t = \phi Y_{t-1} \quad t \in \mathbb{Z}.
\]
This simple recurrence relation has the general solution $Y_t = A\phi^t$, where $A$ is an arbitrary constant. Note that $Y_t \to 0$ as $t \to \infty$, whatever the value of $A$, if and only if $|\phi| < 1$. Thus $\{Y_t\}$ has the stable limit point 0 if and only if $|\phi| < 1$.

We have so far assumed that Equation (5) does indeed specify a stationary process. It turns out that the condition $|\phi| < 1$ is also the necessary and sufficient condition for there to exist a suitably defined stationary process $\{Y_t\}$ which satisfies Equation (5).

### 3.7 Infinite moving average representation of an AR(1) process

Using Equation (5) recursively,

$$Y_t = \phi(\phi Y_{t-2} + \epsilon_{t-1}) + \epsilon_t$$

$$= \phi^2 Y_{t-2} + \phi \epsilon_{t-1} + \epsilon_t$$

$$= \ldots$$

$$= \phi^n Y_{t-n} + \sum_{i=0}^{n-1} \phi^i \epsilon_{t-i}. \quad (7)$$

What we would like to do is to let $n \to \infty$ in Equation (7), state that $\phi^n Y_{t-n} \to 0$, and write

$$Y_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}, \quad (8)$$

which is the infinite moving average representation of the process $\{Y_t\}$. This would appear to be plausible if $|\phi| < 1$. However, we are dealing here with the convergence of sequences of random variables and have to be careful.

In the present context we use the concept of convergence in mean square:

**Definition 3.7.1 (Convergence in Mean Square)**

A sequence of r.v.s $\{\xi_n\}$ is said to converge in mean square to a r.v. $\xi$ if $E[(\xi_n - \xi)^2] \to 0$ as $n \to \infty$. We may then use the notation that $\xi_n \xrightarrow{ms} \xi$. \[\square\]

From Equation (7), if we assume that $\{Y_t\}$ is a stationary process and that $|\phi| < 1$,

$$E \left[ \left( \sum_{i=0}^{n-1} \phi^i \epsilon_{t-i} - Y_t \right)^2 \right] = E[\phi^{2n} Y_{t-n}^2]$$

$$= \phi^{2n} \gamma_0 \to 0$$

as $n \to \infty$. Thus, under these assumptions,

$$\sum_{i=0}^{n-1} \phi^i \epsilon_{t-i} \xrightarrow{ms} Y_t$$

and so the infinite moving average representation of Equation (8) is valid in the sense of convergence in mean square. The value of $Y_t$ is represented as a linear function of the innovations $\epsilon_t$ going back into the infinite past. The geometric decline of the coefficients, $\phi^i$, in the representation reflects the relative residual influence of the previous innovations $\epsilon_{t-i}$ on the current process value $Y_t$ as we go back in time.

9