Chapter 2

Linear time series

Prerequisites

- Familiarity with linear models.
- Solve polynomial equations.
- Be familiar with complex numbers.
- Understand under what conditions the partial sum \( S_n = \sum_{j=1}^{n} a_j \) has a well defined limits (ie. if \( \sum_{j=1}^{\infty} |a_j| < \infty \), then \( S_n \rightarrow S \), where \( S = \sum_{j=1}^{\infty} a_j \)).

Objectives

- Understand what causal and invertible is.
- Know what an AR, MA and ARMA time series model is.
- Know how to find a solution of an ARMA time series, and understand why this is important (how the roots determine causality and why this is important to know - in terms of characteristics in the process and also simulations).
- Understand how the roots of the AR can determine ‘features’ in the time series and covariance structure (such as pseudo periodicities).
2.1 Motivation

The objective of this chapter is to introduce the linear time series model. Linear time series models are designed to model the covariance structure in the time series. There are two popular subgroups of linear time models (a) the autoregressive and (a) the moving average models, which can be combined to make the autoregressive moving average models.

We motivate the autoregressive from the perspective of classical linear regression. We recall one objective in linear regression is to predict the response variable given variables that are observed. To do this, typically linear dependence between response and variable is assumed and we model $Y_i$ as

$$Y_i = \sum_{j=1}^{p} a_j X_{ij} + \varepsilon_i,$$

where $\varepsilon_i$ is such that $E[\varepsilon_i|X_{ij}] = 0$ and more commonly $\varepsilon_i$ and $X_{ij}$ are independent. In linear regression once the model has been defined, we can immediately find estimators of the parameters, do model selection etc.

Returning to time series, one major objective is to predict/forecast the future given current and past observations (just as in linear regression our aim is to predict the response given the observed variables). At least formally, it seems reasonable to represent this as

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t, \quad t \in \mathbb{Z} \tag{2.1}$$

where we assume that $\{\varepsilon_t\}$ are independent, identically distributed, zero mean random variables. Model (2.1) is called an autoregressive model of order $p$ (AR($p$) for short). Further, it would appear that

$$E(X_t|X_{t-1}, \ldots, X_{t-p}) = \sum_{j=1}^{p} \phi_j X_{t-j}$$

(the expected value of $X_t$ given that $X_{t-1}, \ldots, X_{t-p}$ have already been observed), thus the past values of $X_t$ have a linear influence on the conditional mean of $X_t$. However this is not necessarily the case; the autoregressive model appears to be a straightforward extension of the linear regression model but don’t be fooled by this, it is a more complex object.

Unlike the linear regression model, (2.1) is an infinite set of linear difference equations. This
means, for this systems of equations to be well defined, it needs to have a solution which is meaningful. To understand why, recall that (2.1) is defined for all \( t \in \mathbb{Z} \), so let us start the equation at the beginning of time \( (t = -\infty) \) and run it on. Without any constraint on the parameters \( \{\phi_j\} \), there is no reason to believe the solution is finite (contrast this with linear regression where these issues are not relevant). Therefore, the first thing to understand is under what conditions will the AR model (2.1) have a well defined stationary solution and what features in a time series is the solution able to capture.

Of course, one could ask why go through to the effort. One could simply use least squares to estimate the parameters. This is possible, but there are two related problems (a) without a proper analysis it is not clear whether model has a meaningful solution (for example in Section 3.3 we show that the least squares estimator can lead to misspecified models), it’s not even possible to make simulations of the process (b) it is possible that \( E(\varepsilon_t|X_{t-p}) \neq 0 \), this means that least squares is not estimating \( \phi_j \) and is instead estimating an entirely different set of parameters! Therefore, there is a practical motivation behind our theoretical treatment.

In this chapter we will be deriving conditions for a strictly stationary solution of (2.1). We will place moment conditions on the innovations \( \{\varepsilon_t\} \), these conditions will be sufficient but not necessary conditions Under these conditions we obtain a strictly stationary solution but not a second order stationary process. In Chapter 3 we obtain conditions for (2.1) to have be both a strictly and second order stationary solution. It is possible to obtain strictly stationary solution under far weaker conditions (see Theorem 4.0.1), but these won’t be considered here.

**Example 2.1.1** How would you simulate from the model

\[
X_t = \phi_1 X_{t-1} + \phi_2 X_{t-1} + \varepsilon_t.
\]

**Remark 2.1.1** The main objective in this Chapter is to look for stationary solutions to (2.1). If, however, we define the equation

\[
X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t, \quad t = 1, 2, \ldots
\]

(2.2)

(note \( t \geq 0 \)), then nonstationary solutions are also possible. Indeed if assume \( E(\varepsilon_t|X_{t-j}) = 0 \) (for \( j \geq 1 \)) then we may obtain a so called explosive (unstable) process. The sampling properties of estimators for such processes are very different to those in the stationary case (see Anderson

46
2.2 Linear time series and moving average models

2.2.1 Infinite sums of random variables

Before defining a linear time series, we define the MA(q) model which is a subclass of linear time series. Let us suppose that \( \{ \varepsilon_t \} \) are iid random variables with mean zero and finite variance. The time series \( \{ X_t \} \) is said to have a MA(q) representation if it satisfies

\[
X_t = \sum_{j=0}^{q} \psi_j \varepsilon_{t-j},
\]

where \( \mathbb{E}(\varepsilon_t) = 0 \) and \( \text{var}(\varepsilon_t) = 1 \). It is clear that \( X_t \) is a rolling finite weighted sum of \( \{ \varepsilon_t \} \), therefore \( \{ X_t \} \) must be well defined (which for finite sums means it is almost surely finite, this you can see because it has a finite variance). We extend this notion and consider infinite sums of random variables. Now, things become more complicated, since care must be always be taken with anything involving infinite sums. More precisely, for the sum

\[
\sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j},
\]

to be well defined (has a finite limit), the partial sums \( S_n = \sum_{j=-n}^{n} \psi_j \varepsilon_{t-j} \) should be (almost surely) finite and the sequence \( S_n \) should converge (ie. \( |S_{n_1} - S_{n_2}| \to 0 \) as \( n_1, n_2 \to \infty \)). Below, we give conditions under which this is true.

**Lemma 2.2.1** Suppose \( \{ X_t \} \) is a strictly stationary time series with \( \mathbb{E}|X_t| < \infty \), then \( \{ Y_t \} \) defined by

\[
Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j},
\]

where \( \sum_{j=0}^{\infty} |\psi_j| < \infty \), is a strictly stationary time series. Furthermore, the partial sum converges almost surely, \( Y_{n,t} = \sum_{j=0}^{n} \psi_j X_{t-j} \rightarrow Y_t \). If \( \text{var}(X_t) < \infty \), then \( \{ Y_t \} \) is second order stationary and converges in mean square (that is \( \mathbb{E}(Y_{n,t} - Y_t)^2 \rightarrow 0 \)).
PROOF. See Brockwell and Davis (1998), Proposition 3.1.1 or Fuller (1995), Theorem 2.1.1 (page 31) (also Shumway and Stoffer (2006), page 86).

Example 2.2.1 Suppose \( \{X_t\} \) is a strictly stationary time series with \( \text{var}(X_t) < \infty \). Define \( \{Y_t\} \) as the following infinite sum

\[
Y_t = \sum_{j=0}^{\infty} j^k \rho^j |X_{t-j}|
\]

where \( |\rho| < 1 \). Then \( \{Y_t\} \) is also a strictly stationary time series with a finite variance.

We will use this example later in the course.

Having derived conditions under which infinite sums are well defined (good), we can now define the general class of linear and MA(\( \infty \)) processes.

Definition 2.2.1 (The linear process and moving average (MA)(\( \infty \))) Suppose that \( \{\varepsilon_t\} \) are iid random variables, \( \sum_{j=0}^{\infty} |\psi_j| < \infty \) and \( E(|\varepsilon_t|) < \infty \).

A time series is said to be a linear time series if it can be represented as

\[
X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j},
\]

where \( \{\varepsilon_t\} \) are iid random variables with finite variance. Note that since that as these sums are well defined by equation (1.12) \( \{X_t\} \) is a strictly stationary (ergodic) time series.

This is a rather strong definition of a linear process. A more general definition is \( \{X_t\} \) has the representation

\[
X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j},
\]

where \( \{\varepsilon_t\} \) are uncorrelated random variables with mean zero and variance one (thus the independence assumption has been dropped). We show in definition (ii) that all second order stationary processes with finite variance have this representation.

For inference usually the stronger condition is used.
(ii) The time series \([X_t]\) has a MA(\(\infty\)) representation if it satisfies

\[
X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}.
\] (2.3)

Note that some authors have a slightly different definition of an MA(\(\infty\)) representation. A time series \([X_t]\) has an MA(\(\infty\)) representation if it can be represented as

\[
X_t = \sum_{j=1}^{\infty} \psi_j Z_{t-j},
\] (2.4)

where \([Z_t = X_t = P_{X_{t-1},X_{t-2},\ldots}(X_t)]\) (where \(P_{X_{t-1},X_{t-2},\ldots}(X_t)\) is the best linear predictor of \(X_t\) given the past, \(X_{t-1}, X_{t-2}, \ldots\)).

Note that by Wold’s representation theorem (see Section 5.7) any (non-deterministic) second order stationary time series with such a representation. Note further that if \(X_t\) has an MA(\(\infty\)) as defined by (2.3) it may not have the same MA(\(\infty\)) representation given in (2.4).

The representation in (2.4) has many practical advantages. For example Krampe et al. (2016) recently used it to define the so called “MA bootstrap”.

The difference between an MA(\(\infty\)) process and a linear process is quite subtle. A linear process involves both past, present and future innovations \(\{\varepsilon_t\}\), whereas the MA(\(\infty\)) uses only past and present innovations.

**Definition 2.2.2 (Causal and invertible)**

(i) An ARMA\((p, q)\) process \(X_t + \sum_{j=1}^{p} \psi_j X_{t-j} = \sum_{i=1}^{q} \theta_i \varepsilon_t\) (where \(\{\varepsilon_t\}\) are uncorrelated random variables with mean zero and constant variance) is said to be causal if it has the representation

\[
X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}.
\]

(ii) An ARMA\((p, q)\) process \(X_t + \sum_{j=1}^{p} \psi_j X_{t-j} = \sum_{i=1}^{q} \theta_i \varepsilon_t\) (where \(\{\varepsilon_t\}\) are uncorrelated random variables with mean zero and constant variance) is said to be invertible if it has the representation

\[
X_t = \sum_{j=1}^{\infty} b_j X_{t-j} + \varepsilon_t.
\]
Note that define ARMA below. Further as a little taster as to what is to come, and ARMA process may not be causal or invertible; however under certain conditions on the parameters and by changing \( \{\varepsilon_t\} \) such a representation is possible.

Causal and invertible solutions are useful in both estimation and forecasting (predicting the future based on the current and past).

A very interesting class of models which have MA(\( \infty \)) representations are autoregressive and autoregressive moving average models. In the following sections we prove this.

2.3 The autoregressive model and the solution

In this section we will examine under what conditions the AR process has a stationary solution.

2.3.1 Difference equations and back-shift operators

The autoregressive model is defined in terms of inhomogeneous difference equations. Difference equations can often be represented in terms of backshift operators, so we start by defining them and see why this representation may be useful (and why it should work).

The time series \( \{X_t\} \) is said to be an autoregressive (AR(\( p \))) if it satisfies the equation

\[
X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = \varepsilon_t, \quad t \in \mathbb{Z},
\]

where \( \{\varepsilon_t\} \) are zero mean, finite variance random variables. As we mentioned previously, the autoregressive model is a difference equation (which can be treated as a infinite number of simultaneous equations). Therefore for it to make any sense it must have a solution. To obtain a general solution we write the autoregressive model in terms of backshift operators:

\[
X_t - \phi_1 BX_{t-1} - \ldots - \phi_p B^p X_t = \varepsilon_t, \quad \Rightarrow \quad \phi(B)X_t = \varepsilon_t
\]

where \( \phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j \), \( B \) is the backshift operator and is defined such that \( B^k X_t = X_{t-k} \).

Simply rearranging \( \phi(B)X_t = \varepsilon_t \), gives the ‘solution’ of the autoregressive difference equation to be \( X_t = \phi(B)^{-1}\varepsilon_t \), however this is just an algebraic manipulation, below we investigate whether it really has any meaning. To do this, we start with an example.
### 2.3.2 Solution of two particular AR(1) models

Below we consider two different AR(1) models and obtain their solutions.

(i) Consider the AR(1) process

\[ X_t = 0.5X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}. \quad (2.5) \]

Notice this is an equation (rather like \(3x^2 + 2x + 1 = 0\), or an infinite number of simultaneous equations), which may or may not have a solution. To obtain the solution we note that \(X_t = 0.5X_{t-1} + \varepsilon_t\) and \(X_{t-1} = 0.5X_{t-2} + \varepsilon_{t-1}\). Using this we get \(X_t = \varepsilon_t + 0.5(0.5X_{t-2} + \varepsilon_{t-1}) = \varepsilon_t + 0.5\varepsilon_{t-1} + 0.5^2X_{t-2}\). Continuing this backward iteration we obtain at the \(k\)th iteration, \(X_t = \sum_{j=0}^{k}(0.5)^j\varepsilon_{t-j} + (0.5)^{k+1}X_{t-k}\). Because \((0.5)^{k+1} \to 0\) as \(k \to \infty\) by taking the limit we can show that \(X_t = \sum_{j=0}^{\infty}(0.5)^j\varepsilon_{t-j}\) is almost surely finite and a solution of (2.5). Of course like any other equation one may wonder whether it is the unique solution (recalling that \(3x^2 + 2x + 1 = 0\) has two solutions). We show in Section 2.3.3 that this is the unique stationary solution of (2.5).

Let us see whether we can obtain a solution using the difference equation representation. We recall, that by crudely taking inverses, the solution is \(X_t = (1 - 0.5B)^{-1}\varepsilon_t\). The obvious question is whether this has any meaning. Note that \((1 - 0.5B)^{-1} = \sum_{j=0}^{\infty}(0.5B)^j\), for \(|B| \leq 2\), hence substituting this power series expansion into \(X_t\) we have

\[ X_t = (1 - 0.5B)^{-1}\varepsilon_t = \left(\sum_{j=0}^{\infty}(0.5B)^j\right)\varepsilon_t = \left(\sum_{j=0}^{\infty}(0.5^jB^j)\right)\varepsilon_t = \sum_{j=0}^{\infty}(0.5)^j\varepsilon_{t-j}, \]

which corresponds to the solution above. Hence the backshift operator in this example helps us to obtain a solution. Moreover, because the solution can be written in terms of past values of \(\varepsilon_t\), it is causal.

(ii) Let us consider the AR model, which we will see has a very different solution:

\[ X_t = 2X_{t-1} + \varepsilon_t. \quad (2.6) \]

Doing what we did in (i) we find that after the \(k\)th back iteration we have \(X_t = \sum_{j=0}^{k}2^j\varepsilon_{t-j} + 2^{k+1}X_{t-k}\). However, unlike example (i) \(2^k\) does not converge as \(k \to \infty\). This suggest that if
we continue the iteration $X_t = \sum_{j=0}^{\infty} 2^j \varepsilon_{t-j}$ is not a quantity that is finite (when $\varepsilon_t$ are iid). Therefore $X_t = \sum_{j=0}^{\infty} 2^j \varepsilon_{t-j}$ cannot be considered as a solution of (2.6). We need to write (2.6) in a slightly different way in order to obtain a meaningful solution.

Rewriting (2.6) we have $X_{t-1} = 0.5X_t + 0.5\varepsilon_t$. Forward iterating this we get $X_{t-1} = -(0.5) \sum_{j=0}^{k}(0.5)^j \varepsilon_{t+j} - (0.5)^{t+k+1}X_{t+k}$. Since $(0.5)^{t+k+1} \to 0$ we have

$$X_{t-1} = -(0.5) \sum_{j=0}^{\infty}(0.5)^j \varepsilon_{t+j}$$

as a solution of (2.6).

Let us see whether the difference equation can also offer a solution. Since $(1 - 2B)X_t = \varepsilon_t$, using the crude manipulation we have $X_t = (1 - 2B)^{-1} \varepsilon_t$. Now we see that $(1 - 2B)^{-1} = \sum_{j=0}^{\infty}(2B)^j$ for $|B| < 1/2$. Using this expansion gives $X_t = \sum_{j=0}^{\infty} 2^j B^j X_t$, but as pointed out above this sum is not well defined. What we find is that $\phi(B)^{-1} \varepsilon_t$ only makes sense (is well defined) if the series expansion of $\phi(B)^{-1}$ converges in a region that includes the unit circle $|B| = 1$.

What we need is another series expansion of $(1 - 2B)^{-1}$ which converges in a region which includes the unit circle $|B| = 1$ (as an aside, we note that a function does not necessarily have a unique series expansion, it can have difference series expansions which may converge in different regions). We now show that a convergent series expansion needs to be defined in terms of negative powers of $B$ not positive powers. Writing $(1 - 2B) = -(2B)(1 - (2B)^{-1})$, therefore

$$(1 - 2B)^{-1} = -(2B)^{-1} \sum_{j=0}^{\infty}(2B)^{-j},$$

which converges for $|B| > 1/2$. Using this expansion we have

$$X_t = -\sum_{j=0}^{\infty}(0.5)^{j+1} B^{-j-1} \varepsilon_t = -\sum_{j=0}^{\infty}(0.5)^{j+1} \varepsilon_{t+j+1},$$

which we have shown above is a well defined solution of (2.6).
In summary \((1 - 2B)^{-1}\) has two series expansions

\[
\frac{1}{1 - 2B} = \sum_{j=0}^{\infty} (2B)^{-j}
\]

which converges for \(|B| < 1/2\) and

\[
\frac{1}{1 - 2B} = -(2B)^{-1} \sum_{j=0}^{\infty} (2B)^{-j},
\]

which converges for \(|B| > 1/2\). The one that is useful for us is the series which converges when \(|B| = 1\).

It is clear from the above examples how to obtain the solution of a general AR(1). We now show that this solution is the unique stationary solution.

### 2.3.3 The unique solution of a general AR(1)

Consider the AR(1) process \(X_t = \phi X_{t-1} + \varepsilon_t\), where \(|\phi| < 1\). Using the method outlined in (i), it is straightforward to show that \(X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}\) is its stationary solution, we now show that this solution is unique.

We first show that \(X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}\) is well defined (that it is almost surely finite). We note that \(|X_t| \leq \sum_{j=0}^{\infty} |\phi^j| \cdot |\varepsilon_{t-j}|\). Thus we will show that \(\sum_{j=0}^{\infty} |\phi^j| \cdot |\varepsilon_{t-j}|\) is almost surely finite, which will imply that \(X_t\) is almost surely finite. By montone convergence we can exchange sum and expectation and we have

\[
E(|X_t|) \leq E(\lim_{n \to \infty} \sum_{j=0}^{n} |\phi^j| |\varepsilon_{t-j}|) = \lim_{n \to \infty} E(\sum_{j=0}^{n} |\phi^j| E|\varepsilon_{t-j}|) = E(\lim_{\varepsilon_0|} \sum_{j=0}^{\infty} |\phi^j| < \infty.
\]

Therefore since \(E|X_t| < \infty\), \(\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}\) is a well defined solution of \(X_t = \phi X_{t-1} + \varepsilon_t\).

To show that it is the unique stationary causal solution, let us suppose there is another (causal) solution, call it \(Y_t\) (note that this part of the proof is useful to know as such methods are often used when obtaining solutions of time series models). Clearly, by recursively applying the difference equation to \(Y_t\), for every \(s\) we have

\[
Y_t = \sum_{j=0}^{s} \phi^j \varepsilon_{t-j} + \phi^s Y_{t-s-1}.
\]

Evaluating the difference between the two solutions gives \(Y_t - X_t = A_s - B_s\) where \(A_s = \phi^s Y_{t-s-1}\) and \(B_s = \sum_{j=s+1}^{\infty} \phi^j \varepsilon_{t-j}\) for all \(s\). To show that \(Y_t\) and \(X_t\) coincide almost surely we will show that...
for every $\epsilon > 0$, $\sum_{s=1}^{\infty} P(|A_s - B_s| > \epsilon) < \infty$ (and then apply the Borel-Cantelli lemma). We note if $|A_s - B_s| > \epsilon$), then either $|A_s| > \epsilon/2$ or $|B_s| > \epsilon/2$. Therefore $P(|A_s - B_s| > \epsilon) \leq P(|A_s| > \epsilon/2) + P(|B_s| > \epsilon/2)$. To bound these two terms we use Markov’s inequality. It is straightforward to show that $P(|B_s| > \epsilon/2) \leq C\phi^s/\epsilon$. To bound $E|A_s|$, we note that $|Y_s| \leq |\phi| \cdot |Y_{s-1}| + |\epsilon_s|$, since $\{Y_t\}$ is a stationary solution then $E|Y_s|/(1 - |\phi|) \leq E|\epsilon_s|$, thus $E|Y_t| \leq E|\epsilon_t|/(1 - |\phi|) < \infty$. Altogether this gives $P(|A_s - B_s| > \epsilon) \leq C\phi^s/\epsilon$ (for some finite constant $C$). Hence $\sum_{s=1}^{\infty} P(|A_s - B_s| > \epsilon) \leq \sum_{s=1}^{\infty} C\phi^s/\epsilon < \infty$. Thus by the Borel-Cantelli lemma, this implies that the event $\{|A_s - B_s| > \epsilon\}$ happens only finitely often (almost surely). Since for every $\epsilon$, $\{|A_s - B_s| > \epsilon\}$ occurs (almost surely) only finite often for all $\epsilon$, then $Y_t = X_t$ almost surely. Hence $X_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$ is (almost surely) the unique causal solution.

### 2.3.4 The solution of a general AR($p$)

Let us now summarise our observation for the general AR(1) process $X_t = \phi X_{t-1} + \epsilon_t$. If $|\phi| < 1$, then the solution is in terms of past values of $\{\epsilon_t\}$, if on the other hand $|\phi| > 1$ the solution is in terms of future values of $\{\epsilon_t\}$.

Now we try to understand this in terms of the expansions of the characteristic polynomial $\phi(B) = 1 - \phi B$ (using the AR(1) as a starting point). From what we learnt in the previous section, we require the characteristic polynomial of the AR process to have a convergent power series expansion in the region including the ring $|B| = 1$. In terms of the AR(1) process, if the root of $\phi(B)$ is greater than one, then the power series of $\phi(B)^{-1}$ is in terms of positive powers, if it is less than one, then $\phi(B)^{-1}$ is in terms of negative powers.

Generalising this argument to a general polynomial, if the roots of $\phi(B)$ are greater than one, then the power series of $\phi(B)^{-1}$ (which converges for $|B| = 1$) is in terms of positive powers (hence the solution $\phi(B)^{-1}\epsilon_t$ will be in past terms of $\{\epsilon_t\}$). On the other hand, if the roots are both less than and greater than one (but do not lie on the unit circle), then the power series of $\phi(B)^{-1}$ will be in both negative and positive powers. Thus the solution $X_t = \phi(B)^{-1}\epsilon_t$ will be in terms of both past and future values of $\{\epsilon_t\}$. We summarize this result in a lemma below.

**Lemma 2.3.1** Suppose that the AR($p$) process satisfies the representation $\phi(B)X_t = \epsilon_t$, where none of the roots of the characteristic polynomial lie on the unit circle and $E|\epsilon_t| < \infty$. Then $\{X_t\}$ has a stationary, almost surely unique, solution.
the AR process. We will show in Sections 2.3.6 and 3.1.2 that it not only defines the solution but determines some of the characteristics of the time series.

**Exercise 2.1** Suppose \( \{X_t\} \) satisfies the AR\((p)\) representation

\[
X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t,
\]

where \( \sum_{j=1}^{p} |\phi_j| < 1 \) and \( \mathbb{E}|\varepsilon_t| < \infty \). Show that \( \{X_t\} \) will always have a causal stationary solution.

### 2.3.5 Explicit solution of an AR\((2)\) model

**Specific example**

Suppose \( \{X_t\} \) satisfies

\[
X_t = 0.75X_{t-1} - 0.125X_{t-2} + \varepsilon_t,
\]

where \( \{\varepsilon_t\} \) are iid random variables. We want to obtain a solution for the above equations.

It is not easy to use the backward (or forward) iterating technique for AR processes beyond order one. This is where using the backshift operator becomes useful. We start by writing \( X_t = 0.75X_{t-1} - 0.125X_{t-2} + \varepsilon_t \) as \( \phi(B)X_t = \varepsilon_t \), where \( \phi(B) = 1 - 0.75B + 0.125B^2 \), which leads to what is commonly known as the characteristic polynomial \( \phi(z) = 1 - 0.75z + 0.125z^2 \). If we can find a power series expansion of \( \phi(B)^{-1} \), which is valid for \( |B| = 1 \), then the solution is \( X_t = \phi(B)^{-1}\varepsilon_t \).

We first observe that \( \phi(z) = 1 - 0.75z + 0.125z^2 = (1 - 0.5z)(1 - 0.25z) \). Therefore by using partial fractions we have

\[
\frac{1}{\phi(z)} = \frac{1}{(1 - 0.5z)(1 - 0.25z)} = \frac{-1}{1 - 0.5z} + \frac{2}{1 - 0.25z}.
\]

We recall from geometric expansions that

\[
\frac{-1}{1 - 0.5z} = - \sum_{j=0}^{\infty} (0.5)^j z^j \quad |z| \leq 2, \quad \frac{2}{1 - 0.25z} = 2 \sum_{j=0}^{\infty} (0.25)^j z^j \quad |z| \leq 4.
\]
Putting the above together gives

\[
\frac{1}{(1 - 0.5z)(1 - 0.25z)} = \sum_{j=0}^{\infty} \{-0.5^j + 2(0.25)^j\} z^j, \quad |z| < 2.
\]

The above expansion is valid for $|z| = 1$, because $\sum_{j=0}^{\infty} | -0.5^j + 2(0.25)^j | < \infty$ (see Lemma 2.3.2). Hence

\[
X_t = \{(1 - 0.5B)(1 - 0.25B)\}^{-1} \varepsilon_t = (\sum_{j=0}^{\infty} \{-0.5^j + 2(0.25)^j\} B^j) \varepsilon_t = \sum_{j=0}^{\infty} \{-0.5^j + 2(0.25)^j\} \varepsilon_{t-j},
\]

which gives a stationary solution to the AR(2) process (see Lemma 2.2.1). Moreover since the roots lie outside the unit circle the solution is causal.

The discussion above shows how the backshift operator can be applied and how it can be used to obtain solutions to AR(p) processes.

**The solution of a general AR(2) model**

We now generalise the above to general AR(2) models

\[
X_t = (a + b)X_{t-1} - abX_{t-2} + \varepsilon_t,
\]

the characteristic polynomial of the above is $1 - (a + b)z + abz^2 = (1 - az)(1 - bz)$. This means the solution of $X_t$ is

\[
X_t = (1 - Ba)^{-1}(1 - Bb)^{-1} \varepsilon_t,
\]

thus we need an expansion of $(1 - Ba)^{-1}(1 - Bb)^{-1}$. Assuming that $a \neq b$, as above we have

\[
\frac{1}{(1 - za)(1 - zb)} = \frac{1}{b - a} \left( \frac{b}{1 - bz} - \frac{a}{1 - az} \right).
\]

**Cases:**

(i) $|a| < 1$ and $|b| < 1$, this means the roots lie outside the unit circle. Thus the expansion is

\[
\frac{1}{(1 - za)(1 - zb)} = \frac{1}{b - a} \left( b \sum_{j=0}^{\infty} b^j z^j - a \sum_{j=0}^{\infty} a^j z^j \right),
\]

56
which leads to the causal solution

\[ X_t = \frac{1}{b-a} \left( \sum_{j=0}^{\infty} \left( b^{j+1} - a^{j+1} \right) \varepsilon_{t-j} \right). \hspace{1cm} \text{(2.7)} \]

(ii) Case that \(|a| > 1\) and \(|b| < 1\), this means the roots lie inside and outside the unit circle and we have the expansion

\[
\frac{1}{(1-za)(1-zb)} = \frac{1}{b-a} \left( \frac{b}{1-bz} - \frac{a}{(az)^{-1} - 1} \right) = \frac{1}{(b-a)} \left( b \sum_{j=0}^{\infty} b^j z^j + z^{-1} \sum_{j=0}^{\infty} a^{-j} z^{-j} \right), \hspace{1cm} \text{(2.8)}
\]

which leads to the non-causal solution

\[ X_t = \frac{1}{b-a} \left( \sum_{j=0}^{\infty} b^{j+1} \varepsilon_{t-j} + \sum_{j=0}^{\infty} a^{-j} \varepsilon_{t+1+j} \right). \hspace{1cm} \text{(2.9)} \]

Later we show that the non-causal solution has the same correlation structure as the causal solution with \(a^{-1}\) replacing \(a\). Furthermore, it is possible to generate uncorrelated innovations, that are not independent, but which generate the same process i.e.

\[(1 - a^{-1}B)(1 - bB)X_t = \tilde{\varepsilon}_t,\]

thus \(X_t\) has the solution

\[ X_t = \frac{1}{b-a} \left( \sum_{j=0}^{\infty} \left( b^{j+1} - a^{j+1} \right) \tilde{\varepsilon}_{t-j} \right). \hspace{1cm} \text{(2.10)} \]

Returning to (2.9), we see that this solution throws up additional interesting results. Let us return to the expansion in (2.8) and apply it to \(X_t\)

\[
X_t = \frac{1}{(1-Ba)(1-Bb)} \varepsilon_t = \frac{1}{b-a} \left( \frac{b}{1-bB} \varepsilon_t + \frac{1}{B(1-a^{-1}B^{-1})} \tilde{\varepsilon}_t \right) \quad \text{causal AR(1)} + \quad \text{noncausal AR(1)}
\]

\[
= \frac{1}{b-a} (Y_t + Z_{t+1})
\]
where \( Y_t = bY_{t-1} + \varepsilon_t \) and \( Z_{t+1} = a^{-1}Z_{t+2} + \varepsilon_{t+1} \). In other words, the noncausal AR(2) process is the sum of a causal and a ‘future’ AR(1) process. This is true for all noncausal time series (except when there is multiplicity in the roots) and is discussed further in Section 2.6.

Several authors including Richard Davis, Jay Breidt and Beth Andrews argue that noncausal time series can model features in data which causal time series cannot.

(iii) \( a = b < 1 \) (both roots are the same and lie outside the unit circle). The characteristic polynomial is \((1 - az)^2\). To obtain the convergent expansion when \(|z| = 1\) we note that \((1 - az)^{-2} = (-1)^{d(1-az)^{-1}} \). Thus

\[
\frac{(-1)}{(1 - az)^2} = (-1) \sum_{j=0}^{\infty} j(az)^{j-1}.
\]

This leads to the causal solution

\[
X_t = (-1) \sum_{j=1}^{\infty} ja^{j-1}\varepsilon_{t-j}.
\]

In many respects this is analogous to Matern covariance defined over \( \mathbb{R}^d \) (and used in spatial statistics). However, unlike autocovariance defined over \( \mathbb{R}^d \) the behaviour of the autocovariance at zero is not an issue.

**Exercise 2.2** Show for the AR(2) model \( X_t = \phi_1X_{t-1} + \phi_2X_{t-2} + \varepsilon_t \) to have a causal stationary solution the parameters \( \phi_1, \phi_2 \) must lie in the region

\[
\phi_2 + \phi_1 < 1, \quad \phi_2 - \phi_1 < 1, \quad |\phi_2| < 1.
\]

**Exercise 2.3** (a) Consider the AR(2) process

\[
X_t = \phi_1X_{t-1} + \phi_2X_{t-2} + \varepsilon_t,
\]

where \( \{\varepsilon_t\} \) are iid random variables with mean zero and variance one. Suppose the roots of the characteristic polynomial \( 1 - \phi_1z - \phi_2z^2 \) are greater than one. Show that \(|\phi_1| + |\phi_2| < 4\).
(b) Now consider a generalisation of this result. Consider the AR\(p\) process

\[ X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t. \]

Suppose the roots of the characteristic polynomial \(1 - \phi_1 z - \ldots - \phi_p z^p\) are greater than one. Show that \(|\phi_1| + \ldots + |\phi_p| \leq 2^p\).

### 2.3.6 Features of a realisation from an AR(2)

We now explain why the AR(2) (and higher orders) can characterise some very interesting behaviour (over the rather dull AR(1)). For now we assume that \(X_t\) is a causal time series which satisfies the AR(2) representation

\[ X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t \]

where \(\{\varepsilon_t\}\) are iid with mean zero and finite variance. The characteristic polynomial is \(\phi(B) = 1 - \phi_1 B - \phi_2 B^2\). Let us assume the roots of \(\phi(B)\) are complex, since \(\phi_1\) and \(\phi_2\) are real, the roots are complex conjugates. Thus by using case (i) above we have

\[
\frac{1}{1 - \phi_1 B - \phi_2 B^2} = \frac{1}{\lambda - \bar{\lambda}} \left( \frac{\lambda}{1 - \lambda B} - \frac{\bar{\lambda}}{1 - \lambda \bar{B}} \right),
\]

where \(\lambda^{-1}\) and \(\bar{\lambda}^{-1}\) are the roots of the characteristic. Thus

\[
X_t = C \sum_{j=0}^{\infty} \lambda^j \varepsilon_{t-j} - \bar{C} \sum_{j=0}^{\infty} \bar{\lambda}^j \varepsilon_{t-j},
\]

(2.11)

where \(C = [\lambda - \bar{\lambda}]^{-1}\). Since \(\lambda\) and \(C\) are complex we use the representation \(\lambda = r \exp(i\theta)\) and \(C = \alpha \exp(i\beta)\) (noting that \(|r| < 1\)), and substitute these expressions for \(\lambda\) and \(C\) into (2.11) to give

\[
X_t = \alpha \sum_{j=0}^{\infty} r^j \cos(\theta j + \beta) \varepsilon_{t-j}.
\]

We can see that \(X_t\) is effectively the sum of cosines with frequency \(\theta\) that have been modulated by the iid errors and exponentially damped. This is why for realisations of autoregressive processes you will often see periodicities (depending on the roots of the characteristic). These arguments can
Exercise 2.4  (a) Obtain the stationary solution of the AR(2) process

\[ X_t = \frac{7}{3} X_{t-1} - \frac{2}{3} X_{t-2} + \varepsilon_t, \]

where \( \{\varepsilon_t\} \) are iid random variables with mean zero and variance \( \sigma^2 \).

Does the solution have an MA(\( \infty \)) representation?

(b) Obtain the stationary solution of the AR(2) process

\[ X_t = \frac{4 \times \sqrt{3}}{5} X_{t-1} - \frac{4^2 \sigma^2}{5^2} X_{t-2} + \varepsilon_t, \]

where \( \{\varepsilon_t\} \) are iid random variables with mean zero and variance \( \sigma^2 \).

Does the solution have an MA(\( \infty \)) representation?

(c) Obtain the stationary solution of the AR(2) process

\[ X_t = X_{t-1} - 4 X_{t-2} + \varepsilon_t, \]

where \( \{\varepsilon_t\} \) are iid random variables with mean zero and variance \( \sigma^2 \).

Does the solution have an MA(\( \infty \)) representation?

Exercise 2.5  Construct a causal stationary AR(2) process with pseudo-period 17. Using the R function arima.sim simulate a realisation from this process (of length 200) and make a plot of the periodogram. What do you observe about the peak in this plot?

Below we now consider solutions to general AR(\( \infty \)) processes.

2.3.7 Solution of the general AR(\( \infty \)) model

AR(\( \infty \)) models are more general than the AR(p) model and are able to model more complex behaviour, such as slower decay of the covariance structure. It is arguable how useful these models are in modelling data, however recently it has become quite popular in time series bootstrap methods.
In order to obtain the stationary solution of an AR(∞), we need to define an analytic function and its inverse.

**Definition 2.3.1 (Analytic functions in the region Ω)** Suppose that \( z \in \mathbb{C} \). \( \phi(z) \) is an analytic complex function in the region \( \Omega \), if it has a power series expansion which converges in \( \Omega \), that is \( \phi(z) = \sum_{j=-\infty}^{\infty} \phi_j z^j \).

If there exists a function \( \tilde{\phi}(z) = \sum_{j=-\infty}^{\infty} \tilde{\phi}_j z^j \) such that \( \tilde{\phi}(z)\phi(z) = 1 \) for all \( z \in \Omega \), then \( \tilde{\phi}(z) \) is the inverse of \( \phi(z) \) in the region \( \Omega \).

Well known examples of analytic functions include

(i) Finite order polynomials such as \( \phi(z) = \sum_{j=0}^{p} \phi_j z^j \) for \( \Omega = \mathbb{C} \).

(ii) The expansion \( (1 - 0.5z)^{-1} = \sum_{j=0}^{\infty} (0.5z)^j \) for \( \Omega = \{z; |z| \leq 2\} \).

We observe that for AR processes we can represent the equation as \( \phi(B)X_t = \varepsilon_t \), which formally gives the solution \( X_t = \phi(B)^{-1}\varepsilon_t \). This raises the question, under what conditions on \( \phi(B)^{-1} \) is \( \phi(B)^{-1}\varepsilon_t \) a valid solution. For \( \phi(B)^{-1}\varepsilon_t \) to make sense \( \phi(B)^{-1} \) should be represented as a power series expansion. Below, we give conditions on the power series expansion which give a stationary solution. It is worth noting this is closely related to Lemma 2.2.1.

**Lemma 2.3.2** Suppose that \( \phi(z) = \sum_{j=-\infty}^{\infty} \phi_j z^j \) is finite on a region that includes \( |z| = 1 \) (hence it is analytic) and \( \{X_t\} \) is a strictly stationary process with \( \text{E}|X_t| < \infty \). Then \( \sum_{j=-\infty}^{\infty} |\phi_j| < \infty \) and \( Y_t = \phi(B)X_{t-j} = \sum_{j=-\infty}^{\infty} \phi_j X_{t-j} \) is almost surely finite and strictly stationary time series.

**Proof.** It can be shown that if \( \sup_{|z|=1} |\phi(z)| < \infty \), in other words on the unit circle \( \sum_{j=-\infty}^{\infty} |\phi_j z^j| < \infty \), then \( \sum_{j=-\infty}^{\infty} |\phi_j| < \infty \). Since the coefficients are absolutely summable, then by Lemma 2.2.1 we have that \( Y_t = \phi(B)X_{t-j} = \sum_{j=-\infty}^{\infty} \phi_j X_{t-j} \) is almost surely finite and strictly stationary.

Using the above we can obtain the solution of an AR(∞) (which includes an AR(p) as a special case).

**Corollary 2.3.1** Suppose that

\[
X_t = \sum_{j=1}^{\infty} \phi_j X_{t-j} + \varepsilon_t
\]
and φ(z) has an inverse \( \psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j \) which is analytic in a region including \( |z| = 1 \), then the solution of \( X_t \) is

\[
X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}.
\]

**Corollary 2.3.2** Let \( X_t \) be an AR\((p)\) time series, where

\[
X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t.
\]

Suppose the roots of the characteristic polynomial \( \phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j \) do not lie on the unit circle \( |B| = 1 \), then \( X_t \) admits a strictly stationary solution.

In addition suppose the roots of \( \phi(B) \) all lie outside the unit circle, then \( X_t \) admits a strictly stationary, causal solution.

This summarises what we observed in Section 2.3.4.

**Rules of the back shift operator:**

(i) If \( a(z) \) is analytic in a region \( \Omega \) which includes the unit circle \( |z| = 1 \) in its interior and \( \{Y_t\} \) is a well defined time series, then \( X_t \) defined by \( Y_t = a(B)X_t \) is a well defined random variable.

(ii) The operator is commutative and associative, that is \( [a(B)b(B)]X_t = a(B)[b(B)X_t] = [b(B)a(B)]X_t \) (the square brackets are used to indicate which parts to multiply first). This may seems obvious, but remember matrices are not commutative!

(iii) Suppose that \( a(z) \) and its inverse \( \frac{1}{a(z)} \) are both have solutions in the region \( \Omega \) which includes the unit circle \( |z| = 1 \) in its interior. If \( a(B)X_t = Z_t \), then \( X_t = \frac{1}{a(B)}Z_t \).

**Example 2.3.1 (Analytic functions)**

(i) Clearly \( a(z) = 1 - 0.5z \) is analytic for all \( z \in \mathbb{C} \), and has no zeros for \( |z| < 2 \). The inverse is \( \frac{1}{a(z)} = \sum_{j=0}^{\infty} (0.5z)^j \) is well defined in the region \( |z| < 2 \).

(ii) Clearly \( a(z) = 1 - 2z \) is analytic for all \( z \in \mathbb{C} \), and has no zeros for \( |z| > 1/2 \). The inverse is \( \frac{1}{a(z)} = (\frac{-2z}{1(1/2z)})^{-1}(\sum_{j=0}^{\infty} (1/(2z))^j) \) well defined in the region \( |z| > 1/2 \).

(iii) The function \( a(z) = \frac{1}{(1-0.5z)(1-2z)} \) is analytic in the region \( 0.5 < z < 2 \).
(iv) $a(z) = 1 - z$, is analytic for all $z \in \mathbb{C}$, but is zero for $z = 1$. Hence its inverse is not well defined for regions which involve $|z| = 1$ (see Example 2.3.2).

Example 2.3.2 (Unit root/integrated processes and non-invertible processes)

(i) If the difference equation has a root which is one, then an (almost sure) stationary solution of the AR model does not exist. The simplest example is the ‘random walk’ $X_t = X_{t-1} + \varepsilon_t$ ($\phi(z) = (1-z)$). This is an example of an Autoregressive Integrated Moving Average ARIMA($0,1,0$) model $(1 - B)X_t = \varepsilon_t$.

To see that it does not have a stationary solution, we iterate the equation $n$ steps backwards; $X_t = \sum_{j=0}^{n} \varepsilon_{t-j} + X_{t-n}$. $S_{t,n} = \sum_{j=0}^{n} \varepsilon_{t-j}$ is the partial sum, but it is clear that the partial sum $S_{t,n}$ does not have a limit, since it is not a Cauchy sequence, i.e. $|S_{t,n} - S_{t,m}|$ does not have a limit. However, given some initial value $X_0$, for $t > 0$ the so called ‘unit process’ $X_t = X_{t-1} + \varepsilon_t$ is well defined. Notice that the nonstationary solution of this sequence is $X_t = X_0 + \sum_{j=1}^{t} \varepsilon_{t-j}$ which has variance $\text{var}(X_t) = \text{var}(X_0) + t$ (assuming that $\{\varepsilon_t\}$ are iid random variables with variance one and independent of $X_0$).

We observe that we can ‘stationarize’ the process by taking first differences, i.e. defining $Y_t = X_t - X_{t-1} = \varepsilon_t$.

(ii) The unit process described above can be generalised to taking $d$ differences (often denoted as an ARIMA($0,d,0$)) where $(1 - B)^d X_t = \varepsilon_t$ (by taking $d$-differences we can remove $d$-order polynomial trends). We elaborate on this below.

To stationarize the sequence we take $d$ differences, i.e. let $Y_{t,0} = X_t$ and for $1 \leq i \leq d$ define the iteration

$$Y_{t,i} = Y_{t,i-1} - Y_{t-1,i-1}$$

and $Y_t = Y_{t,d}$ will be a stationary sequence. Note that this is equivalent to

$$Y_t = \sum_{j=0}^{d} \frac{d!}{j!(d-j)!} (-1)^j X_{t-j}.$$

(iii) The general ARIMA($p,d,q$) is defined as $(1 - B)^d \phi(B) X_t = \theta(B) \varepsilon_t$, where $\phi(B)$ and $\theta(B)$ are $p$ and $q$ order polynomials respectively and the roots of $\phi(B)$ lie outside the unit circle.

Another way of describing the above model is that after taking $d$ differences (as detailed in (ii))
the resulting process is an ARMA\((p,q)\) process (see Section 2.5 for the definition of an ARMA model).

To illustrate the difference between stationary ARMA and ARIMA processes, in Figure 2.1

Suppose \((1 - B)\phi(B)X_t = \varepsilon_t\) and let \(\tilde{\phi}(B) = (1 - B)\phi(B)\). Then we observe that \(\tilde{\phi}(1) = 0\). This property is useful when checking for unit root behaviour (see Section 2.7).

(iv) In examples (i) - (iii) a stationary solution does not exist. We now consider an example where the process is stationary but an autoregressive representation does not exist (this matters when we want to forecast).

Consider the MA\((1)\) model \(X_t = \varepsilon_t - \varepsilon_{t-1}\). We recall that this can be written as \(X_t = \phi(B)\varepsilon_t\) where \(\phi(B) = 1 - B\). From Example 2.3.1(iv) we know that \(\phi(z)^{-1}\) does not exist, therefore it does not have an AR\((\infty)\) representation since \((1 - B)^{-1}X_t = \varepsilon_t\) is not well defined.

(a) \(X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t\)  
(b) \((1 - B)Y_t = X_t\), where \(X_t\) is defined in (a)

Figure 2.1: Realisations from an AR process and its corresponding integrated process, using \(N(0,1)\) innovations (generated using the same seed).
2.4 An explanation as to why the backshift operator method works

To understand why the magic backshift operator works, we use matrix notation to rewrite the AR\((p)\) model as an infinite set of difference equations

\[
\begin{pmatrix}
  \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
  \ldots & 0 & 1 & -\phi_1 & \ldots & -\phi_p & \ldots \\
  \ldots & 0 & 0 & 1 & -\phi_1 & \ldots & -\phi_p \\
  \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots
\end{pmatrix}
\begin{pmatrix}
  \vdots \\
  X_t \\
  X_{t-1} \\
  X_{t-2} \\
  \vdots
\end{pmatrix}
= \begin{pmatrix}
  \vdots \\
  \varepsilon_t \\
  \varepsilon_{t-1} \\
  \varepsilon_{t-2} \\
  \vdots
\end{pmatrix}.
\]

The above is an infinite dimensional equation (and the matrix is an infinite upper triangular matrix). Formally to obtain a simulation we invert the matrix to get a solution of \(X_t\) in terms of \(\varepsilon_t\). Of course in reality it is not straightforward to define this inverse. Instead let us consider a finite (truncated) version of the above matrix equation. Except for the edge effects this is a circulant matrix (where the rows are repeated, but each time shifted by one, see wiki for a description). Truncating the matrix to have dimension \(n\), we approximate the above by the finite set of \(n\)-equations

\[
\begin{pmatrix}
  1 & -\phi_1 & \ldots & -\phi_p & 0 & \ldots \\
  0 & 1 & -\phi_1 & \ldots & -\phi_p \\
  \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
  -\phi_1 & -\phi_2 & \ldots & \ldots & 0 & 1
\end{pmatrix}
\begin{pmatrix}
  X_n \\
  X_{n-1} \\
  \vdots \\
  X_0
\end{pmatrix}
= \begin{pmatrix}
  \varepsilon_n \\
  \varepsilon_{n-1} \\
  \vdots \\
  \varepsilon_0
\end{pmatrix},
\]

\(\Rightarrow C_n X_n \approx \varepsilon_n\).

The approximation of the AR\((p)\) equation only arises in the first \(p\)-equations, where

\[
X_0 - \sum_{j=1}^{p} \phi_j X_{n-j} = \varepsilon_0
\]

\[
X_1 - \phi_1 X_0 - \sum_{j=2}^{p} \phi_j X_{n+1-j} = \varepsilon_1
\]

\[\vdots \vdots\]

\[
X_p - \sum_{j=1}^{p-1} \phi_j X_{p-j} - \phi_p X_n = \varepsilon_p.
\]
We now define the $n \times n$ matrix $U_n$, where

$$U_n = \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \ldots & 0 \end{pmatrix}.$$

We observe that $U_n$ is a ‘deformed diagonal matrix’ where all the ones along the diagonal have been shifted once to the right, and the ‘left over’ one is placed in the bottom left hand corner. $U_n$ is another example of a circulant matrix, moreover $U_n^2$ shifts once again all the ones to the right

$$U_n^2 = \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \ldots & 0 \end{pmatrix}.$$

$U_n^3$ shifts the ones to the third off-diagonal and so forth until $U_n^n = I$. Thus all circulant matrices can be written in terms of powers of $U_n$ (the matrix $U_n$ can be considered as the building blocks of circulant matrices). In particular

$$C_n = I_n - \sum_{j=1}^{p} \phi_j U_n^j,$$

$[I_n - \sum_{j=1}^{p} \phi_j U_n^j]X_n = \xi_n$ and the solution to the equation is

$$X_n = (I_n - \sum_{j=1}^{p} \phi_j U_n^j)^{-1}\xi_n.$$

Our aim is to write $(I_n - \sum_{j=1}^{p} \phi_j U_n^j)^{-1}$ as a power series in $U_n$, with $U_n$ playing the role of the backshift operator.

To do this we recall the similarity between the matrix $I_n - \sum_{j=1}^{p} \phi_j U_n^j$ and the characteristic equation $\phi(B) = 1 - \sum_{j=1}^{p} \phi_j z^j$. In particular since we can factorize the characteristic equation as $\phi(B) = \prod_{j=1}^{p}[1 - \lambda_j B]$, we can factorize the matrix $I_n - \sum_{j=1}^{p} \phi_j U_n^j = \prod_{j=1}^{p}[I_n - \lambda_j U_n]$. To obtain the inverse, for simplicity, we assume that the roots of the characteristic function are greater than
one (i.e. \( |\lambda_j| < 1 \), which we recall corresponds to a causal solution) and are all different. Then there
exists constants \( c_j \) where

\[
[I_n - \sum_{j=1}^{p} \phi_j U_n^j]^{-1} = \sum_{j=1}^{p} c_j (I_n - \lambda_j U_n)^{-1}
\]

(just as in partial fractions) - to see why multiply the above by \([I_n - \sum_{j=1}^{p} \phi_j U_n^j] \). Finally, we
recall that if the eigenvalues of \( A \) are less than one, then \((1 - A)^{-1} = \sum_{j=0}^{\infty} A^j\). The eigenvalues
of \( U_n \) are \( \{\exp(\frac{2\pi ij}{n}); j = 1, \ldots, n\} \), thus the eigenvalues of \( \lambda_j U_n \) are less than one. This gives
\((I_n - \lambda_j U_n)^{-1} = \sum_{k=0}^{\infty} \lambda_j^k U_n^k \) and

\[
[I_n - \sum_{j=1}^{p} \phi_j U_n^j]^{-1} = \sum_{j=1}^{p} c_j \sum_{k=0}^{\infty} \lambda_j^k U_n^k.
\]  

(2.12)

Therefore, the solution of \( C_n X_n = \varepsilon_n \) is

\[
X_n = C_n^{-1} \varepsilon_n = \left( \sum_{j=1}^{p} c_j \sum_{k=0}^{\infty} \lambda_j^k U_n^k \right) \varepsilon_n.
\]

Let us focus on the first element of the vector \( X_n \), which is \( X_n \). Since \( U_n^k \varepsilon_n \) shifts the elements of
\( \varepsilon_n \) up by \( k \) (note that this shift is with wrapping of the vector) we have

\[
X_n = \sum_{j=1}^{p} c_j \sum_{k=0}^{n} \lambda_j^k \varepsilon_{n-k} + \sum_{j=1}^{p} c_j \sum_{k=n+1}^{\infty} \lambda_j^k \varepsilon_{n-k} \pmod{n}.
\]  

(2.13)

Note that the second term decays geometrically fast to zero. Thus giving the stationary solution
\( X_n = \sum_{j=1}^{p} c_j \sum_{k=0}^{\infty} \lambda_j^k \varepsilon_{n-k} \).

To recollect, we have shown that \([I_n - \sum_{j=1}^{p} \phi_j U_n^j]^{-1}\) admits the solution in (2.12) (which is
the same as the solution of the inverse of \( \phi(B)^{-1} \)) and that \( U_n^j \varepsilon_n \) plays the role of the backshift
operator. Therefore, we can use the backshift operator in obtaining a solution of an AR process
because it plays the role of the matrix \( U_n \).
Example 2.4.1 The AR(1) model, $X_t - \phi_1 X_{t-1} = \epsilon_t$ is written as

$$
\begin{pmatrix}
1 & -\phi_1 & \ldots & 0 & 0 & 0 & 0 \\
0 & 1 & -\phi_1 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
-\phi_1 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
X_n \\
X_{n-1} \\
\vdots \\
X_0
\end{pmatrix}
= 
\begin{pmatrix}
\epsilon_n \\
\epsilon_{n-1} \\
\vdots \\
\epsilon_0
\end{pmatrix}
\Rightarrow C_n X_n = \tilde{\epsilon}_n.
$$

The approximation of the AR(1) is only for the first equation, where $X_0 - \phi_1 X_n = \epsilon_0$. Using the matrix $U_n$, the above equation can be written as $(I_n - \phi_1 U_n) X_n = \tilde{\epsilon}_n$, which gives the solution

$$
X_n = (I_n - \phi_1 U_n)^{-1} \tilde{\epsilon}_n.
$$

Let us suppose that $|\phi_1| > 1$ (i.e., the root lies inside the unit circle and the solution is noncausal), then to get a convergent expansion of $(I_n - \phi_1 U_n)^{-1}$ we rewrite $(I_n - \phi_1 U_n) = -\phi_1 U_n (I_n - \phi_1^{-1} U_n^{-1})$.

Thus we have

$$(I_n - \phi_1 U_n)^{-1} = - \left[ \sum_{k=0}^{\infty} \phi_1^{-k} U_n^{-k} \right] (\phi_1 U_n)^{-1}.$$

Therefore the solution is

$$
X_n = \left( - \sum_{k=0}^{\infty} \phi_1^{-k+1} U_n^{-k+1} \right) \tilde{\epsilon}_n,
$$

which in its limit gives the same solution as Section 2.3.2(ii).

Notice that $U_n^j$ and $B^j$ are playing the same role.

A rigorous explanation on extending this argument to stationary time series defined on $\mathbb{Z}$ can be found in Pourahmadi (2001), Sections 5.3 and 9.5.3. The rough argument is that one defines a Hilbert space $H(X)$ which is the closure of all linear combinations of $\{X_t\}$. Note that the metric on this Hilbert space is simply the covariance i.e. if $Y = \sum_{j \in \mathbb{Z}} a_j X_j$, $Z = \sum_{j \in \mathbb{Z}} b_j X_j$ and $Y, Z \in H(X)$ then $\langle Y, Z \rangle = \text{cov}[Y, Z] = \sum_{j_1, j_2} a_{j_1} b_{j_2} c(j_1 - j_2)$ where $c(\cdot)$ is an autocovariance function of $\{X_t\}$.

We define the operator $U$ where $UX_t = X_{t+1}$ and $U(\sum_{j=1}^m a_j X_{t+j}) = \sum_{j=1}^m a_j X_{t+j+1}$. It can be shown that $U$ extends to $H(X)$ and is a continuous, linear, surjective operator (see Pourahmadi (2001), Section 9.5.3). Moreover since $U$ is an isometric operator (i.e. it is measure preserving;
cov[UY, UZ] = cov[Y, Z] if Y, Z ∈ H(X), this is easy to show), then it is a unitary operator (this means its adjoint operator is also its inverse i.e. U*U = I). It is clear that U*Xt = Xt−1. All this implies if Y ∈ H(X) and Z = UY then Y = U*Z. To jump between ARMA and its solutions we need to extend these arguments to two processes \{Xt, εt\} (see Section Pourahmadi (2001), Section 5.3.2). Once these details are clarified we can jump between φ(B)Xt = εt and Xt = φ(B)−1εt and back again.

### 2.4.1 Representing the AR(p) as a vector AR(1)

Let us suppose Xt is an AR(p) process, with the representation

\[ X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t. \]

For the rest of this section we will assume that the roots of the characteristic function, \(φ(z)\), lie outside the unit circle, thus the solution causal. We can rewrite the above as a Vector Autoregressive (VAR(1)) process

\[ \begin{pmatrix} X_t \\ \varepsilon_t \end{pmatrix} = \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} X_{t-1} \\ \varepsilon_{t-1} \end{pmatrix} \]

where

\[ A = \begin{pmatrix} \phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}, \]

\(X'_t = (X_t, \ldots, X_{t-p+1})\) and \(\varepsilon'_t = (\varepsilon_t, 0, \ldots, 0)\). It is straightforward to show that the eigenvalues of A are the inverse of the roots of \(φ(z)\) (since

\[ \det(A - zI) = z^p - \sum_{i=1}^{p} \phi_i z^{p-i} = z^p (1 - \sum_{i=1}^{p} \phi_i z^{-i})), \]

thus the eigenvalues of A lie inside the unit circle. It can be shown that for any \(|\lambda_{max}(A)| < \delta < 1\), there exists a constant \(C_δ\) such that \(\|A^j\|_{spec} \leq C_δ \delta^j\) (see Appendix A). Note that result is
extremely obvious if the eigenvalues are distinct (in which case the spectral decomposition can be used), in which case \( \|A\|_{\text{spec}} \leq C \lambda_{\max}(A) \) (note that \( \|A\|_{\text{spec}} \) is the spectral norm of \( A \), which is the largest eigenvalue of the symmetric matrix \( AA' \)).

We can apply the same back iterating that we did for the AR(1) to the vector AR(1). Iterating (2.14) backwards \( k \) times gives

\[
X_t = \sum_{j=0}^{k-1} A^j \varepsilon_{t-j} + A^k X_{t-k}.
\]

Since \( \|A^k X_{t-k}\|_2 \leq \|A^k\|_{\text{spec}} \|X_{t-k}\| \overset{P}{\rightarrow} 0 \) we have

\[
X_t = \sum_{j=0}^{\infty} A^j \varepsilon_{t-j}.
\]

### 2.5 The ARMA model

Up to now, we have defined the moving average and the autoregressive model. The MA(\( q \)) average has the feature that after \( q \) lags there isn’t any correlation between two random variables. On the other hand, there are correlations at all lags for an AR(\( p \)) model. In addition as we shall see later on, it is much easier to estimate the parameters of an AR model than an MA. Therefore, there are several advantages in fitting an AR model to the data (note that when the roots are of the characteristic polynomial lie inside the unit circle, then the AR can also be written as an MA(\( \infty \)), since it is causal). However, if we do fit an AR model to the data, what order of model should we use? Usually one uses the AIC (BIC or similar criterion) to determine the order. But for many data sets, the selected order tends to be relative large, for example order 14. The large order is usually chosen when correlations tend to decay slowly and/or the autocorrelations structure is quite complex (not just monotonically decaying). However, a model involving 10-15 unknown parameters is not particularly parsimonious and more parsimonious models which can model the same behaviour would be useful. A very useful generalisation which can be more flexible (and parsimonious) is the ARMA(\( p, q \)) model, in this case \( X_t \) satisfies

\[
X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}.
\]

### Definition 2.5.1 (Summary of AR, ARMA and MA models) (i) The autoregressive AR(\( p \))
model: \( \{X_t\} \) satisfies

\[
X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t. \tag{2.16}
\]

Observe we can write it as \( \phi(B)X_t = \varepsilon_t \)

(ii) The moving average MA(\(q\)) model: \( \{X_t\} \) satisfies

\[
X_t = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}. \tag{2.17}
\]

Observe we can write \( X_t = \theta(B)\varepsilon_t \)

(iii) The autoregressive moving average ARMA(\(p, q\)) model: \( \{X_t\} \) satisfies

\[
X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}. \tag{2.18}
\]

We observe that we can write \( X_t \) as \( \phi(B)X_t = \theta(B)\varepsilon_t \).

Below we give conditions for the ARMA to have a causal solution and also be invertible. We also show that the coefficients of the MA(\(\infty\)) representation of \( X_t \) will decay exponentially.

Lemma 2.5.1 Let us suppose \( X_t \) is an ARMA(\(p, q\)) process with representation given in Definition 2.5.1.

(i) If the roots of the polynomial \( \phi(z) \) lie outside the unit circle, and are greater than \( (1 + \delta) \) (for some \( \delta > 0 \)), then \( X_t \) almost surely has the solution

\[
X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \tag{2.19}
\]

where for \( j > q \), \( a_j = [A^j]_{1,1} + \sum_{i=1}^{q} \theta_i [A^{j-i}]_{1,1}, \) with

\[
A = \begin{pmatrix}
\phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\
1 & 0 & \ldots & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \ldots & \ldots & 1 & 0
\end{pmatrix}.
\]
where $\sum_j |a_j| < \infty$ (we note that really $a_j = a_j(\phi, \theta)$ since its a function of $\{\phi_i\}$ and $\{\theta_i\}$). Moreover for all $j$,

$$|a_j| \leq K \rho^j$$  \hspace{1cm} (2.20)

for some finite constant $K$ and $1/(1 + \delta) < \rho < 1$.

(ii) If the roots of $\phi(z)$ lie both inside or outside the unit circle and are larger than $(1 + \delta)$ or less than $(1 + \delta)^{-1}$ for some $\delta > 0$, then we have

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j},$$  \hspace{1cm} (2.21)

(a vector AR(1) is not possible), where

$$|a_j| \leq K \rho^{|j|}$$  \hspace{1cm} (2.22)

for some finite constant $K$ and $1/(1 + \delta) < \rho < 1$.

(iii) If the absolute value of the roots of $\theta(z) = 1 + \sum_{j=1}^{q} \theta_j z^j$ are greater than $(1 + \delta)$, then (2.18) can be written as

$$X_t = \sum_{j=1}^{\infty} b_j X_{t-j} + \varepsilon_t,$$  \hspace{1cm} (2.23)

where

$$|b_j| \leq K \rho^j$$  \hspace{1cm} (2.24)

for some finite constant $K$ and $1/(1 + \delta) < \rho < 1$.

PROOF. We first prove (i) There are several way to prove the result. The proof we consider here, uses the VAR expansion given in Section 2.4.1; thus we avoid using the Backshift operator (however the same result can easily proved using the backshift). We write the ARMA process as a vector difference equation

$$X_t = AX_{t-1} + \varepsilon_t$$  \hspace{1cm} (2.25)
where $X_t' = (X_t, \ldots, X_{t-p+1})$, $\varepsilon_t' = (\varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, 0, \ldots, 0)$. Now iterating (2.25), we have

$$X_t =\sum_{j=0}^{\infty} A^j \varepsilon_{t-j},$$

concentrating on the first element of the vector $X_t$ we see that

$$X_t = \sum_{i=0}^{\infty} [A^i]_{1,1} (\varepsilon_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-i-j}).$$

Comparing (2.19) with the above it is clear that for $j > q$, $a_j = [A^j]_{1,1} + \sum_{i=1}^q \theta_i [A^{j-i}]_{1,1}$. Observe that the above representation is very similar to the AR(1). Indeed as we will show below the $A^j$ behaves in much the same way as the $\phi^j$ in AR(1) example. As with $\phi^j$, we will show that $A^j$ converges to zero as $j \to \infty$ (because the eigenvalues of $A$ are less than one). We now show that $|X_t| \leq K \sum_{j=1}^{\infty} \rho^j |\varepsilon_{t-j}|$ for some $0 < \rho < 1$, this will mean that $|a_j| \leq K \rho^j$. To bound $|X_t|$ we use (2.26)

$$|X_t| \leq \|X_t\|_2 \leq \sum_{j=0}^{\infty} \|A^j\|_{\text{spec}} \|\varepsilon_{t-j}\|_2.$$

Hence, by using Gelfand’s formula (see Appendix A) we have $\|A^j\|_{\text{spec}} \leq C \rho^j$ (for any $|\lambda_{\text{max}}(A)| < \rho < 1$, where $\lambda_{\text{max}}(A)$ denotes the largest maximum eigenvalue of the matrix $A$), which gives the corresponding bound for $|a_j|$.

To prove (ii) we use the backshift operator. This requires the power series expansion of $\frac{\theta(z)}{\phi(z)}$. If the roots of $\phi(z)$ are distinct, then it is straightforward to write $\phi(z)^{-1}$ in terms of partial fractions which uses a convergent power series for $|z| = 1$. This expansion immediately gives the the linear coefficients $a_j$ and show that $|a_j| \leq C (1 + \delta)^{-|j|}$ for some finite constant $C$. On the other hand, if there are multiple roots, say the roots of $\phi(z)$ are $\lambda_1, \ldots, \lambda_s$ with multiplicity $m_1, \ldots, m_s$ (where $\sum_{j=1}^s m_s = p$) then we need to adjust the partial fraction expansion. It can be shown that $|a_j| \leq C |j|^{\max_s |m_s|} (1 + \delta)^{-|j|}$. We note that for every $(1 + \delta)^{-1} < \rho < 1$, there exists a constant such that $|j|^{\max_s |m_s|}(1 + \delta)^{-|j|} \leq C \rho^{|j|}$, thus we obtain the desired result.

To show (iii) we use a similar proof to (i), and omit the details.

\[\square\]

**Corollary 2.5.1** An ARMA process is invertible if the roots of $\theta(B)$ (the MA coefficients) lie outside the unit circle and causal if the roots of $\phi(B)$ (the AR coefficients) lie outside the unit circle.
An AR(p) process and an MA(q) process is identifiable (meaning there is only one model associated to one solution). However, the ARMA is not necessarily identifiable. The problem arises when the characteristic polynomial of the AR and MA part of the model share common roots. A simple example is $X_t = \varepsilon_t$, this also satisfies the representation $X_t - \phi X_{t-1} = \varepsilon_t - \phi \varepsilon_{t-1}$ etc. Therefore it is not possible to identify common factors in the polynomials.

One of the main advantages of the invertibility property is in prediction and estimation. We will consider this in detail below. It is worth noting that even if an ARMA process is not invertible, one can generate a time series which has identical correlation structure but is invertible (see Section 3.3).

### 2.6 Simulating from an Autoregressive process

#### Simulating from a Gaussian AR process

It is straightforward to simulate from an AR process with Gaussian innovations, $\{\varepsilon_t\}$. Given the autoregressive structure we can deduce the correlation structure (see Chapter 3) (regardless of the distribution of the innovations). Furthermore, from Lemma 2.5.1(ii) we observe that all AR processes can be written as the infinite sum of the innovations. Thus if the innovations are Gaussian, so is the AR process. This allows us to deduce the joint distribution of $X_1, \ldots, X_p$, which in turn allows us generate the AR($p$) process.

We illustrate the details with with an AR(1) process. Suppose $X_t = \phi_1 X_{t-1} + \varepsilon_t$ where $\{\varepsilon_t\}$ are iid standard normal random variables (note that for Gaussian processes it is impossible to discriminate between causal and non-causal processes - see Section 3.3, therefore we will assume $|\phi_1| < 1$). We will show in Section 3.1, equation (3.1) that the autocovariance of an AR(1) is

$$c(r) = \phi_1^r \sum_{j=0}^{\infty} \phi_1^{2j} = \frac{\phi_1^r}{1 - \phi_1^2}.$$ 

Therefore, the marginal distribution of $X_t$ is Gaussian with variance $(1 - \phi_1^2)^{-1}$. Therefore, to simulate an AR(1) Gaussian time series, we draw from a Gaussian time series with mean zero and variance $(1 - \phi_1^2)^{-1}$, calling this $X_1$. We then iterate for $2 \leq t$, $X_t = \phi_1 X_{t-1} + \varepsilon_t$. This will give us a stationary realization from an AR(1) Gaussian time series.
Note the function `arima.sim` is a routine in R which does the above. See below for details.

**Simulating from a non-Gaussian AR model**

Unlike the Gaussian AR process it is difficult to simulate a non-Gaussian model, but we can obtain a very close approximation. This is because if the innovations are non-Gaussian but known it is not clear what the distribution of \( X_t \) will be. Here we describe how to obtain a close approximation in the case that the AR process is causal.

Again we describe the method for the AR(1). Let \( \{ X_t \} \) be an AR(1) process, \( X_t = \phi_1 X_{t-1} + \varepsilon_t \), which has stationary, causal solution

\[
X_t = \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}.
\]

To simulate from the above model, we set \( \tilde{X}_1 = 0 \). Then obtain the iteration \( \tilde{X}_t = \phi_1 \tilde{X}_{t-1} + \varepsilon_t \) for \( t \geq 2 \). We note that the solution of this equation is

\[
\tilde{X}_t = \sum_{j=0}^{t} \phi_1^j \varepsilon_{t-j}.
\]

We recall from Lemma 2.5.1 that \( |X_t - \tilde{X}_t| \leq |\phi_1|^t \sum_{j=0}^{\infty} |\phi_1^j \varepsilon_{-j}| \), which converges geometrically fast to zero. Thus if we choose a large \( n \) to allow ‘burn in’ and use \( \{ \tilde{X}_t; t \geq n \} \) in the simulations we have a simulation which is close to a stationary solution from an AR(1) process.

**Simulating from an Integrated process**

To simulate from an integrated process ARIMA\((p, 1, q)\) \((1 - B)Y_t = X_t\), where \( X_t \) is a causal ARMA\((p, q)\) process. We first simulate \( \{ X_t \} \) using the method above. Then we define the recursion \( Y_1 = X_1 \) and for \( t > 1 \)

\[
Y_t = Y_{t-1} + X_t.
\]

Thus giving a realisation from an ARIMA\((p, 1, q)\).
Simulating from a non-Gaussian non-causal model

Suppose that $X_t$ satisfies the representation

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t,$$

whose characteristic function have roots both inside and outside the unit circle. Thus, the stationary solution of this equation is not causal. It is not possible to simulate from this equation. To see why, consider directly simulating from $X_t = 2X_{t-1} + \varepsilon_t$ without rearranging it as $X_{t-1} = \frac{1}{2} X_t - \frac{1}{2} \varepsilon_t$, the solution would explode. Now if the roots are both inside and outside the unit circle, there would not be a way to rearrange the equation to iterate a stationary solution. There are two methods to remedy this problem:

(i) From Lemma 2.5.1(ii) we recall that $X_t$ has the solution

$$X_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j},$$

(2.27)

where the coefficients $a_j$ are determined from the characteristic equation. Thus to simulate the process we use the above representation, though we do need to truncate the number of terms in (2.27) and use

$$\tilde{X}_t = \sum_{j=-M}^{M} a_j \varepsilon_{t-j}.$$

(ii) The above is a brute force method is an approximation which is also difficult to evaluate. There are simpler methods, if one studies the roots of the characteristic equation.

Let us suppose that $\{\lambda_j; j_1 = 1, \ldots, p_1\}$ are the roots of $\phi(z)$ which lie outside the unit circle and $\{\mu_j; j_2 = 1, \ldots, p_2\}$ are the roots which lie inside the unit circle. For ease of calculation we will assume the roots are distinct.
We can rewrite $\phi(z)^{-1}$ as

$$\phi(z)^{-1} = \frac{1}{\prod_{j_1=1}^{p_1}(1-\lambda_{j_1}z) \cdot \prod_{j_2=1}^{p_2}(1-\mu_{j_2}z)}$$

$$= \sum_{j_1=1}^{p_1} \frac{c_{j_1}}{(1-\lambda_{j_1}z)} + \sum_{j_2=1}^{p_2} \frac{d_{j_d}}{(1-\mu_{j_d}z)}$$

$$= \sum_{j_1=1}^{p_1} \frac{c_{j_1}}{(1-\lambda_{j_1}z)} - \sum_{j_2=1}^{p_2} \frac{d_{j_d}}{\mu_{j_d}z(1-\mu_{j_d}^{-1}z^{-1})}$$

Thus the solution of $X_t$ is

$$X_t = \phi(B)^{-1}\varepsilon_t = \sum_{j_1=1}^{p_1} \frac{c_{j_1}}{(1-\lambda_{j_1}B)}\varepsilon_t - \sum_{j_2=1}^{p_2} \frac{d_{j_d}}{\mu_{j_d}B(1-\mu_{j_d}^{-1}B^{-1})}\varepsilon_t$$

Let $Y_{j_1,t} = \lambda_{j_1}Y_{j_1,t-1} + \varepsilon_t$ and $Z_{j_2,t} = \mu_{j_2}Z_{j_2,t-1} + \varepsilon_t$ (thus the stationary solution is generated with $Z_{j_2,t-1} = \mu_{j_2}^{-1}Z_{j_2,t} - \mu_{j_2}^{-1}\varepsilon_t$). Generate the time series $\{Y_{j_1,t}; j_1 = 1,\ldots,p_1\}$ and $\{Y_{j_1,t}; j_1 = 1,\ldots,p_1\}$ using the method described above. Then the noncausal time series can be generated by using

$$X_t = \sum_{j_1=1}^{p_1} c_{j_1}Y_{j_1,t} - \sum_{j_2=1}^{p_2} d_{j_2}Z_{j_2,t}.$$  

An even easier method is represent $\phi(z)$ as the product of two polynomial, one whose roots are outside the unit circle ($\phi_1(z) = \prod_{i=1}^{p_1}(1-\lambda_{j_1}z)$) and one whose roots are inside the unit circle ($\phi_2(z) = \prod_{i=1}^{p_1}(1-\mu_{j_2}z)$). Then

$$\underbrace{\phi_1(B)\phi_2(B)}X_t = \varepsilon_t \Leftrightarrow \phi_2(B)X_t = \phi_1(B)^{-1}\varepsilon_t.$$  

Thus first define a causal stationary time series defined using the equation

$$\phi_1(B)Y_t = \varepsilon_t.$$  

Next, using $\{Y_t\}$ as the innovations, define a noncausal stationary time series defined using the recursion

$$\phi_2(B)X_t = Y_t.$$  

77
Comments:

- Remember $Y_{j,t}$ is generated using the past $\varepsilon_t$ and $Z_{j,t}$ is generated using future innovations. Therefore to ensure that the generated $\{Y_{j,t}\}$ and $\{Z_{j,t}\}$ are close to the stationary we need to ensure that the initial value of $Y_{j,t}$ is far in the past and the initial value for $Z_{j,t}$ is far in the future.

- If the roots are complex conjugates, then the corresponding $\{Y_{j,t}\}$ or $\{Z_{j,t}\}$ should be written as AR(2) models (to avoid complex processes).

R functions

Shumway and Stoffer (2006) and David Stoffer’s website gives a comprehensive introduction to time series R-functions.

The function `arima.sim` simulates from a Gaussian ARIMA process. For example, `arima.sim(list(order=c(2,0,0), ar = c(1.5, -0.75)), n=150)` simulates from the AR(2) model $X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t$, where the innovations are Gaussian.

**Exercise 2.6** In the following simulations, use non-Gaussian innovations.

(i) Simulate an AR(4) process with characteristic function

$$
\phi(z) = \left[1 - 0.8 \exp\left(i\frac{2\pi}{13}\right)z\right] \left[1 - 0.8 \exp\left(-i\frac{2\pi}{13}\right)z\right] \left[1 - 1.5 \exp\left(i\frac{2\pi}{5}\right)z\right] \left[1 - 1.5 \exp\left(-i\frac{2\pi}{5}\right)z\right].
$$

(ii) Simulate an AR(4) process with characteristic function

$$
\phi(z) = \left[1 - 0.8 \exp\left(i\frac{2\pi}{13}\right)z\right] \left[1 - 0.8 \exp\left(-i\frac{2\pi}{13}\right)z\right] \left[1 - \frac{2}{3} \exp\left(i\frac{2\pi}{5}\right)z\right] \left[1 - \frac{2}{3} \exp\left(-i\frac{2\pi}{5}\right)z\right].
$$

Do you observe any differences between these realisations?

2.7 Some diagnostics

Here we discuss some guidelines which allows us to discriminate between a pure autoregressive process and a pure moving average process; both with low orders. And also briefly discuss how to identify a “unit root” in the time series and whether the data has been over differenced.
2.7.1 ACF and PACF plots for checking for MA and AR behaviour

The ACF and PACF plots are the autocorrelations and partial autocorrelations estimated from the time series data (estimated assuming the time series is second order stationary). The ACF we came across is Chapter 1, the PACF we define in Chapter 3, however roughly it is the correlation between two time points after removing the linear dependence involving the observations inbetween. In R the functions are `acf` and `pacf`. Note that the PACF at lag zero is not given (as it does not make any sense).

The ACF and PACF of an AR(1), AR(2), MA(1) and MA(2) are given in Figures 2.2-2.5.

We observe from Figure 2.2 and 2.3 (which give the ACF of and AR(1) and AR(2) process) that there is correlation at all lags (though it reduces for large lags). However, we see from the PACF for the AR(1) has only one large coefficient at lag one and the PACF plot of the AR(2) has two large coefficients at lag one and two. This suggests that the ACF and PACF plot can be used to diagnose autoregressive behaviour and its order.

Similarly, we observe from Figures 2.4 and 2.5 (which give the ACF of and MA(1) and MA(2) process) that there is no real correlation in the ACF plots after lag one and two respectively, but the PACF plots are more ambiguous (there seems to be correlations at several lags).

![Figure 2.2: ACF and PACF plot of an AR(1), $X_t = 0.5X_{t-1} + \varepsilon_t$, $n = 400$](image)

Figure 2.2: ACF and PACF plot of an AR(1), $X_t = 0.5X_{t-1} + \varepsilon_t$, $n = 400$
2.7.2 Checking for unit roots

We recall that for an AR(1) process, the unit root corresponds to $X_t = X_{t-1} + \epsilon_t$ i.e. $\phi = 1$. Thus to check for unit root type behaviour we estimate $\phi$ and see how close $\phi$ is to one. We can formally turn this into a statistical test $H_0 : \phi = 1$ vs. $H_A : |\phi| < 1$ and there several tests for this, the most famous is the Dickey-Fuller test. Rather intriguingly, the distribution of $\hat{\phi}$ (using the least squares estimator) does not follow a normal distribution with a $\sqrt{n}$-rate!

Extending the the unit root to the AR($p$) process, the unit root corresponds to $(1-B)\phi(B)X_t = \epsilon_t$ where $\phi(B)$ is an order $(p-1)$-polynomial (this is the same as saying $X_t - X_{t-1}$ is a stationary AR($p-1$) process). Checking for unit root is the same as checking that the sum of all the AR coefficients is equal to one. This is easily seen by noting that $\bar{\phi}(1) = 0$ where $\bar{\phi}(B) = (1 - B)\phi(B)$
Figure 2.5: ACF and PACF plot of an MA(2), $n = 400$

or

$$(1 - B)\phi(B)X_t = X_t - (\phi_1 - 1)X_{t-1} - (\phi_2 - \phi_1)X_{t-2} - (\phi_{p-1} - \phi_{p-2})X_{t-p+1} + \phi_{p-1}X_{t-p} = \varepsilon_t.$$ 

Thus we see that the sum of the AR coefficients is equal to one. Therefore to check for unit root behaviour in AR($p$) processes one can see how close the sum of the estimate AR coefficients $\sum_{j=1}^{p-1} \hat{\phi}_j$ is to one. Again this can be turned into a formal test.

In order to remove stochastic or deterministic trend one may difference the data. But if the data is over differenced one can induce spurious dependence in the data which is best avoided (estimation is terrible and prediction becomes a nightmare). One indicator of over differencing is the appearance of negative correlation at lag one in the data. This is illustrated in Figure 2.6, where for both data sets (difference of iid noise and differenced of an AR(2) process) we observe a large negative correlation at lag one.