Chapter 2

Stochastic Models for Time Series

2.1 Introduction

As we know, the subject of Statistics is primarily about modelling things. In this chapter we describe some of the most useful models which are adopted to model time series data, and which can be broadly termed *stochastic processes*.

Remark 1 *The word "stochastic" may be new to you. It is Greek in origin, and roughly means "pertaining to chance", or "random". The study of stochastic processes is a large subject in mathematical Statistics and applied probability, and we only touch on some of the issues here relevant to modelling time series.*

Models in statistics live in the world of mathematics, and are necessarily simplifications of reality. A useful sentence to recall when undertaking mathematical modelling of any kind is the following: "All models are wrong, some are useful." In essence this says that nothing that can be simply written down in mathematics can sensibly model exactly all the complexities of a system in the real world, but having said that, a mathematical model *may* capture the important features of the real–world phenomenon in question. The model may then facilitate predictions, for example.

The building blocks for statistical models are things called *random variables*, an idea that you have met in earlier courses. These exist in the world of mathematics, and are usually denoted by capital letters, such as X. Recall that random variables can be either discrete or continuous, and are defined by a probability density function (if continuous, often called a probability mass function if discrete), which itself defines the cumulative distribution function of the random variable.

Remark 2 *It is a matter of taste whether you think of a random variable defining its probability distribution or the other way around. Either way they are intrinsically linked.*

We do not need much in the way of properties of random variables in what follows, but there are several numbers associated with a random variable which we should understand.

2.2 Operators on random variables

We are familiar with what the terms "mean" and "variance" refer to when related to a set of numbers. The words have a meaning when related to a random variable as well, and we briefly discuss the ideas, along with a third which follows naturally.

2.2.1 Expectation

The *mean* (often called the *expectation*) of a random variable X is denoted $E(X)$, and is a number. You might think of it as the "average" value that X can take when sampled from. Formally, it is defined as

$$E(X) = \sum_x xP(X = x)$$

if X is discrete, where the sum is over all possible values of X, or as

$$E(X) = \int xf(x)\,dx$$

if X is continuous, where $f(x)$ is the density function of X and again the integral is over all possible values.

Example 1 *You have come across means of certain distributions. For example, the mean of a random variable with the Normal distribution $N(\mu, \sigma^2)$ is μ. The mean of the Binomial distribution $B(n, p)$ is np.*

You will not be requested to work out any expectations from "first principles" (i.e., using the definitions above) in what follows, but some properties of the expectation operator $E(\cdot)$ should be remembered, and are quite easily seen from the definition.

1. If $X \geq 0$ (strictly this means $P(X \geq 0) = 1$) then $E(X) \geq 0$.
2. If a is a constant, then $E(a) = a$.
3. If a is a constant, then $E(aX) = aE(X)$.
4. If X and Y are two random variables, $E(X+Y) = E(X) + E(Y)$.

2.2.2 Variance

The variance of a random variable is defined in terms of an expectation. Specifically, the variance of X is defined to be

$$\operatorname{Var}(X) = E(X - E(X))^2 .$$

In words the variance is the "mean of the squared difference from the mean". It is a measure of how "spread out" the distribution of X is, in relation to its mean.

Expanding the square gives (check this)

$$\operatorname{Var}(X) = E\left(X^2\right) - E(X)^2 ,$$

so the variance is also described in words as "the mean of the square minus the square of the mean." Either way it is a number. Its square root is often called the *standard deviation.*

Example 2 *You have come across the variance of certain distributions. For example, the variance of a random variable with the Normal distribution* $N(\mu, \sigma^2)$ *is* σ^2. *The variance of the Binomial distribution* $B(n,p)$ *is* $np(1-p)$.

Some useful properties of the variance operator are given below, all of which follow from the definition and the properties of the expectation operator.

1. For any X, $\operatorname{Var}(X) \geq 0$.

2. If a is a constant, then $\text{Var}(a) = 0$.

3. If a is a constant and X any random variable, then $\text{Var}(aX) = a^2\text{Var}(X)$.

A further property holds if two random variables X and Y are *independent* (which you should take as meaning that any event relating to the random variable X is independent of any event relating to Y, recalling the definition of independent events). In this case

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

2.2.3 Covariance and correlation

The final comment above begs the question: what is the variance of the sum of two variables which are *not* independent? Well, following the definition of a variance, we see

$$\begin{aligned}
\text{Var}(X+Y) &= E([(X+Y) - E(X+Y)]^2) \\
&= E([(X - E(X)) + (Y - E(Y))]^2) \\
&= E\left((X - E(X))^2 + (Y - E(Y))^2 + 2(X - E(X))(Y - E(Y))\right) \\
&= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)
\end{aligned}$$

where $\text{Cov}(X,Y)$ is called the *covariance* between X and Y, and is defined as

$$\text{Cov}(X,Y) := E((X - E(X))(Y - E(Y))).$$

On expanding the brackets above, we see

$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y).$$

A covariance is a number. Some properties of a covariance follow, and hold for any random variables X, Y, and Z.

1. $\text{Cov}(X,Y) = \text{Cov}(Y,X)$

2. If X and Y are independent, then $\text{Cov}(X,Y) = 0$.

3. $\text{Cov}(X+Z,Y) = \text{Cov}(X,Y) + \text{Cov}(Z,Y)$.

4. If a and b are constants, and X and Y random variables, then

$$\text{Cov}(aX, bY) = ab\text{Cov}(X,Y).$$

5. A covariance is not dimensionless. That is to say, its units are determined by the units of X and Y.

This final point makes a covariance hard to interpret as it stands. It is often standardised by dividing by the product of the standard deviations to define the *correlation* between X and Y. This will be denoted $\rho_{X,Y}$, where

$$\rho_{X,Y} := \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}.$$

Some properties of a correlation are:

1. A correlation has no units, it is dimensionless.
2. $-1 \leq \rho_{X,Y} \leq +1$.
3. If $\rho_{X,Y} = \pm 1$, then $X = aY + b$ for some constants a and b. That is to say, X and Y are exactly linearly related.

2.3 Describing models for time series

As mentioned, the models for time series we consider are called stochastic processes. In general, think of a stochastic process as a random variable $X(t)$ which varies with time – that is, its distribution may change as time changes. Equivalently in discrete time, think of $X(t)$ as a sequence of random variables $X(1), X(2), \ldots$. Naturally certain properties of the distribution of $X(t)$ may vary with time, such as its mean and variance. The mean (or expectation, as is usual we use these terms interchangeably) of the process $X(t)$ is $E(X(t))$, and the variance at time t is denoted $\operatorname{Var}(X(t))$.

As we are interested in how the process depends on previous values, it is natural to define the *autocovariance function*, being the covariance between values of $X(t)$ at two different time points, t_1 and t_2 say. This is denoted

$$\gamma(t_1, t_2) := \operatorname{Cov}(X(t_1), X(t_2)).$$

Note that when $t_1 = t_2$, the covariance is just the variance at time t_1. So the variance of a model contains no information which is not contained in the autocovariance function. In general we can think of the difference $t_2 - t_1$ as being the *lag.*

The two functions, the mean and autocovariance, contain much information about a stochastic model. They do not, however, contain total information about how the process evolves over time. The following definition classifies the type of stochastic models we will consider in this course.

Definition 3 *A stochastic process $X(t)$ is called* stationary *if its mean is constant, that is*

$$E(X(t)) = \mu$$

for all t, and its autocovariance function depends only on the lag, that is

$$Cov(X(t), X(t+\tau)) = \gamma(\tau)$$

for all t and τ.

In other words a process is to be called stationary if its expectation does not vary over time, and its autocovariance function depends only on the *distance* between the two time points, the lag, not their place in time. This implies that the variance is a constant, σ^2 say, as well.

Remark 3 *Actually the definition of stationarity given above is not the only one that is available. A stochastic process is called* strongly stationary *if the joint distribution of the sequence $X(t_1), \ldots, X(t_n)$ is the same as that of $X(t_1+\tau), \ldots, X(t_n+\tau)$ for all $t_1, \ldots, t_n$ and all τ. This is a very strong condition, and difficult to even begin to check in practice from a realisation of a stochastic process. In cases where the first two moments are not defined, a stochastic process that is not (weakly) stationary may be strongly stationary.*

We will, in general, consider only stationary processes in what follows.

2.3.1 The autocorrelation function

For a stationary stochastic process $X(t)$ the *autocorrelation function* (acf for short) is denoted $\rho(\tau)$ and given by

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \frac{\gamma(\tau)}{\sigma^2},$$

where $\gamma(\tau)$ is the autocovariance function (acvf for short) at lag τ. Some properties of the acf follow:

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

2. We have that $\rho(\tau) = \rho(-\tau)$ for all τ, since

$$\begin{aligned}
\rho(\tau) &= \frac{\gamma(\tau)}{\sigma^2} \\
&= \frac{\operatorname{Cov}(X(t), X(t+\tau))}{\sigma^2} \\
&= \frac{\operatorname{Cov}(X(t-\tau), X(t))}{\sigma^2} \\
&\qquad\qquad \text{(due to stationarity)} \\
&= \frac{\gamma(-\tau)}{\sigma^2} \\
&= \rho(-\tau).
\end{aligned}$$

3. $|\rho(\tau)| \leq 1$ – the proof of this is an exercise.

4. The autocorrelation function does not uniquely specify a stochastic process; two different processes could have the same acf.

2.4 Examples of stochastic processes

We conclude this chapter with a survey of the most popular stochastic processes used to model time series data. Note that the models depend on *parameters*, such as the variance σ^2. Recall that "$\sim$" is read "is distributed as".

2.4.1 White noise

The process $Z(t)$ will be used in future to denote a purely random sequence of i.i.d. variables – so all values of Z are from the same distribution, each value independent of the others. This is a basic "building block" for other processes, and is called *white noise.*

Since there is no dependence between successive values we have

$$\gamma(k) = 0$$

for all $k \neq 0$, and moreover

$$\gamma(0) = \sigma^2,$$

the variance of the process.

Hence

$$\rho(k) = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Example 4 *If $Z(t) \sim N(0, \sigma^2)$ for all $t = 1, 2, \ldots$ independently, this defines a white noise process. Think of this process to fix ideas where you see $Z(t)$ in future.*

A white noise process is (strongly) stationary.

2.4.2 Random walk

Let $Z(t)$ be a white noise process with mean μ and variance σ^2. Let $X(t)$ be defined by

$$X(t) = X(t-1) + Z(t),$$

called a *random walk* process. To start with, it is usual to set $X(1) = Z(1)$, so that

$$X(t) = \sum_{i=1}^{t} Z(i).$$

From this we see that

$$E(X(t)) = t\mu$$

and

$$\mathrm{Var}(X(t)) = t\sigma^2$$

so the process is *not* stationary. However, the process

$$\begin{aligned} \nabla X(t) &= X(t) - X(t-1) \\ &= Z(t) \end{aligned}$$

is just white noise, so *is* stationary.

Random walk models are often used to model the price of a stock over successive days.

2.4.3 Moving average processes

Let $Z(t)$ denote a white noise process, with mean 0 and variance σ^2. The process $X(t)$ is said to be a *moving average process of order* q if

$$X(t) = \beta_0 Z(t) + \beta_1 Z(t-1) + \cdots + \beta_q Z(t-q) \tag{2.1}$$

for some constants $\beta_0, \beta_1, \ldots, \beta_q$, with usually $\beta_0 = 1$.

In words, a moving average process of order q (an MA(q) for short) is a weighted sum of the last q values of a white noise process Z, plus (when $\beta_0 = 1$) the new value from Z. Such models have quite widespread usage, particularly in modelling financial markets, where the effect of some "random" event (sometimes called an *innovation*) can have an immediate impact and also a "shock wave" effect on later time periods.

Some properties of an MA(q) process $X(t)$ are as follows:

1. $E(X(t)) = 0$.

2. $$\mathrm{Var}(X(t)) = \sigma^2 \sum_{i=0}^{q} \beta_i^2.$$

3. The autocovariance function at lag k is

$$\begin{aligned}
\gamma(k) &= \mathrm{Cov}(X(t), X(t+k)) \\
&= \mathrm{Cov}(\beta_0 Z(t) + \beta_1 Z(t-1) + \cdots + \beta_q Z(t-q), \\
&\qquad \beta_0 Z(t+k) + \beta_1 Z(t+k-1) + \cdots + \beta_q Z(t+k-q)) \\
&= \begin{cases} 0 & \text{if } k > q, \\ \sigma^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k} & \text{for } k = 0, 1, \ldots, q \\ \gamma(-k) & \text{for } k < 0. \end{cases}
\end{aligned}$$

4. As neither the mean nor acvf depend on time, the process is stationary.

5. From 2. and 3. above we see that the acf is

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

and so "cuts off" at lag q. This property in part characterises an MA(q) process.

Example 5 *Consider the MA(2) process*

$$\begin{aligned} X(t) &= Z(t) + Z(t-1) + Z(t-2) \\ &= \sum_{j=0}^{2} Z(t-j). \end{aligned}$$

Then

$$\begin{aligned} E(X(t)) &= E\left(\sum_{j=0}^{2} Z(t-j)\right) \\ &= \sum_{j=0}^{2} E(Z(t-j)) \\ &= 0. \end{aligned}$$

Further,

$$\begin{aligned} \gamma(k) &= E(X(t)X(t+k)) \\ &= E\left(\sum_{j=0}^{2} Z(t-j) \sum_{j=0}^{2} Z(t+k-j)\right) \\ &= E((Z(t) + Z(t-1) + Z(t-2))(Z(t+k) + Z(t+k-1) + Z(t+k-2))). \end{aligned}$$

Now when $k = 0$, this is

$$E((Z(t) + Z(t-1) + Z(t-2))(Z(t) + Z(t-1) + Z(t-2))) = 3\sigma^2.$$

For $k = 1$, we have

$$E((Z(t) + Z(t-1) + Z(t-2))(Z(t+1) + Z(t) + Z(t-1))) = 2\sigma^2.$$

When $k = 2$, $\gamma(k)$ is

$$E((Z(t) + Z(t-1) + Z(t-2))(Z(t+2) + Z(t+1) + Z(t))) = \sigma^2,$$

but for $k > 2$ we have

$$\gamma(k) = 0.$$

So in summary

$$\gamma(k) = \begin{cases} 0 & \text{if } k > 2, \\ \sigma^2(3-k) & \text{for } k = 0, 1, 2 \\ \gamma(-k) & \text{for } k < 0. \end{cases}$$

As a special case, consider the MA(1) process, with $\beta_0 = 1$. This has acf

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

Letting $\beta_1 = \theta$ say, this process is written

$$X(t) = Z(t) + \theta Z(t-1). \tag{2.2}$$

Consider the process

$$X(t) = Z(t) + \frac{1}{\theta} Z(t-1). \tag{2.3}$$

This process has acf

$$\begin{aligned} \rho(k) &= \frac{1/\theta}{1 + 1/\theta^2} \\ &= \frac{\theta}{1+\theta^2}, \end{aligned}$$

so it seems that the processes defined by (2.2) and (2.3) have the same acf. By substitution, we see that process (2.2) can be written

$$\begin{aligned} Z(t) &= X(t) - \theta Z(t-1) \\ &= X(t) - \theta X(t-1) + \theta^2 Z(t-2) \\ &= X(t) - \theta X(t-1) + \theta^2 X(t-2) - \cdots \end{aligned}$$

whereas for process (2.3) the same procedure leads to the infinite sum

$$Z(t) = X(t) - \frac{1}{\theta} X(t-1) + \frac{1}{\theta^2} X(t-2) - \cdots .$$

Now if $|\theta| < 1$ the sum for (2.2) *converges*, meaning essentially that it is well–defined (and not, say $\pm\infty$), whereas the sum for (2.3) does not converge. The converse is true when $|\theta| > 1$, with the sum for (2.3) being the one to converge. The two processes cannot sensibly both be defined.

Definition 6 *An MA(1) process*

$$X(t) = Z(t) + \theta Z(t-1)$$

is said to be invertible *if* $|\theta| < 1$.

Invertibility is a condition to impose on an MA(1) process to ensure that it is uniquely specified by its acf.

We can use the *backward shift operator* to define MA(q) processes. Let

$$BX(t) := X(t-1).$$

Further, let

$$B^j X(t) := X(t-j),$$

for $j = 1, 2, \ldots$. To define an MA(q) let

$$\begin{aligned} X(t) &= \left(\beta_0 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q\right) Z(t) \\ &= \theta(B) Z(t) \end{aligned}$$

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

Theorem 7 *An MA(q) process is invertible if the roots of the polynomial $\theta(B)$ are outside the unit circle. That is, all solutions to the equation $\theta(B) = 0$ are greater than unity in absolute value.*

Going back to our MA(1) case, for model (2.2) we have

$$\theta(B) = 1 + \theta B,$$

which has root $-1/\theta$. For model (2.3) we have

$$\theta(B) = 1 + \frac{1}{\theta} B,$$

which has root $-\theta$. Hence if $|\theta| < 1$ then model (2.2) is better and |root of $\theta(B)$|>1, whereas if $|\theta| > 1$ then model (2.3) is best and |root of $\theta(B)$|>1.

Example 8 *Is the MA(2) process*

$$X(t) = Z(t) - 0.8Z(t-1) - 0.6Z(t-2)$$

invertible? Well in this case

$$\theta(B) = 1 - 0.8B - 0.6B^2,$$

and the roots of this are

$$\frac{0.8 \pm \sqrt{0.64 + 2.4}}{-1.2},$$

i.e., –2.12 and 0.79. So the process is not invertible.

More generally, a stochastic process $X(t)$ is defined as being invertible if the random change it undergoes at time t, the "innovation" $Z(t)$, can be written

$$Z(t) = \sum_{j=0}^{\infty} \pi_j X(t-j),$$

for some constants $\{\pi_0, \pi_1, \ldots\}$ which absolutely converge, i.e., with

$$\sum_{j=0}^{\infty} |\pi_j| < \infty.$$

In effect invertibility dictates that the changes to a process can be written as a so–called "autoregressive process". We define this term fully in the next section.

Remark 4 *An arbitrary constant, μ say, could be added to the right hand side of (2.1) to give a process $X(t)$ with mean μ. This makes no difference to the acf, so has not been included here for simplicity.*

2.4.4 Autoregressive processes

As usual let $Z(t)$ denote a white noise process with mean zero and variance σ^2. Then an *autoregressive process of order* p can be written

$$X(t) = \alpha_1 X(t-1) + \alpha_2 X(t-2) + \cdots + \alpha_p X(t-p) + Z(t)$$

for some constants $\alpha_1, \ldots, \alpha_p$. You have met the term *regression* before – note that in an autoregressive process of order p (AR(p) for short) the value at time t, $X(t)$, is a linear combination of the previous p values, plus some white noise. So the next value is obtained by *regressing* on the previous p values, up to a random error.

The first order autoregressive process, AR(1), is

$$X(t) = \alpha X(t-1) + Z(t).$$

Now by continued back substitution in the right hand side of the above we have

$$
\begin{aligned}
X(t) &= \alpha\left(\alpha X(t-2) + Z(t-1)\right) + Z(t) \\
&= \alpha^2 X(t-2) + \alpha Z(t-1) + Z(t) \\
&= \alpha^2\left(\alpha X(t-3) + Z(t-2)\right) + \alpha Z(t-1) + Z(t) \\
&\vdots \\
&= Z(t) + \alpha Z(t-1) + \alpha^2 Z(t-2) + \cdots,
\end{aligned}
$$

which is an MA(∞) process (provided $|\alpha| < 1$, otherwise the final equation above is not properly defined). The two processes are said to be *dual.*

An AR(1) can be written

$$X(t) - \alpha X(t-1) = Z(t),$$

and so using the backward shift operator we have

$$(1 - \alpha B) X(t) = Z(t),$$

or

$$
\begin{aligned}
X(t) &= (1 - \alpha B)^{-1} Z(t) \\
&= \left(1 + \alpha B + \alpha^2 B^2 + \cdots\right) Z(t) \\
&= Z(t) + \alpha Z(t-1) + \alpha^2 Z(t-2) + \cdots.
\end{aligned}
$$

So we see that

$$
\begin{aligned}
E(X(t)) &= 0, \\
\operatorname{Var}(X(t)) &= \sigma^2\left(1 + \alpha^2 + \alpha^4 + \cdots\right),
\end{aligned}
$$

the latter being finite only if $|\alpha| < 1$, in which case (summing the series in the brackets above) we have

$$
\begin{aligned}
\operatorname{Var}(X(t)) &= \frac{\sigma^2}{1 - \alpha^2} \\
&= \sigma_X^2
\end{aligned}
$$

say.

The acvf at lag $k \geq 0$ is

$$
\begin{aligned}
\gamma(k) &= E\left(X(t)\,X(t+k)\right) \\
&= E\left(\sum_{i=0}^{\infty} \alpha^i Z(t-i) \sum_{j=0}^{\infty} \alpha^j Z(t+k-j)\right) \\
&= \sigma^2 \sum_{i=0}^{\infty} \alpha^i \alpha^{k+i} \\
&= \frac{\alpha^k \sigma^2}{1-\alpha^2} \\
&= \alpha^k \sigma_X^2,
\end{aligned}
$$

again provided $|\alpha| < 1$. So an AR(1) is stationary if $|\alpha| < 1$. The acf is simply

$$\rho(k) = \alpha^{|k|},$$

for $k = 0, \pm 1, \pm 2, \ldots$.. This simple form for the acf of an AR(1) process can often lead to its identification as a suitable model for a data set based on the acf of an observed series. We describe what the acf looks like for some types of cases:

1. If α is quite close to $+1$, 0.8 for instance, the acf will decay to zero quite slowly. This is due to there being relatively long–term dependence in the process.

2. If α is positive and quite close to zero, 0.2 for example, the acf cuts off quite quickly, due to there being only quite weak long–term dependence in the model.

3. For negative values of α, the acf will oscillate either side of zero on consecutive values.

Exercise 2.4.1 *Plot* $\alpha^{|k|}$ *for some values of* α *for* $k = 0, 1, 2, \ldots$ *and verify the statements above.*

For the general autoregressive model, an AR(p) can be written

$$
\begin{aligned}
\left(1 - \alpha_1 B - \cdots - \alpha_p B^p\right) X(t) &= Z(t) \\
\phi(B)\, X(t) &= Z(t)
\end{aligned}
$$

say, where $\phi(B)$ is the polynomial in B, $(1 - \alpha_1 B - \cdots - \alpha_p B^p)$.

In fact an AR(p) process $X(t) = \phi(B)^{-1} Z(t)$ can be expressed as an MA(∞) process for any p, which follows since it is always possible to express $\phi(B)^{-1}$ as some infinite sum

$$1 + \beta_1 B + \beta_2 B^2 + \cdots$$

for some constants $\beta_1, \beta_2, \ldots$. Therefore, for an AR(p) process $X(t)$ we have:

1. $E(X(t)) = 0$.

2. Following the result for the acvf of an MA process (with $\beta_0 = 1$), the acvf of an AR(p) is

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+k}.$$

3. In fact the model is stationary when

$$\sum_{i=1}^{\infty} |\beta_i| < \infty.$$

It is useful for a subsequent theorem to establish under what conditions the process $\phi(B)^{-1} Z(t)$ is stationary.

Proposition 9 *The process*

$$X(t) = \beta_0 Z(t) + \beta_1 Z(t-1) + \beta_2 Z(t-2) + \cdots$$

is stationary if and only if

$$\sum_{j=0}^{\infty} \beta_j^2 < \infty.$$

Proof. That the condition is necessary is clear, since recalling that the variance of the process is

$$\gamma(0) = \sigma^2 \sum_{j=0}^{\infty} \beta_j^2,$$

$X(t)$ can only have finite variance when $\sum_j \beta_j^2 < \infty$, and vice versa. To show that the condition is sufficient, assume $\sum_j \beta_j^2 < \infty$ and recall that

$$\left| \frac{\operatorname{Cov}(X(t), X(t-k))}{\sqrt{\operatorname{Var}(X(t)) \operatorname{Var}(X(t-k))}} \right| \leq 1$$

for all k and t. Now we know that

$$\gamma(0) = \operatorname{Var}(X(t)) = \operatorname{Var}(X(t-k)) < \infty$$

for all k, and so

$$|\gamma(k)| \leq \gamma(0) < \infty,$$

so $X(t)$ is stationary. ■

The following result is important, in that it provides a condition for stationarity which is reasonably easy to verify.

Theorem 10 *An AR(p) process $X(t)$ defined by*

$$\phi(B) X(t) = Z(t)$$

is stationary if and only if all the roots of $\phi(x)$ lie outside the unit circle (i.e., the solutions to the equation $\phi(x) = 0$ are greater than unity in absolute value).

Proof. Let us write the polynomial $\phi(x)$ as

$$\phi(x) = \prod_{r=1}^{p} (1 - b_r x),$$

so that the roots of $\phi(x)$ are $b_1^{-1}, \ldots, b_p^{-1}$. We assume for convenience that these roots are distinct, and note that they possibly are complex. Now we can find constants $c_1, \ldots, c_p$ such that

$$\frac{1}{\phi(x)} = \sum_{r=1}^{p} \frac{c_r}{(1 - b_r x)}.$$

For x small enough that $|b_r x| < 1$ for all r, we can expand $(1 - b_r x)^{-1}$ as a power series to give

$$\begin{aligned}\frac{1}{\phi(x)} &= \sum_{r=1}^{p} c_r \sum_{j=0}^{\infty} b_r^j x^j \\ &= \sum_{j=0}^{\infty} \beta_j x^j\end{aligned}$$

where $\beta_0 = 1$ and

$$\beta_j = \sum_{r=1}^{p} c_r b_r^j.$$

So we have

$$X(t) = \sum_{j=0}^{\infty} \beta_j Z(t-j)$$

which is stationary if and only if $\sum_j \beta_j^2 < \infty$ (by the earlier proposition). Now if for all r we have $|b_r| \leq b < 1$ for some b, then all roots of $\phi(x)$ lie outside the unit circle and

$$\beta_j^2 \leq \left(\sum_{r=1}^{p} c_r\right)^2 b^{2j}$$

and so

$$\sum_{j=0}^{\infty} \beta_j^2 \leq \frac{\left(\sum_{r=1}^{p} c_r\right)^2}{(1-b^2)} < \infty.$$

Alternatively, if $|b_k| \geq 1$ for at least one k, then $\sum_j \beta_j^2$ diverges, as the term involving b_k explodes. So for stationarity we require that the roots of $\phi(x)$ lie outside the unit circle. ■

We have met the special case of this result for the AR(1) process, which is only stationary when $|\alpha| < 1$, $\phi(B)$ in this case being $1 - \alpha B$, having the single root $1/\alpha$.

Example 11 *Consider the AR(2) process*

$$X(t) = 0.4X(t-1) + 0.2X(t-2) + Z(t).$$

This is

$$\phi(B) X(t) = Z(t)$$

where

$$\phi(B) = 1 - 0.4B - 0.2B^2.$$

Now the roots of $\phi(B)$ *are* -3.449 *and* 1.449, *so the process* $X(t)$ *is stationary.*

Remark 5 *Sometimes it can be easier to check whether the roots of* $\phi(B^{-1})$ *are* inside *the unit circle. For example,*

$$\phi\left(B^{-1}\right) = 1 - \frac{0.5}{B} + \frac{0.06}{B^2}$$

equals zero when

$$B^2 - 0.5B + 0.06 = (B - 0.3)(B - 0.2)$$

is zero. This occurs when B *is 0.3 and 0.2, so the process is stationary.*

The Yule–Walker equations

In principle it is possible to find the β's in the MA form of the AR from the α's (by equating coefficients, for example), but this can be rather tedious. An easier way is to start by *assuming* the process is stationary (or check it is by the theorem above). Take the defining equation

$$X(t) = \alpha_1 X(t-1) + \alpha_2 X(t-2) + \cdots + \alpha_p X(t-p) + Z(t)$$

and then

1. Multiply both sides by $X(t-k)$, for $k = 1, 2, \ldots, p$. This gives p equations.

2. Take expectations on both sides of each equation.

3. Divide both sides of each equation by σ_X^2, the variance of the process (which is finite and independent of t provided the process is stationary).

The left hand side of each equation is therefore of the form

$$\frac{E\left(X\left(t\right)X\left(t-k\right)\right)}{\sigma_X^2}=\rho\left(k\right),$$

and the k th equation can be written

$$\rho\left(k\right)=\alpha_1\rho\left(k-1\right)+\alpha_2\rho\left(k-2\right)+\cdots+\alpha_p\rho\left(k-p\right).$$

Now since $\rho\left(k\right)=\rho\left(-k\right)$, the p equations are

$$\begin{aligned}
\rho\left(1\right) &= \alpha_1\rho\left(0\right)+\alpha_2\rho\left(1\right)+\cdots+\alpha_p\rho\left(p-1\right)\\
\rho\left(2\right) &= \alpha_1\rho\left(1\right)+\alpha_2\rho\left(0\right)+\cdots+\alpha_p\rho\left(p-2\right)\\
&\vdots\\
\rho\left(p\right) &= \alpha_1\rho\left(p-1\right)+\alpha_2\rho\left(p-2\right)+\cdots+\alpha_p\rho\left(0\right),
\end{aligned}$$

which are known as the *Yule–Walker equations.* Note the $Z\left(t\right)$ term does not contribute to the above equations, being further ahead in time than (and therefore uncorrelated with) each $X\left(t-k\right)$.

If p is reasonably small, the Yule–Walker equations can be solved directly to give the acf. In fact the general solution is of the form

$$\rho\left(k\right)=A_1d_1^{|k|}+\cdots+A_pd_p^{|k|}$$

where $d_1,\ldots,d_p$ are the roots of the polynomial in D,

$$D^p-\alpha_1D^{p-1}-\cdots-\alpha_pD^0=0$$

and the A_i's are constants subject to the constraint $\sum_{i=1}^{p}A_i=1$ (since $\rho\left(0\right)=1$).

Example 12 *Consider the AR(2) process*

$$X\left(t\right)=\frac{1}{11}X\left(t-1\right)+\frac{1}{11}X\left(t-2\right)+Z\left(t\right).$$

To check that this process is stationary, we consider the roots of $\phi\left(B\right)=1-\frac{1}{11}B-\frac{1}{11}B^2$, *which are seen to be*

$$-\frac{\left(1+\sqrt{45}\right)}{2}\ \textit{and}\ -\frac{\left(1-\sqrt{45}\right)}{2}$$

and are both outside the unit circle. Hence the process is stationary. The solution of the Yule–Walker equations for the acf is of the form

$$\rho(k) = A_1 d_1^{|k|} + A_2 d_2^{|k|}$$

for $k = 0, \pm 1, \pm 2, \ldots,$ *where* A_1 *and* A_2 *are constants and* d_1 *and* d_2 *are roots of the quadratic*

$$D^2 - \frac{1}{11}D - \frac{1}{11} = 0.$$

Specifically, the roots are $\left(1 \pm \sqrt{45}\right)/22$. *Now since* $\rho(0) = 1,\ A_1 + A_2 = 1$ *and*

$$\begin{aligned} \rho(1) &= \frac{1}{11} + \frac{1}{11}\rho(-1) \\ &= \frac{1}{11} + \frac{1}{11}\rho(1), \end{aligned}$$

and so $\rho(1) = 1/10$. *Substituting this value into the solution of the Yule–Walker equation for* $\rho(1)$, *we have*

$$\frac{1}{10} = A_1\left(\frac{1+\sqrt{45}}{22}\right) + A_2\left(\frac{1-\sqrt{45}}{22}\right)$$

and so

$$\frac{11}{5} = A_1\left(1+\sqrt{45}\right) + (1 - A_1)\left(1-\sqrt{45}\right)$$

from which we deduce that

$$\begin{aligned} A_1 &= \frac{2\sqrt{45}+75}{150}, \\ A_2 &= \frac{75-2\sqrt{45}}{150}. \end{aligned}$$

Hence the acf is given by

$$\rho(k) = \left(\frac{2\sqrt{45}+75}{150}\right)\left(\frac{1+\sqrt{45}}{22}\right)^{|k|} + \left(\frac{75-2\sqrt{45}}{150}\right)\left(\frac{1-\sqrt{45}}{22}\right)^{|k|}.$$

Check that this gives $\rho(1) = 0.1$, *and find the next two values* $\rho(2)$ *and* $\rho(3)$.

Remark 6 *Matters are slightly more complicated if roots of*

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

are complex. If so, the complex roots occur in conjugate pairs of the polar form $a \pm bi = c(\cos(\theta) \pm i \sin(\theta))$ *for some* c *and* θ. *Considering these the* j *and* $(j+1)$ *th roots respectively, in such cases solutions to the set of difference equations include terms of the form* $A_j c^k \cos(\theta k) + A_{j+1} c^k \sin(\theta k)$ *corresponding to these complex roots, rather than terms of the form* $A_i d_i^{|k|}$ *for when a root* d_i *is real.*

2.4.5 ARMA models

Often called *mixed* models, these are combinations of an MA(q) and an AR(p) process, and are usually denoted ARMA(p,q). The model for $X(t)$ is of the following form

$$\begin{aligned} X(t) &= \alpha_1 X(t-1) + \cdots + \alpha_p X(t-p) + Z(t) \\ &\quad + \beta_1 Z(t-1) + \beta_2 Z(t-2) + \cdots + \beta_q Z(t-q) \end{aligned}$$

where $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ are constants and $Z(t)$ is white noise with zero mean and variance σ^2.

Expressed in terms of the backward shift operator B, an ARMA(p,q) can be written

$$\phi(B) X(t) = \theta(B) Z(t)$$

where

$$\begin{aligned} \phi(B) &= 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p, \\ \theta(B) &= 1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q. \end{aligned}$$

The facts below follow from properties of AR and MA models met earlier:

1. An ARMA(p,q) process is stationary if the solutions of $\phi(B) = 0$ are outside the unit circle.

2. An ARMA(p,q) process is said be invertible if the solutions of $\theta(B) = 0$ are outside the unit circle.

3. It is possible in principle to work out an equation for the acf of an ARMA(p,q), if rather tedious. See exercises 2 for an example.

ARMA models are of importance as on occasions it may be possible to fit a mixed model to a time series which is smaller (in terms of the number of parameters) than either a pure MA or a pure AR. This is an example of what is sometimes termed “The Principle of Parsimony”, in that pragmatically we seek a model which is as “small” as possible, in terms of the number of parameters required.

Since we know that an AR process can be written as a pure MA, it is therefore possible to express an ARMA model in the form

$$X(t) = \psi(B) Z(t)$$

for some polynomial

$$\psi(B) = \sum_{i=0}^{\infty} \psi_i B^i.$$

This representation, as will be seen next chapter, can be useful for creating confidence intervals for forecasting. Noting that

$$\psi(B) = \frac{\theta(B)}{\phi(B)}$$

it is possible to write an ARMA process as a pure AR, in that

$$\pi(B) X(t) = Z(t)$$

where

$$\pi(B) = \frac{\phi(B)}{\theta(B)}.$$

As we would typically write an AR process in the form

$$X(t) = \sum_{i=1}^{\infty} \pi_i X(t-i) + Z(t)$$

it is common to define $\pi(B)$ as

$$\pi(B) = 1 - \sum_{i=1}^{\infty} \pi_i B^i.$$

Obviously we have that

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

The weights $\{\psi_i\}$ or $\{\pi_i\}$ can be found either by direct division or more typically by equating coefficients in powers of B in either

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

or

$$\theta(B)\,\pi(B) = \phi(B)\,.$$

Example 13 *Consider the ARMA*$(1,1)$ *process*

$$X(t) = 0.8X(t-1) + Z(t) - 0.2Z(t-1)$$

in which

$$\phi(B) = 1 - 0.8B$$

and

$$\theta(B) = 1 - 0.2B$$

define a stationary and invertible process. Now

$$\begin{aligned}
\psi(B) &= \theta(B)\,\phi(B)^{-1} \\
&= (1-0.2B)(1-0.8B)^{-1} \\
&= (1-0.2B)\left(1+0.8B+0.8^2B^2+0.8^3B^3+\cdots\right) \\
&= 1+0.6B+0.48B^2+0.384B^3+\cdots+\left(0.8^j-0.2\times 0.8^{j-1}\right)B^j+\cdots
\end{aligned}$$

and so

$$\begin{aligned}
\psi_j &= 0.8^j - 0.2\times 0.8^{j-1} \\
&= 0.6\times 0.8^{j-1}
\end{aligned}$$

for $j = 1, 2, \ldots$ *. Similarly*

$$\begin{aligned}
\pi(B) &= \phi(B)\,\theta(B)^{-1} \\
&= (1-0.8B)\left(1+0.2B+0.2^2B^2+\cdots\right) \\
&= 1+(-0.8+0.2)\,B+\left(-0.8\times 0.2+0.2^2\right)B^2+\cdots \\
&\qquad +\left(-0.8\times 0.2^{j-1}+0.2^j\right)B^j+\cdots
\end{aligned}$$

and hence for $j = 1, 2, \ldots,$

$$\pi_j = -0.6\times 0.2^{j-1}.$$

2.4.6 ARIMA models

A general class of models arises when first we have to difference our initial model, d times say, before an ARMA(p, q) model is appropriate. This reflects the fact that most observed time series are non-stationary, and differencing can remove many types of non-stationary effects.

Starting with our model $X(t)$ define a new series $Y(t)$ by applying the difference operator d times to $X(t)$, that is

$$Y(t) = \nabla^d X(t) = (1 - B)^d X(t).$$

An ARMA(p, q) model for $Y(t)$ is then defined by

$$\begin{aligned} Y(t) &= \alpha_1 Y(t-1) + \cdots + \alpha_p Y(t-p) + Z(t) \\ &\quad + \beta_1 Z(t-1) + \beta_2 Z(t-2) + \cdots + \beta_q Z(t-q) \end{aligned}$$

and termed an ARIMA(p, d, q) model for $X(t)$, where the "I" stands for *integrated.*

This class of models is useful to fit for series which are non-stationary, as differencing often converts non-stationary series into stationary ones, as we have seen. Usually, $d = 1$, as first differencing often suffices.

Clearly we may write the model in the form

$$\phi(B) Y(t) = \theta(B) Z(t)$$

or equivalently

$$\phi(B)(1 - B)^d X(t) = \theta(B) Z(t).$$

Note that this represents a non-stationary model for $X(t)$, since roots of $\phi(B)(1 - B)^d$ include $B = 1$ which lies on, not outside, the unit circle.

2.5 SARIMA models

Many time series exhibit seasonal effects, of period s say, where for quarterly data $s = 4$, for example. Box and Jenkins extended their ARMA family to include models for periodic data, by defining a seasonal ARIMA model to be of the form

$$\phi(B) \Phi(B^s) W(t) = \theta(B) \Theta(B^s) Z(t)$$

where ϕ, Φ, θ and Θ are polynomials of orders p, P, q and Q respectively, and where

$$W(t) = \nabla^d \nabla_s^D X(t)$$

recalling that

$$\nabla_s X(t) = X(t) - X(t-s).$$

This defines a SARIMA model of order $(p,d,q) \times (P,D,Q)_s$. Typically $d, D \in \{0,1,2\}$, and in cases where both d and D are non-zero we apply the difference operators in "leftmost first" mode.

The above model appears rather complicated at first sight, and it is worth dwelling on its components in turn.

1. We obtain $W(t)$ from $X(t)$ by differencing. So for example if $d = D = 1$ and $s = 4$ we have

$$\begin{aligned} W(t) &= \nabla \nabla_4 X(t) \\ &= \nabla_4 X(t) - \nabla_4 X(t-1) \\ &= (X(t) - X(t-4)) - (X(t-1) - X(t-5)). \end{aligned}$$

2. For the seasonal AR term, $\Phi(B^s)$, suppose $P = 1$ for simplicity, then

$$\Phi(B^s) W(t) = W(t) - \alpha W(t-s)$$

for some constant α.

3. The seasonal MA component is governed by $\Theta(B^s)$, and when $Q = 1$ this acts on $Z(t)$ as follows:

$$\Theta(B^s) Z(t) = Z(t) + \beta Z(t-s)$$

for some constant β.

Example 14 *Consider the SARIMA model of order* $(0,0,1) \times (1,1,0)_4$. *This is*

$$\left(1 - \alpha B^4\right) W(t) = (1 + \beta B) Z(t)$$

for some constants α *and* β, *where*

$$W(t) = \nabla_4 X(t).$$

Written in terms of $X(t)$ the model is

$$\left(1-\alpha B^4\right)\left(X(t)-X(t-4)\right)=Z(t)+\beta Z(t-1),$$

which can be expressed as

$$X(t)=(\alpha+1)\,X(t-4)-\alpha X(t-8)+\beta Z(t-1)+Z(t).$$

Notice how $X(t)$ depends on both $X(t-4)$ and $X(t-8)$, as well as the two most recent values of Z. This is of course a special case of an $ARMA(8,1)$ process.

2.6 Summary

We have built up a family of models, a family which has proved useful in the modelling of time series data. All involve a white noise process as a building block; indeed, we have revealed that (possibly after differencing) all of our models can be written as (possibly infinite) MA processes. That is, any stationary AR, MA, or ARMA process can be written as

$$X(t)=\sum_{i=0}^{\infty}\psi_i Z(t-i),$$

for some constants $\psi_0, \psi_1, \ldots$. The process written above is sometimes called a *general linear process.*

This chapter has covered the mathematical theory of the models that we are hoping to fit to suitable data sets. All these models depend on *parameters* (the α's and β's, as well as σ^2), and to fit a model to the data we must estimate these parameters, in some sensible way, using the data. We must then decide whether the model chosen is appropriate. Having fitted a model to a given time series which appears suitable, we are likely to be interested in predicting future values. These issues concern us in the next chapter.

2.7 Learning outcomes

On completion of this chapter, learners should be able to:

1. Define the autocovariance and autocorrelation functions for a time series model.

2. Define and explain what it means to say that a process is (weakly) stationary.
3. Define what is meant by a *white noise* process.
4. Define a *random walk* model, and derive the basic properties of such a model.
5. Define a *moving average process of order* p, i.e., an MA(q).
6. Derive the mean, variance and autocovariance function of a stationary MA(q) process.
7. Define an MA(q) in terms of the backward shift operator B, and hence define when an MA(q) is invertible.
8. Recall conditions that ensure that an MA(∞) process is stationary.
9. Define an *autoregressive process of order* p, i.e., an AR(p).
10. Derive properties for an AR(1), including the mean, variance and autocorrelation function.
11. Define an AR(p) in terms of the backward shift operator B, and hence define when an AR(p) is stationary.
12. Derive the Yule–Walker equations for an AR(p) process.
13. Recall the general solution to the Yule–Walker equations, and solve these equations where computationally feasible without the aid of a computer.
14. Define an ARMA(p, q) process in terms of the backward shift operator B, and hence define when an ARMA(p, q) is stationary and/or invertible.
15. Express an ARMA(p, q) model as a pure MA process (when $p < 2$) or a pure AR process (when $q < 2$).
16. Define an ARIMA(p, d, q) process in terms of the backward shift operator B and the difference operator ∇.

17. Define a SARIMA$(p,d,q) \times (P,D,Q)_s$ process in terms of the backward shift operator B, the difference operator ∇ and the seasonal difference operator ∇_s.

18. Express a SARIMA$(p,d,q) \times (P,D,Q)_s$ process as an ARMA(p,q) process.

2.8 Exercises 2

In the following, assume $Z(t)$ is a white noise process of zero mean and variance σ^2.

1. Prove the following results for the covariance operator:

$$\begin{array}{cc} \text{(i)} & \operatorname{Cov}(X,Y) = \operatorname{Cov}(Y,X) \\ \text{(ii)} & \operatorname{Cov}(aX,bY) = ab\operatorname{Cov}(X,Y) \\ \text{(iii)} & \operatorname{Cov}(X+Z,Y) = \operatorname{Cov}(X,Y) + \operatorname{Cov}(Z,Y). \end{array}$$

2. Show that if $\rho(\tau)$ is the acf of a stochastic process $X(t)$ then $|\rho(\tau)| \le 1$. (Hint: Note that $\operatorname{Var}(\alpha X(t) + \beta X(t+\tau)) \ge 0$ for all α, β.)

3. The stationary process $\{X(t) : t = 1, 2, \ldots\}$ has autocovariance function $\gamma(k)$ say. The process $Y(t)$ is defined by taking the first differenced series of $X(t)$, i.e.,

$$Y(t) = \nabla X(t) = X(t) - X(t-1).$$

   Is the process $Y(t)$ stationary? Find the autocovariance of $Y(t)$ in terms of $\gamma(k)$.

4. Decide whether the following MA(2) process is invertible

$$X(t) = Z(t) - 0.7Z(t-1) + 0.2Z(t-2).$$

5. Is the MA(3) process

$$X(t) = Z(t) - 0.10Z(t-1) - 1.16Z(t-2) + 0.48Z(t-3)$$

   invertible? (Hint: Look for an integer solution to the associated polynomial, or else use a computer.)

6. Is the AR(2) process

$$X(t) = 0.9X(t-1) + 0.4X(t-2) + Z(t)$$

stationary?

7. Consider the following AR(2) process:

$$X(t) = \frac{1}{2}X(t-1) + \frac{1}{4}X(t-2) + Z(t).$$

Check that it is stationary, and using the Yule–Walker equations, find its autocorrelation function.

8. For what values of α is the AR(2) process

$$X(t) = -\alpha X(t-1) + 6\alpha^2 X(t-2) + Z(t)$$

stationary?

9. Find the autocorrelation function of the ARMA(1,1) model

$$X(t) = \alpha X(t-1) + Z(t) + \beta Z(t-1).$$

10. For the MA(1) process

$$X(t) = Z(t) + \theta Z(t-1)$$

show that

$$\frac{\operatorname{Cov}(X(t) - \rho(1)X(t-1), X(t-2) - \rho(1)X(t-1))}{\operatorname{Var}(X(t) - \rho(1)X(t-1))} = -\frac{\theta^2}{(1+\theta^2+\theta^4)}.$$

11. Show that the process

$$X(t) = 0.7X(t-1) + Z(t) - 0.1Z(t)$$

is both invertible and stationary. Express the process $X(t)$ as (i) a pure MA and (ii) a pure AR process.

12. Write the SARIMA model of order $(0,1,1) \times (1,0,0)_{12}$ as an ARMA process.