Chapter 3

Some Probability Models for Stationary Series

3.1 Orientation

The techniques discussed so far are largely descriptive. If we wish to

- forecast future values of a series and say how precise the forecasts are;
- understand the series (eg *how* were the data generated?);
- do inference (eg *is* there really a trend?)

we need a more precise specification of a time series. Since we want our discussion to apply to time series that are not completely predictable, a key requirement is to be precise about how randomness enters.

We embark on that next. It entails constructing *probability models* for time series, ie models framed explicitly in terms of random variables. It is simplest to do this first for *stationary* series. It will then turn out to be fairly easy to extend the ideas to non-stationary cases.

With probability models available, a general strategy for analyzing time series is:

- plot the series and note its features;
- remove any trend and seasonal components;
- choose a model to fit what's left, the residuals;
- for forecasting, forecast the residuals, then invert the transformations above (ie put back the trend and seasonal components) to get forecasts of the original series.

Residuals might often be stationary even when the original time series isn't. That's why it makes sense to concentrate first on stationary series.

3.2 Preliminaries

Model: Suppose that

$$X_1, X_2, \ldots, X_t, \ldots \text{ are random variables}$$

and that the observed time series $x_1, x_2, \ldots, x_t, \ldots$ consists of a realization of their values. This is an example of a basic *model for data*. It's the underlying idea in much of Statistics.
In many other areas of Statistics it's appropriate to suppose that the X_t are independent. For time series analysis, however, we want to allow the possibility that an observation at one time might be influenced by values at other times. We also want to capture the idea that the form of this possible dependence stays the same through time. With this in view, define:

Stationarity (strong): When the joint distribution of $X_{t_1}, X_{t_2}, \ldots, X_{t_k}$ is the same as that of $X_{t_1+s}, X_{t_2+s}, \ldots, X_{t_k+s}$, for any s, k and $t_1, \ldots, t_k$, ie the probability properties of the sequence do not change over time, then the model (or the random process $\{X_t\}$) is said to be *strongly stationary*.

Consequences: in particular X_t then has the same distribution as X_{t+s}, so that, if the expectations and variances exist,

$$\begin{array}{lllll} \text{a)} \quad \mu_t & = E(X_t) & = E(X_{t+s}) & = \mu_{t+s} & = \mu, \text{ say} \\ \text{b)} \quad \sigma_t^2 & = \mathrm{Var}(X_t) & = \mathrm{Var}(X_{t+s}) & = \sigma_{t+s}^2 & = \sigma^2, \text{ say} \end{array}$$

and the dependence between variables depends only on their *relative* times: thus,

$$\text{c)} \quad \mathrm{Cov}(X_t, X_{t+h}) \quad = \quad \mathrm{Cov}(X_{t+s}, X_{t+s+h}) \quad = \quad \gamma_h, \text{ say}$$

is a function only of h. γ_h is called the *covariance function at lag h*. Note that knowing it for all h tells us what the variance σ^2 is too, since $\gamma_0 = \sigma^2$.

Stationarity (weak, or second order): A model is said to be weakly stationary if its means, variances and covariances are finite and the properties (a)–(c) above hold. Unlike *strong stationarity*, weak stationarity says nothing about other aspects of the distributions of the X_ts. Thus for a weakly stationary process:

$$\begin{aligned} E(X_t) &= \mu \\ \mathrm{Var}(X_t) &= \sigma^2 \\ \mathrm{Cov}(X_t, X_{t+h}) &= \gamma_h, \end{aligned}$$

whatever the value of t.

We will take *stationarity* in what follows to mean weak stationarity.

Throughout the current chapter we will also suppose that $\mu = 0$. This is equivalent to concentrating attention on $X_t - \mu$, and is reasonable from the point of view of our strategy of modelling only the residual component of a time series, leaving the others aside till later.

(Theoretical) Autocorrelation Function: the (theoretical) auto-correlation at lag h is defined, for a weakly stationary process, as

$$\rho_h = \text{cor}(X_t, X_{t+h}) = \frac{\text{Cov}(X_t, X_{t+h})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+h})}} = \frac{\gamma_h}{\gamma_0} \quad h = 0, \pm 1, \ldots$$

and ρ_h regarded as a function of h is called the (theoretical) autocorrelation function (acf) of the process. ρ_h is the theoretical analogue of the sample autocorrelation function r_h (though the latter can be calculated even when the series is not stationary).

Since ρ_h is a correlation, it has the simple properties:

$$\begin{array}{lrcll} \text{Properties:} & \text{a)} & -1 \leq & \rho_h & \leq 1 \\ & \text{b)} & \rho_h & = & \rho_{-h} \\ & \text{c)} & \rho_0 & = & 1 \\ & \text{d)} & \rho_h & = & 0 \quad \text{if } X_t \text{ and } X_{t+h} \text{ independent} \end{array}$$

(The converse of (d) – that a zero correlation ρ_h implies independent X_t and X_{t+h} – is not true in general, though it is in the special case when X_t and X_{t+h} are Normally distributed variables.)

Review Question: Is a sequence of random variables X_t that each take the value 1 with probability 1 a stationary model? If so, what are μ, σ^2 and γ_h?

3.3 Purely Random Series (White Noise)

This is the simplest genuinely random stationary model. It is defined as:

$$X_t = \epsilon_t$$

for uncorrelated random variables ϵ_t with

$$E(\epsilon_t) = 0, \quad \text{and}$$
$$\text{Var}(\epsilon_t) = \sigma^2.$$

A sequence $X_t = \epsilon_t$ with these properties is called a *white noise* sequence.
Notation:

$$X_t \text{ or } \epsilon_t \quad \frown \quad \text{WN}(0, \sigma^2)$$

For a white noise sequence the auto-correlation function is

$$\rho_h = \begin{cases} 0 & h \neq 0, \\ 1 & h = 0. \end{cases} \tag{3.1}$$

Data can rarely be described so simply, but this model is a basis for others with more structure.

Review Exercise: Make sure you can justify (3.1).

3.4 Autoregressive Models

What if there is dependence between successive observations in a time series, in the sense that the value of one observation is likely to influence the value of the next?
In autoregressive models the current value is influenced by previous values and by independent random perturbations as follows:

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \epsilon_t \tag{3.2}$$

where the αs are parameters and the ϵ_ts are white noise $\mathrm{WN}(0, \sigma^2)$. In this context the ϵ_ts are often called *innovations*.

Compare the defining relation (3.2) with *regression* of a response variable X_t on various explanatory variables. Here the explanatory variables are earlier observations of X, hence the *auto* prefix.

> *Example: Unemployment Figures.* If the unemployment total in month t is Y_t and a proportion $1-\alpha$ of those unemployed in any month find a job before the next month, then $Y_t = \alpha Y_{t-1} + \eta_t$, where η_t is the number of new unemployed this month. So, if $E(Y_t) = \mu$ for all t,
>
> $$Y_t - \mu = \alpha Y_{t-1} - \mu + \eta_t = \alpha(Y_{t-1} - \mu) + (\alpha - 1)\mu + \eta_t,$$
>
> which can be written as
>
> $$X_t \qquad\qquad\qquad = \alpha X_{t-1} \quad + \qquad \epsilon_t,$$
>
> for $X_t = Y_t - \mu$ and $(\alpha-1)\mu + \eta_t = \epsilon_t$, which is of the form (3.2) provided $E(X_t) = (\alpha-1)\mu + E(\eta_t) = 0$ and another condition holds.

Terminology:

(i) any stationary sequence X_t satisfying (3.2) will be called an *autoregressive process of order p*,

(ii) Notation: 'X_t is AR(p)',

(iii) the process is called *Gaussian* if the joint distribution of the X_t is Normal (which is so if the ϵ_t are Normal.)

Review Question: What else must be true of the η_t sequence in the Unemployment Figures example for X_t to be an AR(1) sequence?

Note that (3.2) *describes* how values of X_t are related to each other. It doesn't immediately show how to *construct* a process for which those relationships hold, or even to establish whether there are any processes for which they hold. A first question therefore is: 'Do processes X_t having properties (3.2) exist?' The answer of course is Yes, otherwise we wouldn't be paying any attention to them. Nevertheless it's worth exploring the question a little, since the thinking that shows that autoregressive processes *exist* (essentially by seeing how to construct them) reveals facts about their structure important when it comes to practical data analysis. To begin, we see what properties AR processes would have if they existed. This helps in showing that indeed they do exist. To fix ideas we look at the simplest case first, that in which $p = 1$.

3.4.1 AR(1)

This is the special case in which regression is on only the previous value:

$$X_t = \alpha X_{t-1} + \epsilon_t, \tag{3.3}$$

where the ϵ_t are WN$(0, \sigma^2)$. An AR(1) process is any stationary sequence X_t satisfying (3.3).

Suppose for the moment that such a sequence exists. If so, what properties should it have besides (3.3)?

Properties:

(a) **MA(∞) Representation**:
This is a way of writing an AR(1) X_t by expressing it entirely in terms of the innovation sequence $\{\epsilon_t\}$. It is the key to clarifying the question of existence of X_t and it aids calculation of the properties of the process.
Substitution in (3.3) gives:

$$\begin{aligned} X_t &= \alpha(\alpha X_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \alpha(\alpha[\alpha X_{t-3} + \epsilon_{t-2}] + \epsilon_{t-1}) + \epsilon_t \\ &\vdots \\ &= \alpha^k X_{t-k} + \sum_{i=0}^{k-1} \alpha^i \epsilon_{t-i} \\ &\to \sum_{i=0}^{\infty} \alpha^i \epsilon_{t-i} \quad \text{as } k \to \infty \quad \text{if } |\, \alpha \,| < 1. \end{aligned}$$

Thus, provided $|\, \alpha \,| < 1$,

$$X_t = \sum_{i=0}^{\infty} \alpha^i \epsilon_{t-i} \tag{3.4}$$

This is called the *infinite moving average* (MA(∞)) representation of the AR(1) sequence X_t. The reason for the terminology *moving average* will appear later.

An alternative derivation of the MA(∞) representation is as follows:
Use the backward shift operator B to write

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

as

$$X_t = \alpha B X_t + \epsilon_t,$$

so that

$$(1 - \alpha B) X_t = \epsilon_t,$$

and hence formally

$$\begin{aligned} X_t = (1 - \alpha B)^{-1} \epsilon_t &= \sum_{i=0}^{\infty} \alpha^i B^i \epsilon_t \\ &= \sum_{0}^{\infty} \alpha^i \epsilon_{t-i}, \quad \text{if } |\, \alpha \,| < 1. \end{aligned} \tag{3.5}$$

Here we have used the formal substitution

$$(1-\alpha B)^{-1} = \sum_{i=0}^{\infty} (\alpha B)^i,$$

which would certainly hold true if αB were an algebraic quantity with modulus less than 1. (Recall that for any number z with $|z| < 1$, $(1-z)^{-1} = \sum_0^\infty z^i$.) Though it isn't (it's an operator), the fact that (3.5) is the same as (3.4) suggests that nevertheless the usual rules of algebra apply to it.

(b) **Variance and acf**

From the representation above:

$$E(X_t) = \sum_0^\infty \alpha^i E(\epsilon_{t-i}) = 0. \tag{3.6}$$

$$\begin{aligned} \mathrm{Var}(X_t) &= \sum_0^\infty \alpha^{2i} \mathrm{Var}(\epsilon_{t-i}) \\ &= \frac{\sigma^2}{1-\alpha^2}, \quad |\alpha| < 1 \end{aligned} \tag{3.7}$$

$$\begin{aligned} \gamma_h = E(X_t X_{t+h}) &= E\left(\sum_{i,j=0}^{\infty} \alpha^i \epsilon_{t-i}\, \alpha^j\, \epsilon_{t+h-j} \right) \\ &= \sigma^2 \sum_{i=0}^{\infty} \alpha^i\, \alpha^{i+h} \\ &= \frac{\sigma^2\, \alpha^h}{1-\alpha^2} \end{aligned} \tag{3.8}$$

So the autocorrelation function of the AR(1) sequence is

$$\rho_h = \alpha^h, \qquad h = 0, \pm 1, \ldots. \tag{3.9}$$

From (3.6)–(3.9) we see that the mean, variance and autocorrelations do not depend on time t, and so the X_t given by (3.4) satisfy the conditions for stationarity.

Review Exercise: Make sure you can justify the steps in the derivation of (3.8) and (3.9). Sketch the form of the autocorrelation function for AR(1) when $0 < \alpha < 1$ and when $-1 < \alpha < 0$.

The derivation of (3.8) and (3.9) assumes that $|\alpha| < 1$ (where?). To see what happens if this isn't true, look at the following example.

Example: Random Walk.
If $\alpha = 1$ in the definition of AR(1) (3.3) then

$$X_t = X_{t-1} + \epsilon_t,$$

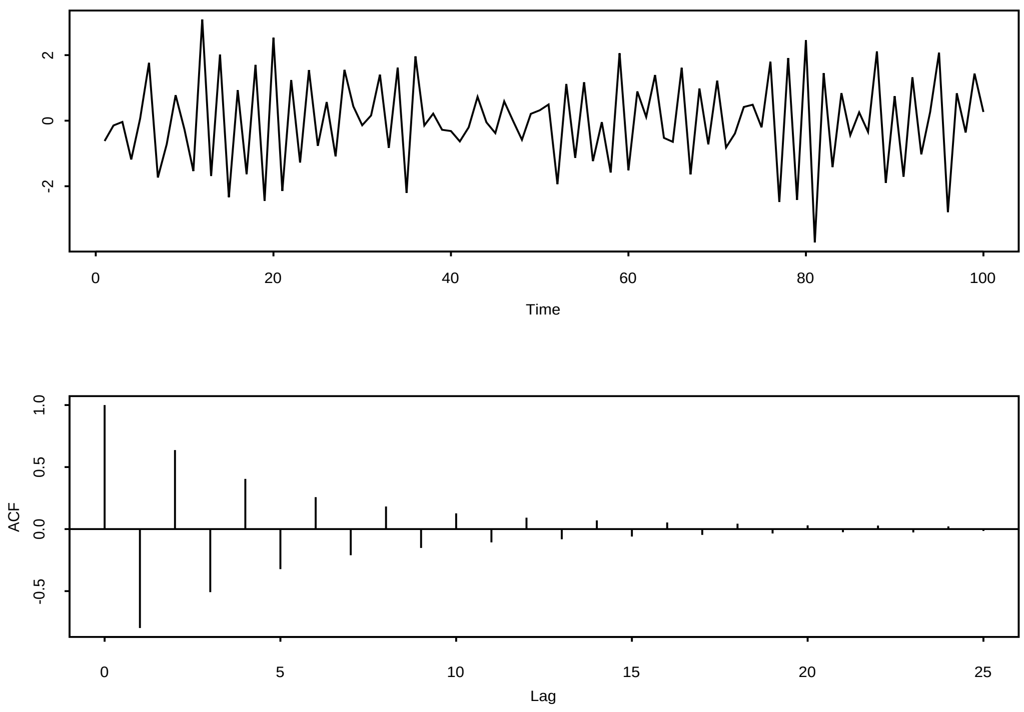
so
$$E(X_t) = constant$$
but
$$\mathrm{Var}(X_t) = \mathrm{Var}(X_{t-1}) + \sigma^2, \; \textit{contradicting stationarity unless } \sigma = 0.$$

So X_t non-stationary.
Thus in general for AR(1) the condition $|\alpha| < 1$ is closely bound up with stationarity: the example shows that if the condition is relaxed then stationarity may fail.
Note that nevertheless for the Random Walk differencing reduces X_t to stationarity.

(Here, it's clear that the only way for Var to be constant when $\sigma > 0$ is for it to be infinite. For AR(1) in general it turns out to be true that Var $< \infty$ is necessary and sufficient for stationarity.)

Example: An AR(1) process. The following Figure shows the acf and a realization of 100 observations from the AR(1) process $X_t = -0.8X_{t-1} + \epsilon_t$.

Figure 3.1: The process $X_t = -0.8X_{t-1} + \epsilon_t$



3.4.2 AR(p), $p > 1$

(a) **MA(∞) Representation**:
In this more general case we can again try to write X_t in terms only of the innovations

$\epsilon_t, \epsilon_{t-1}, \ldots$. Write the defining relation (3.2)

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \epsilon_t$$

as

$$X_t - \sum_{i=1}^{p} \alpha_i B^i X_t = \epsilon_t,$$

that is, as

$$\phi(B) X_t = \epsilon_t$$

where ϕ is the polynomial $\phi(x) = 1 - \sum_1^p \alpha_i x^i$, called the *autoregressive polynomial.* So, formally,

$$X_t = (\phi(B))^{-1} \epsilon_t.$$

In the special case of AR(1) $\phi(B)$ was $1 - \alpha B$ and we saw that $(\phi(B))^{-1}$ could be interpreted as $\sum (\alpha B)^i$. How should $(\phi(B))^{-1}$ be interpreted now? Suppose we think about the algebraic polynomial $\phi(x)$, where x is a (real or complex) number rather than an operator. If $\theta_1, \theta_2, \ldots, \theta_p$ are the roots of $\phi(x) = 0$, then we can write

$$\phi(x) = (1 - \theta_1^{-1} x) \ldots (1 - \theta_p^{-1} x) \tag{3.10}$$

so that

$$\begin{aligned}
\frac{1}{\phi(x)} &= \frac{1}{(1 - \theta_1^{-1} x) \ldots (1 - \theta_p^{-1} x)} \\
&= \frac{c_1}{(1 - \theta_1^{-1} x)} + \ldots + \frac{c_p}{(1 - \theta_p^{-1} x)} \qquad (3.11) \\
&\qquad \text{for constants } c_i \text{, if } \theta_i \text{ distinct} \\
&\qquad \text{(a } \textit{partial fraction} \text{ expansion)} \\
&= \sum_{j=1}^{p} c_j \sum_{i=0}^{\infty} (\theta_j^{-1} x)^i \epsilon_t, \quad \text{if } \mid \theta_j^{-1} x \mid < 1 \\
&\qquad \text{on expanding each of the terms } (1 - \theta_i^{-1} x)^{-1} \\
&\qquad \text{as for AR(1)}
\end{aligned}$$

So, formally

$$X_t = \sum_{j=1}^{p} c_j \sum_{i=0}^{\infty} (\theta_j^{-1} B)^i \epsilon_t = \sum_{i=0}^{\infty} \left(\sum_{j=1}^{p} c_j \theta_j^{-i} \right) \epsilon_{t-i},$$

that is, X_t is of the form

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \tag{3.12}$$

where the ψ_i coefficients depend on the α_is in (3.2):

$$\psi_i = \sum_{j=1}^{p} c_j \theta_j^{-i}. \tag{3.13}$$

Equation (3.12) is the MA(∞) representation for the AR(p) model, generalizing (3.4). Strictly, it applies in the case when the roots of $\phi(x) = 0$ are distinct, but there's a straightforward modification (details omitted) when two or more roots are equal.

Review Question: Through which part in the sum in (3.13) *does* ψ_i *depend on the* α_i*s?*
Review Exercise: Work out c_1 *and* c_2 *in* (3.11) *for the case* $\phi(x) = (1-0.5x)(1-0.25x)$.

(b) **Stationarity Condition: Causality**:
The representation in (3.12) was obtained by formal manipulations. When is it valid? That is, when does it give a (weakly) stationary process? To answer this question rigorously would take us too far astray (into conditions for convergence of infinite sums of random variables). But to get an informal idea we can argue as follows.
Stationarity is bound up with $\mathrm{Var}(X_t)$ being finite and not depending on t. (Other conditions are involved too but let's focus on the variance for now.) Whether stationary or not, X_t as defined in (3.12) is a sum of uncorrelated terms $\psi_i \epsilon_{t-i}$, and so its variance is the sum of their variances. Thus

$$\mathrm{Var}(X_t) = \sum_{i=0}^{\infty} \psi_i^2 \sigma^2 = \sum_{i=0}^{\infty} \left(\sum_{j=1}^{p} c_j \theta_j^{-i} \right)^2 \sigma^2 \tag{3.14}$$

This clearly doesn't depend on t, so the real question becomes: 'Is (3.14) $< \infty$?'. If it isn't, then X_t can't be stationary: if it is, then it turns out that so are $E(X_t)$ and $\mathrm{Cov}(X_t X_{t+h})$, and they don't depend on t, so X_t is stationary.

So when is (3.14) finite?

Note that the ith term in (3.14) is proportional to

$$\psi_i^2 = (c_1\theta_1^{-i} + \cdots + c_p\theta_p^{-i})^2$$

which

$$\begin{aligned} &\to 0 \qquad \text{geometrically fast as } i \to \infty \text{ if each } |\theta_j| > 1, \text{ and} \\ &\to \infty \qquad \text{as } i \to \infty \text{ if any } |\theta_j| \leq 1, \end{aligned}$$

suggesting that whether (3.14) holds or not is bound up with whether the θ_js are all greater than 1 in modulus or not.

(In a bit more detail, the argument here is that if every $|\theta_j| > 1$, then, as i gets larger, each θ_j^{-i} will tend to zero. Those θ_j^{-i} with the largest $|\theta_j|$s will go to zero the fastest, and those with the smallest $|\theta_j|$s the slowest, and in fact when i is large enough the former will be completely negligible compared to the latter. The sum $c_1\theta_1^{-i} + \cdots + c_p\theta_p^{-i}$ will therefore, for large i, be more or less equal to the term coming from the smallest $|\theta_j|$, $c_s\theta_s^{-i}$ say, and will be going to zero at the rate this term goes to zero. Thus

$$\psi_i^2 \sim c_s^2/\theta_s^{2i} \qquad \text{as } i \to \infty$$

where $\sim$ means the two sides are equivalent in the limit. The sum $\sum_i^\infty \psi_i^2\sigma^2$ will therefore be finite if $\sum_i^\infty c_s^2/\theta_s^{2i}$ is finite, that is, since c_s is a constant, if $\sum_i^\infty 1/\theta_s^{2i}$ is finite. But we're assuming all the $|\theta_j|$s are greater than 1, so this last sum is of a geometric progression with common ratio $1/\theta_s^2 < 1$, and so converges. Thus we've shown that when all the θ_js are greater than 1 in modulus the condition (3.14) will be satisfied, and X_t in (3.12) will be stationary.

On the other hand, if any of the $|\theta_j|$s are less than 1 then the corresponding θ_j^{-i} will not tend to zero as i gets larger, and therefore the sum $c_1\theta_1^{-i} + \cdots + c_p\theta_p^{-i}$ will not tend to zero either. But in that case the series in (3.14) can't converge (the individual terms must at least go to zero for it to stand any chance of converging), and so X_t can't be stationary.)

The conclusion from the above is that whether X_t given by (3.12) is stationary or not depends on whether or not all the θ_js are greater than 1 in modulus. Recall that the θ_js are the roots of the equation $\phi(x) = 0$, where ϕ is the autoregressive polynomial. In general, roots of polynomial equations like this are complex numbers (counting real numbers as a simple kind of complex number), and for a complex number, saying it is greater than 1 in modulus is the same as saying it lies outside the unit circle in the complex plane (the circle with unit radius centred at the origin). In this terminology our conclusion is that **the process given by** (3.12) **is stationary iff all roots of** $\phi(x) = 0$ **lie outside the unit circle.** This both settles the question of existence of AR(p) processes and gives a useful means to manipulate them.

The situation is important enough to justify some further terminology:

Definition: an AR(p) process satisfying (3.2) is said to be **causal** if it is stationary and can be represented in terms of the white noise variables ϵ_i in the MA(∞) form

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i},$$

for appropriate constants ψ_i.

We then have the following official version of the condition for causality:

$$\boxed{\textbf{AR}(p) \textbf{ causal iff all roots of } \phi(x) = 0 \textbf{ lie outside the unit circle.}}$$

Example: AR(1)
For the simple AR(1) model, $X_t = \alpha X_{t-1} + \epsilon_t$ and we know already from §3.4.1 that causality depends on the condition $|\alpha| < 1$.
In terms of the current general approach, for this model the autoregressive polynomial is $\phi(x) = 1 - \alpha x$, and so the root of $\phi(x) = 0$ is $\theta = 1/\alpha$. According to the condition above, therefore, the process is causal iff $|\theta| > 1$, that is, iff $|\alpha| < 1$, in agreement with the earlier finding.

Review Exercise: Explain how *the above settles the question of existence of AR(p) processes.*

(c) **ACF**:
Multiply the defining relation (3.2) for X_t by X_{t-k} and take expectations:

$$E(X_{t-k}X_t) = E(\alpha_1 X_{t-k}X_{t-1} + \ldots + \alpha_p X_{t-k}X_{t-p} + X_{t-k}\epsilon_t),$$

giving

$$\gamma_k = \alpha_1\gamma_{k-1} + \ldots + \alpha_p\gamma_{k-p}$$

and hence, dividing by $\gamma_0 = \text{Var}X_t$,

$$\rho_k = \sum_{j=1}^{p} \alpha_j \rho_{k-j}, \qquad \text{for } k \geq 1$$

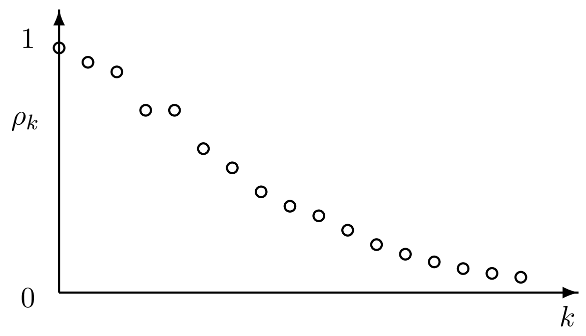
with boundary conditions $\rho_0 = 1$, and $\rho_{-k} = \rho_k$.
These are the *Yule-Walker difference equations* for the auto-correlations γ_k.

It is known that the general solution of difference equations is of the form

$$\rho_k = \sum_{j=1}^{p} \xi_j \theta_j^{-k} \tag{3.15}$$

where the θ_i are the roots of $\phi(x) = 1 - \sum_1^p \alpha_j x^j = 0$ (the autoregressive polynomial!) and the ξ_j are constants which ensure that the boundary conditions are satisfied. (If the roots are repeated the solution is slightly different.)

If the causality condition is satisfied, then all the θ_j are greater than 1 in modulus, so ρ_k given by (3.15) is a linear combination of geometrically-decaying terms – as for AR(1) – and the ACF will decay to zero for large lags approximately like that for AR(1), ie geometrically.



Example: An AR(2) process. Figure 3.2 overleaf shows a realization of the AR(2) process $X_t = 0.5x_{t-1} + 0.25X_{t-2}$, *and the process ACF.*

Review Question: What is the difference between causality and stationarity for an AR(p) process?

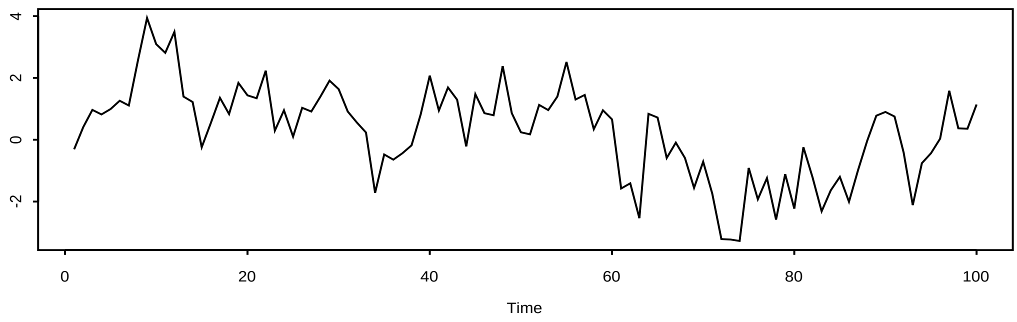
3.5 Moving Average Processes

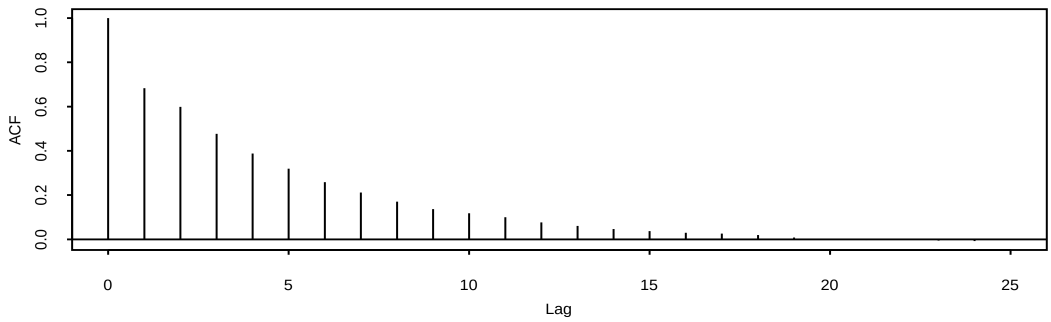
These are defined by

$$X_t = \beta_0 \epsilon_t + \beta_1 \epsilon_{t-1} + \ldots + \beta_q \epsilon_{t-q} \tag{3.16}$$

where the ϵ_ts are white noise $\text{WN}(0, \sigma^2)$.

Figure 3.2: Realization and ACF of the AR(2) model $X_t = 0.5X_{t-1} + 0.25X_{t-2} + \epsilon_t$





The coefficient β_0 can be, and usually is, taken to be 1.[1]

Terminology: A process defined as in (3.16) is called a *Moving Average Process of order q*, and denoted by MA(q).

(Such a process might be appropriate to represent an economic series influenced by independent factors, not all acting immediately. The value of the series might be expected to be a weighted sum of the most recent influences. Another example might be the response of a measuring instrument with inertia (an anemometer for example) for which the reading at any instant is an average of recent influences.)

Note that for MA(q) processes the question of *existence* which figured in connection with AR(p) does not arise: (3.16) is already a constructive definition.

[1]by scaling the ϵ_ts and modifying σ^2: ie by defining $\epsilon'_t = \beta_0\epsilon_t$ so that $\epsilon'_t \frown \mathrm{WN}(0, \beta_0^2\sigma^2)$ and $X_t = \epsilon'_t + (\beta_1/\beta_0)\epsilon'_{t-1} + \cdots + (\beta_q/\beta_0)\epsilon'_{t-q}$.

Properties of MA(q):

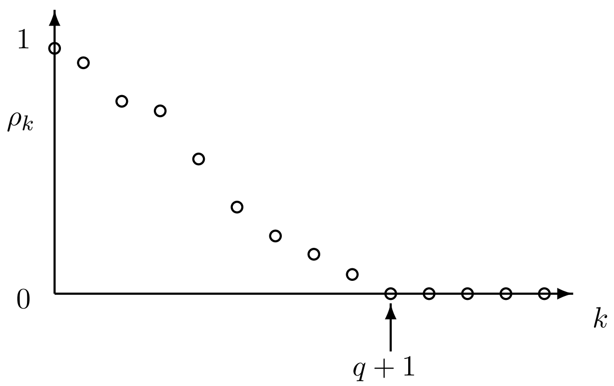
$$\begin{aligned}
E(X_t) &= 0,\\
\mathrm{Var}(X_t) &= \sigma^2 \sum_{i=0}^{q} \beta_i^2,\\
\mathrm{Cov}(X_t, X_{t+k}) &= E\left(\sum_{i=0}^{q} \beta_i \epsilon_{t-i} \sum_{j=0}^{q} \beta_j \epsilon_{t+k-j}\right)\\
&= \sum_{i,j} \beta_i \beta_j E(\epsilon_{t-i}\epsilon_{t+k-j})\\
&= \begin{cases} 0 & \text{if } k > q \\ \sigma^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k} & \text{if } k = 1, \ldots, q \end{cases}
\end{aligned}$$

because $E(\epsilon_{t-i}\epsilon_{t+k-j}) = \sigma^2$
if $i = j - k$ and $= 0$ if not.

So the acf has the form

$$\rho_k = \begin{cases} 0 & \text{for } k > q \\ \sum_0^{q-k} \beta_i \beta_{i+k} / \sum_0^q \beta_i^2 & \text{for } k = 1, \ldots, q. \end{cases} \tag{3.17}$$

– cutting off to zero beyond q.



Notes:

(a) MA(∞). The MA(∞) process of §3.4.1 is formally defined by letting $q \to \infty$ in (3.16).

(b) Stationarity holds iff $\mathrm{Var}(X_t) < \infty$, ie iff $\sum_i \beta_i^2 < \infty$.

(c) Non-uniqueness of the ACF. Look at the particular case of MA(1). For $X_t = \epsilon_t + \beta_1 \epsilon_{t-1}$, expression (3.17) gives $\rho_1 = \beta_1/(1 + \beta_1^2)$ and $\rho_k = 0$ for $k > 1$. This is the same acf as for $X_t = \epsilon_t + \frac{1}{\beta_1}\epsilon_{t-1}$. Thus there's a problem if we were hoping the acf could be used to identify a process.

(d) Invertible processes. The notion of an **invertible** process allows us to get round the difficulty in (c). When an MA process can be expressed as a stationary AR process it's said to be *invertible*. So, for example, the MA process above, $X_t = \epsilon_t + \beta_1 \epsilon_{t-1}$, can be written as

$$\begin{aligned}
X_t = \epsilon_t + \beta_1(X_{t-1} - \beta_1 \epsilon_{t-2}) &= \epsilon_t + \beta_1(X_{t-1} - \beta_1(X_{t-2} - \beta \epsilon_{t-3}))\\
&= \ldots\\
&= -\sum_{i=1}^{\infty} (-\beta_1)^i X_{t-i} + \epsilon_t
\end{aligned}$$

provided $|\beta_1| < 1$. Note that if $|\beta_1| \geq 1$ this manoeuvre doesn't work.

In general suppose

$$X_t = h(B)\epsilon_t$$

$$\text{where} \qquad h(x) = 1 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_q x^q.$$

Here, h is called the *moving average polynomial.* Suppose that the equation $h(x) = 0$ has q distinct roots $\nu_1, \ldots, \nu_q$ so that $h(x) = \prod_1^q (1 - (\nu_i^{-1} x))$. Then, from a partial fraction expansion as in §3.4.2,

$$\epsilon_t = (h(B))^{-1} X_t = \prod_{i=1}^{q} \left(1 - \frac{B}{\nu_i}\right)^{-1} X_t = \sum_{j=0}^{\infty} \delta_j X_{t-j}$$

for some constants δ_j, which is stationary iff (as in §3.4.2) the roots ν_i all lie outside the unit circle (ie $|\nu_i| > 1$ for all i). Thus this condition is necessary and sufficient for invertibility of the MA process.

MA(q) invertible iff all roots of $h(x) = 0$ lie outside the unit circle.

There is (at most) one *invertible* MA process corresponding to any particular autocorrelation function. Once attention is confined to invertible MA processes, therefore, estimation and prediction work smoothly (as will be seen). Note that invertibility ensures that recent values are more important than more distant ones in producing the next member in the series. This is a very natural restriction.

Example: An MA(1) process. Figure 3.3 shows a realization of 100 values of the MA(1) process $X_t = \epsilon_t + 0.8\epsilon_{t-1}$*, and the process ACF.*

Review Question: Is the MA(1) process in the example invertible?

3.6 ARMA Processes

These combine the characteristics of AR and MA processes.

Definition: X_t is an ARMA(p, q) process if it is stationary and satisfies

$$X_t = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \ldots + \beta_q \epsilon_{t-q} \tag{3.18}$$

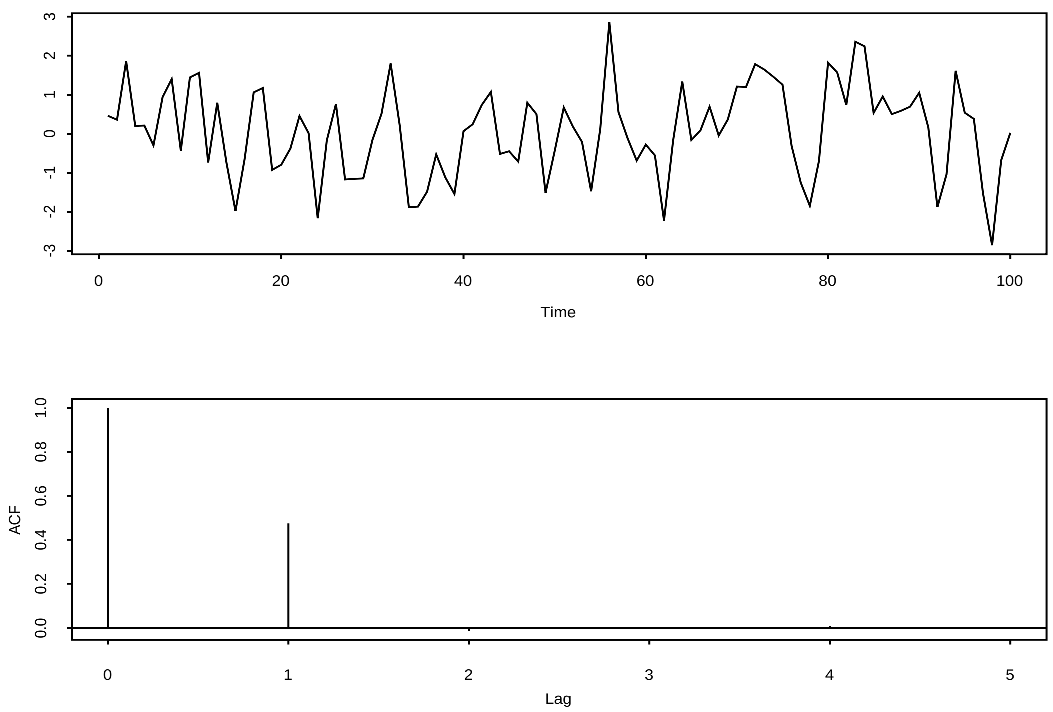
ie

$$\phi(B) X_t = h(B)\epsilon_t$$

with $\phi(x) = 1 - \alpha_1 x - \ldots - \alpha_p x^p$ and $h(x) = 1 + \beta_1 x + \ldots + \beta_q x^q$.

Notes:

Figure 3.3: Realization and ACF of the MA(1) model $X_t = \epsilon_t + 0.8\epsilon_{t-1}$



(a) We get a stationary MA(∞) representation $X_t = (\phi(B))^{-1}h(B)\epsilon_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ iff all roots of $\phi(x) = 0$ lie outside the unit circle. Extending earlier terminology, the process in this case is called *causal*.
We get a stationary AR representation $\epsilon_t = (h(B))^{-1}\phi(B)X_t = \sum_{j=0}^{\infty} \delta_j X_{t-j}$ iff all roots of $h(x) = 0$ lie outside the unit circle. Again extending earlier terminology, the process in this case is said to be *invertible*.

(b) ACF. As in §3.4.2 we can get Yule-Walker equations for the γ_k from the defining relation (3.18) above. Write

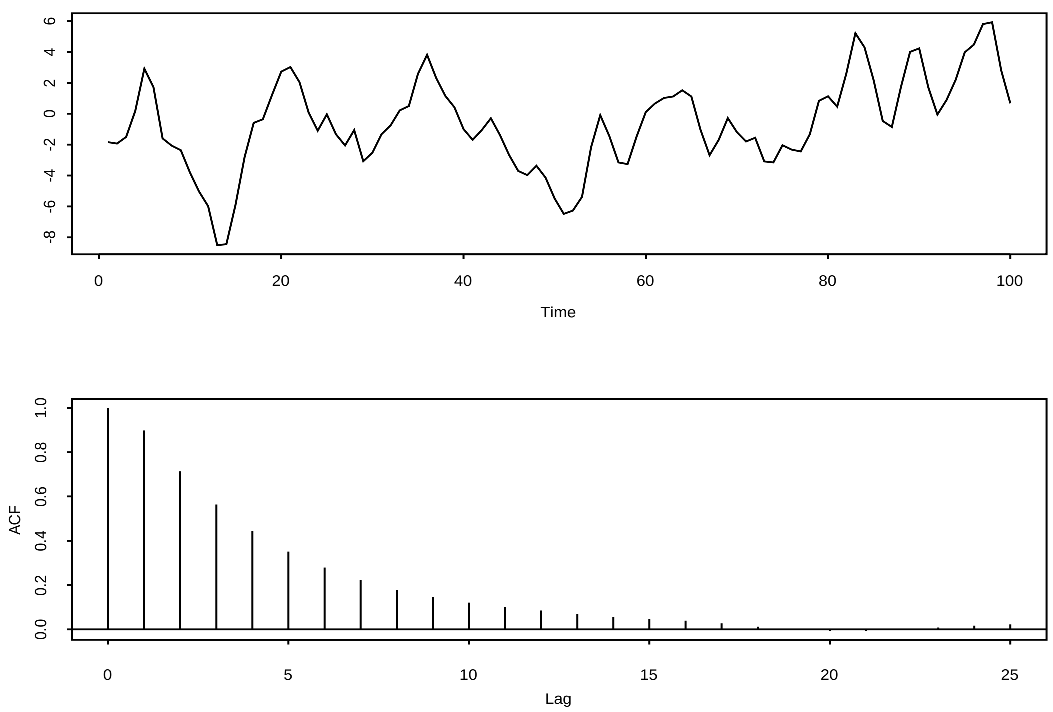
$$X_{t-k}X_t = X_{t-k}(\alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p}) + X_{t-k}(\epsilon_t + \beta_1 \epsilon_{t-1} + \ldots + \beta_q \epsilon_{t-q})$$

and take expectations. For a causal process the X_{t-k} for $k > q$ are uncorrelated with $\epsilon_t, \ldots, \epsilon_{t-q}$, so we get the same equations for ρ_k when $k > q$ as in §3.4.2. (For $k \leq q$ however they will depend on the β_is too.) This implies that the acf of the ARMA(p, q) for large lags decays like that of an AR(p) process.

(c) It turns out that the ARMA processes can approximate a wide range of behaviour using only a small number of parameters, so they provide attractive models for observed time series.

Example: An ARMA(1,1) process. Figure 3.4 shows a realization of 100 values of the ARMA(1,1) process $x_t = 0.8X_{t-1} + \epsilon_t + 0.9\epsilon_{t-1}$, *and the process ACF.*

Figure 3.4: Realization and ACF of the ARMA(1,1) model $X_t = 0.8X_{t-1} + \epsilon_t + 0.9\epsilon_{t-1}$.



Review Exercise: Explain why for a causal ARMA(1,1) process, $X_s, s < t$ are uncorrelated with $\epsilon_t, \epsilon_{t+1}, \ldots$.

Appendix: Difference Equations

We begin by considering a simple second-order difference equation (or *recurrence relation* for an unknown sequence $(y_n, n = 0, 1, 2, \ldots)$,

$$ay_n + by_{n-1} + cy_{n-2} = 0. \tag{3.19}$$

Here a, b and c are given real constants.

We look for solutions of the form $y_n = r^{-n}$ for some $r > 0$. Substituting this in (3.19), we get

$$ar^{-n+2} + br^{-n+1} + cr^{-n} = 0,$$

and on dividing through by r^{-n}, we obtain the *auxiliary equation*

$$ar^2 + br + c = 0. \tag{3.20}$$

This is a quadratic equation, which will in general have two solutions r_1 and r_2 (which may coincide). It can be shown rigorously that the general solution of (3.19), at least in the case where $r_1 \neq r_2$ is

$$y_n = Ar_1^{-n} + Br_2^{-n},$$

where A and B are arbitrary constants. A and B can be explicitly computed if we are given some additional information, e.g. numerical values of y_0 and y_1.

This procedure generalises to the case where the shift is from n to $n-p$ with $p > 2$, as in the Yule-Walker equations:-

$$\rho_k + \alpha_1\rho_{k-1} + \alpha_2\rho_{k-2} + \cdots + \alpha_p\rho_{k-p}. \tag{3.21}$$

Try a solution of the form $\rho_k = \theta^{-k}$, to obtain the auxiliary equation

$$\alpha_p\theta^p + \alpha_{p-1}\theta^{p-1} + \cdots + \alpha_1\theta + 1 = 0. \tag{3.22}$$

If (3.22) has p distinct solutions $\theta_1, \theta_2, \ldots, \theta_p$, the unique solution to (3.21) is

$$\rho_k = \sum_{j=1}^{p} \xi_j\theta_j^{-k},$$

where $\xi_1, \xi_2, \ldots, \xi_p$ are arbitrary constants.

Example Find the general solution of

$$y_{n-3} - 2y_{n-2} - y_{n-1} + 2y_n = 0$$

Find also the particular solution for which $y_0 = 0, y_1 = 1$ and $y_2 = -1$.

Solution Auxiliary equation is $r^3 - 2r^2 - r + 2 = 0$.

By inspection $r = 1$ is a solution. Factorise

$$\begin{aligned} r^3 - 2r^2 - r + 2 &= (r-1)(r^2 - r - 2) \\ &= (r-1)(r+1)(r-2), \end{aligned}$$

so solutions are $r = 1, -1, 2$. Hence the general solution of the difference equation is

$$y_n = A + B(-1)^n + C2^{-n},$$

where A, B and C are arbitrary constants. Substituting in the data, we obtain the three simultaneous equations:

$$A + B + C = 0 \ldots \text{(i)}$$

$$A - B + \frac{C}{2} = 1 \ldots \text{(ii)}$$

$$A + B + \frac{C}{4} = -1 \ldots \text{(iii)}$$

Hence $A = -1/2, B = -5/6$ and $C = 4/3$. So

$$y_n = -\frac{1}{2} - \frac{5}{6}(-1)^n + \frac{4}{3}2^{-n}.$$

Summary of Ideas in Chapter 3

AR(p) :

- $\phi(B)X_t = \epsilon_t$
- ACF decays to zero at large lags approximately geometrically
- causal iff $\phi(x) = 0$ has all roots outside unit circle

MA(q) :

- $X_t = h(B)\epsilon_t$
- ACF identically zero beyond q, where q = order of polynomial h
- invertible iff $h(x) = 0$ has all roots outside unit circle

ARMA(p, q) :

- $\phi(B)X_t = h(B)\epsilon_t$
- ACF similar to that of AR
- causal, invertible iff $\phi(x) = 0$ and $h(x) = 0$ have all roots outside unit circle
- flexible models