STAT 153 & 248 - Time Series
Lecture Twenty

Fall 2025, UC Berkeley

Aditya Guntuboyina

November 06, 2025

In the last lecture, we introduced the notion of stationary time series models, and also started discussing stationarity of AutoRegressive models.

A time series model $\{y_t\}$ is said to be stationary if:

1. $\mathbb{E}y_t$ does not change with t
2. $\text{var}(y_t)$ does not change with t
3. $\text{cov}(y_t, y_{t+h})$ does not change with t for every h.

For a stationary time series, the AutoCovariance Function (ACVF) is defined as

$$\gamma(h) := \text{cov}(y_t, y_{t+h}).$$

By stationarity,

$$\gamma(-h) = -\text{cov}(y_t, y_{t-h}) = \text{cov}(y_{t-h}, y_t) = \text{cov}(y_{t-h}, y_{t-h+h}) = \gamma(h).$$

Note also that $\gamma(0)$ equals the variance of y_t. The Autocorrelation Function (ACF) is given by:

$$\rho(h) = \text{corr}(y_t, y_{t+h}) = \frac{\text{cov}(y_t, y_{t+h})}{\sqrt{\text{var}(y_t)\text{var}(y_{t+h})}} = \frac{\gamma(h)}{\sqrt{\gamma(0) \times \gamma(0)}} = \frac{\gamma(h)}{\gamma(0)}.$$

An important class of stationary time series models are the Moving Average Models.

1 Moving Average (MA) Models

Given a positive integer $q \geq 1$, the Moving Average model with order q (denoted by $\text{MA}(q)$) is defined by the equation:

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \tag{1}$$

where $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. The $\text{MA}(q)$ model has $q+2$ unknown parameters which are estimated from observed data: $\mu, \theta_1, \dots, \theta_q, \sigma$.

The $\text{MA}(q)$ model has been called the "Summation of Random Causes" by its inventor Slutzky in the original paper titled "The summation of random causes as the source of cyclic processes" published in Econometrica in 1937. Basically the ϵ_t's can be treated as random causes which are assumed to be independently and identically distributed. The actual observations y_t's are consequences of these causes. The consequence for tine t depends on the cause for time t as well as the causes for times $t-1, \dots, t-q$. These different causes affect the consequence at time t differently depending on the values of $\theta_1, \dots, \theta_q$. Note that successive observations y_t share some common causes leading to dependence between the successive values of y_t.

The simplest of these MA(q) models is MA(1) (i.e., $q = 1$):

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1}.$$

It is easy to check that each MA(q) model is stationary. Here is the proof for MA(1) (the proof for stationarity of MA(q) for $q \geq 1$ is left as exercise). The mean of y_t is clearly $\mathbb{E} y_t = \mu$ which does not change with t. The variance of y_t is

$$\text{var}(y_t) = \text{var}(\mu + \epsilon_t + \theta_1 \epsilon_{t-1}) = \sigma^2 + \theta_1^2 \sigma^2$$

which also does not change with t. The covariance between y_t and y_{t+1} is

$$\text{cov}(y_t, y_{t+1}) = \text{cov}(\mu + \epsilon_t + \theta_1 \epsilon_{t-1}, \mu + \epsilon_{t+1} + \theta_1 \epsilon_t) = \text{cov}(\epsilon_t, \theta_1 \epsilon_t) = \theta_1 \sigma^2$$

which does not depend on t. The covariance between y_t and y_{t+2} is

$$\text{cov}(y_t, y_{t+2}) = \text{cov}(\mu + \epsilon_t + \theta_1 \epsilon_{t-1}, \mu + \epsilon_{t+2} + \theta_1 \epsilon_{t+1}) = 0$$

Similarly, it is easy to see that the covariance between y_t and y_{t+h} equals zero for every $h \geq 2$. The ACVF of MA(1) is therefore

$$\gamma(h) = \begin{cases} \sigma^2(1 + \theta_1^2) & : h = 0 \\ \sigma^2 \theta_1 & : |h| = 1 \\ 0 & : |h| > 1 \end{cases}$$

The ACF of MA(1) is:

$$\rho(h) = \begin{cases} 1 & : h = 0 \\ \frac{\theta_1}{1+\theta_1^2} & : h = 1 \\ 0 & : h > 1 \end{cases}$$

2 Stationarity of AR(1)

In the last lecture, we discussed whether the AR(1) model is stationary. The AR(1) difference equation is $y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$. This is an implicit equation because y appears in both sides of the equation. To be able to calculate mean, variance and covariance etc., we need a more explicit formula for y_t.

One way to get such an explicit formula is to apply the difference equation recursively on the right hand side as:

$$\begin{aligned} y_t &= \phi_0 + \phi_1 y_{t-1} + \epsilon_t \\ &= \phi_0 + \phi_1 \left(\phi_0 + \phi_1 y_{t-2} + \epsilon_{t-1}\right) + \epsilon_t \\ &= \phi_0(1 + \phi_1) + \phi_1^2 y_{t-2} + \epsilon_t + \phi_1 \epsilon_{t-1} \\ &= \phi_0(1 + \phi_1) + \phi_1^2 \left(\phi_0 + \phi_1 y_{t-3} + \epsilon_{t-2}\right) + \epsilon_t + \phi_1 \epsilon_{t-1} \\ &= \phi_0(1 + \phi_1 + \phi_1^2) + \phi_1^3 y_{t-3} + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2}. \end{aligned}$$

We continue in this way writing $y_{t-3} = \phi_0 + \phi_1 y_{t-4} + \epsilon_{t-3}$ and then $y_{t-4} = \phi_0 + \phi_1 y_{t-5} + \epsilon_{t-4}$ and so on. This leads to

$$y_t = \phi_0 \sum_{j=0}^{M} \phi_1^j + \phi_1^{M+1} y_{t-M-1} + \sum_{j=0}^{M} \phi_1^j \epsilon_{t-j}. \tag{2}$$

This formula is true for every value of $M \geq 0$ and every ϕ_0, ϕ_1. The right hand side of (2) still depends on a y-value (specifically y_{t-M-1}).

Suppose now that $|\phi_1| < 1$. Then ϕ_1^{M+1} decays rapidly to zero. In this case, we can let $M \to \infty$ in (2) and use

$$\begin{aligned}
&\phi_0 \sum_{j=0}^{M} \phi_1^j \to \phi_0 \sum_{j=0}^{\infty} \phi_1^j = \frac{\phi_0}{1-\phi_1} \qquad \text{as } M \to \infty \\
&\phi_1^{M+1} y_{t-M-1} \to 0 \qquad \text{as } M \to \infty \\
&\sum_{j=0}^{M} \phi_1^j \epsilon_{t-j} \to \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j} \qquad \text{as } M \to \infty
\end{aligned}$$

Thus taking $M \to \infty$ in (2) leads to

$$y_t = \frac{\phi_0}{1-\phi_1} + \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}. \tag{3}$$

This expression is well-defined when $|\phi_1| < 1$. It is easy to check that y_t defined as (3) satisfies the AR(1) equation $y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$. In the last lecture, we saw that (3) is a stationary time series model with:

$$\mathbb{E} y_t = \frac{\phi_0}{1-\phi_1} \quad \text{and} \quad \text{cov}(y_t, y_{t+h}) = \sigma^2 \frac{\phi_1^{|h|}}{1-\phi_1^2}.$$

In the formula (3), it is clear that y_t is determined by $\epsilon_t, \epsilon_{t-1}, \dots$ i.e., by the present and past values of ϵ_t. For this reason, (3) is called **Causal, Stationary AR(1)** (causal here roughly means that y_t is fully determined by present and past values of $\{\epsilon_t\}$). One consequence of this kind of causality is that ϵ_t is independent of the past y-values $y_{t-1}, y_{t-2}, y_{t-3} \dots$ (this is because these past y-values only depend on $\epsilon_{t-1}, \epsilon_{t-2}, \dots$ which are all independent of ϵ_t).

When $|\phi_1| \geq 1$, the terms in the right hand side of (2) do not converge when $M \to \infty$. In this case, it is not possible to get a stationary solution y_t of the AR equation which depends only on present and past ϵ-values $\epsilon_t, \epsilon_{t-1}, \dots$. In other words, there is no causal stationary solution to the AR(1) equation when $|\phi_1| \geq 1$.

When $|\phi_1| > 1$ (note the strict inequality), there is a non-causal stationary solution for the AR(1) equation. This is obtained by rewriting the AR(1) equation as:

$$y_t = -\frac{\phi_0}{\phi_1} + \frac{1}{\phi_1} y_{t+1} - \frac{\epsilon_{t+1}}{\phi_1}.$$

and then recursively plugging in $y_{t+1}, y_{t+2}, \ldots$ as:

$$
\begin{aligned}
y_t &= -\frac{\phi_0}{\phi_1} + \frac{1}{\phi_1}\left(-\frac{\phi_0}{\phi_1} + \frac{1}{\phi_1} y_{t+2} - \frac{\epsilon_{t+2}}{\phi_1}\right) - \frac{\epsilon_{t+1}}{\phi_1} \\
&= -\frac{\phi_0}{\phi_1}\left(1 + \frac{1}{\phi_1}\right) + \frac{1}{\phi_1^2} y_{t+2} - \frac{\epsilon_{t+1}}{\phi_1} - \frac{\epsilon_{t+2}}{\phi_1^2} \\
&= -\frac{\phi_0}{\phi_1}\left(1 + \frac{1}{\phi_1} + \frac{1}{\phi_1^2}\right) + \frac{1}{\phi_1^3} y_{t+3} - \frac{\epsilon_{t+1}}{\phi_1} - \frac{\epsilon_{t+2}}{\phi_1^2} - \frac{\epsilon_{t+3}}{\phi_1^3} \\
&= \ldots \\
&= -\frac{\phi_0}{\phi_1}\left(1 + \frac{1}{\phi_1} + \frac{1}{\phi_1^2} + \cdots + \frac{1}{\phi_1^M}\right) + \frac{y_{t+M+1}}{\phi_1^{M+1}} - \frac{\epsilon_{t+1}}{\phi_1} - \frac{\epsilon_{t+2}}{\phi_1^2} - \frac{\epsilon_{t+3}}{\phi_1^3} - \cdots - \frac{\epsilon_{t+M+1}}{\phi_1^{M+1}} \quad (4)
\end{aligned}
$$

Letting $M \to \infty$ and using $|\phi_1| > 1$ (so that $|1/\phi_1| < 1$), we get

$$
y_t = \frac{\phi_0}{1 - \phi_1} - \sum_{j=1}^{\infty} \frac{\epsilon_{t+j}}{\phi_1^j} \quad (5)
$$

It can be checked that (5) is also stationary (note now that $|\phi_1| > 1$) and satisfies the AR(1) equation $y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$. This AR(1) process is called **non-causal** because y_t depends on future values $\epsilon_{t+1}, \epsilon_{t+2}, \ldots$. Here ϵ_t is **not** independent of the past y-values $y_{t-1}, y_{t-2}, \ldots$.

If $|\phi_1| = 1$ (i.e., if $\phi_1 = 1$ or $\phi_1 = -1$), then neither (2) nor (4) converge as $M \to \infty$. In fact, in this case ($|\phi_1| = 1$), there cannot be a stationary solution to the AR(1) equation.

2.1 Practical Implications

In practice, while fitting the AR(1) model, we use the `AutoReg` function (from the `statsmodels` library) which works with the likelihood:

$$
\prod_{t=2}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_t - \phi_0 - \phi_1 y_{t-1})^2}{2\sigma^2}\right) \quad (6)
$$

This likelihood fixes the value of y_1 and assumes that each ϵ_t is independent of the past values $y_{t-1}, y_{t-2}, \ldots$. In other words, the model considered is:

$$
y_1 = \text{fixed at observed value} \quad \text{and} \quad y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t \quad (7)
$$

where $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$ and independent of $y_{t-1}, \ldots, y_1$ for each $t = 2, \ldots, n$. The parameters ϕ_0, ϕ_1, σ are estimated in the usual way (as is done in linear regression). For predictions, one works with the fitted model where the parameters are replaced by their estimates $\hat{\phi}_0, \hat{\phi}_1$ (and also $\hat{\sigma}$).

Based on the value of $\hat{\phi}_1$, the following things can happen:

1. **Case One:** $|\hat{\phi}_1| > 1$. This happens often when we fit the AR(1) model directly to economic data (such as GDP or GNP). Here the fitted model (7) (with $\phi_0 = \hat{\phi}_0$ and $\phi_1 = \hat{\phi}_1$) is very different from the non-causal stationary AR(1) model given by (5). Note again that for (5), ϵ_t becomes dependent on the past values $y_{t-1}, y_{t-2}, \ldots$. The fitted model (7) is quite far from being stationary and future predictions will often exhibit explosive behavior when the prediction horizon increases.

2. **Case Two:** $|\hat{\phi}_1| < 1$. This also happens very often. For economic data, this often happens after some preprocessing (e.g., by taking logarithms of the raw data and then differences once or twice). In this case, the fitted model (7) is also non-stationary and distinct from the causal stationary model (3). However, the discrepancy is minimal and they behave similarly in many ways:

   a) The likelihood for (3) is

   $$\frac{\sqrt{1-\phi_1^2}}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1-\phi_1^2}{2\sigma^2}\left(y_1-\frac{\phi_0}{1-\phi_1}\right)^2\right)\prod_{t=2}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y_t-\phi_0-\phi_1 y_{t-1})^2}{2\sigma^2}\right) \tag{8}$$

   We saw this form of the likelihood in Lecture 17. The only difference between this likelihood and the likelihood (6) is the presence of the term involving y_1. When n is large, this single term (which only involves the first data point) does not affect the overall likelihood significantly, so both likelihood maximizations lead to similar answers.

   b) Recursing the difference equation in (7) from $t, \dots, 2$, we obtain (basically taking $M = t - 2$ in (2))

   $$y_t = \phi_0 \sum_{j=0}^{t-2} \phi_1^j + \phi_1^{t-1} y_1 + \sum_{j=0}^{t-2} \phi_1^j \epsilon_{t-j} \qquad \text{for every } t = 2, \dots, n.$$

   Because $|\phi_1| < 1$ (so that $|\phi_1|^j$ decreases rapidly), we can expect the above to be close to (3) as long as t is not very small.

   c) Future predictions for the models (7) and (3) work in identical fashion (for fixed identical parameter values). These predictions only use independence of ϵ_t and past y-values $y_{t-1}, y_{t-2}, \dots$ which is true in both (7) and (3).

   Because of these reasons, when $|\hat{\phi}_1| < 1$, even though `AutoReg` is fitting (7), it is common to pretend as if we are working with the causal stationary AR(1) model (3).

3. **Case Three:** $|\hat{\phi}_1| = 1$. This is either $\hat{\phi}_1 = 1$ or $\hat{\phi}_1 = -1$. These models are all non-stationary. The case $\hat{\phi}_1 = -1$ almost never arises in practice (unless the data has a strange wild oscillatory pattern from each time point to the next). The case $\hat{\phi}_1 = 1$ is much more applicable. `AutoReg` may not give $\hat{\phi}_1 = 1$ exactly but it might give $\hat{\phi}_1$ that is close to 1. Note that when $\phi_1 = 1$, the AR(1) model equation can be rewritten as $y_t - y_{t-1} = \phi_0 + \epsilon_t$; this suggests preprocessing the data by taking successive differences $y_t - y_{t-1}$ and trying to fit models to this differenced data. AR(1) predictions with $\hat{\phi}_1 = 1$ grow linearly which may be well-suited for many datasets.

Thus from the practical perspective where we only consider causal models, stationary only corresponds to $|\phi_1| < 1$. (When $|\phi_1| > 1$, mathematically speaking, there is a stationary AR(1) model but this is non-causal and does not correspond to our likelihoods and fitting technique).

Next we shall look at AR(p) models for $p \geq 2$ and discuss the analogue of the causal-stationarity condition $|\phi_1|$ when $p \geq 2$.

Before we do that however, let us first go over an alternative method of deriving the causal stationary AR(1) model formula (3) using a formal technique involving Backshift notation.

Before describing this technique, we need to introduce the backshift notation.

3 Backshift Notation

A convenient piece of notation used while working with AR and MA models is the Backshift notation. Let B denote the *backshift operator* defined by

$$By_t = y_{t-1}, B^2 y_t = y_{t-2}, B^3 y_t = y_{t-3}, \dots$$

and similarly

$$B\epsilon_t = \epsilon_{t-1}, B^2\epsilon_t = \epsilon_{t-2}, B^3\epsilon_t = \epsilon_{t-3}, \dots.$$

Also let I denote the identity operator: $Iy_t = y_t$. More generally, we can define polynomial functions of the Backshift operator by, for example,

$$(I + B + 3B^2)y_t = Iy_t + By_t + 3B^2 y_t = y_t + y_{t-1} + 3y_{t-2}.$$

In general, for every polynomial $f(z)$, we can define $f(B)$. One can even extend this notation to negative powers of B which correspond to forward shifts. For example, $B^{-1}y_t = y_{t+1}, B^{-5}y_t = y_{t+5}$ and $(B^3 + 9B^{-2})y_t = y_{t-3} + 9y_{t+2}$ etc.

In this notation, the defining equation $y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$ for the $AR(p)$ model can be written as $\phi(B)y_t = \phi_0 + \epsilon_t$ for the polynomial $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$.

The defining equation $y_t = \epsilon_t + \theta\epsilon_{t-1}$ for the MA(1) model can be written as $y_t = \theta(B)\epsilon_t$ for the polynomial $\theta(z) = 1 + \theta_1 z$.

The defining equation $y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \dots + \theta_q\epsilon_{t-q}$ for the MA(q) model becomes $y_t = \theta(B)\epsilon_t$ for the polynomial $\theta(z) = 1 + \theta_1 z + \dots \theta_q z^q$.

4 Causal Stationary $AR(1)$ formula using Backshift

We derived the formula (3) by recursing the AR(1) equation $y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$ successively into the past (as in (2)) and taking the limit $M \to \infty$. This method is difficult to carry out for AR(p) when $p \geq 2$. Instead there is an alternative method (using Backshift) of directly arriving at (3) from $y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$. This alternative method is very easy to generalize to higher p.

Here is the description of the backshift method for AR(1). We will tackle higher p in the next section. First note that AR(1) difference equation in backshift notation is

$$\phi(B)y_t = \phi_0 + \epsilon_t \qquad \text{where } \phi(z) = 1 - \phi_1 z.$$

Thus we can formally write

$$y_t = \frac{1}{\phi(B)}(\phi_0 + \epsilon_t).$$

Using

$$\frac{1}{\phi(z)} = \frac{1}{1 - \phi_1 z} = 1 + \phi_1 z + \phi_1^2 z^2 + \phi_1^3 z^3 + \dots, \tag{9}$$

we obtain

$$\begin{aligned} y_t &= \left(I + \phi_1 B + \phi_1^2 B^2 + \dots\right)(\phi_0 + \epsilon_t) \\ &= \left(I + \phi_1 B + \phi_1^2 B^2 + \dots\right)\phi_0 + \left(I + \phi_1 B + \phi_1^2 B^2 + \dots\right)\epsilon_t \\ &= \left(1 + \phi_1 + \phi_1^2 + \dots\right)\phi_0 + \sum_{j=0}^{\infty}\phi_1^j \epsilon_{t-j} = \frac{\phi_0}{1 - \phi_1} + \sum_{j=0}^{\infty}\phi_1^j \epsilon_{t-j} \end{aligned}$$

which gives (3). This formal method is sometimes called Backshift Calculus and it works for higher order AR models as well.

5 AR(p) for $p \geq 1$

Now consider the AR(p) difference equationl In backshift notation, this is

$$\phi(B)y_t = \phi_0 + \epsilon_t \qquad \text{where } \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p.$$

We shall call $\phi(z) := 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ the AR polynomial.

We "solve" the AR(p) by writing:

$$y_t = \frac{1}{\phi(B)}\epsilon_t = \frac{1}{1 - \phi_1 B - \cdots - \phi_p B^p}\epsilon_t$$

The next step is to make sense of $1/(1 - \phi_1 B - \cdots - \phi_p B^p)$. It is natural here to factorize the polynomial $1 - \phi_1 B - \cdots - \phi_p B^p$ into monomials, and then use (9). So we write

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = (1 - a_1 z) \ldots (1 - a_p z). \tag{10}$$

so that

$$\phi(B) = (1 - a_1 B) \ldots (1 - a_p B).$$

The numbers $a_1, \ldots, a_p$ appearing in (10) are simply the reciprocals of the roots of $\phi(z)$ i.e., the roots of $\phi(z)$ are given by $1/a_1, \ldots, 1/a_p$. Note here that some of the a_j's can be complex because the polynomial $1 - \phi_1 z - \cdots - \phi_p z^p$ can have complex roots (even though all its coefficients are real).

We then get

$$y_t = \frac{1}{(1 - a_1 B) \ldots (1 - a_p B)}(\phi_0 + \epsilon_t) = \prod_{k=1}^{p} \frac{1}{1 - a_k B}(\phi_0 + \epsilon_t).$$

For each $1/(1 - a_j B)$, we use the formula (9) to get:

$$y_t = \prod_{k=1}^{p} \left(\sum_{j=0}^{\infty} a_k^j B^j \right) (\phi_0 + \epsilon_t). \tag{11}$$

Multiplying out the product $\prod_{k=1}^{p} \left(\sum_{j=0}^{\infty} a_k^j B^j \right)$, we get

$$\begin{aligned}
y_t &= \prod_{k} \left(\sum_{j=0}^{\infty} a_k^j B^j \right) (\phi_0 + \epsilon_t) \\
&= \left(\sum_{j_1=0}^{\infty} a_1^{j_1} B^{j_1} \right) \cdots \left(\sum_{j_p=0}^{\infty} a_p^{j_p} B^{j_p} \right) (\phi_0 + \epsilon_t) \\
&= \left(\sum_{j_1=0}^{\infty} \cdots \sum_{j_p=0}^{\infty} a_1^{j_1} \ldots a_p^{j_p} B^{j_1 + \cdots + j_p} \right) (\phi_0 + \epsilon_t) \\
&= \phi_0 \sum_{j_1=0}^{\infty} \cdots \sum_{j_p=0}^{\infty} a_1^{j_1} \ldots a_p^{j_p} + \sum_{j_1=0}^{\infty} \cdots \sum_{j_p=0}^{\infty} a_1^{j_1} \ldots a_p^{j_p} \epsilon_{t - j_1 - \cdots - j_p}.
\end{aligned} \tag{12}$$

The above expression involves powers of $a_1, \dots, a_p$. For these powers to not explode, we need

$$|a_i| < 1 \qquad \text{for each } i = 1, \dots, p. \tag{13}$$

Note that a_i can be complex so $|a_i|$ represents the modulus of a_j. When $|a_i| < 1$, the powers $|a_i|^j$ decay rapidly in j which makes the infinite sums above well-defined.

The formula above writes y_t in terms of $\epsilon_t, \epsilon_{t-1}, \dots$. By collecting terms where $j_1 + \cdots + j_p = j$ for each $j = 0, 1, \dots$, we can write this solution as

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \tag{14}$$

for some $\mu, \psi_1, \psi_2, \dots$. It can be checked that this is a stationary time series (note it is also causal as y_t only depends on $\epsilon_t, \epsilon_{t-1}, \dots$).

The key condition here is (13). This is the analogue of the AR(1) condition $|\phi_1| < 1$ for AR(p) when $p \geq 1$. Because the roots of the AR polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ are $1/a_1, \dots, 1/a_p$, the condition (13) is equivalent to assuming that all roots of the AR polynomial are strictly larger than 1 in modulus.

This analysis can be made rigorous to show the following:

1. When all roots of the AR polynomial are strictly larger than 1 in modulus, then there exists a unique causal stationary process $\{y_t\}$ which satisfies the AR(p) equation: $y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t$. This causal stationary solution is of the form (14) where the coefficients $\psi_1, \psi_2, \dots$ are derived from (12).

2. When even one root of the AR polynomial has modulus ≤ 1, then there cannot exist a causal stationary solution to the AR equation.

3. `AutoReg` fits the model

$$y_1 = \text{fixed at observed value} \quad \text{and} \quad y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

with $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. When the condition (13) is true, this model has similar behavior to (14) in the same way as the similarity between (3) and (7) for $p = 1$.

For assessing causal-stationarity of the fitted model, the output summary of `AutoReg` gives the values of the moduli of the roots of the fitted AR polynomial.

6 Determination of the order p of AR(p)

How to determine the correct order p for fitting the AR(p) model? For this, one commonly uses a quantity called the sample PACF (PACF stands for Partial AutoCorrelation Function).

The sample PACF is defined as follows: for $h \geq 1$,

$$\text{sample PACF}(h) = \text{estimate } \hat{\phi}_h \text{ of } \phi_h \text{ when AR}(h) \text{ is fit to the data}$$

If the sample PACF(h) becomes negligibly small after a particular p, this suggests that AR(p) is a good model for the data. This method is similar to the heuristic technique that we used in Lab 9 for selecting the order p to fit AR(p). There we were looking at the uncertainty interval for ϕ_p to see if it contains zero when AR(p) is fit to the data. This is the same as checking whether the sample PACF(h) is small at $h = p$.

It can happen (we will see examples of this later) that the sample PACF(h) for $h = 1, 2, \dots, 11$ are all negligible but at $h = 12$, it is nonnegligible. In that case, we would be using AR(12). More specifically, we will use that value of p for which sample PACF(h) is negligible for all $h > p$.

Why should the quantity $\hat{\phi}_h$ (obtained by fitting AR(h) to the data) be called the Sample Partial Autocorrelation? We will understand this in the next lecture.