Lecture 6: Autoregressive Integrated Moving Average Models

Introduction to Time Series, Fall 2023

Ryan Tibshirani

Related reading: Chapters 3.1 and 3.3 Shumway and Stoffer (SS); Chapters 9.1–9.5 and 9.8–9.9 of Hyndman and Athanasopoulos (HA).

1 AR models

- The *autoregressive* (AR) model is one of the foundational legs of ARIMA models, which we'll cover bit by bit in this lecture. (Recall, you've already learned about AR models, which were introduced all the way back in our first lecture)
- Precisely, an AR model of order $p \geq 1$, denoted AR(p), is of the form

$$x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + \epsilon_t \tag{1}$$

  where ϵ_t, $t = 0, \pm 1 \pm 2, \pm 3, \dots$ is a white noise sequence. Note that we allow the time index to be negative here (we extend time back to $-\infty$), which will useful in what follows
- The coefficients $\phi_1, \dots, \phi_p$ in (1) are fixed (nonrandom), and we assume $\phi_p \neq 0$ (otherwise the order here would effectively be less than p). Note that in (1), we have $\mathbb{E}(x_t) = 0$ for all t
- If we wanted to allow for a nonzero but constant mean, then we could add an intercept to the model in (1). We'll omit this for simplicity in this lecture
- A useful tool for expressing and working with AR models is the *backshift operator*: this is an operator we denote by B that takes a given time series and shifts it back in time by one index,

$$Bx_t = x_{t-1}$$

- We can extend this to powers, as in $B^2 x_t = BBx_t = x_{t-2}$, and so on, thus

$$B^k x_t = x_{t-k}$$

- Returning to (1), note now that we can rewrite this as

$$x_t - \phi_1 x_{t-1} - \phi_2 - x_{t-2} - \cdots - \phi_p x_{t-p} = \epsilon_t$$

  or in other words, using backshift notation

$$\Big(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\Big) x_t = \epsilon_t \tag{2}$$

- Hence (2) is just a compact way to represent the AR(p) model (1) using the backshift operator B. Often, authors will write this model even more compactly as

$$\phi(B) x_t = \epsilon_t \tag{3}$$

  where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is called the *autoregressive operator* of order p, associated with the coefficients $\phi_1, \dots, \phi_p$
- Figure 1 shows two simple examples of AR processes

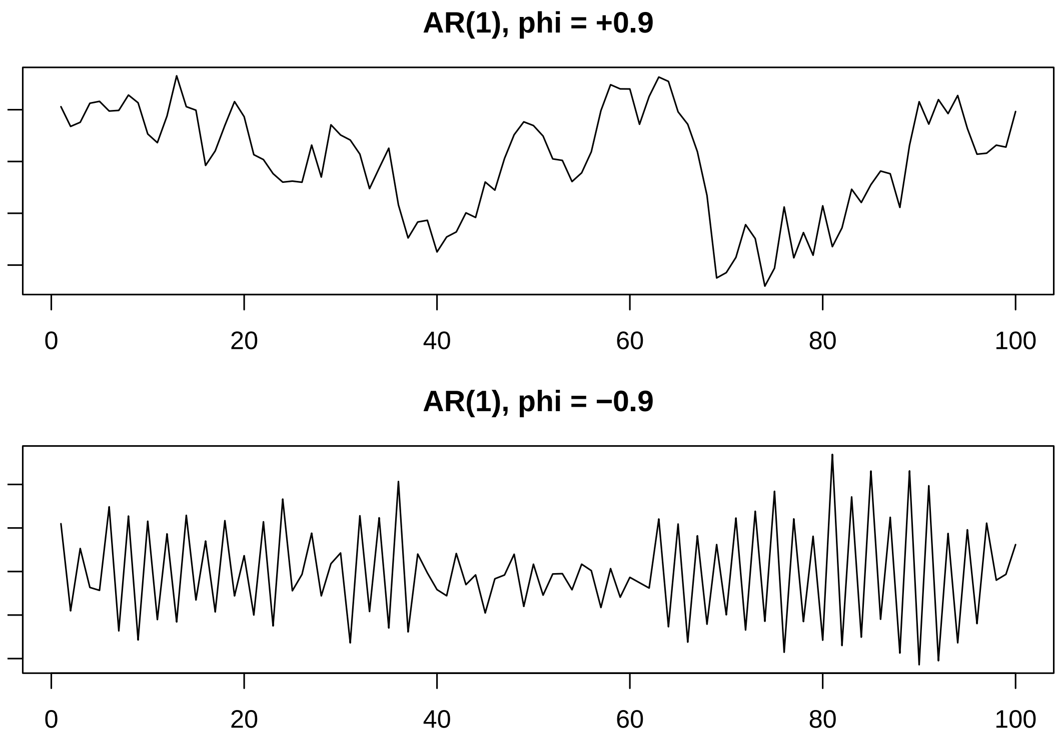

Figure 1: Two examples of AR(1) processes, with $\phi = \pm 0.9$.

1.1 AR(1): auto-covariance and stationarity

- A key question for us will be: *under what conditions does the AR model in* (1)*, or equivalently* (3)*, define a stationary process?*
- The answer will turn out to be fairly sophisticated, but we can glean some intuition by starting with the AR(1) case:

$$x_t = \phi x_{t-1} + \epsilon_t \tag{4}$$

  for $t = 0, \pm 1 \pm 2, \pm 3, \ldots$
- Unraveling the iterations, we get

$$\begin{aligned} x_t &= \phi^2 x_{t-2} + \phi \epsilon_{t-1} + \epsilon_t \\ &= \phi^3 x_{t-3} + \phi^2 \epsilon_{t-2} + \phi \epsilon_{t-1} + \epsilon_t \\ &\vdots \\ &= \phi^k x_{t-k} + \sum_{j=0}^{k} \phi^j \epsilon_{t-j} \end{aligned}$$

- If $|\phi| < 1$, then we can send $k \to \infty$ in the last display to get

$$x_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} \tag{5}$$

  This is called the *stationary representation* of the AR(1) process (4)

- Why is it called this? We can compute the auto-covariance function, writing $\sigma^2 = \mathrm{Var}(\epsilon_t)$ for the noise variance, as

$$\begin{aligned}
\mathrm{Cov}(x_t, x_{t+h}) &= \mathrm{Cov}\left(\sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}, \sum_{\ell=0}^{\infty} \phi^\ell \epsilon_{t+h-\ell}\right) \\
&= \sum_{j,\ell=0}^{\infty} \phi^j \phi^\ell \,\mathrm{Cov}(\epsilon_{t-j}, \epsilon_{t+h-\ell}) \\
&= \sum_{j=0}^{\infty} \phi^j \phi^{j+h} \sigma^2 \\
&= \sigma^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j} \\
&= \sigma^2 \frac{\phi^h}{1-\phi^2}
\end{aligned}$$

  where we used the fact that $\sum_{j=0}^{\infty} b^j = 1/(1-b)$ for $|b| < 1$. Since the auto-covariance in the last line only depends on h, we can see that the AR(1) process is indeed stationary

- To reiterate, the representation (5), and the auto-covariance calculation just given, would have not been possible unless $|\phi| < 1$. This condition is required in order for the AR(1) process to have a stationary representation. We will see later that we can general this come up with a condition that applies to an AR(p), and beyond

1.2 Causality (no, not the usual kind)

- In order to study what conditions on the coefficients render a general AR(p) model (1) stationary, we will introduce a concept called *causality*
- (This is a slightly unfortunate bit of nomenclature that seems to be common in the time series literature, but has really nothing to do with causality used in the broader sense in statistics. We will ... somewhat begrudgingly ... stick with the standard nomenclature in time series here)
- We say that a series x_t, $t = 0, \pm 1, \pm 2, \ldots$ is *causal* provided that it can be written in the form

$$x_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \tag{6}$$

  for a white noise sequence ϵ_t, $t = 0, \pm 1 \pm 2, \pm 3, \ldots$, and coefficients such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$

- You should think of this as a generalization of (5), where we allow for arbitrary coefficients $\psi_0, \psi_1, \psi_2, \ldots$, subject to an absolute summability condition
- It is straightforward to check that causality actually implies stationarity: we can just compute the auto-covariance function in (6), similar to the above calculation:

$$\begin{aligned}
\mathrm{Cov}(x_t, x_{t+h}) &= \mathrm{Cov}\left(\sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \sum_{\ell=0}^{\infty} \psi_\ell \epsilon_{t+h-\ell}\right) \\
&= \sum_{j,\ell=0}^{\infty} \psi_j \psi_\ell \,\mathrm{Cov}(\epsilon_{t-j}, \epsilon_{t+h-\ell}) \\
&= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}
\end{aligned}$$

  The summability condition ensures that these calculations are well-defined and that the last display is finite. Since this only depends on h, we can see that the process is indeed stationary

- Thus, to emphasize, causality actually tells us *more* than stationary: it is stationary "plus" a representation a linear filter of past white noise variates, with summable coefficients
- Note that when $\psi_j = \phi^j$, the summability condition $\sum_{j=0}^{\infty} |\psi_j| < \infty$ is true if and only if $|\phi| < 1$. Hence what we actually proved above for AR(1) was that it is causal if and only if $|\phi| < 1$. And it is this condition—for causality—that we will actually generalize for AR(p) models, and beyond

2 MA models

- A *moving average* (MA) model is dual, in a colloquial sense, to the AR model. Instead of having x_t evolve according to a linear combination of the recent past, the *errors* in the model evolve according to a linear combination of white noise
- Precisely, an MA model of order $q \geq 1$, denoted MA(q), is of the form

$$x_t = \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{7}$$

  where ϵ_t, $t = 0, \pm 1 \pm 2, \pm 3, \ldots$ is a white noise sequence
- The coefficients $\theta_1, \ldots, \theta_q$ in (7) are fixed (nonrandom), and we assume $\theta_q \neq 0$ (otherwise the order here would effectively be less than q). Note that in (7), we have $\mathbb{E}(x_t) = 0$ for all t
- Again, we can rewrite (7), using backshift notation, as

$$x_t = \Big(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q\Big)\epsilon_t \tag{8}$$

- Often, authors will write (8) even more compactly as

$$x_t = \theta(B)\epsilon_t \tag{9}$$

  where $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$ is called the *moving average operator* of order q, associated with the coefficients $\theta_1, \ldots, \theta_q$
- Figure 2 shows two simple examples of MA processes

2.1 Stationarity

- Unlike AR processes, an MA process (7) is stationary *for any values of the parameters* $\theta_1, \ldots, \theta_q$
- To check this, we compute the auto-covariance function using a similar calculation to those we've done before, writing $\theta_0 = 1$ for convenience:

$$\begin{aligned} \mathrm{Cov}(x_t, x_{t+h}) &= \mathrm{Cov}\left(\sum_{j=0}^{q} \theta_j \epsilon_{t-j}, \sum_{\ell=0}^{q} \theta_\ell \epsilon_{t+h-\ell}\right) \\ &= \sum_{j,\ell=0}^{q} \theta_j \theta_\ell \,\mathrm{Cov}(\epsilon_{t-j}, \epsilon_{t+h-\ell}) \\ &= \sigma^2 \sum_{j=0}^{q} \theta_j \theta_{j+h} \end{aligned}$$

  Since this only depends on h, we can see that the process is indeed stationary
- Actually, this (similarity between this and previous calculations) brings us to emphasize the following connection: *an AR(1) model is actually an MA(∞) model*, as we saw in the stationary representation (5). But writing it as an AR(1) process is simpler
- TODO statement about connection in general?

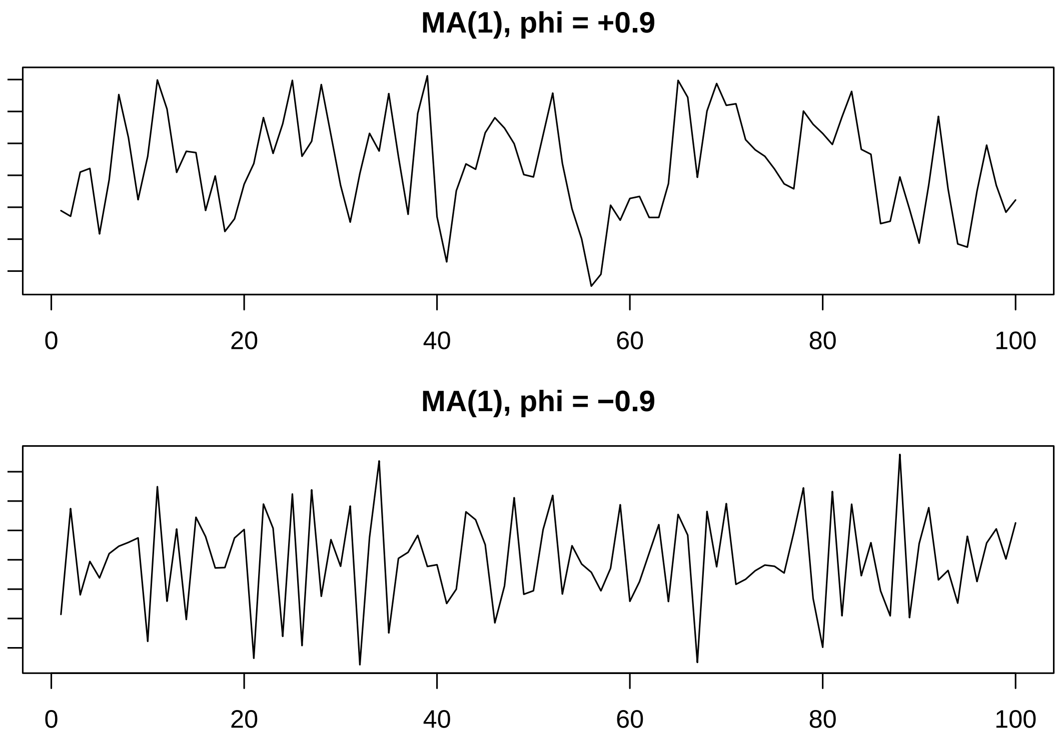

Figure 2: Two examples of MA(1) processes, with $\theta = \pm 0.9$.

2.2 MA(1): non-uniqueness and invertibility

- There are a number of issues with the MA processes that are worth elucidating.

3 ARMA models

ARMA(p,q) Backshift notation

point out special cases: ARMA(0,0): white noise ARMA(1,0): random walk

-

3.1 Parameter redundancy

Parameter redundancy Recap of problems

3.2 Causality and invertibility

Forget about causality entirely

3.3 Difference equations

Diff equations –> don't cover. Auto-covariance

3.4 Partial auto-correlation function

4 ARIMA models

Stationarity and differencing ARIMA(p,d,q) Backshift notation Seasonality extensions

4.1 Parameter estimation

complicated –> don't cover.

4.2 Regression with correlated errors

complicated –> mostly don't cover

5 Forecasting