DS-GA 1018: Lecture 4 Notes

Sebastian Wagner-Carena

September 2024

1 Review – Operators, Invertibility, and Causality

We reviewed the following concepts at the beginning of class:

Definition 1 *The autoregressive operator, $P(B)$, for an AR(p) model is defined as:*

$$P(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p, \tag{1}$$

such that:

$$P(B)X_t = W_t. \tag{2}$$

Definition 2 *The moving average operator for an MA(p) process is defined as:*

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_p B^p, \tag{3}$$

such that:

$$X_t = \Theta(B)W_t. \tag{4}$$

Definition 3 *An Autoregressive Moving Average process of order p,q - ARMA(p,q) is a process with the form:*

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \ldots + \theta_q W_{t-q} \tag{5}$$
$$P(B)X_t = \Theta(B)W_t \tag{6}$$

where W_t is drawn from $\mathcal{N}(0, \sigma_W^2)$, $\phi_1, \ldots, \phi_p$ are constants, $\theta_1, \ldots, \theta_q$ are constants, and both $\theta_q \neq 0$ and $\phi_p \neq 0$. The autoregressive and moving average operators must not share any roots.

Definition 4 *An ARMA(p,q) process is invertible if the time series can be written as:*

$$\pi(B)X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} = W_t \tag{7}$$

with the infinite sum $\sum_{j=0}^{\infty} |\pi_j| < \infty$. We can determine $\pi(B)$ as:

$$\pi(B) = \frac{P(B)}{\Theta(B)}. \tag{8}$$

The conditions for invertability hold so long as the roots of $\Theta(z)$ lie outside the unit circle. See Shumway and Stoffer for the proof of this statement.

By our definition, an AR process of finite order will always be invertible.

Definition 5 *An ARMA(p,q) process is causal if the time series can be written as:*

$$X_t = \psi(B)W_t = \sum_{j=0}^{\infty} \psi_j W_{t-j}, \tag{9}$$

with the infinite sum $\sum_{j=0}^{\infty} |\psi_j| < \infty$. We can determine $\psi(B)$ as:

$$\psi(B) = \frac{\Theta(B)}{P(B)}. \tag{10}$$

The conditions for causality hold so long as the roots of $P(z)$ lie outside the unit circle. Again, see Shumway and Stoffer for the proof of this statement.

By our definition, an MA process of finite order will always be causal.

In the last lecture, we gave the definition of invertibility and causality, but did not develop the intuition for these definitions. The full proof can be found ins Appendix B of Shumway and Stoffer, but let's go through half of it as a means of building understanding. Let's start from the assumption that was have a causal process:

$$X_t = \psi(B)W_t = \sum_{j=0}^{\infty} \psi_j W_{t-j}, \tag{11}$$

with the condition that $\sum_{j=0}^{\infty} |\psi_j| < \infty$. We can then multiply both sides by our autoregressive operator:

$$P(B)X_t = P(B)\psi(B)W_t. \tag{12}$$

Our model is ARMA, so we can also write:

$$P(B)X_t = \Theta(B)W_t. \tag{13}$$

Putting those two facts we can state:

$$P(B)\psi(B)W_t = \Theta(B)W_t. \tag{14}$$

Now we can go from our operators to the complex numbers. Let's start by substituting in complex numbers to write:

$$a(z) = P(z)\psi(z) \tag{15}$$

$$= \sum_{j=0}^{\infty} a_j z^j, \quad |z| \leq 1 \tag{16}$$

we have taken advantage of the fact the product of two infinite power series can be written as an infinite power series. The radius of convergence we have enforced, $|z| \leq 1$, may not be immediately clear. To understand how we got it, note that the radius of convergence of the original $\psi(z)$ power series is guaranteed to be at least 1:

$$\sum_{j=0}^{\infty} \psi_j z^j \leq \sum_{j=0}^{\infty} |\psi_j||z|^j \tag{17}$$

$$\leq \sum_{j=0}^{\infty} |\psi_j|, \quad |z| \leq 1 \tag{18}$$

$$< \infty, \quad |z| \leq 1. \tag{19}$$

Where the last line comes from our definition of causality. The product of two power series has a radius of convergence that is the minimum of the two previous radii of convergence. $P(z)$ is a finite power series by construction, so it has an infinite radius of convergence. Therefore we can assert that Equation 14 is a convergent power series for $|z| \leq 1$. The coefficients a_j are also the coefficients we get from multiplying out our two operators in Equation 14. So we can also state:

$$P(B)\psi(B)W_t = \Theta(B)W_t \tag{20}$$

$$\sum_{j=0}^{\infty} a_j W_{t-j} = \sum_{j=0}^{q} \theta_j W_{t-j}. \tag{21}$$

Now we can multiply both sides by W_{t-h} for lags h spanning from 1 to ∞. If we take the expectation the equality must still hold and so:

$$\mathbb{E}[\sum_{j=0}^{\infty} a_j W_{t-j} W_{t-h}] = \mathbb{E}[\sum_{j=0}^{q} \theta_j W_{t-j} W_{t-h}] \tag{22}$$

$$\tag{23}$$

But remember that W_t is just white noise, so we know that this gives:

$$a_h = \theta_h, \quad h = 0, 1, \ldots, q \tag{24}$$

$$a_h = 0, \quad h > q. \tag{25}$$

So now we can say that:

$$P(z)\psi(z) = a(z) = \Theta(z), \quad |z| \leq 1. \tag{26}$$

Since $|\psi(z)| < \infty$ for $|z| \leq 1$, we can write:

$$|\psi(z)| = \left|\frac{\Theta(z)}{P(z)}\right| < \infty, \quad |z| \leq 1. \tag{27}$$

Since we know $\Theta(z)$ and $P(z)$ share no roots, this requires that $P(z) \neq 0$ for $|z| \leq 1$. So causality requires that the roots lie outside the unit circle (it is necessary). The second half of the proof, that if the roots lie outside the unit circle we are guaranteed causality (sufficient), can be found in Shumway and Stoffer.

2 Review – Difference Equations

Last lecture we introduced the difference equations as a means of solving for the auto-correlation function of our AR process. We started by noting that the equations for our covariance followed the form:

$$\gamma(h) = \sum_{j=1}^{p} \phi_j \gamma(h-j). \tag{28}$$

We then noted that this is a difference equation of order p:

$$u_n - \alpha_1 u_{n-1} - \ldots - \alpha_p u_{n-p} = 0 \tag{29}$$

whose solutions are given by:

$$u_n = z_1^{-n} P_1(n) + z_2^{-n} P_2(n) + \ldots + z_r^{-n} P_r(n). \tag{30}$$

where r is the number of distinct roots and $P_j(n)$ is a polynomial in n of degree $m_j - 1$ where m_j is the multiplicity of the root z_j. Given Equation 30 it is not difficult to solve for our ACF of an AR(2) process, but let's step back and convince ourselves of these solutions. Let's start with the case where our autoregressive operator has two distinct roots. Following Equation 30, the solution should have the form:

$$\rho(h) = c_1 z_1^{-|h|} + c_2 z_2^{-|h|}. \tag{31}$$

Let's plug that into our equation for the AR(2) auto-correlation function:

$$\rho(h) = \phi_1 \rho(h-1) + \phi_2 \rho(h-2) \tag{32}$$

$$c_1 z_1^{-|h|} + c_2 z_2^{-|h|} = \phi_1 c_1 \left(z_1^{-|h-1|} + c_2 z_2^{-|h-1|} \right) + \phi_2 \left(c_1 z_1^{-|h-2|} + c_2 z_2^{-|h-2|} \right). \tag{33}$$

Moving our terms to the right-hand side:

$$c_1 z_1^{-|h|} \left(1 - \phi_1 z_1 - \phi_2 z_1^2\right) + c_2 z_2^{-|h|} \left(1 - \phi_1 z_2 - \phi_2 z_2^2\right) = 0 \tag{34}$$

$$c_1 z_1^{-|h|} P(z_1) + c_2 z_2^{-|h|} P(z_2) = 0 \tag{35}$$

$$0 = 0. \tag{36}$$

If instead the roots are equal we have:

$$\rho(h) = z_1^{-|h|}(c_1 + c_2|h|). \tag{37}$$

We can plug this in:

$$\rho(h) = \phi_1 \rho(h-1) + \phi_2 \rho(h-2) \tag{38}$$

$$z_1^{-|h|}(c_1 + c_2|h|) = \phi_1 c_1 \left(z_1^{-|h-1|}(c_1 + c_2|h-1|) \right) + \phi_2 \left(z_1^{-|h-2|}(c_1 + c_2|h-2|) \right). \tag{39}$$

Moving it all to the left-hand side:

$$0 = z_1^{-|h|}(c_1 + c_2|h|)(1 - \phi_1 z_1 - \phi_2 z_1^2) + c_2 z_1^{-|h-1|} \left(\phi_1 + 2\phi_2 z_1\right) \tag{40}$$

$$= c_2 z_1^{-|h-1|} \left(\phi_1 + 2\phi_2 z_1\right) \tag{41}$$

$$= 0. \tag{42}$$

where we took advantage of the fact that $1 - \phi_1 z - \phi_2 z^2 = (1 - z_1^{-1} z)^2$, took a derivative, and plugged in z_1 to get the final term to cancel.

3 PACF of AR Process and ACF of MA Process

In the last lecture we introduced the PACF and gained some intuition for its value in identifying an AR(p) process. Let's quickly show mathematically that we can guarantee that the PACF drops to 0 for lags greater than p. If we remember the equation defining an AR process, the best regression for X_{t+h} given $X_{t+1:t+h-1}$ for $h > p$ is simply:

$$\hat{X}_t = \sum_{j=1}^{p} \phi_j X_{t+h-j}, \tag{43}$$

where the remainder is fixed to W_t. We don't even need the estimator, $\hat{X}_t$, since we can then write:

$$\begin{aligned} \phi_{hh} &= \mathbb{E}[(X_{t+h} - \hat{X}_{t+h})(X_t - \hat{X}_t)] && (44) \\ &= \mathbb{E}[(W_{t+h})(X_t - \hat{X}_t)] && (45) \\ &= 0, && (46) \end{aligned}$$

where we are taking advantage of causality in the final line. While it is a little more involved, it's also possible to show that the PACF coefficients for an AR(p) process respect:

$$\phi_{hh} = \phi_h, \tag{47}$$

that is they are a measurement of the fundamental coefficients of that AR process.

Similarly, we have asserted that for an MA process that ACF is a good measurement of its order. Let us circle back to that claim quickly by considering the auto-covariance function of an MA process for order q:

$$\begin{aligned} \gamma(h) &= \mathbb{E}[(X_{t+h})(X_t)] && (48) \\ &= \mathbb{E}[(\sum_{j=0}^{q} \theta_j W_{t+h-j})(\sum_{j=0}^{q} \theta_j W_{t-j})] && (49) \\ &= \begin{cases} \sigma_w^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h} & |h| \leq q \\ 0 & |h| > q \end{cases} && (50) \end{aligned}$$

and to get the ACF we only need to divide by $\gamma(0)$:

$$\rho(h) = \begin{cases} \frac{\sigma_w^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{1 + \sum_{j=1}^{q} \theta_j^2} & |h| \leq q \\ 0 & |h| > q \end{cases}, \tag{51}$$

where I have simplified our notation by setting $\theta_0 = 1$.

An AR(p) process will have an ACF that tails off and a PACF with a sharp cut at lag p. An MA(q) process will have an ACF with a sharp cut at lag q and a PACF that tails off. An ARMA(p, q) that is invertible and causal can be expressed as an infinite AR or MA process. Therefore, it will have both an ACF and a PACF that tails off, although if one of the two processes is more dominant a sharp cut may still be observed.

4 Conditional Prediction

Imagine that we have some time series $\{X_1, \ldots, X_t\}$ that is ARMA and we want to predict $\{X_{k+1}, \ldots X_t\}$ given $\{X_1, \ldots X_k\}$. How would we do that? Well, this is where our linearized Gaussians come into play. First, since every random variable in our ARMA process is linear sum of Gaussian random vairable, we know that the full time series can be described as a multivariate Gaussian. If we think of our time series as a vector, we can write our joint likelihood as:

$$\mathcal{N}(\boldsymbol{X} = \boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right]. \tag{52}$$

Now conditioning on the first k observations doesn't seem so hard: that's just the conditional distribution for a multivariate Gaussian. As a reminder from Lecture 2 that was:

$$p(\boldsymbol{x}_a|\boldsymbol{x}_b) = \mathcal{N}\left(\boldsymbol{x}_a|\boldsymbol{\mu}_{a|b}, \boldsymbol{\Sigma}_{a|b}\right) \tag{53}$$

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\boldsymbol{x}_b - \boldsymbol{\mu}_b) \tag{54}$$

$$\boldsymbol{\Sigma}_{a|b} = \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}, \tag{55}$$

with:

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{x}_a \\ \boldsymbol{x}_b\end{bmatrix} \mid \begin{bmatrix}\boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb}\end{bmatrix}\right). \tag{56}$$

The mean of our ARMA process is 0 since it is stationary, so all we need to do is to calculate our covariance matrix entries $\gamma(t, t+h) = \gamma(h)$. This may seem daunting, but for a causal process we can write an ARMA process as an infinite MA process:

$$X_t = \sum_{j=0}^{\infty} \psi_j W_{t-j}. \tag{57}$$

To solve for the coefficients we can take advantage of the fact that $P(z)\psi(z) = \theta(z)$:

$$(1 - \phi_1 z - \ldots)(\psi_0 + \psi_1 z + \ldots) = (1 + \theta_1 z + \ldots). \tag{58}$$

For this to be true for all z, the coefficients on the left- and right-hand sides for every power of z must be equal. That is:

$$\psi_0 = 1 \tag{59}$$
$$\psi_1 - \phi_1\psi_0 = \theta_1 \tag{60}$$
$$\vdots \tag{61}$$

Getting the generic equation from these solutions is highly non-trivial, but we can iteratively solve for each of the ψ_j up to some arbitrary order computationally. Once we have our ψ_j, we can use them to calculate the auto-correlation function using equation 51.

5 ARIMA Revisited

Let's go back to the integrated ARMA model where we introduce a non-stationary mean. We can break our time series into:

$$X_t = \mu_t + Y_t \tag{62}$$

where Y_t is stationary and μ_t is a time-dependent mean. We can extend ARMA to **integrated ARMA (ARIMA)**:

Definition 6 *A process is said to be ARIMA(p, d, q) if:*

$$\nabla^d X_t = (1 - B)^d X_t, \tag{63}$$

is ARMA(p, q). This means that we can write the process as:

$$P(B)(1 - B)^d X_t = \Theta(B) W_t. \tag{64}$$

The simplest case of this is a linear process $\mu_t = \beta_1 t$. In this case you can write:

$$X_t = \mu_t + Y_t \tag{65}$$
$$= \beta_1 t + Y_t \tag{66}$$

and therefore

$$\begin{aligned}\nabla X_t &= X_t - X_{t-1} &(67)\\ &= \beta_1 + Y_t - Y_{t-1} &(68)\\ &= \beta_1 + \nabla Y_t. &(69)\end{aligned}$$

Assuming that ∇Y_t is ARMA(p, q), then we have that $\nabla X_t = (1 - B)X_t$ is ARMA(p, q) with a non-zero but constant mean.

6 Auto-Regressive Conditionally Heteroscedastic Models

Auto-regressive conditionally heteroscedastic models (ARCH) models focus not on the value of the random variable in the time series, but on its percent change between time steps. That is:

$$\frac{X_t - X_{t-1}}{X_{t-1}} = Y_t, \tag{70}$$

where Y_t is modeled as:

$$\begin{aligned}Y_t &= \sigma_t W_t &(71)\\ \sigma_t^2 &= \alpha_0 + \alpha_1 Y_{t-1}^2 + \dots, &(72)\end{aligned}$$

with $W_t \sim \mathcal{N}(0, 1)$. The idea here is to model the volatility as time dependent. It turns out this is a reasonable model for financial data, although the resulting time series does not fit neatly into our definitions. However, it does have some parallelism to ARMA in its formal structure:

Definition 7 *An ARCH model of order m is a time series defined by:*

$$\begin{aligned}\frac{X_t - X_{t-1}}{X_{t-1}} &= Y_t &(73)\\ Y_t &= \sigma_t W_t &(74)\\ \sigma_t^2 &= \alpha_0 + \alpha_1 Y_{t-1}^2 + \dots + \alpha_m Y_{t-m}^2 &(75)\end{aligned}$$

where $W_t \sim \mathcal{N}(0, 1)$ and $\alpha_m \neq 0$.

We can extend this ARCH model to a generalized ARCH model (GARCH):

Definition 8 *A GARCH model of order m, n is a time series defined by:*

$$\frac{X_t - X_{t-1}}{X_{t-1}} = Y_t \tag{76}$$

$$Y_t = \sigma_t W_t \tag{77}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \ldots + \alpha_m Y_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_n \sigma_{t-n}^2, \tag{78}$$

where $W_t \sim \mathcal{N}(0,1)$ and $\alpha_m, \beta_n \neq 0$.