STAT 153 - Introduction to Time Series
Lecture Fourteen

Fall 2022, UC Berkeley

Aditya Guntuboyina

October 12, 2022

1 Fitting Autoregressive Models to Data

The autoregressive model of order p (referred to by $AR(p)$) is given by:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t \tag{1}$$

The unknown parameters of this model are $\theta := (\phi_0, \ldots, \phi_p, \sigma)$ where σ denotes the standard deviation of the noise variable Z_t.

We study how to fit this model to observed time series data $y_1, \ldots, y_n$. Let us assume that p is known and focus on estimating the parameters in θ along with uncertainty quantification (the problem of selecting p is part of model selection which we shall study later). The likelihood is given by:

$$\begin{aligned}
& f_{Y_1,\ldots,Y_n|\theta}(y_1, \ldots, y_n) \\
&= f_{Y_{p+1},\ldots,Y_n|Y_1=y_1,\ldots,Y_p=y_p,\theta}(y_{p+1}, \ldots, y_n) f_{Y_1,\ldots,Y_p|\theta}(y_1, \ldots, y_p) \\
&= \left[\prod_{t=p+1}^{n} f_{Y_t|Y_{t-1}=y_{t-1},\ldots,Y_1=y_1}(y_t)\right] f_{Y_1,\ldots,Y_p|\theta}(y_1, \ldots, y_p) \\
&= \left[\prod_{t=p+1}^{n} f_{\phi_0+\phi_1 Y_{t-1}+\cdots+\phi_p Y_{t-p}+Z_t|Y_{t-1}=y_{t-1},\ldots,Y_1=y_1}(y_t)\right] f_{Y_1,\ldots,Y_p|\theta}(y_1, \ldots, y_p) \\
&= \left[\prod_{t=p+1}^{n} f_{\phi_0+\phi_1 y_{t-1}+\cdots+\phi_p y_{t-p}+Z_t|Y_{t-1}=y_{t-1},\ldots,Y_1=y_1}(y_t)\right] f_{Y_1,\ldots,Y_p|\theta}(y_1, \ldots, y_p) \\
&= \left[\prod_{t=p+1}^{n} f_{Z_t|Y_{t-1}=y_{t-1},\ldots,Y_1=y_1}(y_t - \phi_0 - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p})\right] f_{Y_1,\ldots,Y_p|\theta}(y_1, \ldots, y_p).
\end{aligned}$$

In order to proceed further, we shall make the following assumption:

$$Z_t \mid Y_{t-1} = y_{t-1}, \ldots, Y_1 = y_1 \sim N(0, \sigma^2) \qquad \text{for each } t = p+1, \ldots, n. \tag{2}$$

We then get

$$
\begin{aligned}
&f_{Y_1,\dots,Y_n|\theta}(y_1,\dots,y_n) \\
&= \left[\prod_{t=p+1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_t-\phi_0-\phi_1 y_{t-1}-\dots-\phi_p y_{t-p})^2}{2\sigma^2}\right)\right] f_{Y_1,\dots,Y_p|\theta}(y_1,\dots,y_p) \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n-p} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{n}(y_t-\phi_0-\phi_1 y_{t-1}-\dots-\phi_p y_{t-p})^2\right) f_{Y_1,\dots,Y_p|\theta}(y_1,\dots,y_p).
\end{aligned}
$$

Observe that, in order to write the above formula, we only used the model equation (1) for $t = p+1,\dots,n$. It is more tricky to derive $f_{Y_1,\dots,Y_p|\theta}(y_1,\dots,y_p)$. Here the simplest thing to do is to ignore the model equation (1) for $t \le p$ (i.e., only make the model assumption (1) for $t = p+1,\dots,n$) and directly assume that

$$f_{Y_1,\dots,Y_p|\theta}(y_1,\dots,y_p) \text{ does not depend on } \theta. \tag{3}$$

This allows us to ignore this term in the likelihood to write

$$f_{Y_1,\dots,Y_n|\theta}(y_1,\dots,y_n) \propto \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n-p} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{n}(y_t-\phi_0-\phi_1 y_{t-1}-\dots-\phi_p y_{t-p})^2\right).$$

This likelihood is familiar to us from linear regression. Indeed, we can rewrite it as

$$f_{Y_1,\dots,Y_n|\theta}(y_1,\dots,y_n) \propto \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n-p} \exp\left(-\frac{\|Y-X\beta\|^2}{2\sigma^2}\right)$$

where

$$
Y_{(n-p)\times 1} = \begin{pmatrix} y_{p+1} \\ y_{p+2} \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \qquad
X_{(n-p)\times(p+1)} = \begin{pmatrix} 1 & y_p & y_{p-1} & \dots & y_1 \\ 1 & y_{p+1} & y_{p+2} & \dots & y_2 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & y_{n-1} & y_{n-2} & \dots & y_{n-p} \end{pmatrix} \qquad
\beta_{(p+1)\times 1} = \begin{pmatrix} \phi_0 \\ \phi_1 \\ \cdot \\ \cdot \\ \cdot \\ \phi_p \end{pmatrix}
$$

Therefore, under the assumptions (2) and (3), we can perform inference in the $AR(p)$ model using the same methodology as in the linear regression model (see Lectures 2 and 3). Specifically, we make the prior assumption

$$\phi_0, \phi_1, \dots, \phi_p, \log\sigma \overset{\text{i.i.d}}{\sim} \text{Unif}(-C, C)$$

for a large C. This will allow us to write down the joint posterior of (β, σ). Integrating over σ leads to the posterior for β alone. As in Lectures 2 and 3, this leads to

$$\beta \mid \text{data} \sim t_{n-2p-1}\left(\hat{\beta}, \hat{\sigma}^2 (X'X)^{-1}\right)$$

where

$$\hat{\beta} := (X'X)^{-1}X'Y \quad \text{and} \quad \hat{\sigma} = \sqrt{\frac{\|Y-X\beta\|^2}{n-2p-1}}.$$

Note that the degrees of freedom of the t-distribution above is $n-2p-1$ as the number of observations equals $n-p$ and the number of components of β is $p+1$.

2 Predictions given by $AR(p)$ models

One important goal of time series analysis is prediction also known as forecasting: given the observed data $y_1, \dots, y_n$, what can we say about the future observations $y_{n+1}, \dots, y_{n+k}$ for some $k \geq 1$? In the Bayesian context, prediction is done via the joint probability distribution of

$$Y_{n+1}, \dots, Y_{n+k}$$

conditional on the observed data $Y_1 = y_1, \dots, Y_n = y_n$. For example, point predictions can be obtained by the conditional expectations:

$$\mathbb{E}\left(Y_{n+1} \mid Y_1 = y_1, \dots, Y_n = y_n\right), \dots, \mathbb{E}\left(Y_{n+k} \mid Y_1 = y_1, \dots, Y_n = y_n\right)$$

Uncertainty quantification for the predictions can be done via the conditional variances:

$$\text{var}\left(Y_{n+1} \mid Y_1 = y_1, \dots, Y_n = y_n\right), \dots, \text{var}\left(Y_{n+k} \mid Y_1 = y_1, \dots, Y_n = y_n\right)$$

Let us focus on point predictions using conditional expectations for now. We shall deal with uncertainty quantification for the predictions in the next class. The conditional expectations can be written as

$$\mathbb{E}\left(Y_{n+i} \mid Y_1 = y_1, \dots, Y_n = y_n\right) = \int \mathbb{E}\left(Y_{n+i} \mid Y_1 = y_1, \dots, Y_n = y_n, \theta\right) f_{\theta \mid Y_1 = y_1, \dots, Y_n = y_n}(\theta) d\theta$$

for $i = 1, \dots, n$. Usually the posterior density $f_{\theta \mid Y_1 = y_1, \dots, Y_n = y_n}(\theta)$ will be highly concentrated around $\hat{\theta} = (\hat{\beta}, \hat{\sigma})$ so that

$$\begin{aligned}
\mathbb{E}\left(Y_{n+i} \mid Y_1 = y_1, \dots, Y_n = y_n\right) &= \int \mathbb{E}\left(Y_{n+i} \mid Y_1 = y_1, \dots, Y_n = y_n, \theta\right) f_{\theta \mid Y_1 = y_1, \dots, Y_n = y_n}(\theta) d\theta \\
&\approx \mathbb{E}\left(Y_{n+i} \mid Y_1 = y_1, \dots, Y_n = y_n, \hat{\theta}\right)
\end{aligned}$$

Thus point predictions are essentially given by

$$\hat{Y}_{n+i} := \mathbb{E}\left(Y_{n+i} \mid Y_1 = y_1, \dots, Y_n = y_n, \hat{\theta}\right) \qquad \text{for } i = 1, \dots, k.$$

For the $AR(p)$ model, there is a simple way of computing these predictions recursively for $i = 1, 2, \dots$. This is because (assuming the validity of the model equation (1) also for $t > n$)

$$\begin{aligned}
\hat{Y}_{n+i} &= \mathbb{E}\left(Y_{n+i} \mid Y_1 = y_1, \dots, Y_n = y_n, \hat{\theta}\right) \\
&= \mathbb{E}\left(\phi_0 + \phi_1 Y_{n+i-1} + \phi_2 Y_{n+i-2} + \cdots + \phi_p Y_{n+i-p} \mid Y_1 = y_1, \dots, Y_n = y_n, \hat{\theta}\right) \\
&= \hat{\phi}_0 + \hat{\phi}_1 \hat{Y}_{n+i-1} + \hat{\phi}_2 \hat{Y}_{n+i-2} + \cdots + \phi_p \hat{Y}_{n+i-p}.
\end{aligned}$$

Note that we replaced $\phi_0, \dots, \phi_p$ by $\hat{\phi}_0, \dots, \hat{\phi}_p$ because we have conditioned on $\theta = \hat{\theta}$. We thus have the following recursion for the predictions $\hat{Y}_{n+i}$:

$$\hat{Y}_{n+i} = \hat{\phi}_0 + \hat{\phi}_1 \hat{Y}_{n+i-1} + \hat{\phi}_2 \hat{Y}_{n+i-2} + \cdots + \phi_p \hat{Y}_{n+i-p} \qquad \text{for } i = 1, 2, \dots. \tag{4}$$

If we initialize this recursion with

$$\hat{Y}_j = y_j \qquad \text{for } j = n, n-1, \dots, n+1-p, \tag{5}$$

then (4) can be evaluated in sequence for $i = 1, 2, \dots$ to calculate $\hat{Y}_{n+i}$ for all $i \geq 1$.

3 Difference Equations and their Solutions

The behavior of the predictions (1) given by the $AR(p)$ model can be understood by looking at difference equations. A difference equation is of the form:

$$u_k = \alpha_0 + \alpha_1 u_{k-1} + \cdots + \alpha_p u_{k-p} \qquad \text{for } k = p, p+1, p+2, \dots. \tag{6}$$

This is initialized by specifying the values of $u_0, u_1, \dots, u_{p-1}$. Clearly the prediction recursion (4) of the $AR(p)$ model along with the initial condition (5) is similar to (6) (basically take $u_j = \hat{Y}_{n+1-p+j}$). (6) is called a difference equation of order p. In order to understand its solutions, let us start with the case $p = 1$.

3.1 First Order

Here $p = 1$ so the difference equation becomes:

$$u_k = \alpha_0 + \alpha_1 u_{k-1} \qquad \text{for } k = 1, 2, \dots$$

along with an initial value specification for u_0. We first convert this equation into a **homogeneous** difference equation (a homogeneous equation is one with no intercept term) by taking

$$v_k = u_k - \frac{\alpha_0}{1-\alpha_1}$$

so that

$$v_k = u_k - \frac{\alpha_0}{1-\alpha_1} = \alpha_0 + \alpha_1 u_{k-1} - \frac{\alpha_0}{1-\alpha_1} = \alpha_0 + \alpha_1\left(v_{k-1} + \frac{\alpha_0}{1-\alpha_1}\right) - \frac{\alpha_0}{1-\alpha_1} = \alpha_1 v_{k-1}.$$

Thus v_k satisfies the homogenous equation:

$$v_k = \alpha_1 v_{k-1}.$$

It is now easy to see that the solution is given by

$$v_k = \alpha_1^k v_0 \qquad \text{for } k = 0, 1, 2, \dots.$$

The solution for u_k is thus given by

$$\begin{aligned} u_k &= \frac{\alpha_0}{1-\alpha_1} + \alpha_1^k\left(u_0 - \frac{\alpha_0}{1-\alpha_1}\right) \\ &= \left(1-\alpha_1^k\right)\frac{\alpha_0}{1-\alpha_1} + \alpha_1^k u_0 \\ &= \left(1 + \alpha_1 + \alpha_1^2 + \cdots + \alpha_1^{k-1}\right)\alpha_0 + \alpha_1^k u_0 \qquad \text{for } k = 0, 1, 2, \dots. \end{aligned}$$

The last expression above also makes sense when $\alpha_1 = 1$ (note that, when $\alpha_1 = 1$, some of the previous expressions do not make sense because $1 - \alpha_1$ appearing in the denominator). The behavior of u_k will then be of three kinds depending on the precise value of α_1:

1. $|\alpha_1| < 1$: Here u_k converges exponentially to $\alpha_0/(1-\alpha_1)$.

2. $|\alpha_1| > 1$: Here, when k gets large, u_k is essentially equal to $\alpha_1^k u_0$ which is exploding to infinity exponentially in magnitude.

3. $\alpha_1 = 1$: Here $u_k = k\alpha_0 + u_0$ which is linear

4. $\alpha_1 = -1$: Here u_k oscillates between the two values u_0 and $\alpha_0 - u_0$.

3.2 Second Order

Here $p = 2$ so the difference equation becomes

$$u_k = \alpha_0 + \alpha_1 u_{k-1} + \alpha_2 u_{k-2} \qquad \text{for } k = 2, 3, \ldots$$

along with an initial value specification for u_0 and u_1. To write down the general solution to this equation, first let us convert it into a homogeneous equation by taking

$$v_k = u_k - \frac{\alpha_0}{1 - \alpha_1 - \alpha_2}$$

which is well-defined as long as $\alpha_1 + \alpha_2 \neq 1$. It is then clear that

$$v_k = \alpha_1 v_{k-1} + \alpha_2 v_{k-2} \qquad \text{for } k = 2, 3, \ldots. \tag{7}$$

In order to write down the general solution to this homogeneous difference equation, consider the characteristic equation (see, for example, `https://en.wikipedia.org/wiki/Characteristic_equation_(calculus)`) of this difference equation which is defined by:

$$\phi(z) = 1 - \alpha_1 z - \alpha_2 z^2.$$

Suppose z_1 is a root of this polynomial i.e., $\phi(z_1) = 1 - \alpha_1 z_1 - \alpha_2 z_1^2 = 0$. Then it is easy to check that z_1^{-k} is a solution of (7) because

$$z_1^{-k} - \alpha_1 z_1^{-(k-1)} - \alpha_2 z_1^{-(k-2)} = z_1^{-k}\left(1 - \alpha_1 z_1 - \alpha_2 z_1^2\right) = z_1^{-k}\phi(z_1) = 0.$$

As a result, we can write down the general solution to (7) by considering the following separate cases. Let z_1 and z_2 be the two roots of the polynomial $\phi(z)$.

Case 1: z_1 and z_2 are real and distinct: Here

$$v_k = c_1 z_1^{-k} + c_2 z_2^{-k}$$

is a solution to (7) for every real c_1 and c_2. If we now choose c_1 and c_2 so as to satisfy the initial values given for v_0 and v_1, we would obtain the solution we are seeking.

Case 2: z_1 and z_2 are complex: As the polynomial $\phi(z)$ has real coefficients, we must have $z_2 = \bar{z}_1$ here (i.e., z_2 is the complex conjugate of z_1). Now

$$v_k = c_1 z_1^{-k} + \bar{c}_1 z_2^{-k} = c_1 z_1^{-k} + \bar{c}_1 (\bar{z}_1)^{-k}$$

satisfies (7) for every complext number c_1. We just need to choose c_1 so as to satisfy the initial conditions. Writing $z_1 = |z_1|e^{i\theta}$ and $\bar{c}_1 = ae^{ib}$, we can simplify the expression for v_k as

$$\begin{aligned} v_k &= |z_1|^{-k} e^{-ik\theta} a e^{-ib} + |z_1|^{-k} e^{ik\theta} a e^{ib} \\ &= |z_1|^{-k} a \left(e^{i(b+k\theta)} + e^{-i(b+k\theta)}\right) = |z_1|^{-k} 2a \cos(b + k\theta) \end{aligned} \tag{8}$$

This expression has two constants a and b which need to be chosen so as to satisfy the initial conditions for v_0 and v_1. The interesting thing to note is that, in this case where the roots of $\phi(z)$ are complex, the predictions will be oscillating. Here is a concrete example.

Example 3.1. *Fitting the AR(2) model to the sunspots data for the time period* $1700 - 1969$ *led to the model:*

$$Y_{t+2} = 24.11 + 1.38 Y_{t+1} - 0.69 Y_t + Z_{t+2}$$

which gives the prediction equation:

$$U_t = 24.11 + 1.38U_{t-1} - 0.69U_{t-2}$$

for the future values of sunspot numbers from 1970 onwards. This equation can also be written as

$$U_t - 77.77 = 1.38\left(U_{t-1} - 77.77\right) - 0.69\left(U_{t-2} - 77.77\right)$$

Thus the predictions for $V_t := U_t - 77.77$ *are given by recursing the equation:*

$$V_t = 1.38V_{t-1} - 0.69V_{t-2} \tag{9}$$

for $t = n+1, n+2, \dots$ *(note that* V_{n-1} *and* V_n *are observed from the data). The characteristic eqution of this difference equation is*

$$1 - 1.38z + 0.69z^2$$

whose roots are complex and given by $(1.2)e^{\pm i0.59}$. *Therefore, from* (8)*, the general solution for* (9) *is given by*

$$c_1\left(1.2\right)^{-t}\cos\left(0.59t + c_2\right)$$

for two constants c_1 *and* c_2*. The above is clearly a damped sinusoid (the sinusoid* $\cos\left(0.59t + c_2\right)$ *is damped by the factor* $(1.2)^{-t}$*). This is why the predictions of the* $AR(2)$ *model for the sunspots data having a sinusoidal pattern initially by they quickly decay to a constant value.*

Finally we consider the third case for writing the solution to (7).

Case 3: $z_1 = z_2$. Here z_1 and z_2 must necessarily be real. One solution to (7), as before, is given by z_1^{-k}. Another solution is given by $v_k = kz_1^{-k}$. To see this, write

$$kz_1^{-k} - \alpha_1(k-1)z_1^{-(k-1)} - \alpha_2(k-2)z_1^{-(k-2)} = kz_1^{-k}\left(1 - \alpha_1 z_1 - \alpha_2 z_1^2\right) + z_1^{-(k-1)}\left(\alpha_1 + 2\alpha_2 z_1\right)$$

The first term on the right hand side above equals zero because $\phi(z_1) = 0$ (as z_1 is a root of ϕ). The second term is also zero because z_1 (being a common root) is also a root of $\frac{d\phi(z)}{dz} = -\alpha_1 - 2\alpha_2 z$. Thus a general solution to (7) is given by

$$v_k = c_1 z_1^{-k} + c_2 k z_1^{-k} = z_1^{-k}\left(c_1 + kc_2\right).$$

We shall write down the general solution for the difference equation of order p in the next class.

4 Recommended Reading for Today

1. For more on fitting $AR(p)$ models to data, see Section 3.5 of the book by Shumway and Stoffer titled *Time Series Analysis and its applications* (Fourth Edition). The method we discussed is known as **conditional least squares estimation**.
2. For more on difference equations, see Section 3.2 of the Shumway-Stoffer book.

References

[1] Yule, G. U. (1927). On a method of investigating periodicities disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 226*(636-646), 267–298.