# STAT 153 & 248 - Time Series
# Lecture Sixteen

**Fall 2025, UC Berkeley**

Aditya Guntuboyina

October 23, 2025

## 1 Auto or Lagged Regression

Our next topic of study is AutoRegression. The main models here are called ARIMA. ARIMA is an acronym standing for Auto-Regressive Integrated Moving Average. We will first study Auto-Regressive (AR) models, then we shall include the MA part to get ARMA models, finally we see what "Integrated" means.

## 2 AR (Auto-Regressive) Models

We observe time series $y_1, \dots, y_n$. AR models are simply linear regression models where the covariates are chosen to be past (or lagged) values of $y_t$.

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

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t \tag{1}$$

for $t = p+1, \dots, n$. In matrix notation,

$$Y = X\beta + \epsilon$$

where

$$Y = \begin{pmatrix} y_{p+1} \\ y_{p+2} \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & y_p & y_{p-1} & \cdot & \cdot & \cdot & y_1 \\ 1 & y_{p+1} & y_p & \cdot & \cdot & \cdot & y_2 \\ 1 & y_{p+2} & y_{p+1} & \cdot & \cdot & \cdot & y_3 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & y_{n-1} & y_{n-2} & \cdot & \cdot & \cdot & y_{n-p} \end{pmatrix} \quad \beta = \begin{pmatrix} \phi_0 \\ \phi_1 \\ \phi_2 \\ \cdot \\ \cdot \\ \cdot \\ \phi_p \end{pmatrix} \quad \epsilon = \begin{pmatrix} \epsilon_{p+1} \\ \epsilon_{p+2} \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{pmatrix}$$

This regression model is called AutoRegression because the responses as well as the covariates are both formed from the same time series: the time series $y_t$ is regressed on its own lagged values $y_{t-1}, \dots, y_{t-p}$.

The parameters $\phi_0, \dots, \phi_p$ are estimated in the usual way by minimizing $\|Y - X\beta\|^2$. Let the estimates by $\hat{\phi}_0, \dots, \hat{\phi}_p$.

AR models are useful for predicting future values of the time series. For predicting $y_{n+1}$, we plug $t = n + 1$ in (1) to get

$$y_{n+1} = \hat{\phi}_0 + \hat{\phi}_1 y_n + \hat{\phi}_2 y_{n-1} + \cdots + \hat{\phi}_p y_{n+1-p}.$$

Note that $y_n, y_{n-1}, \ldots, y_{n+1-p}$ are all observed and they are the last $p$ observations. For predicting $y_{n+2}$, we plug $t = n + 2$ in (1) to get

$$y_{n+2} = \hat{\phi}_0 + \hat{\phi}_1 y_{n+1} + \hat{\phi}_2 y_n + \cdots + \hat{\phi}_p y_{n+2-p}.$$

In the above, $y_{n+1}$ is not observed. But we can replace it by the predicted value $\hat{y}_{n+1}$. This gives

$$y_{n+2} = \hat{\phi}_0 + \hat{\phi}_1 \hat{y}_{n+1} + \hat{\phi}_2 y_n + \cdots + \hat{\phi}_p y_{n+2-p}.$$

More generally, we predict $y_{n+i}$ by the recursion

$$\hat{y}_{n+i} = \hat{\phi}_0 + \hat{\phi}_1 \hat{y}_{n+i-1} + \cdots + \hat{\phi}_p \hat{y}_{n+i-p} \qquad \text{for } i = 1, 2, \ldots$$

where the recursion is initialized with

$$\hat{y}_j = y_j \qquad \text{for } j = n, n-1, \ldots, n+1-p.$$

We will look at AR models in more details in the coming lectures. Today, we shall provide a motivation for their use through the sunspots dataset. This was how the AR models were originally invented by Yule [1].

## 3 AR Models for the Sunspots Data

For the sunspots dataset, we previously employed the model

$$y_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t) + \epsilon_t \qquad \text{for } t = 1, \ldots, n \tag{2}$$

We used a Bayesian method to infer the frequency parameter $f$ (which is the main parameter of interest) and this led to an estimated period of close to 11 years (which is often cited as the period of the solar cycle). Note however that (2) is not ideal for the sunspots dataset for at least two reasons: (a) the fit to the data is not very good (some of the oscillations have a much higher amplitude than that explained by the single sinusoid), (b) data generated from the model (2) look much more "noisy" compared to the actual sunspots data. Starting with these observations, Yule [1] proposed an alternative model that is also based on a single sinuosoid. This alternative model is based on the idea of AR modeling.

Yule started with the following basic observation. Let $s_t$ denote the sinusoid:

$$s_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t) \tag{3}$$

The same sinusoid can be understood as the solution to a specific *difference equation*. To derive the difference equation, let us first note that, in continuous time, $s(t)$ satisfies

$$s''(t) = -(2\pi f)^2 \left(\beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)\right) = -(2\pi f)^2 \left(s(t) - \beta_0\right). \tag{4}$$

In discrete time (where $t \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$), the sequence (3) satisfies the following difference equation that is analogous to (4):

$$s_t - 2s_{t-1} + s_{t-2} = 2(\cos(2\pi f) - 1)\left(s_{t-1} - \beta_0\right). \tag{5}$$

To see this, note that (below we take $\omega = 2\pi f$ for notational simplicity)

$$\begin{aligned} s_t - 2s_{t-1} + s_{t-2} &= \beta_1 \left(\cos(\omega t) - 2\cos(\omega(t-1)) + \cos(\omega(t-2))\right) \\ &\quad + \beta_2 \left(\sin(\omega t) - 2\sin(\omega(t-1)) + \sin(\omega(t-2))\right) \end{aligned}$$

Writing $A = \omega(t-1)$ and $B = \omega$, we get

$$\begin{aligned} \cos(\omega t) - 2\cos(\omega(t-1)) + \cos(\omega(t-2)) &= \cos(A+B) - 2\cos A + \cos(A-B) \\ &= 2\cos A(\cos B - 1) \\ &= 2(\cos\omega - 1)\cos(w(t-1)) \end{aligned}$$

and similarly

$$\sin(\omega t) - 2\sin(\omega(t-1)) + \sin(\omega(t-2)) = 2(\cos\omega - 1)\sin(\omega(t-1)).$$

This proves

$$\begin{aligned} s_t - 2s_{t-1} + s_{t-2} &= 2(\cos\omega - 1)\left(\beta_1\cos(\omega(t-1)) + \beta_2\sin(\omega(t-1))\right) \\ &= 2(\cos\omega - 1)(s_{t-1} - \beta_0) \end{aligned}$$

thereby establishing (5).

The converse is also true in the sense that every solution $\{s_t\}$ to the difference equation (5) say, for $t = 1, 2, 3, \dots$, with given values of $s_1$ and $s_2$ (initial conditions) is of the form (3) for some $\beta_1$ and $\beta_2$. To see this, let $g_t = s_t - \beta_0$ and note that $\{g_t\}$ satisfies

$$g_t - 2g_{t-1} + g_{t-2} = 2(\cos\omega - 1)g_{t-1}.$$

We find $\beta_1$ and $\beta_2$ such that (note again that $\omega = 2\pi f$)

$$h_t := \beta_1\cos(\omega t) + \beta_2\sin(\omega t)$$

matches $g_t$ for $t = 1, 2$. Now if $g_{t-1} = h_{t-1}$ and $g_{t-2} = h_{t-2}$, then

$$\begin{aligned} g_t &= (2\cos\omega)g_{t-1} - g_{t-2} \\ &= (2\cos\omega)h_{t-1} - h_{t-2} \\ &= (2\cos\omega)\left(\beta_1\cos(\omega(t-1)) + \beta_2\sin(\omega(t-1))\right) - \left(\beta_1\cos(\omega(t-2)) + \beta_2\sin(\omega(t-2))\right) \\ &= \beta_1\left(2\cos\omega\cos(\omega(t-1)) - \cos(\omega(t-2))\right) + \beta_2\left(2\cos\omega\sin(\omega(t-1)) - \sin(\omega(t-2))\right). \end{aligned}$$

Verify that

$$2\cos\omega\cos(\omega(t-1)) - \cos(\omega(t-2)) = \cos(\omega t)$$

and

$$2\cos\omega\sin(\omega(t-1)) - \sin(\omega(t-2)) = \sin(\omega t),$$

which gives

$$g_t = \beta_1\cos(\omega t) + \beta_2\sin(\omega t) = h_t.$$

We thus proved that if $g_{t-1} = h_{t-1}$ and $g_{t-2} = h_{t-2}$, then $g_t = h_t$. Using this for $t = 1, 2, \dots$ proves that (3) is the unique solution to (5).

To summarize, an alternative way of describing a sinusoid of frequency $\omega = 2\pi f$ is via the difference equation (5) which is equivalent to

$$s_t = (2\cos\omega)s_{t-1} - s_{t-2} + 2(1 - \cos\omega)\beta_0.$$

Based on this equation, Yule proposed the model:

$$y_t = \phi_0 + \phi_1 y_{t-1} - y_{t-2} + \epsilon_t \tag{6}$$

with two parameters $\phi_0$ and $\phi_1$ (and the additional noise parameter $\sigma$ in $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$). Note that (6) is also a single sinusoid plus noise model but now the noise is in a different place.

To better understand the difference between (6) and the earlier model (2), consider the following physical situation where sinusoids naturally arise (see e.g., page 2 of the Fourier Analysis book by Stein and Shakarchi). Consider a mass $m$ that is attached to a horizontal spring, which itself is attached to fixed wall, and assume that the system lies on a frictionless surface. Suppose that $\beta_0$ is the location of the center of the mass when the spring is neither compressed or stretched. When the spring is compressed or stretched and released, the mass undergoes simple harmonic motion.

Let $s(t)$ denote the position of the mass at time $t$. Hooke's law says that the force exerted by the spring on the mass is given by $F = -\kappa (s(t) - \beta_0)$ where $\kappa > 0$ is the spring constant. By Newton's law (note that the acceleration is given by $s''(t)$), we have

$$-\kappa (s(t) - \beta_0) = ms''(t)$$

This is same as

$$s''(t) = -\omega^2 (s(t) - \beta_0) \qquad \text{where } \omega := \sqrt{\frac{k}{m}}$$

whose general solution is the sinusoid $s(t) = \beta_0 + \beta_1 \cos(\omega t) + \beta_2 \sin(\omega t)$. In the context of this physical situation, the two different noisy sinusoid models ((2) and (6)) can be understood as follows. We are taking measurements of the displacement $y_t$ at various times $t$.

**Model** (2): Here our measurements are noisy and every measurement is corrupted by an unknown noise which we are terming $\epsilon_t$ and modeling as $N(0, \sigma^2)$.

**Model** (6): Here there is no measurement error and our measurement mechanism is perfect. However the actual oscillation of the mass is not perfectly sinusoidal and is affected by noise. For example, imagine, as Yule put it, that some kids are randomly throwing stones at the mass (sometimes from the left and sometimes from the right) while it is oscillating.

It is very interesting to note that observations generated from Model (6) are much smoother compared to observations generated from Model (2). Yule used this to argue that (6) is a better model for the sunspots data compared to (2).

The AR(2) model is:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \tag{7}$$

(6) can be seen as a simpler version of the above model where the $\phi_2$ parameter is set to the value $-1$. Yule fit both models ((6) and (7)) to the sunspots dataset. It is interesting that these two models give different predictions for future sunspots values: (6) gives sinusoidal predictions while (7) gives *damped* sinusoidal predictions. In the coming lectures, we shall discuss how different parameter settings for AR models lead to different predictions.

A very nice account of Yule's influential 1927 paper is Chapter 6 of the 2011 book "The Foundations of Modern Time Series Analysis" by T. C. Mills. (available for free from the library website). Yule's paper [1] itself is available freely online.

## 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.