STAT 153 & 248 - Time Series
Lecture Seventeen

Spring 2025, UC Berkeley

Aditya Guntuboyina

March 20, 2025

1 AutoRegressive Models

In the last lecture, we started discussing AutoRegressive models as a first step towards learning the more general ARIMA models. Methodologically, AutoRegression is simply regression of the observed time series on lagged versions of itself. Suppose the observed dataset is $y_1, \ldots, y_n$. From this data, we create a $(n-p) \times 1$ vector Y and a $(n-p) \times (p+1)$ design matrix X as follows:

$$Y = \begin{pmatrix} y_{p+1} \\ y_{p+2} \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & y_p & y_{p-1} & \cdot & \cdot & \cdot & y_1 \\ 1 & y_{p+1} & y_p & \cdot & \cdot & \cdot & y_2 \\ 1 & y_{p+2} & y_{p+1} & \cdot & \cdot & \cdot & y_3 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & y_{n-1} & y_{n-2} & \cdot & \cdot & \cdot & y_{n-p} \end{pmatrix}$$

p here is an integer which represents the order of the AutoRegressive (AR) model. We then regress Y on X (in the standard way using least squares or OLS) to obtain fitted regression coefficients $\phi_0, \phi_1, \ldots, \phi_p$. Since the response variable is y_t and the regressors are $1, y_{t-1}, \ldots, y_{t-p}$ for $t = p+1, \ldots, n$, the fitted regression model can be written as

$$y_t = \hat{\phi}_0 + \hat{\phi}_1 y_{t-1} + \cdots + \hat{\phi}_p y_{t-p}. \tag{1}$$

This is how data is typically analyzed in AutoRegression. In the next section, we shall write down the structure of the AR **model** and we shall see how the above estimation procedure is related to Maximum Likelihood.

One of the main uses of AR (and more generally ARIMA) models is for prediction (also known as forecasting).

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

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

When this method is applied on some time series datasets, the predictions obtained can vary quite significantly with p. We will try to obtain some intuition for the structure of the predictions later in this lecture.

2 The AR Model

Let us start with the AR(1) model (we will later address AR(p) for $p \geq 2$).

The AR(1) model is given by:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t \qquad \text{for } t = 2, \ldots, n. \tag{2}$$

2.1 Detour: usual regression

The model (2) looks just like a usual regression model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{3}$$

except we are using ϕ instead of β for the coefficients, and the index is now t as opposed to i. Let us recall how one writes the likelihood (for parameter estimation) in (3). The data is $(x_i, y_i), i = 1, \ldots, m$ (m is the number of data points). The likelihood is the probability density function of the data treated as a function of the parameters $\theta = (\beta_0, \beta_1, \sigma)$ (σ is the standard deviation of the errors):

$$\text{Likelihood for model (3)} = f_{x_1, y_1, x_2, y_2, \ldots, x_n, y_n | \theta}(x_1, y_1, \ldots, x_n, y_n).$$

We first assume independence across i (i.e., $(x_1, y_1), \ldots, (x_n, y_n)$ are independent) to get

$$\text{Likelihood for model (3)} = \prod_{i=1}^{n} f_{x_i, y_i | \theta}(x_i, y_i).$$

Because the model (3) specifies an equation for y_i in terms of x_i, it is natural to condition first on x_i leading to:

$$\text{Likelihood for model (3)} = \prod_{i=1}^{n} f_{y_i | x_i, \theta}(y_i) f_{x_i | \theta}(x_i).$$

Now we use the model equation to replace y_i by $\beta_0 + \beta_1 x_i + \epsilon_i$:

$$\begin{aligned} \text{Likelihood for model (3)} &= \prod_{i=1}^{n} f_{\beta_0 + \beta_1 x_i + \epsilon_i | x_i, \theta}(y_i) f_{x_i | \theta}(x_i) \\ &= \prod_{i=1}^{n} f_{\epsilon_i | x_i, \theta}(y_i - \beta_0 - \beta_1 x_i) f_{x_i | \theta}(x_i) \end{aligned}$$

We now assume that ϵ_i is independent of x_i and that $\epsilon_i \sim N(0, \sigma^2)$. These result in

$$\begin{aligned}
\text{Likelihood for model (3)} &= \prod_{i=1}^{m} f_{\epsilon_i|x_i,\theta}(y_i - \beta_0 - \beta_1 x_i) f_{x_i|\theta}(x_i) \\
&= \prod_{i=1}^{m} f_{\epsilon_i|\theta}(y_i - \beta_0 - \beta_1 x_i) f_{x_i|\theta}(x_i) \\
&= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right) f_{x_i|\theta}(x_i) \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^m \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(y_i - \beta_0 - \beta_1 x_i)^2\right) \prod_{i=1}^{m} f_{x_i|\theta}(x_i).
\end{aligned}$$

How do we deal with the last term $\prod_{i=1}^{m} f_{x_i|\theta}(x_i)$? We simply assume that this term does not depend on θ so it only becomes a constant (in terms of θ) multiplicative factor in the likelihood that can be omitted leading to

$$\text{Likelihood for model (3)} \propto \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^m \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(y_i - \beta_0 - \beta_1 x_i)^2\right).$$

This is the standard form of the likelihood in linear regression that we worked with previously (see e.g., Lectures 2 and 3). To sum up, we used the following assumptions to derive this likelihood:

1. Independence of $(x_1, y_1), \ldots, (x_n, y_n)$
2. The model equation (3)
3. Independence of ϵ_i and x_i
4. $\epsilon_i \sim N(0, \sigma^2)$
5. The density of x_i does not depend on $\theta = (\beta_0, \beta_1, \sigma)$.

2.2 Back to AR(1)

Let us now get back to the AR(1) model (2). Superficially, (2) looks the same as (3) with $i = t$, $x_i = y_{t-1}$ and $\beta_0 = \phi_0$ and $\beta_1 = \phi_1$. However, some of the other regression assumptions listed above do not hold for (2):

1. Independence of (x_i, y_i) across i no longer holds. This is because $(x_t, y_t) = (y_{t-1}, y_t)$ so one observation is shared between (x_t, y_t) for successive values of t.
2. The density of $x_t = y_{t-1}$ will depend on θ (because $y_{t-1} = \phi_0 + \phi_1 y_{t-2} + \epsilon_{t-1}$ so ϕ_0, ϕ_1 and σ certainly affect y_{t-1}).

As a result, we cannot use the same principles as in usual linear regression to write the likelihood for AR models. Instead we shall proceed as follows. As the data is $y_1, \ldots, y_n$, the

likelihood is given by (below $\theta = (\phi_0, \phi_1, \sigma)$ denotes the set of parameters)

$$\begin{aligned}
&\text{Likelihood for Model (2)} \\
&= f_{y_1,\ldots,y_n|\theta}(y_1,\ldots,y_n) \\
&= f_{y_1|\theta}(y_1) f_{y_2|y_1,\theta}(y_2) f_{y_3|y_1,y_2,\theta}(y_3)\ldots f_{y_n|y_1,\ldots,y_{n-1},\theta}(y_n) \\
&= f_{y_1|\theta}(y_1)\prod_{t=2}^{n} f_{y_t|y_1,\ldots,y_{t-1},\theta}(y_t) \\
&= f_{y_1|\theta}(y_1)\prod_{t=2}^{n} f_{\phi_0+\phi_1 y_{t-1}+\epsilon_t|y_1,\ldots,y_{t-1},\theta}(y_t) \\
&= f_{y_1|\theta}(y_1)\prod_{t=2}^{n} f_{\epsilon_t|y_1,\ldots,y_{t-1},\theta}(y_t-\phi_0-\phi_1 y_{t-1}).
\end{aligned}$$

Now we assume that ϵ_t is independent of $y_1,\ldots y_{t-1}$. This gives

$$\begin{aligned}
&\text{Likelihood for Model (2)} \\
&= f_{y_1|\theta}(y_1)\prod_{t=2}^{n} f_{\epsilon_t|y_1,\ldots,y_{t-1},\theta}(y_t-\phi_0-\phi_1 y_{t-1}) = f_{y_1|\theta}(y_1)\prod_{t=2}^{n} f_{\epsilon_t}(y_t-\phi_0-\phi_1 y_{t-1}).
\end{aligned}$$

With $\epsilon_t \sim N(0,\sigma^2)$, we get

$$\text{Likelihood for Model (2)} = f_{y_1|\theta}(y_1)\prod_{t=2}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(y_t-\phi_0-\phi_1 y_{t-1})^2\right)$$

which is equivalent to:

$$\text{Likelihood for (2)} = f_{y_1|\theta}(y_1)\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n-1}\exp\left(-\frac{1}{2\sigma^2}\sum_{t=2}^{n}(y_t-\phi_0-\phi_1 y_{t-1})^2\right). \qquad (4)$$

To sum up, we used the following assumptions to derive the likelihood (4):

1. The model equation (2).
2. Independence of ϵ_t and $y_1,\ldots,y_{t-1}$ for each $t = 2,\ldots,n-1$.
3. $\epsilon_t \sim N(0,\sigma^2)$.

The likelihood (4) has the term $f_{y_1|\theta}(y_1)$ which we should make explicit before we can compute estimators. Note that the model equation (2) is only for $t = 2,\ldots,n$ which means that y_1 never appears on the left side. So it is not possible to compute $f_{y_1|\theta}(y_1)$ using (2). There are two approaches of dealing with $f_{y_1|\theta}(y_1)$.

1. **Approach One**: Here one simply assumes that $f_{y_1|\theta}(y_1)$ does not depend on θ. Then $f_{y_1|\theta}(y_1)$ becomes a constant factor in (4) that can be ignored in proportionality leading to

$$\text{Likelihood for (2)} \propto \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n-1}\exp\left(-\frac{1}{2\sigma^2}\sum_{t=2}^{n}(y_t-\phi_0-\phi_1 y_{t-1})^2\right). \qquad (5)$$

   It is easy to verify that maximizing the above likelihood leads to estimates $\hat{\phi}_0, \hat{\phi}_1$ that are identical to those obtained by regression Y on X as described in Section 1. The right hand side of (5) is actually equal to the conditional density of $y_2,\ldots,y_n$ given y_1, θ (under the aforementioned assumptions: model (2), $\epsilon_t \sim N(0,\sigma^2)$ and independence of ϵ_t and $y_1,\ldots,y_{t-1}$). For this reason, (5) is called "Conditional Likelihood" and the resulting maximizers "Conditional MLEs" or "Conditional Least Squares Estimators". "Conditional" here refers to conditional on y_1.

2. **Approach Two**: Here one extends the model equation (2) to $t = 1, 0, -1, -2, \ldots$. This allows possible computation of $f_{y_1|\theta}(y_1)$ in the following way. Applying (2) for $t = 1, 0, -1, -2, \ldots$ recursively, we get

$$\begin{aligned} y_1 &= \phi_0 + \phi_1 y_0 + \epsilon_1 \\ &= \phi_0 + \phi_1 (\phi_0 + \phi_1 y_{-1} + \epsilon_0) + \epsilon_1 \\ &= \phi_0(1 + \phi_1) + \phi_1^2 y_{-1} + \phi_1 \epsilon_0 + \epsilon_1 \\ &= \phi_0(1 + \phi_1) + \phi_1^2 (\phi_0 + \phi_1 y_{-2} + \epsilon_{-1}) + \phi_1 \epsilon_0 + \epsilon_1 \\ &= \phi_0 \left(1 + \phi_1 + \phi_1^2\right) + \phi_1^3 y_{-2} + \phi_1^2 \epsilon_{-1} + \phi \epsilon_0 + \epsilon_1. \end{aligned}$$

Continuing this way with using (2) for $t = -2, -3, \ldots, -M$ (for some large M), we get

$$\begin{aligned} y_1 &= \phi_0 \left(1 + \phi_1 + \phi_1^2 + \cdots + \phi_1^M\right) + \phi_1^{M+1} y_{-M} + \phi_1^M \epsilon_{-M+1} + \phi_1^{M-1} \epsilon_{-M+2} + \cdots + \phi \epsilon_0 + \epsilon_1 \\ &= \phi_0 \sum_{j=0}^{M} \phi_1^j + \phi_1^{M+1} y_{-M} + \sum_{j=0}^{M} \phi_1^j \epsilon_{1-j}. \end{aligned}$$

This equation is not enough to allow us to deduce $f_{y_1|\theta}(y_1)$ because it involves the unknown quantity y_{-M}. If $|\phi_1| < 1$, then the coefficient ϕ_1^{M+1} in front of y_{-M} is very small. In this case, it might make sense to ignore the term $\phi_1^{M+1} y_{-M}$ when M is large. This allows us to write

$$y_1 \approx \phi_0 \sum_{j=0}^{M} \phi_1^j + \sum_{j=0}^{M} \phi_1^j \epsilon_{1-j} \approx \phi_0 \sum_{j=0}^{\infty} \phi_1^j + \sum_{j=0}^{\infty} \phi_1^j \epsilon_{1-j} = \frac{\phi_0}{1 - \phi_1} + \sum_{j=0}^{\infty} \phi_1^j \epsilon_{1-j}.$$

The term $\sum_{j=0}^{\infty} \phi_1^j \epsilon_{1-j}$ is the sum of independent normal random variables, so it is Normal with mean zero (as each ϵ_{1-j} has mean zero) and with variance:

$$\text{var}\left(\sum_{j=0}^{\infty} \phi_1^j \epsilon_{1-j}\right) = \sum_{j=0}^{\infty} \text{var}\left(\phi_1^j \epsilon_{1-j}\right) = \sum_{j=0}^{\infty} \phi_1^{2j} \text{var}(\epsilon_{1-j}) = \sigma^2 \sum_{j=0}^{\infty} \phi_1^{2j} = \frac{\sigma^2}{1 - \phi_1^2}.$$

Thus when $|\phi_1| < 1$, we can write

$$y_1 \sim N\left(\frac{\phi_0}{1 - \phi_1}, \frac{\sigma^2}{1 - \phi_1^2}\right).$$

which gives

$$f_{y_1|\theta}(y_1) = \frac{\sqrt{1 - \phi_1^2}}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1 - \phi_1^2}{2\sigma^2}\left(y_1 - \frac{\phi_0}{1 - \phi_1}\right)^2\right).$$

Plugging this in (4), we get

Likelihood for (2)

$$= \frac{\sqrt{1 - \phi_1^2}}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1 - \phi_1^2}{2\sigma^2}\left(y_1 - \frac{\phi_0}{1 - \phi_1}\right)^2\right) \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=2}^{n} (y_t - \phi_0 - \phi_1 y_{t-1})^2\right). \tag{6}$$

This is a more complicated likelihood compared to (5). This is applicable only when $|\phi_1| < 1$. We shall see later the implications of this assumption. (6) is referred to as the full likelihood for AR(1), and the estimates obtained by maximization of (6) as full MLEs (as opposed to conditional MLEs obtained by maximizing (5)). In cases where the assumption $|\phi_1| < 1$ is reasonable, full MLEs will be different from conditional MLEs although when n is large, they will generally be quite close to each other.

2.3 AR(p)

The AR(p) model is given by:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t \tag{7}$$

We can write the likelihood as

$$f_{y_1,\ldots,y_n|\theta}(y_1,\ldots,y_n) = f_{y_{p+1},\ldots,y_n|y_1,\ldots,y_p,\theta}(y_{p+1},\ldots,y_n) f_{y_1,\ldots,y_p|\theta}(y_1,\ldots,y_p).$$

The conditional likelihood is calculated as

$$\begin{aligned}
&f_{y_{p+1},\ldots,y_n|y_1,\ldots,y_p,\theta}(y_{p+1},\ldots,y_n) \\
&= \prod_{t=p+1}^{n} f_{y_t|y_{t-1},\ldots,y_1}(y_t) \\
&= \prod_{t=p+1}^{n} f_{\phi_0+\phi_1 y_{t-1}+\cdots+\phi_p y_{t-p}+\epsilon_t|y_{t-1},\ldots,y_1}(y_t) \\
&= \prod_{t=p+1}^{n} f_{\epsilon_t|y_{t-1},\ldots,y_1}(y_t - \phi_0 - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p}).
\end{aligned}$$

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

$$\epsilon_t \mid y_{t-1},\ldots,y_1 \sim N(0,\sigma^2) \qquad \text{for each } t = p+1,\ldots,n. \tag{8}$$

This is equivalent to assuming that $\epsilon_t \sim N(0,\sigma^2)$ and that ϵ_t is independent of $y_1,\ldots,y_{t-1}$. With (8), we get

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

Observe that, in order to write the above formula, we only used the model equation (7) for $t = p+1,\ldots,n$.

To obtain the parameter estimates, we can directly maximize this conditional likelihood. The resulting estimates, which are identical to the OLS method described in Section 1, are known as Conditional MLEs or Conditional Least Squares Estimates.

The full likelihood is

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

If we assume that $f_{y_1,\ldots,y_p|\theta}(y_1,\ldots,y_p)$ does not depend on θ, then maximizing the full likelihood is equivalent to maximizing the conditional likelihood. If we want to derive $f_{y_1,dots,y_p|\theta}(y_1,\ldots,y_p)$ in a more principled way, then we have to use the model equation (7) for smaller values of t (i.e., $t = p, p-1, p-2, \ldots, 0, -1, \ldots$). We shall see later how this is done (this will also require some assumptions similar to $|\phi_1| < 1$ for $p = 1$).

3 Predictions and Difference Equations

Given a fitted AR(p) model with parameter estimates $\hat{\phi}_0, \ldots, \hat{\phi}_p$, predictions $\hat{y}_{n+i}$ for $i = 1, 2, \ldots$ are obtained by the recursion:

$$\hat{y}_{n+i} = \hat{\phi}_0 + \hat{\phi}_1 \hat{y}_{n+i-1} + \cdots + \hat{\phi}_p \hat{y}_{n+i-p} \qquad \text{for } i = 1, 2, \ldots \tag{9}$$

where the recursion is initialized with

$$\hat{y}_j = y_j \qquad \text{for } j = n, n-1, \ldots, n+1-p. \tag{10}$$

The behavior of the predictions (9) given by the $AR(p)$ model can be understood by looking at difference equations. A difference equation is of the form:

$$u_k = \alpha_0 + \alpha_1 u_{k-1} + \cdots + \alpha_p u_{k-p} \qquad \text{for } k = p, p+1, p+2, \ldots. \tag{11}$$

This is initialized by specifying the values of $u_0, u_1, \ldots, u_{p-1}$. Clearly the prediction recursion (9) of the $AR(p)$ model along with the initial condition (10) is similar to (11) (basically take $u_j = \hat{Y}_{n+1-p+j}$). (11) 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 ($p = 1$)

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

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

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

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

The solution for u_k is thus given by

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

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 $\alpha_0 - u_0$.

We shall see formulae for solutions of the difference equation for $p \geq 2$ in the next lecture.

3.2 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).
2. For more on difference equations, see Section 3.2 of the Shumway-Stoffer book.