Autoregressive moving average models

The classical regression model was developed for the static case, namely, we only allow the dependent variable to be influenced by current values of the independent variables. In the time series case, it is desirable to allow the dependent variable to be influenced by the past values of the independent variables and possibly by its own past values. If the present can be plausibly modeled in terms of only the past values of the independent inputs, we have the enticing prospect that forecasting will be possible.

Autoregressive models are based on the idea that the current value of the series, x_t, can be explained as a function of p past values, x_{t-1}, x_{t-2}, $\ldots$, x_{t-p}, where p determines the number of steps into the past needed to forecast the current value. As a typical case

$$x_t = x_{t-1} - 0.9x_{t-2} + w_t$$

Definición 1. *An autoregressive model of order p, abbreviated $\mathbf{AR}(p)$, is of the form*

$$
\begin{aligned}
x_t - \mu &= \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \cdots + \phi_p(x_{t-p} - \mu) + w_t \\
&= \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t
\end{aligned}
$$

where x_t is stationary, $w_t \sim \mathcal{N}(0, \sigma_w)$, and $\phi_1, \ldots \phi_p$ are constants with $\phi_p \neq 0$.

We note that the autoregressive model is similar to the regression model, and hence the term auto (or self) regression. Some technical difficulties, however, develop from applying that model because the regressors, $x_{t-1}, \ldots, x_{t-p}$, are random components, whereas z_t was assumed to be fixed.

Why $x_{t-1}, \ldots, x_{t-p}$ are random components?

A useful form follows by using the backshift operator to write the $\mathbf{AR}(p)$ model, as

$$\phi(\boldsymbol{B})x_t = (1 - \phi_1\boldsymbol{B} - \phi_2\boldsymbol{B}^2 - \cdots - \phi_p\boldsymbol{B}^p)x_t = w_t$$

Definición 2. *The autoregressive operator is defined to be*

$$\phi(\boldsymbol{B}) = (1 - \phi_1\boldsymbol{B} - \phi_2\boldsymbol{B}^2 - \cdots - \phi_p\boldsymbol{B}^p)$$

The $\mathbf{AR}(1)$ model

Consider the first-order model, $\mathbf{AR}(1)$, given by $x_t = \phi x_{t-1} + w_t$. Iterating backwards k times, we get

$$
\begin{aligned}
x_t &= \phi x_{t-1} + w_t = \phi(\phi x_{t-2} + w_{t-1}) + w_t \\
&= \phi^2 x_{t-2} + \phi w_{t-1} + w_t \\
&\ldots \\
&= \phi^k x_{t-k} + \sum_{j=0}^{k-1} \phi^j w_{t-j}
\end{aligned}
$$

Stationary solution

This method suggests that, by continuing to iterate backward, and provided that $|\phi| < 1$ and $\sup_t \mathrm{var}(x_t) < \infty$, we can represent an $\mathbf{AR}(1)$ model as a linear process given by

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j}$$

This representation is called the stationary solution of the model.

- Compute the the mean of the $\mathbf{AR}(1)$ process.
- Compute the autocovariance $\gamma(h)$ of the $\mathbf{AR}(1)$ process [Hint: $\sum_{k=0}^{\infty} a\, r^k = \frac{a}{1-r}$ for $|r| < 1$]
- Is the $\mathbf{AR}(1)$ process stationary?
- Compute the autocorrelation function $\rho(h) = \gamma(h) / \gamma(0)$
- Show that $\rho(h) = \phi\rho(h-1)$ (See ARIMA models notebook)

Statistical properties

The $\mathrm{AR}(1)$ process is stationary with:

- $\mathbb{E}[x_t] = 0$
- $\gamma(h) = \frac{\sigma^2 \phi^h}{1 - \phi^2}$
- $\rho(h) = \phi \rho(h-1) = \phi^h$

Explosive AR models

We might wonder whether there is a stationary $\mathrm{AR}(1)$ process with $|\phi| > 1$. Such processes are called explosive because the values of the time series quickly become large in magnitude. We can, however, modify that argument to obtain a stationary model as follows. Write $x_{t+1} = \phi x_t + w_{t+1}$, in which case,

$$
\begin{aligned}
x_t &= \phi^{-1}x_{t+1} - \phi^{-1}w_{t+1} = \phi^{-1}(\phi^{-1}x_{t+2} + \phi^{-1}w_{t+2}) - \phi^{-1}w_{t+1} \\
&\dots \\
&= \phi^{-k}x_{t+k} - \sum_{j=1}^{k-1} \phi^{-j}w_{t+j}
\end{aligned}
$$

By iterating forward k steps.

Causality

Because $\phi^{-1} < 1$, this result suggests the stationary future dependent $\mathrm{AR}(1)$ model

$$x_t = -\sum_{j=1}^{\infty} \phi^{-1} w_{t+j}$$

Unfortunately, this model is useless because it requires us to know the future to be able to predict the future. When a process does not depend on the future, such as the $\mathrm{AR}(1)$ when $|\phi| < 1$, we will say the process is *causal*.

Every explosion has a cause

Excluding explosive models from consideration is not a problem because the models have causal counterparts. For example, if

$$x_t = \phi x_{t-1} + w_t$$

with $|\phi| > 1$, $w_t \sim N(0, \sigma_w^2)$ is a non-causal stationary process with

- $\mathbb{E}[x_t] = 0$
- $\gamma(h) = \frac{\sigma_w^2 \phi^{-2} \phi^{-h}}{1 - \phi^{-2}}$

Thus, the causal process defined by

$$y_t = \phi^{-1} y_{t-1} + v_t$$

where $v_t \sim N(0, \sigma_w^2 \phi^{-2})$ is stochastically equal to the x_t process.

For example, if $x_t = 2x_{t-1} + w_t$ with $\sigma_w^2 = 1$, then $y_t = \frac{1}{2} y_{t-1} + v_t$ with $\sigma_v^2 = \frac{1}{4}$.

This concept generalizes to higher orders.

General stationary solution

The technique of iterating backward to get an idea of the stationary solution of AR models works well when $p = 1$, but not for larger orders. A general technique is that of matching coefficients.

Consider the $\mathrm{AR}(1)$ model in operator form

$$(1 - \phi B)x_t = \phi(B)x_t = w_t$$

where $|\phi| < 1$. Also, we can write the model as

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B)w_t$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ and $\psi_j = \phi^j$.

Inverse operator

Consider the $\mathrm{AR}(1)$ model in the operator form

$$\phi(B)x_t = w_t$$

Now multiply both sides by $\phi^{-1}(B)$ (assuming that the inverse operator exists)

$$x_t = \phi^{-1}(B)w_t$$

but, from the stationary form we know that $\phi^{-1}(B) = 1 + \phi B + \phi^2 B^2 + \cdots +$, that is $\phi^{-1}(B) = \psi(B)$. Thus, we notice that working with operators is like working with polynomials.

Operator as polynomial

Consider the polynomial $x_t = 1 - \phi z$, where z is a complex number and $|\phi| < 1$. Then

$$\phi^{-1}(z) = \frac{1}{1 - \phi z} = 1 + \phi z + \phi^2 z^2 + \cdots$$

and the coefficients of B^j in $\phi^{-1}(B)$ are the same as the coefficients of z^j in $\phi^{-1}(z)$. In other words, we may treat the backshift operator, B, as a complex number, z.

Estimating of $\mathrm{AR}(p)$ parameters

Recall that an $\mathrm{AR}(p)$ model can be written as

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

where $w_t \sim \mathrm{WN}(0, \sigma^2)$. We would like to estimate the parameters $\boldsymbol{\phi} = (\phi_1, \ldots \phi_p)^T$ and σ^2. To estimate these parameters we have some known methods:

- Least squares method
- Maximum Likelihood method (leads to least squares method)
- Yule-Walker equations

Least squares method

Consider a dataset of N samples (examples). Our goal is to minimize

$$\sum_{i=p+1}^{N} w_i^2 = \sum_{i=p+1}^{N} \left(x_i - \phi_0 + \sum_{j=1}^{p} \phi_j x_{i-j} \right)^2$$

Let $A = [a_{ij}] \in \mathbb{R}^{(N-p)\times(p+1)}$ where $a_{i1} = [1] \forall i$, and the ith row $a_{ij} = x_{i-j+1} \forall j > 1$. Let $\boldsymbol{x} = [x_{p+1}, x_{p+2}, \ldots, x_N]^T$, $\boldsymbol{\phi} = [\phi_0, \phi_1, \ldots, \phi_p]^T$, and $\boldsymbol{w} = [w_{p+1}, w_{p+2}, \ldots, w_N]$. Then the $\mathrm{AR}(p)$ process can be represented by

$$\boldsymbol{x} = A\boldsymbol{\phi} + \boldsymbol{w}$$

And then the error can be written as

$$E(\boldsymbol{\phi}) = \|\boldsymbol{w}\|^2 = \|\boldsymbol{x} - A\boldsymbol{\phi}\|^2 = \boldsymbol{x}^T\boldsymbol{x} - 2\boldsymbol{\phi}^T A^T \boldsymbol{x} + \boldsymbol{\phi}^T A^T A \boldsymbol{\phi}$$

Optimizing least squares

To minimize the error, we computhe the gradient and set to $\mathbf{0}$

$$\nabla E(\boldsymbol{\phi}) = 2(A^T A \boldsymbol{\phi} - A^T \boldsymbol{x}) = 0$$

Then we get

$$\boldsymbol{\phi} = (A^T A)^{-1} A^T \boldsymbol{x}$$

Once we get $\boldsymbol{\phi}$ we can estimate the variance

$$\sigma^2 = \frac{1}{N-p-1} \sum_{t=p+1}^{N} (x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p})^2$$

Yule-Walker technique

Consider the next $\mathrm{AR}(p)$ process with mean $\mu = 0$

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

Multiplying by x_{t-h} with $h > 0$ and take the expectations

$$\mathbb{E}[x_t x_{t-h}] = \phi_1 \mathbb{E}[x_{t-1} x_{t-h}] + \cdots + \phi_p \mathbb{E}[x_{t-p} x_{t-h}] + \mathbb{E}[w_t x_{t-h}]$$

Then

$$\gamma(h) = \phi_1 \gamma(h-1) + \cdots + \phi_p \gamma(h-p)$$

Divining by $\gamma(0)$

$$\rho(h) = \phi_1 \rho(h-1) + \cdots + \phi_p \rho(h-p)$$

Solving equations

Since $\rho(h) = \rho(-h)$, taking $h = 1, \ldots, p$

$$\left.\begin{array}{l} \rho(1) = \phi_1\rho(0) + \phi_2\rho(1) + \cdots + \phi_p\rho(p-1) \\ \ldots \\ \rho(p) = \phi_1\rho(p-1) + \phi_2\rho(p-2) + \cdots + \phi_p\rho(0) \end{array}\right\} \text{Yule-Walker equations}$$

If the $\rho(h)$'s are known solving these equations the estimate can be obtained.

To estimate the variance, multiply both sides by x_t and take the expectations, to get

$$\sigma^2 = \gamma(0) - \phi_1\gamma(1) - \cdots - \phi_p\gamma(p)$$

However, in general the $\rho(h)$'s would be unknown and needs to be estimated from the sample.

As an alternative to the autoregressive representation in which the x_t on the left-hand side of the equation are assumed to be combined linearly, the moving average model of order q, abbreviated as $\mathbf{MA}(q)$, assumes the white noise w_t on the right-hand side of the defining equation are combined linearly to form the observed data.

Definición 3. *The moving average model of order q, or $\mathrm{MA}(q)$ model, is defined to be*

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}$$

where $w_t \sim \mathcal{N}(0, \sigma_w)$, and $\theta_1, \theta_2, \ldots, \theta_q$ are the parameters with $\theta_q \neq 0$.

We may also write the $\mathbf{MA}(q)$ process in the equivalent form $x_t = \theta(\boldsymbol{B}) w_t$

Definición 4. *The moving average operator is*

$$\theta(\boldsymbol{B}) = 1 + \theta_1 \boldsymbol{B} + \theta_2 \boldsymbol{B}^2 + \cdots + \theta_q \boldsymbol{B}^q$$

Unlike the autoregressive process, the moving average process is stationary for any values of the parameters $\theta_1, \ldots, \theta_q$.

The $\mathbf{MA}(1)$ process

Consider the $\mathbf{MA}(1)$ model $x_t = w_t + \theta w_{t-1}$. Then:

- $\mathbb{E}[x_t] = 0$
- $\gamma(h) = \begin{cases} (1+\theta^2)\sigma_w^2 & h = 0 \\ \theta\sigma_w^2 & h = 1 \\ 0 & h > 1 \end{cases}$
- $\rho(h) = \begin{cases} \frac{\theta}{1+\theta^2} & h = 1 \\ 0 & h > 1 \end{cases}$

Note that for an $\mathrm{MA}(1)$ model, $\rho(h)$ is the same for θ and $\frac{1}{\theta}$. For example, compute $\rho(5)$ and $\rho(\frac{1}{5})$. In addition, the pair $\sigma_w = 1$ and $\theta = 5$ yield the same autocovariance funtion as the pair $\sigma_w = 25$ and $\theta = \frac{1}{5}$.

Thus, the $\mathrm{MA}(1)$ processes

$$x_t = w_t + \frac{1}{5}w_{t-1}, \quad w_t \sim \mathcal{N}(0, 25)$$

and

$$y_t = v_t + 5v_{t-1}, \quad v_t \sim \mathcal{N}(0, 1)$$

are the same. We can only observe the time series, x_t or y_t, and not the noise, w_t or v_t, so we cannot distinguish between the models. Hence, we will have to choose only one of them.

Invertibility of MA models

For convenience, by mimicking the criterion of causality for AR models, we will choose the model with an infinite AR representation. Such a process is called an *invertible process*.

To discover which model is the invertible model, we can reverse the roles of x_t and w_t (because we are mimicking the AR case) and write the $\mathrm{MA}(1)$ model as

$$w_t = -\theta w_{t-1} + x_t$$

So, if $|\theta| < 1$, then

$$w_t = \sum_{j=0}^{\infty} (-\theta)^j x_{t-j}$$

Hence, given a choice, we will choose the model with $\sigma_w = 25$, and $\theta = \frac{1}{5}$ because it is invertible.

MA polynomial

As in the AR case, the polynomial, $\theta(z)$, corresponding to the moving average operators, $\theta(B)$, will be useful in exploring general properties of MA processes.

For example, we can write the $\mathrm{MA}(1)$ model as $x_t = \theta(B) w_t$, where $\theta(B) = 1 + \theta B$.

If $|\theta| < 1$, then we can write the model as $\pi(B) x_t = w_t$, where $\pi(B) = \theta^{-1}(B)$.

Let $\theta(z) = 1 + \theta z$, for $|z| \leq 1$, then $\theta^{-1}(z) = 1/(1 + \theta z) = \sum_{j=0}^{\infty} (-\theta)^j z^j$, and we determine that $\pi(B) = \sum_{j=0}^{\infty} (-\theta)^j B^j$.

Estimating the MA parameters

Estimate the MA parameters is difficult because the regressors are unknown innovations (white noises).

An ad-hoc method is as follows:

- Invert the $\mathrm{MA}(q)$to an $\mathrm{AR}(\infty)$.
- Cut-off the AR at some suitable order.
- Use any of theprevious methods toestimate the AR parameters.
- Solve the MA parameters from their relationship with the AR parameters.

Definición 5. *A time series* $\{x_t; t=0, \pm 1, \pm 2, \dots\}$ *is* $\mathrm{ARMA}(p,q)$ *if it is stationary and*

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

with $\phi_p \neq 0, \theta_q \neq 0$, *and* $\sigma_w^2 > 0$. *The parameters* p *and* q *are called the autoregressive and the moving average orders, respectively. If* x_t *has a non-zero mean* μ, *we set* $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$ *and write the model as:*

$$x_t = \alpha + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

where $w_t \sim \mathcal{N}(0, \sigma_w)$.

ARMA with operators

In particular, the $\mathrm{ARMA}(p, q)$ model can then be written in concise form as

$$\phi(B)x_t = \theta(B)w_t$$

The concise form of the model points to a potential problem in that we can unnecessarily complicate the model by multiplying both sides by another operator.

$$\eta(B)\phi(B)x_t = \eta(B)\phi(B)w_t$$

without changing the dynamics.

Parameter redundancy

Consider a white noise process $x_t = w_t$. If we multiply both sides of the equation by $\eta(B) = 1 - 0.5B$, then the model becomes $(1-0.5B)x_t = (1-0.5B)w_t$, or

$$x_t = 0.5x_{t-1} - 0.5w_{t-1} + w_t$$

which looks like an $\mathrm{ARMA}(1,1)$ model, but of course, x_t is still white noise; nothing has changed in this regard, but we have hidden the fact that x_t is white noise because of the parameter redundancy or over-parameterization.

Problems summary

Previous slides point to a number of problems with the general definition of $\mathrm{ARMA}(p, q)$ models:

i. Parameter redundant models

ii. Stationary AR models that depend on the future, and

iii. MA models that are not unique.

To overcome these problems, we will require some additional restrictions on the model parameters.

Definición 6. *The AR and MA polynomials are defined as*

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p, \quad \phi_p \neq 0$$

and

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q, \quad \theta_q \neq 0$$

respectively, where z *is a complex number.*

To address the first problem, we will henceforth refer to an $\mathrm{ARMA}(p, q)$ model to mean that it is in its simplest form. That is, in addition to the original definition of an ARMA process, we will also require that $\phi(z)$ and $\theta(z)$ have no common factors.

So, the process, $x_t = 0.5x_{t-1} - 0.5w_{t-1} + w_t$, discussed before is not referred to as an $\mathrm{ARMA}(1, 1)$ process because, in its reduced form, x_t is white noise.

Definición 7. *An* $\mathrm{ARMA}(p,q)$ *model is said to be causal, if the time series* $\{x_t; t=0,\pm 1, \pm 2,\dots\}$ *can be written as a one-sided linear process:*

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B) w_t$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$*, and* $\sum_{j=0}^{\infty} |\psi_j| < \infty$*; we set* $\psi_0 = 1$*.*

In the $\mathrm{AR}(1)$ process, $x_t = \phi x_t + w_t$, is causal only when $|\phi| < 1$. Equivalently, the process is causal only when the root of $\phi(z) = 1 - \phi z$ is bigger than one in absolute value. That is, the root, say, z_0, of $\phi(z)$ is $z_0 = \frac{1}{\phi}$ (because $\phi(z_0) = 0$) and $|z_0| > 1$ because $|\phi| < 1$. In general, we have the following property.

Causality of an $\mathrm{ARMA}(p, q)$ process

An $\mathrm{ARMA}(p, q)$ model is causal if and only if $\phi(z) \neq 0$ for $|z| \leq 1$. The coefficients of the previous linear process can be determined by solving

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1$$

Another way to phrase this property is that *an ARMA process is causal only when the roots of $\phi(z)$ lie outside the unit circle; that is, $\phi(z) = 0$ only when $|z| > 1$.*

Definición 8. *An* $\mathrm{ARMA}(p,q)$ *model is said to be invertible, if the time series* $\{x_t; t=0,\pm 1, \pm 2, \ldots\}$ *can be written as*

$$\pi(B)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} = w_t$$

where $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$*, and* $\sum_{j=0}^{\infty} |\pi_j| < \infty$*; we set* $\pi_0 = 1$*.*

Invertibility of an $\mathrm{ARMA}(p, q)$ process

An $\mathrm{ARMA}(p, q)$ model is invertible if and only if $\theta(z) \neq 0$ for $|z| \leq 1$. The coefficients of the previous linear process can be determined by solving

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1$$

Another way to phrase this property is that *an ARMA process is invertible only when the roots of* $\theta(z)$ *lie outside the unit circle; that is,* $\theta(z) = 0$ *only when* $|z| > 1$.

Example: Parameter Redundancy

Consider the process

$$x_t = 0.4x_{t-1} + 0.45x_{t-2} + w_t + w_{t-1} + 0.25w_{t-2}$$

or, in the operator form

$$(1 - 0.4B - 0.45B^2)x_t = (1 + B + 0.25B^2)w_t$$

At first, x t appears to be an $\mathrm{ARMA}(2,2)$ process. But notice that:

$$\begin{aligned}\phi(B) &= 1 - 0.4B - 0.45B^2 = (1 + 0.5B)(1 - 0.9B)\\ \theta(B) &= 1 + B + 0.25B^2 = (1 + 0.5B)^2\end{aligned}$$

There is a common factor that can be canceled(Parameter redundacy). So the model is an $\mathrm{ARMA}(1,1)$.

Example: Causality and invertibility

The $\mathrm{ARMA}(1,1)$ model is

$$x_t = 0.9x_{t-1} + 0.5w_{t-1} + w_t$$

The model is causal because $\phi(z) = 1 - 0.9z = 0$ when $z = \frac{10}{9}$, which is outside the unit circle.

The model is also invertible because the root of $\theta(z) = 1 + 0.5z = 0$ when $z = -2$, which is outside the unit circle.

Example: Causality form

To write the model as a linear process (causality), we can obtain the ψ-weights using the fact that $\phi(z)\psi(z) = \theta(z)$, or

$$(1 - 0.9z)(1 + \psi_1 z + \psi_2 z^2 + \cdots + \psi_j z^j + \cdots) = 1 + 0.5z$$

Rearranging, we get

$$1 + (\psi_1 - 0.9)z + (\psi_2 - 0.9\psi_1)z^2 + \cdots + (\psi_j - 0.9\psi_{j-1})z^j + \cdots = 1 + 0.5z$$

Matching the coefficients of z on the left and right sides we get $\psi_1 - 0.9 = 0.5$ and $\psi_j - 0.9\psi_{j-1} = 0$ for $j > 1$. Thus, $\psi_j = 1.4(0.9)^{j-1}$ for $j \geq 1$ and then $x_t = 0.9x_{t-1} + 0.5w_{t-1} + w_t$ can be written as

$$x_t = w_t + 1.4 \sum_{j=1}^{\infty} 0.9^{j-1} w_{t-j}$$

Example: Invertible form

The invertible representation is obtained by matching coefficients in $\theta(z)\pi(z) = \phi(z)$,

$$(1 + 0.5z)(1 + \pi_1 z + \pi_2 z^2 + \cdots) = 1 - 0.9z$$

In this case, the π-weights are given by $\pi_j = (-1)^j (1.4)(0.5)^j$, for $j \geq 1$, and hence, because $w_t = \sum_{j=0}^{\infty} \pi_j x_{t-j}$, we can write $x_t = 0.9x_{t-1} + 0.5w_{t-1} + w_t$ as

$$x_t = 1.4 \sum_{j=1}^{\infty} (-0.5)^{j-1} x_{t-j} + w_t$$

Difference equations

The study of the behavior of ARMA processes and their ACFs is greatly enhanced by a basic knowledge of difference equations.

Suppose we have a sequence of numbers $u_0, u_1, \ldots$ such that

$$u_n - \alpha u_{n-1} = 0, \quad \alpha \neq 0, \quad n = 1, 2, \ldots$$

To solve the *homogeneous difference equation of order 1*, we write:

$$\begin{aligned} u_1 &= \alpha u_0 \\ u_2 &= \alpha u_1 = \alpha^2 u_0 \\ &\ldots \\ u_n &= \alpha u_{n-1} = \alpha^n u_0 \end{aligned}$$

Given an initial condition $u_0 = c$, we may solve the equation, namely, $u_n = \alpha^n c$

Solutions as polynomial roots

We can write $u_n - \alpha u_{n-1} = 0$ in operator notation

$$(1 - \alpha B) u_n = 0$$

The polynomial associated is $\alpha(z) = 1 - \alpha z$, and the root, say, $z_0 = 1/\alpha$.

Then, we can write the solution with initial condition $u_0 = c$ as

$$u_n = \alpha^n c = (z_0^{-1})^n c$$

That is, the solution of the difference equation depends only on the initial condition and the inverse of the root to the associated polynomial $\alpha(z)$.

Second order

Now suppose that the sequence satisfies:

$$u_n - \alpha_1 u_{n-1} - \alpha_2 u_{n-2} = 0, \quad \alpha_2 \neq 0, \quad n = 2, 3, \ldots$$

This equation is a *homogeneous difference equation of order 2*. The corresponding polynomial is

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

which has two roots, say, z_1 and z_2; that is, $\alpha(z_1) = \alpha(z_2) = 0$.

Solution cases

Lets consider the next cases for $\alpha(z) = 1 - \alpha_1 z - \alpha_2 z^2$ with roots z_1, z_2

- If $z_1 \neq z_2$ then the general solution is $u_n = c_1 z_1^{-n} + c_2 z_2^{-n}$.
- If $z_1 = z_2 (= z_0)$ then the general solution is $u_n = z_0^{-n}(c_1 + c_2 n)$.

To summarize these results, the solution to the homogeneous difference equation of degree two was:

$$\begin{aligned} u_n &= z_1^{-n} \times (\text{a polynomial in } n \text{ of degree } m_1 - 1) \\ &\quad + z_2^{-n} \times (\text{a polynomial in } n \text{ of degree } m_2 - 1) \end{aligned}$$

where m_1 is the multiplicity of the root z_1 and m_2 is the multiplicity of the root z_2.

These results generalize to the homogeneous difference equation of order p:

$$u_n - \alpha_1 u_{n-1} - \cdots - \alpha_p u_{n-p} = 0, \quad \alpha_p \neq 0, \quad n = p, p+1, \ldots$$

The associated polynomial is $\alpha(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$. Suppose $\alpha(z)$has r distinct roots, z_1 with multiplicity m_1, z_2 with multiplicity m_2, ..., and z_r with multiplicity m_r, such that $m_1 + m_2 + \cdots + m_r = p$. The general solution to the difference equation is:

$$u_n = z_1^{-n} P_1(n) + z_2^{-n} P_2(n) + \cdots + z_r^{-n} P_r(n)$$

where $P_j(n)$, for $j = 1, 2, \ldots, r$, is a polynomial in n, of degree $m_j - 1$. Given p initial conditions $u_0, \ldots u_{p-1}$, we can solve for the $P_j(n)$ explicitly.