# Notes on times series analysis

Jiahua Tian

This notes is not intended to be exhaustive and is still under construction.

## Contents



## 1 What is time series?

*A times series is a set of observations, each one being recorded at a specific time.*

What are the objectives of *time series analysis*?

- Forecast the future.
- Control and detect whether there is something wrong in a process.

- Understand the features of data, including periodicity, trend, etc.

A time series is a stochastic process. A stochastic process is an ordered set of random variables labelled by time

$$\{X_t, t \in T\}.$$

Usually we have either $\{X_t, t \in (-\infty, \infty)\}$ or $\{X_t, t \in (1, 2, 3, \dots)\}$.

A time series is a sequence of observations ordered by time therefore they depend on each other. Ultimately we want to learn the joint probability distribution

$$p(X_1, X_2, \dots, X_T)$$

where each $X_i$ is drawn from a distribution. As we have only one observation for each distribution the ordinary statistical inference is infeasible, therefore we need to impose certain very restrictive assumption to the time series to simplify the problem. This assumption is known as *Stationarity*.

## 2 Stationarity

Stationarity means the probability laws do not change over time, i.e., the process is in *statistical equilibrium*. Now that since each observation is drawn from the same distribution, this basically falls in the realm of statistical mechanics in physics. For example, since each distribution has the same mean, this mean value can be estimated based a slice of time series of size $n \gg 1$. This is exactly the idea of *ensemble* in statistical mechanics.

Now we define $n^{th}$ order stationarity. For a time series to be $n^{th}$ order stationary, we have:

$$p(X_{t_1} = x_{t_1}, \dots, X_{t_n} = x_{t_n}) = p(X_{t_1+k} = x_{t_1}, \dots, X_{t_n+k} = x_{t_n}). \tag{1}$$

Strong ($n^{th}$) order stationarity means that shifting the time origin by $k$ has no effect on the joint distribution so that it depends only the time intervals between $t_1, t_2, \dots, t_n$. Usually we assume only weak (second) order stationarity, i.e.,

- $\mathbb{E}[X_t] = \mu$ is constant.
- $Var[X_t] = \sigma^2$ is constant.
- $Cov(X_t, X_{t+k}) = \gamma(k)$ is not a function of $t$.
- $Cor(X_t, X_{t+k}) = \rho(k)$ is not a function of $t$.

In general we have the following definitions.

**Definition** The mean of a time series is $\mu_t = \mathbb{E}[X_t]$. The autocovariance function (ACVF) of a times series is $\gamma_{t,s} = Cov(X_t, X_s)$. The autocorrelation function (ACF) $\rho_{t,s} = Cor(X_t, X_s)$.

We have

$$\rho_{t,s} = \frac{Cov(X_t, X_s)}{\sqrt{Var(X_t)Var(X_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}}. \tag{2}$$

For a (weak) stationary process we have

$$\rho_k = \frac{\gamma_k}{\gamma_0}.$$

## 2.1 Examples

### 2.1.1 Random walk

Now let us look at an example. A (random walk) is defined to be

$$X_t = X_{t-1} + e_t$$

where $e_t \sim N(0, \sigma^2)$. Therefore we have $\mu_t = \sum \mu(e_t) = 0$ and $Var(X_t) = t\sigma^2$. Hence this process is not weak stationary as its variance depends linearly on time.

It is interesting to compute the ACF of random walk. Anticipating the results in a later section, we have

$$X_t = e_t + e_{t-1} + e_{t-2} + \cdots .$$

Therefore we have $\gamma(t,s) = \sum^s \sigma^2$ $(s < t)$ and $\gamma(t,t) = \sum^t \sigma^2$ and $\gamma(s,s) = \sum^s \sigma^2$ therefore

$$\rho(t,s) = \sqrt{\frac{s}{t}}.$$

We could put it in a more familiar form:

$$\rho(t, t+k) = \sqrt{\frac{t}{t+k}} = \frac{1}{\sqrt{1+k/t}}.$$

### 2.1.2 Moving average

Suppose $X_t = \frac{1}{2}(e_t + e_{t-1})$ and $\mu(e_t) = 0$ and $Var(e_t) = \sigma^2$. We have $\mu(X_t) = 0$ and

$$Var(X_t) = Var\left(\frac{e_t + e_{t-1}}{2}\right) = \frac{1}{2}\mathbb{E}[e_t^2] = \frac{1}{2}\sigma^2.$$

It is not hard to compute that

$$\begin{aligned}
\gamma_{t,s} &= \frac{1}{2}\sigma^2,\ t = s,\\
\gamma_{t,s} &= \frac{1}{4}\sigma^2,\ t = s+1,\\
\gamma_{t,s} &= 0,\ t > s+1.
\end{aligned}$$

This is an example of a stationary process.

## 2.2 Properties

Recall $Var(aX + bY) \geq 0$ we have:

$$(a^2 + b^2)\sigma^2 \geq -2ab \times Cov(X, Y).$$

For $X = X_t$ and $Y = X_{t+k}$ we have:

$$-2ab\rho(k) \leq a^2 + b^2.$$

Therefore for a (weak) stationary process we have $|\rho_k| \leq 1$. We could apply this criterion to check the stationarity of various processes.

### 2.2.1 General moving average process

An $MA(q)$ process is defined to be

$$X_t = a_0 e_t + a_1 e_{t-1} + \cdots + a_q e_{t-q}.$$

We assume $e_i$ are independent and $\mathbb{E}[e_i] = 0$, $Var(e_i) = \sigma^2$ therefore $\mathbb{E}[X_t] = 0$ and $Var(X_t) = \sigma^2 \sum a_i^2$. Now we wish to compute $Cov(X_t, X_{t+k})$. We have

$$Cov(X_t, X_{t+k}) = Cov(a_0 e_t + \cdots + a_q e_{t-q}, a_0 e_{t+k} + \cdots + a_q e_{t-q+k}).$$

Apparently when $k > q$ there is no overlap between the two terms therefore the ACVF vanishes. For $k \leq q$ we have

$$Cov(X_t, X_{t+k}) = \sigma^2 \sum_{i=0}^{q-k} a_i a_{i+k}.$$

We see that $Cov(X_t, X_{t+k})$ does not depend on time either therefore we conclude that an $MA(q)$ process is stationary. Finally it is easy to show that

$$\rho(k) = \frac{\sum_{i=0}^{q-k} a_i a_{i+k}}{\sum_{i=0}^{q} a_i^2}.$$

## 2.3 Backward shift operator and invertibility

We define the backward shift operator to be $B$ such that $BX_t = X_{t-1}$. Now we could write various processes in terms of backward shift operator.

For a random walk, we have $X_t = X_{t-1} + e_t$, using the backward shift operator we have

$$(1 - B)X_t = e_t.$$

For an $MA(2)$ process $X_t = e_t + 0.2e_{t-1} + 0.1e_{t-2}$ we have

$$X_t = (1 + 0.2B + 0.1B^2)e_t.$$

In general for an $AR(p)$ process we have

$$\phi(B)X_t = e_t$$

while for an $MA(q)$ process we have

$$X_t = \psi(B)e_t.$$

### 2.3.1 Invertibility

Now we consider the following two models:

$$X_t = e_t + 2e_{t-1}$$

and

$$X_t = e_t + \frac{1}{2}e_{t-1}.$$

The ACVF of the first model is

$$\begin{aligned} \gamma(0) &= 5\sigma^2, \\ \gamma(1) &= 2\sigma^2 \end{aligned}$$

while the ACVF of the latter is:

$$\begin{aligned} \gamma(0) &= \frac{5}{4}\sigma^2, \\ \gamma(1) &= \frac{1}{2}\sigma^2. \end{aligned}$$

Though the ACVF of the two models are different, it is easy to see that the ACF are the same as follows:

$$\begin{aligned} \rho(0) &= 1, \\ \rho(1) &= \frac{2}{5}. \end{aligned}$$

That means for the same set of ACF there can be multiple processes.

Let us consider an $MA(1)$ process

$$X_t = e_t + \alpha e_{t-1}.$$

We have:

$$e_t = X_t - \alpha e_{t-1} = X_t - \alpha(X_{t-1} - \alpha e_{t-2}) = X_t - \alpha X_{t-1} + \alpha^2 e_{t-2}.$$

In this manner we have:

$$e_t = X_t - \alpha X_{t-1} + \alpha^2 X_{t-2} - \alpha^3 X_{t-3} + \cdots.$$

Therefore

$$X_t = e_t + \alpha X_{t-1} - \alpha^2 X_{t-2} + \alpha^3 X_{t-3} + \cdots$$

which is an $AR(\infty)$ process.

In general, for an $MA(q)$ process we have

$$X_t = \psi(B)e_t$$

therefore formally we have

$$\psi(B)^{-1}X_t = e_t.$$

In the special case there $\psi(B) = 1 + \alpha B$ we have

$$\psi(B)^{-1} = \frac{1}{1+\alpha B} = 1 - \alpha B + \alpha^2 B^2 - \cdots.$$

Therefore

$$e_t = \sum (-\alpha)^n X_{t-n}.$$

For the RHS to be convergent, we need $|\alpha| < 1$.

**Definition** For a stochastic process $\{X_t\}$ with random disturbances $\{e_t\}$ and the following relation

$$e_t = \sum \pi_k X_{t-k},$$

$\{X_t\}$ is invertible iff $\sum |\pi_k|$ is convergent.

Therefore, the model $X_t = e_t + 2e_{t-1}$ is not invertible as $\sum 2^k$ diverges while the model $X_t = e_t + \frac{1}{2}e_{t-1}$ is invertible as $\sum 2^{-k}$ is a convergent geometric series.

We conclude that, by imposing invertibility, a set of ACF uniquely determines an $MA$ process.

### 2.3.2 Condition for invertibility of $MA(q)$

For a general $MA(q)$ process we have

$$X_t = \psi(B)e_t$$

hence

$$\psi(B)^{-1}X_t = e_t.$$

Suppose that $\psi(B) = \alpha_0 + \alpha_1 B + \alpha_2 B^2 + \cdots \alpha_q B^q$ we have

$$\psi(B)^{-1} = \frac{1}{\alpha_0 + \alpha_1 B + \alpha_2 B^2 + \cdots \alpha_q B^q}.$$

Consider the auxiliary function $\psi(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots \alpha_q x^q$. We have

$$\psi(x) = \alpha_q (x - x_1) \cdots (x - x_q).$$

Therefore we have

$$\frac{1}{\psi(x)} = \frac{\delta_1}{1 - \frac{1}{x_1}x} + \frac{\delta_2}{1 - \frac{1}{x_2}x} + \cdot + \frac{\delta_q}{1 - \frac{1}{x_q}x}.$$

Therefore

$$\frac{1}{\psi(B)} = \sum \left( \delta_1 \left(\frac{1}{x_1}\right)^k + \delta_2 \left(\frac{1}{x_2}\right)^k + \cdots + \delta_q \left(\frac{1}{x_q}\right)^k \right) B^k$$

hence

$$\pi_k = \delta_1 \left(\frac{1}{x_1}\right)^k + \delta_2 \left(\frac{1}{x_2}\right)^k + \cdots + \delta_q \left(\frac{1}{x_q}\right)^k.$$

For $\sum |\pi_k|$ to be convergent we must require each of $\sum \delta_i \left(\frac{1}{x_i}\right)^k$ to be convergent hence we must have $|x_i| > 1, \ \forall i$. Therefore we conclude that for an $MA(q)$ process to be invertible all roots of the auxiliary function $\psi(x)$ must lie outside the unit circle.

## 2.4 Stationarity condition for autoregressive process and duality

Consider the $AR(1)$ process

$$X_t = \alpha_1 X_{t-1} + e_t.$$

We have $e_t = (1-\alpha_1 B)X_t$ therefore $X_t = (1-\alpha B)^{-1} e_t = (1+\alpha B+\alpha^2 B^2+\cdots)e_t$ which is an $MA(\infty)$ process. Apparently $\mathbb{E}[X_t] = 0$ and when $|\alpha| < 1$ we have $Vat(X_t) = \sigma^2/(1-\alpha^2)$ while when $|\alpha| \geq 1$, $Var(X_t)$ diverges. It is also not hard to show that $\gamma(k) = \alpha^k \sigma^2/(1-\alpha^2)$ and $\rho(k) = \alpha^k$ the convergence of which both require $|\alpha| < 1$.

Now we consider a general $AR(p)$ process. We have $X_t = \phi(B)^{-1} e_t$ where $\phi(B) = 1 - \alpha_1 B - \cdots \alpha_p B^p$. Consider the auxiliary function

$$\phi(x) = 1 - \alpha_1 x - \cdots - \alpha_p x^p.$$

We have

$$\phi(x) = -\alpha_p (x - x_1) \cdots (x - x_p)$$

hence

$$\phi(x)^{-1} = \frac{\delta_1}{1 - \frac{1}{x_1}x} + \frac{\delta_2}{1 - \frac{1}{x_2}x} + \cdot + \frac{\delta_q}{1 - \frac{1}{x_q}x}.$$

Therefore we have

$$X_t = \sum \left( \delta_1 \left(\frac{1}{x_1}\right)^k + \delta_2 \left(\frac{1}{x_2}\right)^k + \cdots + \delta_q \left(\frac{1}{x_q}\right)^k \right) B^k e_t.$$

For simplicity we let

$$\pi_k = \delta_1 \left(\frac{1}{x_1}\right)^k + \delta_2 \left(\frac{1}{x_2}\right)^k + \cdots + \delta_q \left(\frac{1}{x_q}\right)^k.$$

Therefore

$$X_t = (\pi_0 + \pi_1 B + \pi_2 B^2 + \cdots)e_t.$$

Hence $\mathbb{E}[X_t] = 0$ and

$$Var(X_t) = \sigma^2(\pi_0^2 + \pi_1^2 + \cdots)$$

and

$$\begin{aligned}\gamma(k) =& Var\left(\left(\sum \pi_i B^i\right) e_t, \left(\sum \pi_i B^i\right) e_{t-k}\right)\\ =& Var\left(\pi_k e_{t-k} + \pi_{k+1} e_{t-k-1} + \cdots, \pi_0 e_{t-k} + \pi_1 e_{t-k-1} + \cdots\right)\\ =& \sigma^2 \sum \pi_i \pi_{k+i}.\end{aligned}$$

For $Var(X_t)$ and $\gamma(k)$ to be well-defined we have to require that each $\sum \delta_i \left(\frac{1}{x_i}\right)^k$ converge therefore $|x_i| > 1$.

Therefore we conclude that for an $AR(p)$ process to be stationary, the roots of $\phi(x) = 1 - \alpha_1 x - \cdots - \alpha_p x^p$ must all lie outside the unit circle.

**Duality** Now that we have learnt an important fact:

- With invertibility condition: $MA(q)$ is $AR(\infty)$
- With stationarity condition: $AR(p)$ is $MA(\infty)$.

It is worth mentioning the duality condition is nothing but the condition that a polynomial function $f(x)$ has a well-defined reciprocal $1/f(x)$ as a power series expansion of $x$.

## 2.5 Difference equations

For a $k^{th}$ order difference equation

$$a_n = \alpha_1 a_{n-1} + \cdots \alpha_k a_{n-k}$$

whose characteristic equation is

$$\lambda^k - \alpha_1 \lambda^{k-1} - \cdots - \alpha_k = 0,$$

we look for roots of the auxiliary equation $\lambda_1, \lambda_2, \ldots, \lambda_k$, then we have:

$$a_n = c_1 \lambda_1^n + c_2 \lambda_2^n + \cdots + c_k \lambda_k^n$$

where $c_i$ are determined by the initial conditions.

### 2.5.1 Yule-Walker equations

Now we introduce Yule-Walker equations to calculate the ACF of a stationary $AR(p)$ process. For an $AR(p)$ process we have

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + e_t.$$

Multiply the above equation by $X_{t-k}$ we have

$$\gamma(k) = \alpha_1 \gamma(k-1) + \alpha_2 \gamma(k-2) + \cdots + \alpha_p \gamma(k-p)$$

hence

$$\rho(k) = \alpha_1 \rho(k-1) + \alpha_2 \rho(k-2) + \cdots + \alpha_p \rho(k-p)$$

which is a difference equation hence can be solved using the characteristic equation. The free parameters can be thus determined by the initial conditions.

For example, let us consider the $AR(2)$ model

$$X_t = \frac{1}{6} X_{t-1} + \frac{1}{6} X_{t-2} + e_t.$$

We have:

$$\phi(x) = 1 - \frac{1}{6}x - \frac{1}{6}x^2 \sim (x+3)(x-2)$$

both roots of which lie outside the unit circle therefore this process is stationary.

The Yule-Walker equation is

$$\rho(k) = \frac{1}{6}\rho(k-1) + \frac{1}{6}\rho(k-2)$$

whose characteristic equation is

$$x^2 - \frac{1}{6}x - \frac{1}{6} = 0$$

with the two roots $x_1 = -\frac{1}{3}, x_2 = \frac{1}{2}$. Therefore we have

$$\rho(k) = c_1 \left(-\frac{1}{3}\right)^k + c_2 \left(\frac{1}{2}\right)^k.$$

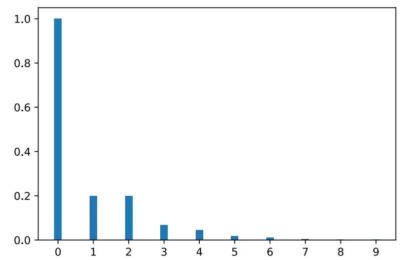
With the following two constraints:

$$\rho(0) = 1 \Rightarrow c_1 + c_2 = 1$$
$$\rho(k) = \rho(-k) \Rightarrow 5\rho(1) = \rho(0)$$

we have

$$\rho(k) = \frac{9}{25}\left(-\frac{1}{3}\right)^k + \frac{16}{25}\left(\frac{1}{2}\right)^k.$$

The ACF is plotted in the following graph




It is also illuminating to look at an $AR(1)$ process $X_t = \alpha X_{t-1} + e_t$. From the Yule-Walker equation we have $\rho(k) = \alpha\rho(k-1)$ therefore it is easy to obtain $\rho(k) = \alpha^k$. We could also compute it the hard way via the recursive relation

$$X_t = \alpha(\alpha X_{t-2} + e_{t-1}) + e_t = e_t + \alpha e_{t-1} + \alpha^2 e_{t-2} + \cdots$$

hence $\gamma(k) = Cov(X_t, X_{t-k}) = \alpha^k\sigma^2$. Since $\gamma(0) = Cov(X_t, X_t) = \sigma^2$ we have $\rho(k) = \gamma(k)/\gamma(0) = \alpha^k$.

### 2.6 Partial autocorrelation function (PACF)

The use of PACF, or more general a partial correlation function is to quantify the correlation between two random variables with the effect of other variables controlled. The most general form of partial correlation function is

$$\begin{aligned} Cov(X, Y|Z) &= Cov(X|Z, Y|Z) \\ &= \mathbb{E}[(X|Z - \mu_{X|Z})(Y|Z - \mu_{Y|Z})] \\ &= \mathbb{E}[XY|Z] - \mu_{X|Z}\mu_{Y|Z} \end{aligned}$$

More specifically, suppose that random variables $X_1, X_2, \ldots, X_n$ have a joint distribution and let $\overline{X}_{1|3,\ldots,n}$ and $\overline{X}_{2|3,\ldots,n}$ be the best linear approximations of the variables $X_1$ and $X_2$ using the variables $X_3, \ldots, X_n$. Then the partial correlation coefficient between $X_1$ and $X_2$ is defined to be:

$$\phi_{12} = \frac{Cov(Y_1, Y_2)}{\sqrt{Var(Y_1)Var(Y_2)}}$$

where $Y_i = X_i - \overline{X}_{i|3,\ldots,n}$ for $i = 1, 2$.

## 3 Estimate the parameters of model

### 3.1 $AR(p)$ model and Yule-Walker equations

Consider an $AR(p)$ process with $\mu = 0$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + e_t$$

where $e_t \sim N(0, \sigma^2)$. The Yule-Walker equation for this model is

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + \cdots + \phi_p \rho(k-p).$$

The autocorrelation function obeys the following set of equations

$$\begin{aligned}
&\rho(1) = \phi_1 \rho(0) + \phi_2 \rho(-1) + \cdots + \phi_p \rho(1-p), \\
&\rho(2) = \phi_1 \rho(1) + \phi_2 \rho(0) + \cdots + \phi_p \rho(2-p), \\
&\cdots \\
&\rho(p) = \phi_1 \rho(p-1) + \phi_2 \rho(p-2) + \cdots + \phi_p \rho(0).
\end{aligned}$$

Recall that $\rho$ is an even function and $\rho(0) = 0$ we have the following matrix form

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ \cdots \\ \rho(p) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(p-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(p-2) \\ \cdots & & & & \\ \rho(p-1) & \rho(p-2) & \rho(p-3) & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \cdots \\ \phi_p \end{pmatrix}.$$

More compactly we could write the above as $\mathrm{P} = R\Phi$. We estimate $r_k \sim \rho(k)$ therefore we have $\mathbf{r} = R\Phi$.

For example let us estimate the parameters of the model $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + e_t$. Given a time series we could estimate its ACF then plug into the equation

$$\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} 1 & r_1 \\ r_1 & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$$

to compute $\phi_i$. We could also estimate $\sigma^2$ of $e_t$ as follows. We have

$$Var(X_t) = \phi_1^2 Var(X_t - 1) + \phi_2^2 Var(X_{t-2}) + 2\phi_1\phi_2 Cov(X_{t-1}, X_{t-2}) + \sigma^2.$$

So that we have

$$\sigma^2 = \gamma(0)\left(1 - \phi_1^2 - \phi_2^2 - 2\phi_1\phi_2\rho(1)\right).$$

Since $\rho(1) = \phi_1 + \rho(1)\phi_2$ and $\rho(2) = \phi_1\rho(1) + \phi_2$, we have

$$1 - \phi_1^2 - \phi_2^2 - 2\phi_1\phi_2\rho(1) = 1 - \phi_1\rho(1) - \phi_2\rho(2).$$

Therefore we estimate $\sigma^2$ to be

$$\sigma^2 = c_0\left(1 - \phi_1 r_1 - \phi_2 r_2\right).$$

For an $AR(p)$ model we could estimate $\sigma^2$ to be

$$\sigma^2 = c_0 \sum_{i=1}^{p} \phi_i r_i.$$

# 4 $ARMA$ and $ARIMA$ model

Combining an $MA(q)$ process and an $AR(p)$ process we have schematically

$$X_t = Noise + AR(p) + MA(q).$$

More precisely we have

$$X_t = e_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i e_{t-i}.$$

This is called an $ARMA(p,q)$ process. For an $ARMA(p,q)$ process we have write it in the following form using the backward shift operator

$$\phi(B)X_t = \theta(B)e_t.$$

Therefore an $ARMA(p,q)$ process can either be understood as an $MA(\infty)$ process

$$X_t = \frac{\theta(B)}{\phi(B)} e_t$$

or an $AR(\infty)$ process

$$\frac{\phi(B)}{\theta(B)} X_t = e_t.$$

Let us consider an $ARMA(1,1)$ process as a concrete example. Suppose we have

$$(1-\alpha B)X_t = (1-\beta B)e_t$$

with $|\alpha| < 1$ and $|\beta| < 1$. Put it in an $AR(\infty)$ process we have

$$(1-\alpha B)(1+\beta B + \beta^2 B^2 + \beta^3 B^3 + \cdots)X_t = e_t.$$

It is not hard to see that the above equation can be expanded to

$$\left(1 + \sum_{i=1} (\beta - \alpha)\alpha^{i-1} B^i\right) X_t = e_t$$

We could then truncate the above expression at a suitable $i$ and apply Yule-Walker equation to compute the $ACF$ and $PACF$ of the process.

Note that $ARMA(p,q)$ process is stationary and invertible, what if the data is non-stationary? In that case we will relax the constraint that the roots of $\phi(B)$ and $\theta(B)$ lie outside the unit circle. Therefore effectively the backward shift operator part can be written as

$$\pi(B) = \widetilde{\pi}(B)(1-B)^d$$

where the roots of $\tilde{\pi}(B)$ lie outside the unit circle. Immediately we recognize that $(1-B)^d$ is the $d^{th}$ order differencing operator. We define

$$Y_t = (1-B)^d X_t.$$

Therefore if $Y_t \sim ARMA(p,q)$ is equivalent to $X_t \sim ARIMA(p,d,q)$.

## 5 $SARIMA$ model and prediction