Chapter 2

Real-Valued discrete time stationary processes

Denote the process by \{X_t\}. For fixed \(t\), \(X_t\) is a random variable (r.v.), and hence there is an associated cumulative probability distribution function (cdf):

\[ F_t(a) = P(X_t \leq a), \]

and

\[
E\{X_t\} = \int_{-\infty}^{\infty} x \, dF_t(x) \\
\text{var}\{X_t\} = \int_{-\infty}^{\infty} (x - \mu_t)^2 \, dF_t(x).
\]

But we are interested in the relationships between the various r.v.s that form the process. For example, for any \(t_1\) and \(t_2 \in T\),

\[ F_{t_1, t_2}(a_1, a_2) = P(X_{t_1} \leq a_1, X_{t_2} \leq a_2) \]

gives the bivariate cdf. More generally for any \(t_1, t_2, \ldots, t_n \in T\),

\[ F_{t_1, t_2, \ldots, t_n}(a_1, a_2, \ldots, a_n) = P(X_{t_1} \leq a_1, \ldots, X_{t_n} \leq a_n) \]

**Stationarity**

The class of all stochastic processes is too large to work with in practice. We consider only the subclass of stationary processes.

**COMPLETE/STRONG/STRICT stationarity**

\{X_t\} is said to be completely stationary if, for all \(n \geq 1\), for any \(t_1, t_2, \ldots, t_n \in T\),
and for any \( \tau \) such that \( t_1 + \tau, t_2 + \tau, \ldots, t_n + \tau \in T \) are also contained in the index set, the joint cdf of \( \{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\} \) is the same as that of \( \{X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau}\} \) i.e.,

\[
F_{t_1,t_2,\ldots,t_n}(a_1, a_2, \ldots, a_n) = F_{t_1+\tau,t_2+\tau,\ldots,t_n+\tau}(a_1, a_2, \ldots, a_n),
\]

so that the probabilistic structure of a completely stationary process is invariant under a shift in time.

**SECOND-ORDER/WEAK/COVARIANCE stationarity**

\( \{X_t\} \) is said to be second-order stationary if, for all \( n \geq 1 \), for any \( t_1, t_2, \ldots, t_n \in T \), and for any \( \tau \) such that \( t_1 + \tau, t_2 + \tau, \ldots, t_n + \tau \in T \) are also contained in the index set, all the joint moments of orders 1 and 2 of \( \{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\} \) exist, are finite, and equal to the corresponding joint moments of \( \{X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau}\} \). Hence,

\[
E\{X_t\} \equiv \mu \quad ; \quad \text{var}\{X_t\} \equiv \sigma^2 \quad (= E\{X_t^2\} - \mu^2),
\]

are constants independent of \( t \). If we let \( \tau = -t_1 \),

\[
E\{X_{t_1}X_{t_2}\} = E\{X_{t_1+\tau}X_{t_2+\tau}\}
\]

\[
= E\{X_0X_{t_2-t_1}\},
\]

and with \( \tau = -t_2 \),

\[
E\{X_{t_1}X_{t_2}\} = E\{X_{t_1+\tau}X_{t_2+\tau}\}
\]

\[
= E\{X_{t_1-t_2}X_0\}.
\]

Hence, \( E\{X_{t_1}X_{t_2}\} \) is a function of the absolute difference \( |t_2 - t_1| \) only, similarly, for the covariance between \( X_{t_1} \) & \( X_{t_2} \):

\[
\text{cov}\{X_{t_1}, X_{t_2}\} = E\{(X_{t_1} - \mu)(X_{t_2} - \mu)\} = E\{X_{t_1}X_{t_2}\} - \mu^2.
\]

For a discrete time second-order stationary process \( \{X_t\} \) we define the autocovariance sequence (acvs) by

\[
s_\tau \equiv \text{cov}\{X_t, X_{t+\tau}\} = \text{cov}\{X_0, X_\tau\}.
\]

Note,

1. \( \tau \) is called the lag.
2. \( s_0 = \sigma^2 \) and \( s_{-r} = s_r \).

3. The autocorrelation sequence (acs) is given by

\[
\rho_r = \frac{s_r}{s_0} = \frac{\text{cov}\{X_t, X_{t+r}\}}{\sigma^2}.
\]

The sample or estimated autocorrelation sequence (acs), \( \{\hat{\rho}_r\} \), for each of our time series are given in Figs. 6 and 7. [We shall see how to compute these in Chapter 4.] Note e.g., that for the Willamette river data \( X_t \) and \( X_{t+6} \) seem to be negatively correlated, while \( X_t \) and \( X_{t+12} \) seem positively correlated (consistent with the river flow varying with a period of roughly 12 months).

4. Since \( \rho_r \) is a correlation coefficient, \( |s_r| \leq s_0 \).

5. The sequence \( \{s_r\} \) is positive semidefinite, i.e., for all \( n \geq 1 \), for any \( t_1, t_2, \ldots, t_n \) contained in the index set, and for any set of nonzero real numbers \( a_1, a_2, \ldots, a_n \)

\[
\sum_{j=1}^{n} \sum_{k=1}^{n} s_{t_j - t_k} a_j a_k \geq 0.
\]

Proof

Let

\[
a = (a_1, a_2, \ldots, a_n)^T, \quad V = (X_{t_1}, X_{t_2}, \ldots, X_{t_n})^T
\]

and let \( \Sigma \) be the variance-covariance matrix of \( V \). Its \( j, k \)-th element is given by \( s_{t_j - t_k} = \text{E}\{(X_{t_j} - \mu)(X_{t_k} - \mu)\} \). Define the r.v.

\[
w = \sum_{j=1}^{n} a_j X_{t_j} = a^T V.
\]

Then

\[
0 \leq \text{var}\{w\} = \text{var}\{a^T V\} = a^T \text{var}\{V\} a = a^T \Sigma a = \sum_{j=1}^{n} \sum_{k=1}^{n} s_{t_j - t_k} a_j a_k.
\]

6. The variance-covariance matrix of equispaced \( X \)'s, \( (X_1, X_2, \ldots, X_N)^T \) has the form

\[
\begin{bmatrix}
  s_0 & s_1 & \ldots & s_{N-2} & s_{N-1} \\
  s_1 & s_0 & \ldots & s_{N-3} & s_{N-2} \\
  \vdots & \ddots & \ddots & \ddots & \ddots \\
  s_{N-2} & s_{N-3} & \ldots & s_0 & s_1 \\
  s_{N-1} & s_{N-2} & \ldots & s_1 & s_0
\end{bmatrix}
\]
which is known as a symmetric Toeplitz matrix – all elements on a diagonal are the same. Note the matrix has only \( N \) unique elements, \( s_0, s_1, \ldots, s_{N-1} \).

7. A stochastic process \( \{X_t\} \) is called Gaussian if, for all \( n \geq 1 \) and for any \( t_1, t_2, \ldots, t_n \) contained in the index set, the joint cdf of \( X_{t_1}, X_{t_2}, \ldots, X_{t_n} \) is multivariate Gaussian.

- 2nd-order stationary \textbf{Gaussian} \( \Rightarrow \) complete stationarity (since MVN completely characterized by 1st and 2nd moments). It is not true in general that 2nd-order stationary \( \Rightarrow \) complete stationarity.
- Complete stationarity \( \Rightarrow \) 2nd-order stationary in general.

**Examples of discrete stationary processes**

[1] \textbf{White noise process}

Also known as a purely random process. Let \( \{X_t\} \) be a sequence of uncorrelated r.v.s such that

\[
E\{X_t\} = \mu, \quad \text{var}\{X_t\} = \sigma^2 \quad \forall t
\]

and

\[
s_\tau = \begin{cases} 
\sigma^2 & \tau = 0 \\
0 & \tau \neq 0
\end{cases}
\]

or

\[
\rho_\tau = \begin{cases} 
1 & \tau = 0 \\
0 & \tau \neq 0
\end{cases}
\]

forms a basic building block in time series analysis. Very different realizations of white noise can be obtained for different distributions of \( \{X_t\} \). Examples are given in Figures 8 and 9 for processes with (a) Gaussian, (b) exponential, (c) uniform and (d) truncated Cauchy distributions.

[2] \textbf{q-th order moving average process} \quad \text{MA}(q)

\( X_t \) can be expressed in the form

\[
X_t = \mu - \theta_{0,q} \epsilon_t - \theta_{1,q} \epsilon_{t-1} - \ldots - \theta_{q,q} \epsilon_{t-q}
\]

\[
= \mu - \sum_{j=0}^{q} \theta_{j,q} \epsilon_{t-j},
\]

where \( \mu \) and \( \theta_{j,q} \)'s are constants \((\theta_{0,q} \equiv -1, \theta_{q,q} \neq 0)\), and \( \{\epsilon_t\} \) is a zero-mean white noise process with variance \( \sigma^2_\epsilon \).
W.l.o.g. assume \( E\{X_t\} = \mu = 0 \).

Then \( \text{cov}\{X_t, X_{t+\tau}\} = E\{X_t X_{t+\tau}\} \).

Recall: \( \text{cov}(X, Y) = E\{(X - E\{X\})(Y - E\{Y\})\} \).

Since \( E\{\epsilon_t\epsilon_{t+\tau}\} = 0 \ \forall \ \tau \neq 0 \) we have for \( \tau \geq 0 \).

\[
\text{cov}\{X_t, X_{t+\tau}\} = \sum_{j=0}^{q} \sum_{k=0}^{q} \theta_{j,q} \theta_{k,q} E\{\epsilon_{t-j}\epsilon_{t+\tau-k}\} \\
= \sigma^2 \sum_{j=0}^{q-\tau} \theta_{j,q} \theta_{j+\tau,q} \ \ (k = j + \tau) \\
\equiv s_\tau,
\]

which does not depend on \( t \). Since \( s_\tau = s_{-\tau} \), \( \{X_t\} \) is a stationary process with

acvs given by

\[
s_\tau = \begin{cases} 
\sigma^2 \sum_{j=0}^{\min(q, |\tau|)} \theta_{j,q} \theta_{j+|\tau|,q} & |\tau| \leq q \\
0 & |\tau| > q
\end{cases}
\]

N.B. No restrictions were placed on the \( \theta_{j,q} \)'s to ensure stationarit, though obviously, \(|\theta_{j,q}| < \infty, j = 1, \ldots, q\).

Examples (see Figures 10 and 11)

\[
X_t = \epsilon_t - \theta_{1,1}\epsilon_{t-1} \quad \text{MA}(1)
\]

\[
\sigma^2 \sum_{j=0}^{1-|\tau|} \theta_{1,1} \theta_{j+|\tau|,1} \quad |\tau| \leq 1,
\]

so,

\[
s_0 = \sigma^2 (\theta_{0,1}\theta_{0,1} + \theta_{1,1}\theta_{1,1}) = \sigma^2 (1 + \theta^2_{1,1});
\]

and,

\[
s_1 = \sigma^2 \theta_{0,1}\theta_{1,1} = -\sigma^2 \theta_{1,1}.
\]

\[
\rho_\tau = \frac{s_\tau}{s_0}.
\]
\[ \rho_0 = 1.0; \quad \rho_1 = \frac{-\theta_{1,1}}{1 + \theta_{1,1}^2}; \quad \rho_2 = \rho_3 = \cdots = 0. \]

(a) \( \theta_{1,1} = 1.0, \sigma^2 = 1.0, \)

we have,

\[ s_0 = 2.0; \quad s_1 = -1.0; \quad s_2 = s_3 = \cdots = 0.0, \]

giving,

\[ \rho_0 = 1.0; \quad \rho_1 = -0.5; \quad \rho_2 = \rho_3 = \cdots = 0.0. \]

(b) \( \theta_{1,1} = -1.0, \sigma^2 = 1.0, \)

we have,

\[ s_0 = 2.0; \quad s_1 = 1.0; \quad s_2 = s_3 = \cdots = 0.0, \]

giving,

\[ \rho_0 = 1.0; \quad \rho_1 = 0.5; \quad \rho_2 = \rho_3 = \cdots = 0.0. \]

Note: if we replace \( \theta_{1,1} \) by \( \theta_{1,1}^{-1} \) the model becomes

\[ X_t = \epsilon_t - \frac{1}{\theta_{1,1}} \epsilon_{t-1} \]

and the autocorrelation becomes

\[ \rho_1 = \frac{-\frac{1}{\theta_{1,1}}}{1 + \left(\frac{1}{\theta_{1,1}}\right)^2} = -\theta_{1,1}, \]

i.e., is unchanged!!!

We cannot identify the MA(1) process uniquely from its autocorrelation!

[3] **p-th order autoregressive process** \( \text{AR}(p) \)

\( \{X_t\} \) is expressed in the form

\[ X_t = \phi_{1,p}X_{t-1} + \phi_{2,p}X_{t-2} + \cdots + \phi_{p,p}X_{t-p} + \epsilon_t, \]

where \( \phi_{1,p}, \phi_{2,p}, \ldots, \phi_{p,p} \) are constants \( (\phi_{p,p} \neq 0) \) and \( \{\epsilon_t\} \) is a zero mean white noise process with variance \( \sigma^2 \). In contrast to the parameters of an MA(\( q \)) process, the \( \{\phi_{k,p}\} \) must satisfy certain conditions for \( \{X_t\} \) to be a stationary process – i.e., not all \( \text{AR}(p) \) processes are stationary (more later).
Examples  (Figures 12 and 13)

\[ X_t = \phi_{1,1} X_{t-1} + \epsilon_t \quad \text{AR}(1) - \text{Markov process} \quad (2.1) \]

\[ = \phi_{1,1} \{ \phi_{1,1} X_{t-2} + \epsilon_{t-1} \} + \epsilon_t \]
\[ = \phi_{1,1}^2 X_{t-2} + \phi_{1,1} \epsilon_{t-1} + \epsilon_t \]
\[ = \phi_{1,1}^3 X_{t-3} + \phi_{1,1}^2 \epsilon_{t-2} + \phi_{1,1} \epsilon_{t-1} + \epsilon_t \]
\[ \vdots \]
\[ = \sum_{k=0}^{\infty} \phi_{1,1}^k \epsilon_{t-k}. \]

Here we take the initial condition \( X_{-N} = 0 \) and let \( N \to \infty \).

Note \( E\{X_t\} = 0 \).

\[ \text{var}\{X_t\} = \text{var}\left\{ \sum_{k=0}^{\infty} \phi_{1,1}^k \epsilon_{t-k} \right\} = \sum_{k=0}^{\infty} \text{var}\{\phi_{1,1}^k \epsilon_{t-k}\} = \sigma^2 \phi_{1,1}^{2k} \]

For \( \text{var}\{X_t\} < \infty \) we must have \( |\phi_{1,1}| < 1 \), in which case

\[ \text{var}\{X_t\} = \frac{\sigma^2}{1 - \phi_{1,1}^2}. \]

To find the form of the acvs, we notice that for \( \tau > 0 \), \( X_{t-\tau} \) is a linear function of \( \epsilon_{t-\tau}, \epsilon_{t-\tau-1}, \ldots \) and is therefore uncorrelated with \( \epsilon_t \). Hence

\[ E\{\epsilon_t X_{t-\tau}\} = 0, \]

so, assuming stationarity and multiplying the defining equation (2.1) by \( X_{t-\tau} \):

\[ X_t X_{t-\tau} = \phi_{1,1} X_t X_{t-\tau} + \epsilon_t X_{t-\tau} \]

\[ \Rightarrow E\{X_t X_{t-\tau}\} = \phi_{1,1} E\{X_{t-1} X_{t-\tau}\} \]

i.e., \( s_\tau = \phi_{1,1} s_{\tau-1} = \phi_{1,1}^2 s_{\tau-2} = \ldots = \phi_{1,1}^\tau s_0 \)

\[ \Rightarrow \rho_\tau = \frac{s_\tau}{s_0} = \phi_{1,1}^\tau. \]

But \( \rho_\tau \) is an even function of \( \tau \), so we obtain an exponentially decaying sequence

\[ \rho_\tau = \phi_{1,1}^{\lfloor \tau \rfloor} \quad \tau = 0, \pm 1, \pm 2, \ldots \]
[4] \((p, q)\)’th order autoregressive-moving average process

ARMA\((p, q)\)

Here \(\{X_t\}\) is expressed as

\[X_t = \phi_{1,p}X_{t-1} + \ldots + \phi_{p,p}X_{t-p} + \epsilon_t - \theta_{1,q}\epsilon_{t-1} - \ldots - \theta_{q,q}\epsilon_{t-q},\]

where the \(\phi_{j,p}\)’s and the \(\theta_{j,q}\)’s are all constants \((\phi_{p,p} \neq 0; \theta_{q,q} \neq 0)\) and again \(\{\epsilon_t\}\) is a zero mean white noise process with variance \(\sigma^2\). The ARMA class is important as many data sets may be approximated in a more parsimonious way (meaning fewer parameters are needed) by a mixed ARMA model than by a pure AR or MA process.

[5] \(p\)’th order autoregressive conditionally heteroscedastic

model

ARCH\((p)\)

Standard linear models were found to be inappropriate for describing the dependence of financial log-return series of stock indices, share prices, exchange rates etc. New multiplicative noise models were developed. One such is the ARCH\((p)\) model.

Assume we have a time series \(\{X_t\}\) that has a variance (volatility) that changes through time,

\[X_t = \sigma_t \epsilon_t\]  \(\text{ (2.2)}\)

where here \(\{\epsilon_t\}\) is a sequence of independent and identically distributed (iid) r.v.s with zero mean and unit variance. (This is stronger than simply uncorrelated). Here, \(\sigma_t\) represents the local conditional standard deviation of the process. \(\{\sigma_t\}\) and \(\{\epsilon_t\}\) are independent.

\(\{X_t\}\) is ARCH\((p)\) if it satisfies equation (2.2) and

\[\sigma_t^2 = \alpha + \beta_{1,p}X_{t-1}^2 + \ldots + \beta_{p,p}X_{t-p}^2,\]  \(\text{ (2.3)}\)

where \(\alpha > 0\) and \(\beta_{j,p} \geq 0, j = 1, \ldots, p\) (to ensure \(\sigma_t^2\) is positive).

Notes:

(a) the absence of an error term in equation (2.3).

(b) unconstrained estimation often leads to violation of the non-negativity constraints that are needed to ensure positive variance.
(c) the modelling of $\sigma_t^2$ can prevent modelling of asymmetry in volatility (i.e. volatility tends to be higher after a decrease than after an equal increase and ARCH cannot account for this).

**Example:** ARCH(1)

$$\sigma_t^2 = \alpha + \beta_{1,1} X_{t-1}^2$$

Define,

$$v_t = X_t^2 - \sigma_t^2, \quad \Rightarrow \quad \sigma_t^2 = X_t^2 - v_t.$$  

So $X_t^2 = \sigma_t^2 + v_t$ and the model can be written as

$$X_t^2 = \alpha + \beta_{1,1} X_{t-1}^2 + v_t,$$

i.e., as an AR(1) model for $\{X_t^2\}$. The errors, $\{v_t\}$, have zero mean, but as $v_t = \sigma_t^2(\epsilon_t^2 - 1)$ the errors are heteroscedastic.

[6] **Harmonic with random amplitude** (see Figures 14 and 14a)

Here $\{X_t\}$ is expressed as

$$X_t = \epsilon_t \cos(2\pi f_0 t + \phi)$$

$f_0$ is a fixed frequency and $\{\epsilon_t\}$ is zero mean white noise with variance $\sigma_t^2$.

**Case (a)** $\phi$ is constant.

$$E\{X_t\} = E\{\epsilon_t \cos(2\pi f_0 t + \phi)\}$$

$$= E\{\epsilon_t\} \cos(2\pi f_0 t + \phi) = 0.$$  

$$\text{var}\{X_t\} = E\{X_t^2\}$$

$$= E\{\epsilon_t^2\} \cos^2(2\pi f_0 t + \phi)$$

$$= \sigma_t^2 \cos^2(2\pi f_0 t + \phi).$$

So the variance depends on $t$ and the process is nonstationary.

**Case (b)** $\phi \sim U[-\pi, \pi]$ and indep. of $\{\epsilon_t\}$.

$$E\{X_t\} = E\{\epsilon_t \cos(2\pi f_0 t + \phi)\} = E\{\epsilon_t\} E\{\cos(2\pi f_0 t + \phi)\} = 0.$$  

15
\[
\text{cov}(X_t, X_{t+\tau}) = E\{X_tX_{t+\tau}\} = E\{\epsilon_t \epsilon_{t+\tau}\} E\{\cos(2\pi f_0 t + \phi) \cos(2\pi f_0 (t + \tau) + \phi)\} \]

Since \(\{\epsilon_t\}\) is white noise we have,

\[
E\{\epsilon_t \epsilon_{t+\tau}\} = \begin{cases} \sigma_\epsilon^2 & \text{if } \tau = 0, \\ 0 & \text{if } \tau \neq 0, \end{cases}
\]

So, for \(\tau \neq 0\), \(\text{cov}(X_t, X_{t+\tau}) = 0\), while for \(\tau = 0\),

\[
\text{cov}(X_t, X_t) = s_0 = \sigma_\epsilon^2 E\{\cos^2(2\pi f_0 t + \phi)\}.
\]

Now,

\[
E\{\cos^2(2\pi f_0 t + \phi)\} = \int_{-\pi}^{\pi} \cos^2(2\pi f_0 t + \phi) \frac{1}{2\pi} \, d\phi
\]

\[
= \frac{1}{2} \int_{-\pi}^{\pi} [1 + \cos(4\pi f_0 t + 2\phi)] \frac{1}{2\pi} \, d\phi
\]

\[
= \frac{1}{2}.
\]

So,

\[
s_0 = \frac{\sigma_\epsilon^2}{2},
\]

and the process is stationary.

The random phase idea is illustrated in Figure 14a: the random point at which data collection started corresponds to breaking-in to the 'sinusoidal-like' behaviour at a random point, which equates to a random phase.

**Trend removal and seasonal adjustment**

There are certain, quite common, situations where the observations exhibit a trend – a tendency to increase or decrease slowly steadily over time – or may fluctuate in a periodic/seasonal manner. The model is modified to

\[
X_t = \mu_t + Y_t
\]

\(\mu_t\) = time dependent mean.

\(Y_t\) = zero mean stationary process.

**Example** CO\textsubscript{2} data
\[ X_t = \text{monthly atmospheric CO}_2 \text{ concentrations expressed in parts per million (ppm)} \]
derived from in situ air samples collected at Mauna Loa observatory, Hawaii.

Monthly data from May 1988 – December 1998, giving \( N = 128 \).

The data is plotted in Figure 15. We can see both a trend and periodic/seasonal effects.

**Trend adjustment**

Represent a simple linear trend by \( \alpha + \beta t \). So take \( X_t = \alpha + \beta t + Y_t \). At least two possible approaches:

(a) Estimate \( \alpha \) and \( \beta \) by least squares, and work with the residuals

\[ \hat{Y}_t = X_t - \hat{\alpha} - \hat{\beta} t. \]

For the CO\(_2\) data these are shown in the middle plot of figure 15.

(b) Take first differences:

\[ X_t^{(1)} = X_t - X_{t-1} = \alpha + \beta t + Y_t - (\alpha + \beta(t - 1) + Y_{t-1}) \]
\[ = \beta + Y_t - Y_{t-1}. \]

For the CO\(_2\) data these are shown in the bottom plot of figure 15.

**Note:** if \( \{ Y_t \} \) is stationary so is \( \{ Y_t^{(1)} \} \)

In the case of linear trend, if we difference again:

\[ X_t^{(2)} = X_t^{(1)} - X_{t-1}^{(1)} = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \]
\[ = (\beta + Y_t - Y_{t-1}) - (\beta + Y_{t-1} - Y_{t-2}) \]
\[ = Y_t - 2Y_{t-1} + Y_{t-2}, \quad (\equiv Y_t^{(1)} - Y_{t-1}^{(1)} = Y_t^{(2)}), \]

so that the effect of \( \mu_t(= \alpha + \beta t) \) has been completely removed.

If \( \mu_t \) is a polynomial of degree \( (d - 1) \) in \( t \), then \( d \)th differences of \( \mu_t \) will be zero (\( d = 2 \) for linear trend). Further,

\[ X_t^{(d)} = \sum_{k=0}^{d} \binom{d}{k} (-1)^k X_{t-k} \]
\[ = \sum_{k=0}^{d} \binom{d}{k} (-1)^k Y_{t-k}. \]
There are other ways of writing this. Define the difference operator

$$\Delta = (1 - B)$$

where $BX_t = X_{t-1}$ is the backward shift operator (sometimes known as the lag operator $L$ – especially in econometrics). Then,

$$X_t^{(d)} = \Delta^d X_t = \Delta^d Y_t.$$

For example, for $d = 2$:

$$X_t^{(2)} = (1 - B)^2 X_t = (1 - B)(X_t - X_{t-1})$$
$$= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})$$
$$= (\beta + Y_t - Y_{t-1}) - (\beta + Y_{t-1} - Y_{t-2})$$
$$= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$$
$$= (1 - B)^2 Y_t = \Delta^2 Y_t.$$

This notation can be incorporated into the ARMA set up. Recall if $\{X_t\}$ is ARMA$(p,q)$,

$$X_t = \phi_{1,p}X_{t-1} + \ldots + \phi_{p,p}X_{t-p} + \epsilon_t - \theta_{1,q}\epsilon_{t-1} - \ldots - \theta_{q,q}\epsilon_{t-q},$$

$$X_t - \phi_{1,p}X_{t-1} - \ldots - \phi_{p,p}X_{t-p} = \epsilon_t - \theta_{1,q}\epsilon_{t-1} - \ldots - \theta_{q,q}\epsilon_{t-q}$$

$$(1 - \phi_{1,p}B - \phi_{2,p}B^2 - \ldots - \phi_{p,p}B^p)X_t = (1 - \theta_{1,q}B - \theta_{2,q}B^2 - \ldots - \theta_{q,q}B^q)\epsilon_t$$

$$\Phi(B)X_t = \Theta(B)\epsilon_t.$$ 

Here

$$\Phi(B) = 1 - \phi_{1,p}B - \phi_{2,p}B^2 - \ldots - \phi_{p,p}B^p$$

and

$$\Theta(B) = 1 - \theta_{1,q}B - \theta_{2,q}B^2 - \ldots - \theta_{q,q}B^q$$

are known as the associated or characteristic polynomials.

Further, we can generalize the class of ARMA models to include differencing to account for certain types of non-stationarity, namely, $X_t$ is called ARIMA$(p,d,q)$ if

$$\Phi(B)(1 - B)^d X_t = \Theta(B)\epsilon_t,$$

or

$$\Phi(B)\Delta^d X_t = \Theta(B)\epsilon_t.$$
Seasonal adjustment

The model is

\[ X_t = \nu_t + Y_t \]

where

\[ \nu_t = \text{seasonal component}, \]

\[ Y_t = \text{zero mean stationary process}. \]

Presuming that the seasonal component maintains a constant pattern over time with period \( s \), there are again several approaches to removing \( \nu_t \). A popular approach used by Box & Jenkins is to use the operator \((1 - B^s)\):

\[
X_t^{(s)} = (1 - B^s)X_t = X_t - X_{t-s} \\
= (\nu_t + Y_t) - (\nu_{t-s} + Y_{t-s}) \\
= Y_t - Y_{t-s}
\]

since \( \nu_t \) has period \( s \) (and so \( \nu_{t-s} = \nu_t \)).

Figure 16 shows this technique applied to the CO\(_2\) data – most of the seasonal structure and trend has been removed by applying the seasonal operator after the differencing operator:

\[ (1 - B^{12})(1 - B)X_t. \]

The General Linear Process

Consider a process of the form

\[ X_t = \sum_{k=-\infty}^{\infty} g_k \epsilon_{t-k}, \]

where \( \{\epsilon_t\} \) is a purely random process, and \( \{g_k\} \) is a given sequence of real-valued constants satisfying \( \sum_{k=-\infty}^{\infty} g_k^2 < \infty \), which ensures that \( \{X_t\} \) has finite variance.

Now we know \(|\rho_r| \leq 1\), so

\[
|s_r| = |\text{cov}\{X_t, X_{t+r}\}| \leq \sigma_X^2 = \sigma_X^2 \sum_k g_k^2 < \infty,
\]
so the covariance is bounded also. If
\[ g_{-1}, g_{-2}, \ldots = 0, \]
then we obtain what is called the General Linear Process
\[ X_t = \sum_{k=0}^{\infty} g_k \epsilon_{t-k}, \]
where \( X_t \) depends only on present and past values \( \epsilon_t, \epsilon_{t-2}, \epsilon_{t-2}, \ldots \) of the purely random process.

Introduce the “\( z \)-polynomial”
\[ G(z) = \sum_{k=0}^{\infty} g_k z^k, \]
where \( z \in \mathbb{C} \). Note \( X_t = G(B) \epsilon_t \).

Then
\[ G(z) = \frac{G_1(z)}{G_2(z)}, \quad \text{say.} \]
Call the roots of \( G_2(z) \) (the “poles” of \( G(z) \)) in the complex plane \( z_1, z_2, \ldots, z_p \), where the zeros are ordered so that \( z_1, \ldots, z_k \) are inside and \( z_{k+1}, \ldots, z_p \) are outside the unit circle \( |z| = 1 \). Then,
\[
\frac{1}{G_2(z)} = \sum_{j=1}^{p} \frac{A_j}{z - z_j} = \sum_{j=1}^{k} \frac{A_j}{z} \times \frac{1}{1 - \frac{z}{z_j}} + \sum_{j=k+1}^{p} \frac{A_j}{z_j} \times \frac{-1}{1 - \frac{z}{z_j}} \\
= \sum_{j=1}^{k} \frac{A_j}{z} \sum_{l=0}^{\infty} \left( \frac{z_j}{z} \right)^l - \sum_{j=k+1}^{p} \frac{A_j}{z_j} \sum_{l=0}^{\infty} \left( \frac{z}{z_j} \right)^l
\]
which is uniformly convergent for \( |z| = 1 \). Replace \( z \) by the backshift operator \( B \) and apply to \( \{\epsilon_t\} \):
\[
\left\{ \frac{1}{G_2(B)} \right\} \epsilon_t = \left\{ \sum_{j=1}^{k} A_j B^{-l} \sum_{l=0}^{\infty} z_j^l B^{-l} - \sum_{j=k+1}^{p} A_j z_j^{-l} \sum_{l=0}^{\infty} z_j^{-l} B^l \right\} \epsilon_t \\
= \sum_{j=1}^{k} A_j \sum_{l=0}^{\infty} z_j^l \epsilon_{t+l+1} - \sum_{j=k+1}^{p} A_j \sum_{l=0}^{\infty} z_j^{-l-1} \epsilon_{t-l}.\]
Hence, if all the roots of \( G_2(z) \) are outside the unit circle (i.e. all the poles of \( G(z) \) are outside the unit circle) only past and present values of \( \{\epsilon_t\} \) are involved and the General Linear Process exists.
Another way of stating this is that

\[ G(z) < \infty \quad |z| \leq 1 \]

i.e., \( G(z) \) is analytic inside and on the unit circle.
So, all the

poles of \( G(z) \) lie outside the unit circle
roots (zeros) of \( G^{-1}(z) \) lie outside the unit circle

Consider the MA(\( q \)) model

\[ X_t = \Theta(B) \epsilon_t, \]

then,

\[ \Theta^{-1}(B)X_t = \epsilon_t \]

and in general, the expansion of \( \Theta^{-1}(B) \) is a polynomial of infinite order. Similarly,
consider the AR(\( p \)) model

\[ \Phi(B)X_t = \epsilon_t, \]

then,

\[ X_t = \Phi^{-1}(B) \epsilon_t \]

Hence,

\[ \text{MA (finite order)} \equiv \text{AR (infinite order)} \]

\[ \text{AR (finite order)} \equiv \text{MA (infinite order)} \]

provided the infinite order expansions exist!

**Invertibility**

Consider inverting the general linear process into autoregressive form

\[ X_t = \sum_{k=0}^{\infty} g_k \epsilon_{t-k} \]

\[ = \sum_{k=0}^{\infty} g_k B^k \epsilon_t \]

\[ X_t = G(B) \epsilon_t \]

\[ \Rightarrow G^{-1}(B)X_t = \epsilon_t \]
The expansion of $G^{-1}(B)$ in powers of $B$ gives the required autoregressive form provided $G^{-1}(B)$ admits a power series expansion

$$G^{-1}(z) = \sum_{k=0}^{\infty} h_k z^k$$

i.e. if $G^{-1}(z)$ is analytic, $|z| \leq 1$. Thus the model is invertible if

$$G^{-1}(z) < \infty, \quad |z| \leq 1.$$ 

⇒ All the poles of $G^{-1}(z)$ are outside the unit circle.

For the MA($q$) process, $G(z) = \Theta(z)$, and so the invertibility condition is that $\Theta(z)$ has no roots inside or on the unit circle; i.e. all the roots of $\Theta(z)$ lie outside the unit circle.

**Example**

Consider the following process

$$X_t = \epsilon_t - 1.3\epsilon_{t-1} + 0.4\epsilon_{t-2}$$

Writing this in $B$ notation:

$$X_t = (1 - 1.3B + 0.4B^2)\epsilon_t = \Theta(B)\epsilon_t$$

to check if invertible, find roots of $\Theta(z) = 1 - 1.3z + 0.4z^2$,

$$1 - 1.3z + 0.4z^2 = 0$$
$$4z^2 - 13z + 10 = 0$$
$$(4z - 5)(z - 2) = 0$$

roots of $\Theta(z)$ are $z = 2$ and $z = 5/4$, which are both outside the unit circle ⇒ invertible.

**Stationarity**

For the AR($p$) process

$$\Phi(B)X_t = \epsilon_t$$

⇒ $X_t = \Phi^{-1}(B)\epsilon_t = G(B)\epsilon_t$,
so that \( G(z) = \Phi^{-1}(z) \). Thus the model is stationary if

\[ G(z) < \infty, \quad |z| \leq 1. \]

⇒ All the poles of \( G(z) \) are outside the unit circle.

Hence the requirement for stationarity is that all the roots of \( G^{-1}(z) = \Phi(z) \) must lie outside the unit circle.

For the MA\( (q) \) process

\[ X_t = \Theta(B)\epsilon_t = G(B)\epsilon_t \]

and since \( G(B) = \Theta(B) \) is a polynomial of finite order \( G(z) < \infty, |z| \leq 1 \), automatically.

### SUMMARY

<table>
<thead>
<tr>
<th></th>
<th>AR((p))</th>
<th>MA((q))</th>
<th>ARMA((p,q))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stationarity</td>
<td>Roots of ( \Phi(z) ) outside (</td>
<td>z</td>
<td>\leq 1 )</td>
</tr>
<tr>
<td>Invertibility</td>
<td>Always invertible</td>
<td>Roots of ( \Theta(z) ) outside (</td>
<td>z</td>
</tr>
</tbody>
</table>

### Example

Determine whether the following model is stationary and/or invertible,

\[ X_t = 1.3X_{t-1} - 0.4X_{t-2} + \epsilon_t - 1.5\epsilon_{t-1}. \]

Writing in \( B \) notation:

\[ (1 - 1.3B + 0.4B^2)X_t = (1 - 1.5B)\epsilon_t \]

we have

\[ \Phi(z) = 1 - 1.3z + 0.4z^2 \]

with roots \( z = 2 \) and \( 5/4 \) (from previous example), so the roots of \( \Phi(z) = 0 \) both lie outside the unit circle, therefore model is stationary, and

\[ \Theta(z) = 1 - 1.5z, \]
so the root of $\Theta(z) = 0$ is given by $z = 2/3$ which lies inside the unit circle and the model is not invertible.

**Directionality and Reversibility**

Consider again the general linear model

$$X_t = \sum_{k=0}^{\infty} g_k \epsilon_{t-k}$$

$$= \sum_{k=0}^{\infty} g_k B^k \epsilon_t$$

$$= G(B) \epsilon_t$$

The reversed form is clearly,

$$X_t = \sum_{k=0}^{\infty} g_k \epsilon_{t+k}$$

$$= \sum_{k=0}^{\infty} g_k B^{-k} \epsilon_t$$

$$= G \left( \frac{1}{B} \right) \epsilon_t,$$

with some stationarity condition.

Now consider the ARMA($p, q$) model given by

$$\Phi(B)X_t = \Theta(B)\epsilon_t,$$

where,

$$\Phi(B) = 1 - \phi_{1,p}B - \phi_{2,p}B^2 - \ldots - \phi_{p,p}B^p$$

$$\Theta(B) = 1 - \theta_{1,q}B - \theta_{2,q}B^2 - \ldots - \theta_{q,q}B^q.$$ 

The reversed form of the ARMA($p, q$) model is,

$$\Phi \left( \frac{1}{B} \right) X_t = \Theta \left( \frac{1}{B} \right) \epsilon_t,$$

$$\left( 1 - \phi_{1,p} \frac{1}{B} - \phi_{2,p} \frac{1}{B^2} - \ldots - \phi_{p,p} \frac{1}{B^p} \right) X_t = \left( 1 - \theta_{1,q} \frac{1}{B} - \theta_{2,q} \frac{1}{B^2} - \ldots - \theta_{q,q} \frac{1}{B^q} \right) \epsilon_t$$

$$\frac{1}{B^p}(B^p - \phi_{1,p}B^{p+1} - \ldots \phi_{p,p})X_t = \frac{1}{B^q}(B^q - \theta_{1,q}B^{q+1} - \ldots \theta_{q,q})\epsilon_t$$

$$\Phi^R(B)X_t = B^{p-q}\Theta^R(B)\epsilon_t$$
where,

\[
\Phi^R(B) = B^p - \phi_{1,p}B^{p-1} - \phi_{2,p}B^{p-2} - \ldots - \phi_{p,p}
\]

\[
\Theta^R(B) = B^q - \theta_{1,q}B^{q-1} - \theta_{2,q}B^{q-2} - \ldots - \theta_{q,q}.
\]

For example, for the ARMA(1,1) model,

\[
(1 - \phi_{1,1}B)X_t = (1 - \theta_{1,1}B)\epsilon_t,
\]

reversed form is

\[
(B - \phi_{1,1})X_t = (B - \theta_{1,1})\epsilon_t.
\]

Now \(\Phi(z) = 1 - \phi_{1,1}z\), and a root is the solution of \(1 - \phi_{1,1}z = 0\), i.e.,

\[
|z| = \left| \frac{1}{\phi_{1,1}} \right| > 1 \Rightarrow |\phi_{1,1}| < 1.
\]

But, \(\Phi^R(z) = z - \phi_{1,1}\), and so a root is the solution of \(z - \phi_{1,1} = 0\), i.e., \(z = \phi_{1,1}\).

But, since for stationarity \(|\phi_{1,1}| < 1\) we have

\[
|z| = |\phi_{1,1}| < 1,
\]

so the root of \(\Phi^R(z)\) is inside the unit circle. Hence the standard assumption for stationarity (roots outside the unit circle) has within it an assumption of directionality. [N.B. only if the roots of \(\Phi(z)\) are on the unit circle is model ALWAYS non-stationary].

Figure 17 shows two time series which have different characteristics when time reversed.