Chapter 3: Autoregressive and moving average processes

Yining Chen∗

26 January 2015†

1 Two operators

Definition. We define the backshift operator by $BX_t = X_{t-1}$ and extend it to powers, i.e. $B^2X_t = B(BX_t) = BX_{t-1} = X_{t-2}$ and so on. Thus, $B^kX_t = X_{t-k}$.

Definition. Differences of order $d$ are defined as $\nabla^d = (1 - B)^d$, where we may expand the operator $(1 - B)^d$ algebraically to evaluate for higher integer values of $d$. When $d = 1$, we usually drop $d$ from the notation.

2 Moving average models

Definition. The moving average model of order $q$, or MA($q$), is defined to be

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q},$$

where $\epsilon_t \sim i.i.d. N(0, \sigma^2)$.

Remarks:
1. Without loss of generality, we assume the mean of the process to be zero.
2. Here $\theta_1, \ldots, \theta_q$ ($\theta_q \neq 0$) are the parameters of the model.
3. Sometimes it suffices to assume that $\epsilon_t \sim WN(0, \sigma^2)$. Here we assume normality mainly to simplify our discussion.
4. By defining the moving average operator as

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$$

We may also write the MA($q$) process in the equivalent form

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

$\Theta(z)$ is also known as the MA polynomial for $z \in \mathbb{C}$.

Proposition. Let $\{X_t\}$ follow the MA($q$) model. Then

1. $EX_t = 0$,
2. $\text{var} X_t = (1 + \theta_1^2 + \cdots + \theta_q^2) \sigma^2$.

∗Please send any comments and corrections to Y.Chen@statslab.cam.ac.uk.
†Updated on 28 Jan.
3. \( p_k = \begin{cases} 
1, & k = 0, \\
\sum_{i=0}^{k-1} \theta_i \theta_{i+k}, & 1 \leq k \leq q, \text{ where we denote } \theta_0 = 1 \text{ for notational convenience.} \\
0, & k > q.
\end{cases} \)

Remarks:

1. For an MA(q) model, its ACF vanishes after lag q.
2. It is both weakly and strongly stationary.

Now consider an MA(1) model, \( X_t = \epsilon_t + \theta \epsilon_{t-1} \). We have shown previously that

\[
\gamma_k = \begin{cases} 
(1 + \theta^2)\sigma^2 & k = 0, \\
\theta \sigma^2 & k = 1, \text{ and } \rho_k = \begin{cases} 
1 & k = 0, \\
\frac{\theta}{1 + \theta^2}, & k = 1, \\
0, & k > 1.
\end{cases}
\end{cases}
\]

Observe that \(|\rho_1| \leq 1/2\) and \(\rho_1\) is the same for \(\theta\) as for \(1/\theta\). Moreover, the pair \((\theta, \sigma^2)\) and \((1/\theta, \sigma^2\theta^2)\) yield the same autocovariance function \(\gamma_k\). For example, the following two models

\[
\begin{align*}
X_t & = \epsilon_t + 1/5 \epsilon_{t-1} & \epsilon_t & \overset{i.i.d.}{\sim} N(0, 25) \\
X_t & = \eta_t + 5 \eta_{t-1} & \eta_t & \overset{i.i.d.}{\sim} N(0, 1)
\end{align*}
\]

are the same because of normality.

Let’s express \(\{\epsilon_t\}\) backward in terms of the data for both models. For the pair \((\theta, \sigma^2)\), we get

\[
\epsilon_t = X_t - \theta \epsilon_{t-1} = X_t - \theta (X_{t-1} - \theta \epsilon_{t-2}) = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \theta^3 X_{t-3} + \cdots,
\]

while for the pair \((1/\theta, \sigma^2\theta^2)\), we get

\[
\epsilon_t = X_t - \theta^{-1} X_{t-1} + \theta^{-2} X_{t-2} - \theta^{-3} X_{t-3} + \cdots.
\]

Note that if \(|\theta| < 1\), the first expression converges while the second diverges. When we want to interpret or estimate the noises \(\epsilon_t\), it is more desirable to deal with a convergent expression, so in the previous MA(1) example, we prefer \(\theta = 1/5\) to \(\theta = 5\).

In the first case, if \(\theta < 1\), we can also derive \(\Pi\) via the moving average operator, i.e.

\[
\epsilon_t = \Theta^{-1}(B)X_t = (1 + \theta B)^{-1}X_t = (1 - \theta B + \theta^2 B^2 - \cdots)X_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \theta^3 X_{t-3} + \cdots.
\]

In general, if the contemporaneous error term of an MA(q) can be expressed as a linear combination of its past observations \(\{X_s, s \leq t\}\), then the process is said to be invertible. More formally,

**Definition.** An MA(q) (or more generally, ARMA (p,q)) process is said to be invertible, if the time series \(\{X_t\}\) can be written as

\[
\Pi(B)X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} = \epsilon_t,
\]

where \(\Pi(B) = \sum_{j=0}^{\infty} \pi_j B^j\), and \(\sum_{j=0}^{\infty} |\pi_j| < \infty\). Here we set \(\pi_0 = 1\).

\footnote{This convergence should be interpreted in the sense of mean-square limit. Rigorous treatment of such issues proceeds by representing all the variables as elements of a suitable Hilbert space, with norm given by \(|X|^2 = \text{var}(X)\) and prove that \(Z_t = X_t + \sum_{i=1}^{\infty} (-\theta)^i X_{t-i}\) is a Cauchy sequence, and so converges. See also the example sheet.}
For the following theorem, the proof of its “if” part is given in the lecture. See the example sheet for a proof of its “only if” part.

**Theorem.** An MA(q) process is invertible if and only if the roots of the equation $\Theta(z) = 0$ all lie outside the unit circle.

Remarks:

1. In case $q = 1$, the equation $\Theta(B) = 0$ becomes $1 + \theta z = 0$. Requiring the root to be outside the unit circle is equivalent to asking for $|−θ| > 1$, which is equivalent to $|θ| < 1$.

2. The coefficients $\pi_j$ of $\Pi$ given in (2) can be determined by solving

   \[
   \Pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{1}{\Theta(z)}, \quad |z| \leq 1.
   \]

3. Autoregressive models

   **Definition.** The autoregressive model of order $p$, or AR(p), is of the form

   \[
   X_t = \epsilon_t + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p}, \quad (3)
   \]

   where $\epsilon_t \sim N(0, \sigma^2)$, and where $\{X_t\}$ is stationary.

   Remarks:

   1. (3) implies that the mean of $X_t$, $\mu$, is zero. If we want to model a series with non-zero mean by AR, we can replace $X_t$ by $X_t - \mu$ in the previous definition, so

      \[
      X_t - \mu = \epsilon_t + \phi_1 (X_{t-1} - \mu) + \phi_2 (X_{t-2} - \mu) + \cdots + \phi_p (X_{t-p} - \mu),
      \]

      or equivalently, we can write

      \[
      X_t = c + \epsilon_t + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p},
      \]

      where $c = \mu(1 - \phi_1 - \cdots - \phi_p)$.

   2. Here $\phi_1, \ldots, \phi_p$ ($\phi_p \neq 0$) are the parameters of the model.

   3. By defining the autoregressive operator as

      \[
      \Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p
      \]

      We may write (3) as $\Phi(B)X_t = \epsilon_t$. $\Phi(z)$ is also known as the AR polynomial for $z \in \mathbb{C}$.

   4. The existence of such a process is guaranteed. The basic idea should be clear when we study AR(1).

   Now consider AR(1) model with $X_t = \phi X_{t-1} + \epsilon_t$ and $\epsilon_t \sim N(0, \sigma^2)$. Iterating backwards $k$ times, we get

   \[
   X_t = \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j \epsilon_{t-j}.
   \]

   Provided that $|\phi| < 1$ and $X_t$ is stationary, we can represent AR(1) as

   \[
   X_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}. \quad (4)
   \]

   again in the mean square sense, see the previous footnote.
The mean of AR(1) is zero. The autocovariance function is \( \gamma_k = \sigma^2 \phi^k / (1 - \phi^2) \) for \( k \geq 0 \), and the autocorrelation is \( \rho_k = \phi^k \). Therefore, AR(1) is weakly stationary by construction via (4).

Note that for AR(1) with \( |\phi| > 1 \), the process is explosive because the values of the series can quickly become large in magnitude. A stationary solution still exists by the following construction:

\[
X_t = \phi^{-1}X_{t+1} - \phi^{-1} \epsilon_{t+1} = \phi^{-1}(\phi^{-1}X_{t+2} - \phi^{-1} \epsilon_{t+2}) - \phi^{-1} \epsilon_{t+1} = \phi^{-k}X_{t+k} - \sum_{j=1}^{k-1} \phi^{-j} \epsilon_{t+j} = -\sum_{j=1}^{\infty} \phi^{-j} \epsilon_{t+j}.
\]

However, this solution is future-dependent.

**Definition.** An AR\((p)\) (or more generally, ARMA \((p,q)\)) process is said to be causal, if the time series \( \{X_t\} \) can be written as

\[
X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} = \Psi(B) \epsilon_t,
\]

where \( \Psi(B) = \sum_{j=0}^{\infty} \psi_j B^j \), and \( \sum_{j=0}^{\infty} |\psi_j| < \infty \). Here we set \( \psi_0 = 1 \).

The following theorem should be compared to its MA analogue stated in the previous section. It can be proved using exactly the same techniques shown in the lecture and the example sheet.

**Theorem.** An AR\((p)\) process is causal if and only if the roots of the equation \( \Phi(z) = 0 \) all lie outside the unit circle.

### 4 Autoregressive moving average models

**Definition.** The autoregressive moving average model of orders \( p \) and \( q \), or ARMA\((p,q)\), is of the form

\[
X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q},
\]

where \( \epsilon_t \sim i.i.d. N(0, \sigma^2) \), and where \( \{X_t\} \) is stationary.

**Remarks:**

1. Using the AR operator and the MA operator, the ARMA model in (6) can be written in concise form as

\[
\Phi(B)X_t = \Theta(B) \epsilon_t.
\]

2. The process reduces to AR\((p)\) if \( q = 0 \), or to MA\((q)\) if \( p = 0 \).

3. The usefulness of ARMA models lies in their parsimonious representation.

The problem of parameter redundancy (or over-parameterisation) can arise in ARMA models. Consider a white noise process \( X_t = \epsilon_t \). It also satisfies the following

\[
X_t = 0.5X_{t-1} + \epsilon_t - 0.5 \epsilon_{t-1},
\]

which looks like an ARMA\((1,1)\) model. When writting the model in operator form, we have

\[
(1 - 0.5B)X_t = (1 - 0.5B) \epsilon_t.
\]
We can easily detect this problem by observing that the AR polynomial $\Phi(z) = 1 - 0.5z$ and the MA polynomial $\Theta(z) = 1 - 0.5z$ have a common factor $1 - 0.5z$. Discarding the common factor in each leaves $\Phi(z) = \Theta(z) = 1$, and we deduce that the model is actually white noise.

More generally, the problem of parameter redundancy arises only if there are common roots in $\Phi(z) = 0$ and $\Theta(z) = 0$. If so, this issue can be resolved by discarding the common factors in $\Phi(z)$ and $\Theta(z)$.

**Theorem.** An ARMA($p$) process is causal and invertible if there is no common factors in $\Phi(z)$ and $\Theta(z)$, and the roots of $\Phi(z) = 0$ and $\Theta(z) = 0$ all lie outside the unit circle.

Remark: $\epsilon_t = \Pi(B)X_t$ is called the invertible representation of ARMA, where the coefficients $\pi_j$ of $\Pi$ given in (2) can be determined by solving

$$\Pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\Phi(z)}{\Theta(z)}, \quad |z| \leq 1.$$ 

On the other hand, $X_t = \Psi(B)\epsilon_t$ is called the causal representation of ARMA, where the coefficients $\psi_j$ of $\Psi$ given in (5) can be determined by solving

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\Theta(z)}{\Phi(z)}, \quad |z| \leq 1.$$ 

5 Autoregressive integrated moving average models

Often we process a time series before data analysis (e.g. differencing). Therefore, it is natural to consider the following generalisation of ARMA:

**Definition.** $\{X_t\}$ is said to be an ARIMA($p, d, q$) process if $\nabla^d X_t$ is an ARMA($p, q$) process.

Remarks:

1. In the operator form, $\Phi(B)(1 - B)^d X_t = \Theta(B)\epsilon_t$.
2. Typically, we pick $d$ to be a small integer ($\leq 2$).