Chapter 1 ARMA Models

1. Structure of ARMA models

1.1 Probability Background

Definition 1.1

Let $\xi_1, \xi_2, \cdots, \xi_n \cdots$ be a sequence of random variables

If, for any $\varepsilon > 0$, $\exists N$ such that, as $n > N$,

$$P(|\xi_n - \xi| > \varepsilon) < \varepsilon;$$

or $$P(|\xi_n - \xi| > \varepsilon) \longrightarrow 0 \quad \text{as } n \to \infty,$$

We say that ξ_n converges to ξ in probability, denoted by $\xi_n \xrightarrow{P} \xi$.

Definition 1.2.

Let $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ be a sequence of random variables, and ξ is a r.v.

Denote $\Omega = \{\omega : \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\}$.

If $P(\Omega) = 1$, then we say that $\xi_n \to \xi$ almost surely. denoted by $\xi_n \xrightarrow{a.s.} \xi$.

Definition 1.3.

Let $\xi_1, \xi_2, \ldots, \xi_n$ be a sequence of a r.v.s. If, for any $\varepsilon > 0$, $\exists N$, such that, as $n > N$,

$$E|\xi_n - \xi|^p < \varepsilon$$

or $\lim\limits_{n\to\infty} E|\xi_n - \xi|^p = 0$, where p is some constant > 0.

We say that $\xi_n \to \xi$ in L^p, denoted by

$$\xi_n \xrightarrow{L^p} \xi .$$

In particular, as $p = 2$, we say that $\xi_n \to \xi$ in mean square, denoted by $\xi_n \xrightarrow{m.s.} \xi$.

Theorem 1.1. The following relationships holds:

<i>. $\xrightarrow{a.s} \Longrightarrow \xrightarrow{P}$

<ii>. $\xrightarrow{L^P} \Longrightarrow \xrightarrow{P}$

<iii>. $\xrightarrow{m.s} \Longrightarrow \xrightarrow{L^P}$ if $0<p\le 2$.

<iv>. $\xrightarrow{L^P} \Longrightarrow$ a.s. if $E\|\xi_{n+1}-\xi_n\|^p \le c_n$

and $\sum_{n=1}^{\infty} c_n < \infty$.

□

Theorem 1.2. (Fatou's Lemma) If $X_n \ge 0$, then

$$\liminf_{n\to\infty} E X_n \ge E\left(\liminf_{n\to\infty} X_n\right)$$

□

Theorem 1.3. Let $\mathcal{F}_t$ be the σ-field generated by a sequence of $\xi_t, \xi_{t-1}, \dots$, and $y_t = f(\xi_t, \xi_{t-1}, \dots)$.

Then y_t is $\mathcal{F}_t$-measurable.

$$\mathcal{F}_t = \sigma\{\xi_t, \xi_{t-1}, \dots\}$$

Theorem 1.4. Let $\xi_t, \xi_{t-1}, \dots$ be a strictly stationary and ergodic time series, and $y_t = f(\xi_t, \xi_{t-1}, \dots)$. Then the time series

$$\{ y_t, t = 0, \pm 1, \pm 2, \dots \} \text{ is strictly stationary and ergodic.} \quad \square$$

In particular, if $\xi_t\ \xi_{t-1} \dots$ is i.i.d, then $\xi_t, \xi_{t-1} \dots$ are strictly stationary and ergodic.

1.2 ARMA Models.

We say that $\{ y_t : t = 0, \pm 1, \pm 2 \dots \}$ follows an ARMA model, if it satisfies the following equation:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \psi_1 \varepsilon_{t-1} + \dots + \psi_q \varepsilon_{t-q} + \varepsilon_t, \quad (1.1)$$

where $\varepsilon_t \sim$ i.i.d with mean zero and variance σ^2.

We first investigate whether or not there is a solution $\{y_t\}$ to model (1.1). For this, we rewrite Model (1.1) as the following vector form:

$$Y_t = A Y_{t-1} + \tilde{\varepsilon}_t \quad (1.2)$$

where $Y_t = (y_t, y_{t-1}, \dots, y_{t-p+1}, \varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q})'$,

$$\tilde{\varepsilon}_t = (\varepsilon_t, 0 \dots 0, \varepsilon_t, 0, \dots, 0)',$$

and

$$A = \left(\begin{array}{cccc|cccc}
\phi_1 & \cdots & \cdots & \phi_p & \psi_1 & \cdots & \cdots & \psi_q \\
1 & 0 & \cdots & 0 & & & & \\
 & & & & & 0 & & \\
0 & \cdots & 1 & 0 & & & & \\
\hline
 & & & & 0 & \cdots & \cdots & 0 \\
 & 0 & & & 1 & 0 & \cdots & 0 \\
 & & & & & & & \\
 & & & & 0 & \cdots & 1 & 0
\end{array}\right).$$

Now, we want to prove the following theorem:

Theorem 1.5. If All the roots of the polynomial $\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p = 0$ lie inside the unit circle, then there exists a unique, and the second-order stationary and $\mathcal{F}_t$-measurable solution to model (1.1). This solution has the following expansion:

$$y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \text{ in m.s. and in a.s.}$$

and hence y_t is strictly stationary and ergodic, where $\mathcal{F}_t = \sigma\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$, and $\psi_i = O(\rho^i)$ with $\rho \in (0,1)$.

Proof. Note that Model (1.1) and model (1.2) are equivalent. We first prove that there exists a solution to model (1.2).

Iterating (1.2), we have

$$Y_t = \hat{\varepsilon}_t + A\hat{\varepsilon}_{t-1} + \cdots + A^n \hat{\varepsilon}_{t-n} + A^{n+1} Y_{t-n-1}.$$

Let $S_t^{(n)} = \hat{\varepsilon}_t + A\hat{\varepsilon}_{t-1} + \cdots + A^n \hat{\varepsilon}_{t-n}$. We need to prove that $S_t^{(n)}$ converges to r.v. in mean square.

By a ~~direction~~ direct calculation, we know that

$$|I\lambda - A| = \left|\begin{array}{cccc|cccc} \lambda-\phi_1 & \cdots & \cdots & -\phi_p & -\theta_1 & \cdots & \cdots & -\theta_q \\ -1 & & & 0 & & & & \\ & \ddots & & & & & O & \\ 0 & & -1 & 0 & & & & \\ \hline & & & & \lambda & 0 & \cdots & 0 \\ & & O & & -1 & \ddots & & \\ & & & & 0 & & -1 & \lambda \end{array}\right|$$

$$= \lambda^q (\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p)$$

Thus, all the roots of $\lambda^p - \phi_1\lambda^{p-1} - \cdots - \phi_p = 0$ lie inside the unit circle if and only if all the eigenvalues of A lie inside the unit circle.

Denote $\rho = \max\{|\text{the eigenvalue}| \text{ of } A\}$. Then $0<\rho<1$

By Corollary A.2 in Johansen (1995, pp. 220), there is a constant $c > 0$ such that

$$\| A^n \| \le c \rho^{n/2}. \tag{1.3}$$

By (1.3), we have that

$$E \| S_t^{(n)} - S_t^{(n+m)} \|^2$$

$$= E \| A_t^{n+1} \tilde{\varepsilon}_{t-n-1} + \cdots + A^{n+m} \hat{\varepsilon}_{t-n-m} \|^2$$

$$\le (\| A^{n+1} \|^2 + \cdots + \| A_i \|^{n+m}) E \| \hat{\varepsilon}_{t-n-m} \|^2$$

$$\le 2c (\rho^{n+1} + \cdots + \rho^{n+m}) \sigma^2 \tag{1.4}$$

$$\longrightarrow 0 \quad \text{as} \quad n, m \to \infty.$$

By (1.4) and the Cauchy criterion, we know that there ~~as~~ is a r.v. S_t such that

$$S_t^{(n)} \xrightarrow{m.s.} S_t \tag{1.5}$$

By the Factou's Lemma,

$$E \| S_t - \sum_{t=0}^{\infty} A^i \hat{\varepsilon}_{ti} \|^2 = E \liminf_{n \to \infty} \| S_t - \sum_{t=0}^{n} A^i \hat{\varepsilon}_{ti} \|^2$$

$$\le \liminf_{n \to \infty} E \| S_t - S_t^{(n)} \|^2 = 0 \tag{1.6}$$

By (1.6), we have that

$$S_t \overset{a.s.}{=} \sum_{i=0}^{\infty} A^i \hat{\varepsilon}_{t-i} \tag{1.7}$$

By (1.5) and (1.7), we know that

$$S_t^{(n)} \xrightarrow{a.s.} \sum_{i=0}^{\infty} A^i \hat{\varepsilon}_{t-i} = S_t$$

and

$$S_t^{(n)} \xrightarrow{m.s.} \sum_{i=0}^{\infty} A^i \hat{\varepsilon}_{t-i} = S_t. \tag{1.8}$$

By (1.8), it is easy to check that S_t in (1.8) is a solution of model (1.2).

Let $y_t = (1, 0, \cdots, 0) S_t$. Then

$$y_t = \sum_{i=0}^{\infty} \nu' A^i \nu_1' \varepsilon_{t-i}$$ is a solution

of model (1.1), where $\nu = (1, 0, \cdots, 0)'$ and $\nu_1 = (1, 0, \cdots, 0, 1, 0, \cdots, 0)$. It is obvious that y_t is strictly stationary and $\mathcal{F}_t$-measurable by Theorem 1.4 and Theorem 1.3, respectively.

It is easy to show that $E y_t^2 < \infty$. Thus $E y_t^2 =$ a constant since y_t is strictly stationary, and hence

Y_t is the second-order stationary (i.e. weak stationary).

In the following, we prove that the ~~not~~ uniqueness.

If there are two solutions Y_t' and Y_t'' to model (1.2), then

$$Y_t' = A Y_{t-1}' + \tilde{\Sigma}_t$$
$$Y_t'' = A Y_{t-1}'' + \hat{\Sigma}_t \quad -)$$

$$Y_t' - Y_t'' = A(Y_{t-1}' - Y_{t-1}'').$$

Let $U_t = Y_t' - Y_t''$. Then we have that

$$U_t = A U_{t-1} = A^2 U_{t-1} = \cdots = A^n U_{t-n}.$$

Thus

$$E\|U_t\|^2 \le \|A^n\|^2 E\|U_{t-n}\|^2$$
$$\le C\rho^n E\|U_{t-n}\|^2 \qquad (1.9)$$

Note that $E\|U_{t-n}\|^2 \le E(\|Y_t'\|^2 + \|Y_t''\|^2)$

$\le c_1$ by the second-order stationarity of Y_t' and Y_t'', where c_1 is a constant.

By (1.9), we have that

$$E\|U_t\|^2 \le C C_1 \rho^n \longrightarrow 0$$

Thus

$$E\|U_t\|^2 = \lim_{n\to\infty} E\|U_t\|^2 \le \lim_{n\to\infty}(C C_1 \rho^n) = 0. \quad (1.10)$$

By (1.10), we know that

$$U_t \overset{a.s.}{=} 0 \text{ and hence } Y_t' \overset{a.s.}{=} Y_t''. \quad (1.11)$$

By (1.11), the solution y_t is unique for model (1.1).

□

Remark. 1.1. (a). In theorem 1.5, the condition $E\varepsilon_t^2 = \sigma < \infty$ can be relaxed to be the condition $E|\varepsilon_t|^p < \infty$ with $p \in (0,2)$. In this case,

$$y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \quad \text{in } L^p \text{ and in a.s.}$$

(b). In Th. 1.5, ε_t i.i.d. can be relaxed to that ε_t is an uncorrelated white noise. In this case,

$$y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \quad \text{in m.s.}$$

but y_t may not be strictly stationary.

□

Theorem 1.6. Suppose that model (1.1) has a second-order stationary solution y_t. If $z^p - \phi_1 z^{p-1} - \cdots - \phi_p = 0$ and $z^q + \varphi_1 z^{q-1} + \cdots + \varphi_q = 0$ have no common roots, then it is neccessary that all the roots of $z^p - \phi_1 z^{p-1} - \cdots - \phi_p = 0$ lie inside the unit circle.

Proof. Since y_t is the second-order stationary solution of model (1.1), and hence Y_t is the second-order stationary solution of model (1.2). Thus,

$$E Y_t Y_t' \equiv \Omega, \text{ a constant matrix.} \quad (1.12)$$

By (1.2), we can obtain that

$$E Y_t Y_t' = A E Y_{t-1} Y_{t-1}' A + E \hat{\varepsilon}_t \hat{\varepsilon}_t' \quad (1.13)$$

Note that

$$E \hat{\varepsilon}_t \hat{\varepsilon}_t' = \sigma^2 \Gamma \Gamma' \text{ with } \Gamma = (1, 0, \cdots, 0, 1, 0, \cdots, 0)'. \quad (1.14)$$

By (1.12)-(1.14), we have that

$$\Omega = A \Omega A' + \sigma^2 \Gamma \Gamma' \quad (1.15)$$

Suppose that A has an eigenvalue λ with corresponding left eigenvector $z = (z_1, z_2, \ldots, z_{p+q})'$, i.e. $z'A = \lambda z$, and $\bar{z}'A = \bar{\lambda}\bar{z}$, where $\bar{z}$ is the conjugate vector of z.

Thus

$$
\begin{aligned}
z'\Omega\bar{z} &= z'A\,\Omega A'\bar{z} + z'\Gamma\Gamma'\bar{z}\,\sigma^2 \\
&= (\lambda z)\,\Omega\,(\bar{z}'A)' + \sigma^2 |z_1 + z_{p+k}|^2 \\
&= \lambda z\,\Omega\,\bar{z}\,\bar{\lambda} + \sigma^2 |z_1 + z_{p+k}|^2 \\
&= |\lambda|^2 z\,\Omega\,\bar{z} + \sigma^2 |z_1 + z_{p+k}|^2 \qquad (1.16)
\end{aligned}
$$

If we can prove that $|z_1 + z_{p+k}| \neq 0$, then $|\lambda| < 1$ by (1.16).

From the equation:

$$
(z_1, \ldots, z_{p+q})
\left[\begin{array}{cccc|cccc}
\phi_1 & \cdots & & \phi_p & \theta_1 & \cdots & & \theta_q \\
1 & 0 & \cdots & 0 & & & & \\
\vdots & & & & & & O & \\
0 & \cdots & 1 & 0 & & & & \\
\hline
 & & & & 0 & \cdots & & 0 \\
 & & O & & 1 & \cdots & & 0 \\
 & & & & \vdots & & & \\
 & & & & 0 & \cdots & 1 & 0
\end{array}\right]
= \lambda (z_1, \ldots, z_{p+q}),
$$

We obtain that

$$z_1\phi_1 + z_2 = \lambda z_1$$
$$z_1\phi_2 + z_3 = \lambda z_2$$
$$\vdots$$
$$z_1\phi_p + z_p = \lambda z_{p-1}$$
$$z_1\phi_p = \lambda z_p$$

(A)

$$z_1\psi_1 + z_{p+2} = \lambda z_{p+1}$$
$$z_1\psi_2 + z_{p+3} = \lambda z_{p+2}$$
$$\vdots$$
$$z_1\psi_{q-1} + z_{p+q} = \lambda z_{p+q-1}$$
$$z_1\psi_q = \lambda z_{p+q}$$

(B)

From the equations in (A), we have that

$$z_1(\lambda^p - \phi_1\lambda^{p-1} - \cdots - \phi_p) = 0 \qquad (1.17)$$

If $z_1 + z_{p+1} = 0$, then $z_1 = -z_{p+1}$. From the equations in (B), we have that

$$-z_{p+1}(\lambda^q + \psi_1\lambda^{q-1} + \cdots + \psi_q) = 0 \quad \text{, i.e.}$$

$$z_1(\lambda^q + \psi_1\lambda^{q-1} + \cdots + \psi_q) = 0 \qquad (1.18)$$

By (1.17) and (1.18), and the Assumption that $z^p - \phi_1 z^{p-1} - \cdots - \phi_p = 0$ and $z^q + \psi_1 q z^{q-1} + \cdots + z\psi_q = 0$ have no common root, we know that $z_1 = 0$ and hence $z_{p+1} = 0$.

Since $z_1 = z_{p+1} = 0$, by the equations (A) and (B), We can claim that $z = (z_1, z_2, \cdots, z_{p+q})' = 0$.

However, z is the eigenvalue of A, so $z \neq 0$. This is a contradiction. Thus

$$z_1 + z_{p+1} \neq 0 \implies |\lambda| < 1.$$

$\square$

2. Estimation

2.1 Probability Background

<u>Theorem 2.1.</u> Let $\{y_t : t = 0, \pm 1, \pm 2 \dots\}$ be a strictly stationary and ergodic and $E|y_t| < \infty$. Then

$$\frac{1}{n} \sum_{t=1}^{n} y_t \xrightarrow{a.s.} E y_t .$$

□

Theorem 2.1 is called the <u>ergodic theorem</u>

<u>Definition 2.1</u> Let $\xi_1, \xi_2 \dots \xi_n \dots$ be a sequence r.v.s with distributions: $F_1(x), F_2(x), \dots, F_n(x), \dots$, respectively, and ξ be another r.v. with distribution $F(x)$. If

$$F_n(x) \longrightarrow F(x)$$

at any value x at which $F(x)$ is continuous. Then, we say that ξ_n converges to ξ in distribution, denoted by

$$\xi_n \xrightarrow{\mathcal{L}} \xi .$$

□

Relationship:

$$\left.\begin{array}{l} \xi_n \xrightarrow{P} \xi \\ \xi_n \xrightarrow{L^p} \xi \\ \xi_n \xrightarrow{a.s.} \xi \end{array}\right\} \Longrightarrow \xi_n \xrightarrow{\mathcal{L}} \xi .$$

Definition 2.2. Let $\{\xi_t : t = 0, \pm 1, \pm 2, \dots\}$ be a sequence of r.v.s., $\mathcal{F}_t$ be an increasing sequence of σ-fields, i.e.

$$\mathcal{F}_t \supset \mathcal{F}_{t-1} \supset \mathcal{F}_{t-2}, \dots, \quad \text{and} \quad \xi_t \text{ is } \mathcal{F}_t\text{-measurable}.$$

If $E(\xi_t \mid \mathcal{F}_{t-1}) = 0$, then we say that $\{\xi_t\}$ is a martingale difference sequence with respect to $\mathcal{F}_t$. □

Theorem 2.2. (central limit theorem). Suppose that $\xi_1, \xi_2 \dots \xi_n \dots$ is an ergodic stationary sequence with

$$\sigma^2 = E\xi_n^2 < \infty \quad \text{and} \quad E(\xi_n \mid \mathcal{F}_{n-1}) = 0,$$

where $\mathcal{F}_n = \sigma\{\xi_t : t \le n\}$. Let $S_n = \xi_1 + \dots + \xi_n$. Then

$$\sqrt{n}\, S_n \xrightarrow{\mathcal{L}} N(0, \sigma^2)$$

as $n \to \infty$ □

Theorem 2.3 (Cramér-Wold device). Let $\{W_n\}$ be a sequence of random vector. Then

$$W_n \xrightarrow{\mathcal{L}} W \quad \text{if and only if}$$

$$\lambda' W_n \xrightarrow{\mathcal{L}} \lambda W$$

for any constant vector $\lambda = (\lambda_1, \dots, \lambda_m)'$ with $\lambda'\lambda \neq 0$, where m is the dimension of the vector W_n. □

2.2. Statistical Background

Suppose that $\{y_t\}$ is generated by the following model:

$$y_t = f(\tilde{Y}_t, \lambda) + \varepsilon_t, \qquad (2.1)$$

with the true parameter λ_0, where $\tilde{Y}_t = \{y_{t-1}, y_{t-2}, \cdots\}$ and ε_t is i.i.d. white noise.

$y_t = f(\tilde{Y}_t, \lambda_0) + \varepsilon_t$ is called the true model.

Model (2.1) is called the unknown parameter model and λ is called the unknown parameter.

Suppose that $\lambda_0 \in \Theta$ and $\lambda \in \Theta$. Then Θ is called the parameter space.

Given $y_1, \cdots, y_n$ which come from model (2.1) with the true parameter λ_0. We want to find an estimator of λ_0 in Θ, denoted this estimator by $\hat{\lambda}_n$.

This $\hat{\lambda}_n$ is a measurable function of $\{y_1, \cdots, y_n\}$.

Definition 2.3. (Consistency)

<i> If $\hat{\lambda}_n \xrightarrow{P} \lambda_0$, we call that $\hat{\lambda}_n$ is an consistent estimator of λ_0 in probability.

<ii> If $\hat{\lambda}_n \xrightarrow{a.s.} \lambda_0$, we say that $\hat{\lambda}_n$ is a consistent estimator of λ_0, almost surely.

Definition 2.4 (Asympotic nomality).

If

$$\sqrt{n}\,(\hat{\lambda}_n - \lambda_0) \xrightarrow{\mathcal{L}} N(0, \Sigma),$$

we say that the estimator $\hat{\lambda}_n$ of λ_0 is asymptotically normal with the asymptotic Variance $\underline{\underline{\Sigma}}$.

Definition 2.5. (a general definition):

If $\quad D_n\,(\hat{\lambda}_n - \lambda_0) \xrightarrow{\mathcal{L}} W$, we say that the estimator $\hat{\lambda}_n$ converges to W in distribution with the rate of convergence D_n.

$\square$

2.3 The estimator of Mean.

Theorem 2.4. Let y_t be generated by the AR(1) model:

$$y_t = \phi y_{t-1} + \varepsilon_t ,$$

where $|\phi|<1$ and ε_t ~~be~~ is i.i.d. white noise with mean zero and variance σ^2. Then

$$\frac{1}{n}\sum_{t=1}^{n} y_t \xrightarrow{a.s.} E y_t = 0 .$$

Proof. Since $|\phi|<1$ and ε_t is i.i.d., by Theorem 1.5, y_t is strictly and ergodic, and $E y_t^2 < \infty$.

By Theorem 1.1 (iii),

$$E|y_t| < \infty \quad \left(\text{or } E|y_t| \le \sqrt{E|y_t|^2} = \sqrt{\sigma^2} < \infty \right).$$

Furthermore, by ergodic Theorem 2.1, we have that

$$\frac{1}{n}\sum_{t=1}^{n} y_t \xrightarrow{a.s.} E y_t = 0$$

$\square$

Theorem 2.5. Let y_t be generated by the AR(1) model:

$$y_t - \mu = \phi (y_{t-1} - \mu) + \varepsilon_t ,$$

where $|\phi| < 1$ and ε_t is i.i.d. white noise with mean zero and variance σ^2. Then $E y_t = \mu$. Let $\hat{\mu}_n = \frac{1}{n} \sum_{t=1}^{n} y_t$. Then

$$\hat{\mu}_n \xrightarrow{a.s.} \mu .$$

Proof. Let $\tilde{y}_t = y_t - \mu$. Then

$$\hat{y}_t = \phi \hat{y}_t + \varepsilon_t . \qquad (2.2)$$

By Assumption given in this theorem and Theorem 1.5, $\tilde{y}_t$ is strictly stationary and ergodic, and $E \hat{y}_t^2 < \infty$. By ergodic theorem 2.1, we have that

$$\frac{1}{n} \sum_{t=1}^{n} \hat{y}_t \xrightarrow{a.s.} E \hat{y}_t = E y_t - \mu = 0 \qquad (2.3)$$

Note that $\frac{1}{n} \sum_{t=1}^{n} \hat{y}_t = \frac{1}{n} \sum_{t=1}^{n} y_t - \mu$. By (2.3), it follows that

$$\hat{\mu}_n = \frac{1}{n} \sum_{t=1}^{n} y_t \xrightarrow{a.s.} \mu$$

$\square$

Theorem 2.6. Let y_t be generated by the ARMA (p,q) model:

$$y_t - \mu = \sum_{i=1}^{p} \phi_i (y_{t-i} - \mu) + \sum_{j=1}^{q} \varphi_j \varepsilon_{t-j} + \varepsilon_t,$$

where all the roots of $\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p = 0$ lie inside the unit circle and ε_t is i.i.d. white noise with mean zero and variance σ^2. Then $E y_t = \mu$. Let $\hat{\mu}_n = \frac{1}{n} \sum_{t=1}^{n} y_t$. Then

$$\hat{\mu}_n \xrightarrow{a.s.} \mu.$$

Proof. It is similar to that of Theorem 2.5 and hence is omitted. □

Remark 2.1. The condition that ε_t is i.i.d. white noise can be relaxed to be the condition that ε_t is uncorrelated white noise. In the latter case, y_t is ~~th~~ second-order stationary. □

Remark 2.2. In practice, when we have a data set $\{X_1, \ldots, X_n\}$, if it is second-order stationary, then

$$\frac{1}{n} \sum_{t=1}^{n} X_t \to \mu = E X_t.$$

Let $y_t = X_t - \frac{1}{n} \sum_{t=1}^{n} X_t$. Then $E y_t = 0$. The new data set $\{y_t : t = 1, \ldots, n\}$ is called the centralized data. □

Remark 2.3. In Theorems 2.5 – 2.6, we can prove that

$$\sqrt{n}\,(\hat{\mu}_n - \mu) \xrightarrow{\mathcal{L}} N(0, \nu),$$

where $\nu = \sigma^2 \left(\sum_{j=0}^{\infty} \psi_j\right)^2$ and ψ_j is the coefficient in theorem 1.5. $\square$

See proposition 7.11 (Hamilton, P_{195})

and Theorem 7.1.2 in Brockwell and Davis (P_{219})

2.3. The Estimaters of Parameters in AR model.

Suppose that $y_1, \cdots, y_n$ are generated by the AR (p) model:

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t \qquad (2.3)$$

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

Let $\tilde{Y}_{t-1} = (y_{t-1}, \cdots, y_{t-p})$. Then the density y_t given $\{y_{t-1}, \cdots, y_1\}$ is

$$f(y_{p+1} \mid \tilde{Y}_p) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_{p+1} - \phi' \tilde{Y}_{p})^2}{2\sigma^2}}$$

$$f(y_{p+2} \mid \tilde{Y}_{p+1}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_{p+2} - \phi' \tilde{Y}_{p+1})^2}{2\sigma^2}}$$

$$\vdots \qquad\qquad \vdots$$

$$f(y_n \mid \tilde{Y}_{n-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_n - \phi' \tilde{Y}_{n-1})^2}{2\sigma^2}}.$$

Thus, the joint density of $\{y_n, y_{n-1}, \cdots, y_{p+1}\}$ given $\tilde{Y}_p$ is

$$f(y_n, \cdots, y_{p+1} \mid \tilde{Y}_p) = \left[\frac{1}{\sqrt{2\pi\sigma^2}}\right]^{n-p} e^{-\sum_{t=p+1}^{n} \frac{(y_t - \phi' \tilde{Y}_{t-1})^2}{2\sigma^2}}$$

Thus, the conditional log-likelihood function is

$$\mathcal{L}(\lambda) = -(n-p)\log(2\pi\sigma^2)^{\frac{1}{2}} - \sum_{t=p+1}^{n} \frac{(y_t - \phi' \tilde{Y}_{t-1})^2}{2\sigma^2}$$

where $\lambda = (\phi', \sigma^2)'$ and $\phi = (\phi_1, \cdots, \phi_p)'$.

$$\frac{\partial \mathcal{L}(\lambda)}{\partial \phi} = -\frac{1}{\sigma^2} \sum_{t=p+1}^{n} (y_t - \phi' \tilde{Y}_{t-1}) \tilde{Y}_{t-1}. \qquad (2.4)$$

Denote the maximum likelihood estimator (MLE) of ϕ_0 is $\hat{\phi}_n$, where ϕ_0 is the true parameter. By (2.4)

$$\sum_{t=p+1}^{n} (y_t - \hat{\phi}_n' \tilde{Y}_{t-1}) \tilde{Y}_{t-1} = 0 \qquad (2.5)$$

By (2.5), we have that

$$\hat{\phi}_n = \left(\sum_{t=p+1}^{n} \tilde{Y}_{t-1} \tilde{Y}_{t-1}' \right)^{-1} \left(\sum_{t=p+1}^{n} y_t \tilde{Y}_{t-1} \right) \qquad (2.6)$$

Lemma 2.1. Suppose that $y_1, \cdots, y_n$ are generated by model (2.3) and all the roots of $z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \cdots - \phi_p = 0$ lie inside the unit circle. Then

$$\frac{1}{n} \sum_{t=p+1}^{n} \tilde{Y}_{t-1} \tilde{Y}_{t-1}' \xrightarrow{a.s.} E(\tilde{Y}_{t-1} \tilde{Y}_{t-1}')$$

and $E(\tilde{Y}_{t-1} \tilde{Y}_{t-1}') > 0$ (i.e. positive definite).

Proof. By Theorem 1.5, y_t is strictly stationary and ergodic, and $E y_t^2 < \infty$.

By Theorem 1.4, $\tilde{Y}_{t-1}' \tilde{Y}_{t-1}'$ is also strictly stationary and ergodic, and

$$E\| \tilde{Y}_{t-1}' \tilde{Y}_{t-1}' \| \le E \| \tilde{Y}_{t-1} \|^2$$
$$= E(y_{t-1}^2 + \cdots + y_{t-p}^2) < \infty .$$

By ergodic theorem 2.1, we have that

$$\frac{1}{n} \sum_{t=p+1}^{n} \tilde{Y}_{t-1}' \tilde{Y}_{t-1}' = \frac{n-p}{n} \cdot \left[\frac{1}{n-p} \sum_{t=p+1}^{n} \tilde{Y}_{t-1}' \tilde{Y}_{t-1}' \right]$$
$$\xrightarrow{a.s.} E(\tilde{Y}_{t-1}' \tilde{Y}_{t-1}').$$

$\square$

Theorem 2.7. Under the assumptions of Lemma 2.1,

$$\hat{\phi}_n \xrightarrow{a.s.} \phi_0 \quad , \text{ as } n \to \infty .$$

Proof. First,

$$\hat{\phi}_n - \phi_0 = \left(\sum_{t=p+1}^{n} \tilde{Y}_{t-1}' \tilde{Y}_{t-1}' \right)^{-1} \cdot$$
$$\left[\sum_{t=p+1}^{n} y_t \tilde{Y}_{t-1}' - \sum_{t=p+1}^{n} \tilde{Y}_{t-1}' \tilde{Y}_{t-1}' \phi_0 \right]$$

$$= \left(\sum_{t=p+1}^{n} \hat{Y}_{t-1} \tilde{Y}'_{t-1} \right)^{-1} \left[\sum_{t=p+1}^{n} (y_t - \phi_0' \tilde{Y}_{t-1}) \hat{Y}'_{t-1} \right]$$

$$= \left(\sum_{t=p+1}^{n} \hat{Y}_{t-1} \hat{Y}'_{t-1} \right)^{-1} \sum_{t=p+1}^{n} \hat{Y}_{t-1} \varepsilon_t$$

$$= \left(\frac{1}{n} \sum_{t=p+1}^{n} \hat{Y}_{t-1} \tilde{Y}'_{t-1} \right)^{-1} \left(\frac{1}{n} \sum_{t=p+1}^{n} \hat{Y}_{t-1} \varepsilon_t \right). \qquad (2.7)$$

By Lemma 2.1, $\frac{1}{n} \sum_{t=p+1}^{n} \hat{Y}_{t-1} \hat{Y}'_{t-1} \xrightarrow{a.s.} E(\hat{Y}_{t-1} \hat{Y}'_{t-1}) > 0$.

Similar to Lemma 2.1, by ergodic theorem and Th. 1.5, we can show that

$$\frac{1}{n} \sum_{t=p+1}^{n} \hat{Y}_{t-1} \varepsilon_t \xrightarrow{a.s.} E(\hat{Y}_{t-1} \varepsilon_t) = 0.$$

Furthermore, by (2.7), we have that

$$\hat{\phi}_n - \phi_0 \xrightarrow{a.s.} 0.$$

$\square$

Theorem 2.8. Under the assumptions of Lemma 2.1,

$$\sqrt{n} (\hat{\phi}_n - \phi_0) \xrightarrow{\mathcal{L}} N(0, \sigma^2 E^{-1}(\hat{Y}_t \hat{Y}'_t)).$$

Proof. By (2.7), we have that

$$\sqrt{n} (\hat{\phi}_n - \phi_0) = \left(\frac{1}{n} \sum_{t=p+1}^{n} \tilde{Y}_{t-1} \hat{Y}'_{t-1} \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \hat{Y}_{t-1} \varepsilon_t \right) \qquad (2.8)$$

We first prove that

$$W_n = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \hat{Y}_{t-1} \varepsilon_t \xrightarrow{\mathcal{L}} N(0, \sigma^2 E(\hat{Y}_{t-1} \hat{Y}_{t-1}')) \qquad (2.9)$$

$$\underset{W_n}{\equiv} \qquad\qquad \equiv W$$

For any constant c with $c'c \neq 0$, by Theorem 2.3 (Cramér-Wold device), if we can prove that

$$c' W_n \xrightarrow{\mathcal{L}} c W, \qquad (2.10)$$

then (2.9) holds. Now, we prove (2.10).

$$c' W_n = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} c' \hat{Y}_{t-1} \varepsilon_t \qquad (2.11)$$

Let $\xi_t = c' \hat{Y}_{t-1} \varepsilon_t$. By Theorem 1.5 and Theorem 1.4, ξ_t is strictly stationary and ergodic, and

$$E \xi_t^2 = c' E(\tilde{Y}_{t-1} \varepsilon_t^2 \hat{Y}_{t-1}')c = c' E(\hat{Y}_{t-1} \tilde{Y}_{t-1}' \varepsilon_t^2)c$$

$$= c' E[E(\hat{Y}_{t-1} \hat{Y}_{t-1}' \varepsilon_t^2 | \mathcal{F}_{t-1})] c$$

$$= \sigma^2 c' E(\hat{Y}_{t-1} \hat{Y}_{t-1}')c.$$

By central limit theorem 2.3,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \xi_t \xrightarrow{\mathcal{L}} N(0, \sigma^2 c' E(\hat{Y}_{t-1} \hat{Y}_{t-1}')c)$$

$$= c W$$

Thus, (2.10) holds and hence (2.9) holds. Furthermore by Lemma 2.1, we know that $\sqrt{n}(\hat{\Phi}_n - \Phi_0) \xrightarrow{\mathcal{L}} N(0, \sigma^2 E^{-1}(\hat{Y}_t \hat{Y}_t'))$. $\square$