Chapter 6

Sampling properties of ARMA parameter estimators

In this section we obtain the sampling properties of estimates of the parameters in an ARMA process

\[ X_t = \sum_{i=1}^{p} \phi_i^{(0)} X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j^{(0)} \varepsilon_{t-j}, \quad (6.1) \]

where \( \{\varepsilon_t\} \) are iid random variables with mean zero and \( \text{var}(\varepsilon_t) = \sigma^2 \). Let \( \phi_0 = (\phi_1^{(0)}, \ldots, \phi_p^{(0)}) \) and \( \theta_0 = (\theta_1^{(0)}, \ldots, \theta_q^{(0)}) \) and \( \varphi_0 = (\phi_0, \theta_0) \).

6.1 Asymptotic properties of the Hannan and Rissanen estimation method

In this section we will derive the sampling properties of the Hannan-Rissanen estimator. We will obtain an almost sure rate of convergence (this will be the only estimator where we obtain an almost sure rate). Typically obtaining only sure rates can be more difficult than obtaining probabilistic rates, moreover the rates can be different (worse in the almost sure case). We now illustrate why that is with a small example. Suppose \( \{X_t\} \) are iid random variables with mean zero and variance one. Let \( S_n = \sum_{t=1}^{n} X_t \). It can easily be shown that

\[ \text{var}(S_n) = \frac{1}{n} \text{ therefore } S_n = O_p\left(\frac{1}{\sqrt{n}}\right). \quad (6.2) \]

However, from the law of iterated logarithm we have for any \( \varepsilon > 0 \)

\[ P(S_n \geq (1 + \varepsilon)\sqrt{2n \log \log n} \text{ infinitely often}) = 0P(S_n \geq (1 - \varepsilon)\sqrt{2n \log \log n} \text{ infinitely often}) = 1 \quad (6.3) \]

Comparing (6.2) and (6.3) we see that for any given trajectory (realisation) most of the time \( \frac{1}{n} S_n \) will be within the \( O\left(\frac{1}{\sqrt{n}}\right) \) bound but there will be excursions above when it to the \( O\left(\frac{\log \log n}{\sqrt{n}}\right) \) bound. In other words we cannot say that \( \frac{1}{n} S_n = \left(\frac{1}{\sqrt{n}}\right) \) almost surely, but we can say that This basically means that

\[ \frac{1}{n} S_n = O\left(\frac{\sqrt{2 \log \log n}}{\sqrt{n}}\right) \text{ almost surely.} \]
Hence the probabilistic and the almost sure rates are (slightly) different. Given this result is true for the average of iid random variables, it is likely that similar results will hold true for various estimators.

In this section we derive an almost sure rate for Hannan-Rissanen estimator, this rate will be determined by a few factors (a) an almost sure bound similar to the one derived above (b) the increasing number of parameters \( p_n \) (c) the bias due to estimating only a finite number of parameters when there are an infinite number in the model.

We first recall the algorithm:

(i) Use least squares to estimate \( \{b_j\}^{p_n}_{j=1} \) and define

\[
\hat{b}_n = \hat{R}_n^{-1}\hat{r}_n,
\]

where \( \hat{b}_n' = (\hat{b}_{1,n}, \ldots, \hat{b}_{p_n,n}) \),

\[
\hat{R}_n = \sum_{t=p_n+1}^{T} X_{t-1}X'_{t-1} \quad \hat{r}_n = \sum_{t=p_n+1}^{T} X_t X_{t-1}
\]

and \( X'_{t-1} = (X_{t-1}, \ldots, X_{t-p}) \).

(ii) Estimate the residuals with

\[
\tilde{\epsilon}_t = X_t - \sum_{j=1}^{p_n} \hat{b}_{j,n}X_{t-j},
\]

(iii) Now use as estimates of \( \phi_0 \) and \( \theta_0 \) \( \tilde{\phi}_n, \tilde{\theta}_n \) where

\[
\tilde{\phi}_n, \tilde{\theta}_n = \arg \min \sum_{t=p_n+1}^{T} (X_t - \sum_{j=1}^{p} \phi_jX_{t-j} - \sum_{i=1}^{q} \theta_i\tilde{\epsilon}_{t-i})^2.
\]

We note that the above can easily be minimised. In fact

\[
(\tilde{\phi}_n, \tilde{\theta}_n) = \hat{R}_n^{-1}\tilde{s}_n
\]

where

\[
\hat{R}_n = \frac{1}{n} \sum_{t=p_n+1}^{T} \hat{Y}_t\hat{Y}_t' \quad \hat{s}_n = \frac{1}{T} \sum_{t=p_n+1}^{T} \hat{Y}_t X_t,
\]

\[
\hat{Y}_t' = (X_{t-1}, \ldots, X_{t-p}, \tilde{\epsilon}_{t-1}, \ldots, \tilde{\epsilon}_{t-q}). \text{ Let } \hat{\phi}_n = (\tilde{\phi}_n, \tilde{\theta}_n).
\]

We observe that in the second stage of the scheme where the estimation of the ARMA parameters are done, it is important to show that the empirical residuals are close to the true residuals. That is \( \tilde{\epsilon}_t = \epsilon_t + o(1) \). We observe that from the definition of \( \tilde{\epsilon}_t \), this depends on the rate of convergence of the AR estimators \( \hat{b}_{j,n} \)

\[
\tilde{\epsilon}_t = X_t - \sum_{j=1}^{p_n} \hat{b}_{j,n}X_{t-j}
\]

\[
= \epsilon_t + \sum_{j=1}^{p_n} (\hat{b}_{j,n} - b_j)X_{t-j} - \sum_{j=p_n+1}^{\infty} b_jX_{t-j}.
\]
Hence
\[ |\hat{\varepsilon}_t - \varepsilon_t| \leq \left| \sum_{j=1}^{p_n} (\hat{b}_{j,n} - b_j) X_{t-j} \right| + \left| \sum_{j=p_n+1}^{\infty} b_j X_{t-j} \right|. \]  \hspace{1cm} (6.7)

Therefore to study the asymptotic properties of \( \hat{\varphi}_n = \hat{\phi}_n, \hat{\theta}_n \) we need to

\begin{itemize}
  \item Obtain a rate of convergence for \( \sup_j |\hat{b}_{j,n} - b_j| \).
  \item Obtain a rate for \( |\hat{\varepsilon}_t - \varepsilon_t| \).
  \item Use the above to obtain a rate for \( \hat{\varphi}_n = (\hat{\phi}_n, \hat{\theta}_n) \).
\end{itemize}

We first want to obtain the uniform rate of convergence for \( \sup_j |\hat{b}_{j,n} - b_j| \). Deriving this is technically quite challenging. We state the rate in the following theorem, an outline of the proof can be found in Section 6.1.1. The proofs uses results from mixingale theory which can be found in Chapter 9.

**Theorem 6.1.1** Suppose that \( \{X_t\} \) is from an ARMA process where the roots of the true characteristic polynomials \( \phi(z) \) and \( \theta(z) \) both have absolute value greater than \( 1 + \delta \). Let \( \hat{b}_n \) be defined as in (6.4), then we have almost surely
\[ \|\hat{b}_n - b_n\|_2 = O\left( p_n^2 \sqrt{\frac{(\log \log n)^{1+\gamma} \log n}{n}} + \frac{p_n^3}{n} + p_n \rho_n \right) \]
for any \( \gamma > 0 \).

**PROOF.** See Section 6.1.1.

**Corollary 6.1.1** Suppose the conditions in Theorem 6.1.1 are satisfied. Then we have
\[ |\hat{\varepsilon}_t - \varepsilon_t| \leq p_n \max_{1 \leq j \leq p_n} |\hat{b}_{j,n} - b_j| Z_{t,p_n} + K \rho_n Y_{t-p_n}, \]  \hspace{1cm} (6.8)

where \( Z_{t,p_n} = \frac{1}{p_n} \sum_{t=1}^{p_n} |X_{t-j}| \) and \( Y_t = \sum_{t=1}^{p_n} \rho^j |X_t| \),
\[ \frac{1}{n} \sum_{t=p_n+1}^{n} |\hat{\varepsilon}_{t-i} X_{t-j} - \varepsilon_{t-i} X_{t-j}| = O(p_n Q(n) + \rho_n) \]  \hspace{1cm} (6.9)

\[ \frac{1}{n} \sum_{t=p_n+1}^{n} |\hat{\varepsilon}_{t-i} \hat{\varepsilon}_{t-j} - \varepsilon_{t-i} \varepsilon_{t-j}| = O(p_n Q(n) + \rho_n) \]  \hspace{1cm} (6.10)

where \( Q(n) = p_n^2 \sqrt{\frac{(\log \log n)^{1+\gamma} \log n}{n}} + \frac{p_n^3}{n} + p_n \rho_n \).
Now in the \( \tilde{\epsilon}_t \) gives us an error term is a bound for \( \tilde{\epsilon}_t \).

To prove (6.10) we use a similar method, hence we omit the details.

**Proof.** Using (6.7) we immediately obtain (6.8).

To obtain (6.9) we use (6.7) to obtain

\[
\frac{1}{n} \sum_{t=p_n+1}^{n} |\tilde{\epsilon}_{t-1} X_{t-j} - \epsilon_{t-1} X_{t-j}| \leq \frac{1}{n} \sum_{t=p_n+1}^{n} |\tilde{X}_{t-j}||\tilde{\epsilon}_{t-i} - \epsilon_{t-i}|
\]

\[
\leq O(p_n Q(n)) \frac{1}{n} \sum_{t=p_n+1}^{n} |X_t||Z_{t,p_n}| + O(p^p n) \frac{1}{n} \sum_{t=p_n+1}^{n} |X_t||Y_{t-p_n}|
\]

\[
= O(p_n Q(n) + p^{p_n}).
\]

To prove (6.10) we use a similar method, hence we omit the details.

We apply the above result in the theorem below.

**Theorem 6.1.2** Suppose the assumptions in Theorem 6.1.1 are satisfied. Then

\[
\|\tilde{\phi}_n - \phi_0\|_2 = O \left( \frac{p_n^2}{n} \sqrt{\frac{(\log \log n)^{1+\gamma} \log n}{n}} + \frac{p_n^4}{n} + p_n^2 p^{p_n} \right),
\]

for any \( \gamma > 0 \), where \( \tilde{\phi}_n = (\tilde{\phi}_n, \tilde{\theta}_n) \) and \( \phi_0 = (\phi_0, \theta_0) \).

**Proof.** We note from the definition of \( \tilde{\phi}_n \) that

\[
(\tilde{\phi}_n - \phi_0) = \tilde{R}_n^{-1}(\tilde{s}_n - \tilde{R}_n \phi_0).
\]

Now in the \( \tilde{R}_n \) and \( \tilde{s}_n \) we replace the estimated residuals \( \tilde{\epsilon}_n \) with the true unobserved residuals. This gives us

\[
(\tilde{\phi}_n - \phi_0) = \tilde{R}_n^{-1}(s_n - R_n \phi_0) + (\tilde{R}_n^{-1} s_n - \tilde{R}_n^{-1} \tilde{s}_n)
\]

\[
\tilde{R}_n = \frac{1}{n} \sum_{t = \max(p,q)}^{n} Y_t Y_t^\prime s_n = \frac{1}{n} \sum_{t = \max(p,q)}^{n} Y_t X_t,
\]

\( Y_t^\prime = (X_{t-1}, \ldots, X_{t-p}, \epsilon_{t-1}, \ldots, \epsilon_{t-q}) \) (recalling that \( Y_t^\prime = (X_{t-1}, \ldots, X_{t-p}, \tilde{\epsilon}_{t-1}, \ldots, \tilde{\epsilon}_{t-q}) \). The error term is

\[
(\tilde{R}_n^{-1} s_n - \tilde{R}_n^{-1} \tilde{s}_n) = \tilde{R}_n^{-1}(\tilde{R}_n - R_n) \tilde{R}_n^{-1} s_n + \tilde{R}_n^{-1}(s_n - \tilde{s}_n).
\]

Now, almost surely \( \tilde{R}_n^{-1}, R_n^{-1} = O(1) \) (if \( \mathbb{E}(R_n) \) is non-singular). Hence we only need to obtain a bound for \( \tilde{R}_n - R_n \) and \( s_n - \tilde{s}_n \). We recall that

\[
\tilde{R}_n - R_n = \frac{1}{n} \sum_{t = p_n+1}^{n} (\tilde{Y}_t \tilde{Y}_t^\prime - Y_t Y_t^\prime),
\]

hence the terms differ where we replace the estimated \( \tilde{\epsilon}_t \) with the true \( \epsilon_t \), hence by using (6.9) and (6.10) we have almost surely

\[
|\tilde{R}_n - R_n| = O(p_n Q(n) + p^{p_n}) \text{ and } |\tilde{s}_n - s_n| = O(p_n Q(n) + p^{p_n}).
\]

67
Therefore by substituting the above into (6.12) we obtain

\[
(\varphi_n - \varphi_0) = R_n^{-1} (s_n - R_n \varphi_0) + O(p_n Q(n) + p^p_n).
\]  

(6.12)

Finally using straightforward algebra it can be shown that

\[
s_n - R_n \varphi_n = \frac{1}{n} \sum_{t=\max(p,q)}^n \varepsilon_t Y_t.
\]

By using Theorem 6.1.3, below, we have

\[
\|s_n - R_n \varphi_n\| = O((p+q)\sqrt{\frac{(\log \log n)^{1+\gamma} \log n}{n}}). 
\]

Substituting the above bound into (6.12), and noting that $O(Q(n))$ dominates $O(\sqrt{\frac{(\log \log n)^{1+\gamma} \log n}{n}})\|s_n - R_n \varphi_n\|$ gives

\[
\|\varphi_n - \varphi_n\|_2 = O\left(\frac{p_n}{n} (\log \log n)^{1+\gamma} \frac{\log n}{n} + p_n^2 p_n^3\right)
\]

and the required result. \(\Box\)

6.1.1 Proof of Theorem 6.1.1 (A rate for \(\|\hat{b}_T - b_T\|_2\))

We observe that

\[
\hat{b}_n - b_n = R_n^{-1} (\hat{r}_n - \hat{R}_n b_n) + (\hat{R}_n^{-1} - R_n^{-1})(\hat{r}_n - \hat{R}_n b_n)
\]

where \(b, R_n\), and \(r_n\) are deterministic, with \(b_n = (b_1, \ldots, b_{p_n})\), \((R_n)_{i,j} = \mathbb{E}(X_i X_j)\) and \((r_n)_{i,j} = \mathbb{E}(X_{i} X_{-j})\). Evaluating the Euclidean distance we have

\[
\|b_n - b_n\|_2 \leq \|R_n^{-1}\|_{\text{spec}} \|\hat{r}_n - \hat{R}_n b_n\|_2 + \|R_n^{-1}\|_{\text{spec}} \|\hat{R}_n^{-1}\|_{\text{spec}} \|\hat{R}_n - R_n\|_2 \|\hat{r}_n - \hat{R}_n b_n\|_2
\]

(6.13)

where we used that \(\hat{R}_n^{-1} - \hat{R}_n^{-1} = \hat{R}_n^{-1}(R_n - \hat{R}_n)R_n^{-1}\) and the norm inequalities. Now by using Lemma 3.3.1 we have \(\lambda_{\min}(R_n^{-1}) > \delta/2\) for all \(T\). Thus our aim is to obtain almost sure bounds for \(\|\hat{r}_n - \hat{R}_n b_n\|_2\) and \(\|\hat{R}_n - R_n\|_2\), which requires the lemma below.

Theorem 6.1.3 Let us suppose that \(\{X_t\}\) has an ARMA representation where the roots of the characteristic polynomials \(\phi(z)\) and \(\theta(z)\) lie are greater than 1 + \(\delta\). Then

\begin{align*}
(i) & & \frac{1}{n} \sum_{t=r+1}^n \varepsilon_t X_{t-r} = O\left(\sqrt{\frac{(\log \log n)^{1+\gamma} \log n}{n}}\right) & \quad (6.14) \\
(ii) & & \frac{1}{n} \sum_{t=\max(i,j)}^n X_{t-i} X_{t-j} = O\left(\sqrt{\frac{(\log \log n)^{1+\gamma} \log n}{n}}\right). & \quad (6.15)
\end{align*}

for any \(\gamma > 0\).
PROOF. The result is proved in Chapter 9.2. □

To obtain the bounds we first note that if the there wasn’t an MA component in the ARMA process, in other words \{X_t\} was an AR(\(p\)) process with \(p_n \geq p\), then \(\hat{r}_n - \check{R}_n b_n = \frac{1}{n} \sum_{t=p+1}^{n} \varepsilon_t X_{t-r}\), which has a mean zero. However because an ARMA process has an AR(\(\infty\)) representation and we are only estimating the first \(p_n\) parameters, there exists a ‘bias’ in \(\hat{r}_n - \check{R}_n b_n\). Therefore we obtain the decomposition

\[
(\hat{r}_n - \check{R}_n b_n)_r = \frac{1}{n} \sum_{t=p_n+1}^{n} (X_t - \sum_{j=1}^{\infty} b_j X_{t-j}) X_{t-r} + \frac{1}{n} \sum_{t=p_n+1}^{n} \sum_{j=1}^{\infty} b_j X_{t-j} X_{t-r} \quad (6.16)
\]

\[
= \frac{1}{n} \sum_{t=p_n+1}^{n} \varepsilon_t X_{t-r} + \frac{1}{n} \sum_{t=p_n+1}^{n} \sum_{j=p_n+1}^{\infty} b_j X_{t-j} X_{t-r} \quad (6.17)
\]

Therefore we can bound the bias with

\[
\left| (\hat{r}_n - \check{R}_n b_n)_r - \frac{1}{n} \sum_{t=p_n+1}^{n} \varepsilon_t X_{t-r} \right| \leq K \rho^{p_n} \frac{1}{n} \sum_{t=1}^{n} |X_{t-r}| \sum_{j=1}^{\infty} \rho^j |X_{t-p_n-j}|. \quad (6.18)
\]

Let \(Y_t = \sum_{j=1}^{\infty} \rho^j |X_{t-j}\) and \(S_{n,k,r} = \frac{1}{n} \sum_{t=1}^{n} |X_{t-r}| \sum_{j=1}^{\infty} \rho^j |X_{t-k-j}|\). We note that \(\{Y_t\}\) and \(\{X_t\}\) are ergodic sequences. By applying the ergodic theorem we can show that for a fixed \(k\) and \(r\), \(S_{n,k,r} \xrightarrow{a.s.} \mathbb{E}(X_{t-r} Y_{t-k})\). Hence \(S_{n,k,r}\) are almost surely bounded sequences and

\[
\rho^{p_n} \frac{1}{n} \sum_{t=1}^{n} |X_{t-r}| \sum_{j=1}^{\infty} \rho^j |X_{t-p_n-j}| = O(\rho^{p_n}).
\]

Therefore almost surely we have

\[
\|\hat{r}_n - \check{R}_n b_n\|_2 = \|\frac{1}{n} \sum_{t=p_n+1}^{n} \varepsilon_t X_{t-1}\|_2 + O(p_n \rho^{p_n}).
\]

Now by using (6.14) we have

\[
\|\hat{r}_n - \check{R}_n b_n\|_2 = O \left( p_n \left\{ \sqrt{\frac{(\log \log n)^{1+\gamma} \log n}{n}} + \rho^{p_n} \right\} \right) . \quad (6.19)
\]

This gives us a rate for \(\hat{r}_n - \check{R}_n b_n\). Next we consider \(\check{R}_n\). It is clear from the definition of \(\check{R}_n\) that almost surely we have

\[
(\check{R}_n)_{i,j} - \mathbb{E}(X_i X_j) = \frac{1}{n} \sum_{t=p_n+1}^{n} X_{t-i} X_{t-j} - \mathbb{E}(X_i X_j)
\]

\[
= \frac{1}{n} \sum_{t=\min(i,j)}^{n} [X_{t-i} X_{t-j} - \mathbb{E}(X_i X_j)] - \frac{1}{n} \sum_{t=\min(i,j)}^{p_n} X_{t-i} X_{t-j} + \frac{\min(i,j)}{n} \mathbb{E}(X_i X_j)
\]

\[
= \frac{1}{n} \sum_{t=\min(i,j)}^{T} [X_{t-i} X_{t-j} - \mathbb{E}(X_i X_j)] + O\left( \frac{p_n}{n} \right).
\]

69
Now by using (6.15) we have almost surely
\[ |(\hat{R}_n)_{i,j} - \mathbb{E}(X_i X_j)| = O\left(\frac{p_n}{n} + \sqrt{\frac{\log \log n}{n}} \right). \]
Therefore we have almost surely
\[ \|\hat{R}_n - R_n\|_2 = O\left(\frac{p_n^2}{n} + \sqrt{\frac{\log \log n}{n}} \right) \quad (6.20). \]

We note that by using (6.13), (6.19) and (6.20) we have
\[ \|\hat{b}_n - b_n\|_2 \leq \|R_n^{-1}\|_\text{spec} \|\hat{R}_n^{-1}\|_\text{spec} O\left(\frac{p_n^2}{n} \sqrt{\frac{\log \log n}{n}} \right). \]

As we mentioned previously, because the spectrum of \( X_t \) is bounded away from zero, \( \lambda_{\text{min}}(R_n) \) is bounded away from zero for all \( T \). Moreover, since \( \lambda_{\text{min}}(\hat{R}_n) \geq \lambda_{\text{min}}(R_n) - \lambda_{\text{max}}(\hat{R}_n - R_n) \geq \lambda_{\text{min}}(R_n) - tr((\hat{R}_n - R_n)^2) \), which for a large enough \( n \) is bounded away from zero. Hence we obtain almost surely
\[ \|\hat{b}_n - b_n\|_2 = O\left(\frac{p_n^2}{n} + \sqrt{\frac{\log \log n}{n}} \right) \quad (6.21), \]
thus proving Theorem 6.1.1 for any \( \gamma > 0 \).

### 6.2 Asymptotic properties of the GMLE

Let us suppose that \( \{X_t\} \) satisfies the ARMA representation
\[ X_t - \sum_{i=1}^{p} \phi_i^{(0)} X_{t-i} = \varepsilon_t + \sum_{j=1}^{q} \theta_j^{(0)} \varepsilon_{t-j}, \quad (6.22) \]
and \( \theta_0 = (\theta_1^{(0)}, \ldots, \theta_q^{(0)}) \), \( \phi_0 = (\phi_1^{(0)}, \ldots, \phi_p^{(0)}) \) and \( \sigma_0^2 = \text{var}(\varepsilon_t) \). In this section we consider the sampling properties of the GML estimator, defined in Section 4.3.2. We first recall the estimator. We use as an estimator of \((\theta_0, \phi_0)\), \( \hat{\theta}_n = (\hat{\phi}_n, \hat{\theta}_n, \hat{\sigma}_n) = \arg \min_{(\theta, \phi) \in \Theta} L_n(\phi, \theta, \sigma) \), where
\[ \frac{1}{n} L_n(\phi, \theta, \sigma) = \frac{1}{n} \sum_{t=1}^{n-1} \log r_{t+1}(\sigma, \phi, \theta) + \frac{1}{n} \sum_{t=1}^{n-1} \frac{(X_{t+1} - X_{t+1}^{(\phi, \theta)})^2}{r_{t+1}(\sigma, \phi, \theta)}. \quad (6.23) \]
To show consistency and asymptotic normality we will use the following assumptions.

**Assumption 6.2.1**

(i) \( X_t \) is both invertible and causal.

(ii) The parameter space should be such that all \( \phi(z) \) and \( \theta(z) \) in the parameter space have roots whose absolute value is greater than \( 1 + \delta \). \( \phi_0(z) \) and \( \theta_0(z) \) belong to this space.
Assumption 6.2.1 means for some finite constant $K$ and \( \frac{1}{1 - \delta} \leq \rho < 1 \), we have \( |\phi(z)^{-1}| \leq K \sum_{j=0}^{\infty} |\rho^j||z^j| \) and \( |\phi(z)^{-1}| \leq K \sum_{j=0}^{\infty} |\rho^j||Z^j| \).

To prove the result, we require the following approximations of the GML. Let

\[
\hat{X}_{t+1|t,...}^{(\phi,\theta)} = \sum_{j=1}^{t} b_j(\phi,\theta) X_{t+1-j}.
\]  

(6.24)

This is an approximation of the one-step ahead predictor. Since the likelihood is constructed from the one-step ahead predictors, we can approximated the likelihood \( \frac{1}{n} L_n(\phi,\theta,\sigma) \) with the above and define

\[
\hat{L}_n(\phi,\theta,\sigma) = \log \sigma^2 + \frac{1}{n\sigma^2} \sum_{t=1}^{T-1} (X_{t+1} - \hat{X}_{t+1|t,...}^{(\phi,\theta)})^2.
\]

(6.25)

We recall that \( \hat{X}_{t+1|t,...}^{(\phi,\theta)} \) was derived from \( X_{t+1|t,...}^{(\phi,\theta)} \), which is the one-step ahead predictor of \( X_{t+1} \) given \( X_t, X_{t-1}, \ldots \), this is

\[
X_{t+1|t,...}^{(\phi,\theta)} = \sum_{j=1}^{\infty} b_j(\phi,\theta) X_{t+1-j}.
\]

(6.26)

Using the above we define a approximation of \( \frac{1}{n} L_n(\phi,\theta,\sigma) \) which in practice cannot be obtained (since the infinite past of \( \{X_t\} \) is not observed). Let us define the criterion

\[
\int \frac{1}{n} L_n(\phi,\theta,\sigma) = \log \sigma^2 + \frac{1}{n\sigma^2} \sum_{t=1}^{T-1} (X_{t+1} - X_{t+1|t,...}^{(\phi,\theta)})^2.
\]

(6.27)

In practice \( \frac{1}{n} L_n(\phi,\theta,\sigma) \) can not be evaluated, but it proves to be a convenient tool in obtaining the sampling properties of \( \hat{\phi}_n \). The main reason is because \( \frac{1}{n} L_n(\phi,\theta,\sigma) \) is a function of \( \{X_t\} \) and \( \{X_{t+1|t,...}^{(\phi,\theta)} = \sum_{j=1}^{\infty} b_j(\phi,\theta) X_{t+1-j} \} \) both of these are ergodic (since the ARMA process is ergodic when its roots lie outside the unit circle and the roots of \( \phi,\theta \in \Theta \) are such that they lie outside the unit circle). In contrast looking at \( L_n(\phi,\theta,\sigma) \), which is comprised of \( \{X_{t+1|t,...}^{(\phi,\theta)} \} \), which not an ergodic random variable because \( X_{t+1} \) is the best linear predictor of \( X_{t+1} \) given \( X_t, \ldots, X_1 \) (see the number of elements in the prediction changes with \( t \)). Using this approximation really simplifies the proof, though it is possible to prove the result without using these approximations.

First we obtain the result for the estimators \( \hat{\phi}_n = (\theta_n^*, \phi_n^*, \hat{\sigma}_n) = \arg \min_{(\theta,\phi)\in \Theta} L_n(\phi,\theta,\sigma) \) and then show the same result can be applied to \( \hat{\phi}_n \).

**Proposition 6.2.1** Suppose \( \{X_t\} \) is an ARMA process which satisfies (6.22), and Assumption 6.2.1 is satisfied. Let \( X_{t+1|t,...}^{(\phi,\theta)}, \hat{X}_{t+1|t,...}^{(\phi,\theta)} \) and \( X_{t+1|t,...}^{(\phi,\theta)} \) be the predictors defined in (4.10), (6.24) and (6.26), obtained using the parameters \( \phi = \{\phi_j\} \) and \( \theta = \{\theta_i\} \), where the roots the corresponding characteristic polynomial \( \phi(z) \) and \( \theta(z) \) have absolute value greater than \( 1 + \delta \). Then

\[
|X_{t+1|t,...}^{(\phi,\theta)} - \hat{X}_{t+1|t,...}^{(\phi,\theta)}| \leq \frac{\rho^t}{1 - \rho} \sum_{i=1}^{t} \rho^i |X_i|,
\]

(6.28)
\[ \mathbb{E}(X_{t+1}^{(\phi, \theta)} - \tilde{X}_{t+1}^{(\phi, \theta)})^2 \leq K \rho^t, \quad (6.29) \]

\[ |\tilde{X}_{t+1|t,...}(1) - X_{t+1|t,...}| = \left| \sum_{j=t+1}^{\infty} b_j(\phi, \theta)X_{t+1-j} \right| \leq K \rho^t \sum_{j=0}^{\infty} \rho^j |X_{-j}|, \quad (6.30) \]

\[ \mathbb{E}(X_{t+1}^{(\phi, \theta)} - \tilde{X}_{t+1}^{(\phi, \theta)})^2 \leq K \rho^t \quad (6.31) \]

and

\[ |r_t(\sigma, \phi, \theta) - \sigma^2| \leq K \rho^t \quad (6.32) \]

for any $1/(1 + \delta) < \rho < 1$ and $K$ is some finite constant.

PROOF. The proof follows closely the proof of Proposition 6.2.1. First we define a separate ARMA process \{\(Y_t\)\}, which is driven by the parameters $\theta$ and $\phi$ (recall that \{\(X_t\)\} is drive by the parameters $\theta_0$ and $\phi_0$). That is $Y_t$ satisfies $Y_t - \sum_{j=1}^{\rho} \phi_j Y_{t-j} = \varepsilon_t + \sum_{j=1}^{\theta} \theta_j \varepsilon_{t-j}$. Recalling that $X_{t+1|t}^{(\phi, \theta)}$ is the best linear predictor of $X_{t+1}$ given $X_t, \ldots, X_1$ and the variances of $\{Y_t\}$ (noting that it is the process driven by $\theta$ and $\phi$), we have

\[ X_{t+1|t}^{(\phi, \theta)} = \sum_{j=1}^{t} b_j(\phi, \theta)X_{t+1-j} + (\sum_{j=t+1}^{\infty} b_j(\phi, \theta)r_{t,j}(\phi, \theta)\Sigma_t(\phi, \theta)^{-1})X_t, \quad (6.33) \]

where $\Sigma_t(\phi, \theta)_{s,t} = \mathbb{E}(Y_s Y_t)$, \((r_{t,j})_i = \mathbb{E}(Y_{t-i}Y_{-j})\) and \(X_t' = (X_t, \ldots, X_1)\). Therefore

\[ X_{t+1|t}^{(\phi, \theta)} - \tilde{X}_{t+1|t,...} = (\sum_{j=t+1}^{\infty} b_j(\phi, \theta)r_{t,j}\Sigma_t(\phi, \theta)^{-1})X_t. \]

Since the largest eigenvalue of $\Sigma_t(\phi, \theta)^{-1}$ is bounded (see Lemma 3.3.1) and $|(r_{t,j})_i| = |\mathbb{E}(Y_{t-i}Y_{-j})| \leq K \rho^{t-1} + j$ we obtain the bound in (6.28). Taking expectations, we have

\[ \mathbb{E}(X_{t+1|t}^{(\phi, \theta)} - \tilde{X}_{t+1|t,...})^2 = (\sum_{j=t+1}^{\infty} b_j(\phi, \theta)r_{t,j})^2 \Sigma_t(\phi, \theta)^{-2} \Sigma_t(\phi, \theta_0)^{-1} \Sigma_t(\phi, \theta)^{-1}(\sum_{j=t+1}^{\infty} b_{t+j}r_{t,j}). \]

Now by using the same arguments given in the proof of (3.10) we obtain (6.29).

To prove (6.31) we note that

\[ \mathbb{E}(\tilde{X}_{t+1|t,...}(1) - X_{t+1|t,...})^2 = \mathbb{E}(\sum_{j=t+1}^{\infty} b_j(\phi, \theta)X_{t+1-j})^2 = \mathbb{E}(\sum_{j=1}^{\infty} b_{t+j}(\phi, \theta)X_{-j})^2, \]

now by using (2.9), we have $|b_{t+j}(\phi, \theta)| \leq K \rho^{t+j}$, for $\frac{1}{1+\delta} < \rho < 1$, and the bound in (6.30).

Using this we have $\mathbb{E}(\tilde{X}_{t+1|t,...}(1) - X_{t+1|t,...})^2 \leq K \rho^t$, which proves the result. \(\square\)
Using \( \varepsilon_t = X_t - \sum_{j=1}^{\infty} b_j(\phi_0, \theta_0)X_{t-j} \) and substituting this into \( \mathcal{L}_n(\phi, \theta, \sigma) \) gives

\[
\frac{1}{n} \mathcal{L}_n(\phi, \theta, \sigma) = \log \sigma^2 + \frac{1}{n \sigma^2} (X_t - \sum_{j=1}^{\infty} b_j(\phi, \theta)X_{t+1-j})^2
\]

\[
= \frac{1}{n} \mathcal{L}_n(\phi, \theta, \sigma) \log \sigma^2 + \frac{1}{n \sigma^2} \sum_{t=1}^{T-1} \{ \theta(B)^{-1}\phi(B)X_t \} \{ \theta(B)^{-1}\phi(B)X_t \}
\]

\[
= \log \sigma^2 + \frac{1}{n \sigma^2} \sum_{t=1}^{n} \varepsilon_t^2 + \frac{2}{n} \sum_{t=1}^{\infty} \varepsilon_t \left( \sum_{j=1}^{\infty} b_j(\phi, \theta)X_{t-j} \right)
\]

\[
+ \frac{1}{n} \sum_{t=1}^{n} \left( \sum_{j=1}^{\infty} (b_j(\phi, \theta) - b_j(\phi_0, \theta_0))X_{t-j} \right)^2.
\]

**Remark 6.2.1 (Derivatives involving the Backshift operator)** Consider the transformation

\[
\frac{1}{1 - \theta B} X_t = \sum_{j=0}^{\infty} \theta^j B^j X_t = \sum_{j=0}^{\infty} \theta^j X_{t-j}.
\]

Suppose we want to differentiate the above with respect to \( \theta \), there are two ways this can be done. Either differentiate \( \sum_{j=0}^{\infty} \theta^j X_{t-j} \) with respect to \( \theta \) or differentiate \( \frac{1}{1 - \theta B} \) with respect to \( \theta \). In other words

\[
d \frac{1}{1 - \theta B} X_t = \frac{-B}{(1 - \theta B)^2} X_t = \sum_{j=0}^{\infty} j \theta^j X_{t-j}.
\]

Often it is easier to differentiate the operator. Suppose that \( \theta(B) = 1 + \sum_{j=1}^{p} \theta_j B^j \) and \( \phi(B) = 1 - \sum_{j=1}^{q} \phi_j B^j \), then we have

\[
d \frac{\phi(B)}{\theta(B)} X_t = -\frac{B^j \phi(B)}{\theta(B)^2} X_t = -\phi(B) X_{t-j}
\]

\[
d \frac{\phi(B)}{\theta_j(B)} X_t = -\frac{B^j}{\theta(B)^2} X_t = -\frac{1}{\theta(B)^2} X_{t-j}.
\]

Moreover in the case of squares we have

\[
d \frac{(\phi(B) X_t)^2}{\theta(B)} = -2(\phi(B) X_t)(\frac{\phi(B)}{\theta(B)^2} X_{t-j}), \quad d \frac{(\phi(B) X_t)^2}{\theta(B) X_t} = -2(\phi(B) X_t)(\frac{1}{\theta(B)^2} X_{t-j}).
\]

Using the above we can easily evaluate the gradient of \( \frac{1}{n} \mathcal{L}_n \)

\[
\frac{1}{n} \nabla_{\theta} \mathcal{L}_n(\phi, \theta, \sigma) = -\frac{2}{\sigma^2} \sum_{t=1}^{n} (\theta(B)^{-1}\phi(B)X_t) \frac{\phi(B)}{\theta(B)^2} X_{t-i}
\]

\[
\frac{1}{n} \nabla_{\phi} \mathcal{L}_n(\phi, \theta, \sigma) = -\frac{2}{n \sigma^2} \sum_{t=1}^{n} (\theta(B)^{-1}\phi(B)X_t) \frac{1}{\theta(B)} X_{t-j}
\]

\[
\frac{1}{n} \nabla_{\sigma^2} \mathcal{L}_n(\phi, \theta, \sigma) = \frac{1}{\sigma^2} - \frac{1}{n \sigma^4} \sum_{t=1}^{n} \left( X_t - \sum_{j=1}^{\infty} b_j(\phi, \theta)X_{t-j} \right)^2.
\]
Let $\nabla = (\nabla \phi, \nabla \theta, \nabla \sigma^2)$. We note that the second derivative $\nabla^2 \mathcal{L}_n$ can be defined similarly.

**Lemma 6.2.1** Suppose Assumption 6.2.1 holds. Then

$$\sup_{\phi, \theta \in \Theta} \| \frac{1}{n} \nabla \mathcal{L}_n \|_2 \leq KS_n \sup_{\phi, \theta \in \Theta} \| \frac{1}{n} \nabla^3 \mathcal{L}_n \|_2 \leq KS_n$$

(6.35)

for some constant $K$,

$$S_n = \frac{1}{n} \sum_{r_1, r_2 = 0}^{\max(p, q)} \sum_{t=1}^{n} Y_{t-r_1} Y_{t-r_2}$$

(6.36)

where

$$Y_t = K \sum_{j=0}^{\infty} \rho^j \cdot |X_{t-j}|.$$ for any $\frac{1}{1+\delta} < \rho < 1$.

**PROOF.** The proof follows from the the roots of $\phi(z)$ and $\theta(z)$ having absolute value greater than $1 + \delta$. \hfill \Box

Define the expectation of the likelihood $\mathcal{L}(\phi, \theta, \sigma) = \mathbb{E}(\frac{1}{n} \mathcal{L}_n(\phi, \theta, \sigma))$. We observe

$$\mathcal{L}(\phi, \theta, \sigma) = \log \sigma^2 + \frac{\sigma_0^2}{\sigma^2} + \frac{1}{\sigma^2} \mathbb{E}(Z_t(\phi, \theta)^2)$$

where

$$Z_t(\phi, \theta) = \sum_{j=1}^{\infty} (b_j(\phi, \theta) - b_j(\phi_0, \theta_0)) X_{t-j}$$

**Lemma 6.2.2** Suppose that Assumption 6.2.1 are satisfied. Then for all $\theta, \phi, \theta \in \Theta$ we have

(i) $\frac{1}{n} \nabla^i \mathcal{L}_n(\phi, \theta, \sigma)$ $\xrightarrow{a.s.} \nabla^i \mathcal{L}(\phi, \theta, \sigma)$ for $i = 0, 1, 2, 3$.

(ii) Let $S_n$ defined in (6.36), then $S_n \xrightarrow{a.s.} \mathbb{E}(\sum_{r_1, r_2 = 0}^{\max(p, q)} \sum_{t=1}^{n} Y_{t-r_1} Y_{t-r_2})$.

**PROOF.** Noting that the ARMA process $\{X_t\}$ are ergodic random variables, then $\{Z_t(\phi, \theta)\}$ and $\{Y_t\}$ are ergodic random variables, the result follows immediately from the Ergodic theorem. We use these results in the proofs below.

**Theorem 6.2.1** Suppose that Assumption 6.2.1 is satisfied. Let $(\hat{\theta}_n, \hat{\phi}_n, \hat{\sigma}_n) = \arg \min \mathcal{L}_n(\theta, \phi, \sigma)$ (noting the practice that this cannot be evaluated). Then we have

(i) $(\hat{\theta}_n, \hat{\phi}_n, \hat{\sigma}_n) \xrightarrow{a.s.} (\theta_0, \phi_0, \sigma_0)$. 

74
convergence and equicontinuity of $\nabla L$ where $\hat{L}$ surely bounded. This implies that $\Box$

Now by using (6.35) we have $L$ unique minimum of $(\cdot)$ (i). To show stochastic equicontinuity we note that for any $(x, \theta_0, \sigma_0)$ we have $\sigma_2^2$ is positive and zero at $(\theta_0, \sigma_0)$ it is clear that $\theta_0, \sigma_0$ is the minimum of $L$. We will assume for now it is the unique minimum. Pointwise convergence is an immediate consequence of Lemma 6.2.2(i). To show stochastic equicontinuity we note that for any $\varphi_1 = (\phi_1, \theta_1, \sigma_1)$ and $\varphi_2 = (\phi_2, \theta_2, \sigma_2)$ we have by the mean value theorem

$$L_n(\phi_1, \theta_1, \sigma_1) - L_n(\phi_2, \theta_2, \sigma_2) = (\varphi_1 - \varphi_2) \nabla L_n(\hat{\phi}, \hat{\theta}, \hat{\sigma}).$$

Now by using (6.35) we have

$$L_n(\phi_1, \theta_1, \sigma_1) - L_n(\phi_2, \theta_2, \sigma_2) \leq S_T||((\phi_1 - \phi_2), (\theta_1 - \theta_2), (\sigma_1 - \sigma_2)||_2.$$
Now we show that $\nabla L_n(\varphi_0)$ is asymptotically normal. By using (6.34) and replacing $X_{t-i} = \phi_0(B)^{-1} \theta_0(B) \varepsilon_{t-i}$ we have

$$\frac{1}{n} \nabla_{\theta_i} L_n(\theta_0, \sigma_0) = \frac{2}{\sigma^2 n} \sum_{t=1}^{n} \varepsilon_t \phi_0(B)^{-1} \varepsilon_{t-i} = \frac{-2}{\sigma^2 n} \sum_{t=1}^{n} \varepsilon_t V_{t-i}, \quad i = 1, \ldots, q$$

$$\frac{1}{n} \nabla_{\phi_j} L_n(\theta_0, \sigma_0) = \frac{2}{\sigma^2 n} \sum_{t=1}^{n} \varepsilon_t \phi_0(B) \varepsilon_{t-j} = \frac{2}{\sigma^2 n} \sum_{t=1}^{T} \varepsilon_t U_{t-j}, \quad j = 1, \ldots, p$$

$$\frac{1}{n} \nabla_{\sigma^2} L_n(\theta_0, \sigma_0) = \frac{1}{\sigma^2} - \frac{1}{\sigma^4} \sum_{t=1}^{T} \varepsilon_t^2 = \frac{1}{\sigma^4} \sum_{t=1}^{T} (\sigma^2 - \varepsilon_t^2),$$

where $U_t = \frac{1}{\phi_0(B)} \varepsilon_t$ and $V_t = \frac{1}{\phi_0(B)} \varepsilon_t$. We observe that $\frac{1}{n} \nabla L_n$ is the sum of vector martingale differences. If $E(\varepsilon_t^4) < \infty$, it is clear that $E((\varepsilon_t U_{t-j})^4) = E((\varepsilon_t V_{t-j})^4) = E((\varepsilon_t^4)E(V_{t-j})^4) < \infty$ and $E((\varepsilon^2 - \varepsilon_t^2)^2) < \infty$. Hence Lindeberg’s condition is satisfied (see the proof given in Section 5.8, for why this is true). Hence we have

$$\sqrt{n} \nabla L_n(\theta_0, \sigma_0) \xrightarrow{D} N(0, \Lambda).$$

Now by using the above and (6.37) we have

$$\sqrt{n}(\hat{\varphi}_n - \varphi_0) = \sqrt{n} \nabla^2 L_n(\hat{\varphi}_n)^{-1} \nabla L_n(\varphi_0) \Rightarrow \sqrt{n}(\hat{\varphi}_n - \varphi_0) \xrightarrow{D} N(0, \sigma^4 \Lambda^{-1}).$$

Thus we obtain the required result. \(\square\)

The above result proves consistency and asymptotically normality of $(\hat{\theta}_n, \hat{\phi}_n, \hat{\sigma}_n)$, which is based on $L_n(\theta, \phi, \sigma)$, which in practice is impossible to evaluate. However we will show below that the gaussian likelihood, $L_n(\theta, \phi, \sigma)$ and is derivatives are sufficiently close to $L_n(\theta, \phi, \sigma)$ such that the estimators $(\hat{\theta}_n, \hat{\phi}_n, \hat{\sigma}_n)$ and the GMLE, $(\tilde{\theta}_n, \tilde{\phi}_n, \tilde{\sigma}_n) = \arg \min \ L_n(\theta, \phi, \sigma)$ are asymptotically equivalent. We use Lemma 6.2.1 to prove the below result.

**Proposition 6.2.2** Suppose that Assumption 6.2.1 hold and $L_n(\theta, \phi, \sigma)$, $\tilde{L}_n(\theta, \phi, \sigma)$ and $L_n(\theta, \phi, \sigma)$ are defined as in (6.23), (6.25) and (6.27) respectively. Then we have for all $(\theta, \phi) \in \text{Theta}$ we have almost surely

$$\sup_{(\phi, \theta, \sigma)} \frac{1}{n} |\nabla(\tilde{L}(\phi, \theta, \sigma) - \nabla L_n(\phi, \theta, \sigma)| = O(\frac{1}{n}) \quad \sup_{(\phi, \theta, \sigma)} \frac{1}{n} |\tilde{L}_n(\phi, \theta, \sigma) - L(\phi, \theta, \sigma)| = O(\frac{1}{n}),$$

for $k = 0, 1, 2, 3$.

**PROOF.** The proof of the result follows from (6.28) and (6.30). We show that result for $\sup_{(\phi, \theta, \sigma)} \frac{1}{n} |\tilde{L}(\phi, \theta, \sigma) - L_n(\phi, \theta, \sigma)|$, a similar proof can be used for the rest of the result.

Let us consider the difference

$$L_n(\phi, \theta) - L_n(\phi, \theta) = \frac{1}{n}(I_n + II_n + III_n),$$

76
where

\[ I_n = \sum_{t=1}^{n-1} \{ r_t(\phi, \theta, \sigma) - \sigma^2 \}, \quad II_n = \sum_{t=1}^{n-1} \frac{1}{r_t(\phi, \theta, \sigma)} (X_{t+1}^{(\phi, \theta)} - X_{t+1|t}^{(\phi, \theta)})^2 \]

\[ III_n = \sum_{t=1}^{n-1} \frac{1}{\sigma^2} \left\{ 2X_{t+1} (X_{t+1|t}^{(\phi, \theta)} - \tilde{X}_{t+1|t, \ldots}^{(\phi, \theta)}) + \left( (X_{t+1|t}^{(\phi, \theta)})^2 - (\tilde{X}_{t+1|t, \ldots}^{(\phi, \theta)})^2 \right) \right\}. \]

Now we recall from Proposition 6.2.1 that

\[ |X_{t+1|t}^{(\phi, \theta)} - \tilde{X}_{t+1|t, \ldots}^{(\phi, \theta)}| \leq K \cdot V_t \frac{\rho^t}{(1 - \rho)} \]

where \( V_t = \sum_{i=1}^t \rho^i |X_i| \). Hence since \( \mathbb{E}(X_t^2) < \infty \) and \( \mathbb{E}(V_t^2) < \infty \) we have that \( \text{sup}_n \mathbb{E}|I_n| < \infty \), \( \text{sup}_n \mathbb{E}|II_n| < \infty \) and \( \text{sup}_n \mathbb{E}|III_n| < \infty \). Hence the sequence \( \{|I_n + II_n + III_n|\}_n \) is almost surely bounded. This means that almost surely

\[ \text{sup}_{\phi, \theta, \sigma} |\mathcal{L}_n(\phi, \theta) - \mathcal{L}_n(\phi, \theta)| = O\left(\frac{1}{n}\right). \]

Thus giving the required result. □

Now by using the above proposition the result below immediately follows.

**Theorem 6.2.2** Let \((\hat{\theta}, \hat{\phi}) = \arg \min L_T(\theta, \phi, \sigma)\) and \((\tilde{\theta}, \tilde{\phi}) = \arg \min \tilde{L}_T(\theta, \phi, \sigma)\)

(i) \((\hat{\theta}, \hat{\phi}) \overset{a.s.}{\to} (\theta_0, \phi_0)\) and \((\tilde{\theta}, \tilde{\phi}) \overset{a.s.}{\to} (\theta_0, \phi_0)\).

(ii) \(\sqrt{T}(\hat{\theta}_T - \theta_0, \hat{\phi}_T - \theta_0) \overset{D}{\to} N(0, \sigma_0^4 \Lambda^{-1})\)

and \(\sqrt{T}(\tilde{\theta}_T - \theta_0, \tilde{\phi}_T - \theta_0) \overset{D}{\to} N(0, \sigma_0^4 \Lambda^{-1})\).

**PROOF.** The proof follows immediately from Proposition 6.2.1. □