6.3.3 Parameter Estimation

In this section we will discuss methods of parameter estimation for ARMA(p,q) assuming that the orders p and q are known.

Method of Moments

In this method we equate the population moments with the sample moments to obtain a set of equations whose solution gives the required estimators. For example, the first population moment is $\mu_1 = \mathrm{E}(X)$ and its sample counterpart is $m_1 = \bar{X}$. This immediately gives

$$\hat{\mu}_1 = \bar{X}.$$

The method of moments gives good estimators for AR models but less efficient estimators for MA or ARMA processes. Hence we will present the method for AR time series. As usual we denote an AR(p) model by

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t.$$

This is a zero-mean model, but the estimation of the mean is straightforward and we will not discuss it further. Here we use the difference equations, where we replace the population autocovariance (central moment of order two) with the sample autocovariance. The first $p+1$ difference equations are

$$\begin{aligned}
&\gamma(0) = \phi_1\gamma(1) + \ldots + \phi_p\gamma(p) + \sigma^2 \\
&\gamma(\tau) = \phi_1\gamma(\tau-1) + \ldots + \phi_p\gamma(\tau-p), \quad \tau = 1, 2, \ldots, p.
\end{aligned}$$

Note, that $q = 0$, so the sum on the right hand side of (6.15) is zero.

In matrix notation we can write

$$\begin{aligned}
&\sigma^2 = \gamma(0) - \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\gamma}_p \\
&\boldsymbol{\Gamma}_p\boldsymbol{\phi} = \boldsymbol{\gamma}_p
\end{aligned}$$

where

$$\begin{aligned}
&\boldsymbol{\Gamma}_p = \{\gamma(i-j)\}_{i,j=1,\ldots,p} \\
&\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^{\mathrm{T}} \\
&\boldsymbol{\gamma}_p = (\gamma(1), \ldots, \gamma(p))^{\mathrm{T}}.
\end{aligned}$$

Replacing $\gamma(\tau)$ by the sample ACVF

$$\widehat{\gamma}(\tau) = \frac{1}{n}\sum_{t=1}^{n-\tau}(X_{t+\tau} - \bar{X})(X_t - \bar{X})$$

we obtain the solution

$$\begin{aligned} \widehat{\sigma}^2 &= \widehat{\gamma}(0) - \widehat{\boldsymbol{\gamma}}_p^{\mathrm{T}} \widehat{\boldsymbol{\Gamma}}_p^{-1} \widehat{\boldsymbol{\gamma}}_p \\ \widehat{\boldsymbol{\phi}} &= \widehat{\boldsymbol{\Gamma}}_p^{-1} \widehat{\boldsymbol{\gamma}}_p. \end{aligned} \tag{6.34}$$

These equations are called **Yule-Walker estimators**. They are often expressed in terms of autocorrelation function rather than autocovariance function. Then we have

$$\begin{aligned} \widehat{\sigma}^2 &= \widehat{\gamma}(0) \left(1 - \widehat{\boldsymbol{\rho}}_p^{\mathrm{T}} \widehat{\boldsymbol{R}}_p^{-1} \widehat{\boldsymbol{\rho}}_p\right) \\ \widehat{\boldsymbol{\phi}} &= \widehat{\boldsymbol{R}}_p^{-1} \widehat{\boldsymbol{\rho}}_p, \end{aligned} \tag{6.35}$$

where

$$\widehat{\boldsymbol{R}}_p = \{\widehat{\rho}(i-j)\}_{i,j=1,2,\ldots,p}$$

is the matrix of the sample autocorrelations and

$$\widehat{\boldsymbol{\rho}}_p = (\widehat{\rho}(1), \ldots, \widehat{\rho}(p))^{\mathrm{T}}$$

is the vector of sample autocorrelations.

Proposition 6.3. *The distribution of the Yule-Walker estimators $\widehat{\boldsymbol{\phi}}$ of the model parameters of a causal AR(p) process*

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t.$$

is asymptotically (as $n \longrightarrow \infty$) normal, in the sense that

$$\sqrt{n}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{d} \mathcal{N}(0, \sigma^2 \boldsymbol{\Gamma}_p^{-1}),$$

and

$$\widehat{\sigma^2} \xrightarrow{p} \sigma^2.$$

Remark 6.12. Note that the matrix equation (6.22) is of the same form as (6.35). Hence, we can use the Durbin-Lewinson algorithm to calculate the estimates. This will give us the values of the sample PACF as well as the estimates of $\boldsymbol{\phi}$.

Proposition 6.4. *The distribution of the sample PACF of a causal AR(p) process is asymptotically normal, that is*

$$\sqrt{n}\widehat{\phi}_{\tau\tau} \xrightarrow{d} \mathcal{N}(0,1) \quad \textit{for } \tau > p.$$

Example 6.8. Consider an AR(2) zero-mean causal process

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t.$$

Then the Yule-Walker estimators are

$$\widehat{\sigma}^2 = \widehat{\gamma}(0)\left(1 - \widehat{\boldsymbol{\rho}}_2^{\mathrm{T}}\widehat{\boldsymbol{R}}_2^{-1}\widehat{\boldsymbol{\rho}}_2\right)$$
$$\widehat{\boldsymbol{\phi}} = \widehat{\boldsymbol{R}}_2^{-1}\widehat{\boldsymbol{\rho}}_2,$$

where

$$\widehat{\boldsymbol{R}}_2 = \begin{pmatrix} \widehat{\rho}(0) & \widehat{\rho}(1) \\ \widehat{\rho}(1) & \widehat{\rho}(0) \end{pmatrix}$$

and

$$\widehat{\boldsymbol{\rho}}_2 = (\widehat{\rho}(1), \widehat{\rho}(2))^{\mathrm{T}}$$
$$\widehat{\boldsymbol{\phi}} = (\widehat{\phi}_1, \widehat{\phi}_2)^{\mathrm{T}}.$$

We can easily invert a 2×2 matrix and calculate the estimators, or we can use the Durbin-Levinson algorithm directly to obtain

$$\widehat{\phi}_{11} = \widehat{\rho}(1) = \frac{\widehat{\phi}_1}{1 - \widehat{\phi}_2}$$
$$\widehat{\phi}_{22} = \frac{\widehat{\rho}(2) - \widehat{\rho}^2(1)}{1 - \widehat{\rho}^2(1)} = \widehat{\phi}_2$$
$$\widehat{\phi}_{21} = \widehat{\rho}(1)[1 - \widehat{\phi}_{22}] = \widehat{\phi}_1.$$

Also, we get

$$\widehat{\sigma}^2 = \gamma(0)\left[1 - (\widehat{\rho}(1), \widehat{\rho}(2))\begin{pmatrix} \widehat{\phi}_1 \\ \widehat{\phi}_2 \end{pmatrix}\right]$$
$$= \gamma(0)[1 - (\widehat{\rho}(1)\widehat{\phi}_1 + \widehat{\rho}(2)\widehat{\phi}_2)]$$

Furthermore, from Proposition 6.3 we can derive the confidence interval for ϕ_i. The proposition says that

$$\sqrt{n}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{d} \mathcal{N}(0, \sigma^2\boldsymbol{\Gamma}_p^{-1}),$$

that is the variance of $\sqrt{n}(\widehat{\phi}_i - \phi_i)$ is the i-th diagonal element of the matrix $\sigma^2\boldsymbol{\Gamma}_p^{-1}$, say v_{ii},

$$v_{ii} = \operatorname{var}[\sqrt{n}(\widehat{\phi}_i - \phi_i)] = n\operatorname{var}(\widehat{\phi}_i - \phi_i) = n\operatorname{var}(\widehat{\phi}_i).$$

Hence,

$$\operatorname{var}(\widehat{\phi}_i) = \frac{1}{n}v_{ii}$$

and the confidence interval is

$$\left[\widehat{\phi}_i - u_\alpha\sqrt{\frac{1}{n}v_{ii}},\ \widehat{\phi}_i + u_\alpha\sqrt{\frac{1}{n}v_{ii}}\right].$$

To calculate the confidence interval for a given data set we replace v_{ii} by its estimate $\hat{v}_{ii}$.

Also, from Proposition 6.4 we have

$$\sqrt{n}\widehat{\phi}_{\tau\tau} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{for } \tau > p,$$

that is

$$\text{var}(\sqrt{n}\widehat{\phi}_{\tau\tau}) \longrightarrow 1 \quad \text{for } \tau > p.$$

This gives the asymptotic result

$$\text{var}(\widehat{\phi}_{\tau\tau}) \longrightarrow \frac{1}{n}.$$

However, we know that the PACF for $\tau > p$ is zero. It means that with probability $1-\alpha$ we have

$$-u_\alpha < \frac{\widehat{\phi}_{\tau\tau} - 0}{\sqrt{\frac{1}{n}}} < u_\alpha.$$

It can be interpreted that the estimate of the PACF indicates a non-significant value of $\phi_{\tau\tau}$ if it is in the interval

$$[-u_\alpha/\sqrt{n},\, u_\alpha/\sqrt{n}].$$

Then, the interval

$$[\widehat{\phi}_{\tau\tau} - u_\alpha/\sqrt{n},\, \widehat{\phi}_{\tau\tau} + u_\alpha/\sqrt{n}].$$

covers zero.

We will do the calculations for the simulated AR(2) process given in Figure 6.2. For these data we have the following values of the sample variance $\widehat{\gamma}(0)$ and the sample autocorrelations $\widehat{\rho}(1)$ and $\widehat{\rho}(2)$

$$\begin{aligned}\widehat{\gamma}(0) &= 1.947669\\ \widehat{\rho}(1) &= 0.66018\\ \widehat{\rho}(2) &= 0.33751.\end{aligned}$$

Then, matrix $\widehat{\boldsymbol{R}}_2$ is equal to

$$\widehat{\boldsymbol{R}}_2 = \begin{pmatrix} 1 & 0.66018 \\ 0.66018 & 1 \end{pmatrix}$$

and its inverse is

$$\widehat{\boldsymbol{R}}_2^{-1} = \begin{pmatrix} 1.77254 & -1.17020 \\ -1.17020 & 1.77254 \end{pmatrix}.$$

Hence, we obtain the following Yule-Walker estimates of the model parameters

$$\begin{pmatrix} \widehat{\phi}_1 \\ \widehat{\phi}_2 \end{pmatrix} = \begin{pmatrix} 1.77254 & -1.17020 \\ -1.17020 & 1.77254 \end{pmatrix} \begin{pmatrix} 0.66018 \\ 0.33751 \end{pmatrix} = \begin{pmatrix} 0.775243 \\ -0.174290 \end{pmatrix}$$

The estimate of the white noise variance is

$$\widehat{\sigma}^2 = 1.947669[1 - (0.66018 \times 0.775243 + 0.33751 \times (-0.174290))] = 1.06542.$$

The series was simulated for $\phi_1 = 0.7$ and $\phi_2 = -0.1$ and a Gaussian White Noise with zero mean and variance equal to 1. These estimates are not far from the true values. Had we not known the true values we would have liked to calculate the confidence intervals for them. There are 200 observations, i.e. $n = 200$, which is big enough to use the asymptotic result given in the Proposition 6.3. To calculate v_{ii} note that

$$\mathbf{\Gamma} = \gamma(0)\boldsymbol{R},$$

which gives

$$\mathbf{\Gamma}^{-1} = \frac{1}{\gamma(0)}\boldsymbol{R}^{-1}.$$

Hence

$$\begin{aligned} \widehat{\sigma}^2\widehat{\mathbf{\Gamma}}^{-1} &= \widehat{\sigma}^2\frac{1}{\widehat{\gamma}(0)}\widehat{\boldsymbol{R}}^{-1} \\ &= \frac{1.06542}{1.947669}\begin{pmatrix} 1.77254 & -1.17020 \\ -1.17020 & 1.77254 \end{pmatrix} \\ &= \begin{pmatrix} 0.969623 & -0.640129 \\ -0.640129 & 0.969623 \end{pmatrix} \end{aligned}$$

and we obtain the estimate of the variance of the parameter estimators

$$\operatorname{var}(\widehat{\phi}_i) = \frac{1}{n}v_{ii} = \frac{1}{200}0.969623 = 0.0048481.$$

The 95% approximate confidence intervals for the model parameters ϕ_1 and ϕ_2 are, respectively

$$\begin{aligned} &[0.775243 - 1.96 \times \sqrt{0.0048481},\ 0.775243 + 1.96 \times \sqrt{0.0048481}] \\ &= [0.638771,\ 0.911714] \end{aligned}$$

$$\begin{aligned} &[-0.17429 - 1.96 \times \sqrt{0.0048481},\ -0.17429 + 1.96 \times \sqrt{0.0048481}] \\ &= [-0.310761,\ -0.037818] \end{aligned}$$

Maximum Likelihood Estimation

The method of Maximum Likelihood Estimation applies to any ARMA(p,q) model

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}.$$

This method requires an assumption on the distribution of the random variable $\boldsymbol{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$. The usual assumption is that the process is Gaussian. Let us denote the p.d.f. of $\boldsymbol{X}$ by

$$f_{\boldsymbol{X}}(X_1, \ldots, X_n; \boldsymbol{\beta}, \sigma^2),$$

where

$$\boldsymbol{\beta} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)^{\mathrm{T}}.$$

Given the values of $\boldsymbol{X}$ the p.d.f. becomes a function of the parameters. It is then denoted by

$$L(\boldsymbol{\beta}, \sigma^2 | x_1, \ldots, x_n)$$

and for the Gaussian process it is

$$L(\boldsymbol{\beta}, \sigma^2 | x_1, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Gamma}_n)}} \exp\left\{ -\frac{1}{2} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{\Gamma}_n^{-1} \boldsymbol{X} \right\}.$$

A more convenient form can be obtained after taking natural logarithm. Then

$$\begin{aligned} l(\boldsymbol{\beta}, \sigma^2 | x_1, \ldots, x_n) &= \ln L(\boldsymbol{\beta}, \sigma^2 | x_1, \ldots, x_n) \\ &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln \det(\boldsymbol{\Gamma}_n) - \frac{1}{2} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{\Gamma}_n^{-1} \boldsymbol{X}. \end{aligned}$$

The Maximum likelihood Estimates are the values of $\boldsymbol{\beta}$ and σ^2 which maximize the function $l(\boldsymbol{\beta}, \sigma^2 | x_1, \ldots, x_n)$. Intuitively, a maximum likelihood estimate is the parameter value for which the observed sample is most likely.

The estimates are usually found numerically using some iterative numerical optimization routines. We will not discuss them here.

The MLE have the property of being asymptotically normally distributed. It is stated in the following proposition.

Proposition 6.5. *The distribution of the MLE $\widehat{\boldsymbol{\beta}}$ of a causal and invertible ARMA(p,q) process is asymptotically normal in the sense that*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \boldsymbol{\Gamma}_{p+q}^{-1}), \tag{6.36}$$

where the $(p+q) \times (p+q)$-dimensional matrix $\boldsymbol{\Gamma}_{p+q}$ depends on the model parameters.

Some Specific Asymptotic Distributions

AR(1): $X_t + \phi X_{t-1} = Z_t$

$$\widehat{\phi} \sim \mathcal{AN}\left[\phi, \frac{1}{n}(1-\phi^2)\right]$$

AR(2): $X_t + \phi_1 X_{t-1} + \phi_2 X_{t-2} = Z_t$

$$\begin{pmatrix} \widehat{\phi}_1 \\ \widehat{\phi}_2 \end{pmatrix} \sim \mathcal{AN}\left[\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \frac{1}{n}\begin{pmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{pmatrix}\right]$$

MA(1): $X_t = Z_t + \theta Z_{t-1}$

$$\widehat{\theta} \sim \mathcal{AN}\left[\theta, \frac{1}{n}(1-\theta^2)\right]$$

MA(2): $X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$

$$\begin{pmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \end{pmatrix} \sim \mathcal{AN}\left[\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \frac{1}{n}\begin{pmatrix} 1-\theta_2^2 & -\theta_1(1+\theta_2) \\ -\theta_1(1+\theta_2) & 1-\theta_2^2 \end{pmatrix}\right]$$

ARMA(1,1): $X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$

$$\begin{pmatrix} \widehat{\phi} \\ \widehat{\theta} \end{pmatrix} \sim \mathcal{AN}\left[\begin{pmatrix} \phi \\ \theta \end{pmatrix}, \frac{1+\phi\theta}{n(\phi+\theta)^2}\begin{pmatrix} (1-\phi^2)(1+\phi\theta) & -(1-\theta^2)(1-\phi^2) \\ -(1-\theta^2)(1-\phi^2) & (1-\theta^2)(1+\phi\theta) \end{pmatrix}\right]$$

Using these results we can construct approximate confidence intervals for the model parameters as in the method of moments.