

3.5: Parameter Estimation

Let $(X_t: t \in \mathbb{Z})$ be a causal and invertible ARMA(p, q)

process with known orders p and q , possibly with mean μ . This section is concerned with estimation procedures for the unknown parameter vector

$$\beta = (\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)^T. \quad (3.5.1)$$

To simplify the estimation procedure, it is assumed that the data has already been adjusted by subtraction of the mean and the discussion is therefore restricted to zero mean ARMA models.

In the following, three estimation methods are introduced. The method of moments works best in case of pure AR processes, while it does not lead to optimal estimation procedures for general ARMA processes. For the latter, more efficient estimators are provided by the maximum likelihood and least squares methods which will be discussed subsequently.

Method 1 (Method of Moments) Since this method is only efficient in their case, the presentation here is restricted to AR(p) processes

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t, t \in \mathbb{Z},$$

where $(Z_t: t \in \mathbb{Z}) \sim \text{WN}(0, \sigma^2)$. The parameter vector β consequently reduces to $(\phi, \sigma^2)^T$ with $\phi = (\phi_1, \dots, \phi_p)^T$ and can be estimated using the *Yule-Walker equations*

$$\Gamma_p \phi = \gamma_p \quad \text{and} \quad \sigma^2 = \gamma(0) - \phi^T \gamma_p,$$

where $\Gamma_p = (\gamma(k-j))_{k,j=1,\dots,p}$ and $\gamma_p = (\gamma(1), \dots, \gamma(p))^T$. Observe that the equations are obtained by the same arguments applied to derive the Durbin-Levinson algorithm in the previous section. The method of moments suggests to replace every quantity in the Yule-Walker equations with their estimated counterparts, which yields the *Yule-Walker estimators*

$$\hat{\phi} = \hat{\Gamma}_p^{-1} \hat{\gamma}_p = \hat{R}_p^{-1} \hat{\rho}_p \quad (3.5.2)$$

$$\hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\gamma}_p^T \hat{\Gamma}_p^{-1} \hat{\gamma}_p = \hat{\gamma}(0) \left[1 - \hat{\rho}_p^T \hat{R}_p^{-1} \hat{\rho}_p \right]. \quad (3.5.3)$$

Therein, $\hat{R}_p = \hat{\gamma}(0)^{-1} \hat{\Gamma}_p$ and $\hat{\rho}_p = \hat{\gamma}(0)^{-1} \hat{\gamma}_p$ with

$\hat{\gamma}(h)$ defined as in (1.2.1). Using $\hat{\gamma}(h)$ as estimator for the ACVF at lag h , a dependence on the sample size n is obtained in an implicit way. This dependence is suppressed in the notation used here. The following theorem contains the limit behavior of the Yule-Walker estimators as n tends to infinity.

Theorem 3.5.1. If $(X_t: t \in \mathbb{Z})$ is a causal AR(p) process, then

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{\mathcal{D}} N(0, \sigma^2 \Gamma_p^{-1}) \quad \text{and} \quad \hat{\sigma}^2 \xrightarrow{P} \sigma^2$$

as $n \rightarrow \infty$, where \xrightarrow{P} indicates convergence in probability.

A proof of this result is given in Section 8.10 of Brockwell and Davis (1991). Since equations (3.5.2) and (3.5.3) have the same structure as the corresponding equations (3.4.3) and (3.4.4), the Durbin-Levinson algorithm can be used to solve recursively for the estimators $\hat{\phi}_h = (\hat{\phi}_{h1}, \dots, \hat{\phi}_{hh})$. Moreover, since ϕ_{hh} is equal to the value of the PACF of $(X_t: t \in \mathbb{Z})$ at lag h , the estimator $\hat{\phi}_{hh}$ can be used as its proxy. Since it is already known that, in the case of AR(p) processes, $\phi_{hh} = 0$ if $h > p$, Theorem (3.5.1) implies immediately the following corollary.

Corollary 3.5.1 If $(X_t: t \in \mathbb{Z})$ is a causal AR(p) process, then

$$\sqrt{n} \hat{\phi}_{hh} \xrightarrow{\mathcal{D}} Z \quad (n \rightarrow \infty)$$

for all $h > p$, where Z stands for a standard normal random variable.

Example 3.5.1. (Yule-Walker estimates for AR(2) processes). Suppose that $n = 144$ values of the autoregressive process $X_t = 1.5X_{t-1} - .75X_{t-2} + Z_t$ have been observed, where $(Z_t: t \in \mathbb{Z})$ is a sequence of independent standard normal variates.

Assume further that $\hat{\gamma}(0) = 8.434$, $\hat{\rho}(1) = 0.834$ and $\hat{\rho}(2) = 0.476$ have been calculated from the data. The Yule-Walker estimators for the parameters are then given by

$$\hat{\phi} = \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix} = \begin{pmatrix} 1.000 & 0.834 \\ 0.834 & 1.000 \end{pmatrix}^{-1} \begin{pmatrix} 0.834 \\ 0.476 \end{pmatrix} = \begin{pmatrix} 1.439 \\ -0.725 \end{pmatrix}$$

and

$$\hat{\sigma}^2 = 8.434 \left[1 - (0.834, 0.476) \begin{pmatrix} 1.439 \\ -0.725 \end{pmatrix} \right] = 1.215.$$

To construct asymptotic confidence intervals using Theorem 3.5.1, the unknown limiting covariance matrix $\sigma^2 \Gamma_p^{-1}$ needs to be estimated. This can be done using the estimator

$$\frac{\hat{\sigma}^2 \hat{\Gamma}_p^{-1}}{n} = \frac{1}{144} \frac{1.215}{8.434} \begin{pmatrix} 1.000 & 0.834 \\ 0.834 & 1.000 \end{pmatrix}^{-1} = \begin{pmatrix} 0.057^2 & -0.003 \\ -0.003 & 0.057^2 \end{pmatrix}.$$

Then, the $1 - \alpha$ level confidence interval for the parameters ϕ_1 and ϕ_2 are computed as

$$1.439 \pm 0.057 z_{1-\alpha/2} \quad \text{and} \quad -0.725 \pm 0.057 z_{1-\alpha/2},$$

respectively, where $z_{1-\alpha/2}$ is the corresponding normal quantile.

Example 3.5.2 (Recruitment Series).

Let us reconsider the recruitment series of Example 3.3.5. There, an AR(2) model was first established as appropriate for the data and the model parameters were then estimated using an ordinary least squares approach. Here, the coefficients will instead be estimated with the Yule-Walker procedure. The R command is

```
> rec.yw = ar.yw(rec, order=2)}
```

The mean estimate can be obtained from `rec.yw$x.mean` as $\hat{\mu} = 62.26$, while the autoregressive parameter estimates and their standard errors are accessed with the commands `rec.yw$ar` and `sqrt(rec.yw$asy.var.coef)` as $\hat{\phi}_1 = 1.3316(.0422)$ and $\hat{\phi}_2 = -.4445(.0422)$. Finally, the variance estimate is obtained from `rec.yw$var.pred` as $\hat{\sigma}^2 = 94.7991$. All values are close to their counterparts in Example 3.3.5.

Example 3.5.3. Consider the invertible MA(1) process $X_t = Z_t + \theta Z_{t-1}$, where $|\theta| < 1$. Using invertibility, each X_t has an infinite autoregressive representation

$$X_t = \sum_{j=1}^{\infty} (-\theta)^j X_{t-j} + Z_t$$

that is nonlinear in the unknown parameter θ to be estimated. The method of moments is here based on solving

$$\hat{\rho}(1) = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = \frac{\hat{\theta}}{1 + \hat{\theta}^2}.$$

for $\hat{\theta}$. The foregoing quadratic equation has the two solutions

$$\hat{\theta} = \frac{1 \pm \sqrt{1 - 4\hat{\rho}(1)^2}}{2\hat{\rho}(1)},$$

of which we pick the invertible one. Note moreover, that $|\hat{\rho}(1)|$ is not necessarily less or equal to $1/2$ which is required for the existence of real solutions. (The theoretical value $|\rho(1)|$, however, is always less than $1/2$ for any MA(1) process, as an easy computation shows). Hence, θ can not always be estimated from given data samples.

Method 2 (Maximum Likelihood Estimation) The innovations algorithm of the previous section applied to a causal ARMA(p, q) process $(X_t: t \in \mathbb{Z})$ gives

$$\hat{X}_{i+1} = \sum_{j=1}^i \theta_{ij}(X_{i+1-j} - \hat{X}_{i+1-j}), \quad 1 \leq i < \max\{p, q\},$$

$$\hat{X}_{i+1} = \sum_{j=1}^p \phi_j X_{i+1-j} + \sum_{j=1}^q \theta_{ij}(X_{i+1-j} - \hat{X}_{i+1-j}), \quad i \geq \max\{p, q\},$$

with prediction error

$$P_{i+1} = \sigma^2 R_{i+1}.$$

In the last expression, σ^2 has been factored out due to reasons that will become apparent from the form of the likelihood function to be discussed below. Recall that the sequence $(X_{i+1} - \hat{X}_{i+1} : i \in \mathbb{Z})$ consists of uncorrelated random variables if the parameters are known. Assuming normality for the errors, we moreover obtain even independence. This can be exploited to define the *Gaussian maximum likelihood estimation (MLE) procedure*. Throughout, it is assumed that $(X_t : t \in \mathbb{Z})$ has zero mean ($\mu = 0$). The parameters of interest are collected in the vectors $\beta = (\phi, \theta, \sigma^2)^T$ and $\beta' = (\phi, \theta)^T$, where $\phi = (\phi_1, \dots, \phi_p)^T$ and $\theta = (\theta_1, \dots, \theta_q)^T$. Assume finally that we have observed the variables X_1, \dots, X_n . Then, the Gaussian likelihood function for the innovations is

$$L(\beta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \left(\prod_{i=1}^n R_i^{1/2} \right) \exp \left(-\frac{1}{2\sigma^2} \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{R_j} \right). \quad (3.5.4)$$

Taking the partial derivative of $\ln L(\beta)$ with respect to the variable σ^2 reveals that the MLE for σ^2 can be calculated from

$$\hat{\sigma}^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n}, \quad S(\hat{\phi}, \hat{\theta}) = \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{R_j}.$$

Therein, $\hat{\phi}$ and $\hat{\theta}$ denote the MLEs of ϕ and θ obtained from minimizing the *profile likelihood* or *reduced likelihood*

$$\ell(\phi, \theta) = \ln \left(\frac{S(\phi, \theta)}{n} \right) + \frac{1}{n} \sum_{j=1}^n \ln(R_j).$$

Observe that the profile likelihood $\ell(\phi, \theta)$ can be computed using the innovations algorithm. The speed of these computations depends heavily on the quality of initial estimates. These are often provided by the non-optimal Yule-Walker procedure. For numerical methods, such as the *Newton-Raphson* and *scoring algorithms*, see Section 3.6 in Shumway and Stoffer (2006).

The limit distribution of the MLE procedure is given as the following theorem. Its proof can be found in Section 8.8 of Brockwell and Davis (1991).

Theorem 3.5.2. Let $(X_t : t \in \mathbb{Z})$ be a causal and invertible ARMA(p, q) process defined with an iid sequence

$(Z_t : t \in \mathbb{Z})$ satisfying $E[Z_t] = 0$ and

$E[Z_t^2] = \sigma^2$. Consider the MLE $\hat{\beta}'$ of β' that is initialized with the moment estimators of

Method 1. Then,

$$\sqrt{n}(\hat{\beta}' - \beta') \xrightarrow{\mathcal{D}} N(0, \sigma^2 \Gamma_{p,q}^{-1}) \quad (n \rightarrow \infty).$$

The result is optimal. The covariance matrix $\Gamma_{p,q}$ is in block form and can be evaluated in terms of covariances of various autoregressive processes.

Example 3.5.4 (Recruitment Series). The MLE estimation procedure for the recruitment series can be applied in R as follows:

```
>rec.mle = ar.mle(rec, order=2)
```

The mean estimate can be obtained from `rec.mle$x.mean` as $\hat{\mu} = 62.26$, while the autoregressive parameter estimates and their standard errors are accessed with the commands `rec.mle$ar` and `sqrt(rec.mle$sasy.var.coef)` as

$\hat{\phi}_1 = 1.3513(.0410)$ and $\hat{\phi}_2 = -.4099(.0410)$. Finally, the variance estimate is obtained from `rec.yw$var.pred` as $\hat{\sigma}^2 = 89.3360$. All values are very close to their counterparts in Example 3.3.5.

Method 3 (Least Squares Estimation) An alternative to the method of moments and the MLE is provided by the least squares estimation (LSE). For causal and invertible ARMA(p, q) processes, it is based on minimizing the weighted sum of squares

$$S(\phi, \theta) = \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{R_j} \quad (3.5.5)$$

with respect to ϕ and θ , respectively. Assuming that $\tilde{\phi}$ and $\tilde{\theta}$ denote these LSEs, the LSE for σ^2 is computed as

$$\tilde{\sigma}^2 = \frac{S(\tilde{\phi}, \tilde{\theta})}{n - p - q}.$$

The least squares procedure has the same asymptotics as the MLE.

Theorem 3.5.3. The result of Theorem 3.5.2. holds also if $\hat{\beta}'$ is replaced with $\tilde{\beta}'$.

Example 3.5.5 (Recruitment Series). The least squares estimation has already been discussed in Example 3.3.5, including the R commands.

Contributors

- [Alexander Aue](#) (Department of Statistics, University of California, Davis)
- Integrated by [Brett Nakano](#) (statistics, UC Davis)

This page titled [3.5: Parameter Estimation](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Alexander Aue](#).