Chapter 5

Estimation of the Mean and the ACVF

A stationary process $\{X_t\}$ is characterized by its mean and its autocovariance function $\gamma(\cdot)$, and so by the autocorrelation function $\rho(\cdot)$. In this chapter we present the estimators of these statistics obtained from observations of $X_1, \ldots, X_n$ and discuss their properties.

5.1 Estimation of the Mean

Denote by

$$\boldsymbol{X} = (X_1, \ldots, X_n)^{\mathrm{T}},$$

an n-dimensional random vector each of whose components is a random variable with expectation μ_i, that is

$$\mathrm{E}\,\boldsymbol{X} = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^{\mathrm{T}},$$

and whose variance-covariance matrix has the form

$$\boldsymbol{V} = \begin{pmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_1, X_2) & \ldots & \mathrm{cov}(X_1, X_n) \\ \mathrm{cov}(X_2, X_1) & \mathrm{var}(X_2) & \ldots & \mathrm{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_n, X_1) & \mathrm{cov}(X_n, X_2) & \ldots & \mathrm{var}(X_n) \end{pmatrix}$$

If $\{X_t\}$ is a stationary process then

$$\boldsymbol{\mu} = (\mu, \ldots, \mu)^{\mathrm{T}} = \mu(1, \ldots, 1)^{\mathrm{T}}$$

and the variance covariance matrix simplifies to

$$V = \begin{pmatrix} \gamma(0) & \gamma(-1) & \dots & \gamma(-n+1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(-n+2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(0) \end{pmatrix}$$

The mean of a process is not always zero and its estimation is important for further inference. The moment estimator of the mean μ of a stationary process is the sample mean

$$\bar{X}_n = \boldsymbol{b}^{\mathrm{T}} \boldsymbol{X},$$

where

$$\begin{aligned} \boldsymbol{b} &= \left(\frac{1}{n}, \dots, \frac{1}{n}\right)^{\mathrm{T}}, \\ \boldsymbol{X} &= (X_1, \dots, X_n)^{\mathrm{T}}. \end{aligned}$$

That is

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

It is an unbiased estimator since

$$\begin{aligned} \mathrm{E}(\boldsymbol{b}^{\mathrm{T}} \boldsymbol{X}) &= \boldsymbol{b}^{\mathrm{T}} \, \mathrm{E}(\boldsymbol{X}) \\ &= \boldsymbol{b}^{\mathrm{T}} \boldsymbol{\mu} \\ &= \frac{1}{n}(1, \dots, 1) \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} = \mu. \end{aligned}$$

The mean square error of $\bar{X}_n$ is

$$\mathrm{E}(\bar{X}_n - \mu)^2 = \mathrm{var}(\bar{X}_n).$$

Using the matrix notation we can write (see Remark 3.2)

$$\begin{aligned}
\operatorname{var}(\bar{X}_n) &= \operatorname{var}(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}) \\
&= \boldsymbol{b}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{b} \\
&= \frac{1}{n^2}(1,\ldots,1)\begin{pmatrix} \gamma(0) & \gamma(-1) & \ldots & \gamma(-n+1) \\ \gamma(1) & \gamma(0) & \ldots & \gamma(-n+2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \ldots & \gamma(0) \end{pmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \\
&= \frac{1}{n^2}\sum_{\tau=-n+1}^{n-1}(n-|\tau|)\gamma(\tau) \\
&= \frac{1}{n}\sum_{\tau=-n+1}^{n-1}\left(1-\frac{|\tau|}{n}\right)\gamma(\tau) \\
&\le \frac{1}{n}\sum_{|\tau|<n}|\gamma(\tau)|.
\end{aligned}$$

Now, if $\gamma(n) \longrightarrow 0$ as $n \longrightarrow \infty$ then the right hand side converges to zero. It means that $\bar{X}_n$ converges to μ in the mean square sense.

If the series is Gaussian, then by the Remark 3.3, we have the normality of $\bar{X}_n$,

$$\bar{X}_n \sim \mathcal{N}(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mu}, \boldsymbol{b}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{b}).$$

That is we can write

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{1}{n}v\right),$$

where

$$v = \sum_{|\tau|<n}\left(1-\frac{|\tau|}{n}\right)\gamma(\tau).$$

Then a confidence interval for μ can be obtained so as

$$P(-u_\alpha < \frac{\bar{X}_n-\mu}{\sqrt{v/n}} < u_\alpha) = 1-\alpha,$$

what can be rearranged to

$$P\left(\bar{X}_n - u_\alpha\sqrt{\frac{v}{n}} < \mu < \bar{X}_n + u_\alpha\sqrt{\frac{v}{n}}\right) = 1-\alpha.$$

Here u_α is such that $P(|U| < u_\alpha) = 1 - \alpha$ and $U = \frac{\bar{X}_n - \mu}{\sqrt{v/n}} \sim \mathcal{N}(0,1)$. For $\alpha = 0.05$ we have $u_\alpha = 1.96$ and the confidence interval boundaries are

$$\left(\bar{X}_n - 1.96\sqrt{\frac{v}{n}},\ \bar{X}_n + 1.96\sqrt{\frac{v}{n}}\right),$$

This results are obtained assuming that v is known. In practice, usually it is not the case and we need to estimate v. To estimate v the covariance $\gamma(\tau)$ is replaced with $\widehat{\gamma}(\tau)$ and $\hat{v}$ is calculated as

$$\hat{v} = \sum_{|\tau|<n} \left(1 - \frac{|\tau|}{n}\right) \hat{\gamma}(\tau).$$

Example 5.1. Let $\{X_t\}$ be an AR(1) process with mean μ, defined by

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t,$$

where $|\phi| < 1$ and $Z_t \sim WN(0, \sigma^2)$. For this process we have

$$\gamma(\tau) = \frac{\phi^{|\tau|}\sigma^2}{1 - \phi^2}.$$

Hence, taking

$$v = \sum_{|\tau|<\infty} \gamma(\tau)$$

we obtain the result

$$v = \frac{\sigma^2}{1-\phi^2} \sum_{|\tau|<\infty} \phi^{|\tau|} = \frac{\sigma^2}{1-\phi^2}\left(-1 + 2\sum_{\tau=0}^{n} \phi^\tau\right) = \frac{\sigma^2}{(1-\phi)^2}.$$

In this case we need to know σ^2 and ϕ to obtain v or their estimates to obtain $\hat{v}$.

5.2 Estimation of ACVF and ACF

Definition 4.3 gives the following estimators for $\gamma(\tau)$ and $\rho(\tau)$, respectively

$$\widehat{\gamma}(\tau) = \frac{1}{k}\sum_{t=1}^{k-|\tau|} (X_t - \bar{X}_k)(X_{t+|\tau|} - \bar{X}_k), \quad -k < \tau < k \tag{5.1}$$

where

$$\bar{X}_k = \frac{1}{k}\sum_{t=1}^{k} X_t.$$

and

$$\widehat{\rho}(\tau) = \frac{\widehat{\gamma}(\tau)}{\widehat{\gamma}(0)}, \quad -k < \tau < k. \tag{5.2}$$

Both estimators are biased, however for for large k the bias is small. The ACVF has the property that the k-dimensional sample covariance matrix

$$\widehat{\boldsymbol{V}}_k = \begin{pmatrix} \widehat{\gamma}(0) & \widehat{\gamma}(-1) & \dots & \widehat{\gamma}(k-1) \\ \widehat{\gamma}(1) & \widehat{\gamma}(0) & \dots & \widehat{\gamma}(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\gamma}(k-1) & \widehat{\gamma}(k-2) & \dots & \widehat{\gamma}(0) \end{pmatrix}$$

is nonnegative definite. To show it means to show that

$$\boldsymbol{a}^{\mathrm{T}} \widehat{\boldsymbol{V}}_k \boldsymbol{a} \geq 0$$

for any k-dimensional real vector $\boldsymbol{a}$. This can be easily obtained if we can express the matrix $\widehat{\boldsymbol{V}}$ as the following product

$$\widehat{\boldsymbol{V}}_{\boldsymbol{k}} = \frac{1}{k} \boldsymbol{C} \boldsymbol{C}^{\mathrm{T}},$$

for some matrix $\boldsymbol{C}$. Take vector r.vs $\boldsymbol{X} = (X_1, \dots, X_k)^{\mathrm{T}}$ and $\boldsymbol{Y} = (X_1 - \bar{X}_1, \dots, X_k - \bar{X}_k)^{\mathrm{T}}$. Then

$$\boldsymbol{C} = \begin{pmatrix} 0 & \dots & 0 & 0 & Y_1 & Y_2 & \dots & Y_k \\ 0 & \dots & 0 & Y_1 & Y_2 & \dots & Y_k & 0 \\ \vdots & & & & & & & \vdots \\ 0 & Y_1 & Y_n & \dots & Y_k & 0 & \dots & 0 \end{pmatrix}$$

It is easy to see that multiplying $\boldsymbol{C}$ by $\boldsymbol{C}^{\mathrm{T}}$ we obtain a matrix of sums of squares and products of Y_i which when divided by k is the $\widehat{\boldsymbol{V}}_{\boldsymbol{k}}$ matrix. Hence,

$$\begin{aligned} \boldsymbol{a}^{\mathrm{T}} \widehat{\boldsymbol{V}}_{\boldsymbol{k}} \boldsymbol{a} &= \boldsymbol{a}^{\mathrm{T}} \frac{1}{k} \boldsymbol{C} \boldsymbol{C}^{\mathrm{T}} \boldsymbol{a} \\ &= \frac{1}{k} (\boldsymbol{a}^{\mathrm{T}} \boldsymbol{C})(\boldsymbol{C}^{\mathrm{T}} \boldsymbol{a}) \geq 0. \end{aligned}$$

Hence, due to the Theorem 4.1 $\widehat{\gamma}(\tau)$ is an autocovariance function of a stationary process as it is nonnegative definite and even.

We will be using the forms 5.1 and 5.2 as the estimators for the ACVF and ACF. The estimates of $\rho(\tau)$ are good if $\tau << n$, where n is the total number of observations. For τ close to n there are too few pairs $(X_t, X_{t+\tau})$ for the estimate to be reliable. Box and Jenkins (1976) suggest that n should be at least 50 and $\tau \leq n/4$.

For statistical inference based on the $\widehat{\rho}(\tau)$ we need to know its distribution. For large sample size it can be approximated by a normal distribution. For linear models the vector

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

is approximately distributed as

$$\widehat{\boldsymbol{\rho}} \underset{approx}{\sim} \mathcal{N}\left(\boldsymbol{\rho}, \frac{1}{n}\boldsymbol{W}\right), \tag{5.3}$$

where

$$\boldsymbol{\rho} = (\rho(1), \ldots, \rho(k))^{\mathrm{T}}$$

and $\boldsymbol{W}$ is the variance-covariance matrix

$$\boldsymbol{W} = \{w_{ij}\}, \tag{5.4}$$

where w_{ij} is given by Bartlett's formula

$$w_{ij} = \sum_{k=1}^{\infty}[\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)][\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)].$$

Example 5.2. Let $\{X_t\} \sim IID(0, \sigma^2)$. Then $\rho(\tau) = 0$ for all $\tau > 0$ and from 5.4 we obtain

$$w_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Then by (5.3) the estimators $\widehat{\rho}(\tau)$ are approximately independent and identically distributed as

$$\widehat{\rho}(\tau) \underset{approx}{\sim} N\left(0, \frac{1}{n}\right).$$

This gives us the confidence bounds for $\rho(\tau)$ of an IID process, which are

$$(-u_\alpha/\sqrt{n},\ u_\alpha/\sqrt{n}),$$

where u_α is such that $P(|U| < u_\alpha) = 1 - \alpha$, where $U \sim \mathcal{N}(0, 1)$. For $\alpha = 0.05$ we have $u_\alpha \approx 1.96$ and the 95% confidence interval boundaries are

$$(-1.96/\sqrt{n},\ 1.96/\sqrt{n}).$$

Example 5.3. Consider MA(1) process

$$X_t = Z_t + \theta Z_{t-1}, \quad t = 0, \pm 1, \pm 2, \dots,$$

where $\{Z_t\} \sim WN(0, \sigma^2)$. Then from equation (5.4) we obtain

$$w_{ii} = \begin{cases} 1 - 3\rho^2(1) + 4\rho^4(1), & \text{if } i = 1 \\ 1 + 2\rho^2(1), & \text{if } i > 1. \end{cases}$$

Then by (5.3) we get the confidence interval for $\rho(1)$, namely

$$\left(\hat{\rho}(1) - u_\alpha \sqrt{\frac{1}{n}(1 - 3\hat{\rho}^2(1) + 4\hat{\rho}^4(1))},\ \hat{\rho}(1) + u_\alpha \sqrt{\frac{1}{n}(1 - 3\hat{\rho}^2(1) + 4\hat{\rho}^4(1))}\right). \tag{5.5}$$

We know that for $\tau > 1$ the true value of the ACF is zero, hence as a kind of test we can calculate the interval in which the obtained sample values of the ACF are not significant. This is

$$\left(-u_\alpha \sqrt{\frac{1}{n}(1 + 2\hat{\rho}^2(1))},\ u_\alpha \sqrt{\frac{1}{n}(1 + 2\hat{\rho}^2(1))}\right), \tag{5.6}$$

Now, take $\theta = 0.5$ as in Example 4.3. Then the theoretical value of $\rho(1)$ is

$$\rho(1) = \frac{\theta}{1 + \theta^2} = \frac{0.5}{1.25} = 0.4.$$

From the simulation we obtained

$$\widehat{\rho}(1) = 0.4763$$

and so the 95% confidence interval is approximated be

$$\begin{aligned} &(0.4763 - 1.96 \cdot 0.0724762,\ 0.4763 + 1.96 \cdot 0.0724762) \\ &= (0.334247,\ 0.618353) \end{aligned}$$

which includes the theoretical value of $\rho(1)$. For lag $\tau > 1$ we have

$$\begin{aligned} &(-1.96 \cdot 0.12057,\ 1.96 \cdot 0.12057) \\ &= (-0.236318,\ 0.236318) \end{aligned}$$

In fact the bounds are often calculated according to the formula for IID noise, which depends only on n. Here $n = 100$ and we obtain

$$(-u_\alpha/\sqrt{n},\ u_\alpha/\sqrt{n}) = (-0.196,\ 0.196)$$

Figure 4.4 shows such boundaries and indeed all the sample autocorrelations for lag $\tau > 1$ are within these boundaries. For lag $\tau = 1$ we obtained the CI covering the true value of $\rho(1)$. These two facts support the compatibility of the simulated data with MA(1) model with $\theta = 0.5$.

Example 5.4. Consider AR(1) process

$$X_t = \phi X_{t-1} + Z_t,$$

where $\{Z_t\}$ is an i.i.d. noise and $|\phi| < 1$. Then the theoretical ACF is given by

$$\rho(\tau) = \phi^{|\tau|} \quad \text{for any } \tau = 0, \pm 1, \pm 2, \ldots.$$

From the Bartlett's formula (5.4) for the variances and covariances of $\boldsymbol{\rho}$ and the form of $\rho(\tau)$ for AR(1) we obtain

$$\begin{aligned} w_{\tau\tau} &= \sum_{k=1}^{\tau} \phi^{2\tau}(\phi^{-k} - \phi^{k})^2 + \sum_{k=\tau+1}^{\infty} \phi^{2k}(\phi^{-\tau} - \phi^{\tau})^2 \\ &= (1 - \phi^{2\tau})(1 + \phi^2)(1 - \phi^2)^{-1} - 2\tau\phi^{2\tau}, \end{aligned} \tag{5.7}$$

for $\tau = 1, 2, \ldots$.

Then, due to (5.3) the approximate confidence bounds can be computed as

$$\left(\hat{\rho}(\tau) - u_\alpha\sqrt{w_{\tau\tau}/n},\ \hat{\rho}(\tau) + u_\alpha\sqrt{w_{\tau\tau}/n}\right). \tag{5.8}$$

Take $\phi = 0.5$ as in the bottom plot of Figure 4.9. The sample ACF is given in the Figure 4.10. Is this sample ACF compatible with AR(1) for $\phi = 0.5$? What conclusions can you draw from the figure below?

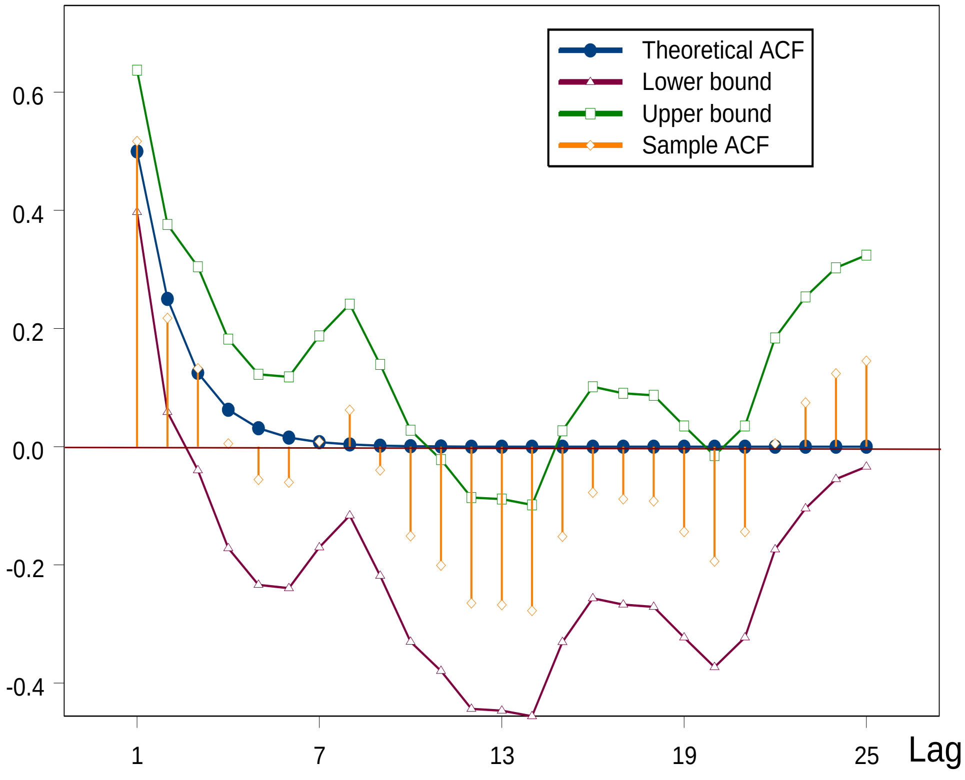

Figure 5.1: Theoretical ACF, sample ACF and the CI bounds for the simulated AR(1) with $\phi = 0.5$.