Chapter 2

The Estimation of Mean and Covariances

In this brief second chapter, some results concerning asymptotic properties of the sample mean and the sample ACVF are collected. Throughout, $(X_t \colon t \in \mathbb{Z})$ denotes a weakly stationary stochastic process with mean μ and ACVF γ. In Section 1.2 it was shown that such a process is completely characterized by these two quantities. The mean μ was estimated by the sample mean $\bar{x}$, and the ACVF γ by the sample ACVF $\hat{\gamma}$ defined in (1.2.1). In the following, some properties of these estimators are discussed in more detail.

2.1 Estimation of the Mean

Assume that an appropriate guess for the unknown mean μ of some weakly stationary stochastic process $(X_t \colon t \in \mathbb{Z})$ has to be found. The sample mean $\bar{x}$, easily computed as the average of n observations $x_1, \ldots, x_n$ of the process, has been identified as suitable in Section 1.2. To investigate its theoretical properties, one needs to analyze the random variable associated with it, that is,

$$\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n).$$

Two facts can be quickly established.

- $\bar{X}_n$ is an *unbiased* estimator for μ, since

  $$E[\bar{X}_n] = E\left[\frac{1}{n}\sum_{t=1}^{n} X_t\right] = \frac{1}{n}\sum_{t=1}^{n} E[X_t] = \frac{1}{n} n\mu = \mu.$$

  This means that "on average", the true but unknown μ is correctly estimated. Notice that there is no difference in the computations between the standard case of independent and identically distributed random variables and the more general weakly stationary process considered here.

- If $\gamma(n) \to 0$ as $n \to \infty$, then $\bar{X}_n$ is a *consistent* estimator for μ, since

  $$\mathrm{Var}(\bar{X}_n) = \mathrm{Cov}\left(\frac{1}{n}\sum_{s=1}^{n} X_s, \frac{1}{n}\sum_{t=1}^{n} X_t\right) = \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n} \mathrm{Cov}(X_s, X_t)$$

$$= \frac{1}{n^2} \sum_{s-t=-n}^{n} (n - |s-t|)\gamma(s-t) = \frac{1}{n} \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right) \gamma(h).$$

Now, the quantity on the right-hand side converges to zero as $n \to \infty$ because $\gamma(n) \to 0$ as $n \to \infty$ by assumption. The first equality sign in the latter equation array follows from the fact that $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$ for any random variable X, the second equality sign uses that the covariance function is linear in both arguments. For the third equality, one can use that $\mathrm{Cov}(X_s, X_t) = \gamma(s-t)$ and that each $\gamma(s-t)$ appears exactly $n - |s-t|$ times in the double summation. Finally, the right-hand side is obtained by replacing $s - t$ with h and pulling one n^{-1} inside the summation.

In the standard case of independent and identically distributed random variables $n\mathrm{Var}(\bar{X}) = \sigma^2$. The condition $\gamma(n) \to 0$ is automatically satisfied. However, in the general case of weakly stationary processes, it cannot be omitted.

More can be proved using an appropriate set of assumptions. The results are formulated as a theorem without giving the proofs.

Theorem 2.1.1. *Let $(X_t \colon t \in \mathbb{Z})$ be a weakly stationary stochastic process with mean μ and ACVF γ. Then, the following statements hold true as $n \to \infty$.*

(a) If $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, then

$$n\mathrm{Var}(\bar{X}_n) \to \sum_{h=-\infty}^{\infty} \gamma(h) = \tau^2;$$

(b) If the process is "close to Gaussianity", then

$$\sqrt{n}(\bar{X}_n - \mu) \sim AN(0, \tau_n^2), \qquad \tau_n^2 = \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right) \gamma(h).$$

Here, $\sim AN(0, \tau_n^2)$ stands for approximately normally distributed with mean zero and variance τ_n^2.

Theorem 2.1.1 can be utilized to construct confidence intervals for the unknown mean parameter μ. To do so, one must, however, estimate the unknown variance parameter τ_n. For a large class of stochastic processes, it holds that τ_n^2 converges to τ^2 as $n \to \infty$. Therefore, we can use τ^2 as an approximation for τ_n^2. Moreover, τ^2 can be estimated by

$$\hat{\tau}_n^2 = \sum_{h=-\sqrt{n}}^{\sqrt{n}} \left(1 - \frac{|h|}{n}\right) \hat{\gamma}(h),$$

where $\hat{\gamma}(h)$ denotes the ACVF estimator defined in (1.2.1). An approximate 95% confidence interval for μ can now be constructed as

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

Example 2.1.1 (Autoregressive Processes)**.** Let $(X_t\colon t\in\mathbb{Z})$ be given by the equations

$$X_t-\mu=\phi(X_{t-1}-\mu)+Z_t,\qquad t\in\mathbb{Z},\tag{2.1.1}$$

where $(Z_t\colon t\in\mathbb{Z})\sim\mathrm{WN}(0,\sigma^2)$ and $|\phi|<1$. It will be shown in Chapter 3 that $(X_t\colon t\in\mathbb{Z})$ defines a weakly stationary process. Utilizing the stochastic difference equations (2.1.1), both mean and autocovariances can be determined. It holds that $E[X_t]=\phi E[X_{t-1}]+\mu(1-\phi)$. Since, by stationarity, $E[X_{t-1}]$ can be substituted with $E[X_t]$, it follows that

$$E[X_t]=\mu,\qquad t\in\mathbb{Z}.$$

In the following we shall work with the process $(X_t^c\colon t\in\mathbb{Z})$ given by letting $X_t^c=X_t-\mu$. Clearly, $E[X_t^c]=0$. From the definition, it follows also that the covariances of $(X_t\colon t\in\mathbb{Z})$ and $(X_t^c\colon t\in\mathbb{Z})$ coincide. First computing the second moment of X_t^c, gives

$$E[\{X_t^c\}^2]=E\big[(\phi X_{t-1}^c+Z_t)^2\big]=\phi^2E[\{X_{t-1}^c\}^2]+\sigma^2$$

and consequently, since $E[\{X_{t-1}^c\}^2]=E[\{X_t^c\}^2]$ by weak stationarity of $(X_t^c\colon t\in\mathbb{Z})$,

$$E[\{X_t^c\}^2]=\frac{\sigma^2}{1-\phi^2},\qquad t\in\mathbb{Z}.$$

It becomes apparent from the latter equation, why the condition $|\phi|<1$ was needed in display (2.1.1). In the next step, the autocovariance function is computed. For $h>0$, it holds that

$$\gamma(h)=E[X_{t+h}^cX_t^c]=E\big[(\phi X_{t+h-1}^c+Z_{t+h})X_t^c\big]=\phi E[X_{t+h-1}^cX_t^c]=\phi\gamma(h-1)=\phi^h\gamma(0)$$

after h iterations. But since $\gamma(0)=E[\{X_t^c\}^2]$, by symmetry of the ACVF, it follows that

$$\gamma(h)=\frac{\sigma^2\phi^{|h|}}{1-\phi^2},\qquad h\in\mathbb{Z}.$$

After these theoretical considerations, a 95% (asymptotic) confidence interval for the mean parameter μ can be constructed. To check if Theorem 2.1.1 is applicable here, one needs to check if the autocovariances are absolutely summable:

$$\begin{aligned}\tau^2=\sum_{h=-\infty}^{\infty}\gamma(h)&=\frac{\sigma^2}{1-\phi^2}\left(1+2\sum_{h=1}^{\infty}\phi^h\right)=\frac{\sigma^2}{1-\phi^2}\left(1+\frac{2}{1-\phi}-2\right)\\&=\frac{\sigma^2}{1-\phi^2}\frac{1}{1-\phi}(1+\phi)=\frac{\sigma^2}{(1-\phi)^2}<\infty.\end{aligned}$$

Therefore, a 95% confidence interval for μ which is based on the observed values $x_1,\dots,x_n$ is given by

$$\left(\bar{x}-1.96\frac{\sigma}{\sqrt{n}(1-\phi)},\bar{x}+1.96\frac{\sigma}{\sqrt{n}(1-\phi)}\right).$$

Therein, the parameters σ and ϕ have to be replaced with appropriate estimators. These will be introduced in Chapter 3 below.

2.2 Estimation of the Autocovariance Function

This section deals with the estimation of the ACVF and ACF at lag h. Recall from equation (1.2.1) that the estimator

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X}_n)(X_t - \bar{X}_n), \qquad h = 0, \pm 1, \ldots, \pm(n-1),$$

may be utilized as a proxy for the unknown $\gamma(h)$. As estimator for the ACF $\rho(h)$,

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \qquad h = 0, \pm 1, \ldots, \pm(n-1),$$

was identified. Some of the theoretical properties of $\hat{\rho}(h)$ are briefly collected in the following. They are not as obvious to derive as in the case of the sample mean, and all proofs are omitted. Note also that similar statements hold for $\hat{\gamma}(h)$ as well.

- The estimator $\hat{\rho}(h)$ is generally biased, that is, $E[\hat{\rho}(h)] \neq \rho(h)$. It holds, however, under non-restrictive assumptions that
$$E[\hat{\rho}(h)] \to \rho(h) \qquad (n \to \infty).$$
This property is called *asymptotic unbiasedness*.

- The estimator $\hat{\rho}(h)$ is consistent for $\rho(h)$ under an appropriate set of assumptions, that is, $\mathrm{Var}(\hat{\rho}(h) - \rho(h)) \to 0$ as $n \to \infty$.

It was already established in Section 1.5 how the sample ACF $\hat{\rho}$ can be used to test if residuals consist of white noise variables. For more general statistical inference, one needs to know the sampling distribution of $\hat{\rho}$. Since the estimation of $\rho(h)$ is based on only a few observations for h close to the sample size n, estimates tend to be unreliable. As a rule of thumb, given by Box and Jenkins (1976), n should at least be 50 and h less than or equal to $n/4$.

Theorem 2.2.1. *For $m \geq 1$, let $\boldsymbol{\rho}_m = (\rho(1), \ldots, \rho(m))^T$ and $\hat{\boldsymbol{\rho}}_m = (\hat{\rho}(1), \ldots, \hat{\rho}(m))^T$, where ^{T} denotes the transpose of a vector. Under a set of suitable assumptions, it holds that*

$$\sqrt{n}(\hat{\boldsymbol{\rho}}_m - \boldsymbol{\rho}_m) \sim AN(\mathbf{0}, \Sigma) \qquad (n \to \infty),$$

where $\sim AN(0, \Sigma)$ stands for approximately normally distributed with mean vector $\mathbf{0}$ and covariance matrix $\Sigma = (\sigma_{ij})$ given by Bartlett's formula

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

The section is concluded with two examples. The first one recollects the results already known for independent, identically distributed random variables, the second deals with the autoregressive process of Example 2.1.1.

Example 2.2.1. Let $(X_t\colon t \in \mathbb{Z}) \sim \mathrm{IID}(0, \sigma^2)$. Then, $\rho(0) = 1$ and $\rho(h) = 0$ for all $h \neq 0$. The covariance matrix Σ is therefore given by

$$\sigma_{ij} = 1 \quad \text{if } i = j \qquad \textit{and} \qquad \sigma_{ij} = 0 \quad \text{if } i \neq j.$$

This means that Σ is a diagonal matrix. In view of Theorem 2.2.1 it holds thus that the estimators $\hat{\rho}(1), \ldots, \hat{\rho}(k)$ are approximately independent and identically distributed normal random variables with mean 0 and variance $1/n$. This was the basis for Methods 1 and 2 in Section 1.6 (see also Theorem 1.2.1).

Example 2.2.2. Reconsider the autoregressive process $(X_t\colon t \in \mathbb{Z})$ from Example 2.1.1 with $\mu = 0$. Dividing $\gamma(h)$ by $\gamma(0)$ yields that

$$\rho(h) = \phi^{|h|}, \qquad h \in \mathbb{Z}.$$

Now the diagonal entries of Σ are computed as

$$\begin{aligned}
\sigma_{ii} &= \sum_{k=1}^{\infty} \big[\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\big]^2 \\
&= \sum_{k=1}^{i} \phi^{2i}(\phi^{-k} - \phi^{k})^2 + \sum_{k=i+1}^{\infty} \phi^{2k}(\phi^{-i} - \phi^{i})^2 \\
&= (1 - \phi^{2i})(1 + \phi^2)(1 - \phi^2)^{-1} - 2i\phi^{2i}.
\end{aligned}$$