4

Stationary Signals

In this chapter we introduce basic concepts necessary to study the time-dependent dynamics of random phenomena. The latter will be modeled as a family of random quantities indexed by a parameter, interpreted in this book as time. The parameter may be either continuous or discrete. Depending on the context and the tradition followed by different authors, such families are called *random signals*, *stochastic processes*, or *random time series*. The emphasis here is on random dynamics which are *stationary*, that is governed by underlying statistical mechanisms that do not change in time, although, of course, particular realizations of such families will be functions that vary with time. Think here about the random signal produced by the proverbial repeated coin tossing; the outomes vary while the fundamental mechanics remains the same.

4.1 Stationarity, autocovariance, and autocorrelation

A *random* (or *stochastic*) *signal* is a time-dependent family of random quantities $X(t)$. Depending on the context, one can consider random signals on the positive time line $t \geq 0$, on the whole time line $-\infty < t < \infty$, or on a finite time interval $t_0 \leq t \leq t_1$. Also it is useful to be able to consider random vector signals and signals with discrete time $t = \dots, -2, -1, 0, 1, 2, \dots$.

In this book we will restrict our attention to signals that are statistically stationary, which means that at least some of their statistical characteristics do not change in time. Several choices are possible here:

First-order strictly stationary signals. In this case, the c.d.f. $F_{X(t)}(x)$ does not change in time (it is time-shift invariant), that is,

$$F_{X(t)}(x) = F_{X(t+\tau)}(x) \quad \text{for all } t, \tau, x. \tag{4.1.1}$$

Second-order strictly stationary signals. In this case, the joint 2D c.d.f. $F_{(X(t_1),X(t_2))}(x_1,x_2)$ does not change in time, that is,

$$F_{(X(t_1),X(t_2))}(x_1,x_2) = F_{(X(t_1+\tau),X(t_2+\tau))}(x_1,x_2) \quad \text{for all } t_1,t_2,\tau,x_1,x_2. \tag{4.1.2}$$

In a similar fashion one can define the *nth-order strict stationarity* of random signal $X(t)$ as the time-shift invariance of the nD joint c.d.f., that is, the requirement that

$$F_{(X(t_1),\ldots,X(t_n))}(x_1,\ldots,x_n) = F_{(X(t_1+\tau),\ldots,X(t_n+\tau))}(x_1,\ldots,x_n) \tag{4.1.3}$$

for all $t_1,\ldots,t_n,\tau,x_1,\ldots,x_n$.

Finally, a random signal $X(t)$ is said to be *strictly stationary* if, for each $n = 1,2,\ldots$, it is nth-order strictly stationary.

Obviously, as n increases, verifying the nth-order stationarity gets more and more difficult, not to mention practical difficulties that arise with checking the full strict stationarity. For this reason, a more modest concept of *second-order weakly stationary signals* is useful. In this case the invariance property is demanded only of the moments of the signal up to order 2. More precisely, a signal $X(t)$ is said to be second-order weakly stationary if its expectations and covariances are time-shift invariant, that is, if

$$\mu_X(t) \equiv \mathbf{E}[X(t)] = \mathbf{E}[X(t+\tau)] \equiv \mu_X(t+\tau) \tag{4.1.4}$$

for all t, τ, and the *autocovariance function* is

$$\begin{aligned}\gamma_X(t_1,t_2) &\equiv \mathrm{Cov}(X(t_1),X(t_2))\\ &= \mathrm{Cov}(X(t_1+\tau),X(t_2+\tau)) \equiv \gamma_X(t_1+\tau,t_2+\tau)\end{aligned} \tag{4.1.5}$$

for all t_1,t_2,τ.

It is a consequence of the above two conditions that, for any second-order weakly stationary signal,

$$\mu_X(t) = \mu_X = \text{constant}, \tag{4.1.6}$$

and the autocovariance function depends only on the time lag $\tau = t_2 - t_1$,

$$\gamma_X(t_1,t_2) = \gamma_X(t_1-t_1,t_2-t_1) = \gamma_X(0,t_2-t_1), \tag{4.1.7}$$

so that, in particular,

$$\mathrm{Var}(X(t)) \equiv \sigma^2_{X(t)} = \mathrm{Cov}(X(t), X(t)) = \gamma_X(0,0) = \sigma^2_X = \text{constant.} \tag{4.1.8}$$

Thus all the first and second moments of the signal can be expressed in terms of just two characteristics, the signal's *mean value* μ_X and signal's autocovariance function

$$\gamma_X(t) := \gamma_X(0,t) = \mathbf{E}[(X(0)-\mu_X)(X(t)-\mu_X)], \tag{4.1.9}$$

which is, as a result of the stationarity assumption, a function of just a single variable.

In the remainder of this discussion, we will restrict our attention to second-order weakly stationary signals $X(t)$, which we will simply call *stationary signals*. We will analyze them assuming only the knowledge of their mean value μ_X and their autocovariance function $\gamma_X(t)$.

The following properties of the autocovariance function follow directly from its definition and the Schwartz inequality (see Section 3.7):

$$\gamma_X(-t) = \gamma_X(t), \tag{4.1.10}$$

and

$$|\gamma_X(t)| \le \gamma_X(0) = \sigma^2_X. \tag{4.1.11}$$

In other words, the covariance function is even and its absolute value is bounded by its value at $t = 0$, where it is simply equal to the signal's variance.

In different situations it is often convenient to use close relatives of the autocovariance function, such as the *autocorrelation function*[26]

$$\begin{aligned}\phi_X(t) &= \mathbf{E}(X(t_1)X(t_1+t)) \\ &= \mathrm{Cov}(X(t_1), X(t_1+t)) + \mathbf{E}(X(t_1))\cdot\mathbf{E}(X(t_1+t)) = \gamma_X(t) + \mu^2_X,\end{aligned} \tag{4.1.12}$$

and the *normalized autocovariance function*

[26] You may have noticed that in signal processing the traditional term "autocorrelation function" is at odds with the previously introduced term "correlation coefficient," which really corresponds to the above-introduced "normalized autocovariance function." But the terminology is so well established that we will stick with it.

$$\xi_X(\tau) = \frac{\gamma_X(t)}{\sigma_X^2} = \frac{\phi_X(t) - \mu_X^2}{\sigma_X^2} \tag{4.1.13}$$

which has the advantage of having its values always contained in the interval $[-1, 1]$.

If the signals' mean value is zero, then, of course, the autocovariance and the autocorrelation functions are identical:

$$\gamma_X(\tau) = \phi_X(\tau).$$

In what follows, unless explicitly stated otherwise, we will always assume that the signals under consideration have zero means so that the autocorrelation and the autocovariance are the same functions.

The reminder of this section is devoted to a series of examples of stationary data. The first, real-life example (see Figure 4.1.1) shows a sample of a 21-channel recording of the sleep electroencephalogram (EEG) of a neonate. The duration of this multidimensional random signal is one minute and the sampling rate is 64 Hz. This particular EEG was taken during the so-called mixed frequency sleep stage and, in addition to the EEG, it also shows related signals such as electrocardiogram (EKG), breathing signal, eye muscle contraction signal, etc. The signal's components seem stationary for some channels while other channels seem to violate the stationarity property. This can be due to some artifacts in the recordings caused, for example, by the physical movements of the infant or by the onset of a different sleep stage (active, passive, rapid eye movement (REM), etc.). The study of EEG signals provides important information on the state of the brain's neural network and, in the case of infants, can be used to assess the maturity level of their brains. In Section 4.2, we will provide a method to estimate the autocorrelation function for such real-life data.

Examples 4.1.1–4.1.6 provide various mathematical models of stationary signals. In those cases, the autocorrelation functions can be explicitely calculated.

Example 4.1.1 (a random harmonic oscillation). Consider a signal which is a simple harmonic oscillation with nonrandom frequency $f_0 = \frac{1}{P}$ but random amplitude A such that the second moment $\mathbf{E}A^2 < \infty$, and random phase Θ uniformly distributed over the period and independent of A. In other words,

$$X(t) = A\cos(2\pi f_0(t + \Theta)).$$

The signal is stationary because its mean value is

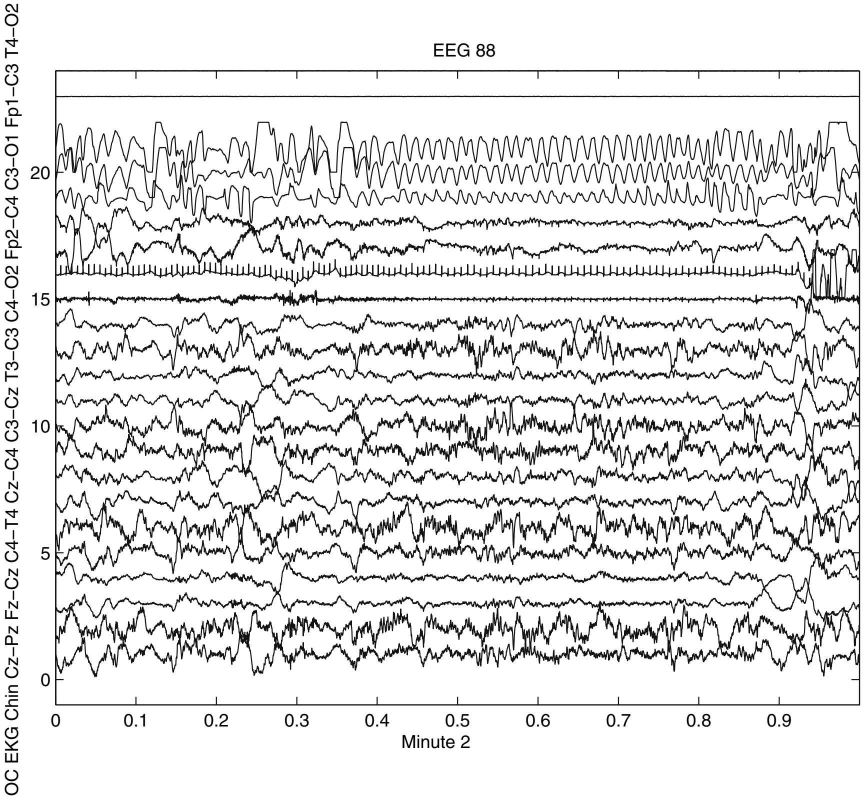

Fig. 4.1.1. A sample of a 21-channel recording of the sleep electroencephalogram (EEG) of a neonate. The duration of this multidimensional random signal is 60 seconds and the sampling rate is 64 Hz. (From A. Piryatinska's Ph.D. dissertation, Department of Statistics, Case Western Reserve University, Cleveland, 2004.)

$$\mathbf{E}X(t) = \mathbf{E}A\cos 2\pi f_0(t+\Theta) = \mathbf{E}A \cdot \int_0^P \cos 2\pi f_0(t+\theta)\frac{d\theta}{P} = \mathbf{E}A \cdot 0 = 0$$

and its autocovariance is

$$\begin{aligned}\gamma_X(t,s) &= \mathbf{E}X(t)X(s) = \mathbf{E}[A\cos 2\pi f_0(t+\Theta)\cdot A\cos 2\pi f_0(s+\Theta)]\\ &= \mathbf{E}A^2 \cdot \int_0^P \cos 2\pi f_0(t+\theta)\cdot\cos 2\pi f_0(s+\theta)\frac{d\theta}{P}\\ &= \mathbf{E}A^2\frac{1}{2}\left(\int_0^P \cos 2\pi f_0(t+s+2\theta)\frac{d\theta}{P} + \int_0^P \cos 2\pi f_0(s-t)\frac{d\theta}{P}\right)\\ &= \frac{\mathbf{E}A^2}{2}\cos 2\pi f_0(s-t),\end{aligned}$$

where we used the independence of the amplitude A and the phase Θ to split the expectations of the product into the product of the expectations. As a result we see that the autocorrelation $\gamma_X(t,s)$ is just a function of the difference $s-t$, which means that the signal is stationary. In particular,

$$\gamma_X(t) = \frac{\mathbf{E}A^2}{2}\cos 2\pi f_0 t.$$

Example 4.1.2 (superposition of random harmonic oscillations). In this example, we consider a signal which is a sum of simple harmonic oscillations with frequencies kf_0, $k = 1,2,\ldots,N$, random amplitudes A_k, $k = 1,2,\ldots,N$, such that $\mathbf{E}A_k^2 < \infty$, and random phases Θ_k, $k = 1,2,\ldots,N$, uniformly distributed over the corresponding periods. All of the above random quantities are assumed to be independent of each other. In other words,

$$X(t) = \sum_{k=1}^{N} A_k \cos(2\pi k f_0(t + \Theta_k)).$$

In this case one can verify (see Section 4.3, problems and exercises) that the signal is again stationary and the covariance function is of the form

$$\gamma_X(t) = \frac{1}{2}\sum_{k=1}^{N} \mathbf{E}A_k^2 \cos(2\pi k f_0 t).$$

Example 4.1.3 (discrete-time white noise). In this example, the time is discrete, that is, $t = n = \ldots,-2,-1,0,1,2,\ldots$ and the random signal $W(n)$ has mean zero and values at different times that are uncorrelated; its variance is σ_W^2. In other words,

$$\mu_W = 0,$$

and

$$\gamma_W(n,k) = \mathbf{E}(W(n)W(k)) = \begin{cases} \sigma_W^2 & \text{if } n-k=0, \\ 0 & \text{if } n-k \neq 0. \end{cases}$$

Note that the above-defined signal is stationary because its autocovariance (autocorrelation, since the mean is zero) is indeed a function of only the time lag and can be written in the form

$$\gamma_W(n,k) = \sigma_W^2 \delta(n-k),$$

where

$$\delta(n) = \begin{cases} 1 & \text{if } n = 0; \\ 0 & \text{if } n \neq 0, \end{cases}$$

is the discrete-time Dirac delta function. This kind of signal is called *discrete-time white noise* and it has mean zero and autocorrelation function

$$\gamma_W(n) = \sigma_W^2 \delta(n).$$

A sample path of a discrete-time white noise with $\sigma_W^2 = \frac{1}{12}$ is shown in Figure 4.1.2. It was produced using a random number generator in *Mathematica*, with the values of W_n uniformly distributed on the interval $[-\frac{1}{2}, \frac{1}{2}]$.

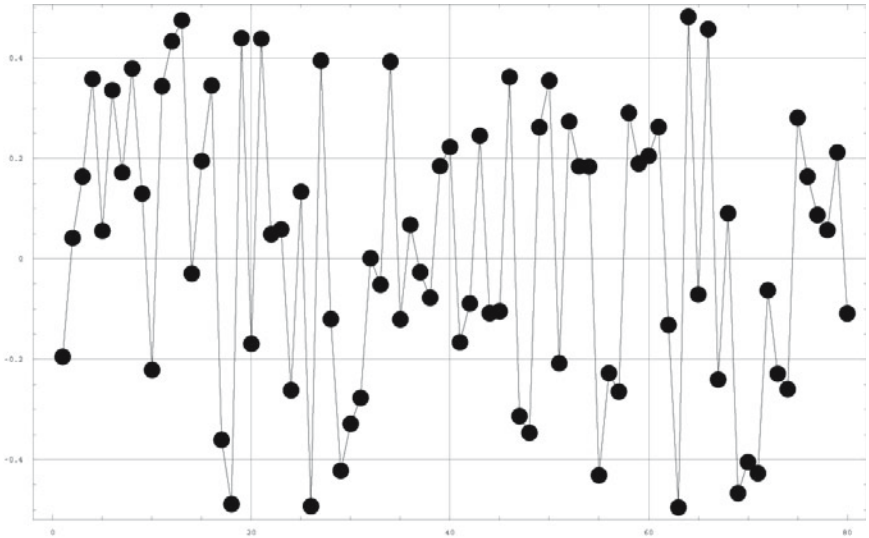

Fig. 4.1.2. A sample discrete-time white noise random signal $W(n)$, $n = 1, 2, \ldots, 50$, with $\sigma_W^2 = \frac{1}{12}$. For the sake of the clarity of the picture, values of $W(n)$ for consecutive integers n were joined by straight line segments.

Example 4.1.4 (moving average of the white noise). The moving average signal $X(n)$ is obtained from the white noise $W(n)$ with variance σ_W^2 by the "windowing" procedure. The windowing procedure mixes values of the white noise, $W(n), W(n-1), \ldots, W(n-q)$, in the time window of fixed width $q+1$, extending into the past, giving values with different time lags different weights, say, $b_0, b_1, \ldots, b_q$. More precisely,

$$X(n) = b_0 W(n) + b_1 W(n-1) + \cdots + b_q W(n-q).$$

You can interpret the moving average signal as a convolution of the white noise with the windowing weight sequence. One immediately obtains that $\mu_X = 0$. Since, for independent random quantities, the variance of the sum is equal to the sum of the variances, the variance is

$$\sigma_X^2 = \sigma_W^2 \sum_{i=0}^{q} b_i^2.$$

Calculation of the autocorrelation function is a little more complicated (see Section 4.3, problems and exercises) and here we will carry it out only in the case of the window of width 2, when

$$X(n) = b_0 W(n) + b_1 W(n-1).$$

Then

$$\begin{aligned}
\gamma_X(n,k) &= \mathbf{E}X(n)X(k)\\
&= \mathbf{E}((b_0W(n) + b_1W(n-1))(b_0W(k) + b_1W(k-1)))\\
&= b_0^2\mathbf{E}(W(n)W(k)) + b_0b_1\mathbf{E}(W(k)W(n-1))\\
&\quad + b_0b_1\mathbf{E}(W(k-1)W(n)) + b_1^2\mathbf{E}(W(n-1)W(k-1))\\
&= \begin{cases} (b_0^2 + b_1^2)\sigma_W^2 & \text{if } n = k \Leftrightarrow n-k = 0;\\ b_0b_1\sigma_W^2 & \text{if } n-1 = k \Leftrightarrow n-k = 1;\\ b_0b_1\sigma_W^2 & \text{if } n = k-1 \Leftrightarrow n-k = -1;\\ 0 & \text{if } |n-k| > 1. \end{cases}
\end{aligned}$$

Since $\gamma_X(n,k)$ depends only on the difference $n-k$, the moving average signal is stationary. For the sample white noise signal from Figure 4.1.2, the moving average signal $X(n) = 2W(n) + 5W(n-1)$ is shown in Figure 4.1.3, and its corresponding autocorrelation function

$$\gamma_X(n) = \begin{cases} \frac{29}{12} & \text{if } n = 0;\\ \frac{10}{12} & \text{if } n = \pm 1;\\ 0 & \text{if } n = \pm 2, \pm 3, \dots. \end{cases}$$

is shown in Figure 4.1.4.

Example 4.1.5 (random switching signal). Consider a signal $X(t)$ switching back and forth between values $+1$ and -1 at random times. More precisely, the intial value of the signal, $X(0)$, is a random quantity with the symmetric Bernoulli distribution, that is, $\mathbf{P}(X(0) = \pm 1) = \frac{1}{2}$, and the interswitching times form a sequence $T_1, T_2, \dots,$ of independent random quantities with the exponential distribution:

$$\mathbf{P}(T_i \le t) = 1 - e^{-t}, \quad t > 0,$$

of mean 1. The initial value $X(0)$ is assumed to be independent of interswitching times T_i. A typical sample of such a signal is shown in Figure 4.1.5.

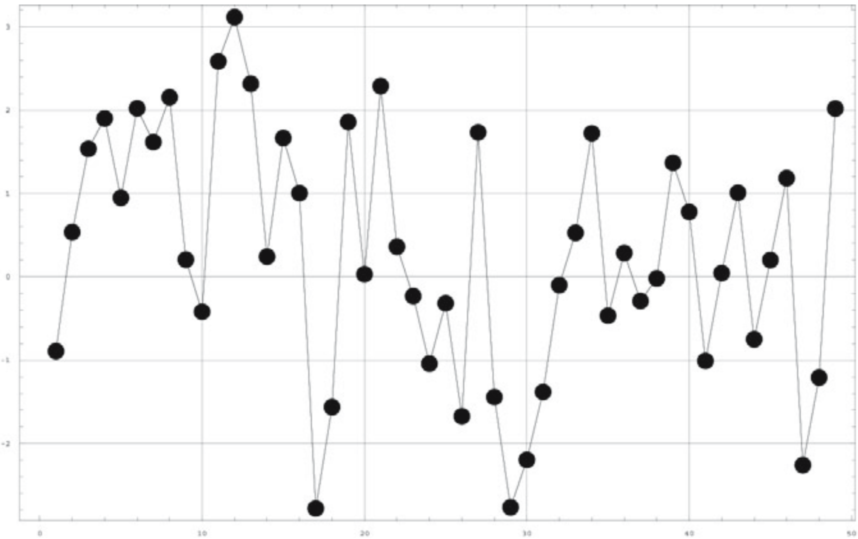

Fig. 4.1.3. Sample moving average signal $X(n) = 2W(n) + 5W(n-1)$ for the sample white noise shown in Figure 4.1.2. Note that the moving average signal appears smoother than the original white noise. The constrained oscillations are a result of nontrivial, although short-term in this example, correlations.

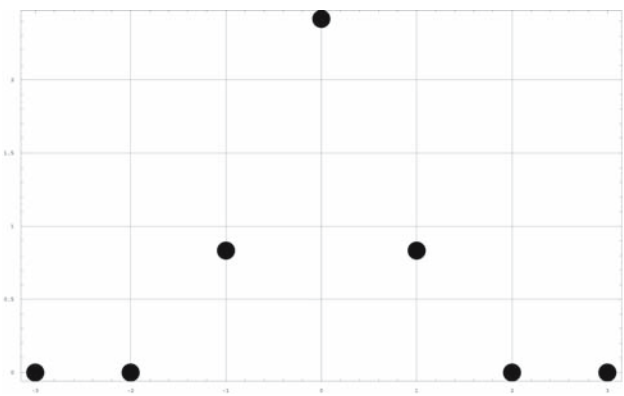

Fig. 4.1.4. Autocovariance function for the moving average signal $X(n) = 2W(n) + 5W(n-1)$. Note that the values of the signal separated by more that one time unit are uncorrelated.

Calculation of the mean and the autocorrelation function of the switching signal depends on the knowledge of the fact that such a random signal can be written in the form

$$X(0) \cdot (-1)^{N(t)},$$

where $N(t)$ is the (nonstationary) random signal counting the number

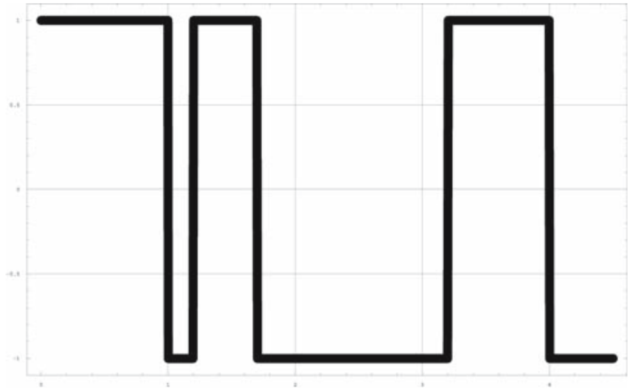

Fig. 4.1.5. A sample of the random switching signal from Example 4.1.5. The values are ± 1 and the initial value is $+1$. The interswitching times are independent and have an exponential c.d.f. of mean 1.

of switches up to time t. One can prove[27] that $N(t)$ has increments in disjoint time intervals that are statistically independent, with the distributions thereof depending only on the interval's length. More strikingly, these increments must have the Poisson probability distribution with mean equal to the interval's length, that is,

$$\mathbf{P}(N(t_0 + t) - N(t_0) = k) = e^{-t} \cdot \frac{t^k}{k!}$$

for any $t, t_0 \geq 0$ and $k = 0, 1, 2, \ldots$.

Armed with this information, we can now easily complete calculations of the mean and the autocorrelation function of the switching signal:

$$\mu_X(t) = \mathbf{E}X(t) = \mathbf{E}X(0) \cdot \mathbf{E}(-1)^{N(t)} = 0,$$

and, for $t < s$,

$$\begin{aligned} \gamma_X(t,s) = \mathbf{E}[X(t)X(s)] &= \mathbf{E}X^2(0) \cdot \mathbf{E}[(-1)^{N(t)}(-1)^{N(s)}] \\ &= 1 \cdot \mathbf{E}[(-1)^{2N(t)}(-1)^{N(s)-N(t)}] \\ &= \mathbf{E}(-1)^{N(s)-N(t)} \\ &= \sum_{k=0}^{\infty} (-1)^k \cdot \frac{e^{-(s-t)}(s-t)^k}{k!} = e^{-2(s-t)}. \end{aligned}$$

Therefore, the random switching signal $X(t)$ is stationary and, because of the symmetry property of all autocorrelation functions, its autocorrelation function

$$\gamma_X(t) = e^{-2|t|}.$$

[27] See, for example, O. Kallenberg, *Foundations of Modern Probability*, Springer-Verlag, New York, 1997.

Example 4.1.6 (solution of a stochastic difference equation). Consider a stochastic difference equation

$$X(n) = \alpha X(n-1) + \beta W(n), \quad n = -2, -1, 0, 1, 2, \ldots,$$

where $W(n)$ is a discrete-time white noise with $\sigma_W^2 = 1$. Observe that the above system, rewritten in the form

$$\frac{X(n) - X(n-1)}{\Delta n} = (\alpha - 1)X(n-1) + \beta W(n), \quad n = -2, -1, 0, 1, 2, \ldots,$$

can be viewed as a discrete-time version of the stochastic differential equation

$$dX(t) = (\alpha - 1)X(t)dt + \beta W(t)dt,$$

where $W(t)$ represents the continuous-time version of the white noise to be discussed in later chapters.

The solution of the above stochastic difference equation can be found by recursion. Therefore,

$$\begin{aligned} X(n) &= \alpha(\alpha X(n-2) + \beta W(n-1)) + \beta W(n) \\ &= \alpha^2 X(n-2) + \alpha\beta W(n-1) + \beta W(n) \\ &= \cdots = \alpha^l X(n-l) + \sum_{k=0}^{l-1} \alpha^k \beta W(n-k). \end{aligned}$$

for any $l = 1, 2, \ldots$. Assuming that $|\alpha| < 1$ and that $X(n-k)$ remain bounded, the first term $\alpha^k X(n-k) \to 0$ as $k \to \infty$. In that case, the second term converges to the infinite sum and the solution is of the form

$$X(n) = \beta \sum_{k=0}^{\infty} \alpha^k W(n-k).$$

This is the special form of the general moving average signal appearing in Problem 4.3.4, with the windowing sequence

$$c_k = \begin{cases} \beta\alpha^k & \text{for } k = 0, 1, 2, \ldots; \\ 0 & \text{for } k = -1, -2, \ldots. \end{cases}$$

Hence its autocorrelation function is

$$\gamma_X(n) = \sum_{k=-\infty}^{\infty} c_k c_{n+k} = \beta^2 \sum_{k=0}^{\infty} \alpha^k \alpha^{n+k} = \beta^2 \frac{\alpha^n}{1-\alpha^2}.$$

Example 4.1.7 (using moving averages to filter noise out of a signal). Consider a signal of the form

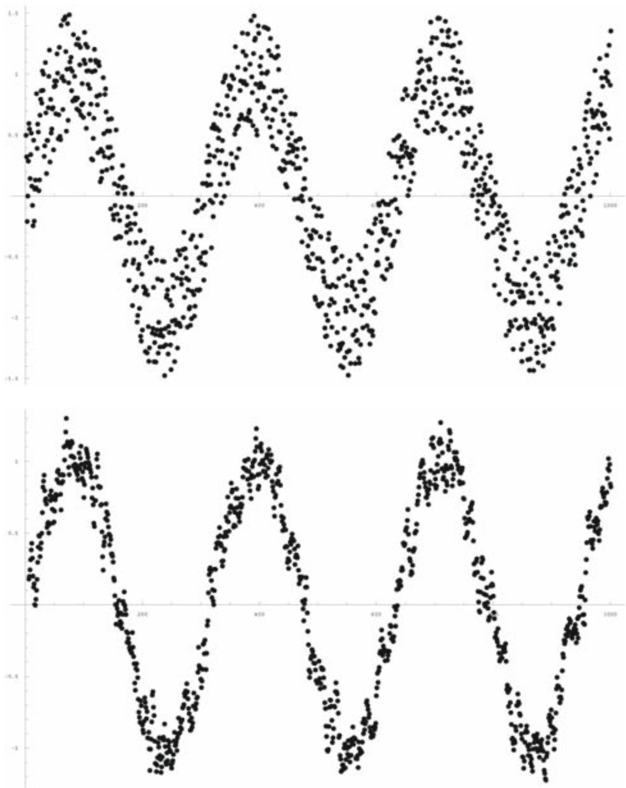

Fig. 4.1.6. *Top*: Signal $X(n)$ from Example 4.1.7 containing a nonrandom harmonic component plus a random white noise. *Bottom*: The same signal after a smoothing, moving average operation filtered out some of the white noise. The figure shows values of both signals for times $n = 1, 2, \ldots, 1000$.

$$X(n) = \sin(0.02n) + W(n),$$

where $W(n)$ is the white noise considered in Example 4.1.3 (shown in Figure 4.1.2), and let $Y(n)$ be a moving average (discrete-time convolution) of signal $X(n)$ with the windowing sequence $b_0 = b_1 = b_2 = b_3 = b_4 = \frac{1}{5}$, that is,

$$Y(n) = \frac{1}{5}X(n) + \frac{1}{5}X(n-1) + \frac{1}{5}X(n-2) + \frac{1}{5}X(n-3) + \frac{1}{5}X(n-4).$$

The values of both signals, $X(n)$ and $Y(n)$, for time instants $n = 1, 2, \ldots, 1000$, are shown in Figure 4.1.6. Clearly, the moving average operation filtered some of the white noise out of the original signal and the transformed signal appears smoother.

4.2 Estimating the mean and the autocorrelation function, ergodic signals

If one can obtain multiple independent samples of the same random stationary signal, then the estimation of its parameters, the mean value and the autocorrelation function, can be based on procedures described in Section 3.6. However, very often the only available information is a single but perhaps long (timewise) sample of the signal; think here about the historical temperature records at a given location, Dow Jones stock market index daily quotations over the past 10 years, or measurements of the sunspot activity over a period of time; these measurements cannot be independently repeated. Estimation of the mean and the autocorrelation function of a stationary signal $X(t)$ based on its single sample is a delicate matter because the standard law of large numbers and the central limit theorem cannot be applied. So one has to proceed with caution, as we now illustrate.

Estimation of the mean μ_X. If a stationary signal $X(t)$ is sampled with the sampling interval T, that is, the known values are

$$X(0), X(T), X(2T), \ldots, X(NT), \ldots,$$

then the obvious candidate for an estimator $\hat{\mu}_X$ of the signal's mean μ_X is

$$\hat{\mu}_X(N) = \frac{1}{N}\sum_{i=0}^{N-1} X(iT).$$

This estimator is easily seen to be unbiased as

$$\mathbf{E}[\hat{\mu}_X(N)] = \frac{1}{N}\sum_{i=0}^{N-1} \mathbf{E}[X(iT)] = \mu_X. \tag{4.2.1}$$

To check whether the estimator $\hat{\mu}_X(N)$ converges to μ_X as the observation interval $NT \to \infty$, that is, to check the estimator's consistency, we will take a look at the mean-square distance (estimation error) between $\hat{\mu}_X(N)$ and μ_X or, equivalently, the variance of their difference:

$$\begin{aligned}
\sigma^2(\hat{\mu}_X(N)) &= \mathbf{E}[(\hat{\mu}_X - \mu_X)^2] \\
&= \frac{1}{N^2}\mathbf{E}\left[\sum_{i=0}^{N-1}(X(iT)-\mu_X)\sum_{k=0}^{N-1}(X(kT)-\mu_X)\right] \\
&= \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{k=0}^{N-1}\gamma_X(iT,kT) = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{k=0}^{N-1}\gamma_C((i-k)T) \\
&= \frac{\sigma_X^2}{N} + \frac{2}{N}\sum_{k=0}^{N-1}\left(1-\frac{k}{N}\right)\gamma_X(kT).
\end{aligned} \tag{4.2.2}$$

So the error of replacing the true value μ_X by the estimator $\hat{\mu}_X$ will converge to zero, as $N \to \infty$, only if the sum in (4.2.2) increases more slowly[28] than N, i.e.,

$$\sum_{k=0}^{N-1} \left(1 - \frac{k}{N}\right) \gamma_X(kT) = o(N) \quad \text{as } N \to \infty. \tag{4.2.3}$$

Thus, for example, if the covariance function $\gamma_X(n)$ vanishes outside a finite interval, as was the case for finite moving averages in Example 4.1.2, then $\hat{\mu}_X$ is a consistent estimator for μ_X.

Example 4.2.1 (consistency of $\hat{\mu}_X$ for solutions of discrete-time stochastic difference equations). Consider the solution $X(n)$ of the stochastic difference equation from Example 4.1.6. Its autocorrelation function was found to be of the form

$$\gamma_X(n) = \beta^2 \frac{\alpha^n}{1-\alpha^2}, \quad |\alpha| < 1.$$

Since it decays exponentially as $n \to \infty$, the sum in (4.2.2) converges and condition (4.2.3) is satisfied. The mean-square error of replacing μ_X by the estimator $\hat{\mu}_X$ can now be controlled:

$$\begin{aligned}
\sigma^2(\hat{\mu}_X(N)) &= \mathbf{E}[(\hat{\mu}_X - \mu_X)^2] \\
&= \frac{\gamma_X(0)}{N} + \frac{2}{N} \sum_{k=0}^{N-1} \left(1 - \frac{k}{N}\right) \beta^2 \frac{\alpha^k}{1-\alpha^2} \\
&\le \frac{\beta^2}{N(1-\alpha^2)} \left(1 + 2\sum_{k=0}^{N-1} \alpha^k\right) \le \frac{\beta^2(3-\alpha)}{N(1-\alpha^2)(1-\alpha)}.
\end{aligned}$$

Estimation of the covariance function $\gamma_X(n)$. For simplicity's sake assume that $\mu_X = 0$, the sampling interval $T = 1$, the signal is real-valued, and that observations $X(0), \dots, X(N)$ are given. The natural candidate for an estimator of the autocorrelation function $\gamma_X(n) = \mathbf{E}X(0)X(n)$ is the time average:

$$\hat{\gamma}_X(n; N) = \frac{1}{N-n} \sum_{k=0}^{N-n-1} X(k)X(k+n). \tag{4.2.4}$$

It is an unbiased estimator since

$$\mathbf{E}[\hat{\gamma}_X(n, N)] = \frac{1}{N-n} \mathbf{E}\left[\sum_{k=0}^{N-n-1} X(k)X(k+n)\right]$$

[28] Here we use Landau's asymptotic notation: we write that $f(x) = o(g(x))$, as $x \to x_0$, and say that $f(x)$ is little "oh" of $g(x)$ at x_0, if $\lim_{x \to x_0} \frac{f(x)}{g(x)} = 0$.

$$= \frac{1}{N-n} \sum_{k=0}^{N-n-1} \gamma_X(n) = \gamma_X(n).$$

One can also prove that if $\gamma_X(n) \to 0$ sufficiently fast,[29] as $n \to \infty$, and if $\gamma_X(0) = \sigma_X^2 < \infty$, then the mean-square distance from $\hat{\gamma}_X(n;N)$ to $\gamma_X(n)$ decreases to 0 as $N \to \infty$. In other words, the estimator (4.2.4) is consistent.

Example 4.2.2. Figure 4.2.1 shows two samples of the central channel recording for a full-term neonate EEG (see Figure 4.1.1 for a sample of the full 21-channel EEG). The duration of each of the samples is three minutes. The data in the top picture were recorded during the active sleep stage, and in the bottom picture during the quiet sleep stage. The estimated autocorrelation functions (ACFs) for both signals were then calculated using formula (4.2.4), and are shown in Figure 4.2.2. The example is taken from A. Piryatinska's Ph.D. dissertation (Department of Statistics, Case Western Reserve University, Cleveland, 2004), mentioned already in Section 4.1. Note that the ACF of the active sleep signal decays much more slowly than the ACF of the quiet sleep, indicating the longer-range dependence structure of the former. Information on the rate of decay in EEG ACFs can then be used to automatically classify stationary segments of the EEG signals as those corresponding to different sleep stages recognized by pediatric neurologists.

Remark 4.2.1 (*ergodicity*). If the estimator $\hat{\mu}_X$ is unbiased and consistent, that is,

$$\mathbf{E}\hat{\mu}_X(N) = \mu_X \quad \text{and} \quad \sigma^2(\hat{\mu}_X(N)) \to 0,$$

as $N \to \infty$, then one often says that the signal is *ergodic in the mean*. Note that, in general, this does not imply that for every sample path of the random signal the estimator converges to the estimated parameter. To guarantee that, for a general test function g, the time averages

$$\frac{g(X(1)) + g(X(2)) + \cdots + g(X(N))}{N}$$

converge to $\mathbf{E}g(X(1))$ as $N \to \infty$, for (almost) every sample path of the random signal, stronger ergodicity and stricter stationarity assumptions are needed. A more detailed analysis of the ergodic behavior of stationary time series can be found in the above-quoted books by M. Denker and W. A. Woyczyński and by P. J. Brockwell and R. A. Davis.

[29] For a thorough exposition of these issues, see, for example, P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1991.

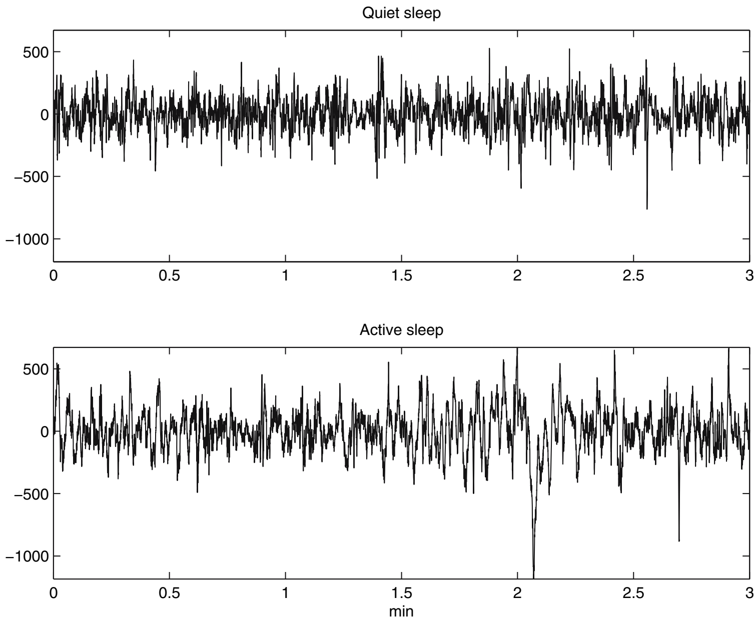

Fig. 4.2.1. *Top*: Three-minute recording of the central channel EEG for an infant in a quiet sleep stage. *Bottom*: Analogous recording for an active sleep stage.

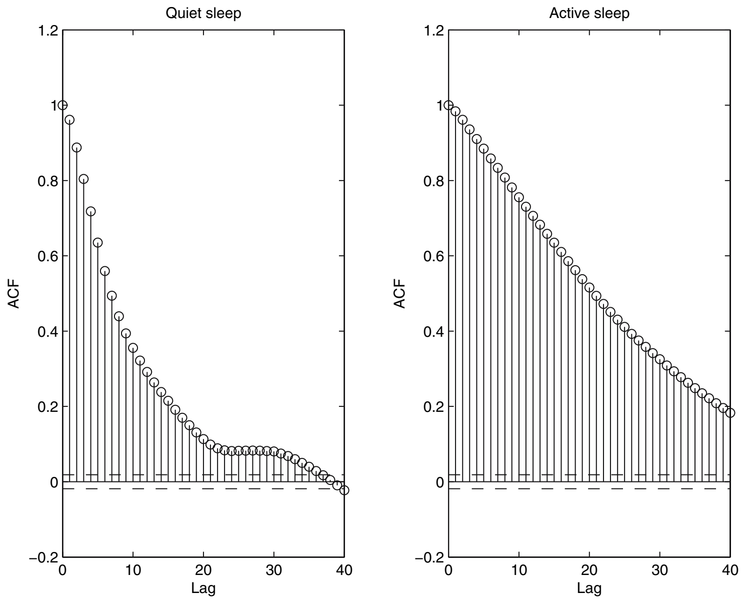

Fig. 4.2.2. *Left*: Estimated autocorrelation function (ACF) for the quiet sleep EEG signal from Figure 4.2.1. *Right*: Analogous estimated ACF for the active sleep stage.

Remark 4.2.2 (*confidence intervals*). Under fairly weak assumptions one can show that the asymptotic distributions ($N \to \infty$) of the suitably rescaled estimators $\hat{\mu}_X(N)$, $\hat{\gamma}_X(n;N)$ are asymptotically normal. Thus the confidence intervals for them can be constructed following the ideas discussed in Section 3.6.

4.3 Problems and exercises

4.3.1. Consider a random signal

$$X(t) = A_1 \cos 2\pi f_0(t + \Theta_1) + \cdots + A_n \cos 2\pi(nf_0)(t + \Theta_n),$$

where $A_1, \Theta_1, \ldots, A_n, \Theta_n$ are independent random variables and $\Theta_1, \ldots, \Theta_n$ are uniformly distributed on the time interval $[0, P = \frac{1}{f_0}]$. Is this signal stationary? Find its mean, autocovariance, and autocorrelation functions.

4.3.2. Consider a random signal

$$X(t) = A_1 \cos 2\pi f_0(t + \Theta_0),$$

where A_1, Θ_0, are independent random variables, and Θ_0 is uniformly distributed on the time interval $[0, \frac{P}{3} = \frac{1}{3f_0}]$. Is this signal stationary? Find its mean, autocovariance, and autocorrelation functions.

4.3.3. Find the mean and autocorrelation functions of the discrete-time signal

$$Y(n) = 3W(n) + 2W(n-1) - W(n-2),$$

where $W(n), n = \ldots, -2, -1, 0, 1, 2, \ldots$, is the discrete-time white noise with $\sigma_W^2 = 4$, that is,

$$EW(n) = 0$$

and

$$\mathbf{E}(W(k)W(n)) = 4\delta(n-k) = \begin{cases} 4 & \text{if } n-k = 0; \\ 0 & \text{if } n-k \neq 0. \end{cases}$$

Use the calculations with the Kronecker δ explicitly.

4.3.4. Consider a general moving average signal

$$X(n) = \sum_{k=-\infty}^{\infty} c_k W_{n-k},$$

where c_k is a "windowing" sequence such that $\sum_k |c_k|^2 < \infty$, and $W(n)$ is the standard white noise signal with $\gamma_W(n) = \delta(n)$. Show that the covariance function is

$$\gamma_X(n) = \sum_{k=-\infty}^{\infty} c_k c_{n+k}.$$

Use the calculations with the Kronecker δ explicitly. Apply this formula to verify the solution to Problem 4.3.3.

4.3.5. *Simulation of white noise with an arbitrary probability distribution.* Formula (3.1.11), $F_Y(y) = F_X(g^{-1}(y))$, describes the c.d.f. $F_Y(y)$ of the random quantity $Y = g(X)$ in terms of the c.d.f. $F_X(x)$ of the random quantity X and the function $g(x)$. It also permits construction of an algorithm to produce random samples from any given probability distribution provided a random sample uniformly distributed on the interval $[0, 1]$ is given. The latter can be obtained by using the random number generator in any computing platform; see Problem 1.4.15.

Let U be a uniformly distributed on $[0, 1]$ random quantity U with the c.d.f.

$$F_U(u) = u, \quad 0 \leq u \leq 1, \tag{4.3.1}$$

Then for a given c.d.f. $F_Z(z)$, the random quantity $Z = F_Z^{-1}(U)$, where $F_Z^{-1}(u)$ is the function inverse to $F_Z(z)$ (that is, a solution of the equation $u = F_Z(F_Z^{-1}(u))$), has the c.d.f. $F_Z(z)$. Indeed, a simple calculation using (4.3.1) shows that

$$\mathbf{P}(F_Z^{-1}(U) \leq z) = \mathbf{P}(U \leq F_Z(z)) = F_Z(z)$$

because $0 \leq F_Z(z) \leq 1$. So, for example, if the desired c.d.f. is exponential, with $F_Z(z) = 1 - e^{-z}$, $z \geq 0$, then $F_Z^{-1}(u) = -\ln(1 - u)$, $0 \leq u \leq 1$, and the random quantity $Z = -\ln(1 - U)$ has the above exponential c.d.f.

The general simulation algorithm is thus as follows:

(i) Choose the sample size N, and produce a random sample, $u_1, u_2, \ldots, u_N$, uniformly distributed on $[0, 1]$.

(ii) Calculate the inverse function $F_Z^{-1}(u)$.

(iii) Substitute the random sample, $u_1, u_2, \ldots, u_N$, into $F_Z^{-1}(u)$ to obtain the random sample

$$z_1 = F_Z^{-1}(u_1), z_2 = F_Z^{-1}(u_2), \ldots, z_N = F_Z^{-1}(u_N),$$

which has the desired c.d.f. $F_Z(z)$.

Use the above algorithm and Problem 1.4.15 to produce and plot examples of the white noise $W(n)$ with (a) the double exponential p.d.f. $f_W(w) = \frac{e^{-|w|}}{2}$ and (b) the Cauchy p.d.f. $f_W(w) = (\pi(1 + w^2))^{-1}$. Start with a calculation of the corresponding c.d.f.s. Check the result graphically by plotting the histograms of the random samples against the theoretical p.d.f.s.

4.3.6. Simulations of stationary random signals. Using the algorithm from Problem 4.3.5, repeat simulations shown in Figures 4.1.2, 4.1.3, and 4.1.6, but replacing the uniformly distributed white noise by (a) a double exponentially distributed white noise and (b) a "white noise" with the Cauchy distribution. Experiment with these simulations by including parameters in the above p.d.f.s, and changing the length of the produced discrete-time random signals.

4.3.7. Using the procedures described in Section 4.2, estimate the means and the autocorrelation functions (ACF) for sample signals obtained in simulation in Problem 4.3.6(a). Then compare graphically the estimated and the theoretical ACFs.

Note. Cauchy random quantities have an infinite variance (check!—cf. Problem 3.7.20), so the correlational definition of the discrete-time white noise is not applicable for them. In such cases, by a discrete-time white noise $W(n), \dots, -2, -1, 0, 1, 2, \dots$, we simply mean a sequence of independent, identically and symmetrically distributed (i.e., $W(n) \sim -W(n)$) random quantities. No moment requirements are made. On the other hand, such a sequence always forms a strictly stationary random signal; cf. (3.3.24) and Problem 3.7.28.