2

Stationary Stochastic Processes

2.0 Introduction

Spectral analysis almost invariably deals with a class of models called stationary stochastic processes. The material in this chapter is a brief review of the theory behind such processes. The reader is referred to Brockwell and Davis (2016), Papoulis and Pillai (2002), Priestley (1981), Shumway and Stoffer (2017) or Yaglom (1987a) for complementary discussions.

2.1 Stochastic Processes

Consider the following experiment (see Figure 22): we hook up a resistor to an oscilloscope in such a way that we can examine the voltage variations across the resistor as a function of time. Every time we press a "reset" button on the oscilloscope, it displays the voltage variations for the one-second interval following the reset. Since the voltage variations are presumably caused by such factors as small temperature variations in the resistor, each time we press the reset button, we will observe a different display on the oscilloscope. Owing to the complexity of the factors that influence the display, there is no way that we can use the laws of physics to predict what will appear on the oscilloscope. However, if we repeat this experiment over and over, we soon see that, although we view a different display each time we press the reset button, the displays resemble each other: there is a characteristic "bumpiness" shared by all the displays.

We can model this experiment by considering a large box in which we have placed pictures of all the oscilloscope displays that we could possibly observe. Pushing the reset button corresponds to reaching into the box and choosing "at random" one of the pictures. Loosely speaking, we call the box of all possible pictures together with the mechanism by which we select the pictures a *stochastic process*. The one particular picture we actually draw out at a given time is called a *realization* of the stochastic process. The collection of all possible realizations is called the *ensemble*.

A more precise definition utilizes the concept of a random variable (RV), defined as a function, or mapping, from the sample space of possible outcomes of a random experiment to the real line (for a real-valued RV), the complex plane (for a complex-valued RV) or m-dimensional Euclidean space (for a vector-valued RV of dimension m). If we let the sample space be the ensemble of all possible realizations for an experiment such as the one we've just described, then for any fixed time t we can define an RV $X(t)$ that describes the outcome of the experiment at time t (this would be the vertical displacement of the picture on the

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

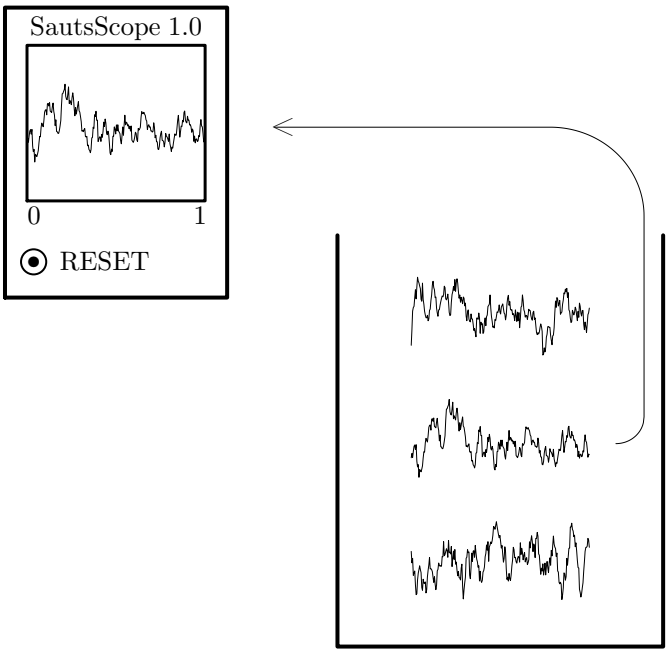

Figure 22 The oscilloscope experiment and a simplified box model. Here the ensemble contains only three possible realizations, the second of which was chosen.

oscilloscope at time t in our example). We can now give the following formal definition of a stochastic process: denoted by $\{X(t) : t \in T\}$, it is a family of RVs indexed by t, where t belongs to some given index set T. For convenience, we refer to t as representing time, although in general it need not.

In the experiment described in this section, the index set is all real numbers between zero and one. Other common index sets are the entire real axis and the set of all integers. If t takes on a continuous range of real values, the stochastic process is called a *continuous parameter* (or *continuous time*) stochastic process; if t takes on a discrete set of real values, it is called a *discrete parameter* (or *discrete time*) stochastic process.

2.2 Notation

Since it is important to be clear about the type of stochastic process we are discussing, we shall use the following notational conventions throughout this book:

[1] $\{X_t\}$ refers to a real-valued *discrete parameter* stochastic process whose tth component is X_t, while

[2] $\{X(t)\}$ refers to a real-valued *continuous parameter* stochastic process whose component at time t is $X(t)$.

[3] When the index set for a stochastic process is not explicitly stated (as is the case for both $\{X_t\}$ and $\{X(t)\}$), we shall assume that it is the set of all integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ for a discrete parameter process and the entire real axis $\mathbb{R} = \{t : -\infty < t < \infty\}$ for a continuous parameter process. Note that "t" is being used in two different ways here: the t in X_t is a unitless index (referring to the tth element of the process $\{X_t\}$), whereas the t in $X(t)$ has physically meaningful units such as seconds or days (hence $X(t)$ is the element occurring at time t of the process $\{X(t)\}$).

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

[4] On occasion we will need to discuss more than one stochastic process at a time. To distinguish among them, we will either introduce another symbol besides X (such as in $\{Y_t\}$) or add another index before the time index. For example, $\{X_{j,t}\}$ and $\{X_{k,t}\}$ refer to the jth and kth discrete parameter processes, while $\{X_j(t)\}$ and $\{X_k(t)\}$ refer to two continuous parameter processes. (Another way to handle multiple processes is to define a vector whose elements are stochastic processes. This approach leads to what are known as vector-valued stochastic processes or multivariate stochastic processes, which we do not deal with in this book.)

[5] We reserve the symbol Z for a complex-valued RV whose real and imaginary components are real-valued RVs. With an index added, $\{Z_t\}$ is a complex-valued discrete parameter stochastic process with a tth component formed from, say, the real-valued RVs $X_{0,t}$ and $X_{1,t}$; i.e., $Z_t = X_{0,t} + \mathrm{i}X_{1,t}$, where $\mathrm{i} \stackrel{\text{def}}{=} \sqrt{-1}$ (i.e., i is "equal by definition" to $\sqrt{-1}$). Likewise, $\{Z(t)\}$ is a complex-valued continuous parameter stochastic process with a tth component formed from two real-valued RVs, say, $X_0(t)$ and $X_1(t)$; i.e., $Z(t) = X_0(t) + \mathrm{i}X_1(t)$.

2.3 Basic Theory for Stochastic Processes

Let us first consider the real-valued discrete parameter stochastic process $\{X_t\}$. Since, for t fixed, X_t is an RV, it has an associated cumulative probability distribution function (CPDF) given by

$$F_t(a) = \mathbf{P}[X_t \leq a],$$

where the notation $\mathbf{P}[A]$ indicates the probability that the event A will occur. Because $\{X_t\}$ is a stochastic process, our primary interest is in the relationships amongst the various RVs that are part of it. These are expressed by various higher-order CPDFs. For example, for any t_0 and t_1 in the index set T,

$$F_{t_0,t_1}(a_0, a_1) = \mathbf{P}[X_{t_0} \leq a_0, X_{t_1} \leq a_1]$$

gives the bivariate CPDF for X_{t_0} and X_{t_1}, where the notation $\mathbf{P}[A_0, A_1]$ refers to the probability of the intersection of the events A_0 and A_1. More generally, for any integer $N \geq 1$ and any $t_0, t_1, \ldots, t_{N-1}$ in the index set, we can define the N-dimensional CPDF by

$$F_{t_0,t_1,\ldots,t_{N-1}}(a_0, a_1, \ldots, a_{N-1}) = \mathbf{P}[X_{t_0} \leq a_0, X_{t_1} \leq a_1, \ldots, X_{t_{N-1}} \leq a_{N-1}].$$

These higher-order CPDFs completely specify the joint statistical properties of the RVs in the stochastic process.

The notions of univariate and higher dimensional CPDFs extend readily to a real-valued continuous parameter stochastic process $\{X(t)\}$ and to complex-valued discrete and continuous parameter stochastic processes. For example, for a complex-valued continuous parameter stochastic process $\{Z(t)\}$ with

$$Z(t) = X_0(t) + \mathrm{i}X_1(t),$$

we can define a univariate CPDF for $Z(t)$ in terms of

$$F_t(a, b) = \mathbf{P}[X_0(t) \leq a, X_1(t) \leq b]$$

and a bivariate CPDF for $Z(t_0)$ and $Z(t_1)$ using

$$F_{t_0,t_1}(a_0, b_0, a_1, b_1) = \mathbf{P}[X_0(t_0) \leq a_0, X_1(t_0) \leq b_0, X_0(t_1) \leq a_1, X_1(t_1) \leq b_1].$$

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

In practice we are often interested in summaries of CPDFs afforded by certain of their moments. If we assume that they exist, the first moment (i.e., expected value or mean) and the second central moment (i.e., variance) associated with the RV X_t having CPDF $F_t(\cdot)$ are given, respectively, by

$$\mu_t \stackrel{\text{def}}{=} E\{X_t\} = \int_{-\infty}^{\infty} x \, \mathrm{d}F_t(x) \tag{24a}$$

and

$$\sigma_t^2 \stackrel{\text{def}}{=} \operatorname{var}\{X_t\} = \int_{-\infty}^{\infty} (x - \mu_t)^2 \, \mathrm{d}F_t(x),$$

where, in general, μ_t and σ_t^2 depend upon the index t. The two integrals above are examples of Riemann–Stieltjes integrals (these are briefly reviewed in the Comments and Extensions [C&Es] later in this section). If $F_t(\cdot)$ is differentiable with derivative $f_t(\cdot)$, which is called a probability density function (PDF), then the integrals reduce to the following Riemann integrals:

$$\int_{-\infty}^{\infty} x \, \mathrm{d}F_t(x) = \int_{-\infty}^{\infty} x f_t(x) \, \mathrm{d}x \quad \text{and} \quad \int_{-\infty}^{\infty} (x - \mu_t)^2 \, \mathrm{d}F_t(x) = \int_{-\infty}^{\infty} (x - \mu_t)^2 f_t(x) \, \mathrm{d}x.$$

If $F_t(\cdot)$ is a step function with steps of size p_k at locations x_k, i.e., $\mathbf{P}[X_t = x_k] = p_k$, then we have

$$\int_{-\infty}^{\infty} x \, \mathrm{d}F_t(x) = \sum_k x_k p_k \quad \text{and} \quad \int_{-\infty}^{\infty} (x - \mu_t)^2 \, \mathrm{d}F_t(x) = \sum_k (x_k - \mu_t)^2 p_k,$$

where $\sum_k p_k = 1$.

While μ_t and σ_t^2 summarize two aspects of univariate CPDFs, the covariance between the RVs X_{t_0} and X_{t_1} summarizes one aspect of their bivariate CPDF $F_{t_0,t_1}(\cdot,\cdot)$:

$$\begin{aligned} \operatorname{cov}\{X_{t_0}, X_{t_1}\} &\stackrel{\text{def}}{=} E\{(X_{t_0} - \mu_{t_0})(X_{t_1} - \mu_{t_1})\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_0 - \mu_{t_0})(x_1 - \mu_{t_1}) \, \mathrm{d}^2 F_{t_0,t_1}(x_0, x_1). \end{aligned} \tag{24b}$$

Note that $\operatorname{cov}\{X_{t_0}, X_{t_0}\} = \operatorname{var}\{X_{t_0}\}$. If $F_{t_0,t_1}(\cdot,\cdot)$ is differentiable and hence is equivalent to the bivariate PDF $f_{t_0,t_1}(\cdot,\cdot)$ in that

$$F_{t_0,t_1}(a_0, a_1) = \int_{-\infty}^{a_1} \int_{-\infty}^{a_0} f_{t_0,t_1}(x_0, x_1) \, \mathrm{d}x_0 \, \mathrm{d}x_1,$$

then we can write

$$\operatorname{cov}\{X_{t_0}, X_{t_1}\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_0 - \mu_{t_0})(x_1 - \mu_{t_1}) f_{t_0,t_1}(x_0, x_1) \, \mathrm{d}x_0 \, \mathrm{d}x_1.$$

On the other hand, suppose that X_{t_0} can only assume the values $x_{t_0,j}$ over some range of the index $j \in \mathbb{Z}$ and that X_{t_1} assumes only $x_{t_1,k}$ over $k \in \mathbb{Z}$. We then have $\mathbf{P}[X_{t_0} = x_{t_0,j}, X_{t_1} = x_{t_1,k}] = p_{j,k}$ with $\sum_j \sum_k p_{j,k} = 1$, and we can write

$$\operatorname{cov}\{X_{t_0}, X_{t_1}\} = \sum_j \sum_k (x_{0,j} - \mu_{t_0})(x_{1,k} - \mu_{t_1}) p_{j,k},$$

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

where

$$\mu_{t_0} = \sum_j x_{t_0,j} \mathbf{P}[X_{t_0} = x_{t_0,j}] = \sum_j x_{t_0,j} \sum_k p_{j,k},$$

with a similar expression for μ_{t_1}.

The above ideas extend in an obvious manner to a real-valued continuous parameter stochastic process $\{X(t)\}$, but there is one key modification required for complex-valued stochastic processes. To set the stage, suppose that $\{Z_t\}$ is a complex-valued discrete parameter stochastic process such that $Z_t = X_{0,t} + \mathrm{i}X_{1,t}$, where $\{X_{0,t}\}$ and $\{X_{1,t}\}$ are real-valued discrete parameter stochastic processes with $E\{X_{0,t}\} = \mu_{0,t}$ and $E\{X_{1,t}\} = \mu_{1,t}$. The obvious definition for the mean value of Z_t is

$$\mu_t \stackrel{\text{def}}{=} E\{Z_t\} = E\{X_{0,t}\} + \mathrm{i}E\{X_{1,t}\} = \mu_{0,t} + \mathrm{i}\mu_{1,t},$$

where now μ_t is in general complex-valued. The key modification is in the definition of the covariance between the complex-valued RVs Z_{t_0} and Z_{t_1}:

$$\mathrm{cov}\,\{Z_{t_0}, Z_{t_1}\} \stackrel{\text{def}}{=} E\{(Z_{t_0} - \mu_{t_0})(Z_{t_1} - \mu_{t_1})^*\} = E\{(Z_{t_0} - \mu_{t_0})(Z_{t_1}^* - \mu_{t_1}^*)\}, \quad (25a)$$

where the asterisk indicates the operation of complex conjugation. In particular,

$$\mathrm{var}\,\{Z_t\} = \mathrm{cov}\,\{Z_t, Z_t\} = E\{(Z_t - \mu_t)(Z_t - \mu_t)^*\} = E\{|Z_t - \mu_t|^2\},$$

where we make use of the fact that $|z|^2 = zz^*$ for any complex-valued variable z. Note that, if we set $X_{1,t} = 0$ for all t so that $\{Z_t\}$ becomes a real-valued process, the covariance defined in Equation (25a) is consistent with the definition of covariance in Equation (24b) because $(Z_{t_1} - \mu_{t_1})^* = Z_{t_1} - \mu_{t_1}$ in this special case.

Exercise [2.1] invites the reader to prove several key properties about covariances.

Comments and Extensions to Section 2.3

[1] We state here one definition for the Riemann–Stieltjes integral and a few facts concerning it (we will need a stochastic version of this integral when we introduce the spectral representation theorem for stationary processes in Section 4.1); for more details, see section 18.9 of Taylor and Mann (1972) or section 4.2 of Greene and Knuth (1990), from which the following material is adapted. Let $g(\cdot)$ and $H(\cdot)$ be two real-valued functions defined over the interval $[L, U]$ with $L < U$, and let P_N be a partition of this interval of size $N + 1$; i.e., P_N is a set of $N + 1$ points x_j such that

$$L = x_0 < x_1 < \cdots < x_{N-1} < x_N = U.$$

Define the "mesh fineness" of the partition P_N as

$$|P_N| \stackrel{\text{def}}{=} \max\{x_1 - x_0, x_2 - x_1, \ldots, x_{N-1} - x_{N-2}, x_N - x_{N-1}\}.$$

Let x_j' be any point in the interval $[x_{j-1}, x_j]$, and consider the summation

$$\mathcal{S}(P_N) \stackrel{\text{def}}{=} \sum_{j=1}^{N} g(x_j') \left[H(x_j) - H(x_{j-1})\right].$$

The Riemann–Stieltjes integral is defined as a limit involving this summation:

$$\int_L^U g(x)\,\mathrm{d}H(x) \stackrel{\text{def}}{=} \lim_{|P_N| \to 0} \mathcal{S}(P_N), \quad (25b)$$

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

provided that $\mathcal{S}(P_N)$ converges to a unique limit as the mesh fineness decreases to zero. Extension of the above to allow $L = -\infty$ and $U = \infty$ is done by the same limiting argument as is used in the case of a Riemann integral.

We now note the following facts (all of which assume that $g(\cdot)$ and $H(\cdot)$ are such that the Riemann–Stieltjes integral over $[L, U]$ exists).

[1] If $H(x) = x$, the Riemann–Stieltjes integral reduces to the ordinary Riemann integral $\int_L^U g(x)\,\mathrm{d}x$.
[2] If $H(\cdot)$ is differentiable everywhere over the interval $[L, U]$ with derivative $h(\cdot)$, the Riemann–Stieltjes integral reduces to the Riemann integral $\int_L^U g(x)h(x)\,\mathrm{d}x$.
[3] Suppose that b is such that $L < b < U$ and that

$$H(x) = \begin{cases} c, & L \le x < b; \\ a + c, & b \le x \le U; \end{cases}$$

i.e., $H(\cdot)$ is a step function with a single step of size a at $x = b$. Then

$$\int_L^U g(x)\,\mathrm{d}H(x) = ag(b).$$

In general, if $H(\cdot)$ is a step function with steps of sizes $a_0, a_1, \ldots, a_{N-1}$ at points $b_0, b_1, \ldots, b_{N-1}$ (all of which are distinct and satisfy $L < b_k < U$), then

$$\int_L^U g(x)\,\mathrm{d}H(x) = \sum_{k=0}^{N-1} a_k g(b_k).$$

The last fact shows that many ordinary summations can be expressed as Riemann–Stieltjes integrals. This gives us a certain compactness in notation. For example, as noted previously, Equation (24a) handles the special cases in which $F_t(\cdot)$ is either differentiable or is a step function. This equation also handles "mixed" CPDFs, a combination of these two special cases often discussed in elementary textbooks on statistics. The Riemann–Stieltjes integral thus gives us some – but by no means all – of the advantages of the Lebesgue integral commonly used in advanced texts on probability and statistics.

2.4 Real-Valued Stationary Processes

The class of all stochastic processes is "too large" to work with in practice. In spectral analysis, we consider only a special subclass called stationary processes. Basically, stationarity requires certain properties of a stochastic process be time-invariant.

There are two common types of stationarity. The first type is *complete* stationarity (sometimes referred to as *strong* stationarity or *strict* stationarity): the process $\{X_t\}$ is said to be completely stationary if, for all $N \ge 1$, for any $t_0, t_1, \ldots, t_{N-1}$ contained in the index set, and for any τ such that $t_0 + \tau, t_1 + \tau, \ldots, t_{N-1} + \tau$ are also contained in the index set, the joint CPDF of $X_{t_0}, X_{t_1}, \ldots, X_{t_{N-1}}$ is the same as that of $X_{t_0+\tau}, X_{t_1+\tau}, \ldots, X_{t_{N-1}+\tau}$; i.e.,

$$F_{t_0,t_1,\ldots,t_{N-1}}(a_0, a_1, \ldots, a_{N-1}) = F_{t_0+\tau,t_1+\tau,\ldots,t_{N-1}+\tau}(a_0, a_1, \ldots, a_{N-1}).$$

In other words, the probabilistic structure of a completely stationary process is invariant under a shift in time.

Unfortunately, completely stationary processes are too difficult to work with as models for most time series of interest since they have to be specified by using N-dimensional CPDFs. A simplifying assumption leads to the second common type of stationarity: the process $\{X_t\}$ is said to be *second-order* stationary (sometimes called *weakly* stationary or *covariance* stationary) if, for all $N \ge 1$, for any $t_0, t_1, \ldots, t_{N-1}$ contained in the index set, and for any τ

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

such that $t_0+\tau, t_1+\tau, \ldots, t_{N-1}+\tau$ are also contained in the index set, all the first moments, second moments and second-order joint moments of $X_{t_0}, X_{t_1}, \ldots, X_{t_{N-1}}$ exist, are finite and are equal to the corresponding moments of $X_{t_0+\tau}, X_{t_1+\tau}, \ldots, X_{t_{N-1}+\tau}$ (second-order moments take the form $E\{X_{t_j}X_{t_k}\}$; more generally, $E\{X_{t_j}^l X_{t_k}^m\}$ is called a joint moment of order $l+m$). Immediate consequences of this definition are that

$$\mu \stackrel{\text{def}}{=} E\{X_t\} \text{ and } \mu_2' \stackrel{\text{def}}{=} E\{X_t^2\}$$

are both constants independent of t. This implies that

$$\sigma^2 \stackrel{\text{def}}{=} \text{var}\,\{X_t\} = \mu_2' - \mu^2$$

is also a constant independent of t. If we allow the shift $\tau = -t_1$, we see that

$$E\{X_{t_0}X_{t_1}\} = E\{X_{t_0-t_1}X_0\}$$

is a function of the difference $t_0 - t_1$ only. Actually, it is a function of the absolute difference $|t_0 - t_1|$ only since, if we now let $\tau = -t_0$, we have

$$E\{X_{t_0}X_{t_1}\} = E\{X_{t_1}X_{t_0}\} = E\{X_{t_1-t_0}X_0\}.$$

The above implies that the covariance between X_{t_0} and X_{t_1} is also a function of the absolute difference $|t_0 - t_1|$, since

$$\text{cov}\,\{X_{t_0}, X_{t_1}\} = E\,\{(X_{t_0} - \mu)(X_{t_1} - \mu)\} = E\{X_{t_0}X_{t_1}\} - \mu^2.$$

For a discrete parameter second-order stationary process $\{X_t\}$, we define the *autocovariance sequence* (ACVS) by

$$s_\tau \stackrel{\text{def}}{=} \text{cov}\,\{X_{t+\tau}, X_t\} = \text{cov}\,\{X_\tau, X_0\}. \tag{27}$$

Likewise, for a continuous parameter second-order stationary process $\{X(t)\}$, we define the *autocovariance function* (ACVF) by

$$s(\tau) \stackrel{\text{def}}{=} \text{cov}\,\{X(t+\tau), X(t)\} = \text{cov}\,\{X(\tau), X(0)\}.$$

Both s_τ and $s(\tau)$ measure the covariance between members of a process that are separated by τ units. The variable τ is called the *lag*.

Here are some further properties of the ACVS and ACVF.

[1] For a discrete parameter process with an index set given by $\mathbb{Z}$, the lag τ can assume any integer value; for a continuous parameter stationary process with an index set given by $\mathbb{R}$, the lag can assume any real value.

[2] Note that, in the discrete parameter case, $s_0 = \sigma^2$ and $s_{-\tau} = s_\tau$; likewise, in the continuous parameter case, $s(0) = \sigma^2$ and $s(-\tau) = s(\tau)$. Thus $\{s_\tau\}$ is an even sequence, and $s(\cdot)$ is an even function.

[3] Assuming the trivial condition $s_0 > 0$, we can define $\rho_\tau = s_\tau/s_0$ as the *autocorrelation sequence* (ACS) for $\{X_t\}$; likewise, we can define $\rho(\tau) = s(\tau)/s(0)$ as the *autocorrelation function* (ACF) for $\{X(t)\}$. Since we have

$$\rho_\tau = \text{corr}\,\{X_{t+\tau}, X_t\} \stackrel{\text{def}}{=} \frac{\text{cov}\,\{X_{t+\tau}, X_t\}}{(\text{var}\,\{X_{t+\tau}\}\,\text{var}\,\{X_t\})^{1/2}} = \frac{\text{cov}\,\{X_{t+\tau}, X_t\}}{\text{var}\,\{X_t\}},$$

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

ρ_τ is the correlation coefficient between pairs of RVs from the process $\{X_t\}$ that are τ units apart. There is an analogous interpretation for $\rho(\tau)$. (As already mentioned in Chapter 1, the definitions we have given for the ACVS, ACS, ACVF and ACF are standard in the statistical literature, but other definitions for these terms are sometimes used in the engineering literature.)

[4] Since ρ_τ and $\rho(\tau)$ are correlation coefficients and hence constrained to lie between -1 and 1, it follows that

$$|s_\tau| \leq s_0 \text{ and } |s(\tau)| \leq s(0) \quad \text{for all } \tau. \tag{28a}$$

[5] A necessary and sufficient condition that an even sequence $\{s_\tau\}$ be the ACVS for some stationary process is that it be *positive semidefinite*; i.e., for all $N \geq 1$, for any t_0, t_1, $\ldots$, t_{N-1} contained in the index set, and for any set of nonzero real numbers a_0, a_1, $\ldots$, a_{N-1},

$$\sum_{j=0}^{N-1} \sum_{k=0}^{N-1} s_{t_j - t_k} a_j a_k \geq 0 \tag{28b}$$

(if this double summation is strictly greater than 0, then $\{s_\tau\}$ is said to be *positive definite*). To see that this condition is necessary, consider the following two column vectors of length N:

$$\boldsymbol{a} \stackrel{\text{def}}{=} [a_0, a_1, \ldots, a_{N-1}]^T \text{ and } \boldsymbol{V} \stackrel{\text{def}}{=} \left[X_{t_0}, X_{t_1}, \ldots, X_{t_{N-1}}\right]^T,$$

where the superscript T denotes the operation of vector transposition. Let $\boldsymbol{\Sigma}$ be the covariance matrix for the vector $\boldsymbol{V}$. Its (j,k)th element is given by

$$E\{(X_{t_j} - \mu)(X_{t_k} - \mu)\} = s_{t_j - t_k}.$$

Define the RV

$$W = \sum_{j=0}^{N-1} a_j X_{t_j} = \boldsymbol{a}^T \boldsymbol{V}.$$

Then we have

$$0 \leq \text{var}\,\{W\} = \text{var}\,\{\boldsymbol{a}^T \boldsymbol{V}\} = \boldsymbol{a}^T \boldsymbol{\Sigma} \boldsymbol{a} = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} s_{t_j - t_k} a_j a_k$$

(see Exercise [2.2]; the sufficiency of the condition is shown in theorem 1.5.1, Brockwell and Davis, 1991). The important point here is that an ACVS cannot just be any arbitrary sequence, a fact that will become important when we discuss estimators for $\{s_\tau\}$. The same comments hold – with obvious modifications – for the function $s(\cdot)$. (For a fascinating discussion on how severely the requirement of positive semidefiniteness constrains a sequence of numbers, see Makhoul, 1990.)

[6] Let us now consider the covariance matrix $\boldsymbol{\Sigma}_N$ for a vector containing N contiguous RVs from the stationary process $\{X_t\}$, say, $[X_0, X_1, \ldots, X_{N-1}]^T$. Since the (j,k)th element of $\boldsymbol{\Sigma}_N$ is given by $\text{cov}\,\{X_j, X_k\} = s_{j-k}$, the elements of this matrix depend upon just the *difference* between the row and column indices. As an example, let us consider the case $N = 5$:

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

$$\boldsymbol{\Sigma}_5 = \begin{bmatrix} s_0 & s_{-1} & s_{-2} & s_{-3} & s_{-4} \\ \underline{s_1} & s_0 & s_{-1} & s_{-2} & s_{-3} \\ s_2 & \underline{s_1} & s_0 & s_{-1} & s_{-2} \\ s_3 & s_2 & \underline{s_1} & s_0 & s_{-1} \\ s_4 & s_3 & s_2 & \underline{s_1} & s_0 \end{bmatrix} = \begin{bmatrix} s_0 & s_1 & s_2 & s_3 & s_4 \\ \underline{s_1} & s_0 & s_1 & s_2 & s_3 \\ s_2 & \underline{s_1} & s_0 & s_1 & s_2 \\ s_3 & s_2 & \underline{s_1} & s_0 & s_1 \\ s_4 & s_3 & s_2 & \underline{s_1} & s_0 \end{bmatrix}. \tag{29a}$$

In the above we have underlined the entries such that $j - k = 1$. A two-dimensional matrix with a constant value along each of its diagonals is known as a *Toeplitz matrix*. Since $s_{j-k} = s_{k-j}$ for a discrete parameter real-valued stationary process, the covariance matrix for any N contiguous RVs from this process is a symmetric Toeplitz matrix of dimension $N \times N$.

It follows from the definitions of complete and second-order stationarity that, if $\{X_t\}$ is a completely stationary process with finite variance, then it is also second-order stationary. In general, second-order stationarity does not imply complete stationarity. An important exception is a *Gaussian process* (also called a *normal process*), defined as follows: the stochastic process $\{X_t\}$ is said to be Gaussian if, for all $N \geq 1$ and for any t_0, t_1, ..., t_{N-1} contained in the index set, the joint CPDF of X_{t_0}, X_{t_1}, ..., $X_{t_{N-1}}$ is multivariate Gaussian. A second-order stationary Gaussian process is also completely stationary due to the fact that the multivariate Gaussian distribution is completely characterized by its moments of first and second order.

Hereafter the unadorned term "stationarity" will mean "second-order stationarity."

2.5 Complex-Valued Stationary Processes

We say that a discrete parameter complex-valued process $\{Z_t\}$, defined via

$$Z_t = X_{0,t} + \mathrm{i}X_{1,t}, \tag{29b}$$

is (second-order) stationary if its real and imaginary parts are *jointly stationary*, by which we mean that $\{X_{0,t}\}$ and $\{X_{1,t}\}$ by themselves are second-order stationary and, in addition, $\mathrm{cov}\{X_{0,t+\tau}, X_{1,t}\}$ is a function of τ only for all t. The stationarity of the component processes $\{X_{0,t}\}$ and $\{X_{1,t}\}$ tells us that

$$\mu \stackrel{\mathrm{def}}{=} E\{Z_t\} = \mu_0 + \mathrm{i}\mu_1 \stackrel{\mathrm{def}}{=} E\{X_{0,t}\} + \mathrm{i}E\{X_{1,t}\},$$

where μ is a complex-valued constant, and μ_0 and μ_1 are real-valued constants, all three of which are independent of t.

▷ **Exercise [29]** Show that

$$\mathrm{cov}\{Z_{t+\tau}, Z_t\} \stackrel{\mathrm{def}}{=} E\{(Z_{t+\tau} - \mu)(Z_t - \mu)^*\} \tag{29c}$$

is independent of t and hence defines the τth element s_τ of the ACVS $\{s_\tau\}$. Show also that the ACVS has the property $s_{-\tau} = s_\tau^*$. Finally, show that

$$r_\tau \stackrel{\mathrm{def}}{=} E\{(Z_{t+\tau} - \mu)(Z_t - \mu)\} \tag{29d}$$

is also independent of t, where we refer to $\{r_\tau\}$ as the *autorelation sequence*. ◁

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

Note that the above also holds for a real-valued stationary process – the complex conjugate of a real number is just the real number itself. For $\tau = 0$, we have

$$s_0 = E\{|Z_t - \mu|^2\} = \text{var}\{Z_t\}.$$

The ACS is defined in the same way as before, so we must have $\rho_{-\tau} = \rho_\tau^*$.

The positive semidefinite property of $\{s_\tau\}$ is now defined slightly differently: $\{s_\tau\}$ is said to be such if, for all $N \geq 1$, for any $t_0, t_1, \ldots, t_{N-1}$ contained in the index set, and for any set of *complex* numbers $c_0, c_1, \ldots, c_{N-1}$,

$$\text{var}\left\{\sum_{j=0}^{N-1} c_j Z_{t_j}\right\} = \sum_{j=0}^{N-1}\sum_{k=0}^{N-1} s_{t_j - t_k} c_j c_k^* \geq 0.$$

While the covariance matrix for a subsequence of a discrete parameter complex-valued stationary process is a Toeplitz matrix, it is not necessarily a symmetric Toeplitz matrix (as in the real-valued case) since $s_{-\tau} \neq s_\tau$ in general; however, because of the condition $s_{-\tau} = s_\tau^*$, the covariance matrix falls in the class of *Hermitian Toeplitz matrices*.

A continuous parameter complex-valued stationary process $\{Z(t)\}$ is defined in a similar way. In particular, we note that its ACVF satisfies $s(-\tau) = s^*(\tau)$ and is positive semidefinite.

We note the following definitions and results for later use.

[1] A complex-valued RV Z_t is said to have a *complex Gaussian* (or *normal*) distribution if its (real-valued) real and imaginary components $X_{0,t}$ and $X_{1,t}$ are bivariate Gaussian; likewise, a collection of N complex-valued RVs $Z_0, Z_1, \ldots, Z_{N-1}$ is said to follow a complex Gaussian distribution if all of its real and imaginary components are multivariate Gaussian (of dimension $2N$).

[2] Complex-valued processes for which the autorelation sequence $\{r_\tau\}$ is everywhere zero are termed *proper* (Schreier and Scharf, 2003, 2010). Note that Miller (1974a) gives a comprehensive discussion of complex-valued stochastic processes, but concentrates on proper processes (Miller, 1974a, p. 42). As a result, he defines $\{Z_t\}$ to be stationary if $\text{cov}\{Z_{t+\tau}, Z_t\}$ is independent of t for any given τ; this is a weaker condition than our definition, which implies that $\text{cov}\{Z_{t+\tau}, Z_t^*\}$ is independent of t for any given τ also, as shown in Exercise [29].

[3] The complex-valued process $\{Z_t\}$ is said to be complex Gaussian if, for all $N \geq 1$ and for any $t_0, t_1, \ldots, t_{N-1}$ contained in the index set, the joint CPDF of the real and imaginary components $X_{0,t_0}, X_{1,t_0}, X_{0,t_1}, X_{1,t_1}, \ldots, X_{0,t_{N-1}}$ and $X_{1,t_{N-1}}$ is multivariate Gaussian.

[4] If Z_0, Z_1, Z_2 and Z_3 are any four complex-valued Gaussian RVs with zero means, then the *Isserlis theorem* (Isserlis, 1918) states that

$$\text{cov}\{Z_0 Z_1, Z_2 Z_3\} = \text{cov}\{Z_0, Z_2\}\,\text{cov}\{Z_1, Z_3\} + \text{cov}\{Z_0, Z_3\}\,\text{cov}\{Z_1, Z_2\} \quad (30)$$

(in fact, the above also holds if we replace Z_0, Z_1, Z_2 and Z_3 with any four real-valued Gaussian RVs with zero means).

Although the main focus of this book is on the application of real-valued stationary processes to time series data, complex-valued processes arise indirectly in several contexts, including the spectral representation theorem (Section 4.1) and the statistical properties of spectral estimators (Section 6.6). Complex-valued stationary processes are used directly to model certain bivariate medical (e.g., Loupas and McDicken, 1990; Rowe, 2005), oceanographic (Chandna and Walden, 2011; Sykulski et al., 2016) and meteorological (Hayashi, 1979; Maitani, 1983) time series.

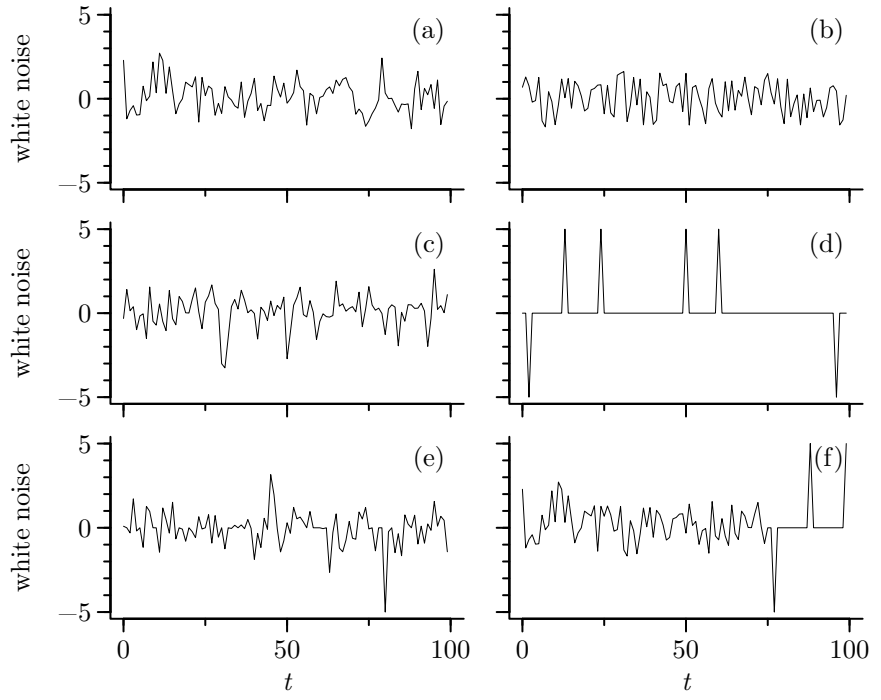

Figure 31 Realizations of length 100 from six white noise processes, all with zero mean and unit variance. The distribution for the process in plot (a) is Gaussian; (b) is uniform over the interval $[-\sqrt{3}, \sqrt{3}]$; (c) is a double exponential with PDF given by $f(x) = \exp(-|x|\sqrt{2})/\sqrt{2}$; and (d) is from a discrete distribution that assumes the values -5, 0 and 5 with probabilities 0.02, 0.96 and 0.02, respectively. The realization in (e) is, at each time index t, a random selection from one of the previous four distributions. Realization (f) was created by pasting together the first 25 values from (a), (b), (c) and (d).

Comments and Extensions to Section 2.5

[1] Picinbono and Bondon (1997) use the term *relation function* to refer to $\{r_\tau\}$ of Equation (29d), but we have added "auto" to emphasize that both RVs are from the same process, and we use "sequence" instead of "function" since we are dealing with a discrete parameter process. This quantity is also called the *complementary covariance* (Schreier and Scharf, 2003); however, note that $\{r_\tau\}$ does not have all the properties of a covariance sequence (for example, while a covariance sequence $\{s_\tau\}$ satisfies $s_{-\tau} = s_\tau^*$, this does not hold for $\{r_\tau\}$ unless it is real-valued).

2.6 Examples of Discrete Parameter Stationary Processes

White Noise Process

Let $\{X_t\}$ be a sequence of uncorrelated RVs such that $E\{X_t\} = \mu$ and $\text{var}\{X_t\} = \sigma^2 > 0$ for all t, where σ^2 is a finite constant. Since uncorrelatedness means that $\text{cov}\{X_{t+\tau}, X_t\} = 0$ for all t and $\tau \neq 0$, it follows that $\{X_t\}$ is stationary with ACVS

$$s_\tau = \begin{cases} \sigma^2, & \tau = 0; \\ 0, & \tau \neq 0, \end{cases} \quad \text{which implies} \quad \rho_\tau = \begin{cases} 1, & \tau = 0; \\ 0, & \tau \neq 0. \end{cases} \tag{31}$$

Despite their simplicity, white noise processes play a central role since they can be manipulated to create many other stationary processes (see the examples that follow).

Six examples of realizations from white noise processes are shown in Figure 31, each with $\mu = 0$ and $\sigma^2 = 1$. The first three examples are realizations of independent and identically distributed (IID) RVs with, respectively, a Gaussian, uniform and double exponential

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

distribution. The fourth example is also based on IID RVs, but with realizations that can only assume three distinct values (-5, 0 and 5). The fifth example was formed by selecting, at each time index t, one of the four previous distributions at random and then generating a realization from the selected distribution. While the distribution of this white noise process is complicated, it is the same for all t. The sixth example was created by pasting together the first 25 points from each of the first four examples. This final example demonstrates that the distribution of a white noise process need not be constant over time; i.e., the only requirements for a process to be white noise are that its RVs are uncorrelated and that the mean and variance for each RV are finite and the same for all t (a process that satisfies the more stringent requirement that its distribution is the same for all t is sometimes called an IID process or IID noise). Although the ACVSs for these six processes are identical, their realizations do not look the same – the characterization of a process by just its first- and second-order moments can gloss over potentially important features.

Finally we note that a sequence of complex-valued RVs $\{Z_t\}$ is deemed to be white noise if $\{Z_t\}$ is stationary with an ACVS that is the same as for real-valued white noise (see Equation (31)). The following exercise considers a special case of complex-valued white noise that has been used widely.

▷ **Exercise [32]** Show that, for *proper* complex-valued white noise, the real and imaginary parts of Z_t are uncorrelated and have the same variance, namely, $\sigma^2/2$ (this latter quantity is sometimes called the *semivariance* – see, for example, Lang and McClellan, 1980). ◁

Moving Average Process

The process $\{X_t\}$ is called a qth-order moving average process – denoted by MA(q) – if it can be expressed in the form

$$X_t = \mu + \epsilon_t - \theta_{q,1}\epsilon_{t-1} - \cdots - \theta_{q,q}\epsilon_{t-q} = \mu - \sum_{j=0}^{q} \theta_{q,j}\epsilon_{t-j}, \quad t \in \mathbb{Z}, \tag{32a}$$

where μ and $\theta_{j,q}$ are constants ($\theta_{q,0} \stackrel{\text{def}}{=} -1$ and $\theta_{q,q} \neq 0$), and $\{\epsilon_t\}$ is a white noise process with zero mean and variance σ_ϵ^2. In other words, a moving average process at time t is the sum of a constant μ, the tth component of a white noise process and a linear combination of the same white noise process at q previous times. Note first that $E\{X_t\} = \mu$, a constant independent of t – we assume it to be zero in what follows. Second, since $E\{\epsilon_{t+\tau}\epsilon_t\} = 0$ for all $\tau \neq 0$, it follows from Equation (32a) and Exercise [2.1e] that, for $\tau \geq 0$,

$$\operatorname{cov}\{X_{t+\tau}, X_t\} = \sum_{k=0}^{q}\sum_{j=0}^{q} \theta_{q,k}\theta_{q,j} \operatorname{cov}\{\epsilon_{t+\tau-k}, \epsilon_{t-j}\} = \sigma_\epsilon^2 \sum_{j=0}^{q-\tau} \theta_{q,j+\tau}\theta_{q,j}$$

depends only on the lag τ. Here we interpret the last summation to be zero when $q - \tau < 0$. Since it is easy to show that $\operatorname{cov}\{X_{t-\tau}, X_t\} = \operatorname{cov}\{X_{t+\tau}, X_t\}$, it follows that $\{X_t\}$ is a stationary process with ACVS given by

$$s_\tau = \begin{cases} \sigma_\epsilon^2 \sum_{j=0}^{q-|\tau|} \theta_{q,j+|\tau|}\theta_{q,j}, & |\tau| \leq q; \\ 0, & |\tau| > q. \end{cases} \tag{32b}$$

Note that we did not need to place any restrictions on the $\theta_{q,j}$ terms to ensure stationarity. Note also that the variance of $\{X_t\}$ is given by

$$s_0 = \sigma_\epsilon^2 \sum_{j=0}^{q} \theta_{q,j}^2.$$

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

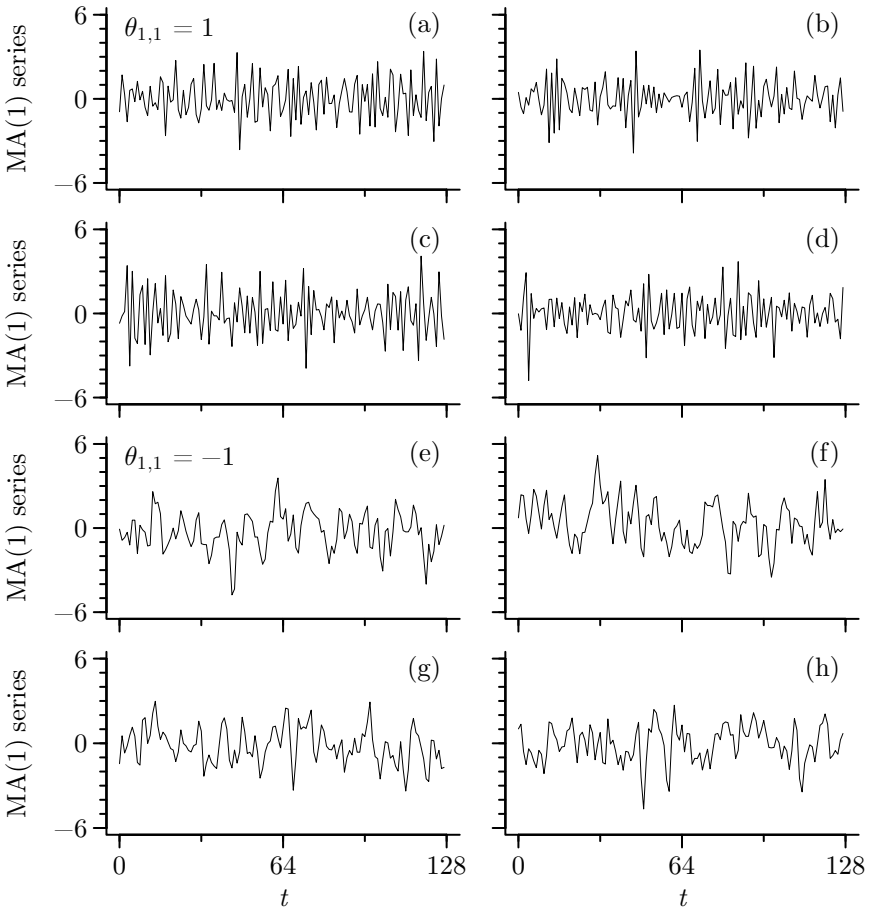

Figure 33 Realizations of length 128 of two first-order Gaussian moving average processes. The top two rows are for $\theta_{1,1} = 1$, and the bottom two, for $\theta_{1,1} = -1$.

Figure 33 shows four realizations from two different Gaussian MA(1) processes of the form $X_t = \epsilon_t - \theta_{1,1}\epsilon_{t-1}$. For the first process (top two rows), $\theta_{1,1} = 1$; for the second (bottom two), $\theta_{1,1} = -1$. Since $\rho_1 = -\theta_{1,1}/(1 + \theta_{1,1}^2)$ here, we see that adjacent values of the first process are negatively correlated ($\rho_1 = -1/2$), while adjacent values of the second process are positively correlated ($\rho_1 = 1/2$). The realizations in Figure 33 agree with this description (see Section 11.1 for a discussion on generating realizations from a moving average process).

An interesting example of a physical process that can be modeled as a moving average process is the thickness of textile slivers (an intermediate stage in converting flax fibers into yarn) as a function of displacement along a sliver (Spencer-Smith and Todd, 1941). Note that the "time" variable here is a displacement (distance) rather than physical time.

Autoregressive Process

The process $\{X_t\}$ with zero mean is called a pth-order autoregressive process – denoted by AR(p) – if it satisfies an equation such as

$$X_t = \phi_{p,1}X_{t-1} + \cdots + \phi_{p,p}X_{t-p} + \epsilon_t, \quad t \in \mathbb{Z}, \tag{33}$$

where $\phi_{p,1}, \ldots, \phi_{p,p}$ are constants (with $\phi_{p,p} \neq 0$) and $\{\epsilon_t\}$ is a white noise process with zero mean and variance σ_p^2. In other words, an AR process at time t is the sum of a white noise

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

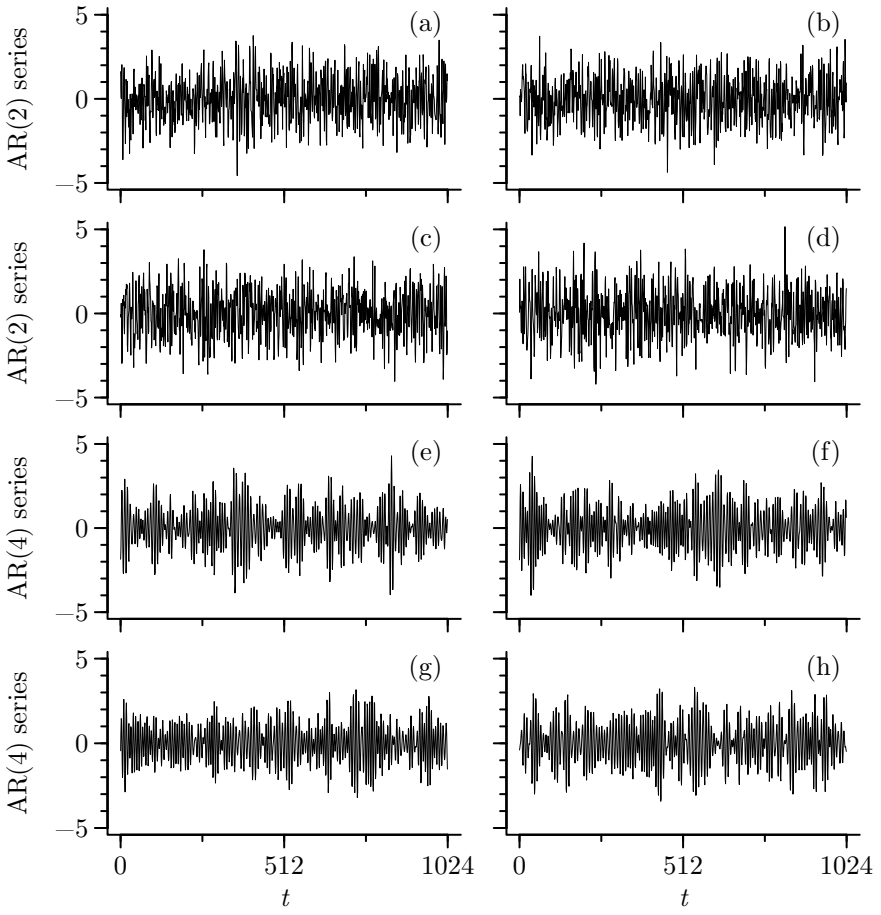

Figure 34 Realizations of length 1024 from two Gaussian autoregressive processes. The top two rows show four realizations of the AR(2) process of Equation (34), while the remaining two rows have four realizations of the AR(4) process of Equation (35a).

process at time t and a linear combination of the AR process itself at p previous times. In contrast to the parameters of an MA(q) process, the $\phi_{p,k}$ terms must satisfy certain conditions for the process $\{X_t\}$ to be stationary and "causal"; i.e., the random variable X_t just depends upon $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots$ and not $\epsilon_{t+1}, \epsilon_{t+2}, \epsilon_{t+3}, \ldots$ (for details, see Section 9.2 and C&E [5] for Section 9.4). For example, when $p = 1$, we must have $|\phi_{1,1}| < 1$ (see Exercise [2.17a]). When $p = 2$, stationarity and causality hold if the three conditions $\phi_{2,2} + \phi_{2,1} < 1$, $\phi_{2,2} - \phi_{2,1} < 1$ and $|\phi_{2,2}| < 1$ are true (Box et al., 2015). If a stationary AR(p) process is causal and nondeterministic (i.e., X_t cannot be perfectly predicted from $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$), it can be written as an infinite-order moving average process:

$$X_t = -\sum_{j=0}^{\infty} \theta_j \epsilon_{t-j},$$

where θ_j can be determined from the $\phi_{p,k}$ terms. AR(p) processes play an important role in modern spectral analysis and are discussed in detail in Chapter 9.

Figure 34 shows realizations from two Gaussian stationary and causal AR processes; the top two rows are of the AR(2) process

$$X_t = 0.75X_{t-1} - 0.5X_{t-2} + \epsilon_t \text{ with } \sigma_2^2 = 1, \tag{34}$$

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

while the two bottom rows are of the AR(4) process

$$X_t = 2.7607X_{t-1} - 3.8106X_{t-2} + 2.6535X_{t-3} - 0.9238X_{t-4} + \epsilon_t \text{ with } \sigma_4^2 = 0.002 \quad (35a)$$

(in both cases $\{\epsilon_t\}$ is a Gaussian white noise process with zero mean). These two AR processes have been used extensively in the literature as test cases (see, for example, Ulrych and Bishop, 1975, and Box et al., 2015). We use the realizations in Figure 34 as examples in Chapters 6, 7, 8 and 9 (Exercises [597] and [11.1] describe how these realizations were generated).

Autoregressive Moving Average Process

The process $\{X_t\}$ with zero mean is called an autoregressive moving average process of order (p, q) – denoted by ARMA(p,q) – if it satisfies an equation such as

$$X_t = \phi_{p,1}X_{t-1} + \cdots + \phi_{p,p}X_{t-p} + \epsilon_t - \theta_{q,1}\epsilon_{t-1} - \cdots - \theta_{q,q}\epsilon_{t-q}, \quad t \in \mathbb{Z}, \quad (35b)$$

where $\phi_{p,j}$ and $\theta_{q,j}$ are constants ($\phi_{p,p} \neq 0$ and $\theta_{q,q} \neq 0$) and again $\{\epsilon_t\}$ is a white noise process with zero mean and variance σ_ϵ^2. With the process parameters appropriately chosen, Equation (35b) describes a rich class of stationary processes that can successfully model a wide range of time series (Box et al., 2015, treat ARMA processes at length).

Harmonic Process

The process $\{X_t\}$ is called a harmonic process if it can be written as

$$X_t = \mu + \sum_{l=1}^{L} A_l \cos(2\pi f_l t) + B_l \sin(2\pi f_l t), \quad t \in \mathbb{Z}, \quad (35c)$$

where μ and $f_l > 0$ are real-valued constants, L is a positive integer, and A_l and B_l are uncorrelated real-valued RVs with zero means such that $\text{var}\{A_l\} = \text{var}\{B_l\}$ (Anderson, 1971; Newton, 1988). We can reexpress a harmonic process as

$$X_t = \mu + \sum_{l=1}^{L} D_l \cos(2\pi f_l t + \phi_l), \quad t \in \mathbb{Z}, \quad (35d)$$

where $D_l^2 = A_l^2 + B_l^2$, $\tan(\phi_l) = -B_l/A_l$, $A_l = D_l \cos(\phi_l)$ and $B_l = -D_l \sin(\phi_l)$ (these relationships are based in part upon the trigonometric identity $\cos(x+y) = \cos(x)\cos(y) - \sin(x)\sin(y)$). If we adopt the convention that $-\pi < \phi_l \leq \pi$, and if we draw a line from the origin of a graph to the point $(A_l, -B_l)$, then D_l is a nonnegative random amplitude describing the length of the line, and ϕ_l is a random phase representing the smaller of the two angles that the line makes with the positive part of the real axis. The RV ϕ_l is related to A_l and B_l via $\arctan(-B_l/A_l)$ if we use the version of this function that pays attention to the quadrant in which $(A_l, -B_l)$ lies. (The model we considered in Equation (8a) of Chapter 1 is a special case of a harmonic process.)

Consider momentarily the case $\mu = 0$, $L = 1$ and $f_1 = 1/20$, and assume that A_1 and B_1 are Gaussian RVs with zero mean and unit variance. Three realizations of length $N = 100$ from this harmonic process are shown in the left-hand column of Figure 36. These realizations are sinusoids oscillating with a period of $1/f_1 = 20$ and differing only in their amplitudes and phases. They are quite different from the realizations of the white noise, MA and AR processes depicted in Figures 31, 33 and 34. In particular, once we have observed a small portion of a realization of this harmonic process, we know what all its other values must be. The same statement cannot be made for realizations of white noise, MA or AR processes. Despite the fact that realizations of this simple harmonic process differ markedly from ones we have seen so far for stationary processes, we have the following remarkable result.

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

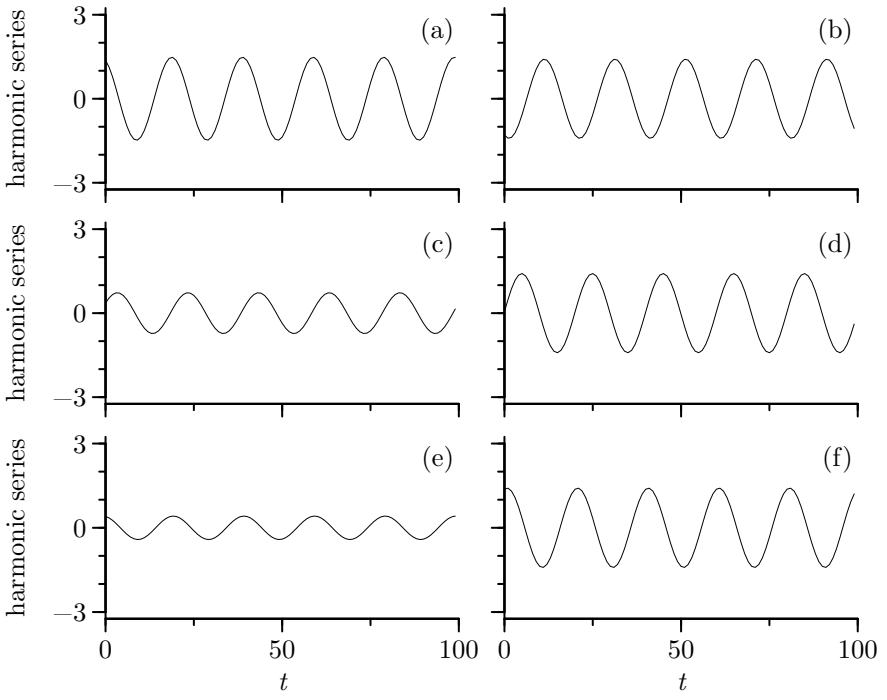

Figure 36 Realizations of length 100 from two harmonic processes for which $\mu = 0$, $L = 1$ and $f_1 = 1/20$. The realizations in the left-hand column are from a process whose distribution for each X_t is Gaussian, while those in the right-hand column have a non-Gaussian distribution that is restricted in its assumable values to the interval $[-\sqrt{2}, \sqrt{2}]$.

▷ **Exercise [36]** Letting $\sigma_l^2 \stackrel{\text{def}}{=} \text{var}\{A_l\} = \text{var}\{B_l\}$, show that the harmonic process of Equation (35c) is a stationary process with mean μ and with an ACVS given by

$$s_\tau = \sum_{l=1}^{L} \sigma_l^2 \cos(2\pi f_l \tau). \tag{36a}$$ ◁

If we compare Equations (35d) and (36a), we see that both the harmonic process and its ACVS consist of sums of cosine waves with exactly the same frequencies, but that all of the cosine terms are "in phase" for the ACVS; i.e., the ϕ_l RVs have been replaced by zeros. Note that the sequence $\{s_\tau\}$ does not damp down to zero as τ gets large. Because

$$s_0 = \sum_{l=1}^{L} \sigma_l^2, \tag{36b}$$

it follows that the ACS is given by

$$\rho_\tau = \sum_{l=1}^{L} \sigma_l^2 \cos(2\pi f_l \tau) \Big/ \sum_{l=1}^{L} \sigma_l^2.$$

Let us consider two specific examples of harmonic processes. For the first example, we assume the RVs A_l and B_l have a Gaussian distribution. Since a linear combination of Gaussian RVs is also a Gaussian RV, it follows from Equation (35c) that the resulting process

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

$\{X_t\}$ is a Gaussian stationary process. Recall that, if $Y_0, Y_1, \ldots, Y_{\nu-1}$ are independent zero mean, unit variance Gaussian RVs, then the RV

$$\chi^2_\nu \stackrel{\text{def}}{=} Y_0^2 + Y_1^2 + \cdots + Y_{\nu-1}^2 \tag{37a}$$

has a chi-square distribution with ν degrees of freedom. Since A_l/σ_l and B_l/σ_l are independent zero mean, unit variance Gaussian RVs, it follows that $D_l^2/\sigma_l^2 = (A_l^2 + B_l^2)/\sigma_l^2$ has a chi-square distribution with two degrees of freedom. The PDF for the RV χ^2_2 is given by

$$f_{\chi^2_2}(u) \stackrel{\text{def}}{=} \begin{cases} \mathrm{e}^{-u/2}/2, & u \geq 0; \\ 0, & u < 0, \end{cases}$$

from which we can deduce that the PDF for D_l^2 is given by

$$f_{D_l^2}(u) \stackrel{\text{def}}{=} \begin{cases} \mathrm{e}^{-u/(2\sigma_l^2)}/(2\sigma_l^2), & u \geq 0; \\ 0, & u < 0. \end{cases}$$

This is a special case of an exponential PDF $f(u) = \exp(-u/\lambda)/\lambda$ with a mean value of $\lambda = 2\sigma_l^2$. The random amplitude D_l is thus the square root of an exponential RV, which is said to obey a Rayleigh distribution. Anderson (1971, p. 376) notes that the symmetry of the bivariate Gaussian distribution for A_l and B_l dictates that ϕ be uniformly distributed over the interval $(-\pi, \pi]$ and be independent of D_l. Thus formulation of a Gaussian harmonic process via Equation (35d) involves Rayleigh distributed D_l and uniformly distributed ϕ_l, with all $2L$ RVs being independent of one another.

Whereas a Gaussian-distributed harmonic process consists of random amplitudes and uniformly distributed random phases, our second example shows that we can dispense with the random amplitudes and still have a harmonic process.

▷ **Exercise [37]** Suppose X_t obeys Equation (35d), but with the stipulation that the D_l terms are real-valued constants, while the ϕ_l terms are independent RVs, each having a uniform distribution on the interval $(-\pi, \pi]$. Show that the process so defined is in fact a harmonic process and that its ACVS is given by

$$s_\tau = \sum_{l=1}^{L} D_l^2 \cos(2\pi f_l \tau)/2, \tag{37b}$$ ◁

Three realizations from this harmonic process for the case of $\mu = 0$, $L = 1$, $f_1 = 1/20$ and $D_1 = \sqrt{2}$ are shown in the right-hand column of Figure 36. This choice of parameters yields an ACVS that is identical to the Gaussian-distributed harmonic process behind the realizations in the left-hand column. Note that each realization has a fixed amplitude so that X_t is restricted to the interval $[-\sqrt{2}, \sqrt{2}]$ both from one realization to the next and within a given realization. The fact that X_t is bounded implies that it cannot have a Gaussian distribution.

Our second example can be called a "random phase" harmonic process and suggests that random phases are more fundamental than random amplitudes in formulating a stationary process (see Exercise [2.19] for another connection between stationarity and a process whose realizations can also be thought of as "randomly phased" versions of a basic pattern). In fact, in place of our more general formulation via Equation (35c), some authors *define* a harmonic process as having fixed amplitudes and random phases (see, e.g., Priestley, 1981). In practical applications, however, a random phase harmonic process is often sufficient – particularly since we would need to have multiple realizations of this process in order to distinguish it from one

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

having both random amplitudes and phases. (Exercise [2.20] explores the consequence of allowing the random phases to come from a nonuniform distribution.)

Note that any given realization from a harmonic process can be regarded as a deterministic function since the realizations of the RVs in right-hand sides of either Equation (35c) or (35d) determine X_t for all t. Assuming the frequencies f_l and mean μ are known, knowledge of any segment of a realization of X_t with a length at least as long as $2L$ is enough to fully specify the entire realization. In one sense, randomizing the phases (or amplitudes and phases) is a mathematical trick allowing us to treat models like Equation (35d) within the context of the theory of stationary processes. In another sense, however, *all* stationary processes can be written as a generalization of a harmonic process with an infinite number of terms – as we shall see, this is the essence of the spectral representation theorem to be discussed in Chapter 4 (harmonic analysis itself is studied in more detail in Chapter 10, and simulation of harmonic processes is discussed in Section 11.1).

It should be noted that the independence of the ϕ_l RVs is a sufficient condition to ensure stationarity (as demonstrated by Exercise [2.21], an assumption of just uncorrelatedness does not suffice in general). Although independence of phases is often assumed for convenience when the random phase model described by Equation (35d) is fit to actual data, this assumption can lead to subtle problems. Walden and Prescott (1983) modeled tidal elevations using the random phase harmonic process approach. The disagreement (often small) between the calculated and measured tidal PDFs is largely attributable to the lack of phase independence of the constituents of the tide (see Section 10.1 for details).

2.7 Comments on Continuous Parameter Processes

All of the stationary processes mentioned in Section 2.6 – with the important exception of harmonic processes – can be constructed by taking (possibly infinite) linear combinations of discrete parameter white noise. One might suppose that we can likewise construct numerous examples of continuous parameter stochastic processes by taking linear combinations of continuous parameter white noise. There is a fundamental technical difficulty with this approach – a continuous parameter white noise process with properties similar to those of a discrete parameter process does not exist; i.e., a continuous parameter process $\{\epsilon(t)\}$ with an ACVF given by

$$s(\tau) = \begin{cases} \sigma^2 > 0, & \tau = 0; \\ 0, & \text{otherwise}, \end{cases}$$

does not exist! It is possible to circumvent these technical problems and to deal with a "fictitious" continuous parameter white noise process to construct a wide range of continuous parameter stationary processes. This fiction is useful in much the same way as the Dirac delta function is a useful fiction (as long as certain precautions are taken). We do not deal with it directly in this book since our primary concern is analyzing time series using techniques appropriate for a digital computer. We thus do not need to construct continuous parameter stationary processes directly – though we do deal with them indirectly through the process of sampling (Section 4.5). For our purposes the spectral representation theorem for continuous parameter stationary processes will serve as the necessary foundation for future study (see C&E [3] for Section 4.1).

2.8 Use of Stationary Processes as Models for Data

The chief use of stationary processes in this book is to serve as models for various time series. Although the concept of stationarity is defined for models and not for data (see Exercise [2.19]), it is proper to ask whether certain time series can be usefully modeled by a given stationary process. We should be careful about drawing conclusions about a time series

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

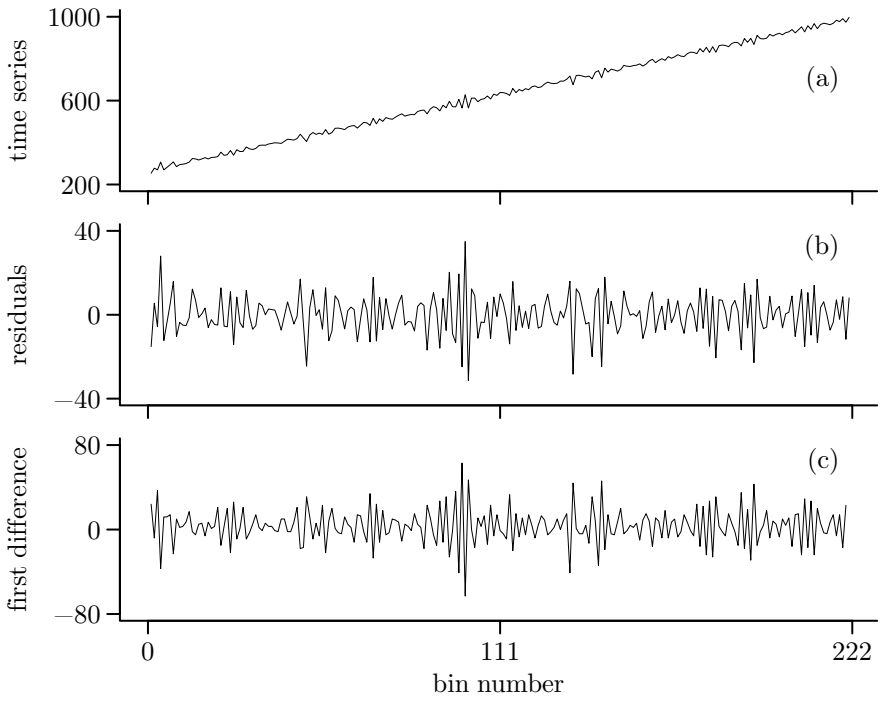

Figure 39 Spinning rotor time series. From top to bottom, the plots are (a) the original data X_t; (b) the residuals $\hat{Y}_t$ from a linear least squares fit; and (c) the first difference $X_t^{(1)}$ of the original data. The data are measured in units of microseconds. This data set was collected by S. Wood at the then National Bureau of Standards (now called the National Institute of Standards and Technology) and made available to the authors by J. Filliben.

based upon analysis techniques (such as spectral analysis) that assume stationarity when the assumed model is not likely to generate realizations that are reasonable alternatives to what was actually observed.

There are many ways in which a time series can be mismatched to a particular stationary process. Fortunately, some common violations are relatively easy to patch up. Here we examine two types of patchable nonstationarities by considering two physical time series for which use of, e.g., a Gaussian-distributed stationary AR or MA process as a model is suspect.

The first example is the time series shown in the top plot of Figure 39. These data concern a spinning rotor that is used to measure air density (i.e., pressure). During the period over which the data were collected, the rotor was always slowing down – the more molecules the rotor hit, the faster it slowed down. The data plotted are the amount of time (measured in microseconds) that it took the rotor to make 400 revolutions. The x-axis is the bin number: the first bin corresponds to the first group of 400 observed revolutions, and the tth bin, to the tth such group. (These data are a good example of a time series for which the "time index" is not actually time – however the values of the series are measured in time!)

As the spinning rotor slowed down, the time it took to make 400 revolutions necessarily increased. If we use a Gaussian-distributed stochastic process $\{X_t\}$ as a model for these data, it is unrealistic to assume that it is stationary. This requires that $E\{X_t\}$ be a constant independent of time, which from physical considerations (and Figure 39) is unrealistic. However, the time dependence of $E\{X_t\}$ appears to be linear, so a reasonable model might be

$$X_t = \alpha + \beta t + Y_t, \tag{39}$$

where α and β are unknown parameters and $\{Y_t\}$ is a stationary process with zero mean.

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

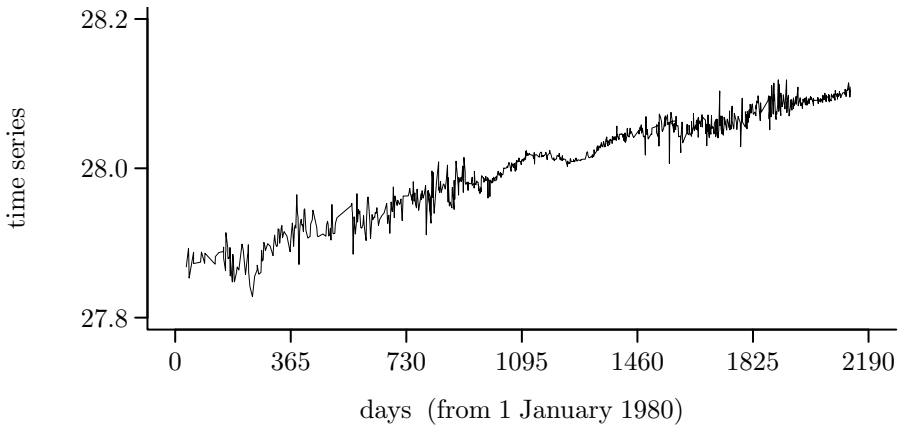

Figure 40 Standard resistor time series. The plot shows the value of a resistor measured over approximately a six-year period. This data set was collected by S. Dziuba at the then National Bureau of Standards (now called the National Institute of Standards and Technology) and made available to the authors by J. Filliben.

How we proceed now depends upon what questions we want answered about the data. For heuristic purposes, let us assume that we want to examine the covariance structure of $\{Y_t\}$. Two important ways of recovering $\{Y_t\}$ from $\{X_t\}$ are shown in the middle and bottom plots of Figure 39. The middle plot shows the residuals from an ordinary linear least squares fit to the parameters α and β; i.e.,

$$\hat{Y}_t \stackrel{\text{def}}{=} X_t - \hat{\alpha} - \hat{\beta}t,$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the linear least squares estimates of α and β. The residual $\hat{Y}_t$ only approximates the unknown Y_t, but nonetheless we can analyze $\{\hat{Y}_t\}$ to get some idea of the covariance structure of $\{Y_t\}$.

A second way of analyzing $\{Y_t\}$ is to examine the first difference of the data:

$$X_t^{(1)} \stackrel{\text{def}}{=} X_t - X_{t-1} = \beta + Y_t^{(1)}, \text{ where } Y_t^{(1)} \stackrel{\text{def}}{=} Y_t - Y_{t-1}. \tag{40}$$

This is shown in the bottom plot of Figure 39 (the sample mean of the $X_t^{(1)}$ values is an estimator of β and is equal to 3.38 here). It can be shown (by arguments similar to those needed for Exercise [2.23]) that the first difference of a stationary process is also a stationary process – hence, if $\{Y_t\}$ is stationary, so is $\{Y_t^{(1)}\}$. Moreover, by using the theory of linear filters (see Chapter 5), it is possible to relate the covariance (or spectral) properties of $\{Y_t^{(1)}\}$ to those of $\{Y_t\}$. Because of this, differencing is often preferred over linear least squares to effectively remove a linear trend in a time series.

An extension to simple differencing for removing a trend can be applied when there is seasonality, or periodicity, in the data with period p (Box et al., 2015). If p is a multiple of the sampling interval, we could choose to examine $X_t - X_{t-p}$, which is called lag p differencing (Brockwell and Davis, 2016). Nature might choose to be less cooperative as in the case of tidal analysis where the dominant tidal period is 12.42 hours and the sampling interval is often 1 hour. Conradsen and Spliid (1981) showed that in this case the periodicity can be largely removed by using $X_t - (1-\delta)X_{t-n} - \delta X_{t-n-1}$, where $n = \lfloor p \rfloor$ and $\delta = p - n$ (these are 12 and 0.42, respectively, in the tidal example). However, as explained in Section 2.6, harmonic processes incorporating such periodicities can be examined within the framework of stationary processes if desired (see also Chapter 10).

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

The second example of a time series for which a Gaussian-distributed stationary process is a questionable model is shown in Figure 40. The data are the resistance of a standard resistor over approximately a six-year period. Here we see a linear or, perhaps, quadratic drift over time. If we assume only a linear drift, an appropriate model for these data might be the same stochastic process considered for the spinning rotor data (Equation (39)). However, here it might be questionable to assume that $\{Y_t\}$, the deviations from the linear drift, is a stationary process. If we compare the data from about days 1095 to 1460 with the rest of the series, it appears that the series has a variance that changes with time – a violation of the stationarity assumption. There are no simple transformations that correct this problem. One common approach is to use different stationary processes to model various chunks of the data. Here two models might work – one for the deviations about a linear drift for days 1095 to 1460, and a second for the data before and after.

2.9 Exercises

[2.1] Here we consider some basic properties of covariances. In what follows, assume for generality that all RVs and constants are complex-valued (the definition of covariance for complex-valued RVs is given in Equation (25a)). All RVs are denoted by Z (or subscripted versions thereof), while c with or without a subscript denotes a constant. Note that, as usual, Z^* denotes the complex conjugate of Z and that $E\{Z^*\} = (E\{Z\})^*$.

(a) Show that $\mathrm{cov}\,\{Z, c\} = 0$.
(b) Show that $\mathrm{cov}\,\{Z_0, Z_1\} = \big(\mathrm{cov}\,\{Z_1, Z_0\}\big)^*$.
(c) Show that $\mathrm{cov}\,\{Z_0 + c_0, Z_1 + c_1\} = \mathrm{cov}\,\{Z_0, Z_1\}$.
(d) Suppose that at least one of the RVs Z_0 and Z_1 has a zero mean. Show that $\mathrm{cov}\,\{Z_0, Z_1\} = E\{Z_0 Z_1^*\}$.
(e) Show that

$$\mathrm{cov}\left\{\sum_j c_{0,j} Z_{0,j}, \sum_k c_{1,k} Z_{1,k}\right\} = \sum_j \sum_k c_{0,j} c_{1,k}^* \,\mathrm{cov}\,\{Z_{0,j}, Z_{1,k}\},$$

where j and k range over finite sets of integers.

Since real-valued RVs and constants are special cases of complex-valued entities, the results above continue to hold when some or all of the RVs and constants in question are real-valued. In particular, when Z_0 and Z_1 are both real-valued, part (b) simplifies to $\mathrm{cov}\,\{Z_0, Z_1\} = \mathrm{cov}\,\{Z_1, Z_0\}$. Also, when Z_1 and $c_{1,k}$ in parts (d) and (e) are real-valued, we can simplify Z_1^* to Z_1, and $c_{1,k}^*$ to $c_{1,k}$.

[2.2] Show that, if $W_0, W_1, \ldots, W_{N-1}$ are any set of $N \geq 2$ real-valued RVs with finite variances, and if $a_0, a_1, \ldots, a_{N-1}$ are any corresponding set of real numbers, then

$$\mathrm{var}\left\{\sum_{j=0}^{N-1} a_j W_j\right\} = \sum_{j=0}^{N-1} a_j^2 \,\mathrm{var}\,\{W_j\} + 2\sum_{j=0}^{N-2}\sum_{k=j+1}^{N-1} a_j a_k \,\mathrm{cov}\,\{W_j, W_k\}. \qquad (41)$$

If we further let $\boldsymbol{a} \stackrel{\mathrm{def}}{=} [a_0, a_1, \ldots, a_{N-1}]^T$ and $\boldsymbol{W} \stackrel{\mathrm{def}}{=} \big[W_0, W_1, \ldots, W_{N-1}\big]^T$, and if we let $\boldsymbol{\Sigma}$ be the covariance matrix for the vector $\boldsymbol{W}$ (i.e., the (j,k)th element of $\boldsymbol{\Sigma}$ is $\mathrm{cov}\,\{W_j, W_k\}$), show that Equation (41) implies that $\mathrm{var}\,\{\boldsymbol{a}^T\boldsymbol{W}\} = \boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a}$.

[2.3] Let the real-valued sequence $\{x_{0,t}\}$ be defined by

$$x_{0,t} = \begin{cases} +1, & t = 0, -1, -2, \ldots; \\ -1, & t = 1, 2, 3, \ldots, \end{cases}$$

and let the real-valued sequence $\{x_{1,t}\}$ be defined by $x_{1,t} = -x_{0,t}$. Suppose we have a discrete parameter stochastic process whose ensemble consists of just $\{x_{0,t}\}$ and $\{x_{1,t}\}$. We generate

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

realizations of this process by picking either $\{x_{0,t}\}$ or $\{x_{1,t}\}$ at random with probability $1/2$ each. Let $\{X_t\}$ represent the process itself.

(a) What is $E\{X_t\}$?
(b) What is $\text{var}\{X_t\}$?
(c) What is $\text{cov}\{X_{t+\tau}, X_t\}$?
(d) Is $\{X_t\}$ a (second-order) stationary process?

[2.4] Suppose that the real-valued RVs X_0, X_1 and X_2 have zero means and a covariance matrix given by

$$\boldsymbol{\Sigma}_3 = \begin{bmatrix} 1.00 & 0.69 & 0.43 \\ 0.69 & 1.00 & 0.30 \\ 0.43 & 0.30 & 1.00 \end{bmatrix}.$$

Is it possible for these RVs to be part of a stationary process $\{X_t\}$?

[2.5] (a) Suppose that $\{X_t\}$ and $\{Y_t\}$ are both stationary processes with zero means and with ACVSs $\{s_{X,\tau}\}$ and $\{s_{Y,\tau}\}$, respectively. If $\{X_t\}$ and $\{Y_t\}$ are uncorrelated with each other, show that the ACVS $\{s_{Z,\tau}\}$ of the complex-valued process defined by $Z_t = X_t + \mathrm{i}Y_t$ must be real-valued. Can $\{s_{X,\tau}\}$ and $\{s_{Y,\tau}\}$ be determined if only $\{s_{Z,\tau}\}$ is known?

(b) Suppose that $Y_t = X_{t+k}$ for some integer k; i.e., the process $\{Y_t\}$ is just a shifted version of $\{X_t\}$. What is $\{s_{Z,\tau}\}$ now?

[2.6] Let $\{Z_t\}$ be a complex-valued stochastic process. For such a process we defined the notion of stationarity by making certain stipulations on the RVs associated with the real and imaginary components of $\{Z_t\}$ (see Section 2.5).

(a) Suppose $\{Z_t\}$ takes the form $Z_t = \epsilon_t \mathrm{e}^{\mathrm{i}t}$, where $\{\epsilon_t\}$ is a real-valued white noise process with mean zero and variance σ_ϵ^2. Find $\text{cov}\{Z_{t+\tau}, Z_t\}$, and show that $\{Z_t\}$ is *not* stationary.

(b) Construct a complex-valued process $\{\widetilde{Z}_t\}$ that is stationary, but is such that

$$\text{cov}\{\widetilde{Z}_{t+\tau}, \widetilde{Z}_t\} = \text{cov}\{Z_{t+\tau}, Z_t\}$$

for all τ, where $\{Z_t\}$ is defined as in part (a). (This example demonstrates that there can be more than one complex-valued stochastic process with the same covariance structure.)

[2.7] (a) Let $\{X_{0,t}\}, \{X_{1,t}\}, \ldots, \{X_{m-1,t}\}$ each be stationary processes with means $\mu_0, \mu_1, \ldots, \mu_{m-1}$ and ACVSs $\{s_{0,\tau}\}, \{s_{1,\tau}\}, \ldots, \{s_{m-1,\tau}\}$, respectively; i.e., $E\{X_{j,t}\} = \mu_j$ and $s_{j,\tau} = \text{cov}\{X_{j,t+\tau}, X_{j,t}\}$. If $X_{j,t}$ and $X_{k,u}$ are uncorrelated for all t, u and $j \neq k$ (i.e., $\text{cov}\{X_{j,t}, X_{k,u}\} = 0$ when $j \neq k$) and if $c_0, c_1, \ldots, c_{m-1}$ are arbitrary real-valued variables, show that

$$X_t \stackrel{\text{def}}{=} \sum_{j=0}^{m-1} c_j X_{j,t}$$

is a stationary process, and determine its ACVS. (The procedure of forming a new process by making a linear combination of several processes is sometimes called *aggregation*.)

(b) Let a be any real-valued constant. Consider the sequence $s_0 = 2 + a^2$, $s_1 = s_{-1} = -a$ and $s_\tau = 0$ for $|\tau| > 1$. Is this a valid ACVS? Hint: consider Equation (32b).

[2.8] Suppose that $\{Y_t\}$ is a stationary process with ACVS $\{s_{Y,\tau}\}$. We subject this process to a lag $p = 12$ seasonal differencing to obtain $X_t \stackrel{\text{def}}{=} Y_t - Y_{t-12}$, following which we form $W_t \stackrel{\text{def}}{=} X_t - X_{t-1}$, i.e., the first difference of X_t. Show that the process $\{W_t\}$ is stationary with mean zero and an ACVS given by

$$\begin{aligned} s_{W,\tau} = {} & 4s_{Y,\tau} - 2(s_{Y,\tau+1} + s_{Y,\tau-1} + s_{Y,\tau+12} + s_{Y,\tau-12}) \\ & + s_{Y,\tau+11} + s_{Y,\tau-11} + s_{Y,\tau+13} + s_{Y,\tau-13}. \end{aligned} \tag{42}$$

(An example of a time series for which it is reasonable to apply both seasonal and first differencing is the CO_2 data discussed in Section 7.12.)

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

[2.9] Suppose that X_0, X_1, X_2 and X_3 are any four real-valued Gaussian RVs with zero means. Use the Isserlis theorem (Equation (30)) to show that

$$E\{X_0X_1X_2X_3\} = E\{X_0X_1\}E\{X_2X_3\} + E\{X_0X_2\}E\{X_1X_3\} + E\{X_0X_3\}E\{X_1X_2\}. \tag{43}$$

[2.10] (a) Let $\{X_t\}$ be a real-valued stationary Gaussian process with mean zero and ACVS $\{s_{X,\tau}\}$. Show that $Y_t = X_t^2$ is a stationary process whose ACVS is given by $s_{Y,\tau} = 2s_{X,\tau}^2$. Hint: make use of the Isserlis theorem (Equation (30)).

(b) Find a real-valued stationary Gaussian process $\{X_t\}$ such that $Y_t = X_t^2$ has an ACVS given by

$$s_{Y,\tau} = \begin{cases} 200, & \tau = 0; \\ 0, & |\tau| = 1; \\ 18, & |\tau| = 2; \\ 0, & \text{otherwise.} \end{cases}$$

Hint: consider Equation (32b), recalling that $\theta_{q,0} \stackrel{\text{def}}{=} -1$.

[2.11] Let W be a real-valued random variable with mean $\mu = 0$ and finite variance $\sigma^2 > 0$, and let c be a real-valued constant. Define a discrete parameter stochastic process $\{X_t\}$ by letting $X_t = W\cos(ct)$ for all integers t.

(a) Show that $\{X_t\}$ is a stationary process if and only if $c = l\pi$ for some integer l, and, for each l, determine the corresponding ACVS $\{s_{l,\tau}\}$. How many unique ACVSs are there in all?

(b) Recall that a sequence $\{s_\tau\}$ is said to be positive semidefinite if the inequality in Equation (28b) holds. Verify directly that this defining inequality holds for each $\{s_{l,\tau}\}$. Is $\{s_{l,\tau}\}$ for any l in fact positive definite as opposed to just positive semidefinite?

(c) Suppose now that $\mu \neq 0$. Is $\{X_t\}$ a stationary process?

[2.12] Show that, if $\{s_\tau\}$ is the ACVS for a stationary process, then, for any $\alpha \geq 0$, the sequence $\{s_\tau^2 + \alpha s_\tau\}$ is the ACVS for some stationary process. (Hint: consider a product of stationary processes with different means, but with the same ACVS.) Show also that, when $\alpha < 0$, the sequence $\{s_\tau^2 + \alpha s_\tau\}$ need not be a valid ACVS.

[2.13] Suppose that $\{\epsilon_t\}$ is a white noise process with zero mean and variance σ_ϵ^2. Define the process $\{X_t\}$ via

$$X_t = \begin{cases} \epsilon_0, & t = 0; \\ X_{t-1} + \epsilon_t, & t = 1, 2, \ldots. \end{cases}$$

Is $\{X_t\}$ a stationary process? (This process is one form of a *random walk*.)

[2.14] In the defining Equation (32a) for a qth-order moving average process, we set $\theta_{q,0} \stackrel{\text{def}}{=} -1$. In terms of the covariance structure for moving average processes, why do we not gain more generality by letting $\theta_{q,0}$ be any arbitrary number?

[2.15] Show that the first-order moving average process defined by $X_t = \epsilon_t - \theta\epsilon_{t-1}$ can be written in terms of previous values of the process as

$$X_t = \epsilon_t - \sum_{j=1}^{p} \theta^j X_{t-j} - \theta^{p+1}\epsilon_{t-p-1}$$

for any positive integer p. In the above, what condition on θ must hold in order that X_t can be expressed as an infinite-order AR process, i.e.,

$$X_t = \epsilon_t - \sum_{j=1}^{\infty} \theta^j X_{t-j}?$$

(A moving average process that can be written as an infinite-order AR process is called *invertible*. For details, see, e.g., Brockwell and Davis, 1991, Section 3.1.)

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

[2.16] Consider the following moving average process of infinite order:

$$X_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j}, \quad \text{where } \theta_j = \begin{cases} 1, & j = 0; \\ 1/\sqrt{j}, & j \geq 1, \end{cases}$$

and $\{\epsilon_t\}$ is a zero mean white noise process with unit variance. Is $\{X_t\}$ a stationary process?

[2.17] (a) Consider the AR(1) process $X_t = \phi X_{t-1} + \epsilon_t$, where, as usual, $\{\epsilon_t\}$ is a zero mean white noise process with variance σ_ϵ^2. Show that this process can be expressed as

$$X_t = \phi^q X_{t-q} + \sum_{j=0}^{q-1} \phi^j \epsilon_{t-j} \tag{44a}$$

for any positive integer q. If $|\phi| < 1$, it can be argued that the term $\phi^q X_{t-q}$ becomes negligible as q gets large and hence that $\{X_t\}$ is a stationary process with an infinite-order moving average representation given by

$$X_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} \tag{44b}$$

(see the discussion on causality in C&E [1] for Section 9.2). Use the above representation to show that the ACVS for $\{X_t\}$ is given by

$$s_\tau = \frac{\phi^{|\tau|} \sigma_\epsilon^2}{1 - \phi^2} \tag{44c}$$

(see Exercise [9.2a] for another way to obtain this ACVS).

(b) Consider the ARMA(1,1) process $X_t = \alpha X_{t-1} + \epsilon_t + \alpha \epsilon_{t-1}$. Show that this process can be expressed as

$$X_t = \alpha^q X_{t-q} + \epsilon_t + 2 \sum_{j=1}^{q-1} \alpha^j \epsilon_{t-j} + \alpha^q \epsilon_{t-q} \tag{44d}$$

for any integer $q \geq 2$. If $|\alpha| < 1$, it can be argued that $\{X_t\}$ is a stationary process with an infinite-order moving average representation given by

$$X_t = \epsilon_t + 2 \sum_{j=1}^{\infty} \alpha^j \epsilon_{t-j}.$$

Using the above representation, show that the ACVS for $\{X_t\}$ is given by

$$s_\tau = \sigma_\epsilon^2 \left(-\delta_\tau + 2\alpha^{|\tau|} \frac{1 + \alpha^2}{1 - \alpha^2} \right),$$

where δ_τ is Kronecker's delta function; i.e., $\delta_\tau = 1$ when $\tau = 0$, and $\delta_\tau = 0$ when $\tau \neq 0$.

[2.18] Let $\{\epsilon_t\}$ be a white noise process with mean zero and variance σ_ϵ^2. Let c and ϕ be real-valued constants. Let $X_0 = c\epsilon_0$ and $X_t = \phi X_{t-1} + \epsilon_t, \ \ t = 1, 2, \ldots$. Let $\boldsymbol{\Sigma}_3$ denote the covariance matrix for X_1, X_2 and X_3; i.e., the (j, k)th element of this matrix is cov $\{X_j, X_k\}$.

(a) Determine the nine elements of $\boldsymbol{\Sigma}_3$ in terms of σ_ϵ^2, c and ϕ.

(b) By considering the case $c = 0$ so that $X_0 = 0$, argue that, unless $\phi = 0$, the matrix $\boldsymbol{\Sigma}_3$ need not be of the form required for X_1, X_2 and X_3 to be three contiguous RVs from a stationary process.

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

(c) Show that, if we make the restriction $|\phi| < 1$ and if we set $c = 1/\sqrt{(1-\phi^2)}$, then $\boldsymbol{\Sigma}_3$ has the required form.

[2.19] This exercise illustrates the point that the concept of stationarity is properly defined only for models and not for data. We do so by claiming that, given any observed time series $\{x_t : t \in T\}$ with $T \stackrel{\text{def}}{=} \{0, 1, \ldots, N-1\}$, we can always construct a stochastic process $\{X_t : t \in T\}$ that is stationary and has $\{x_t\}$ as one of its realizations, with the observed series having a nonzero probability of being selected. To establish this claim, define the ensemble for $\{X_t\}$ to consist of N realizations given by $\{x_t\}$ and all possible circular shifts. For example, if $N = 4$, the ensemble consists of

$$\{x_0, x_1, x_2, x_3\}, \{x_1, x_2, x_3, x_0\}, \{x_2, x_3, x_0, x_1\} \text{ and } \{x_3, x_0, x_1, x_2\}. \tag{45a}$$

If we index the realizations in the ensemble by $k = 0, 1, \ldots, N-1$, we can mathematically describe the kth realization as $\{x_{t+k \bmod N} : t \in T\}$, where, for any integer l, we define $l \bmod N$ as follows. If $0 \le l \le N-1$, then $l \bmod N = l$; otherwise, $l \bmod N = l + mN$, where m is the unique integer such that $0 \le l + mN \le N-1$. For example, when $N = 4$ and $k = 2$,

$$\begin{aligned}\{x_{t+2 \bmod 4} : t = 0, 1, 2, 3\} &= \{x_{2 \bmod 4}, x_{3 \bmod 4}, x_{4 \bmod 4}, x_{5 \bmod 4}\} \\ &= \{x_2, x_3, x_0, x_1\}.\end{aligned}$$

To complete our definition of $\{X_t\}$, we stipulate that the probability of picking any given realization in the ensemble is $1/N$. Thus, if κ is an RV that takes on the values $0, 1, \ldots, N-1$ with equal probability, we can express the stochastic process $\{X_t\}$ as $\{x_{t+\kappa \bmod N}\}$. Show that, for all $s \in T$ and $t \in T$,

(a) $E\{X_t\}$ is a constant that does not depend on t and

(b) $\text{cov}\{X_s, X_t\}$ is a constant that depends upon just the lag $\tau = s - t$;

i.e., $\{X_t\}$ is a stationary process.

[2.20] In the random phase harmonic process $\{X_t\}$ formulated in Exercise [37], we stipulated that the ϕ_l RVs be independent and uniformly distributed over the interval $(-\pi, \pi]$. Suppose that the PDF of ϕ_l is not that of a uniformly distributed RV, but rather is of the form

$$f_{\phi_l}(u) \stackrel{\text{def}}{=} \frac{1}{2\pi}\left(1 + \cos(u)\right), \qquad u \in (-\pi, \pi].$$

A realization of ϕ_l from this PDF is more likely to be close to zero than to $\pm\pi$. With this new stipulation, is $\{X_t\}$ still a stationary process?

[2.21] Let ϕ_1 and ϕ_2 be two RVs whose joint PDF is given by

$$f_{\phi_1,\phi_2}(u, v) = \frac{3}{8\pi^3}\left(|u - v| + |u + v| - \frac{u^2}{\pi} - \frac{v^2}{\pi}\right), \quad \text{where } u, v \in (-\pi, \pi]$$

(Ferguson, 1995).

(a) Show that ϕ_1 is uniformly distributed over $(-\pi, \pi]$; i.e., its univariate (or marginal) PDF is given by

$$f_{\phi_1}(u) = \int_{-\pi}^{\pi} f_{\phi_1,\phi_2}(u, v)\, dv = \frac{1}{2\pi}, \quad u \in (-\pi, \pi].$$

Argue that the PDF $f_{\phi_2}(\cdot)$ for ϕ_2 is also uniform, from which we can conclude that ϕ_1 and ϕ_2 are not independent since $f_{\phi_1}(u) f_{\phi_2}(v) \neq f_{\phi_1,\phi_2}(u, v)$ for all u and v.

(b) Show that, even though they are not independent, the RVs ϕ_1 and ϕ_2 are uncorrelated, i.e., that $\text{cov}\{\phi_1, \phi_2\} = 0$.

(c) Define

$$X_t = D_1 \cos(2\pi f_1 t + \phi_1) + D_2 \cos(2\pi f_2 t + \phi_2), \quad t \in \mathbb{Z}, \tag{45b}$$

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press

where D_1, D_2, f_1 and f_2 are real-valued constants satisfying $D_1 > 0$, $D_2 > 0$ and $0 < f_1 \leq f_2 < 1$. Show that, while $E\{X_t\} = 0$, we have

$$\text{var}\{X_t\} = \frac{D_1^2 + D_2^2}{2} - \frac{3D_1 D_2}{\pi^2} \cos(2\pi f_1 t) \cos(2\pi f_2 t),$$

and hence $\{X_t\}$ is *not* a stationary process.

(d) Suppose that we have a random number generator capable of producing what can be regarded as six independent realizations u_j, $j = 0, 1, \ldots, 5$, from an RV uniformly distributed over the interval $(0, 1)$. The following recipe uses these to create realizations of ϕ_1 and ϕ_2. Let $a = \log(u_0) + \log(u_1)$ and $b = \log(u_2) + \log(u_3)$, and form the ratio $c = a/(a + b)$. Let $d = u_4$. If $u_5 \leq \frac{1}{2}$, let $e = |c - d|$; if $u_5 > \frac{1}{2}$, let $e = 1 - |1 - c - d|$. The realizations of ϕ_1 and ϕ_2 are given by $\pi(2d - 1)$ and $\pi(2e - 1)$. Using this procedure, generate a large number of independent realizations of X_0, X_1, X_2 and X_3 in Equation (45b) with $D_1 = D_2 = 1$, $f_1 = \frac{1}{4}$ and $f_2 = \frac{1}{2}$. Compute the sample means and variances for these realizations, and compare them to the corresponding theoretical values derived in part (c).

[2.22] Suppose that $\{X_t\}$ is a discrete parameter stationary process with zero mean and ACVS $\{s_{X,\tau}\}$, and consider the process defined by $Y_t = X_t \cos(2\pi f_1 t + \phi)$, where $0 < f_1 < 1/2$ is a fixed frequency.

(a) If we regard ϕ as a constant, is $\{Y_t\}$ a stationary process?

(b) If we regard ϕ as an RV that is uniformly distributed over the interval $(-\pi, \pi]$ and is independent of X_t for all t, is $\{Y_t\}$ a stationary process?

[2.23] Suppose that $Y_t = \alpha + \beta t + \gamma t^2 + X_t$, where α, β and γ are nonzero constants and $\{X_t\}$ is a stationary process with ACVS $\{s_{X,\tau}\}$. Show that the first difference of the first difference of Y_t, i.e., $Y_t^{(2)} \stackrel{\text{def}}{=} Y_t^{(1)} - Y_{t-1}^{(1)}$, where $Y_t^{(1)} \stackrel{\text{def}}{=} Y_t - Y_{t-1}$, is a stationary process, and find its ACVS $\{s_{Y^{(2)},\tau}\}$ in terms of $\{s_{X,\tau}\}$.

https://doi.org/10.1017/9781139235723.003 Published online by Cambridge University Press