

Contents

1	Introduction	3
2	Fundamentals of Probability Theory	5
3	Elements of Stationary Processes	7
3.1	The Notion of Stationarity	8
3.1.1	Types of Oscillation	8
3.1.2	Definition of a Stationary Process	11
3.1.3	Complex-Valued Stationary Processes	15
3.1.4	Examples of Stationary Processes	20
3.2	Representation of Stationary Processes	31
3.2.1	A Stochastic Integral	32
3.2.2	Spectral Representation	34
3.3	Decomposition of the Spectral Distribution	36
3.4	Applications of the Spectral Representation	39
3.4.1	Stationary Solutions of Stochastic Difference Equations	39
3.4.2	Effect of a Linear Filter	43
3.4.3	Mean Square Ergodicity	53
3.5	Appendix to Chapter 3	54
3.6	Problems and Complements	56

Chapter 1

Introduction

Chapter 2

Fundamentals of Probability Theory

Chapter 3

Elements of Stationary Processes

Stationary stochastic processes in discrete time provide a natural framework for the development of the theory and applications of higher order crossings. Hence, familiarity with some elementary notions from the theory of stationary processes is essential. This theory will be presented here briefly in its most basic level sufficient for the development in chapters to come, and will be concerned mainly with discrete time processes. All the results carry over to the continuous time case with the appropriate modifications. In this chapter we highlight the basic notions of autocovariance and autocorrelation and their relation to the spectrum. Also, we explain the notion of mixed spectrum and its manifestation in the spectral distribution function and in the process itself. We do not intend to prove much by way of formal proofs. Rather than formal proofs, we will bring home certain useful facts by resorting to discussions and examples. Thus we will explain how to use the spectral representation of a wide sense stationary process but will not prove this celebrated result. For our purposes it is important to know that the spectral representation is as useful as well as convenient operational tool in the study of spectral properties. For example, it can be used in determining the effect of a time invariant linear operation on a process and on its spectrum.

There are many excellent books on stationary processes and their applications in various fields. Here is a scanty description of some of these books. Introductory level treatments with an inclination towards engineering applications can be found in [8], [21]. References that emphasize spectral and correlation analyses and related statistical procedures and applications

are [1],[3],[4], [9],[19], [20], [22], [24], and [30], while more theoretical treatments can be found in [6], and [26].

3.1 The Notion of Stationarity

3.1.1 Types of Oscillation

The physical sciences and the various engineering fields are abundant with examples of what appears to be oscillatory random data in steady-state measured in time. For example, such steady-state oscillation can be observed as a function of time in the velocity of a turbulent flow in a wind tunnel, in any of the components of the earth polar motion, in acoustic signals and in EEG records. When observing such data on an oscilloscope or a time plot, the first thing that strikes the mind is the monotonous repetitiveness or “periodicity” that the data display about a fixed level. Clearly, since there are many different types of mechanisms that generate oscillatory or “periodic” data, some “noisier” than others, from physical considerations alone it is understood that the nature of the observed oscillation cannot possibly be the same in all cases. Thus we are led to the problem of classifying or modeling oscillation observed in time, and proceed to describe a few typical cases.

Consider an infinite sequence $\{\epsilon_t\}$, $t = 1, 2, 3, \dots$, of independent normal random variables each with mean zero and variance σ_ϵ^2 . Such a sequence is referred to as a sequence of *independently and identically distributed* random variables or simply an iid sequence. The probability that ϵ_1 is above 0 is $1/2$, the probability that ϵ_1, ϵ_2 are both above 0 is $1/4$, the probability that $\epsilon_1, \epsilon_2, \epsilon_3$ are all above 0 is $1/8$ and so on. Because the probability decreases exponentially fast, there is an extremely small chance that a long sequence will have all its values above 0. Similarly, there is an equally small chance that a long sequence will have all its value below 0. The infinite random sequence therefore must fluctuate around the 0 level ad infinitum in what we may call completely random oscillation, and will appear to the eye surprisingly “periodic” because of the many passes from positive to negative values and vice versa, an observation made by Slutsky [27], and illustrated in Figure 3.1-a. Observe that since $\{\epsilon_t\}$ is an iid sequence the probability distribution of any subsequence $\epsilon_{t_1}, \epsilon_{t_2}, \dots, \epsilon_{t_n}$ is the same as that of the shifted subsequence $\epsilon_{t_1+\tau}, \epsilon_{t_2+\tau}, \dots, \epsilon_{t_n+\tau}$. This means that in the absence of any external interference, any section of the sequence $\{\epsilon_t\}$ observed in time should appear to the eye quite similar as a quick glance at Figure 3.1-a shows.

In contrast, we consider next a pure random harmonic oscillation. Let A, B be two independent normal random variables each with mean 0 and variance σ^2 , and consider the infinite sequence $\{X_t\}$, $t = 1, 2, 3, \dots$,

$$X_t = A \cos\left(\frac{\pi}{3}t\right) + B \sin\left(\frac{\pi}{3}t\right)$$

When A, B admit values from an appropriate random experiment compatible with the normal assumption, the sequence oscillates about level 0 as a pure nonrandom sinusoid with a period of 6 time units as in Figure 3.1-b. From Figures 3.1-a and 3.1-b it is clear that the two types of oscillation, completely random and sinusoidal, are profoundly different and yet the monotonous steady-state appearance of both as random functions of time points to some common denominator after all.

We can mix the two types of oscillation to create a third. By adding to the pure harmonic sequence the iid random sequence the pure sinusoidal oscillation is altered to produce a hybrid whose graph for $t = 1, 2, \dots, 100$ is shown in Figure 3.1-c. The new sequence of “signal plus noise”,

$$Z_t = A \cos\left(\frac{\pi}{3}t\right) + B \sin\left(\frac{\pi}{3}t\right) + \epsilon_t$$

where A, B are assumed independent of $\{\epsilon_t\}$, will appear more sinusoidal with a period of 6 time units if A, B are large by comparison with ϵ_t , and will appear quite random if the opposite holds. Yet, the monotonous overall oscillation about level 0 still persists. Processes made of the superposition of one or more sinusoids plus “noise” are sometimes called models of hidden periodicities, a term coined last century [25].

It is even possible to construct a fourth type of sequence that oscillates randomly with a *pseudo* period of 6 time units and which is neither a pure sinusoid nor an additive mixture of a sinusoid and an iid sequence. Such a sequence is given by the linear system

$$X_t - .5X_{t-1} + .25X_{t-2} = \epsilon_t$$

driven by the iid sequence ϵ_t , $t = 1, 2, 3, \dots$. A representative plot of $\{X_t\}$, $t = 1, 2, \dots, 100$, is shown in Figure 3.1-d. This random sequence, which consists of *dependent* random variables, has an intrinsic period of 6 time units (obtained by solving the homogeneous difference equation $Y_t - .5Y_{t-1} + .25Y_{t-2} = 0$; the solution is a damped sinusoid with a period of 6 time units [29, p. 35]) and thus displays a monotonous pseudo periodic fluctuation about level 0. The sequence is a special case of the second order stationary autoregressive process introduced by Yule [31] in 1927.

figure=FigThree1a.ps,height=2.2in figure=FigThree1b.ps,height=2.2in
 figure=FigThree1c.ps,height=2.2in figure=FigThree1d.ps,height=2.2in

Figure 3.1: *Realizations from:* (a) ϵ_t , iid $N(0,1)$. (b) $Z_t = 0.975902 \cos(\pi t/3) - 1.34624 \sin(\pi t/3)$. (c) $Z_t = 0.975902 \cos(\pi t/3) - 1.34624 \sin(\pi t/3) + \epsilon_t$. (d) $Z_t = 0.5Z_{t-1} - 0.25Z_{t-2} + \epsilon_t$.

We have discussed four types of oscillation produced by random sequences where the common theme is the monotonous fluctuation about a fixed level. How is it possible to explain or describe the different types of oscillatory behavior illustrated by these examples? A comprehensive answer to this problem is given by the rich theory of stationary stochastic processes, the subject matter of this chapter. From the foregoing discussion we may guess that the displayed monotonous behavior or “sameness” of subsequences and oscillation are connected in some sense, and to a large degree this is indeed the case. In the theory of stationary processes the intuitive “sameness” of subsequences is replaced by the formal stationarity assumption that requires the invariance of joint distributions with respect to time shifts, and as we shall see, the mere notion of stationarity itself leads to an elegant classification of oscillatory phenomena as expressed by the spectral distribution function.

One of the important achievements of the theory of stationary processes is that it is able to put on the same footing different types of models and in particular models of hidden periodicities and pseudo periodic models. Under conditions that guarantee stationarity one theory explains them all. Thus all the four cases discussed above are special cases of stationary processes (both strict and wide sense) that differ in their spectra, and the theory of stationary processes provides the necessary tools for their analysis. The theory tells us that the completely random sequence and the second order autoregressive process possess continuous spectra (in this sense they belong to the same class), the random sinusoid has a discrete spectrum, and the hybrid of a pure random sinusoid plus an independent iid sequence is an example of a mixed spectrum stationary process.

It should be noted that not all steady-state phenomena are stationary and that fluctuations about a fixed level do not guarantee stationarity. However, in many cases it is still possible to stationarize these processes by various tricks such as random time shifts, repeated differencing, or removal of a trend component. The transformation from nonstationary to stationary behavior is useful for estimation purposes, because a great deal is known about estimation in stationary processes. See Example 3.8 and Problem 21.

3.1.2 Definition of a Stationary Process

A stochastic or random process $\{Z_t\}$ is a collection of random variables, real or complex-valued, indexed by t where t takes values in some index set T . Examples of T are $[0, 1]$, $[0, \infty)$, $\{\dots, -2, -1, 0, 1, 2, \dots\}$, and $\{1, 2, 3, \dots\}$. For our purposes it is convenient to think of t as time. When T is an interval, such as $(-\infty, \infty)$, the process is called a continuous time process, and it is called a discrete time process when T is a discrete set such as $\{\dots, -2, -1, 0, 1, 2, \dots\}$. In the sequel we confine ourselves almost exclusively to discrete time processes. But clearly, regardless of the type of the time parameter t , the random variables Z_t may be continuous, discrete, or of mixed type.

Consider first a real-valued process $\{Z_t\}$. As the process evolves or unfolds in time, that is being observed in time, it produces a sequence in discrete time and a function (not necessarily continuous) in continuous time. Both are referred to as *realizations* or *sample functions* or simply as *time records*. Since for each fixed t , Z_t is a random variable, each observation of the whole process at all possible time points results, in most cases of practical interest, almost certainly in a new realization. The collection of all possible realizations is called the *ensemble*. The ensemble enables us to interpret a stochastic process as a collection of functions where randomness is interjected by a “random” experiment whose outcomes are sample functions. From this point of view, for each fixed t , the values of Z_t fall on realizations, thus creating a distribution of all possible values at time t across the ensemble. Shifting between the two interpretations—collection of random variables versus collection of sample functions—is a useful maneuver in the study of stochastic processes.

Sections of infinite realizations are also called realizations or simply *time series*. Thus, Figure 3.1 displays four discrete time finite realizations, or time series, which appear to be continuous because we always connect the resulting time values by line segments.

A remark about notation. In our notation above, $\{Z_t\}$ stands for either discrete or continuous time process. However, the more conventional notation for continuous time processes is $\{Z(t)\}$, while the symbolism $\{Z_t\}$ is reserved for discrete time processes. From now on, with only very few exceptions, we shall comply with the conventional notation.

In the remainder of the chapter we let $T = \{\dots, -2, -1, 0, 1, 2, \dots\}$. In the exceptional cases when reference is made to “continuous time,” we have in mind $T = (-\infty, \infty)$.

For any collection of n time points $t_1 < t_2 < \dots < t_n$, the probability

distribution of the random vector $(Z_{t_1}, Z_{t_2}, \dots, Z_{t_n})$,

$$P(Z_{t_1} \leq z_1, Z_{t_2} \leq z_2, \dots, Z_{t_n} \leq z_n)$$

is called a *finite dimensional (probability) distribution*. In principle, we must specify the finite dimensional distributions in order to describe the stochastic process probabilistically. In practice, however, we are quite content with much less and instead of distributions provide some moment conditions to describe the process.

From the finite dimensional distributions we can obtain probabilities and, when they exist, moments. Suppose $f_{t_1}(z)$ is the probability density function of Z_{t_1} . Then the mean (first moment) of Z_{t_1} is given by

$$E[Z_{t_1}] = \int_{-\infty}^{\infty} z f_{t_1}(z) dz$$

This may be interpreted as an “average across the ensemble” at time t_1 . That is, the average of all the possible values which Z_{t_1} can assume across the ensemble. In general, the mean may be different at different time points, and if we define $m(t) \equiv E[Z_t]$, the result is a function of time called the *mean function*. The mean function may assume any form, but $E[Z_t - m(t)] = 0$ always.

Suppose $f_{t_1, t_2}(z_1, z_2)$ is the joint probability density function of (Z_{t_1}, Z_{t_2}) . Then a second order moment is given by the expectation

$$E[Z_{t_1} Z_{t_2}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z_1 z_2 f_{t_1, t_2}(z_1, z_2) dz_1 dz_2$$

This may be interpreted as the average of the values which the product $Z_{t_1} Z_{t_2}$ assumes across the ensemble. Recall that the covariance $Cov[Z_{t_1}, Z_{t_2}]$ of Z_{t_1} and Z_{t_2} is given by

$$Cov[Z_{t_1}, Z_{t_2}] = E\{[Z_{t_1} - E(Z_{t_1})][Z_{t_2} - E(Z_{t_2})]\} = E[Z_{t_1} Z_{t_2}] - E[Z_{t_1}]E[Z_{t_2}]$$

Clearly, $Cov[Z_{t_1}, Z_{t_1}] = Var[Z_{t_1}]$. The function $R(s, t) \equiv Cov[Z_s, Z_t]$ is called the *covariance* or *autocovariance function*.

The *correlation* or *autocorrelation function*, $\rho(s, t)$, is defined from the normalized covariance function,

$$\rho(s, t) \equiv \frac{R(s, t)}{\sqrt{R(s, s)}\sqrt{R(t, t)}}$$

and satisfies $\rho(s, t) = \rho(t, s)$, and $|\rho(s, t)| \leq \rho(t, t) = 1$.

In the same way we can obtain higher order moments and moments of functions of random variables from the process. For example, let

$$f_{t_1, t_2, \dots, t_n}(z_1, z_2, \dots, z_n)$$

be the n -dimensional probability density of the random vector from the process $(Z_{t_1}, Z_{t_2}, \dots, Z_{t_n})$. Then,

$$E[\varphi(Z_{t_1}, Z_{t_2}, \dots, Z_{t_n})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(z_1, z_2, \dots, z_n) f_{t_1, t_2, \dots, t_n}(z_1, z_2, \dots, z_n) dz_1 dz_2 \dots dz_n$$

provided the expectation exists. Note that in all the expressions involving integrals it is assumed that the random variables are continuous (nothing to do with continuous or discrete time!). When the random variables are discrete, the integrals are replaced by summations.

An important example of a stochastic process is a *Gaussian* or *normal* process. A real-valued process $\{Z_t\}$, $t \in T$, is called Gaussian process if for all $t_1, t_2, \dots, t_n \in T$, the joint distribution of $(Z_{t_1}, Z_{t_2}, \dots, Z_{t_n})$ is multivariate normal. Since a multivariate normal is completely specified by the vector of means and the covariance matrix, the finite dimensional distributions are completely determined from the mean function $m(t) = E[Z_t]$ and the covariance function $R(s, t) = Cov[Z_s, Z_t]$. Thus, for a Gaussian process we need to supply only first and second order moment information in order to obtain the finite dimensional distributions, a fact which makes Gaussian processes easier to handle than many other types of processes.

A stochastic process $\{Z_t\}$ is said to be a *strictly stationary* process if the joint distribution of $(Z_{t_1}, Z_{t_2}, \dots, Z_{t_n})$ is the same as the joint distribution of $(Z_{t_1+\tau}, Z_{t_2+\tau}, \dots, Z_{t_n+\tau})$ for all t_1, t_2, \dots, t_n , n , and τ . In other words, as long as the relative distances between the time points are fixed, the joint distribution does not change.

Strict stationarity implies that Z_t and $Z_{t+\tau}$ have the same distribution for all t and τ . Hence, if first order moments exist, put $\tau = -t$ to see that,

$$E[Z_t] = E[Z_{t+\tau}] = E[Z_0] = m \quad (3.1)$$

where m is a constant. Likewise, stationarity implies that (Z_t, Z_s) and $(Z_{t+\tau}, Z_{s+\tau})$ have the same distribution for all t, s , and τ , and in particular for $\tau = -s$. Therefore, if second order moments exist,

$$E[Z_t Z_s] = E[Z_{t+\tau} Z_{s+\tau}] = E[Z_{t-s} Z_0]$$

or,

$$R(t, s) = \text{Cov}[Z_t, Z_s] = R(t - s)$$

where we use the same symbol R in an obvious way. Thus, the autocovariance function of a real valued stationary process is a function of the time lag τ only,

$$R(\tau) = \text{Cov}[Z_{t+\tau}, Z_t] = \text{Cov}[Z_\tau, Z_0] \quad (3.2)$$

Clearly, $\text{Var}[Z_t] = R(0)$, and $R(\tau) = R(-\tau)$. Likewise, the autocorrelation is a function of τ only,

$$\rho(\tau) = \frac{R(\tau)}{R(0)} \quad (3.3)$$

The autocorrelation $\rho(\tau)$ measures the correlation between $Z_{t+\tau}$ and Z_t as a function of the indices difference independently of t . Obviously, $\rho(\tau) = \rho(-\tau)$, and $|\rho(\tau)| \leq \rho(0) = 1$.

The notation $R(\tau), \rho(\tau)$ is usually reserved for continuous time. We shall mostly use the more suggestive notation R_k, ρ_k , $k = \dots, -2, -1, 0, 1, 2, \dots$, when dealing with discrete time stationary processes. With this convention, R_0 is the variance and ρ_1 is the first order autocorrelation.

In general, strict stationarity and the existence of higher order moments imply,

$$E[Z_t Z_{t+t_1} Z_{t+t_2}] = E[Z_0 Z_{t_1} Z_{t_2}]$$

which is a function of t_1 and t_2 only, and

$$E[Z_t Z_{t+t_1} Z_{t+t_2} Z_{t+t_3}] = E[Z_0 Z_{t_1} Z_{t_2} Z_{t_3}]$$

which is a function of t_1, t_2 and t_3 , and so on. It is worth noting that in recent years there is a growing interest in higher order moments and functions thereof called cumulants [3], [24]. For example, assuming that $E[Z_t] = 0$, the fourth-order cumulant function, $\kappa(t_1, t_2, t_3)$, is given by

$$\kappa(t_1, t_2, t_3) = E[Z_0 Z_{t_1} Z_{t_2} Z_{t_3}] - \{R_{t_1} R_{t_2-t_3} + R_{t_2} R_{t_1-t_3} + R_{t_3} R_{t_1-t_2}\}$$

and comes into play in the statistical estimation of the autocovariance. If the process is Gaussian, $\kappa(t_1, t_2, t_3) \equiv 0$. Hence, for a zero mean stationary Gaussian process,

$$E[Z_0 Z_{t_1} Z_{t_2} Z_{t_3}] = R_{t_1} R_{t_2-t_3} + R_{t_2} R_{t_1-t_3} + R_{t_3} R_{t_1-t_2}$$

A great deal of the theory of stationary processes only requires the fulfillment of the conditions (3.1) and (3.2). In general, a process which satisfies (3.1) and (3.2) is called *weakly stationary* or *stationary in the wide sense* or sometimes is said to be *second order stationary*. A strictly stationary process need not be weakly stationary unless second order moments exist, and conversely, since weak stationarity is less restrictive it does not imply strict stationarity. *However, for the important special case of a Gaussian process, strict and wide sense stationarity coincide* (see Problem 1).

In this book, in order to avoid confusion, *stationary* is taken to mean strict stationarity together with the assumption that second order moments exist. By this convention, a stationary process is stationary in both wide and strict senses. Since the Gaussian assumption plays an important role in the coming chapters, the convention becomes quite palatable. However, the appropriate adjective “strict” or “weak” (or “wide sense”) will be added explicitly whenever needed in emphasizing a particular property associated with one of the stationarity senses.

3.1.3 Complex-Valued Stationary Processes

So far we spoke of real-valued processes, however, there is nothing special about real processes, and the theory can easily be extended to cover complex-valued processes as well. Complex processes may arise in several ways one of which is through filters with a complex impulse response function applied to a real process. In this case, the original process is real but the filtered process is complex. Clearly, every result concerning the complex case automatically holds for real processes as well, and in what follows we find it convenient to introduce new concepts in terms of complex-valued stationary processes.

A complex random variable X has the general form $X = U + iV$ where U, V are real random variables. We have, the complex conjugate $\bar{X} = U - iV$, $|X|^2 = X\bar{X} = U^2 + V^2$, $E[X] \equiv E[U] + iE[V]$, $E[|X|^2] = E[U^2] + E[V^2]$. For two complex-valued random variables X, Y , the covariance is defined by,

$$Cov(X, Y) \equiv E\{[X - E(X)][\overline{Y - E(Y)}]\} = E[X\bar{Y}] - E[X]E[\bar{Y}]$$

and $Var[X] = Cov[X, X] = E[|X - E(X)|^2]$. When $E(X) = E(Y) = 0$, then $Cov(X, Y) = E[X\bar{Y}]$, and when $Cov(X, Y) = 0$, we say that X and Y are uncorrelated.

Similarly, we define a complex-valued stochastic process $\{Z_t\}$ from two real-valued processes $\{U_t\}, \{V_t\}$,

$$Z_t = U_t + iV_t$$

The complex process is described probabilistically from the finite dimensional distributions of the vectors, $(U_{t_1}, U_{t_2}, \dots, U_{t_n}, V_{t_1}, V_{t_2}, \dots, V_{t_n})$, for all possible t_1, t_2, \dots, t_n , and n , and is said to be strictly stationary if the finite dimensional distributions remain unaltered under time shifts, just as in the real case. When $\{Z_t\}$ is strictly stationary, so are the real-valued processes $\{U_t\}$ and $\{V_t\}$ (why?).

Wide or weak sense stationarity is defined as in the real case with the appropriate modifications. Thus we say that $\{Z_t\}$ is weakly stationary if the mean is a (complex) constant m , say,

$$E[Z_t] = E[U_t] + iE[V(t)] = m$$

and the autocovariance is a function of the time lag,

$$R_k = E[Z_{t+k} - m][\overline{Z_t - m}] = E[Z_{t+k}\overline{Z_t}] - |m|^2$$

Observe that the *variance* $R_0 = E[|Z_t - m|^2]$ is always real-valued. As in the real case, the autocorrelation function is obtained as the quotient (assuming of course $R_0 \neq 0$),

$$\rho_k = \frac{R_k}{R_0}$$

We have,

$$R_{-k} = \overline{R_k}, \quad |R_k| \leq R_0 \quad (3.4)$$

and

$$\rho_{-k} = \overline{\rho_k}, \quad |\rho_k| \leq 1$$

Observe that for all complex numbers a_1, a_2, \dots, a_N , and integers t_1, t_2, \dots, t_N , with $N \geq 1$,

$$\begin{aligned} 0 \leq \text{Var}\left[\sum_{j=1}^N a_j Z_{t_j}\right] &= \text{Cov}\left[\sum_{j=1}^N a_j Z_{t_j}, \sum_{l=1}^N a_l Z_{t_l}\right] = \\ &= \sum_{j=1}^N \sum_{l=1}^N a_j \overline{a_l} \text{Cov}[Z_{t_j}, Z_{t_l}] = \sum_{j=1}^N \sum_{l=1}^N a_j \overline{a_l} R_{t_j - t_l} \end{aligned} \quad (3.5)$$

That is, R_k is *nonnegative definite*. The important implication of this ostensibly simple fact is that R_k admits a Fourier or spectral representation in terms of a uniquely defined bounded monotone nondecreasing real function F such that $F(-\pi) = 0$,

$$R_k = \int_{-\pi}^{\pi} e^{ik\lambda} dF(\lambda), \quad k = 0, \pm 1, \pm 2, \dots \quad (3.6)$$

This fact is due to the German mathematician G. Herglotz (1911) [9, p. 34], and $F(\lambda)$ is called *spectral distribution function*. Conversely, it is easy to see that a complex-valued sequence R_k which admits the Fourier representation (3.6) is nonnegative definite (see Problem 7). Note that the Hermitian property (3.4) of the autocovariance implies that $F(\omega)$ is real-valued.

An alternative representation of (3.6), proved in the appendix to this chapter, is as an integral¹ with respect to a unique (see appendix) *spectral distribution* (measure) $F(\Lambda)$ defined over (measurable) subsets Λ of $(-\pi, \pi]$,

$$R_k = \int_{-\pi}^{\pi} e^{ik\lambda} F(d\lambda), \quad k = 0, \pm 1, \pm 2, \dots \quad (3.7)$$

where the relationship between the spectral distribution function and the spectral distribution is,

$$F(\lambda) = F((-\pi, \lambda])$$

with the understanding that we use the same symbol to denote two different, yet closely related, entities one of which is a function and the other a set function. We sometimes refer to $F(\lambda)$ or $F(\Lambda)$ as “spectrum”. Both give complete information as to the distribution of power over frequency, but it is a little easier to interpret $F(\Lambda)$. The spectral measure $F(\Lambda)$ may be interpreted as the amount of power or variance associated with the spectral band Λ , where Λ is a subset of $(-\pi, \pi]$. The total power then is: $R_0 = F(\pi) = F(-\pi, \pi]$.

By extending the definition of $F(\lambda)$ such that $F(\lambda) = 0$ for $\lambda \leq -\pi$, and $F(\lambda) = F(\pi)$ for $\lambda \geq \pi$, $F(\lambda)$ becomes, except for a normalization by a constant, a probability distribution function. *Thus, in summary, the autocovariance of every weakly stationary process in discrete time admits a Fourier representation in terms of a uniquely defined distribution function which, up to a constant, is a probability distribution function supported on $(-\pi, \pi]$.*

As is the case with probability distributions, a spectral distribution function may have a *spectral density* $f(\lambda)$ such that

$$F(\lambda) = \int_{-\pi}^{\lambda} f(\omega) d\omega, \quad f(\lambda) = F'(\lambda), \quad -\pi \leq \lambda \leq \pi$$

A sufficient condition for this is the absolute summability of the autocovari-

¹Both (3.6) and (3.7) are called Lebesgue-Stieltjes integrals. In the symbolism $F(d\lambda)$, $\xi(d\lambda)$, etc., it is helpful to think of $d\lambda$ as a very small interval containing λ .

ance, $\sum_{k=-\infty}^{\infty} |R_k| \leq \infty$, in which case we have the inverse transform,

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-i\lambda k} R_k \quad (3.8)$$

When a spectral density exists, $F(d\lambda) = f(\lambda)d\lambda$, and we can express (3.7) as,

$$R_k = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda \quad (3.9)$$

As an immediate consequence,

$$\text{Var}[Z_t] = R_0 = \int_{-\pi}^{\pi} f(\lambda) d\lambda \quad (3.10)$$

Thus, the spectral density describes how the variance—that is power—is distributed over frequency. Since $F(\lambda)$ is monotone nondecreasing, and since $f(\lambda) = F'(\lambda)$, $f(\lambda) \geq 0$, $-\pi \leq \lambda \leq \pi$.

Everything we said about discrete time extends naturally to continuous time as well. The continuous time version of (3.6) is referred to as the *Wiener-Khintchine relationship* or *Wiener-Khintchine theorem* [30, p. 93],

$$R(\tau) = \int_{-\infty}^{\infty} e^{i\tau\lambda} dF(\lambda) \quad (3.11)$$

or in terms of the spectral distribution,

$$R(\tau) = \int_{-\infty}^{\infty} e^{i\tau\lambda} F(d\lambda) \quad (3.12)$$

with the understanding that,

$$F(\lambda) = F((-\infty, \lambda])$$

When $R(\tau)$ tends to 0 as $|\tau|$ increases such that $\int_{-\infty}^{\infty} |R(\tau)| d\tau < \infty$, then a spectral density $f(\lambda)$ exists and the analog of (3.9) becomes,

$$R(\tau) = \int_{-\infty}^{\infty} e^{ik\lambda} f(\lambda) d\lambda \quad (3.13)$$

It is clear that, except for a normalizing constant, the autocorrelation is related to the spectral distribution in the same way as does the autocovariance. In discrete time this takes the form,

$$\rho_k = \int_{-\pi}^{\pi} e^{ik\lambda} \bar{F}(d\lambda) \quad (3.14)$$

where $\bar{F}(d\lambda) \equiv F(d\lambda)/R_0$ is the normalized spectral distribution. In the sequel we shall refer to any of the relationships between the autocovariance or the autocorrelation and the spectral distribution somewhat loosely as the Wiener-Khintchine relationship, without reference to discrete or continuous time. Note that this is a relationship between *time domain* and *spectral domain* functions. Today we know that the relationship between the spectrum and the autocorrelation also was understood by A. Einstein early in this century. See the reference to and a summary of Einstein's 1914 work on "fluctuating observations" in Notes 13a and 41 of [30, pp. 32, 88, Vol II].

The Real Case

In the real case, $R_k = R_{-k}$, and hence (3.7) reduces to,

$$R_k = \int_{-\pi}^{\pi} \cos(k\lambda) F(d\lambda)$$

and similarly in continuous time $R(\tau) = R(-\tau)$ so that (3.12) becomes,

$$R(\tau) = \int_{-\infty}^{\infty} \cos(\tau\lambda) F(d\lambda)$$

When a spectral density exists, by the symmetry of the autocovariance it is symmetric, $f(\lambda) = f(-\lambda)$, and in the preceding two equations we can replace $F(d\lambda)$ by $f(\lambda)d\lambda$:

$$R_k = \int_{-\pi}^{\pi} \cos(k\lambda) f(\lambda) d\lambda$$

and similarly for continuous time,

$$R(\tau) = \int_{-\infty}^{\infty} \cos(\tau\lambda) f(\lambda) d\lambda$$

3.1.4 Examples of Stationary Processes

Example 3.1: An iid Sequence.

Suppose $\{Z_t\}$ constitutes a sequence of independently and identically distributed complex-valued random variables. Then $\{Z_t\}$ is strictly stationary. If in addition $E[Z_t] = m$, and $Var[Z_t] = \sigma^2 < \infty$, then,

$$R_k = \begin{cases} \sigma^2, & \text{for } k = 0 \\ 0, & \text{for } k = \pm 1, \pm 2, \pm 3, \dots \end{cases}$$

and the process is also wide sense stationary. In this case we can write,

$$R_k = \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} e^{ik\lambda} d\lambda$$

and this defines uniquely the spectral distribution function as,

$$F(\lambda) = \frac{\sigma^2}{2\pi}(\lambda + \pi)$$

which can also be expressed in terms of the spectral density $f(\lambda) = \sigma^2/2\pi$,

$$F(\lambda) = \int_{-\pi}^{\lambda} f(\omega) d\omega = \frac{\sigma^2}{2\pi} \int_{-\pi}^{\lambda} d\omega, \quad -\pi \leq \lambda \leq \pi$$

Note that we also have,

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\lambda} R_k = \dots + 0 + 0 + \frac{\sigma^2}{2\pi} + 0 + 0 + \dots$$

If we extend the definition of $F(\lambda)$ such that $F(\lambda) = 0$, $\lambda \leq -\pi$, and $F(\lambda) = \sigma^2$, $\lambda \geq \pi$, then, except for a normalizing constant, $F(\lambda)$ behaves like a continuous probability distribution function with a density supported on $[-\pi, \pi]$.

We observe, in particular, that any sequence of uncorrelated real- or complex-valued random variables $\{u_t\}$ (this is a much weaker assumption than iid) with mean 0 and variance σ^2 ,

$$R_u(k) \equiv E[u_t \overline{u_{t-k}}] = \begin{cases} \sigma^2, & \text{for } k = 0 \\ 0, & \text{for } k = \pm 1, \pm 2, \pm 3, \dots \end{cases}$$

has the same autocovariance as does the preceding iid sequence. Therefore such a sequence is weakly stationary with a flat spectral density $f(\lambda) =$

$\sigma^2/2\pi$, $-\pi \leq \lambda \leq \pi$. For this reason a sequence of uncorrelated random variables with constant mean and variance is termed *white noise* by analogy with white light which consists of waves with the same power regardless of frequency. As a rule, we always take the mean of white noise to be 0.

♠

Example 3.2: Pure Random Sinusoid.

Let A, B be uncorrelated real-valued random variables with mean 0 and variance σ^2 , and define a real-valued process by,

$$Z_t = A \cos(\omega_1 t) + B \sin(\omega_1 t), \quad t = 0, \pm 1, \pm 2, \dots \quad (3.15)$$

where $\omega_1 \in (0, \pi]$ is a constant frequency. Then since $E[A] = E[B] = 0$, $E[A^2] = E[B^2] = \sigma^2$, and $E[AB] = 0$, and remembering the formula for $\cos(\alpha - \beta)$,

$$\begin{aligned} R_k &= \text{Cov}[Z_{t+k}, Z_t] = E[Z_{t+k} Z_t] \\ &= E\{[A \cos \omega_1(t+k) + B \sin \omega_1(t+k)][A \cos(\omega_1 t) + B \sin(\omega_1 t)]\} \\ &= E\{[A \cos \omega_1(t+k)][A \cos(\omega_1 t)]\} + E\{[B \sin \omega_1(t+k)][B \sin(\omega_1 t)]\} \\ &= \sigma^2[\cos \omega_1(t+k) \cos(\omega_1 t) + \sin \omega_1(t+k) \sin(\omega_1 t)] \\ &= \sigma^2 \cos(\omega_1 k), \quad k = 0, \pm 1, \pm 2, \dots \end{aligned} \quad (3.16)$$

Since the autocovariance depends on the lag k only, the process is weakly stationary. Also,

$$\rho_k = \frac{R_k}{R_0} = \cos(\omega_1 k), \quad k = 0, \pm 1, \pm 2, \dots \quad (3.17)$$

Hence both the autocovariance and autocorrelation are periodic with the same period $2\pi/\omega_1$ as the process itself. However, whereas the process is random, the autocovariance and autocorrelation are nonrandom (deterministic).

Define a function $F(\lambda)$ by,

$$F(\lambda) \equiv \begin{cases} 0, & -\pi \leq \lambda < -\omega_1 \\ \frac{\sigma^2}{2}, & -\omega_1 \leq \lambda < \omega_1 \\ \sigma^2, & \omega_1 \leq \lambda \leq \pi \end{cases} \quad (3.18)$$

Then $F(\lambda)$ is a right-continuous step function with two jumps of size $\frac{1}{2}\sigma^2$ at $\pm\omega_1$, such that $F(-\pi) = 0$, $F(\pi) = \sigma^2$, and we see that

$$\int_{-\pi}^{\pi} \cos(k\lambda) dF(\lambda) = \frac{\sigma^2}{2} \cos(-k\omega_1) + \frac{\sigma^2}{2} \cos(k\omega_1) = \sigma^2 \cos(\omega_1 k) = R_k$$

Therefore, by uniqueness, $F(\lambda)$ must be the spectral distribution function of $\{Z_t\}$.

So far we have not made use of any distributional assumption regarding the amplitudes except for the requirement that they are uncorrelated with mean 0 and variance σ^2 . Assume now in addition that A, B are (jointly) normally distributed. Then, since A, B are uncorrelated, the additional normal assumption makes them independent $\mathcal{N}(0, \sigma^2)$ random variables. Consequently, linear combinations in the Z_t 's are normal and hence the finite dimensional distributions are multivariate normal which makes the process Gaussian. Therefore, the process is also strictly stationary (see Problem 1).

Now, write $A = R \cos \Theta$, $B = R \sin \Theta$ (note the difference between R and the autocovariance). Then

$$\begin{aligned} Z_t &= R \cos \Theta \cos(\omega_1 t) + R \sin \Theta \sin(\omega_1 t) \\ &= R \cos(\omega_1 t - \Theta) \end{aligned}$$

To investigate the distribution of the phase and amplitude in the new representation under the assumption of normal amplitudes, notice that $Y \equiv R^2 = A^2 + B^2$, and $\Theta = \tan^{-1}(B/A)$, and that the Jacobian of the transformation is $J = \frac{1}{2} \cos^2 \theta + \frac{1}{2} \sin^2 \theta = 1/2$. Then the joint probability density of (R^2, Θ) is

$$g(y, \theta) = \frac{1}{4\pi\sigma^2} \exp \left\{ -\frac{y}{2\sigma^2} \right\}$$

Hence, Θ is uniformly distributed in $(0, 2\pi)$ independently of $Y = R^2$ which is distributed as $\sigma^2 \chi_{(2)}^2$. That is, R^2 is proportional to a chi square random variable with 2 degrees of freedom. This in turn implies that the distribution of $R = \sqrt{Y}$ is Rayleigh with probability density,

$$g_R(r) = \begin{cases} \frac{r}{\sigma^2} \exp \left\{ -\frac{r^2}{2\sigma^2} \right\}, & r \geq 0 \\ 0, & r < 0 \end{cases}$$

From this discussion we can see that, *in a sinusoidal model, the assumption of a uniform phase independently distributed of a random amplitude is quite plausible*. In fact, in applications the use of the random phase sinusoidal model

$$Z_t = \beta \cos(\omega_1 t + \phi)$$

where ϕ is uniformly distributed in $(0, 2\pi)$, and β is a *constant*, is widespread. This process is not Gaussian but it is still strictly stationary since a time shift only gives a new phase which is again uniformly distributed over an interval of length 2π (see Problem 4). In the present representation, the only source

of randomness is the uniformly distributed phase with a constant probability density $1/2\pi$ supported over $[0, 2\pi]$, and hence for every t ,

$$E[Z_t] = \frac{\beta}{2\pi} \int_0^{2\pi} \cos(\omega_1 t + \phi) d\phi = 0$$

and the autocovariance is given by,

$$\begin{aligned} R_k = E[Z_t Z_{t-k}] &= \frac{\beta^2}{2\pi} \int_0^{2\pi} \cos(\omega_1 t + \phi) \cos(\omega_1 t - \omega_1 k + \phi) d\phi \\ &= \frac{\beta^2}{4\pi} \int_0^{2\pi} [\cos(2\omega_1 t - \omega_1 k + 2\phi) + \cos(\omega_1 k)] d\phi \\ &= \frac{\beta^2}{2} \cos(\omega_1 k), \quad k = 0, \pm 1, \pm 2, \dots \end{aligned} \quad (3.19)$$

Thus the autocovariance also is a sinusoid whose amplitude depends on the fixed amplitude of the process, while in (3.16) it depends on the variance of the process the latter having a random amplitude. On the other hand, the autocorrelation is identical to (3.17),

$$\rho_k = \frac{R_k}{R_0} = \cos(\omega_1 k), \quad k = 0, \pm 1, \pm 2, \dots \quad (3.20)$$

By employing the same reasoning as in the random amplitude case, we find that $F(\lambda)$ is again a step function with two jumps of size $\beta^2/4$ at $\pm\omega_1$, such that $F(-\pi) = 0$, $F(\pi) = \beta^2/2$. Thus, the spectral distribution functions in the fixed and random amplitude models are of the same type, and when $\beta = \sqrt{2}\sigma$, they are identical. This shows that *two different processes may still have the same spectral distribution*.

Notice that F is very reminiscent of a discrete probability distribution supported at $\pm\omega_1$. For F to be a genuine discrete probability distribution function it has to be extended to the whole real line as in the previous example and normalized such that $F(\pi) = 1$. ♠

Example 3.3: Sum of Random Sinusoids.

The development in Example 3.2 can be easily extended to the case of a sum of random sinusoids. First consider random amplitudes. Let

$$Z_t = \sum_{j=1}^p \{A_j \cos(\omega_j t) + B_j \sin(\omega_j t)\}, \quad t = 0, \pm 1, \pm 2, \dots \quad (3.21)$$

where $A_1, \dots, A_p, B_1, \dots, B_p$ are (pairwise) uncorrelated real-valued random variables with mean 0, and $\text{Var}[A_j] = \text{Var}[B_j] = \sigma_j^2$ for all j . The frequencies ω_j are ordered real constants in $(0, \pi]$, $0 < \omega_1 < \omega_2 < \dots < \omega_p \leq \pi$. Then for all t , $E[Z_t] = 0$, and

$$\begin{aligned} R_k &= E[Z_t Z_{t-k}] = \sum_{j=1}^p \sigma_j^2 \cos(\omega_j k) \\ &= \sum_{j=1}^p \left\{ \frac{1}{2} \sigma_j^2 \cos(\omega_j k) + \frac{1}{2} \sigma_j^2 \cos(-\omega_j k) \right\}, \quad k = 0, \pm 1, \pm 2, \dots \end{aligned}$$

Therefore the process is weakly stationary. The autocorrelation is

$$\rho_k = \frac{R_k}{R_0} = \frac{\sum_{j=1}^p \sigma_j^2 \cos(\omega_j k)}{\sum_{j=1}^p \sigma_j^2}, \quad k = 0, \pm 1, \pm 2, \dots \quad (3.22)$$

By inspection, the spectral distribution function $F(\omega)$ is a nondecreasing step function with jumps of size $\frac{1}{2}\sigma_j^2$ at $\pm\omega_j$, and $F(-\pi) = 0, F(\pi) = R_0 = \sum_{j=1}^p \sigma_j^2$. The thing to remember is that the power, or intensity, at ω_j in (3.21) is proportional to σ_j^2 rather than the size of the amplitudes A_j, B_j .

As in the previous example, when in addition the A 's and B 's are (jointly) normally distributed, the process is Gaussian and thus also strictly stationary. Furthermore, the process can be expressed as a sum of cosines with independent random phases uniformly distributed in $(0, 2\pi)$, and independent of the amplitudes which are independent Rayleigh random variables.

With the frequencies as above, a model of random phases ϕ_j , uniformly and independently distributed in $(0, 2\pi)$, and fixed (nonrandom) amplitudes β_j , is given by

$$Z_t = \sum_{j=1}^p \beta_j \cos(\omega_j t + \phi_j), \quad t = 0, \pm 1, \pm 2, \dots \quad (3.23)$$

The process is stationary with mean 0 and autocovariance,

$$R_k = E[Z_t Z_{t-k}] = \sum_{j=1}^p \frac{\beta_j^2}{2} \cos(\omega_j k), \quad k = 0, \pm 1, \pm 2, \dots$$

The autocorrelation function is

$$\rho_k = \frac{R_k}{R_0} = \frac{\sum_{j=1}^p \beta_j^2 \cos(\omega_j k)}{\sum_{j=1}^p \beta_j^2}, \quad k = 0, \pm 1, \pm 2, \dots \quad (3.24)$$

The spectral distribution function $F(\omega)$ is a nondecreasing step function with jumps of size $\frac{1}{4}\beta_j^2$ at $\pm\omega_j$, and $F(\pi) = R_0 = \sum_{j=1}^p \frac{1}{2}\beta_j^2$. Thus, in the model (3.23) the power at ω_j is proportional to β_j^2 , the square of the amplitude.



Example 3.4: Sum of Random Complex Exponentials.

The complex version of the models (3.21), (3.23), is the almost periodic sequence

$$X_t = \sum_{j=-\infty}^{\infty} \xi_j e^{i\omega_j t}, \quad t = 0, \pm 1, \pm 2, \dots \quad (3.25)$$

where the ω_j are in $(-\pi, \pi]$, and where the ξ_j are zero mean (pairwise) uncorrelated, or *orthogonal*, complex-valued random variables, $E[\xi_j \bar{\xi}_k] = 0$, $j \neq k$, and $E[|\xi_j|^2] = \sigma_j^2$. The sum (3.25) converges in mean square when $\sum_{j=-\infty}^{\infty} \sigma_j^2 < \infty$, and can be made real by appropriate conditions on the frequencies and amplitudes (see Problem 6). By orthogonality we have,

$$\begin{aligned} R_k = E[X_t \bar{X}_{t-k}] &= E\left[\sum_j \xi_j e^{i\omega_j t} \overline{\sum_n \xi_n e^{i\omega_n(t-k)}}\right] \\ &= \sum_j \sum_n e^{i[\omega_j t - \omega_n(t-k)]} E[\xi_j \bar{\xi}_n] \\ &= \sum_j \sigma_j^2 e^{i\omega_j k}, \quad k = 0, \pm 1, \pm 2, \dots \end{aligned} \quad (3.26)$$

Therefore, the process is weakly stationary. As in the real case the spectral distribution function is a step function,

$$F(\omega) = \sum_{\omega_j \leq \omega} \sigma_j^2 \quad (3.27)$$

where the summation is over all j such that $\omega_j \leq \omega$. Indeed this gives

$$R_k = \int_{-\pi}^{\pi} e^{ik\omega} dF(\omega) = \int_{-\pi}^{\pi} e^{ik\omega} F(d\omega)$$

Notice that in the complex case the jump in $F(\omega)$ at ω_j is of size σ_j^2 while in the real case—due to symmetry—it is of size $\frac{1}{2}\sigma_j^2$. ♠

Example 3.5: Signal Plus Noise.

Let $\{X_t\}$ be a “signal” as in (3.25), and let $\{\epsilon_t\}$ be a zero mean wide sense stationary “noise” uncorrelated with $\{X_t\}$ and with a spectrum which possesses a spectral density $f_\epsilon(\omega)$,

$$F_\epsilon(\omega) = \int_{-\pi}^{\omega} f_\epsilon(\lambda) d\lambda$$

Similarly, use $F_x(\omega)$ to denote (3.27). Define a signal plus noise process $\{Z_t\}$ as,

$$Z_t = X_t + \epsilon_t = \sum_{j=-\infty}^{\infty} \xi_j e^{i\omega_j t} + \epsilon_t, \quad t = 0, \pm 1, \pm 2, \dots \quad (3.28)$$

Then with obvious notation, since the signal and noise are uncorrelated,

$$R_k = E[Z_t \bar{Z}_{t-k}] = E[X_t \bar{X}_{t-k}] + E[\epsilon_t \bar{\epsilon}_{t-k}] \equiv R_k^x + R_k^\epsilon$$

We can see that if we define

$$F(\omega) \equiv F_x(\omega) + F_\epsilon(\omega) = \sum_{\omega_j \leq \omega} \sigma_j^2 + \int_{-\pi}^{\omega} f_\epsilon(\lambda) d\lambda \quad (3.29)$$

then $F(\omega)$ must be the spectral distribution function of $\{Z_t\}$. This is an illustration the general fact that *the spectrum of the sum of uncorrelated stationary processes is the sum of the individual spectra.*

Apart from a constant, by extending its definition to the whole real line, we recognize $F(\omega)$ as the distribution function of a mixed probability distribution whose atoms (points of jump) are the frequencies ω_j . Indeed this is an example of a *mixed spectrum*. In the jargon of stationary processes, the spectra of $\{X_t\}$, $\{\epsilon_t\}$, and $\{Z_t\}$ are *discrete*, *continuous*, and *mixed*, respectively.

As for the autocorrelation, it can be written as a convex combination of the autocorrelations of the signal, ρ_k^x , and that of the noise, ρ_k^ϵ ,

$$\rho_k = \frac{R_0^x}{R_0^x + R_0^\epsilon} \rho_k^x + \frac{R_0^\epsilon}{R_0^x + R_0^\epsilon} \rho_k^\epsilon \quad (3.30)$$

As $R_0^x \rightarrow \infty$, $\rho_k \rightarrow \rho_k^x$.

**Example 3.6: The Stationary AR(1) Process.**

Let $\{\epsilon_t\}$, $t = 0, \pm 1, \pm 2, \dots$, be a sequence of uncorrelated real-valued random variables with mean zero and variance σ_ϵ^2 (i.e. white noise) and define a real-valued weakly stationary process $\{Z_t\}$ by the stochastic difference equation

$$Z_t = \phi_1 Z_{t-1} + \epsilon_t, \quad t = 0, \pm 1, \pm 2, \dots \quad (3.31)$$

where $|\phi_1| < 1$. The process (3.31) is called a *first order autoregressive process* and is commonly denoted by $AR(1)$. By repeated substitution we obtain,

$$Z_t = \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \dots + \phi_1^{p-1} \epsilon_{t-(p-1)} + \phi_1^p Z_{t-p}$$

Observe now that since the process is weakly stationary, $E[Z_t^2]$ is equal to a finite constant uniformly in t . This, coupled with the fact that $|\phi_1| < 1$, gives as $p \rightarrow \infty$,

$$E[Z_t - (\epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \dots + \phi_1^{p-1} \epsilon_{t-(p-1)})]^2 = \phi_1^{2p} E[Z_{t-p}^2] \rightarrow 0$$

That is,

$$Z_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j} \quad (3.32)$$

where the sum converges in mean square. In fact, with $|\phi_1| < 1$, the representation (3.32) is *the stationary solution* of the stochastic difference equation (3.31). In other words, with $|\phi_1| < 1$, there exists a stationary solution in terms of present and past ϵ_t , and hence our stationarity assumption is not without foundation. Since limits in mean square and “E” commute, we find

$$E[Z_t] = E \left\{ \lim_{n \rightarrow \infty} \sum_{j=0}^n \phi_1^j \epsilon_{t-j} \right\} = \lim_{n \rightarrow \infty} E \left\{ \sum_{j=0}^n \phi_1^j \epsilon_{t-j} \right\} = 0$$

Thus $E[Z_t] = 0$ for all t . Similarly, by the orthogonality of the ϵ_t ($E[\epsilon_t \epsilon_s] = 0$, $s \neq t$)

$$\begin{aligned} R_0 = E[Z_t^2] &= E \left\{ \lim_{m, n \rightarrow \infty} \sum_{j=0}^m \phi_1^j \epsilon_{t-j} \sum_{k=0}^n \phi_1^k \epsilon_{t-k} \right\} \\ &= \lim_{m, n \rightarrow \infty} E \left\{ \sum_{j=0}^m \phi_1^j \epsilon_{t-j} \sum_{k=0}^n \phi_1^k \epsilon_{t-k} \right\} \end{aligned}$$

$$\begin{aligned}
&= \lim_{m,n \rightarrow \infty} \sum_{j=0}^m \sum_{k=0}^n \phi_1^{j+k} E[\epsilon_{t-j} \epsilon_{t-k}] \\
&= \sum_{j=0}^{\infty} \phi_1^{2j} \sigma_{\epsilon}^2 = \frac{\sigma_{\epsilon}^2}{1 - \phi_1^2}
\end{aligned} \tag{3.33}$$

In the same way, using the representation (3.32) and the orthogonality of the ϵ_t , we obtain $E[\epsilon_t Z_{t-k}] = 0$, $k = 1, 2, \dots$, so that by multiplying both sides of (3.31) by Z_{t-k} and taking expectations,

$$R_k = \phi_1 R_{k-1}$$

This gives $\rho_1 = \phi_1$, and more generally , since $R_k = R_{-k}$,

$$\rho_k = \phi_1^{|k|}, \quad k = 0, \pm 1, \pm 2, \dots \tag{3.34}$$

From this and (3.33) we obtain the autocovariance as,

$$R_k = \frac{\sigma_{\epsilon}^2 \phi_1^{|k|}}{1 - \phi_1^2}, \quad k = 0, \pm 1, \pm 2, \dots \tag{3.35}$$

Since the autocovariance is absolutely summable, a spectral density exists and is given by,

$$\begin{aligned}
f(\lambda) &= \frac{R_0}{2\pi} \sum_{k=-\infty}^{\infty} \phi_1^{|k|} \cos(k\lambda) \\
&= \frac{R_0}{2\pi} \left\{ 1 + \sum_{k=1}^{\infty} \phi_1^k [\exp(ik\lambda) + \exp(-ik\lambda)] \right\} \\
&= \frac{R_0}{2\pi} \left\{ 1 + \frac{\phi_1 e^{i\lambda}}{1 - \phi_1 e^{i\lambda}} + \frac{\phi_1 e^{-i\lambda}}{1 - \phi_1 e^{-i\lambda}} \right\} \\
&= \frac{R_0}{2\pi} \cdot \frac{1 - \phi_1^2}{1 - 2\phi_1 \cos(\lambda) + \phi_1^2} \\
&= \frac{\sigma_{\epsilon}^2}{2\pi} \cdot \frac{1}{1 - 2\phi_1 \cos(\lambda) + \phi_1^2}, \quad -\pi \leq \lambda \leq \pi
\end{aligned} \tag{3.36}$$

Thus, the (weakly) stationary $AR(1)$ process has a continuous spectrum.

The $AR(1)$ process is special in that its parameter $\phi_1 (= \rho_1)$ single-handedly controls both the autocorrelation and the normalized spectral density (i.e. the density divided by R_0 so that the total area under the curve is 1). The autocorrelation and the normalized spectral density for $\phi_1 = 0.8$

figure=FigThree2a.ps,height=2.2infigure=FigThree2b.ps,height=2.2in
figure=FigThree2c.ps,height=2.2infigure=FigThree2d.ps,height=2.2in

Figure 3.2: *Autocorrelation and corresponding spectral density of $AR(1)$ with the indicated parameter.*

and $\phi_1 = -0.8$ are given in Figure 3.2. For $\phi_1 = 0.8$, neighboring observations are positively correlated and we expect dominance of low frequencies. Indeed, the spectral density gives most of its weight to (positive) low frequencies. For $\phi_1 = -0.8$, neighboring observations are negatively correlated which means higher oscillation and an emphasis of (positive) high frequencies. Note that since the process is real, the spectral density is symmetric and hence it is sufficient to consider it over the interval $[0, \pi]$. ♠

Remark. In Example 3.6 we imposed the condition $|\phi_1| < 1$, and showed that the stochastic difference equation (3.31) has a stationary solution in terms of past and present ϵ_t . When $|\phi_1| > 1$, the stochastic difference equation still has a stationary solution in terms of *future* but not in terms of past and present ϵ_t (see Problem 13).

Example 3.7: The Stationary AR(p) Process.

The stationary $AR(1)$ process (3.31) can be generalized by extending the order of the stochastic difference equation,

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + \epsilon_t, \quad t = 0, \pm 1, \pm 2, \dots \quad (3.37)$$

and $\{\epsilon_t\}$ is real-valued white noise as in the $AR(1)$ process. As such this defines a real-valued process referred to as an *autoregressive process of order p*, or simply $AR(p)$. In order to guarantee stationarity, the ϕ_j must satisfy certain restrictions expressed in terms of the roots (zeros) of the *characteristic equation*

$$\phi(z) \equiv 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

Several stationary solutions exist, but this will be addressed later when we introduce the spectral representation of a weakly stationary process. In particular, we shall show then that if the roots of the characteristic polynomial are all outside the unit circle, then (3.37) has a unique (weakly) stationary solution given by the infinite mean square convergent sum,

$$Z_t = \sum_{j=0}^{\infty} h_j \epsilon_{t-j} \quad (3.38)$$

This implies that $E[Z_t] = 0$ for all t . The thing to notice is that only past and present ϵ_t are involved, and that (3.32) is only a special case with $h_j = \phi_1^j$. A representation of the form (3.38) is called an *infinite moving average*. To emphasize the fact that only past and present ϵ_t are involved, we say that the infinite moving average is *one-sided*, in which case the process is said to be *linear*.

From now on, by a (weakly) stationary $AR(p)$ process we shall mean the process (3.37) with the roots of $\phi(z)$ outside the unit circle, and a consequent infinite moving average representation (3.38). Then, $E[Z_t] = 0$, and $E[\epsilon_t Z_{t-k}] = 0$, $k = 1, 2, \dots$.

With $k > 0$, multiply both sides of (3.37) by Z_{t-k} and take expectations. This gives a p 'th order difference equation,

$$R_k = \phi_1 R_{k-1} + \phi_2 R_{k-2} + \cdots + \phi_p R_{k-p}$$

or, dividing by R_0 ,

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, \quad k > 0 \quad (3.39)$$

For $k = 1, 2, \dots, p$, we obtain from (3.39) p linear equations called the *Yule-Walker* equations which allow us to determine the AR parameters $\phi_1, \phi_2, \dots, \phi_p$ from $\rho_1, \rho_2, \dots, \rho_p$, where the latter are estimated from data.

Given a time series, Z_1, Z_2, \dots, Z_N , the ρ_k 's can be estimated by the *sample correlation* [22, p. 330],

$$\hat{\rho}_k = \frac{\sum_{t=1}^{N-k} (Z_t - \hat{m})(Z_{t+k} - \hat{m})}{\sum_{t=1}^{N-k} (Z_t - \hat{m})^2}, \quad k = 1, 2, 3, \dots \quad (3.40)$$

where

$$\hat{m} = \frac{1}{N} \sum_{t=1}^N Z_t$$

is the *sample mean* (see Problem 12). By substituting the $\hat{\rho}_k$'s for the ρ_k 's, we obtain estimates for the ϕ_k 's by solving the Yule-Walker equations.

The general solution of (3.39) is [2, p. 55],

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \dots + A_p G_p^k$$

where the G_j 's are the reciprocal of the roots of $\phi(z) = 0$. Thus $|G_j| < 1$, $j = 1, \dots, p$, which means that ρ_k falls off *exponentially fast* as $|k|$ increases. The fact that ρ_k is absolutely summable implies the existence of a spectral density which we shall derive later on. ♠

3.2 Representation of Stationary Processes

We have seen above that the autocovariance admits a spectral representation in terms of a spectral (measure) distribution $F(\Lambda)$,

$$R_k = \int_{-\pi}^{\pi} e^{ik\lambda} F(d\lambda), \quad k = 0, \pm 1, \pm 2, \dots$$

It turns out that, due to the spectral representation of the autocovariance, the process itself admits a similar representation but in terms of a *random spectral measure* $\xi(\Lambda)$ which assigns uncorrelated (orthogonal) weights to nonoverlapping (Borel) subsets of $(-\pi, \pi]$. This, as we shall see, is a very useful fact.

3.2.1 A Stochastic Integral

In order to introduce our next result, it is helpful first to consider a new mathematical entity, $\xi(\Lambda)$, called *random spectral measure* and defined over subsets² Λ of $(-\pi, \pi]$. The random measure $\xi(\Lambda)$ acts very much like a nonrandom measure such as probability measure, except that the weight assigned to sets is random and complex-valued. Thus, for each fixed Λ , $\xi(\Lambda)$ is a complex-valued random variable. As with probability measures, we assume that for any two nonintersecting subsets Λ_1 and Λ_2 ,

$$\xi(\Lambda_1 \cup \Lambda_2) = \xi(\Lambda_1) + \xi(\Lambda_2), \quad \Lambda_1 \cap \Lambda_2 = \emptyset \quad (3.41)$$

Since $\xi(\Lambda)$ is random, it makes sense to speak of its moments. In particular, it is required that for any subset Λ ,

$$E[\xi(\Lambda)] = 0 \quad (3.42)$$

Now assume that over the subsets of $(-\pi, \pi]$ there is defined a real-valued nonrandom measure (simply think of probability) $F(\Lambda)$. As with probability, $F(\emptyset) = 0$. From the nonrandom $F(\Lambda)$, we formulate the third property of $\xi(\Lambda)$ as,

$$E[\xi(\Lambda_1)\overline{\xi(\Lambda_2)}] = F(\Lambda_1 \cap \Lambda_2) \quad (3.43)$$

The third property implies that

$$E|\xi(\Lambda)|^2 = F(\Lambda)$$

and, when $\Lambda_1 \cap \Lambda_2 = \emptyset$ (disjoint),

$$E[\xi(\Lambda_1)\overline{\xi(\Lambda_2)}] = F(\emptyset) = 0$$

In other words, the random variables $\xi(\Lambda_1)$ and $\xi(\Lambda_2)$ are *uncorrelated* or *orthogonal* when Λ_1 and Λ_2 are disjoint. In short, $\xi(\Lambda)$ is an orthogonal complex-valued set function.

Next, we construct a *stochastic integral* with respect to $\xi(\Lambda)$. Partition $(-\pi, \pi]$ into n disjoint intervals Λ_j , and consider the indicator function $I_{\Lambda_j}(\lambda)$ which is 1 for $\lambda \in \Lambda_j$, and is 0 otherwise. For complex a_j , let,

$$g_n(\lambda) \equiv \sum_{j=1}^n a_j I_{\Lambda_j}(\lambda), \quad \lambda \in (-\pi, \pi]$$

²Technically, all subsets are assumed to be Borel subsets [6, p. 600].

The integral of $g_n(\lambda)$ with respect to $\xi(\Lambda)$ is defined as,

$$\int_{-\pi}^{\pi} g_n(\lambda) \xi(d\lambda) \equiv \sum_{j=1}^n a_j \xi(\Lambda_j)$$

Note that this is a sequence of random variable. We are interested in limits in mean square of such sequences.

Suppose $g(\lambda)$, $\lambda \in (-\pi, \pi]$, is complex-valued (nonrandom) such that $\int_{-\pi}^{\pi} |g(\lambda)|^2 F(d\lambda) < \infty$. Then it can be approximated by functions $g_n(\lambda)$ with respect to F ,

$$\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} |g_n(\lambda) - g(\lambda)|^2 F(d\lambda) = 0$$

We define

$$\int_{-\pi}^{\pi} g(\lambda) \xi(d\lambda) \equiv \lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} g_n(\lambda) \xi(d\lambda) \quad (3.44)$$

where the limit is a limit in mean square. This is what we mean by the stochastic integral of $g(\lambda)$ with respect to $\xi(\Lambda)$. It can be shown that the limit exists and is independent of the particular sequence of approximating functions $g_n(\lambda)$ [6], [9].

The newly defined integral (3.44) is linear,

$$\int_{-\pi}^{\pi} [ag(\lambda) + bh(\lambda)] \xi(d\lambda) = a \int_{-\pi}^{\pi} g(\lambda) \xi(d\lambda) + b \int_{-\pi}^{\pi} h(\lambda) \xi(d\lambda) \quad (3.45)$$

and by the orthogonality of $\xi(\Lambda)$,

$$\begin{aligned} E \left\{ \int_{-\pi}^{\pi} g(\lambda) \xi(d\lambda) \overline{\int_{-\pi}^{\pi} h(\omega) \xi(d\omega)} \right\} &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} g(\lambda) \overline{h(\omega)} E[\xi(d\lambda) \overline{\xi(d\omega)}] \\ &= \int_{-\pi}^{\pi} g(\lambda) \overline{h(\lambda)} F(d\lambda) \end{aligned} \quad (3.46)$$

where we have used the operational convention

$$E[\xi(d\lambda) \overline{\xi(d\omega)}] = \begin{cases} F(d\lambda), & \text{if } \lambda = \omega \\ 0, & \text{if } \lambda \neq \omega \end{cases} \quad (3.47)$$

Here we think of $d\lambda$ and $d\omega$ intuitively as two very small disjoint intervals containing λ and ω , respectively. Since the mean is 0, (3.46) gives

$$\text{Var} \left\{ \int_{-\pi}^{\pi} g(\lambda) \xi(d\lambda) \right\} = E \left| \int_{-\pi}^{\pi} g(\lambda) \xi(d\lambda) \right|^2 = \int_{-\pi}^{\pi} |g(\lambda)|^2 F(d\lambda) \quad (3.48)$$

3.2.2 Spectral Representation

We define a stochastic process $\{Z_t\}$ with mean zero from the integral (3.44) by replacing $g(\lambda)$ by $g(t, \lambda)$, $t = 0, \pm 1, \pm 2, \dots$,

$$Z_t = \int_{-\pi}^{\pi} g(t, \lambda) \xi(d\lambda) \quad (3.49)$$

Then, from (3.46),

$$E[Z_s \overline{Z_t}] = \int_{-\pi}^{\pi} g(s, \lambda) \overline{g(t, \lambda)} F(d\lambda) \quad (3.50)$$

We derived (3.50) from (3.49). Conversely, if (3.50) holds, then so does (3.49) [9, p. 27].

Let $\{Z_t\}$, $t = 0, \pm 1, \pm 2, \dots$, be a zero mean complex-valued weakly stationary process. Then

$$E[Z_s \overline{Z_t}] = R_{s-t} = \int_{-\pi}^{\pi} e^{is\lambda} e^{-it\lambda} F(d\lambda), \quad s, t = 0, \pm 1, \pm 2, \dots \quad (3.51)$$

Therefore, with $g(t, \lambda) = e^{it\lambda}$,

$$Z_t = \int_{-\pi}^{\pi} e^{it\lambda} \xi(d\lambda), \quad t = 0, \pm 1, \pm 2, \dots \quad (3.52)$$

where now the spectral distribution satisfies

$$E[\xi(d\lambda) \overline{\xi(d\omega)}] = \begin{cases} F(d\lambda), & \text{if } \lambda = \omega \\ 0, & \text{if } \lambda \neq \omega \end{cases} \quad (3.53)$$

The Fourier representation (3.52) is the celebrated *spectral representation* of a zero mean weakly stationary process in discrete time. Apparently starting with the works of A. Kolmogorov and H. Cramér in the early 1940's, the representation has been studied by quite a few investigators who devised different methods of proof, some of which are discussed in [22, Sec. 4.11]. In the above discussion we followed [9]. For an interesting historical account of the spectral representation see Note 17 in [30, p. 33, Vol II].

The continuous time version of (3.52) requires the continuity of $R(\tau)$ at $\tau = 0$. This implies that $R(\tau)$ is continuous for all τ . The representation in the continuous time case is entirely analogous to (3.52) except that the limits of integration extend from $-\infty$ to ∞ ,

$$Z(t) = \int_{-\infty}^{\infty} e^{it\lambda} \xi(d\lambda), \quad -\infty < t < \infty \quad (3.54)$$

Spectral Representation for Real-Valued Processes

When $\{Z_t\}$ is real-valued, of course we can still use (3.52) as we shall do on several occasions. However, it is sometimes convenient to rewrite the spectral representation in terms of “sines” and “cosines”. This is done as follows. First note that real $\{Z_t\}$ implies³,

$$\xi(-d\lambda) = \overline{\xi(d\lambda)}$$

If we define real-valued $\xi_1(d\lambda)$ and $\xi_2(d\lambda)$ by

$$\begin{aligned}\xi_1(d\lambda) &= \xi(d\lambda) + \overline{\xi(d\lambda)} \\ \xi_2(d\lambda) &= i\{\xi(d\lambda) - \overline{\xi(d\lambda)}\}\end{aligned}$$

then

$$\xi(d\lambda) = \frac{1}{2}\{\xi_1(d\lambda) - i\xi_2(d\lambda)\}$$

and $\xi_1(d\lambda)$ is even while $\xi_2(d\lambda)$ is odd,

$$\xi_1(-d\lambda) = \xi_1(d\lambda), \quad \xi_2(-d\lambda) = -\xi_2(d\lambda)$$

Also, from the definition,

$$\xi_1(\{0\}) = 2\xi(\{0\}), \quad \xi_2(\{0\}) = 0$$

By making the substitution $\xi(d\lambda) = \frac{1}{2}\{\xi_1(d\lambda) - i\xi_2(d\lambda)\}$ in (3.52), the cross terms vanish and the spectral representation in discrete time in the real case reduces to [19, p. 52],

$$Z_t = \frac{1}{2}\xi_1(\{0\}) + \int_{0+}^{\pi} \cos(t\lambda)\xi_1(d\lambda) + \int_{0+}^{\pi} \sin(t\lambda)\xi_2(d\lambda) \quad (3.55)$$

It is not difficult to see that $\xi_1(d\lambda)$ and $\xi_2(d\lambda)$ are orthogonal random measures, and that they are uncorrelated. We have for all $\lambda, \omega \in (0, \pi]$,

$$E[\xi_1(d\lambda)\overline{\xi_2(d\omega)}] = 0$$

$$E[\xi_1(d\lambda)\overline{\xi_1(d\omega)}] = E[\xi_2(d\lambda)\overline{\xi_2(d\omega)}] = \begin{cases} 2F(d\lambda), & \text{if } \lambda = \omega \\ 0, & \text{if } \lambda \neq \omega \end{cases}$$

This indeed gives

$$R_k = F(\{0\}) + 2 \int_{0+}^{\pi} \cos(k\lambda)F(d\lambda)$$

as it should.

³It is helpful to think of $-d\lambda$ as a small interval containing $-\lambda$.

3.3 Decomposition of the Spectral Distribution

In the previous section we have discussed the interdependency between the spectral representation of the autocovariance and that of the process itself as summarized in equations (3.51) and (3.52). The interdependency manifests itself again through a certain important decomposition of the spectrum and of the random spectral measure.

On several occasions we pointed out that the spectral distribution behaves very much like a probability measure except for a normalizing constant. Indeed, as is the case with probability, the spectral distribution (equivalently the spectral distribution function) can be decomposed into continuous and discrete components⁴,

$$F(d\lambda) = F_c(d\lambda) + F_d(d\lambda) \quad (3.56)$$

where the continuous part F_c can be expressed as an integral of a nonnegative spectral density $f(\lambda)$,

$$F_c(\Lambda) = \int_{\Lambda} f(\lambda) d\lambda$$

while the discrete part F_d can be expressed as a sum of a nonnegative mass function $p(\lambda)$ which assigns positive values only to discrete frequencies λ_j ,

$$F_d(\Lambda) = \sum_{\lambda_j \in \Lambda} p(\lambda_j)$$

Note that $F_d(\Lambda) = 0$ if Λ contains none of the discrete λ_j . It follows that the power associated with a frequency band Λ is given by,

$$F(\Lambda) = \int_{\Lambda} f(\lambda) d\lambda + \sum_{\lambda_j \in \Lambda} p(\lambda_j) \quad (3.57)$$

The mass function $p(\lambda)$, called the *spectral function*, gives the power at discrete frequencies λ_j . That is, $F(\{\lambda_j\}) = F_d(\{\lambda_j\}) = p(\lambda_j)$. Thus, whenever a “spectral line” corresponding to λ_j exists, $p(\lambda_j) > 0$; otherwise $p(\lambda) = 0$. When the process is real-valued, by the symmetry of $F(d\lambda)$, $f(-\lambda) = f(\lambda)$ and $p(-\lambda) = p(\lambda)$.

It then follows from the Wiener-Khintchine relationship that

$$\begin{aligned} R_k &= \int_{-\pi}^{\pi} e^{ik\lambda} F_c(d\lambda) + \int_{-\pi}^{\pi} e^{ik\lambda} F_d(d\lambda) \\ &= \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda + \sum_j e^{ik\lambda_j} p(\lambda_j) \end{aligned} \quad (3.58)$$

⁴A third “singular” component is rather rare and thus assumed absent.

In particular, both components contribute to the total power,

$$R_0 = \int_{-\pi}^{\pi} f(\lambda) d\lambda + \sum_k p(\lambda_k)$$

From this and (3.58) we obtain the general autocorrelation presentation,

$$\rho_k = \frac{\int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda + \sum_j e^{ik\lambda_j} p(\lambda_j)}{\int_{-\pi}^{\pi} f(\lambda) d\lambda + \sum_j p(\lambda_j)} \quad (3.59)$$

When the process is real-valued and $p(0) = 0$, this becomes by symmetry,

$$\rho_k = \frac{\int_0^{\pi} \cos(k\lambda) f(\lambda) d\lambda + \sum_{\lambda_j > 0} \cos(k\lambda_j) p(\lambda_j)}{\int_0^{\pi} f(\lambda) d\lambda + \sum_{\lambda_j > 0} p(\lambda_j)} \quad (3.60)$$

a relationship that we shall exploit when dealing with higher order crossings.

The decomposition (3.56) of the spectral distribution also implies a decomposition of the process itself corresponding to (3.58) [6, p. 488],

$$Z_t = \int_{-\pi}^{\pi} e^{it\lambda} \xi_c(d\lambda) + \sum_j e^{ik\lambda_j} \xi_j, \quad t = 0, \pm 1, \pm 2, \dots \quad (3.61)$$

where

$$E[\xi_c(d\lambda) \overline{\xi_c(d\omega)}] = \begin{cases} F_c(d\lambda) = f(\lambda) d\lambda, & \text{if } \lambda = \omega \\ 0, & \text{if } \lambda \neq \omega \end{cases} \quad (3.62)$$

and $\xi_j = \xi(\{\lambda_j\})$. It follows that the ξ_j are uncorrelated with $\xi_c(d\lambda)$ and,

$$E[\xi_j \overline{\xi_k}] = \begin{cases} F_d(\{\lambda_j\}) = p(\lambda_j), & \text{if } j = k \\ 0, & \text{if } j \neq k \end{cases} \quad (3.63)$$

Thus, the presence of a discrete spectral component $F_d(d\lambda)$ implies the presence of complex exponentials whose amplitudes ξ_j are orthogonal random variables.

From (3.59) we observe that the autocorrelation is a sum of two components corresponding to the continuous and discrete components of $F(d\lambda)$. With obvious notation, rewrite (3.59) as $\rho_k = \rho_c(k) + \rho_d(k)$. It can be shown that the contribution to the autocorrelation from the continuous component falls off to 0 as $|k|$ increases, while the contribution from the discrete part never tends to 0. Therefore, in light of the process decomposition (3.61), the presence of sinusoidal components prevents the autocorrelation from decaying to 0. Moreover, the fact that ρ_k eventually—as k increases—behaves

like $\rho_d(k)$ can serve as the basis for an approach to mixed spectrum analysis [22, p. 626].

In the basic decomposition (3.56), one of the components may be absent. When only $F_c(\Lambda)$ is present, the process is called *continuous spectrum process*. The $AR(1)$ process discussed in Example 3.6 is a continuous spectrum process. The process is called *discrete spectrum process* when only $F_d(\Lambda)$ is present. Such a process was encountered in Example 3.4. When both components are present, as in the signal plus noise process discussed in Example 3.5, the process is called *mixed spectrum process*. Thus (3.59) is the autocorrelation of a mixed spectrum zero-mean wide sense stationary process in discrete time. Mixed spectrum processes do arise in nature and also in man made systems. For example, a diurnal cycle of 24 hours can be observed in many “noisy” meteorological and climatological time series. A specific case of this is discussed in Chapter 7. Another example from geophysics—discussed in Chapter 7—is the Chandler wobble secular motion which scientists believe contains at least one spectral line corresponding, roughly, to 14 months. Also, a vibrating machinery may release signatures containing spectral lines associated with rotating parts.

Similar remarks apply to the spectral distribution function $F(\lambda)$. In terms of the spectral distribution function, the decomposition takes the form,

$$F(\lambda) = F_c(\lambda) + F_d(\lambda) \quad (3.64)$$

where

$$F_c(\lambda) = \int_{-\pi}^{\lambda} f(\omega) d\omega$$

and

$$F_d(\lambda) = \sum_{\lambda_j \leq \lambda} p(\lambda_j)$$

Thus, $F_c(\lambda)$ is absolutely continuous—integral of its derivative—with a non-negative density f , while $F_d(\lambda)$ is a step function with jumps of size $p(\lambda_j)$ at λ_j . This is precisely analogous to the probability distribution function discussed in Chapter 2. From this point of view, the autocovariance plays the role of a characteristic function except for normalization.

Convenient Formalism

It is sometimes convenient to introduce the notion of a density $g(\lambda)$ that puts together the spectral density $f(\lambda)$ and spectral function $p(\lambda)$. This is

done by resorting to the Dirac delta function $\delta(\lambda)$ which satisfies, for $h(\lambda)$ continuous at $\lambda_0 \in (-\pi, \pi]$, the integral relationship,

$$\int_{-\pi}^{\pi} h(\lambda) \delta(\lambda - \lambda_0) d\lambda = h(\lambda_0)$$

Then, if we let

$$g(\lambda) = f(\lambda) + \sum_j p(\lambda_j) \delta(\lambda - \lambda_j)$$

we obtain for $h(\lambda)$ continuous at the λ_j ,

$$\int_{-\pi}^{\pi} h(\lambda) g(\lambda) d\lambda = \int_{-\pi}^{\pi} h(\lambda) f(\lambda) d\lambda + \sum_j p(\lambda_j) h(\lambda_j)$$

In particular,

$$F(\Lambda) = \int_{\Lambda} g(\lambda) d\lambda = \int_{\Lambda} f(\lambda) d\lambda + \sum_{\lambda_j \in \Lambda} p(\lambda_j)$$

and

$$R_k = \int_{-\pi}^{\pi} e^{ik\lambda} g(\lambda) d\lambda = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda + \sum_j p(\lambda_j) e^{ik\lambda_j}$$

Thus, $g(\lambda)$ may serve as a formal spectral density. As an illustration, consider the random real-valued stationary sinusoid in Example 3.2. Then,

$$g(\lambda) = \frac{\sigma^2}{2} [\delta(\omega + \omega_1) + \delta(\omega - \omega_1)]$$

3.4 Applications of the Spectral Representation

General spectral properties of weakly stationary processes can be elicited most conveniently through the spectral representation (3.52). This will be illustrated in this section in studying the effect of linear operations on weakly stationary processes.

3.4.1 Stationary Solutions of Stochastic Difference Equations

As a first application of the spectral representation, we show how to construct stationary autoregressive processes.

Consider the stochastic difference equation of order p ,

$$X_t + a_1 X_{t-1} + \cdots + a_p X_{t-p} = u_t, \quad t = 0, \pm 1, \pm 2, \cdots \quad (3.65)$$

where the a_j are real constants such that $a_p \neq 0$, and $\{u_t\}$ is real-valued white noise with mean 0 and variance σ_u^2 . We are interested in conditions which yield real-valued weakly stationary solutions of (3.65) [9, pp. 37-38].

Clearly, $\{u_t\}$ is weakly stationary so that

$$u_t = \int_{-\pi}^{\pi} e^{it\lambda} \xi_u(d\lambda)$$

Let $\{X_t\}$ be a real-valued stationary solution. Then again,

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} \xi_x(d\lambda)$$

Then we can write,

$$u_t = \int_{-\pi}^{\pi} e^{it\lambda} (1 + a_1 e^{-i\lambda} + \cdots + a_p e^{-ip\lambda}) \xi_x(d\lambda)$$

It is convenient to introduce the polynomial,

$$\phi(z) = 1 + a_1 z + a_2 z^2 + \cdots + a_p z^p$$

Then, for $t = 0, \pm 1, \pm 2, \dots$,

$$\int_{-\pi}^{\pi} e^{it\lambda} \phi(e^{-i\lambda}) \xi_x(d\lambda) = \int_{-\pi}^{\pi} e^{it\lambda} \xi_u(d\lambda)$$

As with Fourier transforms, we can equate the integrands,

$$\phi(e^{-i\lambda}) \xi_x(d\lambda) = \xi_u(d\lambda) \quad (3.66)$$

Therefore, we obtain a useful relationship between the respective spectral distributions $F_x(d\lambda), F_u(d\lambda)$,

$$|\phi(e^{-i\lambda})|^2 F_x(d\lambda) = F_u(d\lambda) = \frac{\sigma_u^2}{2\pi} d\lambda \quad (3.67)$$

If we let $f_x(\lambda)$ be the spectral density of $\{X_t\}$, then (3.67) implies that,

$$f_x(\lambda) = \frac{\sigma_u^2}{2\pi} \cdot \frac{1}{|\phi(e^{-i\lambda})|^2}$$

But since weak stationarity implies finite variances,

$$\int_{-\pi}^{\pi} f_x(\lambda) d\lambda = \text{Var}[X_t] < \infty$$

it follows that $\phi(e^{-i\lambda}) \neq 0$, and we obtain from (3.66)

$$\xi_x(d\lambda) = \frac{1}{\phi(e^{-i\lambda})} \xi_u(d\lambda)$$

or, by integrating both sides,

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} \frac{1}{\phi(e^{-i\lambda})} \xi_u(d\lambda) \quad (3.68)$$

This is the unique general stationary solution of the difference equation (3.65). Evidently, the solution has a continuous spectrum with spectral density

$$f_x(\lambda) = \frac{\sigma_u^2}{2\pi} \cdot \frac{1}{|\phi(e^{-i\lambda})|^2} = \frac{\sigma_u^2}{2\pi} \cdot \frac{1}{|1 + a_1 e^{-i\lambda} + \dots + a_p e^{-ip\lambda}|^2} \quad (3.69)$$

By putting conditions on $\phi(e^{-i\lambda})$ we can deduce from (3.68) some particular forms for X_t .

We saw that under stationarity, $\phi(z)$ has no zeros (roots) on the unit circle $|z| = 1$. But $\phi(z)$, being a polynomial of degree p , has p zeros some of which may lie inside and some outside the unit circle. We shall see momentarily that the case of interest to us is when all the zeros are outside the unit circle.

Denote the roots of $\phi(z) = 0$ by z_1, z_2, \dots, z_p , and suppose all the roots are outside the unit circle. That is, $|z_j| > 1$ for all j . Then we can expand $1/\phi(z)$ in partial fractions for $|z| = 1$. Noting that for $|z| = 1$, $|z/z_j| < 1$, we have for some constants A_j ,

$$\frac{1}{\phi(z)} = \sum_{j=1}^p \frac{A_j}{1 - \frac{z}{z_j}} = \sum_{j=1}^p A_j \sum_{r=0}^{\infty} \left(\frac{z}{z_j} \right)^r$$

and this converges geometrically. We thus have a representation for $1/\phi(e^{-i\lambda})$ which we substitute in the general solution (3.68). By switching the order of summation we have,

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} \sum_{r=0}^{\infty} e^{-ir\lambda} \left[\sum_{j=1}^p \frac{A_j}{z_j^r} \right] \xi_u(d\lambda) \quad (3.70)$$

This however, from the spectral representation of u_t , has the form

$$\sum_{r=-\infty}^{\infty} h_r u_{t-r}$$

where

$$h_r = \begin{cases} \sum_{j=1}^p \frac{A_j}{z_j^r}, & \text{for } r \geq 0 \\ 0, & \text{for } r < 0 \end{cases}$$

Therefore, X_t is a one-sided infinite moving average in terms of past and present u_t ,

$$X_t = \sum_{r=0}^{\infty} h_r u_{t-r}$$

In summary, we have shown that if the roots of $\phi(z) = 0$ all lie outside the unit circle, the stationary solution of (3.65) is a one-sided infinite moving average in terms of $u_t, u_{t-1}, u_{t-2}, \dots$.

In the same way we can show that if all the roots are *inside* the unit circle the stationary solution is an infinite moving average in terms of u_{t+1}, u_{t+2}, \dots . If some roots are inside and some outside the unit circle the stationary solution is a two-sided infinite moving average,

$$X_t = \sum_{r=-\infty}^{\infty} h_r u_{t-r}$$

Going back to the stationary $AR(p)$ process (3.37), we conclude that if the roots of the characteristic equation

$$\phi(z) \equiv 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0$$

all lie outside the unit circle, the process can be represented as an infinite moving average in terms of past and present ϵ_t as in (3.38).

Example 3.8: Stationarizing by Differencing.

Suppose $|a| < 1$, and consider a process $\{X_t\}$, $t = 0, \pm 1, \pm 2, \dots$, defined by a special case of the stochastic difference equation (3.65),

$$X_t - X_{t-1} - aX_{t-1} + aX_{t-2} = u_t$$

Then

$$\phi(z) = 1 - z - az + az^2 = (1 - az)(1 - z)$$

and we see that the root $z_1 = 1/a$ is outside the unit circle while the second root, $z_2 = 1$, is on the unit circle. Thus, $\{X_t\}$ cannot be stationary. However, by defining $Y_t \equiv X_t - X_{t-1}$, the difference equation can be expressed as

$$Y_t = aY_{t-1} + u_t$$

and this has a stationary solution, namely $\sum_{j=0}^{\infty} a^j u_{t-j}$. Thus, in the very special case of a unit root 1, a wide sense stationary process can be obtained by differencing a nonstationary one. The same procedure applies when there are d unit roots all equal to 1, and the rest are outside the unit circle, except that now the process is differenced repeatedly d times. There is practical evidence showing that as little as one or two differences of a nonstationary process may render it stationary [2, Ch. 4]. ♠

3.4.2 Effect of a Linear Filter

By a *time invariant linear filter* applied to a sequence $\{x_t\}$, $t = 0, \pm 1, \pm 2, \dots$, we mean the linear operation or convolution,

$$\mathcal{L}(\{x_t\}) = \sum_{j=-\infty}^{\infty} h_j x_{t-j} \quad (3.71)$$

where $\mathcal{L}(\{x_t\})$ provides the value at time t of the operation, and the h_j are complex constants. The sequence $\{h_j\}$ is called the *impulse response* of the filter and is assumed to fall off to 0 sufficiently fast for the operation to converge in some sense. For example we may require absolute summability. The time invariance refers to the fact that the impulse response does not depend on t (see Problem 22). By linear filter we always mean time invariant linear filter. When $h_j = 0$ for $j < 0$, the filter is called *physically realizable* or *causal*.

Combinations of linear filters are defined in the most straightforward manner. If \mathcal{L}_1 and \mathcal{L}_2 are two linear filters, we define for complex numbers a_1, a_2 ,

$$(a_1\mathcal{L}_1 + a_2\mathcal{L}_2)(\{x_t\}) \equiv a_1\mathcal{L}_1(\{x_t\}) + a_2\mathcal{L}_2(\{x_t\}) \quad (3.72)$$

and

$$\mathcal{L}_2\mathcal{L}_1(\{x_t\}) \equiv \mathcal{L}_2(\mathcal{L}_1(\{x_t\})) \quad (3.73)$$

The application of a filter to the outcome of another as in (3.73) is known as *cascaded* or *sequential* filtering.

A convenient representation of \mathcal{L} in (3.71) is obtained through the *backward shift operator* \mathcal{B} ,

$$\mathcal{B}x_t \equiv \mathcal{B}(x_t) = x_{t-1}, \quad \mathcal{B}^{-1}x_t \equiv \mathcal{B}^{-1}(x_t) = x_{t+1}$$

That is, \mathcal{B} is itself a linear filter whose value at time t when applied to the sequence $\{x_t\}$ is x_{t-1} . Repeated application of \mathcal{B} gives,

$$\mathcal{B}^j x_t \equiv \mathcal{B}^j(x_t) \equiv \mathcal{B}(\mathcal{B}^{j-1}(x_t)) = x_{t-j}$$

Thus we can write,

$$\mathcal{L} \equiv \mathcal{L}(\mathcal{B}) = \sum_{j=-\infty}^{\infty} h_j \mathcal{B}^j \quad (3.74)$$

where $\mathcal{B}^0 \equiv 1$ is the identity operator, $\mathcal{B}^0 x_t \equiv \mathcal{B}^0(x_t) = x_t$. A specific example of (3.74) is the *AR(1)* or “alpha” filter defined for $|\alpha| < 1$,

$$\mathcal{L} \equiv \mathcal{L}(\mathcal{B}) = \frac{1}{1 - \alpha\mathcal{B}} = 1 + \alpha\mathcal{B} + \alpha^2\mathcal{B}^2 + \alpha^3\mathcal{B}^3 + \cdots \quad (3.75)$$

Note that if $h_j = 0$ for all j except for, say, $j = 0, 1, 2, \dots, p$, (3.74) reduces to a polynomial in \mathcal{B} of degree p . In this case

$$(1 + h_1\mathcal{B} + h_2\mathcal{B}^2 + \cdots + h_p\mathcal{B}^p)Z_t = Z_t + h_1Z_{t-1} + h_2Z_{t-2} + \cdots + h_pZ_{t-p}$$

In general, polynomials in \mathcal{B} are well defined and the usual operations of addition, multiplication, and division apply, provided we put restrictions on the coefficients when the result is an infinite polynomial.

A new sequence $\{y_t\}$ can be constructed by the linear operation

$$y_t = \mathcal{L}(\{x_t\}) = \sum_{j=-\infty}^{\infty} h_j \mathcal{B}^j(\{x_t\}) = \sum_{j=-\infty}^{\infty} h_j x_{t-j} \quad (3.76)$$

In (3.76), following the commonplace abuse of notation, y_t stands for both the value of $\mathcal{L}(\{x_t\})$ at time t and the new sequence $\{y_t\}$. Thus (3.76) defines an input-output linear relationship with $\{x_t\}$ being the input and $\{y_t\}$ the output. By a proper choice or *design* of the impulse response sequence, outputs with certain desired properties are obtained.

The first step in understanding linear filters, is to realize the effect of the filter on the complex exponential $x_t = e^{it\lambda}$. For $0 < \lambda \leq \pi$,

$$\mathcal{L}(\{e^{it\lambda}\}) = e^{it\lambda} \sum_{j=-\infty}^{\infty} h_j e^{-ij\lambda}$$

Thus, when the complex exponential $x_t = e^{it\lambda}$ passes through a linear filter the output is the same complex exponential except that now it is multiplied by $H(\lambda)$ defined by

$$H(\lambda) \equiv \sum_{j=-\infty}^{\infty} h_j e^{-ij\lambda}, \quad 0 < \lambda \leq \pi$$

The new function $H(\lambda)$, called the *transfer function* or *frequency response* of the filter, describes the frequency domain characteristics of the filter. Clearly, different impulse responses lead to different transfer functions. A closer look reveals that the complex exponential is an eigenfunction of any linear filter with the transfer function being the corresponding eigenvalue.

Under suitable conditions the impulse response can be recovered from the transfer function. For example, if h_j is absolutely summable or merely

$$\sum_{j=-\infty}^{\infty} |h_j|^2 < \infty$$

then

$$h_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ij\lambda} H(\lambda) d\lambda, \quad j = 0, \pm 1, \pm 2, \dots$$

and from Parseval's relation,

$$\sum_{j=-\infty}^{\infty} |h_j|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(\lambda)|^2 d\lambda$$

The function $|H(\lambda)|$ is called the *gain* of the linear filter. In general

$$H(\lambda) = |H(\lambda)| e^{i\theta(\lambda)}$$

and $\theta(\lambda)$ is called the *phase function* of the filter. Observe that for real impulse response,

$$H(-\lambda) = \overline{H(\lambda)}, \quad |H(-\lambda)| = |H(\lambda)|$$

The effect of combinations of filters can be studied again by inputting the complex exponential $x_t = e^{it\lambda}$. Thus if H_j is the transfer function of \mathcal{L}_j , $j = 1, 2$,

$$(a_1\mathcal{L}_1 + a_2\mathcal{L}_2)(\{e^{it\lambda}\}) = a_1\mathcal{L}_1(\{e^{it\lambda}\}) + a_2\mathcal{L}_2(\{e^{it\lambda}\}) = e^{it\lambda}[a_1H_1(\lambda) + a_2H_2(\lambda)] \quad (3.77)$$

We find that $a_1\mathcal{L}_1 + a_2\mathcal{L}_2$ has transfer function

$$a_1H_1(\lambda) + a_2H_2(\lambda)$$

Similarly,

$$\mathcal{L}_2\mathcal{L}_1(\{e^{it\lambda}\}) = \mathcal{L}_2(\mathcal{L}_1(\{e^{it\lambda}\})) = e^{it\lambda}H_2(\lambda)H_1(\lambda) \quad (3.78)$$

Thus, the transfer function of the sequential filter $\mathcal{L}_2\mathcal{L}_1$ is

$$H_2(\lambda)H_1(\lambda)$$

with gain $|H_2(\lambda)||H_1(\lambda)|$, and the transfer function of $\mathcal{L}_3\mathcal{L}_2\mathcal{L}_1$ is

$$H_3(\lambda)H_2(\lambda)H_1(\lambda)$$

with gain $|H_3(\lambda)||H_2(\lambda)||H_1(\lambda)|$ etc. This translates into *convolution* for the corresponding impulse response sequences. Thus if $h_{1,j}$ and $h_{2,j}$ are the impulse response sequences corresponding to \mathcal{L}_1 and \mathcal{L}_2 , respectively, then the impulse response of $\mathcal{L}_2\mathcal{L}_1$ is convolution

$$h_j = \sum_{k=-\infty}^{\infty} h_{2,k}h_{1,j-k}$$

provided the sum converges.

Time invariant linear filters are used routinely in a wide range of engineering and scientific applications whenever a manipulation of the spectrum is sought. That is, whenever we wish to derive new sequences which possess certain spectral properties. The fact that this can be done is a consequence of the following important result.

Let $\{Z_t\}$, $t = 0, \pm 1, \pm 2, \dots$, be a zero mean weakly stationary process and consider the filter (3.71) with $\sum_j |h_j| < \infty$. This guarantees a weakly stationary output with mean zero and finite variance (see Problem 26). Define a new process by

$$Y_t = \mathcal{L}(\{Z_t\})$$

Then $\{Y_t\}$ admits the spectral representation

$$Y_t = \int_{-\pi}^{\pi} e^{it\lambda} \xi_y(d\lambda) = \int_{-\pi}^{\pi} e^{it\lambda} H(\lambda) \xi_z(d\lambda) \quad (3.79)$$

Therefore, as with Fourier transforms,

$$\xi_y(d\lambda) = H(\lambda) \xi_z(d\lambda) \quad (3.80)$$

and we arrive at an important relationship between the input and output spectra,

$$F_y(d\lambda) = E[|\xi_y(d\lambda)|^2] = |H(\lambda)|^2 F_z(d\lambda) \quad (3.81)$$

From this, the spectral densities and spectral functions obey the relationships,

$$f_y(\lambda) = |H(\lambda)|^2 f_z(\lambda) \quad (3.82)$$

and

$$p_y(\lambda) = |H(\lambda)|^2 p_z(\lambda) \quad (3.83)$$

Thus, by controlling $|H(\lambda)|^2$ we can control the output spectral characteristics. We note however that, from (3.81), if for some Λ , $F_z(\Lambda) = 0$ then also $F_y(\Lambda) = 0$ and filtering cannot change this. In particular, a discrete spectrum cannot be rendered continuous.

From (3.81), by shaping $|H(\lambda)|$ we can let certain frequency bands “pass” while stopping others. When high positive frequencies pass while low positive frequencies are attenuated or cut out completely, the filter is termed *high-pass*. It is termed *low-pass* when low frequencies are enhanced and high ones are attenuated. A filter is called *band-pass* if it passes only a certain band of frequencies. By combining filters sequentially and in parallel we can construct different types of desired filters with special characteristics (see Problem 28).

As alluded to earlier, (3.81) together with $\sum_j |h_j| < \infty$, imply

$$\text{Var}[Y_t] = \int_{-\pi}^{\pi} |H(\lambda)|^2 F_z(d\lambda) < \infty \quad (3.84)$$

This is the so called *matching condition* which we require if the output is to have a finite power [19].

For a better understanding of the meaning of the gain and phase functions of a linear filter, suppose the input $\{Z_t\}$ is a random complex exponential with frequency λ_1 ,

$$Z_t = \int_{-\pi}^{\pi} e^{it\lambda} \xi_z(d\lambda) = \xi_z(\{\lambda_1\}) e^{it\lambda_1}$$

With,

$$\xi_z(\{\lambda_1\}) = |\xi_z(\{\lambda_1\})| e^{i\psi(\lambda_1)}$$

we have,

$$Z_t = |\xi_z(\{\lambda_1\})| e^{i(\lambda_1 t + \psi(\lambda_1))}$$

Then the output becomes,

$$\begin{aligned} Y_t &= \int_{-\pi}^{\pi} e^{it\lambda} H(\lambda) \xi_z(d\lambda) = H(\lambda_1) \xi_z(\{\lambda_1\}) e^{it\lambda_1} \\ &= |H(\lambda_1)| |\xi_z(\{\lambda_1\})| e^{i(\lambda_1 t + \psi(\lambda_1) + \theta(\lambda_1))} \end{aligned}$$

where we have used $H(\lambda_1) = |H(\lambda_1)| e^{i\theta(\lambda_1)}$. Thus, on output, the modulus of the input is multiplied by $|H(\lambda_1)|$ and the phase is incremented by $\theta(\lambda_1)$. When the input consists of a sum of random complex exponentials, the same holds for each term in the sum. Similarly, if the input to a filter with a real impulse response is a real-valued sinusoid with frequency λ_1 and certain (positive) amplitude and phase, the output is the same sinusoid (i.e. with the same frequency) multiplied by $|H(\lambda_1)|$ and with a phase altered additively by an amount $\theta(\lambda_1)$.

In what follows we discuss some specific filters. Many more useful examples can be found in [10],[19],[20],[30].

The Backward Shift Operator

As noted earlier, the shift $\mathcal{B}Z_t = Z_{t-1}$ is a linear filter. Since $\mathcal{B}e^{it\lambda} = e^{it\lambda} e^{-i\lambda}$, the transfer function is $e^{-i\lambda}$ and the gain is equal to 1.

The Difference Operator

With 1 being the identity operator, the difference operator is

$$\nabla Z_t \equiv (1 - \mathcal{B})Z_t = Z_t - Z_{t-1}$$

Figure 3.3: *The squared gains of the difference $1 - \mathcal{B}$ and sum $1 + \mathcal{B}$ filters.*

Figure 3.4: *The squared gains of the repeated difference $(1 - \mathcal{B})^5$ and the repeated sum $(1 + \mathcal{B})^5$.*

This is a linear filter with $h_0 = 1, h_1 = -1$, and $h_j = 0$ otherwise. Since $(1 - \mathcal{B})e^{it\lambda} = e^{it\lambda}(1 - e^{-i\lambda})$ the transfer function is,

$$H(\lambda) = 1 - e^{-i\lambda}$$

and the squared gain is

$$|H(\lambda)|^2 = |1 - e^{-i\lambda}|^2 = 2(1 - \cos \lambda)$$

In $[0, \pi]$ the gain is monotone increasing and hence this is a high-pass filter. See Figure 3.3.

The squared gain of the second difference $(1 - \mathcal{B})^2$ is $4(1 - \cos \lambda)^2$, and hence this is a more pronounced high-pass filter. Repeated or sequential differencing (two or more differences) is a simple way to obtain high-pass filters. See Figure 3.4.

The Summation Operator

The sum $(1 + \mathcal{B})Z_t = Z_t + Z_{t-1}$ has squared gain $2(1 + \cos \lambda)$. It is a low-pass filter as illustrated in Figure 3.3. Repeated summation produces more pronounced low-pass filters. Thus $(1 + \mathcal{B})^5$ with squared gain $[2(1 + \cos \lambda)]^5$ is a much more pronounced low-pass filter than a single sum $(1 + \mathcal{B})$. See Figure 3.4.

A Complex Filter

Consider the filter [11],

$$\mathcal{L}(\mathcal{B}) = (1 + e^{i\theta}\mathcal{B})^n \quad (3.85)$$

where n is a nonnegative integer and $\theta \in [0, \pi]$. The squared gain is

$$|H(\lambda)|^2 = 4^n \cos^{2n} \left(\frac{\theta - \lambda}{2} \right) \quad (3.86)$$

For sufficiently large n the filter behaves as a band-pass filter (“centered” at θ), and is used as such in Section 7.3.4. In fact, n controls the bandwidth

to such an extent that the sequence of probability densities

$$f_n(\lambda; \theta) = \frac{\cos^{2n} \left(\frac{\theta - \lambda}{2} \right)}{\int_{-\pi}^{\pi} \cos^{2n} \left(\frac{\theta - \omega}{2} \right) d\omega}, \quad -\pi < \lambda \leq \pi$$

approaches a Dirac delta as $n \rightarrow \infty$,

$$\int_{-\pi}^{\pi} f_n(\lambda; \theta) g(\lambda) d\lambda \rightarrow g(\theta)$$

for g continuous at θ .

Recursive Filters

The class of recursive filters is of great practical interest as it enables the design of useful transfer functions.

Consider the input-output relationship,

$$Y_t = -a_1 Y_{t-1} - \cdots - a_p Y_{t-p} + b_0 X_t + b_1 X_{t-1} + \cdots + b_q X_{t-q} \quad (3.87)$$

This defines a *recursive* filter with input $\{X_t\}$ and output $\{Y_t\}$. Define the polynomials in \mathcal{B} ,

$$\mathcal{P}(\mathcal{B}) = 1 + a_1 \mathcal{B} + \cdots + a_p \mathcal{B}^p$$

and

$$\mathcal{Q}(\mathcal{B}) = b_0 + b_1 \mathcal{B} + \cdots + b_q \mathcal{B}^q$$

Then,

$$\mathcal{P}(\mathcal{B}) Y_t = \mathcal{Q}(\mathcal{B}) X_t$$

This gives the *formal* equation,

$$Y_t = \frac{\mathcal{Q}(\mathcal{B})}{\mathcal{P}(\mathcal{B})} X_t$$

For this to make sense, we must be able to expand $\mathcal{Q}(\mathcal{B})/\mathcal{P}(\mathcal{B})$ in a power series as in (3.74) such that h_j vanishes fast enough. Now, if the roots z_j of $\mathcal{P}(z) = a_p \prod_{j=1}^p (z - z_j) = 0$ all lie *outside* the unit circle, $|z_j| > 1$ for all j , then by partial fraction expansion, $1/\mathcal{P}(z)$ admits a power series representation,

$$\frac{1}{\mathcal{P}(z)} = \sum_{j=0}^{\infty} \alpha_j z^j$$

which converges for $|z| \leq 1$. It follows that

$$\mathcal{L}(\mathcal{B}) \equiv \frac{\mathcal{Q}(\mathcal{B})}{\mathcal{P}(\mathcal{B})} = \sum_{j=0}^{\infty} h_j \mathcal{B}^j$$

is a well defined power series expansion in \mathcal{B} (filter) where the coefficients are absolutely summable. This gives a *one-sided* representation,

$$Y_t = \sum_{j=0}^{\infty} h_j X_{t-j}$$

where $\sum_{j=0}^{\infty} |h_j| < \infty$. By plugging $e^{-i\lambda}$ instead of \mathcal{B} we obtain the transfer function of $\mathcal{L}(\mathcal{B})$ as

$$H(\lambda) = \frac{\mathcal{Q}(e^{-i\lambda})}{\mathcal{P}(e^{-i\lambda})} \quad (3.88)$$

A filter is said to be *stable* when it gives a bounded output whenever the input itself is bounded. We find that for a recursive filter to be both stable and realizable it is sufficient that all the zeros of $\mathcal{P}(z)$ lie outside the unit circle.

Clearly, we can reverse the roles of X_t and Y_t . Thus, if the roots of $\mathcal{Q}(z)$ are outside the unit circle, we can express X_t in terms of a linear filter with $\{Y_t\}$ being the input.

Example 3.9: The AR(1) Filter.

The $AR(1)$ (or α) filter is the recursive filter

$$Y_t = \alpha Y_{t-1} + X_t$$

where $|\alpha| < 1$. The transfer function is

$$H(\lambda) = \frac{1}{1 - \alpha e^{-i\lambda}}$$

Since $|\alpha| < 1$,

$$\mathcal{L}(\mathcal{B}) = \sum_{j=0}^{\infty} \alpha^j \mathcal{B}^j$$

is well defined and

$$Y_t = \sum_{j=0}^{\infty} \alpha^j X_{t-j}$$

Figure 3.5: The squared gain of the AR(1) filter with $\alpha = 0.5, -0.5$.
 figure=FigThree4.ps,height=2.95in

From Figure 3.5, for $\alpha > 0$ the filter is a low-pass filter, and a high-pass for $\alpha < 0$. We sometimes refer to $Y_t = \alpha Y_{t-1} + X_t$ and the related filter $Y_t = \alpha Y_{t-1} + (1 - \alpha)X_t$ as *exponential smoothing*. In the case of $Y_t = \alpha Y_{t-1} + (1 - \alpha)X_t$, $|H(0)|^2 = 1$ for all $|\alpha| < 1$.

Example 3.10: ARMA Processes.

An *autoregressive moving average* (ARMA) process is defined by a recursive filter applied to white noise. Let $\{\epsilon_t\}$ be (real-valued) white noise with mean 0 and variance σ_ϵ^2 , and consider the process $\{Z_t\}$ defined as the output of the recursive filter,

$$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q} \quad (3.89)$$

The process $\{Z_t\}$, $t = 0, \pm 1, \pm 2, \dots$, is called *autoregressive moving average of order (p, q)* abbreviated to *ARMA (p, q)* or *mixed ARMA (p, q)* [2]. With

$$\mathcal{P}(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$$

$$\mathcal{Q}(z) = 1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q$$

(3.89) becomes,

$$\mathcal{P}(\mathcal{B})Z_t = \mathcal{Q}(\mathcal{B})\epsilon_t, \quad t = 0, \pm 1, \pm 2, \dots \quad (3.90)$$

A sufficient condition for (wide sense) stationarity and realizability is that the roots of $\mathcal{P}(z)$ lie outside the unit circle. In this case, since the spectral density of $\{\epsilon_t\}$ is $\sigma_\epsilon^2/2\pi$, (3.82) and (3.88) imply that the spectral density of the output $\{Z_t\}$ is

$$f_z(\lambda) = \frac{|\mathcal{Q}(e^{-i\lambda})|^2}{|\mathcal{P}(e^{-i\lambda})|^2} f_\epsilon(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \cdot \frac{|\mathcal{Q}(e^{-i\lambda})|^2}{|\mathcal{P}(e^{-i\lambda})|^2}, \quad -\pi \leq \lambda \leq \pi \quad (3.91)$$

This clearly is a *rational* function of $e^{-i\lambda}$. When also the roots of $\mathcal{Q}(z)$ lie outside the unit circle, the process is *invertible* meaning that $\epsilon_t = [\mathcal{P}(\mathcal{B})/\mathcal{Q}(\mathcal{B})]Z_t$, which is seen to be an infinite autoregression.

Example 3.11: Butterworth Sine Filter.

A useful recursive filter is the sine-Butterworth low-pass filter

$$Y_t = \alpha_1 Y_{t-1} + \cdots + \alpha_k Y_{t-k} + \beta_0 X_t$$

Figure 3.6: *The squared gain of the Butterworth sine filter with $k = 6$ and $\omega_b = 1.2$.*

whose weights are determined from the filter squared gain,

$$|H(\omega)|^2 = \frac{1}{1 + \left\{ \frac{\sin(\omega/2)}{\sin(\omega_b/2)} \right\}^{2k}}, \quad \omega, \omega_b \in [0, \pi] \quad (3.92)$$

Here ω_b serves as an ideal cutoff frequency. The method of obtaining the weights is described in [20, Chs. 4,5]. A particular case of (3.92) with $k = 6$ and $\omega_b = 1.2$ is shown in Figure 3.6.

3.4.3 Mean Square Ergodicity

Let $\{Y_t\}$, $t = 0, \pm 1, \pm 2, \dots$, be a weakly stationary process with $E[Y_t] = m$, and spectral distribution $F(d\lambda)$. If Y_1, Y_2, \dots, Y_N is a time series from the process, the sample average or sample mean is

$$\bar{Y} = \frac{1}{N} \sum_{t=1}^N Y_t$$

We shall study the mean square convergence of \bar{Y} .

It is convenient to introduce the centered process

$$Z_t = Y_t - m, \quad t = 0, \pm 1, \pm 2, \dots$$

Then $\{Z_t\}$ is a zero mean weakly stationary process, with the same $F(d\lambda)$, possessing the spectral representation (3.52). The spectral representation of $\{Z_t\}$ provides a very suggestive representation for the sample average \bar{Z} ,

$$\bar{Z} = \frac{1}{N} \sum_{t=1}^N Z_t = \int_{-\pi}^{\pi} \frac{1}{N} \sum_{t=1}^N e^{it\lambda} \xi(d\lambda) = \int_{-\pi}^{\pi} \varphi_N(\lambda) \xi(d\lambda)$$

where

$$\varphi_N(\lambda) = \frac{1}{N} \sum_{t=1}^N e^{it\lambda} = \begin{cases} e^{i\lambda(N+1)/2} \cdot \frac{\sin \frac{1}{2}\lambda N}{N \sin \frac{1}{2}\lambda}, & \lambda \neq 0 \\ 1, & \lambda = 0 \end{cases}$$

Clearly, on $(-\pi, \pi]$, as $N \rightarrow \infty$, $\varphi_N(\lambda)$ converges to 0 except for $\lambda = 0$, $\varphi_N(0) = 1$ for all N . We therefore are tempted to guess that \bar{Z} converges as $N \rightarrow \infty$ to the dc component $\xi(\{0\})$.

To show that this intuition is in fact correct, we appeal to the inequality,

$$\frac{\sin x}{x} \geq \frac{2}{\pi}, \quad -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$$

from which $|\varphi_N(\lambda)| \leq \pi/2$. Then, (3.48) and dominated convergence give,

$$\lim_{N \rightarrow \infty} E \left| \int_{-\pi}^{\pi} \varphi_N(\lambda) \xi(d\lambda) \right|^2 = F(\{0\})$$

From this and the orthogonality of $\xi(d\lambda)$ we find that

$$\lim_{N \rightarrow \infty} E |\bar{Z} - \xi(\{0\})|^2 = 0$$

and therefore, $\bar{Z} \xrightarrow{m.s.} \xi(\{0\})$. Thus, switching back to Y_t ,

$$\bar{Y} = \bar{Z} + m \xrightarrow{m.s.} m + \xi(\{0\}), \quad N \rightarrow \infty \quad (3.93)$$

To get rid of the dc component $\xi(\{0\})$, it is necessary and sufficient that $F(\{0\}) = 0$. That is, that $F(\lambda)$, the spectral distribution function, be continuous at $\lambda = 0$. *We have thus shown that a necessary and sufficient condition for the mean square convergence of the sample mean to the true mean m , is that $F(\lambda)$ has no jump at $\lambda = 0$.* This statement is called *the mean square ergodic theorem* [26, p. 410].

An useful consequence of (3.93) is that

$$Var[\bar{Y}] \rightarrow Var[\xi(\{0\})] = F(\{0\}), \quad N \rightarrow \infty \quad (3.94)$$

and thus, any approximation to $F(\{0\})$ also estimates $Var[\bar{Y}]$ and vice versa. Clearly, $E[\bar{Y}] = m$ for all N .

The convergence result (3.93) is a mean square type of the law of large numbers for weakly stationary processes. Under some conditions on the spectral measure $\xi(d\lambda)$, the result also holds with probability one. Basically what is needed is that $\xi(d\lambda)$ assigns no weight to small neighborhoods of the origin excluding 0. That is, neighborhoods of the form $(2^{-n}, 2^n) \setminus \{0\}$ for large n [7], [15].

3.5 Appendix to Chapter 3

Proof of Herglotz's Theorem

First note that for any function g ,

$$\sum_{j=1}^N \sum_{l=1}^N g(j-l) = \sum_{k=-(N-1)}^{N-1} (N-|k|)g(k) \quad (3.95)$$

Now, in (3.5) take $a_j = e^{-ij\lambda}$ and $t_j = j$. Then for $N \geq 1$ and $\lambda \in [-\pi, \pi]$, the function

$$f_N(\lambda) \equiv \frac{1}{2\pi N} \sum_{j=1}^N \sum_{l=1}^N e^{-ij\lambda} e^{il\lambda} R_{j-l} \geq 0$$

is nonnegative. Rearranging terms and using (3.95), we have

$$f_N(\lambda) = \frac{1}{2\pi} \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N}\right) e^{-ik\lambda} R_k \geq 0, \quad \lambda \in [-\pi, \pi]$$

Define a measure $F_N(\Lambda)$ by the integral

$$F_N(\Lambda) \equiv \int_{\Lambda} f_N(\lambda) d\lambda$$

where Λ is a (measurable) subset of $[-\pi, \pi]$. Then

$$\int_{-\pi}^{\pi} e^{in\lambda} F_N(d\lambda) = \int_{-\pi}^{\pi} e^{in\lambda} f_N(\lambda) d\lambda = \begin{cases} \left(1 - \frac{|n|}{N}\right) R_n, & |n| < N \\ 0, & |n| \geq N \end{cases} \quad (3.96)$$

Since the measures F_N are supported on $[-\pi, \pi]$, there exists (because no spectral mass “escapes” as N grows; this is called “tightness” [26, p. 394]) a subsequence F_{N_k} which converges (weakly) to a limit measure F and hence from (3.96),

$$R_n = \lim_{k \rightarrow \infty} \int_{-\pi}^{\pi} e^{in\lambda} F_{N_k}(d\lambda) = \int_{-\pi}^{\pi} e^{in\lambda} F(d\lambda) \quad (3.97)$$

We can interpret the integration as one over a unit circle in which case we identify the points $-\pi$ and π and transfer the mass or power $F(\{-\pi\})$ at the frequency $-\pi$ to π . This leaves the representation (3.97) unchanged except that now we think of F in (3.97) (really a new F) as supported on $(-\pi, \pi]$.

Uniqueness of the Spectral Measure

The autocovariance defines the spectral measure F uniquely. To see that, suppose F_1 and F_2 are spectral measures supported on $(-\pi, \pi]$ such that for all integers k ,

$$R_k = \int_{-\pi}^{\pi} e^{ik\lambda} F_1(d\lambda) = \int_{-\pi}^{\pi} e^{ik\lambda} F_2(d\lambda)$$

Let $g(\lambda)$ be any continuous bounded function on $(-\pi, \pi]$. Then it can be approximated uniformly on $(-\pi, \pi]$ by trigonometric polynomials of the form [26, p. 395],

$$h_n(\lambda) = \sum_{|k| \leq n} a_k \exp(ik\lambda)$$

where the sum is finite. It follows that

$$\int_{-\pi}^{\pi} g(\lambda) F_1(d\lambda) = \int_{-\pi}^{\pi} g(\lambda) F_2(d\lambda)$$

and this holds in particular for indicators. Therefore, $F_1(\Lambda) = F_2(\Lambda)$ for all (measurable) subsets $\Lambda \in (-\pi, \pi]$.

3.6 Problems and Complements

1. Explain why for a Gaussian process the notions of strict and wide sense stationarity are equivalent.
(Hint: For a Gaussian process, first and second order moments exist.)
2. *Functions of stationary processes* [30, p. 51]. Let $\{Z_t\}$ be a stationary Gaussian process, and define a new process $\{Y_t\}$ by the transformation $Y_t = g(Z_t)$. Show that $\{Y_t\}$ is strictly stationary but that it is not necessarily Gaussian. (Hint: Take $Y_t = 1$ if $Z_t \geq 0$ and $Y_t = 0$ otherwise.)
3. Argue that when R_0 is finite, $F(\lambda)$ must be bounded.
4. *Random phase models* [30, p. 53]. Consider the process $Z_t = A \cos(\omega_1 t + \phi)$ where ϕ is uniformly distributed in $(0, 2\pi)$, independently of the random variable A . Show that $\{Z_t\}$ is strictly stationary. Argue that if ϕ is not uniformly distributed in $(0, 2\pi)$, then $\{Z_t\}$ is not strictly stationary. (Hint: Take $\phi = \text{constant} \in (0, 2\pi)$.)
5. *Random frequency* [1, p. 379]. Let Ω be a random variable uniformly distributed in $(0, 2\pi)$, and define for $t = 1, 2, 3 \dots$, $Z_t = \cos(\Omega t)$. Show that $\{Z_t\}$ is stationary in the wide sense but not in the strict sense.
6. Show that when $\omega_{-j} = -\omega_j$, and $\xi_{-j} = \bar{\xi}_j$, the process (3.25) is real-valued.

7. Show that a complex-valued sequence R_k^* which admits the representation (3.6) in terms of a nondecreasing distribution function $F(\lambda)$ is nonnegative definite. That is, show that for all complex numbers a_1, a_2, \dots, a_N , and integers t_1, t_2, \dots, t_N , with $N \geq 1$,

$$\sum_{j=1}^N \sum_{l=1}^N a_j \overline{a_l} R_{t_j - t_l}^* \geq 0$$

8. Let X be a random variable with characteristic function

$$\varphi(t) = E[e^{itX}] = E[\cos tX] + iE[\sin tX], \quad -\infty < t < \infty$$

Show that $\varphi(t)$ is nonnegative definite.

9. Let $\{Z(t)\}$ be a real-valued wide sense stationary process with mean 0 and autocorrelation $R(\tau) = e^{-|\tau|}$. Find the variance of $\sum_{t=1}^{10} Z_t$.
10. Let $\{Z_t\}$ be a real-valued wide sense stationary process with mean 0 and variance 1, and assume $\sum_k |\rho_k| < \infty$. Show that

$$F(\omega) = \frac{\omega + \pi}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \rho_k \frac{\sin(k\omega)}{k}, \quad \omega \in [-\pi, \pi]$$

(Hint: Use (3.8).)

11. Consider the sum of sinusoids (3.21) with $p = 3$. Express the spectral distribution $F(\omega)$ explicitly as in (3.18). Argue that $F(\omega)$ is continuous from the right.
(Partial Answer: For $\omega_1 \leq \omega < \omega_2$, $F(\omega) = \frac{1}{2}(\sigma_3^2 + \sigma_2^2 + 2\sigma_1^2)$.)
12. *Estimation of the mean* [1, p. 459]. Let Z_1, Z_2, \dots, Z_N be a time series from a weakly stationary process with mean m and autocovariance R_k . Define the sample mean as $\hat{m} = \frac{1}{N} \sum_{t=1}^N Z_t$.
- (a) Show that $E[\hat{m}] = m$.
- (b) Show that

$$N \text{Var}[\hat{m}] = \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N}\right) R_k$$

- (c) Argue that when the sum converges, then $\text{Var}[\hat{m}]$ decreases as $1/N$, in which case \hat{m} converges in mean square to m . An estimator which converges to a parameter is said to be *consistent*.

(d) Show that if the spectrum is continuous with spectral density $f(\lambda)$, then from the Wiener-Khintchine relationship,

$$NVar[\hat{m}] = \int_{-\pi}^{\pi} \frac{\sin^2 \frac{1}{2}\lambda N}{N \sin^2 \frac{1}{2}\lambda} f(\lambda) d\lambda$$

(e) Show that if in addition $f(\lambda)$ is continuous at $\lambda = 0$, then as $N \rightarrow \infty$,

$$NVar[\hat{m}] \rightarrow 2\pi f(0)$$

(Hint: The kernel tends to a Dirac delta.)

13. *Solutions of the first order stochastic difference equation.* Consider the $AR(1)$ process (3.31) with $|\phi_1| > 1$. Show that there exists a stationary solution in terms of $\epsilon_{t+1}, \epsilon_{t+2}, \epsilon_{t+3}, \dots$, weighted by powers of $1/\phi_1$. Argue however that with $|\phi_1| > 1$ the representation (3.32) leads to explosion. Also, show that if $|\phi_1| = 1$, and $\sigma_\epsilon^2 > 0$, the process is nonstationary. That is (3.31) has no stationary solutions. (Hint: Take $Z_0 = 0$ and compute the variance of Z_t [22, p. 118].)

14. Consider the weakly stationary $AR(1)$ process (3.31) with $|\phi_1| < 1$, and assume that $\{\epsilon_t\}$ is a sequence of iid $\mathcal{N}(0, \sigma_\epsilon^2)$. Argue that $\{Z_t\}$ is Gaussian and hence strictly stationary. In particular, by determining the characteristic function of Z_t , $\varphi(\zeta) = E[e^{i\zeta Z_t}]$, show that for every t ,

$$Z_t \sim \mathcal{N}(0, \frac{\sigma_\epsilon^2}{1 - \phi_1^2})$$

(Hint: Use the fact that if a sequence of normal random variables $\{X_n\}$ converges in mean square to a random variable X , then X also is normal. See Problem 14 in Ch. 2.)

15. *The Galton-Watson Process With Immigration Admits an Asymptotic $AR(1)$ Representation.* [12],[28]. We define a simple branching process with immigration as follows. Let (offspring) $\{Y_{n,i}\}$, $i, n = 1, 2, 3, \dots$, be a family of iid zero-one random variables such that $Y_{n,i} = 1$ with probability m (i.e. independent Bernoulli(m) r.v.'s). Let (immigration) $\{I_n\}$, $n = 1, 2, 3, \dots$, be a sequence of iid Poisson(λ) random variables, independent of the $Y_{n,i}$. Recall that the possible values of I_n are $0, 1, 2, \dots$, and that $E[I_n] = \lambda$ for all n . Put $X_0 = 0$, and $\sum_{i=1}^0 \equiv 0$. The Galton-Watson Process with immigration $\{X_n\}$ is defined by the

equation,

$$X_n = \sum_{i=1}^{X_{n-1}} Y_{n,i} + I_n, \quad n = 1, 2, 3, \dots \quad (3.98)$$

Evidently, this is a *Markov chain* on the nonnegative integers. Let \mathcal{F}_n be the σ -field generated by $X_0, X_1, X_2, \dots, X_n$. Think of \mathcal{F}_{n-1} as representing “past information” relative to time n .

(a) Show that $E[X_n | \mathcal{F}_{n-1}] = mX_{n-1} + \lambda$.

(b) With $\epsilon_n \equiv X_n - E[X_n | \mathcal{F}_{n-1}]$, show that $\{X_n\}$ satisfies the *stochastic regression*,

$$X_n = mX_{n-1} + \lambda + \epsilon_n, \quad n = 1, 2, 3, \dots$$

where $E[\epsilon_n | \mathcal{F}_{n-1}] = 0$ (i.e. $\{\epsilon_n\}$ is a *martingale difference*), and $E[\epsilon_n \epsilon_k] = 0, n \neq k$.

(c) Since $m < 1$, as $n \rightarrow \infty$, $\{X_n\}$ approaches a stationary regime. Suppose $\{X_n\}$ is in its stationary regime. Show that its mean is

$$E[X_n] \equiv \mu = \frac{\lambda}{1-m}$$

and that $Y_n \equiv X_n - \mu$, satisfies the *AR(1)* form,

$$Y_n = mY_{n-1} + \epsilon_n, \quad n = 1, 2, 3, \dots$$

where $E[\epsilon_n Y_{n-k}] = 0, k \geq 1$. What are the possible values of Y_n ?

(d) Show that $\rho_k = m^k$, identical with that of an *AR(1)* process whose values are continuous.

(Hint: For $k \geq 1$, $E[\epsilon_n \epsilon_{n-k}] = E\{\epsilon_{n-k} E[\epsilon_n | \mathcal{F}_{n-1}]\}$.)

(e) Fix $m = .99$, $\lambda = 0.05$. By a computer simulation of the process (3.98), illustrate the fact that the distribution of Y_n is markedly skewed, and hence far from normal (see also [17]).

16. *A Solar Energy Storage Model [14]*. Consider the process defined as

$$X_n = \max(\beta X_{n-1}, U_n), \quad n = 1, 2, 3, \dots$$

where $X(0) = 0$, $0 < \beta < 1$, and $\{U_n\}$ are iid random variables uniformly distributed in $[0, 1]$.

(a) Simulate a time series from the process.

(b) Give an interpretation for β .

- (c) Show that with a sufficiently long time series, the parameter β can be determined *exactly* with no error.
 (d) Argue that $\{X_n\}$ is asymptotically stationary.

17. Consider the $AR(p)$ process (3.37). Show that

$$Var[Z_t] = \frac{Var[\epsilon_t]}{1 - \rho_1\phi_1 - \rho_2\phi_2 - \cdots - \rho_p\phi_p}$$

18. *The $MA(q)$ Process.* Let $\{\epsilon_t\}$, $t = 0, \pm 1, \pm 2, \dots$, be real white noise with mean zero and variance σ_ϵ^2 , and define a real-valued weakly stationary process $\{Z_t\}$ by,

$$Z_t = \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \cdots - \theta_q\epsilon_{t-q}, \quad t = 0, \pm 1, \pm 2, \dots \quad (3.99)$$

The process is called a *moving average of order q* abbreviated to $MA(q)$ [2, p. 67]. Show that the autocorrelation is given by

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2}, & \text{for } k = 1, 2, \dots, q \\ 0, & \text{for } k > q \end{cases} \quad (3.100)$$

19. *Infinite Moving Average.* With $\{\epsilon_t\}$ as in the previous problem, let

$$Z_t = \sum_{j=-\infty}^{\infty} h_j \epsilon_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots \quad (3.101)$$

where $\sum_{j=-\infty}^{\infty} |h_j|^2 < \infty$. This is necessary and sufficient for mean square convergence, and is weaker than $\sum_{j=-\infty}^{\infty} |h_j| < \infty$.

(a) Show that $\{Z_t\}$ so defined is a weakly stationary process with autocovariance,

$$R_k = \sigma_\epsilon^2 \sum_{j=-\infty}^{\infty} h_j h_{j+k}, \quad k = 0, 1, 2, \dots \quad (3.102)$$

and spectral density

$$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \left| \sum_{j=-\infty}^{\infty} h_j e^{-i\lambda j} \right|^2 \quad (3.103)$$

(b) Use the spectral representation to show that *every zero mean continuous spectrum weakly stationary process with a positive spectral density $f(\lambda)$ has an infinite moving average representation* [19, p. 214].

20. Argue that (3.61) can be rewritten as

$$Z_t = \int_{-\pi}^{\pi} e^{it\lambda} \xi_c(d\lambda) + \int_{-\pi}^{\pi} e^{it\lambda} \xi_d(d\lambda), \quad t = 0, \pm 1, \pm 2, \dots \quad (3.104)$$

where

$$\xi_d(\Lambda) = \sum_{\lambda_j \in \Lambda} \xi_j$$

Show that $E[\xi_c(d\lambda)\overline{\xi_d(d\omega)}] = 0$ for all $\lambda, \omega \in (-\pi, \pi]$, and that

$$F_d(\Lambda) = E|\xi_d(\Lambda)|^2 = \sum_{\lambda_j \in \Lambda} p(\lambda_j)$$

21. *Stationarizing PC Processes.* Periodically correlated processes (PC) were defined in Chapter 1 Problem 5. These processes are encountered in such diverse fields as radiophysics and economics [30, p. 469]. Let $\{Z_t\}$, $t = 0, \pm 1, \pm 2, \dots$, be a PC process with mean zero and minimal period T . In general PC processes are not stationary, and $E[Z_u \overline{Z_v}] = R(u, v) = R(u+T, v+T)$ admits a spectral representation in terms of a two dimensional spectrum whose mass is concentrated on a set S of $2T-1$ parallel diagonal lines in $C \times C$, $C \equiv (-\pi, \pi]$,

$$S \equiv \{(\lambda_1, \lambda_2) \in C \times C : \lambda_2 = \lambda_1 - 2\pi s/T, s = 0, \pm 1, \pm 2, \dots, \pm(T-1)\}$$

The result is due to E.G. Gladyshev (1961). The continuous time version is due to H.L. Hurd (1989) [16].

(a) Argue that any zero mean wide sense stationary process is PC with $T = 1$ and a spectrum concentrated on the main diagonal.

(b) Show that $Z_t^\circ \equiv Z_{t+\theta}$, where θ is a random shift uniformly distributed over the integers $0, 1, 2, \dots, T-1$, is weakly stationary. (Hint: PC sequences are *harmonizable*. That is they admit a spectral representation in terms of a random measure $\xi(d\lambda)$, not necessarily orthogonal, such that $E[\xi(d\lambda)\overline{\xi(d\omega)}] = r_\xi(d\lambda, d\omega)$ [30, p. 476].)

22. *Linear but not time invariant.* A general linear filter with input $\{x_t\}$ and output $\{y_t\}$ is defined by,

$$y_t = \sum_{k=-\infty}^{\infty} c_{t,k} x_k, \quad t = 0, \pm 1, \pm 2, \dots$$

The filter becomes time-invariant if $c_{t,k} = h_{t-k}$. Define a filter by

$$y_t = ax_{-t}, \quad t = 0, \pm 1, \pm 2, \dots$$

Show that the filter is linear but not time-invariant [4, p. 153].

23. *A nonlinear representation.* In recent years, due to earlier work by K. Itô, V. Volterra, and N. Wiener, there is a growing interest in nonlinear representations of the form

$$\begin{aligned} X_t = & \sum_j h_j^{(1)} u_{t-j} + \sum_j \sum_k h_{jk}^{(2)} u_{t-j} u_{t-k} \\ & + \sum_j \sum_k \sum_l h_{jkl}^{(3)} u_{t-j} u_{t-k} u_{t-l} + \cdots \end{aligned} \quad (3.105)$$

where $t = 0, \pm 1, \pm 2, \dots$. When $\{u_t\}$ is wide sense stationary with mean zero, we also have

$$\begin{aligned} X_t = & \int_{-\pi}^{\pi} e^{it\omega_1} H_1(\omega_1) \xi_u(d\omega_1) \\ & + \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{it(\omega_1+\omega_2)} H_2(\omega_1, \omega_2) \xi_u(d\omega_1) \xi_u(d\omega_2) \\ & + \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{it(\omega_1+\omega_2+\omega_3)} H_3(\omega_1, \omega_2, \omega_3) \xi_u(d\omega_1) \xi_u(d\omega_2) \xi_u(d\omega_3) \\ & + \cdots \end{aligned} \quad (3.106)$$

where the kernels H_k are Fourier transforms of the $h^{(k)}$. The expansion (3.105) is called a *Volterra series expansion*. In the form (3.106) the integrals are called *Wiener-Itô functionals* or *integrals*. In continuous time the limits of integration extend from $-\infty$ to ∞ .

A problem of interest in the time series and econometrics literature is the estimation of the $h^{(k)}$, or equivalently the H_k , and testing whether the infinite expansion can be truncated after a few terms, or whether only the linear component $\sum_j h_j^{(1)} u_{t-j}$ provides an adequate representation, perhaps plus some additive noise. See for example [13], [18], [23], and the many additional references therein. In probability theory, the Wiener-Itô functionals have been shown to be a powerful tool in studying ergodic and central limit theorems of nonlinear functionals of Gaussian processes [5]. See also Chapter 6.

(a) In (3.105), determine the output X_t if $u_t = Ae^{it\omega}$ and if $u_t = A_1 e^{it\omega_1} + A_2 e^{it\omega_2}$. Compare with the answers when only the linear component $\sum_j h_j^{(1)} u_{t-j}$ is present [23, p. 29].

(b) If the u_t are iid $\mathcal{N}(0, 1)$, show that $\{X_t\}$ is stationary.

(c) Suppose the u_t are iid $\mathcal{N}(0, 1)$. Determine the spectrum of $\{X_t\}$ when (3.105) contains the linear and quadratic components only.

(d) With u_t iid $\mathcal{N}(0, 1)$, simulate a time series from (3.105) once when

only the linear component is present, and once when the linear and quadratic components are present.

24. *Estimation of lags* [18]. Let $\{X_t\}$, $t = 0, \pm 1, \pm 2, \dots$ be a zero mean real-valued stationary process with autocovariance $R_{xx}(k)$, and spectral density $f_{xx}(\lambda)$, and assume all the relevant moments exist. Define a zero mean *lag process* with lag v ,

$$X_v(t) \equiv X_t X_{t-v} - R_{xx}(v), \quad t = 0, \pm 1, \pm 2, \dots$$

For a fixed but *unknown* lag v , consider the model,

$$Y_t = \sum_{k=-\infty}^{\infty} l_k X_{t-k} + \sum_{k=-\infty}^{\infty} b_k X_v(t-k) + \epsilon_t, \quad t = 0, \pm 1, \dots \quad (3.107)$$

where $\{\epsilon_t\}$ is stationary noise with mean zero independent of $\{X_t\}$, and where we assume that the infinite sums converge in mean square. Define, $B(\lambda) = \sum_{k=-\infty}^{\infty} e^{-ik\lambda} b_k$, $-\pi < \lambda \leq \pi$. The *cross covariance* of $\{X_t\}$ and $\{Y_t\}$ (with 0 means) is defined by

$$R_{xy}(k) = E[X_t Y_{t-k}], \quad k = 0, \pm 1, \pm 2, \dots$$

The *cross spectral density* is given by,

$$f_{xy}(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-i\lambda k} R_{xy}(k)$$

provided the sum is finite. The quantity,

$$S_1(\lambda) \equiv \frac{|f_{xy}(\lambda)|^2}{f_{xx}(\lambda) f_{yy}(\lambda)}, \quad -\pi < \lambda \leq \pi$$

is called the (squared) *coherence* between $\{X_t\}$ and $\{Y_t\}$ at frequency λ . It acts like a squared correlation in the frequency domain. A concrete application of the measure of coherence can be found in [3, pp. 274-5]. A quantity analogous to $S_1(\lambda)$ is defined by,

$$S_2(\lambda; v) \equiv S_1(\lambda) + \frac{|B(\lambda)|^2}{f_{yy}(\lambda)} \left[f_{x_v x_v}(\lambda) - \frac{|f_{x x_v}(\lambda)|^2}{f_{xx}(\lambda)} \right], \quad -\pi < \lambda \leq \pi$$

and is called *lagged coherence* (of lag u) between $\{X_t\}$ and $\{Y_t\}$.

(a) Show that if in (3.107) the “quadratic” term $\sum_{k=-\infty}^{\infty} b_k X_v(t-k)$ corresponding to the lag process $\{X_v(t)\}$ is missing, then

$$f_{\epsilon\epsilon}(\lambda) = f_{yy}[1 - S_1(\lambda)], \quad -\pi < \lambda \leq \pi$$

Hence conclude that $0 \leq S_1(\lambda) \leq 1$, for all $\lambda \in (-\pi, \pi]$. Argue that when $S_1(\lambda) \equiv 1$, we have a linear system without noise, and when $S_1(\lambda) \equiv 0$, the output is all noise.

(b) With obvious notation, argue that

$$E[\xi_x(d\lambda)\overline{\xi_y(d\omega)}] = \begin{cases} f_{xy}(\lambda)d\lambda, & \text{if } \lambda = \omega \\ 0, & \text{if } \lambda \neq \omega \end{cases} \quad (3.108)$$

and hence another proof that $S_1(\lambda)$ is a frequency domain squared correlation is implied by the Cauchy-Schwarz inequality,

$$|E[\xi_x(d\lambda)\overline{\xi_y(d\omega)}]|^2 \leq E[|\xi_x(d\lambda)|^2]E[|\xi_y(d\lambda)|^2]$$

(c) Show that for any lag v [18],

$$0 \leq S_1(\lambda) \leq S_2(\lambda; v) \leq 1, \quad -\pi < \lambda \leq \pi$$

(d) Show that

$$f_{ee}(\lambda) = f_{yy}[1 - S_2(\lambda; v)], \quad -\pi < \lambda \leq \pi$$

Hence conclude that a *sensible way to estimate the unknown lag v is to maximize $S_2(\lambda; u)$ over u as to minimize the noise variance.*

(e) Argue that when $S_1(\lambda)$ is small while $S_2(\lambda; v)$ is high for all λ , the “quadratic” term $\sum_{k=-\infty}^{\infty} b_k X_v(t - k)$ is indispensable.

25. *Spectral density estimation.* Consider the stationary $AR(2)$ process,

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \epsilon_t, \quad t = 0, \pm 1, \pm 2, \dots$$

where $\{\epsilon_t\}$ is white Gaussian noise. The spectral density is,

$$f(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \cdot \frac{1}{|1 - \phi_1 e^{-i\lambda} - \phi_2 e^{-i2\lambda}|^2}, \quad \lambda \in [-\pi, \pi] \quad (3.109)$$

Choose any ϕ_1, ϕ_2 such that $\phi_2 + \phi_1 < 1$, $\phi_2 - \phi_1 < 1$, $-1 < \phi_2 < 1$ (this guarantees stationarity [2, p. 58]), and simulate a time series of length $N = 1000$ from the process. The simulation can be carried out as follows. First generate iid $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$. By fixing $Z_0 = Z_{-1} = 0$, we have

$$\begin{aligned} Z_1 &= \epsilon_1 \\ Z_2 &= \phi_1 Z_1 + \epsilon_2 \\ Z_3 &= \phi_1 Z_2 + \phi_2 Z_1 + \epsilon_3 \\ Z_4 &= \phi_1 Z_3 + \phi_2 Z_2 + \epsilon_4 \\ &\vdots \end{aligned}$$

Based on $Z_1, Z_2, \dots, Z_{1000}$, estimate the spectral density $f(\lambda)$ by the three following procedures, and then compare the estimates relative to the true density (3.109).

(a) *A parametric approach: AR spectral estimate.* The *AR spectral estimate* is obtained by replacing ϕ_1, ϕ_2 by their estimates obtained by solving the Yule-Walker equations,

$$\begin{aligned}\hat{\rho}_1 &= \phi_1 + \phi_2 \hat{\rho}_1 \\ \hat{\rho}_2 &= \phi_1 \hat{\rho}_1 + \phi_2\end{aligned}$$

where $\hat{\rho}_1, \hat{\rho}_2$ are sample estimates of ρ_1, ρ_2 , respectively.

(b) *A nonparametric approach: Window estimate.* Let $\bar{Z} = \frac{1}{N} \sum_{t=1}^N Z_t$ be the sample average, and define an autocovariance estimate by

$$\hat{R}_k = \frac{1}{N} \sum_{t=1}^{N-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})$$

We refer to a weighting function $w(u)$ as a *lag window* if $w(u) = w(-u)$, $|w(u)| \leq 1$ for $|u| \leq 1$, and is 0 otherwise. The *window estimate* is given by,

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{k=-M}^M w\left(\frac{k}{M}\right) \cos(\lambda k) \hat{R}_k \quad (3.110)$$

where $M(<< N)$ is the *truncation point* whose choice determines the variability and smoothness of the spectral estimate. There is a whole slew of lag windows to choose from. A popular choice is known as the *Blackman-Tukey* or *Tukey-Hanning window*,

$$w(u) = \begin{cases} \frac{1}{2}(1 + \cos(\pi u)), & |u| \leq 1 \\ 0, & |u| > 1 \end{cases}$$

Another popular choice is the *Parzen window*,

$$w(u) = \begin{cases} 1 - 6u^2 + 6|u|^3, & |u| \leq \frac{1}{2} \\ 2(1 - |u|)^3, & \frac{1}{2} \leq |u| \leq 1 \\ 0, & |u| > 1 \end{cases}$$

For $N = 1000$, a choice of $M = 50$ is quite reasonable.

(c) *Smoothed sample spectral density.* Define the *sample spectral density* by

$$J(\lambda) = \frac{1}{2\pi N} \left| \sum_{t=1}^N e^{-it\lambda} Z_t \right|^2$$

A spectral estimate is obtained by smoothing $J(\lambda)$,

$$\tilde{f}(\lambda) = \frac{1}{n} \sum_j J(\lambda_j)$$

where the $\lambda_j = 2\pi j/N$ are n values symmetric about λ , and n controls the variability and smoothness of the estimate. The estimate is sometimes called “smoothed periodogram,” and is essentially of the type (3.110).

26. *A criterion for mean square convergence* [1, p. 415]. Suppose the sequence of random variables $\{X_n\}$ satisfies as $m, n \rightarrow \infty$,

$$E[X_n X_m] \rightarrow \text{positive constant}$$

Show that $\{X_n\}$ converges in mean square to some random variable X .

27. *Stability.* A filter is said to be *stable* if a bounded input produces a bounded output. Show that a sufficient stability condition is the absolute summability of h_j .
28. *Construction of filters* [19, Sec. 4.3]. Discuss the validity of the following rules of filter design.

$$\begin{aligned} \text{high-pass}(Z_t) &= Z_t - \text{low-pass}(Z_t) \\ \text{band-pass}(Z_t) &= \text{high-pass}(\text{low-pass}(Z_t)) \\ &= \text{low-pass}(\text{high-pass}(Z_t)) \end{aligned}$$

Also, a *notch* filter is obtained from the operation,

$$\mathcal{L}(Z_t) = Z_t - \text{band-pass}(Z_t)$$

29. *Inverse of a linear filter.* Let $\{X_t\}$, $t = 0, \pm 1, \dots$, be weakly stationary with mean 0, and let \mathcal{L} be a linear filter with absolutely summable impulse response, and transfer function $H(\lambda)$ such that $H(\lambda) \neq 0$. Show that the inverse \mathcal{L}^{-1} defined by the condition $\mathcal{L}^{-1}(\mathcal{L}(X_t)) = X_t$, that is $\mathcal{L}^{-1}\mathcal{L}$ is the identity filter, is well defined with transfer function $1/H(\lambda)$.

Bibliography

- [1] Anderson, T.W., *The Statistical Analysis of Time Series*, New York: Wiley, 1971.
- [2] Box, G.E.P., and G.M. Jenkins, *Time Series Analysis, Forecasting and Control*, revised edition, Holden-Day: San Francisco, 1976.
- [3] Brillinger, D.R., *Time Series: Data Analysis and Theory*, Holt, Rinehart, and Winston: New York, 1975.
- [4] Brockwell, P.J. and R.A. Davis, *Time Series: Theory and Methods*, sec. edition, Springer-Verlag: New York, 1991.
- [5] Chambers, D., and E. Slud, "Central limit theorems for nonlinear functionals of stationary Gaussian processes," *Prob. Theory Rel. Fields*, 80, pp. 323-346, 1989.
- [6] Doob, L.J. *Stochastic Processes*, New York, NY: John Wiley, 1953.
- [7] Gaposhkin, V.F., "Criteria for the strong law of large numbers for some classes of second-order stationary processes and homogeneous random fields," *Th. Probability Appl.*, 22, pp. 286-310, 1977.
- [8] Gray, R.M. and Davisson, L.D., *Random Processes: A Mathematical Approach for Engineers*, Englewood Cliffs, NJ: Prentice-Hall, 1986.
- [9] Grenander, U. and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, New York, NY: Wiley, 1957.
- [10] Hamming, R.W., *Digital Filters*, Englewood Cliffs, NJ: Prentice-Hall, 1983.

- [11] He, S. and B. Kedem, "Higher order crossings of an almost periodic random sequence in noise," *IEEE Trans. Infor. Theory*, 35, pp. 360-370, March 1989.
- [12] Heyde, C.C., and E. Seneta, "Estimation theory for growth and immigration rates in multiplicative processes," *Jour. Appl. Probability*, 9, pp. 235-256, 1972.
- [13] Hinich, M., "Testing for Gaussianity and linearity of a stationary time series," *Jour. Time Series Anal.*, 3, pp. 169-176, 1982.
- [14] Hooghiemstra, G., and M. Keane, "Calculation of the equilibrium distribution for a solar energy storage model," *Jour. Appl. Probability*, 22, pp. 852-864, 1985.
- [15] Houdre, C., "On the spectral SLLN and pointwise ergodic theorem in L^α ," *Ann. Probability*, 21, No.1, 1993.
- [16] Hurd, H.L., "Representation of strongly harmonizable periodically correlated processes and their covariances," *Jour. Multivariate Anal.*, 29, pp. 53-67, 1989.
- [17] Kedem, B., and L.S. Chiu, "On the lognormality of rain rate," *Proc. Nat. Acad. Sci. USA*, 84, pp. 901-905, February 1987.
- [18] Kedem-Kimelfeld, B., "Estimating the lags of lag processes," *Jour. Amer. Statist. Assoc.*, 70, pp. 603-605, Sept. 1975.
- [19] Koopmans, L.H., *The Spectral Analysis of Time Series*, New York, NY: Academic Press, 1974.
- [20] Otnes, R.K. and L. Enochson, *Applied Time Series Analysis*, New York, NY: Wiley, 1978.
- [21] Papoulis, A., *Probability, Random variables, and Stochastic Processes (2nd ed.)*, New York, NY: McGraw-Hill, 1984.
- [22] Priestley, M. B., *Spectral Analysis and Time Series*, Vols. 1,2, Academic Press, London, 1981.
- [23] Priestley, M. B., *Non-Linear and Non-Stationary Time Series Analysis*, New York, NY: Academic Press, 1988.

- [24] Rosenblatt, M., *Stationary Sequences and Related Fields*, Boston, Mass: Birkhäuser, 1985.
- [25] Schuster, A., "On the investigation of hidden periodicities with application to a supposed 26-day period of meteorological phenomena," *Terr. Magnet.*, 3, pp.13-41, 1898.
- [26] Shiryayev, A.N., *Probability*, New York, NY: Springer Verlag, 1984.
- [27] Slutsky, E.E. "The summation of random causes as the source of cyclic processes," (Russian) *Problems of Economic Conditions*, 3, No. 1. (English Trans.) *Econometrica*, 5, pp.105-146, 1937.
- [28] Winnicki, J., *A Unified Theory for the Branching Process With Immigration*, Doctoral Dissertation, Department of Mathematics, University of Maryland, College Park, 1986.
- [29] Wold, H.O.A., *Bibliography on Time Series and Stochastic Processes*, Cambridge, Mass.: M.I.T. Press, 1965.
- [30] Yaglom, A.M. *Correlation Theory of Stationary and Related Random Functions*, Vols. I,II. New York, NY: Springer Verlag, 1987.
- [31] Yule, G.U., "On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers," *Philos. Trans. Royal Soc. London*, Ser. A, 226, pp.267-298, 1927.