9

Stationary processes

Summary. The theory of stationary processes, with discrete or continuous parameter, is developed. Autocovariances and spectral distributions are introduced. A theory of stochastic integration is developed for functions integrated against a stochastic process, and this theory is used to obtain a representation of a stationary process known as the spectral theorem. A result of major importance is the ergodic theorem, which explains the convergence of successive averages of a stationary process. Ergodic theorems are presented for weakly and strongly stationary processes. The final section is an introduction to Gaussian processes.

9.1 Introduction

Recall that a process X is *strongly stationary* whenever its *finite-dimensional distributions* are invariant under time shifts; it is (*weakly*) *stationary* whenever it has constant means and its *autocovariance function* is invariant under time shifts. Section 8.2 contains various examples of such processes. Next, we shall explore some deeper consequences of stationarity, in particular the spectral theorem and the ergodic theorem†.

A special class of random processes comprises those processes whose joint distributions are multivariate normal; these are called 'Gaussian processes'. Section 9.6 contains a brief account of some of the properties of such processes. In general, a Gaussian process is not stationary, but it is easy to characterize those which are.

We shall be interested mostly in continuous-time processes $X = \{X(t) : -\infty < t < \infty\}$, indexed by the whole real line, and will indicate any necessary variations for processes with other index sets, such as discrete-time processes. It is convenient to suppose that X takes values in the complex plane $\mathbb{C}$. This entails few extra complications and provides the natural setting for the theory. No conceptual difficulty is introduced by this generalization, since any complex-valued process X can be decomposed as $X = X_1 + iX_2$ where X_1 and X_2 are real-valued processes. However, we must take care when discussing the finite-dimensional distributions (fdds) of X since the distribution function of a complex-valued random variable $C = R + iI$ is no longer a function of a single real variable. Thus, our definition of strong

†The word 'ergodic' has several meanings, and probabilists tend to use it rather carelessly. We conform to this custom here.

stationarity requires revision; we leave this to the reader. The concept of weak stationarity concerns covariances; we must note an important amendment to the real-valued theory in this context. As before, the expectation operator $\mathbb{E}$ is defined by $\mathbb{E}(R + iI) = \mathbb{E}(R) + i\mathbb{E}(I)$.

(1) Definition. The **covariance** of two complex-valued random variables C_1 and C_2 is defined to be

$$\operatorname{cov}(C_1, C_2) = \mathbb{E}\big((C_1 - \mathbb{E}C_1)\overline{(C_2 - \mathbb{E}C_2)}\big)$$

where $\overline{z}$ denotes the complex conjugate of z.

This reduces to the usual definition (3.6.7) when C_1 and C_2 are real. Note that the operator 'cov' is not symmetrical in its arguments, since

$$\operatorname{cov}(C_2, C_1) = \overline{\operatorname{cov}(C_1, C_2)}.$$

Variances are defined as follows.

(2) Definition. The **variance** of a complex-valued random variable C is defined to be

$$\operatorname{var}(C) = \operatorname{cov}(C, C).$$

Decompose C into its real and imaginary parts, $C = R + iI$, and apply (2) to obtain $\operatorname{var}(C) = \operatorname{var}(R) + \operatorname{var}(I)$. We may write

$$\operatorname{var}(C) = \mathbb{E}(|C - \mathbb{E}C|^2).$$

We do not generally speak of complex random variables as being 'uncorrelated', preferring to use a word which emphasizes the geometrical properties of the complex plane.

(3) Definition. Complex-valued random variables C_1 and C_2 are called **orthogonal** if they satisfy $\operatorname{cov}(C_1, C_2) = 0$.

If $X = X_1 + iX_2$ is a complex-valued process with real part X_1 and imaginary part X_2 then $\overline{X}$ denotes the complex conjugate process of X, that is, $\overline{X} = X_1 - iX_2$.

(4) Example. Functions of the Poisson process. Let N be a Poisson process with intensity λ. Let α be a positive number, and define $X(t) = N(t + \alpha) - N(t)$, for $t \geq 0$. It is easily seen (*exercise*) from the definition of a Poisson process that X is a strongly stationary process with mean $\mathbb{E}(X(t)) = \lambda\alpha$ and autocovariance function

$$c(t, t+h) = \mathbb{E}\big(X(t)X(t+h)\big) - (\lambda\alpha)^2 = \begin{cases} 0 & \text{if } h \geq \alpha, \\ \lambda(\alpha - h) & \text{if } h < \alpha, \end{cases}$$

where $t, h \geq 0$.

Here is a second example based on the Poisson process. Let $\beta = e^{2\pi i/m}$ be a complex mth root of unity, where $m \geq 2$, and define $Y(t) = \beta^{Z+N(t)}$ where Z is a random variable that is independent of N with mass function $\mathbb{P}(Z = j) = 1/m$, for $1 \leq j \leq m$. Once again, it is left as an *exercise* to show that Y is a strictly stationary (complex-valued) process with mean $\mathbb{E}(Y(t)) = 0$. Its autocovariance function is given by the following calculation:

$$\begin{aligned} \mathbb{E}\big(Y(t)\overline{Y(t+h)}\big) &= \mathbb{E}\big(\beta^{N(t)}\overline{\beta}^{N(t+h)}\big) = \mathbb{E}\big((\beta\overline{\beta})^{N(t)}\overline{\beta}^{N(t+h)-N(t)}\big) \\ &= \mathbb{E}(\overline{\beta}^{N(h)}) \qquad \text{since } \beta\overline{\beta} = 1 \\ &= \exp\big[\lambda h(\overline{\beta} - 1)\big] \qquad \text{for } t, h \geq 0, \end{aligned}$$

where we have used elementary properties of the Poisson process. ●

Exercises for Section 9.1

1. Let $\ldots, Z_{-1}, Z_0, Z_1, Z_2, \ldots$ be independent real random variables with means 0 and variances 1, and let $\alpha, \beta \in \mathbb{R}$. Show that there exists a (weakly) stationary sequence $\{W_n\}$ satisfying $W_n = \alpha W_{n-1} + \beta W_{n-2} + Z_n$, $n = \ldots, -1, 0, 1, \ldots$, if the (possibly complex) zeros of the quadratic equation $z^2 - \alpha z - \beta = 0$ are smaller than 1 in absolute value.

2. Let U be uniformly distributed on [0, 1] with binary expansion $U = \sum_{i=1}^{\infty} X_i 2^{-i}$. Show that the sequence

$$V_n = \sum_{i=1}^{\infty} X_{i+n} 2^{-i}, \qquad n \geq 0,$$

is strongly stationary, and calculate its autocovariance function.

3. Let $\{X_n : n = \ldots, -1, 0, 1, \ldots\}$ be a stationary real sequence with mean 0 and autocovariance function $c(m)$.

(i) Show that the infinite series $\sum_{n=0}^{\infty} a_n X_n$ converges almost surely, and in mean square, whenever $\sum_{n=0}^{\infty} |a_n| < \infty$.

(ii) Let

$$Y_n = \sum_{k=0}^{\infty} a_k X_{n-k}, \qquad n = \ldots, -1, 0, 1, \ldots$$

where $\sum_{k=0}^{\infty} |a_k| < \infty$. Find an expression for the autocovariance function c_Y of Y, and show that

$$\sum_{m=-\infty}^{\infty} |c_Y(m)| < \infty.$$

4. Let $X = \{X_n : n \geq 0\}$ be a discrete-time Markov chain with countable state space S and stationary distribution $\boldsymbol{\pi}$, and suppose that X_0 has distribution $\boldsymbol{\pi}$. Show that the sequence $\{f(X_n) : n \geq 0\}$ is strongly stationary for any function $f : S \to \mathbb{R}$.

9.2 Linear prediction

Statisticians painstakingly observe and record processes which evolve in time, not merely for the benefit of historians but also in the belief that it is an advantage to know the past when attempting to predict the future. Most scientific schemes (and many non-scientific schemes) for prediction are 'model' based, in that they make some specific assumptions about the process, and then use past data to extrapolate into the future. For example, in the statistical theory of 'time series', one often assumes that the process is some combination of general trend, periodic fluctuations, and random noise, and it is common to suppose that the noise component is a stationary process having an autocovariance function of a certain form.

Suppose that we are observing a sequence $\{x_n\}$ of numbers, the number x_n being revealed to us at time n, and that we are prepared to accept that these numbers are the outcomes of a stationary sequence $\{X_n\}$ with known mean $\mathbb{E}X_n = \mu$ and autocovariance function $c(m) = \operatorname{cov}(X_n, X_{n+m})$. We may be required to estimate the value of X_{r+k} (where $k \geq 1$), given the values $X_r, X_{r-1}, \ldots, X_{r-s}$. We saw in Section 7.9 that the 'best' (that is, the minimum mean-squared-error) predictor of X_{r+k} given $X_r, X_{r-1}, \ldots, X_{r-s}$ is the conditional mean $M = \mathbb{E}(X_{r+k} \mid X_r, X_{r-1}, \ldots, X_{r-s})$; that is to say, the mean squared error $\mathbb{E}((Y - X_{r+k})^2)$

is minimized over all choices of functions Y of $X_r, X_{r-1}, \dots, X_{r-s}$ by the choice $Y = M$. The calculation of such quantities requires a knowledge of the finite-dimensional distributions (fdds) of X which we do not generally possess. For various reasons, it is not realistic to attempt to estimate the fdds in order to *estimate* the conditional mean. The problem becomes more tractable, and its solution more elegant, if we restrict our attention to *linear* predictors of X_{r+k}, which is to say that we seek the best predictor of X_{r+k} amongst the class of linear functions of $X_r, X_{r-1}, \dots, X_{r-s}$.

(1) Theorem. *Let X be a real stationary sequence with zero mean and autocovariance function $c(m)$. Amongst the class of linear functions of the subsequence $X_r, X_{r-1}, \dots, X_{r-s}$, the best predictor of X_{r+k} (where $k \geq 1$) is*

$$\widehat{X}_{r+k} = \sum_{i=0}^{s} a_i X_{r-i} \tag{2}$$

where the a_i satisfy the equations

$$\sum_{i=0}^{s} a_i c(|i-j|) = c(k+j) \quad \text{for} \quad 0 \leq j \leq s. \tag{3}$$

Proof. Let H be the closed linear space of linear functions of $X_r, X_{r-1}, \dots, X_{r-s}$. We have from the projection theorem (7.9.14) that the element M of H for which $\mathbb{E}((X_{r+k} - M)^2)$ is a minimum is the (almost surely) unique M such that

$$\mathbb{E}\big((X_{r+k} - M)Z\big) = 0 \quad \text{for all } Z \in H. \tag{4}$$

Certainly $X_{r-j} \in H$ for $0 \leq j \leq s$. Writing $M = \sum_{i=0}^{s} a_i X_{r-i}$ and substituting $Z = X_{r-j}$ in (4), we obtain

$$\mathbb{E}(X_{r+k}X_{r-j}) = \mathbb{E}(MX_{r-j}) = \sum_{i=0}^{s} a_i \mathbb{E}(X_{r-i}X_{r-j}),$$

whence (3) follows by the assumption of zero mean. ■

Therefore, if we know the autocovariance function c, then equation (3) tells us how to find the best linear predictor of future values of the stationary sequence X. In practice we may not know c, and may instead have to estimate it. Rather than digress further in this direction, the reader is referred to the time series literature, for example Chatfield (1989).

(5) Example. Autoregressive scheme. Let $\{Z_n\}$ be a sequence of independent variables with zero means and unit variances, and let $\{Y_n\}$ satisfy

$$Y_n = \alpha Y_{n-1} + Z_n, \quad -\infty < n < \infty, \tag{6}$$

where α is a real number satisfying $|\alpha| < 1$. We have from Problem (8.7.2) that Y is stationary with zero mean and autocovariance function $c(m) = \mathbb{E}(Y_n Y_{n+m})$ given by

$$c(m) = \frac{\alpha^{|m|}}{1-\alpha^2}, \quad -\infty < m < \infty. \tag{7}$$

Suppose we wish to estimate Y_{r+k} (where $k \geq 1$) from a knowledge of $Y_r, Y_{r-1}, \ldots, Y_{r-s}$. The best linear predictor is $\widehat{Y}_{r+k} = \sum_{i=0}^{s} a_i Y_{r-i}$ where the a_i satisfy equations (3):

$$\sum_{i=0}^{s} a_i \alpha^{|i-j|} = \alpha^{k+j}, \quad 0 \leq j \leq s.$$

A solution is $a_0 = \alpha^k$, $a_i = 0$ for $i \geq 1$, so that the best linear predictor is $\widehat{Y}_{r+k} = \alpha^k Y_r$. The mean squared error of prediction is

$$\begin{aligned}\mathbb{E}\big((Y_{r+k} - \widehat{Y}_{r+k})^2\big) &= \operatorname{var}(Y_{r+k} - \alpha^k Y_r) \\ &= \operatorname{var}(Y_{r+k}) - 2\alpha^k \operatorname{cov}(Y_{r+k}, Y_r) + \alpha^{2k} \operatorname{var}(Y_r) \\ &= c(0) - 2\alpha^k c(k) + \alpha^{2k} c(0) = \frac{1-\alpha^{2k}}{1-\alpha^2}, \quad \text{by (7).}\end{aligned}$$

●

(8) Example. Let $X_n = (-1)^n X_0$ where X_0 is equally likely to take each of the values -1 and $+1$. It is easily checked in this special case that X is stationary with zero mean and autocovariance function $c(m) = (-1)^m \mathbb{E}(X_0^2) = (-1)^m$, $-\infty < m < \infty$. The best linear predictor of X_{r+k} (where $k \geq 1$) based on $X_r, X_{r-1}, \ldots, X_{r-s}$ is obtained by solving the equations

$$\sum_{i=0}^{s} a_i (-1)^{|i-j|} = (-1)^{k+j}, \quad 0 \leq j \leq s.$$

A solution is $a_0 = (-1)^j$, $a_i = 0$ for $i \geq 1$, so that $\widehat{X}_{r+k} = (-1)^k X_r$, and the mean squared error of prediction is zero. ●

Exercises for Section 9.2

1. Let X be a (weakly) stationary sequence with zero mean and autocovariance function $c(m)$.

(i) Find the best linear predictor $\widehat{X}_{n+1}$ of X_{n+1} given X_n.

(ii) Find the best linear predictor $\widetilde{X}_{n+1}$ of X_{n+1} given X_n and X_{n-1}.

(iii) Find an expression for $D = \mathbb{E}\{(X_{n+1} - \widehat{X}_{n+1})^2\} - \mathbb{E}\{(X_{n+1} - \widetilde{X}_{n+1})^2\}$, and evaluate this expression when:

(a) $X_n = \cos(nU)$ where U is uniform on $[-\pi, \pi]$,

(b) X is an autoregressive scheme with $c(k) = \alpha^{|k|}$ where $|\alpha| < 1$.

2. Suppose $|a| < 1$. Does there exist a (weakly) stationary sequence $\{X_n : -\infty < n < \infty\}$ with zero means and autocovariance function

$$c(k) = \begin{cases} 1 & \text{if } k = 0, \\ \dfrac{a}{1+a^2} & \text{if } |k| = 1, \\ 0 & \text{if } |k| > 1. \end{cases}$$

Assuming that such a sequence exists, find the best linear predictor $\widehat{X}_n$ of X_n given X_{n-1}, $X_{n-2}, \ldots$, and show that the mean squared error of prediction is $(1 + a^2)^{-1}$. Verify that $\{\widehat{X}_n\}$ is (weakly) stationary.

9.3 Autocovariances and spectra

Let $X = \{X(t) : -\infty < t < \infty\}$ be a (weakly) stationary process which takes values in the complex plane $\mathbb{C}$. It has autocovariance function c given by

$$c(s, s+t) = \operatorname{cov}\big(X(s), X(s+t)\big) \quad \text{for} \quad s, t \in \mathbb{R}$$

where $c(s, s+t)$ depends on t alone. We think of c as a complex-valued function of the single variable t, and abbreviate it to

$$c(t) = c(s, s+t) \quad \text{for any } s.$$

Notice that the variance of $X(t)$ is constant for all t since

$$\operatorname{var}(X(t)) = \operatorname{cov}\big(X(t), X(t)\big) = c(0). \tag{1}$$

We shall sometimes assume that the mean value $\mathbb{E}(X(t))$ of X equals zero; if this is not true, then define $X'(t) = X(t) - \mathbb{E}(X(t))$ to obtain another stationary process with zero means and the same autocovariance function.

Autocovariances have the following properties.

(2) **Theorem.** *We have that:*

(a) $c(-t) = \overline{c(t)}$,

(b) *c is a non-negative definite function, which is to say that*

$$\sum_{j,k} c(t_k - t_j) z_j \bar{z}_k \geq 0$$

for all real $t_1, t_2, \ldots, t_n$ and all complex $z_1, z_2, \ldots, z_n$.

Proof.

(a) $c(-t) = \operatorname{cov}(X(t), X(0)) = \overline{\operatorname{cov}(X(0), X(t))} = \overline{c(t)}$.

(b) This resembles the proof of Theorem (5.7.3c). Just write

$$\sum_{j,k} c(t_k - t_j) z_j \bar{z}_k = \sum_{j,k} \operatorname{cov}\big(z_j X(t_j), z_k X(t_k)\big) = \operatorname{cov}(Z, Z) \geq 0$$

where $Z = \sum_j z_j X(t_j)$. ∎

Of more interest than the autocovariance function is the 'autocorrelation function' (see Definition (3.6.7)).

(3) **Definition.** The **autocorrelation function** of a weakly stationary process X with autocovariance function $c(t)$ is defined by

$$\rho(t) = \frac{\operatorname{cov}(X(0), X(t))}{\sqrt{\operatorname{var}(X(0))\operatorname{var}(X(t))}} = \frac{c(t)}{c(0)}$$

whenever $c(0) = \operatorname{var}(X(t)) > 0$.

Of course, $\rho(t)$ is just the correlation between $X(s)$ and $X(s+t)$, for any s.

Following the discussion in Section 8.2, we seek to assess the incidence of certain regular oscillations within the random fluctuation of X. For a weakly stationary process this is often a matter of studying regular oscillations in its autocorrelation function.

(4) Theorem. Spectral theorem for autocorrelation functions. *The autocorrelation function $\rho(t)$ of a weakly stationary process X with strictly positive variance is the characteristic function of some distribution function F whenever $\rho(t)$ is continuous at $t = 0$. That is to say,*

$$\rho(t) = \int_{-\infty}^{\infty} e^{it\lambda}\, dF(\lambda). \tag{5}$$

Proof. This follows immediately from the discussion after Theorem (5.7.3), and is a simple application of Bochner's theorem. Following (2), we need only show that ρ is uniformly continuous. Without loss of generality we can suppose that $\mathbb{E}(X(t)) = 0$ for all t. Let $c(t)$ be the autocovariance function of X, and use the Cauchy–Schwarz inequality (3.6.9) to obtain

$$\begin{aligned} |c(t+h) - c(t)| &= \left|\mathbb{E}\big(X(0)\big[X(t+h) - X(t)\big]\big)\right| \\ &\le \mathbb{E}\big(|X(0)||X(t+h) - X(t)|\big) \\ &\le \sqrt{\mathbb{E}\big(|X(0)|^2\big)\mathbb{E}\big(|X(t+h) - X(t)|^2\big)} \\ &= \sqrt{c(0)\big[2c(0) - c(h) - c(-h)\big]}. \end{aligned}$$

Therefore c is uniformly continuous whenever it is continuous at $h = 0$. Thus $\rho(t) = c(t)/c(0)$ is uniformly continuous as claimed, and the result follows. ∎

Think of equation (5) as follows. With any real λ we may associate a complex-valued oscillating function g_λ which has period $2\pi/|\lambda|$ and some non-negative amplitude f_λ, say:

$$g_\lambda(t) = f_\lambda e^{it\lambda};$$

in the less general real-valued theory we might consider oscillations of the form $g'_\lambda(t) = f_\lambda \cos(t\lambda)$ (see equations (8.2.6) and (8.2.7)). With any collection $\lambda_1, \lambda_2, \ldots$ of frequencies we can associate a mixture

$$g_\lambda(t) = \sum_j f_j e^{it\lambda_j} \tag{6}$$

of pure oscillations, where the f_j indicate the relative strengths of the various components. As the number of component frequencies in (6) grows, the summation may approach an integral

$$g(t) = \int_{-\infty}^{\infty} f(\lambda) e^{it\lambda}\, d\lambda \tag{7}$$

where f is some non-negative function which assigns weights to the λ. The progression from (6) to (7) is akin to the construction of the abstract integral (see Section 5.6). We have seen many expressions which are similar to (7), but in which f is the density function of some continuous random variable. Just as continuous variables are only a special subclass of the

larger family of all random variables, so (7) is not the most general limiting form for (6); the general form is

$$g(t) = \int_{-\infty}^{\infty} e^{it\lambda}\, dF(\lambda) \tag{8}$$

where F is a function which maps $\mathbb{R}$ into $[0, \infty)$ and which is right-continuous, non-decreasing, and such that $F(-\infty) = 0$; we omit the details of this, which are very much the same as in part B of Section 5.6. It is easy to see that F is a distribution function if and only if $g(0) = 1$. Theorem (4) asserts that ρ enjoys a decomposition in the form of (8), as a mixture of pure oscillations.

There is an alternative view of (5) which differs slightly from this. If Λ is a random variable with distribution function F, then $g_\Lambda(t) = e^{it\Lambda}$ is a pure oscillation with a random frequency. Theorem (4) asserts that ρ is the mean value of this random oscillation for some special distribution F. Of course, by the uniqueness theorem (5.9.3) there is a unique distribution function F such that (5) holds.

(9) Definition. If the autocorrelation function ρ satisfies

$$\rho(t) = \int_{-\infty}^{\infty} e^{it\lambda}\, dF(\lambda)$$

then F is called the **spectral distribution function** of the process. The **spectral density function** is the density function which corresponds to the distribution function F whenever this density exists.

For a given autocorrelation function ρ, we can find the spectral distribution function by the inversion techniques of Section 5.9.

In general, there may be certain frequency bands which make no contribution to (5). For example, if the spectral distribution function F satisfies $F(\lambda) = 0$ for all $\lambda \leq 0$, then only positive frequencies make non-trivial contributions. If the frequency band $(\lambda - \epsilon, \lambda + \epsilon)$ makes a non-trivial contribution to (5) for all $\epsilon > 0$, then we say that λ belongs to the 'spectrum' of the process.

(10) Definition. The **spectrum** of X is the set of all real numbers λ with the property that

$$F(\lambda + \epsilon) - F(\lambda - \epsilon) > 0 \quad \text{for all } \epsilon > 0$$

where F is the spectral distribution function.

If X is a discrete-time process then the above account is inadequate, since the autocorrelation function ρ now maps $\mathbb{Z}$ into $\mathbb{C}$ and cannot be a characteristic function unless its domain is extended. Theorem (4) remains broadly true, but asserts now that ρ has a representation

$$\rho(n) = \int_{-\infty}^{\infty} e^{in\lambda}\, dF(\lambda) \tag{11}$$

for some distribution function F and all integral n. No condition of continuity is appropriate here. This representation (11) is not unique because the integrand $g_\lambda(n) = e^{in\lambda}$ is periodic

in λ, which is to say that $g_{\lambda+2\pi}(n) = g_\lambda(n)$ for all n. In this case it is customary to rewrite equation (11) as

$$\rho(n) = \sum_{k=-\infty}^{\infty} \int_{((2k-1)\pi,(2k+1)\pi]} e^{in\lambda}\, dF(\lambda),$$

yielding the usual statement of the spectral theorem for discrete-time processes:

(12) $$\rho(n) = \int_{(-\pi,\pi]} e^{in\lambda}\, d\widetilde{F}(\lambda)$$

for some appropriate distribution function $\widetilde{F}$ obtained from F and satisfying $\widetilde{F}(-\pi) = 0$ and $\widetilde{F}(\pi) = 1$. A further simplification is possible if X is real valued, since then $\rho(n) = \rho(-n)$, so that

(13) $$\begin{aligned}\rho(n) &= \tfrac{1}{2}[\rho(n) + \rho(-n)] = \int_{(-\pi,\pi]} \tfrac{1}{2}(e^{in\lambda} + e^{-in\lambda})\, d\widetilde{F}(\lambda) \quad \text{by (12)} \\ &= \int_{(-\pi,\pi]} \cos(n\lambda)\, d\widetilde{F}(\lambda).\end{aligned}$$

Furthermore $\cos(n\lambda) = \cos(-n\lambda)$, and it follows that ρ may be expressed as

(14) $$\rho(n) = \int_{[-\pi,\pi]} \cos(n\lambda)\, dG(\lambda)$$

for some distribution function G of a symmetric distribution on $[-\pi, \pi]$. We note that the validity of (14) for some such G is both necessary and sufficient for ρ to be the autocorrelation function of a real-valued stationary sequence. The necessity of (14) has been shown. For its sufficiency, we shall see at the beginning of Section 9.6 that all symmetric, non-negative definite functions ρ with $\rho(0) = 1$ are autocorrelation functions of stationary sequences whose fdds are multivariate normal.

Equations (12)–(14) express ρ as the Fourier transform of some distribution function. Fourier transforms may be inverted in the usual way to obtain an expression for the spectral distribution in terms of ρ. One such expression is the following.

(15) Theorem. *Let ρ be the autocorrelation function of a stationary sequence. If the function $\widetilde{F}$ in* (12) *is differentiable with derivative f, then*

(16) $$f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} \rho(n)$$

at every point λ at which f is differentiable.

For real-valued sequences, (16) may be written as

(17) $$f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \rho(n) \cos(n\lambda), \quad \pi \le \lambda \le \pi.$$

As in the discussion after Theorem (5.9.1) of characteristic functions, a sufficient (but not necessary) condition for the existence of the spectral density function f is

$$\sum_{n=-\infty}^{\infty} |\rho(n)| < \infty. \tag{18}$$

(19) Example. Independent sequences. Let $X = \{X_n : n \geq 0\}$ be a sequence of independent variables with zero means and unit variances. In Example (8.2.8) we found that the autocorrelation function is given by

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

In order to find the spectral density function, either use (15) or recognize that

$$\rho(n) = \int_{-\pi}^{\pi} e^{in\lambda} \cdot \frac{1}{2\pi}\, d\lambda$$

to see that the spectral density function is the uniform density function on $[-\pi, \pi]$. The spectrum of X is the interval $[-\pi, \pi]$. Such a sequence X is sometimes called 'discrete white noise'. ●

(20) Example. Identical sequences. Let Y be a random variable with zero mean and unit variance, and let $X = \{X_n : n \geq 0\}$ be the stationary sequence given by $X_n = Y$ for all n. In Example (8.2.9) we calculated the autocorrelation function as $\rho(n) = 1$ for all n, and we recognize this as the characteristic function of a distribution which is concentrated at 0. The spectrum of X is the set $\{0\}$. ●

(21) Example. Two-state Markov chains. Let $X = \{X(t) : t \geq 0\}$ be a Markov chain with state space $S = \{1, 2\}$. Suppose, as in Example (6.9.15), that the times spent in states 1 and 2 are exponentially distributed with parameters α and β respectively where $\alpha\beta > 0$. That is to say, X has generator $\mathbf{G}$ given by

$$\mathbf{G} = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}.$$

In our solution to Example (6.9.15) we wrote down the Kolmogorov forward equations and found that the transition probabilities

$$p_{ij}(t) = \mathbb{P}\big(X(t) = j \,\big|\, X(0) = i\big), \quad 1 \leq i, j \leq 2,$$

are given by

$$p_{11}(t) = 1 - p_{12}(t) = \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta} e^{-t(\alpha+\beta)},$$

$$p_{22}(t) = 1 - p_{21}(t) = \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta} e^{-t(\alpha+\beta)},$$

in agreement with Example (6.10.12). Let $t \to \infty$ to find that the chain has a stationary distribution $\boldsymbol{\pi}$ given by

$$\pi_1 = \frac{\beta}{\alpha+\beta}, \quad \pi_2 = \frac{\alpha}{\alpha+\beta}.$$

Suppose now that $X(0)$ has distribution $\boldsymbol{\pi}$. As in Example (8.2.4), X is a strongly stationary process. We are going to find its spectral representation. First, find the autocovariance function. If $t \geq 0$, then a short calculation yields

$$\begin{aligned}\mathbb{E}\big(X(0)X(t)\big) &= \sum_i i\mathbb{E}\big(X(t) \mid X(0) = i\big)\pi_i = \sum_{i,j} ijp_{ij}(t)\pi_i \\ &= \frac{(2\alpha+\beta)^2}{(\alpha+\beta)^2} + \frac{\alpha\beta}{(\alpha+\beta)^2}e^{-t(\alpha+\beta)},\end{aligned}$$

and so the autocovariance function $c(t)$ is given by

$$c(t) = \mathbb{E}\big(X(0)X(t)\big) - \mathbb{E}(X(0))\mathbb{E}(X(t)) = \frac{\alpha\beta}{(\alpha+\beta)^2}e^{-t(\alpha+\beta)} \quad \text{if} \quad t \geq 0.$$

Hence $c(0) = \alpha\beta/(\alpha+\beta)^2$ and the autocorrelation function ρ is given by

$$\rho(t) = \frac{c(t)}{c(0)} = e^{-t(\alpha+\beta)} \quad \text{if} \quad t \geq 0.$$

The process X is real valued, and so ρ is symmetric; thus

$$\rho(t) = e^{-|t|(\alpha+\beta)}. \tag{22}$$

The spectral theorem asserts that ρ is the characteristic function of some distribution. We may use the inversion theorem (5.9.2) to find this distribution; however, this method is long and complicated and we prefer to rely on our experience. Compare (22) with the result of Example (5.8.4), where we saw that if Y is a random variable with the Cauchy density function

$$f(\lambda) = \frac{1}{\pi(1+\lambda^2)}, \quad -\infty < \lambda < \infty,$$

then Y has characteristic function $\phi(t) = e^{-|t|}$. Thus $\rho(t) = \phi(t(\alpha+\beta))$, and ρ is the characteristic function of $(\alpha+\beta)Y$ (see Theorem (5.7.6)). By Example (4.7.2) the density function of $\Lambda = (\alpha+\beta)Y$ is

$$f_\Lambda(\lambda) = \frac{1}{\alpha+\beta}f_Y\left(\frac{\lambda}{\alpha+\beta}\right) = \frac{\alpha+\beta}{\pi[(\alpha+\beta)^2+\lambda^2]}, \quad -\infty < \lambda < \infty,$$

and this is the spectral density function of X. The spectrum of X is the whole real line $\mathbb{R}$. ●

(23) Example. Autoregressive scheme. Let $\{Z_n\}$ be uncorrelated random variables with zero means and unit variances, and suppose that

$$X_n = \alpha X_{n-1} + Z_n, \quad -\infty < n < \infty,$$

where α is real and satisfies $|\alpha| < 1$. We saw in Problem (8.7.2) that X has autocorrelation function

$$\rho(n) = \alpha^{|n|}, \quad -\infty < n < \infty.$$

Use (16) to find the spectral density function f_X of X:

$$\begin{aligned} f_X(\lambda) &= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} \alpha^{|n|} \\ &= \frac{1-\alpha^2}{2\pi |1-\alpha e^{i\lambda}|^2} = \frac{1-\alpha^2}{2\pi(1-2\alpha\cos\lambda+\alpha^2)}, \quad -\pi \le \lambda \le \pi. \end{aligned}$$

More generally, suppose that the process Y satisfies

$$Y_n = \sum_{j=1}^{r} \alpha_j Y_{n-j} + Z_n, \quad -\infty < n < \infty$$

where $\alpha_1, \alpha_2, \ldots, \alpha_r$ are constants. The same techniques can be applied, though with some difficulty, to find that Y is stationary if the complex roots $\theta_1, \theta_2, \ldots, \theta_r$ of the polynomial

$$A(z) = z^r - \alpha_1 z^{r-1} - \cdots - \alpha_r = 0$$

satisfy $|\theta_j| < 1$. If this holds then the spectral density function f_Y of Y is given by

$$f_Y(\lambda) = \frac{1}{2\pi\sigma^2 |A(e^{-i\lambda})|^2}, \quad -\pi \le \lambda \le \pi,$$

where $\sigma^2 = \operatorname{var}(Y_0)$. ●

Exercises for Section 9.3

1. Let $X_n = A\cos(n\lambda) + B\sin(n\lambda)$ where A and B are uncorrelated random variables with zero means and unit variances. Show that X is stationary with a spectrum containing exactly one point.

2. Let U be uniformly distributed on $(-\pi, \pi)$, and let V be independent of U with distribution function F. Show that $X_n = e^{i(U-Vn)}$ defines a stationary (complex) sequence with spectral distribution function F.

3. Find the autocorrelation function of the stationary process $\{X(t) : -\infty < t < \infty\}$ whose spectral density function is:

(i) $N(0, 1)$, (ii) $f(x) = \frac{1}{2}e^{-|x|}$, $-\infty < x < \infty$.

4. Let $X_1, X_2, \ldots$ be a real-valued stationary sequence with zero means and autocovariance function $c(m)$. Show that

$$\operatorname{var}\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right) = c(0) \int_{(-\pi,\pi]} \left(\frac{\sin(n\lambda/2)}{n\sin(\lambda/2)}\right)^2 dF(\lambda)$$

where F is the spectral distribution function. Deduce that $n^{-1}\sum_{j=1}^{n} X_j \xrightarrow{\text{m.s.}} 0$ if and only if $F(0) - F(0-) = 0$, and show that

$$c(0)\{F(0) - F(0-)\} = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} c(j).$$

9.4 Stochastic integration and the spectral representation

Let $X = \{X(t) : -\infty < t < \infty\}$ be a stationary process which takes values in $\mathbb{C}$, as before. In the last section we saw that the autocorrelation function ρ enjoys the representation

$$\rho(t) = \int_{-\infty}^{\infty} e^{it\lambda}\, dF(\lambda) \tag{1}$$

as the characteristic function of some distribution function F whenever ρ is continuous at $t = 0$. This spectral representation is very useful in many contexts, including for example statistical analyses of sequences of data, but it is not the full story. Equation (1) is an analytical result with limited probabilistic content; of more interest to us is the process X, and (1) leads us to ask whether X itself enjoys a similar representation. The answer to this is in the affirmative, but the statement of the result is complicated and draws deeply from abstract theory.

Without much loss of generality we can suppose that $X(t)$ has mean 0 and variance 1 for all t. With each such stationary process X we can associate another process S called the 'spectral process' of X, in much the same way as the spectral distribution function F is associated with the autocorrelation function ρ.

(2) Spectral theorem. *If X is a stationary process with zero mean, unit variance, continuous autocorrelation function, and spectral distribution function F, there exists a complex-valued process $S = \{S(\lambda) : -\infty < \lambda < \infty\}$ such that*

$$X(t) = \int_{-\infty}^{\infty} e^{it\lambda}\, dS(\lambda). \tag{3}$$

Furthermore S has orthogonal increments in the sense that

$$\mathbb{E}\big([S(v) - S(u)][\overline{S}(t) - \overline{S}(s)]\big) = 0 \quad \text{if} \quad u \le v \le s \le t,$$

and in addition $\mathbb{E}(|S(v) - S(u)|^2) = F(v) - F(u)$ if $u \le v$.

The discrete-time stationary process $X = \{X_n : -\infty < n < \infty\}$ has a spectral representation also. The only significant difference is that the domain of the spectral process may be taken to be $(-\pi, \pi]$.

(4) Spectral theorem. *If X is a discrete-time stationary process with zero mean, unit variance, and spectral distribution function F, there exists a complex-valued process $S = \{S(\lambda) : -\pi < \lambda \le \pi\}$ such that*

$$X_n = \int_{(-\pi,\pi]} e^{in\lambda}\, dS(\lambda). \tag{5}$$

Furthermore S has orthogonal increments, and

$$\mathbb{E}\big(|S(v) - S(u)|^2\big) = F(v) - F(u) \quad \text{for} \quad u \le v. \tag{6}$$

A proof of (4) is presented later in this section. The proof of (2) is very similar, Fourier sums being replaced by Fourier integrals; this proof is therefore omitted. The process S in (3) and (5) is called the *spectral process* of X.

Before proving the above spectral representation, we embark upon an exploration of the 'stochastic integral', of which (3) and (5) are examples. The theory of stochastic integration is of major importance in modern probability theory, particularly in the study of diffusion processes.

As amply exemplified by the material in this book, probabilists are very often concerned with partial sums $\sum_{i=1}^{n} X_i$ and weighted sums $\sum_{i=1}^{n} a_i X_i$ of sequences of random variables. If X is a continuous-time process rather than a discrete-time sequence, the corresponding objects are integrals of the form $\int_\alpha^\beta a(u)\,dX(u)$; how should such an integral be defined? It is not an easy matter to discuss the 'stochastic integral' before an audience some of whom have seen little or nothing beyond the Riemann integral. There follows such an attempt.

Let $S = \{S(t) : t \in \mathbb{R}\}$ be a complex-valued continuous-time random process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and suppose that S has the following properties:

$$\mathbb{E}(|S(t)|^2) < \infty \quad \text{for all } t, \tag{7}$$

$$\mathbb{E}\big(|S(t+h) - S(t)|^2\big) \to 0 \quad \text{as} \quad h \downarrow 0, \quad \text{for all } t, \tag{8}$$

(9) the process S has *orthogonal increments* in that

$$\mathbb{E}\big([S(v) - S(u)][\overline{S}(t) - \overline{S}(s)]\big) = 0 \quad \text{whenever} \quad u \le v \le s \le t.$$

Condition (7) is helpful, since we shall work with random variables with finite second moments, and with mean-square convergence. Condition (8) is a continuity assumption which will be useful for technical reasons. Condition (9) will be of central importance in demonstrating the existence of limits necessary for the definition of the stochastic integral.

Let $G(t)$ be defined by

$$G(t) = \begin{cases} \mathbb{E}\big(|S(t) - S(0)|^2\big) & \text{if } t \ge 0, \\ -\mathbb{E}\big(|S(t) - S(0)|^2\big) & \text{if } t < 0. \end{cases} \tag{10}$$

It is an elementary calculation that

$$\mathbb{E}\big(|S(t) - S(s)|^2\big) = G(t) - G(s), \quad \text{for} \quad s \le t. \tag{11}$$

To see that this holds when $0 \le s \le t$, for example, we argue as follows:

$$\begin{aligned} G(t) &= \mathbb{E}\big(\big|[S(t) - S(s)] + [S(s) - S(0)]\big|^2\big) \\ &= \mathbb{E}\big(|S(t) - S(s)|^2\big) + \mathbb{E}\big(|S(s) - S(0)|^2\big) \\ &\qquad + \mathbb{E}\big([S(t) - S(s)][\overline{S}(s) - \overline{S}(0)] + [\overline{S}(t) - \overline{S}(s)][S(s) - S(0)]\big) \\ &= \mathbb{E}\big(|S(t) - S(s)|^2\big) + G(s) \end{aligned}$$

by the assumption of orthogonal increments. It follows from (11) that G is monotonic non-decreasing, and is right-continuous in that

$$G(t+h) \to G(t) \quad \text{as} \quad h \downarrow 0. \tag{12}$$

The function G is central to the analysis which follows.

Let $a_1 < a_2 < \cdots < a_n$, and let $c_1, c_2, \ldots, c_{n-1}$ be complex numbers. Define the step function ϕ on $\mathbb{R}$ by

$$\phi(t) = \begin{cases} 0 & \text{if } t < a_1 \text{ or } t \geq a_n, \\ c_j & \text{if } a_j \leq t < a_{j+1}, \end{cases}$$

and define the integral $I(\phi)$ of ϕ with respect to S by

$$\textbf{(13)} \qquad I(\phi) = \int_{-\infty}^{\infty} \phi(t)\, dS(t) = \sum_{j=1}^{n-1} c_j [S(a_{j+1}) - S(a_j)];$$

this is a finite sum, and therefore there is no problem concerning its existence.

Suppose that ϕ_1 and ϕ_2 are step functions of the type given above. We may assume, by a suitable 'refinement' argument, that ϕ_1 and ϕ_2 are of the form

$$\begin{aligned} &\phi_1(t) = \phi_2(t) = 0 \qquad \text{if} \quad t < a_1 \text{ or } t \geq a_n, \\ &\phi_1(t) = c_j, \ \phi_2(t) = d_j \quad \text{if} \quad a_j \leq t < a_{j+1}, \end{aligned}$$

for some $a_1 < a_2 < \cdots < a_n$. Then, using the assumption of orthogonal increments,

$$\begin{aligned} \mathbb{E}\big(I(\phi_1)\overline{I(\phi_2)}\big) &= \sum_{j,k} c_j \overline{d_k} \mathbb{E}\big([S(a_{j+1}) - S(a_j)][\overline{S}(a_{k+1}) - \overline{S}(a_k)]\big) \\ &= \sum_j c_j \overline{d_j} \mathbb{E}\big(|S(a_{j+1}) - S(a_j)|^2\big) \\ &= \sum_j c_j \overline{d_j} [G(a_{j+1}) - G(a_j)] \quad \text{by (11)}, \end{aligned}$$

which may be written as

$$\textbf{(14)} \qquad \mathbb{E}\big(I(\phi_1)\overline{I(\phi_2)}\big) = \int_{-\infty}^{\infty} \phi_1(t)\overline{\phi_2(t)}\, dG(t).$$

It is now immediate by expansion of the squares that

$$\textbf{(15)} \qquad \mathbb{E}\big(|I(\phi_1) - I(\phi_2)|^2\big) = \int_{-\infty}^{\infty} |\phi_1(t) - \phi_2(t)|^2\, dG(t),$$

which is to say that 'integration is distance preserving' in the sense that

$$\textbf{(16)} \qquad \|I(\phi_1) - I(\phi_2)\|_2 = \|\phi_1 - \phi_2\|,$$

where the first norm is given by

$$\textbf{(17)} \qquad \|U - V\|_2 = \sqrt{\mathbb{E}(|U - V|^2)} \quad \text{for random variables } U, V,$$

and the second by

$$\textbf{(18)} \qquad \|f - g\| = \sqrt{\int_{-\infty}^{\infty} |f(t) - g(t)|^2\, dG(t)} \quad \text{for suitable } f, g : \mathbb{R} \to \mathbb{C}.$$

We are ready to take limits. Let $\psi : \mathbb{R} \to \mathbb{C}$ and let $\{\phi_n\}$ be a sequence of step functions such that $\|\phi_n - \psi\| \to 0$ as $n \to \infty$. Then

$$\|\phi_n - \phi_m\| \leq \|\phi_n - \psi\| + \|\phi_m - \psi\| \to 0 \quad \text{as} \quad m, n \to \infty,$$

whence it follows from (16) that the sequence $\{I(\phi_n)\}$ is mean-square Cauchy convergent, and hence convergent in mean square (see Problem (7.11.11)). That is, there exists a random variable $I(\psi)$ such that $I(\phi_n) \xrightarrow{\text{m.s.}} I(\psi)$; we call $I(\psi)$ the integral of ψ with respect to S, writing

$$I(\psi) = \int_{-\infty}^{\infty} \psi(t)\, dS(t). \tag{19}$$

Note that the integral is not defined uniquely, but only as any mean-square limit of $I(\phi_n)$; any two such limits I_1 and I_2 are such that $\mathbb{P}(I_1 = I_2) = 1$.

For which functions ψ do there exist approximating sequences $\{\phi_n\}$ of step functions? The answer is those (measurable) functions for which

$$\int_{-\infty}^{\infty} |\psi(t)|^2\, dG(t) < \infty. \tag{20}$$

To recap, for any given function $\psi : \mathbb{R} \to \mathbb{C}$ satisfying (20), there exists a random variable

$$I(\psi) = \int_{-\infty}^{\infty} \psi(t)\, dS(t) \tag{21}$$

defined as above. Such integrals have many of the usual properties of integrals, for example:

(a) the integral of the zero function is zero,

(b) $I(\alpha\psi_1 + \beta\psi_2) = \alpha I(\psi_1) + \beta I(\psi_2)$ for $\alpha, \beta \in \mathbb{C}$,

and so on. Such statements should be qualified by the phrase 'almost surely', since integrals are not defined uniquely; we shall omit this qualification here.

Integrals may be defined on *bounded* intervals just as on the whole of the real line. For example, if $\psi : \mathbb{R} \to \mathbb{C}$ and (a, b) is a bounded interval, we define

$$\int_{(a,b)} \psi(t)\, dS(t) = \int_{-\infty}^{\infty} \psi_{ab}(t)\, dS(t)$$

where $\psi_{ab}(t) = \psi(t) I_{(a,b)}(t)$.

The above exposition is directed at integrals $\int \psi(t)\, dS(t)$ where ψ is a given function from $\mathbb{R}$ to $\mathbb{C}$. It is possible to extend this definition to the situation where ψ is itself a random process. Such an integral may be constructed very much as above, but at the expense of adding certain extra assumptions concerning the pair (ψ, S); see Section 13.8.

Proof of Theorem (4). Let H_X be the set of all linear combinations of the X_j, so that H_X is the set of all random variables of the form $\sum_{j=1}^{n} a_j X_{m(j)}$ for $a_1, a_2, \ldots, a_n \in \mathbb{C}$ and integers $n, m(1), m(2), \ldots, m(n)$. The space H_X is a vector space over $\mathbb{C}$ with a natural inner product given by

$$\langle U, V \rangle_2 = \mathbb{E}(U\overline{V}). \tag{22}$$

The *closure* $\overline{H}_X$ of H_X is defined to be the space H_X together with all limits of mean-square Cauchy-convergent sequences in H_X.

Similarly, we let H_F be the set of all linear combinations of the functions $f_n : \mathbb{R} \to \mathbb{C}$ defined by $f_n(x) = e^{inx}$ for $-\infty < x < \infty$. We impose an inner product on H_F by

$$\langle u, v \rangle = \int_{(-\pi,\pi]} u(\lambda)\overline{v(\lambda)}\, dF(\lambda) \quad \text{for} \quad u, v \in H_F, \tag{23}$$

and we write $\overline{H}_F$ for the closure of H_F, being the space H_F together with all Cauchy-convergent sequences in H_F (a sequence $\{u_n\}$ is Cauchy convergent if $\langle u_n - u_m, u_n - u_m \rangle \to 0$ as $m, n \to \infty$).

The two spaces $\overline{H}_X$ and $\overline{H}_F$ are Hilbert spaces, and we place them in one–one correspondence in the following way. Define the linear mapping $\mu : H_F \to H_X$ by $\mu(f_j) = X_j$, so that

$$\mu\Big(\sum_{j=1}^{n} a_j f_j\Big) = \sum_{j=1}^{n} a_j X_j;$$

it is seen easily that μ is one–one, in a formal sense. Furthermore,

$$\langle \mu(f_n), \mu(f_m) \rangle_2 = \langle X_n, X_m \rangle_2 = \int_{(-\pi,\pi]} e^{i(n-m)\lambda}\, dF(\lambda) = \langle f_n, f_m \rangle$$

by equations (9.3.12) and (23); therefore, by linearity, $\langle \mu(u), \mu(v) \rangle_2 = \langle u, v \rangle$ for $u, v \in H_F$, so that μ is 'distance preserving' on H_F. The domain of μ may be extended to $\overline{H}_F$ in the natural way: if $u \in \overline{H}_F$, $u = \lim_{n\to\infty} u_n$ where $u_n \in H_F$, we define $\mu(u) = \lim_{n\to\infty} \mu(u_n)$ where the latter limit is taken in the usual sense for $\overline{H}_X$. The new mapping μ from $\overline{H}_F$ to $\overline{H}_X$ is not quite one–one, since mean-square limits are not defined uniquely, but this difficulty is easily avoided (μ is one–one when viewed as a mapping from equivalence classes of functions to equivalence classes of random variables). Furthermore it may easily be checked that μ is distance preserving on $\overline{H}_F$, and linear in that

$$\mu\Big(\sum_{j=1}^{n} a_j u_j\Big) = \sum_{j=1}^{n} a_j \mu(u_j)$$

for $a_1, a_2, \ldots, a_n \in \mathbb{C}$, $u_1, u_2, \ldots, u_n \in \overline{H}_F$.

The mapping μ is sometimes called an *isometric isomorphism*. We now define the process $S = \{S(\lambda) : -\pi < \lambda \leq \pi\}$ by

$$S(\lambda) = \mu(I_\lambda) \quad \text{for} \quad -\pi < \lambda \leq \pi, \tag{24}$$

where $I_\lambda : \mathbb{R} \to \{0, 1\}$ is the indicator function of the interval $(-\pi, \lambda]$. It is a standard result of Fourier analysis that $I_\lambda \in \overline{H}_F$, so that $\mu(I_\lambda)$ is well defined. We introduce one more piece of notation, defining $J_{\alpha\beta}$ to be the indicator function of the interval $(\alpha, \beta]$; thus $J_{\alpha\beta} = I_\beta - I_\alpha$.

We need to show that X and S are related (almost surely) by (5). To this end, we check first that S satisfies conditions (7)–(9). Certainly $\mathbb{E}(|S(\lambda)|^2) < \infty$ since $S(\lambda) \in \overline{H}_X$. Secondly,

$$\begin{aligned} \mathbb{E}\big(|S(\lambda + h) - S(\lambda)|^2\big) &= \langle S(\lambda + h) - S(\lambda), S(\lambda + h) - S(\lambda) \rangle_2 \\ &= \langle J_{\lambda,\lambda+h}, J_{\lambda,\lambda+h} \rangle \end{aligned}$$

by linearity and the isometry of μ. Now $\langle J_{\lambda,\lambda+h}, J_{\lambda,\lambda+h}\rangle \to 0$ as $h \downarrow 0$, and (8) has been verified. Thirdly, if $u \le v \le s \le t$, then

$$\langle S(v) - S(u), S(t) - S(s)\rangle_2 = \langle J_{uv}, J_{st}\rangle = 0$$

since $J_{uv}(x)J_{st}(x) = 0$ for all x. Thus S has orthogonal increments. Furthermore, by (23),

$$\mathbb{E}\big(|S(v) - S(u)|^2\big) = \langle J_{uv}, J_{uv}\rangle = \int_{(u,v]} dF(\lambda) = F(v) - F(u)$$

since F is right-continuous; this confirms (6), and it remains to check that (5) holds.

The process S satisfies conditions (7)–(9), and it follows that the stochastic integral

$$I(\psi) = \int_{(-\pi,\pi]} \psi(\lambda)\, dS(\lambda)$$

is defined for a broad class of functions $\psi : (-\pi, \pi] \to \mathbb{C}$. We claim that

(25) $$I(\psi) = \mu(\psi) \quad \text{(almost surely)} \quad \text{for } \psi \in \overline{H}_F.$$

The result of the theorem will follow immediately by the choice $\psi = f_n$, for which (25) implies that (almost surely) $I(f_n) = \mu(f_n) = X_n$, which is to say that

$$\int_{(-\pi,\pi]} e^{in\lambda}\, dS(\lambda) = X_n$$

as required.

It remains to prove (25), which we do by systematic approximation. Suppose first that ψ is a step function,

(26) $$\psi(x) = \begin{cases} 0 & \text{if } x < a_1 \text{ or } x \ge a_n, \\ c_j & \text{if } a_j \le x < a_{j+1}, \end{cases}$$

where $-\pi < a_1 < a_2 < \cdots < a_n \le \pi$ and $c_1, c_2, \ldots, c_n \in \mathbb{C}$. Then

$$\begin{aligned} I(\psi) &= \sum_{j=1}^{n} c_j[S(a_{j+1}) - S(a_j)] = \sum_{j=1}^{n} c_j \mu(J_{a_j,a_{j+1}}) \quad \text{by (24)} \\ &= \mu\Big(\sum_{j=1}^{n} c_j J_{a_j,a_{j+1}}\Big) = \mu(\psi) \qquad\qquad \text{by (26).} \end{aligned}$$

Hence $I(\psi) = \mu(\psi)$ for all step functions ψ. More generally, if $\psi \in \overline{H}_F$ and $\{\psi_n\}$ is a sequence of step functions converging to ψ, then $\mu(\psi_n) \to \mu(\psi)$. By the definition of the stochastic integral, it is the case that $I(\psi_n) \to I(\psi)$, and it follows that $I(\psi) = \mu(\psi)$, which proves (25). ■

Exercises for Section 9.4

1. Let S be the spectral process of a stationary process X with zero mean and unit variance. Show that the increments of S have zero means.

2. Moving average representation. Let X be a discrete-time stationary process having zero means, continuous strictly positive spectral density function f, and with spectral process S. Let

$$Y_n = \int_{(-\pi,\pi]} \frac{e^{in\lambda}}{\sqrt{2\pi f(\lambda)}}\, dS(\lambda).$$

Show that $\ldots, Y_{-1}, Y_0, Y_1, \ldots$ is a sequence of uncorrelated random variables with zero means and unit variances.

Show that X_n may be represented as a moving average $X_n = \sum_{j=-\infty}^{\infty} a_j Y_{n-j}$ where the a_j are constants satisfying

$$\sqrt{2\pi f(\lambda)} = \sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda} \qquad \text{for } \lambda \in (-\pi, \pi].$$

3. Gaussian process. Let X be a discrete-time stationary sequence with zero mean and unit variance, and whose fdds are of the multivariate-normal type. Show that the spectral process of X has independent increments having normal distributions.

9.5 The ergodic theorem

The law of large numbers asserts that

$$\frac{1}{n}\sum_{j=1}^{n} X_j \to \mu \tag{1}$$

whenever $\{X_j\}$ is an independent identically distributed sequence with mean μ; the convergence takes place almost surely. This section is devoted to a complete generalization of the law of large numbers, the assumption that the X_j be independent being replaced by the assumption that they form a stationary process. This generalization is called the 'ergodic theorem' and it has more than one form depending on the type of stationarity—weak or strong—and the required mode of convergence; recall the various corresponding forms of the law of large numbers†.

It is usual to state the ergodic theorem for discrete-time processes, and we conform to this habit here. Similar results hold for continuous-time processes, sums of the form $\sum_1^n X_j$ being replaced by integrals of the form $\int_0^n X(t)\, dt$. Here is the usual form of the ergodic theorem.

(2) Theorem. Ergodic theorem for strongly stationary processes. *Let $X = \{X_n : n \geq 1\}$ be a strongly stationary process such that $\mathbb{E}|X_1| < \infty$. There exists a random variable Y with the same mean as the X_n such that*

$$\frac{1}{n}\sum_{j=1}^{n} X_j \to Y \quad \textit{a.s. and in mean.}$$

†The original weak ergodic theorem was proved by von Neumann, and the later strong theorem by Birkhoff.

The proof of this is difficult, as befits a complete generalization of the strong law of large numbers (see Problem (9.7.10)). The following result is considerably more elementary.

(3) Theorem. Ergodic theorem for weakly stationary processes. *If $X = \{X_n : n \geq 1\}$ is a (weakly) stationary process, there exists a random variable Y such that $\mathbb{E}Y = \mathbb{E}X_1$ and*

$$\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{\text{m.s.}} Y.$$

We prove the latter theorem first. The normal proof of the 'strong ergodic theorem' (2) is considerably more difficult, and makes use of harder ideas than those required for the 'weak ergodic theorem' (3). The second part of this section is devoted to a discussion of the strong ergodic theorem, together with a relatively straightforward proof.

Theorems (2) and (3) generalize the laws of large numbers. There are similar generalizations of the central limit theorem and the law of the iterated logarithm, although such results hold only for stationary processes which satisfy certain extra conditions. We give no details of this here, save for pointing out that these extra conditions take the form 'X_m and X_n are "nearly independent" when $|m - n|$ is large'.

We give two proofs of (3). Proof A is conceptually easy but has some technical difficulties; we show that $n^{-1}\sum_1^n X_j$ is a mean-square Cauchy-convergent sequence (see Problem (7.11.11)). Proof B uses the spectral representation of X; we sketch this here and show that it yields an explicit form for the limit Y as the contribution made towards X by 'oscillations of zero frequency'.

Proof A. Recall from (7.11.11) that a sequence $\{Y_n\}$ converges in mean square to some limit if and only if $\{Y_n\}$ is *mean-square Cauchy convergent*, which is to say that

$$\mathbb{E}\big(|Y_n - Y_m|^2\big) \to 0 \quad \text{as} \quad m, n \to \infty. \tag{4}$$

A similar result holds for complex-valued sequences. We shall show that the sequence $\{n^{-1}\sum_1^n X_j\}$ satisfies (4) whenever X is stationary. This is easy in concept, since it involves expressions involving the autocovariance function of X alone; the proof of the mean-square version of the law of large numbers was easy for the same reason. Unfortunately, the verification of (4) is not a trivial calculation.

For any complex-valued random variable Z, define

$$\|Z\| = \sqrt{\mathbb{E}(|Z|^2)};$$

the function $\|\cdot\|$ is a norm (see Section 7.2) when viewed as a function on the collection of equivalence classes of random variables with finite second moment and with $Y \sim Z$ if $\mathbb{P}(Y = Z) = 1$. We wish to show that

$$\|\langle X\rangle_n - \langle X\rangle_m\| \to 0 \quad \text{as} \quad n, m \to \infty \tag{5}$$

where

$$\langle X\rangle_n = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

[Physicists often use the notation $\langle \cdot \rangle$ to denote expectation.] Set

$$\mu_N = \inf_{\lambda} \|\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_N X_N\|$$

where the infimum is calculated over all vectors $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ containing non-negative entries with sum 1. Clearly $\mu_N \geq \mu_{N+1}$ and so

$$\mu = \lim_{N\to\infty} \mu_N = \inf_N \mu_N$$

exists. If $m < n$ then

$$\|\langle X\rangle_n + \langle X\rangle_m\| = 2\left\|\sum_{j=1}^{n} \lambda_j X_j\right\|$$

where

$$\lambda_j = \begin{cases} \dfrac{1}{2}\left(\dfrac{1}{m} + \dfrac{1}{n}\right) & \text{if } 1 \leq j \leq m, \\ \dfrac{1}{2n} & \text{if } m < j \leq n, \end{cases}$$

and so

$$\|\langle X\rangle_n + \langle X\rangle_m\| \geq 2\mu.$$

It is not difficult to deduce (see Exercise (7.1.2b) for the first line here) that

$$\begin{aligned} \|\langle X\rangle_n - \langle X\rangle_m\|^2 &= 2\|\langle X\rangle_n\|^2 + 2\|\langle X\rangle_m\|^2 - \|\langle X\rangle_n + \langle X\rangle_m\|^2 \\ &\leq 2\|\langle X\rangle_n\|^2 + 2\|\langle X\rangle_m\|^2 - 4\mu^2 \\ &= 2\left|\|\langle X\rangle_n\|^2 - \mu^2\right| + 2\left|\|\langle X\rangle_m\|^2 - \mu^2\right| \end{aligned}$$

and (5) follows as soon as we can show that

(6) $$\|\langle X\rangle_n\| \to \mu \quad \text{as} \quad n \to \infty.$$

The remaining part of the proof is devoted to demonstrating (6).

Choose any $\epsilon > 0$ and pick N and $\boldsymbol{\lambda}$ such that

$$\|\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_N X_N\| \leq \mu + \epsilon$$

where $\lambda_i \geq 0$ and $\sum_1^N \lambda_i = 1$. Define the moving average

$$Y_k = \lambda_1 X_k + \lambda_2 X_{k+1} + \cdots + \lambda_N X_{k+N-1};$$

it is not difficult to see that $Y = \{Y_k\}$ is a stationary process (see Problem (8.7.1)). We shall show that

(7) $$\|\langle Y\rangle_n - \langle X\rangle_n\| \to 0 \quad \text{as} \quad n \to \infty$$

where

$$\langle Y\rangle_n = \frac{1}{n}\sum_{j=1}^{n} Y_j.$$

Note first that, by the triangle inequality (7.1.5),

$$\|\langle Y\rangle_n\| \le \|Y_1\| \le \mu + \epsilon \quad \text{for all } n \tag{8}$$

since $\|Y_n\| = \|Y_1\|$ for all n. Now

$$\langle Y\rangle_n = \lambda_1\langle X\rangle_{1,n} + \lambda_2\langle X\rangle_{2,n} + \cdots + \lambda_N\langle X\rangle_{N,n}$$

where

$$\langle X\rangle_{k,n} = \frac{1}{n}\sum_{j=k}^{k+n-1} X_j;$$

now use the facts that $\langle X\rangle_{1,n} = \langle X\rangle_n$, $1-\lambda_1 = \lambda_2+\lambda_3+\cdots+\lambda_N$, and the triangle inequality to deduce that

$$\|\langle Y\rangle_n - \langle X\rangle_n\| \le \sum_{j=2}^{N} \lambda_j \|\langle X\rangle_{j,n} - \langle X\rangle_{1,n}\|.$$

However, by the triangle inequality again,

$$\begin{aligned}\|\langle X\rangle_{j,n} - \langle X\rangle_{1,n}\| &= \frac{1}{n}\big\|(X_j + \cdots + X_{j+n-1}) - (X_1 + \cdots + X_n)\big\| \\ &= \frac{1}{n}\big\|(X_{n+1} + \cdots + X_{j+n-1}) - (X_1 + \cdots + X_{j-1})\big\| \\ &\le \frac{2j}{n}\|X_1\|\end{aligned}$$

since $\|X_n\| = \|X_1\|$ for all n. Therefore,

$$\|\langle Y\rangle_n - \langle X\rangle_n\| \le \sum_{j=2}^{N} \lambda_j \frac{2j}{n}\|X_1\| \le \frac{2N}{n}\|X_1\|;$$

let $n \to \infty$ to deduce that (7) holds. Use (8) to obtain

$$\begin{aligned}\mu \le \|\langle X\rangle_n\| &\le \|\langle X\rangle_n - \langle Y\rangle_n\| + \|\langle Y\rangle_n\| \\ &\le \|\langle X\rangle_n - \langle Y\rangle_n\| + \mu + \epsilon \to \mu + \epsilon \quad \text{as} \quad n \to \infty.\end{aligned}$$

Now ϵ was arbitrary, and we let $\epsilon \downarrow 0$ to obtain (6).

Since $\langle X\rangle_n \xrightarrow{\text{m.s.}} Y$, we have that $\langle X\rangle_n \xrightarrow{1} Y$, which implies that $\mathbb{E}\langle X\rangle_n \to \mathbb{E}Y$. However, $\mathbb{E}\langle X\rangle_n = \mathbb{E}X_1$, whence $\mathbb{E}Y = \mathbb{E}X_1$. ■

Sketch proof B. Suppose that $\mathbb{E}(X_n) = 0$ for all n. The process X has a spectral representation

$$X_n = \int_{(-\pi,\pi]} e^{in\lambda}\, dS(\lambda).$$

Now,

$$\langle X\rangle_n = \frac{1}{n}\sum_{j=1}^{n} X_j = \int_{(-\pi,\pi]} \frac{1}{n}\sum_{j=1}^{n} e^{ij\lambda}\, dS(\lambda) = \int_{(-\pi,\pi]} g_n(\lambda)\, dS(\lambda) \tag{9}$$

where

$$g_n(\lambda) = \begin{cases} 1 & \text{if } \lambda = 0, \\ \dfrac{e^{i\lambda}}{n}\dfrac{1-e^{in\lambda}}{1-e^{i\lambda}} & \text{if } \lambda \neq 0. \end{cases} \tag{10}$$

We have that $|g_n(\lambda)| \leq 1$ for all n and λ, and, as $n \to \infty$,

$$g_n(\lambda) \to g(\lambda) = \begin{cases} 1 & \text{if } \lambda = 0, \\ 0 & \text{if } \lambda \neq 0. \end{cases} \tag{11}$$

It can be shown that

$$\int_{(-\pi,\pi]} g_n(\lambda)\, dS(\lambda) \xrightarrow{\text{m.s.}} \int_{(-\pi,\pi]} g(\lambda)\, dS(\lambda) \quad \text{as} \quad n \to \infty,$$

implying that

$$\langle X \rangle_n \xrightarrow{\text{m.s.}} \int_{(-\pi,\pi]} g(\lambda)\, dS(\lambda) = S(0) - S(0-),$$

by the right-continuity of S, where $S(0-) = \lim_{y \uparrow 0} S(y)$. This shows that $\langle X \rangle_n$ converges in mean square to the random magnitude of the discontinuity of $S(\lambda)$ at $\lambda = 0$ (this quantity may be zero); in other words, $\langle X \rangle_n$ converges to the 'zero frequency' or 'infinite wavelength' contribution of the spectrum of X. This conclusion is natural and memorable, since the average of any oscillation having non-zero frequency is zero. ■

The second proof of Theorem (3) is particularly useful in that it provides an explicit representation for the limit in terms of the spectral process of X. It is easy to calculate the first two moments of this limit.

(12) Lemma. *If X is a stationary process with zero means and autocovariance function $c(m)$ then the limit variable $Y = \lim_{n\to\infty}\{n^{-1}\sum_{j=1}^{n} X_j\}$ satisfies*

$$\mathbb{E}(Y) = 0, \quad \mathbb{E}(|Y|^2) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} c(j).$$

A similar result holds for processes with non-zero means.

Proof. We have that $\langle X \rangle_n \xrightarrow{\text{m.s.}} Y$, and so $\langle X \rangle_n \xrightarrow{1} Y$ by Theorem (7.2.3). The result of Exercise (7.2.1) implies that $\mathbb{E}(\langle X \rangle_n) \to \mathbb{E}(Y)$ as $n \to \infty$; however, $\mathbb{E}(\langle X \rangle_n) = \mathbb{E}(X_1) = 0$ for all n.

In order to prove the second part, either use Exercise (7.2.1) again and expand $\mathbb{E}(\langle X \rangle_n^2)$ in terms of c (see Exercise (2)), or use the method of Proof B of (3). We use the latter method. The autocovariance function $c(m)$ satisfies

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n} c(j) &= c(0) \int_{(-\pi,\pi]} g_n(\lambda)\, dF(\lambda) \\ &\to c(0) \int_{(-\pi,\pi]} g(\lambda)\, dF(\lambda) \quad \text{as} \quad n \to \infty \\ &= c(0)[F(0) - F(0-)] \end{aligned}$$

where g_n and g are given by (10) and (11), F is the spectral distribution function, and $F(0-) = \lim_{y \uparrow 0} F(y)$ as usual. We can now use (9.4.6) and the continuity properties of S to show that

$$c(0)[F(0) - F(0-)] = \mathbb{E}\big(|S(0) - S(0-)|^2\big) = \mathbb{E}(|Y|^2). \qquad \blacksquare$$

We turn now to the strong ergodic theorem (2), which we shall first rephrase slightly. Here is some terminology and general discussion.

A vector $\mathbf{X} = (X_1, X_2, \dots)$ of real-valued random variables takes values in the set of real vectors of the form $\mathbf{x} = (x_1, x_2, \dots)$. We write $\mathbb{R}^T$ for the set of all such real sequences, where T denotes the set $\{1, 2, \dots\}$ of positive integers. The natural σ-field for $\mathbb{R}^T$ is the product $\mathcal{B}^T$ of the appropriate number of copies of the Borel σ-field $\mathcal{B}$ of subsets of $\mathbb{R}$. Let $\mathbb{Q}$ be a probability measure on the pair $(\mathbb{R}^T, \mathcal{B}^T)$. The triple $(\mathbb{R}^T, \mathcal{B}^T, \mathbb{Q})$ is our basic probability space, and we make the following crucial definitions.

There is a natural 'shift operator' τ mapping $\mathbb{R}^T$ onto itself, defined by $\tau(\mathbf{x}) = \mathbf{x}'$ where $\mathbf{x}' = (x_2, x_3, \dots)$; that is, the vector $\mathbf{x} = (x_1, x_2, \dots)$ is mapped to the vector $(x_2, x_3, \dots)$. The measure $\mathbb{Q}$ is called *stationary* if and only if $\mathbb{Q}(A) = \mathbb{Q}(\tau^{-1}A)$ for all $A \in \mathcal{B}^T$ (remember that $\tau^{-1}A = \{\mathbf{x} \in \mathbb{R}^T : \tau(\mathbf{x}) \in A\}$). If $\mathbb{Q}$ is stationary, we call the shift τ 'measure preserving'. Stationary measures correspond to strongly stationary sequences of random variables, as the following example indicates.

(13) Example. Let $\mathbf{X} = (X_1, X_2, \dots)$ be a strongly stationary sequence on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define the probability measure $\mathbb{Q}$ on $(\mathbb{R}^T, \mathcal{B}^T)$ by $\mathbb{Q}(A) = \mathbb{P}(\mathbf{X} \in A)$ for $A \in \mathcal{B}^T$. Now $\mathbf{X}$ and $\tau(\mathbf{X})$ have the same fdds, and therefore

$$\mathbb{Q}(\tau^{-1}A) = \mathbb{P}\big(\tau(\mathbf{X}) \in A\big) = \mathbb{P}(\mathbf{X} \in A) = \mathbb{Q}(A)$$

for all (measurable) subsets A of $\mathbb{R}^T$.

We have seen that every strongly stationary sequence generates a stationary measure on $(\mathbb{R}^T, \mathcal{B}^T)$. The converse is true also. Let $\mathbb{Q}$ be a stationary measure on $(\mathbb{R}^T, \mathcal{B}^T)$, and define the sequence $\mathbf{Y} = (Y_1, Y_2, \dots)$ of random variables by $Y_n(\mathbf{x}) = x_n$, the nth component of the real vector $\mathbf{x}$. We have from the stationarity of $\mathbb{Q}$ that, for $A \in \mathcal{B}^T$,

$$\mathbb{Q}(\mathbf{Y} \in A) = \mathbb{Q}(A) = \mathbb{Q}(\tau^{-1}A) = \mathbb{Q}\big(\tau(\mathbf{Y}) \in A\big)$$

so that $\mathbf{Y}$ and $\tau(\mathbf{Y})$ have the same fdds. Hence $\mathbf{Y}$ is a strongly stationary sequence. ●

There is a certain special class of events in $\mathcal{B}^T$ called *invariant* events.

(14) Definition. An event A in $\mathcal{B}^T$ is called **invariant** if $A = \tau^{-1}A$.

An event A is invariant if

(15) $$\mathbf{x} \in A \quad \text{if and only if} \quad \tau(\mathbf{x}) \in A,$$

for any $\mathbf{x} \in \mathbb{R}^T$. Now (15) is equivalent to the statement '$\mathbf{x} \in A$ if and only if $\tau^n(\mathbf{x}) \in A$ for all $n \geq 0$'; remembering that $\tau^n(\mathbf{x}) = (x_{n+1}, x_{n+2}, \dots)$, we see therefore that the membership by $\mathbf{x}$ of an invariant event A does not depend on any finite collection of the components of $\mathbf{x}$. Here are some examples of invariant events:

$$A_1 = \left\{\mathbf{x} : \limsup_{n \to \infty} x_n \leq 3\right\},$$

$$A_2 = \{\mathbf{x} : \text{the sequence } n^{-1}x_n \text{ converges}\},$$

$$A_3 = \{\mathbf{x} : x_n = 0 \text{ for all large } n\}.$$

We denote by $\mathcal{I}$ the set of all invariant events. It is not difficult to see (Exercise (1)) that $\mathcal{I}$ is a σ-field, and therefore $\mathcal{I}$ is a sub-σ-field of $\mathcal{B}^T$, called the *invariant σ-field*.

Finally, we need the idea of conditional expectation. Let U be a random variable on $(\mathbb{R}^T, \mathcal{B}^T, \mathbb{Q})$ with finite mean $\mathbb{E}(U)$; here, $\mathbb{E}$ denotes expectation with respect to the measure $\mathbb{Q}$. We saw in Theorem (7.9.26) that there exists an $\mathcal{I}$-measurable random variable Z such that $\mathbb{E}|Z| < \infty$ and $\mathbb{E}((U - Z)I_G) = 0$ for all $G \in \mathcal{I}$; Z is usually denoted by $Z = \mathbb{E}(U \mid \mathcal{I})$ and is called the conditional expectation of U given $\mathcal{I}$.

We are now ready to restate the strong ergodic theorem (2) in the following way.

(16) Ergodic theorem. *Let $\mathbb{Q}$ be a stationary probability measure on $(\mathbb{R}^T, \mathcal{B}^T)$, and let Y be a real-valued random variable on the space $(\mathbb{R}^T, \mathcal{B}^T, \mathbb{Q})$. Let $Y_1, Y_2, \dots$ be the sequence of random variables defined by*

$$\textbf{(17)} \qquad Y_i(\mathbf{x}) = Y(\tau^{i-1}(\mathbf{x})) \quad \textit{for} \quad \mathbf{x} \in \mathbb{R}^T.$$

If Y has finite mean, then

$$\textbf{(18)} \qquad \frac{1}{n}\sum_{i=1}^{n} Y_i \to \mathbb{E}(Y \mid \mathcal{I}) \quad \textit{a.s. and in mean.}$$

The sequence $\mathbf{Y} = (Y_1, Y_2, \dots)$ is of course strongly stationary: since $\mathbb{Q}$ is stationary,

$$\mathbb{Q}\big((Y_2, Y_3, \dots) \in A\big) = \mathbb{Q}\big(\tau(\mathbf{Y}) \in A\big) = \mathbb{Q}(\mathbf{Y} \in \tau^{-1}A) = \mathbb{Q}(\mathbf{Y} \in A) \quad \text{for } A \in \mathcal{B}^T.$$

The above theorem asserts that the average of the first n values of $\mathbf{Y}$ converges as $n \to \infty$, the limit being the conditional mean of Y given $\mathcal{I}$; this is a conclusion very similar to that of the strong law of large numbers (7.5.1).

To understand the relationship between Theorems (2) and (16), consider the situation treated by (2). Let $X_1, X_2, \dots$ be a strongly stationary sequence on $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathbb{Q}$ be the stationary measure on $(\mathbb{R}^T, \mathcal{B}^T)$ defined by $\mathbb{Q}(A) = \mathbb{P}(\mathbf{X} \in A)$ for $A \in \mathcal{B}^T$. We define $Y : \mathbb{R}^T \to \mathbb{R}$ by $Y(\mathbf{x}) = x_1$ for $\mathbf{x} = (x_1, x_2, \dots) \in \mathbb{R}^T$, so that Y_i in (17) is given by $Y_i(\mathbf{x}) = x_i$. It is clear that the sequences $\{X_n : n \geq 1\}$ and $\{Y_n : n \geq 1\}$ have the same joint distributions, and it follows that the convergence of $n^{-1}\sum_1^n Y_i$ entails the convergence of $n^{-1}\sum_1^n X_i$.

(19) Definition. The stationary measure $\mathbb{Q}$ on $(\mathbb{R}^T, \mathcal{B}^T)$ is called **ergodic** if each invariant event has probability either 0 or 1, which is to say that $\mathbb{Q}(A) = 0$ or 1 for all $A \in \mathcal{I}$.

Ergodic stationary measures are of particular importance. The simplest example of such a measure is product measure.

(20) Example. Independent sequences. Let $\mathbb{S}$ be a probability measure on $(\mathbb{R}, \mathcal{B})$, and let $\mathbb{Q} = \mathbb{S}^T$, the appropriate product measure on $(\mathbb{R}^T, \mathcal{B}^T)$. Product measures arise in the context of independent random variables, as follows. Let $X_1, X_2, \dots$ be a sequence of independent identically idstributed random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathbb{S}(A) = \mathbb{P}(X_1 \in A)$ for $A \in \mathcal{B}$. Then $\mathbb{S}$ is a probability measure on $(\mathbb{R}, \mathcal{B})$. The probability space $(\mathbb{R}^T, \mathcal{B}^T, \mathbb{S}^T)$ is the natural space for the vector $\mathbf{X} = (X_1, X_2, \dots)$; that is, $\mathbb{S}^T(A) = \mathbb{P}(\mathbf{X} \in A)$ for $A \in \mathcal{B}^T$.

Suppose that A ($\in \mathcal{B}^T$) is invariant. Then, for all n, A belongs to the σ-field generated by the subsequence $(X_n, X_{n+1}, \dots)$, and hence A belongs to the tail σ-field of the X_i. By Kolmogorov's zero–one law (7.3.15), the latter σ-field is trivial, in that all events therein have probability either 0 or 1. Hence all invariant events have probability either 0 or 1, and therefore the measure $\mathbb{S}^T$ is ergodic. ●

The conclusion (18) of the ergodic theorem takes on a particularly simple form when the measure $\mathbb{Q}$ is ergodic as well as stationary. In this case, the random variable $\mathbb{E}(Y \mid \mathcal{I})$ is (a.s.) constant, as the following argument demonstrates. The conditional expectation $\mathbb{E}(Y \mid \mathcal{I})$ is $\mathcal{I}$-measurable, and therefore the event $A_y = \{\mathbb{E}(Y \mid \mathcal{I}) \le y\}$ belongs to $\mathcal{I}$ for all y. However, $\mathcal{I}$ is trivial, in that it contains only events having probability 0 or 1. Hence $\mathbb{E}(Y \mid \mathcal{I})$ takes almost surely the value $\sup\{y : \mathbb{Q}(A_y) = 0\}$. Taking expectations, we find that this value is $\mathbb{E}(Y)$, so that the conclusion (18) becomes

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to \mathbb{E}(Y) \quad \text{a.s. and in mean} \tag{21}$$

in the ergodic case.

Proof of ergodic theorem (16). We give full details of this for the case when $\mathbb{Q}$ is ergodic, and finish the proof with brief notes describing how to adapt the argument to the general case.

Assume then that $\mathbb{Q}$ is ergodic, so that $\mathbb{E}(Y \mid \mathcal{I}) = \mathbb{E}(Y)$. First we prove almost-sure convergence, which is to say that

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to \mathbb{E}(Y) \quad \text{a.s.} \tag{22}$$

It suffices to prove that

$$\text{if} \quad \mathbb{E}(Y) < 0 \quad \text{then} \quad \limsup_{n\to\infty}\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i\right\} \le 0 \quad \text{a.s.} \tag{23}$$

To see that (23) suffices, we argue as follows. Suppose that (23) holds, and that Z is a (measurable) function on $(\mathbb{R}^T, \mathcal{B}^T, \mathbb{Q})$ with finite mean, and let $\epsilon > 0$. then $Y' = Z - \mathbb{E}(Z) - \epsilon$ and $Y'' = -Z + \mathbb{E}(Z) - \epsilon$ have negative means. Applying (23) to Y' and Y'' we obtain

$$\mathbb{E}(Z) - \epsilon \le \liminf_{n\to\infty}\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i\right\} \le \limsup_{n\to\infty}\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i\right\} \le \mathbb{E}(Z) + \epsilon \quad \text{a.s.},$$

where Z_i is the random variable given by $Z_i(\mathbf{x}) = Z(\tau^{i-1}(\mathbf{x}))$. These inequalities hold for all $\epsilon > 0$, and therefore

$$\liminf_{n\to\infty}\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i\right\} = \limsup_{n\to\infty}\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i\right\} = \mathbb{E}(Z) \quad \text{a.s.}$$

as required for almost-sure convergence.

Turning to the proof of (23), suppose that $\mathbb{E}(Y) < 0$, and introduce the notation $S_n = \sum_{i=1}^n Y_i$. Now $S_n \le M_n$ where $M_n = \max\{0, S_1, S_2, \dots, S_n\}$ satisfies $M_n \le M_{n+1}$. Hence $S_n \le M_\infty$ where $M_\infty = \lim_{n\to\infty} M_n$. Therefore

$$\limsup_{n\to\infty} \left\{\frac{1}{n} S_n\right\} \le \limsup_{n\to\infty} \left\{\frac{1}{n} M_\infty\right\}, \tag{24}$$

and (23) will be proved once we know that $M_\infty < \infty$ a.s. It is easily seen that the event $\{M_\infty < \infty\}$ is an invariant event, and hence has probability either 0 or 1; it is here that we use the hypothesis that $\mathbb{Q}$ is ergodic. We must show that $\mathbb{Q}(M_\infty < \infty) = 1$, and to this end we assume the contrary, that $\mathbb{Q}(M_\infty = \infty) = 1$.

Now,

$$\begin{aligned} M_{n+1} &= \max\{0, S_1, S_2, \dots, S_{n+1}\} \\ &= \max\big\{0, S_1 + \max\{0, S_2 - S_1, \dots, S_{n+1} - S_1\}\big\} \\ &= \max\{0, S_1 + M_n'\} \end{aligned} \tag{25}$$

where $M_n' = \max\{0, S_1', S_2', \dots, S_n'\}$, and $S_j' = \sum_{i=1}^j Y_{i+1}$. It follows from (25) that

$$M_{n+1} = M_n' + \max\{-M_n', Y\},$$

since $S_1 = Y$. Taking expectations and using the fact that $\mathbb{E}(M_n') = \mathbb{E}(M_n)$, we find that

$$0 \le \mathbb{E}(M_{n+1}) - \mathbb{E}(M_n) = \mathbb{E}\big(\max\{-M_n', Y\}\big). \tag{26}$$

If $M_n \uparrow \infty$ a.s. then $M_n' \uparrow \infty$ a.s., implying that $\max\{-M_n', Y\} \downarrow Y$ a.s. It follows by (26) (and dominated convergence) that $0 \le \mathbb{E}(Y)$ in contradiction of the assumption that $\mathbb{E}(Y) < 0$. Our initial hypothesis was therefore false, which is to say that $\mathbb{Q}(M_\infty < \infty) = 1$, and (23) is proved.

Having proved almost-sure convergence, convergence in mean will follow by Theorem (7.10.3) once we have proved that the family $\{n^{-1}S_n : n \ge 1\}$ is uniformly integrable. The random variables $Y_1, Y_2, \dots$ are identically distributed with finite mean; hence (see Exercise (5.6.5)) for any $\epsilon > 0$, there exists $\delta > 0$ such that, for all i,

$$\mathbb{E}\big(|Y_i| I_A\big) < \epsilon \quad \text{for all } A \text{ satisfying } \mathbb{Q}(A) < \delta. \tag{27}$$

Hence, for all n,

$$\mathbb{E}\big(|n^{-1}S_n| I_A\big) \le \frac{1}{n} \sum_{i=1}^n \mathbb{E}\big(|Y_i| I_A\big) < \epsilon$$

whenever $\mathbb{Q}(A) < \delta$. We deduce by an appeal to Lemma (7.10.6) that $\{n^{-1}S_n : n \ge 1\}$ is a uniformly integrable family as required.

This completes the proof in the ergodic case. The proof is only slightly more complicated in the general case, and here is a sketch of the additional steps required.

1. Use the definition of $\mathcal{I}$ to show that $\mathbb{E}(Y \mid \mathcal{I}) = \mathbb{E}(Y_i \mid \mathcal{I})$ for all i.
2. Replace (23) by the following statement: on the event $\{\mathbb{E}(Y \mid \mathcal{I}) < 0\}$, we have that
$$\limsup_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} Y_i \right\} \leq 0$$
except possibly for an event of probability 0. Check that this is sufficient for the required result by applying it to the random variables
$$Y' = Z - \mathbb{E}(Z \mid \mathcal{I}) - \epsilon, \quad Y'' = -Z + \mathbb{E}(Z \mid \mathcal{I}) - \epsilon,$$
where $\epsilon > 0$.
3. Moving to (26), prove that $\mathbb{E}(M_n' \mid \mathcal{I}) = \mathbb{E}(M_n \mid \mathcal{I})$, and deduce the inequality $\mathbb{E}\big(\max\{-M_n', Y\} \,\big|\, \mathcal{I}\big) \geq 0$.
4. Continuing from (26), show that $\{M_n \to \infty\} = \{M_n' \to \infty\}$, and deduce the inequality $\mathbb{E}(Y \mid \mathcal{I}) \geq 0$ on the event $\{M_n \to \infty\}$. This leads us to the same contradiction as in the ergodic case, and we conclude the proof as before. ■

Here are some applications of the ergodic theorem.

(28) Example. Markov chains. Let $X = \{X_n\}$ be an irreducible ergodic Markov chain with countable state space S, and let $\boldsymbol{\pi}$ be the unique stationary distribution of the chain. Suppose that $X(0)$ has distribution $\boldsymbol{\pi}$; the argument of Example (8.2.4) shows that X is strongly stationary. Choose some state k and define the collection $I = \{I_n : n \geq 0\}$ of indicator functions by
$$I_n = \begin{cases} 1 & \text{if } X_n = k, \\ 0 & \text{otherwise.} \end{cases}$$
Clearly I is strongly stationary. It has autocovariance function
$$c(n, n+m) = \operatorname{cov}(I_n, I_{n+m}) = \pi_k[p_{kk}(m) - \pi_k], \quad m \geq 0,$$
where $p_{kk}(m) = \mathbb{P}(X_m = k \mid X_0 = k)$. The partial sum $S_n = \sum_{j=0}^{n-1} I_j$ is the number of visits to the state k before the nth jump, and a short calculation gives
$$\frac{1}{n}\mathbb{E}(S_n) = \pi_k \quad \text{for all } n.$$
It is a consequence of the ergodic theorem (2) that
$$\frac{1}{n} S_n \xrightarrow{\text{a.s.}} S \quad \text{as} \quad n \to \infty,$$
where S is a random variable with mean $\mathbb{E}(S) = \mathbb{E}(I_0) = \pi_k$. Actually S is constant in that $\mathbb{P}(S = \pi_k) = 1$; just note that $c(n, n+m) \to 0$ as $m \to \infty$ and use the result of Problem (9.7.9). ●

(29) Example. Binary expansion. Let X be uniformly distributed on $[0, 1]$. The random number X has a binary expansion
$$X = 0 \cdot X_1 X_2 \cdots = \sum_{j=1}^{\infty} X_j 2^{-j}$$

where $X_1, X_2, \ldots$ is a sequence of independent identically distributed random variables, each taking one of the values 0 or 1 with probability $\frac{1}{2}$ (see Problem (7.11.4)). Define

$$Y_n = 0 \cdot X_n X_{n+1} \cdots \quad \text{for} \quad n \geq 1 \tag{30}$$

and check for yourself that $Y = \{Y_n : n \geq 1\}$ is strongly stationary. Use (2) to see that

$$\frac{1}{n} \sum_{j=1}^{n} Y_j \xrightarrow{\text{a.s.}} \frac{1}{2} \quad \text{as} \quad n \to \infty.$$

Generalize this example as follows. Let $g : \mathbb{R} \to \mathbb{R}$ be such that:

(a) g has period 1, so that $g(x + 1) = g(x)$ for all x,

(b) g is uniformly continuous and integrable over $[0, 1]$,

and define $Z = \{Z_n : n \geq 1\}$ by $Z_n = g(2^{n-1}X)$ where X is uniform on $[0, 1]$ as before. The process Y, above, may be constructed in this way by choosing $g(x) = x$ modulo 1. Check for yourself that Z is strongly stationary, and deduce that

$$\frac{1}{n} \sum_{j=1}^{n} g(2^{j-1}X) \xrightarrow{\text{a.s.}} \int_0^1 g(x)\, dx \quad \text{as} \quad n \to \infty.$$

Can you adapt this example to show that

$$\frac{1}{n} \sum_{j=1}^{n} g\big(X + (j-1)\pi\big) \xrightarrow{\text{a.s.}} \int_0^1 g(x)\, dx \quad \text{as} \quad n \to \infty$$

for any fixed positive irrational number π? ●

(31) Example. Range of random walk. Let $X_1, X_2, \ldots$ be independent identically distributed random variables taking integer values, and let $S_n = X_1 + X_2 + \cdots + X_n$; think of S_n as being the position of a random walk after n steps. Let R_n be the *range* of the walk up to time n, which is to say that R_n is the number of distinct values taken by the sequence $S_1, S_2, \ldots, S_n$. It was proved by elementary means in Problem (3.11.27) that

$$\frac{1}{n}\mathbb{E}(R_n) \to \mathbb{P}(\text{no return}) \quad \text{as} \quad n \to \infty \tag{32}$$

where the event $\{\text{no return}\} = \{S_k \neq 0 \text{ for all } k \geq 1\}$ is the event that the walk never revisits its starting point $S_0 = 0$.

Of more interest than (32) is the fact that

$$\frac{1}{n} R_n \xrightarrow{\text{a.s.}} \mathbb{P}(\text{no return}), \tag{33}$$

and we shall prove this with the aid of the ergodic theorem (16).

First, let N be a positive integer, and let Z_k be the number of distinct points visited by $S_{(k-1)N+1}, S_{(k-1)N+2}, \ldots, S_{kN}$; clearly $Z_1, Z_2, \ldots$ are independent identically distributed

variables. Now, if $KN \le n < (K+1)N$, then $|R_n - R_{KN}| \le N$ and $R_{KN} \le Z_1 + Z_2 + \cdots + Z_K$. Therefore

$$\frac{1}{n}R_n \le \frac{1}{KN}(R_{KN} + N) \le \frac{1}{KN}(Z_1 + Z_2 + \cdots + Z_K) + \frac{1}{K}$$
$$\xrightarrow{\text{a.s.}} \frac{1}{N}\mathbb{E}(Z_1) \quad \text{as} \quad K \to \infty$$

by the strong law of large numbers. It is easily seen that $Z_1 = R_N$, and therefore, almost surely,

$$\limsup_{n\to\infty}\left\{\frac{1}{n}R_n\right\} \le \frac{1}{N}\mathbb{E}(R_N) \to \mathbb{P}(\text{no return}) \tag{34}$$

as $N \to \infty$, by (32). This is the required upper bound.

For the lower bound, we must work a little harder. Let V_k be the indicator function of the event that the position of the walk at time k is not revisited subsequently; that is,

$$V_k = \begin{cases} 1 & \text{if } S_j \neq S_k \text{ for all } j > k, \\ 0 & \text{otherwise.} \end{cases}$$

The collection of points S_k for which $V_k = 1$ is a collection of distinct points, and it follows that

$$R_n \ge V_1 + V_2 + \cdots + V_n. \tag{35}$$

On the other hand, V_k may be represented as $Y(X_{k+1}, X_{k+2}, \dots)$ where $Y : \mathbb{R}^T \to \{0, 1\}$ is defined by

$$Y(x_1, x_2, \dots) = \begin{cases} 1 & \text{if } x_1 + \cdots + x_l \neq 0 \text{ for all } l \ge 1, \\ 0 & \text{otherwise.} \end{cases}$$

The X_j are independent and identically distributed, and therefore Theorem (16) may be applied to deduce that

$$\frac{1}{n}(V_1 + V_2 + \cdots + V_n) \xrightarrow{\text{a.s.}} \mathbb{E}(V_1).$$

Note that $\mathbb{E}(V_1) = \mathbb{P}(\text{no return})$.

It follows from (35) that

$$\liminf_{n\to\infty}\left\{\frac{1}{n}R_n\right\} \ge \mathbb{P}(\text{no return}) \quad \text{a.s.},$$

which may be combined with (34) to obtain the claimed result (33). ●

Exercises for Section 9.5

1. Let $T = \{1, 2, \dots\}$ and let $\mathcal{I}$ be the set of invariant events of $(\mathbb{R}^T, \mathcal{B}^T)$. Show that $\mathcal{I}$ is a σ-field.

2. Assume that $X_1, X_2, \dots$ is a stationary sequence with autocovariance function $c(m)$. Show that

$$\operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{2}{n^2}\sum_{j=1}^{n}\sum_{i=0}^{j-1} c(i) - \frac{c(0)}{n}.$$

Assuming that $j^{-1}\sum_{i=0}^{j-1} c(i) \to \sigma^2$ as $j \to \infty$, show that

$$\operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \to \sigma^2 \qquad \text{as } n \to \infty.$$

3. Let $X_1, X_2, \dots$ be independent identically distributed random variables with zero mean and unit variance. Let

$$Y_n = \sum_{i=0}^{\infty} \alpha_i X_{n+i} \qquad \text{for } n \geq 1$$

where the α_i are constants satisfying $\sum_i \alpha_i^2 < \infty$. Use the martingale convergence theorem to show that the above summation converges almost surely and in mean square. Prove that $n^{-1}\sum_{i=1}^{n} Y_i \to 0$ a.s. and in mean, as $n \to \infty$.

9.6 Gaussian processes

Let $X = \{X(t) : -\infty < t < \infty\}$ be a real-valued stationary process with autocovariance function $c(t)$; in line with Theorem (9.3.2), c is a real-valued function which satisfies:

(a) $c(-t) = c(t)$,

(b) c is a non-negative definite function.

It is not difficult to see that a function $c : \mathbb{R} \to \mathbb{R}$ is the autocovariance function of some real-valued stationary process if and only if c satisfies (a) and (b). Subject to these conditions on c, there is an explicit construction of a corresponding stationary process.

(1) Theorem. *If $c : \mathbb{R} \to \mathbb{R}$ and c satisfies* (a) *and* (b) *above, there exists a real-valued strongly stationary process X with autocovariance function c.*

Proof. We shall construct X by defining its finite-dimensional distributions (fdds) and then using the Kolmogorov consistency conditions (8.6.3). For any vector $\mathbf{t} = (t_1, t_2, \dots, t_n)$ of real numbers with some finite length n, let $F_{\mathbf{t}}$ be the multivariate normal distribution function with zero means and covariance matrix $\mathbf{V} = (v_{jk})$ with entries $v_{jk} = c(t_k - t_j)$ (see Section 4.9).

The family $\{F_{\mathbf{t}} : \mathbf{t} \in \mathbb{R}^n,\ n = 1, 2, \dots\}$ satisfies the Kolmogorov consistency conditions (8.6.3) and so there exists a process X with this family of fdds. It is clear that X is strongly stationary with autocovariance function c. ■

A result similar to (1) holds for complex-valued functions $c : \mathbb{R} \to \mathbb{C}$, (a) being replaced by the property that

(2) $$c(-t) = \overline{c(t)}.$$

We do not explore this here, but choose to consider real-valued processes only. The process X which we have constructed in the foregoing proof is an example of a (real-valued) 'Gaussian process'.

(3) **Definition.** A real-valued continuous-time process X is called a **Gaussian** process if each finite-dimensional vector $(X(t_1), X(t_2), \ldots, X(t_n))$ has the multivariate normal distribution $N(\boldsymbol{\mu}(\mathbf{t}), \mathbf{V}(\mathbf{t}))$ for some mean vector $\boldsymbol{\mu}$ and some covariance matrix $\mathbf{V}$ which may depend on $\mathbf{t} = (t_1, t_2, \ldots, t_n)$.

The $X(t_j)$ may have a singular multivariate normal distribution. We shall often restrict our attention to Gaussian processes with $\mathbb{E}(X(t)) = 0$ for all t; as before, similar results are easily found when this fails to hold.

A Gaussian process is not necessarily stationary.

(4) **Theorem.** *The Gaussian process X is stationary if and only if $\mathbb{E}(X(t))$ is constant for all t and the covariance matrix $\mathbf{V}(\mathbf{t})$ in Definition* (3) *satisfies $\mathbf{V}(\mathbf{t}) = \mathbf{V}(\mathbf{t} + h)$ for all $\mathbf{t}$ and $h > 0$, where $\mathbf{t} + h = (t_1 + h, t_2 + h, \ldots, t_n + h)$.*

Proof. This is an easy *exercise*. ■

It is clear that a Gaussian process is strongly stationary if and only if it is weakly stationary.

Can a Gaussian process be a Markov process? The answer is in the affirmative. First, we must rephrase the Markov property (6.1.1) to deal with processes which take values in the real line.

(5) **Definition.** The continuous-time process X, taking values in $\mathbb{R}$, is called a **Markov process** if the following holds:

$$\text{(6)} \quad \mathbb{P}\big(X(t_n) \leq x \,\big|\, X(t_1) = x_1, \ldots, X(t_{n-1}) = x_{n-1}\big) = \mathbb{P}\big(X(t_n) \leq x \,\big|\, X(t_{n-1}) = x_{n-1}\big)$$

for all $x, x_1, x_2, \ldots, x_{n-1}$, and all increasing sequences $t_1 < t_2 < \cdots < t_n$ of times.

(7) **Theorem.** *The Gaussian process X is a Markov process if and only if*

$$\text{(8)} \quad \mathbb{E}\big(X(t_n) \,\big|\, X(t_1) = x_1, \ldots, X(t_{n-1}) = x_{n-1}\big) = \mathbb{E}\big(X(t_n) \,\big|\, X(t_{n-1}) = x_{n-1}\big)$$

for all $x_1, x_2, \ldots, x_{n-1}$ and all increasing sequences $t_1 < t_2 < \cdots < t_n$ of times.

Proof. It is clear from (5) that (8) holds whenever X is Markov. Conversely, suppose that X is Gaussian and satisfies (8). Both the left- and right-hand sides of (6) are normal distribution functions. Any normal distribution is specified by its mean and variance, and so we need only show that the left- and right-hand sides of (6) have equal first two moments. The equality of the first moments is trivial, since this is simply the assertion of (8). Also, if $1 \leq r < n$, then $\mathbb{E}(Y X_r) = 0$ where

$$\text{(9)} \quad Y = X_n - \mathbb{E}(X_n \mid X_1, \ldots, X_{n-1}) = X_n - \mathbb{E}(X_n \mid X_{n-1})$$

and we have written $X_r = X(t_r)$ for ease of notation; to see this, write

$$\begin{aligned} \mathbb{E}(Y X_r) &= \mathbb{E}\big(X_n X_r - \mathbb{E}(X_n X_r \mid X_1, \ldots, X_{n-1})\big) \\ &= \mathbb{E}(X_n X_r) - \mathbb{E}(X_n X_r) = 0. \end{aligned}$$

However, Y and X are normally distributed, and furthermore $\mathbb{E}(Y) = 0$; as in Example (4.5.9), Y and X_r are independent. It follows that Y is independent of the collection $X_1, X_2, \ldots, X_{n-1}$, using properties of the multivariate normal distribution.

Write $A_r = \{X_r = x_r\}$ and $A = A_1 \cap A_2 \cap \cdots \cap A_{n-1}$. By the proven independence, $\mathbb{E}(Y^2 \mid A) = \mathbb{E}(Y^2 \mid A_{n-1})$, which may be written as $\text{var}(X_n \mid A) = \text{var}(X_n \mid A_{n-1})$, by (9). Thus the left- and right-hand sides of (6) have the same second moment also, and the result is proved. ■

(10) Example. A stationary Gaussian Markov process. Suppose X is stationary, Gaussian and Markov, and has zero means. Use the result of Problem (4.14.13) to obtain that

$$c(0)\mathbb{E}[X(s+t) \mid X(s)] = c(t)X(s) \quad \text{whenever} \quad t \geq 0,$$

where c is the autocovariance function of X. Thus, if $0 \leq s \leq s+t$ then

$$\begin{aligned} c(0)\mathbb{E}[X(0)X(s+t)] &= c(0)\mathbb{E}\Big[\mathbb{E}\big(X(0)X(s+t) \,\big|\, X(0), X(s)\big)\Big] \\ &= c(0)\mathbb{E}\big[X(0)\mathbb{E}\big(X(s+t) \,\big|\, X(s)\big)\big] \\ &= c(t)\mathbb{E}\big(X(0)X(s)\big) \end{aligned}$$

by Lemma (7.7.10). Thus

$$c(0)c(s+t) = c(s)c(t) \quad \text{for} \quad s, t \geq 0. \tag{11}$$

This is satisfied whenever

$$c(t) = c(0)e^{-\alpha|t|}. \tag{12}$$

Following Problem (4.14.5) we can see that (12) is the general solution to (11) subject to some condition of regularity such as that c be continuous. We shall see later (see Problem (13.12.4)) that such a process is called a stationary *Ornstein–Uhlenbeck process*. ●

(13) Example. The Wiener process. Suppose that $\sigma^2 > 0$ and define

$$c(s,t) = \sigma^2 \min\{s,t\} \quad \text{whenever} \quad s, t \geq 0. \tag{14}$$

We claim that there exists a Gaussian process $W = \{W(t) : t \geq 0\}$ with zero means such that $W(0) = 0$ and $\text{cov}(W(s), W(t)) = c(s,t)$. By the argument in the proof of (1), it is sufficient to show that the matrix $\mathbf{V}(\mathbf{t})$ with entries (v_{jk}), where $v_{jk} = c(t_k, t_j)$, is positive definite for all $\mathbf{t} = (t_1, t_2, \ldots, t_n)$. In order to see that this indeed holds, let $z_1, z_2, \ldots, z_n$ be complex numbers and suppose that $0 = t_0 < t_1 < \cdots < t_n$. It is not difficult to check that

$$\sum_{j,k=1}^{n} c(t_k, t_j) z_j \bar{z}_k = \sigma^2 \sum_{j=1}^{n} (t_j - t_{j-1}) \left| \sum_{k=j}^{n} z_k \right|^2 > 0$$

whenever one of the z_j is non-zero; this guarantees the existence of W. It is called the *Wiener process*; we explore its properties in more detail in Chapter 13, noting only two facts here.

(15) Lemma. *The Wiener process W satisfies $\mathbb{E}(W(t)^2) = \sigma^2 t$ for all $t \geq 0$.*

Proof. $\mathbb{E}(W(t)^2) = \operatorname{cov}\big(W(t), W(t)\big) = c(t,t) = \sigma^2 t.$ ■

(16) Lemma. *The Wiener process W has stationary independent increments, that is:*

(a) *the distribution of $W(t) - W(s)$ depends on $t - s$ alone,*

(b) *the variables $W(t_j) - W(s_j)$, $1 \leq j \leq n$, are independent whenever the intervals $(s_j, t_j]$ are disjoint.*

Proof. The increments of W are jointly normally distributed; their independence follows as soon as we have shown that they are uncorrelated. However, if $u \leq v \leq s \leq t$,

$$\begin{aligned}\mathbb{E}\big([W(v) - W(u)][W(t) - W(s)]\big) &= c(v,t) - c(v,s) + c(u,s) - c(u,t) \\ &= \sigma^2(v - v + u - u) = 0\end{aligned}$$

by (14).

Finally, $W(t) - W(s)$ is normally distributed with zero mean, and with variance

$$\begin{aligned}\mathbb{E}\big([W(t) - W(s)]^2\big) &= \mathbb{E}(W(t)^2) - 2c(s,t) + \mathbb{E}(W(s)^2) \\ &= \sigma^2(t - s) \quad \text{if} \quad s \leq t.\end{aligned}$$ ■●

Exercises for Section 9.6

1. Show that the function $c(s,t) = \min\{s,t\}$ is positive definite. That is, show that

$$\sum_{j,k=1}^{n} c(t_k, t_j) z_j \bar{z}_k > 0$$

for all $0 \leq t_1 < t_2 < \cdots < t_n$ and all complex numbers $z_1, z_2, \ldots, z_n$ at least one of which is non-zero.

2. Let $X_1, X_2, \ldots$ be a stationary Gaussian sequence with zero means and unit variances which satisfies the Markov property. Find the spectral density function of the sequence in terms of the constant $\rho = \operatorname{cov}(X_1, X_2)$.

3. Show that a Gaussian process is strongly stationary if and only if it is weakly stationary.

4. Let X be a stationary Gaussian process with zero mean, unit variance, and autocovariance function $c(t)$. Find the autocovariance functions of the processes $X^2 = \{X(t)^2 : -\infty < t < \infty\}$ and $X^3 = \{X(t)^3 : -\infty < t < \infty\}$.

9.7 Problems

1. Let $\ldots, X_{-1}, X_0, X_1, \ldots$ be uncorrelated random variables with zero means and unit variances, and define

$$Y_n = X_n + \alpha \sum_{i=1}^{\infty} \beta^{i-1} X_{n-i} \qquad \text{for } -\infty < n < \infty,$$

where α and β are constants satisfying $|\beta| < 1$, $|\beta - \alpha| < 1$. Find the best linear predictor of Y_{n+1} given the entire past $Y_n, Y_{n-1}, \ldots$.

2. Let $\{Y_k : -\infty < k < \infty\}$ be a stationary sequence with variance σ_Y^2, and let

$$X_n = \sum_{k=0}^{r} a_k Y_{n-k}, \qquad -\infty < n < \infty,$$

where $a_0, a_1, \ldots, a_r$ are constants. Show that X has spectral density function

$$f_X(\lambda) = \frac{\sigma_Y^2}{\sigma_X^2} f_Y(\lambda) |G_a(e^{i\lambda})|^2$$

where f_Y is the spectral density function of Y, $\sigma_X^2 = \operatorname{var}(X_1)$, and $G_a(z) = \sum_{k=0}^{r} a_k z^k$.

Calculate this spectral density explicitly in the case of 'exponential smoothing', when $r = \infty$, $a_k = \mu^k(1-\mu)$, and $0 < \mu < 1$.

3. Suppose that $\widehat{Y}_{n+1} = \alpha Y_n + \beta Y_{n-1}$ is the best linear predictor of Y_{n+1} given the entire past $Y_n, Y_{n-1}, \ldots$ of the stationary sequence $\{Y_k : -\infty < k < \infty\}$. Find the spectral density function of the sequence.

4. Recurrent events (5.2.15). Meteorites fall from the sky at integer times $T_1, T_2, \ldots$ where $T_n = X_1 + X_2 + \cdots + X_n$. We assume that the X_i are independent, $X_2, X_3, \ldots$ are identically distributed, and the distribution of X_1 is such that the probability that a meteorite falls at time n is constant for all n. Let Y_n be the indicator function of the event that a meteorite falls at time n. Show that $\{Y_n\}$ is stationary and find its spectral density function in terms of the characteristic function of X_2.

5. Let $X = \{X_n : n \geq 1\}$ be given by $X_n = \cos(nU)$ where U is uniformly distributed on $[-\pi, \pi]$. Show that X is stationary but not strongly stationary. Find the autocorrelation function of X and its spectral density function.

6. (a) Let N be a Poisson process with intensity λ, and let $\alpha > 0$. Define $X(t) = N(t+\alpha) - N(t)$ for $t \geq 0$. Show that X is strongly stationary, and find its spectral density function.

(b) Let W be a Wiener process and define $X = \{X(t) : t \geq 1\}$ by $X(t) = W(t) - W(t-1)$. Show that X is strongly stationary and find its autocovariance function. Find the spectral density function of X.

7. Let $Z_1, Z_2, \ldots$ be uncorrelated variables, each with zero mean and unit variance.

(a) Define the moving average process X by $X_n = Z_n + \alpha Z_{n-1}$ where α is a constant. Find the spectral density function of X.

(b) More generally, let $Y_n = \sum_{i=0}^{r} \alpha_i Z_{n-i}$, where $\alpha_0 = 1$ and $\alpha_1, \ldots, \alpha_r$ are constants. Find the spectral density function of Y.

8. Show that the complex-valued stationary process $X = \{X(t) : -\infty < t < \infty\}$ has a spectral density function which is bounded and uniformly continuous whenever its autocorrelation function ρ is continuous and satisfies $\int_0^{\infty} |\rho(t)|\, dt < \infty$.

9. Let $X = \{X_n : n \geq 1\}$ be stationary with constant mean $\mu = \mathbb{E}(X_n)$ for all n, and such that $\operatorname{cov}(X_0, X_n) \to 0$ as $n \to \infty$. Show that $n^{-1} \sum_{j=1}^{n} X_j \xrightarrow{\text{m.s.}} \mu$.

10. Deduce the strong law of large numbers from an appropriate ergodic theorem.

11. Let $\mathbb{Q}$ be a stationary measure on $(\mathbb{R}^T, \mathcal{B}^T)$ where $T = \{1, 2, \dots\}$. Show that $\mathbb{Q}$ is ergodic if and only if

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to \mathbb{E}(Y) \qquad \text{a.s. and in mean}$$

for all $Y : \mathbb{R}^T \to \mathbb{R}$ for which $\mathbb{E}(Y)$ exists, where $Y_i : \mathbb{R}^T \to \mathbb{R}$ is given by $Y_i(\mathbf{x}) = Y(\tau^{i-1}(\mathbf{x}))$. As usual, τ is the natural shift operator on $\mathbb{R}^T$.

12. The stationary measure $\mathbb{Q}$ on $(\mathbb{R}^T, \mathcal{B}^T)$ is called *strongly mixing* if $\mathbb{Q}(A \cap \tau^{-n}B) \to \mathbb{Q}(A)\mathbb{Q}(B)$ as $n \to \infty$, for all $A, B \in \mathcal{B}^T$; as usual, $T = \{1, 2, \dots\}$ and τ is the shift operator on $\mathbb{R}^T$. Show that every strongly mixing measure is ergodic.

13. Ergodic theorem. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $T : \Omega \to \Omega$ be measurable and measure preserving (i.e., $\mathbb{P}(T^{-1}A) = \mathbb{P}(A)$ for all $A \in \mathcal{F}$). Let $X : \Omega \to \mathbb{R}$ be a random variable, and let X_i be given by $X_i(\omega) = X(T^{i-1}(\omega))$. Show that

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}(X \mid \mathcal{I}) \qquad \text{a.s. and in mean}$$

where $\mathcal{I}$ is the σ-field of invariant events of T.

If T is ergodic (in that $\mathbb{P}(A)$ equals 0 or 1 whenever A is invariant), prove that $\mathbb{E}(X \mid \mathcal{I}) = \mathbb{E}(X)$ almost surely.

14. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega = [0, 1)$, $\mathcal{F}$ is the set of Borel subsets, and $\mathbb{P}$ is Lebesgue measure. Show that the shift $T : \Omega \to \Omega$ defined by $T(x) = 2x \pmod 1$ is measurable, measure preserving, and ergodic (in that $\mathbb{P}(A)$ equals 0 or 1 if $A = T^{-1}A$).

Let $X : \Omega \to \mathbb{R}$ be the random variable given by the identity mapping $X(\omega) = \omega$. Show that the proportion of 1's, in the expansion of X to base 2, equals $\frac{1}{2}$ almost surely. This is sometimes called 'Borel's normal number theorem'.

15. Let $g : \mathbb{R} \to \mathbb{R}$ be periodic with period 1, and uniformly continuous and integrable over $[0, 1]$. Define $Z_n = g\big(X + (n-1)\alpha\big)$, $n \geq 1$, where X is uniform on $[0, 1]$ and α is irrational. Show that, as $n \to \infty$,

$$\frac{1}{n}\sum_{j=1}^{n} Z_j \to \int_0^1 g(u)\,du \qquad \text{a.s.}$$

16. Let $X = \{X(t) : t \geq 0\}$ be a non-decreasing random process such that:
(a) $X(0) = 0$, X takes values in the non-negative integers,
(b) X has stationary independent increments,
(c) the sample paths $\{X(t, \omega) : t \geq 0\}$ have only jump discontinuities of unit magnitude.
Show that X is a Poisson process.

17. Let X be a continuous-time process. Show that:
(a) if X has stationary increments and $m(t) = \mathbb{E}(X(t))$ is a continuous function of t, then there exist α and β such that $m(t) = \alpha + \beta t$,
(b) if X has stationary independent increments and $v(t) = \operatorname{var}(X(t) - X(0))$ is a continuous function of t then there exists σ^2 such that $\operatorname{var}(X(s+t) - X(s)) = \sigma^2 t$ for all s.

18. A Wiener process W is called *standard* if $W(0) = 0$ and $W(1)$ has unit variance. Let W be a standard Wiener process, and let α be a positive constant. Show that:
(a) $\alpha W(t/\alpha^2)$ is a standard Wiener process,
(b) $W(t + \alpha) - W(\alpha)$ is a standard Wiener process,
(c) the process V, given by $V(t) = tW(1/t)$ for $t > 0$, $V(0) = 0$, is a standard Wiener process,
(d) the process $W(1) - W(1 - t)$ is a standard Wiener process on $[0, 1]$.

19. Let W be a standard Wiener process. Show that the stochastic integrals

$$X(t) = \int_0^t dW(u), \quad Y(t) = \int_0^t e^{-(t-u)}\, dW(u), \qquad t \geq 0,$$

are well defined, and prove that $X(t) = W(t)$, and that Y has autocovariance function $\operatorname{cov}(Y(s), Y(t)) = \frac{1}{2}(e^{-|s-t|} - e^{-s-t})$, $s < t$.

20. Let W be a standard Wiener process. Find the means of the following processes, and the autocovariance functions in cases (b) and (c):

(a) $X(t) = |W(t)|$,

(b) $Y(t) = e^{W(t)}$,

(c) $Z(t) = \int_0^t W(u)\, du$.

Which of these are Gaussian processes? Which of these are Markov processes?

21. Let W be a standard Wiener process. Find the conditional joint density function of $W(t_2)$ and $W(t_3)$ given that $W(t_1) = W(t_4) = 0$, where $t_1 < t_2 < t_3 < t_4$.

Show that the conditional correlation of $W(t_2)$ and $W(t_3)$ is

$$\rho = \sqrt{\frac{(t_4 - t_3)(t_2 - t_1)}{(t_4 - t_2)(t_3 - t_1)}}.$$

22. Empirical distribution function. Let $U_1, U_2, \ldots$ be independent random variables with the uniform distribution on $[0, 1]$. Let $I_j(x)$ be the indicator function of the event $\{U_j \leq x\}$, and define

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} I_j(x), \qquad 0 \leq x \leq 1.$$

The function F_n is called the 'empirical distribution function' of the U_j.

(a) Find the mean and variance of $F_n(x)$, and prove that $\sqrt{n}(F_n(x) - x) \xrightarrow{\mathrm{D}} Y(x)$ as $n \to \infty$, where $Y(x)$ is normally distributed.

(b) What is the (multivariate) limit distribution of a collection of random variables of the form $\{\sqrt{n}(F_n(x_i) - x_i) : 1 \leq i \leq k\}$, where $0 \leq x_1 < x_2 < \cdots < x_k \leq 1$?

(c) Show that the autocovariance function of the asymptotic finite-dimensional distributions of $\sqrt{n}(F_n(x) - x)$, in the limit as $n \to \infty$, is the same as that of the process $Z(t) = W(t) - tW(1)$, $0 \leq t \leq 1$, where W is a standard Wiener process. The process Z is called a 'Brownian bridge' or 'tied-down Brownian motion'.