Chapter 2
Elements of Ergodic Theory of Stationary Processes and Strong Mixing

2.1 Basic Definitions and Ergodicity

Let $\big(X_n,\ n \in \mathbb{Z}\big)$ be a discrete-time stationary stochastic process. Consider the space $\mathbb{R}^{\mathbb{Z}}$ of the doubly infinite sequences $\mathbf{x} = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots)$ of real numbers, and equip this space with the usual cylindrical σ-field $\mathcal{B}^{\mathbb{Z}}$. The stochastic process naturally induces a probability measure $\mu_{\mathbf{X}}$ on this space via

$$\mu_{\mathbf{X}}\Big\{\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} : \big(x_i, \ldots, x_j\big) \in B\Big\} = P\Big(\big(X_i, \ldots, X_j\big) \in B\Big) \tag{2.1}$$

for all $i \le j$ and Borel sets $B \in \mathbb{R}^{j-i+1}$. The space $\mathbb{R}^{\mathbb{Z}}$ has a natural *left shift* operation $\theta : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$. For $\mathbf{x} = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots) \in \mathbb{R}^{\mathbb{Z}}$, the shifted sequence $\theta\mathbf{x}$ is the sequence of real numbers whose ith coordinate is the $(i+1)$st coordinate x_{i+1} of $\mathbf{x}$ for each $i \in \mathbb{Z}$. Formally,

$$\theta\big((\ldots, x_{-1}, x_0, x_1, x_2, \ldots)\big) = (\ldots, x_0, x_1, x_2, x_3 \ldots)\,.$$

Clearly, the left shift is a one-to-one transformation of $\mathbb{R}^{\mathbb{Z}}$ onto itself, and both θ and its inverse, the *right shift* θ^{-1}, are measurable with respect to the cylindrical σ-field. Note that the left shift θ leaves the measure $\mu_{\mathbf{X}}$ on $\mathbb{R}^{\mathbb{Z}}$ unchanged, because for all $i \le j$ and Borel sets $B \in \mathbb{R}^{j-i+1}$,

$$\begin{aligned}
\mu_{\mathbf{X}} \circ \theta^{-1}\Big\{\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} : \big(x_i, \ldots, x_j\big) \in B\Big\} &= \mu_{\mathbf{X}}\Big\{\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} : \big(x_{i+1}, \ldots, x_{j+1}\big) \in B\Big\} \\
&= P\Big(\big(X_{i+1}, \ldots, X_{j+1}\big) \in B\Big) = P\Big(\big(X_i, \ldots, X_j\big) \in B\Big) \\
&= \mu_{\mathbf{X}}\Big\{\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} : \big(x_i, \ldots, x_j\big) \in B\Big\}\,,
\end{aligned}$$

G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*, Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-45575-4_2

where the third equality follows from the stationarity of the process. In other words, the left shift preserves the measure $\mu_{\mathbf{X}}$ induced by a stationary process on $\mathbb{R}^{\mathbb{Z}}$. This is, of course, not particularly exciting. On the other hand, in spite of this preservation of the measure $\mu_{\mathbf{X}}$ by the left shift, if we choose a point (sequence) $\mathbf{x} \in \mathbb{R}^{\mathbb{Z}}$ according to the probability measure $\mu_{\mathbf{X}}$, there is no reason to expect that the *trajectory* $\theta^n \mathbf{x}$, $n = 0, 1, 2, \ldots$, of the point $\mathbf{x}$ should be in some way trivial. Here $\theta^n = \theta \circ \ldots \circ \theta$ is the composition of n left shifts, $n = 1, 2, \ldots$ (which is, of course, simply a left shift by n time units), while θ^0 is the identity operator on $\mathbb{R}^{\mathbb{Z}}$.

In fact, for most stationary stochastic processes, a "typical point" $\mathbf{x}$ selected according to the measure $\mu_{\mathbf{X}}$ follows a highly nontrivial trajectory. Such trajectories are, obviously, closely related to interesting probabilistic properties of a stationary process. Therefore, ergodic theory that studies measure-preserving transformations (as well as more general transformations) of a measure space provides an important point of view on stationary processes. In this and the following sections of this chapter, we describe certain basic notions of ergodic theory and discuss what they mean for stationary stochastic processes. Much more detail can be found in, for example, Krengel (1985) and Aaronson (1997).

We commence by noting that the connection between a stationary stochastic process $(X_n,\, n \in \mathbb{Z})$ and the probability measure $\mu_{\mathbf{X}}$ it induces on the cylindrical σ-field on $\mathbb{R}^{\mathbb{Z}}$ is not a one-way affair, in which the stationary process, defined on some probability space $(\Omega, \mathcal{F}, P)$, is the "primary actor" while the induced measure $\mu_{\mathbf{X}}$ is "secondary." In fact, if we start with *any* probability measure μ on $\mathbb{R}^{\mathbb{Z}}$ that is invariant under the left shift θ, then we can define a stochastic process $(X_n,\, n \in \mathbb{Z})$ on the probability space $(\mathbb{Z}, \mathcal{B}^{\mathbb{Z}}, \mu)$ by

$$X_n(\mathbf{x}) = x_n,\ n \in \mathbb{Z},\ \text{for } \mathbf{x} = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots) \in \mathbb{R}^{\mathbb{Z}}, \tag{2.2}$$

and then the invariance of the measure μ causes stationarity of the process $(X_n,\, n \in \mathbb{Z})$. Moreover, the measure $\mu_{\mathbf{X}}$ induced by this process coincides with μ. Recall also that the definition of the cylindrical σ-field shows that there is a one-to-one correspondence between shift-invariant probability measures on $\mathbb{R}^{\mathbb{Z}}$ and collections of the finite-dimensional distributions of stationary stochastic processes indexed by $\mathbb{Z}$.

We conclude that, given a collection of the finite-dimensional distributions of a stationary stochastic process, we can define a stochastic process with these finite-dimensional distributions on the space $\mathbb{R}^{\mathbb{Z}}$ equipped with the cylindrical σ-field and appropriate shift-invariant probability measure via the coordinate evaluation scheme (2.2). Since the ergodic properties of stationary stochastic processes we discuss (such as ergodicity and mixing) depend only on the finite-dimensional distributions of these processes, it is completely unimportant on what probability space a stochastic process is defined. However, the sequence space $\mathbb{R}^{\mathbb{Z}}$ has a built-in left shift operation, which provides a convenient language for discussing ergodic properties. Therefore, in this section we assume, unless stated otherwise, that a stationary process $(X_n,\, n \in \mathbb{Z})$ is defined on the probability space $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu)$

by (2.2), and that the probability measure μ is shift-invariant. We emphasize that our conclusions about ergodic properties of stationary stochastic processes applies regardless of the actual probability space on which a process is defined.

For now, however, we consider an arbitrary σ-finite measure space $(E, \mathcal{E}, m)$. Let $\phi : E \to E$ be a measurable map. The powers of ϕ are defined in the sense of repeated application: ϕ^n is a map from E to E given by $\phi^n(x) = \phi(\phi(\ldots\phi(x)))$ for $n \geq 1$ (applying ϕ n times). The operator ϕ^0 is, by definition, the identity operator on E. A set $A \in \mathcal{E}$ is called ϕ-invariant if $m(A \triangle \phi^{-1}A) = 0$, where $\triangle$ denotes the symmetric difference of two sets. It is easy to check that the collection $\mathcal{I}$ of all ϕ-invariant sets is a sub-σ-field of $\mathcal{E}$ (see Exercise 2.6.2); we call $\mathcal{I}$ the (ϕ-) invariant σ-field. Invariant σ-fields naturally appear in the following key result of ergodic theory.

Theorem 2.1.1 (Birkhoff's pointwise ergodic theorem). *Suppose that $\phi : E \to E$ is a measurable map, preserving the measure m. Let $f \in L_1(m)$. Then there is a function $g_f : E \to \mathbb{R}$, measurable with respect to the invariant σ-field $\mathcal{I}$, such that*

$$\frac{1}{n}\sum_{j=0}^{n-1} f\big(\phi^j(x)\big) \to g_f(x) \ \text{ as } n \to \infty$$

for m-almost every $x \in E$. The function g_f satisfies $g_f \in L_1(m)$, $\|g_f\|_1 \leq \|f\|_1$, and

$$\int_A g_f(x)\, m(dx) = \int_A f(x)\, m(dx)$$

for every set $A \in \mathcal{I}$ of finite measure m.

See, e.g., Theorem 2.3 in Krengel (1985, p. 9). Note, for example, that if the measure m in Theorem 2.1.1 is actually a probability measure, then the properties of the function g_f in the theorem identify that function as the conditional expectation of f (viewed as a random variable on the probability space $(E, \mathcal{E}, m)$) given the invariant σ-field $\mathcal{I}$.

A map $\phi : E \to E$ is called *nonsingular* if ϕ is both onto and one-to-one, both ϕ and its inverse $\phi^{-1} : E \to E$ are measurable, and the induced measure $m \circ \phi^{-1}$ is equivalent to the original measure m. Clearly, if ϕ preserves the measure m, it is also nonsingular, as long as it satisfies the other requirements of nonsingularity.

Definition 2.1.2. A nonsingular map ϕ on $(E, \mathcal{E}, m)$ is called ergodic if every ϕ-invariant set A is such that either $m(A) = 0$ or $m(A^c) = 0$.

Note that every measurable set A such that either $m(A) = 0$ or $m(A^c) = 0$ is invariant for every nonsingular map (see Exercise 2.6.3). They are trivially invariant, so to speak. What distinguishes ergodic nonsingular maps is that no other measurable sets are invariant for these maps.

Example 2.1.3. Let $E = \mathbb{Z}$, let $\mathcal{E}$ be the collection of all subsets of $\mathbb{Z}$, and let m be the counting measure. Consider two nonsingular (actually, measure-preserving) maps: $\phi_1(x) = x + 1$, $\phi_2(x) = x$, $x \in \mathbb{Z}$.

Note that the only ϕ_1-invariant sets are the empty set and the entire space $\mathbb{Z}$. These are trivially invariant sets, and hence ϕ_1 is an ergodic map.

On the other hand, ϕ_2 is the identity map, so that every measurable set is invariant with respect to ϕ_2. Since this includes many nontrivially invariant sets (such as the set of all even numbers, for example), the map ϕ_2 is not ergodic.

If ϕ is a nonsingular measure-preserving ergodic map, then Exercise 2.6.4 tells us that the function g_f in Birkhoff's pointwise ergodic theorem must be a constant function regardless of what function $f \in L_1(m)$ we choose. In particular, if m is an infinite measure, then for every such f, the function g_f must be, up to a set of measure zero, the zero function, since $g_f \in L_1(m)$, and the only constant function in the space $L_1(m)$ of an infinite measure m is the zero function.

Now let $(X_n,\ n \in \mathbb{Z})$ be a stationary stochastic process defined by (2.2) on the probability space $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu)$ with a shift-invariant μ. We say that the stochastic process is ergodic if the left shift θ is an ergodic map, $\mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$, in the sense of Definition 2.1.2. Since ergodicity of the left shift is determined by the probability measure μ, which is, in turn, determined by the finite-dimensional distributions of the process, the latter determine whether a given stationary process is ergodic. Notice that a stationary process $(X_n,\ n \in \mathbb{Z})$ is ergodic if and only if the time-reversed process $(X_{-n},\ n \in \mathbb{Z})$ is ergodic (see Exercise 2.6.3).

Example 2.1.4. Tail and invariant σ-fields In the context of stationary stochastic processes on $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu)$, the ($\theta$-)invariant σ-field $\mathcal{I}$ is a sub-σ-field of $\mathcal{B}^{\mathbb{Z}}$. Another important natural sub-σ-field of $\mathcal{B}^{\mathbb{Z}}$ is the *tail σ-field* $\mathcal{T}$, defined as the completion with respect to the measure μ of the σ-field

$$\bigcap_{n=1}^{\infty} \sigma(x_n, x_{n+1}, \ldots)\,.$$

In general, not every tail event is an invariant event, as can be seen by choosing

$$\mu = \frac{1}{2}\delta_{(\ldots,0,1,0,\ldots)} + \frac{1}{2}\delta_{(\ldots,1,0,1,\ldots)} \tag{2.3}$$

($x_0 = 1$ in the first sequence and $x_0 = 0$ in the second sequence), and

$$A = \Big\{(\ldots, x_{-1}, x_0, x_1, x_2, \ldots) : x_{2n} = 1 \text{ for infinitely many } n = 0, 1, 2, \ldots\Big\}\,.$$

On the other hand, we claim that every invariant event is a tail event, that is,

$$\mathcal{I} \subset \mathcal{T}\,. \tag{2.4}$$

To see this, let A be an invariant cylindrical set. Since the sets in a σ-field can be approximated arbitrarily closely with respect to a probability measure by the sets in a field that generates the σ-field, for every $\varepsilon > 0$ there is a finite-dimensional cylindrical set B such that

$$\mu\big(A \triangle B\big) \le \varepsilon$$

(finite-dimensionality of B means that $B \in \sigma\big(x_k, \dots, x_m\big)$ for some $k \le m$); see, for example, Corollary 1, in Billingsley (1995, p. 169). Since the measure μ is shift-invariant and the set A is invariant as well, we conclude that for every $n \ge 0$,

$$\varepsilon \ge \mu\big(A \triangle B\big) = \mu\Big(\theta^{-1}\big(A \triangle B\big)\Big)$$
$$= \mu\Big(\theta^{-1}(A) \triangle \theta^{-1}(B)\Big) = \mu\Big(A \triangle \theta^{-1}(B)\Big),$$

and iterating this procedure, we see that

$$\mu\Big(A \triangle \theta^{-n}(B)\Big) \le \varepsilon$$

for all $n = 0, 1, \dots$. Note that if $B \in \sigma\big(x_k, \dots, x_m\big)$, then $\theta^{-n}(B) \in \sigma\big(x_{k+n}, \dots, x_{m+n}\big)$.

We conclude that for every $\varepsilon > 0$, there are sets $B_n \in \sigma(x_n, x_{n+1}, \dots)$ for $n = 0, 1, 2, \dots$ such that for every n, $\mu\big(A \triangle B_n\big) \le \varepsilon/2^n$. This implies that for every $m = 0, 1, 2 \dots$,

$$\mu\Big(A \triangle \cup_{n=m}^{\infty} B_n\Big) \le \sum_{n=m}^{\infty} \mu\big(A \triangle B_n\big) \le \varepsilon/2^{m-1},$$

and then also

$$\mu\Big(A \triangle \cap_{m=0}^{\infty} \cup_{n=m}^{\infty} B_n\Big) \le \sum_{m=0}^{\infty} \mu\big(A \triangle \cup_{n=m}^{\infty} B_n\big) \le 4\varepsilon.$$

Therefore, for every $\varepsilon > 0$, there is a set $B \in \mathcal{T}$ such that $\mu\big(A \triangle B\big) \le \varepsilon$, and since $\mathcal{T}$ is a σ-field, there is also $B \in \mathcal{T}$ such that $\mu\big(A \triangle B\big) = 0$. Therefore, A itself is a tail event.

As a corollary, we conclude that every stationary process for which the tail σ-field consists of trivial events (i.e., of the events of probability 0 or 1) must be ergodic, because every invariant event will also have probability 0 or 1. For example, if $\big(X_n,\ n \in \mathbb{Z}\big)$ consists of i.i.d. random variables, then by the Kolmogorov zero–one law, the tail σ-field consists of trivial events, and so every i.i.d. process is ergodic.

If $(X_n,\, n \in \mathbb{Z})$ is a stationary stochastic process defined by (2.2) on the probability space $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu)$ with a shift-invariant μ, then Birkhoff's pointwise ergodic theorem applies, of course, in the usual way, and the limiting function g_f in that theorem is, of course, the conditional expectation of f given the invariant σ-field just because m is a probability measure. It is interesting to see how the pointwise ergodic theorem applies if a stationary process $(X_n,\, n \in \mathbb{Z})$ is defined on an abstract probability space $(\Omega, \mathcal{F}, P)$.

We start by defining the invariant σ-sub-field $\mathcal{I}_{\mathbf{X}}$ of $\mathcal{F}$. Let $T_{\mathbf{X}} : \Omega \to \mathbb{R}^{\mathbb{Z}}$ be the measurable map $T_{\mathbf{X}}(w) = (\ldots, X_{-1}(\omega), X_0(\omega), X_1(\omega), \ldots)$. If $\mathcal{I}$ is the invariant σ-field in $\mathbb{R}^{\mathbb{Z}}$, then we set

$$\mathcal{I}_{\mathbf{X}} = T_{\mathbf{X}}^{-1}\mathcal{I} = \left\{T_{\mathbf{X}}^{-1}(A),\, A \in \mathcal{I}\right\} \subset \mathcal{F}. \tag{2.5}$$

If μ is the shift-invariant probability measure on $\mathbb{R}^{\mathbb{Z}}$ induced by the stationary process via (2.1), then for every measurable function $f : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ such that

$$E\big|f(\ldots, X_{-1}, X_0, X_1, \ldots)\big| < \infty, \tag{2.6}$$

we have

$$\frac{1}{n}\sum_{j=0}^{n-1} f(\ldots, X_{j-1}(\omega), X_j(\omega), X_{j+1}(\omega), \ldots) = \frac{1}{n}\sum_{j=0}^{n-1} f\big(\theta^j(T_{\mathbf{X}}(\omega))\big)$$

$$\to E_\mu\big(f\big|\mathcal{I}\big)(T_{\mathbf{X}}(\omega)) = E\Big(f\big(\ldots, X_{-1}, X_0, X_1, \ldots\big)\Big|\mathcal{I}_{\mathbf{X}}\Big)(\omega) \tag{2.7}$$

with probability 1, where E_μ is the (conditional) expectation with respect to the probability measure μ on $\mathbb{R}^{\mathbb{Z}}$. The convergence follows from Birkhoff's pointwise ergodic theorem, and the last equality follows from the definition of the σ-field $\mathcal{I}_{\mathbf{X}}$.

In particular, a stationary process $(X_n,\, n \in \mathbb{Z})$ is ergodic if and only if the σ-field $\mathcal{I}_{\mathbf{X}}$ defined by (2.5) consists of trivial events. For an ergodic stationary process, one has

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f\big(\ldots, X_{j-1}(\omega), X_j(\omega), X_{j+1}(\omega), \ldots\big) \tag{2.8}$$

$$= Ef\big(\ldots, X_{-1}, X_0, X_1, \ldots\big)$$

with probability 1, for every measurable f satisfying (2.6). Of course, the converse statement is also true: if (2.8) holds for every measurable f satisfying (2.6), then the process is ergodic.

Example 2.1.5. The Strong Law of Large Numbers Suppose that $(X_n,\, n \in \mathbb{Z})$ is an ergodic stationary process with a finite mean. Choosing

$$f\big(\ldots, x_{-1}, x_0, x_1, \ldots\big) = x_0\,,$$

an immediate application of (2.8) proves the strong law of large numbers

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} X_j = EX_0 \tag{2.9}$$

with probability 1. Note, however, that the strong law of large numbers (2.9) may hold for nonergodic stationary processes and hence does not imply ergodicity. For example, let $(X_n,\ n \in \mathbb{Z})$ be a canonical stationary process defined by (2.2) on $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu)$, where a shift-invariant μ is the following modification of the measure in (2.3):

$$\mu = \frac{1}{4}\delta_{(\ldots,0,1,0,\ldots)} + \frac{1}{4}\delta_{(\ldots,1,0,1,\ldots)} + \frac{1}{2}\delta_{(\ldots,1/2,1/2,1/2,\ldots)}\,.$$

In this case, the law of large numbers trivially holds (with $EX_0 = 1/2$), but the process is not ergodic, since the event

$$A = \left\{\text{infinitely many of } x_n \text{ are equal to zero}\right\}$$

is obviously invariant, and its probability is equal to $1/2$.

The following proposition presents a characterization of ergodicity in the context of stationary stochastic processes that is particularly easy to visualize: a stationary process is ergodic unless it can be represented as a mixture of two stationary processes with different finite-dimensional distributions.

Proposition 2.1.6. *A stationary process $(X_n,\ n \in \mathbb{Z})$ is nonergodic if and only if there is a probability space supporting two stationary processes, $(Y_n,\ n \in \mathbb{Z})$ and $(Z_n,\ n \in \mathbb{Z})$, with different finite-dimensional distributions, and a Bernoulli(p) random variable with $0 < p < 1$ independent of them such that*

$$\left(X_n,\ n \in \mathbb{Z}\right) \stackrel{d}{=} \begin{cases} \left(Y_n,\ n \in \mathbb{Z}\right) \text{ with probability } p, \\ \left(Z_n,\ n \in \mathbb{Z}\right) \text{ with probability } 1-p. \end{cases} \tag{2.10}$$

Proof. Suppose first that the process $(X_n,\ n \in \mathbb{Z})$ is not ergodic, and let μ be the probability measure induced by the process on the probability space $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu)$ via (2.1). By the lack of ergodicity, there is a set of sequences $A \in \mathcal{I}$ with $p := \mu(A) \in (0,1)$. Define two new probability measures on $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu)$ by

$$\mu_1(B) = p^{-1}\mu\big(B \cap A\big),\ \ \mu_2(B) = (1-p)^{-1}\mu\big(B \cap A^c\big),\ B \in \mathcal{B}^{\mathbb{Z}}\,.$$

Using first the shift-invariance of A and then the shift-invariance of μ, we see that

$$\begin{aligned}\mu_1\big(\theta^{-1}(B)\big) &= p^{-1}\mu\big(\theta^{-1}(B) \cap A\big) = p^{-1}\mu\big(\theta^{-1}(B) \cap \theta^{-1}(A)\big) \\ &= p^{-1}\mu\big(\theta^{-1}(B \cap A)\big) = p^{-1}\mu\big(B \cap A\big) = \mu_1(B),\end{aligned}$$

and hence μ_1 is shift-invariant. Similarly, the probability measure μ_2 is shift-invariant as well.

Let now $\Omega = \mathbb{R}^{\mathbb{Z}} \cap \{0,1\} \cap \mathbb{R}^{\mathbb{Z}}$, with the product σ-field, and let

$$P = \mu_1 \times \big(p\delta_1 + (1-p)\delta_0\big) \times \mu_2 \,.$$

Using the obvious notation, we define three stochastic processes on this probability space by

$$W_n\big((\ldots, y_{-1}, y_0, y_1 \ldots),\, b,\, (\ldots, z_{-1}, z_0, z_1 \ldots)\big) = \begin{cases} y_n & \text{if } b = 1 \\ z_n & \text{if } b = 0 \end{cases},$$

$$Y_n\big((\ldots, y_{-1}, y_0, y_1 \ldots),\, b,\, (\ldots, z_{-1}, z_0, z_1 \ldots)\big) = y_n \,,$$

$$Z_n\big((\ldots, y_{-1}, y_0, y_1 \ldots),\, b,\, (\ldots, z_{-1}, z_0, z_1 \ldots)\big) = z_n, \quad n \in \mathbb{Z} \,.$$

Since the measures μ_1 and μ_2 are shift-invariant, the processes $\big(Y_n,\, n \in \mathbb{Z}\big)$ and $\big(Z_n,\, n \in \mathbb{Z}\big)$ are stationary. If the two processes had the same finite-dimensional distributions, then the probability measures they generate on $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$ would coincide. However, these probability measures are μ_1 and μ_2 respectively, and they cannot coincide, since they live on disjoint subsets A and A^c of $\mathbb{R}^{\mathbb{Z}}$. Therefore, the two processes have different finite-dimensional distributions. Finally, let A be a cylindrical subset of $\mathbb{R}^{\mathbb{Z}}$. Then

$$\begin{aligned} P\Big(\big(W_n,\, n \in \mathbb{Z}\big) \in A\Big) &= p\mu_1(A) + (1-p)\mu_2(A) \\ &= \mu(A) = P\Big(\big(X_n,\, n \in \mathbb{Z}\big) \in A\Big), \end{aligned}$$

and we conclude that the relation (2.10) holds.

Conversely, suppose that (2.10) holds. Since the processes $\mathbf{Y} = \big(Y_n,\, n \in \mathbb{Z}\big)$ and $\mathbf{Z} = \big(Z_n,\, n \in \mathbb{Z}\big)$ have different finite-dimensional distributions, there is a bounded measurable function $f : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ such that $Ef(\mathbf{Y}) \neq Ef(\mathbf{Z})$ (one can take $f = \mathbf{1}_A$, with A a cylindrical set to which the laws of $\mathbf{Y} = \big(Y_n,\, n \in \mathbb{Z}\big)$ and $\mathbf{Z} = \big(Z_n,\, n \in \mathbb{Z}\big)$ assign different probabilities). If we call the right-hand side of (2.10) $\big(X_n,\, n \in \mathbb{Z}\big)$ (which is legitimate, since ergodicity depends only on the finite-dimensional distributions of a process), then the ergodic theorem (2.7) tells us that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\big(\ldots, X_{j-1}(\omega), X_j(\omega), X_{j+1}(\omega), \ldots\big)$$

$$= \begin{cases} \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\big(\ldots, Y_{j-1}(\omega), Y_j(\omega), Y_{j+1}(\omega), \ldots\big) & \text{with probability } p \\ \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\big(\ldots, Z_{j-1}(\omega), Z_j(\omega), Z_{j+1}(\omega), \ldots\big) & \text{with probability } 1-p \end{cases}$$

$$:= \begin{cases} L_1 & \text{with probability } p \\ L_2 & \text{with probability } 1-p \end{cases},$$

where

$$EL_1 = Ef(\mathbf{Y}) \neq Ef(\mathbf{Z}) = EL_2 .$$

Therefore, L_1 and L_2 cannot be two identical constants, which would be the only possibility if $\big(X_n,\ n \in \mathbb{Z}\big)$ were ergodic. Hence (2.10) implies lack of ergodicity. □

We now switch back to an arbitrary σ-finite measure space $\big(E, \mathcal{E}, m\big)$ and a measurable map $\phi : E \to E$. Let $\big(G, \mathcal{G}\big)$ be another measurable space, and $\varphi : G \to G$ a measurable map. We say that a mapping $f : E \to G$ is compatible with the maps ϕ and φ if $f \circ \phi = \varphi \circ f$, or in other words, if $f(\phi(x)) = \varphi(f(x))$ for all $x \in E$. It turns out that compatible mappings preserve ergodicity, as the following proposition shows.

Proposition 2.1.7. *Let $\phi : E \to E$ be a nonsingular map on a σ-finite measure space $\big(E, \mathcal{E}, m\big)$. Let $\varphi : G \to G$ be a one-to-one and onto map on a measurable space $\big(G, \mathcal{G}\big)$ such that both φ and its inverse are measurable. Let $f : E \to G$ be a measurable map that is compatible with ϕ and φ. If ϕ is ergodic on $\big(E, \mathcal{E}, m\big)$, then φ is ergodic on $\big(G, \mathcal{G}, m \circ f^{-1}\big)$.*

Proof. The compatibility of f with ϕ and φ implies that for every subset B of G,

$$\phi^{-1}\big(f^{-1}(B)\big) = f^{-1}\big(\varphi^{-1}(B)\big) . \tag{2.11}$$

Since ϕ is nonsingular, for every $B \in \mathcal{G}$,

$$\begin{aligned} m \circ f^{-1}(B) = 0 \iff m\big(f^{-1}(B)\big) = 0 \iff m\big(\phi^{-1}\big(f^{-1}(B)\big)\big) = 0 \\ \iff m\big(f^{-1}\big(\varphi^{-1}(B)\big)\big) = 0 \iff m \circ f^{-1}\big(\varphi^{-1}(B)\big) = 0 , \end{aligned}$$

and so φ is nonsingular on $\big(G, \mathcal{G}, m \circ f^{-1}\big)$. Next, let B be a φ-invariant event. Using (2.11) once again, we see that

$$\begin{aligned} 0 = m \circ f^{-1}\big(B \,\triangle\, \varphi^{-1}(B)\big) &= m\big(f^{-1}(B) \,\triangle\, f^{-1}\big(\varphi^{-1}(B)\big)\big) \\ &= m\big(f^{-1}(B) \,\triangle\, \phi^{-1}\big(f^{-1}(B)\big)\big) , \end{aligned}$$

implying that $f^{-1}(B)$ is invariant for ϕ. Since ϕ is ergodic, we conclude that $m\big(f^{-1}(B)\big) = 0$ or 1, which is the same as $m \circ f^{-1}(B) = 0$ or 1, which means that every φ-invariant event is trivially invariant, and hence φ is ergodic. □

Naturally compatible maps are produced by the common transformations of stochastic processes. Let $\big(X_n,\ n \in \mathbb{Z}\big)$ be a stationary stochastic process, and let $g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ be a measurable function. Then

$$Y_n = g\big((\ldots, X_{n-1}, X_n, X_{n+1}, \ldots)\big),\ n \in \mathbb{Z} \tag{2.12}$$

(with X_n in the zeroth position in the definition of Y_n) is obviously a stationary process as well.

We can view this common situation as a special case of compatible maps as follows. Let $E = G = \mathbb{R}^{\mathbb{Z}}$, $\mathcal{E} = \mathcal{G} = \mathcal{B}^{\mathbb{Z}}$, $m = \mu_X$, the law of the process $(X_n,\, n \in \mathbb{Z})$ given by (2.1). Of course, the spaces E and G are identical, and we use the same left shift θ on both. In order to avoid confusion, we will use the notation θ_E when working on the space E, and θ_G when working on the space G.

Define a map $f:\, E \to G$ by

$$f(\mathbf{x}) = \big(\ldots, g(\theta_E^{-1}\mathbf{x}), g(\mathbf{x}), g(\theta_E(\mathbf{x})), \ldots\big) \in G$$

for $\mathbf{x} = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots) \in E$. Clearly, f is measurable. It is also trivially compatible with the shifts θ_E and θ_G. Note also that the probability measure $m \circ f^{-1}$ is simply the law of the process $(Y_n,\, n \in \mathbb{Z})$ given by (2.1).

Therefore, Proposition 2.1.7 applies, and we have proved the following statement.

Corollary 2.1.8. *Let $(X_n,\, n \in \mathbb{Z})$ be an ergodic stationary stochastic process, and let $(Y_n,\, n \in \mathbb{Z})$ be a stationary process given by* (2.12) *for some measurable function $g:\, \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$. Then $(Y_n,\, n \in \mathbb{Z})$ is ergodic as well.*

Example 2.1.9. Moving average processes are ergodic. Let

$$X_n = \sum_{j=-\infty}^{\infty} \varphi_j\, \varepsilon_{n-j}\,, \quad n \in \mathbb{Z}$$

be an infinite moving average process (1.23). Recall that the noise variables $(\varepsilon_n,\, n = \ldots, -1, 0, 1, 2, \ldots)$ are i.i.d. and (φ_n) are deterministic coefficients. We assume that the series defining the process converges with probability 1.

Define a function $g:\, \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ by

$$g\big((\ldots, x_{-1}.x_0, x_1, \ldots)\big) = \begin{cases} \sum_{j=-\infty}^{\infty} \varphi_j\, x_{-j} & \text{if the sum converges,} \\ 0 & \text{if the sum diverges.} \end{cases}$$

Clearly, g is a measurable function, and

$$X_n = g\big((\ldots, \varepsilon_{n-1}.\varepsilon_n, \varepsilon_{n+1}, \ldots)\big),\ n \in \mathbb{Z}\,.$$

Since the i.i.d. process $(\varepsilon_n,\, n \in \mathbb{Z})$ is ergodic (see Example 2.1.4), it follows from Corollary 2.1.8 that every infinite moving average process is ergodic.

2.2 Mixing and Weak Mixing

We start by introducing another basic ergodic-theoretical notion, that of *mixing*. It applies to measure-preserving maps on probability spaces.

Definition 2.2.1. A nonsingular measure-preserving map ϕ on a probability space $\big(E, \mathcal{E}, m\big)$ is called mixing if for every two sets $A, B \in \mathcal{E}$,

$$\lim_{n\to\infty} m\big(A \cap \phi^{-n}B\big) = m(A)\, m(B)\,.$$

Here ϕ^{-n} is the nth power of the inverse operator ϕ^{-1}. An immediate observation is that mixing is a stronger property than ergodicity. Indeed, let ϕ be a mixing map on a probability space $\big(E, \mathcal{E}, m\big)$, and suppose that ϕ is not ergodic. In that case, there is an invariant set $C \in \mathcal{E}$ with $0 < m(C) < 1$. Taking $A = B = C$, we have

$$\begin{aligned} m\big(A \cap \phi^{-n}B\big) &= m\big(C \cap \phi^{-n}C\big) = m(C) \\ &\neq \big(m(C)\big)^2 = m(A)\, m(B)\,, \end{aligned}$$

contradicting the assumed mixing. Therefore, mixing implies ergodicity. On the other hand, a map can be ergodic without being mixing.

Example 2.2.2. An ergodic but not mixing map.

Consider the left shift θ on the sequence space $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$, where μ is the two-point shift-invariant probability measure (2.3). Let $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ be the two points of the support of μ. Note that $\theta\mathbf{x}^{(1)} = \mathbf{x}^{(2)}$ and $\theta\mathbf{x}^{(2)} = \mathbf{x}^{(1)}$. Therefore, every invariant set A contains either both of these points or none, so that $\mu(A) = 1$ or 0. Since all invariant sets are trivial, θ is ergodic.

On the other hand, let

$$A = B = \Big\{(\ldots, x_{-1}, x_0, x_1, x_2, \ldots) : x_0 = 1\Big\}.$$

Then

$$\mu\big(A \cap \theta^{-n}B\big) = \begin{cases} \frac{1}{2} & \text{if } n \text{ is even,} \\ 0 & \text{if } n \text{ is odd} \end{cases} \not\to \frac{1}{4} = \mu(A)\mu(B)\,.$$

Therefore, θ is not mixing.

Let $\big(X_n,\ n \in \mathbb{Z}\big)$ be a stationary stochastic process. We assume first that the process is defined by (2.2) on the probability space $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$ with a shift-invariant μ. We say that the process is mixing if the left shift θ is a mixing map on $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$. As in the case of ergodicity, whether or not a stationary process is mixing is determined by its finite-dimensional distributions, regardless of what probability space the process is really defined on. Explicitly, a stationary process $\big(X_n,\ n \in \mathbb{Z}\big)$ is mixing if for every two cylindrical subsets A, B of $\mathbb{R}^{\mathbb{Z}}$,

$$P\Big(\big(\ldots, X_{-1}, X_0, X_1, \ldots\big) \in A,\ \big(\ldots, X_{n-1}, X_n, X_{n+1}, \ldots\big) \in B\Big) \tag{2.13}$$

$$\to P\Big(\big(\ldots, X_{-1}, X_0, X_1, \ldots\big) \in A\Big) P\Big(\big(\ldots, X_{-1}, X_0, X_1, \ldots\big) \in B\Big)$$

as $n \to \infty$.

Example 2.2.3. An i.i.d. process is mixing

We showed at the end of Example 2.1.4 that an i.i.d. process $(X_n,\, n \in \mathbb{Z})$ is ergodic. Now we will check that such a process is, in fact, mixing. This fact will also follow from Theorem 2.2.7 below, but a direct argument is instructive. We may assume that the process is defined on the sequence space $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu)$ by (2.2).

Take any sets $A, B \in \mathcal{B}^{\mathbb{Z}}$. As in Example 2.1.4, given $\varepsilon > 0$, we may choose finite-dimensional cylindrical sets A_1 and $B_1 \in \sigma(x_k, \ldots, x_m)$ for some $k \le m$ such that

$$\mu(A \triangle A_1) \le \varepsilon, \quad \mu(B \triangle B_1) \le \varepsilon\,.$$

Note that $\theta^{-n}(B_1) \in \sigma(x_{k+n}, \ldots, x_{m+n})$, so that for $n > m - k$, the sets A_1 and $\theta^{-n}(B_1)$ are generated by disjoint sets of coordinates of a point in the sequence space. Since the measure μ is the law of an i.i.d. sequence, different components are independent under μ, which means that A_1 and $\theta^{-n}(B_1)$ are independent events under μ when $n > m - k$. Therefore, for such n,

$$\begin{aligned}
\left|\mu(A \cap \theta^{-n}B) - \mu(A)\,\mu(B)\right| &\le \left|\mu(A \cap \theta^{-n}B) - \mu(A_1 \cap \theta^{-n}B_1)\right| \\
+\left|\mu(A_1 \cap \theta^{-n}B_1) - \mu(A_1)\,\mu(B_1)\right| &+ \left|\mu(A_1)\,\mu(B_1) - \mu(A)\,\mu(B)\right| \\
\le 2\mu(A \triangle A_1) &+ 2\mu(B \triangle B_1) \le 4\varepsilon.
\end{aligned}$$

That is, for every $\varepsilon > 0$,

$$\limsup_{n\to\infty}\left|\mu(A \cap \theta^{-n}B) - \mu(A)\,\mu(B)\right| \le 4\varepsilon,$$

and letting $\varepsilon \to 0$, we conclude that

$$\mu(A \cap \theta^{-n}B) \to \mu(A)\,\mu(B)$$

as $n \to \infty$ for all $A, B \in \mathcal{B}^{\mathbb{Z}}$. Therefore, the shift θ is mixing, and hence so is the i.i.d. process $(X_n,\, n \in \mathbb{Z})$.

Proposition 2.1.7 has a counterpart describing preservation of mixing.

Proposition 2.2.4. *Let $\phi : E \to E$ be a nonsingular measure-preserving map on a probability space $(E, \mathcal{E}, m)$. Let $\varphi : G \to G$ be a one-to-one and onto map on a measurable space $(G, \mathcal{G})$ such that both φ and its inverse are measurable. Let $f : E \to G$ be a measurable map that is compatible with ϕ and φ. If ϕ is mixing on $(E, \mathcal{E}, m)$, then φ is mixing on $(G, \mathcal{G}, m \circ f^{-1})$.*

Proof. Starting with (2.11), an inductive argument shows that the latter relation extends to

$$\phi^{-n}(f^{-1}(B)) = f^{-1}(\varphi^{-n}(B))$$

for all $n \geq 1$. Therefore, for all sets $A, B \in \mathcal{G}$,

$$
\begin{aligned}
m \circ f^{-1}\big(A \cap \varphi^{-n}(B)\big) &= m\big(f^{-1}(A) \cap f^{-1}(\varphi^{-n}(B))\big) \\
= m\big(f^{-1}(A) \cap \phi^{-n}(f^{-1}(B))\big) &\to m\big(f^{-1}(A)\big)\, m\big(f^{-1}(B)\big) \\
&= m \circ f^{-1}(A)\, m \circ f^{-1}(B)\,,
\end{aligned}
$$

and so φ is mixing on $\big(G, \mathcal{G}, m \circ f^{-1}\big)$. $\square$

As in the case of ergodicity, we immediately obtain the following corollary to Proposition 2.2.4.

Corollary 2.2.5. *Let $\big(X_n,\, n \in \mathbb{Z}\big)$ be a mixing stationary stochastic process, and let $\big(Y_n,\, n \in \mathbb{Z}\big)$ be a stationary process given by* (2.12) *for some measurable function $g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$. Then $\big(Y_n,\, n \in \mathbb{Z}\big)$ is mixing as well.*

Example 2.2.6. Moving average processes are mixing.

Let $(X_n,\, n \in \mathbb{Z})$ be the infinite moving process of Example 2.1.9. Applying Corollary 2.2.5, we see that the mixing property of the noise variables (ε_n) says that every infinite moving process is mixing as well.

It turns out that the mixing property of a stationary stochastic process is equivalent to weak convergence to independence of the joint distributions of the blocks of observations of the process separated by increasing time intervals.

Theorem 2.2.7. *A stationary process $\big(X_n,\, n \in \mathbb{Z}\big)$ is mixing if and only if for every $k = 1, 2, \ldots$,*

$$
\big(X_1, \ldots, X_k, X_{n+1}, \ldots, X_{n+k}\big) \Rightarrow \big(X_1, \ldots, X_k, Y_1, \ldots, Y_k\big) \tag{2.14}
$$

as $n \to \infty$, where $(Y_1, \ldots, Y_k)$ is an independent copy of $(X_1, \ldots, X_k)$.

Proof. We start with checking the easier implication, namely the necessity of condition (2.14). Suppose that the process $\big(X_n,\, n \in \mathbb{Z}\big)$ is mixing, and let $k \geq 1$. The weak convergence in (2.14) will follow once we check that for all k-dimensional Borel sets C and D, we have

$$
\begin{aligned}
&P\big((X_1, \ldots, X_k) \in C,\ (X_{n+1}, \ldots, X_{n+k}) \in D\big) \\
&\to P\big((X_1, \ldots, X_k) \in C\big)\, P\big((X_1, \ldots, X_k) \in D\big)
\end{aligned} \tag{2.15}
$$

as $n \to \infty$ (of course, it is really necessary to check (2.15) for continuity sets of the law of $(X_1, \ldots, X_k)$). This statement, however, is a special case of the statement (2.13).

Suppose now that (2.14) holds. We will prove that the process is mixing by checking the condition in the definition for successively more general sets A and B. We treat finite-dimensional cylindrical sets first. Fix a dimension $k \geq 1$. We prove the statement (2.15). The challenge is, of course, in the fact that by default, weak convergence tells us that this statement holds only for continuity sets C and D, and we need to establish (2.15) for all k-dimensional Borel sets.

Consider first the case in which the sets C and D are "southwest corners" of the form

$$C=\Big\{(x_1,x_2,\dots,x_k):\ x_i \le a_i \text{ for } i\in I_1,\ \ x_i < a_i \text{ for } i\in\{1,\dots,k\}\setminus I_1\Big\}, \tag{2.16}$$
$$B=\Big\{(x_1,x_2,\dots,x_k):\ x_i \le b_i \text{ for } i\in I_2,\ \ x_i < b_i \text{ for } i\in\{1,\dots,k\}\setminus I_2\Big\},$$

where I_1, I_2 are subsets of $\{1,\dots,k\}$, and $a_1,\dots,a_k,b_1,\dots,b_k$ are real numbers. Note that if all of these $2k$ numbers are continuity points of the distribution of X_1, then the sets C and D are continuity sets of the vector $(X_1,\dots,X_k,Y_1,\dots,Y_k)$ on the right-hand side of (2.14), so (2.15) follows from the weak convergence in this case. This case forms the basis of an inductive argument. Specifically, let $m=0,1,\dots,2k$ be the number of points among $a_1,\dots,a_k,b_1,\dots,b_k$ that are not continuity points of the distribution of X_1. We have checked that (2.15) holds if $m=0$, and the induction hypothesis is that (2.15) holds if $m<m_0$ for some $m_0=1,,\dots,2k$. Suppose now that $m=m_0$, and choose one point out of the m_0 discontinuity points. We suppose that this point is of the type b_i with $i\in I_2$; all other cases are similar. For ease of notation, we will simply use b_1, with $1\in I_2$.

Choose a sequence $\varepsilon_j\downarrow 0$ such that for each j, $b_1+\varepsilon_j$ is a continuity point of the distribution of X_1. By the induction hypothesis,

$$\begin{aligned}
&\limsup_{n\to\infty} P\big((X_1,\dots,X_k)\in C,\ (X_{n+1},\dots,X_{n+k})\in D\big)\\
&\le \limsup_{n\to\infty} P\big((X_1,\dots,X_k)\in C,\ X_{n+1}\le b_1+\varepsilon_j,\ X_{n+i}\le b_i \text{ for } i\in I_2\setminus\{1\},\\
&\qquad X_{n+i}<b_i \text{ for } i\in\{1,\dots,k\}\setminus I_2\big)\\
&= P\big((X_1,\dots,X_k)\in C\big)P\big(X_1\le b_1+\varepsilon_j,\ X_i\le b_i \text{ for } i\in I_2\setminus\{1\},\\
&\qquad X_i<b_i \text{ for } i\in\{1,\dots,k\}\setminus I_2\big)
\end{aligned}$$

for every $j=1,2,\dots$, and letting $j\to\infty$, we obtain the upper bound

$$\begin{aligned}
&\limsup_{n\to\infty} P\big((X_1,\dots,X_k)\in C,\ (X_{n+1},\dots,X_{n+k})\in D\big)\\
&\le P\big((X_1,\dots,X_k)\in C\big)\,P\big((X_1,\dots,X_k)\in D\big)\,.
\end{aligned}\tag{2.17}$$

In order to obtain a matching lower bound, we write

$$\begin{aligned}
&P\big((X_1,\dots,X_k)\in C,\ (X_{n+1},\dots,X_{n+k})\in D\big)\\
&= P\big((X_1,\dots,X_k)\in C,\ X_{n+1}<b_1,\ X_{n+i}\le b_i \text{ for } i\in I_2\setminus\{1\},\\
&\qquad X_{n+i}<b_i \text{ for } i\in\{1,\dots,k\}\setminus I_2\big)\\
&+P\big((X_1,\dots,X_k)\in C,\ X_{n+1}=b_1,\ X_{n+i}\le b_i \text{ for } i\in I_2\setminus\{1\},\\
&\qquad X_{n+i}<b_i \text{ for } i\in\{1,\dots,k\}\setminus I_2\big) := p_1(n)+p_2(n)\,.
\end{aligned}$$

An argument identical to the one used to obtain the upper bound above (replacing the condition $X_{n+1} < b_1$ by the condition $X_{n+} < b_1 - \varepsilon_j$, where $\varepsilon_j \downarrow 0$ and $b_1 - \varepsilon_j$ is, for each j, a continuity point of the distribution of X_1) shows that

$$\liminf_{n\to\infty} p_1(n) \geq P\big((X_1,\dots,X_k) \in C\big)P\big(X_1 < b_1,\, X_i \leq b_i \text{ for } i \in I_2 \setminus \{1\},$$
$$X_i < b_i \text{ for } i \in \{1,\dots,k\} \setminus I_2\big)\,.$$

Furthermore, given $\varepsilon > 0$, we can find $\delta > 0$ such that both $P(b_1-\delta < X_1 < b_1) \leq \varepsilon$ and $P(b_1 < X_1 < b_1 + \delta) \leq \varepsilon$, and both $b_1 - \delta$ and $b_1 + \delta$ are continuity points of the distribution of X_1. Then

$$p_2(n) \geq P\big((X_1,\dots,X_k) \in C,\, b_1 - \delta < X_{n+1} < b_1 + \delta,\, X_{n+i} \leq b_i \text{ for } i \in I_2 \setminus \{1\},$$
$$X_{n+i} < b_i \text{ for } i \in \{1,\dots,k\} \setminus I_2\big) - 2\varepsilon\,,$$

and by the induction hypothesis, we obtain

$$\liminf_{n\to\infty} p_2(n) \geq P\big((X_1,\dots,X_k) \in C\big)P\big(b_1-\delta < X_1 < b_1+\delta,\, X_i \leq b_i \text{ for } i \in I_2 \setminus \{1\},$$
$$X_i < b_i \text{ for } i \in \{1,\dots,k\} \setminus I_2\big) - 2\varepsilon$$
$$\geq P\big((X_1,\dots,X_k) \in C\big)P\big(X_1 = b_1,\, X_i \leq b_i \text{ for } i \in I_2 \setminus \{1\},$$
$$X_i < b_i \text{ for } i \in \{1,\dots,k\} \setminus I_2\big) - 2\varepsilon\,.$$

Letting $\varepsilon \to 0$ we see that

$$\liminf_{n\to\infty} p_2(n) \geq P\big((X_1,\dots,X_k) \in C\big)P\big(X_1 = b_1,\, X_i \leq b_i \text{ for } i \in I_2 \setminus \{1\},$$
$$X_i < b_i \text{ for } i \in \{1,\dots,k\} \setminus I_2\big)\,,$$

which establishes a lower bound matching (2.17) and hence completes the inductive argument. Therefore, we have proved that (2.15) holds when the sets C and D are "southwest corners" of the form (2.16).

The next step is to show that (2.15) holds when C and D are "rectangles" of the form

$$C = \Big\{(x_1,x_2,\dots,x_k) :\; a_i^{(1)} \leq x_i \leq a_i^{(2)} \text{ for } i \in I_{11},\;\; a_i^{(1)} \leq x_i < a_i^{(2)} \text{ for } i \in I_{12},$$
$$a_i^{(1)} < x_i \leq a_i^{(2)} \text{ for } i \in I_{13},\;\; a_i^{(1)} < x_i < a_i^{(2)} \text{ for } i \in I_{14}\Big\}, \tag{2.18}$$

$$B = \Big\{(x_1,x_2,\dots,x_k) :\; b_i^{(1)} \leq x_i \leq b_i^{(2)} \text{ for } i \in I_{21},\;\; b_i^{(1)} \leq x_i < b_i^{(2)} \text{ for } i \in I_{22},$$
$$b_i^{(1)} < x_i \leq b_i^{(2)} \text{ for } i \in I_{23},\;\; b_i^{(1)} < x_i < b_i^{(2)} \text{ for } i \in I_{24}\Big\}\,,$$

where both $(I_{11}, I_{12}, I_{13}, I_{14})$ and $(I_{21}, I_{22}, I_{23}, I_{24})$ are partitions of $\{1, \ldots, k\}$, and $-\infty \le a_i^{(1)} \le a_i^{(2)} \le \infty$, $i = 1, \ldots, k$, and $b_i^{(1)} \le b_i^{(2)}$, $i = 1, \ldots, k$, are real numbers. For notational ease, we will consider only the case $I_{13} = I_{23} = \{1, \ldots, k\}$, but the other cases are similar.

It is easy to check that

$$P\big((X_1, \ldots, X_k) \in C,\ (X_{n+1}, \ldots, X_{n+k}) \in D\big) \tag{2.19}$$

$$= \sum_{J_1, J_2} (-1)^{|J_1|+|J_2|} P\big(X_i \le a_i^{(1)} \text{ for } i \in J_1,\ X_i \le a_i^{(2)} \text{ for } i \in \{1, \ldots, k\} \setminus J_1,$$

$$X_{n+i} \le b_i^{(1)} \text{ for } i \in J_2,\ X_{n+i} \le b_i^{(2)} \text{ for } i \in \{1, \ldots, k\} \setminus J_2\big),$$

where the sum is taken over all subsets J_1 and J_2 of $\{1, \ldots, k\}$. The right-hand side of (2.19) is a finite sum of probabilities that we have already considered when we proved (2.15) for the "southwest corners." Therefore,

$$\lim_{n\to\infty} P\big((X_1, \ldots, X_k) \in C,\ (X_{n+1}, \ldots, X_{n+k}) \in D\big)$$

$$= \sum_{J_1, J_2} (-1)^{|J_1|+|J_2|} P\big(X_i \le a_i^{(1)} \text{ for } i \in J_1,\ X_i \le a_i^{(2)} \text{ for } i \in \{1, \ldots, k\} \setminus J_1\big)$$

$$P\big(X_i \le b_i^{(1)} \text{ for } i \in J_2,\ X_i \le b_i^{(2)} \text{ for } i \in \{1, \ldots, k\} \setminus J_2\big)$$

$$= \sum_{J_1} (-1)^{|J_1|} P\big(X_i \le a_i^{(1)} \text{ for } i \in J_1,\ X_i \le a_i^{(2)} \text{ for } i \in \{1, \ldots, k\} \setminus J_1\big)$$

$$\sum_{J_2} (-1)^{|J_2|} P\big(X_i \le b_i^{(1)} \text{ for } i \in J_2,\ X_i \le b_i^{(2)} \text{ for } i \in \{1, \ldots, k\} \setminus J_2\big)$$

$$= P\big((X_1, \ldots, X_k) \in C\big)\, P\big((X_1, \ldots, X_k) \in D\big),$$

where at the last step we used (2.19) once again. This proves (2.15) in the case that C and D are "rectangles" of the form (2.18).

Next, denote by $\mathcal{U}_k$ the collection of all disjoint finite unions of "rectangles" of the form (2.18). Note that $\mathcal{U}_k$ forms a field in $\mathbb{R}^k$ that generates the Borel σ-field. Since (2.15) holds for the "rectangles," it extends by linearity to the case that C and D are sets in $\mathcal{U}_k$. Furthermore, if C and D are arbitrary k-dimensional Borel sets, then given $\varepsilon > 0$, we can find sets C_1 and D_1 in $\mathcal{U}_k$ such that

$$P\big((X_1, \ldots, X_k) \in C \triangle C_1\big) \le \varepsilon,\ P\big((X_1, \ldots, X_k) \in D \triangle D_1\big) \le \varepsilon\,;$$

see again Corollary 1, p. 169, in Billingsley (1995). Then

$$\Big| P\big((X_1, \ldots, X_k) \in C,\ (X_{n+1}, \ldots, X_{n+k}) \in D\big)$$

$$- P\big((X_1, \ldots, X_k) \in C\big)\, P\big((X_1, \ldots, X_k) \in D\big) \Big|$$

$$\leq \Big| P\big((X_1,\ldots,X_k)\in C_1,\ (X_{n+1},\ldots,X_{n+k})\in D_1\big)$$
$$-P\big((X_1,\ldots,X_k)\in C_1\big)\,P\big((X_1,\ldots,X_k)\in D_1\big)\Big|$$
$$+\Big| P\big((X_1,\ldots,X_k)\in C,\ (X_{n+1},\ldots,X_{n+k})\in D\big)$$
$$-P\big((X_1,\ldots,X_k)\in C_1,\ (X_{n+1},\ldots,X_{n+k})\in D_1\big)\Big|$$
$$+\Big| P\big((X_1,\ldots,X_k)\in C_1\big)\,P\big((X_1,\ldots,X_k)\in D_1\big)$$
$$-P\big((X_1,\ldots,X_k)\in C\big)\,P\big((X_1,\ldots,X_k)\in D\big)\Big|$$
$$\leq \Big| P\big((X_1,\ldots,X_k)\in C_1,\ (X_{n+1},\ldots,X_{n+k})\in D_1\big)$$
$$-P\big((X_1,\ldots,X_k)\in C_1\big)\,P\big((X_1,\ldots,X_k)\in D_1\big)\Big| + 4\varepsilon\,.$$

Since we have proved that (2.15) holds for sets in $\mathcal{U}_k$, we conclude that

$$\limsup_{n\to\infty}\Big| P\big((X_1,\ldots,X_k)\in C,\ (X_{n+1},\ldots,X_{n+k})\in D\big)$$
$$-P\big((X_1,\ldots,X_k)\in C\big)\,P\big((X_1,\ldots,X_k)\in D\big)\Big| \leq 4\varepsilon$$

for every $\varepsilon > 0$, and letting $\varepsilon \to 0$ shows that (2.15) holds for arbitrary k-dimensional Borel sets.

Now that (2.15) has been established in its full generality, mixing of the process $\big(X_n,\ n\in\mathbb{Z}\big)$ follows because the statement (2.13) holds. To show that this is true, one approximates arbitrary sets in $\mathcal{B}^{\mathbb{Z}}$ by finite-dimensional cylindrical sets, as in Example 2.2.3. □

Example 2.2.8. An immediate conclusion of Theorem 2.2.7 is that a stationary Gaussian process $\big(X_n,\ n\in\mathbb{Z}\big)$ is mixing if and only if its correlation function asymptotically vanishes: $\rho_n := \mathrm{Corr}(X_{j+n}, X_j) \to 0$ as $n\to\infty$.

Example 2.2.9. We now take a second look at the tail σ-field $\mathcal{T}$ of a stationary stochastic process. We saw in Example 2.1.4 that if this σ-field is trivial, then the process is ergodic. We will show now that a trivial tail σ-field implies mixing as well. This statement is an immediate corollary of the following characterization: the tail σ-field $\mathcal{T}$ of a stationary process is trivial if and only if for every set A in $\mathcal{B}^{\mathbb{Z}}$,

$$\lim_{n\to\infty}\ \sup_{B\in\sigma(x_0,x_1,\ldots)} \big|\mu\big(A\cap\theta^{-n}B\big) - \mu(A)\mu(B)\big| = 0\,, \tag{2.20}$$

where μ is the law of the stationary process on $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}})$. This statement makes it possible to view the triviality of the tail σ-field as a kind of uniform mixing and hence akin to the strong mixing properties of Section 2.3.

The sufficiency of (2.20) for triviality of the tail σ-field is clear: if the tail σ-field is not trivial, then there is a tail event $A \in \cap_{n=1}^{\infty}\sigma\big(x_n, x_{n+1}, \ldots\big)$ satisfying $0 < \mu(A) < 1$. This means that for every $n = 1, 2, \ldots$, the set $\theta^n A$ is in $\sigma(x_0, x_1, \ldots)$, so that

$$\sup_{B\in\sigma(x_0,x_1,\ldots)} \big|\mu\big(A\cap\theta^{-n}B\big) - \mu(A)\mu(B)\big| \geq \big|\mu\big(A\cap\theta^{-n}(\theta^n A)\big) - \mu(A)\mu(\theta^n A)\big|$$

$$= \mu(A) - (\mu(A))^2 > 0\,,$$

contradicting (2.20). On the other hand, suppose that the tail σ-field is trivial. For events A and B as in (2.20), define random variables $X = \mathbf{1}_A - \mu(A)$ and $Y_B = \mathbf{1}_{\theta^{-n}B} - \mu(B)$, and note that Y is measurable with respect to the σ-field $\sigma\big(x_n, x_{n+1}, \ldots\big)$. Therefore,

$$\begin{aligned}
&\sup_{B\in\sigma(x_0,x_1,\ldots)} \big|\mu\big(A\cap\theta^{-n}B\big) - \mu(A)\mu(B)\big| \\
&= \sup_{B\in\sigma(x_0,x_1,\ldots)} \big|E(XY_B)\big| = \sup_{B\in\sigma(x_0,x_1,\ldots)} \Big|E\Big(Y_B E\big(X\big|\sigma(x_n, x_{n+1}, \ldots)\big)\Big)\Big| \\
&\leq \Big[E\Big(E\big(X\big|\sigma(x_n, x_{n+1}, \ldots)\big)\Big)^2\Big]^{1/2} \sup_{B\in\sigma(x_0,x_1,\ldots)} (EY_B^2)^{1/2} \\
&\leq \Big[E\Big(E\big(X\big|\sigma(x_n, x_{n+1}, \ldots)\big)\Big)^2\Big]^{1/2}.
\end{aligned}$$

Since the σ-fields $\sigma\big(x_n, x_{n+1}, \ldots\big)$ decrease to the trivial σ-field $\mathcal{T}$, it follows that with probability 1,

$$E\big(X\big|\sigma(x_n, x_{n+1}, \ldots)\big) \to E\big(X\big|\mathcal{T}\big) = EX = 0\,;$$

see, e.g., Theorem 35.9 in Billingsley (1995). Since the random variable $|X|$ is, furthermore, bounded by 1, the convergence in (2.20) follows.

Let us consider once again a nonsingular measure-preserving map ϕ on a probability space $\big(E, \mathcal{E}, m\big)$. Notice that the two ergodic-theoretical properties of ϕ we have already discussed, ergodicity and mixing, can be stated as

$$\phi \text{ is } \begin{cases} \text{mixing if } m\big(A\cap\phi^{-n}B\big) - m(A)\,m(B) \to 0,\ A, B \in \mathcal{E}, \\ \text{ergodic if } \frac{1}{n}\sum_{j=0}^{n-1}\big(m\big(A\cap\phi^{-j}B\big) - m(A)\,m(B)\big) \to 0 A, B \in \mathcal{E}; \end{cases} \tag{2.21}$$

in fact, the first line in (2.21) is just the definition of mixing. Let us check the second line in (2.21). Sufficiency of the condition presented there is clear: if ϕ were not ergodic, then choosing $A = B$ to be an invariant set whose measure takes a value in $(0, 1)$ would provide a counterexample to the condition in (2.21). Let us check the necessity of this condition. Suppose that ϕ is ergodic. By the pointwise ergodic theorem,

$$\frac{1}{n}\sum_{j=0}^{n-1}\mathbf{1}_B(\phi^j x) \to m(B)$$

as $n \to \infty$ for m-almost every $x \in E$. By the bounded convergence theorem, the integral over the set A of the left-hand side above converges to the integral of the right-hand side over the same set, and this gives the condition in (2.21).

Since the usual convergence of a sequence implies its Cesaro convergence, (2.21) provides yet another explanation of the fact that mixing is a stronger property than ergodicity. Furthermore, (2.21) makes it clear that there is an intermediate property of probability measure-preserving maps, weaker than mixing but stronger than ergodicity. We introduce this notion in the following definition.

Definition 2.2.10. A nonsingular measure-preserving map ϕ on a probability space $(E, \mathcal{E}, m)$ is called weakly mixing if for every two sets $A, B \in \mathcal{E}$,

$$\frac{1}{n}\sum_{j=0}^{n-1}\left|m\big(A \cap \phi^{-j}B\big) - m(A)\,m(B)\right| \to 0\,.$$

If $(X_n,\ n \in \mathbb{Z})$ is a stationary stochastic process, then we say that it is weakly mixing if the left shift θ is a weakly mixing map on the probability space $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu)$, where μ is the probability measure generated by the process on $\mathbb{R}^{\mathbb{Z}}$.

By definition,

$$\text{mixing} \;\Rightarrow\; \text{weak mixing} \;\Rightarrow\; \text{ergodicity}\,,$$

either for stationary stochastic processes or for measure-preserving maps. However, neither of these two implications can, in general, be reversed. It is easy to construct an example of an ergodic map that is not weakly mixing.

Example 2.2.11. An ergodic but not weakly mixing map.

Consider the two-point left shift-invariant probability measure (2.3) considered in Example 2.2.2. The left shift θ is ergodic on that probability space, but is not mixing. It is not weakly mixing either, since for A and B as in Example 2.2.2, we have

$$\left|m\big(A \cap \theta^{-n}B\big) - m(A)\,m(B)\right| = \frac{1}{4}$$

for each n, and the Cesaro limit of these numbers is equal to $1/4$, not to 0.

We will see in the sequel examples of weakly mixing stationary processes that fail to be mixing.

An alternative point of view on weak mixing is based on the notion of convergence in density. Recall that a subset K of positive integers $\mathbb{N}$ is said to have density zero if

$$\lim_{n\to\infty}\frac{\left|K \cap \{1,2,\ldots,n\}\right|}{n} = 0\,.$$

A sequence (b_n) in a metric space with a metric d is said to converge in density to b if there is a set K of density zero such that

$$\lim_{n\to\infty,\, n\notin K} b_n = b\,.$$

Explicitly, for every $\varepsilon > 0$, there is $N = N(\varepsilon)$ such that for every $n > N$, $n \notin K$, $d(b_n, b) \le \varepsilon$. The following useful lemma is even more explicit.

Lemma 2.2.12. *A sequence* (b_n) *converges in density to b if and only if for every* $\varepsilon > 0$, *the set* $K_\varepsilon = \{n : d(b_n, b) > \varepsilon\}$ *has density zero.*

Proof. The necessity in the statement is clear, so let us check the sufficiency part. Assume that for every $\varepsilon > 0$, the set K_ε has density zero. The task is to construct a set of density zero away from which the sequence (b_n) converges to b.

We define an increasing sequence of nonnegative integers as follows. Let $N_0 = 0$, and for $k \ge 1$, let

$$N_k = \sup\Bigl\{n \ge N_{k-1} + 1 : \Bigl|\{j = 1,\dots,n : d(b_j, b) > \frac{1}{k}\}\Bigr| > \frac{n}{2^k}\Bigr\}\,.$$

By the assumption, each N_k is a finite number. Let

$$K = \bigcup_{k=1}^{\infty}\{j : j = N_k + 1,\dots,N_{k+1},\, d(b_j, b) > \frac{1}{k}\}\,.$$

By the definition, if $n \notin K$ and $n > N_k$, then $d(b_n, b) \le \frac{1}{k}$, for $k = 1, 2, \dots$. Therefore, $\lim_{n\to\infty,\, n\notin K} b_n = b$. To check that K has density zero, let k_i be the largest k such that $N_k \le i$, and notice that $k_i \to \infty$ as $i \to \infty$. Choose $m \ge 1$ and let n be so large that $k_n > m$. Then

$$\begin{aligned}
\bigl|\bigl(K \cap \{1,\dots,n\}\bigr)\bigr| &\le N_m + \sum_{k=m}^{k_n-1}\Bigl|\Bigl(\{j : j = N_k + 1,\dots,N_{k+1},\, d(b_j, b) > \frac{1}{k}\}\Bigr)\Bigr| \\
&\quad + \Bigl|\Bigl(\{j : j = N_{k_n} + 1,\dots,n,\, d(b_j, b) > \frac{1}{k_n}\}\Bigr)\Bigr| \\
&\le N_m + \sum_{k=m}^{k_n-1}\Bigl|\Bigl(\{j : j = 1,\dots,N_{k+1},\, d(b_j, b) > \frac{1}{k}\}\Bigr)\Bigr| \\
&\quad + \Bigl|\Bigl(\{j : j = 1,\dots,n,\, d(b_j, b) > \frac{1}{k_n}\}\Bigr)\Bigr| \\
&\le N_m + \sum_{k=m}^{k_n-1}\frac{N_{k+1}}{2^k} + \frac{n}{2^{k_n}} \le N_m + \sum_{k=m}^{k_n}\frac{n}{2^k} \le N_m + \frac{n}{2^{m-1}}\,.
\end{aligned}$$

Therefore,

$$\limsup_{n\to\infty} \frac{\big|K \cap \{1,2,\ldots,n\}\big|}{n} \le \frac{1}{2^{m-1}}\,,$$

and letting $m \to \infty$ shows that K has density zero. □

The following proposition is a simple consequence.

Proposition 2.2.13. *A nonsingular measure-preserving map ϕ on a probability space $\big(E,\mathcal{E},m\big)$ is weakly mixing if and only if for every two sets $A,B \in \mathcal{E}$,*

$$m\big(A \cap \phi^{-n}B\big) \to m(A)\,m(B) \ \ \textit{in density.}$$

Proof. The sufficiency for weak mixing of the condition in the proposition is clear. Let us check the necessity. Suppose that the condition in the proposition fails. By Lemma 2.2.12, there is $\varepsilon > 0$ such that the set

$$K_\varepsilon = \Big\{n:\ \big|m\big(A \cap \phi^{-j}B\big) - m(A)\,m(B)\big| > \varepsilon\Big\}$$

does not have density zero. Since

$$\frac{1}{n}\sum_{j=0}^{n-1}\big|m\big(A \cap \phi^{-j}B\big) - m(A)\,m(B)\big|$$

$$\ge \frac{1}{n}\sum_{j\in K_\varepsilon\cap\{0,\ldots,n-1\}}\big|m\big(A \cap \phi^{-j}B\big) - m(A)\,m(B)\big| \ge \varepsilon\,\frac{\big|K_\varepsilon \cap \{1,2,\ldots,n\}\big|}{n}\,,$$

which does not converge to zero, ϕ cannot be weakly mixing, and the proof is complete. □

Proposition 2.2.4 has an immediate counterpart for weak mixing. The argument is the same; just use Proposition 2.2.13 and replace the usual convergence by convergence away from a set of density zero.

Proposition 2.2.14. *Let $\phi : E \to E$ be a nonsingular measure-preserving map on a probability space $\big(E,\mathcal{E},m\big)$. Let $\varphi : G \to G$ be a one-to-one and onto map on a measurable space $\big(G,\mathcal{G}\big)$ such that both φ and its inverse are measurable. Let $f : E \to G$ be a measurable map that is compatible with ϕ and φ. If ϕ is weakly mixing on $\big(E,\mathcal{E},m\big)$, then φ is weakly mixing on $\big(G,\mathcal{G},m\circ f^{-1}\big)$.*

As in the case of ergodicity and mixing, the following corollary is immediate.

Corollary 2.2.15. *Let $\big(X_n,\ n \in \mathbb{Z}\big)$ be a weakly mixing stationary stochastic process, and let $\big(Y_n,\ n \in \mathbb{Z}\big)$ be a stationary process given by* (2.12) *for some measurable function $g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$. Then $\big(Y_n,\ n \in \mathbb{Z}\big)$ is weakly mixing as well.*

Many of the important stochastic processes are constructed as sums of independent, and more elementary, stochastic processes. Therefore, it would be nice to be able to obtain ergodicity of a sum based on the ergodicity of the summands. Unfortunately, this is impossible to do in general, since the sum of two independent stationary ergodic processes does not have to be ergodic; see Exercise 2.6.7. It turns out, however, that strengthening the assumption on one of the two summands from ergodicity to weak mixing is sufficient to guarantee the ergodicity of the sum.

Theorem 2.2.16. *Let $(X_n,\ n \in \mathbb{Z})$ and $(Y_n,\ n \in \mathbb{Z})$ be independent stationary stochastic processes. Assume that $(X_n,\ n \in \mathbb{Z})$ is ergodic, and that $(Y_n,\ n \in \mathbb{Z})$ is weakly mixing. Then the process $Z_n = X_n + Y_n$, $n \in \mathbb{Z}$ is ergodic.*

Proof. Let μ be the probability measure generated by the process $(X_n,\ n \in \mathbb{Z})$ on $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}})$, and let ν be the measure generated by the process $(Y_n,\ n \in \mathbb{Z})$. Consider the product probability space, $(\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}} \times \mathcal{B}^{\mathbb{Z}}, \mu \times \nu)$. The left shift θ on $\mathbb{R}^{\mathbb{Z}}$ extends naturally to the product shift $\theta \times \theta$ operating on the product space: $(\theta \times \theta)(\mathbf{x}, \mathbf{y}) = (\theta\mathbf{x}, \theta\mathbf{y})$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathbb{Z}}$. This is clearly a nonsingular map that preserves the product probability measure. Note that the mapping $f : \mathbb{R}^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ defined by $f(\mathbf{x}, \mathbf{y}) = \mathbf{x} + \mathbf{y}$ (coordinatewise addition) is compatible with the maps $\theta \times \theta$ on $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}}$ and θ on $\mathbb{R}^{\mathbb{Z}}$. Furthermore, the measure $(\mu \times \nu) \circ f^{-1}$ that f induces on $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}})$ is the law of the process $Z_n = X_n + Y_n$, $n \in \mathbb{Z}$. Therefore, in order to prove ergodicity of the latter process, it is sufficient, by Proposition 2.1.7, to prove that the product left shift $\theta \times \theta$ is itself ergodic.

To this end, we will use the criterion for ergodicity given in (2.21). We need to prove that for every two sets $A, B \in \mathcal{B}^{\mathbb{Z}} \times \mathcal{B}^{\mathbb{Z}}$,

$$\frac{1}{n} \sum_{j=0}^{n-1} (\mu \times \nu)\big(A \cap (\theta \times \theta)^{-j} B\big) = (\mu \times \nu)(A)\,(\mu \times \nu)(B) \tag{2.22}$$

as $n \to \infty$. We begin with the case that A and B are measurable rectangles of the type $A = A_1 \times A_2$ and $B = B_1 \times B_2$, for A_1, A_2, B_1, B_2 measurable sets in $\mathcal{B}^{\mathbb{Z}}$. In that case, the left-hand side of (2.22) becomes

$$\frac{1}{n} \sum_{j=0}^{n-1} \mu\big(A_1 \cap \theta^{-j} B_1\big) \nu\big(A_2 \cap \theta^{-j} B_2\big). \tag{2.23}$$

Since the left shift θ is weakly mixing on the probability space $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \nu)$, we know that

$$\nu\big(A_2 \cap \theta^{-n} B_2\big) \to \nu(A_2)\nu(B_2)$$

in density. Since a set of zero density cannot affect the Cesaro limit of a bounded sequence, the limit of the expression in (2.23) is equal to the limit of the expression

$$\frac{1}{n} \sum_{j=0}^{n-1} \mu\big(A_1 \cap \theta^{-j} B_1\big) \nu(A_2)\nu(B_2)$$

(and the two limits exist at the same time). Since the left shift θ is ergodic on the probability space $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu)$, this last limit exists and is equal to

$$\mu(A_1)\mu(B_1)\nu(A_2)\nu(B_2) = (\mu \times \nu)(A)\,(\mu \times \nu)(B)\,,$$

and so we have checked that (2.22) holds when A and B are measurable rectangles. Now we can proceed as in the proof of Theorem 2.2.7. Since (2.22) holds for measurable rectangles, it extends, by linearity, to the case that A and B are each a finite disjoint union of measurable rectangles. Since the latter sets form a field generating the product σ-field $\mathcal{B}^{\mathbb{Z}} \times \mathcal{B}^{\mathbb{Z}}$, we can approximate general sets A and B arbitrarily well (with respect to the product measure $\mu \times \nu$) by finite disjoint unions of measurable rectangles, and this shows that (2.22) holds in full generality. □

We now present an expected counterpart of Theorem 2.2.7, which says that weak mixing of a stationary stochastic process is equivalent to weak convergence *in density* to independence of the joint distributions of the blocks of observations separated by long periods of time. Notice that the weak convergence of probability measures on a Euclidian space is metrizable, for example by the Prokhorov metric; see, e.g., Billingsley (1999).

Theorem 2.2.17. *A stationary process $(X_n,\ n \in \mathbb{Z})$ is weakly mixing if and only if for every $k = 1, 2, \ldots$,*

$$\left(X_1, \ldots, X_k, X_{n+1}, \ldots, X_{n+k}\right) \Rightarrow \left(X_1, \ldots, X_k, Y_1, \ldots, Y_k\right) \tag{2.24}$$

in density, where $(Y_1, \ldots, Y_k)$ is an independent copy of $(X_1, \ldots, X_k)$.

Proof. Sufficiency of (2.24) for weak mixing can be proved in exactly the same way as the sufficiency part in Theorem 2.2.7; the latter proof does not differentiate between the usual convergence and convergence in density. We prove now the necessity of (2.24) for weak mixing.

Fix k, and recall that

$$d(\mu_1, \mu_2) = \sup_{\boldsymbol{\theta} \in \mathbb{R}^{2k}} \frac{|\hat{\mu}_1(\boldsymbol{\theta}) - \hat{\mu}_2(\boldsymbol{\theta})|}{1 + \|\boldsymbol{\theta}\|}$$

is a also metric on the space of probability measures on $\mathbb{R}^{2k}$ that metrizes weak convergence. Here $\hat{\mu}$ is the characteristic function of a probability measure μ. If we denote by ν_n the law of the random vector on the left-hand side of (2.24), and by ν the law of the random vector on the right-hand side, then by Lemma 2.2.12, we need to show only that for every $\varepsilon > 0$, the set $K_\varepsilon = \{n : d(\nu_n, \nu) > \varepsilon\}$ has density zero.

Fix, therefore, $\varepsilon > 0$, and choose $N = 1, 2, \ldots$, $\theta_0 > 0$, and $m = 1, 2, \ldots$ such that

$$P(|X_0| > N) \le \frac{\varepsilon}{12k}, \quad \theta_0 \ge \frac{2}{\varepsilon}, \quad \frac{\theta_0\sqrt{2k}}{m} \le \frac{\varepsilon}{6}\,.$$

In the following decomposition, the probability measure μ is either ν_n or ν. Note that in either case, the one-dimensional marginals of μ are the same, and they coincide with the law of X_0. We may assume that the law of X_0 puts no mass on the rational numbers (otherwise, a global scale change of the process would have this property). We write

$$\hat{\mu}(\boldsymbol{\theta}) = \int_{\mathbf{x}\in T_N} e^{i(\boldsymbol{\theta},\mathbf{x})}\,\mu(d\mathbf{x}) + \sum_{j_1=-m}^{m-1}\cdots\sum_{j_{2k}=-m}^{m-1}\int_{\mathbf{x}\in I_{N,m}(j_1,\ldots,j_{2k})} e^{i(\boldsymbol{\theta},\mathbf{x})}\,\mu(d\mathbf{x}),$$

where

$$T_N = \{\mathbf{x} = (x_1,\ldots,x_{2k}) \in \mathbb{R}^{2k} : |x_i| > N \text{ for some } i = 1,\ldots,2k\},$$

$$I_{N,m}(j_1,\ldots,j_{2k}) = \{\mathbf{x} = (x_1,\ldots,x_{2k}) \in \mathbb{R}^{2k} :$$
$$Nj_i/m < x_i < N(j_i+1)/m,\ i = 1,\ldots,2k\},$$

$j_i = -m,\ldots,m-1,\ i = 1,\ldots,2k$. Note that by the choice of N,

$$\left|\int_{\mathbf{x}\in T_N} e^{i(\boldsymbol{\theta},\mathbf{x})}\,\mu(d\mathbf{x})\right| \le \frac{\varepsilon}{6}. \tag{2.25}$$

Furthermore, for each $j_i = -m,\ldots,m-1,\ i = 1,\ldots,2k$, there is a zero-density subset $K_{N,m}(j_1,\ldots,j_{2k})$ of positive integers such that for all n outside of this set,

$$\left|\nu_n\big(I_{N,m}(j_1,\ldots,j_{2k})\big) - \nu\big(I_{N,m}(j_1,\ldots,j_{2k})\big)\right| \le \frac{\varepsilon}{6(2m)^{2k}}. \tag{2.26}$$

We define K_ε to be the union of the sets $K_{N,m}(j_1,\ldots,j_{2k})$ over all $j_i = -m,\ldots,m-1$, $i = 1,\ldots,2k$. Clearly, K_ε is a zero-density set. It remains to show that outside of this set, we have $d(\nu_n,\nu) \le \varepsilon$.

For each $j_i = -m,\ldots,m-1,\ i = 1,\ldots,2k$, we can write

$$\int_{\mathbf{x}\in I_{N,m}(j_1,\ldots,j_{2k})} e^{i(\boldsymbol{\theta},\mathbf{x})}\,\mu(d\mathbf{x})$$
$$= \Big[\cos\big(\boldsymbol{\theta},\mathbf{x}^{(1)}(\boldsymbol{\theta},\mu,j_1,\ldots,j_{2k})\big) + i\sin\big(\boldsymbol{\theta},\mathbf{x}^{(2)}(\boldsymbol{\theta},\mu,j_1,\ldots,j_{2k})\big)\Big]\mu\big(I_{N,m}(j_1,\ldots,j_{2k})\big)$$

for some points $\mathbf{x}^{(d)}(\boldsymbol{\theta},\mu,j_1,\ldots,j_{2k}) \in I_{N,m}(j_1,\ldots,j_{2k})$, $d = 1,2$. Applying the resulting decomposition of a characteristic function to the measures ν_n and ν, we obtain for $\boldsymbol{\theta} \in \mathbb{R}^k$ with $\|\boldsymbol{\theta}\| \le \theta_0$ and $n \notin K_\varepsilon$, using (2.25),

$$|\hat{\nu}_n(\boldsymbol{\theta}) - \hat{\nu}(\boldsymbol{\theta})| \le \frac{\varepsilon}{3}$$
$$+ \sum_{j_1=-m}^{m-1}\cdots\sum_{j_{2k}=-m}^{m-1}\left|\int_{\mathbf{x}\in I_{N,m}(j_1,\ldots,j_{2k})} e^{i(\boldsymbol{\theta},\mathbf{x})}\,\nu_n(d\mathbf{x}) - \int_{\mathbf{x}\in I_{N,m}(j_1,\ldots,j_{2k})} e^{i(\boldsymbol{\theta},\mathbf{x})}\,\nu(d\mathbf{x})\right|.$$

Each term in the sum on the right-hand side can be bounded from above by

$$\Big|\cos\big(\boldsymbol{\theta}, \mathbf{x}^{(1)}(\boldsymbol{\theta}, \nu_n, j_1, \ldots, j_{2k})\big)\nu_n\big(I_{N,m}(j_1, \ldots, j_{2k})\big) \\ - \cos\big(\boldsymbol{\theta}, \mathbf{x}^{(1)}(\boldsymbol{\theta}, \nu, j_1, \ldots, j_{2k})\big)\nu\big(I_{N,m}(j_1, \ldots, j_{2k})\big)\Big|$$

plus the corresponding term with the sin function replacing the cos function. Note that each of the two terms can be bounded by

$$\begin{aligned} &\nu\big(I_{N,m}(j_1, \ldots, j_{2k})\big)\|\boldsymbol{\theta}\|\,\big\|\mathbf{x}^{(1)}(\boldsymbol{\theta}, \nu_n, j_1, \ldots, j_{2k}) - \mathbf{x}^{(1)}(\boldsymbol{\theta}, \nu, j_1, \ldots, j_{2k})\big\| \\ &\quad + \Big|\nu_n\big(I_{N,m}(j_1, \ldots, j_{2k})\big) - \nu\big(I_{N,m}(j_1, \ldots, j_{2k})\big)\Big| \\ &\le \nu\big(I_{N,m}(j_1, \ldots, j_{2k})\big)\theta_0\frac{\sqrt{2k}}{m} + \frac{\varepsilon}{6(2m)^{2k}} \\ &\le \frac{\varepsilon}{6}\nu\big(I_{N,m}(j_1, \ldots, j_{2k})\big) + \frac{\varepsilon}{6(2m)^{2k}} \end{aligned}$$

by (2.26) and the choice of m. Summarizing, we conclude that

$$|\hat{\nu}_n(\boldsymbol{\theta}) - \hat{\nu}(\boldsymbol{\theta})| \le \varepsilon$$

for all $\boldsymbol{\theta} \in \mathbb{R}^k$ with $\|\boldsymbol{\theta}\| \le \theta_0$ and $n \notin K_\varepsilon$. By the choice of θ_0, this shows that $d(\nu_n, \nu) \le \varepsilon$ for each $n \notin K_\varepsilon$, and the proof is complete. □

Example 2.2.18. Let $\mathbf{X} = \big(X_n,\ n \in \mathbb{Z}\big)$ be a stationary centered Gaussian process. Let μ_X be its spectral measure, i.e., a measure on $(-\pi, \pi]$ such that the covariance function of the process satisfies

$$R_X(n) = \int_{(-\pi,\pi]} e^{inx}\, \mu_X(dx),\ n \in \mathbb{Z};$$

see Theorem 1.2.1. Complementing the characterization of mixing of stationary Gaussian processes in Example 2.2.8, we will show now that $\mathbf{X}$ is weakly mixing if and only if it is ergodic, and that a necessary and sufficient condition for that is that the spectral measure μ_X be atomless.

Since weak mixing implies ergodicity, we need to check only that the presence of an atom in μ_X rules out ergodicity of $\mathbf{X}$, while the absence of atoms implies weak mixing of $\mathbf{X}$.

Suppose that $a \in (-\pi, \pi]$ is an atom of μ_X, i.e., $\mu_X(\{a\}) > 0$. Let μ_1 be a measure on $(-\pi, \pi]$ obtained by removing from μ_X the atoms at $-a$ and a if $a \in (-\pi, \pi)$, or only the atom at a if $a = \pi$. Let μ_2 be the measure on $(-\pi, \pi]$ consisting of these removed atoms. Note that $\mu_X = \mu_1 + \mu_2$, and μ_2 is a nonzero measure. If $\mathbf{X}^{(j)}$ is a stationary centered Gaussian process with spectral measure μ_j, $j = 1, 2$, and if $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are independent, then $\mathbf{X}$ has the same distribution as

$\mathbf{X}^{(1)} + \mathbf{X}^{(2)}$, so lack of ergodicity of the process $\mathbf{X}^{(2)}$ and Proposition 2.1.6 would imply that the process $\mathbf{X}$ is not ergodic. To see that the process $\mathbf{X}^{(2)}$ is not ergodic, observe that the covariance function of this process has the form

$$\operatorname{Cov}\bigl(X_0^{(2)}, X_n^{(2)}\bigr) = \cos(an)\|\mu_2\|, \ n \in \mathbb{Z},$$

so that

$$\bigl(X_n^{(2)}, \, n \in \mathbb{Z}\bigr) \stackrel{d}{=} \Bigl(\|\mu_2\|^{1/2}\bigl(G_1 \cos(an) + G_2 \sin(an)\bigr), \, n \in \mathbb{Z}\Bigr),$$

where G_1 and G_2 are independent standard normal random variables. This shows that for the invariant set

$$A = \bigl\{\sup_{n\in\mathbb{Z}} |x_n| > 1\bigr\},$$

we have $P\bigl(\mathbf{X}^{(2)} \in A\bigr) \in (0,1)$, and hence the process $\mathbf{X}^{(2)}$ is not ergodic.

Suppose now that the spectral measure μ_X is atomless. By Fubini's theorem, we conclude that the finite measure on $(-2\pi, 2\pi]$ given by a convolution of μ_X with itself, $F = \mu_X * \mu_X$, does not have atoms at the origin or at 2π. Note further that

$$R_X(n)^2 = \int_{(-2\pi,2\pi]} e^{inx} \, F(dx), \ n \in \mathbb{Z}.$$

Therefore, for $n \geq 1$,

$$\frac{1}{n}\sum_{j=0}^{n-1} R_X(j)^2 = \int_{(-2\pi,2\pi]} \frac{1}{n}\sum_{j=0}^{n-1} e^{ijx} \, F(dx).$$

The functions

$$\varphi_n(x) = \frac{1}{n}\sum_{j=0}^{n-1} e^{ijx}, \ \ x \in (-2\pi, 2\pi],$$

are uniformly bounded and converge, as $n \to \infty$, to the function $\mathbf{1}\bigl(x = 0 \text{ or } x = 2\pi\bigr)$, which is equal to zero F-a.e. By the bounded convergence theorem, we conclude that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} R_X(j)^2 = 0,$$

and therefore, $R_X(n)$ converges to zero in density as $n \to \infty$. By Theorem 2.2.17, this implies that the process $\mathbf{X}$ is weakly mixing.

2.3 Strong Mixing

In the previous section, we introduced the notions of mixing and weak mixing for stationary stochastic processes. Perhaps unexpectedly, there exist notions of strong mixing of stationary processes that are different from the notion of mixing. These notions of strong mixing are not, strictly speaking, ergodic-theoretical notions in the sense of being properties of nonsingular measure-preserving maps. They are, rather, properties of certain families of σ-fields generated by a stationary process. The ergodic-theoretical notion of mixing of a stationary process turns out to be too weak to be directly applicable in certain limit theorems for stationary processes, and it is natural to view the strong mixing properties as strengthening the notion of mixing in ways that are suitable for different purposes. We discuss several of the common strong mixing notions in this section. An important point to keep in mind is that the notions we present (and a number of notions that we do not present) are collectively known as strong mixing properties. This is in spite of the fact that one of these properties is itself known as the strong mixing property. This is confusing, since often only the difference between the grammatical singular and the grammatical plural clarifies the meaning. Still, this is the existing usage, and we will adhere to this language here.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\mathcal{A}$ and $\mathcal{B}$ two sub-σ-fields of $\mathcal{F}$. We define

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} \left| P(A \cap B) - P(A)P(B) \right| . \tag{2.27}$$

Clearly, this is a measure of dependence between the σ-fields $\mathcal{A}$ and $\mathcal{B}$, and $\alpha(\mathcal{A}, \mathcal{B}) = 0$ if and only if the σ-fields $\mathcal{A}$ and $\mathcal{B}$ are independent. The following lemma lists several elementary properties of this measure of dependence.

Lemma 2.3.1. *For all sub-σ-fields $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{F}$,*

$$0 \le \alpha(\mathcal{A}, \mathcal{B}) \le \frac{1}{4} .$$

Further,

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{B \in \mathcal{B}} E\big(P(B|\mathcal{A}) - P(B)\big)_+ = \frac{1}{2} \sup_{B \in \mathcal{B}} E\big|P(B|\mathcal{A}) - P(B)\big| . \tag{2.28}$$

Proof. The first claim of the lemma follows from the Cauchy–Schwarz inequality: for any two events A and B,

$$\left| P(A \cap B) - P(A)P(B) \right| = |\mathrm{Cov}(\mathbf{1}_A, \mathbf{1}_B)| \le \sqrt{\mathrm{Var}(\mathbf{1}_A)\mathrm{Var}(\mathbf{1}_B)} \le \frac{1}{4} .$$

For the second claim, fix $B \in \mathcal{B}$, and note that

$$\sup_{A\in\mathcal{A}} \left|P(A\cap B) - P(A)P(B)\right|$$

$$= \max\Big(\sup_{A\in\mathcal{A}} \big(P(A\cap B) - P(A)P(B)\big)_+, \sup_{A\in\mathcal{A}} \big(P(A\cap B) - P(A)P(B)\big)_-\Big).$$

Further,

$$\sup_{A\in\mathcal{A}} \big(P(A\cap B) - P(A)P(B)\big)_+ = \sup_{A\in\mathcal{A}} \big(E(\mathbf{1}_A\mathbf{1}_B) - E\mathbf{1}_A E\mathbf{1}_B\big)_+$$

$$= \sup_{A\in\mathcal{A}} \big(E(\mathbf{1}_A P(B|\mathcal{A})) - E\mathbf{1}_A P(B)\big)_+ = \sup_{A\in\mathcal{A}} \Big(E\big(\mathbf{1}_A (P(B|\mathcal{A})) - P(B)\big)\Big)_+$$

$$= E\big(P\big(B|\mathcal{A}\big) - P(B)\big)_+ ,$$

since the last supremum is obviously achieved on the event $A = \{P(B|\mathcal{A}) > P(B)\} \in \mathcal{A}$. Similarly,

$$\sup_{A\in\mathcal{A}} \big(P(A\cap B) - P(A)P(B)\big)_- = E\big(P\big(B|\mathcal{A}\big) - P(B)\big)_- .$$

Since for every random variable X with $EX = 0$ we have $EX_+ = EX_- = E|X|/2$, we obtain (2.28) after optimizing over $B \in \mathcal{B}$. □

Let now $\big(X_n,\, n \in \mathbb{Z}\big)$ be a stationary process defined on some probability space $\big(\Omega, \mathcal{F}, P\big)$. Then

$$\mathcal{F}_{-\infty}^0 = \sigma(X_k,\, k \le 0),\;\; \mathcal{F}_n^\infty = \sigma(X_k,\, k \ge n),\; n \ge 1, \tag{2.29}$$

are sub-σ-fields of $\mathcal{F}$, and we introduce the *strong mixing coefficient* of the process $\big(X_n,\, n \in \mathbb{Z}\big)$ by

$$\alpha_X(n) := \alpha\big(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty\big),\; n = 1, 2, \ldots . \tag{2.30}$$

Definition 2.3.2. A stationary process $\big(X_n,\, n \in \mathbb{Z}\big)$ is called strongly mixing if $\alpha_X(n) \to 0$ as $n \to \infty$.

A strongly mixing stationary process is alternatively known as α-mixing. In spite of defining, as above, the strong mixing property via certain sub-σ-fields in the probability space on which a process is defined, the presence or absence of this property is determined solely by the finite-dimensional distributions of the process (Exercise 2.6.9).

Example 2.3.3. It turns out that a strongly mixing stationary process $\big(X_n,\, n \in \mathbb{Z}\big)$ has a trivial tail σ-field. To see this, let μ be the measure generated by the process on $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}\big)$. In order to prove triviality of the tail σ-field, we need to check that (2.20) holds for every set A in $\mathcal{B}^{\mathbb{Z}}$. If $A \in \sigma\big(\ldots, x_{k-1}, x_k\big)$ for some $k \geq 0$, then $A_1 = \theta^k A \in \sigma\big(\ldots, x_{-1}, x_0\big)$, and so for every $B \in \sigma(x_0, x_1, \ldots)$, by stationarity,

$$\begin{aligned}&\big|\mu\big(A \cap \theta^{-n} B\big) - \mu(A)\mu(B)\big| \\ = &\big|\mu\big(A_1 \cap \theta^{-(n-k)} B\big) - \mu(A_1)\mu(B)\big| \leq \alpha_X(n-k)\,,\end{aligned}$$

and (2.20) holds by the α-mixing of the process. Since sets A as above form a field generating the σ-field $\mathcal{B}^{\mathbb{Z}}$, we can use our usual approximating argument to extend the validity of (2.20) to all A in $\mathcal{B}^{\mathbb{Z}}$. Therefore, the tail σ-field is indeed trivial.

Combining the lessons learned from Examples 2.2.9 and 2.3.3, we immediately obtain the following proposition showing that, as the terminology implies, strong mixing guarantees ergodic-theoretical mixing.

Proposition 2.3.4. *A strongly mixing stationary process is also mixing.*

To introduce the next measure of dependence between two σ-fields, let once again $\big(\Omega, \mathcal{F}, P\big)$ be a probability space, and $\mathcal{A}$ and $\mathcal{B}$ two sub-σ-fields of $\mathcal{F}$. Two natural probability measures on the product space $\big(\Omega \times \Omega, \mathcal{A} \times \mathcal{B}\big)$ are $P_{\rm id}^{\mathcal{A},\mathcal{B}}$ and $P_{\rm ind}^{\mathcal{A},\mathcal{B}}$. Here $P_{\rm id}^{\mathcal{A},\mathcal{B}}$ is the probability measure induced on the product space by the measurable map $T_{\rm id} : \Omega \to \Omega \times \Omega$ with $T_{\rm id}(\omega) = (\omega, \omega)$, $\omega \in \Omega$, and $P_{\rm ind}^{\mathcal{A},\mathcal{B}}$ is the product measure of the restrictions $P_{\mathcal{A}}$ and $P_{\mathcal{B}}$ of P to the σ-fields $\mathcal{A}$ and $\mathcal{B}$ respectively. Note that the two measures $P_{\rm id}^{\mathcal{A},\mathcal{B}}$ and $P_{\rm ind}^{\mathcal{A},\mathcal{B}}$ put the same marginal probability measures (equal to $P_{\mathcal{A}}$ and $P_{\mathcal{B}}$) on the two copies of Ω. We define

$$\beta(\mathcal{A}, \mathcal{B}) = \big\|P_{\rm id}^{\mathcal{A},\mathcal{B}} - P_{\rm ind}^{\mathcal{A},\mathcal{B}}\big\| = \sup_{C \in \mathcal{A} \times \mathcal{B}} \big|P_{\rm id}^{\mathcal{A},\mathcal{B}}(C) - P_{\rm ind}^{\mathcal{A},\mathcal{B}}(C)\big|\,, \tag{2.31}$$

the total variation distance between the probability measures $P_{\rm id}^{\mathcal{A},\mathcal{B}}$ and $P_{\rm ind}^{\mathcal{A},\mathcal{B}}$.

The following lemma lists several basic properties of the measure of dependence between σ-fields defined in (2.31).

Lemma 2.3.5. *Let $\mathcal{A}$ and $\mathcal{B}$ be sub-σ-fields of $\mathcal{F}$.*

(i) We have

$$0 \leq \beta(\mathcal{A}, \mathcal{B}) \leq 1$$

and

$$\beta(\mathcal{A}, \mathcal{B}) \geq 2\alpha(\mathcal{A}, \mathcal{B})\,. \tag{2.32}$$

(ii) An alternative way of representing $\beta(\mathcal{A}, \mathcal{B})$ is

$$\beta(\mathcal{A}, \mathcal{B}) = \frac{1}{2} \sup_{\mathcal{I}, \mathcal{J}} \sum_{i=1}^{I} \sum_{j=1}^{J} \big|P(A_i \cap B_j) - P(A_i)P(B_j)\big|\,, \tag{2.33}$$

where the supremum is taken over all finite partitions $\mathcal{I} = \{A_1, \dots, A_I\}$ *and* $\mathcal{J} = \{B_1, \dots, B_J\}$ *of* Ω *into* $\mathcal{A}$*-measurable sets and* $\mathcal{B}$*-measurable sets, respectively.*

(iii) $\beta(\mathcal{A}, \mathcal{B}) = 0$ *if and only if the* σ*-fields* $\mathcal{A}$ *and* $\mathcal{B}$ *are independent.*

Proof. The first claim in part (i) is an obvious property of the total variation distance between two probability measures. For the second claim in part (i), note that if $A \in \mathcal{A}$ and $B \in \mathcal{B}$, then $P_{\text{id}}^{\mathcal{A},\mathcal{B}}(A \times B) = P(A \cap B)$. If we define for such events A and B the sets $C_1(A, B) = A \times B$ and $C_2(A, B) = A^c \times B^c$, then these product sets are disjoint events in $\mathcal{A} \times \mathcal{B}$, and

$$\begin{aligned} P_{\text{id}}^{\mathcal{A},\mathcal{B}}\big(C_1(A,B)\big) - P_{\text{ind}}^{\mathcal{A},\mathcal{B}}\big(C_1(A,B)\big) &= P_{\text{id}}^{\mathcal{A},\mathcal{B}}\big(C_2(A,B)\big) - P_{\text{ind}}^{\mathcal{A},\mathcal{B}}\big(C_2(A,B)\big) \\ &= P(A \cap B) - P(A)P(B) \,. \end{aligned}$$

Therefore,

$$\begin{aligned} \beta(\mathcal{A}, \mathcal{B}) &\geq \sup_{A \in \mathcal{A}, B \in \mathcal{B}} \left| P_{\text{id}}^{\mathcal{A},\mathcal{B}}\big(C_1(A,B) \cup C_2(A,B)\big) - P_{\text{ind}}^{\mathcal{A},\mathcal{B}}\big(C_1(A,B) \cup C_2(A,B)\big) \right| \\ &= 2 \sup_{A \in \mathcal{A}, B \in \mathcal{B}} \left| P(A \cap B) - P(A)P(B) \right| = 2\alpha(\mathcal{A}, \mathcal{B}) \,, \end{aligned}$$

proving (2.32).

For part (ii), let Q be a probability measure on $\big(\Omega \times \Omega, \mathcal{A} \times \mathcal{B}\big)$ such that both $P_{\text{id}}^{\mathcal{A},\mathcal{B}} \ll Q$ and $P_{\text{ind}}^{\mathcal{A},\mathcal{B}} \ll Q$; an example of such a measure is $Q = \big(P_{\text{id}}^{\mathcal{A},\mathcal{B}} + P_{\text{ind}}^{\mathcal{A},\mathcal{B}}\big)/2$. Setting

$$f_{\text{id}} = \frac{dP_{\text{id}}^{\mathcal{A},\mathcal{B}}}{dQ}, \; f_{\text{ind}} = \frac{dP_{\text{ind}}^{\mathcal{A},\mathcal{B}}}{dQ} \,,$$

we have

$$\begin{aligned} \left\| P_{\text{id}}^{\mathcal{A},\mathcal{B}} - P_{\text{ind}}^{\mathcal{A},\mathcal{B}} \right\| &= \frac{1}{2} \int_{\Omega \times \Omega} |f_{\text{id}}(\omega_1, \omega_2) - f_{\text{ind}}(\omega_1, \omega_2)| \, Q(d(\omega_1, \omega_2)) \\ &= \int_{\Omega \times \Omega} \big(f_{\text{id}}(\omega_1, \omega_2) - f_{\text{ind}}(\omega_1, \omega_2)\big)_+ \, Q(d(\omega_1, \omega_2)) \\ &= \int_C \big(f_{\text{id}}(\omega_1, \omega_2) - f_{\text{ind}}(\omega_1, \omega_2)\big) \, Q(d(\omega_1, \omega_2)) \,, \end{aligned}$$

where

$$C = \{(\omega_1, \omega_2) : f_{\text{id}}(\omega_1, \omega_2) > f_{\text{ind}}(\omega_1, \omega_2)\} \,. \tag{2.34}$$

In particular, for all relevant finite partitions $\mathcal{I} = \{A_1, \ldots, A_I\}$ and $\mathcal{J} = \{B_1, \ldots, B_J\}$ of Ω, we have

$$\begin{aligned}
\beta(\mathcal{A}, \mathcal{B}) &= \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \int_{A_i \times B_j} |f_{\text{id}}(\omega_1, \omega_2) - f_{\text{ind}}(\omega_1, \omega_2)| \, Q(d(\omega_1, \omega_2)) \\
&\geq \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \left| \int_{A_i \times B_j} (f_{\text{id}}(\omega_1, \omega_2) - f_{\text{ind}}(\omega_1, \omega_2)) \, Q(d(\omega_1, \omega_2)) \right| \\
&= \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \left| P_{\text{id}}^{\mathcal{A},\mathcal{B}}(A_i \times B_j) - P_{\text{ind}}^{\mathcal{A},\mathcal{B}}(A_i \times B_j) \right| \\
&= \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \left| (A_j \cap B_j) - P(A_j) P(B_j) \right| .
\end{aligned}$$

Taking the supremum over all relevant partitions yields

$$\beta(\mathcal{A}, \mathcal{B}) \geq \frac{1}{2} \sup_{\mathcal{I}, \mathcal{J}} \sum_{i=1}^{I} \sum_{j=1}^{J} \left| P(A_i \cap B_j) - P(A_i) P(B_j) \right| .$$

On the other hand, using the fact that the disjoint unions of measurable rectangles of the form $A \times B$ with $A \in \mathcal{A}$, $B \in \mathcal{B}$, form a field generating the product σ-field $\mathcal{A} \times \mathcal{B}$, for every $\varepsilon > 0$ we can select such a disjoint union C_0 satisfying $Q(C \triangle C_0) \leq \varepsilon$, where C is the set in (2.34); see Corollary 1, p. 169, in Billingsley (1995). Given $\delta > 0$, we can select $\varepsilon > 0$ so small that this property of C_0 will imply that

$$\begin{aligned}
&\left| \int_{C} \big(f_{\text{id}}(\omega_1, \omega_2) - f_{\text{ind}}(\omega_1, \omega_2)\big) \, Q(d(\omega_1, \omega_2)) \right. \\
&\left. - \int_{C_0} \big(f_{\text{id}}(\omega_1, \omega_2) - f_{\text{ind}}(\omega_1, \omega_2)\big) \, Q(d(\omega_1, \omega_2)) \right| \leq \delta;
\end{aligned}$$

see, e.g., Problem 16.8 in Billingsley (1995). We conclude that for every $\delta > 0$, there are relevant finite partitions $\mathcal{I} = \{A_1, \ldots, A_I\}$ and $\mathcal{J} = \{B_1, \ldots, B_J\}$ of Ω such that

$$\begin{aligned}
\beta(\mathcal{A}, \mathcal{B}) &\leq \int_{C_0} \big(f_{\text{id}}(\omega_1, \omega_2) - f_{\text{ind}}(\omega_1, \omega_2)\big) \, Q(d(\omega_1, \omega_2)) + \delta \\
&\leq \sum_{i=1}^{I} \sum_{j=1}^{J} \big(P(A_i \cap B_j) - P(A_i) P(B_j)\big)_+ + \delta .
\end{aligned}$$

Since for all partitions we have

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\big(P(A_i\cap B_j)-P(A_i)P(B_j)\big)_+=\sum_{i=1}^{I}\sum_{j=1}^{J}\big(P(A_i\cap B_j)-P(A_i)P(B_j)\big)_-\,,$$

we conclude, after switching to the supremum over the relevant partitions, that for every $\delta>0$,

$$\beta(\mathcal{A},\mathcal{B})\le\frac{1}{2}\sup_{\mathcal{I},\mathcal{J}}\sum_{i=1}^{I}\sum_{j=1}^{J}\big|P(A_i\cap B_j)-P(A_i)P(B_j)\big|+\delta\,.$$

Letting $\delta\to 0$ proves (2.33).

Part (iii) of the lemma is an immediate conclusion from (2.33). □

Let, once again, $\big(X_n,\,n\in\mathbb{Z}\big)$ be a stationary process on a probability space $\big(\Omega,\mathcal{F},P\big)$. Using the notation in (2.29) for the σ-fields generated by the process, we define the *beta mixing coefficient* of the process $\big(X_n,\,n\in\mathbb{Z}\big)$ by

$$\beta_X(n):=\beta\big(\mathcal{F}_{-\infty}^{0},\mathcal{F}_n^{\infty}\big),\;n=1,2,\ldots\,. \tag{2.35}$$

Definition 2.3.6. A stationary process $\big(X_n,\,n\in\mathbb{Z}\big)$ is called absolutely regular, or β-mixing, if $\beta_X(n)\to 0$ as $n\to\infty$.

An immediate conclusion of part (i) of Lemma 2.3.5 is the following comparison of the strong mixing property and the absolute regularity property.

Proposition 2.3.7. *An absolutely regular stationary process is also strongly mixing.*

Let, once again, $\big(\Omega,\mathcal{F},P\big)$ be a probability space, and $\mathcal{A}$ and $\mathcal{B}$ two sub-σ-fields of $\mathcal{F}$. We introduce a third notion of dependence between two σ-fields by

$$\phi(\mathcal{A},\mathcal{B})=\sup_{A\in\mathcal{A}:P(A)>0,B\in\mathcal{B}}\big|P(B|A)-P(B)\big|\,. \tag{2.36}$$

An immediate observation is that unlike the measures of dependence introduced in (2.27) and (2.31), this new notion of dependence does not appear to be symmetric, in the sense that there seems to be no reason why $\phi(\mathcal{A},\mathcal{B})$ and $\phi(\mathcal{B},\mathcal{A})$ should coincide. In fact, the two quantities are, in general, different. See, for instance, Problem 2.6.10.

The next lemma summarizes the basic properties of the measure of dependence $\phi(\mathcal{A},\mathcal{B})$.

Lemma 2.3.8. *Let $\mathcal{A}$ and $\mathcal{B}$ be sub-σ-fields of $\mathcal{F}$.*

(i) We have

$$0\le\phi(\mathcal{A},\mathcal{B})\le 1$$

and

$$\phi(\mathcal{A},\mathcal{B}) \geq \beta(\mathcal{A},\mathcal{B}). \tag{2.37}$$

(ii) $\phi(\mathcal{A},\mathcal{B}) = 0$ *if and only if the* σ*-fields* $\mathcal{A}$ *and* $\mathcal{B}$ *are independent.*

Proof. The first statement of part (i) is obvious. For the second statement of part (i) we use the representation of $\beta(\mathcal{A},\mathcal{B})$ given in (2.33). Let $\{A_1,\dots,A_I\}$ and $\{B_1,\dots,B_J\}$ be partitions of Ω into $\mathcal{A}$-measurable sets and $\mathcal{B}$-measurable sets, respectively. Then

$$\begin{aligned}
&\sum_{i=1}^{I}\sum_{j=1}^{J}\left|P(A_i\cap B_j) - P(A_i)P(B_j)\right| \\
&= \sum_{i=1}^{I} P(A_i)\sum_{j=1}^{J}\left|P(B_j|A_i) - P(B_j)\right| \\
&= 2\sum_{i=1}^{I} P(A_i)\left|P(B_{+,i}|A_i) - P(B_{+,i})\right|,
\end{aligned}$$

where $B_{+,i}$ is the union of all B_j, $j = 1,\dots,J$, such that $P(B_j|A_i) \geq P(B_j)$. Therefore,

$$\begin{aligned}
&\sum_{i=1}^{I}\sum_{j=1}^{J}\left|P(A_i\cap B_j) - P(A_i)P(B_j)\right| \\
&\leq 2\phi(\mathcal{A},\mathcal{B})\sum_{i=1}^{I} P(A_i) = 2\phi(\mathcal{A},\mathcal{B})
\end{aligned}$$

for all partitions, and the second statement of part (i) follows from (2.33). The claim of part (ii) is obvious.

For a stationary process $\left(X_n,\ n\in\mathbb{Z}\right)$ defined on a probability space $\left(\Omega,\mathcal{F},P\right)$, the *phi mixing coefficient* is defined by

$$\phi_X(n) := \phi\left(\mathcal{F}_{-\infty}^{0},\mathcal{F}_{n}^{\infty}\right),\ n = 1,2,\dots, \tag{2.38}$$

where the σ-fields are defined in (2.29). The phi mixing coefficient leads to one more notion of strong mixing for stationary processes.

Definition 2.3.9. A stationary process $\left(X_n,\ n\in\mathbb{Z}\right)$ is called ϕ-mixing if $\phi_X(n)\to 0$ as $n\to\infty$.

By Lemma 2.3.8, we immediately obtain the following statement.

Proposition 2.3.10. *A* ϕ*-mixing stationary process is absolutely regular (and hence also strongly mixing).*

The lack of symmetry of the notion of dependence between σ-fields introduced in (2.36) makes it possible to introduce the "time-reversed" versions of the ϕ mixing coefficient and of the ϕ-mixing property. For a stationary process $\big(X_n,\ n \in \mathbb{Z}\big)$, one simply defines

$$\phi_X^{\mathrm{rev}}(n) := \phi\big(\mathcal{F}_n^{\infty}, \mathcal{F}_{-\infty}^{0}\big),\ n = 1, 2, \dots, \tag{2.39}$$

leading to the following definition.

Definition 2.3.11. A stationary process $\big(X_n,\ n \in \mathbb{Z}\big)$ is called reverse-ϕ-mixing if $\phi_X^{\mathrm{rev}}(n) \to 0$ as $n \to \infty$.

Clearly, an obvious version of Proposition 2.3.10 for reverse-ϕ-mixing holds. We mention that there exist stationary processes that are ϕ-mixing but not reverse-ϕ-mixing (and conversely); see Rosenblatt (1971).

2.4 Conservative and Dissipative Maps

In this section, we return to maps on general σ-finite measure spaces and discuss certain properties of recurrence of such maps. Their probabilistic significance will become apparent in Section 3.6.

Let $\big(E, \mathcal{E}, m\big)$ be a σ-finite measure space. In this section we will deal with a nonsingular map $\phi : E \to E$ that preserves the measure m (a measure-preserving map). A set $W \in \mathcal{E}$ is called *wandering* if the sets $\big(\phi^{-n}(W),\ n = 1, 2, \dots\big)$ are mutually disjoint modulo m (recall that two measurable sets A and B are disjoint modulo m if $m(A \cap B) = 0$). This means that for every point x in a wandering set W, apart from a possible subset of W of measure zero, the trajectory $\phi^n(x)$, $n = 1, 2, \dots$, never reenters the set A.

Every set of measure zero is trivially a wandering set. What sets different maps apart is the existence of nontrivial wandering sets.

Definition 2.4.1. A measure-preserving map ϕ on $\big(E, \mathcal{E}, m\big)$ is conservative if it does not admit a wandering set of positive measure.

Observe that every measure-preserving map on a finite measure space is automatically conservative, since a finite measure space cannot contain infinitely many sets of equal and positive measure that are mutually disjoint modulo m.

Example 2.4.2. Let $E = \mathbb{Z}$ and let m be the counting measure. The right shift $\phi(x) = x + 1$ for $x \in \mathbb{Z}$ is measure-preserving. Since the set $W = \{0\}$ is obviously a wandering set of positive measure, the right shift is not conservative.

The following result shows that in general, a map ϕ has both a "conservative part" and a part that is "purely nonconservative."

Theorem 2.4.3. The Hopf decomposition *Let ϕ be a measure-preserving map on a σ-finite measure space $\big(E,\mathcal{E},m\big)$. Then there is a partition of E into ϕ-invariant sets $\mathcal{C}(\phi)$ and $\mathcal{D}(\phi)$ such that*

(i) there is no wandering set of a positive measure that is a subset of $\mathcal{C}(\phi)$;
(ii) there is a wandering set W such that $\mathcal{D}(\phi) = \cup_{n=-\infty}^{\infty}\phi^n(W)$ modulo m.

The partition $E = \mathcal{C}(\phi) \cup \mathcal{D}(\phi)$ is unique in the sense that if $E = \mathcal{C}_1 \cup \mathcal{D}_1$, where $\mathcal{C}_1$ and $\mathcal{D}_1$ are ϕ-invariant sets satisfying (i) and (ii), then $m\big(\mathcal{C}(\phi)\triangle\,\mathcal{C}_1\big) = m\big(\mathcal{D}(\phi)\triangle\,\mathcal{D}_1\big) = 0$.

Proof. We begin with a probability measure P on $\big(E,\mathcal{E}\big)$ equivalent to the σ-finite measure m. We will construct recursively an increasing sequence of ϕ-invariant sets $(I_n,\ n = 0,1,2,\ldots)$ and a sequence of wandering sets $(W_n,\ n = 1,2,\ldots)$ as follows. Let $I_0 = \emptyset$. For $n = 0,1,2,\ldots$, let

$$\alpha_n = \sup\{P(W),\ W \subset I_n^c,\ W \text{ wandering}\}\,.$$

Let $W_{n+1} \subset I_n^c$ be such that $P(W_{n+1}) \geq \alpha_n/2$ (if $\alpha_n = 0$, we can always choose $W_{n+1} = \emptyset$), and let

$$I_{n+1} = I_n \cup \bigcup_{k=-\infty}^{\infty} \phi^k(W_{n+1})\,.$$

Since P is a probability measure, we see that $\alpha_n \to 0$ as $n \to \infty$. Set $W = \cup_{n=1}^{\infty}W_n$. It is clear that W is a wandering set.

Let $\mathcal{D}(\phi) = \cup_{k=-\infty}^{\infty}\phi^k(W)$. Then $\mathcal{D}(\phi)$ is a ϕ-invariant set, and so is its complement, which we denote by $\mathcal{C}(\phi)$. By construction, $\mathcal{C}(\phi) \subset I_n^c$ for each $n = 1,2,\ldots$, and hence so is every wandering subset V of $\mathcal{C}(\phi)$. This implies that $P(V) \leq \alpha_n$ for each n, and since $\alpha_n \to 0$, we conclude that $P(V) = 0$. Since P is equivalent to m, we also have $m(V) = 0$.

To prove uniqueness, suppose, for example, that $m\big(\mathcal{C}_1 \setminus \mathcal{C}(\phi)\big) > 0$. Then for some n, $m\big(\mathcal{C}_1 \cap \phi^n(W)\big) > 0$. Since a subset of a wandering set is a wandering set, this contradicts the fact that $\mathcal{C}_1$ possesses property (i). □

The Hopf decomposition $E = \mathcal{C}(\phi) \cup \mathcal{D}(\phi)$ is called the decomposition of E into the conservative part of E and the dissipative part of E with respect to ϕ. It is usual to say simply that $\mathcal{C}(\phi)$ and $\mathcal{D}(\phi)$ are the conservative part of ϕ and the dissipative part of ϕ respectively. We see immediately that ϕ is conservative if and only if its dissipative part vanishes (modulo m). The picture is completed by the following definition.

Definition 2.4.4. A measure-preserving map ϕ on $\big(E,\mathcal{E},m\big)$ is dissipative if $\mathcal{C}(\phi) = \emptyset$ modulo m.

Since an ergodic map ϕ cannot have two disjoint invariant sets of positive measure, each ergodic map is either conservative or dissipative. For the right shift on the integers in Example 2.4.2, it is immediate that $\mathcal{D}(\phi) = \mathbb{Z}$, and so the right shift is dissipative.

A useful and intuitive criterion for distinguishing between a conservative map and a dissipative map, or, more generally, between the conservative and dissipative parts of a map, is through the difference in the behavior of certain sums.

Theorem 2.4.5. *Let ϕ be a measure-preserving map on a σ-finite measure space $\big(E, \mathcal{E}, m\big)$. For every function $f \in L_1(m)$ such that $f > 0$ m-a.e., we have*

$$\left\{ x \in E : \sum_{n=1}^{\infty} f\big(\phi^n(x)\big) = \infty \right\} = \mathcal{C}(\phi) \ \text{ modulo } m. \tag{2.40}$$

Proof. We first show that the set on the left-hand side of (2.40) is, modulo m, a subset of $\mathcal{C}(\phi)$. By Theorem 2.4.3, it is enough to prove that for every wandering set W,

$$m\left(x \in W : \sum_{n=1}^{\infty} f\big(\phi^n(x)\big) = \infty \right) = 0\,. \tag{2.41}$$

However, for $k = 1, 2, \ldots$, since ϕ is measure-preserving and W is wandering,

$$\int_W \sum_{n=1}^{k} f \circ \phi^n \, dm = \sum_{n=1}^{k} \int_E \mathbf{1}_W f \circ \phi^{k-n+1} \, dm$$

$$= \sum_{n=1}^{k} \int_E \mathbf{1}_W \circ \phi^n f \circ \phi^{k+1} \, dm = \int_E \left(\sum_{n=1}^{k} \mathbf{1}_W \circ \phi^n \right) f \circ \phi^{k+1} \, dm$$

$$\leq \int_E f \circ \phi^{k+1} \, dm = \int_E f \, dm\,.$$

By the monotone convergence theorem,

$$\int_W \sum_{n=1}^{\infty} f \circ \phi^n \, dm \leq \int_E f \, dm < \infty\,,$$

which clearly implies (2.41).

In order to prove the second inclusion in (2.40), suppose that to the contrary,

$$m\left(x \in \mathcal{C}(\phi) : \sum_{n=1}^{\infty} f\big(\phi^n(x)\big) < \infty \right) > 0\,.$$

Let $\varepsilon > 0$ be such that

$$m\left(x \in \mathcal{C}(\phi) : \phi^j(x) > \varepsilon \text{ for some } j \in \mathbb{Z},\ \sum_{n=1}^{\infty} f\big(\phi^n(x)\big) < \infty \right) > 0\,. \tag{2.42}$$

Let $\mathcal{C}_\varepsilon$ be the set of positive measure in (2.42). Define W to be the set of those points x in $\mathcal{C}_\varepsilon$ such that

$$\sup\{j \in \mathbb{Z} : f(\phi^j(x)) > \varepsilon\} = 0 .$$

Then W is a set of positive measure, and it is clearly a wandering set. It is also a subset of $\mathcal{C}(\phi)$, which is impossible by the definition of the conservative part of ϕ. This contradiction proves the second inclusion in (2.40). □

We finish this section with a refinement of the Hopf decomposition that will prove very useful to us in the sequel in a discussion of the memory of stationary non-Gaussian infinitely divisible processes.

Definition 2.4.6. A measure-preserving map ϕ on $(E, \mathcal{E}, m)$ is positive if there is a finite measure $\tilde{m}$ on $(E, \mathcal{E})$ that is preserved by ϕ and is equivalent to the measure m.

The following statement is immediate.

Proposition 2.4.7. *A positive map ϕ does not admit a wandering set of positive measure and hence is conservative.*

Proof. Indeed, if W is a wandering set with $m(W) > 0$, then $\tilde{m}(W) > 0$ as well. Since the sets $(\phi^{-n}(W),\ n = 1, 2, \ldots)$ are disjoint modulo m, they are also disjoint modulo $\tilde{m}$, so that

$$\tilde{m}\left(\bigcup_{n=1}^{\infty} \phi^{-n}(W)\right) = \sum_{n=1}^{\infty} \tilde{m}(\phi^{-n}(W)) = \infty,$$

since ϕ preserves $\tilde{m}$. This contradicts the fact that $\tilde{m}$ is a finite measure.

The fact that a positive map ϕ is also conservative follows now from Theorem 2.4.3, since if ϕ had a nontrivial dissipative part, it would admit a wandering set of positive measure. □

As in the case of the Hopf decomposition, in general a map ϕ has both a "positive part" and a part that is "purely nonpositive."

Theorem 2.4.8. The positive–null decomposition

Let ϕ be a measure-preserving map on a σ-finite measure space $(E, \mathcal{E}, m)$. Then there is a partition of E into ϕ-invariant sets $\mathcal{P}(\phi)$ and $\mathcal{N}(\phi)$ such that

(i) ϕ is positive on $\mathcal{P}(\phi)$;
(ii) no ϕ-invariant measurable set $A \subseteq \mathcal{N}(\phi)$ satisfies $0 < m(A) < \infty$.

The partition $E = \mathcal{P}(\phi) \cup \mathcal{N}(\phi)$ is unique in the sense that if $E = \mathcal{P}_1 \cup \mathcal{N}_1$, where $\mathcal{P}_1$ and $\mathcal{N}_1$ are ϕ-invariant sets satisfying (i) and (ii), then $m(\mathcal{P}(\phi) \triangle \mathcal{P}_1) = m(\mathcal{N}(\phi) \triangle \mathcal{N}_1) = 0$.

Proof. We proceed by "exhaustion," as in the proof of Theorem 2.4.3. We again begin with a probability measure P on $\big(E, \mathcal{E}\big)$ equivalent to the σ-finite measure m and construct recursively an increasing sequence of ϕ-invariant sets $(I_n,\, n = 0, 1, 2, \ldots)$ and a sequence of sets $(F_n,\, n = 1, 2, \ldots)$ as follows. Let $I_0 = \emptyset$. For $n = 0, 1, 2, \ldots$, let

$$\alpha_n = \sup\{P(F),\, F \subset I_n^c,\, F\ \phi\text{-invariant and } m(F) < \infty\}\,.$$

Let $F_{n+1} \subset I_n^c$ be such that $P(F_{n+1}) \geq \alpha_n/2$ (use $F_{n+1} = \emptyset$ if $\alpha_n = 0$), and let $I_{n+1} = I_n \cup F_{n+1}$. Finally, let $\mathcal{P}(\phi) = \cup_{n=1}^{\infty} F_n$. By construction, $\mathcal{P}(\phi)$ is an invariant set, and we denote its complement by $\mathcal{N}(\phi)$. We now check that these sets satisfy (i) and (ii) of the theorem. If $m(\mathcal{P}(\theta)) = 0$, then ϕ is trivially positive on $\mathcal{P}(\phi)$, since we can take $\tilde{m}$ to be the null measure. If $m(\mathcal{P}(\theta)) > 0$, then we can construct a nontrivial ϕ-invariant finite measure on $\mathcal{P}(\phi)$, equivalent to m on that set, by setting

$$\frac{d\tilde{m}}{dm} = 2^{-n} \text{ on } F_n \text{ if } m(F_n) > 0,\ n = 1, 2, \ldots\,.$$

Therefore, ϕ is positive on $\mathcal{P}(\phi)$. The property (ii) holds by the construction of the set $\mathcal{P}(\phi)$.

It remains to prove the uniqueness of a decomposition. Suppose, once again, that $m\big(\mathcal{C}_1 \setminus \mathcal{C}(\phi)\big) > 0$. Then there is a probability measure m_1 supported by $\mathcal{C}_1 \setminus \mathcal{C}(\phi)$, equivalent to m on that set and invariant under ϕ. The function $g = dm_1/dm$ is ϕ-invariant, finite, and positive on $\mathcal{C}_1 \setminus \mathcal{C}(\phi)$ modulo m, and integrates to 1 with respect to m. Therefore, there are $0 < \varepsilon_1 \leq \varepsilon_2 < \infty$ such that

$$\int_{\mathcal{C}_1 \setminus \mathcal{C}(\phi)} g(x)\mathbf{1}\big(\varepsilon_1 \leq g(x) \leq \varepsilon_2\big)\, m(dx) > 0\,.$$

The set

$$\big\{x \in \mathcal{C}_1 \setminus \mathcal{C}(\phi) :\ \varepsilon_1 \leq g(x) \leq \varepsilon_2\big\}$$

is a ϕ-invariant set of finite positive measure m. Since this set is also a subset of $\mathcal{N}(\phi)$, we obtain a contradiction with the fact that $\mathcal{N}(\phi)$ satisfies (ii). □

We call a positive–null decomposition $E = \mathcal{P}(\phi) \cup \mathcal{N}(\phi)$ the decomposition of E into the positive part of E and the null part of E with respect to ϕ. With terminology similar to that in the Hopf decomposition, we say that $\mathcal{P}(\phi)$ and $\mathcal{N}(\phi)$ are the positive part of ϕ and the null part of ϕ respectively. Then ϕ is positive if and only if its null part vanishes (modulo m). We observe that it follows from Proposition 2.4.7 that

$$\mathcal{P}(\phi) \subseteq \mathcal{C}(\phi),\ \mathcal{D}(\phi) \subseteq \mathcal{N}(\phi)\,, \tag{2.43}$$

and the Hopf decomposition can be combined with the positive–null decomposition into a three-way decomposition of E: let $\mathcal{CN}(\phi) = \mathcal{C}(\phi) \cap \mathcal{N}(\phi)$. Then we decompose

$$E = \mathcal{P}(\phi) \cup \mathcal{CN}(\phi) \cup \mathcal{D}(\phi) \tag{2.44}$$

uniquely into ϕ-invariant sets that inherit their properties from the Hopf decomposition and the positive–null decomposition. Finally, we also introduce the following definition.

Definition 2.4.9. A measure-preserving map ϕ on $\big(E, \mathcal{E}, m\big)$ is null if $\mathcal{P}(\phi) = \emptyset$ modulo m.

Our next goal is to understand better the structure of the positive–null decomposition. To this end, we introduce a new notion. A set $W \in \mathcal{E}$ is called *weakly wandering* if there is a sequence $n_k \to \infty$ such that the sets $\big(\phi^{-n_k}(W),\ k = 0, 1, 2, \ldots\big)$, are disjoint modulo m (with $n_0 = 0$). Clearly, every wandering set is weakly wandering as well. Furthermore, Proposition 2.4.7 immediately extends, with the same proof, to the following.

Proposition 2.4.10. *A positive map ϕ does not admit a weakly wandering set of positive measure.*

The following theorem, which we prove only partially, completely clarifies the connection between weakly wandering sets and the positive–null decomposition. It also presents a "weighted version" of the criterion for distinguishing the parts in the Hopf decomposition, presented in Theorem 2.40. The weighted version allows one to distinguish the parts in the positive–null decomposition. Denote by $\mathcal{W}$ the set of sequences $(w_n,\ n = 1, 2, \ldots)$ with the following properties:

$$w_n > 0,\ w_{n+1} \le w_n,\ n = 1, 2, \ldots,\ \sum_{n=1}^{\infty} w_n = \infty\,. \tag{2.45}$$

Theorem 2.4.11. *Let ϕ be a measure-preserving map on a σ-finite measure space $\big(E, \mathcal{E}, m\big)$.*

(i) There exist a set W and a sequence $n_k \to \infty$ such that the sets $\big(\phi^{-n_k}(W),\ k = 0, 1, 2, \ldots\big)$ are disjoint modulo m and $\mathcal{N}(\phi) = \cup_{k=1}^{\infty} \phi^{-n_k}(W)$ modulo m (the set W is automatically weakly wandering).
(ii) For every sequence $(w_n,\ n = 1, 2, \ldots)$ in $\mathcal{W}$ and every function $f \in L_1(m)$ such that $f > 0$ m-a.e,

$$\sum_{n=1}^{\infty} w_n f\big(\phi^n(x)\big) = \infty \text{ } m\text{-a.e. on } \mathcal{P}(\phi). \tag{2.46}$$

Further, there is a sequence $(w_n,\ n = 1, 2, \ldots)$ *in* $\mathcal{W}$ *such that for every function* $f \in L_1(m)$ *such that* $f > 0$ *m-a.e., we have*

$$\left\{ x \in E : \sum_{n=1}^{\infty} w_n f\big(\phi^n(x)\big) < \infty \right\} = \mathcal{N}(\phi) \ \textit{modulo } m. \tag{2.47}$$

Proof. Part (i) of the theorem is in Jones and Krengel (1974). For the first statement of part (ii), fix a sequence in $\mathcal{W}$. We may assume without loss of generality that $w_1 = 1$. Let $p_n = w_n - w_{n+1}$, $n = 1, 2, \ldots$, and $p_\infty = \lim_{n\to\infty} w_n$, so that (p_n) are probabilities on $\mathbb{N} \cup \{+\infty\}$. By the definition of the set $\mathcal{W}$, it follows that this probability distribution has an infinite mean. Then

$$\sum_{n=1}^{\infty} w_n f\big(\phi^n(x)\big) = p_\infty \sum_{n=1}^{\infty} f\big(\phi^n(x)\big) + \sum_{j=1}^{\infty} p_j \sum_{n=1}^{j} f\big(\phi^n(x)\big). \tag{2.48}$$

If $m\big(\mathcal{P}(\phi)\big) > 0$, then there exists a probability measure $\tilde{m}$ supported by $\mathcal{P}(\phi)$ and equivalent to m on that set. By the pointwise ergodic theorem,

$$\frac{1}{n} \sum_{j=0}^{n-1} f\big(\phi^j(x)\big) \to g_f(x) \ \text{ as } n \to \infty$$

for P-almost every $x \in \mathcal{P}(\phi)$, hence for m-almost every $x \in \mathcal{P}(\phi)$. Furthermore, the limit is the conditional expectation of a positive random variable, hence is itself a.s. positive. Now (2.46) follows from (2.48) and the fact that the probabilities (p_n) have an infinite mean.

The second statement of part (ii) is in Theorem 3 in Krengel (1967). □

The following example should clarify the distinction between a wandering set and a weakly wandering set.

Example 2.4.12. Let $0 < p < 1$. The transition rule $p_{i,i+1} = 1 - p_{i,i-1} = p$ for $i \in \mathbb{Z}$ defines a Markov chain on $\mathbb{Z}$, the simple random walk that takes a step to the right with probability p. This Markov chain is transient if $p \neq 1/2$, and it is null recurrent if $p = 1/2$. In any case, it has an infinite invariant measure, which is the counting measure on $\mathbb{Z}$. We use the law of this random walk to construct a σ-finite measure m on $E = \mathbb{Z}^{\mathbb{Z}}$ by

$$m(A) = \sum_{i \in \mathbb{Z}} P_i\big(\text{a realization of the random walk is in } A\big). \tag{2.49}$$

The probability on the right-hand side of (2.49) is computed according to the law of the Markov chain that visits state i at time zero. In the special case of a simple random walk, this law can be constructed very simply: let $(W_n^+,\ n = 0, 1, 2, \ldots)$ and $(W_n^-,\ n = 0, 1, 2, \ldots)$ be two independent simple random walks starting at zero, the first one stepping to the right with probability p, and the second one stepping to the right with probability $1-p$. Then the stochastic process

$$\left(\ldots, i+W_2^-, i+W_1^-, i, i+W_1^+, i+W_2^+, \ldots\right)$$

has the law P_i on E. Since the counting measure is an invariant measure for the random walk, the measure m defined by (2.49) is invariant under the left shift on E.

Suppose first that $p \neq 1/2$, and consider the set

$$A = \left\{\mathbf{x} \in \mathbb{Z}^{\mathbb{Z}} : x_0 = 0,\ x_n \neq 0 \text{ for } n > 0\right\}. \tag{2.50}$$

By construction, the set A in (2.50) is a wandering set, irrespective of the value of p. We claim that for $p \neq 1/2$, we have

$$\bigcup_{n=-\infty}^{\infty} \phi^n(A) = E \ \text{ modulo } m. \tag{2.51}$$

According to Theorem 2.4.3, this would show that the left shift ϕ is dissipative in this case. To see that (2.51) holds, note that

$$\left(\bigcup_{n=-\infty}^{\infty} \phi^n(A)\right)^c = \left\{\mathbf{x} :\in \mathbb{Z} : x_n \neq 0 \text{ for all } n \in \mathbb{Z}\right\}$$
$$\cup \left\{\mathbf{x} :\in \mathbb{Z} : x_n = 0 \text{ for arbitrary large } n \in \mathbb{Z}\right\}.$$

The first set above has zero measure with respect to m, because a simple random walk converges a.s. to the two different infinities as the times goes to $\pm\infty$, and does so without skipping steps. Therefore, regardless of the initial state, it will a.s. visit state zero. The second set above has zero measure with respect to m, because a simple random walk with $p \neq 1/2$ is transient, and so it visits any given state only finitely many times. This proves (2.51).

This argument for the relation (2.51) fails in the symmetric case $p = 1/2$, since in that case, the random walk is recurrent. This actually implies that the left shift ϕ is conservative with respect to m, and hence does not admit any wandering set of positive measure. In order to see that ϕ is now conservative, consider a positive measurable function on $\mathbb{Z}$ given by $f(\mathbf{x}) = 2^{-|x_0|}$. It is clear that $f \in L_1(m)$. However,

$$\sum_{n=1}^{\infty} f\left(\phi^n(\mathbf{x})\right) \geq \sum_{n=1}^{\infty} \mathbf{1}(x_n = 0) = \infty$$

m-a.e., since the simple symmetric random walk is recurrent and hence visits state zero infinitely many times. By Theorem 2.4.5, $\mathcal{C}(\theta) = E$ modulo m, and ϕ is conservative.

It turns out, however, that in the case $p = 1/2$, the left shift admits quite "large" weakly wandering sets, and we will construct a family of such sets. Let $n_1, n_2, \ldots$ and $b_0, b_1, b_2, \ldots$ be two strictly increasing sequences of positive integers. Let

$$B = \left\{ \mathbf{x} \in \mathbb{Z}^{\mathbb{Z}} : x_0 \in (-b_0, b_0),\ x_{-n_k} \in (-b_k, -b_{k-1}] \cup [b_{k-1}, b_k),\ k \geq 1 \right\}. \tag{2.52}$$

Observe that for each $k \geq 1$ and $\mathbf{x} \in \phi^{-n_k}(B)$, we have $x_0 \in (-b_k, -b_{k-1}] \cup [b_{k-1}, b_k)$, so the sets $\big(\phi^{-n_k}(B),\ k = 1, 2, \ldots\big)$ are disjoint, and hence the set B is weakly wandering. We now show that one can choose the sequences $n_1, n_2, \ldots$ and $b_0, b_1, b_2, \ldots$ in such a way as to make this set "large." To this end, we will simply note that by the elementary central limit theorem, for a simple symmetric random walk $(W_n,\ n = \ldots, -1, 0, 1, 2, \ldots)$, the law of $n^{-1/2} W_n$ converges weakly, as $n \to \infty$, to the standard normal law, whence for every $x > 0$,

$$P_0(|W_n| \leq x) \to 0 \text{ as } n \to \infty. \tag{2.53}$$

Let $\varepsilon > 0$ be a small number. We will choose b_0 depending on ε momentarily, but let us keep b_0 fixed for now. By (2.53), we can select n_1 so large that

$$P_i(|W_{n_1}| < b_0) \leq \varepsilon/8 \text{ for every } |i| < b_0,$$

and then we choose $b_1 > b_0$ so large that

$$P_i(|W_{n_1}| \geq b_1) \leq \varepsilon/8 \text{ for every } |i| < b_0.$$

We proceed inductively. Once $n_1, \ldots, n_k$ and $b_0, b_1, \ldots, b_k$ have been chosen, we first use (2.53) to select $n_{k+1} > n_k$ so large that

$$P_i(|W_{n_{k+1}-n_k}| < b_k) \leq 2^{-(k+3)}\varepsilon \text{ for every } |i| < b_k,$$

and then we select $b_{k+1} > b_k$ so large that

$$P_i(|W_{n_{k+1}-n_k}| \geq b_{k+1}) \leq 2^{-(k+3)}\varepsilon \text{ for every } |i| < b_k.$$

In order to convince ourselves that we have obtained a "large" set B in (2.52), we construct a probability measure ν on $\mathbb{Z}^{\mathbb{Z}}$, equivalent to m, by selecting strictly positive probabilities (p_i) on $\mathbb{Z}$, and defining, analogously to (2.49),

$$\nu(A) = \sum_{i \in \mathbb{Z}} p_i P_i\big(\text{a realization of the random walk is in } A\big). \tag{2.54}$$

Select b_0 so large that

$$\sum_{|i| \geq b_0} p_i \leq \varepsilon/2,$$

and note that

$$\nu(B^c) \le \varepsilon/2 + \sum_{|i|<b_0} p_i P_i \left(\bigcup_{k=1}^{\infty} \Big\{ |W_{n_{k-1}}| < b_{k-1},\ W_{n_k} \notin (-b_k, -b_{k-1}] \cup [b_{k-1}, b_k) \Big\} \right)$$

(with $n_0 = 0$). We now use the Markov property to see that by construction, for every $k = 1, 2, \ldots$,

$$\begin{aligned} &P_i\Big(|W_{n_{k-1}}| < b_{k-1},\ W_{n_k} \notin (-b_k, -b_{k-1}] \cup [b_{k-1}, b_k) \Big) \\ &= \sum_{|j|<b_{k-1}} P_i\big(|W_{n_{k-1}}| = j \big) P_j\Big(W_{n_k - n_{k-1}} \notin (-b_k, -b_{k-1}] \cup [b_{k-1}, b_k) \Big) \\ &\le 2 \cdot 2^{-(k+2)} \varepsilon = 2^{-(k+1)} \varepsilon\,, \end{aligned}$$

so that $\nu(B^c) \le \varepsilon$. Selecting $\varepsilon > 0$ small, we can hence obtain a "large" weakly wandering set B.

Notice that we have actually proved that the map ϕ is a null map. Indeed, if we had $m\big(\mathcal{P}(\phi)\big) > 0$, then we would also have $\nu\big(\mathcal{P}(\phi)\big) > 0$ for the probability measure ν in (2.54). This would imply that for $\varepsilon > 0$ small enough, $\nu\big(\mathcal{P}(\phi) \cap B\big) > 0$, hence also $m\big(\mathcal{P}(\phi) \cap B\big) > 0$. Since a subset of weakly wandering set is itself weakly wandering, we obtain a contradiction with Proposition 2.4.10.

2.5 Comments on Chapter 2

Comments on Section 2.1

There are many books on ergodic theory. The notions of ergodicity and mixing are widely used in probability. Among the sources that have been used in probability are Cornfeld et al. (1982) and Walters (1982).

Comments on Section 2.3

A very useful survey on strong mixing conditions is in Bradley (2005).

Comments on Section 2.4

Much of the material on the positive–null decomposition presented in this section is in Section 3.6 of Aaronson (1997) and Section 3.4 of Krengel (1985), and many of the results are due to U. Krengel and his coworkers, e.g., Krengel (1967) and Jones and Krengel (1974).

The phenomenon exhibited in Example 2.4.12 goes much further than the case of a simple random walk. For example, it is shown in Harris and Robbins (1953) that every recurrent real-valued Markov chain with an invariant measure generates similarly a shift-invariant measure on $\mathbb{R}^{\mathbb{Z}}$ with respect to which the left shift is conservative.

2.6 Exercises to Chapter 2

Exercise 2.6.1. *Let $(E, \mathcal{E}, m)$ be a σ-finite measure space, and $\phi : E \to E$ a measurable map. Show that the collection $\mathcal{I}$ of all ϕ-invariant sets $\mathcal{E}$ is a σ-field.*

Exercise 2.6.2. *Let $\phi : E \to E$ be onto and one-to-one. Check that ϕ is nonsingular if and only if its inverse ϕ^{-1} is nonsingular.*

Exercise 2.6.3. *(i) Prove that measurable sets satisfying either $m(A) = 0$ or $m(A^c) = 0$ are invariant sets for every nonsingular map ϕ.*

(ii) Suppose that ϕ is nonsingular. Prove that the σ-field of ϕ-invariant sets coincides with the σ-field of ϕ^{-1}-invariant sets. Does this remain true if ϕ is one-to-one and onto, but not necessarily nonsingular? Conclude that a nonsingular map is ergodic if and only if the inverse map ϕ^{-1} is ergodic.

Exercise 2.6.4. *(i) Let ϕ be a measurable map, and $f : E \to \mathbb{R}$ a function measurable with respect to the σ-field $\mathcal{I}$ of ϕ-invariant sets. Show that f is ϕ-invariant in the sense that $f(\phi(x)) = f(x)$ for m-almost every x.*

(ii) Suppose that ϕ is a nonsingular and ergodic map, and f is as in part (i). Show that f is constant in the sense that there is $a \in \mathbb{R}$ such that $f(x) = a$ for m-almost every x.

Exercise 2.6.5. *Recall that the conditional expectation can be defined for nonnegative random variables without a finite mean; see Billingsley (1995). Prove that the ergodic theorem in the form of* (2.7) *holds for a nonnegative measurable function f even if the integrability assumption* (2.6) *fails.*

Exercise 2.6.6. *Show that a map ϕ is mixing if and only if the inverse map ϕ^{-1} is mixing.*

Exercise 2.6.7. *Show by example that the sum of two independent stationary ergodic processes may not be ergodic.* ***Hint****: What is the stationary process corresponding to the measure in* (2.3)*? Show that on the other hand, the sum of two independent stationary mixing processes must be mixing.*

Exercise 2.6.8. *Weak convergence is not kind to various ergodic-theoretical notions.*

(i) Give an example of a family of nonergodic stationary processes that converge weakly to a mixing process (no need to work hard: a trivial example suffices).

(ii) Give an example of a family of mixing processes that converge weakly to a nonergodic process (an autoregressive process of order 1 with Gaussian innovations can provide an easy example).

Exercise 2.6.9. *Show that the presence or absence of strong mixing is determined by the finite-dimensional distributions of a stationary process.*

Exercise 2.6.10. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* A, B *two events in* $\mathcal{F}$ *such that* $P(A \cap B) = .2, P(A \setminus B) = .3$, $P(B \setminus A) = .4$. *Let* $\mathcal{A}$ *be the* σ*-field generated by the event* A*, and* $\mathcal{B}$ *the* σ*-field generated by the event* B*. Calculate* $\phi(\mathcal{A}, \mathcal{B})$ *and* $\phi(\mathcal{B}, \mathcal{A})$ *and check that they are different.*

Exercise 2.6.11. *Let* $g : \mathbb{R} \to \mathbb{R}$ *be a one-to-one function such that both* g *and* g^{-1} *are measurable. Show that a stationary process* $(X_n, \, n \in \mathbb{Z})$ *is* α*- (*β*-,* ϕ*-) mixing if and only if the process* $(g(X_n), \, n \in \mathbb{Z})$ *is* α*- (*β*-,* ϕ*-) mixing.*

Exercise 2.6.12. *Theorem 2.4.5 requires the function to be strictly positive* m*-a.e. We can relax this assumption somewhat. Let* $B \subset \mathcal{C}(\phi)$ *be such that* $m(B) > 0$. *Prove that*

$$\sum_{n=1}^{\infty} \mathbf{1}_B\big(\phi^n(x)\big) = \infty \;\; m\text{-a.e. on } B.$$

Exercise 2.6.13. *Let* ϕ *be a null map on (an automatically infinite)* σ*-finite measure space* $(E, \mathcal{E}, m)$*, and* $f \in L_1(m)$*. Prove that*

$$\frac{1}{n} \sum_{j=0}^{n-1} f\big(\phi^j(x)\big) \to 0 \;\; \text{as } n \to \infty$$

for m*-almost every* $x \in E$.

Exercise 2.6.14. *A nonnegative sequence* $(g_n, \, n = 1, 2, \ldots)$ *is called subadditive if* $g_{n+m} \le g_n + g_m$ *for all* $n, m \ge 1$. *Prove that for every subadditive sequence, the limit* $\lim_{n\to\infty} g_n/n$ *exists and is equal to* $\inf_{n\ge 1} g_n/n$.

http://www.springer.com/978-3-319-45574-7

Stochastic Processes and Long Range Dependence
Samorodnitsky, G.
2016, XI, 415 p. 5 illus., Hardcover
ISBN: 978-3-319-45574-7