

# Chapter 16

## Stationary Sequences

**Abstract** Section 16.1 contains the definitions and a discussion of the concepts of strictly stationary sequences and measure preserving transformations. It also presents Poincaré’s theorem on the number of visits to a given set by a stationary sequence. Section 16.2 discusses invariance, ergodicity, mixing and weak dependence. The Birkhoff–Khintchin ergodic theorem is stated and proved in Sect. 16.3.

### 16.1 Basic Notions

Let  $(\Omega, \mathfrak{F}, \mathbf{P})$  be a probability space and  $\xi = (\xi_0, \xi_1, \dots)$  an infinite sequence of random variables given on it.

**Definition 16.1.1** A sequence  $\xi$  is said to be strictly *stationary* if, for any  $k$ , the distribution of the vector  $(\xi_n, \dots, \xi_{n+k})$  does not depend on  $n$ ,  $n \geq 0$ .

Along with the sequence  $\xi$ , consider the sequence  $(\xi_n, \xi_{n+1}, \dots)$ . Since the finite-dimensional distributions of these sequences (i.e. the distributions of the vectors  $(\xi_m, \dots, \xi_{m+k})$ ) coincide, the distributions of the sequences will also coincide (one has to make use of the measure extension theorem (see Appendix 1) or the Kolmogorov theorem (see Sect. 3.5). In other words, for a stationary sequence  $\xi$ , for any  $n$  and  $\mathbf{B} \in \mathfrak{B}^\infty$  (for notation see Sect. 3.5), one has

$$\mathbf{P}(\xi \in \mathbf{B}) = \mathbf{P}((\xi_n, \xi_{n+1}, \dots) \in \mathbf{B}).$$

The simplest example of a stationary sequence is given by a sequence of independent identically distributed random variables  $\zeta = (\zeta_0, \zeta_1, \dots)$ . It is evident that the sequence  $\xi_k = \alpha_0 \zeta_k + \dots + \alpha_s \zeta_{k+s}$ ,  $k = 0, 1, 2, \dots$ , will also be stationary, but the variables  $\xi_k$  will no longer be independent. The same holds for sequences of the form

$$\xi_k = \sum_{j=0}^{\infty} \alpha_j \zeta_{k+j},$$

provided that  $\mathbf{E}|\zeta_j| < \infty$ ,  $\sum |\alpha_j| < \infty$ , or if  $\mathbf{E}\zeta_k = 0$ ,  $\text{Var}(\zeta_k) < \infty$ ,  $\sum \alpha_j^2 < \infty$  (the latter ensures a.s. convergence of the series of random variables, see Sect. 10.2). In a

similar way one can consider stationary sequences  $\xi_k = g(\zeta_k, \zeta_{k+1}, \dots)$  “generated” by  $\zeta$ , where  $g(x)$  is an arbitrary measurable functional  $\mathbb{R}^\infty \mapsto \mathbb{R}$ .

Another example is given by stationary Markov chains. If  $\{X_n\}$  is a real-valued Markov chain with invariant measure  $\pi$  and transition probability  $P(\cdot, \cdot)$  then the chain  $\{X_n\}$  with  $X \in \pi$  will form a stationary sequence, because the distribution

$$\mathbf{P}(X_n \in B_0, \dots, X_{n+k} \in B_k) = \int_{B_0} \pi(dx_0) \int_{B_1} P(x_0, dx_1) \cdots \int_{B_k} P(x_{k-1}, dx_k)$$

will not depend on  $n$ .

Any stationary sequence  $\xi = (\xi_0, \xi_1, \dots)$  can always be extended to a stationary sequence  $\bar{\xi} = (\dots \xi_{-1}, \xi_0, \xi_1, \dots)$  given on the “whole axis”.

Indeed, for any  $n, -\infty < n < \infty$ , and  $k \geq 0$  define the joint distributions of  $(\xi_n, \dots, \xi_{n+k})$  as those of  $(\xi_0, \dots, \xi_k)$ . These distributions will clearly be consistent (see Sect. 3.5) and by the Kolmogorov theorem there will exist a unique probability distribution on  $\mathbb{R}_{-\infty}^\infty = \prod_{k=-\infty}^\infty \mathbb{R}_k$  with the respective  $\sigma$ -algebra such that any finite-dimensional distribution is a projection of that distribution on the corresponding subspace. It remains to take the random element  $\bar{\xi}$  to be the identity mapping of  $\mathbb{R}_{-\infty}^\infty$  onto itself.

In some of the subsequent sections it will be convenient for us to use stationary sequences given on the whole axis.

Let  $\xi$  be such a sequence. Define a transformation  $\theta$  of the space  $\mathbb{R}_{-\infty}^\infty$  onto itself with the help of the relations

$$(\theta \mathbf{x})_k = (\mathbf{x})_{k+1} = x_{k+1}, \tag{16.1.1}$$

where  $(\mathbf{x})_k$  is the  $k$ -th component of the vector  $\mathbf{x} \in \mathbb{R}_{-\infty}^\infty, -\infty < k < \infty$ . The transformation  $\theta$  clearly has the following properties:

1. It is a one-to-one mapping,  $\theta^{-1}$  is defined by

$$(\theta^{-1} \mathbf{x})_k = x_{k-1}.$$

2. The sequence  $\theta \bar{\xi}$  is also stationary, its distribution coinciding with that of  $\bar{\xi}$ :

$$\mathbf{P}(\theta \bar{\xi} \in \mathbf{B}) = \mathbf{P}(\bar{\xi} \in \mathbf{B}).$$

It is natural to call the last property of the transformation  $\theta$  the “measure preserving” property.

The above remarks explain to some extent why historically exploring the properties of stationary sequences followed the route of studying measure preserving transforms. Studies in that area constitute a substantial part of the modern analysis. In what follows, we will relate the construction of stationary sequences to measure preserving transformations, and it will be more convenient to regard the latter as “primary” objects.

**Definition 16.1.2** Let  $\langle \Omega, \mathfrak{F}, \mathbf{P} \rangle$  be the basic probability space. A transformation  $T$  of  $\Omega$  into itself is said to be *measure preserving* if:

- (1)  $T$  is measurable, i.e.  $T^{-1}A = \{\omega : T\omega \in A\} \in \mathfrak{F}$  for any  $A \in \mathfrak{F}$ ; and
- (2)  $T$  preserves the measure:  $\mathbf{P}(T^{-1}A) = \mathbf{P}(A)$  for any  $A \in \mathfrak{F}$ .

Let  $T$  be a measure preserving transformation,  $T^n$  its  $n$ -th iteration and  $\xi = \xi(\omega)$  be a random variable. Put  $U\xi(\omega) = \xi(T\omega)$ , so that  $U$  is a *transformation of random variables*, and  $U^k\xi(\omega) = \xi(T^k\omega)$ . Then

$$\xi = \{U^n\xi(\omega)\}_0^\infty = \{\xi(T^n\omega)\}_0^\infty \tag{16.1.2}$$

is a stationary sequence of random variables.

*Proof* Indeed, let  $A = \{\omega : \xi \in \mathbf{B}\}$ ,  $\mathbf{B} \in \mathfrak{B}^\infty$  and  $A_1 = \{\omega : \theta\xi \in \mathbf{B}\}$ . We have

$$\xi = (\xi(\omega), \xi(T\omega), \dots), \quad \theta\xi = (\xi(T\omega), \xi(T^2\omega), \dots).$$

Therefore  $\omega \in A_1$  if and only if  $T\omega \in A$ , i.e. when  $A_1 = T^{-1}A$ . But  $\mathbf{P}(T^{-1}A) = \mathbf{P}(A)$  and hence  $\mathbf{P}(A_1) = \mathbf{P}(A)$ , so that  $\mathbf{P}(A_n) = \mathbf{P}(A)$  for any  $n \geq 1$  as well, where  $A_n = \{\omega : \theta^n\xi \in \mathbf{B}\}$ . □

Stationary sequences defined by (16.1.2) will be referred to as *sequences generated by the transformation  $T$* .

To be able to construct stationary sequences on the whole axis, we will need measure preserving transformations acting both in “positive” and “negative” directions.

**Definition 16.1.3** A transformation  $T$  is said to be *bidirectional measure preserving* if:

- (1)  $T$  is a one-to-one transformation, the domain and range of  $T$  coincide with the whole  $\Omega$ ;
- (2) the transformations  $T$  and  $T^{-1}$  are measurable, i.e.

$$T^{-1}A = \{\omega : T\omega \in A\} \in \mathfrak{F}, \quad TA = \{T\omega : \omega \in A\} \in \mathfrak{F}$$

for any  $A \in \mathfrak{F}$ ;

- (3) the transformation  $T$  preserves the measure:  $\mathbf{P}(T^{-1}A) = \mathbf{P}(A)$ , and therefore  $\mathbf{P}(A) = \mathbf{P}(TA)$  for any  $A \in \mathfrak{F}$ .

For such transformations we can, as before, construct stationary sequences  $\xi$  defined on the whole axis:

$$\xi = \{U^n\xi(\omega)\}_{-\infty}^\infty = \{\xi(T^n\omega)\}_{-\infty}^\infty.$$

The argument before Definition 16.1.2 shows that this approach “exhausts” all stationary sequences given on  $(\Omega, \mathfrak{F}, \mathbf{P})$ , i.e. to any stationary sequence  $\xi$  we can relate a measure preserving transformation  $T$  and a random variable  $\xi = \xi_0$  such that  $\xi_k(\omega) = \xi_0(T^k\omega)$ . In this construction, we consider the “sample probability space”  $(\mathbb{R}^\infty, \mathfrak{B}^\infty, \mathbf{P})$  for which  $\xi(\omega) = \omega, \theta = T$ . The transformation  $\theta = T$  (that is,

transformation (16.1.1)) will be called the *pathwise shift transformation*. It always exists and “generates” any stationary sequence.

Now we will give some simpler examples of (bidirectional) measure preserving transformations.

*Example 16.1.1* Let  $\Omega = \{\omega_1, \dots, \omega_d\}$ ,  $d \geq 2$ , be a finite set,  $\mathfrak{F}$  be the  $\sigma$ -algebra of all its subsets,  $T\omega_i = \omega_{i+1}$ ,  $1 \leq i \leq d - 1$  and  $T\omega_d = \omega_1$ . If  $\mathbf{P}(\omega_i) = 1/d$  then  $T$  and  $T^{-1}$  are measure preserving transformations.

*Example 16.1.2* Let  $\Omega = [0, 1)$ ,  $\mathfrak{F}$  be the  $\sigma$ -algebra of Borel sets,  $\mathbf{P}$  the Lebesgue measure and  $s$  a fixed number. Then  $T\omega = \omega + s \pmod{1}$  is a bidirectional measure preserving transformation.

In these examples, the spaces  $\Omega$  are rather small, which allows one to construct on them only stationary sequences with deterministic or almost deterministic dependence between their elements. If we choose in Example 16.1.1 the variable  $\xi$  so that all  $\xi(\omega_i)$  are different, then the value  $\xi_k(\omega) = \xi(T^k\omega)$  will uniquely determine  $T^k\omega$  and thereby  $T^{k+1}\omega$  and  $\xi_{k+1}(\omega)$ . The same can be said of Example 16.1.2 in the case when  $\xi(\omega)$ ,  $\omega \in [0, 1)$ , is a monotone function of  $\omega$ .

As our argument at the beginning of the section shows, the space  $\Omega = \mathbb{R}^\infty$  is large enough to construct on it any stationary sequence.

Thus, we see that the concept of a measure preserving transformation arises in a natural way when studying stationary processes. But not only in that case. It also arises, for instance, while studying the dynamics of some physical systems. Indeed, the whole above argument remains valid if we consider on  $(\Omega, \mathfrak{F})$  an arbitrary measure  $\mu$  instead of the probability  $\mathbf{P}$ . For example, for  $\Omega = \mathbb{R}^\infty$ , the value  $\mu(A)$ ,  $A \in \mathfrak{F}$ , could be the Lebesgue measure (volume) of the set  $A$ . The measure preserving property of the transformation  $T$  will mean that any set  $A$ , after the transform  $T$  has acted on it (which, say, corresponds to the change of the physical system’s state in one unit of time), will retain its volume. This property is rather natural for incompressible liquids. Many laws to be established below will be equally applicable to such physical systems.

Returning to probabilistic models, i.e. to the case when the measure is a probability distribution, it turns out that, in that case, for any set  $A$  with  $\mathbf{P}(A) > 0$ , the “trajectory”  $T^n\omega$  will visit  $A$  infinitely often for almost all (with respect to the measure  $\mathbf{P}$ )  $\omega \in A$ .

**Theorem 16.1.1** (Poincaré) *Let  $T$  be a measure preserving transformation and  $A \in \mathfrak{F}$ . Then, for almost all  $\omega \in A$ , the relation  $T^n\omega \in A$  holds for infinitely many  $n \geq 1$ .*

*Proof* Put  $N := \{\omega \in A : T^n\omega \notin A \text{ for all } n \geq 1\}$ . Because  $\{\omega : T^n\omega \in A\} \in \mathfrak{F}$ , it is not hard to see that  $N \in \mathfrak{F}$ . Clearly,  $N \cap T^{-n}N = \emptyset$  for any  $n \geq 1$ , and  $T^{-m}N \cap T^{-(m+n)}N = T^{-m}(N \cap T^{-n}N) = \emptyset$ . This means that we have infinitely many sets  $T^{-n}N$ ,  $n = 0, 1, 2, \dots$ , which are disjoint and have one and the same probability. This evidently implies that  $\mathbf{P}(N) = 0$ .

Thus, for each  $\omega \in A \setminus N$ , there exists an  $n_1 = n_1(\omega)$  such that  $T^{n_1}\omega \in A$ . Now we apply this assertion to the measure preserving mapping  $T_k = T^k, k \geq 1$ . Then, for each  $\omega \in A \setminus N_k, \mathbf{P}(N_k) = 0$ , there exists an  $n_k = n_k(\omega) \geq 1$  such that  $(T^k)^{n_k}\omega \in A$ . Since  $kn_k \geq k$ , the theorem is proved.  $\square$

**Corollary 16.1.1** *Let  $\xi(\omega) \geq 0$  and  $A = \{\omega : \xi(\omega) > 0\}$ . Then, for almost all  $\omega \in A$ ,*

$$\sum_{n=0}^{\infty} \xi(T^n \omega) = \infty.$$

*Proof* Put  $A_k = \{\omega : \xi(\omega) \geq 1/k\} \subset A$ . Then by Theorem 16.1.1 the above series diverges for almost all  $\omega \in A_k$ . It remains to notice that  $A = \bigcup_k A_k$ .  $\square$

*Remark 16.1.1* Formally, one does not need condition  $\mathbf{P}(A) > 0$  in Theorem 16.1.1 and Corollary 16.1.1. However, in the absence of that condition, the assertions may become meaningless, since the set  $A \setminus N$  in the proof of Theorem 16.1.1 can turn out to be empty. Suppose, for example, that in the conditions of Example 16.1.2,  $A$  is a one-point set:  $A = \{\omega\}, \omega \in [0, 1)$ . If  $s$  is irrational, then  $T^k\omega$  will never be in  $A$  for  $k \geq 1$ . Indeed, if we assume the contrary, then we will infer that there exist integers  $k$  and  $m$  such that  $\omega + sk - m = \omega, s = m/k$ , which contradicts the irrationality of  $s$ .

## 16.2 Ergodicity (Metric Transitivity), Mixing and Weak Dependence

**Definition 16.2.1** A set  $A \in \mathfrak{F}$  is said to be *invariant* (with respect to a measure preserving transformation  $T$ ) if  $T^{-1}A = A$ . A set  $A \in \mathfrak{F}$  is said to be *almost invariant* if the sets  $T^{-1}A$  and  $A$  differ from each other by a set of probability zero:  $\mathbf{P}(A \oplus T^{-1}A) = 0$ , where  $A \oplus B = A\bar{B} \cup \bar{A}B$  is the symmetric difference.

It is evident that the class of all invariant (almost invariant) sets forms a  $\sigma$ -algebra which will be denoted by  $\mathfrak{I} (\mathfrak{I}^*)$ .

**Lemma 16.2.1** *If  $A$  is an almost invariant set then there exists an invariant set  $B$  such that  $\mathbf{P}(A \oplus B) = 0$ .*

*Proof* Put  $B = \limsup_{n \rightarrow \infty} T^{-n}A$  (recall that  $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$  is the set of all points which belong to infinitely many sets  $A_k$ ). Then

$$T^{-1}B = \limsup_{n \rightarrow \infty} T^{-(n+1)}A = B,$$

i.e.  $B \in \mathfrak{I}$ . It is not hard to see that

$$A \oplus B \subset \bigcup_{k=0}^{\infty} (T^{-k} A \oplus T^{-(k+1)} A).$$

Since

$$\mathbf{P}(T^{-k} A \oplus T^{-(k+1)} A) = \mathbf{P}(A \oplus T^{-1} A) = 0,$$

we have  $\mathbf{P}(A \oplus B) = 0$ . The lemma is proved.  $\square$

**Definition 16.2.2** A measure preserving transformation  $T$  is said to be *ergodic* (or *metric transitive*) if each invariant set has probability zero or one.

A stationary sequence  $\{\xi_k\}$  associated with such  $T$  (i.e. the sequence which generated  $T$  or was generated by  $T$ ) is also said to be ergodic (metric transitive).

**Lemma 16.2.2** A transformation  $T$  is ergodic if and only if each almost invariant set has probability 0 or 1.

*Proof* Let  $T$  be ergodic and  $A \in \mathfrak{I}^*$ . Then by Lemma 16.2.1 there exists an invariant set  $B$  such that  $\mathbf{P}(A \oplus B) = 0$ . Because  $\mathbf{P}(B) = 0$  or 1, the probability  $\mathbf{P}(A) = 0$  or 1. The converse assertion is obvious.  $\square$

**Definition 16.2.3** A random variable  $\zeta = \zeta(\omega)$  is said to be *invariant* (almost *invariant*) if  $\zeta(\omega) = \zeta(T\omega)$  for all  $\omega \in \Omega$  (for almost all  $\omega \in \Omega$ ).

**Theorem 16.2.1** Let  $T$  be a measure preserving transformation. The following three conditions are equivalent:

- (1)  $T$  is ergodic;
- (2) each almost invariant random variable is a.s. constant;
- (3) each invariant random variable is a.s. constant.

*Proof* (1)  $\Rightarrow$  (2). Assume that  $T$  is ergodic and  $\xi$  is almost invariant, i.e.  $\xi(\omega) = \xi(T\omega)$  a.s. Then, for any  $v \in \mathbb{R}$ , we have  $A_v := \{\omega : \xi(\omega) \leq v\} \in \mathfrak{I}^*$  and, by Lemma 16.2.2,  $\mathbf{P}(A_v)$  equals 0 or 1. Put  $V := \sup\{v : \mathbf{P}(A_v) = 0\}$ . Since  $A_v \uparrow \Omega$  as  $v \uparrow \infty$  and  $A_v \downarrow \emptyset$  as  $v \downarrow -\infty$ , one has  $|V| < \infty$  and

$$\mathbf{P}(\xi(\omega) < V) = \mathbf{P}\left(\bigcup_{n=1}^{\infty} \left\{ \xi(\omega) < V - \frac{1}{n} \right\}\right) = 0.$$

Similarly,  $\mathbf{P}(\xi(\omega) > V) = 0$ . Therefore  $\mathbf{P}(\xi(\omega) = V) = 1$ .

(2)  $\Rightarrow$  (3). Obvious.

(3)  $\Rightarrow$  (1). Let  $A \in \mathfrak{I}$ . Then the indicator function  $I_A$  is an invariant random variable, and since it is constant, one has either  $I_A = 0$  or  $I_A = 1$  a.s. This implies that  $\mathbf{P}(A)$  equals 0 or 1. The theorem is proved.  $\square$

The assertion of the theorem clearly remains valid if one considers in (3) only bounded random variables. Moreover, if  $\xi$  is invariant, then the truncated variable  $\xi_{(N)} = \min(\xi, N)$  is also invariant.

Returning to Examples 16.1.1 and 16.1.2, in Example 16.1.1,

$$\Omega = (\omega_1, \dots, \omega_d), \quad T\omega_i = \omega_{i+1 \pmod d}, \quad \mathbf{P}(\omega_i) = 1/d.$$

The transformation  $T$  is obviously metric transitive.

In Example 16.1.2,  $\Omega = [0, 1)$ ,  $T\omega = \omega + s \pmod 1$ , and  $\mathbf{P}$  is the Lebesgue measure. We will now show that  $T$  is ergodic if and only if  $s$  is irrational.

Consider a square integrable random variable  $\xi = \xi(\omega) : \mathbf{E}\xi^2(\omega) < \infty$ . Then by the Parseval equality, the Fourier series

$$\xi(\omega) = \sum_{n=0}^{\infty} a_n e^{2\pi i n \omega}$$

for this function has the property  $\sum_{n=0}^{\infty} |c_n^2| < \infty$ . Assume that  $s$  is irrational, while  $\xi$  is invariant. Then

$$\begin{aligned} a_n &= \mathbf{E}\xi(\omega)e^{-2\pi i n \omega} = \mathbf{E}\xi(T\omega)e^{-2\pi i n T\omega} \\ &= e^{-2\pi i n s} \mathbf{E}\xi(T\omega)e^{-2\pi i n \omega} = e^{-2\pi i n s} \mathbf{E}\xi(\omega)e^{-2\pi i n \omega} = e^{-2\pi i n s} a_n. \end{aligned}$$

For irrational  $s$ , this equality is only possible when  $a_n = 0, n \geq 1$ , and  $\xi(\omega) = a_0 = \text{const}$ . By Theorem 16.2.1 this means that  $T$  is ergodic.

Now let  $s = m/n$  be rational ( $m$  and  $n$  are integers). Then the set

$$A = \bigcup_{k=0}^{n-1} \left\{ \omega : \frac{2k}{2n} \leq \omega < \frac{2k+1}{2n} \right\}$$

will be invariant and  $\mathbf{P}(A) = 1/2$ . This means that  $T$  is not ergodic. □

**Definition 16.2.4** A measure preserving transformation  $T$  is called *mixing* if, for any  $A_1, A_2 \in \mathfrak{F}$ , as  $n \rightarrow \infty$ ,

$$\mathbf{P}(A_1 \cap T^{-n}A_2) \rightarrow \mathbf{P}(A_1)\mathbf{P}(A_2). \tag{16.2.1}$$

Now consider the stationary sequence  $\xi = (\xi_0, \xi_1, \dots)$  generated by the transformation  $T : \xi_k(\omega) = \xi_0(T^k\omega)$ .

**Definition 16.2.5** A stationary sequence  $\xi$  is said to be *weakly dependent* if  $\xi_k$  and  $\xi_{k+n}$  are asymptotically independent as  $n \rightarrow \infty$ , i.e. for any  $B_1, B_2 \in \mathfrak{B}$

$$\mathbf{P}(\xi_k \in B_1, \xi_{k+n} \in B_2) \rightarrow \mathbf{P}(\xi_0 \in B_1)\mathbf{P}(\xi_0 \in B_2). \tag{16.2.2}$$

**Theorem 16.2.2** A measure preserving transformation  $T$  is mixing if and only if any stationary sequence  $\xi$  generated by  $T$  is weakly dependent.

*Proof* Let  $T$  be mixing. Put  $A_i := \xi_0^{-1}(B_i)$ ,  $i = 1, 2$ , and set  $k = 0$  in (16.2.2). Then

$$\mathbf{P}(\xi_0 \in B_1, \xi_n \in B_2) = \mathbf{P}(A_1 \cap T^{-n}A_2) \rightarrow \mathbf{P}(A_1)\mathbf{P}(A_2).$$

Now assume any sequence generated by  $T$  is weakly dependent. For any given  $A_1, A_2 \in \mathfrak{F}$ , define the random variable

$$\xi(\omega) = \begin{cases} 0 & \text{if } \omega \notin A_1 \cup A_2; \\ 1 & \text{if } \omega \in A_1 \bar{A}_2; \\ 2 & \text{if } \omega \in A_1 A_2; \\ 3 & \text{if } \omega \in \bar{A}_1 A_2; \end{cases}$$

and put  $\xi_k(\omega) := \xi(T^k\omega)$ . Then, as  $n \rightarrow \infty$ ,

$$\begin{aligned} \mathbf{P}(A_1 \cap T^{-n}A_2) &= \mathbf{P}(0 < \xi_0 < 3, \xi_n > 2) \rightarrow \mathbf{P}(0 < \xi_0 < 3)\mathbf{P}(\xi_0 > 2) \\ &= \mathbf{P}(A_1)\mathbf{P}(A_2). \end{aligned}$$

The theorem is proved.  $\square$

Let  $\{X_n\}$  be a stationary real-valued Markov chain with an invariant distribution  $\pi$  that satisfies the conditions of the ergodic theorem, i.e. such that, for any  $B \in \mathfrak{B}$  and  $x \in \mathbb{R}$ , as  $n \rightarrow \infty$ ,

$$\mathbf{P}(X_n \in B \mid X_0 = x) \rightarrow \pi(B).$$

Then  $\{X_n\}$  is weakly dependent, and therefore, by Theorem 16.2.2, the respective transformation  $T$  is mixing. Indeed,

$$\mathbf{P}(X_0 \in B_1, X_n \in B_2) = \mathbf{E}I(X_0 \in B_1)\mathbf{P}(X_n \in B_2 \mid X_0),$$

where the last factor converges to  $\pi(B_2)$  for each  $X_0$ . Therefore the above probability tends to  $\pi(B_2)\pi(B_1)$ .

Further characterisations of the mixing property will be given in Theorems 16.2.4 and 16.2.5.

Now we will introduce some notions that are somewhat broader than those from Definitions 16.2.4 and 16.2.5.

**Definition 16.2.6** A transformation  $T$  is called *mixing on the average* if, as  $n \rightarrow \infty$ ,

$$\frac{1}{n} \sum_{k=1}^n \mathbf{P}(A_1 \cap T^{-k}A_2) \rightarrow \mathbf{P}(A_1)\mathbf{P}(A_2). \quad (16.2.3)$$

A stationary sequence  $\xi$  is said to be *weakly dependent on the average* if

$$\frac{1}{n} \sum_{k=1}^n \mathbf{P}(\xi_0 \in B_1, \xi_k \in B_2) \rightarrow \mathbf{P}(\xi_0 \in B_1)\mathbf{P}(\xi_0 \in B_2). \quad (16.2.4)$$



**Theorem 16.2.3** *A measure preserving transformation  $T$  is mixing on the average if and only if any stationary sequence  $\xi$  generated by  $T$  is weakly dependent on the average.*

*The Proof* is the same as for Theorem 16.2.2, and is left to the reader.  $\square$

If  $\{X_n\}$  is a periodic real-valued Markov chain with period  $d$  such that each of the embedded sub-chains  $\{X_{i+nd}\}_{n=0}^{\infty}$ ,  $i = 0, \dots, d-1$ , satisfies the ergodicity conditions with invariant distributions  $\pi^{(i)}$  on disjoint sets  $\mathcal{X}_0, \dots, \mathcal{X}_{d-1}$ , then the “common” invariant distribution  $\pi$  will be equal to  $d^{-1} \sum_{i=0}^{d-1} \pi^{(i)}$ , and the chain  $\{X_n\}$  will be weakly dependent on the average. At the same time, it will clearly not be weakly dependent for  $d > 1$ .

**Theorem 16.2.4** *A measure preserving transformation  $T$  is ergodic if and only if it is mixing on the average.*

*Proof* Let  $T$  be mixing on the average, and  $A_1 \in \mathfrak{F}$ ,  $A_2 \in \mathfrak{J}$ . Then  $A_2 = T^{-k}A_2$  and hence  $\mathbf{P}(A_1 \cap T^{-k}A_2) = \mathbf{P}(A_1A_2)$  for all  $k \geq 1$ . Therefore, (16.2.3) means that  $\mathbf{P}(A_1A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2)$ . For  $A_1 = A_2$  we get  $\mathbf{P}(A_2) = \mathbf{P}^2(A_2)$ , and consequently  $\mathbf{P}(A_2)$  equals 0 or 1.

We postpone the proof of the converse assertion until the next section.  $\square$

Now we will give one more important property of ergodic transforms.

**Theorem 16.2.5** *A measure preserving transformation  $T$  is ergodic if and only if, for any  $A \in \mathfrak{F}$  with  $\mathbf{P}(A) > 0$ , one has*

$$\mathbf{P}\left(\bigcup_{n=0}^{\infty} T^{-n}A\right) = 1. \quad (16.2.5)$$

Note that property (16.2.5) means that the sets  $T^{-n}A$ ,  $n = 0, 1, \dots$ , “exhaust” the whole space  $\Omega$ , which associates well with the term “mixing”.

*Proof* Let  $T$  be ergodic. Put  $B := \bigcup_{n=0}^{\infty} T^{-n}A$ . Then  $T^{-1}B \subset B$ . Because  $T$  is measure preserving, one also has that  $\mathbf{P}(T^{-1}B) = \mathbf{P}(B)$ . From this it follows that  $T^{-1}B = B$  up to a set of measure 0 and therefore  $B$  is almost invariant. Since  $T$  is ergodic,  $\mathbf{P}(B)$  equals 0 or 1. But  $\mathbf{P}(B) \geq \mathbf{P}(A) > 0$ , and hence  $\mathbf{P}(B) = 1$ .

Conversely, if  $T$  is not ergodic, then there exists an invariant set  $A$  such that  $0 < \mathbf{P}(A) < 1$  and, therefore, for this set  $T^{-n}A = A$  holds and

$$\mathbf{P}(B) = \mathbf{P}(A) < 1.$$

The theorem is proved.  $\square$

*Remark 16.2.1* In Sects. 16.1 and 16.2 we tacitly or explicitly assumed (mainly for the sake of simplicity of the exposition) that the components  $\xi_k$  of the stationary sequence  $\xi$  are real. However, we never actually used this, and so we could, as we did while studying Markov chains, assume that the state space  $\mathcal{X}$ , in which  $\xi_k$  take their values, is an arbitrary measurable space. In the next section we will substantially use the fact that  $\xi_k$  are real- or vector-valued.

### 16.3 The Ergodic Theorem

For a sequence  $\xi_1, \xi_2, \dots$  of independent identically distributed random variables we proved in Chap. 11 the strong law of large numbers:

$$\frac{S_n}{n} \xrightarrow{a.s.} \mathbf{E}\xi_1, \quad \text{where } S_n = \sum_{k=0}^{n-1} \xi_k.$$

Now we will prove the same assertion under much broader assumptions—for stationary ergodic sequences, i.e. for sequences that are weakly dependent on the average.

Let  $\{\xi_k\}$  be an arbitrary strictly stationary sequence,  $T$  be the associated measure preserving transformation, and  $\mathcal{J}$  be the  $\sigma$ -algebra of invariant sets.

**Theorem 16.3.1** (Birkhoff–Khintchin) *If  $\mathbf{E}|\xi_0| < \infty$  then*

$$\frac{1}{n} \sum_{k=0}^{n-1} \xi_k \xrightarrow{a.s.} \mathbf{E}(\xi_0 | \mathcal{J}). \quad (16.3.1)$$

*If the sequence  $\{\xi_k\}$  (or transformation  $T$ ) is ergodic, then*

$$\frac{1}{n} \sum_{k=0}^{n-1} \xi_k \xrightarrow{a.s.} \mathbf{E}\xi_0. \quad (16.3.2)$$

Below we will be using the representation  $\xi_k = \xi(T^k \omega)$  for  $\xi = \xi_0$ . We will need the following auxiliary result.

**Lemma 16.3.1** *Set*

$$S_n(\omega) := \sum_{k=0}^{n-1} \xi(T^k \omega), \quad M_k(\omega) := \max\{0, S_1(\omega), \dots, S_k(\omega)\}.$$

*Then, under the conditions of Theorem 16.3.1,*

$$\mathbf{E}[\xi(\omega) \mathbf{I}_{\{M_n > 0\}}(\omega)] \geq 0$$

*for any  $n \geq 1$ .*

*Proof* For all  $k \leq n$ , one has  $S_k(T\omega) \leq M_n(T\omega)$ , and hence

$$\xi(\omega) + M_n(T\omega) \geq \xi(\omega) + S_k(T\omega) = S_{k+1}(\omega).$$

Because  $\xi(\omega) \geq S_1(\omega) - M_n(T\omega)$ , we have

$$\xi(\omega) \geq \max\{\max(S_1(\omega), \dots, S_n(\omega)) - M_n(T\omega)\}.$$

Further, since

$$\{M_n(\omega) > 0\} = \{\max(S_1(\omega), \dots, S_n(\omega)) > 0\},$$

we obtain that

$$\begin{aligned} \mathbf{E}[\xi(\omega)I_{\{M_n > 0\}}(\omega)] &\geq \mathbf{E}(\max(S_1(\omega), \dots, S_n(\omega)) - M_n(T\omega)) I_{\{M_n > 0\}}(\omega) \\ &\geq \mathbf{E}(M_n(\omega) - M_n(T\omega)) I_{\{M_n > 0\}}(\omega) \\ &\geq \mathbf{E}(M_n(\omega) - M_n(T\omega)) = 0. \end{aligned}$$

The lemma is proved.  $\square$

*Proof of Theorem 16.3.1* Assertion (16.3.2) is an evident consequence of (16.3.1), because, for ergodic  $T$ , the  $\sigma$ -algebra  $\mathfrak{I}$  is trivial and  $\mathbf{E}(\xi|\mathfrak{I}) = \mathbf{E}\xi$  a.s. Hence, it suffices to prove (16.3.1).

Without loss of generality, we can assume that  $\mathbf{E}(\xi|\mathfrak{I}) = 0$ , for one can always consider  $\xi - \mathbf{E}(\xi|\mathfrak{I})$  instead of  $\xi$ .

Let  $\bar{S} := \limsup_{n \rightarrow \infty} n^{-1} S_n$  and  $\underline{S} := \liminf_{n \rightarrow \infty} n^{-1} S_n$ . To prove the theorem, it suffices to establish that

$$0 \leq \underline{S} \leq \bar{S} \leq 0 \quad \text{a.s.} \quad (16.3.3)$$

Since  $\bar{S}(\omega) = \bar{S}(T\omega)$ , the random variable  $\bar{S}$  is invariant and hence the set  $A - \varepsilon = \{\bar{S}(\omega) > \varepsilon\}$  is also invariant for any  $\varepsilon > 0$ . Introduce the variables

$$\begin{aligned} \xi^*(\omega) &:= (\xi(\omega) - \varepsilon)I_{A_\varepsilon}(\omega), \\ S_k^*(\omega) &:= \xi^*(\omega) + \dots + \xi^*(T^{k-1}\omega), \\ M_k^*(\omega) &:= \max(0, S_1^*, \dots, S_k^*). \end{aligned}$$

Then, by Lemma 16.3.1, for any  $n \geq 1$ , one has

$$\mathbf{E}\xi^* I_{\{M_n^* > 0\}} \geq 0.$$

But, as  $n \rightarrow \infty$ ,

$$\{M_n^* > 0\} = \left\{ \max_{1 \leq k \leq n} S_k^* > 0 \right\} \uparrow \left\{ \sup_{k \geq 1} S_k^* > 0 \right\}$$

$$= \left\{ \sup_{k \geq 1} \frac{S_k^*}{k} > 0 \right\} = \left\{ \sup_{k \geq 1} \frac{S_k}{k} > \varepsilon \right\} \cap A_\varepsilon = A_\varepsilon.$$

The last equality follows from the observation that

$$A_\varepsilon = \{\bar{S} > \varepsilon\} \subset \left\{ \sup_{k \geq 1} \frac{S_k}{k} > \varepsilon \right\}.$$

Further,  $\mathbf{E}|\xi^*| \leq \mathbf{E}|\xi| + \varepsilon$ . Hence, by the dominated convergence theorem,

$$0 \leq \mathbf{E}\xi^* \mathbf{I}_{\{M_n^* > 0\}} \rightarrow \mathbf{E}\xi^* \mathbf{I}_{A_\varepsilon}.$$

Consequently,

$$\begin{aligned} 0 &\leq \mathbf{E}\xi^* \mathbf{I}_{A_\varepsilon} = \mathbf{E}(\xi - \varepsilon) \mathbf{I}_{A_\varepsilon} = \mathbf{E}\xi \mathbf{I}_{A_\varepsilon} - \varepsilon \mathbf{P}(A_\varepsilon) \\ &= \mathbf{E} \mathbf{I}_{A_\varepsilon} \mathbf{E}(\xi \mid \mathcal{J}) - \varepsilon \mathbf{P}(A_\varepsilon) = -\varepsilon \mathbf{P}(A_\varepsilon). \end{aligned}$$

This implies that  $\mathbf{P}(A_\varepsilon) = 0$  for any  $\varepsilon > 0$ , and therefore  $\mathbf{P}(\bar{S} \leq 0) = 1$ .

In a similar way, considering the variables  $-\xi$  instead of  $\xi$ , we obtain that

$$\limsup_{n \rightarrow \infty} \left( -\frac{S_n}{n} \right) = -\liminf_{n \rightarrow \infty} \frac{S_n}{n} = -\underline{S},$$

and  $\mathbf{P}(-\underline{S} \leq 0) = 1$ ,  $\mathbf{P}(\underline{S} \geq 0) = 1$ . The required inequalities (16.3.3), and therefore the theorem itself, are proved.  $\square$

Now we can complete the

*Proof of Theorem 16.2.4* It remains to show that the ergodicity of  $T$  implies mixing on the average. Indeed, let  $T$  be ergodic and  $A_1, A_2 \in \mathfrak{F}$ . Then, by Theorem 16.3.1, we have

$$\zeta_n = \frac{1}{n} \sum_{k=1}^n \mathbf{I}(T^{-k} A_2) \xrightarrow{a.s.} \mathbf{P}(A_2), \quad \mathbf{I}(A_1) \zeta_n \xrightarrow{a.s.} \mathbf{I}(A_1) \mathbf{P}(A_2).$$

Since  $\zeta_n \mathbf{I}(A_1)$  are bounded, one also has the convergence

$$\mathbf{E} \zeta_n \mathbf{I}(A_1) \rightarrow \mathbf{P}(A_2) \cdot \mathbf{P}(A_1).$$

Therefore

$$\frac{1}{n} \sum_{k=1}^n \mathbf{P}(A_1 \cap T^{-k} A_2) = \mathbf{E} \mathbf{I}(A_1) \zeta_n \rightarrow \mathbf{P}(A_1) \mathbf{P}(A_2).$$

The theorem is proved.  $\square$

Now we will show that convergence in mean also holds in (16.3.1) and (16.3.2).

**Theorem 16.3.2** *Under the assumptions of Theorem 16.3.1, one has along with (16.3.1) and (16.3.2) that, respectively,*

$$\mathbf{E} \left| \frac{1}{n} \sum_{k=0}^{n-1} \xi_k - \mathbf{E}(\xi_0 | \mathcal{J}) \right| \rightarrow 0 \quad (16.3.4)$$

and

$$\mathbf{E} \left| \frac{1}{n} \sum_{k=0}^{n-1} \xi_k - \mathbf{E} \xi_0 \right| \rightarrow 0 \quad (16.3.5)$$

as  $n \rightarrow \infty$ .

*Proof* The assertion of the theorem follows in an obvious way from Theorems 16.3.1, 6.1.7 and the uniform integrability of the sums

$$\frac{1}{n} \sum_{k=0}^{n-1} \xi_k,$$

which follows from Theorem 6.1.6. □

**Corollary 16.3.1** *If  $\{\xi_k\}$  is a stationary metric transitive sequence and  $a = \mathbf{E} \xi_k < 0$ , then  $S(\omega) = \sup_{k \geq 0} S_k(\omega)$  is a proper random variable.*

The proof is obvious since, for  $0 < \varepsilon < -a$ , one has  $S_k < (a + \varepsilon)k < 0$  for all  $k \geq n(\omega) < \infty$ . □

An unusual feature of Theorem 16.3.1 when compared with the strong law of large numbers from Chap. 11 is that the limit of

$$\frac{1}{n} \sum_{k=0}^{n-1} \xi_k$$

can be a random variable. For instance, let  $T\omega_k := \omega_{k+2}$  and  $d = 2l$  be even in the situation of Example 16.1.1. Then the transformation  $T$  will not be ergodic, since the set  $A = \{\omega_1, \omega_3, \dots, \omega_{d-1}\}$  will be invariant, while  $\mathbf{P}(A) = 1/2$ .

On the other hand, it is evident that, for any function  $\xi(\omega)$ , the sum

$$\frac{1}{n} \sum_{k=0}^{n-1} \xi(T^k \omega)$$

will converge with probability 1/2 to

$$\frac{2}{d} \sum_{j=0}^{l-1} \xi(\omega_{2j+1})$$

(if  $\omega = \omega_i$  and  $i$  is odd) and with probability  $1/2$  to

$$\frac{2}{d} \sum_{j=1}^l \xi(\omega_{2j})$$

(if  $\omega = \omega_i$  and  $i$  is even). This limiting distribution is just the distribution of  $\mathbf{E}(\xi|\mathcal{J})$ .