F. Stationary processes and the ergodic theorem

F.1 Stationarity

Let $\{X_n\ ;\ n \geq 1\}$ be a sequence of random variables indexed by the positive integers. We shall refer to such a collection as a *stochastic process* (see Chapter 2, section G) and think of the parameter n as a time index. So far we have studied large number laws when the process consists of a sequence of independent and identically distributed random variables. In this section we shall replace the assumption of independence with a much broader condition called *stationarity*. Informally speaking, a stochastic process $\{X_n\}$ is said to be stationary if its statistical behavior is independent of shifts in the time parameter n; that is, for any positive integer m, the processes $\{X_1, X_2, \ldots\}$ and $\{X_{m+1}, X_{m+2}, \ldots\}$ are statistically indistinguishable one from the other. This means that the process is in a kind of statistical steady state—it looks the same no matter what time one starts observing it. Such a steady state property is very natural in many applications, for example in communication and signal theory, in the study of queues, or in time series coming from economic and physical problems. In particular, stationary processes arise as long time limits of random dynamical systems, after transient phenomenon due to the influence of initial conditions die away. We shall illustrate this in Example 3 below.

To formulate the rigorous definition of stationarity efficiently, it is useful to define the *law* of a stochastic process. This was introduced implicity in Chapter 2, section G, in the construction of canonical processes. Let us repeat the construction, this time without compactifying the real line. The law of the stochastic processes $W_0 := \{X_n; n \geq 1\}$ will be a probability measure $\mathbb{F}_{Z_0}$ on the product space $\mathbb{R}^\infty$, with the product σ-algebra $\otimes_1^\infty \mathcal{B}(\mathbb{R})$, (which for convenience, will be denoted by $\mathcal{B}(\mathbb{R}^\infty)$); namely,

$$\mathbb{F}_{W_0}(U) = \mathbb{P}\left(\, (X_1, X_2, \ldots) \in U\, \right) \qquad U \in \mathcal{B}(\mathbb{R}^\infty).$$

Now given the process W_0 and any positive integer m, let $W_m := \{X_{m+1}, X_{m+2} \ldots\}$ denote the same process restarted at $m+1$.

Definition The process $\{X_n; n \geq 1\}$ is *stationary* if for every positive integer m,

$$\mathbb{F}_{W_0} = \mathbb{F}_{W_m}. \qquad (1)$$

This definition states precisely that the statistical description of the process is invariant with respect to shifts of the starting time.

The class of stationary processes includes the i.i.d. processes we have been working with.

Example 1. Any i.i.d. processes is stationary.

To see this let, $W_0 := \{X_n\}$ be an i.i.d. sequence, and let $\mathbb{F}$ denote the common distribution measure of the X_i's; that is, for any Borel subset G in $\mathbb{R}$ and any positive integer n, $\mathbb{F}(G) = \mathbb{P}(X_n \in G)$. Then $\mathbb{F}_{W_0} = \mathbb{F}^\infty$. But for any positive integer m,

$W_m := \{X_{m+1}, X_{m+2}, \ldots\}$ is again a sequence of i.i.d. random variables with common distribution $\mathbb{F}$, and hence again has law $\mathbb{F}^\infty$. $\diamond$

Before going further, note that the different random variables of any stationary process must be identically distributed. Indeed, if $W_0 := \{X_n\}$ is stationary, then for any Borel subset G of $\mathbb{R}$ and any positive integer m,

$$\mathbb{P}(X_1 \in G) = \mathbb{F}_{W_0}(G \times \mathbb{R}^\infty) = \mathbb{F}_{Z_m}(G \times \mathbb{R}^\infty) = \mathbb{P}(X_m \in G).$$

However, stationary processes need not consist of independent random variables, as we shall see shortly by example.

The next lemma states some easy necessary and sufficient conditions for stationarity. Condition (b) of this lemma will be particularly important for the theoretical development of stationarity and requires a preliminary definition. On $\mathbb{R}^\infty$ define the left shift operator

$$T_\ell(x_1, x_2, x_3 \ldots) = (x_2, x_3, \ldots).$$

T_ℓ is measurable as a map from $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ to itself; indeed, if U is any set belonging to $\mathcal{B}(\mathbb{R}^\infty)$, then $T_\ell^{-1}(U) = \mathbb{R} \times U$, which is again a set in $\mathcal{B}(\mathbb{R}^\infty)$. For any probability measure $\mathbb{F}$ on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$, define

$$\left(\mathbb{F} \circ T_\ell^{-1}\right)(U) = \mathbb{F}\left(T_\ell^{-1}(U)\right) \qquad U \in \mathcal{B}(\mathbb{R}^\infty).$$

Similarly, $\mathbb{F} \circ T_\ell^{-2} = (\mathbb{F} \circ T_\ell^{-1}) \circ T_\ell^{-1}$, and iterating one defines $\mathbb{F} \circ T_\ell^{-n}$ for any positive integer n.

Lemma F.1 The following statements are equivalent:
(a) $W_0 := \{X_n; n \geq 1\}$ is stationary.
(b) $\mathbb{F} \circ T_\ell^{-1} = \mathbb{F}$.
(c) $\mathbb{F}_{W_0} = \mathbb{F}_{W_1}$.
(d) For every positive integer N, Borel set U in $\mathcal{B}(\mathbb{R}^N)$ and non-negative integer n,

$$\mathbb{P}((X_1, \ldots, X_N) \in U) = \mathbb{P}((X_{n+1}, \ldots, X_{n+N}) \in U).$$

Remark. Property (b) is described by saying that the left shift T_ℓ is *measure-preserving* for $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \mathbb{F}_{W_0})$. The notion of measure-preserving transformation will be generalized below and related more broadly to stationary processes.

Property (d) is the most concrete statement of the stationarity property.

Proof: Notice first that

$$\mathbb{F}_{W_m} = \mathbb{F}_{W_0} \circ T_\ell^{-m}$$

for any positive integer m. Setting $m = 1$, it follows that (b) and (c) are equivalent. Statement (d) is the same as saying that $\mathbb{F}_{W_0}$ and $\mathbb{F}_{W_1}$ agree on the cylinder sets of $\mathbb{R}^\infty$, and thus (c) implies (d). But (d) implies (c) because the algebra of cylinder sets generate $\mathcal{B}(\mathbb{R}^\infty)$ and we know from Proposition 4.2 in chapter 1 that two probability measure are equal if they agree on a generating algebra. Hence (b), (c) and (d) are equivalent.

Statement (a) implies statement (c) by taking $m = 1$ in the definition of stationarity. Since (b), (c), and (d) are equivalent, we complete the proof by showing that (b) implies (a). Indeed, if (b) is true, then for any positive integer m,

$$\mathbb{F}_{W_0} \circ T_\ell^{-m} = \left(\mathbb{F}_{W_0} \circ T_\ell^{-1}\right) \circ T_\ell^{-m+1} = \mathbb{F}_{W_0} \circ T_\ell^{-m+1},$$

and, iterating, $\mathbb{F}_{W_m} = \mathbb{F}_{W_0} \circ T_\ell^{-m} = \mathbb{F}_{W_0}$, for any positive integer m. Thus $W_0 = \{X_n\}$ is stationary. ◇

Example 2. Stationary processes built from other stationary processes.

Let $\{X_n\}$ be a stationary process, and let ϕ be any Borel measurable function of m variables, for any finite, positive integer m. Then

$$Y_k := \phi(X_k, \ldots, X_{k-1+m}), \quad k \geq 1, \qquad \text{is stationary.}$$

This is easy to see by criterion (d) and the stationarity of $\{X_n\}$. For any postive integer N, let

$$\Phi(x_1, \ldots, x_{N-1+m}) = \big(\phi(x_1, \ldots, x_{k-1+m}), \phi(x_2, \ldots, x_{k+m}), \ldots, \phi(x_N, \ldots, X_{N-1+m})\big).$$

Then

$$\begin{aligned} (Y_1, \ldots, Y_N) &= \Phi(X_1, \ldots, X_{N-1+m}) \qquad \text{and} \\ (Y_{n+1}, \ldots, Y_{n+N}) &= \Phi(X_{n+1}, \ldots, X_{n+N-1+m}) \end{aligned}$$

It is clear from the stationarity of $\{X_n\}$ that these two random vectors are identically distributed and hence that criterion (d) is satisfied.

There is no need in this construction for ϕ to depend only on a finite number of variables. Suppose instead that ϕ is a $\mathcal{B}(\mathbb{R}^\infty)$-measurable function on $\mathbb{R}^\infty$. Then the same reasoning as above shows that the process $Y_k = \phi(X_k, X_{k+1}, \ldots)$, $k \geq 1$, is stationary.

We can now construct examples of stationary processes that are not i.i.d. Let $\{X_k\}$ be an i.i.d. process, and let ϕ be any multivariate function that depends non-trivially on more than one of its variables. We know that $\{X_k\}$ is stationary. Hence the associate process $\{Y_k\}$ will be stationary also, but the Y_k's will not in general be independent.

A popular example in time series of the processes constructed here is the *moving average* process,

$$Y_k = a_0 X_k + a_1 X_{k-1} + \cdots + a_N X_{k-N}, \qquad k \geq 1$$

(To properly define X_k for $1 \leq k \leq N$, assume that X_n is defined for $n \geq 1 - N$.)

Returning to the general case $W_0 := \{X_k\}_{k\geq 1}$, notice that at time $m > 1$, there is a past to the process W_m. To accomodate the past in the definition of stationarity, it is natural to generalize the definition to doubly infinite sequences of random variables, that is, to processes $W = \{\ldots, X_{-1}, X_0, X_1, \ldots\}$ indexed by the set of all integers, positive and negative. To build the proper canonical space for such a process, let $\mathbb{R}_i$, $-\infty < i < \infty$ be labelled copies of the real line, and consider the product $\otimes_{-\infty}^{\infty} \mathbb{R}_i$ with the σ-algebra

$\mathcal{B}(\otimes_{-\infty}^{\infty} I\!R_i)$ generated by the algebra of cylinder sets in the usual way. Consider the process W as a random vector

$$W(\omega) = (\dots, X_{-1}(\omega), X_0(\omega), X_1(\omega), \dots) \in \otimes_{-\infty}^{\infty} I\!R_i,$$

where $X_i(\omega)$ is the coordinate of $W(\omega)$ in $I\!R_i$. Then the law of W is defined in the usual way by $I\!F_W(U) = I\!P(W \in U)$. Define the left shift T_ℓ on $\otimes_{-\infty}^{\infty} I\!R_i$, so that, if $x = (\dots, x_{-1}, x_0, x_1, \dots)$, one has $(T_\ell(x))_i = x_{i+1}$, where $(T_\ell(x))_i$ is the projection of $T_\ell(x)$ on $I\!R_i$. Then W is called stationary if $I\!F_W = I\!F_W \circ T_\ell^{-1}$. Note that in this case the left shift T_ℓ admits a measurable inverse, namely the right shift, and so T_ℓ^n makes sense for any integer n, positive or negative.

Example 3. This example illustrates how a stationary process can arise as the limit of a random dynamical system running for a long time. Let $\{\xi_j \,;\, -\infty < j < \infty\}$ be a doubly infinite sequence of i.i.d. random variables with mean zero and variance one.

Let a be a number $|a| < 1$. Consider the equation

$$\begin{aligned} X_n^{(N)} &= a X_{n-1}^{(N)} + \xi_n, \qquad n \geq -N+1 \\ X_{-N}^{(N)} &= 0 \end{aligned}$$

It may be checked easily that the solution to this equation is

$$X_n^{(N)} = \sum_{k=-N+1}^{n} a^{n-k} \xi_k.$$

Now take the limit as $N \to \infty$; this corresponds to starting the process further and further in the past. Formally, the limit

$$X_n = \sum_{k=0}^{\infty} a^k \xi_{n-k}.$$

The reader should check that the series defining X_n converges a.s. for every integer n. We claim that $\{X_n \,;\, n \geq 1\}$ is a stationary process. (In fact, the doubly infinite sequence $\{X_n \,;\, -\infty < n < \infty\}$ is stationary in the sense defined above.) Indeed, for $x \in \otimes_{-\infty}^{\infty} I\!R_i$, let

$$H(x) = \begin{cases} \sum_0^\infty a^k x_{-k}, & \text{if the series converges;} \\ 0, & \text{otherwise.} \end{cases}$$

Then for every integer n,

$$X_n(\omega) = H\left(T_\ell^n\left(\dots, \xi_{-1}(\omega), \xi_0(\omega), \xi_1(\omega), \dots\right)\right).$$

Because $\{\xi_j \,;\, -\infty < j < \infty\}$ is stationary it follows that X_n is also stationary by the same argument as used in example 2. The process $\{X_n\}$ is really an example of a moving average process, although now the sum defining X_n is infinite.

F.2 Measure-preserving transformations and stationarity

The concept of a measure preserving transformation is closely related to stationarity and gives a method for constructing stationary processes generalizing the method of examples 2 and 3.

Definition Let $(\Omega, \mathcal{F}, I\!\!P)$ be a probability space and let $T : \Omega \to \Omega$ be a measurable map—that is, $T^{-1}(A) \in \mathcal{F}$ for every $A \in \mathcal{F}$. T is said to *measure-preserving* for $(\Omega, \mathcal{F}, I\!\!P)$ if

$$I\!\!P\left(T^{-1}(A)\right) = I\!\!P(A) \qquad \text{for every } A \in \mathcal{F}.$$

We saw above that the left shift on $I\!\!R^\infty$ is measure-preserving for the law of a stationary process. Here are some other examples from analysis.

Example 4. Let $(\Omega, \mathcal{F}, I\!\!P) = ([0,1), \mathcal{B}, \lambda)$, where λ is Lebesgue measure and $\mathcal{B}$ denotes the Borel sets. Then

$$T_1(x) = 2x \mod 1$$

is measure-preserving. So also is

$$T_2(x) = (x + \alpha) \mod 1$$

for any real number α.

Lemma F.1 showed that stationarity can be expressed in terms of a measure-preserving transformation (the left shift) on a probability space. We show next a kind of converse: given a measure-preserving transformation on a probability space, one can construct from it many stationary processes. The basic idea is this. Suppose T is any measurable map from $(\Omega, \mathcal{F})$ to itself, measure-preserving or not, and let X be any random variable on $(\Omega, \mathcal{F})$. For notational simplicity, denote the action of T on a point ω by $T\omega$ rather than $T(\omega)$. Then

$$X_n(\omega) := X\left(T^{n-1}\omega\right), \qquad n \geq 1, \tag{2}$$

defines a sequence of random variables; for a fixed ω, the sequence $X(\omega)$, $X(T\omega), \ldots$, $X(T^{k-1}\omega), \ldots$ samples the values of X along the orbit $(\omega, T\omega, \ldots)$ of T.

Theorem 3.5.1. Let T be a measure-preserving transformation on $(\Omega, \mathcal{F}, I\!\!P)$ and let X be a random variable on $(\Omega, \mathcal{F}, I\!\!P)$. Define $\{X_n\}$ as in (2). Then $\{X_n\,;\, n \geq 1\}$ is a stationary process on $(\Omega, \mathcal{F}, I\!\!P)$.

Proof: Fix $N \geq 1$ and an arbitrary Borel subset B of $I\!\!R^N$. Let

$$A := \{\omega\,;\, (X_1(\omega), \ldots, X_N(\omega)) \in B\}.$$

Since T is measure-preserving,

$$\begin{aligned} I\!\!P\,(\,(X_1, \ldots, X_N) \in B\,) &= I\!\!P(A) = I\!\!P\left(T^{-1}(A)\right) \\ &= I\!\!P\left(\,\left\{\omega\,;\, (X(T(\omega)), \ldots, X(T^N)(\omega)) \in B\right\}\right) \\ &= I\!\!P\,(\,(X_2, \ldots, X_{N+1}) \in B\,) \end{aligned}$$

Now applying characterization (d) from Lemma 1 proves that $\{X_n\}$ is stationary. $\diamond$

Note that this proof is a generalization of the construction in Example 2.

Exercise. Let T_1 denote the measure-preserving transformation on $([0,1),\mathcal{B},\lambda)$ defined in example 4. Let $X(\omega) = \mathbf{1}_{[1/2,1)}\omega$ for $\omega \in [0,1)$. Then $X_n(\omega) := X(T^{n-1}\omega)$ defines an i.i.d. sequence of Bernoulli random variables.

Here is an alternate statement of the measure-preserving property that shall be used repeatedly.

Lemma F.2 A measurable transformation T on $(\Omega,\mathcal{F},I\!\!P)$ is measure preserving if and only if

$$E[X] = E[X \circ T] \tag{3}$$

for all integrable random variables X on $(\Omega,\mathcal{F},I\!\!P)$.

Proof. Assume (3) and let $A \in \mathcal{F}$ be arbitrary. Setting $X = \mathbf{1}_A$ in (3) gives $I\!\!P(A) = I\!\!P(T^{-1}(A))$, and hence T is measure-preserving. Conversely, if T is measure-preserving, (3) is certainly true for any simple random variable. But an arbitrary integrable r.r. X can be expressed as an a.s. limit of simple functions all of which are bounded by $|X|$, and so (3) follows for X by the dominated convergence theorem. $\diamond$

F.3 The ergodic theorem

The ergodic theorem is a generalization of the law of large numbers to measure-preserving transformations on a probability space, or equivalently, to stationary processes. The set-up of measure-preserving transformations provides the more convenient and elegant framework for the statement and proof of the theorem, and so we adopt this framework here. The next section translates the results into statements about stationary processes.

Two important concepts must first be introduced—invariance and ergodicity.

Definition Let T be a measure-preserving transformation on $(\Omega,\mathcal{F},I\!\!P)$. An event $A \in \mathcal{F}$ is called *invariant* (with respect to T) if $T^{-1}(A) = A$. The class of all invariant sets is denoted by $\mathcal{I}$.

An event A is *quasi-invariant* if it is a.s. invariant in the sense that there exists an invariant set $\tilde{A}$ such that $I\!\!P(A \triangle \tilde{A}) = 0$. The class of all quasi-invariant set is denoted by $\tilde{\mathcal{I}}$.

A random variable W is invariant (resp., almost invariant) if W is $\mathcal{I}$-measurable (resp., $\tilde{\mathcal{I}}$-measurable).

Lemma F.3 Both $\mathcal{I}$ and $\tilde{\mathcal{I}}$ are σ-algebras.

Proof: Exercise.

The importance of invariant sets for limit theorems can be explained by noting that if X is a random variable on $(\Omega, \mathcal{F}, I\!P)$ then

$$Y(\omega) := \limsup_{n\to\infty} \frac{1}{n} \sum_{1}^{n} X\left(T^k(\omega)\right) \quad \text{is an invariant (extended) random variable.}$$

Indeed, it is easy to see $Y(\omega) = Y(T(\omega))$ and this implies $\mathcal{I}$-measurability by the following observation. For any real z and $A := \{\omega\,;\, Y(\omega) < z\}$,

$$T^{-1}(A) = \{\omega\,;\, Y(T(\omega)) < z\} = \{\omega\,;\, Y(\omega) < z\} = A$$

Thus A is invariant. Since sets of the form A generate $\sigma(Y)$, it follows that $\sigma(Y) \subset \mathcal{I}$. Likewise, $\liminf \frac{1}{n} \sum_1^n X\left(T^k(\omega)\right)$ is invariant. These examples are important in the proof of the ergodic theorem.

The proof of invariance in the last example is typical. It is generalized in the next lemma, whose proof is left as an exercise.

Lemma F.4 A r.v. (or extended r.v.) W is invariant if and only if

$$W(\omega) = W\left(T(\omega)\right) \qquad \text{for all } \omega.$$

Definition A measure preserving transformation T on $(\Omega, \mathcal{F}, I\!P)$ is *ergodic* if the σ-algebra of invariant sets is trivial; that is, for any invariant set A, either $I\!P(A) = 0$ or $I\!P(A) = 1$.

Putting this definition together with Lemma F.4 gives the following basic fact: *Ergodicity means that invariant random variables are almost surely constant.*

In general checking ergodicity can be hard. For the moment we give the following example, which associates ergodicity to i.i.d sequences. At the end of this subsection, we show that T_2 defined in Example 4 is ergodic for $([0,1), \mathcal{B}([0,1)), \lambda)$ if α is irrational; T_1 is also ergodic on this space. Let $I\!F$ be a probability distribution on $I\!R$, and let $I\!F^\infty$ be the corresponding infinite product measure on $I\!R^\infty$. $I\!F^\infty$ is the law of an i.i.d. sequence of random variables with common probability distribution measure $I\!F$. In fact, let X denote the function on $I\!R^\infty$ which returns the first coordinate; for every $x = (x_1, x_2, \ldots) \in I\!R^\infty$, $X(x) = x_1$. Define $X_n(x) = X(T^{n-1}x)$ as above and note that for every positive integer n, $X_n(x) = x_n$ is just projection on coordinate n. Then, on the probability space $(I\!R^\infty, \mathcal{B}(I\!R^\infty), I\!F^\infty)$, the process $\{X_n\}_{n\geq 1}$ is a sequence of i.i.d. random variables with common distribution $I\!F$; it is the so-called *canonical process*. Let T denote the left shift on $I\!R^\infty$. We know that T is measure-preserving for $I\!F^\infty$ (see example 1). *In fact, T is ergodic.* This follows by Kolmogorov's Zero-One Law as we will now show. Let A be a left-shift invariant event. Then for any positive integer n,

$$\begin{aligned} A &= \{x \in I\!R^\infty \mid (x_n, x_{n+1}, \ldots) \in A\} \\ &= \{x \in I\!R^\infty \mid (X_n(x), X_{n+1}(x), \ldots) \in A\} \end{aligned}$$

This means that for every n, A is measurable with respect to the σ-algebra $\sigma\{X_n, X_{n+1}, \ldots\}$ generated by $X_n, X_{n+1}, \ldots$. Hence A is in the tail σ-algebra of $\{X_n\}_{n\geq 1}$. This shows that the invariant σ-algebra is contained in the tail σ-algebra . Since $\{X_n\}$ is an i.i.d. sequence on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \mathbb{F}^\infty)$, Kolmogorov's Zero-One Law implies that the tail σ-algebra is trivial with respect to $\mathbb{F}^\infty$, and hence so also is the invariant σ-algebra .

Now we may state the ergodic theorem for measure-preserving transformations.

Theorem F.2 Let T be a measure-preserving transformation on $(\Omega, \mathcal{F}, \mathbb{P})$. Let X be a r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying $E[|X|] < \infty$. Then there is random variable Z such that

$$\text{(a.s.)} \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X(T^{k-1}\omega) = Z(\omega). \tag{4}$$

The convergence holds in L^1 as well, that is

$$\lim_{n\to\infty} E\left[\left| \frac{1}{n} \sum_{k=1}^{n} X \circ T^{k-1}) - Z \right| \right] = 0. \tag{5}$$

The random variable Z is uniquely (up to a.s. equivalence) defined by the following conditions:

(i) Z is $\mathcal{I}$-measurable.

(ii) For every $A \in \mathcal{I}$, $E[\mathbf{1}_A Z] = E[\mathbf{1}_A X]$.

Remark. The limit Z is called the conditional expectation of X given $\mathcal{I}$. Conditional expectation is a very important topic which we take up in a subsequent chapter.

Here is the pay-off of the ergodic theorem for ergodic transformations.

Corollary F.1 If T is ergodic and $E[|X|] < \infty$,

$$\text{(a.s.)} \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X(T^{k-1}\omega) = E[X], \tag{6}$$

and the convergence also holds in L^1.

The reader may wish to explore the implications of the ergodic theorem in the next subsection before engaging the proof. However, the following explanatory remarks should be studied immediately. First, where do conditions (i) and (ii) on Z come from and why do they uniquely characterize Z, at least up to almost-sure equivalence? Let us first derive these conditions from the previous statements of the ergodic theorem. Define

$$Z(\omega) := \begin{cases} \lim \frac{1}{n} \sum_{k=1}^{n} X(T^{k-1}\omega), & \text{when the limit exists and is finite;} \\ 0, & \text{otherwise.} \end{cases}$$

We have shown previously that the limsup and liminf of the empirical mean are invariant. Therefore, the set where the liminf and limsup are equal and its complement are both invariant. The invariance of Z follows and the first statement of the ergodic theorem says that Z is the almost sure limit of the empirical mean.

Now let A be any invariant set. Then $A = T^{-n+1}A$ for any positive integer n, and thus,

$$E\left[\mathbf{1}_A X \circ T^{n-1}\right] = E\left[\left(\mathbf{1}_A \circ T^{n-1}\right) X \circ T^{n-1}\right] = E\left[\mathbf{1}_A X\right].$$

The last equality above used the measure-preserving property of T à la Lemma F.2. It follows that

$$E\left[\mathbf{1}_A \frac{1}{n} \sum_i^n X \circ T^{k-1}\right] = E\left[\mathbf{1}_A X\right],$$

for every positive integer n. Taking limits as $n \to \infty$ and using the L^1 convergence stated in (5),

$$E\left[\mathbf{1}_A Z\right] = E\left[\mathbf{1}_A X\right],$$

thus proving property (ii) of Z.

Second, why do properties (i) and (ii) uniquely characterize Z? Let $\bar{Z}$ be a second random variable also satisfying (i) and (ii). We show that $I\!P(\bar{Z} = Z) = 1$. To this end, let $A := \{Z > \bar{Z}\}$. The event A is invariant since both Z and $\bar{Z}$ are invariant. Applying property (ii) for both Z and $\bar{Z}$,

$$E\left[\mathbf{1}_A Z\right] = E\left[\mathbf{1}_A x\right] = E\left[\mathbf{1}_A \bar{Z}\right],$$

and hence,

$$E\left[\mathbf{1}_{\{Z>\bar{Z}\}}\left(Z - \bar{Z}\right)\right] \geq 0.$$

It follows that $I\!P\left(Z > \bar{Z}\right) = 0$. Reversing the rôles of Z and $\bar{Z}$ gives also $I\!P\left(\bar{Z} > Z\right) = 0$, and so $I\!P(Z = \bar{Z}) = 1$.

At this juncture, it is natural to move to the proof of the corollary.

Proof of Corollary F.1. Let Z be the limit of the empirical mean process provided by the ergodic theorem. Since T is ergodic and Z is invariant, there is a constant c such that $Z = c$, almost surely. Using propery (ii) of Z in the ergodic theorem with $A = \Omega$,

$$c = E[Z] = E[X],$$

and therefore $Z = E[X]$ almost surely. $\diamond$

Before launching into the proof of the ergodic theorem, let us compute the limit Z in the next simplest case beyond ergodic transformations, namely when the invariant σ-algebra $\mathcal{I}$ is just $\{A_1, A_2, \emptyset, \Omega\}$, where A_1 and A_2 constitute a disjoint partition of Ω and both $I\!P(A_1)$ and $I\!P(A_2)$ are strictly positive. In this case, if $\omega \in A_1$, $T^k\omega$ remains in A_1 for all positive integers k, and similarly for A_2. Thus if $\omega \in A_1$, the empirical mean $(1/n)\sum_1^n X(T^{k-1})$ will average X only over the set A_1, and we expect that this average should be an expectation independent of ω in A_1. It is easy to prove this. Consider the reduced probability space $(A_1, \mathcal{F}_{A_1}, I\!P(\cdot/A))$, where $\mathcal{F}_{A_1}$ is the ensemble of events in $\mathcal{F}$ contained in A_1 and $I\!P(\cdot/A_1)$ is the conditional measure defined by

$$I\!P\left(B/A_1\right) = \frac{I\!P(B)}{I\!P(A_1)}, \qquad B \in \mathcal{F}_{A_1}.$$

This is just the measure $I\!P$ restricted to $\mathcal{F}_{A_1}$ and normalized by $I\!P(A_1)$ so that it is a probability measure. Because there are now several probability measures around, it is necessary to distinguish between them carefully when writing expectations. We shall use $E[W]$ to express expectations with respect to the original probability measure $I\!P$; thus, $E[W] = \int_\Omega X\, dI\!P(\omega)$. On the other hand, expectations with respect to $I\!P(\cdot/A_1)$ are denoted by $E[W/A_1]$; thus

$$E[W/A_1] = \int_{A_1} W\, dI\!P(\omega/A_1) = \frac{\int_{A_1} W\, dI\!P(\omega)}{I\!P(A_1)} = \frac{E\left[W\mathbf{1}_{A_1}\right]}{I\!P(A_1)}.$$

With this, we are ready to compute Z on A_1. Since T is measure-preserving for $I\!P$ and A_1 is invariant, T is measure-preserving for $(A_1, \mathcal{F}_{A_1}, I\!P(\cdot/A_1))$ also. Since A_1 does not contain any proper, invariant subevents, T must also be ergodic for $(A_1, \mathcal{F}_{A_1}, I\!P(\cdot/A_1))$, and thus by Corollary F.1, being careful to compute the expectation with respect to $I\!P(\cdot/A_1)$

(7)
$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X(T^{k-1}\omega) = E\left[X/A_1\right] = \frac{E\left[W\mathbf{1}_{A_1}\right]}{I\!P(A_1)} \qquad I\!P(\cdot/A_1)\text{–a.s.}$$

Since $I\!P(B/A_1) = 0$ and $B \subset A_1$ imply that $I\!P(B) = 0$, (7) holds $I\!P$-almost surely on A_1. The analogous calculation may be made on A_2, and, combining the results on the whole of Ω,

$$Z(\omega) = E\left[X/A_1\right]\mathbf{1}_{A_1}(\omega) + E\left[X/A_2\right]\mathbf{1}_{A_2}(\omega) \quad I\!P\text{–a.s.}$$

Note that we could also have obtained this result from the characterization (i) and (ii) of Z. We will see this sort of computation in the theory of conditional expectation.

The proof of the ergodic theorem relies on the ergodic maximal inequality, which we state and prove next. The proof of the maximal inequality is simple, but the result is deep, as we shall see. Throughout, X is a fixed random variable with finite expected value. Set $S_0 = 0$, and for positive integers n, introduce the notations,

$$S_n = \sum_{i}^{n} X \circ T^{k-1}, \qquad M_n = \sup_{1\le k\le n} S_k, \qquad M = \sup_{1\le k<\infty} S_k.$$

The maximal lemma exploits the additive structure of the process $\{S_n\}$, as expressed in the following identity, whose proof is an easy calculation:

(8)
$$S_{n+1}(\omega) = X(\omega) + S_n\left(T\omega\right), \quad n \ge 0.$$

It follows then that

(9)
$$M_{n+1}(\omega) \le X(\omega) + \max_{0\le k\le n} S_k(T\omega) = X(\omega) + \max\{0, M_n(T\omega)\}$$

This is the principal observation.

Lemma F.5 (The Maximal Inequality) For any invariant or almost invariant event A,

$$E\left[\mathbf{1}_A \mathbf{1}_{\{M>0\}} X\right] \geq 0. \tag{10}$$

Proof. Since A is invariant $\mathbf{1}_A = \mathbf{1}_A \circ T$. Inequality (9) implies

$$E\left[\mathbf{1}_A \mathbf{1}_{\{M_n>0\}} X\right] \geq E\left[\mathbf{1}_A \mathbf{1}_{\{M_n>0\}} M_{n+1}\right] - E\left[\mathbf{1}_A \mathbf{1}_{\{M_n>0\}} \max\{0, M_n \circ T\}\right]. \tag{11}$$

However, $\max\{0, M_n \circ T\} = \mathbf{1}_{\{M_n \circ T>0\}} M_n \circ T$, and so

$$\begin{aligned} E\left[\mathbf{1}_A \mathbf{1}_{\{M_n>0\}} \max\{0, M_n \circ T\}\right] &\leq E\left[\mathbf{1}_A \max\{0, M_n \circ T\}\right] \\ &= E\left[\mathbf{1}_A \circ T \mathbf{1}_{\{M_n \circ T>0\}} M_n \circ T\right] \\ &= E\left[\mathbf{1}_A \mathbf{1}_{\{M_n>0\}} M_n\right], \end{aligned}$$

where the last step uses Lemma F.2. Using this in (11) gives

$$E\left[\mathbf{1}_A \mathbf{1}_{\{M_n>0\}} X\right] \geq E\left[\mathbf{1}_A \mathbf{1}_{\{M_n>0\}} \left(M_{n+1} - M_n\right)\right] \geq 0. \tag{12}$$

Notice that the sequence of random variables $\mathbf{1}_{\{M_n>0\}}$, $n \geq 1$, is increasing and converges everywhere to $\mathbf{1}_{\{M>0\}}$. Thus inequality (10) follows from (12) by letting $n \to \infty$ and using dominated convergence. $\diamond$

The remainder of this subsection is devoted to the proof of the ergodic theorem. For every real number b, define the invariant events

$$U_b := \left\{\limsup \frac{S_n}{n} > b\right\} \quad \text{and} \quad L_b := \left\{\liminf \frac{S_n}{n} < b\right\}.$$

We will show the following. First, for any $a < b$,

$$I\!\!P\left(\liminf \frac{S_n}{n} < a < b < \limsup \frac{S_n}{n}\right) = I\!\!P\left(L_a \cap U_b\right) = 0 \tag{13}$$

Second,

$$I\!\!P\left(\liminf \frac{S_n}{n} = -\infty\right) = 0 \quad \text{and} I\!\!P\left(\limsup \frac{S_n}{n} = \infty\right) = 0. \tag{14}$$

Since the event that S_n/n does *not* converge to a finite limit is contained in

$$\left\{\liminf \frac{S_n}{n} = -\infty\right\} \cup \left\{\limsup \frac{S_n}{n} = \infty\right\} \cup \left[\bigcup_{a<b,\ a,b\in Q} L_a \cap U_b\right],$$

where Q denotes the rational numbers, we will have shown that S_n/n converges with probability one. Set

$$Z(\omega) := \begin{cases} \lim S_n(\omega)/n & \text{, if the limit exists;} \\ 0, & \text{otherwise} \end{cases}$$

Then Z is invariant and $S_n/n \to Z$, $I\!\!P$-a.s., thus proving (4) of the ergodic theorem.

Facts (13) and (14) are proved by applying the maximal lemma to $X - b$ and $a - X$. The main step is formulated in the next result.

Lemma F.6 For any invariant set B,

$$E\left[\mathbf{1}_B \mathbf{1}_{U_b}(X - b)\right] \geq 0 \qquad \text{and,} \qquad E\left[\mathbf{1}_B \mathbf{1}_{L_a}(X - a)\right] \leq 0.$$

Proof. Let B be invariant. Let $\bar{X} := X - b$, $\bar{S}_n = S_n - nb$, and $\bar{M} = \sup_n \bar{S}_n$. Since U_b is invariant, the ergodic theorem implies that

$$E\left[\mathbf{1}_B \mathbf{1}_{U_b} \mathbf{1}_{\{\bar{M}>0\}}(X - b)\right] \geq 0. \tag{15}$$

However, since $\limsup \bar{S}_n/n = \limsup S_n/n - b$, it follows that $U_b = \{\limsup \bar{S}_n/n > 0\} \subset \{\bar{M} > 0\}$, and hence that $\mathbf{1}_{U_b}\mathbf{1}_{\{\bar{M}>0\}} = \mathbf{1}_{U_b}$. Thus (15) reduces precisely to the first inequality of the lemma. The second inequality is proved by applying the same argument to $a - X$ instead of $X - b$. $\diamond$

Proof of (14). Notice that as b increases, the events L_b decrease down to the event $\{\limsup S_n/n = \infty\}$ and so $I\!\!P\left(\limsup S_n/n = \infty\right) = \lim_{b\to\infty} I\!\!P(L_b)$. Now apply Lemma F.6 with $B = \Omega$ and rearrange terms. The result is:

$$bI\!\!P\left(L_b\right) = bE\left[\mathbf{1}_{L_b}\right] \leq E\left[\mathbf{1}_{L_b} X\right] \leq E\left[|X|\right].$$

Hence, $I\!\!P(L_b) \leq E[|X|]/b \to 0$ as $b \to \infty$, which completes the proof of the first equality in (14). The second equality is proved by a similar argument.

Proof of (13). Let $a < b$. Apply Lemma F.6 using the invariant set $B = U_a \cap L_b$ and rearrange terms. The result is

$$bI\!\!P\left(U_a \cap L_b\right) \leq E\left[\mathbf{1}_{U_a \cap L_b} X\right] \leq aI\!\!P\left(U_a \cap L_b\right).$$

Since $a < b$, this can only be true if $I\!\!P\left(U_a \cap L_b\right) = 0$.

The proof that the limit of S_n/n exists almost surely is now complete. It remains to prove that the S_n/n converges to Z in L^1 norm. The fact that Z satisfies the conditions (i) and (ii) then follows, as was shown in the discussion above.

Suppose first that $|X|$ is bounded, say by a constant K. Then $|S_n/n|$ and hence Z are also bounded by K and it follows by dominated convergence that $E[|S_n/n - Z|] \to 0$ as $n \to \infty$. This establishes L^1 convergence for the bounded case.

Now assume only that $E[|X|] < \infty$, and let $Z = \lim_n S_N/n$. For any $K > 0$, let $X^K := X\mathbf{1}_{\{|X| \leq K\}}$, $S_n^K = \sum_1^n X^K \circ T^{k-1}$, and

$$Z^K := \lim_n \frac{1}{n} \sum_1^n X^K \circ T^{k-1},$$

which we know exists. By Fatou's lemma, the triangle inequality, and Lemma F.2,

$$E\left[|Z - Z^K|\right] \leq \liminf_n E\left[\left| \frac{1}{n} \sum_1^n (X - X^K) \circ T^{k-1} \right|\right] \leq E\left[|X - X^K|\right].$$

Therefore

$$E\left[\left| \frac{1}{n} S_n - Z \right|\right] \leq E\left[\left| \frac{1}{n}\left(S_n - S_n^K\right) \right|\right] + E\left[\left| \frac{1}{n} S_n^K - Z^K \right|\right] + E\left[|Z - Z^K|\right]$$
$$\leq 2E\left[|X - X^K|\right] + E\left[\left| \frac{1}{n} S_n^K - Z^K \right|\right]$$

Take $n \to \infty$; then the last term goes to 0 since X^K is bounded, and thus

$$\limsup_n E\left[\left| \frac{1}{n} S_n - Z \right|\right] \leq 2E\left[|X - X^K|\right].$$

But this is true for any positive K. Thus take $K \to \infty$ and use dominated convergence to obtain

$$\limsup_n E\left[\left| \frac{1}{n} S_n - Z \right|\right] = 0.$$

The proof of the ergodic theorem is complete. $\diamond$

Before going on to considering the ergodic theorem for general stationary processes, we prove that the transformation T_2 on $([0,1), \mathcal{B}([0,1)), \lambda)$ if α is irrational. The standard slick proof is by use of Fourier series. Let A be a Borel subset of $[0,1)$. The indicator function $\mathbf{1}_A$ has a Fourier series representation

$$\mathbf{1}_A(x) = \sum_{-\infty}^{\infty} c_n e^{i2n\pi x},$$

where the c_n are complex numbers such that $\sum_{-\infty}^{\infty} c_n^2 < \infty$ and the Fourier series converges in the $L^2(\lambda$ sense. Clearly,

$$\left(\mathbf{1}_A \circ T_2\right)(x) = \sum_{-\infty}^{\infty} c_n e^{i2n\pi T_2 x} = \sum_{-\infty}^{\infty} c_n e^{i2n\pi(x+\alpha)}.$$

If A is invariant, $\mathbf{1}_A = \mathbf{1}_A \circ T_2$, and hence the two Fourier series must be the same. This means that $c_n = c_n e^{i2n\pi\alpha}$, for every integer n. But if α is irrational $e^{i2n\pi\alpha} \neq 1$ as long as $n \neq 0$. Hence $c_n = 0$ if $n \neq 0$. It follows that

$$\mathbf{1}_A(x) = c_0 \quad \text{Lebesgue almost everywhere.}$$

Either $c_0 = 0$, in which case $\lambda(A) = 0$, or $c_0 = 1$, in which case $\lambda(A) = 1$. This proves that the invariant σ-algebra is trivial, and so proves the ergodicity of T_2. From this argument one also sees that if α is rational, T_2 is *not* ergodic.

F.4 The ergodic theorem for stationary processes

As stated, the Ergodic Theorem F.2 is a limit theorem for a special class of stationary processes, namely those built by evaluating a random variable along the orbit of a measure-preserving transformation. This can be used to state an ergodic theorem for general stationary processes. Here is the idea. Let $W = \{X_n; n \geq 1\}$ be a stationary process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We will construct the canonical process associated to W in the language of measure-preserving transformations. Let $\mathbb{F}_W$ be the law of the process on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$. Then we know that T_ℓ is a measure preserving transformation on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \mathbb{F}_W)$. Now, define on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ the random variable Y which projects onto the first coordinate of $\mathbb{R}^\infty$; $Y(x) = x_1$ for $x = (x_1, x_2, \ldots) \in \mathbb{R}^\infty$. Then the process

$$Y_n(x) := Y(T^{n-1}x), \quad n \geq 1,$$

is the canonical process for W, because for every n, $Y_n(x) = x_n$ is the projection on coordinate n. To repeat in different words, $\{Y_n\}$, as a process on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \mathbb{F}_W)$, is a stationary process with the same law as $W = \{X_n\}$; that is, for every event U in $\mathcal{B}(\mathbb{R}^\infty)$,

$$\mathbb{P}\left((X_1, X_2, \ldots) \in U\right) = \mathbb{F}_W\left(\{x = (x_1, x_2, \ldots) \in U\}\right) = \mathbb{F}_W\left(\{x; (Y_1(x), Y_2(x), \ldots) \in U\}\right).$$

$\{Y_n\}$ on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \mathbb{F}_W)$ with the left shift T_ℓ is in a form for which Theorem F.2 directly applies. We shall derive the ergodic theorem for stationary process by using Theorem F.2 on the canonical process.

Before starting, it is useful to note explicitly the following. Because $\{X_n\}$ and its canonical process have the same law,

$$(16) \qquad E\left[\Phi\left((X_1, X_2, \ldots)\right)\right] = \int_{\mathbb{R}^\infty} \Phi(x)\, d\mathbb{F}_W(x) = \int_{\mathbb{R}^\infty} \Phi\left(Y_1(x), Y_2(x), \ldots\right)\, d\mathbb{F}_W(x).$$

for any $\mathcal{B}(\mathbb{R}^\infty)$-measurable Φ, as long as either side is well-defined.

Before stating the theorems, it is useful to extend the notions of invariant set and ergodicity to general stationary process. This is done, of course, through the canonical process. An event $A \in \mathcal{F}$ is invariant for the stationary process $\{X_n\}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ if there exists a subset $B \in \mathcal{B}(\mathbb{R}^\infty)$ such that

$$A = \{\omega\ ;\ (X_n(\omega), X_{n+1}(\omega), \ldots) \in B\} = \left\{\omega\ ;\ T^{n-1}\left(X_1(\omega), X_2(\omega)\ \ldots\right) \in B\right\},$$

for all positive integers n. From the second representation of A, one sees that any such B must be at least quasi-invariant with respect to the left shift on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \mathbb{F}_W)$. If B is left-shift invariant, then A is certainly invariant. The ensemble of invariant events for $\{X_n\}$ is a σ-algebra and is denoted by $\mathcal{I}(\{X_n\})$. We define the ensemble of quasi-invariant events for $\{X_N\}$ to be those events which are $\mathbb{P}$-a.s. equal to invariant events. It should be noted that for any invariant event A,

$$E\left[\mathbf{1}_A X_n\right] = E\left[\mathbf{1}_A X_m\right] \quad \text{for all positive integers } n \text{ and } m. \tag{17}$$

This can be proved directly using stationarity. It follows also by computing both sides using (16) and then applying Lemma F.4.

Now ergodicity is easy to define in general. A stationary process is ergodic if its invariant σ-algebra is trivial. It is clear from this definition that $\{X_n\}$ is ergodic if and only if the left shift, T_ℓ, is ergodic for $\mathbb{F}_W$.

Theorem F.3 Let $\{X_n\}$ be a stationary process such that $E[|X_1|] < \infty$. Then there is an $\mathcal{I}(\{X_n\})$-measurable random variable Z such that $\frac{1}{n}\sum_1^n X_k$ converges to Z almost surely and in $L^1(\mathbb{P})$. Moreover, for every $k \geq 1$, Z is the conditional expectation of X_n given $\mathcal{I}(\{X_n\})$; that is, for any $k \geq 1$, Z is uniquely characterized by the conditions

(i) Z is $\mathcal{I}(\{X_n\})$-measurable.

(ii) For every $A \in \mathcal{I}(\{X_n\})$, $E\left[\mathbf{1}_A Z\right] = E\left[\mathbf{1}_A X_k\right]$.

Finally, if $\{X_n\}$ is ergodic, then $Z = E[X_k]$ (which is indpendent of k, by stationarity).

Proof: Apply Theorem F.2 to the canonical process $\{Y_n\}$ of $\{X_n\}$, as defined above. This says there exists a T_ℓ-invariant Ξ such that if

$$C := \left\{ x \in \mathbb{R}^\infty \mid \lim_n \frac{1}{n}\sum_1^n Y_k(x) = \Xi(x) \right\}$$

then

$$\mathbb{F}_W(C) = 1.$$

Now set $Z(\omega) = \Xi\left(X_1(\omega), X_2(\omega), \ldots\right)$. Then Z is invariant for $\{X_n\}$ because Ξ is left-shift invariant. Moreover,

$$\mathbb{P}\left(\left\{\omega\,;\, \lim_n \frac{1}{n}\sum_1^n X_k(\omega) = Z(\omega)\right\}\right) = \mathbb{P}\left(\{\omega\,;\, (X_1(\omega), X_2(\omega), \ldots) \in C\}\right) = F_W(C) = 1.$$

This proves the a.s. existence of the limit of the empirical mean.

Because of (16)

$$E\left[\,\left|\,\frac{1}{n}\sum_1^n X_k(\omega) - Z\,\right|\,\right] = \int_{\mathbb{R}^\infty} \left|\,\sum_1^n Y_k(x) - \Xi(x)\,\right|\, d\mathbb{F}_W(x).$$

But the right hand side converges to 0 as $n \to \infty$ by Theorem F.2 applied to $\{Y_n\}$ on $(I\!R^\infty, \mathcal{B}(I\!R^\infty), I\!F_W)$, thus proving the L^1-convergence of the empirical mean process of $\{X_n\}$ to Z.

The characterization of Z by properties (i) and (ii) is proved with the help of (17) by the same argument used for Theorem F.2; see the previous subsection.

When $\{X_n\}$ is ergodic, any invariant random variable for $\{X_n\}$ is almost-surely constant, and so $Z = E[X]$. $\diamond$

Example 5. If $\{X_n\}$ is an i.i.d. sequence, then it is ergodic. We can argue this in two ways, either directly or via the canonical process. To argue directly, note that the invariant σ-algebra is contained in the tail σ-algebra, which is trivial by Kolmogorov's Zero-One Law. Or, equivalently, the law of an i.i.d. sequence is a product measure, and we showed in the previous section that the left shift is ergodic for any product measure. By applying Theorem F.3 to i.i.d. sequences, we immediately obtain the strong law of large numbers.

We wish now to point out that the ergodic theorem gives a lot more than the straightforward strong law in the case of ergodic processes, even in the i.i.d. case. Recall from Example 2, that if $\{X_n\}$ is stationary, then, given a $\mathcal{B}(I\!R^\infty)$-measurable function Φ on $I\!R^\infty$, the process

$$Y_n := \phi(X_n, X_{n+1}, \ldots), \qquad n \geq 1, \tag{17}$$

is also stationary. In fact, ergodicity also carries over.

Exercise. Show that if $\{X_n\}$ is ergodic, then so is $\{Y_n\}$.

Thus if $\{X_n\}$ is ergodic, in particular, if it is i.i.d., and if $E[X_1] < \infty$, not only does the strong law hold for $\{X_n\}$, it holds for any process $\{Y_n\}$ of the form (17), so long as Y_1 (and hence any Y_n) is integrable.

The observation we have made can be extended into a Theorem.

Theorem F.4 Let $\{X_n\}$ be a stationary process. The following are equivalent:

(a) $\{X_n\}$ is ergodic.

(b) For any $\mathcal{B}(I\!R^\infty)$-measurable function Φ on $I\!R^\infty$,

$$\text{(a.s.)} \lim \frac{1}{n} \sum_{1}^{n} \phi(X_k, X_{k+1}, \ldots) = E\left[\phi(X_1, X_2, \ldots)\right].$$

if the expectation exists and is finite.

(c) For any positive integer N and any bounded, Borel function ϕ on $I\!R^N$,

$$\text{(a.s.)} \lim \frac{1}{n} \sum_{1}^{n} \phi(X_k, X_{k+1}, \ldots, X_{k-1+N}) = E\left[\phi(X_1, X_2, \ldots, X_N)\right].$$

Proof: We have shown already that (a) implies (b). Clearly (b) implies (c).

We show next that (b) implies (a). In fact, knowing that (b) is true only for indicators $\mathbf{1}_B$, $B \in \mathcal{B}(\mathbb{R}^\infty)$ is enough to prove (a). Indeed, suppose (b) holds for indicators. Let A be an invariant set for $\{X_n\}$. Then there is $B \in \mathcal{B}(\mathbb{R}^\infty)$ such that,

$$\mathbf{1}_A(\omega) = \mathbf{1}_B\left(X_k(\omega), X_{k+1}(\omega), \ldots\right)$$

for every positive integer k. It follows that

$$\sum_1^n \mathbf{1}_B\left(X_k(\omega), X_{k+1}(\omega), \ldots\right) = \mathbf{1}_A(\omega) \quad \text{for every } \omega,$$

and hence from (b) that $\mathbf{1}_A = E[\mathbf{1}_A] = \mathbb{P}(A)$ almost surely. But this can only be true if either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. Thus, $\{X_n\}$ is ergodic.

It remains only to prove that (c) implies (b), and for this it suffices to prove that (c) implies (b) holds only for indicator functions (since then (a) holds, which implies (b) for arbitrary ϕ). Thus let $B \in \mathcal{B}(\mathbb{R}^\infty)$, and take an arbitrary $\epsilon > 0$. Since the algebra of cylinder sets of $\mathbb{R}^\infty$ generates $\mathcal{B}(\mathbb{R}^\infty)$, Proposition 4.3 of Chapter 1 implies that there exists a cylinder set C such that $\mathbb{P}(C \triangle B) = E[|\mathbf{1}_C - \mathbf{1}_B|] < \epsilon$. Let

$$Z = (\text{a.s.}) \lim \frac{1}{n} \sum_1^n \mathbf{1}_B(X_k, X_{k+1}, \ldots),$$

which exists by the ergodic theorem, and let

$$Z_C = (\text{a.s.}) \lim \frac{1}{n} \sum_1^n \mathbf{1}_B(X_k, X_{k+1}, \ldots).$$

Assuming (c), $Z_C = E[\mathbf{1}_C]$, $\mathbb{P}$-almost surely. But, by Fatou's Lemma, the triangle inequality, and stationarity,

$$E\left[|Z - Z_C|\right] \le \liminf E\left[\left|\ \frac{1}{n} \sum_1^n \left(\mathbf{1}_B - \mathbf{1}_C\right)(X_k, \ldots)\ \right|\right] \le E\left[\left|\ \mathbf{1}_B - \mathbf{1}_C\ \right|\right].$$

As $\epsilon \to 0$, $Z_C \to \mathbb{P}(B)$, and hence $Z = \mathbb{P}(B) = E\left[\mathbf{1}_B\right]$, as we needed to show. $\diamond$