8 Ergodic Theorem

We have seen in Corollary 7.5.3 that a recurrent and irreducible Markov chain has ergodicity, in this chapter, we will going to discuss the ergodic theory in a more general setting. This can be understood as the strong law of large numbers discussed in Theorem 4.3.1 where we extend the result to sequences of non i.i.d. random variables.

8.1 Measure-Preserving Transformations and Properties

Definition 8.1.1 : A random process $(X_n)_{n\geqslant 0}$ is a *stationary sequence* (平穩序列) or a *stationary process* (平穩過程) if for all non-negative integer $k \geqslant 0$, the random processes $(X_n)_{n\geqslant 0}$ and $(X_{n+k})_{n\geqslant 0}$ have the same distribution.

Example 8.1.2 : Below are a few examples of stationary processes.

(1) Any i.i.d. sequence $(X_n)_{n\geqslant 0}$ of random variables is a stationary process.

(2) Let $(X_n)_{n\geqslant 0}$ be a Markov chain with transition matrix Q. If it possesses a stationary probability measure π, then when $X_0 \sim \pi$, the random process $(X_n)_{n\geqslant 0}$ is stationary, see Remark 7.4.3.

Then, we introduce the notion of *measure-preserving transformations*, generalizing the notion of stationary processes.

Definition 8.1.3 : Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A measurable function $\varphi : \Omega \longrightarrow \Omega$ is said to be a *measure-preserving transformation* (測度守恆變換) on $(\Omega, \mathcal{F}, \mathbb{P})$ if $\mathbb{P}(\varphi^{-1}(A)) = \mathbb{P}(A)$ for all $A \in \mathcal{F}$. In this case, we also say that $\mathbb{P}$ is an *invariant measure* (不變測度) for φ.

More generally speaking, we do not need to assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space in Definition 8.1.3, it is enough to have a σ-finite measured space. Since our class is about probability theory, we will only focus on the case of probability spaces in what follows.

Example 8.1.4 : Below are a few examples of measure-preserving measures.

(1) Consider the measured space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$, then for any $y \in \mathbb{R}$, the translation function

$$\theta_y : \quad x \quad \mapsto \quad x + y$$

is a measure-preserving transformation.

(2) Consider the unit circle $\mathbb{S}^1 = \{x \in \mathbb{C} : |x| = 1\}$ and the uniform probability distribution μ

defined above. For any $\beta \in \mathbb{R}$, we can define the rotation operator

$$\theta_\beta : e^{2\pi \mathrm{i}\,\alpha} \mapsto e^{2\pi \mathrm{i}(\alpha+\beta)}.$$

which is a measure-preserving transformation on $(\mathbb{S}^1, \mathcal{B}(\mathbb{S}^1), \mu)$.

(3) Consider the probability space $([0,1], \mathcal{B}([0,1]), \mathbb{P})$, where $\mathbb{P}$ is the Lebesgue measure on $[0,1]$. Then,

$$\theta(x) := \begin{cases} 2x & \text{if } x \in [0, \frac{1}{2}), \\ 2 - 2x & \text{if } x \in [\frac{1}{2}, 1], \end{cases}$$

is a measure-preserving transformation.

The below proposition shows that, the notion of a stationary sequence defined previously is indeed a special case of invariant measure with respect to a properly chosen measure-preserving transformation.

Proposition 8.1.5 : *Let $E = \mathbb{R}$ or $\mathbb{R}^d$ be the state space, $\Omega = E^{\mathbb{N}_0}$ be the sample space, $\mathcal{F} = \mathcal{B}(E)^{\otimes \mathbb{N}_0}$ be the smallest σ-algebra making all the coordinate functions measurable, $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$. If $\omega \sim \mathbb{P}$ is a stationary process, then the* shift operator **(推移算子)** *$\theta := \theta_1$ defined as $\theta((\omega_n)_{n \geqslant 0}) = (\omega_{n+1})_{n \geqslant 0}$ is a measure-preserving transformation on $(\Omega, \mathcal{F}, \mathbb{P})$.*

Proof : First, let us check that θ is measurable. For any cylinder event A in $\mathcal{F}$, we have $\theta^{-1}(A) = E \times A$ which is also a cylinder event. At the same time, we know that the cylinder events generate $\mathcal{F}$, so θ is measurable. Then, due to the stationarity of $\mathbb{P}$, we have $\mathbb{P}(A) = \mathbb{P}(\theta^{-1}(A))$ for all cylinder event A, so the measures $\mathbb{P}$ and $\mathbb{P} \circ \theta^{-1}$ take the same value on the collection of cylinder events which generates $\mathcal{F}$. We are done with the proof. $\square$

Proposition 8.1.6 : *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If φ is a measure-preserving transformation on $(\Omega, \mathcal{F}, \mathbb{P})$, then for any random variable $X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (G, \mathcal{G})$ taking values in a measurable space $(G, \mathcal{G})$, we can define $X_n(\omega) = X(\varphi^n(\omega))$ for any non-negative integer $n \geqslant 0$, and we have that $(X_n)_{n \geqslant 0}$ is a stationary process.*

Remark 8.1.7 : In Proposition 8.1.6, if we choose $(\Omega, \mathcal{F}) = (E^{\mathbb{N}_0}, \mathcal{B}(E)^{\otimes \mathbb{N}_0})$ and $(G, \mathcal{G}) = (E, \mathcal{B}(E))$ and assume that the shift operator θ is measure-preserving, then by taking $X = X_0$, we get the converse of Proposition 8.1.5. This tells us that a stationary sequence is indeed a special case of invariant measure with respect to a properly chosen measure-preserving transformation.

Proof : Given $n \geqslant 0$ and let $B \in \mathcal{G}^{\otimes(n+1)}$ and $A = \{\omega \in \Omega : (X_0(\omega), \ldots, X_n(\omega)) \in B\}$. Then,

$$\begin{aligned} \mathbb{P}((X_k(\omega), \ldots, X_{k+n}(\omega)) \in B) &= \mathbb{P}(\omega : \varphi^k(\omega) \in A) \\ &= \mathbb{P}(\omega \in A) \\ &= \mathbb{P}((X_0(\omega), \ldots, X_n(\omega)) \in B). \end{aligned}$$

$\square$

Definition 8.1.8 : Let φ be a measure-preserving transformation on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The *invariant set* (不變集合) of φ is defined as

$$\mathcal{I} = \mathcal{I}_\varphi := \{A \in \mathcal{F} : \varphi^{-1}(A) = A\}. \tag{8.1}$$

We can check that $\mathcal{I}_\varphi$ is a σ-algebra, so it is also called *invariant σ-algebra* (不變 σ 代數) of φ. Additionally, if $\mathcal{I}_\varphi$ is a trivial σ-algebra, we say that φ is *ergodic* (遍歷性) .

Question 8.1.9: Check that the invariant set $\mathcal{I}$ defined in Definition 8.1.8 is indeed a σ-algebra.

Remark 8.1.10 : If φ is not ergodic, then we can find $A \in \mathcal{I}$ such that $\varphi(A) = A$, $\varphi(A^c) = A^c$ and $\mathbb{P}(A) > 0$, $\mathbb{P}(A^c) > 0$. In other words, φ is reducible.

Question 8.1.11: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which we consider a measure-preserving transformation φ.

(1) A random variable $X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is $\mathcal{I}$-measurable if and only if $X \circ \varphi = X$.

(2) φ is ergodic if and only if all the $\mathcal{I}$-measurable random variables $X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ are $\mathbb{P}$-a.s. constant.

For a given sequence $(X_n)_{n\geqslant 0}$ of random variables, we recall the *asymptotic σ-algebra* (漸進 σ 代數) defined in Definition 4.2.1, which we also call *tail σ-algebra* (尾端 σ 代數) , detnoed $\mathcal{T}$,

$$\mathcal{T} = \bigcap_{n=0}^{\infty} \mathcal{F}^n, \qquad \mathcal{F}^n = \sigma(X_n, X_{n+1}, \dots), \quad \forall n \geqslant 0.$$

Proposition 8.1.12 : *Let $E = \mathbb{R}$ or $\mathbb{R}^d$ be the state space and $\Omega = E^{\mathbb{N}_0}$ be the sample space. Take $\mathcal{F} = \mathcal{B}(E)^{\otimes \mathbb{N}_0}$ and assume that the shift operator θ is measure-preserving. Then, $\mathcal{I} \subseteq \mathcal{T}$.*

Remark 8.1.13 : Consider the case where $(X_n)_{n\geqslant 0}$ is an i.i.d. sequence. By Kolmogorov's 0-1 law (Theorem 4.2.2), we know that $\mathcal{T}$ is a trivial σ-algebra, so $\mathcal{I}$ is also trivial. In other words, for any i.i.d. sequence of random variables, the shift operator is ergodic.

Proof : Given an invariant set $A \in \mathcal{I}$, we have

$$A = \theta^{-1}(A) = \{\omega : \theta(\omega) \in A\} \in \mathcal{F}^1.$$

If we keep iterating, we obtain $A \in \mathcal{F}^n$ for all $n \geqslant 0$, that is $A \in \mathcal{T}$. □

Question 8.1.14: Please find an example with $\mathcal{I} \subsetneq \mathcal{T}$.

Definition 8.1.15 : We keep the notations in Proposition 8.1.5. If $\omega \sim \mathbb{P}$ is a stationary process and the shift operator θ is ergodic, we say that $\omega = (\omega_n)_{n\geqslant 0}$ is an *ergodic process* (遍歷過程) .

Remark 8.1.16 : We note that, according to Definition 8.1.15, for a random process to be called ergodic, it needs first to be stationary. If we come back to the setting of Markov chains, Remark 7.4.3 says that if a Markov chain evolves from the initial state given by its stationary probability measure (assuming existence), then we do have a stationary process. However, for a recurrent and irreducible Markov chain, even without the assumption of stationarity, for any given initial state $\mathbb{P}$ and bounded (or non-negative) function $f : E \longrightarrow \mathbb{R}$, the ergodic theorem (Corollary 7.5.3)

$$\frac{1}{n}\sum_{k=0}^{n} f(X_k) \longrightarrow \mu(f), \qquad \mathbb{P}\text{-a.s.}$$

holds and sometimes we also say such a process is *ergodic.*

Now, let us discuss the ergodicity of Markov chains in the sense of Definition 8.1.15.

Proposition 8.1.17 : *Given a Markov chain $(X_n)_{n\geqslant 0}$ defined on a countable state space E and assume that it has an invariant probability measure π such that $\pi(x) > 0$ for all $x \in E$. Then, the Markov chain $(X_n)_{n\geqslant 0}$ is irreducible if and only if it is ergodic.*

Proof : First we note that, by Example 8.1.2 (2) and Proposition 8.1.5, we know that the shift operator θ is a measure-preserving transformation for $\mathbb{P}_\pi$.

We can show that all the states are recurrent by adapting the proof of Proposition 7.4.15. Moreover, Theorem 7.3.6 says that $E = \sqcup R_i$ where R_i's are disjoint irreducible sets. Besides, we also know that if $X_0 \in R_i$, then $X_n \in R_i$ for all $n \geqslant 1$, implying

$$\{\omega : X_0(\omega) \in R_i\} \in \mathcal{I}, \qquad \forall i.$$

In consequence, this tells us that if a Markov chain is not irreducible, then the shift operator θ is not ergodic.

Conversely, consider $A \in \mathcal{I}_\theta$, then for any iteration of the shift operator $\theta_n = \theta^n$, we have $\mathbb{1}_A \circ \theta_n = \mathbb{1}_A$. Hence, if we define $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ for $n \geqslant 0$, the simple Markov property gives

$$\mathbb{E}_\pi[\mathbb{1}_A \,|\, \mathcal{F}_n] = \mathbb{E}_\pi[\mathbb{1}_A \circ \theta_n \,|\, \mathcal{F}_n] = \mathbb{E}_{X_n}[\mathbb{1}_A] = h(X_n),$$

where $h(x) = \mathbb{E}_x[\mathbb{1}_A]$. Since $A \in \mathcal{I}$ and the left side of the above formula is an uniformly integrable martingale, Theorem 6.5.6 implies that it converges almost surely to $\mathbb{E}_\pi[\mathbb{1}_A \,|\, \mathcal{F}_\infty] = \mathbb{1}_A$. Besides, since $(X_n)_{n\geqslant 0}$ is irreducible and recurrent, for any $y \in E$, the right side of the above formula can take the value $h(y)$ infinitely many times, so $h \equiv 0$ or $h \equiv 1$, that is $\mathbb{P}_\pi(A) = 0$ or 1. □

Below is a useful criterion to check whether a measure-preserving transformation is ergodic.

Definition 8.1.18 : Let φ be a measure-preserving transformation on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that φ is *mixing* (混合性) if for all $F, G \in \mathcal{F}$,

$$\mathbb{P}(F \cap \varphi^{-n}(G)) \xrightarrow[n\to\infty]{} \mathbb{P}(F)\,\mathbb{P}(G),$$

Lemma 8.1.19 : *Let φ be a measure-preserving transformation on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If φ is mixing, then φ is also ergodic.*

Proof : Let $F \in \mathcal{I}$. Then, we have

$$\mathbb{P}(F) = \mathbb{P}(F \cap \varphi^{-n}(F)) \longrightarrow \mathbb{P}(F)^2.$$

This gives $\mathbb{P}(F) = \mathbb{P}(F)^2$, that is $\mathbb{P}(F) \in \{0, 1\}$. Hence, φ is ergodic. □

Exercise 8.6 is an example where we use Lemma 8.1.19 to check that a measure-preserving transformation is ergodic. Exercise 8.14 also gives another useful lemma to check the ergodicity.

8.2 Birkhoff's Ergodic Theorem

In this section, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which we consider a measure-preserving transformation φ. Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (G, \mathcal{G})$ be a random variable and define the random process $(X_n)_{n \geqslant 0}$ with $X_n(\omega) = X(\varphi^n(\omega))$.

Theorem 8.2.1 (Birkhoff's ergodic theorem) **:** *Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ be an integrable random variable and φ be a measure-preserving transformation on $(\Omega, \mathcal{F}, \mathbb{P})$. Then, the following convergence holds $\mathbb{P}$-a.s. and in L^1,*

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k = \frac{1}{n}\sum_{k=0}^{n-1} X \circ \varphi^k \xrightarrow[n\to\infty]{} \mathbb{E}[X \,|\, \mathcal{I}].$$

Lemma 8.2.2 (Hopf's maximal ergodic lemma) **:** *If we define*

$$S_0 = 0, \quad S_n = \sum_{k=0}^{n-1} X \circ \varphi^k, \qquad \forall n \geqslant 1,$$
$$M_n = \max_{0 \leqslant k \leqslant n} S_k, \qquad \forall n \geqslant 1,$$

then for all $n \geqslant 1$, we have

$$\mathbb{E}[X \mathbb{1}_{M_n > 0}] \geqslant 0.$$

Proof : Given a positive integer $n \geqslant 1$. First we note that

$$S_{k+1} = X + S_k \circ \varphi, \qquad \forall k \geqslant 0.$$

Thus, we have, for $1 \leqslant k \leqslant n$,

$$S_k = X + S_{k-1} \circ \varphi \leqslant X + M_n \circ \varphi.$$

On the event $\{M_n > 0\}$, we have $M_n = \max_{1 \leqslant k \leqslant n} S_k$ (note that the index k starts from 1 instead of

0). We can multiply the above formula by $\mathbb{1}_{M_n>0}$ and take max on the left side for $1 \leqslant k \leqslant n$ to obtain

$$M_n \mathbb{1}_{M_n>0} \leqslant (X + M_n \circ \varphi)\mathbb{1}_{M_n>0}.$$

Finally, we conclude by writing

$$\begin{aligned}\mathbb{E}[X\mathbb{1}_{M_n>0}] &\geqslant \mathbb{E}[M_n\mathbb{1}_{M_n>0}] - \mathbb{E}[(M_n \circ \varphi)\mathbb{1}_{M_n>0}] \\ &\geqslant \mathbb{E}[M_n] - \mathbb{E}[M_n \circ \varphi] = 0,\end{aligned}$$

where in the inequality in the second line, we use the fact that $M_n \equiv 0$ and $M_n \circ \varphi \geqslant 0$ on $\{M_n > 0\}^c$, and in the last equality, we use the assumption that φ is a measure-preserving transformation. □

Now we are ready to prove the ergodic theorem in Theorem 8.2.1.

Proof : We can assume $\mathbb{E}[X \,|\, \mathcal{I}] = 0$ since without loss of generality, we can replace X with $X - \mathbb{E}[X \,|\, \mathcal{I}]$. Now we want to show

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k \xrightarrow[n\to\infty]{} 0. \tag{8.2}$$

Given $\varepsilon > 0$ and let

$$L_\varepsilon = \Big\{ \limsup_{n\to\infty} \frac{S_n}{n} > \varepsilon \Big\}.$$

To prove Eq. (8.2), it is sufficient to show $\mathbb{P}(L_\varepsilon) = 0$ then replace X with $-X$ by symmetry to conclude, since ε is arbitrarily small.

Define the following notations

$$\begin{aligned}&\forall k \geqslant 0, \qquad X_k^\varepsilon = (X_k - \varepsilon)\mathbb{1}_{L_\varepsilon}, \\ &\forall n \geqslant 0, \qquad S_n^\varepsilon = \sum_{k=0}^{n-1} X_k^\varepsilon, \qquad M_n^\varepsilon = \max_{0\leqslant k\leqslant n} S_n^\varepsilon.\end{aligned}$$

We note that $L_\varepsilon = \varphi^{-1}(L_\varepsilon)$, hence $L_\varepsilon \in \mathcal{I}$ and $(X_k^\varepsilon)_{k\geqslant 0}$ is a stationarity process. Using Lemma 8.2.2, we obtain

$$\forall n \geqslant 0, \qquad \mathbb{E}[X_0^\varepsilon \mathbb{1}_{M_n^\varepsilon>0}] \geqslant 0.$$

Taking the limit $n \to \infty$ in the above formula, we get

$$\mathbb{E}[X_0^\varepsilon \mathbb{1}_{\sup_{n\geqslant 0} S_n^\varepsilon>0}] \geqslant 0.$$

But at the same time, we have

$$\Big\{ \sup_{n\geqslant 0} S_n^\varepsilon > 0 \Big\} = \Big\{ \sup_{n\geqslant 0} \frac{S_n}{n} > \varepsilon \Big\} \cap L_\varepsilon = L_\varepsilon,$$

so

$$0 \leqslant \mathbb{E}[(X_0 - \varepsilon)\mathbb{1}_{L_\varepsilon}] = \mathbb{E}[\mathbb{1}_{L_\varepsilon}\,\mathbb{E}[X \,|\, \mathcal{I}]] - \varepsilon\,\mathbb{P}(L_\varepsilon) = -\varepsilon\,\mathbb{P}(L_\varepsilon).$$

This tells us that $\mathbb{P}(L_\varepsilon) = 0$, giving the a.s. convergence in the ergodic theorem.

Next we discuss the L^1 convergence. We may show that $(\frac{S_n}{n})_{n\geqslant 1}$ is uniformly integrable to deduce the L^1 convergence. But we will use the following more straightforward method.

Given $M > 0$ and define

$$X' = X\mathbb{1}_{|X|\leqslant M} \quad \text{and} \quad X'' = X - X'.$$

Using the a.s. ergodic theorem above, we have

$$\frac{1}{n}\sum_{k=0}^{n-1} X' \circ \varphi^k \xrightarrow[n\to\infty]{} \mathbb{E}[X' \,|\, \mathcal{I}], \quad \text{a.s.}$$

Since X' is bounded, from the dominated convergence theorem, the above convergence also holds in L^1. Then we note that

$$\mathbb{E}\left[\left|\Big(\frac{1}{n}\sum_{k=0}^{n-1} X'' \circ \varphi^k\Big) - \mathbb{E}[X'' \,|\, \mathcal{I}]\right|\right] \leqslant 2\,\mathbb{E}[|X''|] \xrightarrow[M\to\infty]{} 0.$$

Therefore, for any $\varepsilon > 0$, we can choose a large enough M such that $\mathbb{E}[|X''|] < \varepsilon$, then a large enough n such that

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{k=0}^{n-1} X' \circ \varphi^k - \mathbb{E}[X' \,|\, \mathcal{I}]\right|\right] \leqslant \varepsilon,$$

to deduce the L^1 convergence of the ergodic theorem. □

Question 8.2.3: We keep the notations in Theorem 8.2.1 and assume additionally that $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$. Prove the L^p convergence of the ergodic theorem.

Example 8.2.4 : Below are a few applications of Theorem 8.2.1.

(1) If $(X_k)_{k\geqslant 0}$ is an i.i.d. sequence of random variables in L^1, Remark 8.1.13 says that $\mathcal{I}$ is a trivial σ-algebra, giving the following a.s. convergence and L^1 convergence,

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k \xrightarrow[n\to\infty]{} \mathbb{E}[X_0].$$

(2) Consider an irreducible Markov chain $(X_n)_{n\geqslant 0}$ on a countable state space E with invariant probability measure μ. In Proposition 8.1.17, we have seen that the shift operator θ is ergodic with respect to $\mathbb{P}_\mu$. Let $f : E \longrightarrow \mathbb{R}$ with $\mu(|f|) < \infty$, then we have the following $\mathbb{P}_\mu$-a.s. convergence and convergence in L^1,

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \xrightarrow[n\to\infty]{} \mathbb{E}_\mu[f \,|\, \mathcal{I}_\theta] = \mu(f).$$

(3) We carry on the discussion of the rotation operator on the unit circle from Example 8.1.4 (2). According to Exercise 8.12, when θ is irrational, θ_β is ergodic. Thus, given $A \in \mathcal{B}(\mathbb{S}^1)$, we have

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\theta_\beta^k(\omega)\in A} \xrightarrow[n\to\infty]{} \mu(A),$$

where μ is the uniform probability measure on $\mathbb{S}^1$. This is Weyl's equidistribution theorem (等

分佈定理).

8.3 Recurrence of Stationary Sequences

In this section, we discuss the recurrence of stationary sequences using the ergodic theorem seen in Section 8.2.

Below, we want to consider random processes indexed by time steps in $\mathbb{N}$ with values in $\mathbb{R}^d$. We define the sample space to be $\Omega = (\mathbb{R}^d)^{\mathbb{N}}$ and the usual σ-algebra $\mathcal{F} := \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}$, which is the smallest σ-algebra making all the coordinate functions measurable. Let $\mathbb{P}$ be a probability measure such that the shift operator θ is measure-preserving on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Take $\omega \sim \mathbb{P}$, that is the sequence $(X_k(\omega))_{k\geqslant 1}$ given by the coordinate functions is stationary with values in $\mathbb{R}^d$. For all positive integer $n \geqslant 1$, let

$$S_n = X_1 + \cdots + X_n,$$
$$R_n = |\{S_1, \cdots, S_n\}|.$$

If we interprete $(X_k)_{k\geqslant 1}$ as increments of a random walk, then S_n represents its position at time n, R_n the number of sites visited by the walk up to time n.

We have the following result.

Theorem 8.3.1 : *Let*

$$A = \{S_k \neq 0,\ \forall k \geqslant 1\}.$$

Then we have the following convergence,

$$\frac{R_n}{n} \xrightarrow{a.s.} \mathbb{P}(A \,|\, \mathcal{I}).$$

Proof : First, we note that R_n can be represented as $R_n = |B_n|$ with

$$B_n = \{m \geqslant 1 : S_{m+k} \neq S_m,\ \forall 1 \leqslant k \leqslant n-m\}, \qquad \forall n \geqslant 1.$$

Since the following equivalence holds for all integers $m \geqslant 1$,

$$\theta^m(\omega) \in A \Leftrightarrow \omega \in \{S_{m+k} - S_m \neq 0,\ \forall k \geqslant 1\},$$

we have

$$R_n \geqslant \sum_{m=1}^{n} \mathbb{1}_A(\theta^m(\omega)), \qquad \forall n \geqslant 1.$$

From Birkhoff's ergodic theorem (Theorem 8.2.1), we have

$$\liminf_{n\to\infty} \frac{R_n}{n} \geqslant \mathbb{E}[\mathbb{1}_A \,|\, \mathcal{I}], \qquad \text{a.s.}$$

Next, we consider, for a given integer $k \geqslant 1$,

$$A_k = \{S_j \neq 0,\ \forall 1 \leqslant j \leqslant k\}.$$

Similar to the above discussion, we have

$$R_n \leqslant k + \sum_{m=1}^{n-k} \mathbb{1}_{A_k}(\theta^m(\omega)).$$

We use Birkhoff's ergodic theorem again to obtain

$$\limsup_{n\to\infty} \frac{R_n}{n} \leqslant \mathbb{E}[\mathbb{1}_{A_k} \,|\, \mathcal{I}], \qquad \text{a.s.}$$

Due to the monotonicity of the sequence $(A_k)_{k\geqslant 1}$ and the fact that A_k decreases to A when $k \to \infty$, we conclude using the monotone convergence theorem. □

The below theorem extends the results on random walks with i.i.d. increments (Example 7.3.9 and Exercise 7.9) to random walks with increments given by a stationary sequence.

Theorem 8.3.2 : *Consider a stationary sequence $(X_k)_{k\geqslant 1}$ in L^1 with values in $\mathbb{Z}$. Then, we have the following properties.*

(1) If $\mathbb{E}[X_1 \,|\, \mathcal{I}] = 0$, then $\mathbb{P}(A) = 0$.

(2) If $\mathbb{P}(A) = 0$, then $\mathbb{P}(S_n = 0 \text{ i.o.}) = 1$.

This means that if the conditional expectation of the increments of a random walk is zero with respect to the invariant σ-algebra, then it is recurrent.

Proof :

(1) If $\mathbb{E}[X_1 \,|\, \mathcal{I}] = 0$, then the ergodic theorem says that $\frac{S_n}{n}$ converges to 0 almost surely. Thus, we have

$$\limsup_{n\to\infty} \left(\max_{1\leqslant k\leqslant n} \frac{|S_k|}{n} \right) = \limsup_{n\to\infty} \left(\max_{K\leqslant k\leqslant n} \frac{|S_k|}{n} \right) \leqslant \max_{k\geqslant K} \frac{|S_k|}{k} \xrightarrow[K\to\infty]{} 0,$$

that is

$$\lim_{n\to\infty} \left(\max_{1\leqslant k\leqslant n} \frac{|S_k|}{n} \right) = 0.$$

Since $R_n \leqslant 1 + 2\max_{1\leqslant k\leqslant n} |S_k|$, we get $\frac{R_n}{n} \longrightarrow 0$ and Theorem 8.3.1 implies $\mathbb{P}(A) = 0$.

(2) We note that $A^c = \{\exists n \geqslant 1 : S_n = 0\}$, which means that if $\mathbb{P}(A) = 0$, then $(S_n)_{n\geqslant 1}$ visits 0 at least once with probability 1. According to the first visit time of $(S_n)_{n\geqslant 1}$ to 0, we decompose A^c into a disjoint union of events $A^c = \sqcup_{k\geqslant 1} F_k$ with

$$F_k = \{S_i \neq 0 \text{ for all } 1 \leqslant i < k \text{ and } S_k = 0\}, \qquad \forall k \geqslant 1.$$

Since we want to look at the probability that $(S_n)_{n\geqslant 1}$ visits 0 at least two times, we define the following events,

$$G_{j,k} = \theta^j(F_k) = \{S_{j+i} - S_j \neq 0 \text{ for all } 1 \leqslant i < k \text{ and } S_{j+k} - S_j = 0\}, \qquad \forall j,k \geqslant 1.$$

Due to the stationarity of $(X_k)_{k\geqslant 1}$, we know that $\mathbb{P}(G_{j,k}) = \mathbb{P}(F_k)$ for all $j,k \geqslant 1$ and when j is fixed, $(G_{j,k})_{k\geqslant 1}$ is a collection of disjoint sets with total measure 1. In other words, we have

$$\sum_{j,k\geqslant 1} \mathbb{P}(F_j \cap G_{j,k}) = \sum_{j\geqslant 1} \mathbb{P}(F_j) = 1.$$

The event $F_j \cap G_{j,k}$ represents $S_j = 0$ and $S_{j+k} = 0$, hence we have shown that $(S_n)_{n\geqslant 1}$ visits 0 at least twice with probability 1. We can repeat this technique to show that for any $m \geqslant 1$, $(S_n)_{n\geqslant 1}$ visits 0 at least m times with probability 1.

□

To close this section, we extend the result of Corollary 7.4.13 to stationary sequences.

Theorem 8.3.3 : *Consider a stationary sequence $(X_n)_{n\geqslant 0}$ in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a measurable set $A \in \mathcal{F}$ and define the stopping times $T_0 = 0$ and*

$$T_{n+1} = \inf\{k > T_n : X_k \in A\}, \qquad \forall n \geqslant 0.$$

If we have $T_1 < \infty$ $\mathbb{P}$-a.s., then under the conditional probability $\mathbb{P}(\cdot \mid X_0 \in A)$, the sequence of random variables $(\tau_n := T_{n+1} - T_n)_{n\geqslant 0}$ is stationary and

$$\mathbb{E}[\tau_0 \mid X_0 \in A] = \frac{1}{\mathbb{P}(X_0 \in A)}. \tag{8.3}$$

Remark 8.3.4 : If $(X_n)_{n\geqslant 0}$ is a positive recurrent and irreducible Markov chain with the unique invariant probability measure μ, then under $\mathbb{P}_\mu$, the process $(X_n)_{n\geqslant 0}$ is starionary. For any $x \in E$, if we take $A = \{x\}$, then Eq. (8.3) simplifies to Eq. (7.19). Thus, Theorem 8.3.3 can be seen as an extension of Corollary 7.4.13, where the assumption of the Markov chain is removed and the starting point needs not be a fixed point almost surely.

Proof : Since we need to take conditional probabilities and conditional expectations with respect to $X_0 \in A$, we can use the following trick to simplify the computations. By Exercise 8.2, we can extend $(X_n)_{n\geqslant 0}$ to a stationary sequence $(X_n)_{n\in\mathbb{Z}}$ indexed by $\mathbb{Z}$. Then, define the following disjoint events

$$C_k = \{X_{-1} \notin A, \dots, X_{-(k-1)} \notin A, X_{-k} \in A\}, \qquad \forall k \geqslant 1.$$

We have

$$\Big(\bigsqcup_{k=1}^{K} C_k\Big)^c = \{X_k \notin A \text{ for all } -K \leqslant k \leqslant -1\}.$$

By stationarity, the above event on the right side has the same probability as $\{X_k \notin A$ for all $1 \leqslant k \leqslant$

$K\}$, taking $K \to \infty$ and using the assumption of the theorem, we have

$$\mathbb{P}\Big(\bigsqcup_{k=1}^{\infty} C_k\Big) = 1. \tag{8.4}$$

Now, we are ready to show that under the conditional probability $\mathbb{P}(\cdot \,|\, X_0 \in A)$, the sequence $(\tau_n)_{n\geqslant 0}$ is stationary. Given a positive integer j, non-negative integers $n_1 < n_2 < \cdots < n_j$ and positive integers $m_1, \cdots, m_j$, we have

$$\begin{aligned}\mathbb{P}(\tau_{n_1+1} = m_1, \dots, \tau_{n_j+1} = m_j, X_0 \in A) &= \sum_{k\geqslant 1} \mathbb{P}(\tau_0 = k, \tau_{n_1+1} = m_1, \dots, \tau_{n_j+1} = m_j, X_0 \in A)\\ &= \sum_{k\geqslant 1} \mathbb{P}(C_k, X_0 \in A, \tau_{n_1} = m_1, \dots, \tau_{n_j} = m_j),\\ &= \mathbb{P}(X_0 \in A, \tau_{n_1} = m_1, \dots, \tau_{n_j} = m_j),\end{aligned}$$

where we use the assumption of stationarity in the second equality, and Eq. (8.4) in the last equality to conclude.

Finally, let us prove Eq. (8.3),

$$\begin{aligned}\mathbb{E}[\tau_0 \,|\, X_0 \in A] &= \sum_{k\geqslant 1} \mathbb{P}(\tau_0 \geqslant k \,|\, X_0 \in A)\\ &= \mathbb{P}(X_0 \in A)^{-1} \sum_{k\geqslant 1} \mathbb{P}(\tau_0 \geqslant k, X_0 \in A)\\ &= \mathbb{P}(X_0 \in A)^{-1} \sum_{k\geqslant 1} \mathbb{P}(C_k) = \mathbb{P}(X_0 \in A)^{-1},\end{aligned}$$

where in the third equality, we use the property of a stationary sequence,

$$\begin{aligned}\mathbb{P}(\tau_0 \geqslant k, X_0 \in A) &= \mathbb{P}(X_0 \in A, X_1 \notin A, \dots, X_{k-1} \notin A)\\ &= \mathbb{P}(\theta^{-k}(X_0 \in A, X_1 \notin A, \dots, X_{k-1} \notin A)) = \mathbb{P}(C_k).\end{aligned}$$

□

Question 8.3.5: We keep the notations from Theorem 8.3.3. Let $B \in \mathcal{F}$ satisfying $A \cap B = \varnothing$. Show

$$\mathbb{E}\left[\sum_{k=0}^{T_1-1} \mathbb{1}_{X_k\in B} \,\middle|\, X_0 \in A\right] = \frac{\mathbb{P}(X_0 \in B)}{\mathbb{P}(X_0 \in A)}.$$

How to compare this to the case of a Markov chain in Theorem 7.4.10?

8.4 Subadditive Ergodic Theorem

In this chapter, we will discuss a generalization of Birkhoff's ergodic theorem. In Theorem 8.2.1, the assumption on the stationarity of the sequence of random variables was necessary to obtain the ergodic theorem. However, in real-world problems, this assumption is still too restrictive and hence unrealistic. Later, we will see in Theorem 8.4.3 that under appropriate conditions, even without stationarity, we are still able to get a similar result as Birkhoff's ergodic theorem.

A real sequence $(x_n)_{n\geqslant 1}$ is said to be *subadditive* (劣加性) if it satisfies

$$x_{n+m} \leqslant x_n + x_m, \qquad \forall n, m \geqslant 1.$$

Lemma 8.4.1 (Fekete 引理) **:** *The following limit exists for any subadditive sequence* $(x_n)_{n\geqslant 1}$,

$$\lim_{n\to\infty} \frac{x_n}{n} = \inf_{n\geqslant 1} \frac{x_n}{n} \in \mathbb{R} \cup \{-\infty\}.$$

Proof : First note that we have

$$\liminf_{n\to\infty} \frac{x_n}{n} \geqslant \inf_{n\geqslant 1} \frac{x_n}{n}.$$

Given $\varepsilon > 0$, take $N \geqslant 1$ such that

$$\frac{x_N}{N} \leqslant \inf_{n\geqslant 1} \frac{x_n}{n} + \varepsilon.$$

For any $n \geqslant N$, we can write $n = kN + r$ with $k \in \mathbb{N}$ and $0 \leqslant r \leqslant N - 1$. By subadditivity, we have

$$\frac{x_n}{n} \leqslant \frac{kx_N}{n} + \frac{x_r}{n}.$$

Taking $\limsup$ in the above formula, we get

$$\limsup_{n\to\infty} \frac{x_n}{n} \leqslant \frac{x_N}{N} \leqslant \inf_{n\geqslant 1} \frac{x_n}{n} + \varepsilon.$$

Since $\varepsilon > 0$ can be arbitrarily small, the proof is complete. □

Example 8.4.2 (self-avoiding walk) **:** We are given a path $(x_k)_{0\leqslant k\leqslant n}$ of length $n \geqslant 1$ on $\mathbb{Z}^d$. It is called a *self-avoiding walk* (自迴避路徑) of length n if

- we have $x_k \sim x_{k+1}$ for all $0 \leqslant k \leqslant n-1$;
- we have $x_j \neq x_k$ for all $0 \leqslant j < k \leqslant n$.

Let us denote by $\mathrm{SAW}(n)$ the set of all the self-avoiding walks starting from 0 of length n and denote its cardinal by $c_n = |\mathrm{SAW}(n)|$. The sequence $(\ln c_n)_{n\geqslant 1}$ is subadditive and one can apply Lemma 8.4.1 to obtain

$$\gamma := \lim_{n\to\infty} \frac{\ln c_n}{n} = \inf_{n\geqslant 1} \frac{\ln c_n}{n} \in \mathbb{R} \cup \{-\infty\}.$$

The constant $\mu = e^\gamma$ is called *connective constant* (連通常數) .

On any periodic lattice in any dimension, we can use Lemma 8.4.1 to show the existence of the connective constant but it is highly non trivial to compute its exact value. The only exact computation available at the current moment is on the planar hexagonal lattice with value $\mu = \sqrt{2+\sqrt{2}}$. [1]

[1] Duminil-Copin, Hugo and Stanislav Smirnov (2012). "The connective constant of the honeycomb lattice equals $\sqrt{2+\sqrt{2}}$". *Annals of Mathematics*, 1653-1665.

The below theorem is the stochastic version of Fekete's lemma.

Theorem 8.4.3 (Subadditive ergodic theorem) **:** *Let $(X_{m,n})_{n>m\geqslant 0}$ be a sequence of random variables satisfying*

(a) $X_{0,n} \leqslant X_{0,m} + X_{m,n}$ *for all* $n > m \geqslant 1$.

(b) The sequence $(X_{nk,(n+1)k})_{n\geqslant 0}$ *is a stationary sequence for all integers* $k \geqslant 1$.

(c) The distribution of the sequence $(X_{m,m+k})_{k\geqslant 1}$ *does not depend on* $m \geqslant 0$.

(d) $X_{0,n} \in L^1$ *for all* $n \geqslant 1$.

$$\inf_{n\geqslant 1} \frac{\mathbb{E}[X_{0,n}]}{n} > -\infty.$$

Then, the following statements holds.

(1) The convergence holds almost surely and in L^1,

$$X_\infty := \lim_{n\to\infty} \frac{X_{0,n}}{n}.$$

(2) The following equality holds,

$$\mathbb{E}[X_\infty] = \lim_{n\to\infty} \frac{\mathbb{E}[X_{0,n}]}{n} = \inf_{n\geqslant 1} \frac{\mathbb{E}[X_{0,n}]}{n}.$$

(3) If the sequences in the assumption (b) are ergodic for all $k \geqslant 0$, *then* $X_\infty = \mathbb{E}[X_\infty]$ *a.s.*

Remark 8.4.4 :

(1) This theorem was first proved in 1973 by Kingman, then improved in 1985 by Liggett. The statement here is the version from Liggett. The difference between two versions lies in the assumption (a), Kingman required the stronger assumption below,

$$X_{l,n} \leqslant X_{l,m} + X_{m,n}, \qquad \forall 0 \leqslant l \leqslant m \leqslant n.$$

(2) If we use the notations from Theorem 8.2.1 and define

$$X_{m,n} = \sum_{k=m}^{n-1} X \circ \varphi^k, \qquad \forall n > m \geqslant 0,$$

then the subadditive ergodic theorem gives the below almost sure convergence and convergence in L^1,

$$X_\infty = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X \circ \varphi^k,$$

where $\mathbb{E}[X_\infty] = \mathbb{E}[X]$. Additionally, for all $A \in \mathcal{I}_\varphi$, we have

$$\mathbb{E}[X_\infty \mathbb{1}_A] = \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n} \sum_{k=0}^{n-1} X \circ \varphi^k \mathbb{1}_A\right] = \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n} \sum_{k=0}^{n-1} X \mathbb{1}_{\varphi^{-k}(A)}\right] = \mathbb{E}[X\mathbb{1}_A],$$

giving that $X_\infty = \mathbb{E}[X \mid \mathcal{I}_\varphi]$, which is the result from Theorem 8.2.1.

Before proving Theorem 8.4.3, we still need the following Definition 8.4.5 and Lemma 8.4.6.

Definition 8.4.5 : Let X and Y be two random variables with values in $\mathbb{R}^d$. We say that X *stochastically dominates* (隨機優勢) Y if

$$\mathbb{E}[f(X)] \geqslant \mathbb{E}[f(Y)],$$

for all non-decreasing measurable functions $f : \mathbb{R}^d \longrightarrow \mathbb{R}_{\geqslant 0}$, where we consider the partial order $\leqslant$ on $\mathbb{R}^d$ to be defined as

$$(x_1, \dots, x_d) \leqslant (y_1, \dots, y_d) \quad \Leftrightarrow \quad x_i \leqslant y_i \quad \forall i = 1, \dots, d.$$

We may also write $X \succeq Y$.

If we make use of the notion of coupling introduced in Definition 7.5.9, then we can find random variables X' and Y' defined on the same probability space satisfying $X \overset{(d)}{=} X'$, $Y \overset{(d)}{=} Y'$ and $X \geqslant Y$ a.s.

Lemma 8.4.6 : *Given a sequence $(X_n)_{n\geqslant 1}$ of real-valued random variables and assume that the sequence $(X_n^-)_{n\geqslant 1}$ of negative parts is uniformly integrable. Then, Fatou's lemma still holds,*

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n].$$

Proof : Let $\varepsilon > 0$ and $a > 0$ such that

$$\mathbb{E}\left[(X_n^-)\mathbb{1}_{X_n^- \geqslant a}\right] < \varepsilon.$$

First, we apply Fatou's lemma to $(X_n + a)^+$,

$$\mathbb{E}\left[\liminf_{n\to\infty}(X_n + a)\right] \leqslant \mathbb{E}\left[\liminf_{n\to\infty}(X_n + a)^+\right] \leqslant \liminf_{n\to\infty} \mathbb{E}[(X_n + a)^+].$$

Besides, we have, for all $n \geqslant 1$, that

$$(X_n + a)^+ = (X_n + a) + (X_n + a)^- \leqslant (X_n + a) + X_n^- \mathbb{1}_{X_n^- \geqslant a}.$$

We combine the above three inequalities together to obtain

$$\mathbb{E}\left[\liminf_{n\to\infty}(X_n + a)\right] \leqslant \liminf_{n\to\infty} \mathbb{E}[(X_n + a) + X_n^- \mathbb{1}_{X_n^- \geqslant a}] = \liminf_{n\to\infty} \mathbb{E}[(X_n + a)] + \varepsilon.$$

We substract a from both sides and take $\varepsilon \to 0$ to complete the proof. □

Now we are ready to prove Theorem 8.4.3.

Proof : We define the following notations,

$$\forall n \geqslant 1, \quad a_n = \mathbb{E}[X_{0,n}], \qquad \overline{X}_\infty = \limsup_{n\to\infty} \frac{X_{0,n}}{n}, \qquad \underline{X}_\infty = \liminf_{n\to\infty} \frac{X_{0,n}}{n},$$

and decompose the prove into four steps.

(i) Use Fekete's lemma to show that a_n/n a.s. converges and denote this limit by γ.

(ii) Show that $\mathbb{E}[\overline{X}_\infty] \leqslant \gamma$.

(iii) Show that $\mathbb{E}[\underline{X}_\infty] \geqslant \gamma$ and hence from (ii) we obtain $X_\infty := \underline{X}_\infty = \overline{X}_\infty$ a.s. and $\mathbb{E}[X_\infty] = \gamma$.

(iv) Show that $X_{0,n}/n$ converges in L^1.

We start with the prove of (i). From assumptions (a) and (c), we have

$$a_{n+m} \leqslant a_n + a_m, \qquad \forall n, m \geqslant 1.$$

So Fekete's subadditive lemma (Lemma 8.4.1) gives

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\geqslant 1} \frac{a_n}{n} =: \gamma. \tag{8.5}$$

Then, let us prove (ii). Given a positive integer $m \geqslant 1$, use the assumption (b) and Birkhoff's ergodic theorem (Theorem 8.2.1), the following converges a.s. and in L^1,

$$\frac{1}{k}\sum_{j=0}^{k-1} X_{jm,(j+1)m} \xrightarrow[n\to\infty]{} \mathbb{E}[X_{0,m} \,|\, \mathcal{I}_m] =: A_m,$$

where $\mathcal{I}_m$ is the invariant σ-algebra for the m-th iterated shift operator θ^m. We make use of the assumption (a) to obtain

$$\limsup_{k\to\infty} \frac{X_{0,km}}{km} \leqslant \frac{A_m}{m}. \tag{8.6}$$

However, the quantity we want to look at is $\overline{X}_\infty = \limsup \frac{X_{0,n}}{n}$, so we need to estimate the corresponding missing terms in Eq. (8.6). Any positive integer $n \geqslant 1$ can be written as $n = km + l$ with $1 \leqslant l \leqslant m$, so we have

$$X_{0,n} \leqslant X_{0,km} + X_{km,km+l}. \tag{8.7}$$

If l is fixed, the assumption (c) guarantees that $X_{km,km+l}$ has the same distribution as $X_{0,l} \in L^1$ and we have, for all $\varepsilon > 0$,

$$\sum_{k\geqslant 1} \mathbb{P}(|X_{km,km+l}| > k\varepsilon) = \sum_{k\geqslant 1} \mathbb{P}(|X_{0,l}| > k\varepsilon) = \frac{1}{\varepsilon}\, \mathbb{E}\left[|X_{o,l}|\right] < \infty.$$

By the Borel-Cantelli lemma, this implies that almost surely, we have

$$\lim_{k\to\infty} \frac{X_{km,km+l}}{k} = 0, \quad \forall m \geqslant 1, 1 \leqslant \ell \leqslant m.$$

Therefore, along with Eq. (8.6) and Eq. (8.7), we find

$$\overline{X}_\infty \leqslant \frac{A_m}{m}. \tag{8.8}$$

This also implies

$$\mathbb{E}[\overline{X}_\infty] \leqslant \frac{\mathbb{E}[A_m]}{m} = \frac{a_m}{m}.$$

We can take inf in m in the right side of the above formula to conclude $\mathbb{E}[\overline{X}_\infty] \leqslant \gamma$.

We can note that, if in the assumption (b), the sequence $(X_{nk,(n+1)k})_{n\geqslant 0}$ is ergodic for all $k \geqslant 1$, then we can deduce directly $\overline{X}_\infty \leqslant \gamma$ a.s.

Next, let us prove (iii). If we can construct a stationary sequence $(Y_k)_{k\geqslant 1}$ satisfying

$$\begin{cases} \mathbb{E}[Y_k] \geqslant \gamma, & \forall k \geqslant 1, \\ (X_{0,k})_{1\leqslant k\leqslant n} \succeq (Y_1 + \dots Y_k)_{1\leqslant k\leqslant n}, \quad \text{a.s.,} & \forall n \geqslant 1, \end{cases} \tag{8.9}$$

then we can easily get (iii).

Given a positive integer $p \geqslant 1$, let U_p be a random variable with uniform distribution on $\{1, \dots, p\}$. Let us assume that $(U_p)_{p\geqslant 1}$ is an independent sequence and that it is independent of $(X_{m,n})_{n>m\geqslant 0}$ as well. For any positive integers $k, p \geqslant 1$, we define

$$Y_k^{(p)} = X_{0,k+U_p} - X_{0,k+U_p-1},$$

then we have

$$\mathbb{E}[Y_k^{(p)}] = \frac{1}{p}\,\mathbb{E}[X_{0,k+p} - X_{0,k}] = \frac{1}{p}(a_{k+p} - a_k).$$

Then, the assumptions (a) and (c) give that

$$\mathbb{E}[(Y_k^{(p)})^+] = \frac{1}{p}\sum_{l=1}^{p}\mathbb{E}[(X_{0,k+l} - X_{0,k+l-1})^+] \leqslant \frac{1}{p}\sum_{l=1}^{p}\mathbb{E}[X^+_{k+l-1,k+l}] = \mathbb{E}[X^+_{0,1}].$$

In consequence, when k is fixed, we obtain

$$\begin{aligned} \sup_{p\geqslant 1}\mathbb{E}[|Y_k^{(p)}|] &= \sup_{p\geqslant 1}\big(2\,\mathbb{E}[(Y_k^{(p)})^+] - \mathbb{E}[Y_k^{(p)}]\big) \\ &\leqslant 2\,\mathbb{E}[X^+_{0,1}] - \inf_{p\geqslant 1}\big(\frac{1}{p}(a_{k+p} - a_k)\big) < \infty, \end{aligned}$$

and also the following using Eq. (8.5),

$$\lim_{p\to\infty}\mathbb{E}[Y_k^{(p)}] = \gamma.$$

By Exercise 1.26, we know that an L^1-bounded sequence of random variables has an almost surely converging subsequence, moreover, the a.s. convergence implies the convergence in distribution, so Cantor's diagonal argument (對角論證法) provides us with a subsequence $(p_i)_{i\geqslant 1}$ such that the sequence $(Y_k^{(p_i)})_{k\geqslant 1}$ of random variables converges in distribution with limit denoted by $(Y_k)_{k\geqslant 1}$. This is equivalent to saying that for any $k \geqslant 1$ and a bounded measurable function f on $\mathbb{R}^k$, the following convergence holds,

$$\mathbb{E}[f(Y_1, \dots, Y_k)] = \lim_{i\to\infty}\frac{1}{p_i}\sum_{l=1}^{p_i}\mathbb{E}[f(X_{0,l+1} - X_{0,l}, \dots, X_{0,l+k} - X_{0,l+k-1})].$$

From this formula we can also conclude that $(Y_k)_{k\geqslant 1}$ is stationary.

Now let us check that the sequence $(Y_k)_{k\geqslant 1}$ constructed above satisfies Eq. (8.9). First, the assumptions (a) and (c) give us

$$Y_1^{(p)} = X_{0,U_p+1} - X_{0,U_p} \leqslant X_{U_p,U_p+1} \overset{(d)}{=} X_{0,1},$$

so the sequence $((Y_1^{(p)})^+)_{p\geqslant 1}$ is uniformly integrable. From Fatou's lemma (Lemma 8.4.6), we have

$$\mathbb{E}[Y_1] = \mathbb{E}\left[\lim_{i\to\infty} Y_1^{(p_i)}\right] \geqslant \limsup_{i\to\infty} \mathbb{E}[Y_1^{(p_i)}] = \gamma,$$

Again using Birkhoff's ergodic theorem (Theorem 8.2.1), we conclude that the limit

$$Y := \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} Y_k, \quad \text{a.s.}$$

exists and $\mathbb{E}[Y] = \mathbb{E}[Y_1] \geqslant \gamma$.

In the end, we still need to check that the sequence $(X_{0,1}, X_{0,2}, \dots, X_{0,n})$ stochastically dominates $(Y_1, Y_1 + Y_2, \dots, Y_1 + \cdots + Y_n)$. Given $n \geqslant 1$ and a non-decreasing and non-negative measurable function f on $\mathbb{R}^n$, we have

$$\begin{aligned}
&\mathbb{E}[f(Y_1, Y_1+Y_2, \dots, Y_1+\cdots+Y_n)] \\
&\quad = \lim_{i\to\infty} \frac{1}{p_i}\sum_{l=1}^{p_i} \mathbb{E}[f(X_{0,l+1} - X_{0,l}, X_{0,l+2} - X_{0,l}, \dots, X_{0,l+n} - X_{0,l})] \\
&\quad \leqslant \lim_{i\to\infty} \frac{1}{p_i}\sum_{l=1}^{p_i} \mathbb{E}[f(X_{l,l+1}, X_{l,l+2}, \dots, X_{l,l+n})] \\
&\quad = \mathbb{E}[f(X_{0,1}, X_{0,2}, \dots, X_{0,n})].
\end{aligned}$$

Finally, let us prove (iv). From the assumption (a), we have

$$(X_{0,n} - nX_\infty)^+ \leqslant \sum_{k=0}^{n-1} (X_{k,k+1} - X_\infty)^+,$$

and since the assumption (c) says that the terms in the sequence $((X_{k,k+1} - X_\infty)^+)_{k\geqslant 0}$ of random variables are identically distributed, $((\frac{X_{0,n}}{n} - X_\infty)^+)_{n\geqslant 0}$ is an uniformly integrable sequence. Using the fact that X_∞ is the a.s. limit of $\frac{X_{0,n}}{n}$, we obtain

$$\mathbb{E}\left[\left(\frac{X_{0,n}}{n} - X_\infty\right)^+\right] \xrightarrow[n\to\infty]{} 0.$$

Since we also have $\mathbb{E}[\frac{X_{0,n}}{n}] \to \mathbb{E}[X_\infty]$, using the identity $|x| = 2x^+ - x$, we find

$$\mathbb{E}\left[\left|\frac{X_{0,n}}{n} - X_\infty\right|\right] \xrightarrow[n\to\infty]{} 0.$$

□