Chapter 13
Sequences of Dependent Trials. Markov Chains

Abstract The chapter opens with in Sect. 13.1 presenting the key definitions and first examples of countable Markov chains. The section also contains the classification of states of the chain. Section 13.2 contains necessary and sufficient conditions for recurrence of states, the Solidarity Theorem for irreducible Markov chains and a theorem on the structure of a periodic Markov chain. Key theorems on random walks on lattices are presented in Sect. 13.3, along with those for a general symmetric random walk on the real line. The ergodic theorem for general countable homogeneous chains is established in Sect. 13.4, along with its special case for finite Markov chains and the Law of Large Numbers and the Central Limit Theorem for the number of visits to a given state. This is followed by a short Sect. 13.5 detailing the behaviour of transition probabilities for reducible chains. The last three sections are devoted to Markov chains with arbitrary state spaces. First the ergodicity of such chains possessing a positive atom is proved in Sect. 13.6, then the concept of Harris Markov chains is introduced and conditions of ergodicity of such chains are established in Sect. 13.7. Finally, the Laws of Large Numbers and the Central Limit Theorem for sums of random variables defined on a Markov chain are obtained in Sect. 13.8.

13.1 Countable Markov Chains. Definitions and Examples. Classification of States

13.1.1 Definition and Examples

So far we have studied sequences of independent trials. Now we will consider the simplest variant of a sequence of *dependent* trials.

Let G be an experiment having a finite or countable set of outcomes $\{E_1, E_2, \ldots\}$. Suppose we keep repeating the experiment G. Denote by X_n the number of the outcome of the n-th experiment.

In general, the probabilities of different values of E_{X_n} can depend on what events occurred in the previous $n-1$ trials. If this probability, *given a fixed outcome* $E_{X_{n-1}}$ *of the* $(n-1)$*-st trial*, does not depend on the outcomes of the preceding $n-2$ trials, then one says that this sequence of trials forms a Markov chain.

A.A. Borovkov, *Probability Theory*, Universitext,
DOI 10.1007/978-1-4471-5201-9_13, © Springer-Verlag London 2013

To give a precise definition of a Markov chain, consider a sequence of integer-valued random variables $\{X_n\}_{n=0}^{\infty}$. If the n-th trial resulted in outcome E_j, we set $X_n := j$.

Definition 13.1.1 A sequence $\{X_n\}_0^{\infty}$ forms a *Markov chain* if

$$\begin{aligned}\mathbf{P}(X_n = j \,|\, X_0 = k_0, X_1 = k_1, \ldots, X_{n-2} = k_{n-2}, X_{n-1} = i)\\ \mathbf{P}(X_n = j \,|\, X_{n-1} = i) =: p_{ij}^{(n)}.\end{aligned} \qquad (13.1.1)$$

These are the so-called countable (or discrete) Markov chains, i.e. Markov chains *with countable state spaces*.

Thus, a Markov chain may be thought of as a system with possible states $\{E_1, E_2, \ldots\}$. Some "initial" distribution of the variable X_0 is given:

$$\mathbf{P}(X_0 = j) = p_j^0, \quad \sum p_j^0 = 1.$$

Next, at integer time epochs the system changes its state, the conditional probability of being at state E_j at time n given the previous history of the system only being dependent on the state of the system at time $n-1$. One can briefly characterise this property as follows: *given the present, the future and the past of the sequence X_n are independent*.

For example, the branching process $\{\zeta_n\}$ described in Sect. 7.7, where ζ_n was the number of particles in the n-th generation, is a Markov chain with possible states $\{0, 1, 2, \ldots\}$.

In terms of conditional expectations or conditional probabilities (see Sect. 4.8), the *Markov property* (as we shall call property (13.1.1)) can also be written as

$$\mathbf{P}\big(X_n = j \bigm| \sigma(X_0, \ldots, X_{n-1})\big) = \mathbf{P}\big(X_n \bigm| \sigma(X_{n-1})\big),$$

where $\sigma(\cdot)$ is the σ-algebra generated by random variables appearing in the argument, or, which is the same,

$$\mathbf{P}(X_n = j \bigm| X_0, \ldots, X_{n-1}) = \mathbf{P}(X_n \mid X_{n-1}).$$

This definition allows immediate extension to the case of a Markov chain with a more general state space (see Sects. 13.6 and 13.7).

The problem of the existence of a sequence $\{X_n\}_0^{\infty}$ which is a Markov chain with given transition probabilities $p_{ij}^{(n)}$ ($p_{ij}^{(n)} \ge 0$, $\sum_j p_{ij}^{(n)} = 1$) and a given "initial" distribution $\{p_k^0\}$ of the variable X_0 can be solved in the same way as for independent random variables. It suffices to apply the Kolmogorov theorem (see Appendix 2) and specify consistent joint distributions by

$$\mathbf{P}(X_0 = k_0, X_1 = k_1, \ldots, X_n = k_n) := p_{k_0}^0 p_{k_0 k_1}^{(1)} p_{k_1 k_2}^{(2)} \cdots p_{k_{n-1} k_n}^{(n)},$$

which are easily seen to satisfy the Markov property (13.1.1).

Definition 13.1.2 A Markov chain $\{X_n\}_0^\infty$ is said to be *homogeneous* if the probabilities $p_{ij}^{(n)}$ do not depend on n.

We consider several examples.

Example 13.1.1 (Walks with absorption and reflection) Let $a > 1$ be an integer. Consider a walk of a particle over integers between 0 and a. If $0 < k < a$, then from the point k with probabilities $1/2$ the particle goes to $k-1$ or $k+1$. If k is equal to 0 or a, then the particle remains at the point k with probability 1. This is the so-called walk with *absorption*. If X_n is a random variable which is equal to the coordinate of the particle at time n, then the sequence $\{X_n\}$ forms a Markov chain, since the conditional expectation of the random variable X_n given $X_0, X_1, \ldots, X_{n-1}$ depends only on the value of X_{n-1}. It is easy to see that this chain is homogeneous.

This walk can be used to describe a fair game (see Example 4.2.3) in the case when the total capital of both gamblers equals a. Reaching the point a means the ruin of the second gambler.

On the other hand, if the particle goes from the point 0 to the point 1 with probability 1, and from the point a to the point $a-1$ with probability 1, then we have a walk with *reflection*. It is clear that in this case the positions X_n of the particle also form a homogeneous Markov chain.

Example 13.1.2 Let $\{\xi_k\}_{k=0}^\infty$ be a sequence of independent integer-valued random variables and $d > 0$ be an integer. The random variables $X_n := \sum_{k=0}^n \xi_k \pmod d$ obtained by adding ξ_k modulo d ($X_n = \sum_{k=0}^n \xi_k - jd$, where j is such that $0 \le X_n < d$) form a Markov chain. Indeed, we have $X_n = X_{n-1} + \xi_n \pmod d$, and therefore the conditional distribution of X_n given $X_1, X_2, \ldots, X_{n-1}$ depends only on X_{n-1}.

If, in addition, $\{\xi_k\}$ are identically distributed, then this chain is homogeneous.

Of course, all the aforesaid also holds when $d = \infty$, i.e. for the conventional summation. The only difference is that the set of possible states of the system is in this case infinite.

From the definition of a homogeneous Markov chain it follows that the probabilities $p_{ij}^{(n)}$ of transition from state E_i to state E_j on the n-th step do not depend on n. Denote these probabilities by p_{ij}. They form the *transition* matrix $P = \|p_{ij}\|$ with the properties

$$p_{ij} \ge 0, \qquad \sum_j p_{ij} = 1.$$

The second property is a consequence of the fact that the system, upon leaving the state E_i, enters with probability 1 one of the states $E_1, E_2, \ldots$.

Matrices with the above properties are said to be *stochastic*.

The matrix P *completely describes the law of change of the state of the system after one step*. Now consider the change of the state of the system after k steps. We introduce the notation $p_{ij}(k) := \mathbf{P}(X_k = j | X_0 = i)$. For $k > 1$, the total probability formula yields

$$p_{ij}(k) = \sum_s \mathbf{P}(X_{k-1} = s | X_0 = i) p_{sj} = \sum_s p_{is}(k-1) p_{sj}.$$

Summation here is carried out over all states. If we denote by $P(k) := \| p_{ij}(k) \|$ the matrix of transition probabilities $p_{ij}(k)$, then the above equality means that $P(k) = P(k-1)P$ or, which is the same, that $P(k) = P^k$. Thus the matrix P *uniquely determines transition probabilities for any number of steps*. It should be added here that, for a homogeneous chain,

$$\mathbf{P}(X_{n+k} = j | X_n = i) = \mathbf{P}(X_k = j | X_0 = i) = p_{ij}(k).$$

We see from the aforesaid that the "distribution" of a chain will be completely determined by the matrix P and the initial distribution $p_k^0 = \mathbf{P}(X_0 = k)$.

We leave it to the reader as an exercise to verify that, for an arbitrary $k \geq 1$ and sets $B_1, \ldots, B_{n-k}$,

$$\mathbf{P}(X_n = j | X_{n-k} = i; X_{n-k-1} \in B_1, \ldots, X_0 \in B_{n-k}) = p_{ij}(k).$$

To prove this relation one can first verify it for $k = 1$ and then make use of induction.

It is obvious that a sequence of independent integer-valued identically distributed random variables X_n forms a Markov chain with $p_{ij} = p_j = \mathbf{P}(X_n = j)$. Here one has $P(k) = P^k = P$.

13.1.2 Classification of States[1]

Definition 13.1.3

K1. A state E_i is called *inessential* if there exist a state E_j and an integer $t_0 > 0$ such that $p_{ij}(t_0) > 0$ and $p_{ji}(t) = 0$ for every integer t.

Otherwise the state E_i is called *essential*.

K2. Essential states E_i and E_j are called *communicating* if there exist such integers $t > 0$ and $s > 0$ that $p_{ij}(t) > 0$ and $p_{ji}(s) > 0$.

Example 13.1.3 Assume a system can be in one of the four states $\{E_1, E_2, E_2, E_4\}$ and has the transition matrix

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

[1] Here and in Sect. 12.2 we shall essentially follow the paper by A.N. Kolmogorov [23].

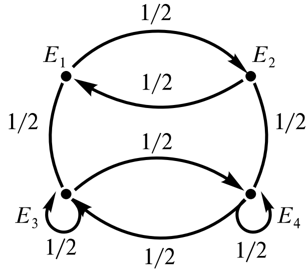

Fig. 13.1 Possible transitions and their probabilities in Example 13.1.3

In Fig. 13.1 the states are depicted by dots, transitions from state to state by arrows, numbers being the corresponding probabilities. In this chain, the states E_1 and E_2 are inessential while E_3 and E_4 are essential and communicating.

In the walk with absorption described in Example 13.1.1, the states $1, 2, \ldots, a-1$ are inessential. The states 0 and a are essential but non-communicating, and it is natural to call them *absorbing*. In the walk with reflection, all states are essential and communicating.

Let $\{X_n\}_{n=0}^{\infty}$ be a homogeneous Markov chain. We distinguish the class S^0 of all inessential states. Let E_i be an essential state. Denote by S_{E_i} the class of states comprising E_i and all states communicating with it. If $E_j \in S_{E_i}$, then E_j is essential and communicating with E_i, and $E_i \in S_{E_j}$. Hence $S_{E_i} = S_{E_j}$. Thus, the whole set of essential states can be decomposed into disjoint classes of communicating states which will be denoted by $S^1, S^2, \ldots$

Definition 13.1.4 If the class S_{E_i} consists of the single state E_i, then this state is called *absorbing*.

It is clear that after a system has hit an essential state E_i, it can never leave the class S_{E_i}.

Definition 13.1.5 A Markov chain consisting of a single class of essential communicating states is said to be *irreducible*. A Markov chain is called *reducible* if it contains more than one such class.

If we enumerate states so that the states from S^0 come first, next come states from S^1 and so on, then the matrix of transition probabilities will have the form shown in Fig. 13.2. Here the submatrices marked by zeros have only zero entries. The cross-hatched submatrices are stochastic.

Each such submatrix corresponds to some irreducible chain. If, at some time, the system is at a state of such an irreducible chain, then the system will never leave this chain in the future. Hence, to study the dynamics of an arbitrary Markov chain, it is sufficient to study the dynamics of irreducible chains. Therefore one of the basic objects of study in the theory of Markov chains is *irreducible Markov chains*. We will consider them now.

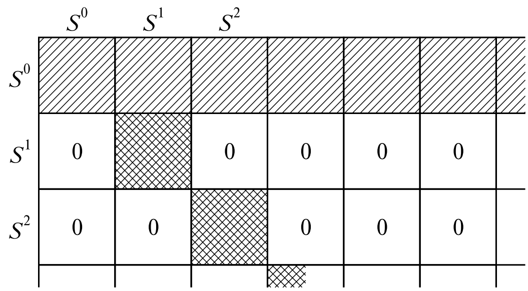

Fig. 13.2 The structure of the matrix of transition probabilities of a general Markov chain. The class S^0 consists of all inessential states, whereas $S^1, S^2, \ldots$ are closed classes of communicating states

We introduce the following notation:

$$f_j(n) := \mathbf{P}(X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j \mid X_0 = j), \qquad F_j := \sum_{n=1}^{\infty} f_j(n);$$

$f_j(n)$ is the probability that the system leaving the j-th state will return to it for the first time after n steps. The probability that the system leaving the j-th state will eventually return to it is equal to F_j.

Definition 13.1.6

K3. A state E_j is said to be *recurrent* (or *persistent*) if $F_j = 1$, and *transient* if $F_j < 1$.
K4. A state E_j is called *null* if $p_{jj}(n) \to 0$ as $n \to \infty$, and *positive* otherwise.
K5. A state E_j is called *periodic* with period d_j if the recurrence with this state has a positive probability only when the number of steps is a multiple of $d_j > 1$, and d_j is the maximum number having such property.

In other words, $d_j > 1$ is the greatest common divisor (g.c.d.) of the set of numbers $\{n : f_j(n) > 0\}$. Note that one can always choose from this set a finite subset $\{n_1, \ldots, n_k\}$ such that d_j is the greatest common divisor of these numbers. It is also clear that $p_{jj}(n) = f_j(n) = 0$ if $n \neq 0 \pmod{d_j}$.

Example 13.1.4 Consider a walk of a particle over integer points on the real line defined as follows. The particle either takes one step to the right or remains on the spot with probabilities 1/2. Here $f_j(1) = 1/2$, and if $n > 1$ then $f_j(n) = 0$ for any point j. Therefore $F_j < 1$ and all the states are transient. It is easily seen that $p_{jj}(n) = 1/2^n \to 0$ as $n \to \infty$ and hence every state is null.

On the other hand, if the particle jumps to the right with probability 1/2 and with the same probability jumps to the left, then we have a chain with period 2, since recurrence to any particular state is only possible in an even number of steps.

13.2 Necessary and Sufficient Conditions for Recurrence of States. Types of States in an Irreducible Chain. The Structure of a Periodic Chain

Recall that the function

$$a(z) = \sum_{n=0}^{\infty} a_n z^n$$

is called the generating function of the sequence $\{a_n\}_{n=0}^{\infty}$. Here z is a complex variable. If the sequence $\{a_n\}$ is bounded, then this series converges for $|z| < 1$.

Theorem 13.2.1 *A state E_j is recurrent if and only if $P_j = \sum_{n=1}^{\infty} p_{jj}(n) = \infty$. For a transient E_j,*

$$F_j = \frac{P_j}{1 + P_j}. \tag{13.2.1}$$

The assertion of this theorem is a kind of expansion of the Borel–Cantelli lemma to the case of dependent events $A_n = \{X_n = j\}$. With probability 1 there occur infinitely many events A_n if and only if

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = P_j = \infty.$$

Proof By the total probability formula we have

$$p_{jj}(n) = f_j(1)p_{jj}(n-1) + f_j(2)p_{jj}(n-2) + \cdots + f_j(n-1)p_{jj}(1) + f_j(n) \cdot 1.$$

Introduce the generating functions of the sequences $\{p_{jj}(n)\}_{n=0}^{\infty}$ and $\{f_j(n)\}_{n=0}^{\infty}$:

$$P_j(z) := \sum_{n=1}^{\infty} p_{jj}(n) z^n, \qquad F_j(z) := \sum_{n=1}^{\infty} f_j(n) z^n.$$

Both series converge inside the unit circle and represent analytic functions. The above formula for $p_{jj}(n)$, after multiplying both sides by z^n and summing up over n, leads (by the rule of convolution) to the equality

$$P_j(z) = z f_1(1)\big(1 + P_j(z)\big) + z^2 f_1(2)\big(1 + P_j(z)\big) + \cdots = \big(1 + P_j(z)\big) F_j(z).$$

Thus

$$F_j(z) = \frac{P_j(z)}{1 + P_j(z)}, \qquad P_j(z) = \frac{F_j(z)}{1 + F_j(z)}.$$

Assume that $P_j = \infty$. Then $P_j(z) \to \infty$ as $z \uparrow 1$ and therefore $F_j(z) \to 1$. Since $F_j(z) < F_j$ for real $z < 1$, we have $F_j = 1$ and hence E_j is recurrent.

Now suppose that $F_j = 1$. Then $F_j(z) \to 1$ as $z \uparrow 1$, and so $P_j(z) \to \infty$. Therefore $P_j(z) = \infty$.

If E_j is transient, it follows from the above that $P_j(z) < \infty$, and setting $z := 1$ we obtain equality (13.2.1). □

The quantity $P_j = \sum_{n=1}^{\infty} p_{jj}(n)$ can be interpreted as the mean number of visits to the state E_j, provided that the initial state is also E_j. It follows from the fact that the number of visits to the state E_j can be represented as $\sum_{n=1}^{\infty} I(X_n = j)$, where, as before, $I(A)$ is the indicator of the event A. Therefore the expectation of this number is equal to

$$\mathbf{E}\sum_{n=1}^{\infty} I(X_n = j) = \sum_{n=1}^{\infty} \mathbf{E}I(X_n = j) = \sum_{n=1}^{\infty} p_{jj}(n) = P_j.$$

Theorem 13.2.1 implies the following result.

Corollary 13.2.1 *A transient state is always null.*

This is obvious, since it immediately follows from the convergence of the series $\sum p_{jj}(n) < \infty$ that $p_{jj}(n) \to 0$.

Thus, based on definitions K3–K5, we could distinguish, in an irreducible chain, 8 possible types of states (each of the three properties can either be present or not). But in reality there are only 6 possible types since transient states are automatically null, and positive states are recurrent. These six types are generated by:

1) Classification by the asymptotic properties of the probabilities $p_{jj}(n)$ (transient, recurrent null and positive states).

2) Classification by the arithmetic properties of the probabilities $p_{jj}(n)$ or $f_j(n)$ (periodic or aperiodic).

Theorem 13.2.2 (Solidarity Theorem) *In an irreducible homogeneous Markov chain all states are of the same type*: *if one is recurrent then all are recurrent*, *if one is null then all are null*, *if one state is periodic with period d then all states are periodic with the same period d.*

Proof Let E_k and E_j be two different states. There exist numbers N and M such that

$$p_{kj}(N) > 0, \quad p_{jk}(M) > 0.$$

The total probability formula

$$p_{kk}(N + M + n) = \sum_{l,s} p_{kl}(N) p_{ls}(n) p_{sk}(M)$$

implies the inequality

$$p_{kk}(N + M + n) \ge p_{kj}(N) p_{jj}(n) p_{jk}(M) = \alpha\beta p_{jj}(n).$$

Here $n > 0$ is an arbitrary integer, $\alpha = p_{jj}(N) > 0$, and $\beta = p_{jj}(M) > 0$. In the same way one can obtain the inequality

$$p_{jj}(N + M + n) \geq \alpha\beta p_{kk}(n).$$

Hence

$$\frac{1}{\alpha\beta} p_{kk}(N + M + n) \geq p_{kk}(n) \geq \alpha\beta p_{kk}(n - M - N). \tag{13.2.2}$$

We see from these inequalities that the asymptotic properties of $p_{kk}(n)$ and $p_{jj}(n)$ are the same. If E_k is null, then $p_{kk}(n) \to 0$, therefore $p_{jj}(n) \to 0$ and E_j is also null. If E_k is recurrent or, which is equivalent, $P_k = \sum_{n=1}^{\infty} p_{kk}(n) = \infty$, then

$$\sum_{n=M+N+1}^{\infty} p_{jj}(n) \geq \alpha\beta \sum_{n=M+N+1}^{\infty} p_{kk}(n - M - N) = \infty,$$

and E_j is also recurrent.

Suppose now that E_k is a periodic state with period d_k. If $p_{kk}(n) > 0$, then d_k divides n. We will write this as $d_k \mid n$. Since $p_{kk}(M + N) \geq \alpha\beta > 0$, then $d_k \mid (M + N)$.

We now show that the state E_j is also periodic and its period d_j is equal to d_k. Indeed, if $p_{jj}(n) > 0$ for some n, then by virtue of (13.2.2), $p_{kk}(n + M + N) > 0$. Therefore $d_k \mid (n + M + N)$, and since $d_k \mid (M + N)$, $d_k \mid n$ and hence $d_k \leq d_j$. In a similar way one can prove that $d_j \leq d_k$. Thus $d_j = d_k$. □

If the states of an irreducible Markov chain are periodic with period $d > 1$, then the chain is called *periodic*.

We will now show that the study of periodic chains can essentially be reduced to the study of aperiodic chains.

Theorem 13.2.3 *If a Markov chain is periodic with period d, then the set of states can be split into d subclasses $\Psi_0, \Psi_1, \ldots, \Psi_{d-1}$ such that, with probability 1, in one step the system passes from Ψ_k to Ψ_{k+1}, and from Ψ_{d-1} the system passes to Ψ_0.*

Proof Choose some state, say, E_1. Based on this we will construct the subclasses $\Psi_0, \Psi_1, \ldots, \Psi_{d-1}$ in the following way: $E_i \in \Psi_\alpha$, $0 \leq \alpha \leq d - 1$, if there exists an integer $k > 0$ such that $p_{1i}(kd + \alpha) > 0$.

We show that no state can belong to two subclasses simultaneously. To this end it suffices to prove that if $E_i \in \Psi_\alpha$ and $p_{1i}(s) > 0$ for some s, then $s = \alpha \pmod d$.

Indeed, there exists a number $t_1 > 0$ such that $p_{i1}(t_1) > 0$. So, by the definition of Ψ_α, we have $p_{11}(kd + \alpha + t_1) > 0$. Moreover, $p_{11}(s + t_1) > 0$. Hence $d \mid (kd + \alpha + t_1)$ and $d \mid (s + t_1)$. This implies $\alpha = s \pmod d$.

Since starting from the state E_1 it is possible with positive probability to enter any state E_i, the union $\bigcup_\alpha \Psi_\alpha$ contains all the states.

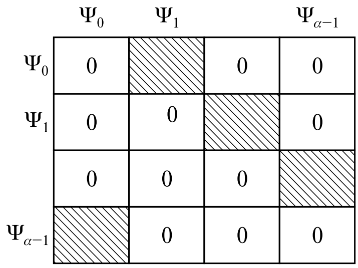

Fig. 13.3 The structure of the matrix of transition probabilities of a periodic Markov chain: an illustration to the proof of Theorem 13.2.3

We now prove that in one step the system goes from Ψ_α with probability 1 to $\Psi_{\alpha+1}$ (here the sum $\alpha + 1$ is modulo d). We have to show that, for $E_i \in \Psi_\alpha$,

$$\sum_{E_j \in \Psi_{\alpha+1}} p_{ij} = 1.$$

To do this, it suffices to prove that $p_{ij} = 0$ when $E_i \in \Psi_\alpha$, $E_j \notin \Psi_{\alpha+1}$. If we assume the opposite ($p_{ij} > 0$) then, taking into account the inequality $p_{1i}(kd + \alpha) > 0$, we have $p_{1j}(kd + \alpha + 1) > 0$ and consequently $E_j \in \Psi_{\alpha+1}$. This contradiction completes the proof of the theorem. □

We see from the theorem that the matrix of a periodic chain has the form shown in Fig. 13.3 where non-zero entries can only be in the shaded cells.

From a periodic Markov chain with period d one can construct d new Markov chains. The states from the subset Ψ_α will be the states of the α-th chain. Transition probabilities are given by

$$p_{ij}^{\alpha} := p_{ij}(d).$$

By virtue of Theorem 13.2.3, $\sum_{E_j \in \Psi_\alpha} p_{ij}^{\alpha} = 1$. The new chains, to which one can reduce in a certain sense the original one, will have no subclasses.

13.3 Theorems on Random Walks on a Lattice

1. A random walk on integer points on the line. Imagine a particle moving on integer points of the real line. Transitions from one point to another occur in equal time intervals. In one step, from point k the particle goes with a positive probability p to the point $k + 1$, and with positive probability $q = 1 - p$ it moves to the point $k - 1$. As was already mentioned, to this physical system there corresponds the following Markov chain:

$$X_n = X_{n-1} + \xi_n = X_0 + S_n,$$

where ξ_n takes values 1 and -1 with probabilities p and q, respectively, and $S_n = \sum_{k=1}^{n} \xi_k$. The states of the chain are integer points on the line.

It is easy to see that returning to a given point with a positive probability is only possible after an even number of steps, and $f_0(2) = 2pq > 0$. Therefore this chain is periodic with period 2.

We now establish conditions under which the random walk forms a recurrent chain.

Theorem 13.3.1 *The random walk $\{X_n\}$ forms a recurrent Markov chain if and only if $p = q = 1/2$.*

Proof Since $0 < p < 1$, the random walk is an irreducible Markov chain. Therefore by Theorem 13.2.2 it suffices to examine the type of any given point, for example, zero.

We will make use of Theorem 13.2.1. In order to do this, we have to investigate the convergence of the series $\sum_{n=1}^{\infty} p_{00}(n)$. Since our chain is periodic with period 2, one has $p_{00}(2k+1) = 0$. So it remains to compute $\sum_1^{\infty} p_{00}(2k)$. The sum S_n is the coordinate of the walking particle after n steps ($X_0 = 0$). Therefore $p_{00}(2k) = \mathbf{P}(S_{2k} = 0)$. The equality $S_{2k} = 0$ holds if k of the random variables ξ_j are equal to 1 and the other k are equal to -1 (k steps to the right and k steps to the left). Therefore, by Theorem 5.2.1,

$$\mathbf{P}(S_{2k} = 0) \sim \frac{1}{\sqrt{\pi k}} e^{-2kH(1/2)} = \frac{1}{\sqrt{\pi k}} (4pq)^k.$$

We now elucidate the behaviour of the function $\beta(p) = 4pq = 4p(1-p)$ on the interval $[0, 1]$. At the point $p = 1/2$ the function $\beta(p)$ attains its only extremum, $\beta(1/2) = 1$. At all the other points of $[0, 1]$, $\beta(p) < 1$. Therefore $4pq < 1$ for $p \neq 1/2$, which implies convergence of the series $\sum_{k=1}^{\infty} p_{00}(2k)$ and hence the transience of the Markov chain. But if $p = 1/2$ then $p_{00}(2k) \sim 1/\sqrt{\pi k}$ and the series $\sum_{k=1}^{\infty} p_{00}(2k)$ diverges, which implies, in turn, that all the states of the chain are recurrent. The theorem is proved. □

Theorem 13.3.1 allows us to make the following remark. If $p \neq 1/2$, then the mean number of recurrences to 0 is finite, as it is equal to $\sum_{k=1}^{\infty} p_{00}(2k)$. This means that, after a certain time, the particle will never return to zero. The particle will "drift" to the right or to the left depending on whether p is greater than $1/2$ or less. This can easily be obtained from the law of large numbers.

If $p = 1/2$, then the mean number of recurrences to 0 is infinite; the particle has no "drift". It is interesting to note that the increase in the mean number of recurrences is not proportional to the number of steps. Indeed, the mean number of recurrences over the first $2n$ steps is equal to $\sum_{k=1}^{n} p_{00}(2k)$. From the proof of Theorem 13.3.1 we know that $p_{00}(2k) \sim 1/\sqrt{\pi k}$. Therefore, as $n \to \infty$,

$$\sum_{k=1}^{n} p_{00}(2k) \sim \sum_{k=1}^{n} \frac{1}{\sqrt{\pi k}} \sim \frac{2\sqrt{n}}{\sqrt{\pi}}.$$

Thus, in the fair game considered in Example 4.2.2, the proportion of ties rapidly decreases as the number of steps increases, and deviations are growing both in magnitude and duration.

13.3.1 Symmetric Random Walks in $\mathbb{R}^k$, $k \ge 2$

Consider the following random walk model in the k-dimensional Euclidean space $\mathbb{R}^k$. If the walking particle is at point $(m_1, \ldots, m_k)$, then it can move with probabilities $1/2^k$ to any of the 2^k vertices of the cube $|x_j - m_j| = 1$, i.e. the points with coordinates $(m_1 \pm 1, \ldots, m_k \pm 1)$. It is natural to call this walk symmetric. Denoting by X_n the position of the particle after the n-th jump, we have, as before, a sequence of k-dimensional random variables forming a homogeneous irreducible Markov chain. We shall show that all states of the walk on the plane are, as in the one-dimensional case, recurrent. In the three-dimensional space, the states will turn out to be transient. Thus we shall prove the following assertion.

Theorem 13.3.2 *The symmetric random walk is recurrent in spaces of one and two dimensions and transient in spaces of three or more dimensions.*

In this context, W. Feller made the sharp comment that the proverb "all roads lead to Rome" is true only for two-dimensional surfaces. The assertion of Theorem 13.3.2 is adjacent to the famous theorem of Pólya on the transience of symmetric walks in $\mathbb{R}^k$ for $k > 2$ when the particle jumps to neighbouring points along the coordinate axes (so that ξ_j assumes $2k$ values with probabilities $1/2k$ each). We now turn to the proof of Theorem 13.3.2.

Proof of Theorem 13.3.2 Let $k = 2$. It is not difficult to see that our walk X_n can be represented as a sum of two independent components

$$X_n = \left(X_n^{-1}, 0\right) + \left(0, X_n^2\right), \quad \left(X_0^1, X_0^2\right) = X_0,$$

where X_n^i, $i = 1, 2, \ldots$, are scalar (one-dimensional) sequences describing symmetric independent random walks on the respective lines (axes). This is obvious, for the two-dimensional sequence admits the representation

$$X_{n+1} = X_n + \xi_n, \tag{13.3.1}$$

where ξ_n assumes 4 values $(\pm 1, 0) + (0, \pm 1) = (\pm 1, \pm 1)$ with probabilities $1/4$ each.

With the help of representation (13.3.1) we can investigate the asymptotic behaviour of the transition probabilities $p_{ij}(n)$. Let X_0 coincide with the origin $(0, 0)$. Then

$$p_{00}(2n) = \mathbf{P}\big(X_{2n} = (0, 0) | X_0 = (0, 0)\big)$$

$$= \mathbf{P}\big(X^1_{2n} = 0 | X^1_0 = 0\big)\mathbf{P}\big(X^2_{2n} = 0 | X^2_0 = 0\big) \sim (1/\sqrt{\pi n})^2 = 1/(\pi n).$$

From this it follows that the series $\sum_{n=0}^{\infty} p_{00}(n)$ diverges and so all the states of our chain are recurrent.

The case $k = 3$ should be treated in a similar way. Represent the sequence X_n as a sum of three independent components

$$X_n = \big(X^1_n, 0, 0\big) + \big(0, X^2_n, 0\big) + \big(0, 0, X^3_n\big),$$

where the X^i_n are, as before, symmetric random walks on the real line. If we set $X_0 = (0, 0, 0)$, then

$$p_{00}(2n) = \big(\mathbf{P}\big(X^1_{2n} = 0 | \; X^1_0 = 0\big)\big)^3 \sim 1/(\pi n)^{3/2}.$$

The series $\sum_{n=1}^{\infty} p_{00}(n)$ is convergent here, and hence the states of the chain are transient. In contrast to the straight line and plane cases, a particle leaving the origin will, with a positive probability, never come back.

It is evident that a similar situation takes place for walks in k-dimensional space with $k \geq 3$, since $\sum_{n=1}^{\infty} (\pi n)^{-k/2} < \infty$ for $k \geq 3$. The theorem is proved. □

13.3.2 Arbitrary Symmetric Random Walks on the Line

Let, as before,

$$X_n = X_0 + \sum_1^n \xi_j, \tag{13.3.2}$$

but now ξ_j are arbitrary independent identically distributed integer-valued random variables. Theorem 13.3.1 may be generalised in the following way:

Theorem 13.3.3 *If the ξ_j are symmetric and the expectation $\mathbf{E}\xi_j$ exists (and hence $\mathbf{E}\xi_j = 0$) then the random walk X_n forms a recurrent Markov chain with null states.*

Proof It suffices to verify that

$$\sum_{n=1}^{\infty} \mathbf{P}(S_n = 0) = \infty,$$

where $S_n = \sum_1^n \xi_j$, and that $\mathbf{P}(S_n = 0) \to 0$ as $n \to \infty$. Put

$$p(z) := \mathbf{E}z^{\xi_1} = \sum_{k=-\infty}^{\infty} z^k \mathbf{P}(\xi_1 = k).$$

Then the generating function of S_n will be equal to $\mathbf{E}z^{S_n} = p^n(z)$, and by the inversion formula (see Sect. 7.7)

$$\mathbf{P}(S_n = 0) = \frac{1}{2\pi i} \int_{|z|=1} p^n z^{-1} dz, \qquad (13.3.3)$$

$$\sum_{n=0}^{\infty} \mathbf{P}(S_n = 0) = \frac{1}{2\pi i} \int_{|z|=1} \frac{dz}{z(1-p(z))} = \frac{1}{\pi} \int_0^{\pi} \frac{dt}{1-p(e^{it})}.$$

The last equality holds since the real function $p(r)$ is even and is obtained by substituting $z = e^{it}$.

Since $\mathbf{E}\xi_1 = 0$, one has $1 - p(e^{it}) = o(t)$ as $t \to 0$ and, for sufficiently small δ and $0 \le t < \delta$,

$$0 \le 1 - p\big(e^{it}\big) < t$$

(the function $p(e^{it})$ is real by virtue of the symmetry of ξ_1). This implies

$$\int_0^{\pi} \frac{dt}{1-p(e^{it})} \ge \int_0^{\delta} \frac{dt}{t} = \infty.$$

Convergence $\mathbf{P}(S_n = 0) \to 0$ is a consequence of (13.3.3) since, for all z on the circle $|z| = 1$, with the possible exclusion of finitely many points, one has $p(z) < 1$ and hence $p^n(z) \to 0$ as $n \to \infty$. The theorem is proved. □

Theorem 13.3.3 can be supplemented by the following assertion.

Theorem 13.3.4 *Under the conditions of Theorem 13.3.3, if the g.c.d. of the possible values of ξ_j equals 1 then the set of values of $\{X_n\}$ constitutes a single class of essential communicating states. This class coincides with the set of all integers.*

The assertion of the theorem follows from the next lemma.

Lemma 13.3.1 *If the g.c.d. of integers $a_1 > 0, \dots, a_r > 0$ is equal to 1, then there exists a number K such that every natural $k \ge K$ can be represented as*

$$k = n_1 a_1 + \cdots + n_r a_r,$$

where $n_i \ge 0$ are some integers.

Proof Consider the function $L(\mathbf{n}) = n_1 a_1 + \cdots + n_r a_r$, where $\mathbf{n} = (n_1, \dots, n_r)$ is a vector with integer (possibly negative) components. Let $d > 0$ be the minimal natural number for which there exists a vector $\mathbf{n}^0$ such that

$$d = L\big(\mathbf{n}^0\big).$$

We show that every natural number that can be represented as $L(\mathbf{n})$ is divisible by d. Suppose that this is not true. Then there exist $\mathbf{n}$, k and $0 < \alpha < d$ such that

$$L(\mathbf{n}) = kd + \alpha.$$

But since the function $L(\mathbf{n})$ is linear,

$$L\big(\mathbf{n} - kx^0\big) = kd + \alpha - kd = \alpha < d,$$

which contradicts the minimality of d in the set of positive integer values of $L(\mathbf{n})$.

The numbers $a_1, \dots, a_r$ are also the values of the function $L(\mathbf{n})$, so they are divisible by d. The greatest common divisor of these numbers is by assumption equal to one, so that $d = 1$.

Let k be an arbitrary natural number. Denoting by $\theta < A$ the remainder after dividing k by $A := a_1 + \cdots + a_r$, we can write

$$\begin{aligned} k &= m(a_1 + \cdots + a_r) + \theta = m(a_1 + \cdots + a_r) + \theta L\big(\mathbf{n}^0\big) \\ &= a_1\big(m + \theta n_1^0\big) + a_2\big(m + \theta n_2^0\big) + \cdots + a_r\big(m + \theta n_r^0\big), \end{aligned}$$

where $n_i := m + \theta n_i^0 > 0$, $i = 1, \dots, r$, for sufficiently large k (or m).

The lemma is proved. □

Proof of Theorem 13.3.4 Put $q_j := \mathbf{P}(\xi = a_j) > 0$. Then, for each $k \ge K$, there exists an $\mathbf{n}$ such that $n_j \ge 0$, $\sum_{j=1}^r a_j n_j = k$, and hence, for $n = \sum_{j=1}^r n_j$, we have

$$p_{0k}(n) \ge q_1^{n_1} \cdots q_r^{n_r} > 0.$$

In other words, all the states $k \ge K$ are reachable from 0. Similarly, all the states $k \le -K$ are reachable from 0. The states $k \in [-K, K]$ are reachable from the point $-2K$ (which is reachable from 0). The theorem is proved. □

Corollary 13.3.1 *If the conditions of Theorems 13.3.3 and 13.3.4 are satisfied, then the chain (13.3.2) with an arbitrary initial state X_0 visits every state k infinitely many times with probability 1. In particular, for any X_0 and k, the random variable $\nu = \min\{n : X_n = k\}$ will be proper.*

If we are interested in investigating the periodicity of the chain (13.3.2), then more detailed information on the set of possible values of ξ_j is needed. We leave it to the reader to verify that, for example, if this set is of the form $\{a + a_k d\}$, $k = 1, 2, \dots, d \ge 1$, g.c.d. $(a_1, a_2, \dots) = 1$, g.c.d. $(a, d) = 1$, then the chain will be periodic with period d.

13.4 Limit Theorems for Countable Homogeneous Chains

13.4.1 Ergodic Theorems

Now we return to arbitrary countable homogeneous Markov chains. We will need the following conditions:

(I) There exists a state E_0 such that the recurrence time $\tau^{(s)}$ to E_s ($\mathbf{P}(\tau^{(s)} = n) = f_s(n)$) has finite expectation $\mathbf{E}\tau^{(s)} < \infty$.
(II) The chain is irreducible.
(III) The chain is aperiodic.

We introduce the so-called "taboo probabilities" $P_i(n, j)$ of transition from E_i to E_j in n steps without visiting the "forbidden" state E_i:

$$P_i(n, j) := \mathbf{P}(X_n = j; X_1 \neq i, \dots, X_{n-1} \neq i \mid X_0 = i).$$

Theorem 13.4.1 (The ergodic theorem) *Conditions* (I)–(III) *are necessary and sufficient for the existence, for all i and j, of the positive limits*

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j > 0, \quad i, j = 0, 1, 2, \dots. \tag{13.4.1}$$

The sequence of values $\{\pi_j\}$ is the unique solution of the system

$$\begin{cases} \sum_{j=0}^{\infty} \pi_j = 1, \\ \pi_j = \sum_{k=0}^{\infty} \pi_k p_{kj}, \quad j = 0, 1, 2, \dots, \end{cases} \tag{13.4.2}$$

in the class of absolutely convergent series.

Moreover, $\mathbf{E}\tau^{(j)} < \infty$ for all j, and the quantities $\pi_j = (\mathbf{E}\tau^{(j)})^{-1}$ admit the representation

$$\pi_j = \left(\mathbf{E}\tau^{(j)}\right)^{-1} = \left(\mathbf{E}\tau^{(s)}\right)^{-1} \sum_{k=1}^{\infty} P_s(k, j) \tag{13.4.3}$$

for any s.

Definition 13.4.1 A chain possessing property (13.4.1) is called *ergodic*.

The numbers π_j are essentially the probabilities that the system will be in the respective states E_j after a long period of time has passed. It turns out that these probabilities lose dependence on the initial state of the system. The system "forgets" where it began its motion. The distribution $\{\pi_j\}$ is called *stationary* or *invariant*. Property (13.4.2) expresses the invariance of the distribution with respect to the transition probabilities p_{ij}. In other words, if $\mathbf{P}(X_n = k) = \pi_k$, then $\mathbf{P}(X_{n+1} = k) = \sum \pi_j p_{jk}$ is also equal to π_k.

Proof of Theorem 13.4.1 *Sufficiency* in the first assertion of the theorem. Consider the "trajectory" of the Markov chain starting at a fixed state E_s. Let $\tau_1 \ge 1$, $\tau_2 \ge 1$, ... be the time intervals between successive returns of the system to E_s. Since after each return the evolution of the system begins anew from the same state, by the Markov property the durations τ_k of the cycles (as well as the cycles themselves) are independent and identically distributed, $\tau_k \stackrel{d}{=} \tau^{(s)}$. Moreover, it is obvious that

$$\mathbf{P}(\tau_k = n) = \mathbf{P}\big(\tau^{(s)} = n\big) = f_s(n).$$

Recurrence of E_s means that the τ_k are proper random variables. Aperiodicity of E_s means that the g.c.d. of all possible values of τ_k is equal to 1. Since

$$p_{ss}(n) = \mathbf{P}\big(\gamma(n) = 0\big),$$

where $\gamma(n)$ is the defect of level n for the renewal process $\{T_k\}$,

$$T_k = \sum_{i=1}^{k} \tau_i,$$

by Theorem 10.3.1 the following limit exists

$$\lim_{n\to\infty} p_{ss}(n) = \lim_{n\to\infty} \mathbf{P}\big(\gamma(n) = 0\big) = \frac{1}{\mathbf{E}\tau_1} > 0. \qquad (13.4.4)$$

Now prove the existence of $\lim_{n\to\infty} p_{sj}(n)$ for $j \ne s$. If $\gamma(n)$ is the defect of level n for the walk $\{T_k\}$ then, by the total probability formula,

$$p_{sj}(n) = \sum_{k=1}^{n} \mathbf{P}\big(\gamma(n) = k\big)\mathbf{P}\big(X_n = j \,|\, X_0 = s, \gamma(n) = k\big). \qquad (13.4.5)$$

Note that the second factors in the terms on the right-hand side of this formula do not depend on n by the Markov property:

$$\begin{aligned}
&\mathbf{P}\big(X_n = j \,|\, X_0 = s, \gamma(n) = k\big) \\
&\qquad = \mathbf{P}(X_n = j \,|\, X_0 = s,\, X_{n-1} \ne s, \ldots, X_{n-k+1} \ne s,\, X_{n-k} = s) \\
&\qquad = \mathbf{P}(X_k = j \,|\, X_0 = s,\, X_1 \ne s, \ldots, X_{k-1} \ne s) = \frac{P_s(k, j)}{\mathbf{P}(\tau_1 \ge k)},
\end{aligned} \qquad (13.4.6)$$

since, for a fixed $X_0 = s$,

$$\begin{aligned}
\mathbf{P}(X_k = j \,|\, X_1 \ne s, \ldots, X_{k-1} \ne s) &= \frac{\mathbf{P}(X_k = j, X_1 \ne s, \ldots, X_{k-1} \ne s)}{\mathbf{P}(X_1 \ne s, \ldots, X_{k-1} \ne s)} \\
&= \frac{P_s(k, j)}{\mathbf{P}(\tau^{(s)} \ge k)}.
\end{aligned}$$

For the sake of brevity, put $\mathbf{P}(\tau_1 > k) = P_k$. The first factors in (13.4.5) converge, as $n \to \infty$, to $P_{k-1}/\mathbf{E}\tau_1$ and, by virtue of the equality

$$\mathbf{P}\big(\gamma(n) = k\big) = \mathbf{P}\big(\gamma(n-k) = 0\big) P_{k-1} \le P_{k-1}, \tag{13.4.7}$$

are dominated by the convergent sequence P_{k-1}. Therefore, by the dominated convergence theorem, the following limit exists

$$\lim_{n\to\infty} p_{sj}(n) = \sum_{k=1}^{\infty} \frac{P_{k-1}}{\mathbf{E}\tau_1} \frac{P_s(k,j)}{\mathbf{P}(\tau_1 \ge k)} = \frac{1}{\mathbf{E}\tau_1} \sum_{k=1}^{\infty} P_s(k,j) =: \pi_j, \tag{13.4.8}$$

and we have, by (13.4.5)–(13.4.7),

$$p_{sj}(n) \le \sum_{k=1}^{n} P_s(k,j) \le \sum_{k=1}^{\infty} P_s(k,j) = \pi_j \mathbf{E}\tau_1. \tag{13.4.9}$$

To establish that, for any i,

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j > 0,$$

we first show that the system departing from E_i will, with probability 1, eventually reach E_s.

In other words, if $f_{is}(n)$ is the probability that the system, upon leaving E_i, hits E_s for the first time on the n-th step then

$$\sum_{n=1}^{\infty} f_{is}(n) = 1.$$

Indeed, both states E_i and E_s are recurrent. Consider the cycles formed by subsequent visits of the system to the state E_i. Denote by A_k the event that the system is in the state E_s at least once during the k-th cycle. By the Markov property the events A_k are independent and $\mathbf{P}(A_k) > 0$ does not depend on k. Therefore, by the Borel–Cantelli zero–one law (see Sect. 11.1), with probability 1 there will occur infinitely many events A_k and hence $\mathbf{P}(\bigcup A_k) = 1$.

By the total probability formula,

$$p_{ij}(n) = \sum_{k=1}^{n} f_{is}(k) p_{sj}(n-k),$$

and the dominated convergence theorem yields

$$\lim_{n\to\infty} p_{ij}(n) = \sum_{n=1}^{\infty} f_{is}(k)\pi_j = \pi_j.$$

Representation (13.4.3) follows from (13.4.8).

Now we will prove the *necessity* in the first assertion of the theorem. That conditions (II)–(III) are necessary is obvious, since $p_{ij}(n) > 0$ for every i and j if n is large enough. The necessity of condition (I) follows from the fact that equalities (13.4.4) are valid for E_s. The first part of the theorem is proved.

It remains to prove the second part of the theorem. Since

$$\sum p_{sj}(n) = 1,$$

one has $\sum_j \pi_j \le 1$. By virtue of the inequalities $p_{sj}(n) \le \pi_j \mathbf{E}\tau_1$ (see (13.4.9)), we can use the dominated convergence theorem both in the last equality and in the equality $p_{sj}(n+1) = \sum_{k=0}^{\infty} p_{sk}(n) p_{kj}$ which yields

$$\sum \pi_j = 1, \quad \pi_j = \sum_{k=0}^{\infty} \pi_k p_{kj}.$$

It remains to show that the system has a unique solution. Let the numbers $\{q_j\}$ also satisfy (13.4.2) and assume the series $\sum |q_j|$ converges. Then, changing the order of summation, we obtain that

$$\begin{aligned} q_j &= \sum_k q_k p_{kj} = \sum_k p_{kj} \left(\sum_l p_{lk} q_l \right) = \sum_l q_l \sum_k p_{lk} p_{kj} = \sum_l q_l p_{lj}(2) \\ &= \sum_l p_{lj}(2) \left(\sum_m p_{ml} q_m \right) = \sum_m q_m p_{mj}(3) = \cdots = \sum_k q_k p_{kj}(n) \end{aligned}$$

for any n. Since $\sum q_k = 1$, passing to the limit as $n \to \infty$ gives

$$q_j = \sum_k q_k \pi_j = \pi_j.$$

The theorem is proved. □

If a Markov chain is periodic with period d, then $p_{ij}(t) = 0$ for $t \ne kd$ and every pair of states E_i and E_j belonging to the same subclass (see Theorem 13.2.3). But if $t = kd$, then from the theorem just proved and Theorem 13.2.3 it follows that the limit $\lim_{k\to\infty} p_{ij}(kd) = \pi_j > 0$ exists and does not depend on i.

Verifying conditions (II)–(III) of Theorem 13.4.1 usually presents no serious difficulties. The main difficulties would be related to verifying condition (I). For finite Markov chains, this condition is always met.

Theorem 13.4.2 *Let a Markov chain have finitely many states and satisfy conditions* (II)–(III). *Then there exist $c > 0$ and $q < 1$ such that*, *for the recurrence time τ to an arbitrary fixed state*, *one has*

$$\mathbf{P}(\tau > n) < cq^n, \quad n \ge 1. \tag{13.4.10}$$

These equalities clearly mean that *condition* (I) *is always met for finite chains and hence the ergodic theorem for them holds if and only if conditions* (II)–(III) *are satisfied*.

Proof Consider a state E_s and put

$$r_j(n) := \mathbf{P}(X_k \neq s, k = 1, 2, \dots, n | X_0 = j).$$

Then, if the chain has m states one has $r_j(m) < 1$ for any j. Indeed, $r_j(n)$ does not grow as n increases. Let N be the smallest number satisfying $r_j(N) < 1$. This means that there exists a sequence of states $E_j, E_{j_1}, \dots, E_{j_N}$ such that $E_{j_N} = E_s$ and the probability of this sequence $p_{jj_1} \cdots p_{j_{N-1}j_N}$ is positive. But it is easy to see that $N \le m$, since otherwise this sequence would contain at least two identical states. Therefore the cycle contained between these states could be removed from the sequence which could only increase its probability. Thus

$$r_j(m) < 1, \quad r(m) = \max_j r_j(m) < 1.$$

Moreover, $r_j(n_1 + n_2) \le r_j(n_1) r(n_2) \le r(n_1) r(n_2)$.

It remains to note that if τ is the recurrence time to E_s, then $\mathbf{P}(\tau > nm) = r_s(nm) \le r(m)^n$. The statement of the theorem follows. □

Remark 13.4.1 Condition (13.4.10) implies the *exponential* rate of convergence of the differences $|p_{ij}(n) - \pi_j|$ to zero. One can verify this by making use of the analyticity of the function

$$F_s(z) = \sum_{n=1}^{\infty} f_s(n) z^n$$

in the domain $|z| < q^{-1}$, $q^{-1} > 1$, and of the equality

$$P_s(z) = \sum p_{ss}(n) z^n = \frac{1}{1 - F_s(z)} - 1 \tag{13.4.11}$$

(see Theorem 13.2.1; we assume that the τ in condition (13.4.10) refers to the state E_s, so that $f_s(n) = \mathbf{P}(\tau = n)$). Since $F_s'(1) = \mathbf{E}\tau = 1/\pi_s$, one has

$$F_s(z) = 1 + \frac{(z-1)}{\pi_s} + \cdots,$$

and from (13.4.11) it follows that the function

$$P_s(z) - \frac{z\pi_s}{1 - z} = \sum_{n=1}^{\infty} \big(p_{ss}(n) - \pi_s\big) z^n$$

is analytic in the disk $|z| \le 1 + \varepsilon$, $\varepsilon > 0$. It evidently follows from this that

$$|p_{ss}(n) - \pi_s| < c(1 + \varepsilon)^{-n}, \quad c = \text{const}.$$

Now we will give two examples of finite Markov chains.

Example 13.4.1 Suppose that the behaviour of two chess players A and B playing in a multi-player tournament can be described as follows. Independently of the outcomes of the previous games, player A wins every new game with probability p, loses with probability q, and makes a tie with probability $r = 1 - p - q$. Player B is less balanced. He wins a game with probabilities $p + \varepsilon$, p and $p - \varepsilon$, respectively, if he won, made a tie, or lost in the previous one. The probability that he loses behaves in a similar way: in the above three cases, it equals $q - \varepsilon$, q and $q + \varepsilon$, respectively. Which of the players A and B will score more points in a long tournament?

To answer this question, we will need to compute the stationary probabilities π_1, π_2, π_3 of the states E_1, E_2, E_3 which represent a win, tie, and loss in a game, respectively (cf. the law of large numbers at the end of this section).

For player A, the Markov chain with states E_1, E_2, E_3 describing his performance in the tournament will have the matrix of transition probabilities

$$P_A = \begin{pmatrix} p & r & q \\ p & r & q \\ p & r & q \end{pmatrix}.$$

It is obvious that $\pi_1 = p$, $\pi_2 = r$, $\pi_3 = q$ here.

For player B, the matrix of transition probabilities is equal to

$$P_B = \begin{pmatrix} p+\varepsilon & r & q-\varepsilon \\ p & r & q \\ p-\varepsilon & r & q+\varepsilon \end{pmatrix}.$$

Equations for stationary probabilities in this case have the form

$$\begin{aligned} \pi_1(p+\varepsilon) + \pi_2 p + \pi_3(p-\varepsilon) &= \pi_1, \\ \pi_1 r + \pi_2 r + \pi_3 r &= \pi_2, \\ \pi_1 + \pi_2 + \pi_3 &= 1. \end{aligned}$$

Solving this system we find that

$$\pi_2 - r = 0, \qquad \pi_1 - p = \varepsilon \frac{p-q}{1-2\varepsilon}.$$

Thus, the long run proportions of ties will be the same for both players, and B will have a greater proportion of wins if $\varepsilon > 0$, $p > q$ or $\varepsilon < 0$, $p < q$. If $p = q$, then the stationary distributions will be the same for both A and B.

Example 13.4.2 Consider the summation of independent integer-valued random variables $\xi_1, \xi_2, \dots$ modulo some $d > 1$ (see Example 13.1.2). Set $X_0 := 0$, $X_1 := \xi_1 - \lfloor \xi_1/d \rfloor d$, $X_2 := X_1 + \xi_2 - \lfloor (X_1 + \xi_2)/d \rfloor d$ etc. (here $\lfloor x \rfloor$ denotes the integral part of x), so that X_n is the remainder of the division of $X_{n-1}+\xi_n$ by d. Such summation is sometimes also called summation on a circle (points 0 and d are glued together in a single point). Without loss of generality, we can evidently suppose that ξ_k takes the values $0, 1, \dots, d-1$ only. If $\mathbf{P}(\xi_k = j) = p_j$ then

$$p_{ij} = \mathbf{P}(X_n = j | X_{n-1} = i) = \begin{cases} p_{j-i} & \text{if } j \ge i, \\ p_{d+j-i} & \text{if } j < i. \end{cases}$$

Assume that the set of all indices k with $p_k > 0$ has a g.c.d. equal to 1. Then it is clear that the chain $\{X_n\}$ has a single class of essential states without subclasses, and there will exist the limits

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j$$

satisfying the system $\sum_i \pi_i p_{ij} = \pi_j$, $\sum \pi_j = 1$, $j = 0, \dots, d-1$. Now note that the stochastic matrix of transition probabilities $\|p_{ij}\|$ has in this case the following property:

$$\sum_i p_{ij} = \sum_j p_{ij} = 1.$$

Such matrices are called *doubly stochastic*. Stationary distributions for them are always uniform, since $\pi_j = 1/d$ satisfy the system for final probabilities.

Thus summation of arbitrary random variables on a circle leads to the *uniform limit distribution*. The rate of convergence of $p_{ij}(k)$ to the stationary distribution is exponential.

It is not difficult to see that the convolution of two uniform distributions under addition modulo d is also uniform. The uniform distribution is in this sense stable. Moreover, the convolution of an arbitrary distribution with the uniform distribution will also be uniform. Indeed, if η is uniformly distributed and independent of ξ_1 then (addition and subtraction are modulo d, $p_j = \mathbf{P}(\xi_1 = j)$)

$$\mathbf{P}(\xi_1 + \eta = k) = \sum_{j=0}^{d-1} p_j \mathbf{P}(\eta = k - j) = \sum_{j=0}^{d-1} p_j \frac{1}{d} = \frac{1}{d}.$$

Thus, if one transmits a certain signal taking d possible values (for example, letters) and (uniform) "random" noise is superimposed on it, then the received signal will also have the uniform distribution and therefore will contain no information about the transmitted signal. This fact is widely used in cryptography.

This example also deserves attention as a simple illustration of laws that appear when summing random variables taking values not in the real line but in some group (the set of numbers $0, 1, \dots, d-1$ with addition modulo d forms a finite Abelian group). It turns out that the phenomenon discovered in the example—the uniformity of the limit distribution—holds for a much broader class of groups.

We return to arbitrary countable chains. We have already mentioned that the main difficulties when verifying the conditions of Theorem 13.4.1 are usually related to condition (I). We consider this problem in Sect. 13.7 in more detail for a wider class of chains (see Theorems 13.7.2–13.7.3 and corollaries thereafter). Sometimes condition (I) can easily be verified using the results of Chaps. 10 and 12.

Example 13.4.3 We saw in Sect. 12.5 that waiting times in the queueing system satisfy the relationships

$$X_{n+1} = \max(X_n + \xi_{n+1}, 0), \qquad w_1 = 0,$$

where the ξ_n are independent and identically distributed. Clearly, X_n form a homogeneous Markov chain with the state space $\{0, 1, \ldots\}$, provided that the ξ_k are integer-valued. The sequence X_n may be interpreted as a *walk with a delaying screen* at the point 0. If $\mathbf{E}\xi_k < 0$ then it is not hard to derive from the theorems of Chap. 10 (see also Sect. 13.7) that the recurrence time to 0 has finite expectation. Thus, applying the ergodic theorem we can, independently of Sect. 11.4, come to the conclusion that there exists a limiting (stationary) distribution for X_n as $n \to \infty$ (or, taking into account what we said in Sect. 11.4, conclude that $\sup_{k \ge 0} S_k$ is finite, where $S_k = \sum_{j=1}^{k} \xi_j$, which is essentially the assertion of Theorem 10.2.1).

Now we will make several remarks allowing us to state one more criterion for ergodicity which is related to the existence of a solution to Eq. (13.4.2).

First of all, note that Theorem 13.2.2 (the solidarity theorem) can now be complemented as follows. A state E_j is said to be *ergodic* if, for any i, $p_{ij}(n) \to \pi_j > 0$ as $n \to \infty$. A state E_j is said to be *positive recurrent* if it is recurrent and non-null (in that case, the recurrence time $\tau^{(j)}$ to E_j has finite expectation $E\tau^{(j)} < \infty$). It follows from Theorem 13.4.1 that, for an irreducible aperiodic chain, a *state E_j is ergodic if and only if it is positive recurrent. If at least one state is ergodic, all states are*.

Theorem 13.4.3 *Suppose a chain is irreducible and aperiodic* (*satisfies conditions* (II)–(III)). *Then only one of the following two alternatives can take place*: *either all the states are null or they are all ergodic. The existence of an absolutely convergent solution to system* (13.4.2) *is necessary and sufficient for the chain to be ergodic.*

Proof The first assertion of the theorem follows from the fact that, by the local renewal Theorem 10.2.2 for the random walk generated by the times of the chain's hitting the state E_j, the limit $\lim_{n\to\infty} p_{jj}(n)$ always exists and equals $(E\tau^{(j)})^{-1}$.

Therefore, to prove sufficiency in the second assertion (the necessity follows from Theorem 13.4.1) we have, in the case of the existence of an absolutely convergent solution $\{\pi_j\}$, to exclude the existence of null states. Assume the contrary, $p_{ij}(n) \to 0$. Choose j such that $\pi_j > 0$. Then

$$0 < \pi_j = \sum \pi_i p_{ij}(n) \to 0$$

as $n \to \infty$ by dominated convergence. This contradiction completes the proof of the theorem. □

13.4.2 *The Law of Large Numbers and the Central Limit Theorem for the Number of Visits to a Given State*

In conclusion of this section we will give two assertions about the limiting behaviour, as $n \to \infty$, of the number $m_j(n)$ of visits of the system to a fixed state E_j by the time n. Let $\tau^{(j)}$ be the recurrence time to the state E_j.

Theorem 13.4.4 *Let the chain be ergodic and, at the initial time epoch, be at an arbitrary state E_s. Then, as $n \to \infty$,*

$$\frac{\mathbf{E}m_j(n)}{n} \to \pi_j, \qquad \frac{m_j(n)}{n} \xrightarrow{a.s.} \pi_j.$$

If additionally $\operatorname{Var}(\tau^{(j)}) = \sigma_j^2 < \infty$ *then*

$$\mathbf{P}\left(\frac{m_j(n) - n\pi_j}{\sigma_j\sqrt{n\pi_j^3}} < x \,|\, X_0 = s\right) \to \Phi(x)$$

as $n \to \infty$, where $\Phi(x)$ is, as before, the distribution function of the normal law with parameters $(0, 1)$.

Proof Note that the sequence $m_j(n) + 1$ coincides with the renewal process formed by the random variables $\tau_1, \tau_2, \tau_3, \ldots$, where τ_1 is the time of the first visit to the state E_j by the system which starts at E_s and $\tau_k \stackrel{d}{=} \tau^{(j)}$ for $k \ge 2$. Clearly, by the Markov property all τ_j are independent. Since $\tau_1 \ge 0$ is a proper random variable, Theorem 13.4.4 is a simple consequence of the generalisations of Theorems 10.1.1, 11.5.1, and 10.5.2 that were stated in Remarks 10.1.1, 11.5.1 and 10.5.1, respectively.

The theorem is proved. □

Summarising the contents of this section, one can note that studying the sequences of dependent trials forming homogeneous Markov chains with discrete sets of states can essentially be carried out with the help of results obtained for sequences of independent random variables. Studying other types of dependent trials requires, as a rule, other approaches.

13.5* The Behaviour of Transition Probabilities for Reducible Chains

Now consider a *finite* Markov chain *of the general type*. As we saw, its state space consists of the class of inessential states S^0 and several classes $S^1, \ldots, S^l$ of essential states. To clarify the nature of the asymptotic behaviour of $p_{ij}(n)$ for such

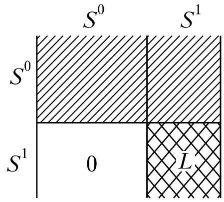

Fig. 13.4 The structure of the matrix of transition probabilities of a periodic Markov chain with the class S^0 of inessential states: an illustration to the proof of Theorem 13.2.3

chains, it suffices to consider the case where essential states constitute a single class without subclasses ($l = 1$). Here, the matrix of transition probabilities $p_{ij}(n)$ has the form depicted in Fig. 13.4.

By virtue of the ergodic theorem, the entries of the submatrix L have positive limits π_j. Thus it remains to analyse the behaviour of the entries in the upper part of the matrix.

Theorem 13.5.1 *Let $E_i \in S^0$. Then*

$$\lim_{t\to\infty} p_{ij}(t) = \begin{cases} 0, & \text{if } E_j \in S^0, \\ \pi_j > 0, & \text{if } E_j \in S^1. \end{cases}$$

Proof Let $E_j \in S^0$. Set

$$A_j(t) := \max_{E_i \in S^0} p_{ij}(t).$$

For any essential state E_r there exists an integer t_r such that $p_{ir}(t_r) > 0$. Since transition probabilities in L are all positive starting from some step, there exists an s such that $p_{il}(s) > 0$ for $E_i \in S^0$ and all $E_l \in S^1$. Therefore, for sufficiently large t,

$$p_{ij}(t) = \sum_{E_k \in S^0} p_{ik}(s) p_{kj}(t-s) \le A_j(t-s) \sum_{E_k \in S^0} p_{ik}(s),$$

where

$$q(i) := \sum_{E_k \in S^0} p_{ik}(s) = 1 - \sum_{E_k \in S^1} p_{ik}(s) < 1.$$

If we put $q := \max_{E_i \in S^0} q(i)$, then the displayed inequality implies that

$$A_j(t) \le q A_j(t-s) \le \cdots \le q^{[t/s]}.$$

Thus $\lim_{t\to\infty} p_{ij}(t) \le \lim_{t\to\infty} A_j(t) = 0$.

Now let $E_i \in S^0$ and $E_j \in S^1$. One has

$$p_{ij}(t+s) = \sum_k p_{ik}(t) p_{kj}(s) = \sum_{E_k \in S^0} p_{ik}(t) p_{kj}(s) + \sum_{E_k \in S^1} p_{ik}(t) p_{kj}(s).$$

Letting t and s go to infinity, we see that the first sum in the last expression is $o(1)$. In the second sum,

$$\sum_{E\in S^1} p_{ik}(t) = 1 + o(1); \qquad p_{kj}(t) = \pi_j + o(1).$$

Therefore

$$p_{ij}(t+s) = \pi_j \sum_{E_k\in S} p_{ik}(t) + o(t) = \pi_j + o(1)$$

as $t, s \to \infty$. The theorem is proved. □

Using Theorem 13.5.1, it is not difficult to see that the existence of the limit

$$\lim_{t\to\infty} p_{ij}(n) = \pi_j \ge 0$$

is a necessary and sufficient condition for the chain to have two classes S^0 and S^1, of which S^1 contains no subclasses.

13.6 Markov Chains with Arbitrary State Spaces. Ergodicity of Chains with Positive Atoms

13.6.1 Markov Chains with Arbitrary State Spaces

The Markov chains $X = \{X_n\}$ considered so far have taken values in the countable sets $\{1, 2, \dots\}$ or $\{0, 1, \dots\}$; such chains are called *countable* (*denumerable*) or *discrete*. Now we will consider Markov chains with values in an arbitrary set of states $\mathcal{X}$ endowed with a σ-algebra $\mathfrak{B}_{\mathcal{X}}$ of subsets of $\mathcal{X}$. The pair $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$ forms a (measurable) *state space* of the chain $\{X_n\}$. Further let $(\Omega, \mathfrak{F}, \mathbf{P})$ be the underlying probability space. A measurable mapping Y of the space $(\Omega, \mathfrak{F})$ into $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$ is called an $\mathcal{X}$-*valued random element*. If $\mathcal{X} = \mathbb{R}$ and $\mathfrak{B}_{\mathcal{X}}$ is the σ-algebra of Borel sets on the line, then Y will be a conventional random variable. The mapping Y could be the identity, in which case $(\Omega, \mathfrak{F}) = (\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$ is also called a *sample space*.

Consider a sequence $\{X_n\}$ of $\mathcal{X}$-valued random elements and denote by $\mathcal{F}_{k,m}$, $m \ge k$, the σ-algebra generated by the elements $X_k, \dots, X_m$ (i.e. by events of the form $\{X_k \in B_k\}, \dots, \{X_m \in B_m\}$, $B_i \in \mathfrak{B}_{\mathcal{X}}$, $i = k, \dots, m$). It is evident that $\mathfrak{F}_n := \mathfrak{F}_{0,n}$ form a non-decreasing sequence $\mathfrak{F}_0 \subset \mathfrak{F}_1 \dots \subset \mathfrak{F}_n \dots$. The conditional expectation $\mathbf{E}(\xi|\mathfrak{F}_{k,m})$ will sometimes also be denoted by $\mathbf{E}(\xi|X_k, \dots, X_m)$.

Definition 13.6.1 An $\mathcal{X}$-*valued Markov chain* is a sequence of $\mathcal{X}$-valued elements X_n such that, for any $B \in \mathfrak{B}_{\mathcal{X}}$,

$$\mathbf{P}(X_{n+1} \in B \mid \mathfrak{F}_n) = \mathbf{P}(X_{n+1} \in B \mid X_n) \quad \text{a.s.} \tag{13.6.1}$$

In the sequel, the words "almost surely" will, as a rule, be omitted.

By the properties of conditional expectations, relation (13.6.1) is clearly equivalent to the condition: for any measurable function $f : \mathcal{X} \to \mathbb{R}$, one has

$$\mathbf{E}\big(f(X_{n+1}) \mid \mathcal{F}_n\big) = \mathbf{E}\big(f(X_{n+1}) \mid X_n\big). \tag{13.6.2}$$

Definition 13.6.1 is equivalent to the following.

Definition 13.6.2 A sequence $X = \{X_n\}$ forms a Markov chain if, for any $A \in \mathfrak{F}_{n+1,\infty}$,

$$\mathbf{P}(A|\mathfrak{F}_n) = \mathbf{P}(A|X_n) \tag{13.6.3}$$

or, which is the same, for any $\mathfrak{F}_{n+1,\infty}$-measurable function $f(\omega)$,

$$\mathbf{E}\big(f(\omega)|\mathfrak{F}_n\big) = \mathbf{E}\big(f(\omega)|X_n\big). \tag{13.6.4}$$

Proof of equivalence We have to show that (13.6.2) implies (13.6.3). First take any $B_1, B_2 \in \mathfrak{B}_{\mathcal{X}}$ and let $A := \{X_{n+1} \in B_1, X_{n+2} \in B_2\}$. Then, by virtue of (13.6.2),

$$\begin{aligned}\mathbf{P}(A|\mathfrak{F}_n) &= \mathbf{E}\big[\mathrm{I}(X_{n+1} \in B_1)\mathbf{P}(X_{n+2} \in B_2|\mathfrak{F}_{n+1})|\mathfrak{F}_n\big] \\ &= \mathbf{E}\big[\mathrm{I}(X_{n+1} \in B_1)\mathbf{P}(X_{n+2} \in B_2|X_{n+1})|\mathfrak{F}_n\big] \\ &= \mathbf{E}(A|X_n).\end{aligned}$$

This implies inequality (13.6.3) for any $A \in \mathcal{A}_{n+1,n+2}$, where $\mathcal{A}_{k,m}$ is the algebra generated by sets $\{X_k \in B_k, \dots, X_m \in B_m\}$. It is clear that $\mathcal{A}_{n+1,n+2}$ generates $\mathfrak{F}_{n+1,n+2}$. Now let $A \in \mathfrak{F}_{n+1,n+2}$. Then, by the approximation theorem, there exist $A_k \in \mathcal{A}_{n+1,n+2}$ such that $d(A, A_k) \to 0$ (see Sect. 3.4). From this it follows that $\mathrm{I}(A_k) \xrightarrow{p} \mathrm{I}(A)$ and, by the properties of conditional expectations (see Sect. 4.8.2),

$$\mathbf{P}\big(A_k|\mathfrak{F}^*\big) \xrightarrow{p} \mathbf{P}\big(A|\mathfrak{F}^*\big),$$

where $\mathfrak{F}^* \subset \mathfrak{F}$ is some σ-algebra. Put $P_A = P_A(\omega) := \mathbf{P}(A|X_n)$. We know that, for $A_k \in \mathcal{A}_{n+1,n+2}$,

$$\mathbf{E}(P_{A_k}; B) = \mathbf{P}(A_k B) \tag{13.6.5}$$

for any $B \in \mathfrak{F}_n$ (this just means that $P_{A_k}(\omega) = \mathbf{P}(A_k|\mathfrak{F}_n)$). Again making use of the properties of conditional expectations (the dominated convergence theorem, see Sect. 4.8.2) and passing to the limit in (13.6.5), we obtain that $\mathbf{E}(P_A;\ B) = \mathbf{P}(AB)$. This proves (13.6.3) for $A \in \mathfrak{F}_{n+1,n+2}$.

Repeating the above argument m times, we prove (13.6.3) for $A \in \mathfrak{F}_{n+1,m}$. Using a similar scheme, we can proceed to the case of $A \in \mathfrak{F}_{n+1,\infty}$. □

Note that (13.6.3) can easily be extended to events $A \in \mathfrak{F}_{n,\infty}$. In the above proof of equivalence, one could work from the very beginning with $A \in \mathfrak{F}_{n,\infty}$ (first with $A \in \mathcal{A}_{n,n+2}$, and so on).

We will give one more equivalent definition of the Markov property.

Definition 13.6.3 A sequence $\{X_n\}$ forms a Markov chain if, for any events $A \in \mathfrak{F}_n$ and $B \in \mathfrak{F}_{n,\infty}$,

$$\mathbf{P}(AB|X_n) = \mathbf{P}(A|X_n)\mathbf{P}(B|X_n). \tag{13.6.6}$$

This property means that the future is conditionally independent of the past given the present (conditional independence of $\mathfrak{F}_n$ and $\mathfrak{F}_{n,\infty}$ given X_n).

Proof of the equivalence Assume that (13.6.4) holds. Then, for $A \in \mathfrak{F}_n$ and $B \in \mathfrak{F}_{n,\infty}$,

$$\begin{aligned}\mathbf{P}(AB|X_n) &= \mathbf{E}\big[\mathbf{E}(\mathrm{I}_A\mathrm{I}_B|\mathfrak{F}_n)|X_n\big] = \mathbf{E}\big[\mathrm{I}_A\mathbf{E}(\mathrm{I}_B|\mathfrak{F}_n)|X_n\big]\\ &= \mathbf{E}\big[\mathrm{I}_A\mathbf{E}(\mathrm{I}_B|X_n)|X_n\big] = \mathbf{E}(\mathrm{I}_B|X_n)\mathbf{E}(\mathrm{I}_A|X_n),\end{aligned}$$

where I_A is the indicator of the event A.

Conversely, let (13.6.6) hold. Then

$$\begin{aligned}\mathbf{P}(AB) &= \mathbf{E}\mathbf{P}(AB|X_n) = \mathbf{E}\mathbf{P}(A|X_n)\mathbf{P}(B|X_n)\\ &= \mathbf{E}\mathbf{E}[\mathrm{I}_A\mathbf{P}(B|X_n)|X_n] = \mathbf{E}\mathrm{I}_A\mathbf{P}(B|X_n).\end{aligned} \tag{13.6.7}$$

On the other hand,

$$\mathbf{P}(AB) = \mathbf{E}\mathrm{I}_A\mathrm{I}_B = \mathbf{E}\mathrm{I}_A\,\mathbf{P}(B|\mathfrak{F}_n). \tag{13.6.8}$$

Since (13.6.7) and (13.6.8) hold for any $A \in \mathfrak{F}_n$, this means that

$$\mathbf{P}(B|X_n) = \mathbf{P}(B|\mathfrak{F}_n).$$

□

Thus, let $\{X_n\}$ be an $\mathcal{X}$-valued Markov chain. Then, by the properties of conditional expectations,

$$\mathbf{P}(X_{n+1} \in B|X_n) = P_{(n)}(X_n, B),$$

where the function $P_n(x, B)$ is, for each $B \in \mathfrak{B}_{\mathcal{X}}$, measurable in x with respect to the σ-algebra $\mathfrak{B}_{\mathcal{X}}$. In what follows, we will assume that the functions $P_{(n)}(x, B)$ are conditional distributions (see Definition 4.9.1), i.e., for each $x \in \mathcal{X}$, $P_{(n)}(x, B)$ is a probability distribution in B. Conditional distributions $P_{(n)}(x, B)$ always exist if the σ-algebra $\mathfrak{B}_{\mathcal{X}}$ is countably-generated, i.e. generated by a countable collection of subsets of $\mathcal{X}$ (see [27]). This condition is always met if $\mathcal{X} = \mathbb{R}^k$ and $\mathfrak{B}_{\mathcal{X}}$ is the σ-algebra of Borel sets. In our case, there is an additional problem that the "null probability" sets $\mathcal{N} \subset \mathcal{X}$, on which one can arbitrarily vary $P_{(n)}(x, B)$, can depend on the distribution of X_n, since the "null probability" is with respect to the distribution of X_n.

Definition 13.6.4 A Markov chain $X = \{X_n\}$ is called *homogeneous* if there exist conditional distributions $P_{(n)}(x, B) = P(x, B)$ independent of n and the initial value X_0 (or the distributions of X_n). The function $P(x, B)$ is called the *transition* *probability* (or *transition function*) of the homogeneous Markov chain. It can be graphically written as

$$P(x, B) = \mathbf{P}(X_1 \in B \mid X_0 = x). \tag{13.6.9}$$

If the Markov chain is countable, $\mathcal{X} = \{1, 2, \ldots\}$, then, in the notation of Sect. 13.1, one has $P(i, \{j\}) = p_{ij} = p_{ij}(1)$.

The transition probability and initial distribution (of X_0) completely determine the joint distribution of $X_0, \ldots, X_n$ for any n. Indeed, by the total probability formula and the Markov property

$$\mathbf{P}(X_0 \in B_0, \ldots, X_n \in B_n) = \int_{y_0 \in B_0} \cdots \int_{y_n \in B_n} \mathbf{P}(X_0 \in dy_0)\, P(y_0, dy_1) \cdots P(y_{n-1}, dy_n). \tag{13.6.10}$$

A Markov chain with the initial value $X_0 = x$ will be denoted by $\{X_n(x)\}$.

In applications, Markov chains are usually given by their conditional distributions $P(x, B)$ or—in a "stronger form"—by explicit formulas expressing X_{n+1} in terms X_n and certain "control" elements (see Examples 13.4.2, 13.4.3, 13.6.1, 13.6.2, 13.7.1–13.7.3) which enable one to immediately write down transition probabilities. In such cases, as we already mentioned, the joint distribution of $(X_0, \ldots, X_n)$ can be defined in terms of the initial distribution of X_0 and the transition function $P(x, B)$ by formula (13.6.10). It is easily seen that the sequence $\{X_n\}$ with so defined joint distributions satisfy all the definitions of a Markov chain and has transition function $P(x, B)$. In what follows, wherever it is needed, we will assume condition (13.6.10) is satisfied. It can be considered as one more definition of a Markov chain, but a stronger one than Definitions 13.6.2–13.6.4, for it explicitly gives (or uses) the transition function $P(x, B)$.

One of the main objects of study will be the asymptotic behaviour of the n step transition probability:

$$P(x, n, B) := \mathbf{P}\big(X_n(x) \in B\big) = \mathbf{P}(X_n \in B \mid X_0 = x).$$

The following recursive relation, which follows from the total probability formula (or from (13.6.10)), holds for this function:

$$\mathbf{P}(X_{n+1} \in B) = \mathbf{E}\mathbf{E}\big(\mathrm{I}(X_{n+1} \in B) \mid \mathfrak{F}_n\big) = \int \mathbf{P}(X_n \in dy) P(y, B),$$

$$P(x, n+1, B) = \int P(x, n, dy) P(y, B). \tag{13.6.11}$$

Now note that the Markov property (13.6.3) of homogeneous chains can also be written in the form

$$\mathbf{P}(X_{n+k} \in B_k \mid \mathfrak{F}_n) = P(X_n, k, B_k),$$

or, more generally,

$$\mathbf{P}(X_{n+1} \in B_1, \ldots, X_{n+k} \in B_k | \mathfrak{F}_n) = \mathbf{P}\big(X_1^{\text{new}}(X_n) \in B_1, \ldots, X_k^{\text{new}}(X_n) \in B_k\big), \tag{13.6.12}$$

where $\{X_k^{\text{new}}(x)\}$ is a Markov chain independent of $\{X_n\}$ and having the same transition function as $\{X_n\}$ and the initial value x. Property (13.6.12) can be extended to a random time n. Recall the definition of a stopping time.

Definition 13.6.5 A random variable $\nu \ge 0$ is called a *Markov* or *stopping time* with respect to $\{\mathfrak{F}_n\}$ if $\{\nu \le n\} \in \mathfrak{F}_n$. In other words, that the event $\{\nu \le n\}$ occurred or not is completely determined by the trajectory segment $X_0, X_1, \ldots, X_n$.

Note that, in Definition 13.6.5, by $\mathfrak{F}_n$ one often understands wider σ-algebras, the essential requirements being the relations $\{\nu \le n\} \in \mathcal{F}_n$ and measurability of $X_0, \ldots, X_n$ with respect to $\mathfrak{F}_n$.

Denote by $\mathfrak{F}_\nu$ the σ-algebra of events B such that $B \cap \{\nu = k\} \in \mathfrak{F}_k$. In other words, $\mathcal{F}_\nu$ can be thought of as the σ-algebra generated by the sets $\{\nu = k\}B_k$, $B_k \in \mathfrak{F}_k$, i.e. by the trajectory of $\{X_n\}$ until time ν.

Lemma 13.6.1 (The Strong Markov Property) *For any $k \ge 1$ and $B_1, \ldots, B_k \in \mathfrak{B}_\mathcal{X}$,*

$$\mathbf{P}(X_{\nu+1} \in B_1, \ldots, X_{\nu+k} \in B_k | \mathfrak{F}_\nu) = \mathbf{P}\big(X_1^{\text{new}}(X_\nu) \in B_1, \ldots, X_k^{\text{new}}(X_\nu) \in B_k\big),$$

where the process $\{X_k^{\text{new}}\}$ is defined in (13.6.12).

Thus, after a random stopping time ν, the trajectory $X_{\nu+1}, X_{\nu+2}, \ldots$ will evolve according to the same laws as $X_1, X_2, \ldots$, but with the initial condition X_ν. This property is called the *strong Markov property*. It will be used below for the first hitting times $\nu = \tau_V$ of certain sets $V \subset \mathcal{X}$ by $\{X_n\}$. We have already used this property tacitly in Sect. 13.4, when the set V coincided with a point, which allowed us to cut the trajectory of $\{X_n\}$ into independent cycles.

Proof of Lemma 13.6.1 For the sake of simplicity, consider one-dimensional distributions. We have to prove that

$$\mathbf{P}(X_{\nu+1} \in B_1 | \mathfrak{F}_\nu) = P(X_\nu, B_1).$$

For any $A \in \mathfrak{F}_\nu$,

$$\begin{aligned}
\mathbf{E}\big(P(X_\nu, B_1); A\big) &= \sum_n \mathbf{E}\big(P(X_n, B_1); A\{\nu = n\}\big) \\
&= \sum_n \mathbf{E}\mathbf{E}\big(I\big(A\{\nu = n\}\{X_{n+1} \in B_1\}\big) | \mathfrak{F}_n\big) \\
&= \sum_n \mathbf{P}\big(A\{\nu = n\}\{X_{n+1} \in B_1\}\big) = \mathbf{P}\big(A\{X_{\nu+1} \in B_1\}\big).
\end{aligned}$$

But this just means that $P(X_\nu, B_1)$ is the required conditional expectation. The case of multi-dimensional distributions is dealt with in the same way, and we leave it to the reader. □

Now we turn to consider the asymptotic properties of distributions $P(x,n,B)$ as $n \to \infty$.

Definition 13.6.6 A distribution $\boldsymbol{\pi}(\cdot)$ on $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$ is called *invariant* if it satisfies the equation

$$\boldsymbol{\pi}(B) = \int \boldsymbol{\pi}(dy) P(y, B), \qquad B \in \mathfrak{B}_{\mathcal{X}}. \tag{13.6.13}$$

It follows from (13.6.11) that if $X_n \Subset \boldsymbol{\pi}$, then $X_{n+1} \Subset \pi$. The distribution $\boldsymbol{\pi}$ is also called *stationary*.

For Markov chains in arbitrary state spaces $\mathcal{X}$, a simple and complete classification similar to the one carried out for countable chains in Sect. 13.1 is not possible, although some notions can be extended to the general case.

Such natural and important notions for countable chains as, say, irreducibility of a chain, take in the general case another form.

Example 13.6.1 Let $X_{n+1} = X_n + \xi_n \pmod 1$ (X_{n+1} is the fractional part of $X_n + \xi_n$), ξ_n be independent and identically distributed and take with positive probabilities the two values 0 and $\sqrt{2}$. In this example, the chain "splits", according to the initial state x, into a continual set of "subchains" with state spaces of the form $M_x = \{x + k\sqrt{2} \pmod 1,\ k = 0, 1, 2 \ldots\}$. It is evident that if $x_1 - x_2$ is not a multiple of $\sqrt{2}$ (mod 1), then M_{x_1} and M_{x_2} are disjoint, $\mathbf{P}(X_n(x_1) \in M_{x_2}) = 0$ and $\mathbf{P}(X_n(x_2) \in M_{x_1}) = 0$ for all n. Thus the chain is clearly reducible. Nevertheless, it turns out that the chain is ergodic in the following sense: for any x, $X_n(x) \Leftrightarrow U_{0,1}$ ($P(x,n,[0,t]) \to t$) as $n \to \infty$ (see, e.g., [6], [18]). For the most commonly used irreducibility conditions, see Sect. 13.7.

Definition 13.6.7 A chain is called *periodic* if there exist an integer $d \ge 2$ and a set $\mathcal{X}_1 \subset \mathcal{X}$ such that, for $x \in \mathcal{X}_1$, one has $P(x,n,\mathcal{X}_1) = \mathbf{P}(X_n(x) \in \mathcal{X}_1) = 1$ for $n = kd$, $k = 1, 2, \ldots$, and $P(x,n,\mathcal{X}_1) = 0$ for $n \ne kd$.

Periodicity means that the whole set of states $\mathcal{X}$ is decomposed into subclasses $\mathcal{X}_1, \ldots, \mathcal{X}_d$, such that $\mathbf{P}(X_1(x) \in \mathcal{X}_{k+1}) = 1$ for $x \in \mathcal{X}_k$, $k = 1, \ldots, d$, $\mathcal{X}_{d+1} = \mathcal{X}_1$. In the absence of such a property, the chain will be called *aperiodic*.

A state $x_0 \in \mathcal{X}$ is called an *atom* of the chain X if, for any $x \in \mathcal{X}$,

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty} \{X_n(x) = x_0\} \right) = 1.$$

Example 13.6.2 Let $X_0 \ge 0$ and, for $n \ge 0$,

$$X_{n+1} = \begin{cases} (X_n + \xi_{n+1})^+ & \text{if } X_n > 0, \\ \eta_{n+1} & \text{if } X_n = 0, \end{cases}$$

where ξ_n and $\eta_n \geq 0$, $n = 1, 2, \ldots$, are two sequences of independent random variables, identically distributed in each sequence. It is clear that $\{X_n\}$ is a Markov chain and, for $\mathbf{E}\xi_k < 0$, by the strong law of large numbers, this chain has an atom at the point $x_0 = 0$:

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty}\{X_n(x) = 0\}\right) = \mathbf{P}\Big(\inf_k S_k \leq -x\Big) = 1,$$

where $S_k = \sum_{j=1}^{k} \xi_j$. This chain is a generalisation of the Markov chain from Example 13.4.3.

Markov chains in an arbitrary state space $\mathcal{X}$ are rather difficult to study. However, if a chain has an atom, the situation may become much simpler, and the ergodic theorem on the asymptotic behaviour of $P(x, n, B)$ as $n \to \infty$ can be proved using the approaches considered in the previous sections.

13.6.2 Markov Chains Having a Positive Atom

Let x_0 be an atom of a chain $\{X_n\}$. Set

$$\tau := \min\big\{k > 0 : X_k(x_0) = x_0\big\}.$$

This is a proper random variable ($\mathbf{P}(\tau < \infty) = 1$).

Definition 13.6.8 The atom x_0 is said to be *positive* if $\mathbf{E}\tau < \infty$.

In the terminology of Sect. 13.4, x_0 is a recurrent non-null (positive) state.

To characterise convergence of distributions in arbitrary spaces, we will need the notions of the total variation distance and convergence in total variation. If $\mathbf{P}$ and $\mathbf{Q}$ are two distributions on $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$, then the *total variation distance* between them is defined by

$$\|\mathbf{P} - \mathbf{Q}\| = 2 \sup_{B \in \mathfrak{B}_{\mathcal{X}}} \big|\mathbf{P}(B) - Q(B)\big|.$$

One says that a sequence of distributions $\mathbf{P}_n$ on $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$ converges in total variation to $\mathbf{P}$ ($\mathbf{P}_n \xrightarrow{TV} \mathbf{P}$) if $\|\mathbf{P}_n - \mathbf{P}\| \to 0$ as $n \to \infty$. For more details, see Sect. 3.6.2 of Appendix 3.

As in Sect. 13.4, denote by $P_{x_0}(k, B)$ the "taboo probability"

$$P_{x_0}(k, B) := \mathbf{P}\big(X_k(x_0) \in B, X_1(x_0) \neq x_0, \ldots, X_{k-1}(x_0) \neq x_0\big)$$

of transition from x_0 into B in k steps without visiting the "forbidden" state x_0.

Theorem 13.6.1 *If the chain $\{X_n\}$ has a positive atom and the g.c.d. of the possible values of τ is* 1, *then the chain is ergodic in the convergence in total variation sense*:

there exists a unique invariant distribution $\boldsymbol{\pi}$ *such that, for any* $x \in \mathcal{X}$, *as* $n \to \infty$,

$$\left\| P(x, n, \cdot) - \boldsymbol{\pi}(\cdot) \right\| \to 0. \tag{13.6.14}$$

Moreover, for any $B \in \mathfrak{B}_{\mathcal{X}}$,

$$\boldsymbol{\pi}(B) = \frac{1}{\mathbf{E}\tau} \sum_{k=1}^{\infty} P_{x_0}(k, B). \tag{13.6.15}$$

If we denote by $X_n(\boldsymbol{\mu}_0)$ *a Markov chain with the initial distribution* $\boldsymbol{\mu}_0$ $(X_0 \Subset \boldsymbol{\mu}_0)$ *and put*

$$P(\boldsymbol{\mu}_0, n, B) := \mathbf{P}\big(X_n(\boldsymbol{\mu}_0) \in B\big) = \int \boldsymbol{\mu}_0(dx)\, P(x, n, B),$$

then, as well as (13.6.14), *we will also have that, as* $n \to \infty$,

$$\left\| P(\boldsymbol{\mu}_0, n, \cdot) - \boldsymbol{\pi}(\cdot) \right\| \to 0 \tag{13.6.16}$$

for any initial distribution $\boldsymbol{\mu}_0$.

The condition that there exists a positive atom is an analogue of conditions (I) and (II) of Theorem 13.4.1. A number of conditions sufficient for the finiteness of $\mathbf{E}\tau$ can be found in Sect. 13.7. The condition on the g.c.d. of possible values of τ is the aperiodicity condition.

Proof We will effectively repeat the proof of Theorem 13.4.1. First let $X_0 = x_0$. As in Theorem 13.4.1 (we keep the notation of that theorem), we find that

$$\begin{aligned}
&P(x_0, n, B) \\
&\quad = \sum_{k=1}^{n} \mathbf{P}\big(\gamma(n) = k\big)\, \mathbf{P}(X_n \in B \,|\, X_{n-k} = x_0, X_{n-k+1} \neq x_0, \dots, X_{n-1} \neq x_0) \\
&\quad = \sum_{k=1}^{n} \frac{\mathbf{P}(\gamma(n) = k)}{\mathbf{P}(\tau \geq k)} \mathbf{P}(\tau \geq k)\, \mathbf{P}(X_k \in B \,|\, X_0 = x_0, X_1 \neq x_0, \dots, X_{k-1} \neq x_0) \\
&\quad = \sum_{k=1}^{n} \frac{\mathbf{P}(\gamma(n) = k)}{\mathbf{P}(\tau \geq k)}\, P_{x_0}(k, B).
\end{aligned}$$

For the measure $\boldsymbol{\pi}$ defined in (13.6.15) one has

$$\begin{aligned}
&P(x_0, n, B) - \boldsymbol{\pi}(B) \\
&\quad = \sum_{k=1}^{n} \left(\frac{\mathbf{P}(\gamma(n) = k)}{\mathbf{P}(\tau \geq k)} - \frac{1}{\mathbf{E}\tau} \right) P_{x_0}(k, B) - \frac{1}{\mathbf{E}\tau} \sum_{k > n} P_{x_0}(k, B).
\end{aligned}$$

Since $\mathbf{P}(\gamma(n)=k) \le \mathbf{P}(\tau \ge k)$ and $P_{x_0}(k, B) \le \mathbf{P}(\tau \ge k)$ (see the proof of Theorem 13.4.1), one has, for any N,

$$\sup_B \bigl| P(x_0, n, B) - \boldsymbol{\pi}(B) \bigr| \le \sum_{k=1}^{N} \left(\frac{\mathbf{P}(\gamma(n)=k)}{\mathbf{P}(\tau \ge k)} - \frac{1}{\mathbf{E}\tau} \right) + 2 \sum_{k>N} \mathbf{P}(\tau \ge k). \tag{13.6.17}$$

Further, since

$$\mathbf{P}\bigl(\gamma(n)=k\bigr) \to \mathbf{P}(\tau \ge k)/\mathbf{E}\tau, \qquad \sum_{k=1}^{\infty} \mathbf{P}(\tau \ge k) = \mathbf{E}\tau < \infty,$$

the right-hand side of (13.6.17) can be made arbitrarily small by choosing N and then n. Therefore,

$$\lim_{n\to\infty} \sup_B \bigl| P(x_0, n, B) - \boldsymbol{\pi}(B) \bigr| = 0.$$

Now consider an arbitrary initial state $x \in \mathcal{X}$, $x \ne x_0$. Since x_0 is an atom, for the probabilities

$$F(x, k, x_0) := \mathbf{P}\bigl(X_k(x) = x_0,\ X_1 \ne x_0, \ldots, X_{k-1} \ne x_0\bigr)$$

of hitting x_0 for the first time on the k-th step, one has

$$\sum_k F(x, k, x_0) = 1, \qquad P(x, n, B) = \sum_{k=1}^{n} F(x, k, x_0) P(x_0, n-k, B),$$

$$\begin{aligned} &\bigl\| P(x, n, \cdot) - \boldsymbol{\pi}(\cdot) \bigr\| \\ &\quad \le \sum_{k \le n/2} F(x, k, x_0) \bigl\| P(x_0, n-k, \cdot) - \boldsymbol{\pi}(\cdot) \bigr\| + 2 \sum_{k>n/2} F(x, k, x_0) \to 0 \end{aligned}$$

as $n \to \infty$.

Relation (13.6.16) follows from the fact that

$$\bigl\| P(\boldsymbol{\mu}_0, n, \cdot) - \boldsymbol{\pi}(\cdot) \bigr\| \le \int \boldsymbol{\mu}_0(dx) \bigl\| P(x, n, \cdot) - \boldsymbol{\pi}(\cdot) \bigr\| \to 0$$

by the dominated convergence theorem.

Further, from the convergence of $P(x, n, \cdot)$ in total variation it follows that

$$\int P(x, n, dy) P(y, B) \to \int \boldsymbol{\pi}(dy) P(y, B).$$

Since the left hand-side of this relation is equal to $P(x, n+1, B)$ by virtue of (13.6.11) and converges to $\boldsymbol{\pi}(B)$, one has (13.6.13), and hence $\boldsymbol{\pi}$ is an invariant measure.

Now assume that $\boldsymbol{\pi}_1$ is another invariant distribution. Then

$$\boldsymbol{\pi}_1(\cdot) = \mathbf{P}(\boldsymbol{\pi}_1, n, \cdot) \xrightarrow{TV} \boldsymbol{\pi}(\cdot), \quad \boldsymbol{\pi}_1 = \boldsymbol{\pi}.$$

The theorem is proved. □

Returning to Example 13.6.2, we show that the conditions of Theorem 13.6.1 are met provided that $\mathbf{E}\xi_k < 0$ and $\mathbf{E}\eta_k < \infty$. Indeed, put

$$\eta(-x) := \min\left\{k \ge 1 : S_k = \sum_{j=1}^{k} \xi_j \le -x\right\}.$$

By the renewal Theorem 10.1.1,

$$H(x) = \mathbf{E}\eta(-x) \sim \frac{x}{|\mathbf{E}\xi_1|} \quad \text{as} \quad x \to \infty$$

for $\mathbf{E}\xi_1 < 0$, and therefore there exist constants c_1 and c_2 such that $H(x) < c_1 + c_2 x$ for all $x \ge 0$. Hence, for the atom $x_0 = 0$, we obtain that

$$\mathbf{E}\tau = \int_0^\infty \mathbf{P}(\eta_1 \in dx)\, H(x) \le c_1 + c_2 \int_0^\infty x\mathbf{P}(\eta_1 \in dx) = c_1 + c_2 \mathbf{E}\eta_1 < \infty.$$

13.7* Ergodicity of Harris Markov Chains

13.7.1 The Ergodic Theorem

In this section we will consider the problem of establishing ergodicity of Markov chains in arbitrary state spaces $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$. A lot of research has been done on this problem, the most important advancements being associated with the names of W. Döblin, J.L. Doob, T.E. Harris and E. Omey. Until recently, this research area had been considered as a rather difficult one, and not without reason. However, the construction of an artificial atom suggested by K.B. Athreya, P.E. Ney and E. Nummelin (see, e.g. [6, 27, 29]) greatly simplified considerations and allowed the proof of ergodicity by reducing the general case to the special case discussed in the last section.

In what follows, the notion of a "Harris chain" will play an important role. For a fixed set $V \in \mathfrak{B}_{\mathcal{X}}$, define the random variable

$$\tau_V(x) = \min\left\{k \ge 1 : X_k(x) \in V\right\},$$

the time of the first hitting of V by the chain starting from the state x (we put $\tau_V(x) = \infty$ if all $X_k(x) \notin V$).

Definition 13.7.1 A Markov chain $X = \{X_n\}$ in $(\mathcal{X}, \mathfrak{B}_\mathcal{X})$ is said to be a *Harris chain* (or *Harris irreducible*) if there exists a set $V \in \mathfrak{B}_\mathcal{X}$, a probability measure $\boldsymbol{\mu}$ on $(\mathcal{X}, \mathfrak{B}_\mathcal{X})$, and numbers $n_0 \ge 1$, $p \in (0, 1)$ such that

(I_0) $\mathbf{P}(\tau_V(x) < \infty) = 1$ for all $x \in \mathcal{X}$; and
(II) $P(x, n_0, B) \ge p\boldsymbol{\mu}(B)$ for all $x \in V$, $B \in \mathfrak{B}_\mathcal{X}$.

Condition (I_0) plays the role of an irreducibility condition: starting from any point $x \in \mathcal{X}$, the trajectory of X_n will sooner or later visit the set V. Condition (II) guarantees that, after n_0 steps since hitting V, the distribution of the walking particle will be minorised by a common "distribution" $p\boldsymbol{\mu}(\cdot)$. This condition is sometimes called a "mixing condition"; it ensures a "partial loss of memory" about the trajectory's past. This is not the case for the chain from Example 13.6.1 for which condition (II) does not hold for any V, $\boldsymbol{\mu}$ or n_0 ($P(x, \cdot)$ form a collection of mutually singular distributions which are singular with respect to Lebesgue measure).

If a chain has an atom x_0, then conditions (I_0) and (II) are always satisfied for $V = \{x_0\}$, $n_0 = 1$, $p = 1$, and $\boldsymbol{\mu}(\cdot) = P(x_0, \cdot)$, so that such a chain is a Harris chain.

The set V is usually chosen to be a "compact" set (if $\mathcal{X} = \mathbb{R}^k$, it will be a bounded set), for otherwise one cannot, as a rule, obtain inequalities in (II). If the space $\mathcal{X}$ is "compact" itself (a finite or bounded subset of $\mathbb{R}^k$), condition (II) can be met for $V = \mathcal{X}$ (condition (I_0) then always holds). For example, if $\{X_n\}$ is a finite, irreducible and aperiodic chain, then by Theorem 13.4.2 there exists an n_0 such that $P(i, n_0, j) \ge p > 0$ for all i and j. Therefore condition (II) holds for $V = \mathcal{X}$ if one takes $\boldsymbol{\mu}$ to be a uniform distribution on $\mathcal{X}$.

One could interpret condition (II) as that of the presence, in all distributions $P(x, n_0, \cdot)$ for $x \in V$, of a component which is absolutely continuous with respect to the measure $\boldsymbol{\mu}$:

$$\inf_{x \in V} \frac{P(x, n_0, dy)}{\boldsymbol{\mu}(dy)} \ge p > 0.$$

We will also need a condition of "positivity" (positive recurrence) of the set V (or that of "positivity" of the chain):

(I) $\sup_{x\in V} \mathbf{E}\tau_V(x) < \infty$,

and the aperiodicity condition which will be written in the following form. Let $X_k(\boldsymbol{\mu})$ be a Markov chain with an initial value $X_0 \Subset \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is from condition (II). Put

$$\tau_V(\boldsymbol{\mu}) := \min\bigl\{k \ge : X_k(\boldsymbol{\mu}) \in V\bigr\}.$$

It is evident that $\tau_V(\boldsymbol{\mu})$ is, by virtue of (I_0), a proper random variable. Denote by $n_1, n_2, \ldots$ the possible values of $\tau_V(\boldsymbol{\mu})$, i.e. the values for which

$$\mathbf{P}\bigl(\tau_V(\boldsymbol{\mu}) = n_k\bigr) > 0, \quad k = 1, 2, \ldots.$$

Then the aperiodicity condition will have the following form.

(III) *There exists a* $k \ge 1$ *such that*

$$\text{g.c.d.}\{n_0 + n_1, n_0 + n_2, \dots, n_0 + n_k\} = 1,$$

where n_0 *is from condition* (II).

Condition (III) is always satisfied if (II) holds for $n_0 = 1$ and $\boldsymbol{\mu}(V) > 0$ (then $n_1 = 0$, $n_0 + n_1 = 1$).

Verifying condition (I) usually requires deriving bounds for $\mathbf{E}\tau_V(x)$ for $x \notin V$ which would automatically imply (I_0) (see the examples below).

Theorem 13.7.1 *Suppose conditions* (I_0), (I), (II) *and* (III) *are satisfied for a Markov chain* X, *i.e. the chain is an aperiodic positive Harris chain. Then there exists a unique invariant distribution* π *such that, for any initial distribution* $\boldsymbol{\mu}_0$, *as* $n \to \infty$,

$$\big\| P(\boldsymbol{\mu}_0, n, \cdot) - \pi(\cdot) \big\| \to 0. \tag{13.7.1}$$

The proof is based on the use of the above-mentioned construction of an "artificial atom" and reduction of the problem to Theorem 13.6.1. This allows one to obtain, in the course of the proof, a representation for the invariant measure π similar to (13.6.15) (see (13.7.5)).

A remarkable fact is that the conditions of Theorem 13.7.1 are necessary for convergence (13.7.1) (for more details, see [6]).

Proof of Theorem 13.7.1 For simplicity's sake, assume that $n_0 = 1$. First we will construct an "extended" Markov chain $X^* = \{X_n^*\} = \{\widetilde{X}_n, \omega(n)\}$, $\omega(n)$ being a sequence of independent identically distributed random variables with

$$\mathbf{P}\big(\omega(n) = 1\big) = p, \qquad \mathbf{P}\big(\omega(n) = 0\big) = 1 - p.$$

The joint distribution of $(\widetilde{X}(n), \omega(n))$ in the state space

$$\mathcal{X}^* := \mathcal{X} \times \{0, 1\} = \big\{x^* = (x, \delta) : x \in \mathcal{X}; \delta = 0, 1\big\}$$

and the transition function P^* of the chain X^* are defined as follows (the notation $X_n^*(x^*)$ has the same meaning as $X_n(x)$):

$$\mathbf{P}\big(X_1^*\big(x^*\big) \in (B, \delta)\big) =: P^*\big(x^*, (B, \delta)\big) = P(x, B)\, \mathbf{P}\big(\omega(1) = \delta\big) \quad \text{for } x \notin V$$

(i.e., for $\widetilde{X}_n \notin V$, the components of X_{n+1}^* are "chosen at random" independently with the respective marginal distributions). But if $x \in V$, the distribution of $X^*(x^*, 1)$ is given by

$$\mathbf{P}(X_1^*\big((x, 1) \in (B, \delta)\big) = P^*\big((x, 1), (B, \delta)\big) = \boldsymbol{\mu}(B)\, \mathbf{P}\big(\omega(1) = \delta\big),$$
$$\mathbf{P}(X_1^*\big((x, 0) \in (B, \delta)\big) = P^*\big((x, 0), (B, \delta)\big) = Q(x, B)\, \mathbf{P}\big(\omega(1) = \delta\big),$$

where

$$Q(x, B) := \big(P(x, B) - p\boldsymbol{\mu}(B)\big)/(1 - p),$$

so that, for any $B \in \mathfrak{B}_{\mathcal{X}}$,

$$p\boldsymbol{\mu}(B) + (1 - p)Q(x, B) = P(x, B). \tag{13.7.2}$$

Thus $\mathbf{P}(\omega(n + 1) = 1 | X_n^*) = p$ for any values of X_n^*. However, when "choosing" the value $\widetilde{X}_{n+1}$ there occurs (only when $\widetilde{X}_n \in V$) a partial randomisation (or splitting): for $\widetilde{X}_n \in V$, we let $\mathbf{P}(\widetilde{X}_{n+1} \in B | X_n^*)$ be equal to the value $\boldsymbol{\mu}(B)$ (not depending on $\widetilde{X}_n \in V$!) provided that $\omega(n) = 1$. If $\omega(n) = 0$, then the value of the probability is taken to be $Q(\widetilde{X}_n, B)$. It is evident that, by virtue of condition (II) (for $n_0 = 1$), $\boldsymbol{\mu}(B)$ and $Q(x, B)$ are probability distributions, and by equality (13.7.2) the first component $\widetilde{X}_n$ of the process X_n^* has the property $\mathbf{P}(\widetilde{X}_{n+1} \in B | \widetilde{X}_n) = P(\widetilde{X}_n, B)$, and therefore the distributions of the sequences X and $\widetilde{X}$ coincide.

As we have already noted, the "extended" process $X^*(n)$ possesses the following property: the conditional distribution $\mathbf{P}(X_{n+1}^* \in (B, \delta) | X_n^*)$ does not depend on $X^*(n)$ on the set $X_n^* \in V^* := (V, 1)$ and is there the known distribution $\boldsymbol{\mu}(B)\mathbf{P}(\omega(1) = \delta)$. This just means that visits of the chain X^* to the set V^* divide the trajectory of X^* into independent cycles, in the same way as it happens in the presence of a positive atom.

We described above how one constructs the distribution of X^* from that of X. Now we will give obvious relations reconstructing the distribution of X from that of the chain X^*:

$$\mathbf{P}\big(X_n(x) \in B\big) = p\,\mathbf{P}(X_n^*\big((x, 1) \in B^*\big) + (1 - p)\,\mathbf{P}\big(X_n^*(x, 0) \in B^*\big), \tag{13.7.3}$$

where $B^* := (B, 0) \cup (B, 1)$. Note also that, if we consider $X_n = \widetilde{X}_n$ as a component of X_n^*, we need to write it as a function $X_n(x^*)$ of the initial value $x^* \in \mathcal{X}^*$.

Put

$$\tau^* := \min\big\{k \ge 1 : X_k^*\,\big(x^*\big) \in V^*\big\}, \quad x^* \in V^* = (V, 1).$$

It is obvious that τ^* does not depend on the value $x^* = (x, 1)$, since $X_1(x^*)$ has the distribution $\boldsymbol{\mu}$ for any $x \in V$. This property allows one to identify the set V^* with a single point. In other words, one needs to consider one more state space $\mathcal{X}^{**}$ which is obtained from $\mathcal{X}^*$ if we replace the set $V^* = (V, 1)$ by a point to be denoted by x_0. In the new state space, we construct a chain X^{**} equivalent to X^* using the obvious relations for the transition probability P^{**}:

$$P^{**}\big(x^*, (B, \delta)\big) := P^*\big(x^*, (B, \delta)\big) \quad \text{for } x^* \ne (V, 1) = V^*, \ (B, d) \ne V^*,$$
$$P^{**}\big(x_0, (B, \delta)\big) := p\boldsymbol{\mu}(B), \qquad P^{**}\big(x^*, x_0\big) := P^*\big(x^*, V^*\big).$$

Thus we have constructed a chain X^{**} with the transition function P^{**}, and this chain has atom x_0. Clearly, $\tau^* = \min\{k \ge 1 : X_k^{**}(x_0) = x_0\}$. We now prove that this atom is positive. Put

$$E := \sup_{x \in V} \mathbf{E}\tau_V(x).$$

Lemma 13.7.1 $\mathbf{E}\tau^* \le \frac{2}{p}E$.

Proof Consider the evolution of the first component $X_k(x^*)$ of the process $X_k^*(x^*)$, $x^* \in V^*$. Partition the time axis $k \ge 0$ into intervals by hitting the set V by $X_k(x^*)$. Let $\tau_1 \ge 1$ be the first such hitting time (recall that $X_1(x^*) \stackrel{d}{=} X_0(\boldsymbol{\mu})$ has the distribution $\boldsymbol{\mu}$, so that $\tau_1 = 1$ if $\boldsymbol{\mu}(V) = 1$). Prior to time τ_1 (in the case $\tau_1 > 1$) transitions of $X_k(x^*)$, $k \ge 2$, were governed by the transition function $P(y, B)$, $y \in V^c = \mathcal{X} \setminus V$. At time τ_1, according to the definition of X^*, one carries out a Bernoulli trial independent of the past history of the process with success (which is the event $\omega(\tau_1) = 1$) probability p. If $\omega(\tau_1) = 1$ then $\tau^* = \tau_1$. If $\omega(\tau_1) = 0$ then the transition from $X_{\tau_1}(x^*)$ to $X_{\tau_1+1}(x^*)$ is governed by the transition function $Q(y, B) = (P(y, B) - p\boldsymbol{\mu}(B))/(1 - p)$, $y \in V$. The further evolution of the chain is similar: if $\tau_1 + \tau_2$ is the time of the second visit of $X(x^*, k)$ to V (in the case $\omega(\tau_1) = 0$) then in the time interval $[\tau_1 + 1, \tau_2]$ transitions of $X(x^*, k)$ occur according to the transition function $P(y, B)$, $y \in V^c$. At time $\tau_1 + \tau_2$ one carries out a new Bernoulli trial with the outcome $\omega(\tau_1 + \tau_2)$. If $\omega(\tau_1 + \tau_2) = 1$, then $\tau^* = \tau_1 + \tau_2$. If $\omega(\tau_1 + \tau_2) = 0$, then the transition from $X(x^*, \tau_1 + \tau_2)$ to $X(x^*, \tau_1 + \tau_2 + 1)$ is governed by $Q(y, B)$, and so on.

In other words, the evolution of the component $X_k(x^*)$ of the process $X_k^*(x^*)$ is as follows. Let $\widetilde{X} = \{\widetilde{X}_k\}$, $k = 1, 2, \ldots$, be a Markov chain with the distribution $\boldsymbol{\mu}$ at time $k = 1$ and transition probability $Q(x, B)$ at times $k \ge 2$,

$$Q(x, B) = \begin{cases} (P(x, B) - p\boldsymbol{\mu}(B))/(1 - p) & \text{if } x \in V, \\ P(x, B) & if x \in V^c. \end{cases}$$

Define T_i as follows:

$$T_0 := 0, \qquad T_1 = \tau_1 = \min\{k \ge 1 : \widetilde{X}_k \in V\},$$
$$T_i := \tau_1 + \cdots + \tau_i = \min\{k > T_{i-1} : \widetilde{X}_k \in V\}, \quad i \ge 2.$$

Let, further, ν be a random variable independent of $\widetilde{X}$ and having the geometric distribution

$$\mathbf{P}(\nu = k) = (1 - p)^{k-1}p, \quad k \ge 1, \quad \nu = \min\bigl\{k \ge 1 : \omega(T_k) = 1\bigr\}. \tag{13.7.4}$$

Then it follows from the aforesaid that the distribution of $X_1(x^*), \ldots, X_{\tau^*}(x^*)$ coincides with that of $\widetilde{X}_1, \ldots, \widetilde{X}_\nu$; in particular, $\tau^* = T_\nu$, and

$$\mathbf{E}\tau^* = \sum_{k=1}^{\infty} p(1 - p)^{k-1}\mathbf{E}T_k.$$

Further, since $\boldsymbol{\mu}(B) \le P(x, B)/p$ for $x \in V$, then, for any $x \in V$,

$$\mathbf{E}\tau_1 = \boldsymbol{\mu}(V) + \int_{V^c} \boldsymbol{\mu}(du)\bigl(1 + \mathbf{E}\tau_V(u)\bigr)$$

$$\leq \frac{1}{p}\left[P(x,V)+\int_{V^c}P(x,du)\big(1+\mathbf{E}\tau_V(u)\big)\right]=\frac{\mathbf{E}\tau_V(x)}{p}\leq\frac{E}{p}.$$

To bound $\mathbf{E}\tau_i$ for $i \geq 2$, we note that $Q(x,B) \leq (1-p)^{-1}P(x,B)$ for $x \in V$. Therefore, if we denote by $\mathcal{F}_{(i)}$ the σ-algebra generated by $\{\widetilde{X}_k,\, \omega(\tau_k)\}$ for $k \leq T_i$, then

$$\begin{aligned}\mathbf{E}(\tau_i|\mathcal{F}_{(i-1)}) &\leq \sup_{x\in V}\left[Q(x,V)+\int_{V^c}Q(x,du)\big(1+\mathbf{E}\tau_V(u)\big)\right]\\ &\leq \frac{1}{1-p}\sup_{x\in V}\left[P(x,V)+\int_{V^c}P(x,du)\big(1+\mathbf{E}\tau_V(u)\big)\right]\\ &=(1-p)^{-1}\sup_{x\in V}\mathbf{E}\tau_V(x)=E(1-p)^{-1}.\end{aligned}$$

This implies the inequality $\mathbf{E}T_k \leq E(1/p+(k-1)/(1-p))$, from which we obtain that

$$\mathbf{E}\tau^* \leq E\left(1/p+p\sum_{k=1}^{\infty}(k-1)(1-p)^{k-2}\right)=2E/p.$$

The lemma is proved. □

We return to the proof of the theorem. To make use of Theorem 13.6.1, we now have to show that $\mathbf{P}(\tau^*(x^*)<\infty)=1$ for any $x^* \in \mathcal{X}^*$, where

$$\tau^*\big(x^*\big):=\min\big\{k\geq 1: X_k^*\big(x^*\big)\in V^*\big\}.$$

But the chain X visits V with probability 1. After ν visits to V (ν was defined in (13.7.4)), the process $X^*=(X(n),\omega(n))$ will be in the set V^*.

The aperiodicity condition for $n_0=1$ will be met if $\boldsymbol{\mu}(V)>0$. In that case we obtain by virtue of Theorem 13.6.1 that there exists a unique invariant measure π^* such that, for any $x^* \in \mathcal{X}^*$,

$$\begin{gathered}\big\|P^*\big(x^*,n,\cdot\big)-\boldsymbol{\pi}^*(\cdot)\big\|\to 0,\qquad \boldsymbol{\pi}^*\big((B,\delta)\big)=\frac{1}{\mathbf{E}\tau^*}\sum_{k=1}^{\infty}P_{V^*}^*\big(k,(B,d)\big),\\ P_{V^*}^*\big(k,(B,\delta)\big)=\mathbf{P}\big(X_k^*\big(x^*\big)\in(B,\delta),\,X_1^*\big(x^*\big)\notin V^*,\dots,X_{k-1}^*\big(x^*\big)\notin V^*\big).\end{gathered}\tag{13.7.5}$$

In the last equality, we can take any point $x^* \in V^*$; the probability does not depend on the choice of $x^* \in V^*$.

From this and the "inversion formula" (13.7.3) we obtain assertion (13.7.1) and a representation for the invariant measure π of the process X.

The proof of the convergence $\|P(\boldsymbol{\mu}_0,n,\cdot)-\pi(\cdot)\|\to 0$ and uniqueness of the invariant measure is exactly the same as in Theorem 13.6.1 (these facts also follow from the respective assertions for X^*).

Verifying the conditions of Theorem 13.6.1 in the case where $n_0 > 1$ or $\boldsymbol{\mu}(V) = 0$ causes no additional difficulties and we leave it to the reader.

The theorem is proved. □

Note that in a way similar to that in the proof of Theorem 13.4.1, one could also establish the uniqueness of the solution to the integral equation for the invariant measure (see Definition 13.6.6) in a wider class of signed finite measures.

The main and most difficult to verify conditions of Theorem 13.7.1 are undoubtedly conditions (I) and (II). Condition (I_0) is usually obtained "automatically", in the course of verifying condition (I), for the latter requires bounding $\mathbf{E}\tau_V(x)$ for all x. Verifying the aperiodicity condition (III) usually causes no difficulties. If, say, recurrence to the set V is possible in m_1 and m_2 steps and g.c.d. $(m_1, m_2) = 1$, then the chain is aperiodic.

13.7.2 On Conditions (I) and (II)

Now we consider in more detail the main conditions (I) and (II). Condition (II) is expressed directly in terms of local characteristics of the chain (transition probabilities in one or a fixed number of steps $n_0 > 1$), and in this sense it could be treated as a "final" one. One only needs to "guess" the most appropriate set V and measure $\boldsymbol{\mu}$ (of course, if there are any). For example, for multi-dimensional Markov chains in $\mathcal{X} = \mathbb{R}^d$, condition (II) will be satisfied if at least one of the following two conditions is met.

(II_a) *The distribution of $X_{n_0}(x)$ has, for some n_0 and $N > 0$ and all $x \in V_N := \{y : |y| \le N\}$, a component which is absolutely continuous with respect to Lebesgue measure* (*or to the sum of the Lebesgue measures on $\mathbb{R}^d$ and its "coordinate" subspaces*) *and is "uniformly" positive on the set V_M for some $M > 0$*. In this case, one can take $\boldsymbol{\mu}$ to be the uniform distribution on V_M.

(II_l) $\mathcal{X} = \mathbb{Z}^d$ *is the integer lattice in* $\mathbb{R}^d$. In this case the chain is countable and everything simplifies (see Sect. 13.4).

We have already noted that, in the cases when a chain has a positive atom, which is the case in Example 13.6.2, no assumptions about the structure (smoothness) of the distribution of $X_{n_0}(x)$ are needed.

The "positivity" condition (I) is different. It is given in terms of rather complicated characteristics $\mathbf{E}\tau_V(x)$ requiring additional analysis and a search for conditions in terms of local characteristics which would ensure (I). The rest of the section will mostly be devoted to this task.

First of all, we will give an "intermediate" assertion which will be useful for the sequel. We have already made use of such an assertion in Example 13.6.2.

Theorem 13.7.2 *Suppose there exists a nonnegative measurable function $g : \mathcal{X} \to \mathbb{R}$ such that the following conditions* (I^g) *are met*:

($\mathrm{I}^g)_1$ $\mathbf{E}\tau_V(x) \le c_1 + c_2 g(x)$ *for* $x \in V^c = \mathcal{X} \setminus V$, $c_1, c_2 = \mathit{const}$.

$(\mathrm{I}^g)_2$ $\sup_{x\in V}$ $\mathbf{E}g(X_1(x)) < \infty$.
Then conditions (I_0) *and* (I) *are satisfied.*

The function g from Theorem 13.7.2 is often called the *test*, or *Lyapunov, function*. For brevity's sake, put $\tau_V(x) := \tau(x)$.

Proof If (I^g) holds then, for $x \in V$,

$$\begin{aligned}\mathbf{E}\tau(x) &\le 1 + \mathbf{E}\big[\tau\big(X_1(x)\big); X_1(x) \in V^c\big]\\ &\le 1 + \mathbf{E}\big(\mathbf{E}\big[\tau\big(X_1(x)\big)\,|\,X_1(x)\big]; X_1(x) \in V^c\big)\\ &\le 1 + \mathbf{E}\big(c_1 + c_2 g\big(X_1(x)\big); X_1(x) \in V^c\big)\\ &\le 1 + c_1 + c_2 \sup_{x\in V} \mathbf{E}g\big(X_1(x)\big) < \infty.\end{aligned}$$

The theorem is proved. □

Note that condition $(\mathrm{I}^g)_2$, like condition (II), refers to "local" characteristics of the system, and in that sense it can also be treated as a "final" condition (up to the choice of function g).

We now consider conditions ensuring $(\mathrm{I}^g)_1$. The processes

$$\{X_n\} = \big\{X_n(x)\big\}, \quad X_0(x) = x,$$

to be considered below (for instance, in Theorem 13.7.3) do not need to be Markovian. We will only use those properties of the processes which will be stated in conditions of assertions.

We will again make use of nonnegative trial functions $g : \mathcal{X} \to R$ and consider a set V "induced" by the function g and a set U which in most cases will be a bounded interval of the real line:

$$V := g^{-1}(U) = \big\{x \in \mathcal{X} : g(x) \in U\big\}.$$

The notation $\tau(x) = \tau_U(x)$ will retain its meaning:

$$\tau(x) := \min\big\{k \ge 1 : g\big(X_k(x)\big) \in U\big\} = \min\big\{k \ge 1 : X_k(x) \in V\big\}.$$

The next assertion is an essential element of Lyapunov's (or the test functions) approach to the proof of positive recurrence of a Markov chain.

Theorem 13.7.3 *If* $\{X_n\}$ *is a Markov chain and, for* $x \in V^c$,

$$\mathbf{E}g\big(X_1(x)\big) - g(x) \le -\varepsilon, \tag{13.7.6}$$

then $\mathbf{E}\tau(x) \le g(x)/\varepsilon$ *and therefore* $(\mathrm{I}^g)_1$ *holds.*

To prove the theorem we need

Lemma 13.7.2 *If, for some* $\varepsilon > 0$, *all* $n = 0, 1, 2, \ldots$, *and any* $x \in V^c$,

$$\mathbf{E}\big(g(X_{n+1}) - g(X_n)|\tau(x) > n\big) \le -\varepsilon, \tag{13.7.7}$$

then

$$\mathbf{E}\tau(x) \le \frac{g(x)}{\varepsilon}, \qquad x \in V^c,$$

and therefore $(\mathrm{I}^g)_1$ *holds.*

Proof Put $\tau(x) := \tau$ for brevity and set

$$\tau_{(N)} := \min(\tau, N), \qquad \Delta(n) := g(X_{n+1}) - g(X_n).$$

We have

$$\begin{aligned} -g(x) &= -\mathbf{E}g(X_0) \le \mathbf{E}\big(g(X_{\tau_{(N)}}) - g(X_0)\big) \\ &= \mathbf{E}\sum_{n=0}^{\tau_{(N)}-1} \Delta(n) = \sum_{n=0}^{N} \mathbf{E}\Delta(n) I(\tau > n) \\ &= \sum_{n=0}^{N} \mathbf{P}(\tau > n)\mathbf{E}\big(\Delta(n)|\tau > n\big) \le -\varepsilon \sum_{n=0}^{N} \mathbf{P}(\tau > n). \end{aligned}$$

This implies that, for any N,

$$\sum_{n=0}^{N} \mathbf{P}(\tau > n) \le \frac{g(x)}{\varepsilon}.$$

Therefore this inequality will also hold for $N = \infty$, so that $\mathbf{E}\tau \le g(x)/\varepsilon$. The lemma is proved. □

Proof of Theorem 13.7.3 The proof follows in an obvious way from the fact that, by (13.7.6) and the homogeneity of the chain, $\mathbf{E}(g(X_{n+1}) - g(X_n)|\, X_n) \le -\varepsilon$ holds on $\{X_n \in V^c\}$, and from inclusion $\{\tau > n\} \subset \{X_n \in V^c\}$, so that

$$\mathbf{E}\big(g(X_{n+1}) - g(X_n); \tau > n\big) = \mathbf{E}\big[\mathbf{E}\big(g(X_{n+1}) - g(X_n)|X_n\big); \tau > n\big] \le -\varepsilon\mathbf{P}(\tau > n).$$

The theorem is proved. □

Theorem 13.7.3 is a modification of the positive recurrence criterion known as the Foster–Moustafa–Tweedy criterion (see, e.g., [6, 27]).

Consider some applications of the obtained results. Let X be a Markov chain on the real half-axis $\mathbb{R}_+ = [0, \infty)$. For brevity's sake, put $\xi(x) := X_1(x) - x$. This is the one-step increment of the chain starting at the point x; we could also define $\xi(x)$ as a random variable with the distribution

$$\mathbf{P}\big(\xi(x) \in B\big) = P(x, B - x) \quad \big(B - x = \{y \in \mathcal{X} : y + x \in B\}\big).$$

Corollary 13.7.1 *If, for some* $N \ge 0$ *and* $\varepsilon > 0$,

$$\sup_{x \le N} \mathbf{E}\xi(x) < \infty, \qquad \sup_{x > N} \mathbf{E}\xi(x) \le -\varepsilon, \tag{13.7.8}$$

then conditions (I_0) *and* (I) *hold for* $V = [0, N]$.

Proof Make use of Theorems 13.7.2, 13.7.3 and Corollary 13.3.1 with $g(x) \equiv x$, $V = [0, N]$. Conditions $(\mathrm{I}^g)_2$ and (13.7.6) are clearly satisfied. □

Thus the presence of a "negative drift" in the region $x > N$ guarantees positivity of the chain. However, that condition (I) is met could also be ensured when the "drift" $\mathbf{E}\xi(x)$ vanishes as $x \to \infty$.

Corollary 13.7.2 *Let* $\sup_x \mathbf{E}\xi^2(x) < \infty$ *and*

$$\mathbf{E}\xi^2(x) \le \beta, \qquad \mathbf{E}\xi(x) \le -\frac{c}{x} \quad \text{for } x > N.$$

If $2c > \beta$ *then conditions* (I_0) *and* (I) *hold for* $V = [0, N]$.

Proof We again make use of Theorems 13.7.2 and 13.7.3, but with $g(x) = x^2$. We have for $x > N$:

$$\mathbf{E}g\big(X_1(x)\big) - g(x) = \mathbf{E}\big(2x\xi(x) + \xi^2(x)\big) \le -2c + \beta < 0. \qquad \square$$

Before proceeding to examples related to ergodicity we note the following. The "larger" the set V the easier it is to verify condition (I), and the "smaller" that set, the easier it is to verify condition (II). In this connection there arises the question of when one can consider two sets: a "small" set W and a "large" set $V \supset W$ such that if (I) holds for V and (II) holds for W then both (I) and (II) would hold for W. Under conditions of Sect. 13.6 one can take W to be a "one-point" atom x_0.

Lemma 13.7.3 *Let sets* V *and* W *be such that the condition*

$$(\mathrm{I}_V) \qquad E := \sup_{x \in V} \mathbf{E}\tau_V(x) < \infty$$

holds and there exists an m *such that*

$$\inf_{x \in V} \mathbf{P}\left(\bigcup_{j=1}^{m} \big\{X_j(x) \in W\big\}\right) \ge q > 0.$$

Then the following condition is also met:

$$(\mathrm{I}_W) \qquad \sup_{x \in W} \mathbf{E}\tau_W(x) \le \sup_{x \in V} \mathbf{E}\tau_W(x) \le \frac{mE}{q}.$$

Thus, under the assumptions of Lemma 13.7.3, if condition (I) holds for V and condition (II) holds for W, then conditions (I) and (II) hold for W.

To prove Lemma 13.7.3, we will need the following assertion extending (in the form of an inequality) the well-known Wald identity.

Assume we are given a sequence of nonnegative random variables $\tau_1, \tau_2, \dots$ which are measurable with respect to σ-algebras $\mathfrak{U}_1 \subset \mathfrak{U}_2 \subset \cdots$, respectively, and let $T_n := \tau_1 + \cdots + \tau_n$. Furthermore, let ν be a given stopping time with respect to $\{\mathfrak{U}_n\}$: $\{\nu \le n\} \in \mathfrak{U}_n$.

Lemma 13.7.4 *If* $\mathbf{E}(\tau_n|\mathfrak{U}_{n-1}) \le a$ *then* $\mathbf{E}T_\nu \le a\mathbf{E}\nu$.

Proof We can assume without loss of generality that $\mathbf{E}\nu < \infty$ (otherwise the inequality is trivial). The proof essentially repeats that of Theorem 4.4.1. One has

$$\mathbf{E}\tau_\nu = \sum_{k=1}^{\infty} \mathbf{E}(T_k;\, \nu = k) = \sum_{k=1}^{\infty} \mathbf{E}(\tau_k,\, \nu \ge k). \tag{13.7.9}$$

Changing the summation order here is well-justified, for the summands are nonnegative. Further, $\{\nu \le k-1\} \in \mathfrak{U}_{k-1}$ and hence $\{\nu \ge k\} \in \mathfrak{U}_{k-1}$. Therefore

$$\mathbf{E}(\tau_k;\, \nu \ge k) = \mathbf{E}\mathrm{I}(\nu \ge k)\mathbf{E}(\tau_k|\mathfrak{U}_{k-1}) \le a\mathbf{P}(\nu \ge k).$$

Comparing this with (13.7.9) we get

$$\mathbf{E}T_\nu \le a \sum_{k=1}^{\infty} \mathbf{P}(\nu \ge k) = a\mathbf{E}nu.$$

The lemma is proved. □

Proof of Lemma 13.7.3 Suppose the chain starts at a point $x \in V$. Consider the times $T_1, T_2, \dots$ of successive visits of X to V, $T_0 = 0$. Put $Y_0 := x$, $Y_k := X_{T_k}(x)$, $k = 1, 2, \dots$. Then, by virtue of the strong Markov property, the sequence (Y_k, T_k) will form a Markov chain. Set $\mathfrak{U}_k := \sigma(T_1, \dots, T_k; Y_1, \dots, Y_k)$, $\tau_k := T_k - T_{k-1}$, $k = 1, 2 \dots$. Then $\nu := \min\{k : Y_k \in W\}$ is a stopping time with respect to $\{\mathfrak{U}_k\}$. It is evident that $\mathbf{E}(\tau_k|\mathfrak{U}_{k-1}) \le E$. Bound $\mathbf{E}\nu$. We have

$$p_k := \mathbf{P}(\nu \ge km) \le \mathbf{P}\left(\bigcap_{j=1}^{T_{km}} \{X_j \notin W\}\right)$$

$$= \mathbf{E}\mathrm{I}\left(\bigcap_{j=1}^{T_{(k-1)m}} \{X_j \notin W\}\right)\mathbf{E}\left(\mathrm{I}\left(\bigcap_{j=T_{(k-1)m+1}}^{T_{km}} \{X_j \notin W\}\right)\Big|\mathfrak{U}_{(k-1)m}\right).$$

Since $\tau_j \ge 1$, the last factor, by the assumptions of the lemma and the strong Markov property, does not exceed

$$\mathbf{P}\left(\bigcap_{j=1}^{m}\left\{X_j^{\text{new}}\left(X_{T_{(k-1)m}}\right) \notin W\right\}\right) \le (1-q),$$

where, as before, $X_k^{\text{new}}(x)$ is a chain with the same distribution as $X_k(x)$ but independent of the latter chain. Thus $p_k \le (1-q)p_{k-1} \le (1-q)^k$, $\mathbf{E}\nu \le m/q$, and by Lemma 13.7.4 we have $\mathbf{E}T_\nu \le E_m/q$. It remains to notice that $\tau_W(x) = T_\nu$. The lemma is proved. □

Example 13.7.1 A random walk with reflection. Let $\xi_1, \xi_2, \dots$ be independent identically distributed random variables,

$$X_{n+1} := |X_n + \xi_{n+1}|, \quad n = 0, 1, \dots. \tag{13.7.10}$$

If the ξ_k and hence the X_k are non-arithmetic, then the chain X has, generally speaking, no atoms. If, for instance, ξ_k have a density $f(t)$ with respect to Lebesgue measure then $\mathbf{P}(X_k(x) = y) = 0$ for any $x, y, k \ge 1$. We will assume that a broader condition (A) holds:

(A). *In the decomposition*

$$\mathbf{P}(\xi_k < t) = p_a F_a(t) + p_c F_c(t)$$

of the distribution of ξ_k *into the absolutely continuous* (F_a) *and singular* (F_c) (*including discrete*) *components, one has* $p_a > 0$.

Corollary 13.7.3 *If condition* (A) *holds,* $a = \mathbf{E}\xi_k < 0$, *and* $\mathbf{E}|\xi_k| < \infty$, *then the Markov chain defined in* (13.7.10) *satisfies the conditions of Theorem* 13.7.2 *and therefore is ergodic in the sense of convergence in total variation.*

Proof We first verify that the chain satisfies the conditions of Corollary 13.7.1. Since in our case $|X_1(x) - x| \le |\xi_1|$, the first of conditions (13.7.8) is satisfied. Further,

$$\mathbf{E}\xi(x) = \mathbf{E}|x + \xi_1| - x = \mathbf{E}(\xi_1; \xi_1 \ge -x) - \mathbf{E}(2x + \xi_1; \xi_1 < -x) \to \mathbf{E}\xi_1$$

as $x \to \infty$, since

$$x\mathbf{P}(\xi_1 < -x) \le \mathbf{E}\big(|\xi_1|, |\xi_1| > x\big) \to 0.$$

Hence there exists an N such that $\mathbf{E}\xi(x) \le a/2 < 0$ for $x \ge N$. This proves that conditions (I_0) and (I) hold for $V = [0, N]$.

Now verify that condition (II) holds for the set $W = [0, h]$ with some h. Let $f(t)$ be the density of the distribution F_a from condition (A). There exist an $f_0 > 0$ and a segment $[t_1, t_2]$, $t_2 > t_1$, such that $f(t) > f_0$ for $t \in [t_1, t_2]$. The density of $x + \xi_1$

will clearly be greater than f_0 on $[x+t_1, x+t_2]$. Put $h := (t_2 - t_1)/2$. Then, for $0 \le x \le h$, one will have $[t_2 - h, t_2] \subset [x+t_1, x+t_2]$.

Suppose first that $t_2 > 0$. The aforesaid will then mean that the density of $x + \xi_1$ will be greater than f_0 on $[(t_2 - h)^+, t_2]$ for all $x \le h$ and, therefore,

$$\inf_{x \le h} \mathbf{P}\big(X_1(x) \in B\big) \ge p_1 \int_B f_0(t)\,dt,$$

where

$$f_0(t) = \begin{cases} f_0 & \text{if} t \in [(t_2 - h)^+, t_2], \\ 0 & \text{otherwise.} \end{cases}$$

This means that condition (II) is satisfied on the set $W = [0, h]$. The case $t_2 \le 0$ can be considered in a similar way.

It remains to make use of Lemma 13.7.3 which implies that condition (I) will hold for the set W. The condition of Lemma 13.7.3 is clearly satisfied (for sufficiently large m, the distribution of $X_m(x)$, $x \le N$, will have an absolutely continuous component which is positive on W). For the same reason, the chain X cannot be periodic. Thus all conditions of Theorem 13.7.2 are met. The corollary is proved. □

Example 13.7.2 An oscillating random walk. Suppose we are given two independent sequences $\xi_1, \xi_2, \dots$ and $\eta_1, \eta_2, \dots$ of independent random variables, identically distributed in each of the sequences. Put

$$X_{n+1} := \begin{cases} X_n + \xi_{n+1} & \text{if } X_n \ge 0, \\ X_n + \eta_{n+1} & \text{if } X_n < 0. \end{cases} \tag{13.7.11}$$

Such a random walk is called *oscillating*. It clearly forms a Markov chain in the state space $\mathcal{X} = (-\infty, \infty)$.

Corollary 13.7.4 *If at least one of the distributions of ξ_k or η_k satisfies condition* (A) *and* $-\infty < \mathbf{E}\xi_k < 0$, $\infty > \mathbf{E}\eta_k > 0$, *then the chain* (13.7.11) *will satisfy the conditions of Theorem* 13.7.2 *and therefore will be ergodic.*

Proof The argument is quite similar to the proof of Corollary 13.7.3. One just needs to take, in order to verify condition (I), $g(x) = |x|$ and $V = [-N, N]$. After that it remains to make use of Lemma 13.7.3 with $W = [0, h]$ if condition (A) is satisfied for ξ_k (and with $W = [-h, 0)$ if it is met for η_k). □

Note that condition (A) in Examples 13.7.1 and 13.7.2 can be relaxed to that of the existence of an absolutely continuous component for the distribution of the sum $\sum_{j=1}^m \xi_j$ (or $\sum_{j=1}^m \eta_j$) for some m. On the other hand, if the distributions of these sums are singular for all m, then convergence of distributions $P(x, n, \cdot)$ in total variation cannot take place. If, for instance, one has $\mathbf{P}(\xi_k = -\sqrt{2}) = \mathbf{P}(\xi_k = 1) = 1/2$ in Example 13.7.1, then $\mathbf{E}\xi_k < 0$ and condition (I) will be met, while condition (II) will not. Convergence of $P(x,n,\cdot)$ in total variation to the limiting distribution $\boldsymbol{\pi}$ is also impossible. Indeed, it follows from the equation for the invariant distribution $\boldsymbol{\pi}$ that this distribution is necessarily continuous. On the other hand, say, the distributions $P(0,n,\cdot)$ are concentrated on the countable set $\mathcal{N}$ of the numbers $|-k\sqrt{2}+l|$; $k,l=1,2,\ldots$. Therefore $\mathbf{P}(0,n,\mathcal{N})=1$ for all n, $\pi(\mathcal{N})=0$. Hence only weak convergence of the distributions $\mathbf{P}(x,n,\cdot)$ to $\boldsymbol{\pi}(\cdot)$ may take place. And although this convergence does not raise any doubts, we know no reasonably simple proof of this fact.

Example 13.7.3 (continuation of Examples 13.4.2 and 13.6.1) Let $\mathcal{X}=[0,1]$, $\xi_1,\xi_2,\ldots$ be independent and identically distributed, and $X_{n+1}:=X_n+\xi_{n+1}$ (mod 1) or, which is the same, $X_{n+1}:=\{X_n+\xi_{n+1}\}$, where $\{x\}$ denotes the fractional part of x. Here, condition (I) is clearly met for $V=\mathcal{X}=[0,1]$. If the ξ_k satisfy condition (A) then, as was the case in Example 13.7.1, condition (II) will be met for the set $W=[0,h]$ with some $h>0$, which, together with Lemma 13.7.3, will mean, as before, that the conditions of Theorem 13.7.2 are satisfied. The invariant distribution $\boldsymbol{\pi}$ will in this example be uniform on $[0,1]$. For simplicity's sake, we can assume that the distribution of ξ_k has a density $f(t)$, and without loss of generality we can suppose that $\xi_k\in[0,1]$ ($f(t)=0$ for $t\notin[0,1]$). Then the density $p(x)\equiv 1$ of the invariant measure π will satisfy the equation for the invariant measure:

$$p(x)=1=\int_0^x dy\, f(x-y)+\int_x^1 dy\, f(x-y+1)=\int_0^1 f(y)\,dy.$$

Since the stationary distribution is unique, one has $\boldsymbol{\pi}=U_{0,1}$. Moreover, by Theorem A3.4.1 of Appendix 3, along with convergence of $P(x,n,\cdot)$ to $U_{0,1}$ in total variation, convergence of the densities $P(x,n,dt)/dt$ to 1 in (Lebesgue) measure will take place.

The fact that the invariant distribution is uniform remains true for arbitrary non-lattice distributions of ξ_k. However, as we have already mentioned in Example 13.6.1, in the general case (without condition (A)) only weak convergence of the distributions $P(x,n,\cdot)$ to the uniform distribution is possible (see [6, 18]).

13.8 Laws of Large Numbers and the Central Limit Theorem for Sums of Random Variables Defined on a Markov Chain

13.8.1 Random Variables Defined on a Markov Chain

Let, as before, $X=\{X_n\}$ be a Markov Chain in an arbitrary measurable state space $\langle\mathcal{X},\mathfrak{B}_{\mathcal{X}}\rangle$ defined in Sect. 13.6, and let a measurable function $f\colon\mathcal{X}\to\mathbb{R}$ be given on $\langle\mathcal{X},\mathfrak{B}_{\mathcal{X}}\rangle$. The sequence of sums

$$S_n:=\sum_{k=1}^n f(X_k) \tag{13.8.1}$$

is a generalisation of the random walks that were studied in Chaps. 8 and 11. One can consider an even more general problem on the behaviour of *sums of random variables defined on a Markov chain*. Namely, we will assume that a collection of distributions $\{\mathbf{F}_x\}$ is given which depend on the parameter $x \in \mathcal{X}$. If $F_x^{(-1)}(t)$ is the quantile transform of $\mathbf{F}_x$ and $\omega \Subset \mathbf{U}_{0,1}$, then $\xi_x := F_x^{(-1)}(\omega)$ will have the distribution $\mathbf{F}_x$ (see Sect. 3.2.4).

The mapping $\mathbf{F}_x$ of the space $\mathcal{X}$ into the set of distributions is assumed to be such that the function $\xi_x(t) = F_x^{(-1)}(t)$ is measurable on $\mathcal{X} \times \mathbb{R}$ with respect to $\mathfrak{B}_{\mathcal{X}} \times \mathfrak{B}$, where $\mathfrak{B}$ is the σ-algebra of Borel sets on the real line. In this case, $\xi_x(\omega)$ will be a random variable such that the moments

$$\mathbf{E}\xi_x^s = \int_{-\infty}^{\infty} v^s d\mathbf{F}_x(v) = \int_0^1 \left[F_x^{(-1)}(u)\right]^s du$$

are measurable with respect to $\mathfrak{B}_{\mathcal{X}}$ (and hence will be random variables themselves if we set a distribution on $\langle \mathcal{X}, \mathfrak{B}_{\mathcal{X}} \rangle$).

Definition 13.8.1 If $\omega_i \Subset \mathbf{U}_{0,1}$ are independent then the sequence

$$\xi_{X_n} := F_{X_n}^{(-1)}(\omega_n), \quad n = 0, 1, \ldots,$$

is called a *sequence of random variables defined on the Markov chain* $\{X_n\}$.

The basic objects of study in this section are the asymptotic properties of the distributions of the sums

$$S_n := \sum_{k=0}^{n} \xi_{X_k}. \tag{13.8.2}$$

If the distribution $\mathbf{F}_x$ is degenerate and concentrated at the point $f(x)$ then (13.8.2) turns into the sum (13.8.1). If the chain X is countable with states $E_0, E_1, \ldots$ and $f(x) = \mathrm{I}(E_j)$ then $S_n = m_j(n)$ is the number of visits to the state E_j by the time n considered in Theorem 13.4.4.

13.8.2 Laws of Large Numbers

In this and the next subsection we will confine ourselves to Markov chains satisfying the ergodicity conditions from Sects. 13.6 and 13.7. As was already noticed, ergodicity conditions for Harris chains mean, in essence, the existence of a positive atom (possibly in the extended state space). Therefore, for the sake of simplicity, we will assume from the outset that the chain X has a positive atom at a point x_0 and put, as before,

$$\tau(x) := \min\bigl\{k \ge 0 : X_k(x) = x_0\bigr\}, \quad \tau(x_0) = \tau.$$

Summing up the conditions sufficient for (I_0) and (I) to hold (the finiteness of $\tau(x)$ and $\mathbf{E}\tau$) studied in Sect. 13.7, we obtain the following assertion in our case.

Corollary 13.8.1 *Let there exist a set* $V \in \mathfrak{B}_{\mathcal{X}}$ *such that, for the stopping time* $\tau_V(x) := \min\{k : X_k(x) \in V\}$, *we have*

$$E := \sup_{x \in V} \mathbf{E}\tau_V(x) < \infty. \tag{13.8.3}$$

Furthermore, let there exist an $m \geq 1$ *such that*

$$\inf_{x \in V} \mathbf{P}\Bigg(\bigcup_{j=1}^{m} \big\{X_j(x) = x_0\big\}\Bigg) \geq q > 0.$$

Then

$$\mathbf{E}\tau \leq \frac{mE}{q}.$$

This assertion follows from Lemma 13.7.2. One can justify conditions (I_0) and (13.8.3) by the following assertion.

Corollary 13.8.2 *Let there exist an* $\varepsilon > 0$ *and a nonnegative measurable function* $g : \mathcal{X} \to \mathbb{R}$ *such that*

$$\sup_{x \in V} \mathbf{E}g\big(X_1(x)\big) < \infty$$

and, for $x \in V^c$,

$$\mathbf{E}g\big(X_1(x)\big) - g(x) \leq -\varepsilon.$$

Then conditions (I_0) *and* (13.8.3) *are met.*

In order to formulate and prove the law of large numbers for the sums (13.8.2), we will use the notion of the increment of the sums (13.8.2) on a cycle between consequent visits of the chain to the atom x_0. Divide the trajectory $X_0, X_1, X_2, \ldots, X_n$ of the chain X on the time interval $[0, n]$ into segments of lengths $\tau_1 := \tau(x), \tau_2, \tau_3, \ldots$ ($\tau_j \overset{d}{=} \tau$ for $j \geq 2$) corresponding to the visits of the chain to the atom x_0. Denote the increment of the sum S_n on the k-th cycle (on $(T_{k-1}, T_k]$) by ζ_k:

$$\begin{aligned} \zeta_1 &:= \sum_{j=0}^{\tau_1} \xi_{X_j}, \\ \zeta_k &:= \sum_{j=T_{k-1}+1}^{T_k} \xi_{X_j}, \ k \geq 2, \quad \text{where } T_k := \sum_{j=1}^{k} \tau_j, \ k \geq 1, \ T_0 = 0. \end{aligned} \tag{13.8.4}$$

The vectors (τ_k, ζ_k), $k \geq 2$, are clearly independent and identically distributed. For brevity, the index k will sometimes be omitted: $(\tau_k, \zeta_k) \overset{d}{=} (\tau, \zeta)$ for $k \geq 2$.

Now we can state the law of large numbers for the sums (13.8.2).

Theorem 13.8.1 *Let* $\mathbf{P}(\tau(x) < \infty) = 1$ *for all* x, $\mathbf{E}\tau < \infty$, $\mathbf{E}|\zeta| < \infty$, *and the g.c.d. of all possible values of* τ *equal* 1. *Then*

$$\frac{S_n}{n} = \frac{1}{n}\sum_{k=1}^{n} \xi_{X_k} \xrightarrow{p} \frac{\mathbf{E}\zeta}{\mathbf{E}\tau} \quad as \quad n \to \infty.$$

Proof Put

$$\nu(n) := \max\{k : T_k \le n\}.$$

Then the sum S_n can be represented as

$$S_n = \zeta_1 + Z_{\nu(n)} + z_n, \tag{13.8.5}$$

where

$$Z_k := \sum_{j=2}^{k} \zeta_j, \quad z_n := \sum_{j=T_{\nu(n)}+1}^{n} \xi_{X_j}.$$

Since τ_1 and ζ_1 are proper random variables, we have, as $n \to \infty$,

$$\frac{\zeta_1}{n} \xrightarrow{a.s.} 0. \tag{13.8.6}$$

The sum z_n consists of $\gamma(n) := n - T_{\nu(n)}$ summands. Theorem 10.3.1 implies that the distribution of $\gamma(n)$ converges to a proper limiting distribution, and the same is true for z_n. Hence, as $n \to \infty$,

$$\frac{z_n}{n} \xrightarrow{p} 0. \tag{13.8.7}$$

The sums $Z_{\nu(n)}$, being the main part of (13.8.5), are nothing else but a generalised renewal process corresponding to the vectors (τ, ζ) (see Sect. 10.6).

Since $\mathbf{E}\tau < \infty$, by Theorem 11.5.2, as $n \to \infty$,

$$\frac{Z_{\nu(n)}}{n} \xrightarrow{p} \frac{\mathbf{E}\zeta}{\mathbf{E}\tau}. \tag{13.8.8}$$

Together with (13.8.6) and (13.8.7) this means that

$$\frac{S_n}{n} \xrightarrow{p} \frac{\mathbf{E}\zeta}{\mathbf{E}\tau}. \tag{13.8.9}$$

The theorem is proved. □

As was already noted, sufficient conditions for $\mathbf{P}(\tau(x) < \infty) = 1$ and $\mathbf{E}\tau < \infty$ to hold are contained in Corollaries 13.8.1 and 13.8.2. It is more difficult to find conditions sufficient for $\mathbf{E}\zeta < \infty$ that would be adequate for the nature of the problem.

Below we will obtain certain relations which clarify, to some extent, the connection between the distributions of ζ and τ and the stationary distribution of the chain X.

Theorem 13.8.2 (A generalisation of the Wald identity) *Assume* $\mathbf{E}\tau < \infty$, *the g.c.d. of all possible values of* τ *be* 1, $\boldsymbol{\pi}$ *be the stationary distribution of the chain* X, *and*

$$\mathbf{E}_{\boldsymbol{\pi}}\mathbf{E}|\xi_x| := \int \mathbf{E}|\xi_x|\boldsymbol{\pi}(dx) < \infty. \tag{13.8.10}$$

Then

$$\mathbf{E}\zeta = \mathbf{E}\tau\mathbf{E}_{\boldsymbol{\pi}}\mathbf{E}\xi_x. \tag{13.8.11}$$

The value of $\mathbf{E}_{\boldsymbol{\pi}}\mathbf{E}\xi_x$ is the "doubly averaged" value of the random variable ξ_x: over the distribution $\mathbf{F}_x$ and over the stationary distribution $\boldsymbol{\pi}$.

Theorem 13.8.2 implies that the condition $\sup_x \mathbf{E}|\xi_x| < \infty$ is sufficient for the finiteness of $\mathbf{E}|\zeta|$.

Proof [of Theorem 13.8.2] First of all, we show that condition (13.8.10) implies the finiteness of $\mathbf{E}|\zeta|$. If $\xi_x \ge 0$ then $\mathbf{E}\zeta$ is always well-defined. If we assume that $\mathbf{E}\zeta = \infty$ then, repeating the proof of Theorem 13.8.1, we would easily obtain that, in this case, $S_n/n \xrightarrow{p} \infty$, and hence necessarily $\mathbf{E}S_n/n \to \infty$ as $n \to \infty$. But

$$\mathbf{E}S_n = \sum_{j=0}^{n} \mathbf{E}\xi_{X_j} = \sum_{j=0}^{n} \int (\mathbf{E}\xi_x)\mathbf{P}(X_j \in dx),$$

where the distribution $\mathbf{P}(X_j \in \cdot)$ converges in total variation to $\boldsymbol{\pi}(\cdot)$ as $j \to \infty$,

$$\int (\mathbf{E}\xi_x)\mathbf{P}(X_j \in dx) \to \int (\mathbf{E}\xi_x)\boldsymbol{\pi}(dx),$$

and hence

$$\frac{1}{n}\mathbf{E}S_n \to \mathbf{E}_{\boldsymbol{\pi}}\mathbf{E}\xi_x < \infty. \tag{13.8.12}$$

This contradicts the above assumption, and therefore $\mathbf{E}\zeta < \infty$. Applying the above argument to the random variables $|\xi_x|$, we conclude that condition (13.8.10) implies $\mathbf{E}|\zeta| < \infty$.

Let, as above, $\eta(n) := \nu(n) + 1 = \min\{k : T_k > n\}$. We will need the following.

Lemma 13.8.5 *If* $\mathbf{E}|\zeta| < \infty$ *then*

$$\mathbf{E}\zeta_{\eta(n)} = o(n). \tag{13.8.13}$$

If $\mathbf{E}\zeta^2 < \infty$ *then*

$$\mathbf{E}\zeta^2_{\eta(n)} = o(n) \tag{13.8.14}$$

as $n \to \infty$.

Proof Without losing generality, assume that $\xi_x \ge 0$ and $\zeta \ge 0$. Since $\tau_j \ge 1$, we have

$$h(k) := \sum_{j=0}^{k} \mathbf{P}(T_j = k) \le 1 \quad \text{for all } k.$$

Therefore,

$$\mathbf{P}(\zeta_{\eta(n)} > v) = \sum_{k=0}^{n} h(k)\mathbf{P}(\zeta > v, \tau > n-k) \le \sum_{k=0}^{n} \mathbf{P}(\zeta > v, \tau > k).$$

If $\mathbf{E}\zeta < \infty$ then

$$\mathbf{E}\zeta_{\eta(n)} \le \sum_{k=0}^{n} \int_0^\infty \mathbf{P}(\zeta > v; \tau > k)\, dv = \sum_{k=0}^{n} \mathbf{E}(\zeta; \tau > k), \tag{13.8.15}$$

where $\mathbf{E}(\zeta; \tau > k) \to 0$ as $k \to \infty$. This follows from Lemma A3.2.3 of Appendix 3. Together with (13.8.15) this proves (13.8.13).

Similarly, for $\mathbf{E}\zeta^2 < \infty$,

$$\mathbf{E}\zeta^2_{\eta(n)} \le 2\sum_{k=0}^{n} \int_0^\infty v\mathbf{P}(\zeta > v, \tau > k)\, dv = \sum_{k=0}^{n} \mathbf{E}\bigl(\zeta^2, \tau > k\bigr) = o(n).$$

The lemma is proved. □

Now we continue the proof of Theorem 13.8.2. Consider representation (13.8.5) for $X_0 = x_0$ and assume again that $\xi_x \ge 0$. Then $\zeta_1 = \xi_{x_0}$,

$$S_n = \zeta_1 + Z_{\eta(n)} + z_n - \zeta_{\eta(n)},$$

where by the Wald identity

$$\mathbf{E}Z_{\eta(n)} = \mathbf{E}\eta(n)\mathbf{E}\zeta \sim n\frac{\mathbf{E}\zeta}{\mathbf{E}\tau}.$$

Since $\boldsymbol{\pi}(\{x_0\}) = 1/\mathbf{E}\tau > 0$, we have, by (13.8.10), $\mathbf{E}|\xi_{x_0}| < \infty$. Moreover, for $\xi_x \ge 0$,

$$|\zeta_{\eta(n)} - z_n| < \zeta_{\eta(n)}.$$

Hence, by Lemma 13.8.5,

$$\mathbf{E}S_n = n\frac{\mathbf{E}\zeta}{\mathbf{E}\tau} + o(n). \tag{13.8.16}$$

Combining this with (13.8.12), we obtain the assertion of the theorem.

It remains to consider the case where ξ_x can take values of both signs. Introduce new random variables ξ^*_x on the chain X, defined by the equalities $\xi^*_x := |\xi_x|$, and endow with the superscript $*$ all already used notations that will correspond to the new random variables. Since all $\xi_x^* \geq 0$, by condition (13.8.10) we can apply to them all the above assertions and, in particular, obtain that

$$\mathbf{E}\zeta^* < \infty, \qquad \mathbf{E}\zeta^*_{\eta(n)} = o(n). \tag{13.8.17}$$

Since

$$|\zeta| \leq \zeta^*, \qquad |\zeta_{\eta(n)}| \leq \zeta^*_{\eta(n)}, \qquad |\zeta_{\eta(n)} - z_n| < \zeta^*_{\eta(n)},$$

it follows from (13.8.17) that

$$\mathbf{E}|\zeta| < \infty, \qquad \mathbf{E}|\zeta_{\eta(n)} - z_n| = o(n)$$

and relation (13.8.16) is valid along with identity (13.8.11).

The theorem is proved. □

Now we will prove the strong law of large numbers.

Theorem 13.8.3 *Let the conditions of Theorem 13.8.1 be satisfied. Then*

$$\frac{S_n}{n} \xrightarrow{a.s.} \mathbf{E}_\pi \mathbf{E}\xi_x \quad \text{as } n \to \infty.$$

Proof Since in representation (13.8.5) one has $\zeta_1/n \xrightarrow{a.s.} 0$ as $n \to \infty$, we can neglect this term in (13.8.5).

The strong laws of large numbers for $\{Z_k\}$ and $\{T_k\}$ mean that, for a given $\varepsilon > 0$, the trajectory of $\{S_{T_k}\}$ will lie within the boundaries $k\mathbf{E}\zeta(1 \pm \varepsilon)$ and $\frac{\mathbf{E}\zeta}{\mathbf{E}\tau} T_k(1 \pm 2\varepsilon)$ for all $k \geq n$ and n large enough. (We leave a more formal formulation of this to the reader.)

We will prove the theorem if we verify that the probability of the event that, between the times T_k, $k \geq n$, the trajectory of S_j will cross at least once the boundaries $rj(1 \pm 3\varepsilon)$, where $r = \dfrac{\mathbf{E}\zeta}{\mathbf{E}\tau}$, tends to zero as $n \to \infty$. Since

$$\max_{T_{k-1} < j \leq T_k} |S_j - S_{T_k}| \leq \zeta_k^* \tag{13.8.18}$$

(in the notation of the proof of Theorem 13.8.1), it is sufficient to verify that $\mathbf{P}(A_n) \to 0$ as $n \to \infty$, where $A_n := \bigcup_{k=n}^{\infty} \{\zeta_k^* > \varepsilon r T_k\}$. But

$$\mathbf{P}(A_n) = \mathbf{P}(A_n B_n) + \mathbf{P}(A_n \overline{B}_n), \tag{13.8.19}$$

where

$$B_n = \bigcap_{k=n}^{\infty} \bigl\{T_k > k\mathbf{E}\tau(1-\varepsilon)\bigr\}, \qquad \mathbf{P}(\overline{B}_n) \to 0 \quad \text{as } n \to \infty,$$

so the second summand in (13.8.19) tends to zero. The first summand on the right-hand side of (13.8.19) does not exceed (for $c=\varepsilon(1-\varepsilon)\mathbf{E}\zeta$)

$$\mathbf{P}\Biggl(\bigcup_{k=n}^{\infty}\bigl\{\zeta_k^*>\varepsilon\mathbf{E}\zeta k(1-\varepsilon)\bigr\}\Biggr)\le\sum_{k=n}^{\infty}\mathbf{P}\bigl(\zeta_k^*>ck\bigr)\to 0$$

as $n\to\infty$, since $\mathbf{E}\zeta^*<\infty$ (see (13.8.17)). The theorem is proved. □

13.8.3 The Central Limit Theorem

As in Theorem 13.8.1, first we will prove the main assertion under certain conditions on the moments of ζ and τ, and then we will establish a connection of these conditions to the stationary distribution of the chain X. Below we retain the notation of the previous section.

Theorem 13.8.4 *Let* $\mathbf{P}(\tau(x)<\infty)=1$ *for any* x, $\mathbf{E}\tau^2<\infty$, *the g.c.d. of all possible values of* τ *is* 1, *and* $\mathbf{E}\zeta^2<\infty$. *Then, as* $n\to\infty$,

$$\frac{S_n-rn}{d\sqrt{n/a}}\Leftrightarrow\mathbf{\Phi}_{0,1},$$

where $r:=a_\zeta/a$, $a_\zeta:=\mathbf{E}\zeta$, $a:=\mathbf{E}\tau$ *and* $d^2:=\mathbf{D}(\zeta-r\tau)$.

Proof We again make use of representation (13.8.5), where clearly

$$\frac{\zeta_1}{\sqrt{n}}\xrightarrow{p}0,\qquad\frac{z_n}{\sqrt{n}}\xrightarrow{p}0$$

(see the proof of Theorem 13.8.1). This means that the problem reduces to that of finding the limiting distribution of $Z_{\nu(n)}=Z_{\eta(n)}-\zeta_{\eta(n)}$, where by Lemma 10.6.1 $\zeta_{\eta(n)}$ has a proper limiting distribution, and so $\zeta_{\eta(n)}/\sqrt{n}\xrightarrow{p}0$ as $n\to\infty$. Furthermore, by Theorem 10.6.3,

$$\frac{Z_{\eta(n)}}{\sigma_S\sqrt{n}}\Leftrightarrow\mathbf{\Phi}_{0,1},$$

where $\sigma_S^2:=a^{-1}\mathbf{D}(\zeta-r\tau)$, $r=\frac{\mathbf{E}\zeta}{\mathbf{E}\tau}$. The theorem is proved. □

Now we will establish relations between the moment characteristics used for normalising S_n and the stationary distribution $\boldsymbol{\pi}$. The answer for the number r was given in Theorem 13.8.2: $r=\mathbf{E}_{\boldsymbol{\pi}}\mathbf{E}\xi_x$. For the number σ_S^2 we have the following result.

Theorem 13.8.5 *Let*

$$\sigma^2 := \int \mathbf{D}\xi_x \boldsymbol{\pi}(dx) + 2\sum_{j=1}^{\infty} \mathbf{E}(\xi_{X_0} - r)(\xi_{X_j} - r)$$

be well-defined and finite, where $X_0 \Subset \boldsymbol{\pi}$. *Then*

$$\sigma_S^2 := a^{-1}d^2 = \sigma^2.$$

Note that here the expectation under the sum sign is a "triple averaging": over the distribution $\boldsymbol{\pi}(dy)\mathbf{P}(y, j, dz)$ and the distributions of ξ_y and ξ_z.

Proof We have

$$\mathbf{E}(S_n - rn)^2 = \mathbf{E}\left[\sum_{k=0}^{n}(\xi_{X_k} - r)\right]^2$$

$$= \sum_{k=0}^{n} \mathbf{E}(\xi_{X_k} - r)^2 + 2\sum_{k<j} \mathbf{E}(\xi_{X_k} - r)(\xi_{X_j} - r), \quad (13.8.20)$$

where

$$\sum_{k=0}^{n} \mathbf{E}(\xi_{X_k} - r)^2 = \sum_{k=0}^{n} \mathbf{E}(\xi_{X_k} - \mathbf{E}\xi_{X_k})^2 + \sum_{k=0}^{n}(\mathbf{E}\xi_{X_k} - r)^2. \quad (13.8.21)$$

The summands in the first sum on the right-hand side of (13.8.21) converge to $\sigma_\xi^2 := \int \mathbf{D}\xi_x \boldsymbol{\pi}(dx)$, the summands in the second sum converging to zero. Therefore, the left-hand side of (13.8.21) is asymptotically equivalent to $n\sigma_\xi^2$.

Further,

$$\sum_{k<j} \mathbf{E}(\xi_{X_k} - r)(\xi_{X_j} - r) = \sum_{k=0}^{n} \sum_{j \ge k+1} \mathbf{E}(\xi_{X_k} - r)(\xi_{X_j} - r), \quad (13.8.22)$$

where the distribution of X_k converges in total variation to the stationary distribution $\boldsymbol{\pi}$ of the chain. Hence the inner sums on the right-hand side of (13.8.22), for large k and $n - k$ (say, for $\sqrt{n} < k < n - \sqrt{n}$ when $n \to \infty$), will be close to

$$E := \sum_{j=1}^{\infty} \mathbf{E}(\xi_{X_0} - r)(\xi_{X_j} - r),$$

where $X_0 = \boldsymbol{\pi}$ and the whole sum on the right-hand side of (13.8.22) is asymptotically equivalent, as $n \to \infty$, to nE (or will be $o(n)$ if $E = 0$).

Thus

$$\frac{1}{n}\mathbf{E}(S_n - rn)^2 \sim \sigma_\xi^2 + 2E. \quad (13.8.23)$$

We now show that the existence of σ_ξ^2 and E implies the finiteness of $d^2 = \mathbf{E}(\zeta - r\tau)^2$.

Consider the truncated random variables

$$\xi_x^{(N)} := \begin{cases} \xi_x & \text{if } \xi_x \in [-N, N], \\ N & \text{if } \xi_x > N, \\ -N & \text{if } \xi_x < -N. \end{cases}$$

Since $\sigma_\xi^2 < \infty$, we have $\mathbf{E}\xi_x^2 < \infty$ (a.e. with respect to the measure $\boldsymbol{\pi}$) and

$$r^{(N)} \to r, \quad \big(\sigma_\xi^{(N)}\big)^2 \to \sigma_\xi^2, \quad E^{(N)} \to E \quad \text{as } N \to \infty,$$

where the superscript (N) means that the notation corresponds to the truncated random variables. By virtue of Theorem 13.8.4,

$$\liminf_{n\to\infty} \frac{1}{n} \mathbf{E}\big(S_n^{(N)} - r^{(N)}\big)^2 \ge a^{-1}\big(d^{(N)}\big)^2.$$

If we assume that $d = \infty$ then we will get that the lim inf on the left-hand side of this relation is infinite. But this contradicts relation (13.8.23), by which the above lim inf equals $(\sigma_\xi^{(N)})^2 + 2\mathbf{E}^{(N)}$ and remains bounded. We have obtained a contradiction, which shows that $d < \infty$.

On the other hand, for $d < \infty$, $\mathbf{E}\zeta^2 < \infty$ and, for the initial value x_0, by (13.8.5) we have

$$\begin{aligned} \mathbf{E}(S_n - rn)^2 &= \mathbf{E}(Z_{\nu(n)} + z_n - rn)^2 \\ &= \mathbf{E}(Z_{\eta(n)} - rn)^2 + 2\mathbf{E}(Z_{\eta(n)} - rn)(z_n - \zeta_{\eta(n)}) + \mathbf{E}(z_n - \zeta_{\eta(n)})^2, \end{aligned} \tag{13.8.24}$$

where $n = T_{\eta(n)} - \chi(n)$. Therefore, putting $Y_n := Z_n - rT_n = \sum_{k=1}^n (\zeta_k - r\tau_k)$, we obtain

$$\mathbf{E}(Z_{\eta(n)} - rn)^2 = \mathbf{E}Y_{\eta(n)}^2 - 2\mathbf{E}Y_{\eta(n)}\chi(n) + \mathbf{E}\chi^2(n).$$

By virtue of (10.4.7), $\mathbf{E}\chi^2(n) = o(n)$. By (10.6.4) (with a somewhat different notation),

$$\mathbf{E}Y_{\eta(n)}^2 = d^2\mathbf{E}\eta(n),$$

where $d^2 := \mathbf{D}(\zeta - r\tau)$, $\mathbf{E}\eta(n) \sim n/a$ and $a = \mathbf{E}\tau$. Hence, applying the Cauchy–Bunjakovsky inequality, we get

$$\big|\mathbf{E}Y_{\eta(n)}\chi(n)\big| = o(n), \qquad \mathbf{E}(Z_{\eta(n)} - rn)^2 \sim nd^2a^{-1}. \tag{13.8.25}$$

It remains to estimate the last two terms on the right-hand side of (13.8.24). But

$$\big|\zeta_{\eta(n)} - z_n\big| \le \zeta_{\eta(n)}^*,$$

where ζ^* corresponds to the summands $\xi^*_{X_k} = |\xi_{X_k}|$ and where, by Lemma 13.8.5 applied to $\xi^*_x = |\xi_x|$, we have

$$\mathbf{E}\big(\zeta^*_{\eta(n)}\big)^2 = o(n).$$

Therefore $\mathbf{E}(\zeta_{\eta(n)} - z_n) = o(n)$ and, by the Cauchy–Bunjakovsky inequality and relation (13.8.25), the same relation is valid for the shifted moment in (13.8.24). Thus,

$$\mathbf{E}(S_n - rn)^2 \sim a^{-1}d^2n.$$

Combining this relation with (13.8.23), we obtain the assertion of the theorem. □