

Chapter 3

Markov Chains

In the early twentieth century, Markov (1856–1922) introduced in [67] a new class of models called Markov chains, applying sequences of dependent random variables that enable one to capture dependencies over time. Since that time, Markov chains have developed significantly, which is reflected in the achievements of Kolmogorov, Feller, Doob, Dynkin, and many others. The significance of the extensive theory of Markov chains and the continuous-time variant called Markov processes is that it can be successfully applied to the modeling behavior of many problems in, for example, physics, biology, and economics, where the outcome of one experiment can affect the outcome of subsequent experiments. The terminology is not consistent in the literature, and many authors use the same name (Markov chain) for both discrete and continuous cases. We also apply this terminology.

Heuristically, the property that characterizes Markov chains can be expressed by the so-called memoryless notion (Markov property) as follows: a Markov chain is a stochastic process for which future behavior, given the past and the present, depends only on the present and not on the past.

This chapter presents a brief introduction to the theory of discrete-time Markov chains (DTMCs) and to the continuous-time variant, continuous-time Markov chains (CTMCs), that will be applied to the modeling and analysis of queueing systems. Note that DTMCs and CTMCs taking values in a set of countable elements have many similar properties; however, in contrast to discrete-time processes, the characteristics of a sample path essentially differ in continuous cases.

We limit ourselves here to the definition of Markov processes and to their basic properties with countable state space in discrete time $\mathcal{T} = \{0, 1, \dots\}$ and continuous time $\mathcal{T} = [0, \infty)$. In connection with the classic results discussed in this chapter, we refer mainly to the classic works [35, 36].

Consider a discrete-time or continuous-time stochastic process $X = (X_t, t \in \mathcal{T})$ given on a probability space (Ω, \mathcal{A}, P) and taking values in a countable set, called the **state space**, $\mathcal{X} = \{x_0, x_1, \dots\}$. The state space \mathcal{X} is called **finite** if it consists of a finite number of elements. The sample path of a discrete-time process with discrete

sample space is defined in the space of sequences $\mathcal{S} = \{x_{k_0}, x_{k_1}, \dots\}$, $x_{k_i} \in \mathcal{X}$, while it is an element of the space of all functions $\mathcal{S} = \{x_t : x_t \in \mathcal{X}, t \geq 0\}$ in continuous-time cases.

We say that the process is **in the state** $x \in \mathcal{X}$ at the time $t \in \mathcal{T}$ if $X_t = x$. The process starts from a state $x_0 \in \mathcal{X}$ determined by the distribution of the random variable X_0 , which is the **initial distribution** of the process. If there exists a state $x_0 \in \mathcal{X}$ for which $\mathbf{P}(X_0 = x_0) = 1$, then the state x_0 is called the **initial state**. The state of the process can change from time to time, and these changes in state are known as **transitions**. The probabilities of these state changes are called **transition probabilities**, which with the initial distribution determine the statistical behavior of the process.

If we denote by \mathcal{B}_X the σ -algebra of all subsets of the state space \mathcal{X} , then the pair $(\mathcal{X}, \mathcal{B}_X)$ is a measurable space and the connection $\{X_t \in A\} \in \mathcal{A}$ holds for all $t \in \mathcal{T}$ and $\{A \in \mathcal{B}_X\} \in \mathcal{A}$.

Definition 3.1. A stochastic process $(X_t, t \in \mathcal{T})$ with the discrete state space \mathcal{X} is called a **Markov chain** if for every nonnegative integer n and for all $t_0 < \dots < t_n < t_{n+1}$, $t_i \in \mathcal{T}$, $x_0, \dots, x_{n+1} \in \mathcal{X}$

$$\mathbf{P}(X_{t_{n+1}} = x_{n+1} \mid X_{t_0} = x_0, \dots, X_{t_n} = x_n) = \mathbf{P}(X_{t_{n+1}} = x_{n+1} \mid X_{t_n} = x_n), \quad (3.1)$$

provided that this conditional probability exists. Let $x, y \in \mathcal{X}$, $s \leq t$, $s, t \in \mathcal{T}$; then the function

$$p_{x,y}(s, t) = \mathbf{P}(X_t = y \mid X_s = x)$$

is called a **transition probability function** of a Markov chain. If the equation $p_{x,y}(s, t) = p_{x,y}(t - s)$ holds for all $x, y \in \mathcal{X}$, $s \leq t$, $s, t \in \mathcal{T}$, then the Markov chain is called **(time) homogeneous**; otherwise it is known as **inhomogeneous**.

In both discrete- and continuous-time cases, this definition expresses the aforementioned memoryless property of a Markov chain, and it ensures that the transition probabilities depend only on the present state X_s , not on how the present state was reached. We start with a discussion of DTMCs.

3.1 Discrete-Time Markov Chains with Discrete State Space

Given a Markov chain $X = (X_t, t \in \mathcal{T})$, $\mathcal{T} = \{0, 1, \dots\}$ on a probability space (Ω, \mathcal{A}, P) taking values in a finite or countably infinite set of elements \mathcal{X} . It is conventional to denote the finite state space by the set $\mathcal{X} = \{0, 1, \dots, K\}$ ($0 < K < \infty$) and the countably infinite one by $\mathcal{X} = \{0, 1, \dots\}$. This notation is quite reasonable for queueing systems, and in general, it does not lead to a separate problem if the elements of \mathcal{X} serve to distinguish the states only; otherwise, the state space is chosen based on practical requirements. Assume that the events $\{X_t = i\}$, $i \in \mathcal{X}$, are disjoint for all $t \in \mathcal{T}$.

In the discrete-time case we can give an alternative definition of a Markov chain instead of Eq. (3.1).

Definition 3.2. A discrete-time stochastic process X with state space \mathcal{X} is called a Markov chain if for every $n = 0, 1, \dots$ and for all states $i_0, \dots, i_{n+1} \in \mathcal{X}$

$$\begin{aligned} p_{i_n, i_{n+1}}(n, n+1) &= \mathbf{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n) \\ &= \mathbf{P}(X_{n+1} = i_{n+1} \mid X_n = i_n), \end{aligned} \quad (3.2)$$

provided that a conditional probability exists. The probability

$$p_{i,j}(n, n+1) = \mathbf{P}(X_{n+1} = j \mid X_n = i), \quad i, j \in \mathcal{X}, n = 0, 1, \dots,$$

is called a **one-step transition probability**, which is the probability of a transition from a state i to a state j in a single step from time n to time $n+1$.

Relation (3.2) is simpler in our case than that of Eq. (3.1), but it can be easily checked that they are equivalent to each other. Here we can define, from a practical point of view, the transition probability $p_{i,j}(s, t) = 0$, when the probability of the event $\{X_s = i\}$ equals 0 at the time point s because if $\mathbf{P}\{X_s = i\} = 0$ holds, then the sample path arrives at the state i with probability 0 at time s ; therefore, the quantity $p_{i,j}(s, t)$ can be defined freely in this case.

Definition 3.3. We say that a stochastic process X with state space \mathcal{X} is a **Markov chain of m -order** (or a **Markov chain with memory m**) if for every $n = 1, 2, \dots$ and for arbitrary states $i_k \in \mathcal{X}, k = 0, \dots, n+m$,

$$\begin{aligned} &\mathbf{P}(X_{n+m} = i_{n+m} \mid X_0 = i_0, \dots, X_{n+m-1} = i_{n+m-1}) \\ &= \mathbf{P}(X_{n+m} = i_{n+m} \mid X_n = i_n, \dots, X_{n+m-1} = i_{n+m-1}), \end{aligned}$$

provided that a conditional probability exists.

It is not difficult to verify that an m -order Markov chain can be represented as a first-order one if we introduce a new m -dimensional process as follows. Define the vector-valued process $Y = (Y_0, Y_1, \dots)$,

$$Y_n = (X_n, \dots, X_{n+m-1}), \quad n = 0, 1, \dots,$$

with state space

$$\mathcal{X}' = \{(k_1, \dots, k_m) : k_1, \dots, k_m \in \mathcal{X}\}.$$

Then the process Y is a first-order Markov chain because

$$\begin{aligned} &\mathbf{P}(Y_{n+1} = (i_{n+1}, \dots, i_{n+m}) \mid Y_0 = (i_0, \dots, i_{m-1}), \dots, Y_n = (i_n, \dots, i_{n+m-1})) \\ &= \mathbf{P}(X_{n+m} = i_{n+m}, \dots, X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_{n+m-1} = i_{n+m-1}) \\ &= \mathbf{P}(X_{n+m} = i_{n+m}, \dots, X_{n+1} = i_{n+1} \mid X_n = i_n, \dots, X_{n+m-1} = i_{n+m-1}) \\ &= \mathbf{P}(Y_{n+1} = (i_{n+1}, \dots, i_{n+m}) \mid Y_n = (i_n, \dots, i_{n+m-1})). \end{aligned}$$

This is why we consider only first-order Markov chains and why, later on, we will write only *Markov chain* instead of *Markov chain of first order*.

In the theory of DTMCs, the initial distribution

$$P = (p_i, i \in \mathcal{X}), \text{ where } p_i = \mathbf{P}(X_0 = i),$$

and the transition probabilities [see Eq. (3.2)]

$$p_{ij}(n, n+1), i, j \in \mathcal{X}, n = 0, 1, \dots,$$

play a fundamental role because the statistical behavior of a Markov chain is completely determined by them (Theorem 3.4).

The states i and j , which play a role in Definition 3.2, can be identical, which means that the process can remain in the same state at the next time point. We say that a Markov chain X is **(time) homogeneous** if the transition probabilities do not depend on time shifting, that is,

$$p_{ij} = \mathbf{P}(X_{n+1} = j | X_n = i) = \mathbf{P}(X_1 = j | X_0 = i), \quad i, j \in \mathcal{X}, n = 0, 1, \dots$$

If a Markov chain is not homogeneous, then it is called **inhomogeneous**.

3.1.1 Homogeneous Markov Chains

From a practical point of view, the class of homogeneous Markov chains plays a significant role; therefore, in this chapter we will investigate the properties of this class of processes. However, many results for homogeneous cases remain valid in the inhomogeneous case, too.

By definition, for a homogeneous Markov chain the one-step transition probability (or simply transition probability) $p_{i,j}$, $i, j \in \mathcal{X}$, equals the probability that, starting from the initial state $X_0 = i$ at time 0, the process will be in the state j at the next time point 1, and this probability does not change if we take the transition probability in arbitrary time $n = 1, 2, \dots$,

$$p_{ij} = \mathbf{P}\{X_1 = j | X_0 = i\} = \mathbf{P}\{X_{n+1} = j | X_n = i\}.$$

The transition probabilities satisfy the equation

$$\sum_{j \in \mathcal{X}} p_{ij} = 1.$$

This equation expresses the obvious fact that starting in a state i at the next time point the process takes certainly some state $j \in \mathcal{X}$. The following theorem states that the initial distribution and the transition probabilities determine the finite-dimensional distribution of a homogeneous Markov chain, and as a consequence we obtain that a Markov chain can be given in a statistical sense with the state space, the initial distribution, and the transition probabilities.

Theorem 3.4. *The finite-dimensional distributions of a Markov chain X are uniquely determined by the initial distribution and the transition probabilities and*

$$\mathbf{P}(X_0 = i_0, \dots, X_n = i_n) = p_{i_{n-1}i_n} p_{i_{n-2}i_{n-1}} \cdots p_{i_0 i_1} p_{i_0}. \quad (3.3)$$

Proof. Let n be a positive integer, and let $i_0, \dots, i_n \in \mathcal{X}$. First, assume that $\mathbf{P}(X_0 = i_0, \dots, X_n = i_n) > 0$. By the definition of conditional probability,

$$\begin{aligned} \mathbf{P}(X_0 = i_0, \dots, X_n = i_n) &= \mathbf{P}(X_n = i_n \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \mathbf{P}(X_0 = i_0, \dots, X_{n-1} = i_{n-1}) = \dots \\ &= \mathbf{P}(X_n = i_n \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \\ &\quad \times \mathbf{P}(X_{n-1} = i_{n-1} \mid X_0 = i_0, \dots, X_{n-2} = i_{n-2}) \cdots \mathbf{P}(X_1 = i_1 \mid X_0 = i_0) \\ &\quad \mathbf{P}(X_0 = i_0). \end{aligned}$$

Using the Markov property we can rewrite this formula in the form

$$\begin{aligned} \mathbf{P}(X_0 = i_0, \dots, X_n = i_n) &= \mathbf{P}(X_n = i_n \mid X_{n-1} = i_{n-1}) \cdots \mathbf{P}(X_1 = i_1 \mid X_0 = i_0) \mathbf{P}(X_0 = i_0) \\ &= p_{i_{n-1}i_n} p_{i_{n-2}i_{n-1}} \cdots p_{i_0 i_1} p_{i_0}. \end{aligned}$$

If $\mathbf{P}(X_0 = i_0, \dots, X_n = i_n) = 0$, then either $\mathbf{P}(X_0 = i_0) = p_{i_0} = 0$ or there exists an index m , $0 \leq m \leq n-1$, for which

$$\mathbf{P}(X_0 = i_0, \dots, X_m = i_m) > 0 \quad \text{and} \quad \mathbf{P}(X_0 = i_0, \dots, X_{m+1} = i_{m+1}) = 0.$$

Consequently,

$$\begin{aligned} \mathbf{P}(X_0 = i_0, \dots, X_{m+1} = i_{m+1}) &= \mathbf{P}(X_{m+1} = i_{m+1} \mid X_0 = i_0, \dots, X_m = i_m) \mathbf{P}(X_0 = i_0, \dots, X_m = i_m) \\ &= p_{i_m i_{m+1}} \mathbf{P}(X_0 = i_0, \dots, X_m = i_m), \end{aligned}$$

and therefore $p_{i_m i_{m+1}} = 0$. This means that the product $p_{i_{n-1}i_n} p_{i_{n-2}i_{n-1}} \cdots p_{i_0 i_1} p_{i_0}$ equals 0 in both cases, and so assertion (3.3) of the theorem is true. \square

Comment 3.5. *From relation (3.4) it immediately follows that for any $A_i \subset \mathcal{X}$, $0 \leq i \leq n$ the probability $\mathbf{P}(X_0 \in A_0, \dots, X_n \in A_n)$ can be given in the form*

$$\mathbf{P}(X_0 \in A_0, \dots, X_n \in A_n) = \sum_{i_0 \in A_0} \cdots \sum_{i_n \in A_n} \mathbf{P}(X_0 = i_0, \dots, X_n = i_n),$$

where the probabilities are determined by relation (3.3), that is, with the help of the initial distribution and the transition probabilities.

The following remark clarifies an essential property of the homogeneous Markov chain, and on that basis limit theorems can be proved. This property relates the behavior of Markov chains to renewal and regenerative processes, which we will discuss later on in Sects. 4.1 and 4.2.

Comment 3.6. *From the memoryless property of a Markov chain X it follows that we can divide the time access into disjoint parts where the process behavior is mutually independent and follows the same probabilistic rules. We define the limits of these independent parts by the time instants when the process stays in the state $i_0 \in \mathcal{X}$.*

Formally, we define the sequence of random time points $0 \leq \tau_1 < \tau_2 < \dots$ by the condition $X_{\tau_n} = i_0$, $n = 1, 2, \dots$, and $X_s \neq i_0$ if $s \notin \{\tau_1, \tau_2, \dots\}$. In this way $0 \leq \tau_1 < \tau_2 < \dots$ are the times of the first, second, etc. visits to the state i_0 , and i_0 is not visited between τ_n and τ_{n+1} , $n = 1, 2, \dots$. We define Y_n and $Z_{n,k}$ by $Y_n = \tau_{n+1} - \tau_n$ and $Z_{n,k} = X_{\tau_n+k}$, $0 \leq k < Y_n$. Y_n is the time between the n th and the $n+1$ th visits to i_0 , and $Z_{n,k}$ is the state of the process at k steps after the n th visit to i_0 , having that the next visit to i_0 is after $\tau_n + k$. Using the memoryless property of the Markov chain X we obtain that the random variables $(Y_n, Z_{n,k}, 0 \leq k < Y_n)$, $n = 1, 2, \dots$, are independent and their stochastic behaviors are identical. This fact ensures that the process is regenerative (Sect. 4.2).

In many cases, the study of Markov chains will be made simpler by the use of transition probability matrices.

Definition 3.7. The matrices associated with the transition probabilities of a Markov chain \mathcal{X} with finite or countably infinite elements are

$$\Pi = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0N} \\ p_{10} & p_{11} & \cdots & p_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N0} & p_{N1} & \cdots & p_{NN} \end{bmatrix} \quad \text{and} \quad \Pi = \begin{bmatrix} p_{00} & p_{01} & p_{12} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

These matrices are called (one-step) **transition probability matrices**.

A matrix with nonnegative entries $\mathbf{A} = [a_{ij}]_{i,j \in \mathcal{X}}$ is called a **stochastic matrix** if for every row the sum of row elements equals 1. Then all transition probability matrices are stochastic ones:

(a) The elements of Π are obviously nonnegative,

$$p_{ij} \geq 0, \quad i, j \in \mathcal{X}.$$

(b) For every i the sum of the i th row elements of Π equals 1,

$$\sum_{j \in \mathcal{X}} p_{ij} = 1, \quad i \in \mathcal{X}.$$

The first of the following three examples shows that a sequence of independent and identically distributed discrete random variables is a homogeneous Markov chain. The second one shows that the sequence of sums of these random variables also constitutes a homogeneous Markov chain. If in the second case the random variables are independent, but not identically distributed, then the defined sequences will be an inhomogeneous Markov chain. The third example describes the stochastic behavior of a random walk on the real number line; in this case it is reasonable to choose the state space to be the set of all integer numbers, that is, $\mathcal{X} = \{0, \pm 1, \pm 2, \dots\}$.

Let Z_0, Z_1, \dots be a sequence of independent and identically distributed random variables with a common CDF

$$\mathbf{P}(Z_m = k) = p_k, \quad p_k \geq 0, \quad k = 0, 1, \dots, \quad m = 0, 1, \dots$$

Example 3.8. Define the discrete-time stochastic process X with the relation $X_n = Z_n$, $n = 0, 1, \dots$. Then X is a homogeneous Markov chain with initial distribution $\mathbf{P}(X_0 = k) = p_k$, $k = 0, 1, \dots$ and transition probability matrix

$$\Pi = \begin{bmatrix} p_0 & p_1 & p_2 & \cdots \\ p_0 & p_1 & p_2 & \cdots \\ p_0 & p_1 & p_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Example 3.9. Consider the process $X_n = Z_1 + \dots + Z_n$, $n = 0, 1, \dots$, with the initial distribution $\mathbf{P}(X_0 = 0) = 1$, i.e., the initial state is 0. The one-step transition probabilities are

$$\begin{aligned} p_{ij}(n, n + 1) &= \mathbf{P}(X_{n+1} = j \mid X_n = i) \\ &= \mathbf{P}(Z_1 + \dots + Z_{n+1} = j \mid Z_1 + \dots + Z_n = i) \\ &= \mathbf{P}(Z_{n+1} = j - i) = \begin{cases} p_{j-i}, & \text{ha } j \geq i, \\ 0, & \text{ha } j < i. \end{cases} \end{aligned}$$

This means that the process X is a homogeneous Markov chain with the transition probability matrix

$$\Pi = \begin{bmatrix} p_0 & p_1 & p_2 & p_3 & p_4 & \cdots \\ 0 & p_0 & p_1 & p_2 & p_3 & \cdots \\ 0 & 0 & p_0 & p_1 & p_2 & \cdots \\ 0 & 0 & 0 & p_0 & p_1 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}.$$

Example 3.10. Now let

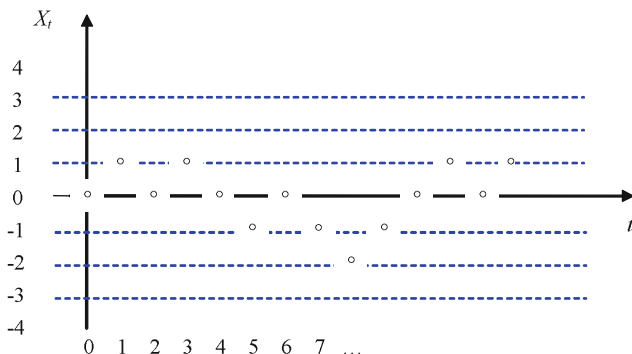


Fig. 3.1 Random walk

$$\mathbf{P}(Z_i = +1) = p, \mathbf{P}(Z_i = -1) = 1 - p \quad (0 < p < 1), i = 1, 2, \dots,$$

be the common distribution function of a sequence of independent random variables Z_0, Z_1, \dots , and define the process $X_n = Z_1 + \dots + Z_n$, $n = 1, 2, \dots$. Let $\mathbf{P}(X_0 = 0) = 1$ be the initial distribution of the process X . Then the process X is a homogeneous Markov chain with initial state $X_0 = 0$ and transition probability matrix

$$\begin{aligned} p_{ij}(n, n+1) &= \mathbf{P}(X_{n+1} = j \mid X_n = i) \\ &= \mathbf{P}(Z_1 + \dots + Z_{n+1} = j \mid Z_1 + \dots + Z_n = i) \\ &= \mathbf{P}(Z_{n+1} = j - i) = \begin{cases} p, & \text{if } j = i + 1, \\ 1 - p, & \text{if } j = i - 1, \\ 0, & \text{if } |i - j| \neq 1. \end{cases} \end{aligned}$$

The process X describes the random walk on the number line starting from the origin and moves at every step one unit to the right with probability p and to the left with probability $(1 - p)$, with these moves being independent of each other. The case $p = 1/2$ corresponds to the symmetric random walk.

Figure 3.1 demonstrates the transitions of the random walk, while Fig. 3.2 shows the transitions of a Markov chain with a finite state space.

3.1.2 The m -Step Transition Probabilities

Let X be a DTMC with discrete state space \mathcal{X} . Denote by

$$p_{ij}(s, t) = \mathbf{P}(X_t = j \mid X_s = i)$$

the transition probabilities of X and by

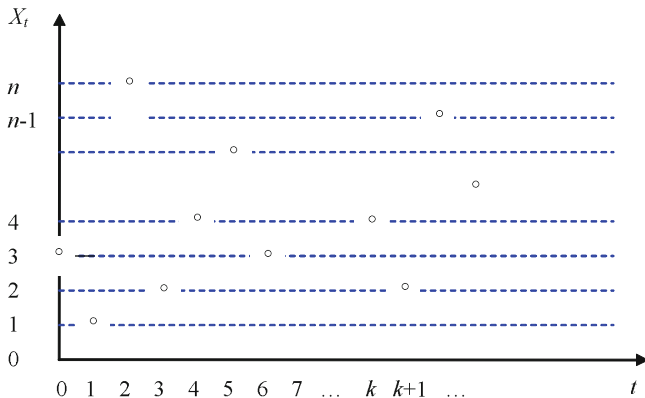


Fig. 3.2 Markov chain with finite state space

$$\Pi(s, t) = [p_{ij}(s, t)]_{i, j \in \mathcal{X}}, \quad i, j \in \mathcal{X} \text{ and } 0 \leq s \leq t < \infty$$

the transition probability matrices. We set for $s = t$

$$p_{ij}(s, s) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

If the Markov chain X is homogeneous, then the transition probability $p_{ij}(s, t)$ depends only on the difference $t - s$. Thus, using the notation $t = s + m$, we have

$$p_{ij}(s, s + m) = p_{ij}(m), \quad s, m = 0, 1, \dots, \quad i, j \in \mathcal{X}.$$

Definition 3.11. The quantities $p_{ij}(m)$, $m = 0, 1, \dots$, $i, j \in \mathcal{X}$ are called the **m -step transition probabilities** of the Markov chain X , and the matrix $\Pi(m) = [p_{ij}(m)]_{i, j \in \mathcal{X}}$ associated with them is called an **m -step transition probability matrix**.

Theorem 3.12 (Chapman–Kolmogorov equation). For every nonnegative integer number r, s , the $(r + s)$ -step transition probabilities of the homogeneous Markov chain satisfy the equation

$$p_{ij}(r + s) = \sum_{k \in \mathcal{X}} p_{ik}(r) p_{kj}(s). \tag{3.4}$$

Proof. Assume the initial state of the process is i , that is, the process starts from the state i at the time point 0. First we note that the relation

$$p_{ik}(r) = \mathbf{P}(X_r = k \mid X_0 = i) = \frac{\mathbf{P}(X_0 = i, X_r = k)}{\mathbf{P}(X_0 = i)} = 0$$

holds for some state k if and only if $\mathbf{P}(X_0 = i, X_r = k) = 0$. On the other hand, since $\{X_r = k\}$, $k \in \mathcal{X}$ form a complete system of events, $\sum_{k \in \mathcal{X}} \mathbf{P}(X_r = k) = 1$, and, in accordance with the definitions of the $(r + s)$ -step transition probability and the conditional probability, we obtain

$$\begin{aligned}
 p_{ij}(r + s) &= \mathbf{P}(X_{r+s} = j \mid X_0 = i) \\
 &= \frac{\mathbf{P}(X_{r+s} = j, X_0 = i)}{\mathbf{P}(X_0 = i)} = \sum_{k \in \mathcal{X}} \frac{\mathbf{P}(X_{r+s} = j, X_0 = i, X_r = k)}{\mathbf{P}(X_0 = i)} \\
 &= \sum_{k \in \mathcal{X}} \mathcal{I}_{\{p_{ik} \neq 0\}} \frac{\mathbf{P}(X_0 = i, X_r = k)}{\mathbf{P}(X_0 = i)} \frac{\mathbf{P}(X_{r+s} = j, X_0 = i, X_r = k)}{\mathbf{P}(X_0 = i, X_r = k)} \\
 &= \sum_{k \in \mathcal{X}} \mathcal{I}_{\{p_{ik} \neq 0\}} \mathbf{P}(X_r = k \mid X_0 = i) \mathbf{P}(X_{r+s} = j \mid X_r = k, X_0 = i) \\
 &= \sum_{k \in \mathcal{X}} \mathcal{I}_{\{p_{ik} \neq 0\}} p_{ik}(0, r) p_{kj}(r, r + s) = \sum_{k \in \mathcal{X}} p_{ik}(r) p_{kj}(s).
 \end{aligned}$$

□

If we use the matrix notation $\Pi(s, t) = [p_{ij}(s, t)]_{i, j \in \mathcal{X}}$, then the Chapman–Kolmogorov equation can be rewritten in the matrix form

$$\Pi(s, t) = \Pi(s, r) \Pi(r, t),$$

where s, r, t , and n are integer numbers satisfying the inequality $0 \leq s \leq r \leq t, n \geq 1$. Successively repeating this relation we have

$$\Pi(0, n) = \Pi(0, 1) \Pi(1, n) = \dots = \Pi(0, 1) \Pi(1, 2) \dots \Pi(n - 1, n).$$

Consequently, the m -step transition probability matrix of a homogeneous Markov chain can be given in the form

$$\Pi(m) = \Pi^m,$$

where $\Pi = \Pi(0, 1)$ is the (one-step) transition probability matrix of the Markov chain.

3.1.3 Classification of States of Homogeneous Markov Chains

The behavior of a Markov chain and its asymptotic properties essentially depend on the transition probabilities, which reflect the connections among the different states.

Denote by $P_i(t) = \mathbf{P}(X_t = i)$, $i \in \mathcal{X}$ the distribution of the Markov chain X at the time $t \geq 0$. One of the most important questions in the theory of Markov chains

concerns the conditions under which a limit distribution exists for all initial states $X_0 = k \in \mathcal{X}$,

$$\lim_{t \rightarrow \infty} P(t) = \pi = (\pi_i, i \in \mathcal{X}),$$

of the time-dependent distribution $P(t) = (P_i(t), i \in \mathcal{X})$, where $\pi_i \geq 0$, $\sum_{i \in \mathcal{X}} \pi_i = 1$.

In the answer to this question, the arithmetic properties of the transition probabilities play an important role.

To demonstrate this fact, consider the case where the sample space \mathcal{X} can be divided into two disjoint (nonempty) sets \mathcal{X}_1 and \mathcal{X}_2 such that

$$p_{ij} = p_{ji} = 0, \text{ for all } i \in \mathcal{X}_1 \text{ and } j \in \mathcal{X}_2.$$

Obviously, if $X_0 = i_0 \in \mathcal{X}_1$ is the initial state, then the relation $X_t \in \mathcal{X}_1$ is valid for all $t \geq 0$, and in the opposite case, $X_t \in \mathcal{X}_2$ for all $t \geq 0$ holds if the initial state i_0 satisfies the condition $i_0 \in \mathcal{X}_2$. This means that in this case we can in fact consider two Markov chains $(\mathcal{X}_k, (P_i(t), i \in \mathcal{X}_k), \Pi_k)$, $k = 1, 2$, that can be investigated independently of each other.

Definition 3.13. The state $j \in \mathcal{X}$ is **accessible** from the state $i \in \mathcal{X}$ (denoted by $i \rightarrow j$) if there exists a positive integer m such that $p_{ij}(m) > 0$. If the states $i, j \in \mathcal{X}$ are mutually accessible from each other, then we say that they **communicate** (denoted by $i \longleftrightarrow j$).

$p_{ii}(0) = 1$, $i \in \mathcal{X}$ represents the assumption that “every state is accessible in 0 steps from itself.” If the state $j \in \mathcal{X}$ is not accessible from the state $i \in \mathcal{X}$ (denoted by $i \not\rightarrow j$), then $p_{ij}(m) = 0$, $m \geq 1$. It is easy to check that $i \longleftrightarrow j$ is an **equivalence relation**: it is reflexive, transitive, and symmetric. Furthermore, if the states i and j do not communicate, then either $p_{ij}(m) = 0$, $m \geq 1$, or $p_{ji}(m) = 0$, $m \geq 1$. If a state i satisfies the condition $p_{ii} = p_{ii}(1) = 1$, then the state i is called **absorbing**. This means that if the process visits an absorbing state at time t , then it remains there forever and no more state transitions occur.

If the state space \mathcal{X} does not consist of the states i and j such that $i \rightarrow j$, but $j \not\rightarrow i$, then \mathcal{X} can be given as a union of finite or countable disjoint sets

$$\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \dots,$$

where for every k the states of \mathcal{X}_k communicate, while for every k, n , $k \neq n$, the states of \mathcal{X}_k cannot be accessible from the states of \mathcal{X}_n .

Definition 3.14. A set of states is called **irreducible** if all pairs of its elements communicate.

In the theory of Markov chains, irreducible classes play an important role because they can be independently analyzed.

Definition 3.15. A Markov chain is called **irreducible** if all pairs of its states communicate.

Clearly, if a Markov chain is irreducible, then it consists of only one irreducible class of states, that is, for every $i, j \in \mathcal{X}$ there exists an integer $m \geq 1$ (depending on i and j) such that $p_{ij}(m) > 0$.

Definition 3.16. For every i denote by $d(i)$ the greatest common divisor of integer numbers $m \geq 1$ for which $p_{ii}(m) > 0$. If $p_{ii}(m) = 0$ for every m , then we set $d(i) = 0$. Then the number $d(i)$ is called the **period** of the Markov chain. If $d(i) = 1$ for every state, then the Markov chain is called **aperiodic**.

Example 3.17 (Periodic Markov chain). Consider the random walk on the number line demonstrated earlier in Example 3.10. Starting from an arbitrary state i we can return to state i with positive probabilities in steps $2, 4, \dots$ only. It is clear that in this case, $p_{ii}(2k) > 0$ and $p_{ii}(2(k-1)+1) = 0$ for every $i \in \mathcal{X}$ and $k = 1, 2, \dots$; therefore, $d(i) = 2$. At the same time, the Markov chain is obviously irreducible.

Theorem 3.18. Let X be a homogeneous Markov chain with state space \mathcal{X} , and let $\mathcal{X}' \subset \mathcal{X}$ be a nonempty irreducible class. Then for every $i, j \in \mathcal{X}'$, the periods of i and j are the same, i.e., $d(i) = d(j)$.

Proof. Let $i, j \in \mathcal{X}'$, $i \neq j$, be two arbitrary states. Since \mathcal{X}' is an irreducible class, there exist $t, s \geq 1$ integers such that the inequalities $p_{ij}(t) > 0$ and $p_{ji}(s) > 0$ hold. From this, by the Chapman–Kolmogorov equation, we obtain

$$p_{ii}(t+s) \geq p_{ij}(t)p_{ji}(s) > 0 \quad \text{and} \quad p_{jj}(t+s) \geq p_{ji}(s)p_{ij}(t) > 0;$$

therefore, the numbers $d(i)$ and $d(j)$ differ from 0. Choose arbitrarily an integer $m \geq 1$ such that $p_{ii}(m) > 0$. Repeatedly applying the Chapman–Kolmogorov equation, we have for any $k \geq 1$

$$p_{jj}(t+s+km) \geq p_{ji}(s)p_{ii}(km)p_{ij}(t) \geq p_{ji}(s)(p_{ii}(m))^k p_{ij}(t) > 0.$$

Thus by the definition of the period of the state j , $d(j)$ is a divisor of both $(t+s+m)$ and $(t+s+2m)$, and hence it is also a divisor of their difference $(t+s+2m) - (t+s+m) = m$. From this it immediately follows that $d(j)$ is a divisor of every m for which $p_{ii}(m) > 0$, and thus it is a divisor of $d(i)$; therefore, $d(j) \leq d(i)$. Changing the role of i and j we get the reverse inequality $d(j) \geq d(i)$, and consequently $d(j) = d(i)$. \square

Notice that from this theorem it follows that the states of an irreducible class have a common period $d(\mathcal{X}')$ called the **period of the class**. As a consequence, we have the following assertion.

Corollary 3.19. If the Markov chain X is homogeneous and irreducible with state space \mathcal{X} , then every state has the same period $d = d(\mathcal{X}) > 0$ and is periodic or aperiodic depending on $d > 1$ or $d = 1$, respectively.

The main property of the numbers for which the probabilities of returning to a state i in k steps are positive, i.e., $p_{ii}(k) > 0$, is given by the following assertion.

Theorem 3.20. *Let X be a homogeneous irreducible Markov chain with state space \mathcal{X} . Then for every state $i \in \mathcal{X}$ there exists an integer M_i such that $p_{ii}(d(i)m) > 0$ if $m \geq M_i$.*

Proof. By the previous theorem, $d(i) \geq 1$. Let m_1, \dots, m_L be different positive integer numbers such that, on the one hand, $p_{ii}(m_k) > 0$, $1 \leq k \leq L$, and on the other hand, $d(i)$ can be given as the greatest common divisor of integers m_1, \dots, m_L . Then, using the well-known assertion from the number theory that there exists an integer M_i such that for every integer $m \geq M_i$, the equation $md(i) = r_1 m_1 + \dots + r_L m_L$ has a solution with nonnegative integers i_1, \dots, i_L . Applying this fact and the Chapman–Kolmogorov equation we obtain

$$p_{ii}(md(i)) \geq (p_{ii}(m_1))^{r_1} \cdot \dots \cdot (p_{ii}(m_L))^{r_L} > 0,$$

and consequently the assertion of the theorem is true. \square

Consider now the homogeneous irreducible Markov chain with period $d(\mathcal{X}) > 1$. We show that for the transitions among the states there exists a cyclic property, demonstrated in Example 3.28, of the random walk on a number line: if the walk starts from state 0, then the process can take only even integers in even steps and only odd integers in odd steps. The cyclic property in this case means that after even numbers follow odd numbers and after odd numbers follow even numbers as states. This division of states is generalized subsequently for Markov chains with arbitrary period d .

Let $i_0 \in \mathcal{X}$ be an arbitrarily fixed state, and define the sets

$$\mathcal{X}_k = \{j \in \mathcal{X} : p_{i_0 j}(k + md) > 0, \text{ for some } m \geq 0\}, k = 0, 1, \dots, d - 1.$$

That is, \mathcal{X}_k is the set of states that are available from i_0 in $k + md$ ($m = 0, 1, \dots$) steps.

Theorem 3.21. *The sets $\mathcal{X}_1, \dots, \mathcal{X}_{d-1}$ are disjoint, $\mathcal{X} = \mathcal{X}_0 \cup \dots \cup \mathcal{X}_{d-1}$, and the Markov chain allows for only the following cyclic transitions among the sets \mathcal{X}_k :*

$$\mathcal{X}_0 \rightarrow \mathcal{X}_1 \rightarrow \dots \rightarrow \mathcal{X}_{d-1} \rightarrow \mathcal{X}_0. \quad (3.5)$$

Proof. First we prove that the sets $\mathcal{X}_0, \dots, \mathcal{X}_{d-1}$ are disjoint and their union is \mathcal{X} . In contrast, assume that there exist integers k_1, k_2, m_1, m_2 such that $0 \leq k_1 < k_2 \leq d - 1$, $m_1, m_2 \geq 1$, $p_{i_0 j}(k_1 + m_1 d) > 0$, and $p_{i_0 j}(k_2 + m_2 d) > 0$. Since the Markov chain is irreducible, there exists an integer $K \geq 1$ such that $p_{i_0 i_0}(K) > 0$. Using the Chapman–Kolmogorov equation we have

$$\begin{aligned} p_{i_0 i_0}(k_1 + m_1 d + K) &\geq p_{i_0 j}(k_1 + m_1 d) p_{j i_0}(K) > 0, \\ p_{i_0 i_0}(k_2 + m_2 d + K) &\geq p_{i_0 j}(k_2 + m_2 d) p_{j i_0}(K) > 0. \end{aligned}$$

By the definition of the period d , d is a divisor of both $(k_1 + m_1d + K)$ and $(k_2 + m_2d + K)$, thus it is also a divisor of their difference, that is, $(k_2 - k_1) + (m_2 - m_1)d$. Consequently, d is a divisor of the difference $(k_2 - k_1)$, which is a contradiction, because $0 < k_2 - k_1 \leq d - 1$. The irreducibility condition ensures that if all states $i \in \mathcal{X}$ are accessible from the state i_0 , then $\mathcal{X} = \mathcal{X}_0 \cup \dots \cup \mathcal{X}_{d-1}$.

We now verify that for every k , $0 \leq k \leq d - 1$, $i \in \mathcal{X}_k$ and $j \in \mathcal{X}$ such that $p_{ij} > 0$, the relation $j \in \mathcal{X}_K$, $0 \leq K < d$, is true, where

$$K = \begin{cases} k + 1, & \text{if } 0 \leq k < d - 1, \\ 0, & \text{if } k = d - 1. \end{cases}$$

This property guarantees the transitions between the states in (3.5).

Since $j \in \mathcal{X}_k$, then, by the definition of the sets \mathcal{X}_k , there exists an integer $m \geq 0$ such that $p_{i_0j}(k + md) > 0$. From this, by the use of the Chapman–Kolmogorov equality, we have

$$p_{i_0\ell}(k + 1 + md) \geq p_{i_0j}(k + md)p_{j\ell} > 0.$$

In view of the fact that

$$k + 1 + md = \begin{cases} K + m d, & \text{if } 0 \leq k < d - 1, \\ 0 + (m + 1) d, & \text{if } k = d - 1, \end{cases}$$

from the definition of \mathcal{X}_K follows the relation $j \in \mathcal{X}_K$. □

As a consequence of Theorem 3.21, we have the next important corollary, which allows us to consider an aperiodic Markov chain instead of a periodic one.

Corollary 3.22. *Theorem 3.21 states that starting from a state of \mathcal{X}_k , $k = 0, 1, \dots, d - 1$, after exactly d steps the process returns to a state of \mathcal{X}_k . If we define the quantities*

$$p_{ij}^{(k)} = \mathbf{P}(X_d = i \mid X_0 = j), \quad i, j \in \mathcal{X}_k,$$

then $\sum_{j \in \mathcal{X}_k} p_{ij}^{(k)} = 1$, $i \in \mathcal{X}_k$ follows. This means that the matrices $\mathbf{P}^{(k)} = [p_{ij}^{(k)}]_{i, j \in \mathcal{X}_k}$ are stochastic; they can be interpreted as one-step transition probability matrices, and consequently the processes

$$Y^{(k)} = (Y_0, Y_1, \dots), \quad k = 0, 1, \dots, d - 1,$$

with the state space \mathcal{X}_k and transition probability matrix $\mathbf{P}^{(k)}$, are homogeneous and irreducible Markov chains, and so, instead of the original chain, d homogeneous irreducible Markov chains can be considered independently.

If the states of the Markov chain are numbered according to the $\mathcal{X}_k, k = 0, 1, \dots, d - 1$, sets, then the transition probability matrix has the following structure:

$\mathcal{X}_0\{$	0	$P_{0 \rightarrow 1}$			
$\mathcal{X}_1\{$		0	$P_{1 \rightarrow 2}$		
			\ddots	\ddots	
$\mathcal{X}_{d-2}\{$				0	$P_{d-2 \rightarrow d-1}$
$\mathcal{X}_{d-1}\{$	$P_{d-1 \rightarrow 0}$				0

3.1.4 Recurrent Markov Chains

We consider the question of what conditions ensure the existence of limit theorems for homogeneous aperiodic Markov chains, that is, under what conditions does there exist the limit distribution $\pi = (\pi_i, i \in \mathcal{X}), (\pi_i \geq 0, \sum_{i \in \mathcal{X}} \pi_i = 1)$, such that, independently of the initial distribution $(p_i, i \in \mathcal{X})$, the limit is

$$\lim_{n \rightarrow \infty} P_i(n) = \lim_{n \rightarrow \infty} \mathbf{P}(X_n = i) = \pi_i, i \in \mathcal{X}?$$

To provide an answer to this question, it is necessary to consider some quantities such as the probability and the expected value of returning to a given state of a Markov chain or arriving at a state j from another state i . Let $i, j \in \mathcal{X}$ be two arbitrary states, and introduce the following notations:

$$T_{ij} = \inf\{n : n > 1, X_n = j \mid X_0 = i\},$$

$$f_{ij}(0) = 0,$$

$$f_{ij}(1) = \mathbf{P}(X_1 = j \mid X_0 = i),$$

$$f_{ij}(n) = \mathbf{P}(X_1 \neq j, X_2 \neq j, \dots, X_{n-1} \neq j, X_n = j \mid X_0 = i), n = 2, 3, \dots$$

If $i \neq j$, then the quantities $f_{ij}(n) = \mathbf{P}\{T_{ij} = n\}$ mean the **first hit** (or **first passage**) **probabilities** for the state j from i , which is the probability that starting from the state i at time point 0, the process will be first in the state j during n steps (or in time n). If $i = j$, then the quantity $f_{ii}(n)$ means the **first return probability** in the state i in n steps.

Denote $f_{ij} = \sum_{k=1}^{\infty} f_{ij}(k), i, j \in \mathcal{X}$. Obviously, the quantity f_{ij} means the probability that the Markov chain starts from a state i at time 0 and at some time arrives at the state j , that is, $f_{ij} = \mathbf{P}\{T_{ij} < \infty\}$.

Definition 3.23. A state i is called **recurrent** if the process returns to the state i with probability 1, that is, $f_{ii} = \mathbf{P}\{T_{ii} < \infty\} = 1$. If $f_{ii} < 1$, then the state i is called **transient**.

From the definition it follows that when i is a transient state, then a process with positive probability will never return to the state i . The following theorem describes the connection between the return probabilities and the m -step transition probabilities of a Markov chain in the form of a so-called discrete renewal equation.

Theorem 3.24. For every $i, j \in \mathcal{X}$, $n = 1, 2, \dots$,

$$p_{ij}(n) = \sum_{k=1}^n f_{ij}(k)p_{jj}(n-k). \quad (3.6)$$

Proof. By the definition $p_{jj}(0) = 1$, in the case $n = 1$ we have $p_{ij}(1) = f_{ij}(1)p_{jj}(0) = f_{ij}(1)$. Now let $n \geq 2$. Using conditional probability and the Markov property we get

$$\begin{aligned} \mathbf{P}(X_n = j, X_1 = j \mid X_0 = i) &= \mathbf{P}(X_n = j \mid X_1 = j, X_0 = i)\mathbf{P}(X_1 = j \mid X_0 = i) \\ &= p_{jj}(n-1)p_{ij}(1) = f_{ij}(1)p_{jj}(n-1). \end{aligned}$$

Similarly, we obtain

$$\begin{aligned} \mathbf{P}(X_n = j, X_k = j, X_m \neq j, 1 \leq m \leq k-1 \mid X_0 = i) \\ = f_{ij}(k)p_{jj}(n-k), \quad n = 1, 2, \dots \end{aligned}$$

On the basis of the last two equations, it follows that

$$\begin{aligned} p_{ij}(n) &= \mathbf{P}(X_n = j, X_1 = j \mid X_0 = i) \\ &\quad + \sum_{k=2}^n \mathbf{P}(X_n = j, X_k = j, X_m \neq j, 1 \leq m \leq k-1 \mid X_0 = i) \\ &= f_{ij}(1)p_{jj}(n-1) + \sum_{k=2}^n f_{ij}(k)p_{jj}(n-k), \quad n = 1, 2, \dots \end{aligned}$$

□

The notion of the recurrence of a state is defined by the return probabilities, but the following theorem makes it possible to provide a condition for it with the use of n -step transition probabilities $p_{ii}(n)$ and to classify the Markov chains.

Theorem 3.25. (a) The state $i \in \mathcal{X}$ is recurrent if and only if

$$\sum_{n=1}^{\infty} p_{ii}(n) = \infty.$$

(b) If i and j are communicating states and i is recurrent, then j is also recurrent.

(c) If a state $j \in \mathcal{X}$ is transient, then for arbitrary $i \in \mathcal{X}$

$$\sum_{n=1}^{\infty} p_{ij}(n) < \infty \quad \text{and consequently} \quad \lim_{n \rightarrow \infty} p_{ij}(n) = 0.$$

Proof. (a) By the definition $p_{ii}(0) = 1$ and using relation (3.6) of the preceding theorem we obtain

$$\begin{aligned} \sum_{n=1}^{\infty} p_{ii}(n) &= \sum_{n=1}^{\infty} \sum_{k=1}^n f_{ii}(k) p_{ii}(n-k) = \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} f_{ii}(k) p_{ii}(n-k) \\ &= \sum_{k=1}^{\infty} f_{ii}(k) \left(p_{ii}(0) + \sum_{n=1}^{\infty} p_{ii}(n) \right). \end{aligned}$$

From this equation, if the sum $\sum_{n=1}^{\infty} p_{ii}(n)$ is finite, then we get

$$f_{ii} = \left(1 + \sum_{n=1}^{\infty} p_{ii}(n) \right)^{-1} \sum_{n=1}^{\infty} p_{ii}(n) < 1;$$

consequently, i is not a recurrent state.

If $\sum_{n=1}^{\infty} p_{ii}(n) = \infty$, then obviously $\lim_{N \rightarrow \infty} \sum_{n=1}^N p_{ii}(n) = \infty$. Since for all positive integers N the relation

$$\begin{aligned} \sum_{n=1}^N p_{ii}(n) &= \sum_{n=1}^N \sum_{k=1}^n f_{ii}(k) p_{ii}(n-k) \\ &= \sum_{k=1}^N \sum_{n=k}^N f_{ii}(k) p_{ii}(n-k) \leq \sum_{k=1}^N f_{ii}(k) \sum_{n=0}^N p_{ii}(n) \\ &\leq \left(1 + \sum_{n=1}^N p_{ii}(n) \right) \sum_{k=1}^N f_{ii}(k) \end{aligned}$$

holds, from the limit $\sum_{n=1}^N p_{ii}(n) \rightarrow \infty$

$$1 \geq f_{ii} = \sum_{k=1}^{\infty} f_{ii}(k) \geq \sum_{k=1}^N f_{ii}(k) \geq \left(1 + \sum_{k=1}^N p_{ii}(k) \right)^{-1} \sum_{k=1}^N p_{ii}(k) \rightarrow 1, \quad N \rightarrow \infty$$

follows. Consequently, $f_{ii} = 1$, and thus the state i is recurrent.

- (b) Since the states i and j communicate, there exist integers $n, m \geq 1$ such that $p_{ij}(m) > 0$ and $p_{ji}(n) > 0$. By the Chapman–Kolmogorov equation for every integer $k \geq 1$,

$$\begin{aligned} p_{ii}(m+k+n) &\geq p_{ij}(m)p_{jj}(k)p_{ji}(n), \\ p_{jj}(m+k+n) &\geq p_{ji}(n)p_{ii}(k)p_{ij}(m). \end{aligned}$$

From this

$$\begin{aligned} \sum_{k=1}^{\infty} p_{ii}(k) &\geq \sum_{k=1}^{\infty} p_{ii}(m+n+k) \geq p_{ij}(m)p_{ji}(n) \sum_{k=1}^{\infty} p_{jj}(k), \\ \sum_{k=1}^{\infty} p_{jj}(k) &\geq \sum_{k=1}^{\infty} p_{jj}(m+n+k) \geq p_{ij}(m)p_{ji}(n) \sum_{k=1}^{\infty} p_{ii}(k). \end{aligned}$$

Both series $\sum_{k=1}^{\infty} p_{ii}(k)$ and $\sum_{k=1}^{\infty} p_{jj}(k)$ are simultaneously convergent or divergent because $p_{ij}(m) > 0$ and $p_{ji}(n) > 0$; thus, by assertion (a) of the theorem, the states i and j are recurrent or transient at the same time.

- (c) Applying the discrete renewal Eq. (3.6) and result (a), assertion (c) immediately follows. □

Definition 3.26. A Markov chain is called **recurrent** or **transient** if every state is recurrent or transient.

Comment 3.27. Using the n -step transition probabilities $p_{ii}(n)$, a simple formula can be given for the expected value of the number of returns to a state $i \in \mathcal{X}$. Let $X_0 = i$ be the initial state of the Markov chain. The expected value of the return number is expressed as

$$\begin{aligned} \mathbf{E} \left(\sum_{k=1}^{\infty} \mathcal{I}_{\{X_k=i\}} | X_0 = i \right) &= \sum_{k=1}^{\infty} \mathbf{E} (\mathcal{I}_{\{X_k=i\}} | X_0 = i) \\ &= \sum_{k=1}^{\infty} \mathbf{P} (X_k = i | X_0 = i) = \sum_{k=1}^{\infty} p_{ii}(k). \end{aligned}$$

The assertion of Theorem 3.25 can be interpreted in another way: a state $i \in \mathcal{X}$ is recurrent if and only if the expected value of the number of returns equals infinity.

Example 3.28 (Recurrent Markov chain). Consider the random walk process $X = (X_n, n = 0, 1, \dots)$ described in Example 3.10. The process, starting from the origin, at all steps moves one unit to the right with probability p and to the left

with probability $(1 - p)$, independently of each other. We have proved earlier that the process X is a homogeneous, irreducible, and periodic Markov chain with period 2. Here we discuss the conditions under which the Markov chain will be recurrent.

By the condition $X_0 = 0$, it is clear that $p_{00}(2k + 1) = 0$, $k = 0, 1, \dots$. The process can return in $2k$ steps to the state 0 only if it moves, in some way, k times to the left and k times to the right, the probability of which is

$$p_{00}(2k) = \binom{2k}{k} p^k (1 - p)^k = \frac{(2k)!}{k!k!} [p(1 - p)]^k.$$

Using the well-known Stirling's formula, which gives an asymptotic relation for $k!$ as $k \rightarrow \infty$ as follows (see p. 616 of [5]):

$$\sqrt{2\pi} k^{k+1/2} e^{-k} < k! < \sqrt{2\pi} k^{k+1/2} e^{-k} \left(1 + \frac{1}{4k}\right);$$

then

$$k! \approx \left(\frac{k}{e}\right)^k \sqrt{2\pi k};$$

and thus we have

$$p_{00}(2k) \approx \left(\frac{2k}{e}\right)^{2k} \sqrt{2\pi(2k)} \left(\left(\frac{k}{e}\right)^k \sqrt{2\pi k}\right)^{-2} [p(1 - p)]^k = \frac{[4p(1 - p)]^k}{\sqrt{\pi k}}.$$

By the inequality between arithmetic and geometrical means, the numerator has an upper bound

$$4[p(1 - p)] \leq 4 \left[\frac{p + (1 - p)}{2} \right]^2 = 1,$$

where the equality holds if and only if $p = 1 - p$, that is, $p = 1/2$. In all other cases the product is less than 1; consequently, the sum of return probabilities $p_{00}(2k)$ is divergent if and only if $p = 1/2$ (symmetric random walk); otherwise, it is convergent. As a consequence of Theorem 3.25, we obtain that the state 0, and together with it all states of the Markov chain, is recurrent if and only if $p = 1/2$.

Note that a similar result is valid if we consider the random walk with integer coordinates in a plane. It can be verified that only in the case of a symmetric random walk will the state $(0, 0)$ be recurrent, when the probabilities of the movements left-right-up-down are $1/4 - 1/4$. In addition, if a random walk is defined in a similar way in higher-dimensional (≥ 3) spaces, then the Markov chain will no longer be recurrent.

3.2 Fundamental Limit Theorem of Homogeneous Markov Chains

3.2.1 Positive Recurrent and Null Recurrent Markov Chains

Let X be a homogeneous Markov chain with the finite ($N < \infty$) or countably infinite ($N = \infty$) state space $\mathcal{X} = \{0, 1, \dots, N\}$ and (one-step) transition probability matrix $\Pi = [p_{ij}]_{i \in \mathcal{X}}$. Let $P = (p_i = \mathbf{P}(X_0 = i), i \in \mathcal{X})$ be the initial distribution. Denote by $P(n) = (P_i(n) = \mathbf{P}(X_n = i), i \in \mathcal{X}), n = 0, 1, \dots$, the time-dependent distribution of the Markov chain; then $P(0) = P$.

The main question to be investigated here concerns the conditions under which there exists a limit distribution of m -step transition probabilities

$$\lim_{m \rightarrow \infty} p_{ij}(m) = \pi_j, \text{ where } \pi_j \geq 0 \text{ and } \sum_{i \in \mathcal{X}} \pi_i = 1$$

and how it can be determined. The answer is closely related to the behavior of the recurrent states i of a Markov chain. Note that the condition of recurrence $f_{ii} = \sum_{k=1}^{\infty} f_{ii}(k) = 1$ does not ensure the existence of a limit distribution. The main characteristics are the expected values of the return times $\mu_i = T_{ii} = \sum_{k=1}^{\infty} k f_{ii}(k)$, and the recurrent states will be classified according to whether or not the μ_i are finite because the condition $\mu_i < \infty, i \in \mathcal{X}$, guarantees the existence of a limit distribution.

Definition 3.29. A recurrent state $i \in \mathcal{X}$ is called **positive recurrent** (or **nonnull recurrent**) if the return time has a finite expected value μ_i ; otherwise, if $\mu_i = \infty$, then it is called **null recurrent**.

Theorem 3.30. Let X be a homogeneous, irreducible, aperiodic, and recurrent Markov chain. Then for all states $i, j \in \mathcal{X}$,

$$\lim_{m \rightarrow \infty} p_{ij}(m) = \frac{1}{\mu_j}.$$

Note that this theorem not only gives the limit of the m -step transition probabilities with the help of the expected value of the return times, but it interprets the notion of positive and null recurrence. By definition, a recurrent state j is positive recurrent if $1/\mu_j > 0$ and null recurrent if $1/\mu_j = 0$ (here and subsequently, we write $1/\infty = 0$). The assertion given in the theorem is closely related to the discrete renewal Eq. (3.6), and using it we can prove a limit theorem, as the following lemma shows (see [29] and Chap. XIII of [31]).

Lemma 3.31 (Erdős, Feller, Pollard). Let $(q_i, i \geq 0)$ be an arbitrary distribution on the natural numbers, i.e., $q_i \geq 0, \sum_{i=0}^{\infty} q_i = 1$. Assume that the distribution $(q_i, i \geq 0)$ is not latticed, that is, the greatest common divisor of the indices with

the probabilities $q_i > 0$ equals 1. If the sequence $\{v_n, n \geq 0\}$, satisfies the discrete renewal equation

$$v_0 = 1, \quad v_n = \sum_{k=1}^n q_k v_{n-k}, \quad n \geq 1,$$

then

$$\lim_{n \rightarrow \infty} v_n = \frac{1}{\mu},$$

where $\mu = \sum_{k=1}^{\infty} k q_k$ and $\frac{1}{\mu} = 0$ if $\mu = \infty$.

The proof of Theorem 3.30 uses the following result from analysis.

Lemma 3.32. Assume that the sequence (q_1, q_2, \dots) of nonnegative real numbers satisfies the condition $\sum_{i=0}^{\infty} q_i = 1$. If the sequence of real numbers $(w_n, n \geq 0)$ is convergent, $\lim_{n \rightarrow \infty} w_n = w$, then

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n q_{n-k} w_k = w.$$

Proof. It is clear that the elements of $\{w_n\}$ are bounded; then there exists a number W such that $|w_n| \leq W, n \geq 0$. From the conditions $\lim_{n \rightarrow \infty} w_n = w$ and $\sum_{i=0}^{\infty} q_i = 1$ it follows for any $\varepsilon > 0$ that there exist integers $N(\varepsilon)$ and $K(\varepsilon)$ such that

$$|w_n - w| < \varepsilon \quad \text{and} \quad \sum_{k=K(\varepsilon)}^{\infty} q_k < \varepsilon.$$

It is easy to check that for every $n > n(\varepsilon) = \max(N(\varepsilon), K(\varepsilon))$,

$$\begin{aligned} |w_n - w| &\leq \left| \sum_{k=0}^n q_k w_{n-k} - \sum_{k=0}^n q_k w \right| \\ &\leq \sum_{k=0}^{n(\varepsilon)} q_k |w_{n-k} - w| + \sum_{k=n(\varepsilon)+1}^n q_k |w_{n-k} - w| + \sum_{k=n+1}^{\infty} q_k |w| \\ &\leq \sum_{k=0}^{n(\varepsilon)} q_k \varepsilon + \sum_{k=n(\varepsilon)+1}^n q_k (W + |w|) + \sum_{k=n+1}^{\infty} q_k |w| \\ &\leq \varepsilon + \varepsilon(W + |w|) + \varepsilon |w| = \varepsilon(1 + W + 2|w|). \end{aligned}$$

Since $\varepsilon > 0$ can be chosen freely, we get the convergence $w_n \rightarrow w, n \rightarrow \infty$. \square

Proof (Theorem 3.30).

- (a) We prove firstly the assertion for the case $i = j$. By the discrete renewal equation

$$p_{ii}(0) = 1, \quad p_{ii}(n) = \sum_{k=1}^n f_{ii}(k)p_{ii}(n-k), \quad n = 1, 2, \dots,$$

where the state i is recurrent, $f_{ii} = \sum_{k=1}^{\infty} f_{ii}(k) = 1$ ($f_{ii} \geq 0$). Using the assertion of Lemma 3.31 we have

$$\lim_{n \rightarrow \infty} p_{ii}(n) = \frac{1}{\mu_i}.$$

- (b) Now let $i \neq j$, and apply Lemma 3.32. Since the Markov chain is irreducible and recurrent, $f_{ij} = \sum_{k=1}^{\infty} f_{ij}(k) = 1$ ($f_{ij} \geq 0$). Then, as $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} p_{ij}(n) = \lim_{n \rightarrow \infty} \sum_{k=1}^n f_{ij}(k)p_{jj}(n-k) = \sum_{k=1}^{\infty} f_{ij}(k) \frac{1}{\mu_j} = \frac{1}{\mu_j}. \quad \square$$

Similar results can be easily proven for periodic cases. Let X be a homogeneous, irreducible, and recurrent Markov chain with period $d > 1$. Then the state space \mathcal{X} can be decomposed into disjoint subsets $\mathcal{X}_0, \dots, \mathcal{X}_{d-1}$ [see Eq. (3.21)] such that the Markov chain allows only for cyclic transitions between the states of the sets \mathcal{X}_i : $\mathcal{X}_0 \rightarrow \mathcal{X}_1 \rightarrow \dots \rightarrow \mathcal{X}_{d-1} \rightarrow \mathcal{X}_0$. Let $0 \leq k, m \leq d-1$ be arbitrarily fixed integers; then, starting from a state $i \in \mathcal{X}_k$, the process arrives at a state of \mathcal{X}_k in exactly

$$\ell = \begin{cases} m-k, & \text{if } k < m, \\ m-k+d, & \text{if } m \leq k, \end{cases}$$

steps. From this follows $p_{ij}(s) = 0$ if $s-1$ is divisible by d .

Theorem 3.33. *Let X be a homogeneous, irreducible, and recurrent Markov chain with period $d > 1$ and $i \in \mathcal{X}_k, j \in \mathcal{X}_m$ arbitrarily fixed states. Then*

$$\lim_{n \rightarrow \infty} p_{ij}(\ell + nd) = \frac{d}{\mu_j},$$

where $\mu_j = \sum_{k=1}^{\infty} k f_{jj}(k) = \sum_{r=1}^{\infty} rd f_{jj}(rd)$.

Proof. First assume $k = m$, and consider the transition probabilities $p_{ij}(nd)$ for $i, j \in \mathcal{X}_k$. This is equivalent to the case (see Conclusion 3.22 according to

the cyclic transitions of a Markov chain) where we investigate the Markov chain \bar{X} with the state space \mathcal{X}_k and it has the (one-step) transition probability matrix $\bar{\Pi} = [\bar{p}_{ij}]_{i,j \in \mathcal{X}_k}$, $\bar{p}_{ij} = p_{ij}(d)$, $i, j \in \mathcal{X}_k$. Obviously, the Markov chain \bar{X} that originated from X is a homogeneous, irreducible, recurrent, and aperiodic Markov chain. Using the limit theorem 3.31 we obtain

$$\lim_{n \rightarrow \infty} \bar{p}_{ii}(n) = \lim_{n \rightarrow \infty} p_{ii}(nd) = \frac{1}{\sum_{k=1}^{\infty} k f_{ii}(kd)} = \frac{d}{\sum_{k=1}^{\infty} k d f_{ii}(kd)} = \frac{d}{\sum_{k=1}^{\infty} k f_{ii}(k)} = \frac{d}{\mu_i},$$

where $f_{ii}(r) = 0$ if $r \neq d, 2d, \dots$

Assume now that $k \neq m$. Then $f_{ij}(k) = 0$ and $p_{ij}(k) = 0$ if $k \neq \ell + nd$, $n \geq 0$; moreover, the Markov chain \bar{X} is recurrent because

$$f_{ij} = \sum_{s=1}^{\infty} f_{ij}(s) = \sum_{k=1}^{\infty} f_{ij}(\ell + rd) = 1;$$

then

$$p_{ij}(\ell + nd) = \sum_{k=1}^{\ell+nd} f_{ij}(k) p_{jj}(\ell + nd - k) = \sum_{r=1}^{\ell+nd} f_{ij}(\ell + rd) p_{jj}(rd) \rightarrow \frac{d}{\mu_j}, \quad n \rightarrow \infty.$$

□

Theorem 3.34. *If the homogeneous Markov chain X is irreducible and has a positive recurrent state $i \in \mathcal{X}$, then all its states are positive recurrent.*

Proof. Let $j \in \mathcal{X}$ be arbitrary. Since the Markov chain is irreducible, there exist integers $s, t > 0$ such that $p_{ij}(s) > 0$, $p_{ji}(t) > 0$. Denote by d the period of the Markov chain. It is clear that $d > 0$ because $p_{ii}(s + t) \geq p_{ij}(s) p_{ji}(t) > 0$. Moreover,

$$\begin{aligned} p_{ii}(s + nd + t) &\geq p_{ij}(s) p_{jj}(nd) p_{ji}(t), \\ p_{jj}(s + nd + t) &\geq p_{ji}(t) p_{ii}(nd) p_{ij}(s). \end{aligned}$$

Applying Theorem 3.33 and taking the limit as $n \rightarrow \infty$ we have

$$\frac{1}{\mu_i} \geq p_{ij}(s) \frac{1}{\mu_j} p_{ji}(t), \quad \frac{1}{\mu_j} \geq p_{ij}(s) \frac{1}{\mu_i} p_{ji}(t);$$

thus

$$\frac{1}{\mu_i} \geq p_{ij}(s) p_{ji}(t) \frac{1}{\mu_j} \geq [p_{ij}(s) p_{ji}(t)]^2 \frac{1}{\mu_i}.$$

From the last inequality it immediately follows that when the state i is recurrent, at the same time j is also recurrent. □

Summing up the results derived previously, we can state the following theorem.

Theorem 3.35. *Let X be a homogeneous irreducible Markov chain; then*

1. *All states are aperiodic or all states are periodic with the same period,*
2. *All states are transient or all states are recurrent, and in the latter case*
 - *All are positive recurrent or all are null recurrent.*

3.2.2 Stationary Distribution of Markov Chains

Retaining the notations introduced previously, $P(n) = (P_i(n) = \mathbf{P}(X_n = i), i \in \mathcal{X})$ denotes the distribution of a Markov chain depending on the time $n \geq 0$. Then $P(0) = (P_i(0) = p_i, i \in \mathcal{X})$ is the initial distribution.

Definition 3.36. Let $\pi = (\pi_i, i \in \mathcal{X})$ be a distribution, i.e., $\pi_i \geq 0$ and $\sum_{i \in \mathcal{X}} \pi_i = 1$. π is called a **stationary distribution** of the Markov chain X if by choosing $P(0) = \pi$ as the initial distribution, the distribution of the process does not depend on time, that is,

$$P(n) = \pi, n \geq 0.$$

A stationary distribution is also called an **equilibrium distribution** of a chain.

With Markov chains, the main problem is the existence and determination of stationary distributions. Theorem 3.30 deals with the convergence of the sequence of n -step transition probabilities $P(n)$ as $n \rightarrow \infty$, and if it converges, then the limit gives the stationary distribution of the chain. The proofs of these results are not too difficult but consist of many technical steps [35, 36], and so we omit them here.

Theorem 3.37. *Let X be a homogeneous, irreducible, recurrent, and aperiodic Markov chain. Then the following assertions hold:*

(A) *The limit*

$$\pi_i = \lim_{n \rightarrow \infty} P_i(n) = \frac{1}{\mu_i}, i \in \mathcal{X},$$

exists and does not depend on the initial distribution.

(B) *If all states are recurrent null states, then the stationary distribution does not exist and $\pi_i = 0$ for all $i \in \mathcal{X}$.*

(C) *If all states are positive recurrent, then the stationary distribution $\pi = (\pi_i, i \in \mathcal{X})$ does exist and $\pi_i = 1/\mu_i > 0$ for all $i \in \mathcal{X}$ and $P(n) \rightarrow \pi$, as $n \rightarrow \infty$. The stationary distribution is unique and satisfies the system of linear equations*

$$\sum_{i \in \mathcal{X}} \pi_i = 1, \tag{3.7}$$

$$\pi_i = \sum_{j \in \mathcal{X}} \pi_j p_{ji}, i \in \mathcal{X}. \tag{3.8}$$

Comment 3.38. *Since the Markov chain is irreducible, it is enough to require in part (B) the existence of a positive recurrent state because from the existence of a single positive recurrent state and the fact that the Markov chain is irreducible it follows that all states are positive recurrent.*

Equation (3.8) of Theorem 3.37 can be rewritten in the more concise form $\pi = \pi\Pi$, where Π is the one-step transition probability matrix of the chain.

The initial distribution does not play a role in Eqs. (3.7) and (3.8); therefore, when the stationary distribution π exists, it does not depend on the initial distribution, only on the transition probability matrix Π .

Given that the stationary distribution π exists, it can be easily proven that π satisfies the system of linear Eq. (3.8), and at the same time, these circumstances lead to an iterative method of solution [see Eq. (3.9) below]. This iterative procedure to determine the stationary distribution can be applied to chains with finite state spaces.

The time-dependent distribution $P(n) = (P_0(n), P_1(n), \dots)$ satisfies the equation for all $n = 0, 1, \dots$,

$$P(n) = P(n-1)\Pi. \quad (3.9)$$

Repeating this equation n times, we have

$$P(n) = P(0)\Pi^n, \quad n = 0, 1, \dots$$

Since it is assumed that the stationary distribution π exists, we can write

$$\pi = \lim_{n \rightarrow \infty} P(n);$$

thus from the equation

$$\lim_{n \rightarrow \infty} P(n) = \lim_{n \rightarrow \infty} P(n-1)\Pi$$

it follows that

$$\pi = \pi\Pi.$$

Definition 3.39. A state i of an irreducible homogeneous Markov chain X is called **ergodic** if the state i is aperiodic and positive recurrent, i.e., $d(i) = 1$, $\mu_i < \infty$. If all states of the chain are ergodic, then the Markov chain is called ergodic.

Here we define the ergodic property only of Markov chains. This property can be defined for much more complex stochastic processes as well.

By Theorem 3.37, a homogeneous, aperiodic, positive recurrent Markov chain is always ergodic. Since an irreducible Markov chain with finite state space is positive recurrent, the following statement is also true.

Theorem 3.40. *A homogeneous, irreducible, aperiodic Markov chain with finite state space is ergodic.*

In practical applications, the equilibrium distributions of Markov chains play an essential role. In what follows, we give two theorems without proofs whose conditions ensure the existence of the stationary distribution of a homogeneous, irreducible, aperiodic Markov chain X with state space $\mathcal{X} = \{0, 1, \dots\}$. The third theorem gives an upper bound for the convergence rate to the stationary distribution of the iterative procedure (3.9).

Theorem 3.41 (Klimov [56]). *If there exists a function $g(i)$, $i \in \mathcal{X}$, a state $i_0 \in \mathcal{X}$, and a positive constant ε such that the relations*

$$\mathbf{E}(g(X_{n+1}) \mid X_n = i) \leq g(i) - \varepsilon, \quad i \geq i_0, \quad n \geq 0,$$

$$\mathbf{E}(g(X_{n+1}) \mid X_n = i) < \infty, \quad i \geq 0, \quad n \geq 0,$$

hold, then the chain X is ergodic.

Theorem 3.42 (Foster [33]). *Assume that there exist constants $a, b > 0$ and $\ell \geq 0$ such that the inequalities*

$$\mathbf{E}(X_{n+1} \mid X_n = i) \leq a, \quad i \leq \ell,$$

$$\mathbf{E}(X_{n+1} \mid X_n = i) \leq i - b, \quad i > \ell,$$

are valid. Then the Markov chain X is ergodic.

Theorem 3.43 (Bernstein [10]). *Assume that there exist a state $i_0 \in \mathcal{X}$ and a constant $\lambda > 0$ such that for all $i \in \mathcal{X}$ the inequality $p_{ii_0} \geq \lambda$ holds. Then*

$$\lim_{n \rightarrow \infty} p_{ij}(n) = \pi_j, \quad i, j \in \mathcal{X},$$

where $\pi = (\pi_i, i \in \mathcal{X})$ denotes the stationary distribution of the Markov chain; moreover,

$$\sum_{i \in \mathcal{X}} |p_{ij}(n) - \pi_j| \leq 2(1 - \lambda)^n, \quad n \geq 1.$$

3.2.3 Ergodic Theorems for Markov Chains

Let X be a homogeneous irreducible and positive recurrent Markov chain with state space $\mathcal{X} = \{0, 1, \dots\}$ and i a fixed state. Compute the time and the relative frequencies when the process stays in the state i on the time interval $[0, T]$ as follows:

$$S_i(T) = \sum_{n=0}^T \mathcal{I}_{\{X_n=i\}},$$

$$\bar{S}_i(T) = \frac{1}{T} \sum_{n=0}^T \mathcal{I}_{\{X_n=i\}} = \frac{1}{T} S_i(T).$$

Let us consider when and in what sense there exists a limit of the relative frequencies $\bar{S}_i(T)$ as $T \rightarrow \infty$ and, if it exists, how it can be determined. This problem has, in particular, practical importance when applying simulation methods. To clarify the stochastic background of the problem, we introduce the following notations.

Assume that a process starts at time 0 from the state i . Let $0 = T_0^{(i)} < T_1^{(i)} < T_2^{(i)} < \dots$ be the sequence of the consecutive random time points when a Markov chain arrives at the state i , that is, $T_k^{(i)}$, $k = 1, 2, \dots$, are the return time points to the state i of the chain. This means that

$$X(T_n^{(i)}) = i, \quad n = 0, 1, \dots \quad \text{and} \quad X(k) \neq i, \quad \text{if } k \neq T_0^{(i)}, T_1^{(i)}, \dots$$

Denote by

$$\tau_k^{(i)} = T_k^{(i)} - T_{k-1}^{(i)}, \quad k = 1, 2, \dots,$$

the time length between the return time points. Since the Markov chain has the memoryless property, these random variables are independent; moreover, from the homogeneity of the Markov chain it follows that $\tau_n^{(i)}$, $n \geq 1$, are also identically distributed. The common distribution of these random variables $\tau_n^{(i)}$ is the distribution of the return times from the state i to i , namely, $(f_{ii}(n), n \geq 1)$.

Heuristically, it is clear that when the return time has a finite expected value μ_i , then during the time T the process returns to the state i on average T/μ_i times. This means that the quantity $\bar{S}_i(T)$ fluctuates around the value $1/\mu_i$ and has the same limit as $T \rightarrow \infty$. This result can be given in exact mathematical form on the basis of the law of large numbers as follows.

Theorem 3.44. *If X is an ergodic Markov chain, then, with probability 1,*

$$\lim_{T \rightarrow \infty} \bar{S}_i(T) = \frac{1}{\mu_i}, \quad i \in \mathcal{X}. \quad (3.10)$$

If the Markov property is satisfied, then not only are the return times independent and identically distributed, but the stochastic behaviors of the processes on the return periods are identical as well. This fact allows us to prove more general results for an ergodic Markov chain as Eq. (3.10).

Theorem 3.45. *Let X be an ergodic Markov chain and $g(i)$, $i \in \mathcal{X}$, be a real-valued function such that $\sum_{i \in \mathcal{X}} \pi_i |g(i)| < \infty$. Then the convergence*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{n=1}^T g(X_n) = \sum_{i \in \mathcal{X}} \pi_i g(i)$$

is true with probability 1, where π_i , $i \in \mathcal{X}$, denotes the stationary distribution of the Markov chain, which exists under the given condition.

3.2.4 Estimation of Transition Probabilities

In modeling ergodic Markov chains an important question is to estimate the transition probabilities by the observation of the chain. The relative frequencies give corresponding estimates of the probabilities because by Theorem 3.44 they tend to them with probability 1 under the given conditions. Note that from the heuristic approach discussed previously it follows under quite general conditions that not only can the law of large numbers be derived for the relative frequencies, but the central limit theorems can as well.

Consider now the estimate of transition probabilities with the maximum likelihood method. Let X be an ergodic Markov chain with finite state space $\mathcal{X} = \{0, 1, \dots, N\}$ and with the (one-step) transition probability matrix $\Pi = (p_{ij})_{i,j \in \mathcal{X}}$. Assume that we have an observation of n elements $X_1 = i_1, \dots, X_n = i_n$ starting from the initial state $X_0 = i_0$, and we will estimate the entries of the matrix Π . By the Markov property, the conditional likelihood function can be given in the form

$$\mathbf{P}(X_1 = i_1, \dots, X_n = i_n \mid X_0 = i_0) = p_{i_0 i_1} \cdots p_{i_{n-1} i_n}.$$

Denote by n_{ij} , $i, j \in \mathcal{X}$, the number of one-step transitions from the state i to j in the sample path i_0, i_1, \dots, i_n , and let $0^0 = 1$, $0/0 = 0$. Then the conditional likelihood function given the $X_0 = i_0$ initial state is

$$L(i_1, \dots, i_n; \Pi \mid i_0) = \prod_{i=0}^N \left(\prod_{j=0}^N p_{ij}^{n_{ij}} \right). \quad (3.11)$$

Applying the maximum likelihood method, maximize the expression in p_{ij} under the conditions

$$p_{ij} \geq 0, \quad i, j \in \mathcal{X}, \quad \sum_{j \in \mathcal{X}} p_{ij} = 1, \quad i \in \mathcal{X}.$$

It is clear that there are no relations between the products playing a role in the parentheses of Eq. (3.11) for different i ; therefore, the maximization problem can be solved by means of $N + 1$ different, but similar, optimization problems:

$$\max \left\{ \prod_{j=0}^N p_{ij}^{n_{ij}} : p_{ij} \geq 0, \sum_{j \in \mathcal{X}} p_{ij} = 1 \right\}, \quad i = 0, 1, \dots, N.$$

Obviously it is enough to solve it only for one state i since the others can be derived analogously to that one.

Let $i \in \mathcal{X}$ be a fixed state, and denote $n_i = \sum_{j \in \mathcal{X}} n_{ij}$. Apply the Lagrange multiplier method; then for every $m = 0, \dots, N$,

$$\frac{\partial}{\partial p_{im}} \left(\prod_{j=0}^N p_{ij}^{n_{ij}} + \lambda(p_{i0} + p_{i1} + \dots + p_{iN} - 1) \right) = \frac{n_{im}}{p_{im}} \prod_{j=0}^N p_{ij}^{n_{ij}} + \lambda = 0;$$

consequently, for a constant λ_i we have

$$\frac{n_{im}}{p_{im}} = \lambda \prod_{j=0}^N p_{ij}^{-n_{ij}} = \lambda_i, \quad m = 0, \dots, N.$$

From this it follows that the equations

$$n_{im} = \lambda_i p_{im}, \quad m = 0, \dots, N,$$

hold; then

$$\sum_{m=0}^N n_{im} = n_i = \lambda_i \sum_{m=0}^N p_{im} = \lambda_i.$$

These relations lead to the conditional maximum likelihood estimates for the transition probabilities p_{im} as follows:

$$\hat{p}_{im} = \frac{n_{im}}{\lambda_i} = \frac{n_{im}}{n_i}, \quad 0 \leq i, m \leq N.$$

It can be verified that these estimates \hat{p}_{im} converge to p_{im} with probability 1 as $\rightarrow \infty$.

3.3 Continuous-Time Markov Chains

Like the case of the DTMCs, we assume that the state space \mathcal{X} is a finite $\{0, 1, \dots, N\}$ or countably infinite set $\{0, 1, \dots\}$ and assume that the time parameter varies in $\mathcal{T} = [0, \infty)$. According to the general definition (3.1), a process $X = (X_t, t \geq 0)$ is said to be CTMC with state space \mathcal{X} if for every positive integer n and $0 \leq t_0 < t_1 < \dots < t_n, i_0, \dots, i_n \in \mathcal{X}$, the equation

$$\begin{aligned} & \mathbf{P}(X_{t_n} = i_n \mid X_{t_{n-1}} = i_{n-1}, \dots, X_{t_0} = i_0) \\ &= \mathbf{P}(X_{t_n} = i_n \mid X_{t_{n-1}} = i_{n-1}) = p_{i_{n-1}, i_n}(t_{n-1}, t_n) \end{aligned}$$

holds, provided that a conditional probability exists. The Markov chain X is (time) homogeneous if the transition probability function $p_{ij}(s, t)$ satisfies the condition $p_{ij}(s, t) = p_{ij}(t - s)$ for all $i, j \in \mathcal{X}, 0 \leq s \leq t$. Denote by $\Pi(s, t) = [p_{ij}(s, t), i, j \in \mathcal{X}]$ the transition probability matrix.

In the case of a CTMC the time index $t \in [a, b]$ can take uncountably many values for arbitrary $0 \leq a < b < \infty$; therefore, the collection of random variables X_t , $t \in (a, b]$, is also uncountable. If we consider the questions in accordance with the sample paths of the chain, then these circumstances can lead to measurability problems (discussed later). However, the Markov processes that will be investigated later are the so-called stepwise processes, and they ensure the necessary measurability property.

We will deal mainly with the part of the theory that is relevant to queueing theory, and we touch upon only some questions in general cases showing the root of the measurability problems. A discussion of jumping processes, which is more general than the investigation of stepwise Markov chains, can be found in [36, Chap. III].

If the Markov chain $\{X_t, t \geq 0\}$, is homogeneous, then the transition probability functions $p_{ij}(s, t)$ can be given in a simpler form:

$$p_{ij}(s, s+t) = p_{ij}(t), \quad i, j \in \mathcal{X}, \quad s, t \geq 0,$$

and thus the matrix form of transition probabilities is

$$\Pi(s, s+t) = \Pi(t), \quad s, t \geq 0.$$

As was done previously, denote by

$$P(t) = (P_0(t), P_1(t), \dots), \quad t \geq 0,$$

the time-dependent distribution of the chain, where $P_i(t) = \mathbf{P}(X_t = i)$, $i \in \mathcal{X}$; then $P(0)$ means the initial distribution, while if there exists a state $k \in \mathcal{X}$ such that $\mathbf{P}(X_0 = k) = 1$, then k is the initial state.

3.3.1 Characterization of Homogeneous Continuous-Time Markov Chains

We now deal with the main properties of homogeneous CTMCs. Similarly to the discrete-time case, the transition probabilities satisfy the following conditions.

- (A) $p_{ij}(s) \geq 0$, $s \geq 0$, $p_{ij}(0) = \delta_{ij}$, $i, j \in \mathcal{X}$, where δ_{ij} is the Kronecker δ -function (which equals 1 if $i = j$ and 0 if $i \neq j$).
- (B) $\sum_{j \in \mathcal{X}} p_{ij}(s) = 1$, $s \geq 0$, $i \in \mathcal{X}$.
- (C) $p_{ij}(s+t) = \sum_{k \in \mathcal{X}} p_{ik}(s)p_{kj}(t)$, $s, t \geq 0$, $i, j \in \mathcal{X}$.

An additional condition is needed for our considerations.

- (D) The transition probabilities of the Markov chain X satisfy the conditions

$$\lim_{h \rightarrow 0^+} p_{ij}(h) = p_{ij}(0) = \delta_{ij}, \quad i, j \in \mathcal{X}. \quad (3.12)$$

Comment 3.46. Condition (B) expresses that $\Pi(s)$, $s \geq 0$, is a stochastic matrix. We will not consider the so-called **killed Markov chains**, where the lifetime $[0, \tau]$ of the chain is random (where the process is defined) and with probability 1 is finite, i.e., $\mathbf{P}\{\tau < \infty\} = 1$. It should be noted that condition (B) ensures that the chain is defined on the whole interval $[0, \infty)$ because the process will be certainly in some state $i \in \mathcal{X}$ for any time $s \geq 0$.

Condition (C) is the Chapman–Kolmogorov equation related to the continuous-time case. It can be given in matrix form as follows:

$$\Pi(s + t) = \Pi(s)\Pi(t), \quad s, t \geq 0.$$

Similarly to the discrete-time case, the time-dependent distribution of the chain satisfies the equation

$$P(s + t) = P(s)\Pi(t), \quad s, t \geq 0,$$

and thus for all $t > 0$

$$P(t) = P(0)\Pi(t).$$

The last relation means that the initial distribution and the transition probabilities uniquely determine the distribution of the chain at all time points $t \geq 0$.

Instead of (D) it is enough to assume that the condition

$$\lim_{h \rightarrow 0+} p_{ii}(h) = 1, \quad i \in \mathcal{X},$$

holds, because for every $i, j \in \mathcal{X}$, $i \neq j$, the relation

$$0 \leq p_{ij}(h) \leq \sum_{j \neq i} p_{ij}(h) = 1 - p_{ii}(h) \rightarrow 0, \quad h \rightarrow 0+,$$

is true.

Under conditions (A)–(D), the following relations are valid.

Theorem 3.47. The transition probabilities $p_{ij}(t)$, $0 \leq t < \infty$, $i \neq j$, are uniformly continuous.

Proof. Using conditions (A)–(D) we obtain

$$\begin{aligned} |p_{ij}(t + h) - p_{ij}(t)| &= \left| \sum_{k \in \mathcal{X}} p_{ik}(h) p_{kj}(t) - \sum_{k \in \mathcal{X}} \delta_{ik} p_{kj}(t) \right| \\ &\leq \sum_{k \in \mathcal{X}} |p_{ik}(h) - \delta_{ik}| p_{kj}(t) \\ &\leq 1 - p_{ii}(h) + \sum_{k \neq i} p_{ik}(h) = 2(1 - p_{ii}(h)) \rightarrow 0, \quad h \rightarrow 0+. \end{aligned}$$

□

Theorem 3.48 ([36, p. 200]). For all $i, j \in \mathcal{X}$, $i \neq j$, the finite limit

$$q_{ij} = \lim_{h \rightarrow 0^+} \frac{p_{ij}(h)}{h}$$

exists.

For every $i \in \mathcal{X}$ there exists a finite or infinite limit

$$q_i = \lim_{h \rightarrow 0^+} \frac{1 - p_{ii}(h)}{h} = -p'_{ii}(0).$$

The quantities q_{ij} and q_i are the most important characteristics of a homogeneous continuous-time Markov chain. Subsequently we will use the notation $q_{ii} = -q_i$, $i \in \mathcal{X}$, also and interpret the meaning of these quantities.

Definition 3.49. The quantity q_{ij} is called the **transition rate** of intensity from the state i to the state j , while q_i is called the **transition rate** from the state i .

We classify the states in accordance with whether or not the rate q_i is finite. If $q_i < \infty$, then i is called a **stable state**, while if $q_i = +\infty$, then we say that i is an **instantaneous** state. Note that there exists a Markov chain with the property $q_i = +\infty$ [36, pp. 207–210].

Definition 3.50. A stable noninstantaneous state i is called **regular** if

$$\sum_{i \neq j} q_{ij} = -q_{ii} = q_i,$$

and a Markov chain is **locally regular** if all its states are regular.

Corollary 3.51. As a consequence of Theorem 3.48, we obtain that locally regular Markov chains satisfy the following asymptotic properties as $h \rightarrow 0^+$:

$$\mathbf{P}(X_{t+h} \neq i \mid X_t = i) = q_i h + o(h),$$

$$\mathbf{P}(X_{t+h} = i \mid X_t = i) = 1 - q_i h + o(h),$$

$$\mathbf{P}(X_{t+h} = j \mid X_t = i) = q_{ij} h + o(h), \quad j \neq i.$$

From Theorem 3.48 it also follows that Markov chains with a finite state space are locally regular because all q_{ij} , $i \neq j$, are finite and, consequently, all q_i are also finite.

The condition

$$q = \sup_{i \in \mathcal{X}} q_i < \infty \tag{3.13}$$

will play an important role in our subsequent investigations. We introduce the notation

$$Q = [q_{ij}]_{i,j \in \mathcal{X}} = [p'_{ij}(0)]_{i,j \in \mathcal{X}} = \Pi'(0)$$

for locally regular Markov chains. Recall that

$$\lim_{t \rightarrow 0^+} \Pi(t) = \Pi(0) = I, \quad (3.14)$$

where I is the identity matrix with suitable dimension.

Definition 3.52. The matrix Q is called a **rate** or **infinitesimal matrix** of a continuous-time Markov chain.

The following assertions hold for all locally regular Markov chains under the initial condition (3.14) [36, pp. 204–206].

Theorem 3.53. *The transition probabilities of a locally regular Markov chain satisfy the Kolmogorov backward differential equation*

$$\Pi'(t) = Q \Pi(t), \quad t \geq 0 \quad (I).$$

If condition (3.13) is fulfilled, then the Kolmogorov forward differential equation

$$\Pi'(t) = \Pi(t) Q, \quad t \geq 0 \quad (II)$$

is valid. Under condition (3.13) differential Eqs. (I) and (II), referred to as first- and second-system Kolmogorov equations, have unique solutions.

3.3.2 Stepwise Markov Chains

The results of Theorem 3.53 are related to the analytical properties of transition probabilities and do not deal with the stochastic behavior of sample paths. In this part we investigate the so-called stepwise Markov chains and their sample paths. We introduce the embedded Markov chain and consider the transition probabilities and holding times. In the remaining part of this chapter we assume that the Markov chain is locally regular and condition (3.13) holds.

Definition 3.54. A Markov chain X is a **jump process** if for any $t \geq 0$ there exists a random time $\Delta = \Delta(t, \omega) > 0$ such that

$$X_s = X_t, \quad \text{if } s \in [t, t + \Delta).$$

In the definition, Δ can be the remaining time the process stays at state $X(t)$, and the definition requires that this time be positive.

Definition 3.55. We say that a Markov chain has a **jump** at time $t_0 > 0$ if there exists a monotonically increasing sequence t_1, t_2, \dots such that $t_n \rightarrow t_0, n \rightarrow \infty$ and at the same time $X_{t_n} \neq X_{t_0}, n = 1, 2, \dots$. A Markov chain is called a **stepwise process** if it is a jump process and the number of jumps is finite for all sample paths on all finite intervals $[0, t]$.

It should be noted that a stepwise process is continuous from the right and has a limit from the left at all jumping points.

Denote by $(\tau_0 =) 0 < \tau_1 < \tau_2 < \dots$ the sequence of consecutive jumping points; then all finite time intervals consist, at most, of finite jumping points. Between two jumping points the state of the process does not change, and this time is called the **holding time**.

Definition 3.56. A stepwise Markov chain is called **regular** if the sequence of holding times $\zeta_k = \tau_{k+1} - \tau_k, k = 0, 1, \dots$, satisfies the condition

$$\mathbf{P} \left(\sum_{k=0}^{\infty} \zeta_k = \infty \right) = 1.$$

By the definition of stepwise process, we have

$$X_s \equiv X_{\tau_i}, s \in [\tau_i, \tau_{i+1}), i = 0, 1, \dots$$

Denote by $Y_k = X_{\tau_k}, k = 0, 1, \dots$, the states at time points where the transitions change, and define for $i \neq j$

$$\pi_{ij} = \begin{cases} \frac{q_{ij}}{q_i}, & \text{if } q_i > 0, \\ 0, & \text{if } q_i = 0. \end{cases} \quad (3.15)$$

In addition, let

$$\pi_{ii} = 1 - \sum_{j \neq i} \pi_{ij}. \quad (3.16)$$

By the Markov property, the process $(Y_k, k \geq 0)$, is a discrete-time homogeneous Markov chain with the state space $\mathcal{X} = \{0, 1, \dots\}$ and the transition probabilities

$$\mathbf{P}(Y_{n+1} = j \mid Y_n = i) = \pi_{ij}, ij \in \mathcal{X}, n \geq 0.$$

The process $(Y_k, k \geq 0)$ is called an **embedded Markov chain** of the continuous-time stepwise Markov chain X .

Note that the condition $q_i = 0$ corresponds to the case where i is an absorbing state, and in other cases the holding times for arbitrary state i have an exponential distribution with parameter q_i whose density function is $q_i e^{-q_i x}, x > 0$.

3.3.3 Construction of Stepwise Markov Chains

The construction derived here gives a method for simulating stepwise Markov chains at the same time. Thus, we construct a CTMC $\{X_t, t \geq 0\}$, with initial distribution $P(0) = (P_0(0), P_1(0), \dots)$ and transition probability matrix $\Pi(t) = [p_{ij}(t)]$, $t \geq 0$, satisfying condition (3.13).

Using notations (3.15) and (3.16), define the random time intervals with length S_0, S_1, \dots , nonnegative random variables K_0, K_1, \dots , taking integer numbers and the random jumping points $\tau_m = S_0 + \dots + S_{m-1}$, $m = 1, 2, \dots$, by the following procedure.

- (a) Generate a random variable K_0 with distribution $P(0)$ [i.e., $\mathbf{P}(K_0 = k) = P_k(0)$, $k \in \mathcal{X}$] and a random variable S_0 distributed exponentially with parameter q_{K_0} conditionally dependent on K_0 . Define $X_t = K_0$ if $0 \leq t < S_0$.
- (b) In the m th steps ($m = 1, 2, \dots$) generate a random variable K_m with distribution $P^{(m)} = (\pi_{K_{m-1}, j}, j \in \mathcal{X})$, and a random variable S_m distributed exponentially with the parameter q_{K_m} . Define $X_t = K_m$ if $\tau_m \leq t < \tau_{m+1}$, $m = 0, 1, \dots$

Then the stochastic process $\{X_t, t \geq 0\}$ is a stepwise Markov chain with initial distribution $P(0)$ and transition probability matrix $\Pi(t)$, $t \geq 0$.

3.3.4 Some Properties of the Sample Path of Continuous-Time Markov Chains

By the considerations of the sample paths of CTMCs, there are problems that cannot arise in the case of discrete-time chains. For example, X_t , $t \geq 0$, are random variables; therefore, $\{X_t \leq x\} \in \mathcal{A}$ is an event for all $t \geq 0$ and $x \in \mathbb{R}$. But at the same time, for example, the set

$$\bigcap_{a \leq t < b} \{\omega \in \Omega : X_t(\omega) \leq x\}$$

is not necessarily an event (element of \mathcal{A}). This question is closely connected to the separability property of the processes (see, for example, [35, Chap. III]). The root essence is whether a countable and everywhere dense subset $\mathcal{S} \subset [0, \infty)$ exists such that the statistical behavior of the process X can be characterized by a countable set of the random variables X_t , $t \in \mathcal{S}$. The notion of separability is given in general by the following definition.

Definition 3.57. A process $X = (X_t, t \geq 0)$ is called **separable** if there exists an event $N \in \mathcal{A}$ with probability 0 and a countable subset $\mathcal{S} = \{r_i, i = 1, 2, \dots\}$ of $\mathbb{R}_+ = [0, \infty)$ that is always dense in \mathbb{R}_+ such that for any open set $G \subset \mathbb{R}_+$ and for any closed set $F \subset \mathcal{X}$ the sets $\{\omega : X_{r_i} \in F, r_i \in G\}$ and $\{\omega : X_t \in F, t \in G\}$ can differ only on the subset of N .

With the help of transition probabilities one can easily give a simple condition that ensures the continuity in probability of the process and, at the same time, the separability property.

Definition 3.58. A stochastic process $(X_t, t \geq 0)$ is called **continuous in probability** (or **stochastically**) at the point $t_0 \geq 0$ if for all positive numbers ε the convergence

$$\lim_{t \rightarrow t_0} \mathbf{P}(|X_t - X_{t_0}| > \varepsilon) = 0$$

holds. A process is said to be continuous in probability if it is continuous in probability everywhere.

Theorem 3.59. *If a Markov chain X is locally regular and condition (3.13) is satisfied, then it is continuous in probability.*

Proof. First we check that

$$\delta(h) = \sup_{k \in \mathcal{X}} (1 - p_{kk}(h)) \rightarrow 0, \quad h \rightarrow 0+. \quad (3.17)$$

Since by the relation in [36, p. 201]

$$\frac{1 - p_{kk}(h)}{h} \leq \lim_{h \rightarrow 0+} \frac{1 - p_{kk}(h)}{h} = q_k \leq q,$$

then

$$\sup_{k \in \mathcal{X}} (1 - p_{kk}(h)) \leq qh \rightarrow 0, \quad h \rightarrow 0+.$$

It is not difficult to see that for arbitrary $u, h \geq 0$ and $\varepsilon > 0$ we have

$$\begin{aligned} \mathbf{P}(|X_{u+h} - X_u| > \varepsilon) &\leq \mathbf{P}(|X_{u+h} - X_u| > 0) \\ &= \sum_{k \in \mathcal{X}} \mathbf{P}(|X_{u+h} - X_u| > 0 \mid X_u = k) \mathbf{P}(X_u = k) \\ &= \sum_{k \in \mathcal{X}} [1 - \mathbf{P}(|X_{u+h} - X_u| = 0 \mid X_u = k)] \mathbf{P}(X_u = k) \\ &= \sum_{k \in \mathcal{X}} (1 - p_{kk}(h)) \mathbf{P}(X_u = k) \leq \delta(h) \rightarrow 0, \quad h \rightarrow 0+, \end{aligned}$$

which means actually the continuity in probability of the chain X . □

Definition 3.60. The stochastic processes $(X_t, t \geq 0)$ and $(X'_t, t \geq 0)$, given on the same probability space, are said to be equivalent if

$$\mathbf{P}(X_t = X'_t) = 1, \quad t \geq 0.$$

The following theorem ensures that under the condition of continuity in probability, one can consider the separable version of the original process.

Theorem 3.61. *If a process $(X_t, t \geq 0)$ is continuous in probability, then there exists a continuous-in-probability separable version $(X'_t, t \geq 0)$ that is stochastically equivalent to $(X_t, t \geq 0)$.*

Theorem 3.62. *If a Markov chain satisfies condition (3.13), then there exists a separable and stochastically equivalent version of this Markov chain.*

Proof. From Theorem 3.59 it follows that the Markov chain is continuous in probability; therefore, as a consequence of Theorem 3.61, we have the assertion of the present theorem. \square

We assume later on that condition (3.13) is fulfilled because this condition with Theorem 3.62 guarantees that the Markov chain has a stochastically equivalent separable version. Assuming that condition (3.13) holds, one can bypass the measurability problems that can arise in the case of CTMCs, and the holding times are positive for all states.

Theorem 3.63. *If a homogeneous Markov chain X satisfies condition (3.13), then X has an equivalent stepwise version.*

Proof. Since from condition (3.13) follows Eq. (3.17), then by the use of the theorem of [36, p. 281], we obtain that there exists a stepwise version of the Markov chain that is equivalent to the original Markov chain. \square

3.3.5 Poisson Process as Continuous-Time Markov Chain

Theorem 3.64. *Let $(N_t, t \geq 0)$ be a homogeneous Poisson process with intensity rate λ , $N_0 = 0$. Then the process N_t is a homogeneous Markov chain.*

Proof. Choose arbitrarily a positive integer n , integers $0 \leq i_1 \leq \dots \leq i_{n+1}$, and real numbers $t_0 = 0 < t_1 < \dots < t_{n+1}$. It can be seen that

$$\begin{aligned} & \mathbf{P}(N_{t_{n+1}} = i_{n+1} \mid N_{t_n} = i_n, \dots, N_{t_1} = i_1) \\ &= \frac{\mathbf{P}(N_{t_{n+1}} = i_{n+1}, N_{t_n} = i_n, \dots, N_{t_1} = i_1)}{\mathbf{P}(N_{t_n} = i_n, \dots, N_{t_1} = i_1)} \\ &= \frac{\mathbf{P}(N_{t_{n+1}} - N_{t_n} = i_{n+1} - i_n, \dots, N_{t_2} - N_{t_1} = i_2 - i_1, N_{t_1} = i_1)}{\mathbf{P}(N_{t_n} - N_{t_{n-1}} = i_n - i_{n-1}, \dots, N_{t_2} - N_{t_1} = i_2 - i_1, N_{t_1} = i_1)}. \end{aligned}$$

Since the increments of the Poisson process are independent, the last fraction can be written in the form

$$\begin{aligned} & \frac{\mathbf{P}(N_{t_{n+1}} - N_{t_n} = i_{n+1} - i_n) \cdot \dots \cdot \mathbf{P}(N_{t_2} - N_{t_1} = i_2 - i_1) \mathbf{P}(N_{t_1} = i_1)}{\mathbf{P}(N_{t_n} - N_{t_{n-1}} = i_n - i_{n-1}) \cdot \dots \cdot \mathbf{P}(N_{t_2} - N_{t_1} = i_2 - i_1) \mathbf{P}(N_{t_1} = i_1)} \\ &= \mathbf{P}(N_{t_{n+1}} - N_{t_n} = i_{n+1} - i_n). \end{aligned}$$

From the independence of the increments $N_{t_{n+1}} - N_{t_n}$ and $N_{t_n} = N_{t_n} - N_0$ it follows that the events $\{N_{t_{n+1}} - N_{t_n} = i_{n+1} - i_n\}$ and $\{N_{t_n} = i_n\}$ are also independent, and thus

$$\begin{aligned} \mathbf{P}(N_{t_{n+1}} - N_{t_n} = i_{n+1} - i_n) &= \mathbf{P}(N_{t_{n+1}} - N_{t_n} = i_{n+1} - i_n | N_{t_n} = i_n) \\ &= \mathbf{P}(N_{t_{n+1}} = i_{n+1} | N_{t_n} = i_n), \end{aligned}$$

and finally we have

$$\mathbf{P}(N_{t_{n+1}} = i_{n+1} | N_{t_n} = i_n, \dots, N_{t_1} = i_1) = \mathbf{P}(N_{t_{n+1}} = i_{n+1} | N_{t_n} = i_n).$$

□

It is easy to determine the rate matrix of a homogeneous Poisson process with intensity λ . Clearly, the transition probability of the process is

$$p_{ij}(h) = \mathbf{P}(N_{t+h} = j | N_t = i) = \mathbf{P}(N_h = j - i) = \frac{(\lambda h)^{j-i}}{(j-i)!} e^{-\lambda h}, \quad j \geq i,$$

and

$$p_{ij}(h) \equiv 0, \quad j < i.$$

If $j < i$, then obviously $q_{ij} \equiv 0$. Let now $i \leq j$; then

$$q_{ij} = \lim_{h \rightarrow 0^+} \frac{p_{ij}(h)}{h} = \lim_{h \rightarrow 0^+} \frac{1}{h} \frac{(\lambda h)^{j-i}}{(j-i)!} e^{-\lambda h} = \begin{cases} \lambda, & \text{if } j = i + 1, \\ 0, & \text{if } j > i + 1. \end{cases}$$

Finally, let $i = j$. By the use of the L'Hospital rule

$$q_i = \lim_{t \rightarrow 0^+} \frac{1 - p_{ii}(h)}{h} = \lim_{t \rightarrow 0^+} \frac{1 - e^{-\lambda h}}{h} = \lambda.$$

Thus, summing up the obtained results, we have the rate matrix

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \cdot \\ 0 & -\lambda & \lambda & 0 & \cdot \\ 0 & 0 & -\lambda & \lambda & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}. \quad (3.18)$$

The Poisson process is regular because for all $i \in \mathcal{X}$

$$\sum_{j \neq i} q_{ij} = \lambda = q_i < \infty.$$

3.3.6 Reversible Markov Chains

Definition 3.65. A discrete-time Markov process is called **reversible** if for every state i, j the equation

$$\pi_i p_{ij} = \pi_j p_{ji}$$

holds, where π_i is the equilibrium probability of the states $i \in \mathcal{X}$.

The equation of the definition is usually called a **local** (or **detailed**) **balance condition** because of its similarity to the (global) balance Eq. (3.8) or, more precisely, to its form

$$\sum_{j \in \mathcal{X}} \pi_i p_{ij} = \sum_{j \in \mathcal{X}} \pi_j p_{ji}, \quad i \in \mathcal{X}.$$

The notation of reversibility of Markov chains originates from the fact that if the initial distribution of the chain equals the stationary one, then the forward and reverse conditional transition probabilities are identical, that is,

$$\mathbf{P}(X_n = i \mid X_{n+1} = j) = \mathbf{P}(X_{n+1} = i \mid X_n = j).$$

Indeed,

$$\begin{aligned} \mathbf{P}(X_n = i \mid X_{n+1} = j) &= \frac{\mathbf{P}(X_n = i, X_{n+1} = j)}{\mathbf{P}(X_{n+1} = j)} \\ &= \frac{\mathbf{P}(X_n = i) \mathbf{P}(X_{n+1} = j \mid X_n = i)}{\mathbf{P}(X_{n+1} = j)} \\ &= \frac{\pi_i p_{ij}}{\pi_j} = \frac{\pi_j p_{ji}}{\pi_j} = p_{ji} \\ &= \mathbf{P}(X_{n+1} = i \mid X_n = j). \end{aligned}$$

In the case of CTMCs, a definition can be applied analogously to the discrete-time case.

Definition 3.66. A CTMC is called **reversible** if for all pairs i, j of states the equation

$$\pi_i q_{ij} = \pi_j q_{ji}$$

holds, where π_i is the equilibrium probability of the state $i \in \mathcal{X}$.

The reversibility property and the local balance equations are often valid for Markov chains describing the processes in queueing networks (Sect. 10.1); in consequence the equilibrium probabilities can be computed in a simple, so-called **product form**.

3.4 Birth-Death Processes

Definition 3.67. The right-continuous stochastic process $\{v(t), t \geq 0\}$ is a *birth-death process* if

1. Its set of states is $I = \{0, 1, 2, \dots\}$ [that is, $v(t) \in I$];
2. The sojourn time in the state $k \in I, k > 0$, is exponentially distributed with the parameter

$$\alpha_k = a_k + b_k, \quad a_k, b_k \geq 0, k > 0,$$

and it is independent of the trajectory before arriving at the state k ;

3. After the state $k \in I, k \geq 1$, the process visits the state $k + 1$ with probability

$$p_k = \frac{a_k}{\alpha_k} \text{ and state } k - 1 \text{ with probability } q_k = 1 - p_k = \frac{b_k}{\alpha_k};$$

4. For the state 0 we consider the following two cases:

- The process stays an exponentially distributed amount of time in state 0 with parameter $\alpha_0 = a_0 > 0$ and after that visits state 1 (with probability $p_0 = 1$).
- Once the process arrives at state 0 it remains there forever ($q_0 = 1, p_0 = 0$).

$P_k(0) = \mathbf{P}(v(0) = k) = \varphi_k, k \in I$, denotes the initial distribution of the process.

If $\{v(t), t \geq 0\}$ is a birth-death process, then it is an infinite-state continuous-time (time-homogeneous) Markov chain. The parameters a_k and b_k are referred to as the *birth rate* and the *death rate* in the state k , respectively, and k is referred to as the *population*. The special case where $b_k \equiv 0$ is referred to as the *birth process* and where $a_k \equiv 0$ as the *death process*.

Let $T_0 = 0 < T_1 < T_2 < \dots$ denote the time instants of the population changes (birth and death). The discrete-time $\{v_n, n \geq 0\}$ process, where $v_n = v(T_n)$ is the population after the n th change in population [n th jump of $v(t)$], is referred to as the Markov chain embedded in the population changes of $\{v(t), t \geq 0\}$. The state-transition probability matrix of the embedded Markov chain is

$$\begin{bmatrix} q_0 & p_0 & 0 & 0 & 0 & \dots \\ q_1 & 0 & p_1 & 0 & 0 & \dots \\ 0 & q_2 & 0 & p_2 & 0 & \dots \\ 0 & 0 & q_3 & 0 & p_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

3.4.1 Some Properties of Birth-Death Processes

The transient state probability, its Laplace transform, and the initial probabilities for $k \geq 0, t \geq 0$, and $\text{Re } s > 0$ are denoted by

$$P_k(t) = \mathbf{P}(v(t) = k), \quad p_k^*(s) = \int_0^{\infty} e^{-st} P_k(t) dt, \quad P_k(0) = \mathbf{P}(v(0) = k) = \varphi_k.$$

In special cases, the following theorems are true. ([69])

Theorem 3.68. *If $p_0 = 1$, $0 < p_k < 1$, $k \geq 1$, then the following statements hold:*

1. $P_k(t)$ satisfies the following ordinary differential equations:

$$P_0'(t) = -a_0 P_0(t) + b_1 P_1(t),$$

$$P_k'(t) = a_{k-1} P_{k-1}(t) - (a_k + b_k) P_k(t) + b_{k+1} P_{k+1}(t), \quad k \geq 1.$$

2. For φ_k , $k \geq 0$, and $\operatorname{Re} s > 0$ the following linear system defines $p_k^*(s)$:

$$s p_0^*(s) - \varphi_0 = -a_0 p_0^*(s) + b_1 p_1^*(s),$$

$$s p_k^*(s) - \varphi_k = a_{k-1} p_{k-1}^*(s) - (a_k + b_k) p_k^*(s) + b_{k+1} p_{k+1}^*(s), \quad k \geq 1.$$

3. For $k \geq 0$ the limits

$$\lim_{t \rightarrow \infty} P_k(t) = \pi_k$$

exist and are independent of the initial distribution of the process.

$$\pi_k = 0, \quad k \geq 0,$$

if

$$\sum_{k=0}^{\infty} \rho_k < \infty, \tag{3.19}$$

where $\rho_0 = 1$ and $\rho_k = \frac{a_0 a_1 \cdots a_{k-1}}{b_1 b_2 \cdots b_k}$, $k \geq 1$. Otherwise, $\pi_k > 0$, $k \geq 0$, and

$$\pi_0 = \left(\sum_{j=0}^{\infty} \rho_j \right)^{-1}, \tag{3.20}$$

$$\pi_k = \rho_k \pi_0. \tag{3.21}$$

Theorem 3.69 (Finite birth-death process). *Let the state space of $v(t)$ be $\{0, 1, 2, \dots, n\}$, $p_0 = 1$, $0 < p_k < 1$, for $1 \leq k \leq n-1$ and $p_n = 0$; then the following statements hold:*

1. $P_k(t)$ satisfies the following ordinary differential equations:

$$P_0'(t) = -a_0 P_0(t) + b_1 P_1(t),$$

$$P_k'(t) = a_{k-1} P_{k-1}(t) - (a_k + b_k) P_k(t) + b_{k+1} P_{k+1}(t), \quad 1 \leq k \leq n-1,$$

$$P_n'(t) = a_{n-1} P_{n-1}(t) - b_n P_n(t).$$

2. If the initial distribution of the process is $\varphi_k = \mathbf{P}(v(0) = k)$, $0 \leq k \leq n$, then for $\operatorname{Re} s > 0$ the Laplace transforms of the transient state probabilities $p_k^*(s)$ satisfy

$$\begin{aligned} sp_0^*(s) - \varphi_0 &= -a_0 p_0^*(s) + b_1 p_1^*(s), \\ sp_k^*(s) - \varphi_k &= a_{k-1} p_{k-1}^*(s) - (a_k + b_k) p_k^*(s) + b_{k+1} p_{k+1}^*(s), \quad 1 \leq k \leq n-1, \\ sp_n^*(s) - \varphi_n &= a_{n-1} p_{n-1}^*(s) - b_n p_n^*(s). \end{aligned}$$

3. For $0 \leq k \leq n$ the

$$\lim_{t \rightarrow \infty} P_k(t) = \pi_k > 0$$

limit exists and is independent of the initial distribution:

$$\pi_j = \rho_j \pi_0, \quad \pi_0 = \left(\sum_{j=0}^{\infty} \rho_j \right)^{-1},$$

where

$$\rho_0 = 1, \quad \rho_j = \frac{a_0 a_1 \cdots a_{j-1}}{b_1 b_2 \cdots b_j}, \quad 1 \leq j \leq n.$$

Theorem 3.70. The following equations hold.

1. Let $p_0 = 0$, $0 < p_k < 1$, $k \geq 1$; then for $P_k(t)$ we have

$$\begin{aligned} P_0'(t) &= b_1 P_1(t), \\ P_1'(t) &= -(a_1 + b_1) P_1(t) + b_2 P_2(t), \\ P_k'(t) &= a_{k-1} P_{k-1}(t) - (a_k + b_k) P_k(t) + b_{k+1} P_{k+1}(t), \quad k \geq 2, \end{aligned}$$

and for $\operatorname{Re} s > 0$ and the initial distribution φ_k , $k \geq 0$, we have

$$\begin{aligned} sp_0^*(s) - \varphi_0 &= b_1 p_1^*(s), \\ sp_1^*(s) - \varphi_1 &= -(a_1 + b_1) p_1^*(s) + b_2 p_2^*(s), \\ sp_k^*(s) - \varphi_k &= a_{k-1} p_{k-1}^*(s) - (a_k + b_k) p_k^*(s) + b_{k+1} p_{k+1}^*(s), \quad k \geq 2. \end{aligned}$$

2. Let $v(t) \in \{0, 1, 2, \dots, n\}$, $p_0 = 0$, $0 < p_k < 1$ if $1 \leq k \leq n-1$, and $p_n = 0$; then for $P_k(t)$ we have

$$\begin{aligned} P_0'(t) &= b_1 P_1(t), \\ P_1'(t) &= -(a_1 + b_1) P_1(t) + b_2 P_2(t), \end{aligned}$$

$$P'_k(t) = a_{k-1}P_{k-1}(t) - (a_k + b_k)P_k(t) + b_{k+1}P_{k+1}(t), \quad 2 \leq k \leq n-1,$$

$$P'_n(t) = a_{n-1}P_{n-1}(t) - b_nP_n(t),$$

and for $p_k^*(s)$, $\text{Re } s > 0$, we have $[\varphi_k = \mathbf{P}(v(0) = k), 0 \leq k \leq n]$

$$sp_0^*(s) - \varphi_0 = b_1p_1^*(s),$$

$$sp_1^*(s) - \varphi_1 = -(a_1 + b_1)p_1^*(s) + b_2p_2^*(s),$$

$$sp_k^*(s) - \varphi_k = a_{k-1}p_{k-1}^*(s) - (a_k + b_k)p_k^*(s) + b_{k+1}p_{k+1}^*(s), \quad 2 \leq k \leq n-1,$$

$$sp_n^*(s) - \varphi_n = a_{n-1}p_{n-1}^*(s) - b_np_n^*(s).$$

Comment 3.71. In Theorems 3.68–3.70 the differential equations for $P_j(t)$ are indeed the Kolmogorov (forward) differential equations for the given systems. The equations for $p_j^*(s)$ can be obtained from the related differential equations for $P_j(t)$ using

$$\int_0^{\infty} e^{-st} P'_j(t) dt = sp_j^*(s) - P'_j(0).$$

In Theorem 3.70 state 0 is an absorbing state. In this way, the theorem allows one to compute the parameters of the busy period of birth-death Markov chains starting from state k ($\varphi_k = 1$), where the busy period is the time to reach state 0 (which commonly represents the idle state of a system, where the server is not working, in contrast to the $i > 0$ states, where the server is commonly busy). Let Π_k denote the length of the busy period starting from state k ; then

$$\Pi_k(t) = \mathbf{P}(\Pi_k \leq t) = \mathbf{P}(v(t) = 0) = P_0(t)$$

defines the distribution of the length of the busy period, and from Theorem 3.70.1 we have

$$\Pi'_k(t) = P'_0(t) = b_1P_1(t),$$

from which the Laplace–Stieltjes transform of the distribution of $\Pi_k(t)$, $\pi_k(s)$, is

$$\begin{aligned} \pi_k(s) &= \int_0^{\infty} e^{-st} d\Pi_k(t) = \int_0^{\infty} e^{-st} \Pi'_k(t) dt \\ &= \int_0^{\infty} e^{-st} b_1P_1(t) dt = b_1p_1^*(s). \end{aligned}$$

If the arrival intensity is constant in all states, i.e., $a_k = \lambda > 0$ ($\forall k \geq 0$), then the arrival process is a Poisson process at rate λ . Further results on the properties of special birth-death processes can be obtained, e.g., in [48].

3.5 Exercises

Exercise 3.1. Compute the probability that a CTMC with the generator matrix $\begin{pmatrix} -1 & 0.5 & 0.5 \\ 1 & -2 & 1 \\ 1 & 0 & -1 \end{pmatrix}$ stays in state 1 after the second state transition if the initial distribution is $(0.5, 0.5, 0)$.

Exercise 3.2. Compute the stationary distribution of a CTMC with the generator matrix $\begin{pmatrix} -3 & 3 & 0 \\ 4 & -4 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ if the initial distribution is $(0.5, 0, 0.5)$.

Exercise 3.3. Z_n and $Y_n, n = 1, 2, \dots$, are discrete independent random variables. $\mathbf{P}(Z_n = 0) = 1 - p$, $\mathbf{P}(Z_n = 1) = p$ and $\mathbf{P}(Y_n = 0) = 1 - q$, $\mathbf{P}(Y_n = 1) = q$. Define the transition probability matrix of the DTMC X_n if

$$X_{n+1} = (X_n - Y_n)^+ + Z_n,$$

where $(x)^+ = \max(x, 0)$. This equation is commonly referred to as the evolution equation of a DTMC.

Exercise 3.4. $X_n, n = 1, 2, \dots$, is a DTMC with the transition probability matrix $P = \begin{pmatrix} 3/6 & 1/6 & 2/6 \\ 3/4 & 0 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix}$. Compute $\mathbf{E}(X_0 X_1)$ and $\text{corr}(X_0, X_1)$ if the initial distribution is $(0.5, 0, 0.5)$ and the state space is $S = \{0, 1, 2\}$.

Exercise 3.5. The generator of a CTMC is defined by

$$q_{0j} = \begin{cases} \frac{1}{3} & \text{if } j = 1, \\ \frac{1}{3} & \text{if } j = 2, \\ -\frac{2}{3} & \text{if } j = 0, \\ 0 & \text{otherwise;} \end{cases} \quad q_{ij} = \begin{cases} \frac{1}{3^i} & \text{if } j = i + 1, \\ \frac{1}{3^i} & \text{if } j = i + 2, \\ -\frac{2}{3^i} - \mu & \text{if } j = i, \\ \mu & \text{if } j = i - 1, \\ 0 & \text{otherwise,} \end{cases} \quad \text{for } i = 1, 2, \dots$$

Evaluate the properties of this Markov chain using, e.g., the Foster theorem.

Exercise 3.6. Show examples of

- Reducible
- Periodic (and irreducible)
- Transient (and irreducible)

DTMCs. Evaluate $\lim_{n \rightarrow \infty} \mathbf{P}(X_n = i)$ for these DTMCs, where i is a state of the Markov chain.

Exercise 3.7. Two players, A and B , play with dice according to the following rule. They throw the dice, and if the number is 1, then A gets £2 from B ; if the number is 2 or 3, then A gets £1 from B ; and if the number is greater than 3, then B gets £1 from A . At the beginning of the game both A and B have £3. The game lasts until one of the players can no longer pay. What is the probability that A wins?

Exercise 3.8. Two players, A and B , play with dice according to the following rule. They throw the dice, and if the number is 1, then A gets £2 from B ; if the number is 2 or 3, then A gets £1 from B ; and if the number is greater than 3, then B gets £1 from A . At the beginning of the game both A and B have £3. If one of them cannot pay the required amount, then he must give all his money to the other player and the game goes on. What is the expected amount of money A will have after a very long run? What is the probability that B will not be able to pay the required amount in the next step of the game after a very long run?

Exercise 3.9. There are two machines, A and B , at a production site. Their failure times are exponentially distributed with the parameters λ_A and λ_B , respectively. Their repair times are also exponentially distributed with the parameters μ_A and μ_B , respectively. A single repairman is associated with the two machines; he can work on only one machine at a time. Compute the probability that at least one of the machines works.

Exercise 3.10. Let $X = (X_0, X_1, \dots)$ be a two-state Markov chain with the state space $\mathcal{X} = \{0, 1\}$ and with the probability transition matrix $P = \begin{bmatrix} a & 1-a \\ 1-b & b \end{bmatrix}$, where $0 < a, b < 1$. Prove that $P^n = \frac{1}{2-a-b}\Pi + \frac{(a+b-1)^n}{2-a-b}(I - P)$, where $\Pi = \begin{bmatrix} 1-b & 1-a \\ 1-b & 1-a \end{bmatrix}$ and $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.