

# Chapter 21

## Markov Processes

**Abstract** This chapter presents the fundamentals of the theory of general Markov processes in continuous time. Section 21.1 contains the definitions and a discussion of the Markov property and transition functions, and derives the Chapman–Kolmogorov equation. Section 21.2 studies Markov processes in countable state spaces, deriving systems of backward and forward differential equations for transition probabilities. It also establishes the ergodic theorem and contains examples illustrating the presented theory. Section 21.3 deals with continuous time branching processes. Then the elements of the general theory of semi-Markov processes are presented in Sect. 21.4, including the ergodic theorem and some other related results for such processes. Section 21.5 discusses the so-called regenerative processes, establishing their ergodicity and the Laws of Large Numbers and Central Limit Theorem for integrals of functions of their trajectories. Section 21.6 is devoted to diffusion processes. It begins with the classical definition of diffusion, derives the forward and backward Kolmogorov equations for the transition probability function of a diffusion process, and gives a couple of examples of using the equations to compute important characteristics of the respective processes.

### 21.1 Definitions and General Properties

Markov processes in discrete time (Markov chains) were considered in Chap. 13. Recall that their main property was independence of the “future” of the process of its “past” given its “present” is fixed. The same principle underlies the definition of Markov processes in the general case.

#### 21.1.1 Definition and Basic Properties

Let  $(\Omega, \mathfrak{F}, \mathbf{P})$  be a probability space and  $\{\xi(t) = \xi(t, \omega), t \geq 0\}$  a random process given on it. Set

$$\mathfrak{F}_1 := \sigma(\xi(u); u \leq t), \quad \mathfrak{F}_{[t, \infty)} := \sigma(\xi(u); u \geq t),$$

so that the variable  $\xi(u)$  is  $\mathfrak{F}_t$ -measurable for  $u \leq t$  and  $\mathfrak{F}_{[t,\infty)}$ -measurable for  $u \geq t$ . The  $\sigma$ -algebra  $\sigma(\mathfrak{F}_t, \mathfrak{F}_{[t,\infty)})$  is generated by the variables  $\xi(u)$  for all  $u$  and may coincide with  $\mathfrak{F}$  in the case of the sample probability space.

**Definition 21.1.1** We say that  $\xi(t)$  is a *Markov process* if, for any  $t$ ,  $A \in \mathfrak{F}_t$ , and  $B \in \mathfrak{F}_{[t,\infty)}$ , we have

$$\mathbf{P}(AB|\xi(t)) = \mathbf{P}(A|\xi(t))\mathbf{P}(B|\xi(t)). \quad (21.1.1)$$

This expresses precisely the fact that the future is independent of the past when the present is fixed (conditional independence of  $\mathfrak{F}_t$  and  $\mathfrak{F}_{[t,\infty)}$  given  $\xi(t)$ ).

We will now show that the above definition is equivalent to the following.

**Definition 21.1.2** We say that  $\xi(t)$  is a *Markov process* if, for any bounded  $\mathfrak{F}_{[t,\infty)}$ -measurable random variable  $\eta$ ,

$$\mathbf{E}(\eta|\mathfrak{F}_t) = \mathbf{E}(\eta|\xi(t)). \quad (21.1.2)$$

It suffices to take  $\eta$  to be functions of the form  $\eta = f(\xi(s))$  for  $s \geq t$ .

*Proof of the equivalence* Let (21.1.1) hold. By the monotone convergence theorem it suffices to prove (21.1.2) for simple functions  $\eta$ . To this end it suffices, in turn, to prove (21.1.2) for  $\eta = I_B$ , the indicator of the set  $B \in \mathfrak{F}_{[t,\infty)}$ . Let  $A \in \mathfrak{F}_t$ . Then, by (21.1.1),

$$\begin{aligned} \mathbf{P}(AB) &= \mathbf{E}\mathbf{P}(AB|\xi(t)) = \mathbf{E}[\mathbf{P}(A|\xi(t))\mathbf{P}(B|\xi(t))] \\ &= \mathbf{E}\mathbf{E}[I_A\mathbf{P}(B|\xi(t))|\xi(t)] = \mathbf{E}[I_A\mathbf{P}(B|\xi(t))]. \end{aligned} \quad (21.1.3)$$

On the other hand,

$$\mathbf{P}(AB) = \mathbf{E}[I_A I_B] = \mathbf{E}[I_A \mathbf{P}(B|\mathfrak{F}_t)]. \quad (21.1.4)$$

Because (21.1.3) and (21.1.4) hold for any  $A \in \mathfrak{F}_t$ , this means that  $\mathbf{P}(B|\mathfrak{F}_t) = \mathbf{P}(B|\xi(t))$ .

Conversely, let (21.1.2) hold. Then, for  $A \in \mathfrak{F}_t$  and  $B \in \mathfrak{F}_{[t,\infty)}$ , we have

$$\begin{aligned} \mathbf{P}(AB|\xi(t)) &= \mathbf{E}[\mathbf{E}(I_A I_B|\mathfrak{F}_t)|\xi(t)] = \mathbf{E}[I_A \mathbf{E}(I_B|\mathfrak{F}_t)|\xi(t)] \\ &= \mathbf{E}[I_A \mathbf{E}(I_B|\xi(t))|\xi(t)] = \mathbf{P}(B|\xi(t))\mathbf{P}(A|\xi(t)). \end{aligned} \quad \square$$

It remains to verify that it suffices to take  $\eta = f(\xi(s))$ ,  $s \geq t$ , in (21.1.2). In order to do this, we need one more equivalent definition of a Markov process.

**Definition 21.1.3** We say that  $\xi(t)$  is a *Markov process* if, for any bounded function  $f$  and any  $t_1 < t_2 < \dots < t_n \leq t$ ,

$$\mathbf{E}(f(\xi(t))|\xi(t_1), \dots, \xi(t_n)) = \mathbf{E}(f(\xi(t))|\xi(t_n)). \quad (21.1.5)$$

*Proof of the equivalence* Relation (21.1.5) follows in an obvious way from (21.1.2). Now assume that (21.1.5) holds. Then, for any  $A \in \sigma(\xi(t_1), \dots, \xi(t_n))$ ,

$$\mathbf{E}(f(\xi(t)); A) = \mathbf{E}[\mathbf{E}(f(\xi(t)) | \xi(t_n)); A]. \quad (21.1.6)$$

Both parts of (21.1.6) are measures coinciding on the algebra of cylinder sets. Therefore, by the theorem on uniqueness of extension of a measure, they coincide on the  $\sigma$ -algebra generated by these sets, i.e. on  $\mathfrak{F}_{t_n}$ . In other words, (21.1.6) holds for any  $A \in \mathfrak{F}_{t_n}$ , which is equivalent to the equality

$$\mathbf{E}[f(\xi(t)) | \mathfrak{F}_{t_n}] = \mathbf{E}[f(\xi(t)) | \xi(t_n)]$$

for any  $t_n \leq t$ . Relation (21.1.2) for  $\eta = f(\xi(t))$  is proved.  $\square$

We now prove that in (21.1.2) it suffices to take  $\eta = f(\xi(s))$ ,  $s \geq t$ . Let  $t \leq u_1 < \dots < u_n$ . We prove that then (21.1.2) is true for

$$\eta = \prod_{i=1}^n f_i(\xi(u_i)). \quad (21.1.7)$$

We will make use of induction and assume that equality (21.1.2) holds for the functions

$$\gamma = \prod_{i=1}^{n-1} f_i(\xi(u_i))$$

(for  $n = 1$  relation (21.1.2) is true). Then, putting  $g(u_{n-1}) := \mathbf{E}[f_n(\xi(u_n)) | \xi(u_{n-1})]$ , we obtain

$$\begin{aligned} \mathbf{E}(\eta | \mathfrak{F}_t) &= \mathbf{E}[\mathbf{E}(\eta | \mathfrak{F}_{u_{n-1}}) | \mathfrak{F}_t] = \mathbf{E}[\gamma \mathbf{E}(f_n(\xi(u_n)) | \mathfrak{F}_{u_{n-1}}) | \mathfrak{F}_t] \\ &= \mathbf{E}[\gamma \mathbf{E}(f_n(\xi(u_n)) | \xi(u_{n-1})) | \mathfrak{F}_t] = \mathbf{E}[\gamma g(\xi(u_{n-1})) | \mathfrak{F}_t]. \end{aligned}$$

By the induction hypothesis this implies that

$$\mathbf{E}(\eta | \mathfrak{F}_t) = \mathbf{E}[\gamma g(\xi(u_{n-1})) | \xi(t)]$$

and, therefore, that  $\mathbf{E}(\eta | \mathfrak{F}_t)$  is  $\sigma(\xi(t))$ -measurable and

$$\mathbf{E}(\eta | \xi(t)) = \mathbf{E}(\mathbf{E}(\eta | \mathfrak{F}_t) | \xi(t)) = \mathbf{E}(\eta | \mathfrak{F}_t).$$

We proved that (21.1.2) holds for  $\sigma(\xi(u_1), \dots, \xi(u_n))$ -measurable functions of the form (21.1.7). By passing to the limit we establish first that (21.1.2) holds for simple functions, and then that it holds for any  $\mathfrak{F}_{[t, \infty)}$ -measurable functions.  $\square$

### 21.1.2 Transition Probability

We saw that, for a Markov process  $\xi(t)$ , the conditional probability

$$\mathbf{P}(\xi(t) \in B | \mathfrak{F}_s) = \mathbf{P}(\xi(t) \in B | \xi(s)) \quad \text{for } t > s$$

is a Borel function of  $\xi(s)$  which we will denote by

$$P(s, \xi(s); t, B) := \mathbf{P}(\xi(t) \in B | \xi(s)).$$

One can say that  $P(s, x; t, B)$  as a function of  $B$  and  $x$  is the conditional distribution (see Sect. 4.9) of  $\xi(t)$  given that  $\xi(s) = x$ . By the Markov property, it satisfies the relation ( $s < u < t$ )

$$P(s, x; t, B) = \int P(s, x; u, dy)P(u, y; t, B), \tag{21.1.8}$$

which follows from the equality

$$\begin{aligned} & \mathbf{P}(\xi(t) \in B | \xi(s) = x) \\ &= \mathbf{E}[\mathbf{P}(\xi(t) \in B | \mathfrak{F}_u) | \xi(s) = x] = \mathbf{E}[P(u, \xi(u); t, B) | \xi(s) = x]. \end{aligned}$$

Equation (21.1.8) is called the *Chapman–Kolmogorov equation*.

The function  $P(s, x; t, B)$  can be used in an analytic definition of a Markov process. First we need to clarify what properties a function  $P_{x,B}(s, t)$  should possess in order that there exists a Markov process  $\xi(t)$  for which

$$P_{x,B}(s, t) = P(s, x; t, B).$$

Let  $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$  be a measurable space.

**Definition 21.1.4** A function  $P_{x,B}(s, t)$  is said to be a *transition function on*  $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$  if it satisfies the following conditions:

- (1) As a function of  $B$ ,  $P_{x,B}(s, t)$  is a probability distribution for each  $s \leq t, x \in \mathcal{X}$ .
- (2)  $P_{x,B}(s, t)$  is measurable in  $x$  for each  $s \leq t$  and  $B \in \mathfrak{B}_{\mathcal{X}}$ .
- (3) For  $0 \leq s < u < t$  and all  $x$  and  $B$ ,

$$P_{x,B}(s, t) = \int P_{x,dy}(s, u)P_{y,B}(u, t)$$

(the Chapman–Kolmogorov equation).

- (4)  $P_{x,B}(s, t) = \mathbf{I}_B(x)$  for  $s = t$ .

Here properties (1) and (2) ensure that  $P_{x,B}(s, t)$  can be a conditional distribution (cf. Sect. 4.9).

Now define, with the help of  $P_{x,B}(s, t)$ , the finite-dimensional distributions of a process  $\xi(t)$  with the initial condition  $\xi(0) = a$  by the formula

$$\begin{aligned} & \mathbf{P}(\xi(t) \in dy_1, \dots, \xi(t_n) \in dy_n) \\ &= P_{a,dy_1}(0, t_1)P_{y_1,dy_2}(t_1, t_2) \cdots P_{y_{n-1},dy_n}(t_{n-1}, t_n). \end{aligned} \tag{21.1.9}$$

By virtue of properties (3) and (4), these distributions are consistent and therefore by the Kolmogorov theorem define a process  $\xi(t)$  in  $(\mathbb{R}^T, \mathfrak{B}_{\mathbb{R}}^T)$ , where  $T = [0, \infty)$ .

By formula (21.1.9) and rule (21.1.5),

$$\begin{aligned} \mathbf{P}(\xi(t_n) \in B_n | (\xi(t_1), \dots, \xi(t_{n-1})) = (y_1, \dots, y_{n-1})) \\ = P_{y_{n-1}, B_n}(t_{n-1}, t_n) = \mathbf{P}(\xi(t_n) \in B_n | \xi(t_{n-1}) = y_{n-1}) \\ = P(t_{n-1}, y_{n-1}; t_n, B_n). \end{aligned}$$

We could also verify this equality in a more formal way using the fact that the integrals of both sides over the set  $\{\xi(t_1) \in B_1, \dots, \xi(t_{n-1}) \in B_{n-1}\}$  coincide.

Thus, by virtue of Definition 21.1.3, we have constructed a Markov process  $\xi(t)$  for which

$$P(s, x; t, B) = P_{x, B}(s, t).$$

This function will also be called the *transition function* (or *transition probability*) of the process  $\xi(t)$ .

**Definition 21.1.5** A Markov process  $\xi(t)$  is said to be *homogeneous* if  $P(s, x; t, B)$ , as a function of  $s$  and  $t$ , depends on the difference  $t - s$  only:

$$P(s, x; t, B) = P(t - s; x, B).$$

This is the probability of transition during a time interval of length  $t - s$  from  $x$  to  $B$ . If

$$P(u; t, B) = \int_B p(u; t, y) dy$$

then the function  $p(u; x, y)$  is said to be a *transition density*.

It is not hard to see that the Wiener and Poisson processes are both homogeneous Markov processes. For example, for the Wiener process,

$$P(u; x, y) = \frac{1}{\sqrt{2\pi u}} e^{-(x-y)^2/2u}.$$

## 21.2 Markov Processes with Countable State Spaces. Examples

### 21.2.1 Basic Properties of the Process

Assume without loss of generality that the “discrete state space”  $\mathcal{X}$  coincides with the set of integers  $\{0, 1, 2, \dots\}$ . For simplicity’s sake we will only consider homogeneous Markov processes.

The transition function of such a process is determined by the collection of functions  $P(t; i, j) = p_{ij}(t)$  which form a stochastic matrix  $P(t) = \|p_{ij}(t)\|$  (with  $p_{ij}(t) \geq 0$ ,  $\sum_j p_{ij}(t) = 1$ ). Chapman–Kolmogorov’s equation now takes the form

$$p_{ij}(t + s) = \sum_k p_{ik}(t) p_{kj}(s),$$

or, which is the same, in the matrix form,

$$P(t + s) = P(t)P(s) = P(s)P(t). \tag{21.2.1}$$

In what follows, we consider only *stochastically continuous processes* for which  $\xi(t + s) \xrightarrow{P} \xi(t)$  as  $s \rightarrow 0$ , which is equivalent in the case under consideration to each of the following three relations:

$$\mathbf{P}(\xi(t + s) \neq \xi(t)) \rightarrow 0, \quad P(t + s) \rightarrow P(t), \quad P(s) \rightarrow P(0) \equiv E \tag{21.2.2}$$

as  $s \rightarrow 0$  (component-wise;  $E$  is the unit matrix).

We will also assume that convergence in (21.2.2) is *uniform* (for a finite  $\mathcal{X}$  this is always the case).

According to the separability requirement, we will assume that  $\xi(t)$  cannot change its state in “zero time” more than once (thus excluding the effects illustrated in Example 18.1.1, i.e. assuming that if  $\xi(t) = j$  then, with probability 1,  $\xi(t + s) = j$  for  $s \in [0, \tau)$ ,  $\tau = \tau(\omega) > 0$ ). In that case, the trajectories of the processes will be piece-wise constant (right-continuous for definiteness), i.e. the time axis is divided into half-intervals  $[0, \tau_1)$ ,  $[\tau_1, \tau_1 + \tau_2)$ ,  $\dots$ , on which  $\xi(t)$  is constant. Put

$$q_j(t) := \mathbf{P}(\xi(u) = j, 0 \leq u < t | \xi(0) = j) = \mathbf{P}(\tau_1 \geq t).$$

**Theorem 21.2.1** *Under the above assumptions (stochastic continuity and separability),*

$$q_i(t) = e^{-q_i t},$$

where  $q_i < \infty$ ; moreover,  $q_i > 0$  if  $p_{ii}(t) \neq 1$ . There exist the limits

$$\lim_{t \rightarrow 0} \frac{1 - p_{ii}(t)}{t} = q_i, \quad \lim_{t \rightarrow 0} \frac{p_{ij}(t)}{t} = q_{ij}, \quad i \neq j, \tag{21.2.3}$$

where  $\sum_{j:j \neq i} q_{ij} = q_i$ .

*Proof* By the Markov property,

$$q_i(t + s) = q_i(t)q_i(s),$$

and  $q_i(t) \downarrow$ . Therefore there exists a unique solution  $q_i(t) = e^{-q_i t}$  of this equation, where  $q_i < \infty$ , since  $\mathbf{P}(\tau_1 > 0) = 1$  and  $q_i > 0$ , because  $q_i(t) < 1$  when  $p_{ii}(t) \neq 1$ .

Let further  $0 < t_0 < t_1 \dots < t_n < t$ . Since the events

$$\{\xi(u) = i \text{ for } u \leq t_r, \xi(t_{r+1}) = j\}, \quad r = 0, \dots, n - 1; \quad j \neq i,$$

are disjoint,

$$p_{ii}(t) = q_i(t) + \sum_{r=0}^{n-1} \sum_{j:j \neq i} q_i(t_r) p_{ij}(t_{r+1} - t_r) p_{ji}(t - t_{r+1}). \tag{21.2.4}$$

Here, by condition (21.2.2),  $p_{ji}(t - t_{r+1}) < \varepsilon_t$  for all  $j \neq i$ , and  $\varepsilon_t \rightarrow 0$  as  $t \rightarrow 0$ , so that the sum in (21.2.4) does not exceed

$$\varepsilon_t \sum_{r=0}^{n-1} \sum_{j:j \neq i} q_i(t_r) p_{ij}(t_{r+1} - t_r) = \varepsilon_t \mathbf{P} \left( \bigcup_{r=1}^n \{ \xi(t_r) \neq i \} \mid \xi(0) = i \right) < \varepsilon_t (1 - q_i(t)),$$

$$p_{ii}(t) \leq q_i(t) + \varepsilon_t (1 - q_i(t)).$$

Together with the obvious inequality  $p_{ii}(t) \geq q_i(t)$  this gives

$$1 - q_i(t) \geq 1 - p_{ii}(t) \geq (1 - q_i(t))(1 + \varepsilon_t)$$

(i.e. the asymptotic behaviour of  $1 - q_i(t)$  and  $1 - p_{ii}(t)$  as  $t \rightarrow \infty$  is identical). This implies the second assertion of the theorem (i.e., the first relation in (21.2.3)).

Now let  $t_r := rt/n$ . Consider the transition probabilities

$$p_{ij}(t) \geq \sum_{r=0}^{n-1} q_i(t_r) p_{ij}(t/n) q_j(t - t_{r+1})$$

$$\geq (1 - \varepsilon_t) p_{ij}(t/n) \sum_{r=0}^{n-1} e^{-q_i r t/n} \geq (1 - \varepsilon_t) p_{ij}(t/n) \frac{(1 - e^{-q_i t})n}{q_i t}.$$

This implies that

$$p_{ij}(t) \geq (1 - \varepsilon_t) \left( \frac{1 - e^{-q_i t}}{q_i} \right) \limsup_{\delta \rightarrow 0} \frac{p_{ij}(\delta)}{\delta},$$

and that the upper limit on the right-hand side is bounded. Passing to the limit as  $t \rightarrow 0$ , we obtain

$$\liminf_{t \rightarrow 0} \frac{p_{ij}(t)}{t} \geq \limsup_{\delta \rightarrow 0} \frac{p_{ij}(\delta)}{\delta}.$$

Since  $\sum_{j:j \neq i} p_{ij}(t) = 1 - p_{ii}(t)$ , we have  $\sum_{j:j \neq i} q_{ij} = q_i$ . The theorem is proved. □

The theorem shows that the quantities

$$p_{ij} = \frac{q_{ij}}{q_i}, \quad j \neq i, \quad p_{ii} = 0$$

form a stochastic matrix and give the probabilities of transition from  $i$  to  $j$  during an infinitesimal time interval  $\Delta$  given the process  $\xi(t)$  left the state  $i$  during that time interval:

$$\mathbf{P}(\xi(t + \Delta) = j \mid \xi(t) = i, \xi(t + \Delta) \neq i) = \frac{p_{ij}(\Delta)}{1 - p_{ii}(\Delta)} \rightarrow \frac{q_{ij}}{q_i}$$

as  $\Delta \rightarrow 0$ .

Thus the evolution of  $\xi(t)$  can be thought of as follows. If  $\xi(0) = X_0$ , then  $\xi(t)$  stays at  $X_0$  for a random time  $\tau_1 \in \Gamma_{q_{X_0}}$ . Then  $\xi(t)$  passes to a state  $X_1$  with probability  $p_{X_0 X_1}$ . Further,  $\xi(t) = X_1$  over the time interval  $[\tau_1, \tau_1 + \tau_2)$ ,  $\tau_2 \in \Gamma_{q_{X_1}}$ , after which the system changes its state to  $X_2$  and so on. It is clear that  $X_0, X_1, \dots$  is a homogeneous Markov chain with the transition matrix  $\|p_{ij}\|$ . Therefore the

further study of  $\xi(t)$  can be reduced in many aspects to that of the Markov chain  $\{X_n; n \geq 0\}$ , which was carried out in detail in Chap. 13.

We see that the evolution of  $\xi(t)$  is completely specified by the quantities  $q_{ij}$  and  $q_i$  forming the matrix

$$Q = \|q_{ij}\| = \lim_{t \rightarrow 0} \frac{P(t) - P(0)}{t}, \quad (21.2.5)$$

where we put  $q_{ii} := -q_i$ , so that  $\sum_j q_{ij} = 0$ . We can also justify this claim using an analytical approach. To simplify the technical side of the exposition, we will assume, where it is needed, that the entries of the matrix  $Q$  are bounded and convergence in (21.2.3) is uniform in  $i$ .

Denote by  $e^A$  the matrix-valued function

$$e^A = E + \sum_{k=1}^{\infty} \frac{1}{k!} A^k.$$

**Theorem 21.2.2** *The transition probabilities  $p_{ij}(t)$  satisfy the systems of differential equations*

$$P'(t) = P(t)Q, \quad (21.2.6)$$

$$P'(t) = QP(t). \quad (21.2.7)$$

Each of the systems (21.2.6) and (21.2.7) has a unique solution

$$P(t) = e^{Qt}.$$

It is clear that the solution can be obtained immediately by formally integrating equation (21.2.6).

*Proof* By virtue of (21.2.1), (21.2.2) and (21.2.5),

$$P'(t) = \lim_{s \rightarrow 0} \frac{P(t+s) - P(t)}{s} = \lim_{s \rightarrow 0} P(t) \frac{P(s) - E}{s} = P(t)Q. \quad (21.2.8)$$

In the same way we obtain, from the equality

$$P(t+s) - P(t) = (P(s) - E)P(t),$$

the second equation in (21.2.7). The passage to the limit is justified by the assumptions we made.

Further, it follows from (21.2.6) that the function  $P(t)$  is infinitely differentiable, and

$$\begin{aligned} P^{(k)}(t) &= P(t)Q^k, \\ P(t) - P(0) &= \sum_{k=1}^{\infty} P^{(k)}(0) \frac{t^k}{k!} = \sum_{k=1}^{\infty} \frac{Q^k t^k}{k!}, \\ P(t) &= P(0)e^{Qt}. \end{aligned}$$

The theorem is proved. □



Because of the derivation method, (21.2.6) is called the *backward Kolmogorov equation*, and (21.2.7) is known as the *forward Kolmogorov equation* (the time increment is taken *after* or *before* the basic time interval).

The difference between these equations becomes even more graphical in the case of inhomogeneous Markov processes, when the transition probabilities

$$\mathbf{P}(\xi(t) = j | \xi(s) = i) = p_{ij}(s, t), \quad s \leq t,$$

depend on two time arguments:  $s$  and  $t$ . In that case, (21.2.1) becomes the equality  $P(s, t + u) = P(s, t)P(t, t + u)$ , and the backward and forward equations have the form

$$\frac{\partial P(s, t)}{\partial s} = P(s, t)Q(s), \quad \frac{\partial P(s, t)}{\partial t} = Q(t)P(s, t),$$

respectively, where

$$Q(t) = \lim_{u \rightarrow 0} \frac{P(t, t + u) - E}{u}.$$

The reader can derive these relations independently.

What are the general conditions for existence of a stationary limiting distribution? We can use here an approach similar to that employed in Chap. 13.

Let  $\xi^{(i)}(t)$  be a process with the initial value  $\xi^{(i)}(0) = i$  and right-continuous trajectories. For a given  $i_0$ , put

$$\begin{aligned} v^{(i)} &:= \min\{t \geq 0 : \xi^{(i)}(t) = i_0\} =: v_0, \\ v_k &:= \min\{t \geq v_{k-1} + 1 : \xi^{(i)}(t) = i_0\}, \quad k = 1, 2, \dots \end{aligned}$$

Here in the second formula we consider the values  $t \geq v_{k-1} + 1$ , since for  $t \geq v_{k-1}$  we would have  $v_k \equiv v_{k-1}$ . Clearly,  $\mathbf{P}(v_k - v_{k-1} = 1) > 0$ , and  $\mathbf{P}(v_k - v_{k-1} \in (t, t + h)) > 0$  for any  $t \geq 1$  and  $h > 0$  provided that  $p_{i_0 i_0}(t) \neq 1$ .

Note also that the variables  $v_k, k = 0, 1, \dots$ , are not defined for all elementary outcomes. We put  $v_0 = \infty$  if  $\xi^{(i)}(t) \neq i_0$  for all  $t \geq 0$ . A similar convention is used for  $v_k, k \geq 1$ . The following ergodic theorem holds.

**Theorem 21.2.3** *Let there exist a state  $i_0$  such that  $\mathbf{E}v_1 < \infty$  and  $\mathbf{P}(v^{(i)} < \infty) = 1$  for all  $i \in \mathcal{X}_0 \subset \mathcal{X}$ . Then there exist the limits*

$$\lim_{t \rightarrow \infty} p_{ij}(t) = p_j \tag{21.2.9}$$

which are independent of  $i \in \mathcal{X}_0$ .

*Proof* As was the case for Markov chains, the epochs  $v_1, v_2, \dots$  divide the time axis into independent cycles of the same nature, each of them being completed when the system returns for the first time (after one time unit) to the state  $i_0$ . Consider the renewal process generated by the sums  $v_k, k = 0, 1, \dots$ , of independent random variables  $v_0, v_k - v_{k-1}, k = 1, 2, \dots$ . Let

$$\eta(t) := \min\{k : v_k > t\}, \quad \gamma(t) := t - v_{\eta(t)-1}, \quad H(t) := \sum_{k=0}^{\infty} \mathbf{P}(v_k < t).$$

The event  $A_{dv} := \{\gamma(t) \in [v, v + dv)\}$  can be represented as the intersection of the events

$$B_{dv} := \bigcup_{k \geq 0} \{v_k \in (t - v - dv, t - v)\} \in \mathfrak{F}_{1-v}$$

and  $C_v := \{\xi(u) \neq i_0 \text{ for } u \in [t - v + 1, t]\} \in \mathfrak{F}_{[t-v, \infty)}$ . We have

$$\begin{aligned} p_{ij}(t) &= \int_0^t \mathbf{P}(\xi^{(i)}(t) = j, \gamma(t) \in [v, v + dv)) = \int_0^t \mathbf{P}(\xi^{(i)} = j, B_{dv}C_v) \\ &= \int_0^t \mathbf{E}[\mathbf{1}_{B_{dv}} \mathbf{P}(\xi^{(i)}(t) = j, C_v | \mathfrak{F}_{t-v})] \\ &= \int_0^t \mathbf{E}[\mathbf{1}_{B_{dv}} \mathbf{P}(\xi^{(i)}(t) = j, C_v | \xi(t - v))]. \end{aligned}$$

On the set  $B_{dv}$ , one has  $\xi(t - v) = i_0$ , and hence the probability inside the last integral is equal to

$$\mathbf{P}(\xi^{(i_0)}(v) = j, \xi(u) \neq i_0 \text{ for } u \in [1, v]) =: g(v)$$

and is independent of  $t$  and  $i$ . Since  $\mathbf{P}(B_{dv}) = dH(t - v)$ , one has

$$p_{ij}(t) = \int_0^t g(v) \mathbf{P}(B_{dv}) = \int_0^t g(v) dH(t - v).$$

By the key renewal theorem, as  $t \rightarrow \infty$ , this integral converges to

$$\frac{1}{\mathbf{E}v_1} \int_0^\infty g(v) dv.$$

The existence of the last integral follows from the inequality  $g(v) \leq \mathbf{P}(v_1 > v)$ . The theorem is proved. □

**Theorem 21.2.4** *If the stationary distribution*

$$P = \lim_{t \rightarrow \infty} P(t)$$

*exists with all the rows of the matrix  $P$  being identical, then it is a unique solution of the equation*

$$PQ = 0. \tag{21.2.10}$$

It is evident that Eq. (21.2.10) is obtained by setting  $P'(t) = 0$  in (21.2.6). Equation (21.2.7) gives the trivial equality  $QP = 0$ .

*Proof* Equation (21.2.10) is obtained by passing to the limit in (21.2.8) first as  $t \rightarrow \infty$  and then as  $s \rightarrow 0$ . Now assume that  $P_1$  is a solution of (21.2.10), i.e.  $P_1Q = 0$ . Then  $P_1P(t) = P_1$  for  $t < 1$ , since

$$P_1(P(t) - P(0)) = P_1 \sum_{k=1}^\infty \frac{Q^k t^k}{k!} = 0.$$

Further,  $P_1 = P_1 P^k(t) = P_1 P(kt)$ ,  $P(kt) \rightarrow P$  as  $k \rightarrow \infty$ , and hence  $P_1 = P_1 P = P$ . The theorem is proved.  $\square$

Now consider a Markov chain  $\{X_n\}$  in discrete time with transition probabilities  $p_{ij} = q_{ij}/q_i$ ,  $i \neq j$ ,  $p_{ii} = 0$ . Suppose that this chain is ergodic (see Theorem 13.4.1). Then its stationary probabilities  $\{\pi_j\}$  satisfy Eqs. (13.4.2). Now note that Eq. (21.2.10) can be written in the form

$$p_j q_j = \sum_k p_k q_k p_{kj}$$

which has an obvious solution  $p_j = c\pi_j/q_j$ ,  $c = \text{const}$ . Therefore, if

$$\sum \frac{\pi_j}{q_j} < \infty \tag{21.2.11}$$

then there exists a solution to (21.2.10) given by

$$p_j = \frac{\pi_j}{q_j} \left( \sum \frac{\pi_j}{q_j} \right)^{-1}. \tag{21.2.12}$$

In Sects. 21.4 and 21.5 we will derive the ergodic theorem for processes of a more general form than the one in the present section. That theorem will imply, in particular, that ergodicity of  $\{X_n\}$  and convergence (21.2.11) imply (21.2.9). Recall that, for ergodicity of  $\{X_n\}$ , it suffices, in turn, that Eqs. (13.4.2) have a solution  $\{\pi_j\}$ . Thus the existence of solution (21.2.12) implies the ergodicity of  $\xi(t)$ .

### 21.2.2 Examples

*Example 21.2.1* The Poisson process  $\xi(t)$  with parameter  $\lambda$  is a Markov process for which  $q_i = \lambda$ ,  $q_{i,i+1} = \lambda$ , and  $p_{i,i+1} = 1$ ,  $i = 1, 0, \dots$ . For this process, the stationary distribution  $p = (p_0, p_1, \dots)$  does not exist (each trajectory goes to infinity).

*Example 21.2.2* Birth-and-death processes. These are processes for which, for  $i \geq 1$ ,

$$p_{ij}(\Delta) = \begin{cases} \lambda_i \Delta + o(\Delta) & \text{for } j = i + 1, \\ \mu_i \Delta + o(\Delta) & \text{for } j = i - 1, \\ o(\Delta) & \text{for } |j - i| \geq 2, \end{cases}$$

so that

$$p_{ij} = \begin{cases} \frac{\lambda_i}{\lambda_i + \mu_i} & \text{for } j = i + 1, \\ \frac{\mu_i}{\lambda_i + \mu_i} & \text{for } j = i - 1 \end{cases}$$

are probabilities of birth and death, respectively, of a particle in a certain population given that the population consisted of  $i$  particles and changed its composition. For

$i = 0$  one should put  $\mu_0 := 0$ . Establishing conditions for the existence of a stationary regime is a rather difficult problem (related mainly to finding conditions under which the trajectory escapes to infinity). If the stationary regime exists, then according to Theorem 21.2.4 the stationary probabilities  $p_j$  can be uniquely determined from the recursive relations (see Eq. (21.2.10), in our case  $q_{ii} = -q_i = -(\lambda_i + \mu_i)$ )

$$\begin{aligned} & -p_0\lambda_0 + p_1\mu_1 = 0, \\ & p_0\lambda_0 - p_1(\lambda_1 + \mu_1) + p_2\mu_2 = 0, \\ & \dots\dots\dots \\ & p_{k-1}\lambda_{k-1} - p_k(\lambda_k + \mu_k) + p_{k+1}\mu_{k+1} = 0, \\ & \dots\dots\dots \end{aligned} \tag{21.2.13}$$

and condition  $\sum p_j = 1$ .

*Example 21.2.3 The telephone lines problem* from queueing theory. Suppose we are given a system consisting of infinitely many communication channels which are used for telephone conversations. The probability that, for a busy channel, the transmitted conversation terminates during a small time interval  $(t, t + \Delta)$  is equal to  $\lambda\Delta + o(\Delta)$ . The probability that a request for a new conversation (a new call) arrives during the same time interval is  $\mu\Delta + o(\Delta)$ . Thus the ‘‘arrival flow’’ of calls is nothing else but the Poisson process with parameter  $\lambda$ , and the number  $\xi(t)$  of busy channels at time  $t$  is the value of the birth-and-death process for which  $\lambda_i = \lambda$  and  $\mu_i = i\mu$ .

In that case, it is not hard to verify with the help of Theorem 21.2.3 that there always exists a stationary limiting distribution, for which Eqs. (21.2.13) have the form

$$\begin{aligned} & \lambda p_0 = \mu p_1, \\ & \dots\dots\dots \\ & (\lambda + \mu k)p_k = \lambda p_{k-1} + (k + 1)\mu p_{k+1}, \\ & \dots\dots\dots \end{aligned} \tag{21.2.14}$$

From this we get that

$$p_1 = p_0 \frac{\lambda}{\mu}, \quad p_2 = \frac{p_0}{2} \left(\frac{\lambda}{\mu}\right)^2, \quad \dots, \quad p_k = \left(\frac{\lambda}{\mu}\right)^k \frac{p_0}{k!}, \tag{21.2.15}$$

so that  $p_0 = e^{-\lambda/\mu}$ , and the limiting distribution will be the Poisson law with parameter  $\lambda/\mu$ .

If the number of channels  $n$  is finite, the calls which find all the lines busy will be rejected, and in (21.2.13) one has to put  $\lambda_n = 0$ ,  $p_{n+1} = p_{n+2} = \dots = 0$ . In that case, the last equation in (21.2.14) will have the form  $\mu n p_n = \lambda p_{n-1}$ . Since the formulas (21.2.15) will remain true for  $k \leq n$ , we obtain the so-called Erlang formulas for the stationary distribution:

$$p_k = \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} \left[ \sum_{j=0}^n \frac{1}{j!} \left(\frac{\lambda}{\mu}\right)^j \right]^{-1}$$

(the truncated Poisson distribution).

The next example will be considered in a separate section.

### 21.3 Branching Processes

The essence of the mathematical model describing a branching process remains roughly the same as in Sect. 7.7.2. A continuous time branching process can be defined as follows. Let  $\xi^{(i)}(t)$  denote the number of particles at time  $t$  with the initial condition  $\xi^{(i)}(0) = i$ . Each particle, independently of all others, splits during the time interval  $(t, t + \Delta)$  with probability  $\mu\Delta + o(\Delta)$  into a random number  $\eta \neq 1$  of particles (if  $\eta = 0$ , we say that the particle dies). Thus,

$$\xi^{(i)}(t) = \xi_1^{(1)}(t) + \dots + \xi_i^{(1)}(t), \tag{21.3.1}$$

where  $\xi_k^{(1)}(t)$  are independent and distributed as  $\xi^{(1)}(t)$ . Moreover,

$$\begin{aligned} p_{ij}(\Delta) &= i\mu\Delta h_{j-i+1} + o(\Delta), \quad j \neq i; \quad h_k = \mathbf{P}(\eta = k); \quad h_1 = 0; \\ p_{ii}(\Delta) &= 1 - i\mu\Delta + o(\Delta), \end{aligned} \tag{21.3.2}$$

so that here  $q_{ij} = i\mu h_{j-i+1}$ ,  $q_{ii} = -i\mu$ .

By formula (21.3.2),  $i\mu\Delta$  is the principal part of the probability that at least one particle will split. Clearly, the state 0 is absorbing. It will not be absorbing any more if one considers processes with *immigration* when a Poisson process (with intensity  $\lambda$ ) of “outside” particles is added to the process  $\xi^{(i)}(t)$ . Then

$$\begin{aligned} p_{ij}(\Delta) &= i\mu\Delta h_{j-i+1} + o(\Delta) \quad \text{for } j - i \neq 0, 1, \\ p_{i,i+1}(\Delta) &= \Delta(i\mu h_2 + \lambda) + o(\Delta). \end{aligned}$$

We return to the branching process (21.3.1), (21.3.2). By (21.3.1) we have

$$r^{(i)}(t, z) := \mathbf{E}z^{\xi^{(i)}(t)} = [\mathbf{E}z^{\xi^{(1)}(t)}]^i = r^i(t, z) = \sum_{k=0}^{\infty} z^k p_{ik}(t),$$

where

$$r(t, z) := \mathbf{E}z^{\xi^{(1)}(t)} = \sum_{k=0}^{\infty} z^k p_{1k}(t). \tag{21.3.3}$$

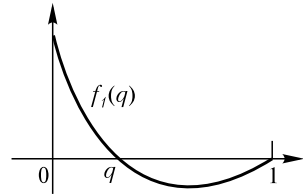
Equation (21.2.7) implies

$$p'_{1k}(t) = \sum_{l=0}^{\infty} q_{1l} p_{lk}(t).$$

Therefore, differentiating (21.3.3) with respect to  $t$ , we find that

$$\begin{aligned} r'_i(t, z) &= \sum_{k=0}^{\infty} z^k p'_{ik}(t) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} q_{1l} p_{lk}(t) z^k \\ &= \sum_{l=0}^{\infty} q_{1l} \sum_{k=0}^{\infty} z^k p_{lk}(t) = \sum_{l=0}^{\infty} q_{1l} r^l(t, z). \end{aligned} \tag{21.3.4}$$

**Fig. 21.1** The form of the plot of the function  $f_1$ . The smaller root of the equation  $f_1(q) = q$  gives the probability of the eventual extinction of the branching process



But  $q_{1l} = \mu p_l$  for  $l \neq 1$ ,  $q_{11} = -\mu$ , and putting

$$f(s) := \sum_{l=0}^{\infty} q_{1l} s^l = \mu(\mathbf{E}s^\eta - s) = \mu \left( \sum_{l=0}^{\infty} h_l s^l - s \right),$$

we can write (21.3.4) in the form

$$r'_t(t, z) = f(r(t, z)).$$

We have obtained a differential equation for  $r = r(t, z)$  (equivalent to (21.3.2)) which is more convenient to write in the form

$$\frac{dr}{f(r)} = dt, \quad t = \int_{r(0,z)}^{r(t,z)} \frac{dy}{f(y)} = \int_z^{r(t,z)} \frac{dy}{f(y)}.$$

Consider the behaviour of the function  $f_1(y) = \mathbf{E}y^\eta - y$  on  $[0, 1]$ . Clearly,  $f_1(0) = \mathbf{P}(\eta = 0)$ ,  $f_1(1) = 0$ , and

$$f'_1(1) = \mathbf{E}\eta - 1, \quad f''_1(y) = \mathbf{E}\eta(\eta - 1)y^{\eta-2} > 0.$$

Consequently, the function  $f_1(y)$  is convex and has no zeros in  $(0, 1)$  if  $\mathbf{E}\eta \leq 1$ . When  $\mathbf{E}\eta > 1$ , there exists a point  $q \in (0, 1)$  such that  $f_1(q) = 0$ ,  $f'_1(q) < 0$  (see Fig. 21.1), and  $f_1(y) = (y - q)f'_1(q) + O((y - q)^2)$  in the vicinity of this point.

Thus if  $\mathbf{E}\eta > 1$ ,  $z < q$  and  $r \uparrow q$ , then, by virtue of the representation

$$\frac{1}{f_1(y)} = \frac{1}{(y - q)f'_1(q)} + O(1),$$

we obtain

$$t = \int_z^r \frac{dy}{f(y)} = \frac{1}{\mu f'_1(q)} \ln \left( \frac{r - q}{z - q} \right) + O(1).$$

This implies that, as  $t \rightarrow \infty$ ,

$$\begin{aligned} r(t, z) - q &= (z - q)e^{\mu t f'_1(q) + O(1)} \sim (z - q)e^{\mu t f'_1(q)}, \\ r(t, z) &= q + O(e^{-\alpha t}), \quad \alpha = -\mu f'_1(q) > 0. \end{aligned} \tag{21.3.5}$$

In particular, the extinction probability

$$p_{10}(t) = r(t, 0) = q + O(e^{-\alpha t})$$

converges exponentially fast to  $q$ ,  $p_{10}(\infty) = q$ . Comparing our results with those from Sect. 7.7, the reader can see that the extinction probability for a discrete time

branching process had the same value (we could also come to this conclusion directly). Since  $p_{k0}(t) = [p_{10}(t)]^k$ , one has  $p_{k0}(\infty) = q^k$ .

It follows from (21.3.5) that the remaining “probability mass” of the distribution of  $\xi(t)$  quickly moves to infinity as  $t \rightarrow \infty$ .

If  $E\eta < 1$ , the above argument remains valid with  $q$  replaced with 1, so that the extinction probability is  $p_{10}(\infty) = p_{k0}(\infty) = 1$ .

If  $E\eta = 1$ , then

$$f_1(y) = \frac{(y-1)^2}{2} f_1''(1) + O((y-1)^3),$$

$$t = \int_z^r \frac{dy}{f(y)} \sim -\frac{2}{\mu f_1''(1)} \cdot \frac{1}{r-1}, \quad r(t, z) - 1 \sim -\frac{2}{\mu t f_1''(1)}.$$

Thus the extinction probability  $r(t, 0) = p_{10}(t)$  also tends to 1 in this case.

## 21.4 Semi-Markov Processes

### 21.4.1 Semi-Markov Processes on the States of a Chain

*Semi-Markov processes* can be described as follows. Let an aperiodic discrete time irreducible Markov chain  $\{X_n\}$  with the state space  $\mathcal{X} = \{0, 1, 2, \dots\}$  be given. To each state  $i$  we put into correspondence the distribution  $F_i(t)$  of a positive random variable  $\zeta^{(i)}$ :

$$F_i(t) = \mathbf{P}(\zeta^{(i)} < t).$$

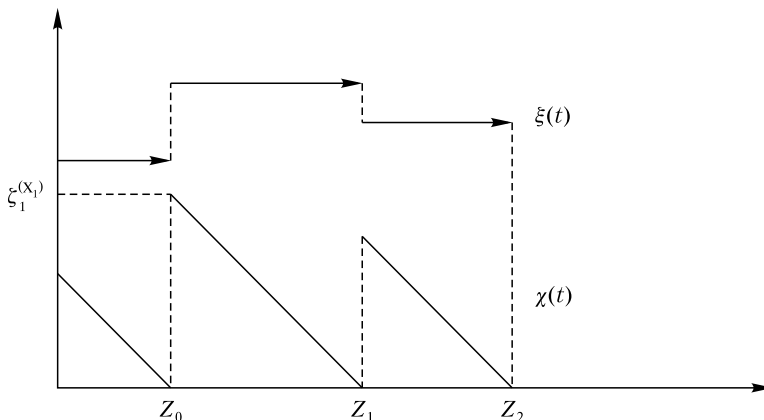
Consider independent of the chain  $\{X_n\}$  and of each other the sequences  $\zeta_1^{(i)}, \zeta_2^{(i)}, \dots; \zeta_j^{(i)} \stackrel{d}{=} \zeta^{(i)}$ , of independent random variables with the distribution  $F_i$ . Let, moreover, the distribution of the initial random vector  $(X_0, \zeta_0)$ ,  $X_0 \in \mathcal{X}$ ,  $\zeta_0 \geq 0$ , be given. The evolution of the semi-Markov process  $\xi(u)$  is described as follows:

$$\begin{aligned} \xi(u) &= X_0 && \text{for } 0 \leq u < \zeta_0, \\ \xi(u) &= X_1 && \text{for } \zeta_0 \leq u < \zeta_0 + \zeta_1^{(X_1)}, \\ \xi(u) &= X_2 && \text{for } \zeta_0 + \zeta_1^{(X_1)} \leq u < \zeta_0 + \zeta_1^{(X_1)} + \zeta_2^{(X_2)}, \\ &\dots, \\ \xi(u) &= X_n && \text{for } Z_{n-1} \leq u < Z_n, \quad Z_n = \zeta_0 + \zeta_1^{(X_1)} + \dots + \zeta_n^{(X_n)}, \end{aligned} \tag{21.4.1}$$

and so on. Thus, upon entering state  $X_n = j$ , the trajectory of  $\xi(u)$  remains in that state for a random time  $\zeta_n^{(X_n)} = \zeta_n^{(j)}$ , then switches to state  $X_{n+1}$  and so on. It is evident that such a process is, generally speaking, not Markovian. It will be a Markov process only if

$$1 - F_i(t) = e^{-q_i t}, \quad q_i > 0,$$

and will then coincide with the process described in Sect. 21.2.



**Fig. 21.2** The trajectories of the semi-Markov process  $\xi(t)$  and of the residual sojourn time process  $\chi(t)$

If the distribution  $F_i$  is not exponential, then, given the value  $\xi(t) = i$ , the time between  $t$  and the next jump epoch will depend on the epoch of the preceding jump of  $\xi(\cdot)$ , because

$$\mathbf{P}(\zeta^{(i)} > v + u \mid \zeta^{(i)} > v) = \frac{1 - F_i(v + u)}{1 - F_i(v)}$$

for non-exponential  $F_i$  depends on  $v$ . It is this property that means that the process is non-Markovian, for fixing the “present” (i.e. the value of  $\xi(t)$ ) does not make the “future” of the process  $\xi(u)$  independent of the “past” (i.e. of the trajectory of  $\xi(u)$  for  $u < t$ ).

The process  $\xi(t)$  can be “complemented” to a Markov one by adding to it the component  $\chi(t)$  of which the value gives the time  $u$  for which the trajectory  $\xi(t + u)$ ,  $u \geq 0$ , will remain in the current state  $\xi(t)$ . In other words,  $\chi(t)$  is the excess of level  $t$  for the random walk  $Z_0, Z_1, \dots$  (see Fig. 21.2):

$$\chi(t) = Z_{v(t)+1} - t, \quad v(t) = \max\{k : Z_k \leq t\}.$$

The process  $\chi(t)$  is Markovian and has “saw-like” trajectories deterministic inside the intervals  $(Z_k, Z_{k+1})$ . The process  $X(t) = (\xi(t), \chi(t))$  is obviously Markovian, since the value of  $X(t)$  uniquely determines the law of evolution of the process  $X(t + u)$  for  $u \geq 0$  whatever the “history”  $X(v)$ ,  $v < t$ , is. Similarly, we could consider the Markov process  $Y(t) = (\xi(t), \gamma(t))$ , where  $\gamma(t)$  is the defect of level  $t$  for the walk  $Z_0, Z_1, \dots$ :

$$\gamma(t) = t - Z_{v(t)}.$$

### 21.4.2 The Ergodic Theorem

In the sequel, we will distinguish between the following two cases.



(A) The *arithmetic* case when the possible values of  $\zeta^{(i)}$ ,  $i = 0, 1, \dots$ , are multiples of a certain value  $h$  which can be assumed without loss of generality to be equal to 1. In that case we will also assume that the g.c.d. of the possible values of the sums of the variables  $\zeta^{(i)}$  is also equal to  $h = 1$ . This is clearly equivalent to assuming that the g.c.d. of the possible values of recurrence times  $\theta^{(i)}$  of  $\xi(t)$  to the state  $i$  is equal to 1 for any fixed  $i$ .

(NA) The *non-arithmetic* case, when condition (A) does not hold.  
 Put  $a_i := \mathbf{E}\zeta^{(i)}$ .

**Theorem 21.4.1** *Let the Markov chain  $\{X_n\}$  be ergodic (satisfy the conditions of Theorem 13.4.1) and  $\{\pi_j\}$  be the stationary distribution of that chain. Then, in the non-arithmetic case (NA), for any initial distribution  $(\zeta_0, X_0)$  there exists the limit*

$$\lim_{t \rightarrow \infty} \mathbf{P}(\xi(t) = i, \chi(t) > v) = \frac{\pi_i}{\sum \pi_j a_j} \int_v^\infty \mathbf{P}(\zeta^{(i)} > u) du. \tag{21.4.2}$$

In the arithmetic case (A), (21.4.2) holds for integer-valued  $v$  (the integral becomes a sum in that case). It follows from (21.4.2) that the following limit exists

$$\lim_{t \rightarrow \infty} \mathbf{P}(\xi(t) = i) = \frac{\pi_i a_i}{\sum \pi_j a_j}.$$

*Proof* For definiteness we restrict ourselves to the non-arithmetic case (NA). In Sect. 13.4 we considered the times  $\tau^{(i)}$  between consecutive visits of  $\{X_n\}$  to state  $i$ . These times could be called “embedded”, as well as the chain  $\{X_n\}$  itself in regard to the process  $\xi(t)$ . Along with the times  $\tau^{(i)}$ , we will need the “real” times  $\theta^{(i)}$  between the visits of the process  $\xi(t)$  to the state  $i$ . Let, for instance,  $X_1 = 1$ . Then

$$\theta^{(1)} \stackrel{d}{=} \zeta_1^{(X_1)} + \zeta_2^{(X_2)} + \dots + \zeta_\tau^{(X_\tau)},$$

where  $\tau = \tau^{(1)}$ . For definiteness and to reduce notation, we fix for the moment the value  $i = 1$  and put  $\theta^{(1)} =: \theta$ . Let first

$$\zeta_0 \stackrel{d}{=} \zeta^{(1)}, \quad X_0 = 1. \tag{21.4.3}$$

Then the whole trajectory of the process  $X(t)$  for  $t \geq 0$  will be divided into identically distributed independent cycles by the epochs when the process hits the state  $\xi(t) = 1$ . We denote the lengths of these cycles by  $\theta_1, \theta_2, \dots$ ; they are independent and identically distributed. We show that

$$\mathbf{E}\theta = \frac{1}{\pi_1} \sum a_j \pi_j. \tag{21.4.4}$$

Denote by  $\theta(n)$  the “real” time spent on  $n$  transitions of the governing chain  $\{X_n\}$ . Then

$$\theta_1 + \dots + \theta_{\eta(n)-1} \leq \theta(n) \leq \theta_1 + \dots + \theta_{\eta(n)}, \tag{21.4.5}$$

where  $\eta(n) := \min\{k : T_k > n\}$ ,  $T_k = \sum_{j=1}^k \tau_j$ ,  $\tau_j$  are independent and distributed as  $\tau$ . We prove that, as  $n \rightarrow \infty$ ,

$$\mathbf{E}\theta(n) \sim n\pi_1 \mathbf{E}\theta. \tag{21.4.6}$$

By Wald's identity and (21.4.5),

$$\mathbf{E}\theta(n) \leq \mathbf{E}\theta\mathbf{E}\eta(n), \tag{21.4.7}$$

where  $\mathbf{E}\eta(n) \sim n/\mathbf{E}\tau = n\pi_1$ .

Now we bound from below the expectation  $\mathbf{E}\theta(n)$ . Put  $m := \lfloor n\pi_1 - \varepsilon n \rfloor$ ,  $\Theta_n := \sum_{j=1}^n \theta_j$ . Then

$$\begin{aligned} \mathbf{E}\theta(n) &\geq \mathbf{E}(\theta(n); \eta(n) > m) \\ &\geq \mathbf{E}(\Theta_m; \eta(n) > m) = m\mathbf{E}\theta - \mathbf{E}(\Theta_m; \eta(n) \leq m). \end{aligned} \tag{21.4.8}$$

Here the random variable  $\Theta_m/m \geq 0$  possesses the properties

$$\Theta_m/m \xrightarrow{P} \mathbf{E}\theta \quad \text{as } m \rightarrow \infty, \quad \mathbf{E}(\Theta_m/m) = \mathbf{E}\theta.$$

Therefore it satisfies the conditions of part 4 of Lemma 6.1.1 and is uniformly integrable. This, in turn, by Lemma 6.1.2 and convergence  $\mathbf{P}(\eta(n) \leq m) \rightarrow 0$  means that the last term on the right-hand side of (21.4.8) is  $o(m)$ . By virtue of (21.4.8), since  $\varepsilon > 0$  is arbitrary, we obtain that

$$\liminf_{n \rightarrow \infty} n^{-1}\mathbf{E}\theta(n) \geq \pi_1\mathbf{E}\theta.$$

This together with (21.4.7) proves (21.4.6).

Now we will calculate the value of  $\mathbf{E}\theta(n)$  using another approach. The variable  $\theta(n)$  admits the representation

$$\theta(n) = \sum_j (\zeta_1^{(j)} + \dots + \zeta_{N(j,n)}^{(j)}),$$

where  $N(j, n)$  is the number of visits of the trajectory of  $\{X_k\}$  to the state  $j$  during the first  $n$  steps. Since  $\{\zeta_k^{(j)}\}_{k=1}^\infty$  and  $N(j, n)$  are independent for each  $j$ , we have

$$\mathbf{E}\theta(n) = \sum_j a_j \mathbf{E}N(j, n), \quad \mathbf{E}N(j, n) = \sum_{k=1}^n p_{1j}(k).$$

Because  $p_{1j}(k) \rightarrow \pi_j$  as  $k \rightarrow \infty$ , one has

$$\lim_{n \rightarrow \infty} n^{-1}\mathbf{E}N(j, n) = \pi_j.$$

Moreover,

$$\pi_j = \sum_l \pi_l p_{lj}(k) \geq \pi_1 p_{1j}(k)$$

and, therefore,

$$p_{1j}(k) \leq \pi_j/\pi_1.$$

Hence

$$n^{-1}\mathbf{E}N(j, n) \leq \pi_j/\pi_1,$$

and in the case when  $\sum_j a_j \pi_j < \infty$ , the series  $\sum_j a_j n^{-1} \mathbf{E}N(j, n)$  converges uniformly in  $n$ . Consequently, the following limit exists

$$\lim_{n \rightarrow \infty} n^{-1} \mathbf{E}\theta(n) = \sum_j a_j \pi_j.$$

Comparing this with (21.4.6) we obtain (21.4.4). If  $\mathbf{E}\theta = \infty$  then clearly  $\mathbf{E}\theta(n) = \infty$  and  $\sum_j a_j \pi_j = \infty$ , and vice versa, if  $\sum_j a_j \pi_j = \infty$  then  $\mathbf{E}\theta = \infty$ .

Consider now the random walk  $\{\Theta_k\}$ . To the  $k$ -th cycle there correspond  $T_k$  transitions. Therefore, by the total probability formula,

$$\mathbf{P}(\xi(t) = 1, \chi(t) > v) = \sum_{k=1}^{\infty} \int_0^t \mathbf{P}(\Theta_k \in du, \zeta_{T_{k+1}}^{(1)} > t - u + v),$$

where  $\zeta_{T_{k+1}}^{(1)}$  is independent of  $\Theta_k$  and distributed as  $\zeta^{(1)}$  (see Lemma 11.2.1 or the strong Markov property). Therefore, denoting by  $H_\theta(u) := \sum_{k=1}^{\infty} \mathbf{P}(\Theta_k < u)$  the renewal function for the sequence  $\{\Theta_k\}$ , we obtain for the non-arithmetic case (NA), by virtue of the renewal theorem (see Theorem 10.4.1 and (10.4.2)), that, as  $t \rightarrow \infty$ ,

$$\begin{aligned} \mathbf{P}(\xi(t) = 1, \chi(t) > v) &= \int_0^t dH_\theta(u) \mathbf{P}(\zeta^{(1)} > t - u + v) \\ &\rightarrow \frac{1}{\mathbf{E}\theta} \int_0^\infty \mathbf{P}(\zeta^{(1)} > u + v) dv = \frac{1}{\mathbf{E}\theta} \int_v^\infty \mathbf{P}(\zeta^{(1)} > u) du. \end{aligned} \quad (21.4.9)$$

We have proved assertion (21.4.2) for  $i = 1$  and initial conditions (21.4.3). The transition to arbitrary initial conditions is quite obvious and is done in exactly the same way as in the proof of the ergodic theorems of Chap. 13.

If  $\sum a_i \pi_i = \infty$  then, as we have already observed,  $\mathbf{E}\theta = \infty$  and, by the renewal theorem and (21.4.9), one has  $\mathbf{P}(\xi(t) = 1, \chi(t) > v) \rightarrow 0$  as  $t \rightarrow \infty$ . It remains to note that instead of  $i = 1$  we can fix any other value of  $i$ . The theorem is proved.  $\square$

In the same way we could also prove that

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbf{P}(\xi(t) = i, \gamma(t) > v) &= \frac{\pi_i}{\sum a_j \pi_j} \int_v^\infty \mathbf{P}(\zeta^{(i)} > y) dy, \\ \lim_{t \rightarrow \infty} \mathbf{P}(\xi(t) = i, \chi(t) > u, \gamma(t) > v) &= \frac{\pi_i}{\sum a_j \pi_j} \int_{u+v}^\infty \mathbf{P}(\zeta^{(i)} > y) dy \end{aligned}$$

(see Theorem 10.4.3).

### 21.4.3 Semi-Markov Processes on Chain Transitions

Along with the semi-Markov processes  $\xi(t)$  described at the beginning of the present section, one sometimes considers semi-Markov processes “given on the

transitions” of the chain  $\{X_n\}$ . In that case, the distributions  $F_{ij}$  of random variables  $\zeta^{(ij)} > 0$  are given and, similarly to (21.4.1), for the initial condition  $(X_0, X_1, \zeta_0)$  one puts

$$\begin{aligned} \xi(u) &:= (X_0, X_1) && \text{for } 0 \leq u < \zeta_0 \\ \xi(u) &:= (X_1, X_2) && \text{for } \zeta_0 \leq u < \zeta_0 + \zeta_1^{(X_0, X_1)} \\ \xi(u) &:= (X_2, X_3) && \text{for } \zeta_0 + \zeta_1^{(X_0, X_1)} \leq u < \zeta_0 + \zeta_1^{(X_0, X_1)} + \zeta_2^{(X_1, X_2)}, \end{aligned} \tag{21.4.10}$$

and so on. Although at first glance this is a very general model, it can be completely reduced to the semi-Markov processes (21.4.1). To that end, one has to notice that the “two-dimensional” sequence  $Y_n = (X_n, X_{n+1})$ ,  $n = 0, 1, \dots$ , also forms a Markov chain. Its transition probabilities have the form

$$\begin{aligned} p^{(ij)(kl)} &= \begin{cases} p_{jl} & \text{for } k = j, \\ 0 & \text{for } k \neq j, \end{cases} \\ p^{(ij)(kl)}(n) &= p_{jk}(n)p_{kl} \quad \text{for } n > 1, \end{aligned}$$

so that if the chain  $\{X_n\}$  is ergodic, then  $\{Y_n\}$  is also ergodic and

$$p^{(ij)(kl)}(n) \rightarrow \pi_k p_{kl}.$$

This enables one to restate Theorem 21.4.1 easily for the semi-Markov processes (21.4.10) given on the transitions of the Markov chain  $\{X_n\}$ , since the process (21.4.10) will be an ordinary semi-Markov process given on the chain  $\{Y_n\}$ .

**Corollary 21.4.1** *If the chain  $\{X_n\}$  is ergodic then, in the non-arithmetic case,*

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbf{P}(\xi(t) = (i, j), \chi(t) > v) \\ = \frac{\pi_i p_{ij}}{\sum_{k,l} a_{kl} \pi_k p_{kl}} \int_v^\infty \mathbf{P}(\zeta^{(ij)} > u) du, \quad a_{kl} = \mathbf{E}\zeta^{(kl)}. \end{aligned}$$

*In the arithmetic case  $v$  must be a multiple of the lattice span.*

We will make one more remark which could be helpful when studying semi-Markov processes and which concerns the so-called semi-Markov renewal functions  $H_{ij}(t)$ . Denote by  $T_{ij}(n)$  the epoch (in the “real time”) of the  $n$ -th jump of the process  $\xi(t)$  from state  $i$  to  $j$ . Put

$$H_{ij}(t) := \sum_{n=1}^\infty \mathbf{P}(T_{ij}(n) < t).$$

If  $v_{ij}(t)$  is the number of jumps from state  $i$  to  $j$  during the time interval  $[0, t)$ , then clearly  $H_{ij}(t) = \mathbf{E}v_{ij}(t)$ .

Set  $\Delta f(t) := f(t + \Delta) - f(t)$ ,  $\Delta > 0$ .

**Corollary 21.4.2** *In the non-arithmetic case,*

$$\lim_{t \rightarrow \infty} \Delta H_{ij}(t) = \frac{\pi_i p_{ij} \Delta}{\sum_l a_l \pi_l}. \tag{21.4.11}$$

*In the arithmetic case  $v$  must be a multiple of the lattice span.*

*Proof* Denote by  $v_{ij}^{(k)}(u)$  the number of transitions of the process  $\xi(t)$  from  $i$  to  $j$  during the time interval  $(0, u)$  given the initial condition  $(k, 0)$ . Then, by the total probability formula,

$$\mathbf{E}\Delta v_{ij}(t) = \int_0^\Delta \sum_{k=0}^\infty \mathbf{P}(\xi(t) = k, \chi(t) \in du) \mathbf{E}v_{ij}^{(k)}(\Delta - u).$$

Since  $v_{ij}^{(k)}(u) \leq v_{ij}^{(i)}(u)$ , by Theorem 21.4.1 one has

$$h_{ij}(\Delta) := \lim_{t \rightarrow \infty} \mathbf{E}\Delta v_{ij}(t) = \frac{1}{\sum_l a_l \pi_l} \sum_{k=0}^\infty \pi_k \int_0^\Delta \mathbf{P}(\zeta^{(k)} > u) \mathbf{E}v_{ij}^{(k)}(\Delta - u) du. \tag{21.4.12}$$

Further,

$$\mathbf{P}(\zeta^{(i)} < \Delta - u) \leq F_i(\Delta) \rightarrow 0$$

as  $\Delta \rightarrow 0$ , and

$$\begin{aligned} \mathbf{P}(v_{ij}^{(k)}(\Delta - u) = s) &\leq (p_{ij} F_i(\Delta))^s, \quad k \neq i, \\ \mathbf{P}(v_{ij}^{(i)}(\Delta - u) = s + 1) &\leq (p_{ij} F_i(\Delta))^s, \quad s \geq 1, \\ \mathbf{P}(v_{ij}^{(i)}(\Delta - u) = 1) &= p_{ij} + o(F_i(\Delta)). \end{aligned}$$

It follows from the aforesaid that

$$\mathbf{E}v_{ij}^{(k)}(\Delta - u) = o(F_i(\Delta)), \quad \mathbf{E}v_{ij}^{(i)}(\Delta - u) = p_{ij} + o(F_i(\Delta)).$$

Therefore,

$$h_{ij}(\Delta) = \frac{\pi_i p_{ij} \Delta}{\sum_l a_l \pi_l} + o(\Delta). \tag{21.4.13}$$

Further, from the equality

$$H_{ij}(t + 2\Delta) - H_{ij}(t) = \Delta H_{ij}(t) + \Delta H_{ij}(t + \Delta)$$

we obtain that  $h_{ij}(2\Delta) = 2h_{ij}(\Delta)$ , which means that  $h_{ij}(\Delta)$  is linear. Together with (21.4.13) this proves (21.4.11). The corollary is proved.  $\square$

The class of processes for which one can prove ergodicity using the same methods as the one used for semi-Markov processes and also in Chap. 13, can be somewhat extended. For this broader class of processes we will prove in the next section the ergodic theorem, and also the laws of large numbers and the central limit theorem for integrals of such processes.

## 21.5 Regenerative Processes

### 21.5.1 Regenerative Processes. The Ergodic Theorem

Let  $X(t)$  and  $X_0(t); t \geq 0$ , be processes given in the space  $D(0, \infty)$  of functions without discontinuities of the second type (the state space of these processes could be any metric space, not necessarily the real line). The process  $X(t)$  is said to be *regenerative* if it possesses the following properties:

(1) There exists a state  $x_0$  which is visited by the process  $X$  with probability 1. After each such visit, the evolution of the process starts anew as if it were the original process  $X(t)$  starting at the state  $X(0) = x_0$ . We will denote this new process by  $X_0(t)$  where  $X_0(0) = x_0$ . To state this property more precisely, we introduce the time  $\tau_0$  of the first visit to  $x_0$  by  $X$ :

$$\tau_0 := \inf\{t \geq 0 : X(t) = x_0\}.$$

However, it is not clear from this definition whether  $\tau_0$  is a random variable. For definiteness, assume that the process  $X$  is such that for  $\tau_0$  one has

$$\{\tau_0 > t\} = \bigcup_n \bigcap_{t_k \in S} \{|X(t_k) - x_0| > 1/n\},$$

where  $S$  is a countable set everywhere dense in  $[0, t]$ . In that case the set  $\{\tau_0 > t\}$  is clearly an event and  $\tau_0$  is a random variable. The above stated property means that  $\tau_0$  is a proper random variable:  $\mathbf{P}(\tau_0 < \infty) = 1$ , and that the distribution of  $X(\tau_0 + u), u \geq 0$ , coincides with that of  $X_0(u), u \geq 0$ , whatever the ‘‘history’’ of the process  $X(t), t \leq \tau_0$ .

(2) The recurrence time  $\tau$  of the state  $x_0$  has finite expectation  $\mathbf{E}\tau < \infty$ ,  $\tau := \inf\{t : X_0(t) = x_0\}$ .

The aforesaid means that the evolution of the process is split into independent identically distributed cycles by its visits to the state  $x_0$ . The visit times to  $x_0$  are called *regeneration times*. The behaviour of the process inside the cycles may be arbitrary, and no further conditions, including Markovity, are imposed.

We introduce the so-called ‘‘taboo probability’’

$$P(t, B) := \mathbf{P}(X_0(t) \in B, \tau > t).$$

We will assume that, as a function of  $t$ ,  $P(t, B)$  is measurable and Riemann integrable.

**Theorem 21.5.1** *Let  $X(t)$  be a regenerative process and the random variable  $\tau$  be non-lattice. Then, for any Borel set  $B$ , as  $t \rightarrow \infty$ ,*

$$\mathbf{P}(X(t) \in B) \rightarrow \pi(B) = \frac{1}{\mathbf{E}\tau} \int_0^\infty \mathbf{P}(u, B) du.$$

*If  $\tau$  is a lattice variable (which is the case for processes  $X(t)$  in discrete time), the assertion holds true with the following obvious changes:  $t \rightarrow \infty$  along the lattice and the integral is replaced with a sum.*

*Proof* Let  $T_0 := 0$ ,  $T_k := \tau_1 + \cdots + \tau_k$  be the epoch of the  $k$ -th regeneration of the process  $X_0(t)$ , and

$$H(u) := \sum_{k=0}^{\infty} \mathbf{P}(\tau_k < u)$$

( $\tau_k \stackrel{d}{=} \tau$  are independent). Then, using the total probability formula and the key renewal theorem, we obtain, as  $t \rightarrow \infty$ ,

$$\begin{aligned} \mathbf{P}(X_0(t) \in B) &= \sum_{k=0}^{\infty} \int_0^t \mathbf{P}(T_k \in du) P(t-u, B) \\ &= \int_0^t dH(u) P(t-u, B) \rightarrow \frac{1}{\mathbf{E}\tau} \int_0^{\infty} P(u, B) du = \pi(B). \end{aligned}$$

For the process  $X(t)$  one gets

$$\mathbf{P}(X(t) \in B) = \int_0^t \mathbf{P}(t_0 \in du) \mathbf{P}(X_0(t-u) \in B) \rightarrow \pi(B).$$

The theorem is proved.  $\square$

### 21.5.2 The Laws of Large Numbers and Central Limit Theorem for Integrals of Regenerative Processes

Consider a measurable mapping  $f : \mathcal{X} \rightarrow \mathbb{R}$  of the state space  $\mathcal{X}$  of a process  $X(t)$  to the real line  $\mathbb{R}$ . As in Sect. 21.4.2, for the sake of simplicity, we can assume that  $\mathcal{X} = \mathbb{R}$  and the trajectories of  $X(t)$  lie in the space  $D(0, \infty)$  of functions without discontinuities of the second kind. In this case the paths  $f(X(u))$ ,  $u \geq 0$ , will be measurable functions, for which the integral

$$S(t) = \int_0^t f(X(u)) du$$

is well defined. For such integrals we have the following law of large numbers. Set

$$\zeta := \int_0^{\tau} f(X_0(u)) du, \quad a := \mathbf{E}\tau.$$

**Theorem 21.5.2** *Let the conditions of Theorem 21.5.1 be satisfied and there exist  $a_{\zeta} := \mathbf{E}\zeta$ . Then, as  $t \rightarrow \infty$ ,*

$$\frac{S(t)}{t} \xrightarrow{p} \frac{a_{\zeta}}{a}.$$

For conditions of existence of  $\mathbf{E}\zeta$ , see Theorem 21.5.4 below.

*Proof* The proof of the theorem largely repeats that of the similar assertion (Theorem 13.8.1) for sums of random variables defined on a Markov chain. Divide the domain  $u \geq 0$  into half-intervals

$$(0, T_0], \quad (T_{k-1}, T_k], \quad k \geq 1, \quad T_0 = \tau_0,$$

where  $T_k$  are the epochs of hitting the state  $x_0$  by the process  $X(t)$ ,  $\tau_k = T_k - T_{k-1}$  for  $k \geq 1$  are independent and distributed as  $\tau$ . Then the random variables

$$\zeta_k = \int_{T_{k-1}}^{T_k} f(X(u)) du, \quad k \geq 1$$

are independent, distributed as  $\zeta$ , and have finite expectation  $a_\zeta$ . The integral  $S(t)$  can be represented as

$$S(t) = z_0 + \sum_{k=1}^{v(t)} \zeta_k + z_t,$$

where

$$v(t) := \max\{k : T_k \leq t\}, \quad z_0 := \int_0^{T_0} f(X(u)) du, \quad z_t := \int_{T_{v(t)}}^t f(X(u)) du.$$

Since  $\tau_0$  is a proper random variable,  $z_0$  is a proper random variable as well, and hence  $z_0/t \xrightarrow{a.s.} 0$  as  $t \rightarrow \infty$ . Further,

$$z_t \stackrel{d}{=} \int_0^{\gamma(t)} f(X_0(u)) du,$$

where  $\gamma(t) = t - T_{v(t)}$  has a proper limiting distribution as  $t \rightarrow \infty$  (see Chap. 10), so  $z_t/t \xrightarrow{p} 0$  as  $t \rightarrow \infty$ . The sum  $S_{v(t)} = \sum_{k=1}^{v(t)} \zeta_k$  is nothing else but the generalised renewal process studied in Chaps. 10 and 11. By Theorem 11.5.2, as  $t \rightarrow \infty$ ,

$$\frac{S_{v(t)}}{t} \xrightarrow{p} \frac{a_\zeta}{a}.$$

The theorem is proved. □

In order to prove the strong law of large numbers we need a somewhat more restrictive condition than that in Theorem 21.5.2. Put

$$\zeta^* := \int_0^\tau |f(X_0(u))| du.$$

**Theorem 21.5.3** *Let the conditions of Theorem 21.5.1 be satisfied and  $\mathbf{E}\zeta^* < \infty$ . Then*

$$\frac{S(t)}{t} \xrightarrow{a.s.} \frac{a_\zeta}{a}.$$



The proof essentially repeats (as was the case for Theorem 21.5.2) that of the law of large numbers for sums of random variables defined on a Markov chain (see Theorem 13.8.3). One only needs to use, instead of (13.8.18), the relation

$$\sup_{T_k \leq u \leq T_{k+1}} \left| \int_{T_k}^u f(X(v)) dv \right| \leq \zeta_k^* = \int_{T_k}^{T_{k+1}} |f(X(v))| dv$$

and the fact that  $\mathbf{E} \zeta_k^* < \infty$ . The theorem is proved. □

Here an analogue of Theorem 13.8.2, in which the conditions of existence of  $\mathbf{E} \zeta^*$  and  $\mathbf{E} \zeta$  are elucidated, is the following.

**Theorem 21.5.4** (Generalisation of Wald’s identity) *Let the conditions of Theorem 21.5.1 be met and there exist*

$$\mathbf{E} |f(X(\infty))| = \int |f(x)| \pi(dx),$$

where  $X(\infty)$  is a random variable with the stationary distribution  $\pi$ . Then there exist

$$\mathbf{E} \zeta^* = \mathbf{E} \tau \mathbf{E} |f(X(\infty))|, \quad \mathbf{E} \zeta = \mathbf{E} \tau \mathbf{E} f(X(\infty)).$$

The proof of Theorem 21.5.4 repeats, with obvious changes, that of Theorem 13.8.2. □

**Theorem 21.5.5** (The central limit theorem) *Let the conditions of Theorem 21.5.1 be met and  $\mathbf{E} \tau^2 < \infty, \mathbf{E} \zeta^2 < \infty$ . Then*

$$\frac{S(t) - rt}{d\sqrt{t/a}} \Rightarrow \Phi_{0,1}, \quad t \rightarrow \infty,$$

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

The proof, as in the case of Theorems 21.5.2–21.5.4, repeats, up to evident changes, that of Theorem 13.8.4. □

Here an analogue of Theorem 13.8.5 (on the conditions of existence of variance and on an identity for  $a^{-1}d^2$ ) looks more complicated than under the conditions of Sect. 13.8 and is omitted.

## 21.6 Diffusion Processes

Now we will consider an important class of Markov processes with continuous trajectories.

**Definition 21.6.1** A homogeneous Markov process  $\xi(t)$  with state space  $(\mathbb{R}, \mathfrak{B})$  and the transition function  $P(t, x, B)$  is said to be a *diffusion process* if, for some finite functions  $a(x)$  and  $b^2(x) > 0$ ,

- (1)  $\lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int (y - x) P(\Delta, x, dy) = a(x),$
- (2)  $\lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int (y - x)^2 P(\Delta, x, dy) = b^2(x),$
- (3) for some  $\delta > 0$  and  $c < \infty,$

$$\int |y - x|^{2+\delta} P(\Delta, x, dy) < c\Delta^{1+\delta/2}.$$

Put  $\Delta\xi(t) := \xi(t + \Delta) - \xi(t).$  Then the above conditions can be written in the form:

$$\begin{aligned} \mathbf{E}[\Delta\xi(t) | \xi(t) = x] &\sim a(x)\Delta, \\ \mathbf{E}[(\Delta\xi(t))^2 | \xi(t) = x] &\sim b^2(x)\Delta, \\ \mathbf{E}[|\Delta\xi(t)|^{2+\delta} | \xi(t) = x] &< c\Delta^{1+\delta/2} \quad \text{as } \Delta \rightarrow 0. \end{aligned}$$

The coefficients  $a(x)$  and  $b(x)$  are called the *shift* and *diffusion coefficients*, respectively. Condition (3) is an analogue of the Lyapunov condition. It could be replaced with a Lindeberg type condition:

(3a)  $\mathbf{E}[(\Delta\xi(t))^2; |\Delta\xi(t)| > \varepsilon] = o(\Delta)$  for any  $\varepsilon > 0$  as  $\Delta \rightarrow 0.$

It follows immediately from condition (3) and the Kolmogorov theorem that a diffusion process  $\xi(t)$  can be thought of as a process with continuous trajectories.

The standard Wiener process  $w(t)$  is a diffusion process, since in that case

$$\begin{aligned} P(t; x, B) &= \frac{1}{\sqrt{2\pi t}} \int_B e^{-(x-y)^2/(2t)} dy, \\ \mathbf{E}\Delta w(t) &= 0, \quad \mathbf{E}[\Delta w(t)]^2 = \Delta, \quad \mathbf{E}[\Delta w(t)]^4 = 3\Delta^2. \end{aligned}$$

Therefore the Wiener process has zero shift and a constant diffusion coefficient. Clearly, the process  $w(t) + at$  will have shift  $a$  and the same diffusion coefficient.

We saw in Sect. 21.2 that the “local” characteristic  $Q$  of a Markov process  $\xi(t)$  with a discrete state space  $\mathcal{X}$  specifies uniquely the evolution law of the process. A similar situation takes place for diffusion processes: the *distribution of the process is determined uniquely by the coefficients  $a(x)$  and  $b(x)$ .* The way to establishing this fact again lies via the Chapman–Kolmogorov equation.

**Theorem 21.6.1** *If the transition probability  $P(t; x, B)$  of a diffusion process is twice continuously differentiable with respect to  $x$ , then  $P(t; x, B)$  is differentiable with respect to  $t$  and satisfies the equation*

$$\frac{\partial P}{\partial t} = a \frac{\partial P}{\partial x} + \frac{b^2}{2} \frac{\partial^2 P}{\partial x^2} \tag{21.6.1}$$

with the initial condition

$$P(0; x, B) = \mathbf{I}_B(x). \tag{21.6.2}$$

*Remark 21.6.1* The conditions of the theorem on smoothness of the transition function  $P$  can actually be proved under the assumption that  $a$  and  $b$  are continuous,  $b \geq b_0 > 0,$   $|a| \leq c(|x| + 1)$  and  $b^2 \leq c(|x| + 1).$

*Proof of Theorem 21.6.1* For brevity's sake denote by  $P'_t$ ,  $P'_x$ , and  $P''_x$  the partial derivatives  $\frac{\partial P}{\partial t}$ ,  $\frac{\partial P}{\partial x}$  and  $\frac{\partial^2 P}{\partial x^2}$ , respectively, and make use of the relation

$$\begin{aligned} P(t; y, B) - P(t; x, B) &= (y - x)P'_x + \frac{(y - x)^2}{2}P''_x + \frac{(y - x)^2}{2}[P''_x(t; y_x, B) - P''_x(t; x, B)], \\ y_x &\in (x, y). \end{aligned} \tag{21.6.3}$$

Then by the Chapman–Kolmogorov equation

$$\begin{aligned} P(t + \Delta; x, B) - P(t; x, B) &= \int P(\Delta; x, dy)[P(t; y, B) - P(t; x, B)] \\ &= a(x)P'_x\Delta + \frac{b^2(x)}{2}P''_x\Delta + o(\Delta) + R, \end{aligned} \tag{21.6.4}$$

where

$$R = \int \frac{(y - x)^2}{2}[P''_x(t; y_x, B) - P''_x(t; x, B)]P(\Delta; x, dy) = \int_{|y-x| \leq \varepsilon} + \int_{|y-x| > \varepsilon}.$$

The first integral, by virtue of the continuity of  $P''_x$ , does not exceed

$$\delta(\varepsilon) \left[ \frac{b^2(x)}{2} \Delta + o(\Delta) \right],$$

where  $\delta(\varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ ; the second integral is  $o(\Delta)$  by condition (3a). Since  $\varepsilon$  is arbitrary, one has  $R = o(\Delta)$  and it follows from the above that

$$P'_t = \lim_{\Delta \rightarrow 0} \frac{P(t + \Delta; x, B) - P(t; x, B)}{\Delta} = a(x)P'_x + \frac{b^2(x)}{2}P''_x.$$

This proves (21.6.1). The theorem is proved. □

It is known from the theory of differential equations that, under wide assumptions about the coefficients  $a$  and  $b$  and for  $B = (-\infty, z)$ , the Cauchy problem (21.6.1)–(21.6.2) has a unique solution  $P$  which is infinitely many times differentiable with respect to  $t$ ,  $x$  and  $z$ . From this it follows that  $P(t; x, B)$  has a density  $p(t; x, z)$  which is the fundamental solution of (21.6.1).

It is also not difficult to derive from Theorem 21.6.1 that, along with  $P(t; x, B)$ , the function

$$u(t, x) = \int g(z)P(t; x, dz) = \mathbf{E}[g(\xi^{(x)}(t))]$$

will also satisfy Eq. (21.6.1) for any smooth function  $g$  with a compact support,  $\xi^{(x)}(t)$  being the diffusion process with the initial value  $\xi^{(x)}(0) = x$ .

In the proof of Theorem 21.6.1 we considered (see (21.6.4)) the time increment  $\Delta$  preceding the main time interval. In this connection Eqs. (21.6.1) are called *backward* Kolmogorov equations. *Forward* equations can be derived in a similar way.

**Theorem 21.6.2** (Forward Kolmogorov equations) *Let the transition density  $p(t; x, y)$  be such that the derivatives*

$$\frac{\partial}{\partial y}[a(y)p(t; x, y)] \quad \text{and} \quad \frac{\partial^2}{\partial y^2}[b^2(y)p(t; x, y)]$$

*exist and are continuous. Then  $p(t, x, y)$  satisfies the equation*

$$Dp := \frac{\partial p}{\partial t} + \frac{\partial}{\partial y}[a(y)p(t; x, y)] - \frac{1}{2} \frac{\partial^2}{\partial y^2}[b^2(y)p(t; x, y)] = 0. \quad (21.6.5)$$

*Proof* Let  $g(y)$  be a smooth function with a bounded support,

$$u(t, x) := \mathbf{E}g(\xi^{(x)}(t)) = \int g(y)p(x; t, y) dy.$$

Then

$$\begin{aligned} &u(t + \Delta, x) - u(t, x) \\ &= \int p(t; x, z) \left[ p(\Delta; z, y)g(y) dy - \int p(\Delta, z, y)g(z) dy \right] dz. \end{aligned} \quad (21.6.6)$$

Expanding the difference  $g(y) - g(z)$  into a series, we obtain in the same way as in the proof of Theorem 21.4.1 that, by virtue of properties (1)–(3), the expression in the brackets is

$$\left[ a(z)g'(z) + \frac{b^2(z)}{2}g''(z) \right] \Delta + o(\Delta).$$

This implies that there exists the derivative

$$\frac{\partial u}{\partial t} = \int p(t; x, z) \left[ a(z)g'(z) dz + \frac{1}{2} \frac{b^2(z)}{2}g''(z) \right] dz.$$

Integrating by parts we get

$$\frac{\partial u}{\partial t} = \int \left\{ -\frac{\partial}{\partial z}[a(z)p(t; x, z)] + \frac{1}{2} \frac{\partial}{\partial z^2}[b^2(z)p(t; x, z)] \right\} g(z) dz = 0$$

or, which is the same,

$$\int Dp(t; x, z)g(z) dz = 0.$$

Since  $g$  is arbitrary, (21.6.5) follows. The theorem is proved. □

As in the case of discrete  $\mathcal{X}$ , the difference between the forward and backward Kolmogorov equations becomes more graphical for non-homogeneous diffusion processes, when the transition probabilities  $P(s, x; t, B)$  depend on two time variables, while  $a$  and  $b$  in conditions (1)–(3) are functions of  $s$  and  $x$ . Then the backward Kolmogorov equation (for densities) will relate the derivatives of the transition densities  $p(s, x; t, y)$  with respect to the first two variables, while the forward equation will hold for the derivatives with respect to the last two variables.

We return to homogeneous diffusion processes. One can study conditions ensuring the existence of the limiting stationary distribution of  $\xi^{(x)}(t)$  as  $t \rightarrow \infty$  which is independent of  $x$  using the same approach as in Sect. 21.2. Theorem 21.2.3 will remain valid (one simply has to replace  $i_0$  in it with  $x_0$ , in agreement with the notation of the present section). The proof of Theorem 21.2.3 also remains valid, but will need a somewhat more precise argument (in the new situation, on the event  $B_{dv}$  one has  $\xi(t - v) \in dx_0$  instead of  $\xi(t - v) = x_0$ ).

If the stationary distribution density

$$\lim_{t \rightarrow \infty} p(t; x, y) = p(y) \tag{21.6.7}$$

exists, how could one find it? Since the dependence of  $p(t; x, y)$  of  $t$  and  $x$  vanishes as  $t \rightarrow \infty$ , the backward Kolmogorov equations turn into the identity  $0 = 0$  as  $t \rightarrow \infty$ . Turning to the forward equations and passing in (21.6.6) to the limit first as  $t \rightarrow \infty$  and then as  $\Delta \rightarrow 0$ , we come, using the same argument as in the proof of Theorem 21.2.3, to the following conclusion.

**Corollary 21.6.1** *If (21.6.7) and the conditions of Theorem 21.6.2 hold, then the stationary density  $p(y)$  satisfies the equation*

$$-[a(y)p(y)]' + \frac{1}{2}[b^2(y)p(y)]'' = 0$$

(which is obtained from (21.6.5) if we put  $\frac{\partial p}{\partial t} = 0$ ).

*Example 21.6.1* The Ornstein–Uhlenbeck process

$$\xi^{(x)}(t) = xe^{at} + \sigma e^{at} w\left(\frac{1 - e^{-2at}}{2a}\right),$$

where  $w(u)$  is the standard Wiener process, is a homogeneous diffusion process with the transition density

$$p(t; x, y) = \frac{1}{\sqrt{2\pi}\sigma(t)} \exp\left\{-\frac{(y - xe^{at})^2}{2\sigma^2(t)}\right\}, \quad \sigma^2(t) = \frac{\sigma^2}{2a}(e^{2at} - 1). \tag{21.6.8}$$

We leave it to the reader to verify that this process has coefficients  $a(x) = ax$ ,  $b(x) = \sigma = \text{const}$ , and that function (21.6.8) satisfies the forward and backward equations. For  $a < 0$ , there exists a stationary process (the definition is given in the next chapter)

$$\xi(t) = \sigma e^{at} w\left(\frac{e^{-2at}}{2a}\right),$$

of which the density (which does not depend on  $t$ ) is equal to

$$p(y) = \lim_{t \rightarrow \infty} p(x; t, y) = \frac{1}{\sqrt{2\pi}\sigma(\infty)} \exp\left\{-\frac{y^2}{2\sigma^2(\infty)}\right\}, \quad \sigma(\infty) = -\frac{\sigma^2}{2a}.$$

In conclusion of this section we will consider the problem, important for various applications, of finding the probability that the trajectory of a diffusion process will not leave a given strip. For simplicity's sake we confine ourselves to considering this problem for the Wiener process. Let  $c > 0$  and  $d < 0$ .

Put

$$\begin{aligned}
 U(t; x, B) &:= \mathbf{P}(w^{(x)}(u) \in (d, c) \text{ for all } u \in [0, t]; w^{(x)}(t) \in B) \\
 &= \mathbf{P}\left(\sup_{u \leq t} w^{(x)}(u) < c, \inf_{u \leq t} w^{(x)}(u) > d, w^{(x)}(t) \in B\right).
 \end{aligned}$$

Leaving out the verification of the fact that the function  $U$  is twice continuously differentiable, we will only prove the following proposition.

**Theorem 21.6.3** *The function  $U$  satisfies Eq. (21.6.1) with the initial condition*

$$U(0; x, B) = \mathbf{I}_B(x) \tag{21.6.9}$$

and boundary conditions

$$U(t; c, B) = U(t; d, B) = 0. \tag{21.6.10}$$

*Proof* First of all note that the function  $U(t; x, B)$  for  $x \in (d, c)$  satisfies conditions (1)–(3) imposed on the transition function  $P(t; x, B)$ . Indeed, consider, for instance, property (1).

We have to verify that

$$\int_d^c (y - x)U(\Delta; x, dy) = \Delta a(x) + o(\Delta) \tag{21.6.11}$$

(with  $a(x) = 0$  in our case). But  $U(t, x, B) = P(t; x, B) - V(t; x, B)$ , where

$$V(t; x, B) = \mathbf{P}\left(\left\{\sup_{u \leq t} w^{(x)}(u) \geq c \text{ or } \inf_{u \leq t} w^{(x)}(u) \leq d\right\} \cap \{w^{(x)}(t) \in B\}\right),$$

and

$$\begin{aligned}
 &\int_d^c (y - x)V(\Delta; x, dy) \\
 &\leq \max(c, -d)\left[\mathbf{P}\left(\sup_{u \leq \Delta} w^{(x)}(u) \geq c\right) + \mathbf{P}\left(\inf_{u \leq \Delta} w^{(x)}(u) \leq d\right)\right].
 \end{aligned}$$

The first probability in the brackets is given, as we know (see (20.2.1) and Theorem 19.2.2), by the value

$$2\mathbf{P}(w^{(x)}(\Delta) > c) = 2\mathbf{P}\left(w(1) > \frac{c - x}{\sqrt{\Delta}}\right) \sim \frac{2}{\sqrt{2\pi z}} e^{-z^2/2}, \quad z = \frac{c - x}{\sqrt{\Delta}}.$$

For any  $x < c$  and  $k > 0$ , it is  $o(\Delta^k)$ . The same holds for the second probability. Therefore (21.6.11) is proved. In the same way one can verify properties (2) and (3).

Further, because by the total probability formula, for  $x \in (d, c)$ ,

$$U(t + \Delta; x, B) = \int_d^c U(\Delta; x, dy)U(t; y, B),$$

using an expansion of the form (21.6.3) for the function  $U$ , we obtain in the same way as in (21.6.4) that

$$\begin{aligned} U(t + \Delta; x, B) - U(t; x, B) &= \int U(\Delta; x, dy) [U(t; y, B) - U(t; x, B)] \\ &= a(x) \frac{\partial U}{\partial x} \Delta + \frac{b^2(x)}{2} \frac{\partial^2 U}{\partial x^2} \Delta + o(\Delta). \end{aligned}$$

This implies that  $\frac{\partial U}{\partial t}$  exists and that Eq. (21.6.1) holds for the function  $U$ .

That the boundary and initial conditions are met is obvious. The theorem is proved.  $\square$

The reader can verify that the function

$$u(t; x, y) := \frac{\partial}{\partial y} U(t; x, (-\infty, y)), \quad y \in (d, c),$$

playing the role of the fundamental solution to the boundary problem (21.6.9)–(21.6.10) (the function  $u$  satisfies (21.6.1) with the boundary conditions (21.6.10) and the initial conditions degenerating into the  $\delta$ -function), is equal to

$$\begin{aligned} u(t; x, y) &= \frac{1}{\sqrt{2\pi t}} \left[ \sum_{k=-\infty}^{\infty} \exp\left\{-\frac{[y + 2k(c-d)]^2}{2t}\right\} \right. \\ &\quad - \sum_{k=0}^{\infty} \exp\left\{-\frac{[y - 2c - 2k(c-d)]^2}{2t}\right\} \\ &\quad \left. - \sum_{k=0}^{\infty} \exp\left\{-\frac{[y - 2d - 2k(c-d)]^2}{2t}\right\} \right]. \end{aligned}$$

This expression can also be obtained directly from probabilistic considerations (see, e.g., [32]).