CHAPTER 1

Markov processes

1. General theory

We start with some basic definitions and notation, which will be used in our later considerations. Let T be the time set: in general, it is an arbitrary partially ordered set, but throughout we will study two important examples:

· $T = \mathbb{Z}_+ = 0, 1, 2, \dots$ (discrete time);

· $T = \mathbb{R}_+ = [0, \infty)$, $T = \mathbb{R}$ or $T = [a, b]$ (continuous time).

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a fixed probability space and let $(E, \mathcal{E})$ be a given measure space.

Definition 1.1. *A stochastic process* with state space $(E, \mathcal{E})$ is a collection $X = (X_t)_{t\in T}$ of random variables $X_t : \Omega \to E$. *A filtration generated by* X is the family $(\mathcal{F}_t^X)_{t\in T} = \big(\sigma(X_s \,:\, s \in T,\, s \le t)\big)_{t\in T}$ of sub-σ-fields of $\mathcal{F}$. For each $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is called a *trajectory* (or *sample-path*, or *realization*) of X.

Definition 1.2. Let $X = (X_t)_{t\in T}$ be a stochastic process taking values in a measurable space $(E, \mathcal{E})$. We say that X is *a Markov process* if for all $t \in T$ and $h > 0$ such that $t + h \in T$, and for all $\Gamma \in \mathcal{E}$, we have

$$\mathbb{P}(X_{t+h} \in \Gamma \mid \mathcal{F}_t^X) = \mathbb{P}(X_{t+h} \in \Gamma \mid X_t) \quad \text{a.s.}$$

Equivalently, for all $t \in T$, $h > 0$ such that $t + h \in T$, and all measurable functions $f : E \to \mathbb{R}$ such that $\mathbb{E}|f(X_{t+h})| < \infty$, we have

$$\mathbb{E}\big(f(X_{t+h}) \mid \mathcal{F}_t^X\big) = \mathbb{E}\big(f(X_{t+h}) \mid X_t\big) \quad \text{a.s.}$$

There are many equivalent formulations of Markov property: let us present here three of them. The proof is left as an easy exercise.

Proposition 1.3. *Let $X = (X_t)_{t\in T}$ be a stochastic process taking values in a measurable space $(E, \mathcal{E})$. Then X is a Markov process if and only if one of the following three conditions holds.*

(i) For all $t \in T$, $A \in \mathcal{F}_{\ge t}^X = \sigma(X_s \,:\, s \in T,\, s \ge t)$ and $B \in \mathcal{F}_t^X$ we have

$$\mathbb{P}(A \cap B \mid X_t) = \mathbb{P}(A \mid X_t)\,\mathbb{P}(B \mid X_t) \quad \text{a.s.}$$

(ii) For all $t \in T$ and $A \in \mathcal{F}_{\ge t}^X$, we have

$$\mathbb{P}(A \mid \mathcal{F}_t^X) = \mathbb{P}(A \mid X_t) \quad \text{a.s.}$$

(iii) For all $t \in T$, $h > 0$ such that $t + h \in T$, and $s_1 < s_2 < \dots < s_n = t$ with $s_j \in T$, we have

$$\mathbb{P}(X_{t+h} \in \Gamma \mid X_{s_1}, \dots, X_{s_n}) = \mathbb{P}(X_{t+h} \in \Gamma \mid X_t) \quad \text{a.s.}$$

Here is a more general definition of a Markov process relative to a filtration.

Definition 1.4. Let $X = (X_t)_{t\in T}$ be a stochastic process with values in $(E, \mathcal{E})$ and let $(\mathcal{F}_t)_{t\in T}$ be a filtration. Then X is *a Markov process with respect to* $(\mathcal{F}_t)_{t\in T}$, if

· the process $(X_t)_{t\in T}$ is $(\mathcal{F}_t)_{t\in T}$-adapted (i.e., for each t, X_t is $\mathcal{F}_t$-measurable),
· for all $t \in T$ and $h > 0$ such that $t + h \in T$, and for all $\Gamma \in \mathcal{E}$, we have

$$\mathbb{P}(X_{t+h} \in \Gamma \mid \mathcal{F}_t) = \mathbb{P}(X_{t+h} \in \Gamma \mid X_t) \quad \text{a.s.}$$

One verifies easily that if X is a Markov process relative to some $(\mathcal{F}_t)_{t\in T}$, then it is automatically Markovian with respect to its natural filtration (i.e., satisfies the condition in Definition 1.2).

Markov processes can be classified according to the properties of the nature of time and the properties of their state space. We distinguish the following classes:

(i) discrete time, finite state space (discrete-space Markov chains);
(ii) discrete time, countable state space (discrete-space Markov chains);
(iii) discrete time, general state space (Markov chains);
(iv) continuous time, countable state space (continuous Markov chains);
(v) continuous time, general state space (Markov process).

The case (i) is elementary and can be studied by means of an elementary linear algebra. The case (ii) is much more interesting, and introduces new concepts such as recurrence and transience. The case (iii) is a further complication (but not to a much extent), there are new issues arising due to ergodicity problems. The case (iv) is close to the case (ii): many continuous Markov chains are obtained from discrete time, discrete case Markov processes where each unit of time is replaced by an exponentially distributed random time, whose parameter depends on the position in space. However, it should be emphasized that fundamentally new issues can arise if these parameters are unbounded. The case (v) is really new, and poses challenging new problems that require some serious tools from functional analysis. A key new problem here is how to describe such a process in simple terms.

We return to the description of the background. The next foundational concept is the following.

Definition 1.5. A family of functions $(P_{s,t}(x,\Gamma))_{s,t\in T, s\le t, x\in E, \Gamma\in\mathcal{E}}$ is called *a Markovian transition function* (or *Markovian transition kernel*) if the following four properties hold:

(i) For all $s, t \in T$, $s \le t$, and $x \in E$, $P_{s,t}(x,\cdot)$ is a measure on $(E, \mathcal{E})$ with $P_{s,t}(x, E) \le 1$;

(ii) For all $s, t \in T$, $s \le t$, and $\Gamma \in \mathcal{E}$, the mapping $x \mapsto P_{s,t}(x,\Gamma)$ is measurable from $(E, \mathcal{E})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$;

(iii) For all $s \in T$, all $x \in E$, and all $\Gamma \in \mathcal{E}$,

$$P_{s,s}(x,\Gamma) = \delta_x(\Gamma),$$

where δ_x is the Dirac measure at x.

(iv) For all $s, t, u \in T$, $s < t < u$, and all $x \in E$, $\Gamma \in \mathcal{E}$,

$$P_{s,u}(x,\Gamma) = \int_E P_{t,u}(y,\Gamma)\, P_{s,t}(x, dy).$$

The condition (iv) is referred to as *the Chapman-Kolmogorov equation*.

Definition 1.6. Let $(P_{s,t}(\cdot,\cdot))$ be a Markovian transition function. It is *a transition function for a Markov process* $X = (X_t)_{t\in T}$, if for all $s,\, t \in T,\, s \le t$ and $\Gamma \in \mathcal{E}$,

$$\mathbb{P}(X_t \in \Gamma \mid X_s) = P_{s,t}(X_s, \Gamma) \quad \text{a.s.} \tag{1.1}$$

Equivalently, for all $s,\, t \in T,\, s \le t$ and all $f : E \to \mathbb{R}$ satisfying $\mathbb{E}|f(X_t)| < \infty$,

$$\mathbb{E}\big[f(X_t) \mid X_s\big] = \int_E f(x) P_{s,t}(X_s, \mathrm{d}x) \quad \text{a.s.} \tag{1.2}$$

Let us distinguish two special cases.

I. Suppose that E is at most countable (i.e., X is a continuous Markov chain). Then the transition function is uniquely determined by the transition matrices

$$(P_{s,t})_{s\le t} = (p_{ij}(s,t))_{s\le t,\, x,\, y\in E}, \qquad p_{ij}(s,t) = P_{s,t}(i, \{j\}).$$

Specifically, we have $P_{s,t}(x,\Gamma) = \sum_{y\in\Gamma} p_{xy}(s,t)$. The Chapman-Kolmogorov equation is equivalent to saying that for all $s,\, t,\, u$ as previously and $x,\, z \in E$,

$$p_{xz}(s,u) = \sum_{y\in E} p_{xy}(s,t) p_{yz}(t,u).$$

That is, in the language of matrices, we have $P_{s,u} = P_{s,t}P_{t,u}$.

II. Let $(E,\mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ for some $d \ge 1$. A family $(p_{s,t}(\cdot,\cdot))_{s,\, t\in T,\, s<t}$ is a transition density, if for all $s,\, t \in T,\, s < t$, and any $\Gamma \in \mathcal{E}$ we have

$$P_{s,t}(x,\Gamma) = \int_\Gamma p_{s,t}(x,y)\mathrm{d}y.$$

In this setting, the Chapman-Kolmogorov equation is equivalent to saying that for all $s,\, t,\, u$ as before and all $x \in E$ we have

$$p_{s,u}(x,z) = \int_E p_{s,t}(x,y) p_{t,u}(y,z)\mathrm{d}y \qquad \text{for almost all } z \in E.$$

Definition 1.7. We say that a transition function is homogeneous if $P_{s,t} = P_{s+h,t+h}$ for all $s,\, t,\, h$. In this case, $P_{s,t}(x,\Gamma) = P_{t-s}(x,\Gamma)$, where the function $P_t(x,\Gamma)$, $t \in T, x \in E, \Gamma \in \mathcal{E}$ satisfies the following conditions:

(i) For all $t \in T$ and $x \in E$, $P_t(x,\cdot)$ is a measure on $(E,\mathcal{E})$, with $P_t(x,E) \le 1$;

(ii) For all $t \in T$ and $\Gamma \in \mathcal{E}$, the mapping $x \mapsto P_t(x,\Gamma)$ is measurable from $(E,\mathcal{E})$ to $(\mathbb{R},\mathcal{B}(\mathbb{R}))$;

(iii) For all $x \in E$ and $\Gamma \in \mathcal{E}$, we have $P_0(x,\Gamma) = \delta_x(\Gamma)$.

(iv) For all $s,\, t \in T$ and $x \in E,\, \Gamma \in \mathcal{E}$, we have

$$P_{s+t}(x,\Gamma) = \int_E P_s(y,\Gamma)\, P_t(x,dy).$$

A Markov process is called *homogeneous* if it has a homogeneous transition function.

Definition 1.8. Suppose that $0 \in T \subseteq [0,\infty)$. Then *the initial distribution* of a process $X = (X_t)_{t\in T}$ is the law of X_0.

2. Continuous time Markov chains

2.1. Definition, examples. Throughout this section, we assume that $(X_t)_{t\geq 0}$ is a *time-homogeneous* continuous Markov chain on some countable space E. In this setup, the Markov property can be equivalently restated as follows: for any $0 \leq t_0 < t_1 < t_2 < \ldots < t_n$ and any $i_0, i_1, i_2, \ldots, i_n \in E$,

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

The transition kernel $P_t = (p_{ij}(t))_{i,\, j\in E}$, $t \geq 0$, satisfies the following properties:

(i) For all $t \geq 0$ and any $i \in E$ we have $\sum_{j\in E} p_{ij}(t) \leq 1$.

(ii) For all $i, j \in E$ we have $p_{ij}(0) = \delta_{i,j}$. (That is, $P_0 = I$.)

(iii) For all $s,\, t \geq 0$ and any $i,\, j \in E$, we have

$$p_{ij}(s + t) = \sum_{k\in E} p_{ik}(s)p_{kj}(t).$$

(That is, we have the semigroup property $P_{s+t} = P_s P_t$.)

The distribution of the variable X_t can be identified with the column vector $\mu(t) = \{\mu_i(t)\}_{i\in E}$, where $\mu_i(t) = \mathbb{P}(X_t = i)$. It is obtained from the initial distribution $\mu(0)$ by

$$\mathbb{P}(X_t = j) = \sum_{i\in E} \mathbb{P}(X_t = j|X_0 = i)\mathbb{P}(X_0 = i) = \sum_{i\in E} \mu_i(0)p_{ij}(t),$$

or $\mu(t)^T = \mu(0)^T P_t$ in vectorial notation. Furthermore, for all $0 = t_0 \leq t_1 < t_2 < \ldots < t_k$ and all states $i_1,\, i_2,\, \ldots,\, i_k$ we have

$$\mathbb{P}(X_{t_1} = i_1,\, X_{t_2} = i_2,\, \ldots,\, X_{t_k} = i_k) = \sum_{i_0\in E} \mu_{i_0}(0) \cdot \prod_{j=1}^{k} p_{i_{j-1}i_j}(t_j - t_{j-1}).$$

Example 2.1. Let $\lambda > 0$ be a positive parameter. A *Poisson process with intensity* λ is a process $N = (N_t)_{t\geq 0}$ satisfying the following requirements:

1° $N_0 = 0$ almost surely;

2° N has independent increments: for any $t_0 < t_1 < t_2 < \ldots < t_n$, the random variables N_{t_0}, $N_{t_1} - N_{t_0}$, $N_{t_2} - N_{t_1}$, …, $N_{t_n} - N_{t_{n-1}}$ are independent.

3° for any $0 \leq s < t$, the variable $N_t - N_s$ has the distribution $\mathrm{Pois}(\lambda(t - s))$.

4° almost surely, N has right-continuous trajectories with left limits.

Formally, the above definition works also for $\lambda = 0$ (then we have $N \equiv 0$). Poisson process is a homogeneous Markov chain. Indeed, for any $0 \leq t_0 < t_1 < t_2 < \ldots < t_n$ and any states $i_0,\, i_1,\, i_2,\, \ldots,\, i_n$ we have, by the independence of increments,

$$\begin{aligned}&\mathbb{P}\Big(N_{t_n} = i_n \mid N_{t_{n-1}} = i_{n-1}, N_{t_{n-2}} = i_{n-2},\, \ldots,\, N_{t_0} = i_0\Big) \\ &= \mathbb{P}\Big(N_{t_n} - N_{t_{n-1}} = i_n - i_{n-1} \mid N_{t_{n-1}} = i_{n-1}, N_{t_{n-2}} = i_{n-2},\, \ldots,\, N_{t_0} = i_0\Big) \\ &= \mathbb{P}\Big(N_{t_n} - N_{t_{n-1}} = i_n - i_{n-1}\Big) \\ &= \mathbb{P}\Big(N_{t_n} - N_{t_{n-1}} = i_n - i_{n-1} \mid N_{t_{n-1}} = i_{n-1}\Big) = \mathbb{P}\Big(N_{t_n} = i_n \mid N_{t_{n-1}} = i_{n-1}\Big).\end{aligned}$$

Furthermore, the above verification immediately gives the transition probabilities

$$p_{t_{n-1}t_n}(i_{n-1}, i_n) = \mathbb{P}(N_{t_n} - N_{t_{n-1}} = i_n - i_{n-1}) = e^{-\lambda(t_n - t_{n-1})} \cdot \frac{(\lambda(t_n - t_{n-1}))^{i_n - i_{n-1}}}{(i_n - i_{n-1})!}.$$

Since the dependence on t_{n-1} and t_n is through $t_n - t_{n-1}$, N is homogeneous and

$$p_{ij}(t) = \begin{cases} e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!} & \text{if } j \geq i, \\ 0 & \text{if } j < i. \end{cases}$$

Remark 2.2. There is an alternative definition, which is actually more natural from the viewpoint of our further considerations. Let ξ_1, ξ_2, …, be a sequence of i.i.d. random variables with expontial distribution $\mathrm{Exp}(\lambda)$. Let $T_0 = 0$, $T_n = \sum_{k=1}^{n} \xi_k$, $n = 1, 2, \ldots$, be the associated *clock process*. Then

$$N_t = \sum_{n=1}^{\infty} \mathbb{1}_{\{T_n \leq t\}}, \qquad t \geq 0,$$

is a Poisson process with intensity λ.

Example 2.3. The example above can be generalized to the setting in which the states 0, 1, 2, … are assigned with different intensities. Suppose that ξ_1, ξ_2, … are independent random variables such that $\xi_n \sim \mathrm{Exp}(\lambda_n)$, where $(\lambda_n)_{n\geq 1}$ is a *bounded* sequence of nonnegative numbers. Let $(T_n)_{n\geq 0}$ and $N = (N_t)_{t\geq 0}$ be as in Remark 2.2 above. Then N is a homogeneous continuous Markov chain. One can consider also unbounded sequences $(\lambda_n)_{n\geq 1}$, but this might result in an explosion of the process N: see below.

Example 2.4. Let $\{Y_n\}_{n\geq 0}$ be a discrete-time homogeneous Markov chain with countable state space E and transition matrix Γ. Let N be an independent process as in Example 2.3 above. The process $\{X(t)\}_{t\geq 0}$ with values in E, defined by

$$X(t) = Y_{N(t)}, \tag{1.3}$$

is a homogeneous continuous Markov chain. The Poisson process N is called the *clock*, and the chain $\{Y_n\}_{n\geq 0}$ is called the *subordinated chain* or *embedded chain*. In the case when N is a Poisson process (i.e., comes from Example 2.1), X is referred to as the *uniform Markov chain*.

2.2. Infinitesimal generator.

We need to introduce further notions.

Definition 2.5. A transition function P is called

- *honest (regular)*, if equality holds in (i): $\sum_{j\in E} p_{ij}(t) = 1$ for all $i \in E$, $t \geq 0$.
- *standard*, if we have $\lim_{t\to 0} P_t = I$ entrywise: for any i, $j \in E$, $\lim_{t\to 0} p_{ij}(t) = \delta_{ij}$.

Remark 2.6. There are non-standard transition kernels, e.g. let $E = \{1, 2\}$,

$$P_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad P_t = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad \text{for } t > 0.$$

From now on, we will restrict ourselves to standard transition kernels, which, as we will show now, have nice differentiability properties. Let us first record, without a proof, the following elementary fact.

Lemma 2.7. *Let $\varphi : (0,\infty) \to [0,\infty)$ be a function satisfying*

- $\varphi(s+t) \le \varphi(s) + \varphi(t)$ *for all* $s, t > 0$;
- $\lim_{t\to 0} \varphi(t) = 0$.

Then the limit $\lim_{t\downarrow 0} \frac{\varphi(t)}{t} = q$ *exists (it may be* $+\infty$*). Furthermore,* $q = \sup_{t>0} \frac{\varphi(t)}{t}$.

Equipped with the above lemma, we can now establish the following key result.

Theorem 2.8. *Let P be a standard transition kernel. For any state i, the limit*

$$q_i \overset{\text{def}}{=} \lim_{h\downarrow 0} \frac{1 - p_{ii}(h)}{h} \in [0,\infty], \tag{2.12}$$

exists. Furthermore, for any pair of distinct states i, j, there exists

$$q_{ij} \overset{\text{def}}{=} \lim_{h\downarrow 0} \frac{p_{ij}(h)}{h} \in [0,\infty). \tag{1.4}$$

That is, the function $t \mapsto p_{ij}(t)$ is differentiable at $t = 0$ and we have $p'_{ij}(0) = q_{ij}$ (where $q_{ii} = -q_i$).

Proof. It is convenient to split the reasoning into a few steps.

Step 1. First we show that for any $i \in E$ and any $t > 0$ we have $p_{ii}(t) > 0$. To this end, note that we can choose n so that $p_{ii}(t/n) > 0$ (since P is standard). It remains to note that $p_{ii}(t) \ge [p_{ii}(t/n)]^n$, by the Markov property.

Step 2. Define $\varphi(t) = -\log p_{ii}(t)$. By the previous step, φ is well-defined; furthermore, since $p_{ii}(s+t) \ge p_{ii}(s)p_{ii}(t)$ and $p_{ii}(t) \to 1$ as $t \to 0$, the function φ satisfies the hypotheses of the previous lemma. Thus, the limit

$$q_i = \lim_{t\downarrow 0} \frac{\varphi(t)}{t}$$

exists and equals $\sup_{t>0} \varphi(t)/t$. Now, if $q_i = 0$, then $\varphi(t)/t = 0$ for all $t > 0$, so $p_{ii}(t) = 1$ for all $t \ge 0$, and hence

$$\lim_{t\to 0} \frac{1 - p_{ii}(t)}{t} = 0 = q_i,$$

and we are finished. Otherwise, suppose that $q_i > 0$. By continuity of φ, there is $\delta > 0$ such that $\varphi(t) > 0$ for $t \in (0,\delta)$. Then

$$\lim_{t\to 0} \frac{1 - p_{ii}(t)}{t} = \lim_{t\to 0} \frac{1 - e^{-\varphi(t)}}{\varphi(t)} \cdot \lim_{t\to 0} \frac{\varphi(t)}{t} = 1 \cdot q_i = q_i.$$

Step 3. Now, fix arbitrary distinct states i, j. Take $c \in (\frac{1}{2}, 1)$ and $\delta > 0$ such that $p_{ii}(t), p_{jj}(t) > c$ for all $t \in [0,\delta)$. Finally, pick $h > 0$ and $n \in \mathbb{N}$ such that $0 < nh < \delta$. We will establish the auxiliary bound

$$p_{ij}(nh) \ge c(2c-1) \cdot np_{ij}(h). \tag{1.5}$$

To this end, consider finite discrete chain $Y_k = X_{kh}$, $k = 0, 1, 2, \dots, n$ and let

$$A_k = \Big\{Y_0 = i,\, Y_1 \neq j,\, Y_2 \neq j, \dots, Y_{k-2} \neq j,\, Y_{k-1} = i,\, Y_k = j,\, Y_n = j\Big\},$$
$$B_k = \Big\{Y_0 = i,\, Y_1 \neq j,\, Y_2 \neq j, \dots, Y_{k-2} \neq j,\, Y_{k-1} = i\Big\}$$

for $k = 1, 2, \dots, n$. Then $A_1, A_2, \dots, A_n$ are pairwise disjoint and $\bigcup_{k=1}^n A_k \subseteq \{X_0 = i,\, X_{nh} = j\}$. Furthermore, we have $\mathbb{P}(A_k) = \mathbb{P}(B_k) p_{ij}(h) p_{jj}((n-k)h) \geq c p_{ij}(h) \mathbb{P}(B_k)$ by Markov property. Consequently,

$$p_{ij}(nh) = \mathbb{P}(Y_n = j \mid Y_0 = i) \geq \sum_{k=1}^n \mathbb{P}(A_k | Y_0 = i) \geq c p_{ij}(h) \sum_{k=1}^n \mathbb{P}(B_k | Y_0 = i). \tag{1.6}$$

Now we write

$$\begin{aligned}\mathbb{P}(B_k|Y_0 = i) &= \mathbb{P}(Y_k = i | Y_0 = i) \\ &\quad - \sum_{\ell=1}^{k-1} \mathbb{P}(Y_0 = i, Y_1 \neq j, Y_2 \neq j, \dots, Y_{\ell-1} \neq j, Y_\ell = j,\, Y_k = i | Y_0 = i)\end{aligned}$$

and observe that

$$\begin{aligned}&\mathbb{P}\Big(Y_0 = i, Y_1 \neq j, Y_2 \neq j, \dots, Y_{\ell-1} \neq j, Y_\ell = j,\, Y_k = i | Y_0 = i\Big) \\ &= \mathbb{P}\Big(Y_0 = i, Y_1 \neq j, Y_2 \neq j, \dots, Y_{\ell-1} \neq j, Y_\ell = j | Y_0 = i\Big) \mathbb{P}(Y_k = i | Y_\ell = j) \\ &\leq (1-c) \mathbb{P}\Big(Y_0 = i, Y_1 \neq j, Y_2 \neq j, \dots, Y_{\ell-1} \neq j, Y_\ell = j | Y_0 = i\Big),\end{aligned}$$

since $\mathbb{P}(Y_k = i | Y_\ell = j) \leq 1 - \mathbb{P}(Y_k = j | Y_\ell = j) \leq 1 - c$. This gives

$$\begin{aligned}\mathbb{P}(B_k|Y_0 = i) &= \mathbb{P}(Y_k = i | Y_0 = i) \\ &\geq c - (1-c) \sum_{\ell=1}^{k-1} \mathbb{P}\Big(Y_0 = i, Y_1 \neq j, Y_2 \neq j, \dots, Y_{\ell-1} \neq j, Y_\ell = j | Y_0 = i\Big) \\ &\geq 2c - 1.\end{aligned}$$

Plugging this into (1.6), we get (1.5).

Step 4. Finally, fix $t \in (0, \delta)$ and let $n = \lfloor t/h \rfloor$. The bound (1.5) can be rewritten in the form

$$\frac{p_{ij}(h)}{h} \leq \frac{1}{c(2c-1)} \cdot \frac{p_{ij}(\lfloor t/h \rfloor h)}{\lfloor t/h \rfloor h}.$$

Letting $h \to 0$, we obtain

$$\limsup_{h \downarrow 0} \frac{p_{ij}(h)}{h} \leq \frac{1}{c(2c-1)} \frac{p_{ij}(t)}{t},$$

in particular, the upper limit is finite. Since t was chosen arbitrarily, this yields

$$\limsup_{h \downarrow 0} \frac{p_{ij}(h)}{h} \leq \frac{1}{c(2c-1)} \liminf_{t \downarrow 0} \frac{p_{ij}(t)}{t} < \infty.$$

But c can be chosen arbitrarily close to 1; this proves that the upper and the lower limits coincide (and are finite). This is the claim. □

Recall the notation $q_{ii} = -q_i$.

Definition 2.9. The matrix $Q = (q_{ij})_{i,j\in E}$ is called the *q-matrix* (or *infinitesimal matrix, infinitesimal generator*) of P. In compact notation, we have

$$Q = \lim_{h\downarrow 0} \frac{P_h - I}{h}$$

entrywise; that is, Q is the derivative at zero of the matrix function $t \mapsto P_t$.

Example 2.10. The q-matrix of the Poisson process with intensity λ is given by

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & 0 & \dots \\ 0 & 0 & -\lambda & \lambda & 0 & \dots \\ 0 & 0 & 0 & -\lambda & \lambda & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

For the more general process from Example 2.3, we have

$$Q = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & 0 & \dots \\ 0 & -\lambda_2 & \lambda_2 & 0 & 0 & \dots \\ 0 & 0 & -\lambda_3 & \lambda_3 & 0 & \dots \\ 0 & 0 & 0 & -\lambda_4 & \lambda_4 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

Definition 2.11. A state $i \in E$ is said to be

- *stable*, if $q_i < +\infty$;
- *instantaneous*, if $q_i = +\infty$;
- *absorbing* if $q_i = 0$ (equivalently, if $p_{ii}(t) = 1$ for all $t \geq 0$).
- *conservative* if $q_i = \sum_{j\neq i} q_{ij}$.

The transition function $P = (P_t)_{t\geq 0}$, and the matrix Q, are called *stable* if all states $i \in E$ are stable; they are called *conservative*, if all states are conservative.

The reason for the introduction of the conservation property comes from the following result.

Proposition 2.12. *Let P be a standard transition function. Then for any $i \in E$,*

$$\sum_{j\neq i} q_{ij} \leq q_i.$$

Proof. Fix $i \in E$ and let A be a finite subset of $E \setminus \{i\}$. Observe that

$$\frac{1 - p_{ii}(h)}{h} \geq \frac{1}{h}\sum_{j\neq i} p_{ij}(h) \geq \frac{1}{h}\sum_{j\in A} p_{ij}(h)$$

for all $h > 0$. Letting $h \to 0$ we get

$$\sum_{j\in A} q_{ij} \leq q_i.$$

Since A is arbitrary, the result follows. □

Remark 2.13. Directly from Theorem 2.8, if a state i is stable, then

$$\mathbb{P}(X_{t+h} = i \mid X_t = i) = 1 - q_i h + o(h), \qquad h \to 0,$$

and for $j \neq i$,

$$\mathbb{P}(X_{t+h} = j \mid X_t = i) = q_{ij} h + o(h), \qquad h \to 0.$$

A very general class of Markov chains, namely regular jump Markov chains, are stable and conservative.

Definition 2.14. A stochastic process $X = (X_t)_{t \geq 0}$ taking its values in the (not necessarily countable) state space E is called a *jump process* if for almost all $\omega \in \Omega$ and all $t \geq 0$, there exists $\varepsilon(t, \omega) > 0$ such that

$$X_{t+s}(\omega) = X_t(\omega) \quad \text{for all } s \in [0, \varepsilon(t, \omega)].$$

It is called a *regular jump process* if, in addition, for almost all $\omega \in \Omega$, the set $A(\omega)$ of discontinuities of the function $t \mapsto X_t(\omega)$ is σ-discrete, that is, for all $c \geq 0$,

$$|A(\omega) \cap [0, c]| < \infty.$$

A continuous Markov chain which is a (regular) jump process is called a *(regular) jump Markov chain.*

Observe that for a jump process (not necessarily regular), there exists a sequence of times $(\tau_n)_{n \geq 0}$ such that $0 = \tau_0 < \tau_1 < \tau_2 < \tau_3 < \dots$ and a sequence $(Y_n)_{n \geq 0}$ such that

$$X_t = Y_n \quad \text{if } \tau_n \leq t < \tau_{n+1}.$$

This describes X on the interval $[0, \tau_\infty)$, where $\tau_\infty = \lim_{n \to \infty} \tau_n$ is the explosion time. If the process is regular, then $\tau_\infty = \infty$ almost surely and X is right-continuous.

Example 2.15. A continuous-time birth-and-death process is a homogeneous jump Markov chain taking its values in $\mathbb{Z}_+$, with the infinitesimal generator

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & \dots \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \lambda_3 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

The parameters λ_i and μ_i are the *birth* and *death* parameters, respectively: we have

$$\mathbb{P}\big(X_{t+h} = i + 1 \mid X_t = i\big) = \lambda_i h + o(h),$$
$$\mathbb{P}\big(X_{t+h} = i - 1 \mid X_t = i\big) = \mu_i \mathbf{1}_{\{i \geq 1\}} h + o(h)$$

and

$$\mathbb{P}\big(X_{t+h} = i \mid X_t = i\big) = 1 - (\lambda_i + \mu_i \mathbf{1}_{\{i \geq 1\}}) h + o(h).$$

The next result describes precisely the structure of jump Markov chains. In particular, it shows, in a sense, that the construction in Example 2.4 is canonical.

Theorem 2.16. *Let $X = (X_t)_{t\geq 0}$ be a regular jump Markov chain, with q-matrix Q, the transition times sequence $(\tau_n)_{n\geq 0}$, and the embedded process $(Y_n)_{n\geq 0}$. Then*

(a) X is stable and conservative.

(b) $(Y_n)_{n\geq 0}$ is a discrete-time homogeneous Markov chain with state space $E^{\Delta} = E \cup \{\Delta\}$ and transition matrix $\Gamma = [\gamma_{i,j}]_{i,j\in E^{\Delta}}$ given by

$$\gamma_{\Delta,\Delta} = 1, \qquad \gamma_{i,\Delta} = \begin{cases} 1, & \text{if } i \in E \text{ and } q_i = 0, \\ 0, & \text{if } i \in E \text{ and } q_i > 0, \end{cases}$$

and if $q_i > 0$ and $j \neq i$, then $\gamma_{ij} = \frac{q_{ij}}{q_i}$.

(c) Given $(Y_n)_{n\geq 0}$, the sequence $(\tau_{n+1} - \tau_n)_{n\geq 0}$ is independent, and for all $n \geq 0$ and all $a \in \mathbb{R}_+$,

$$\mathbb{P}\big(\tau_{n+1} - \tau_n > a \mid Y_n = i\big) = e^{-q_i a}$$

The proof of this result is rather technical and will be omitted. In particular, the theorem above shows that the q-matrix uniquely determines the distribution of the regular jump Markov chain. That is, we have the following.

Corollary 2.17. *Two regular jump Markov chains with the same infinitesimal generator have the same transition semigroup.*

Corollary 2.18 (Excursions)**.** *Fix a state $i \in E$ and consider the return times by $R_0^{(i)} = 0$ and, for $n \geq 1$,*

$$R_n^{(i)} = \inf\Big\{t > R_{n-1}^{(i)} \,:\, X_t = i \text{ and } X_s \neq i \text{ for some } s \in (R_{n-1}^{(i)}, t]\Big\}.$$

Then $\Big((X_t)_{R_n^{(i)} < t \leq R_{n+1}^{(i)}},\, R_{n+1}^{(i)} - R_n^{(i)}\Big)_{n\geq 0}$ is an i.i.d. collection of pieces of path.

2.3. Kolmogorov's Differential Systems. We have seen above that a standard transition kernel leads to an infinitesimal generator Q. There is a natural question whether the semigroup can be extracted from Q. We consider two cases, corresponding to finite and infinite spaces.

Case I: Finite state spaces. By the semigroup property, we have ,

$$\frac{P_{t+h} - P_t}{h} = P_t \frac{P_h - I}{h} = \frac{P_h - I}{h} P_t \tag{1.7}$$

for all $t \geq 0$ and all $h \geq 0$. Consequently, if the passage to the limit above is allowed (which is the case when the state space E is finite), we obtain the differential system

$$\frac{d}{dt} P_t = P_t Q = Q P_t, \tag{1.8}$$

where Q is the infinitesimal generator. The equation

$$\frac{d}{dt} P_t = Q P_t \tag{1.9}$$

can be written explicitly as follows: for all $i, j \in E$,

$$\frac{d}{dt} p_{ij}(t) = -q_i p_{ij}(t) + \sum_{k\neq i} q_{ik} p_{kj}(t),$$

where the summation is over all $k \in E$ with $k \neq i$. This system is referred to as *Kolmogorov's backward equations.* The *forward differential system* is

$$\frac{d}{dt}P_t = P_t Q, \tag{1.10}$$

that is, for all $i, j \in E$,

$$\frac{d}{dt}p_{ij}(t) = -p_{ij}(t)q_j + \sum_{k\in E, k\neq j} p_{ik}(t)q_{kj}.$$

It follows from the theory of ordinary differential equations that the solution of (1.9) or (1.10), with the initial condition $P_0 = I$, is

$$P_t = e^{Qt} = \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!}. \tag{1.11}$$

Case II: Infinite state spaces. When the state space is not finite, difficulties may arise in the passage to the limit $h \downarrow 0$ in (1.7), because of the possibly infinite sums involved. However, for the backward system, there is a positive result.

Theorem 2.19. *If the semigroup P is standard, stable and conservative, Kolmogorov's backward differential system* (1.9) *is satisfied.*

Proof. Let us identify E with $\mathbb{Z}_+$. We start with

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = \frac{p_{ii}(h) - 1}{h} p_{ij}(t) + \sum_{k\neq i} \frac{p_{ik}(h)}{h} p_{kj}(t) \tag{1.12}$$

and hence, by Fatou's lemma,

$$\liminf_{h\downarrow 0} \sum_{k\neq i} \frac{p_{ik}(h)}{h} p_{kj}(t) \geq \sum_{k\neq i} q_{ik}p_{kj}(t). \tag{1.13}$$

On the other hand, for $N > i$,

$$\begin{aligned}\sum_{k\neq i} \frac{p_{ik}(h)}{h} p_{kj}(t) &\leq \sum_{k\leq N,\, k\neq i} \frac{p_{ik}(h)}{h} p_{kj}(t) + \sum_{k>N} \frac{p_{ik}(h)}{h} \\ &= \sum_{k\leq N,\, k\neq i} \frac{p_{ik}(h)}{h} p_{kj}(t) + \frac{1 - p_{ii}(h)}{h} + \sum_{k\leq N, k\neq i} \frac{p_{ik}(h)}{h}\end{aligned}$$

and therefore

$$\limsup_{h\downarrow 0} \sum_{k\neq i} \frac{p_{ik}(h)}{h} p_{kj}(t) \leq \sum_{k\leq N,\, k\neq i} q_{ik}p_{kj}(t) + q_i + \sum_{k\leq N, k\neq i} q_{ik}.$$

Letting $N \to \infty$ and using the conservation property, we get

$$\limsup_{h\downarrow 0} \sum_{k\neq i} \frac{p_{ik}(h)}{h} p_{kj}(t) \leq \sum_{k\neq i} q_{ik}p_{kj}(t),$$

which combined with (1.13) gives

$$\lim_{h\downarrow 0} \sum_{k\neq i} \frac{p_{ik}(h)}{h} p_{kj}(t) = \sum_{k\neq i} q_{ik}p_{kj}(t).$$

It remains to let $h \to 0$ in (1.12) to get the assertion. □

For the forward system, the result is considerably less general.

Theorem 2.20. *Suppose that a standard semigroup P is stable and conservative. Assume in addition that for all states i and all $t \geq 0$,*

$$\sum_{k \in E} p_{ik}(t) q_k < \infty. \tag{1.14}$$

Then Kolmogorov's forward differential system (1.10) *is satisfied.*

Proof. For $j \neq k$ we have

$$\frac{p_{kj}(h)}{h} \leq \frac{1 - p_{kk}(h)}{h} \leq q_k.$$

Now we write the first equality in (1.7) in the form

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = p_{ij}(t) \frac{p_{jj}(h) - 1}{h} + \sum_{k \neq j} p_{ik}(t) \frac{p_{kj}(h)}{h}.$$

It remains to let $h \to 0$ and apply Lebesgue's dominated convergence theorem. □

Remark 2.21. Condition (1.14) is trivially satisfied when $\sup_{i \in E} q_i < \infty$.

Example 2.22 (M/M/1 queue)**.** Suppose that customers arrive at a single-server service station in accordance with a Poisson process with intensity λ. Upon arrival, each customer goes directly into service if the server is free; if not, then the customer joins the queue and waits for his/her turn. When the server finishes serving a customer, he/she leaves the system and the next customer in line, if any, enters the service (that is, the service follows the order of the arrival of customers: first-in-queue, first-out-of-queue, FIFO). The successive service times are assumed to be independent random variables following the exponential distribution $\mathrm{Exp}(\mu)$.

Let X_t denote the number of customers in the system at time t, where the system refers to the line plus the service area: for instance, $X_t = 2$ means there is one customer in service and one waiting in line. A transition can only occur at customer arrival or departure times, and departures occur whenever a service is completed. At an arrival time, X_t jumps up by 1, and at a departure time X_t jumps down by 1. Therefore, this is an example of a birth-and-death process, with the intensity matrix

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \dots \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \dots \\ 0 & \mu & -(\lambda+\mu) & \lambda & 0 & \dots \\ 0 & 0 & \mu & -(\lambda+\mu) & \lambda & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

By Theorem 2.16, the embedded chain has the transition matrix

$$\Gamma = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \dots \\ \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & 0 & 0 & \dots \\ 0 & \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & 0 & \dots \\ 0 & 0 & \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix},$$

so it is an asymmetric random walk over nonnegative integers, with reflection at 0.

It turns out that this model, though simple, leads to complicated transition kernels. Suppose we want to compute explicitly the probability $p_{ij}(t)$ for given i, j and t. The Kolmogorov forward equations

$$\begin{aligned}\frac{d}{dt}p_{ij}(t) &= \sum_{k\geq 0} p_{ik}(t)q_{kj}\\ &= \begin{cases} -\lambda p_{00}(t) + \mu p_{01}(t) & \text{if } i=j=0,\\ \mu p_{i,j-1}(t) - (\lambda+\mu)p_{ij}(t) + \lambda p_{i,j+1}(t) & \text{if } j\geq 1.\end{cases}\end{aligned}$$

can be explicitly solved with the use of moment generating functions. The formula for $p_{ij}(t)$ is

$$e^{-(\lambda+\mu)}\left[\rho^{(k-i)/2}I_{|k-i|}(at) + \rho^{(k-i-1)/2}I_{|k+i+1|}(at) + (1-\rho)\rho^k \sum_{j=k+i+2}^{\infty} \rho^{-j/2}I_{|j|}(at)\right],$$

where $\rho = \lambda/\mu$, $a = 2\mu\rho^{1/2}$ and

$$I_k(x) = \sum_{m=0}^{\infty} \frac{(x/2)^{k+2m}}{(k+m)!m!}, \qquad k\geq 0,$$

is the modified Bessel function of the first kind of order k. For $\lambda = \mu$, we have $\rho = 1$, $a = 2\mu$ and the formula simplifies to

$$e^{-(\lambda+\mu)}\Big(I_{|k-i|}(2\mu t) + I_{|k+i+1|}(2\mu t)\Big),$$

still involving Bessel functions.

2.4. Recurrence and Invariant Measures. Throughout this subsection, we assume that $X = (X_t)_{t\geq 0}$ is a regular jump Markov chain with a transition function $P = (P_t)_{t\geq 0}$, q-matrix Q and the embedded Markov chain $Y = (Y_n)_{n\geq 0}$ with a transition matrix Γ.

Definition 2.23. (a) The process X is said to be *irreducible*, if the embedded Markov chain Y is irreducible.

(b) The state $i \in E$ is called *recurrent*, if it is recurrent for the embedded Markov chain Y. Otherwise, it is called a *transient* state.

(c) The vector $\nu = \{\nu_i\}_{i\in E}$ with nonnegative entries is an *invariant measure*, if

$$\nu P_t = \nu \quad \text{for all } t\geq 0.$$

If the invariant measure is a probability distribution, it is called an *invariant distribution*, *stationary distribution* or *steady-state distribution.*

Remark 2.24. If the initial distribution μ of X is stationary, then for any t the law of X_t is $\mu P_t = \mu$.

Theorem 2.25. *Let $X = (X_t)_{t\geq 0}$ be a regular jump Markov chain and suppose that the embedded chain $(Y_n)_{n\geq 0}$ is irreducible and recurrent. Then X has an invariant measure ν, which is unique up to multiplicative factors and satisfies $0 < \nu_j < \infty$ for all $j \in E$. Furthermore, a stationary distribution exists for X if and only if*

$$\sum_{i\in E} \nu_i < \infty,$$

in which case the stationary distribution is $\pi = \left\{\nu_j / \sum_{i\in E} \nu_j\right\}_{j\in E}$.

Proof. We will use the theory of discrete-time Markov chains. Since Y is irreducible and recurrent, there is a unique invariant measure μ for Y (up to multiplicative constants), i.e., there is a unique vector $(\mu_j)_{j\in E}$ satisfying $\mu(\Gamma - I) = 0$ and $0 < \mu_j < \infty$ for all j. Furthermore, one might take

$$\mu_i = \mathbb{E}\left[\sum_{n=1}^{T_1^{(k)}} 1_{\{Y_n = i\}} \mid Y_0 = k\right],$$

where k is an arbitrary state and $T_1^{(k)}$ is the first return to k of the embedded chain. Since X is recurrent, no states are absorbing: $q_j > 0$ for all j, and hence by Theorem 2.16, the j-th row of $\Gamma - I$ coincides with the j-th row of the matrix Q/q_j. Thus, the equation $\mu(\Gamma - I) = 0$ is equivalent to $\nu Q = 0$, with $\nu_j = \mu_j/q_j$ for all j. Note that $0 < \nu_j < \infty$ for all j. Clearly, this measure will induce a stationary distribution if its total mass if finite.

It remains to address the fact that ν is an invariant measure for X; we will prove this only in the case when $\sum_{j\in E} \nu_j < \infty$. Arguing as in the proof of backward Kolmogorov equations, we show that $(\nu P_t)' = \nu P_t' = \nu Q P_t = 0$. Consequently, we get $\nu P_t = \nu P_0 = \nu$ and the invariance follows. □

Remark 2.26. As the above proof shows, the invariant measure ν can be defined by two different approaches:

- Solving for $\nu Q = 0$;
- Letting $\nu_j = \mu_j/q_j$ for all j, where μ is stationary for Y.

There is a third formula, referring to the regenerative structure of continuous Markov chains:

$$\nu_j = \mathbb{E}\left[\int_0^{R_1^{(i)}} 1_{\{X_t = j\}} \mathrm{d}t \mid X_0 = i\right], \qquad j \in E, \tag{1.15}$$

where i is an arbitrary state and $R_1^{(i)}$ is the first return to i of X. For different starting states, the formula (1.15) might produce different measures ν (proportional to each other).

Example 2.27. (i) Let N be the Poisson process with intensity λ. The associated q-matrix is

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & 0 & \dots \\ 0 & 0 & -\lambda & \lambda & 0 & \dots \\ 0 & 0 & 0 & -\lambda & \lambda & \dots \\ \dots & \dots & \dots & \dots & \dots & \end{pmatrix}$$

and the only solution to the equation $\nu Q = 0$ is the zero measure ν.

(ii) Let X be a pure-birth process on $\mathbb{Z}$, with the q matrix

$$Q = \begin{pmatrix} \dots & \dots & \dots & \dots & \dots & \dots & \\ \dots & -\lambda & \lambda & 0 & 0 & 0 & \dots \\ \dots & 0 & -\lambda & \lambda & 0 & 0 & \dots \\ \dots & 0 & 0 & -\lambda & \lambda & 0 & \dots \\ \dots & 0 & 0 & 0 & -\lambda & \lambda & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \end{pmatrix}.$$

Then the solutions to $\mu Q = 0$ are the constant multiples of the counting measure.

Definition 2.28. The recurrent, irreducible regular jump Markov chain X is called *ergodic*, if a stationary distribution exists for X.

Corollary 2.29. *Suppose that X is recurrent and irreducible. Then the following are equivalent:*

- *The expected cycle lengths are finite: for each i,*

$$m_i := \mathbb{E}\big(R_1^{(i)}|X_0 = i\big) < \infty. \tag{1.16}$$

- *The equation $\nu A = 0$ has a probability distribution as a solution.*
- *X is ergodic.*

Proof. By (1.15) we have $\mathbb{E}R_1^{(i)} = \sum_{j\in E} \nu_j$, for some nonzero measure ν satisfying $\nu Q = 0$. Thus the first two conditions in the statement are equivalent. The equivalence of the second and the third condition is evident. □

Remark 2.30. In the literature, the states $i \in E$ satisfying (1.16) are called t-positive recurrent.

Remark 2.31. If X is ergodic, there is a nice formula for its stationary distribution ν. Fix an arbitrary starting state i and apply (1.15) to obtain

$$c_i\nu_j = \mathbb{E}\left[\int_0^{R_1^{(i)}} 1_{\{X_t=j\}}\mathrm{d}t \mid X_0 = i\right] \tag{1.17}$$

for some positive c_i. Summing over all j, we get $c_i = c_i \sum_j \nu_j = \mathbb{E}\big[R_1^{(i)} \mid X_0 = i\big] = m_i$. Furthermore,

$$c_i\nu_i = \mathbb{E}\left[\int_0^{R_1^{(i)}} 1_{\{X_t=i\}}\mathrm{d}t \mid X_0 = i\right] = \frac{1}{q_i},$$

since the integral - the time spent at i - is exponentially distributed with parameter q_i. This yields the formula

$$\nu_i = \frac{1}{q_i m_i}.$$

Plugging this back into (1.17) gives

$$\mathbb{E}\left[\int_0^{R_1^{(i)}} 1_{\{X_t=j\}}\mathrm{d}t \mid X_0 = i\right] = \frac{m_i}{q_j m_j}.$$

We conclude with a version of the ergodic theorem for continuous Markov chains.

Theorem 2.32. *Let $X = (X_t)_{t \geq 0}$ be an ergodic jump Markov chain with the stationary distribution ν.*

(i) For any i we have the convergence

$$\lim_{t \to \infty} p_{ij}(t) = \nu_j, \quad j \in E.$$

(ii) If a function $f : E \to \mathbb{R}$ satisfies $\sum_{i \in S} |f(i)| \nu_i < \infty$, then for any initial distribution μ we have

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t f(X_s)\, ds = \sum_{i \in S} f(i) \nu_i \quad \textit{almost surely.}$$

In particular, setting $f = \chi_{\{i\}}$, we see that the fraction of time spent in state i converges almost surely to ν_i.

Proof. This follows from the corresponding statement from discrete-time setting. □