Markov and Semi-Markov Processes

5

This chapter is devoted to jump Markov processes and finite semi-Markov processes. In both cases, the index is considered as the calender time, continuously counted over the positive real line. Markov processes are continuous-time processes that share the Markov property with the discrete-time Markov chains. Their future evolution conditional to the past depends only on the last occupied state. Their extension to the so-called semi-Markov processes naturally arises in many types of applications. The future evolution of a semi-Markov process given the past depends on the occupied state too, but also on the time elapsed since the last transition.

Detailed notions on homogeneous jump Markov processes with discrete (countable) state spaces will be presented. Some basic notions on semi-Markov processes with finite state spaces will follow, illustrated through typical examples.

5.1 Jump Markov Processes

We will investigate in this section mainly the jump Markov process. Since a jump Markov process is a Markov process which is constant between two successive jumps, we will first present some basic notions on Markov processes with continuous time.

Thereafter, $\mathbf{X} = (X_t)_{t\in\mathbb{R}_+}$ will denote a process defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbb{P})$, taking values in a finite or enumerable set E. The filtration will be supposed to be the natural filtration of the process, even if, of course, $\mathbf{X}$ can satisfy the Markov property below with respect to a larger filtration.

5.1.1 Markov Processes

A stochastic process is called a Markov process if its future values given the past and the present depend only on the present. Especially, processes with independent increments—Brownian motion, Poisson processes—are Markov processes.

V. Girardin, N. Limnios, *Applied Probability*,
https://doi.org/10.1007/978-3-319-97412-5_5

Definition 5.1 A stochastic process $\mathbf{X} = (X_t)_{t\in\mathbb{R}_+}$, with state space E, is called a Markov process—with respect to its natural filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if it satisfies the Markov property

$$\mathbb{P}(X_{s+t} = j \mid \mathcal{F}_s) = \mathbb{P}(X_{s+t} = j \mid X_s) \quad \text{a.s.},$$

for all non-negative real numbers t and s and all $j \in E$.

The Markov property can also be written

$$\mathbb{P}(X_{s+t} = j \mid X_{s_1} = i_1, \dots, X_{s_n} = i_n, X_s = i) = \mathbb{P}(X_{s+t} = j \mid X_s = i),$$

for all $n \in \mathbb{N}$, all $0 \le s_1 < \cdots < s_n < s$, $0 \le t$, and all states $i_1, \dots, i_n, i, j$ in E.

If moreover the above conditional probability does not depend on s, then

$$\mathbb{P}(X_{s+t} = j \mid X_s = i) = P_t(i, j), \quad i \in E, j \in E \text{ and } s \ge 0, t \ge 0,$$

and the process $\mathbf{X}$ is said to be homogeneous with respect to time.

We will study here only homogeneous processes. The trajectories $t \longrightarrow X_t(\omega)$ will be assumed to be continuous on the right for the discrete topology, with a.s. finite limits on the left; such a process is said to be cadlag. Similarly to Markov chains, the distribution of X_0 is called the initial distribution of the process, and we will set $\mathbb{P}_i(\cdot) = \mathbb{P}(\cdot \mid X_0 = i)$ and $\mathbb{E}_i(\cdot) = \mathbb{E}(\cdot \mid X_0 = i)$.

▷ *Example 5.2 (Process with Independent Increments)* Any continuous-time process $\mathbf{X} = (X_t)_{t\in\mathbb{R}_+}$ with independent increments and taking values in $\mathbb{Z}$ is a Markov process. Indeed, for all nonnegative real numbers s and t,

$$\begin{aligned}
\mathbb{P}(X_{s+t} = j \mid \mathcal{F}_s) &= \sum_{i\in\mathbb{Z}} \mathbb{P}(X_{s+t} = j, X_s = i \mid \mathcal{F}_s) \\
&\overset{(1)}{=} \sum_{i\in\mathbb{Z}} \mathbb{1}_{(X_s=i)} \mathbb{P}(X_{s+t} - X_s = j - i \mid X_s) \\
&= \sum_{i\in\mathbb{Z}} \mathbb{P}(X_{s+t} = j, X_s = i \mid X_s) = \mathbb{P}(X_{s+t} = j \mid X_s).
\end{aligned}$$

Note that we can also write

$$\begin{aligned}
\mathbb{P}(X_{s+t} = j \mid \mathcal{F}_s) &= \sum_{i\in\mathbb{Z}} \mathbb{E}\,[\mathbb{1}_{(X_s=i)} \mathbb{1}_{(X_{s+t}-X_s=j-i)} \mid \mathcal{F}_s] \\
&\overset{(1)}{=} \sum_{i\in\mathbb{Z}} \mathbb{1}_{(X_s=i)} \mathbb{E}\,[\mathbb{1}_{(X_{s+t}-X_s=j-i)}].
\end{aligned} \tag{5.1}$$

(1) because X_s is $\mathcal{F}_s$-measurable and $X_{s+t} - X_s$ is independent of $\mathcal{F}_s$.

The random variable defined in (5.1) is measurable for the σ-algebra generated by X_s, so is equal to $\mathbb{P}(X_{s+t} = j \mid X_s)$, and the process satisfies the Markov property. $\triangleleft$

Proposition 5.3 *The past and the future of a Markov process are independent given the present, that is, for all $A \in \mathcal{F}_t = \sigma(X_s, s \leq t)$ and $B \in \sigma(X_s, s \geq t)$,*

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

Proof We compute

$$\begin{aligned}\mathbb{P}(A \cap B \mid X_t) &= \mathbb{E}\left[\mathbb{E}\left(\mathbb{1}_A \mathbb{1}_B \mid \mathcal{F}_t\right) \mid X_t\right] = \mathbb{E}\left[\mathbb{1}_A \mathbb{E}\left(\mathbb{1}_B \mid \mathcal{F}_t\right) \mid X_t\right] \\ &= \mathbb{E}\left[\mathbb{1}_A \mathbb{E}\left(\mathbb{1}_B \mid X_t\right) \mid X_t\right] = \mathbb{E}\left[\mathbb{1}_A \mid X_t\right] \mathbb{E}\left(\mathbb{1}_B \mid X_t\right) \\ &= \mathbb{P}(A \mid X_t)\mathbb{P}(B \mid X_t),\end{aligned}$$

where all equalities hold a.s. $\square$

Definition 5.4 A Markov process $\mathbf{X}$ is said to satisfy the strong Markov property if for any stopping time T adapted to $\mathbf{X}$ and all $s \geq 0$, on $(T < +\infty)$,

$$\mathbb{P}(X_{T+s} = j \mid \mathcal{F}_T) = \mathbb{P}_{X_T}(X_s = j) \quad \text{a.s.},$$

where $\mathcal{F}_T = \{A \in \mathcal{F} : \forall n \in \overline{\mathbb{N}},\ A \cap (T = n) \in \mathcal{F}_n\}$ is the σ-algebra of events previous to T. Then, the Markov process is said to be strong.

Theorem 5.5 *Let $\mathbf{X}$ be a strong Markov process, with state space E. If $f : E^{\mathbb{R}_+} \longrightarrow \mathbb{R}_+^d$ is a Borel function and T is a finite stopping time for $\mathbf{X}$, then*

$$\mathbb{E}_i(f \circ \mathbf{X} \circ \theta_T \mid \mathcal{F}_T) = \mathbb{E}_{X_T}(f \circ \mathbf{X}), \quad a.s.,\ i \in E.$$

where θ_s is the shift operator such that $X_t \circ \theta_s = X_{t+s}$.

Proof For a function $f \circ \mathbf{X} = f(X_{t_1}, \ldots, X_{t_n})$ with $0 \leq t_1 < \cdots < t_n$.

We compute

$$\begin{aligned}&\mathbb{E}_i[f(X_{t_1+T}, \ldots, X_{t_n+T}) \mid \mathcal{F}_T] = \\ &= \sum_{(i_1,\ldots,i_n)\in e^n} \mathbb{P}_i(X_{t_1+T} = i_1, \ldots, X_{t_n+T} = i_n \mid \mathcal{F}_T) f(i_1, \ldots, i_n).\end{aligned}$$

Thanks to the strong Markov property,

$$
\begin{aligned}
&\mathbb{P}_i(X_{t_1+T} = i_1, \dots, X_{t_n+T} = i_n \mid \mathcal{F}_T) = \\
&= \mathbb{P}_i(X_{t_n+T} = i_n \mid X_{t_{n-1}+T} = i_{n-1}) \dots \mathbb{P}_i(X_{t_1+T} = i_n \mid X_T) \\
&= \mathbb{P}_i(X_{t_n} = i_n \mid X_{t_{n-1}} = i_{n-1}) \dots \mathbb{P}_{X_T}(X_{t_1} = i_n) \\
&= \mathbb{P}_{X_T}(X_{t_1} = i_1, \dots, X_{t_n} = i_n),
\end{aligned}
$$

Therefore,

$$
\mathbb{E}_i[f(X_{t_1+T}, \dots, X_{t_n+T}) \mid \mathcal{F}_T] = \mathbb{E}_{X_T}[f(X_{t_1}, \dots, X_{t_n})],
$$

and the result follows. □

5.1.2 Transition Functions

Definition 5.6 A Markov process **X** is said to be a (pure) jump Markov process if, whatever be its initial state, it evolves through isolated jumps from state to state, and its trajectories are a.s. constant between the jumps.

To be exact, a jump time is a time where a change of state occurs. If (T_n) is the sequence of successive jump times of a jump Markov process **X**, for all $n \in \mathbb{N}$, we have $T_n < T_{n+1}$ if $T_n < +\infty$ and $T_n = T_{n+1}$ if $T_n = +\infty$, and

$$
X_t = X_{T_n}, \quad T_n \leq t < T_{n+1}, \quad T_n < +\infty, \quad n \geq 0.
$$

A typical trajectory of such a process is shown in Fig. 5.1.

Clearly, any jump Markov process is a strong Markov process.

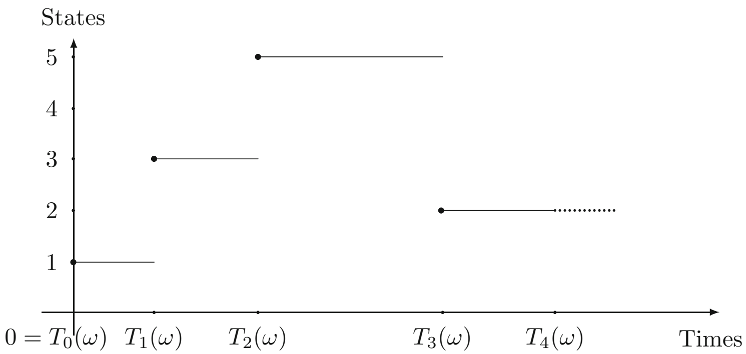

Fig. 5.1 A trajectory of a jump Markov process

Definition 5.7 A jump Markov process is said to be regular, or non explosive, if for any initial state, the number of jumps in any finite time interval is a.s. finite.

Let $\zeta = \sup_{n\geq 1} T_n$ be the time duration of the process, a random variable taking values in $\overline{\mathbb{R}}_+$. If $\mathbb{P}(\zeta = +\infty) = 1$, the process is regular; otherwise, that is if $\mathbb{P}(\zeta < +\infty) > 0$, it is explosive.

For regularity criteria, see Theorem 5.31 and Proposition 5.33 below. The next example explains the phenomenon of explosion.

$\triangleright$ *Example 5.8* Let (X_n) be a sequence of independent random variables with exponential distributions with parameters (λ_n). Setting $S = \sum_{n\geq 1} X_n$, we have

$$\mathbb{P}(S = +\infty) > 0 \quad \text{if and only if} \quad \sum_{n\geq 1} \frac{1}{\lambda_n} = +\infty.$$

Indeed, if $\mathbb{P}(S = +\infty) > 0$, then $\sum_{n\geq 1} 1/\lambda_n = \mathbb{E}\, S = +\infty$, and

$$\mathbb{E}\,(e^{-S}) = \frac{1}{\prod_{n\geq 1}(1 + 1/\lambda_n)} \leq \frac{1}{\sum_{n\geq 1} 1/\lambda_n}$$

yields the converse. $\triangleleft$

Definition 5.9 Let $\mathbf{X}$ be a homogeneous jump Markov process. The family of functions defined on $\mathbb{R}_+$ by $t \longrightarrow P_t(i, j) = \mathbb{P}(X_{t+h} = j \mid X_h = i)$, for i and j in E, are called the transition functions of the process on E.

We will denote by P_t the (possibly infinite) matrix $(P_t(i, j))_{(i,j)\in E\times E}$, and consider only processes such that $P_0(i, i) = 1$.

Properties of Transition Functions

1. $0 \leq P_t(i, j) \leq 1$, for all i and j in E and $t \geq 0$, because $P_t(i, \cdot)$ is a probability.
2. $\sum_{j\in E} P_t(i, j) = 1$, for all $i \in E$ and $t \geq 0$, because E is the set of all values taken by $\mathbf{X}$.
3. **(Chapman-Kolmogorov equation)**

$$\sum_{k\in E} P_t(i, k) P_s(k, j) = P_{t+s}(i, j), \quad i, j \in E,\ s \geq 0,\ t \geq 0.$$

4. If $\lim_{t\to 0^+} P_t(i, j) = \delta_{ij}$ for all $i \in E$, the process is said to be stochastically continuous. If this property is satisfied uniformly in i, the transition function (or semi-group) is said to be uniform.

Thanks to Chapman-Kolmogorov equation, the family $\{P_t : t \geq 0\}$ equipped with the composition $P_t P_h = P_{t+h}$ is a semi-group. Indeed, the operation is commutative and associative, and I (the identity $|E| \times |E|$-matrix) is a neutral element; on the contrary, in general, a given element has no inverse in the family, which therefore is not a group.

$\triangleright$ *Example 5.10 (Birth Process)* Let $\mathbf{X}$ be a stochastic process with state space $E = \mathbb{N}$ and transition function such that, when $h \to 0^+$,

$$P_h(i, j) = \begin{cases} \lambda_i h + o(h) & \text{if } j = i + 1, \\ 1 - \lambda_i h + o(h) & \text{if } j = i, \\ 0 & \text{otherwise,} \end{cases}$$

This is indeed a transition function satisfying the four above properties.

If, for instance, $\lambda_i = \lambda / i$, then P_t is uniform; on the contrary, it is not uniform if $\lambda_i = i\lambda$. If $\lambda_i = \lambda$ for all i, the process is a Poisson process. $\triangleleft$

The following proposition is a straightforward consequence of both the compound probabilities formula and Markov property.

Proposition 5.11 *For all $n \in \mathbb{N}$, all nonnegative reals $0 \leq t_0 < t_1 < \cdots < t_n$, and all finite sequence of states $i_0, i_1, \ldots, i_n$, we have*

$$\mathbb{P}(X_{t_1} = i_1, \ldots, X_{t_n} = i_n \mid X_{t_0} = i_0) = P_{t_1 - t_0}(i_0, i_1) \ldots P_{t_n - t_{n-1}}(i_{n-1}, i_n),$$

and

$$\mathbb{P}(X_0 = i_0, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) = \alpha(i_0) P_{t_1}(i_0, i_1) \ldots P_{t_n - t_{n-1}}(i_{n-1}, i_n),$$

where α denotes the initial distribution of the process.

Thus, the finite-dimensional distributions of a jump Markov process are characterized by its initial distribution and its transition function.

We state without proof the next result, necessary for proving the following one.

Theorem 5.12 (Lévy) *For all given states i and j, $P_t(i, j)$ is either identically null, or never null on $\mathbb{R}_+$.*

Proposition 5.13 *Let P_t be the transition function of a jump Markov process.*

1. *If some $t > 0$ exists such that $P_t(i, i) = 1$, then $P_s(i, i) = 1$ for all $s \in \mathbb{R}_+$.*
2. *$| P_{t+\varepsilon}(i, j) - P_t(i, j) | \leq 1 - P_{|\varepsilon|}(i, i)$ for all $t \geq 0$ and $(i, j) \in E \times E$, and hence $P_t(i, j)$ is uniformly continuous with respect to t.*

Proof

1. Let $t > 0$ be such that $P_t(i,i) = 1$. Then, for any $s < t$,

$$0 = 1 - P_t(i,i) = \sum_{j \neq i} P_t(i,j) \geq P_{t-s}(i,i) P_s(i,j) \geq 0.$$

But, thanks to Lévy's theorem, $P_{t-s}(i,i) > 0$, so $P_s(i,j) = 0$ for all $j \neq i$, and hence $P_s(i,i) = 1$.

For $s > t$, it is sufficient to choose n such that $s/n < t$ for getting $P_s(i,i) \geq [P_{s/n}(i,i)]^n \geq 1$.

2. Let $\varepsilon > 0$. We deduce from

$$P_{t+\varepsilon}(i,j) - P_t(i,j) = \sum_{k \neq i} P_\varepsilon(i,k) P_t(k,j) - P_t(i,j)[1 - P_\varepsilon(i,i)]$$

that

$$\begin{aligned} -[1 - P_\varepsilon(i,i)] \leq -P_t(i,j)[1 - P_\varepsilon(i,i)] &\leq P_{t+\varepsilon}(i,j) - P_t(i,j) \\ &\leq \sum_{k \neq i} P_\varepsilon(i,k) P_t(k,j) \leq \sum_{k \neq i} P_\varepsilon(i,k) \leq 1 - P_\varepsilon(i,i), \end{aligned}$$

and hence

$$\mid P_{t+\varepsilon}(i,j) - P_t(i,j) \mid \leq 1 - P_\varepsilon(i,i). \tag{5.2}$$

Replacing t by $t - \varepsilon$ in the above inequality for $0 < \varepsilon < t$, we get

$$\mid P_{t-\varepsilon}(i,j) - P_t(i,j) \mid = \mid P_t(i,j) - P_{t-\varepsilon}(i,j) \mid \leq 1 - P_\varepsilon(i,i). \tag{5.3}$$

The result follows from (5.2) and (5.3). □

5.1.3 Infinitesimal Generators and Kolmogorov's Equations

The transition function of a Markov process is identified from its generator through the Kolmogorov's equations.

Definition 5.14 The (infinitesimal) generator $A = (a_{ij})_{(i,j) \in E \times E}$ of a Markov process $\mathbf{X}$ is given by the derivative on the right of the transition function P_t at time $t = 0$, or

$$a_{ij} = \lim_{t \to 0^+} \frac{P_t(i, j) - I(i, j)}{t},$$

where I is the identity matrix.

These quantities are always well-defined, but when $i = j$ they may be equal to $-\infty$. The generator of a Markov process is such that $a_{ij} \geq 0$ for all $i \neq j$ and $\sum_{j \in E} a_{ij} = 0$ for all i. We set $a_i = -a_{ii} \geq 0$.

Definition 5.15 A state i is said to be stable if $0 < a_i < +\infty$, instantaneous if $a_i = +\infty$, absorbing if $a_i = 0$, and conservative if $\sum_{j \in E} a_{ij} = 0$.

A generator—or the associated process—of which all states are stable (instantaneous, conservative) is said to be stable (instantaneous, conservative).

At each passage in a stable state, the process will spend a.s. a non null and finite time. On the contrary, it will a.s. jump instantaneously from an instantaneous state. Finally, reaching an absorbing state, the process will remain there forever.

We state without proof the next result.

Theorem 5.16 *Let* $\mathbf{X}$ *be a jump Markov process.*

1. *The trajectories of* $\mathbf{X}$ *are a.s. continuous on the right if and only if no instantaneous states exist.*
2. *If* E *is finite, then no instantaneous states exist.*

▷ *Example 5.17 (Birth Process on* $\mathbb{N}$*)* The generator of this process is

$$a_{ij} = \begin{cases} \lambda_i & \text{if } j = i + 1, \\ -\lambda_i & \text{if } j = i, \\ 0 & \text{otherwise,} \end{cases}$$

for all integers i and j. Hence it is a conservative process. ◁

Theorem 5.18 (Kolmogorov's Equations) *If* $\mathbf{X}$ *is regular, then the transition functions* $t \longrightarrow P_t(i, j)$ *are continuously differentiable on* $\mathbb{R}_+^*$ *for all states* i *and* j*, and satisfy the equations*

$$P_t^{'}(i, j) = \sum_{k \in E} a_{ik} P_t(k, j) \quad \textit{and} \quad P_t^{'}(i, j) = \sum_{k \in E} P_t(i, k) a_{kj}.$$

In matrix form, the above equations become

$$\frac{d}{dt}P_t = AP_t \quad \text{and} \quad \frac{d}{dt}P_t = P_t A, \tag{5.4}$$

and are respectively called backward and forward Kolmogorov's equations.

Proof We prove only the second part of the theorem for a finite E.

We deduce by differentiating with respect to s the Chapman-Kolmogorov equation that

$$P'_{s+t}(i, j) = \sum_{k \in E} P'_s(i, k) P_t(k, j),$$

or, when $s \to 0^+$,

$$P'_t(i, j) = \sum_{k \in E} a_{ik} P_t(k, j).$$

The second equation is obtained symmetrically. □

The definition of uniformisable processes is necessary for stating the following result.

Definition 5.19 A jump Markov process is said to be uniformisable if

$$\sup_{i \in E} a_i < +\infty.$$

The next result is a straightforward consequence of the theory of linear differential equations.

Theorem 5.20 *When the process is uniformisable, the common solution of Kolmogorov's equations (5.4) is*

$$P_t = e^{tA} = I + \sum_{k \geq 1} \frac{t^k}{k!} A^k. \tag{5.5}$$

Numerous methods for computing numerically the above solution of Kolmogorov's equations exist: direct computation of the series (5.5) truncated at a certain value of k, uniformisation—see Example 5.51 below, Laplace transform, determination of the eigen-values and eigen-vectors, …

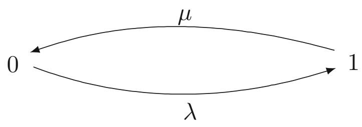

Fig. 5.2 Graph of the conservative process of Example 5.21

▷ *Example 5.21* Consider a conservative process with two states, say $E = \{0, 1\}$, with generator

$$A = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix},$$

with $\lambda > 0$ and $\mu > 0$ (Fig. 5.2). The two eigen-values of the generator A are $s_1 = 0$ and $s_2 = -\lambda - \mu$, and $(1, 1)'$ and $(\lambda, -\mu)'$ are two associated eigen-vectors. Therefore,

$$e^{tA} = QDQ^{-1},$$

where

$$Q = \begin{pmatrix} 1 & \lambda \\ 1 & -\mu \end{pmatrix}, \quad Q^{-1} = \frac{1}{\lambda + \mu} \begin{pmatrix} \mu & \lambda \\ 1 & -1 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 1 & 0 \\ 0 & e^{-(\lambda+\mu)t} \end{pmatrix}.$$

So, finally,

$$P_t = e^{tA} = \frac{1}{\lambda + \mu} \begin{pmatrix} \mu & \lambda \\ \mu & \lambda \end{pmatrix} + \frac{e^{-(\lambda+\mu)t}}{\lambda + \mu} \begin{pmatrix} \lambda & -\lambda \\ -\mu & \mu \end{pmatrix},$$

that is a closed-form expression. ◁

5.1.4 Embedded Chains and Classification of States

Let us begin by showing that the exit time of a given state of a Markov process has an exponential distribution. The nature of the parameter and its connection with the generator of the process will be specified later in Corollary 5.27. Note the exit time of i is also the sojourn time in i (before first exit), or the hitting time of $E \setminus \{i\}$.

Proposition 5.22 *If the state i is not an absorbing state, then the exit time of i has an exponential distribution with respect to $\mathbb{P}_i$, with parameter λ_i depending on i.*

Proof Let T_1 denote this first jump time. If θ_s denotes the shift operator, then $T_1 \circ \theta_s$ is the time of the first jump after time s, and we have

$$\begin{aligned}
\mathbb{P}_i(T_1 > s+t) &= \mathbb{P}_i(T_1 > s, T_1 \circ \theta_s > t) = \mathbb{E}_i[\mathbb{P}_i(T_1 > s, T_1 \circ \theta_s > t \mid \mathcal{F}_s)] \\
&= \mathbb{E}_i[\mathbb{1}_{(T_1>s)}\mathbb{P}_i(T_1 \circ \theta_s > t \mid \mathcal{F}_s)] \\
&\overset{(1)}{=} \mathbb{E}_i[\mathbb{1}_{(T_1>s)}\mathbb{P}_i(T_1 > t)] = \mathbb{P}_i(T_1 > s)\mathbb{P}_i(T_1 > t).
\end{aligned}$$

(1) by the Markov property.

Set $R(t) = \mathbb{P}_i(T_1 > t)$. The above equation can be written $R(s+t) = R(s)R(t)$, for all nonnegative reals s and t. This is Cauchy functional equation on $\mathbb{R}_+$, whose solution is well-known to be an exponential function, precisely

$$R(u) = \mathbb{P}_i(T_1 > u) = e^{-\lambda_i u}\mathbb{1}_{\mathbb{R}_+}(u),$$

with $\lambda_i \geq 0$; in other words, $T_1 \sim \mathcal{E}(\lambda_i)$. □

Note that the expected sojourn time in an instantaneous state is null.

Proposition 5.23 *The random variables T_1 and X_{T_1} are independent with respect to $\mathbb{P}_i$ for all non absorbing $i \in E$.*

Proof Let B denote a subset of $E \setminus \{i\}$. We have

$$\begin{aligned}
\mathbb{P}_i(X_{T_1} \in B, T_1 > s) &= \mathbb{E}_i[\mathbb{P}_i(X_{T_1} \circ \theta_s \in B, T_1 > s \mid \mathcal{F}_s)] \\
&= \mathbb{E}_i[\mathbb{1}_{(T_1>s)}\mathbb{P}_i(X_{T_1} \circ \theta_s \in B \mid \mathcal{F}_s)] \\
&= \mathbb{P}_i(T_1 > s)\mathbb{P}_i(X_{T_1} \in B),
\end{aligned}$$

if $s > 0$. □

Now, let us consider the sequence of random variables (J_n) defined by

$$J_n = X_{T_n}, \quad n \text{ such that } T_n < +\infty.$$

This is the sequence of the successive states visited by the process $\mathbf{X}$. Clearly, it is defined up to the explosion of the process—if explosion occurs; in this regard, it is said to be minimal.

Theorem 5.24 *For all $n \in \mathbb{N}$, all $i, j \in E$ and all $t \geq 0$, we have:*

1. $\mathbb{P}(J_{n+1} = j, T_{n+1} - T_n \leq t \mid \mathcal{F}_{T_n}) = \mathbb{P}(J_{n+1} = j, T_{n+1} - T_n \leq t \mid J_n)$;
2. $\mathbb{P}(J_{n+1} = j, T_{n+1} - T_n \leq t \mid J_n = i) = \mathbb{P}_i(X_{T_1} = j)(1 - e^{-\lambda_i t})$.
3. *Moreover, the sequence (J_n) is a Markov chain.*

Proof

1. We compute

$$\begin{aligned}\mathbb{P}(J_{n+1} = j, T_{n+1} - T_n \le t \mid \mathcal{F}_{T_n}) &= \mathbb{P}(X_{T_{n+1}} = j, T_{n+1} - T_n \le t \mid \mathcal{F}_{T_n})\\ &\overset{(1)}{=} \mathbb{P}(X_{T_{n+1}} = j, T_{n+1} - T_n \le t \mid X_{T_n}).\end{aligned}$$

   (1) by the strong Markov property.
2. Since $\mathbb{P}(J_{n+1} = j, T_{n+1} - T_n \le t \mid J_n = i) = \mathbb{P}(X_{T_{n+1}} = j, T_{n+1} - T_n \le t \mid X_{T_n} = i)$, we have

$$\begin{aligned}\mathbb{P}(J_{n+1} = j, T_{n+1} - T_n \le t \mid J_n = i) = &\overset{(1)}{=} \mathbb{P}_i(X_{T_1} = j, T_1 \le t)\\ &\overset{(2)}{=} \mathbb{P}_i(X_{T_1} = j)\mathbb{P}_i(T_1 \le t)\\ &\overset{(3)}{=} \mathbb{P}_i(X_{T_1} = j)(1 - e^{-\lambda_i t}).\end{aligned}$$

   (1) by homogeneity, (2) by independence of X_{T_1} and T_1 and (3) by Proposition 5.22.
3. Letting t go to infinity in 2. yields the result. □

The chain (J_n) is called the embedded Markov chain of the process, with transition function P defined by $P(i, j) = \mathbb{P}_i(X_{T_1} = j)$; see Corollary 5.27 below for a closed-form expression. We check that $P(i, j) \ge 0, P(i, i) = 0$ and $\sum_{j \in E} P(i, j) = 1$.

The sojourn times in the different states are mutually independent given the successive states visited by the process. Applying iteratively Theorem 5.24 and Proposition 5.22 yields the following closed-form expression.

Corollary 5.25 *For all $n \in \mathbb{N}^*$, all $i_0, \ldots, i_{n-1} \in E$, all $k = 1, \ldots, n$, and all $t_k \ge 0$, we have*

$$\begin{aligned}&\mathbb{P}(T_1 - T_0 \le t_1, \ldots, T_n - T_{n-1} \le t_n \mid J_k = i_k, k \ge 0) =\\ &= \mathbb{P}(T_1 - T_0 \le t_1 \mid J_0 = i_0) \ldots \mathbb{P}(T_n - T_{n-1} \le t_n \mid J_{n-1} = i_{n-1})\\ &= \prod_{k=0}^{n-1} (1 - e^{-\lambda_{i_k} t_{k+1}}).\end{aligned}$$

Theorem 5.26 (Kolmogorov's Integral Equation) *For any non absorbing state i,*

$$P_t(i, j) = I(i, j)e^{-\lambda_i t} + \sum_{k \in E} \int_0^t \lambda_i e^{-\lambda_i s} P(i, k) P_{t-s}(k, j) ds, \quad t \ge 0, \ j \in E.$$

For an absorbing state i, the above theorem amounts to $P_t(i, j) = I(i, j)$.

Proof We have

$$P_t(i, j) = \mathbb{P}_i(X_t = j, T_1 > t) + \mathbb{P}_i(X_t = j, T_1 \leq t).$$

We compute $\mathbb{P}_i(X_t = j, T_1 > t) = I(i, j)e^{-\lambda_i t}$ and

$$\begin{aligned}
\mathbb{P}_i(X_t = j, T_1 \leq t) &= \mathbb{E}_i[\mathbb{P}_i(X_t = j, T_1 \leq t \mid \mathcal{F}_{T_1})] \\
&= \mathbb{E}_i[\mathbb{1}_{(T_1 \leq t)}\mathbb{P}_{X_{T_1}}(X_{t-T_1} = j)] \\
&= \int_0^{+\infty} \mathbb{1}_{]0,t]}(s) \sum_{k \in E} \mathbb{P}_k(X_{t-s} = j)\mathbb{P}_i(T_1 \in ds, X_{T_1} = k) \\
&= \int_0^t \sum_{k \in E} P_{t-s}(k, j)P(i, k)\lambda_i e^{-\lambda_i s} ds.
\end{aligned}$$
□

We can now link the distribution of the first jump time to the distribution of the hitting time of the complementary set of i, in other words compute the transition matrix P of the embedded chain in terms of the generator of the jump Markov process.

Corollary 5.27 *For any state i, we have $\lambda_i = a_i$. If i is non absorbing, then*

$$P(i, j) = \mathbb{P}_i(X_{T_1} = j) = \begin{cases} a_{ij}/a_i & \text{if } j \neq i \\ 0 & \text{if } j = i. \end{cases} \tag{5.6}$$

Proof According to Proposition 5.13, the transition function is continuous with respect to t. Differentiating the Kolmogorov's integral equation yields

$$P_t'(i, j) = -\lambda_i e^{-\lambda_i t} I(i, j) + P(i, j)\lambda_i e^{-\lambda_i t}. \tag{5.7}$$

Thus, for $t \to 0^+$ and $i = j$, we get $a_{ii} = -\lambda_i$, and, for $t \to 0^+$ and $i \neq j$, we get $a_{ij} = P(i, j)a_i$. □

Equation (5.7) implies that $a_{ij} = -a_i I(i, j) + a_i P(i, j)$, or under matrix form,

$$A = \operatorname{diag}(a_i)(P - I). \tag{5.8}$$

The above corollary yields the stochastic simulation of a trajectory of a jump Markov process in a given interval of time $[0, T]$. Indeed, the method presented in Sect. 3.1.1 applies to the embedded chain, and simulation of the sojourn times amounts to simulation of the exponential distribution as follows.

1. Let x_0 be the realization of a random variable $J_0 \sim \alpha$. $n := 0$. $T_0(\omega) := 0$.
2. $n := n + 1$. Let $W_n(\omega)$ be the realization of an $\mathcal{E}(a_{x_{n-1}})$-distributed random variable. $T_n(\omega) := T_{n-1}(\omega) + W_n(\omega)$.
3. If $T_n(\omega) \geq T$, then end.
4. Let $J_n(\omega)$ be the realization of a random variable whose distribution is given by (5.6); set $x_n := J_n(\omega)$.
5. Continue at Step 2.

▷ *Example 5.28 (Birth Process (Continuation of Example 5.10))* The transition matrix of the embedded chain of a birth process is given by

$$P(i, j) = \begin{cases} 1 & \text{if } j = i + 1, \\ 0 & \text{otherwise.} \end{cases}$$

The sojourn times have exponential distributions with parameters λ_i. ◁

The following example is an application of Markov processes to reliability. Another example is presented in Exercise 5.4.

▷ *Example 5.29 (Cold Stand-by System)* A cold stand-by system generally contains one component (or sub-system) functioning and one or several components (or sub-systems) in stand-by, all identical. The stand-by component is tried when the functioning component fails, and then begins working successfully or not according to a given probability.

Consider a stand-by system with two components. When the functioning component fails, the stand-by component is connected through a switching device. This latter commutes successfully with probability $p \in]0, 1]$. The failure rate of the functioning component is λ. The failure rate of the stand-by component is null. This system can be modelled by a Markov process with three states:

state 1: one component works, the second is in stand-by;
state 2: one component is failed, the second works;
state 3: either both components or one component and the commutator are failed.

The repairing rate between state 3 and state 2, and between state 2 and state 1 is μ; the direct transition from state 3 to state 1 is impossible; see Fig. 5.3.

The generator of the process is

$$A = \begin{pmatrix} -\lambda & p\lambda & (1-p)\lambda \\ \mu & -(\lambda + \mu) & \lambda \\ 0 & \mu & -\mu \end{pmatrix},$$

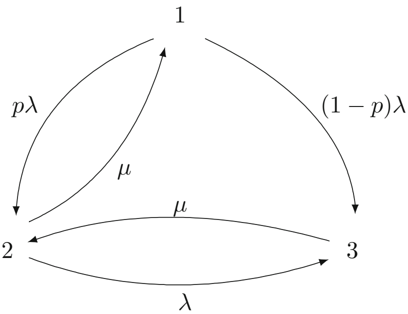

Fig. 5.3 Graph of the 3-state Markov process of Example 5.29

and

$$P = \begin{pmatrix} 0 & p & (1-p) \\ \mu/(\lambda+\mu) & 0 & \lambda/(\lambda+\mu) \\ 0 & 1 & 0 \end{pmatrix}$$

is the transition matrix of its embedded chain. $\lhd$

Let us now state some criterion linked to regularity.

Proposition 5.30 *For a uniformisable jump Markov process, the probability of more than one jump occurring in a time interval* $[0, h]$ *is* $o(h)$ *when* $h \to 0^+$, *uniformly in* $i \in E$.

Proof We have

$$\mathbb{P}_i(T_2 \le h) = \mathbb{P}_i[T_1 + (T_2 - T_1) \le h] \le \mathbb{P}_i(T_1 \le h, T_2 - T_1 \le h)$$

and

$$\begin{aligned} \mathbb{P}_i(T_1 \le h, T_2 - T_1 \le h) &= \mathbb{E}_i[\mathbb{P}_i(T_1 \le h, T_2 - T_1 \le h \mid \mathcal{F}_{T_1})] \\ &= \mathbb{E}_i[\mathbb{1}_{(T_1 \le h)} \mathbb{P}_i(T_2 - T_1 \le h \mid \mathcal{F}_{T_1})] \\ &\overset{(1)}{=} \mathbb{E}_i(\mathbb{1}_{(T_1 \le h)}[1 - \exp(-a_{X_{T_1}} h)]) \\ &\le \mathbb{E}_i[\mathbb{1}_{(T_1 \le h)}(1 - e^{-ah})] \le (1 - e^{-ah})^2 = o(h), \end{aligned}$$

where $a_{X_{T_1}} = a_j$ on the event $(X_{T_1} = j)$ and $a = \sup_{i \in E} a_i < +\infty$.
(1) by Proposition 5.22. $\square$

The following two criterion for a jump Markov process to be regular are stated without proofs.

Theorem 5.31 *A jump Markov process is regular if and only if one of the following conditions is fulfilled:*

1. ***(Reuter's condition of explosion)*** *the only bounded nonnegative solution of the equation $Ay = y$ is the null solution.*
2. $\sum_{n \geq 1} 1/a_{J_n} = +\infty$ *a.s., where $a_{J_n} = a_j$ on the event $(J_n = j)$.*

▷ *Example 5.32 (Birth Process (Continuation of Example 5.10))* If $\sum_{i \in E} 1/a_i$ is infinite, the birth process is regular; if the sum is finite, it is explosive. ◁

Proposition 5.33 *The process is regular if one of the two following conditions is fulfilled:*

1. *The process is uniformisable.*
2. *Its embedded chain is recurrent.*

Thus, a finite (such that $|E| < +\infty$) process is regular.

As for Markov chains, the states of a jump Markov process are classified according to their nature, and a communication relation can be defined. Recall that T_1 denotes the first jump time of the process.

Definition 5.34 Let $i \in E$. If $\mathbb{P}_i(\sup\{t \geq 0 : X_t = i\} = +\infty) = 1$, the state i is said to be recurrent. Otherwise, that is if $\mathbb{P}_i(\sup\{t \geq 0 : X_t = i\} < +\infty) = 1$, it is said to be transient.

If i is recurrent, then either $\mu_i = \mathbb{E}_i(\inf\{t \geq T_1 : X_t = i\}) < +\infty$ and i is said to be positive recurrent, or $\mu_i = +\infty$ and i is said to be null recurrent. The quantity μ_i is called the mean recurrence time to i.

If all states are positive recurrent, the process is said to be positive recurrent too. Such a process is regular.

Theorem 5.35 *A state is recurrent (transient) for the jump Markov process if and only if it is recurrent (transient) for its embedded Markov chain.*

Proof The absorbing case is clear. If i is recurrent and not absorbing, then $\sup\{t \in \mathbb{R}_+^* : X_t = i\} = +\infty$, a.s.

If $N = \sup\{n \in \mathbb{N}^* : X_{T_n} = i\}$ was a.s. finite, then T_{N+1} would be finite too, hence a contradiction.

Similar arguments yield the converse and the transient case. □

If i and j are two states in E, then i is said to lead to j if $P_t(i, j) > 0$ for some $t > 0$. If i leads to j and j leads to i, the states i and j are said to be communicating. The communication relation is an equivalence relation on E. If all states are communicating, the process is said to be irreducible.

Considering the embedded Markov chain, the following result is clear.

Corollary 5.36 *Recurrence and transience are class properties.*

5.1.5 Stationary Distribution and Asymptotic Behavior

The stationary measures are linked to the asymptotic behavior of the jump Markov processes, exactly as for Markov chains. Note that measures and distributions are represented for finite state spaces by line vectors.

Definition 5.37 Let $\mathbf{X}$ be a jump Markov process, with generator A and transition function P_t. A measure π on $(E, \mathcal{P}(E))$ is said to be stationary or invariant (for A or $\mathbf{X}$) if $\pi P_t = \pi$ for all real numbers $t \geq 0$. If, moreover π is a probability measure, it is called a stationary distribution of the process.

For uniformisable processes, stationary measures are solutions of a homogeneous linear system.

Proposition 5.38 *If the process is uniformisable, a measure π is stationary if and only if $\pi A = 0$.*

Proof Thanks to (5.5) p. 223, $\pi = \pi P_t$ is equivalent to

$$\pi = \pi\Big(I + \sum_{k\geq 1} \frac{t^t}{k!} A^k\Big) = \pi + \pi \sum_{k\geq 1} \frac{t^t}{k!} A^k, \quad t \geq 0.$$

This is satisfied if and only if

$$\sum_{k\geq 1} \frac{t^k}{k!} \pi A^k = 0, \quad t \geq 0,$$

or $\pi A^k = 0$ for all k, from which the result follows. □

▷ *Example 5.39 (Cold Stand-by System (Continuation of Example 5.29))* The stationary distribution of the Markov process modeling the cold stand-by system is solution of $\pi A = 0$. In other words,

$$\pi(1) = \frac{2\mu^2}{d}, \quad \pi(2) = \frac{2\lambda\mu}{d}, \quad \pi(3) = \frac{\lambda(\lambda + \mu - p\mu)}{d},$$

where $d = 3\lambda\mu - p\mu\lambda + 2\mu^2 + \lambda^2$. ◁

Definition 5.40 A measure (or distribution) λ on E is said to be reversible (for A or $\mathbf{X}$) if

$$\lambda(i)a_{ij} = \lambda(j)a_{ji}, \quad i \in E,\ j \in E. \tag{5.9}$$

Proposition 5.41 *All reversible distributions are stationary.*

Proof If λ is reversible then summing both sides of (5.9) on $i \in E$, we get $\sum_{i\in E} \lambda(i)a_{ij} = 0$, and hence, according to Proposition 5.38, λ is stationary. □

The stationary distributions of the process $\mathbf{X}$ and its embedded chain (J_n) are not equal, but they are closely linked.

Proposition 5.42 *If π is the stationary distribution of the process $\mathbf{X}$ and ν the stationary distribution of the embedded chain (J_n), then*

$$\pi(i)a_i = \nu(i) \sum_{k\in E} a_k \pi(k), \quad i \in E,$$

Proof We deduce from both $\pi A = 0$ and (5.8) p. 227 that $\pi D P = \pi D$, where $D = \operatorname{diag}(a_i)$. Therefore, πD is an invariant measure of P, and hence $\nu = \pi D / \sum_{i\in E} a_i \pi(i)$. □

▷ *Example 5.43 (Continuation of Example 5.21)* This two-state process has a reversible distribution satisfying the equations $\pi(0)\lambda = \pi(1)\mu$ and $\pi(1) + \pi(2) = 1$, so

$$\pi(0) = \frac{\mu}{\lambda+\mu} \quad \text{and } \pi(1) = \frac{\lambda}{\lambda+\mu}.$$

This distribution is stationary for the process. ◁

Definition 5.44 An irreducible jump Markov process whose all states are positive recurrent is said to be ergodic.

Note that the embedded Markov chain of an ergodic jump Markov process is not ergodic itself in general, as shown in the next example.

▷ *Example 5.45 (Continuation of Example 5.21)* An irreducible two-state jump Markov process is ergodic, but its embedded chain is never ergodic because it is 2-periodic. ◁

The entropy rate of an ergodic jump Markov process has an explicit expression given by the following result that we state without proof.

Proposition 5.46 *Let* $\mathbf{X}$ *be a jump Markov process with generator* (a_{ij}) *and stationary distribution* π*. Its entropy rate is*

$$\mathbb{H}(\mathbf{X}) = -\sum_{i\in E}\pi(i)\sum_{j\neq i}a_{ij}\log a_{ij} + \sum_{i\in E}\pi(i)\sum_{j\neq i}a_{ij},$$

if this quantity is finite.

The Lagrange multipliers method yields that the jump Markov process with finite state space E having the maximum entropy rate is that with a uniform generator.

The next result leads to characterize the asymptotic behavior of an ergodic jump Markov process in the following two theorems.

Lemma 5.47 *If* X *is an ergodic Markov process with stationary distribution* π*, then the mean recurrence time of any state* $i \in E$ *is given by*

$$\mu_i = \frac{1}{a_i\pi(i)}.$$

Proof The embedded chain (J_n) is an irreducible and recurrent Markov chain. By Proposition 5.42, its stationary distribution ν is given by $\nu(j) = \pi(j)a_j$ for all $j \in E$.

Suppose J_n starts from state j. The expectation of the first jump time—or mean sojourn time in j—is $1/a_j$. Further, the expectation of the number of visits of (J_n) to state i before return to j, is $\nu(j)/\nu(i) = \pi(j)a_j/\pi(i)a_i$; see Theorem 3.44 and Proposition 3.45.

Therefore

$$\mu_i = \sum_{j\in E}\frac{\pi(j)a_j}{\pi(i)a_i}\frac{1}{a_j} = \frac{1}{\pi(i)a_i}$$

for all states $i \in E$. □

Theorem 5.48 *Let* $\mathbf{X}$ *be an ergodic jump Markov process. For all states* i *and* j*, we have*

$$P_t(i,j) \longrightarrow \frac{1}{a_j\mu_j} = \pi(j), \quad t \to +\infty,$$

where μ_j *is the mean recurrence time of state* j*.*

Proof Thanks to Chapman-Kolmogorov equation, $P_{nh}(i,j) = (P_h)^n(i,j)$ for any fixed $h > 0$ and $n \in \mathbb{N}$. Thanks to the ergodic Theorem 3.50, we know that $(P_h)^n(i,j)$ converges to $\pi(j)$ when n tends to infinity.

For any $\varepsilon > 0$, some integer N exists such that

$$|P_{nh}(i,j) - \pi(j)| \leq \varepsilon/2 \qquad \text{for all} \quad n \geq N.$$

From Lévy's theorem, for any $t \geq 0$, some $h > 0$ exists such that

$$|P_{t+h}(i,j) - P_t(i,j)| \leq 1 - P_h(i,i) \leq \varepsilon/2.$$

Thus, for $nh \leq t < (n+1)h$ and $n > N$, we get $|P_t(i,j) - P_{nh}(i,j)| \leq \varepsilon/2$.
Finally,

$$|P_t(i,j) - \pi(j)| \leq |P_t(i,j) - P_{nh}(i,j)| + |P_{nh}(i,j) - \pi(j)| \leq \varepsilon,$$

and the result follows. □

Theorem 5.49 (Ergodic) *If* $\mathbf{X}$ *is an ergodic jump Markov process, then, for all states* i *and* j*,*

$$\frac{1}{t}\int_0^t \mathbb{1}_{(X_u=i)}du \longrightarrow \pi(i), \quad t \to +\infty,\ \mathbb{P}_j - \text{a.s.}$$

Proof Let (W_n) denote the i.i.d. sequence of successive sojourn times and let $N_i(t)$ be the number of visits in the time interval $]0,t]$ to a given state i. Then

$$\frac{W_1 + \cdots + W_{N_i(t)-1}}{t} \leq \frac{1}{t}\int_0^t \mathbb{1}_{(X_u=i)}du \leq \frac{W_1 + \cdots + W_{N_i(t)}}{t}$$

or

$$\frac{W_1 + \cdots + W_{N_i(t)-1}}{N_i(t) - 1}\,\frac{N_i(t) - 1}{t} \leq \frac{1}{t}\int_0^t \mathbb{1}_{(X_u=i)}du \tag{5.10}$$

and

$$\frac{1}{t}\int_0^t \mathbb{1}_{(X_u=i)}du \leq \frac{W_1 + \cdots + W_{N_i(t)}}{N_i(t)}\,\frac{N_i(t)}{t}. \tag{5.11}$$

Thanks to Theorem 4.17,

$$\frac{W_1 + \cdots + W_{N_i(t)}}{N_i(t)} \xrightarrow{\text{a.s.}} \mathbb{E}\, W_1 = \frac{1}{a_i}.$$

Thanks to Proposition 4.59, $N_i(t)/t$ converges a.s. to $1/\mu_i$. The result follows from inequalities (5.10) and (5.11) for t tending to infinity. □

Therefore, for ergodic jump Markov processes, if g is a real function defined on E, the time mean is equal to the space mean of the function, that is

$$\frac{1}{t}\int_0^t g(X_u)du = \sum_{i\in E} g(i)\frac{1}{t}\int_0^t \mathbb{1}_{(X_u=i)}du \longrightarrow \sum_{i\in E} g(i)\pi(i), \quad t\to+\infty,\ \mathbb{P}_j-\text{a.s.},$$

provided that $\sum_{i\in E}|g(i)|\pi(i) < +\infty$.

Thanks to the dominated convergence theorem, the following result is clear.

Corollary 5.50 *If* $\mathbf{X}$ *is an ergodic jump Markov process, then, for all states* i *and* j,

$$\frac{1}{t}\int_0^t P_u(j,i)du \longrightarrow \pi(i), \quad t\to+\infty.$$

$\triangleright$ *Example 5.51 (Uniformisation method)* Using the stationary distribution provides a numerical method for solving Kolmogorov's equations for an ergodic process.

Let $\mathbf{X}$ be an ergodic uniformisable process, with stationary distribution π, with $a=\sup_{i\in E} a_i < +\infty$. The matrix $Q = I + a^{-1}A$ is stochastic. We compute

$$P_t = e^{tA} = e^{ta(Q-I)} = e^{-at}e^{atQ} = e^{-at}\sum_{n\geq 0}\frac{(at)^n}{n!}Q^n.$$

Let Π be the $E\times E$-matrix defined by $\Pi(i,j)=\pi(j)$, for all i and j. We can write

$$P_t = \Pi + e^{-at}\sum_{n\geq 0}\frac{(at)^n}{n!}(Q^n-\Pi).$$

The system $\pi A = 0$ is equivalent to $\pi a(Q-I)=0$, or to $\pi Q = \pi$. Thus, Q has the same invariant distribution π as P_t. Therefore, $Q^n - \Pi$ converges to zero when n tends to infinity.

If α is a distribution on E, one can show that

$$\sup_{t\geq 0}\|\alpha P_t - \alpha P_t(k)\| \longrightarrow 0, \quad k\to+\infty,$$

where

$$P_t(k) = \Pi + e^{-at}\sum_{n=0}^{k}\frac{(at)^n}{n!}(Q^n-\Pi).$$

Note that the truncating level k can be chosen such that the error is bounded for some t by an ε, and then it will be bounded for all $t\geq 0$. $\triangleleft$

5.2 Semi-Markov Processes

This section is dedicated to the investigation of semi-Markov processes, mainly with finite state spaces. The semi-Markov processes constitute a natural generalization of the Markov and renewal processes. Their future evolution depends on both the occupied state and the time elapsed since the last transition. This time, called local time, is measured by a watch that comes back to zero at each transition. Of course, if the watch is considered as an integral part of the system—in other words if $E \times \mathbb{R}_+$ becomes the state space (where E is the state space of the semi-Markov process), then the process becomes a Markov process.

5.2.1 Markov Renewal Processes

In order to define semi-Markov processes, it is easier first to define Markov renewal processes.

Definition 5.52 Let $(J_n, T_n)_{n\in\mathbb{N}}$ be a process defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that (J_n) is a random sequence taking values in a discrete set E and (T_n) is an increasing random sequence taking values in $\mathbb{R}_+$, with $T_0 = 0$. Set $\mathcal{F}_n = \sigma(J_k, T_k; k \leq n)$. The process (J_n, T_n) is called a Markov renewal process with discrete state space E if

$$\mathbb{P}(J_{n+1} = j, T_{n+1} - T_n \leq t \mid \mathcal{F}_n) = \mathbb{P}(J_{n+1} = j, T_{n+1} - T_n \leq t \mid J_n) \quad \text{a.s.},$$

for all $n \in \mathbb{N}$, all $j \in E$ and all $t \in \mathbb{R}_+$.

If the above conditional probability does not depend on n, the process is said to be homogeneous, and we set

$$Q_{ij}(t) = \mathbb{P}(J_{n+1} = j, T_{n+1} - T_n \leq t \mid J_n = i) \tag{5.12}$$

for all $n \in \mathbb{N}$, $i \in E$, $j \in E$ and $t \in \mathbb{R}_+$. The family $Q = \{Q_{ij}(t);\ i, j \in E, t \in \mathbb{R}_+\}$ is called a semi-Markov kernel on E and the square matrix $Q(t) = (Q_{ij}(t))_{(i,j)\in E\times E}$ is a semi-Markov matrix.

We will study here only homogeneous Markov renewal processes with finite state spaces, say $E = \{1, \ldots, e\}$.

The process (J_n, T_n) is a two-dimensional Markov chain, with state space $E \times \mathbb{R}_+$; its transition function is the semi-Markov kernel Q. Letting t go to infinity in (5.12) shows that (J_n) is a Markov chain with state space E and transition matrix $P = (P(i, j))$ where

$$P(i, j) = \lim_{t\to+\infty} Q_{ij}(t), \quad i, j \in E.$$

▷ *Example 5.53* The following processes are Markov renewal processes:

1. For a renewal process, only one state is visited—say $E = \{1\}$, and $Q(t) = Q_{11}(t) = F(t)$, which is a scalar function.
2. For an alternated renewal process, $E = \{1, 2\}$, $P(1, 2) = 1$, $P(2, 1) = 1$ and

$$Q(t) = \begin{pmatrix} 0 & F(t) \\ G(t) & 0 \end{pmatrix};$$

see Exercise 4.4 below for further details on this process. ◁

Definition 5.54 The stochastic process $\mathbf{Z} = (Z_t)_{t \in \mathbb{R}_+}$, defined by

$$Z_t = J_n \quad \text{if} \quad T_n \leq t < T_{n+1}$$

is called the semi-Markov process associated with the Markov renewal process (J_n, T_n).

The process $\mathbf{Z}$ is continuous on the right. Clearly $J_n = Z_{T_n}$, meaning that (J_n) is the sequence of the successive states visited by $\mathbf{Z}$. As for Markov processes, (J_n) is called the embedded chain of the process.

The initial distribution α of $\mathbf{Z}$ (that is the distribution of Z_0) is also the initial distribution of (J_n). Knowledge of both α and the kernel Q characterizes the distribution of $\mathbf{Z}$.

Properties of Semi-Markov Kernels

1. For all $i, j \in E$, the function $t \longrightarrow Q_{ij}(t)$ is a defective distribution function on $\mathbb{R}_+$. On the contrary, $H_i(t) = \sum_{j \in E} Q_{ij}(t)$ for $t \geq 0$ defines the distribution function of the total time spent by $\mathbf{Z}$ in $[0, t]$ at i, called sojourn time. We will write $H(t) = \operatorname{diag}(H_i(t))_{i \in E}$, and denote by m_i the mean sojourn time in state i, that is $m_i = \mathbb{E}_i(T_1) = \int_0^{+\infty} (1 - H_i(t))\, dt$.
2. $Q_{ij}(t) = P(i, j) F_{ij}(t)$, where $F_{ij}(t) = \mathbb{P}(T_{n+1} - T_n \leq t \mid J_n = i, J_{n+1} = j)$ is the distribution function of the time spent by $\mathbf{Z}$ in state i conditional on transition to state j.
3. $\mathbb{P}(J_n = j, T_n \leq t \mid J_0 = i) = Q_{ij}^{*(n)}(t)$ for $n > 0$, where $Q_{ij}^{*(n)}$ is the n-th Lebesgue-Stieltjes convolution of Q_{ij}, that is

$$Q_{ij}^{*(n)}(t) = \sum_{k \in E} \int_0^t Q_{ik}(ds) Q_{kj}^{*(n-1)}(t - s), \quad n \geq 2,$$

with $Q^{*(1)} = 0$ and $Q_{ij}^{*(0)} = \delta_{ij}$.

▷ *Example 5.55 (Kernel of a jump Markov process)* For a jump Markov process—with generator (a_{ij}), the sequences (J_n) and (T_n) are $\mathbb{P}_i$-independent. We obtain from Theorem 5.24 that

$$Q_{ij}(t) = \mathbb{P}_i(J_1 = j)(1 - e^{-a_i t}) = \frac{a_{ij}}{a_i}(1 - e^{-a_i t}), \quad i \neq j \in E,$$

and $Q_{ii}(t) = 0$ for $t \in \mathbb{R}_+$, where $a_i = -a_{ii} = \sum_{j \neq i} a_{ij}$. ◁

Let $N_j(t)$ be the number of visits of $\mathbf{Z}$ to state j in the time interval $]0, t]$, and set $N(t) = (N_1(t), \dots, N_e(t))$. Note that the Markov renewal process may alternatively be defined from the process $\mathbf{N} = (N(t))_{t \in \mathbb{R}_+}$.

Definition 5.56 The function

$$t \longrightarrow \psi_{ij}(t) = \mathbb{E}_i N_j(t) = \sum_{n \geq 0} Q_{ij}^{*(n)}(t)$$

is called a Markov renewal function; we will write $\psi(t) = (\psi_{ij}(t))_{(i,j) \in E \times E}$.

We compute

$$\psi_{ij}(t) = I(i, j)\mathbb{1}_{\mathbb{R}_+}(t) + \sum_{k \in E} Q_{ik} * \psi_{kj}(t),$$

or, under matrix form, $\psi(t) = \operatorname{diag}(\mathbb{1}_{\mathbb{R}_+}(t)) + Q * \psi(t)$. This equation is a particular case of the Markov renewal equation

$$L(t) = G(t) + Q * L(t), \quad t \in \mathbb{R}_+,$$

where G (given) and L (unknown) are matrix functions null on $\mathbb{R}_-$ and bounded on the finite intervals of $\mathbb{R}_+$. When it exists, the solution takes the form

$$L(t) = \psi * G(t), \quad t \in \mathbb{R}_+.$$

We assume here that none of the functions H_i, for $i \in E$, is degenerated (that is $H_i(t) \neq \mathbb{1}_{(t \geq 0)}$).

Definition 5.57 The transition function of $\mathbf{Z}$ is defined by

$$P_t(i, j) = \mathbb{P}(Z_t = j \mid Z_0 = i), \quad i, j \in E, \ t \in \mathbb{R}_+,$$

and we will write in matrix form $P(t) = (P_t(i, j))_{(i,j) \in E \times E}$.

Proposition 5.58 *The transition function* $P(t)$ *of* $\mathbf{Z}$ *is solution of the Markov renewal equation*

$$P(t) = \mathrm{diag}(1 - H_i(t)) + Q * P(t). \tag{5.13}$$

Proof We have

$$P_t(i, j) = \mathbb{P}(Z_t = j \mid Z_0 = i) = \mathbb{P}_i(Z_t = j, T_1 > t) + \mathbb{P}_i(Z_t = j, T_1 \leq t).$$

Since $\mathbb{P}_i(Z_t = j, T_1 > t) = [1 - H_i(t)]I(i, j)$ and

$$\begin{aligned}
\mathbb{P}_i(Z_t = j, T_1 \leq t) &= \mathbb{E}_i\left[\mathbb{P}_i(Z_t = j, T_1 \leq t \mid \mathcal{F}_{T_1})\right] \\
&= \mathbb{E}_i\left[\mathbb{1}_{(T_1 \leq t)}\mathbb{P}_{Z_{T_1}}(Z_{t-T_1} = j)\right] \\
&= \sum_{k \in E} \int_0^t Q_{ik}(ds)P_{t-s}(k, j),
\end{aligned}$$

(5.13) follows, under matrix form. □

The unique solution of (5.13) is given by

$$P(t) = [I - Q(t)]^{*(-1)} * [I - H(t)], \tag{5.14}$$

where $[I - Q(t)]^{*(-1)} = \psi(t) = \sum_{n \geq 0} Q^{*(n)}(t)$.

5.2.2 Classification of States and Asymptotic Behavior

Let $(T_n^j)_{n \in \mathbb{N}}$ be the sequence of successive times of visit of the semi-Markov process $\mathbf{Z}$ to state $j \in E$. It is a renewal process, possibly modified. Thus, T_0^j is the time of the first visit to j and $G_{ij}(t) = \mathbb{P}(T_0^j \leq t \mid Z_0 = i)$ is the distribution function of the time of the first transition from state i to state j. If $i = j$, then $T_0^j = 0$, and hence $G_{jj}(t) = \mathbb{P}(T_1^j \leq t \mid Z_0 = j)$ is the distribution function of the time between two successive visits to j. We have

$$\psi_{jj}(t) = \sum_{n \geq 0} G_{jj}^{*(n)}(t),$$

and for $i \neq j$,

$$\psi_{ij}(t) = \sum_{n \geq 0} G_{ij} * G_{jj}^{*(n)}(t) = G_{ij} * \psi_{jj}(t).$$

The expectation of the hitting time of j starting from i at time 0 is

$$\mu_{ij} = \int_0^{+\infty} (1 - G_{ij}(t))\, dt, \quad i, j \in E. \tag{5.15}$$

When $i = j$, it is the mean return time to i.

As for Markov processes, two states i and j are said to be communicating if $G_{ij}(+\infty)G_{ji}(+\infty) > 0$ or if $i = j$. The communication relation is an equivalence relation on E. If all states are communicating, the process is said to be irreducible.

Definition 5.59 If $G_{ii}(+\infty) = 1$ or $\psi_{ii}(+\infty) = +\infty$, the state i is said to be recurrent. Otherwise, it is said to be transient.

If i is recurrent, then either its mean return time μ_{ii} is finite and i is said to be positive recurrent, or it is infinite and i is said to be null recurrent.

A state i is said to be periodic with period $h > 0$ if G_{ii} is arithmetic with period h. Then, $\psi_{ii}(t)$ is constant on the intervals of the form $[nh, nh + h[$, where h is the largest number sharing this property. Otherwise, i is said to be aperiodic. If all states are aperiodic, the process is said to be aperiodic. Note that this notion of periodicity for semi-Markov processes is different from the notion of periodicity seen in Chap. 3 for Markov chains.

The following result is a straightforward consequence of Blackwell's renewal theorem applied to the renewal process (T_n^i).

Theorem 5.60 (Markov Renewal) *If the state i is aperiodic, then for any $c > 0$,*

$$\psi_{ii}(t) - \psi_{ii}(t - c) \longrightarrow \frac{c}{\mu_{ii}}, \quad t \to +\infty. \tag{5.16}$$

When i is periodic with period h, the result remains valid if c is a multiple of h.

We state the next result without proof.

Theorem 5.61 (Key Markov Renewal) *If (J_n) is irreducible and aperiodic, if ν is an invariant measure for P and if $m_i < +\infty$ for all $i \in E$, then for all direct Riemann integrable real functions g_i defined on $\mathbb{R}_+$ for $i \in E$,*

$$\int_0^t g_i(t - y)\psi_{ji}(dy) \longrightarrow \frac{\nu(i)}{< \nu, m >} \int_0^{+\infty} g_i(y)dy, \quad t \to +\infty.$$

Thus, thanks to the key Markov renewal theorem and (5.14) p. 239, we get

$$\pi(j) = \lim_{t \to +\infty} P_t(i, j) = \frac{\nu(j)m_j}{< \nu, m >}, \quad i, j \in E,$$

which defines the limit distribution π of $\mathbf{Z}$.

The entropy rate of a semi-Markov process has an explicit form under suitable conditions. It is given by the next proposition stated without proof.

Proposition 5.62 *Let* $\mathbf{Z}$ *be a semi-Markov process such that*

$$Q_{ij}(t) = \int_0^t q_{ij}(x)dx, \quad \textit{with} \quad \int_{\mathbb{R}_+} q_{ij}(t)|\log q_{ij}(t)|dt < +\infty, \quad i, j \in E.$$

If $m_i < +\infty$ *for all* $i \in E$*, then*

$$\mathbb{H}(\mathbf{X}) = \frac{-1}{< \nu, m >} \sum_{i,j \in E} \nu(i) \int_0^{+\infty} q_{ij}(x) \log q_{ij}(x)dx.$$

▷ *Example 5.63 (Analysis of Seismic Risk)* We consider here two simplified models. The intensity of an earthquake is classified according to a discrete ladder of states in $E = \{1, \dots, N\}$. The process $\mathbf{Z} = (Z(t))_{t \in \mathbb{R}_+}$ is defined by $Z(t) = i$ for $1 \leq i < N$ if the intensity of the last earthquake before time t was $f \in [i, i+1[$ and finally $Z(t) = N$ if $f \geq N$.

For a time predictable model, it is assumed that the stronger is an earthquake, the longer is the time before the next occurs. The stress accumulated on a given rift has a minimum bound. When a certain level of stress is reached, an earthquake occurs. Then the stress decreases to the minimum level; see Fig. 5.4. The semi-Markov kernel of $\mathbf{Z}$ is then

$$Q_{ij}(t) = \nu(j)F_j(t), \quad t \in \mathbb{R}_+.$$

For a slip predictable model, it is assumed that the longer is the time elapsed since the last earthquake, the stronger is the next one. The stress has a maximum bound. When this level is reached, an earthquake occurs. Then the stress decreases of a certain quantity; see Fig. 5.4. The semi-Markov kernel of $\mathbf{Z}$ is

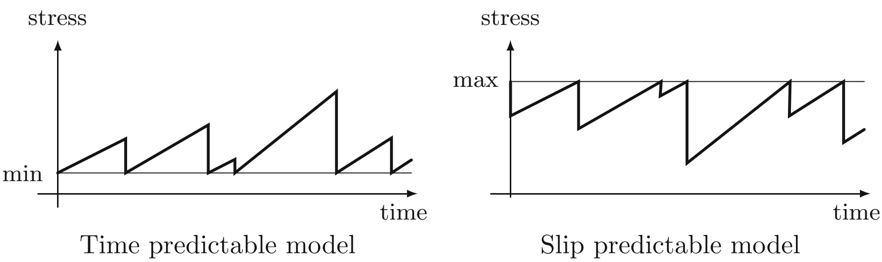

Fig. 5.4 Two semi-Markov models for seismic risk analysis

then

$$Q_{ij}(t) = \nu(j)F_i(t), \quad t \in \mathbb{R}_+.$$

In both cases, since F_{ij} depends only on one of the two states i and j, necessarily the probability ν is the stationary distribution of the embedded chain of the process. If the functions F_i are differentiable for all $i \in E$, the entropy rate of the process can be computed explicitly.

For the slip predictable model,

$$\mathbb{H}(\mathbf{Z}) = \frac{-1}{< \nu, m >} \sum_{i,j \in E} \nu(i) \int_{\mathbb{R}_+} \nu(j) f_i(t) \log[\nu(j) f_i(t)]\, dt.$$

Since

$$\int_{\mathbb{R}_+} \nu(j) f_i(t) \log[\nu(j) f_i(t)]\, dt = \nu(j) \log[\nu(j)] + \nu(j) \int_{\mathbb{R}_+} f_i(t) \log[f_i(t)]\, dt,$$

we get

$$\mathbb{H}(\mathbf{Z}) = \frac{-1}{< \nu, m >} \sum_{i \in E} \nu(i) \Big[\log \nu(i) + \int_{\mathbb{R}_+} f_i(t) \log f_i(t)\, dt \Big].$$

Similar computation for the time predictable model yields the same formula. $\triangleleft$

Other applications of semi-Markov processes, linked to reliability, will be studied in Exercises 5.5 and 5.6.

5.3 Exercises

∇ **Exercise 5.1 (Birth-and-death Process)** The Markov process $(X_t)_{t \geq 0}$ with state space $E = \mathbb{N}$ and generator $A = (a_{ij})$ defined by

$$a_{ij} = \begin{cases} \lambda_i & \text{if } j = i+1,\ i \geq 0, \\ \mu_i & \text{if } j = i-1,\ i \geq 1, \\ -(\lambda_i + \mu_i) & \text{if } j = i,\ i \geq 0, \\ 0 & \text{otherwise}, \end{cases}$$

with $\mu_0 = 0$, is called a birth-and-death process. If $\mu_i = 0$ for all i, the process is a birth process; if $\lambda_i = 0$ for all i, it is a death process.

Assume that $\lambda_i \mu_{i+1} > 0$ for all $i \geq 0$. Give a necessary and sufficient condition for the process to have a reversible distribution and give then its stationary distribution.

Solution A reversible distribution π satisfies $\pi(i-1)\lambda_{i-1} = \pi(i)\mu_i$ for $i \geq 1$, that is

$$\pi(i) = \frac{\lambda_{i-1}}{\mu_i}\pi(i-1) = \frac{\lambda_{i-1}\lambda_{i-2}}{\mu_i \mu_{i-1}}\pi(i-2) = \cdots = \frac{\lambda_{i-1}\ldots\lambda_0}{\mu_i\ldots\mu_1}\pi(0) = \gamma_i \pi(0),$$

where $\gamma_0 = 1$ and

$$\gamma_i = \frac{\lambda_0 \ldots \lambda_{i-1}}{\mu_1 \ldots \mu_i}, \quad i \geq 1.$$

Summing on $i \geq 0$ yields

$$1 = \sum_{i\geq 0} \pi(i) = \sum_{i \geq 0} \gamma_i \pi(0).$$

Therefore, the convergence of the sum $\sum_{i\geq 0} \gamma_i$ is a necessary and sufficient condition for the process to have a reversible distribution. We compute then

$$\pi(0) = \Big(\sum_{i\geq 0} \gamma_i\Big)^{-1},$$

and hence

$$\pi(i) = \frac{\gamma_i}{\sum_{k\geq 0} \gamma_k}, \quad i \geq 0.$$

Thus, according to Proposition 5.41, the reversible distribution is also a stationary distribution. △

∇ **Exercise 5.2 ($M/M/1$ Queueing Systems)** At the post office, only one customer can be served at a time. The time of service has an exponential distribution $\mathcal{E}(\mu)$. The times of arrivals of the customers form a homogeneous Poisson process with intensity λ. When a customer arrives, either he is immediately served if the server is available, or he joins the (possibly infinite) queue. Such a system is called an $M/M/1$ queueing system (Fig. 5.5). Let X_t be the random variable equal to the number of customers present in the post office at time t, for $t \in \mathbb{R}_+$.

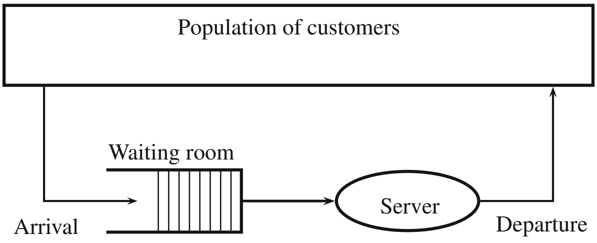

Fig. 5.5 $M/M/1$ queueing system

1. Show that $\mathbf{X} = (X_t)$ is a Markov process. Give its generator.
2. Determine the stationary distribution π of $\mathbf{X}$, when it exists.
3. The initial distribution of $\mathbf{X}$ is assumed to be π.
   a. Compute the average number of customers in the post office at a fixed time t.
   b. Determine the distribution of the time spent in the post office by a customer.
4. Compute the average time during which the post office is empty, for $\lambda/\mu = 1/2$.

Solution

1. The process $\mathbf{X}$ is a birth-death Markov process with state space $E = \mathbb{N}$ and generator determined by $\lambda_i = \lambda$, for $i \geq 0$ and $\mu_i = \mu$, for $i \geq 1$; see Exercise 5.1 for notation.
2. With the same notation,

$$\gamma_0 = 1 \quad \text{and} \quad \gamma_i = \left(\frac{\lambda}{\mu}\right)^i, \quad i \geq 1.$$

   Set $a = \lambda/\mu$. For $a < 1$, we have $\sum_{k\geq 0} \gamma_k = 1/(1-a)$, and then—and only then—$\mathbf{X}$ has a stationary distribution π, given by $\pi_i = a^i(1-a)$, for $i \geq 0$. In other words, π is a geometric distribution on $\mathbb{N}$ with parameter a.
3. a. When the initial distribution is π, the process is stationary and the expectation of X_t is the expectation of the geometric distribution $\mathcal{G}(a)$, that is $\mathbb{E}\, X_t = a/(1-a)$.
   b. Let W be the total time passed in the system by some given customer, arriving at time T_0. Clearly, the (exit) process $(M(t))$ of other customers' exit times, after T_0 and until the customer's exit, is a homogeneous Poisson process with intensity μ. We compute

$$\begin{aligned}\mathbb{P}(W > t) &= \sum_{n\geq 0} \mathbb{P}(W > t, X_{T_0^-} = n) = \sum_{n\geq 0} \mathbb{P}(W > t \mid X_{T_0^-} = n)\mathbb{P}(X_{T_0^-} = n)\\ &= \sum_{n\geq 0} \mathbb{P}(M_t \leq n)\pi(n) = \sum_{n\geq 0}\left[\sum_{k=0}^{n} e^{-\mu t}\frac{(\mu t)^k}{k!}\right] a^n(1-a) = e^{-\mu(1-a)t},\end{aligned}$$

   meaning that W has an exponential distribution $\mathcal{E}(\mu(1-a))$.

c. We have $\mathbb{E}\, X_t = \lambda \mathbb{E}\, W$. Note that this formula is a particular case of Little's formula, that characterizes ergodic queueing systems.

4. If $\lambda/\mu = 1/2$, then $a = 1/2$, and hence $\pi_0 = 1/2$. Therefore, the post office will be empty half the time.

   See Problem 5.7 for a feed back queue. △

▽ **Exercise 5.3 (Epidemiological Models)** Consider a population of m individuals. Suppose that exactly one individual is contaminated at time $t = 0$. The others can then be contaminated and the affection is incurable. Suppose that in any time length h, one infected individual can infect a healthy individual with probability $\alpha h + o(h)$ for $h \to 0^+$, where $\alpha > 0$. Let X_t be the number of individuals contaminated at time $t \geq 0$, and let T_i be the time necessary to pass from i contaminated individuals to $i + 1$, for $1 \leq i \leq m - 1$.

1. a. Suppose $X_t = i$. Compute the probability that only one individual is contaminated in the time interval $[t, t + h]$. Show that the probability that two or more individuals are contaminated in the same time interval is $o(h)$.
   b. Show that $\mathbf{X} = (X_t)$ is a Markov process; give its state space and generator.
2. Show that T_i has an exponential distribution; give its parameter.
3. Let T be the time necessary for the whole population to be contaminated; compute its mean and variance.
4. Numerical application: compute an approximate value of the mean of T for $m = 6 \times 10^7$, $\alpha = 6 \times 10^{-8}$ per day, and $h = 1$ day.

Solution

1. a. If i individuals are infected, then each individual among the $m - i$ healthy ones can be contaminated in $]t, t + h]$ with probability $i\alpha h + o(h)$. Thus, the probability that one individual among the $m - i$ will be contaminated in $]t, t + h]$ is

$$\binom{m-i}{1}[i\alpha h + o(h)]^1[1 - i\alpha h + o(h)]^{m-i-1} = (m - i)i\alpha h + o(h).$$

   Similarly, for $k \geq 2$,

$$\binom{m-i}{k}[i\alpha h + o(h)]^k[1 - i\alpha h + o(h)]^{m-i-k} = o(h).$$

   b. Therefore, (X_t) is a Markov process with state space $E = [\![1, m]\!]$ and generator $A = (a_{ij})_{(i,j)\in E\times E}$, where

$$a_{ij} = \begin{cases} (m-i)i\alpha, & \text{if } j = i+1, \\ -(m-i)i\alpha, & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases}$$

This process is a birth process, also called Yule process.

2. Proposition 5.22 and Corollary 5.27 together yield that $T_i \sim \mathcal{E}(-a_{ii})$.
3. Since $T = T_1 + \cdots + T_{m-1}$,

$$\mathbb{E}\,T = \sum_{i=1}^{m-1} \frac{1}{(m-i)i\alpha}.$$

The variables T_i, for $1 \le i \le m-1$, are independent, so

$$\mathbb{V}\mathrm{ar}\,T = \sum_{i=1}^{m-1} \frac{1}{[(m-i)i\alpha]^2}.$$

4. We compute

$$\mathbb{E}\,T = \frac{1}{m\alpha}\sum_{i=1}^{m-1}\left(\frac{1}{m-i}+\frac{1}{i}\right) \approx \frac{1}{m\alpha}\int_1^{m-1}\left(\frac{1}{m-t}+\frac{1}{t}\right)dt,$$

and $\int_1^{m-1}\left(\frac{1}{m-t}+\frac{1}{t}\right)dt = 2\log(m-1)$, so $\mathbb{E}\,T = 2\log(m-1)/m\alpha$. For the data, $\mathbb{E}\,T \approx 358$ days, around 1 year. △

∇ **Exercise 5.4 (Reliability of a Markov System)** Consider a system whose stochastic behavior is modelled by a Markov process, $\mathbf{X} = (X_t)_{t\in\mathbb{R}_+}$, with finite state space $E = [\![1, e]\!]$, generator A, transition function $P_t(i, j)$, and initial distribution α. Let $U = [\![1, m]\!]$ be the set of functioning states and $D = [\![m+1, e]\!]$ the set of failed states, for some $m \in [\![2, e-1]\!]$.

1. a. Compute the instantaneous availability $A(t)$ of the system for $t > 0$; see Exercise 4.4 for definition.
   b. Use a. to compute the limit availability when $\mathbf{X}$ is ergodic.
2. Let $T_D = \inf\{t \ge 0 : X_t \in D\}$ be the hitting time of the set D of failed states of $\mathbf{X}$, with the convention $\inf\phi = +\infty$. Consider the process $\mathbf{Y}$ with state space $U \cup \{\Delta\}$—where Δ is an absorbing state, defined by

$$Y_t = \begin{cases} X_t & \text{if } t < T_D,\\ \Delta & \text{if } t \ge T_D.\end{cases}$$

   a. Give the initial distribution and the generator of $\mathbf{Y}$, which is a Markov process.
   b. Use $\mathbf{Y}$ to compute the reliability function of the system, defined by $R(t) = \mathbb{P}(T_D > t)$ for $t \ge 0$.

Solution

1. a. The instantaneous availability is

$$A(t) = \mathbb{P}(X_t \in U) = \sum_{j\in U} \mathbb{P}(X_t = j) = \sum_{j\in U}\sum_{i\in E} \mathbb{P}(X_t = j, X_0 = i)$$
$$= \sum_{j\in U}\sum_{i\in E} \mathbb{P}(X_t = j \mid X_0 = i)\mathbb{P}(X_0 = i)$$
$$= \sum_{j\in U}\sum_{i\in E} \alpha(i)P_t(i, j) = \alpha P_t \mathbf{1}_{e,m} = \alpha e^{tA}\mathbf{1}_{e,m},$$

where $\mathbf{1}_{e,m} = (1, \dots, 1, 0, \dots, 0)'$ is the e-dimensional column vector of which the first m components are equal to 1 and the $e - m$ others are equal to 0.

b. Therefore, the limit availability is

$$A = \lim_{t\to\infty} A(t) = \alpha\Pi\mathbf{1}_{e,m} = \sum_{k\in U} \pi(k).$$

2. a. For computing the reliability, it is necessary to consider the partition of A and α between U and D, that is, $\alpha = (\alpha_1, \alpha_2)$ and

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

We can write $Y_t = X_{t\wedge T_D}$, and $\mathbf{Y}$ is indeed a Markov process with generator

$$B = \begin{pmatrix} A_{11} & A_{12}\mathbf{1} \\ \mathbf{0} & 0 \end{pmatrix}.$$

Its initial distribution is $\beta = (\alpha_1, b)$ with $b = \alpha_2\mathbf{1}$.

b. Let Q_t be the transition function of $\mathbf{Y}$. We have

$$R(t) = \mathbb{P}(\forall u \in [0, t], X_u \in U) = \mathbb{P}(Y_t \in U) = \sum_{j\in U} \mathbb{P}(Y_t = j).$$

We compute for all $j \in U$

$$\mathbb{P}(Y_t = j) = \sum_{i\in U} \mathbb{P}(Y_t = j, Y_0 = i)$$
$$= \sum_{i\in U} \mathbb{P}(Y_t = j \mid Y_0 = i)\mathbb{P}(Y_0 = i) = \sum_{i\in U} \alpha(i)Q_t(i, j),$$

that is $R(t) = (\alpha_1, 0)Q_t\mathbf{1}_{s,m} = \alpha_1 e^{tA_{11}}\mathbf{1}_m$. △

∇ Exercise 5.5 (A Binary Semi-Markov System and Its Entropy Rate)

1. Model by a semi-Markov process the system of Exercise 4.4.
2. Write the Markov renewal equation and determine the transition function of the system.
3. Show again the availability results.
4. Assume that X_1 (Y_1) has a density f (g) and a finite expected value a (b). Compute the entropy rate of the process. Determine the sojourn times distributions maximizing this rate.

Solution With the notation of Sect. 5.2.

1. If the functioning states are represented by 0 and the failed states by 1, the system can be modelled by a semi-Markov process $\mathbf{Z}$ defined by

$$Z_t = \sum_{n\geq 0} \mathbf{1}_{(S_n \leq t < S_n + X_{n+1})}, \quad t \geq 0,$$

with semi-Markov kernel Q given by

$$Q(t) = \begin{pmatrix} 0 & F(t) \\ G(t) & 0 \end{pmatrix}.$$

2. The transition function P_t satisfies the Markov renewal equation $P = I - H + Q * P$, or $[I - Q] * P = H$, where

$$H(t) = \begin{pmatrix} F(t) & 0 \\ 0 & G(t) \end{pmatrix}.$$

Its solution is $P(t) = [I - Q(t)]^{*(-1)} * [I - H(t)]$.

On the one hand,

$$\begin{aligned}
[I - Q(t)]^{*(-1)} &= \begin{pmatrix} 1 & -F(t) \\ -G(t) & 1 \end{pmatrix}^{*(-1)} \\
&= [1 - F * G(t)]^{*(-1)} * \begin{pmatrix} 1 & F(t) \\ G(t) & 1 \end{pmatrix}
\end{aligned}$$

and the renewal function of the alternated renewal process of Exercise 4.4 is

$$m(t) = [1 - F * G(t)]^{*(-1)} = \sum_{n\geq 0} (F * G)^{*(n)}(t).$$

On the other hand,

$$I - H(t) = \begin{pmatrix} 1 - F(t) & 0 \\ 0 & 1 - G(t) \end{pmatrix}.$$

Finally,

$$\begin{aligned} P(t) &= m * \begin{pmatrix} 1 & F(t) \\ G(t) & 1 \end{pmatrix} * \begin{pmatrix} 1 - F(t) & 0 \\ 0 & 1 - G(t) \end{pmatrix} \\ &= m * \begin{pmatrix} 1 - F(t) & F * (1 - G)(t) \\ G * (1 - F)(t) & 1 - G(t) \end{pmatrix}. \end{aligned} \tag{5.17}$$

3. Taking as initial distribution $(1, 0)$, we obtain the availability $A(t) = P_{00}(t) = m * (1 - F)(t)$.

   Note that this approach is much more general than the one obtained by using the alternated renewal process, because all results are given by (5.17). For example,

$$A(t) = P_{01}(t) = m * G * (1 - F)(t),$$

   if the system is assumed to be failed at time 0.
4. We compute

$$< \nu, m >= \int_0^{+\infty} \frac{t}{2}[f(t) + g(t)]\, dt = \frac{a + b}{2},$$

   and

$$\begin{aligned} &\sum_{i,j=1}^{2} \nu(i) \int_0^{+\infty} q_{ij}(t) \log q_{ij}(t)\, dt = \\ &= \frac{1}{2}\left[\int_0^{+\infty} f(t) \log f(t)\, dt + \int_0^{+\infty} g(t) \log g(t)\, dt\right], \end{aligned}$$

   so the entropy rate of $\mathbf{Z}$ is

$$\mathbb{H}(\mathbf{Z}) = \frac{1}{a + b}[\mathcal{I}(X_1) + \mathcal{I}(Y_1)].$$

   If f and g are exponential distributions with respective parameters λ and μ, the entropy rate is

$$\mathbb{H}(\mathbf{Z}) = \frac{\lambda\mu}{\lambda + \mu}[2 - \log(\lambda\mu)],$$

   and is clearly maximum. △

∇ **Exercise 5.6 (A Treatment Station)** A factory discharges polluting waste at a known flow rate. A treatment station is constructed for sparing environment. A tank is provided for stocking the waste during the failures of the station; this avoids to stop the factory if the repairing is finished before the tank is full. Both the time for emptying the tank and the time necessary for the treatment of its content by the station are assumed to be negligible. The random variable τ equal to the time for filling the tank is called the delay.

This system is modelled by a semi-Markov process with three states.

1. Give the states and the semi-Markov kernel of the process.
2. Determine the transition matrix and the stationary distribution of its embedded chain.
3. Give the limit distribution of the process.
4. Assuming that the system works perfectly at time $t = 0$, determine its reliability R; see Exercise 5.4 for definition.

Solution

1. The states of the process are the following:

   state 1 : the factory is functioning;
   state 2 : the factory has been failed shorter than τ;
   state 3 : the factory has been failed longer than τ.

   The semi-Markov kernel is

$$Q = \begin{pmatrix} 0 & F & 0 \\ Q_{21} & 0 & Q_{23} \\ B & 0 & 0 \end{pmatrix},$$

   with

$$Q_{21}(t) = \int_0^t [1 - C(x)]dA(x) \quad \text{and} \quad Q_{23}(t) = \int_0^t [1 - A(x)]dC(x),$$

   where F is the distribution function of the life time of the station, A that of the repairing time of the station, C that of the delay, and B that of the repairing time of the factory.
2. The transition matrix of the embedded chain (J_n) is

$$P = \begin{pmatrix} 0 & 1 & 0 \\ q & 0 & p \\ 1 & 0 & 0 \end{pmatrix}$$

where $p + q = 1$, with

$$p = \int_0^{+\infty} [1 - A(x)]dC(x).$$

The stationary distribution of (J_n) is $\nu = (1/(2 + p), 1/(2 + p), p/(2 + p))$.

3. The limit distribution of the process is

$$\pi = \frac{1}{< \nu, m >} \mathrm{diag}(\nu(i))m = \frac{1}{m_1 + m_2 + pm_3}(m_1, m_2, pm_3),$$

where $m_1 = \int_0^{+\infty}[1 - F(x)]dx$, $m_2 = \int_0^{+\infty}[1 - Q_{21}(x) - Q_{23}(x)]dx$, and $m_3 = \int_0^{+\infty}[1 - B(x)]dx$.

4. We compute

$$R(t) = M * [1 - F + F * (1 - F - Q_{23})](t),$$

where $M(t) = \sum_{n\geq 0}(F * Q_{21})^{(n)}(t)$. △

∇ **Exercise 5.7 (A Feed Back Queue)** Let us consider again the queue of Exercise 5.2. When a customer arrives, if more than N customers already queue, then he leaves the system. Moreover, once served, either he comes back queueing, with probability $p \in]0, 1[$, or he leaves the system, with probability $1 - p$; see Fig. 5.6. Let X_t be the random variable equal to the number of customers present in the post office at time t.

1. Of which type is the semi-Markov process $\mathbf{X} = (X_t)$? Give its semi-Markov kernel and determine the transition matrix of its embedded chain.
2. Compute the stationary distribution of this chain, the average sojourn times in each state and the limit distribution of the process.

Solution The number $\mathbf{X} = (X_t)$ of customers present in the system is a birth-death semi-Markov process with state space $E = [\![0, N]\!]$ and exponentially distributed sojourn times, that is a Markov process again. The only non-zero entries of its

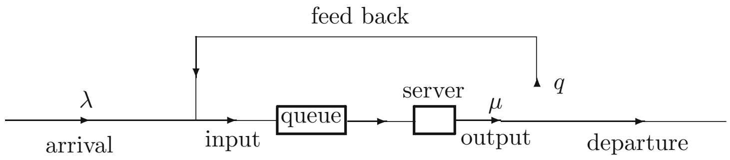

Fig. 5.6 A feed back queue—Exercise 5.7

semi-Markov kernel are

$$
\begin{aligned}
Q_{i,i-1}(t) &= \frac{\mu_i q_i}{\lambda_i + \mu_i}[1 - e^{-(\lambda_i+\mu_i)t}], \quad 1 \le i \le N, \\
Q_{i,i}(t) &= \frac{\mu_i p_i}{\lambda_i + \mu_i}[1 - e^{-(\lambda_i+\mu_i)t}], \quad 1 \le i \le N, \\
Q_{i,i+1}(t) &= \frac{\lambda_i}{\lambda_i + \mu_i}[1 - e^{-(\lambda_i+\mu_i)t}], \quad 0 \le i \le N-1,
\end{aligned}
$$

2. Its mean sojourn times are given by $m_i = 1/(\lambda_i + \mu_i)$. The stationary distribution ν of its embedded chain is given by

$$
\begin{aligned}
\nu_i &= \lambda_1 \dots \lambda_{i-1}(1 + \lambda_i/\mu_i + 1/q_i)\nu_0/\mu_1 \dots \mu_{i-1}, \quad 2 \le i \le N, \\
\nu_1 &= (1 + \lambda_i/\mu_i + 1/q_i)\nu_0, \\
\nu_0 &= \Big[\sum_{i=1}^{N} \lambda_1 \dots \lambda_{i-1}(1 + 1/q_i)/\mu_1 \dots \mu_{i-1}\Big]^{-1}.
\end{aligned}
$$

Note that for $p = 0$, this system amounts to the system described in Exercise 5.2, for a finite queue, that is an M/M/1/$(N + 1)$ queue. △