# 14

# Markov Processes with a Finite State Space

## 14.1 Definition of a Markov Process

In this section we define a homogeneous Markov process with values in a finite state space. We can assume that the state space $X$ is the set of the first $r$ positive integers, that is $X = \{1, \ldots, r\}$.

Let $P(t)$ be a family of $r \times r$ stochastic matrices indexed by the parameter $t \in [0, \infty)$. The elements of $P(t)$ will be denoted by $P_{ij}(t)$, $1 \leq i, j \leq r$. We assume that the family $P(t)$ forms a semi-group, that is $P(s)P(t) = P(s+t)$ for any $s, t \geq 0$. Since $P(t)$ are stochastic matrices, the semi-group property implies that $P(0)$ is the identity matrix. Let $\mu$ be a distribution on $X$.

Let $\widetilde{\Omega}$ be the set of all functions $\widetilde{\omega} : \mathbb{R}^+ \rightarrow X$ and $\mathcal{B}$ be the $\sigma$-algebra generated by all the cylindrical sets. Define a family of finite-dimensional distributions $\mathrm{P}_{t_0,\ldots,t_k}$, where $0 = t_0 \leq t_1 \leq \ldots \leq t_k$, as follows

$$\mathrm{P}_{t_0,\ldots,t_k}(\widetilde{\omega}(t_0) = i_0, \widetilde{\omega}(t_1) = i_1, \ldots, \widetilde{\omega}(t_k) = i_k)$$

$$= \mu_{i_0} P_{i_0 i_1}(t_1) P_{i_1 i_2}(t_2 - t_1) \ldots P_{i_{k-1} i_k}(t_k - t_{k-1}).$$

It can be easily seen that this family of finite-dimensional distributions satisfies the consistency conditions. By the Kolmogorov Consistency Theorem, there is a process $X_t$ with values in $X$ with these finite-dimensional distributions. Any such process will be called a homogeneous Markov process with the family of transition matrices $P(t)$ and the initial distribution $\mu$. (Since we do not consider non-homogeneous Markov processes in this section, we shall refer to $X_t$ simply as a Markov process).

**Lemma 14.1.** *Let $X_t$ be a Markov process with the family of transition matrices $P(t)$. Then, for $0 \leq s_1 \leq \ldots \leq s_k$, $t \geq 0$, and $i_1, \ldots, i_k, j \in X$, we have*

$$\mathrm{P}(X_{s_k+t} = j | X_{s_1} = i_1, \ldots, X_{s_k} = i_k) = \mathrm{P}(X_{s_k+t} = j | X_{s_k} = i_k) = P_{i_k j}(t) \quad (14.1)$$

*if the conditional probability on the left-hand side is defined.*

L. Koralov and Y.G. Sinai, *Theory of Probability and Random Processes*, Universitext, DOI 10.1007/978-3-540-68829-7_14,

The proof of this lemma is similar to the arguments in Sect. 5.2, and thus will not be provided here. As in Sect. 5.2, it is easy to see that for a Markov process with the family of transition matrices $P(t)$ and the initial distribution $\mu$ the distribution of $X_t$ is $\mu P(t)$.

**Definition 14.2.** *A distribution $\pi$ is said to be stationary for a semi-group of Markov transition matrices $P(t)$ if $\pi P(t) = \pi$ for all $t \geq 0$.*

As in the case of discrete time we have the Ergodic Theorem.

**Theorem 14.3.** *Let $P(t)$ be a semi-group of Markov transition matrices such that for some $t$ all the matrix entries of $P(t)$ are positive. Then there is a unique stationary distribution $\pi$ for the semi-group of transition matrices. Moreover, $\sup_{i,j \in X} |P_{ij}(t) - \pi_j|$ converges to zero exponentially fast as $t \to \infty$.*

This theorem can be proved similarly to the Ergodic Theorem for Markov chains (Theorem 5.9). We leave the details as an exercise for the reader.

## 14.2 Infinitesimal Matrix

In this section we consider semi-groups of Markov transition matrices which are differentiable at zero. Namely, assume that there exist the following limits

$$Q_{ij} = \lim_{t \downarrow 0} \frac{P_{ij}(t) - I_{ij}}{t}, \quad 1 \leq i, j \leq r, \tag{14.2}$$

where $I$ is the identity matrix.

**Definition 14.4.** *If the limits in (14.2) exist for all $1 \leq i, j \leq r$, then the matrix $Q$ is called the infinitesimal matrix of the semigroup $P(t)$.*

Since $P_{ij}(t) \geq 0$ and $I_{ij} = 0$ for $i \neq j$, the off-diagonal elements of $Q$ are non-negative. Moreover,

$$\sum_{j=1}^{r} Q_{ij} = \sum_{j=1}^{r} \lim_{t \downarrow 0} \frac{P_{ij}(t) - I_{ij}}{t} = \lim_{t \downarrow 0} \frac{\sum_{j=1}^{r} P_{ij}(t) - 1}{t} = 0,$$

or, equivalently,

$$Q_{ii} = -\sum_{j \neq i} Q_{ij}.$$

**Lemma 14.5.** *If the limits in (14.2) exist, then the transition matrices are differentiable for all $t \in \mathbb{R}^+$ and satisfy the following systems of ordinary differential equations.*

$$\frac{dP(t)}{dt} = P(t)Q \qquad (\textit{forward system}).$$

$$\frac{dP(t)}{dt} = QP(t) \qquad (\textit{backward system}).$$

*The derivatives at $t = 0$ should be understood as one-sided derivatives.*

*Proof.* Due to the semi-group property of $P(t)$,

$$\lim_{h \downarrow 0} \frac{P(t+h) - P(t)}{h} = P(t) \lim_{h \downarrow 0} \frac{P(h) - I}{h} = P(t)Q. \qquad (14.3)$$

This shows, in particular, that $P(t)$ is right-differentiable. Let us prove that $P(t)$ is left-continuous. For $t > 0$ and $0 \leq h \leq t$,

$$P(t) - P(t-h) = P(t-h)(P(h) - I).$$

All the elements of $P(t-h)$ are bounded, while all the elements of $(P(h) - I)$ tend to zero as $h \downarrow 0$. This establishes the continuity of $P(t)$.

For $t > 0$,

$$\lim_{h \downarrow 0} \frac{P(t) - P(t-h)}{h} = \lim_{h \downarrow 0} P(t-h) \lim_{h \downarrow 0} \frac{P(h) - I}{h} = P(t)Q. \qquad (14.4)$$

Combining (14.3) and (14.4), we obtain the forward system of equations.

Due to the semi-group property of $P(t)$, for $t \geq 0$,

$$\lim_{h \downarrow 0} \frac{P(t+h) - P(t)}{h} = \lim_{h \downarrow 0} \frac{P(h) - I}{h} P(t) = QP(t),$$

and similarly, for $t > 0$,

$$\lim_{h \downarrow 0} \frac{P(t) - P(t-h)}{h} = \lim_{h \downarrow 0} \frac{P(h) - I}{h} \lim_{h \downarrow 0} P(t-h) = QP(t).$$

This justifies the backward system of equations. □

The system $dP(t)/dt = P(t)Q$ with the initial condition $P_0 = I$ has the unique solution $P(t) = \exp(tQ)$. Thus, the transition matrices can be uniquely expressed in terms of the infinitesimal matrix.

Let us note another property of the infinitesimal matrix. If $\pi$ is a stationary distribution for the semi-group of transition matrices $P(t)$, then

$$\pi Q = \lim_{t \downarrow 0} \frac{\pi P(t) - \pi}{t} = 0.$$

Conversely, if $\pi Q = 0$ for some distribution $\pi$, then

$$\pi P(t) = \pi \exp(tQ) = \pi\left(I + tQ + \frac{t^2 Q^2}{2!} + \frac{t^3 Q^3}{3!} + \cdots\right) = \pi.$$

Thus, $\pi$ is a stationary distribution for the family $P(t)$.

## 14.3 A Construction of a Markov Process

Let $\mu$ be a probability distribution on $X$ and $P(t)$ be a differentiable semi-group of transition matrices with the infinitesimal matrix $Q$. Assume that $Q_{ii} < 0$ for all $i$.

On an intuitive level, a Markov process with the family of transition matrices $P(t)$ and initial distribution $\mu$ can be described as follows. At time $t = 0$ the process is distributed according to $\mu$. If at time $t$ the process is in a state $i$, then it will remain in the same state for time $\tau$, where $\tau$ is a random variable with exponential distribution. The parameter of the distribution depends on $i$, but does not depend on $t$. After time $\tau$ the process goes to another state, where it remains for exponential time, and so on. The transition probabilities depend on $i$, but not on the moment of time $t$.

Now let us justify the above description and relate the transition times and transition probabilities to the infinitesimal matrix. Let $Q$ be an $r \times r$ matrix with $Q_{ii} < 0$ for all $i$. Assume that there are random variables $\xi$, $\tau_i^n$, $1 \leq i \leq r$, $n \in \mathbb{N}$, and $\eta_i^n$, $1 \leq i \leq r$, $n \in \mathbb{N}$, defined on a common probability space, with the following properties.

1. The random variable $\xi$ takes values in $X$ and has distribution $\mu$.
2. For any $1 \leq i \leq r$, the random variables $\tau_i^n$, $n \in \mathbb{N}$, are identically distributed according to the exponential distribution with parameter $r_i = -Q_{ii}$.
3. For any $1 \leq i \leq r$, the random variables $\eta_i^n$, $n \in \mathbb{N}$, take values in $X \setminus \{i\}$ and are identically distributed with $\mathrm{P}(\eta_i^n = j) = -Q_{ij}/Q_{ii}$ for $j \neq i$.
4. The random variables $\xi$, $\tau_i^n$, $\eta_i^n$, $1 \leq i \leq r$, $n \in \mathbb{N}$, are independent.

We inductively define two sequences of random variables: $\sigma^n$, $n \geq 0$, with values in $\mathbb{R}^+$, and $\xi^n$, $n \geq 0$, with values in $X$. Let $\sigma^0 = 0$ and $\xi^0 = \xi$. Assume that $\sigma^m$ and $\xi^m$ have been defined for all $m < n$, where $n \geq 1$, and set

$$\sigma^n = \sigma^{n-1} + \tau^n_{\xi^{n-1}}.$$

$$\xi^n = \eta^n_{\xi^{n-1}}.$$

We shall treat $\sigma^n$ as the time till the $n$-th transition takes place, and $\xi^n$ as the $n$-th state visited by the process. Thus, define

$$X_t = \xi^n \quad \text{for} \quad \sigma^n \leq t < \sigma^{n+1}. \tag{14.5}$$

**Lemma 14.6.** *Assume that the random variables $\xi$, $\tau_i^n$, $1 \leq i \leq r$, $n \in \mathbb{N}$, and $\eta_i^n$, $1 \leq i \leq r$, $n \in \mathbb{N}$, are defined on a common probability space and satisfy assumptions 1–4 above. Then the process $X_t$ defined by (14.5) is a Markov process with the family of transition matrices $P(t) = \exp(tQ)$ and initial distribution $\mu$.*

*Sketch of the Proof.* It is clear from (14.5) that the initial distribution of $X_t$ is $\mu$. Using the properties of $\tau_i^n$ and $\eta_i^n$ it is possible to show that, for $k \neq j$,

$$\begin{aligned} &\mathrm{P}(X_0 = i, X_t = k, X_{t+h} = j) \\ &= \mathrm{P}(X_0 = i, X_t = k)(\mathrm{P}(\tau_k^1 < h)\mathrm{P}(\xi_k^1 = j) + o(h)) \\ &= \mathrm{P}(X_0 = i, X_t = k)(Q_{kj}h + o(h)) \quad \text{as} \quad h \downarrow 0. \end{aligned}$$

In other words, the main contribution to the probability on the left-hand side comes from the event that there is exactly one transition between the states $k$ and $j$ during the time interval $[t, t+h)$.

Similarly,

$$\begin{aligned} &\mathrm{P}(X_0 = i, X_t = j, X_{t+h} = j) \\ &= \mathrm{P}(X_0 = i, X_t = j)(\mathrm{P}(\tau_j^1 \geq h) + o(h)) \\ &= \mathrm{P}(X_0 = i, X_t = j)(1 + Q_{jj}h + o(h)) \quad \text{as} \quad h \downarrow 0, \end{aligned}$$

that is, the main contribution to the probability on the left-hand side comes from the event that there are no transitions during the time interval $[t, t+h]$.

Therefore,

$$\begin{aligned} &\sum_{k=1}^{r} \mathrm{P}(X_0 = i, X_t = k, X_{t+h} = j) \\ &= \mathrm{P}(X_0 = i, X_t = j) + h \sum_{k=1}^{r} \mathrm{P}(X_0 = i, X_t = k) Q_{kj} + o(h). \end{aligned}$$

Let $R_{ij}(t) = \mathrm{P}(X_0 = i, X_t = j)$. The last equality can be written as

$$R_{ij}(t+h) = R_{ij}(t) + h \sum_{k=1}^{r} R_{ik}(t) Q_{kj} + o(h).$$

Using matrix notation,

$$\lim_{h \downarrow 0} \frac{R(t+h) - R(t)}{h} = R(t)Q.$$

The existence of the left derivative is justified similarly. Therefore,

$$\frac{dR(t)}{dt} = R(t)Q \quad \text{for} \quad t \geq 0.$$

Note that $R_{ij}(0) = \mu_i$ for $i = j$, and $R_{ij}(0) = 0$ for $i \neq j$. These are the same equation and initial condition that are satisfied by the matrix-valued function $\mu_i P_{ij}(t)$. Therefore,

$$R_{ij}(t) = \mathrm{P}(X_0 = i, X_t = j) = \mu_i P_{ij}(t). \tag{14.6}$$

In order to prove that $X_t$ is a Markov process with the family of transition matrices $P(t)$, it is sufficient to demonstrate that

$$\mathrm{P}(X_{t_0} = i_0, X_{t_1} = i_1, \ldots, X_{t_k} = i_k)$$

$$= \mu_{i_0} P_{i_0 i_1}(t_1) P_{i_1 i_2}(t_2 - t_1) \ldots P_{i_{k-1} i_k}(t_k - t_{k-1}).$$

for $0 = t_0 \leq t_1 \leq \ldots \leq t_k$. The case $k = 1$ has been covered by (14.6). The proof for $k > 1$ is similar and is based on induction on $k$. □

## 14.4 A Problem in Queuing Theory

Markov processes with a finite or countable state space are used in the Queuing Theory. In this section we consider one basic example.

Assume that there are $r$ identical devices designed to handle incoming requests. The times between consecutive requests are assumed to be independent exponentially distributed random variables with parameter $\lambda$. At a given time, each device may be either free or busy servicing one request. An incoming request is serviced by any of the free devices and, if all the devices are busy, the request is rejected. The times to service each request are assumed to be independent exponentially distributed random variables with parameter $\mu$. They are also assumed to be independent of the arrival times of the requests.

Let us model the above system by a process with the state space $X = \{0, 1, \ldots, r\}$. A state of the process corresponds to the number of devices busy servicing requests. If there are no requests in the system, the time till the first one arrives is exponential with parameter $\lambda$. If there are $r$ requests in the system, the time till the first one of them is serviced is an exponential random variable with parameter $r\mu$. If there are $1 \leq i \leq r - 1$ requests in the system, the time till either one of them is serviced, or a new request arrives, is an exponential random variable with parameter $\lambda + i\mu$. Therefore, the process remains in a state $i$ for a time which is exponentially distributed with parameter

$$\gamma(i) = \begin{cases} \lambda & \text{if } i = 0, \\ \lambda + i\mu & \text{if } 1 \leq i \leq r - 1, \\ i\mu & \text{if } i = r. \end{cases}$$

If the process is in the state $i = 0$, it can only make a transition to the state $i = 1$, which corresponds to an arrival of a request. From a state $1 \leq i \leq r-1$ the process can make a transition either to state $i-1$ or to state $i+1$. The former corresponds to completion of one of $i$ requests being serviced before the arrival of a new request. Therefore, the probability of transition from $i$ to $i - 1$ is equal to the probability that the smallest of $i$ exponential random variables with parameter $\mu$ is less than an exponential random variable with parameter $\lambda$ (all the random variables are independent). This probability is equal to $i\mu/(i\mu+\lambda)$. Consequently, the transition probability from $i$ to $i+1$ is equal to $\lambda/(i\mu+\lambda)$. Finally, if the process is in the state $r$, it can only make a transition to the state $r-1$.

Let the initial state of the process $X_t$ be independent of the arrival times of the requests and the times it takes to service the requests. Then the process $X_t$ satisfies the assumptions of Lemma 14.6 (see the discussion before Lemma 14.6). The matrix $Q$ is the $(r+1)\times(r+1)$ tridiagonal matrix with the vectors $\gamma(i)$, $0 \le i \le r$, on the diagonal, $u(i) \equiv \lambda$, $1 \le i \le r$, above the diagonal, and $l(i) = i\mu$, $1 \le i \le r$, below the diagonal. By Lemma 14.6, the process $X_t$ is Markov with the family of transition matrices $P(t) = \exp(tQ)$.

It is not difficult to prove that all the entries of $\exp(tQ)$ are positive for some $t$, and therefore the Ergodic Theorem is applicable. Let us find the stationary distribution for the family of transition matrices $P(t)$. As noted in Sect. 14.2, a distribution $\pi$ is stationary for $P(t)$ if and only if $\pi Q = 0$. It is easy to verify that the solution of this linear system, subject to the conditions $\pi(i) \ge 0$, $0 \le i \le r$, and $\sum_{i=0}^{r}\pi(i) = 1$, is

$$\pi(i) = \frac{(\lambda/\mu)^i/i!}{\sum_{j=0}^{r}(\lambda/\mu)^j/j!}, \quad 0 \le i \le r.$$

## 14.5 Problems

**1.** Let $P(t)$ be a differentiable semi-group of Markov transition matrices with the infinitesimal matrix $Q$. Assume that $Q_{ij} \ne 0$ for $1 \le i,j \le r$. Prove that for every $t > 0$ all the matrix entries of $P(t)$ are positive. Prove that there is a unique stationary distribution $\pi$ for the semi-group of transition matrices. (Hint: represent $Q$ as $(Q+cI)-cI$ with a constant $c$ sufficiently large so that to make all the elements of the matrix $Q+cI$ non-negative.)

**2.** Let $P(t)$ be a differentiable semi-group of transition matrices. Prove that if all the elements of $P(t)$ are positive for some $t$, then all the elements of $P(t)$ are positive for all $t > 0$.

**3.** Let $P(t)$ be a differentiable semi-group of Markov transition matrices with the infinitesimal matrix $Q$. Assuming that $Q$ is self-adjoint, find a stationary distribution for the semi-group $P(t)$.

**4.** Let $X_t$ be a Markov process with a differentiable semi-group of transition matrices and initial distribution $\mu$ such that $\mu(i) > 0$ for $1 \le i \le r$. Prove that $\mathrm{P}(X_t = i) > 0$ for all $i$.

**5.** Consider a taxi station where taxis and customers arrive according to Poisson processes. The taxis arrive at the rate of 1 per min, and the customers at the rate of 2 per min. A taxi will wait only if there are no other taxis waiting already. A customer will wait no matter how many other customers are in line. Find the probability that there is a taxi waiting at a given moment and the average number of customers waiting in line.

**6.** A company gets an average of five calls an hour from prospective clients. It takes a company representative an average of 20 min to handle one call (the distribution of time to handle one call is exponential). A prospective client who cannot immediately talk to a representative never calls again. For each prospective client that talks to a representative the company makes \$1,000. How many representatives should the company maintain if each is paid \$10 an hour?