7 Discrete Time Markov chains

7.1 Definitions and Basic Properties

In this section, we discuss the discrete time Markov chains. These are special stochastic processes satisfying the property that at a given time, the future evolution depends only on the state at the *current time* instead of the past of the whole process.

We take E to be a finite or countable set equipped with the σ-algebra $\mathcal{P}(E)$.

Definition 7.1.1 : A function $Q : E \times E \longrightarrow \mathbb{R}$ is called *transition matrix* (轉移矩陣) or *stochastic matrix* (隨機矩陣) on E if it satisfies

(1) $0 \leqslant Q(x, y) \leqslant 1$ for all $x, y \in E$;

(2) $\sum_{y \in E} Q(x, y) = 1$ for all $x \in E$.

Remark 7.1.2 : The notion of transition matrix on E is equivalent to the notion of transition probability from E to E. Indeed, given a transition matrix Q on E, we may define $\nu : E \times \mathcal{P}(E) \longrightarrow \mathbb{R}$ by

$$\nu(x, A) = \sum_{y \in A} Q(x, y), \quad \forall x \in E, A \subseteq E.$$

Then, one can easily check that ν is a transition probability from E to E. Conversely, given a transition probability ν, we can obtain a transition matrix on E by letting $Q(x, y) = \nu(x, \{y\})$.

Definition 7.1.3 : We define $Q_n = Q^n$ for all positive integer $n \geqslant 1$. In other words, $Q_1 = Q$ and the following recurrence relation holds for all $n \geqslant 1$,

$$Q_{n+1}(x, y) = \sum_{z \in E} Q_n(x, z) Q(z, y). \tag{7.1}$$

This equality is called *Chapman-Kolmogorov equality* (Chapman-Kolmogorov 等式) . We can check that Q_n is still a transition matrix on E. In order to extend the validity of to the case $n = 0$, we may define $Q_0(x, y) = \mathbb{1}_{x=y}$, which is a diagonal transition matrix.

Definition 7.1.4 : For any function $f : E \longrightarrow \mathbb{R}_{\geqslant 0}$, we define $Qf : E \longrightarrow \mathbb{R}_{\geqslant 0}$ to be

$$Qf(x) = \sum_{y \in E} Q(x, y) f(y), \quad \forall x \in E. \tag{7.2}$$

Remark 7.1.5 :

(1) The summation in (7.2) is always well defined, being a countable summation of non-negative terms.

(2) If we see f as a column vector, then the way the function Qf is defined is exactly through a matrix multiplication.

Definition 7.1.6 : Let Q be a transition matrix on E and $(X_n)_{n\geqslant 0}$ be a stochastic process with values in E. We say that $(X_n)_{n\geqslant 0}$ is a *Markov chain* (馬可夫鏈) with transition matrix Q if for all $n\geqslant 0$, the conditional distribution of X_{n+1} knowing $(X_0, X_1, \ldots, X_n)$ is $Q(X_n, \cdot)$. Since E is a discrete space, this condition is equivalent to the follows,

$$\mathbb{P}(X_{n+1} = y \,|\, X_0 = x_0, \ldots, X_n = x_n) = Q(x_n, y),$$

for all $x_0, x_1, \ldots, x_n, y \in E$ with $\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) > 0$.

Remark 7.1.7 :

(1) In general, the conditional distribution of X_{n+1} knowing $(X_0, X_1, \ldots, X_n)$ does not only depend on X_n, but on $X_0, X_1, \ldots, X_n$. In the case of a Markov chain, the fact that the above conditional distribution only depends on X_n is called *Markov property* (馬可夫性質). This should be understood as follows: knowing the whole past $(X_0, X_1, \ldots, X_n)$ does not give more information than knowing only the current state X_n.

(2) The conditional distribution $Q(x, \cdot)$ mentioned above does not depend on the time n. In this case, we say that the Markov chain is *homogeneous* (均勻). We may also consider a transition matrix that evolve with time n.

Proposition 7.1.8 : *Given a stochastic process $(X_n)_{n\geqslant 0}$ with values in E. Then, $(X_n)_{n\geqslant 0}$ is a Markov chain with transition matrix Q if and only if for all $n\geqslant 0$ and $x_0, x_1, \ldots, x_n \in E$, the following holds,*

$$\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) = \mathbb{P}(X_0 = x_0) Q(x_0, x_1) \ldots Q(x_{n-1}, x_n). \tag{7.3}$$

Besides, when $\mathbb{P}(X_0 = x_0) > 0$, we have

$$\mathbb{P}(X_n = x_n \,|\, X_0 = x_0) = Q_n(x_0, x_n). \tag{7.4}$$

Proof : Let $(X_n)_{n\geqslant 0}$ be a Markov chain with transition matrix Q. Then, we have

$$\begin{aligned}
&\mathbb{P}(X_0 = x_0, \ldots, X_{n+1} = x_{n+1}) \\
&= \mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) \times \mathbb{P}(X_{n+1} = x_{n+1} \,|\, X_0 = x_0, \ldots, X_n = x_n) \\
&= \mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) \times Q(x_n, x_{n+1}).
\end{aligned}$$

Hence by induction, we obtain Eq. (7.3). Conversely, if Eq. (7.3) holds, we can check

$$\mathbb{P}(X_{n+1} = y \,|\, X_0 = x_0, \ldots, X_n = x_n) = \frac{\mathbb{P}(X_{n+1} = y, X_0 = x_0, \ldots, X_n = x_n)}{\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n)} = Q(x_n, y).$$

Finally, the marginal distribution Eq. (7.4) can be obtained from the following,

$$Q_n(x_0, x_n) = \sum_{x_1,\ldots,x_{n-1}\in E} Q(x_0, x_1)\ldots Q(x_{n-1}, x_n).$$

□

Remark 7.1.9 : We know from Eq. (7.3) that the initial condition, or the distribution of X_0, along with the transition matrix Q, determines the distribution of the whole Markov chain $(X_n)_{n\geqslant 0}$.

The following proposition contains different properties of a Markov chain.

Proposition 7.1.10 : *Let $(X_n)_{n\geqslant 0}$ be a Markov chain with transition matrix Q.*

(1) For all $n \geqslant 0$ and non-negative measurable function $f : E \longrightarrow \mathbb{R}_{\geqslant 0}$, we have

$$\mathbb{E}[f(X_{n+1}) \mid X_0, X_1, \ldots, X_n] = \mathbb{E}[f(X_{n+1}) \mid X_n] = Qf(X_n).$$

More generally speaking, for any finite subset $\{i_1, \ldots, i_k\}$ of $\{0, \ldots, n-1\}$, we have

$$\mathbb{E}[f(X_{n+1}) \mid X_{i_1}, \ldots, X_{i_k}, X_n] = Qf(X_n).$$

(2) For any $n \geqslant 0$, $p \geqslant 1$ and $y_1, \ldots y_p \in E$, we have

$$\mathbb{P}(X_{n+1} = y_1, \ldots, X_{n+p} = y_p \mid X_0, \ldots, X_n) = Q(X_n, y_1)Q(y_1, y_2)\ldots Q(y_{p-1}, y_p), \tag{7.5}$$

and in consequence,

$$\mathbb{P}(X_{n+p} = y_p \mid X_n) = Q_p(X_n, y_p). \tag{7.6}$$

If we set $Y_p = X_{n+p}$ for all $p \geqslant 0$, then $(Y_p)_{p\geqslant 0}$ is still a Markov chain with transition matrix Q.

Proof :

(1) From the definition, we have

$$\mathbb{E}[f(X_{n+1}) \mid X_0, \ldots, X_n] = \sum_{y\in E} Q(X_n, y)f(y) = Qf(X_n).$$

Moreover, if $\{i_1, \ldots, i_k\}$ is a finite subset of $\{0, \ldots, n-1\}$, we have

$$\begin{aligned}\mathbb{E}[f(X_{n+1}) \mid X_{i_1}, \ldots, X_{i_k}, X_n] &= \mathbb{E}[\mathbb{E}[f(X_{n+1}) \mid X_1, \ldots, X_n] \mid X_{i_1}, \ldots, X_{i_k}, X_n] \\ &= \mathbb{E}[Qf(X_n) \mid X_{i_1}, \ldots, X_{i_k}, X_n] \\ &= Qf(X_n).\end{aligned}$$

(2) Eq. (7.5) is a direct application of Eq. (7.3). Similarly to the proof of Eq. (7.4), Eq. (7.6) can be obtained by summing all the possible values taken by $y_1, \ldots, y_{p-1}$. To conclude, we obtain the following from Eq. (7.5),

$$\mathbb{P}(Y_0 = y_0, \ldots, Y_n = y_n) = \mathbb{P}(X_n = y_0)Q(y_0, y_1)\ldots Q(y_{p-1}, y_p),$$

and use Proposition 7.1.8.

□

Example 7.1.11 : Below are examples of Markov chains.

(1) (Independent random variables) If $(X_n)_{n\geqslant 0}$ is an i.i.d. sequence of random variables with values in E, let us denote by μ the distribution of any term, then $(X_n)_{n\geqslant 0}$ is a Markov chain with transition matrix

$$Q(x,y)=\mu(y),\quad \forall x,y\in E.$$

(2) (Random walk on $\mathbb{Z}^d$) Let $\eta,\xi_1,\dots,\xi_n$ be independent random variables with values in $\mathbb{Z}^d$. Assume that $(\xi_n)_{n\geqslant 1}$ is an i.i.d. sequence with distribution μ. For all $n\geqslant 0$, let

$$X_n=\eta+\xi_1+\cdots+\xi_n.$$

Then, $(X_n)_{n\geqslant 0}$ is a Markov chain with transition matrix

$$Q(x,y)=\mu(y-x),\quad \forall x,y\in E.$$

Using the independence of ξ_{n+1} with $(X_0,\dots,X_n)$, we have

$$\begin{aligned}&\mathbb{P}(X_{n+1}=y\,|\,X_0=x_0,\dots,X_n=x_n)\\&=\mathbb{P}(\xi_{n+1}=y-x_n\,|\,X_0=x_0,\dots,X_n=x_n)\\&=\mathbb{P}(\xi_{n+1}=y-x_n)=\mu(y-x_n).\end{aligned}$$

Let $(e_1,\dots,e_d)$ be the canonical basis (正則基底) of $\mathbb{R}^d$. The Markov chain $(X_n)_{n\geqslant 0}$ is called *symmetric simple random walk* (對稱簡單隨機漫步) on $\mathbb{Z}^d$ if μ satisfies

$$\mu(e_i)=\mu(-e_i)=\frac{1}{2d},\quad \forall i\in\{1,\dots,d\}.$$

(3) (Simple random walk on graphs) Let E be any set, $\mathcal{P}_2(E)$ be the set consisting of all the pairwise disjoint elements from E and F be a subset of $\mathcal{P}_2(E)$. For all $x\in E$, set

$$F_x=\{y\in E:\{x,y\}\in F\}.$$

We assume that the set F_x is not empty and is finite for all $x\in E$. Let us define Q to be the transition matrix on E,

$$\forall x,y\in E,\quad Q(x,y)=\begin{cases}|F_x|^{-1} & \text{if } \{x,y\}\in F,\\ 0 & \text{otherwise.}\end{cases}$$

A Markov chain with transition matrix Q is called *simple random walk* (簡單隨機漫步) on the graph (E,F).

(4) (Branching process) We recall the branching process discussed in Example 6.3.8 and Exercise 6.24.

Let μ be a probability distribution on non-negative integers. Assume that μ is integrable and denote its expectation $m<\infty$. We exclude the special cases where $\mu=\delta_0$ or δ_1. Consider an i.i.d. sequence $(\xi_{n,j})_{n,j\geqslant 0}$ of random variables with distribution μ. Fix an integer $\ell \geqslant 1$, and define another sequence $(X_n)_{n\geqslant 0}$ of random variables by induction,

$$X_0 = \ell,$$
$$X_{n+1} = \sum_{j=1}^{X_n} \xi_{n,j}, \quad \forall n \geqslant 0.$$

The random process $(X_n)_{n\geqslant 0}$ can be understood as the evolution of a parthenogenetic population where μ is the distribution of the offspring of each family member, and the population at the n-th generation is given by X_n. In this case, $(X_n)_{n\geqslant 0}$ is a Markov chain on $E = \mathbb{N}_0$ with transition matrix

$$Q(x,y) = \mu^{*x}(y), \quad \forall x, y \in \mathbb{N}_0, \tag{7.7}$$

where μ^{*x} is the x-fold convolution (捲積) of μ with itself, which can also be interpreted as the distribution of the sum of x i.i.d. random variables with distribution μ. Eq. (7.7) is a consequence of the below computation,

$$\begin{aligned}\mathbb{P}(X_{n+1} = y \,|\, X_0 = x_0, \dots, X_n = x_n)\\ \mathbb{P}\Big(\sum_{j=1}^{x_n} \xi_{n,j} = y \,\Big|\, X_0 = x_0, \dots, X_n = x_n\Big)\\ \mathbb{P}\Big(\sum_{j=1}^{x_n} \xi_{n,j} = y\Big) = \mu^{*x_n}(y),\end{aligned}$$

where in the second equality, we use the property that $(\xi_{n,j})_{j\geqslant 1}$ and $X_0, \dots, X_n$ are independent.

7.2 Canonical Markov Chain

In Definition 7.1.6, we gave the property that a Markov chain needs to satisfy. In the below proposition, we explain, given a transition matrix, how to construct a corresponding Markov chain. Later in Theorem 7.2.3, we will see the uniqueness of the distribution of such a Markov chain. Thus, we can say *the* Markov chain instead of *a* Markov chain when the transition matrix is given.

Proposition 7.2.1 : *Let Q be a transition matrix on E. We can find a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ such that for any $x \in E$, we can construct a Markov chain $(X_n^x)_{n\geqslant 0}$ with transition matrix Q and initial state $X_0^x = x$.*

Proof : Take $\Omega' = [0,1)$, $\mathcal{F}' = \mathcal{B}(\Omega')$ and $\mathbb{P}'$ to be the Lebesgue measure. For any real number $\omega \in [0,1)$, we have the dyadic expansion,

$$\omega = \sum_{n=0}^{\infty} \varepsilon_n(\omega) 2^{-n-1}, \quad \varepsilon_n(\omega) \in \{0,1\}.$$

The above provides us with an i.i.d. sequence $(\varepsilon_n)_{n\geqslant 0}$ of random variables such that $\mathbb{P}(\varepsilon_n = 1) =$

$\mathbb{P}(\varepsilon_n = 0) = \frac{1}{2}$. If $\varphi : \mathbb{N}_0 \times \mathbb{N}_0 \longrightarrow \mathbb{N}_0$ is an injective function, define $\eta_{i,j} = \varepsilon_{\varphi(i,j)}$ for all $i, j \in \mathbb{N}_0$, then $(\eta_{i,j})_{i,j \geqslant 0}$ is still an i.i.d. sequence of random variables. Let

$$\forall i \geqslant 0, \quad U_i = \sum_{j \geqslant 0} \eta_{i,j} 2^{-j-1}.$$

Then, $(U_i)_{i \geqslant 0}$ is an i.i.d. sequence of random variables with the uniform distribution on $[0, 1]$.

Since E is countable, we can enumerate the elements by $y_1, y_2, \ldots$. Given $x \in E$, let $X_0^x = x$ and we define the value of X_1^x using the uniform random variable U_0,

$$X_1^x = y_k \quad \text{if} \quad \sum_{1 \leqslant j \leqslant k-1} Q(x, y_j) < U_0 \leqslant \sum_{1 \leqslant j \leqslant k} Q(x, y_j).$$

We can easily check that we do have $\mathbb{P}(X_1^x = y) = Q(x, y)$ for all $y \in E$. Then, similarly to the definition of X_1^x, we define the Markov chain $(X_n^x)_{n \geqslant 0}$ by induction: for all $n \geqslant 0$, define

$$X_{n+1}^x = y_k \quad \text{if} \quad \sum_{1 \leqslant j \leqslant k-1} Q(X_n^x, y_j) < U_n \leqslant \sum_{1 \leqslant j \leqslant k} Q(X_n^x, y_j).$$

Then, using the independence of (U_i), we reach at

$$\begin{aligned}
&\mathbb{P}(X_{n+1}^x = y_k \,|\, X_0^x = x_0, \ldots, X_n^x = x_n) \\
&= \mathbb{P}\left(\sum_{1 \leqslant j \leqslant k-1} Q(x_n, y_j) < U_n \leqslant \sum_{1 \leqslant j \leqslant k} Q(x_n, y_j) \,\middle|\, X_0^x = x_0, \ldots, X_n^x = x_n \right) \\
&= \mathbb{P}\left(\sum_{1 \leqslant j \leqslant k-1} Q(x_n, y_j) < U_n \leqslant \sum_{1 \leqslant j \leqslant k} Q(x_n, y_j) \right) \\
&= Q(x_n, y_k).
\end{aligned}$$

As a consequence, $(X_n^x)_{n \geqslant 0}$ is a Markov chain with transition matrix Q. □

In what follows, we will choose a canonical space on which the Markov chain is defined. If the state space is E, the sample space is $\Omega = E^{\mathbb{N}_0}$. An element $\omega \in \Omega$ in the sample space is written as $\omega = (\omega_0, \omega_1, \ldots)$, which is a sequence in E. For all $n \geqslant 0$, we can define the coordinate function (座標函數) X_n to be

$$X_n : \begin{array}{ccc} \Omega = E^{\mathbb{N}_0} & \to & E \\ \omega = (\omega_0, \omega_1, \ldots) & \mapsto & \omega_n \end{array}.$$

On Ω, we consider $\mathcal{F}$ to be the smallest σ-algebra that makes all the coordinate functions X_n measurable, that is

$$\mathcal{F} := \sigma(X_n^{-1}(x) : \forall n \geqslant 0, \forall x \in E).$$

This σ-algebra can also be generated by the following cylindrical events (圓柱事件)

$$\forall n \geqslant 0, \quad \forall x_0, \ldots, x_n \in E, \quad C = \{\omega \in \Omega : \omega_0 = x_0, \ldots, \omega_n = x_n\}.$$

In the rest of this section, we will define a probability measure on Ω so that we can talk about the uniqueness of the Markov chain in the sense of distribution.

Proposition 7.2.1 above gives the existence of the Markov chain once the transition matrix and the initial state are given. The uniqueness, however, does not hold in general. As discussed in Section 2.1.2, the random variable itself is not unique, what is unique is its *distribution*.

Lemma 7.2.2 : *Let $(G, \mathcal{G})$ be a measurable space and $\psi : (G, \mathcal{G}) \longrightarrow (\Omega, \mathcal{F})$ be a function. Then, the two following properties are equivalent.*

(1) $X_n \circ \psi$ is measurable for all $n \geqslant 0$.

(2) ψ is measurable.

Proof : We only need to prove that (1) $\Rightarrow$ (2). Let

$$\mathcal{F}' = \{A \in \mathcal{F} : \psi^{-1}(A) \in \mathcal{G}\}.$$

It is not hard to check that $\mathcal{F}'$ is indeed a σ-algebra on Ω. For any fixed $y \in E$, since E is discrete, $\{y\}$ is measurable. By (1), we know that $\psi^{-1}(X_n^{-1}(\{y\})) = (X_n \circ \psi)^{-1}(\{y\}) \in \mathcal{G}$, so $X_n^{-1}(\{y\}) \in \mathcal{F}'$. Therefore, $\mathcal{F}'$ makes all the coordinate functions X_n measurable, giving the equality $\mathcal{F}' = \mathcal{F}$. $\square$

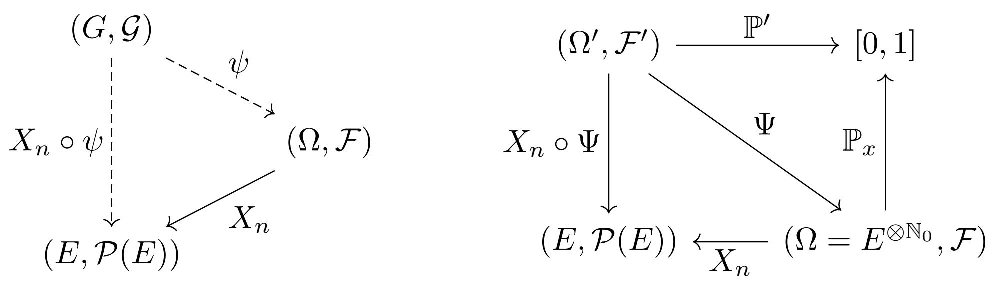

Figure 7.1: **Left**: The coordinate function X_n is measurable for all $n \geqslant 0$. Lemma 7.2.2 states that ψ is measurable if and only if $X_n \circ \psi$ is measurable for all $n \geqslant 0$. **Right**: For the construction of the canonical Markov chain in Theorem 7.2.3, we push forward the measure defined on $(\Omega', \mathcal{F}')$ to the canonical space $(\Omega, \mathcal{F})$.

Theorem 7.2.3 (Canonical Markov chain) **:** *Let Q be a transition matrix on E. Then, for all $x \in E$, there exists a unique probability measure $\mathbb{P}_x$ on $\Omega = E^{\mathbb{N}_0}$ such that the stochastic process defined by the coordinate functions $(X_n)_{n \geqslant 0}$ is a Markov chain with transition matrix Q and $\mathbb{P}_x(X_0 = x) = 1$.*

Proof : Let $x \in E$. Proposition 7.2.1 gives the existence of a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ on which we can construct a Markov chain $(X_n^x)_{n \geqslant 0}$ with transition matrix Q and $X_0^x = x$. From Lemma 7.2.2, the below function is measurable,

$$\begin{array}{rccl} \Psi: & \Omega' & \longrightarrow & \Omega = E^{\otimes \mathbb{N}_0} \\ & \omega' & \longrightarrow & (X_n^x(\omega'))_{n \geqslant 0} \end{array}.$$

Let us define $\mathbb{P}_x$ to be the image measure of $\mathbb{P}'$ under Ψ, that is $\mathbb{P}_x = \mathbb{P}' \circ \Psi^{-1}$, then we have

$$\mathbb{P}_x(X_0 = x) = \mathbb{P}' \circ \Psi^{-1}(X_0 = x) = \mathbb{P}'(X_0^x = x) = 1.$$

Then, for all $x_0, x_1, \ldots, x_n \in E$, we also have

$$\begin{aligned} \mathbb{P}_x(X_0 = x_0, \ldots, X_n = x_n) &= \mathbb{P}'(X_0^x = x_0, \ldots, X_n^x = x_n) \\ &= \mathbb{P}'(X_0^x = x_0) Q(x_0, x_1) \ldots Q(x_{n-1}, x_n) \\ &= \mathbb{P}_x(X_0 = x_0) Q(x_0, x_1) \ldots Q(x_{n-1}, x_n). \end{aligned}$$

Hence, from Proposition 7.1.8, under $\mathbb{P}_x$, $(X_n)_{n \geqslant 0}$ is a Markov chain with transition matrix Q and initial position x.

Then we discuss uniqueness. If $\mathbb{P}_x$ and $\mathbb{P}'_x$ both satisfy the properties required by the theorem, then they take the same value on all cylindrical events. Since the set of cylindrical events is closed under finite intersections and generate the σ-algebra $\mathcal{F}$, the monotone class lemma tells us that $\mathbb{P}_x = \mathbb{P}'_x$ (see Corollary 1.1.19). □

Remark 7.2.4 :

(1) From Proposition 7.1.8, we know that we have, for all $n \geqslant 0$ and $x, y \in E$, that

$$\mathbb{P}_x(X_n = y) = Q_n(x, y).$$

(2) If μ is a probability measure on E, denote

$$\mathbb{P}_\mu = \sum_{x \in E} \mu(x) \mathbb{P}_x,$$

which is a probability measure on Ω. We can check that under $\mathbb{P}_\mu$, the process $(X_n)_{n \geqslant 0}$ of coordinate functions is a Markov chain with transition matrix Q and X_0 has distribution μ.

(3) On the probability space $(\Omega', \mathcal{F}', \mathbb{P}')$, if $(X'_n)_{n \geqslant 0}$ is a Markov chain with transition matrix Q and initial distribution μ, then for any measurable set $B \subseteq \Omega = E^{\mathbb{N}_0}$, we have

$$\mathbb{P}'((X'_n)_{n \geqslant 0} \in B) = \mathbb{P}_\mu(B).$$

In the proof of Theorem 7.2.3, we know that the above holds for all cylindrical events, hence for all the measurable events from the monotone class lemma. This equality tells us that any statement that holds for the canonical Markov chain will also be valid for any Markov chain with the same transition matrix, and the same initial condition.

Next, we will discuss the *Markov properties* (馬可夫性質), and some more notations need to be introduced.

Definition 7.2.5 : First, we define the translation operator (平移算子). For all non-negative integer $k \geqslant 0$, define $\theta_k : \Omega \longrightarrow \Omega$ to be

$$\theta_k((\omega_n)_{n\geqslant 0}) = (\omega_{k+n})_{n\geqslant 0}.$$

From Lemma 7.2.2, we know that these operators are measurable.

Also, let us write $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$ for all $n \geqslant 0$, meaning that $(\mathcal{F}_n)_{n\geqslant 0}$ is the canonical filtration (正則濾鏈) of $(X_n)_{n\geqslant 0}$. Besides, we write $\mathbb{E}_x$ for the expectation under the probability measure $\mathbb{P}_x$.

Theorem 7.2.6 (Simple Markov property) **:** *Let G be a non-negative measurable function on Ω. Fix $n \geqslant 0$. Then, for any fixed $x \in E$, we have*

$$\mathbb{E}_x[F \cdot G \circ \theta_n] = \mathbb{E}_x[F \ \mathbb{E}_{X_n}[G]], \tag{7.8}$$

for every non-negative $\mathcal{F}_n$-measurable function F. Equivalently, for any fixed $x \in E$, the above can be reformulated as follows,

$$\mathbb{E}_x[G \circ \theta_n \mid \mathcal{F}_n] = \mathbb{E}_{X_n}[G], \tag{7.9}$$

which can be interpreted as: the conditional distribution of $\theta_n(\omega)$ knowing $(X_0, \ldots, X_n)$ is $\mathbb{P}_{X_n}$.

Proof : The second part of the theorem is equivalent to the first part, so we only need to show the first part. To prove Eq. (7.8), we only need to check it for all $x_0, x_1, \ldots, x_n \in E$ and F of the form

$$F = \mathbb{1}_{X_0=x_0,\ldots,X_n=x_n}.$$

Similarly, for $p \geqslant 0$ and $y_0, \ldots, y_p \in E$, consider

$$G = \mathbb{1}_{X_0=y_0,\ldots,X_p=y_p}. \tag{7.10}$$

Given $y \in E$, we have

$$\mathbb{E}_y[G] = \mathbb{1}_{y_0=y} Q(y_0, y_1) \cdots Q(y_{p-1}, y_p),$$

and

$$\begin{aligned}\mathbb{E}_x[F \cdot G \circ \theta_n] &= \mathbb{P}_x(X_0 = x_0, \ldots, X_n = x_n, X_n = y_0, X_{n+1} = y_1, \ldots, X_{n+p} = y_p) \\ &= \mathbb{1}_{x_0=x} Q(x_0, x_1) \cdots Q(x_{n-1}, x_n) \mathbb{1}_{y_0=x_n} Q(y_0, y_1) \cdots Q(y_{p-1}, y_p),\end{aligned}$$

so Eq. (7.8) is true for G of the form Eq. (7.10). The monotone class lemma then allows to conclude that Eq. (7.8) holds for all $G = \mathbb{1}_A$ with $A \in \mathcal{F}$. □

The above theorem is a general form of the simple Markov property, meaning that knowing the past $(X_0, \ldots, X_n)$, the distribution $\theta_n(\omega)$ of the Markov chain in the future time only depends on the current state X_n. Below we will talk about the *strong Markov property* (強馬可夫性質), in which case the "current time" needs not to be a deterministic time n but is given by a random variable T. Later in Section 7.3, we will see some applications of the strong Markov property.

Theorem 7.2.7 (Strong Markog property) **:** *Let T be a stopping time with respect to the filtration $(\mathcal{F}_n)_{n\geqslant 0}$ and G be a non-negative measurable function defined on Ω. Then, for any fixed $x \in E$, we have*

$$\mathbb{E}_x[\mathbb{1}_{T<\infty} F \cdot G \circ \theta_T] = \mathbb{E}_x[\mathbb{1}_{T<\infty} F \ \mathbb{E}_{X_T}[G]], \tag{7.11}$$

for every non-negative $\mathcal{F}_T$-measurable function F. The above can also be reformulated as below for any fixed $x \in E$,

$$\mathbb{E}_x[\mathbb{1}_{T<\infty} G \circ \theta_T \mid \mathcal{F}_T] = \mathbb{1}_{T<\infty} \mathbb{E}_{X_T}[G]. \tag{7.12}$$

Remark 7.2.8 : We can note that the random variable X_T is defined on a $\mathcal{F}_T$-measurable set $\{T < \infty\}$ and is $\mathcal{F}_T$-measurable. The expectation $\mathbb{E}_{X_T}[G]$ is a random variable which is also defined on $\{T < \infty\}$. It can be seen as the composition between the functions $\omega \mapsto X_T(\omega)$ and $x \mapsto \mathbb{E}_x[G]$.

Proof : Given $n \geqslant 0$, since $\mathbb{1}_{T=n} F$ is $\mathcal{F}_n$-measurable, we can apply the simple Markov property stated in Theorem 7.2.6, giving

$$\mathbb{E}_x[\mathbb{1}_{T=n} F \cdot G \circ \theta_T] = \mathbb{E}_x[\mathbb{1}_{T=n} F \cdot G \circ \theta_n] = \mathbb{E}_x[\mathbb{1}_{T=n} F \ \mathbb{E}_{X_n}[G]].$$

Then, we take the summation over $n \in \mathbb{N}_0$ and obtain Eq. (7.11). □

Corollary 7.2.9 : *Let T be a stopping time satisfying $\mathbb{P}_x(T < \infty) = 1$. Suppose that there exists $y \in E$ such that $\mathbb{P}_x(X_T = y) = 1$, then under the probability measure $\mathbb{P}_x$, $\theta_T(\omega)$ is independent of $\mathcal{F}_T$ and has the same distribution as $\mathbb{P}_y$.*

Proof : It is a direct application of Theorem 7.2.7,

$$\mathbb{E}_x[F \cdot G(\theta_T(\omega))] = \mathbb{E}_x[F \ \mathbb{E}_{X_T}[G]] = \mathbb{E}_x[F \ \mathbb{E}_y[G]] = \mathbb{E}_x[F] \, \mathbb{E}_y[G].$$

□