3. Markov chains

3.1. **Definitions and examples.**

Definition 3.1. Let $(S,\mathcal{C})$ be a measurable space, $(\mathcal{F}_n)_{n=0}^\infty$ be a filtration and $(X_n)_{n=0}^\infty$ be a stochastic process taking valued on S. X_n is called a *Markov chain* w.r.t. $\mathcal{F}_n$ if X_n is $\mathcal{F}_n$-measurable and

$$\mathbb{P}(X_{n+1}\in B|\mathcal{F}_n)=\mathbb{P}(X_{n+1}\in B|X_n),\quad \forall B\in\mathcal{C},\, n\ge 0.$$

The distribution of X_0 is called the initial distribution.

Lemma 3.1. *A sequence of random elements X_n taking values in $(S,\mathcal{C})$ is a Markov chain w.r.t. a filtration $\mathcal{F}_n$ if and only if*

$$\mathbb{E}(f(X_{n+1})|\mathcal{F}_n)=\mathbb{E}(f(X_{n+1})|X_n), \tag{3.1}$$

where $f:S\to\mathbb{R}$ is any bounded $\mathcal{C}$-measurable function.

Proof. The sufficient condition for Markov chains is clear. For the necessary condition, let H be the class of all bounded $\mathcal{C}$-measurable function f such that (3.1) holds. Obviously, H is a linear space containing the constant function $\mathbf{1}$ and the multiplicative system $\{\mathbf{1}_B : B\in\mathcal{C}\}$. Also, H is closed under the bounded convergence. By the multiplicative system theorem, H contains all bounded $\mathcal{C}$-measurable functions. □

Example 3.1 (Random walks). Let $X_0,\xi_1,\xi_2,...$ be independent random elements taking values in $\mathbb{R}^d$. Let $X_n=X_0+\xi_1+\cdots+\xi_n$ and $\mathcal{F}_n=\mathcal{F}(X_0,\xi_1,...,\xi_n)$. Then X_n is a Markov chain w.r.t. $\mathcal{F}_n$. To see this fact, let μ_n be the distribution of ξ_n. Note that, for any random elements X,Y taking values on $(R,\mathcal{B})$ and $(S,\mathcal{C})$, if Y is $\mathcal{F}$-measurable, $\mathcal{F}(X)$ is independent of $\mathcal{F}$ and φ is a random variable defined on $(R\times S,\mathcal{B}\times\mathcal{C})$ satisfying $\mathbb{E}|\varphi(X,Y)|<\infty$, then $\mathbb{E}(\varphi(X,Y)|\mathcal{F})=\phi(Y)$, where $\phi(y)=\mathbb{E}\varphi(X,y)$. Replacing $X,Y,\mathcal{F},\varphi(x,y)$ with $\xi_{n+1},X_n,\mathcal{F}_n,\mathbf{1}_B(x+y)$ yields

$$\mathbb{P}(X_{n+1}\in B|\mathcal{F}_n)=\mu_{n+1}(B-X_n),$$

where $\mathbb{E}\mathbf{1}_B(\xi_{n+1},x_n)=\mathbb{P}(x_n+\xi_{n+1}\in B)=\mu_{n+1}(B-x_n)$ is used. Similarly, one has $\mathbb{P}(X_{n+1}\in B|X_n)=\mu_{n+1}(B-X_n)$.

Remark 3.1. It follows immediately from (3.1) that, for bounded $\mathcal{C}$-measurable functions $f_0,...,f_k$ with $k\ge 1$,

$$\begin{aligned}\mathbb{E}\left(\prod_{i=0}^k f_i(X_{n+i})\middle|\mathcal{F}_n\right)&=\mathbb{E}\left(\prod_{i=0}^{k-1} f_i(X_{n+i})\mathbb{E}(f_k(X_{n+k})|\mathcal{F}_{n+k-1})\middle|\mathcal{F}_n\right)\\&=\mathbb{E}\left(\prod_{i=0}^{k-2} f_i(X_{n+i})g(X_{n+k-1})|\mathcal{F}_n\right),\end{aligned}$$

where $\prod_{i=0}^{-1}:=1$ and $g(x)=f_{k-1}(x)\mathbb{E}(f_k(X_{n+k})|X_{n+k-1}=x)$, which is bounded and $\mathcal{C}$-measurable. By induction, one has

$$\begin{aligned}\mathbb{E}\left(\prod_{i=0}^{k-2} f_i(X_{n+i})g(X_{n+k-1})|\mathcal{F}_n\right)&=\mathbb{E}\left(\prod_{i=0}^{k-2} f_i(X_{n+i})g(X_{n+k-1})|X_n\right)\\&=\mathbb{E}\left(\prod_{i=0}^{k-1} f_i(X_{n+i})\mathbb{E}(f_k(X_{n+k})|\mathcal{F}_{n+k-1})\middle|X_n\right)\\&=\mathbb{E}\left(\prod_{i=0}^{k} f_i(X_{n+i})\middle|X_n\right).\end{aligned}$$

As a result, we obtain

$$\mathbb{E}\left(\prod_{i=0}^{k} f_i(X_{n+i})\middle|\mathcal{F}_n\right) = \mathbb{E}\left(\prod_{i=0}^{k} f_i(X_{n+i})\middle|X_n\right), \quad \forall k \geq 0, \tag{3.2}$$

where the case of $k = 0$ is obvious.

Thereafter, we need the following notations. For $n \geq 0$, let $\mathcal{C}_{n+1} = \mathcal{C} \otimes \cdots \otimes \mathcal{C}$ be the product σ-field over S^{n+1} and $\mathcal{C}_\infty = \mathcal{C} \otimes \mathcal{C} \otimes \cdots$ be the product σ-field over S^∞.

Lemma 3.2. *Let X_n be a Markov chain on $(S, \mathcal{C})$ w.r.t. $\mathcal{F}_n$ and $\mathcal{C}_\infty$. For any bounded $\mathcal{C}_\infty$-measurable function f, one has*

$$\mathbb{E}(f(X_n, X_{n+1}, ...)|\mathcal{F}_n) = \mathbb{E}(f(X_n, X_{n+1}, ...)|X_n).$$

Proof. By the Lebesgue dominated convergence theorem, it suffices to prove that

$$\mathbb{P}((X_n, X_{n+1}, ...) \in B|\mathcal{F}_n) = \mathbb{P}((X_n, X_{n+1}, ...) \in B|X_n).$$

By the π-λ lemma, it remains to consider the case $B = B_0 \times \cdots \times B_k \times S^\infty$, where $B_0, ..., B_k \in \mathcal{C}$, $S^\infty = S \times S \times \cdots$ and $k \geq 0$, and this is given by (3.2) with $f_i = \mathbf{1}_{B_i}$, as desired. ☐

Remark 3.2. It follows immediately from Lemma 3.2 that, for $n \in \mathbb{N}$ and any bounded $\mathcal{C}_{n+1}$-measurable and $\mathcal{C}_\infty$-measurable functions f, g,

$$\mathbb{E}(f(X_0, ..., X_n)g(X_n, X_{n+1}, ...)|X_n) = \mathbb{E}(f(X_0, ..., X_n)|X_n)\mathbb{E}(g(X_n, X_{n+1}, ...)|X_n).$$

In particular, for $A \in \mathcal{F}(X_0, X_1, ..., X_n)$ and $B \in \mathcal{F}(X_n, X_{n+1}, ...)$,

$$\mathbb{P}(A \cap B|X_n) = \mathbb{P}(A|X_n)\mathbb{P}(B|X_n).$$

Definition 3.2. Let $(S, \mathcal{C})$ be a measurable space. A function $p : S \times \mathcal{C} \to \mathbb{R}$ is said to be a *transition probability* or *transition function* if:

1. For each $x \in S$, $p(x, \cdot)$ is a probability on $(S, \mathcal{C})$;
2. For each $A \in \mathcal{C}$, $p(\cdot, A)$ is a $\mathcal{C}$-measurable function.

A process X_n is said to be a Markov chain w.r.t. $\mathcal{F}_n$ with *transition probabilities p_n* if

$$\mathbb{P}(X_{n+1} \in B|\mathcal{F}_n) = p_n(X_n, B) \quad \forall B \in \mathcal{C},\ n \geq 1.$$

Remark 3.3. Note that if $p : S \times \mathcal{C} \to \mathbb{R}$ is a transition probability and f is a bounded $\mathcal{C}$-measurable function, then the following map

$$x \mapsto \int_S f(y)p(x, dy) \tag{3.3}$$

is $\mathcal{C}$-measurable. To see the details, let H be the class of all bounded $\mathcal{C}$-measurable functions satisfying (3.3). It is obvious that H is a linear space containing $\mathbf{1}_S$ and is closed under bounded convergence. Since $\mathbf{1}_B \in H$ for all $B \in \mathcal{C}$, the multiplicative system theorem implies that H is the class of all bounded $\mathcal{C}$-measurable functions.

Theorem 3.3. *Let X_n be a Markov chain on $(S, \mathcal{C})$ with transition probabilities p_n and initial distribution μ. Fix $m \geq 1$ and let f be a bounded $\mathcal{C}_{m+1}$-measurable function. Set*

$$\varphi(x) = \int_S p_n(x, dy_1) \cdots \int_S p_{n+m-1}(y_{m-1}, dy_m) f(x, y_1, ..., y_m).$$

Then, φ is well-defined, $\mathcal{C}$-measurable and

$$\mathbb{E}(f(X_n, ..., X_{n+m})|\mathcal{F}_n) = \varphi(X_n), \quad \forall n \geq 0.$$

In particular,

$$\mathbb{E}f(X_0,...,X_m)=\int_S\mu(dx_0)\int_S p_0(x_0,dx_1)\cdots\int_S p_{m-1}(x_{m-1},dx_m)f(x_0,...,x_m). \tag{3.4}$$

and, for $B_0,...,B_n\in\mathcal{C}$,

$$\mathbb{P}(X_m\in B_m,\ 0\le m\le n)=\int_{B_0}\mu(dx_0)\int_{B_1}p_0(x_0,dx_1)\cdots\int_{B_n}p_{n-1}(x_{n-1},dx_n). \tag{3.5}$$

Proof. The second and third identities are special cases of the first one. For the first identity, it suffices to consider $f(y_0,...,y_m)=\prod_{i=0}^m f_i(y_i)$, where $f_0,..,f_m$ are bounded $\mathcal{C}$-measurable functions, by the multiplicative system theorem. The well-definedness and $\mathcal{C}$-measurability of φ can be proved by induction and the details are skipped. For the identity, the case of $m=1$ is immediate from Proposition 1.10. For $m>1$, note that $\mathbb{E}(f_m(X_{n+m})|\mathcal{F}_{n+m-1})=g(X_{n+m})$, where $g(x)=\int_S p_{n+m-1}(x,dy_m)f_m(y_m)$. By induction, we may write

$$\begin{aligned}&\mathbb{E}(f(X_n,...,X_{n+m})|\mathcal{F}_n)=\mathbb{E}[\mathbb{E}(f(X_n,...,X_{n+m})|\mathcal{F}_{n+m-1})|\mathcal{F}_n]\\=&\mathbb{E}[f_0(X_n)\cdots f_{m-2}(X_{n+m-2})(f_{m-1}g)(X_{n+m-1})|\mathcal{F}_n]=\widetilde{\varphi}(X_n),\end{aligned}$$

where

$$\widetilde{\varphi}(x)=\int_S p_n(x,dy_1)\cdots\int_S p_{n+m-2}(y_{m-2},dy_{m-1})f_0(x)f_1(y_1)\cdots f_{m-1}(y_{m-1})g(y_{m-1})=\varphi(x),$$

as desired. □

Note that (3.1) can be rewritten as

$$\mathbb{P}(X_m\in B_m,\ 0\le m\le n)=\int_S\mu(dx_0)\int_S p_0(x_0,dx_1)\cdots\int_S\mathbf{1}_{B_0\times\cdots\times B_n}p_{n-1}(x_{n-1},dx_n).$$

Based on this observation, it is natural to consider the following function

$$\mathbb{P}^{(n)}_\mu(A)=\int_S\mu(dx_0)\int_S p_0(x_0,dx_1)\cdots\int_S\mathbf{1}_A p_{n-1}(x_{n-1},dx_n), \tag{3.6}$$

for all $A\in\mathcal{C}_{n+1}$ and $n\ge 0$. By (3.4), (3.6) defines a probability on $(S^{n+1},\mathcal{C}_{n+1})$ and

$$\mathbb{P}^{(n+1)}_\mu(A\times S)=\mathbb{P}^{(n)}_\mu(A),\quad\forall A\in\mathcal{C}_{n+1},\,n\ge 0.$$

Lemma 3.4. *Let $\mathbb{P}^{(n)}_\mu$ be the probability in (3.6). For $n\ge 0$, set $X_n(\omega)=\omega_n$ for $\omega=(\omega_n)_{n=0}^\infty\in S^\infty$ and $\mathcal{F}_n=\mathcal{C}_{n+1}$. Assume that there is a probability $\mathbb{P}_\mu$ on $(S^\infty,\mathcal{C}_\infty)$ satisfying $\mathbb{P}_\mu(A\times S^\infty)=\mathbb{P}^{(n)}_\mu(A)$ for all $A\in\mathcal{C}_{n+1}$ and $n\ge 0$. Then, X_n is a Markov chain w.r.t. $\mathcal{F}_n$ with initial distribution μ and transition probability p_n.*

Proof. It is clear that X_n is adapted to $\mathcal{F}_n$. Note that, for $A=\{X_i\in B_i,\,0\le i\le n\}$ and $B_{n+1}\in\mathcal{C}$,

$$\begin{aligned}&\int_A\mathbf{1}_{\{X_{n+1}\in B_{n+1}\}}d\mathbb{P}_\mu=\mathbb{P}_\mu(A,X_{n+1}\in B_{n+1})=\mathbb{P}_\mu(B_0\times\cdots\times B_{n+1}\times S^\infty)\\=&\mathbb{P}^{(n)}_\mu(B_0\times\cdots\times B_{n+1})=\int_{B_0}\mu(dx_0)\int_{B_1}p_0(x_0,dx_1)\cdots\int_{B_n}p_{n-1}(x_{n-1},dx_n)p_n(x_n,B_{n+1}).\end{aligned}$$

By the multiplicative system theorem, one can show that

$$\int_{B_0}\mu(dx_0)\int_{B_1}p_0(x_0,dx_1)\cdots\int_{B_n}p_{n-1}(x_{n-1},dx_n)f(x_n)=\int_A f(X_n)d\mathbb{P}_\mu,$$

for any bounded $\mathcal{C}$-measurable function f. As $p_n(\cdot, B_{n+1})$ is $\mathcal{C}$-measurable, this implies

$$\int_A \mathbf{1}_{\{X_{n+1}\in B_{n+1}\}} d\mathbb{P}_\mu = \int_A p_n(X_n, B_{n+1}) d\mathbb{P}_\mu.$$

As a consequence of the π-λ lemma, the above identity holds for all $A \in \mathcal{F}_n$ or equivalently $\mathbb{P}_\mu(X_{n+1} \in B|\mathcal{F}_n) = p_n(X_n, B)$. Since $p_n(X_n, B)$ is $\mathcal{F}(X_n)$-measurable, we obtain

$$\mathbb{P}_\mu(X_{n+1} \in B|\mathcal{F}_n) = \mathbb{E}_\mu\left(\mathbf{1}_{\{X_{n+1}\in B\}}|\mathcal{F}_n\right) = \mathbb{E}_\mu\left(\mathbb{E}\left(\mathbf{1}_{\{X_{n+1}\in B\}}|\mathcal{F}_n\right)|X_n\right) = \mathbb{P}_\mu(X_{n+1} \in B|X_n),$$

where $\mathbb{E}_\mu$ denotes the expectation under $\mathbb{P}_\mu$. □

Note that the transition probability is closely related to the regular condition probability and distribution. Recall that $(S, \mathcal{C})$ is a Borel space if there is $R \in \mathcal{B}(\mathbb{R})$ and a bijection $\varphi : (S, \mathcal{C}) \to (R, \mathcal{B}(R))$, where $\mathcal{B}(R) = \{R \cap E | E \in \mathcal{B}(\mathbb{R})\}$, such that φ and φ^{-1} are measurable.

Lemma 3.5. *Let X_n be a Markov chain on $(S, \mathcal{C})$ w.r.t. $\mathcal{F}_n$. If $(S, \mathcal{C})$ is a Borel space, then there exist transition probabilities for X_n.*

Proof. Since $(S, \mathcal{C})$ is a Borel space, there exists a regular conditional distribution for X_{n+1} given $X_n = x$ and we write it as p_n, which means that, for any $x \in S$, $p_n(x, \cdot)$ is a probability and, for any $B \in \mathcal{C}$, $p_n(x, B)$ is a version for $\mathbb{P}(X_{n+1} \in B|X_n = x)$. This implies that p_n is a transition probability and

$$\mathbb{P}(X_{n+1} \in B|\mathcal{F}_n) = \mathbb{P}(X_{n+1} \in B|X_n) = p_n(X_n, B) \quad \forall B \in \mathcal{C}, n \geq 0.$$

□

Theorem 3.6. *Let μ be a probability measure and p_n be a sequence of transition functions on $S \times \mathcal{C}$. If $(S, \mathcal{C})$ is a Borel space, then there exists a Markov chains X_n on $(S, \mathcal{C})$ with transition probability p_n and initial distribution μ.*

Proof. For $n \geq 0$, let $\mathbb{P}_\mu^{(n)}$ be the probability defined by (3.6). Note that $\mathbb{P}_\mu^{(n)}$ possesses the consistency property and $(S, \mathcal{C})$ is a Borel space. By the Kolmogorov extension theorem, there is an extension probability on $(S^\infty, \mathcal{C}_\infty)$, say $\mathbb{P}_\mu$, such that $\mathbb{P}_\mu(A \times S^\infty) = \mathbb{P}_\mu^{(n)}(A)$ for $A \in \mathcal{C}_{n+1}$ and $n \geq 0$. The remaining proof is then given by Lemma 3.4. □

Example 3.2 (Markov chains with discrete state spaces). Assume that S is a countable set and $\mathcal{C} = 2^S$. It is clear that $(S, \mathcal{C})$ is a Borel space. Suppose $p_n(i, j) \geq 0$ and $\sum_j p_n(i, j) = 1$ for all $i \in S$ and $n \geq 0$. Then, $p_n(i, A) = \sum_{j\in A} p_n(i, j)$ defines a transition probability. By Theorem 3.6, there is a Markov chain on S with transition probabilities p_n and this implies $p_n(i, j) = P(X_{n+1} = j|X_n = i)$.

Exercise 3.1. Let $\mathbb{P}_\mu$ and $\mathbf{X} = (X_0, X_1, ...)$ be the probability and the stochastic process created in the proofs of Theorem 3.6 and Lemma 3.4. If $\mu = \delta_x$, the Dirac measure at x, we simply write $\mathbb{P}_x$ for $\mathbb{P}_{\delta_x}$. Prove that, for $B \in \mathcal{C}_\infty$, the map $x \mapsto \mathbb{P}_x(B)$ is $\mathcal{C}$-measurable and

$$\mathbb{P}_\mu(B) = \int_S \mu(dx)\mathbb{P}_x(B), \quad \forall B \in \mathcal{C}_\infty.$$

Use the above equality to conclude that, for $B \in \mathcal{C}_\infty$,

$$\mathbb{P}_\mu(\mathbf{X} \in B|X_0 = x) = \mathbb{P}_x(\mathbf{X} \in B).$$

Hint: The π-λ lemma.

Remark 3.4. It follows immediately from Exercise 3.1 that, for any bounded $\mathcal{C}_\infty$-measurable function f,

$$\mathbb{E}_\mu(f(\mathbf{X})|X_0 = x) = \mathbb{E}_x f(\mathbf{X}), \quad \mathbb{E}_\mu f(\mathbf{X}) = \int_S \mu(dx)\mathbb{E}_x f(\mathbf{X}).$$

Example 3.3 (Branching processes). Let $S = \{0, 1, ...\}$ and ξ_i^n, $i, n \geq 1$, be i.i.d. nonnegative integer-valued random variables. Set

$$p(i,j) = \mathbb{P}\left(\sum_{k=1}^{i} \xi_k^n = j\right).$$

Let Z_n be the number of the population at time n. Then, Z_n forms a (time homogeneous) Markov chain w.r.t. $\mathcal{F}_n = \mathcal{F}(\xi_i^m, i \geq 0, m \leq n)$ with common transition probability p. In details, one has

$$\begin{aligned}&\mathbb{P}(Z_{n+1} = j | Z_1 = i_1, ..., Z_{n-1} = i_{n-1}, Z_n = i)\\ &\qquad = \mathbb{P}\left(\sum_{k=1}^{i} \xi_k^{n+1} = j\right) = p(i,j) = \mathbb{P}(Z_{n+1} = j | Z_n = i)\end{aligned}$$

Example 3.4 (Renewal chains). Let a_k be a sequence of nonnegative real numbers summing up to 1. A renewal chain is a (time homogeneous) Markov chain with common transition probability p given by

$$p(i,j) = \begin{cases} a_{k+1} & \text{if } (i,j) = (0,k),\ k \geq 0 \\ 1 & \text{if } (i,j) = (k, k-1) \\ 0 & \text{o.w.} \end{cases}$$

Concerning the meaning of a renewal chain, let's consider the following setting. Let $\xi_1, \xi_2, ...$ be i.i.d. random variables with $\mathbb{P}(\xi_n = j) = a_j$ and $T_0 = i_0$. For $k > 0$, set $T_k = T_{k-1} + \xi_k$. T_k should be viewed as a sequence of renewal times. It is worthwhile to note that T_k forms a Markov chain. Let

$$Y_m = \begin{cases} 1 & \text{if } m \in \{T_0, T_1, ...\} \\ 0 & \text{o.w.} \end{cases}$$

and set $X_n = \inf\{m - n : Y_m = 1,\ m \geq n\}$. X_n is the amount of time until the first renewal after time n. We shall prove in the following that X_n is a (time homogeneous) Markov chain w.r.t. $\mathcal{F}_n = \mathcal{F}(\xi_1, ..., \xi_n)$ with common transition probability p.

Note that T_k is adapted to $\mathcal{F}_k$. Let $N_n = \inf\{k : T_k \geq n\}$. Clearly, N_n is a stopping time on the filtration $\mathcal{F}_n = \mathcal{F}(\xi_1, ..., \xi_n)$. It is easy to see that $T_{N_n} = X_n + n$ and this implies

$$\mathcal{F}(T_{N_1}, ..., T_{N_n}) = \mathcal{F}(X_1, ..., X_n).$$

Since N_n is nondecreasing in n, for $i_1 \leq \cdots \leq i_{n-1} \leq i_n$ and $k \geq 0$,

$$\{T_{N_j} = i_j, 1 \leq j \leq n, N_n = k\} = \bigcup_{0 \leq \ell_1 \leq \cdots \leq \ell_{n-1} \leq \ell_n = k} \{T_{\ell_j} = i_j, N_j = \ell_j, 1 \leq j \leq n\} \in \mathcal{F}_k.$$

This yields

$$\mathcal{F}(T_{N_1}, ..., T_{N_n}) \subset \mathcal{F}_{N_n} = \mathcal{F}(\xi_1, \xi_2, ..., \xi_{N_n}).$$

If $X_n = i > 0$, then $X_{n+1} = i - 1$. This implies

$$\mathbb{P}(X_{n+1} = i - 1 | X_n = i, X_{n-1} = i_{n-1}, ..., X_1 = i_1) = 1 = \mathbb{P}(X_{n+1} = i - 1 | X_n = i).$$

If $X_n = 0$, then $X_{n+1} = \xi_{N_n+1} - 1$. Since $\mathbb{P}(N_n < \infty) = 1$, $\mathcal{F}_{N_n}$ is independent of $\mathcal{F}(\xi_{N_n+1})$ and ξ_{N_n+1} has the same distribution as ξ_1. (Why?) Hence, we have

$$\mathbb{P}(X_{n+1} = k | X_n = 0, X_{n-1} = i_{n-1}, ..., X_1 = i_1) = a_{k+1} = \mathbb{P}(X_{n+1} = k | X_n = 0).$$

This proves that X_n is a Markov chain with transition probability p.

Example 3.5 (Ehrenfest chain). An Ehrenfest chain is a (time homogeneous) Markov chain on $\{0, 1, ..., r\}$ with the following common transition probability.

$$\begin{cases} p(i, i+1) = 1 - i/r & \text{for } 0 \le i < r \\ p(i, i-1) = i/r & \text{for } 0 < i \le r \\ p(i, j) = 0 & \text{otherwise} \end{cases}$$

Paul Ehrenfest uses this chain to model the diffusion of air molecules between two chambers connected by a small hole and explain the second law of thermodynamics.

Proposition 3.7. *Let S be a countable set and X_n is a Markov chain on S with transition probability p_n and initial distribution μ. Then, for $n \ge 1$, $\mathbb{P}(X_n = j) = (\mu p_0 p_1 \cdots p_{n-1})(j)$ for $j \in S$, where*

$$(p_0 p_1 \cdots p_m)(i, j) = \sum_{k \in S} (p_0 p_1 \cdots p_{m-1})(i, k) p_m(k, j)$$

and μp^n is the multiplication of the row vector μ and p^n.

Proof. Immediate from the following fact.

$$\mathbb{P}(X_k = i_k, 0 \le k \le n) = \mu(x_0) \prod_{k=0}^{n-1} p(x_k, x_{k+1}).$$

□

Exercise 3.2. Let $S = \{0, 1\}$ and X_n be a (time homogeneous) Markov chain on S with common transition probability p given by

$$p = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}.$$

Show that for $n \ge 0$,

$$\mathbb{P}(X_n = 0) = \frac{b}{a+b} + (1-a-b)^n \left(\mu(0) - \frac{b}{a+b} \right).$$

Exercise 3.3. Let $\xi_1, \xi_2, ...$ be i.i.d. random variables taking values on $\{1, 2, ..., N\}$ satisfying $\mathbb{P}(\xi_1 = i) = 1/N$ for $1 \le i \le N$. Set $X_n = |\{\xi, ..., \xi_n\}|$ where $|A|$ denotes the number of different elements in A. Prove that X_n is a Markov chain and describe the transition probability.

Exercise 3.4. Let $\xi_1, \xi_2, ...$ be i.i.d. random variables satisfying $\mathbb{P}(\xi_1 = 1) = P(\xi_1 = -1) = 1/2$. Let $S_0 = 0$, $S_n = \xi_1 + \cdots + \xi_n$ and set $X_n = \max\{S_m : 0 \le m \le n\}$. Show that X_n is not a Markov chain.

3.2. **Markov property and strong Markov property.** A naive way to define a time homogeneous Markov chain X_n is to consider the following identity

$$\mathbb{P}(X_{n+1} \in B | X_n = x) = \mathbb{P}(X_1 \in B | X_0 = x), \quad \forall B \in \mathcal{C},\ n \ge 0, \tag{3.7}$$

where the equality means that there is a common version for $\mathbb{P}(X_{n+1} \in B | X_n = x)$ and $\mathbb{P}(X_1 \in B | X_0 = x)$. Note that such a definition of homogeneity for Markov chains can't be easily fulfilled and any theorem like Lemmas 3.1 and 3.2 may be generated with complicated priori assumptions. However, if X_n possesses a common transition probability, then (3.7) turns out an obvious request. Thus, it is reasonable to consider the following definition.

Definition 3.3. A Markov chain X_n on $(S, \mathcal{C})$ with transition probability p_n is *time homogeneous* if $p_n = p_0$ for all $n \ge 0$.

Throughout the remaining of this section, all Markov chains taking values on $(S,\mathcal{C})$ are restricted to stochastic processes $(X_n)_{n=0}^{\infty}$, where $X_n(\omega) = \omega_n$ for all $\omega = (\omega_n)_{n=0}^{\infty} \in S^{\infty}$ and $n \geq 0$ and $\mathcal{F}_n = \mathcal{F}(X_0, ..., X_n)$. When we say that $(X_n)_{n=0}^{\infty}$ is a Markov chain with transition probability p_n and initial distribution μ, it means that $(S^{\infty}, \mathcal{C}_{\infty})$ is equipped with the probability $\mathbb{P}_{\mu}$ generated in the proof of Theorem 3.6. We will use $\mathbb{E}_{\mu}$ to denote the expectation under $\mathbb{P}_{\mu}$. If $\mu = \delta_x$, we simply write $\mathbb{P}_x, \mathbb{E}_x$, for short. Remember that if $(S,\mathcal{C})$ is a Borel space, then $\mathbb{P}_{\mu}$ always exists for any probability μ on $(S,\mathcal{C})$.

Theorem 3.8 (Markov property)**.** *Let X_n be a time homogeneous Markov chain on $(S,\mathcal{C})$ with respect to $\mathcal{F}_n$ with transition probability p. Then, for any bounded $\mathcal{C}_{\infty}$-measurable function f,*

$$\mathbb{E}_{\mu}(f(X_n, X_{n+1}, ...)|\mathcal{F}_n) = \varphi(X_n) \quad \mathbb{P}_{\mu}\text{-a.s.},$$

for all $n \geq 0$, where $\varphi(x) = \mathbb{E}_x f(X_0, X_1, ...)$.

Proof. By the multiplicative system theorem, it suffices to prove the above identity with $f(x_0, x_1, ...) = g(x_0, ..., x_m)$, where g is a bounded $\mathcal{C}_{m+1}$-measurable function, and $m \geq 0$. By Theorem 3.3, one has

$$\mathbb{E}_{\mu}[g(X_n, , ..., X_{n+m})|\mathcal{F}_n] = \phi(X_n),$$

where

$$\phi(x) = \int_S p(x, dx_1) \int_S p(x_1, dx_2) \cdots \int_S p(x_{m-1}, dx_m) g(x, x_1, ..., x_m) = \varphi(x),$$

as desired. □

Corollary 3.9 (Chapman-Kolmogorov equation)**.** *Let X_n be a time homogeneous Markov chain on $(S,\mathcal{C})$. Then, for $B \in \mathcal{C}$,*

$$\mathbb{P}_x(X_{m+n} \in B) = \int_S \mathbb{P}_y(X_n \in B)\mathbb{P}_x(X_m \in dy).$$

Proof. By the Markov property, one has

$$\mathbb{P}_x(X_{m+n} \in B) = \mathbb{E}_x(\mathbb{P}_x(X_{m+n} \in B|\mathcal{F}_m)) = \mathbb{E}_x \varphi(X_m),$$

where $\varphi(y) = \mathbb{P}_y(X_n \in B)$. This implies

$$\mathbb{P}_x(X_{m+n} \in B) = \int_S \varphi(y)\mathbb{P}_x(X_m \in dy) = \int_S \mathbb{P}_y(X_n \in B)\mathbb{P}_x(X_m \in dy).$$

□

Corollary 3.10. *Let X_n be a time homogeneous Markov chain on $(S,\mathcal{C})$ and $A_n, B_n \in \mathcal{C}$ be events satisfying*

$$\mathbb{P}_{\mu}\left(\bigcup_{i=n+1}^{\infty} \{X_i \in B_i\} \middle| X_n \right) \geq \delta > 0 \quad \text{on } \{X_n \in A_n\}.$$

Then, $\mathbb{P}_{\mu}(\{X_n \in A_n \text{ i.o.}\} \setminus \{X_n \in B_n \text{ i.o.}\}) = 0$.

Proof. Let $A = \{X_n \in A_n \text{ i.o.}\}$, $B = \{X_n \in B_n \text{ i.o.}\}$ and $\widetilde{B}_n = \bigcup_{m>n}\{X_n \in B_n\}$. Then, $\mathbf{1}_{\widetilde{B}_n} \to \mathbf{1}_B$ as $n \to \infty$. By the Markov property and Theorem 2.26, we have

$$\mathbb{E}_{\mu}(\mathbf{1}_{\widetilde{B}_n}|X_n) = \mathbb{E}_{\mu}(\mathbf{1}_{\widetilde{B}_n}|\mathcal{F}_n) \overset{a.s.}{\to} \mathbb{E}_{\mu}(\mathbf{1}_B|\mathcal{F}_{\infty}) = \mathbf{1}_B,$$

where $\mathcal{F}_{\infty} = \sigma(\bigcup_n \mathcal{F}_n)$. Note that, almost surely on A, $\mathbf{1}_{B_n} \geq \delta$ for infinitely many n . This implies $A \subset B$ almost surely. □

Exercise 3.5. Let X_n be a time homogeneous Markov chains. A state $a \in S$ is called an absorbing state if $\mathbb{P}_a(X_1 = a) = 1$. Let $D = \{X_n = a \text{ for some } n\}$ and $h(x) = \mathbb{P}_x(D)$. Show that $h(X_n) \to 0$ $\mathbb{P}_\mu$-a.s. on D^c for any initial distribution μ.

Recall the concept of stopped σ-fileds as follows. Let $\mathcal{F}$ be a σ-field over Ω and $\mathcal{F}_n$ be a filtration contained in $\mathcal{F}$. For any stopping time N for $\mathcal{F}_n$, $\mathcal{F}_N$ is the smallest σ-field containing events $A \in \mathcal{F}$ satisfying $A \cap \{N = n\} \in \mathcal{F}_n$ for all $n < \infty$. Clearly, if X_n is adapted to $\mathcal{F}_n$, then $X_N \mathbf{1}_{\{N<\infty\}}$ is $\mathcal{F}_N$-measurable.

Theorem 3.11 (Strong Markov property)**.** *Let X_n be a time homogeneous Markov chain on $(S, \mathcal{C})$ and N be a stopping time for $\mathcal{F}_n$. Then, for any sequence of uniformly bounded $\mathcal{C}_\infty$-measurable functions, $(f_n)_{n=0}^\infty$,*

$$\mathbb{E}_\mu(f_N(X_N, X_{N+1}, ...)|\mathcal{F}_N) = \varphi(X_N, N) \quad \textit{on } \{N < \infty\},$$

where $\varphi(x, n) := \mathbb{E}_x f_n(X_0, X_1, ...)$.

Remark 3.5. If $f_n = f$ for all $n \geq 0$, then the strong Markov property becomes

$$\mathbb{E}_\mu(f(X_N, X_{N+1}, ...)|\mathcal{F}_N) = \varphi(X_N) \quad \text{on } \{N < \infty\},$$

where $\varphi(x) = \mathbb{E}_x f(X_0, X_1, ...)$.

Proof. Note that, for $A \in \mathcal{F}_N$,

$$\mathbb{E}_\mu(f_N(X_N, X_{N+1}, ...); A \cap \{N < \infty\}) = \sum_{n=0}^\infty \mathbb{E}_\mu(f_n(X_n, X_{n+1}, ...); A \cap \{N = n\}).$$

Since $\varphi(X_N, N)\mathbf{1}_{\{N<\infty\}}$ is $\mathcal{F}_N$-measurable (why?), it remains to show that

$$\mathbb{E}_\mu(f_n(X_n, X_{n+1}, ...); A \cap \{N = n\}) = \mathbb{E}_\mu(\varphi(X_n, n); A \cap \{N = n\}),$$

which is, in fact, given by the Markov property. □

Example 3.6 (Reflection principle). Let $S_0, \xi_1, \xi_2, ...$ be independent random variables and $\xi_1, \xi_2, ...$ are identically distributed with distributions symmetric about 0. Set $S_n = S_0 + \xi_1 + \cdots + \xi_n$. Then, for $a > 0$,

$$\mathbb{P}\left(\max_{1\leq m\leq n} S_m > a\right) \leq 2\mathbb{P}(S_n > a).$$

We prove the inequality by the strong Markov property. Set, for $m \leq n$ and $\omega = (\omega_n)_{n=0}^\infty$,

$$f_m(\omega) = \begin{cases} 1 & \text{if } \omega_{n-m} > a \\ 0 & \text{o.w.} \end{cases}$$

and $N = \inf\{1 \leq m \leq n : S_m > a\}$ with $\inf \emptyset = \infty$. Then,

$$\{N < \infty\} = \{N \leq n\} = \left\{\max_{0\leq m\leq n} S_m > a\right\}.$$

Note that, on $\{N \leq n\}$, $f_N(\omega_N, \omega_{N+1}, ...) = 1$ if $\omega_n > a$ and 0 otherwise. Since S_n is a Markov chain, by the strong Markov property,

$$\mathbb{E}(f_N(S_N, S_{N+1}, ...)|\mathcal{F}_N) = \varphi(S_N, N) \quad \text{on } \{N \leq n\},$$

where $\varphi(y, m) = \mathbb{E}_y f_m(S_0, S_1, ...)$. Observe that, for $y > a$ and $m \leq n$,

$$\begin{aligned} \varphi(y, m) :=& \mathbb{E}_y f_m(S_0, S_1, ...) = \mathbb{P}_y(S_{n-m} > a) \geq \mathbb{P}_y(S_{n-m} \geq y) \\ =& \mathbb{P}_y(S_{n-m} - S_0 \geq 0) \geq 1/2, \end{aligned}$$

where the last inequality uses the symmetry of $\mathbb{P}_0(S_{n-m} - S_0 \in \cdot)$. Thus, on $\{N < \infty\} = \{N \le n\}$, $\varphi(S_N, N) \ge 1/2$. As a consequence, we obtain

$$\begin{aligned}\tfrac{1}{2}\mathbb{P}(N \le n) &\le \mathbb{E}(\varphi(X_N, N); N \le n) = \mathbb{E}(\mathbb{E}(f_N(S_N, S_{N+1}, ...)|\mathcal{F}_N); N \le n)\\ &= \mathbb{P}(S_n > a, N \le n) = \mathbb{P}(S_n > a).\end{aligned}$$

In the following two exercises, we consider Markov chians X_n on countable state spaces S with transition probability p and set

$$\tau_A = \inf\{n \ge 0 : X_n \in A\}, \quad T_A = \inf\{n \ge 1 : X_n \in A\}.$$

Briefly, we write $\tau_y = \tau_{\{y\}}$ and $T_y = T_{\{y\}}$.

Exercise 3.6. [First entrance decomposition] Show that, for $n \ge 1$ and $x, y \in S$,

$$\mathbb{P}_x(X_n = y) = \sum_{m=1}^{n} \mathbb{P}_x(T_y = m)\mathbb{P}_y(X_{n-m} = y)$$

and, for $k \ge 0$,

$$\sum_{m=0}^{n} \mathbb{P}_x(X_m = x) \ge \sum_{m=k}^{n+k} \mathbb{P}_x(X_m = x).$$

Exercise 3.7. Suppose that $S \setminus C$ is a finite set and, for each $x \in S \setminus C$, $\mathbb{P}_x(\tau_C < \infty) > 0$. Show that there exist $N > 0$ and $\epsilon > 0$ such that

$$\mathbb{P}_x(\tau_C > kN) \le (1-\epsilon)^k \quad \forall k \ge 1,\ x \in S \setminus C. \tag{3.8}$$

Use this to conclude that $\mathbb{P}_x(\tau_C < \infty) = 1$ for all $x \notin C$.

Example 3.7. Let S be a countable set and X_n be a time homogeneous Markov chain on S with transition probability p. A function h defined on S is called *harmonic* on $E \subset S$ if

$$h(x) = \sum_{y \in S} h(y)p(x, y) \quad \forall x \in E.$$

Let A, B be disjoint subsets of S such that $(A \cup B)^c$ is finite. By Exercise 3.7, $\mathbb{P}_x(\tau_{A\cup B} < \infty) = 1$ for all $x \in (A \cup B)^c$. We claim that if $\mathbb{P}_x(\tau_{A\cup B} < \infty) > 0$ for all $x \in (A \cup B)^c$, then $\mathbb{P}_x(\tau_A < \tau_B)$ is the unique function h on S, which is harmonic on $(A \cup B)^c$ and satisfies $h = 1$ on A and $h = 0$ on B. First, we shall prove the following statements in order.

(1) The mapping $x \mapsto \mathbb{P}_x(\tau_A < \tau_B)$ is harmonic on $(A \cup B)^c$.
(2) Let h be a bounded function on S. If h is harmonic on $(A \cup B)^c$, then $h(X_{n\wedge\tau_{A\cup B}})$ is a martingale under $\mathbb{P}_x$ for $x \in (A \cup B)^c$.

Proof. Set $f(x) = \mathbb{P}_x(\tau_A < \tau_B)$. Obviously, $f = 1$ on A and $f = 0$ on B. Note that, for $x \notin A \cup B$, $\mathbb{P}_x(\tau_A > 0) = \mathbb{P}_x(\tau_B > 0) = 1$. For $E \subset S$, set

$$f_E(x_0, x_1, ...) = \inf\{k \ge 0 | x_k \in E\}, \quad \forall (x_0, x_1, ...) \in S^\infty.$$

Clearly, f_E is $\mathcal{C}_\infty$-measurable and $\tau_E = f_E(X_0, X_1, ...)$. Note that if $f_E(x_0, x_1, ...) > 0$, then

$$f_E(x_0, x_1, ...) = f_E(x_1, x_2, ...) + 1.$$

By the Markov property, we have that, for $x \notin A \cup B$,

$$\begin{aligned}f(x) &= \mathbb{P}_x(\tau_A < \tau_B) = \mathbb{E}_x(\mathbb{P}_x(f_A(X_1, X_2, ...) < f_B(X_1, X_2, ...)|X_1))\\ &= \mathbb{E}_x f(X_1) = \sum_{y \in S} f(y)p(x, y).\end{aligned}$$

This proves (1).

For (2), let $x \in (A\cup B)^c$. Note that $\tau_{A\cup B}$ is a stopping time for $\mathcal{F}_n$ and, for $D \in \mathcal{F}_{n\wedge\tau_{A\cup B}}$,

$$D\cap\{\tau_{A\cup B}=k\}\in\mathcal{F}_k,\quad \forall k<n,\quad D\cap\{\tau_{A\cup B}\ge n\}\in\mathcal{F}_n.$$

This implies $D\in\mathbb{F}_n$ and, thus, $D\cap\{\tau_{A\cup B}=n\}$ and $D\cap\{\tau_{A\cup B}>n\}$ are in $\mathcal{F}_n$. Write

$$\int_D h(X_{(n+1)\wedge\tau_{A\cup B}})d\mathbb{P}_x=\sum_{k=0}^{n}\int_{D\cap\{\tau_{A\cup B}=k\}}h(X_k)d\mathbb{P}_x+\int_{D\cap\{\tau_{A\cup B}>n\}}h(X_{n+1})d\mathbb{P}_x.$$

Observe that, by the Markov property, $\mathbb{E}_x(h(X_{n+1})|\mathcal{F}_n)=\varphi(X_n)$, where

$$\varphi(y)=\mathbb{E}_y h(X_1)=\sum_{z\in S}h(z)p(y,z)=h(y),\quad \forall y\in(A\cup B)^c.$$

Sice $D\cap\{\tau_{A\cup B}>n\}\in\mathcal{F}_n$, this yields

$$\begin{aligned}\int_{D\cap\{\tau_{A\cup B}>n\}}h(X_{n+1})d\mathbb{P}_x&=\mathbb{E}_x(\mathbb{E}_x(h(X_{n+1})|\mathcal{F}_n);D\cap\{\tau_{A\cup B}>n\})\\&=\mathbb{E}_x(h(X_n);D\cap\{\tau_{A\cup B}>n\})\end{aligned}$$

As a consequence, we obtain

$$\begin{aligned}\int_D h(X_{(n+1)\wedge\tau_{A\cup B}})d\mathbb{P}_x&=\sum_{k=0}^{n-1}\int_{D\cap\{\tau_{A\cup B}=k\}}h(X_k)d\mathbb{P}_x+\int_{D\cap\{\tau_{A\cup B}\ge n\}}h(X_n)d\mathbb{P}_x\\&=\int_D h(X_{n\wedge\tau_{A\cup B}})d\mathbb{P}_x,\end{aligned}$$

which proves (2). □

Back to our example. Let h be a function which is harmonic on $(A\cup B)^c$ and satisfies $h=1$ on A and $h=0$ on B. By (2), we have that $h(x)=\mathbb{E}_x h(X_0)=\mathbb{E}_x h(X_{n\wedge\tau_{A\cup B}})$ for all $n\ge 0$. Since $\mathbb{P}_x(\tau_{A\cap B}<\infty)=1$ for all $x\in S$ and h is bounded, the martingale convergence theorem implies

$$h(x)=\lim_{n\to\infty}\mathbb{E}_x h(X_{n\wedge\tau_{A\cup B}})=\mathbb{E}_x h(X_{\tau_{A\cup B}})=\mathbb{P}_x(\tau_A<\tau_B).$$

3.3. **Asymptotic stationarity.** In this subsection, all Markov chains are assumed to be time homogeneous. As before, let X_n be the coordinate representation process defined on $(S^\infty,\mathcal{C}_\infty)$ and p be a transition probability on $(S,\mathcal{C})$. Assume that, for any probability μ on $(S,\mathcal{C})$, there is a probability $\mathbb{P}_\mu$ on $(S^\infty,\mathcal{C}_\infty)$ such that X_n is a Markov chain on $(S,\mathcal{C})$ with transition probability p and initial distribution μ.

If there is a probability μ on $(S,\mathcal{C})$ such that

$$\lim_{n\to\infty}\mathbb{P}_\mu(X_n\in B)=\pi(B),\quad\forall B\in\mathcal{C},$$

then one can show that π is a finitely additive probability on $(S,\mathcal{C})$ and, for any $\mathcal{C}$-measurable simple function f,

$$\lim_{n\to\infty}\mathbb{E}_\mu f(X_n)=\int_S f(x)\pi(dx).$$

Fix $B\in\mathcal{C}$. Assume in addition that π is a probability. Then, for any bounded $\mathcal{C}$-measurable function g and $\epsilon>0$, we may choose a $\mathcal{C}$-measurable simple function f such that $\sup_x|g(x)-f(x)|<\epsilon$. This implies

$$|\mathbb{E}_\mu g(X_n)-\mathbb{E}_\mu f(X_n)|\le\epsilon,\quad\left|\int_S g(x)\pi(dx)-\int_S f(x)\pi(dx)\right|\le\epsilon.$$

Letting $n \to \infty$ and then $\epsilon \to 0$ yields

$$\lim_{n\to\infty} \mathbb{E}_\mu g(X_n) = \int_S g(x)\pi(dx). \tag{3.9}$$

In particular, for $g = p(\cdot, B)$, the above limit turns out

$$\lim_{n\to\infty} \mathbb{E}_\mu p(X_n, B) = \int_S p(x, B)\pi(dx).$$

Note that, by the Markov property, one has

$$\mathbb{P}_\mu(X_{n+1} \in B) = \int_S p(x, B)\mathbb{P}_\mu(X_n \in dx) = \mathbb{E}_\mu p(X_n, B).$$

As a result, we obtain

$$\pi(B) = \int_S p(x, B)\pi(dx), \quad \forall B \in \mathcal{C}. \tag{3.10}$$

Definition 3.4. Let X_n be a Markov chain on $(S, \mathcal{C})$ with transition probability p. A probability π on $(S, \mathcal{C})$ is said to be a *stationary distribution* if (3.10) holds.

Remark 3.6. If π is a stationary distribution of a Markov chain with transition probability p, then $\mathbb{P}_\pi(X_n = \cdot) = \pi(\cdot)$ for all n.

Lemma 3.12. *Suppose there is a probability μ on $(S, \mathcal{C})$ such that*

$$\lim_{n\to\infty} \mathbb{P}_\mu(X_n \in B) = \pi(B), \quad \forall B \in \mathcal{C}.$$

If π is a probability on $(S, \mathcal{C})$, then it's a stationary distribution.

Definition 3.5. A process X_n taking values on $(S, \mathcal{C})$ is said to be *stationary* if

$$\mathbb{P}((X_0, X_1, ...) \in B) = \mathbb{P}((X_n, X_{n+1}, ...) \in B), \quad \forall B \in \mathcal{C}_\infty,\ n \geq 0.$$

Proposition 3.13. *Let X_n be a Markov chain with transition probability p. If π is a stationary distribution for X_n, then $(X_0, X_1, ...)$ is a stationary process under $\mathbb{P}_\pi$.*

Proof. By the Markov property, we have

$$\mathbb{P}_\pi((X_n, X_{n+1}, ...) \in B | X_n = x) = \mathbb{P}_x((X_0, X_1, ...) \in B), \quad \forall B \in \mathcal{C}_\infty,\ n \geq 0.$$

Since π is a stationary distribution, $\mathbb{P}_\pi(X_n \in \cdot) = \pi(\cdot)$ for all $n \geq 0$. Integrating both sides of the above equation w.r.t. π leads to the desired identity. □

Proposition 3.14. *Let X_n be a Markov chain on $(S, \mathcal{C})$ with transition probability p. Assume that there are probabilities μ, π on $(S, \mathcal{C})$ such that*

$$\lim_{n\to\infty} \mathbb{P}_\mu(X_n \in B) = \pi(B), \quad \forall B \in \mathcal{C}.$$

Then, for any bounded $\mathcal{C}_\infty$-measurable function f,

$$\lim_{n\to\infty} \mathbb{E}_\mu(f(X_n, X_{n+1}, ...)) = \mathbb{E}_\pi f(X_0, X_1, ...).$$

In particular, for $B \in \mathcal{C}_\infty$,

$$\lim_{n\to\infty} \mathbb{P}_\mu((X_n, X_{n+1}, ...) \in B) = \mathbb{P}_\pi((X_0, X_1, ...) \in B).$$

Remark 3.7. The above proposition says that under $\mathbb{P}_\mu$, the process X_n is asymptotically stationary.

Proof of Proposition 3.14. In a similar reasoning for (3.10), it suffices to consider the case that $f = \mathbf{1}_B$ with $B \in \mathcal{C}_\infty$. Fix $B \in \mathcal{C}_\infty$ and set $\varphi(x) = \mathbb{P}_x((X_0, X_1, ...) \in B)$. By the Markov property, $\mathbb{P}_\mu((X_n, X_{n+1}, ...) \in B|\mathcal{F}_n) = \varphi(X_n)$. This implies

$$\begin{aligned}\mathbb{P}_\mu((X_n, X_{n+1}, ...) \in B) &= \mathbb{E}_\mu(\mathbb{P}_\mu((X_n, X_{n+1}, ...) \in B|\mathcal{F}_n)) \\ &= \mathbb{E}_\mu\varphi(X_n) = \int_S \varphi(y)\mathbb{P}_\mu(X_n \in dy).\end{aligned}$$

As a result of (3.9), we obtain

$$\lim_{n\to\infty} \int_S \varphi(y)\mathbb{P}_\mu(X_n \in dy) = \int_S \mathbb{P}_y((X_0, X_1, ...) \in B)\pi(dy) = \mathbb{P}_\pi((X_0, X_1, ...) \in B).$$

□

One natural question arises. How many stationary distributions a Markov chain may possess?

Lemma 3.15. *Let* Π *be the set of all stationary distributions for a Markov chain on* $(S, \mathcal{C})$. *If* Π *is nonempty, then* Π *forms a convex set in the space of all probability measures on* $(S, \mathcal{C})$.

Proof. Suppose p is the transition probability of the Markov chain. Let $\mu, \nu \in \Pi$ and set, for $a \in (0, 1)$, $\pi_a = a\mu + (1 - a)\nu$. Then, for $B \in \mathcal{C}$,

$$\begin{aligned}\int_S p(x, B)\pi_a(dx) &= a \int_S p(x, B)\mu(dx) + (1 - a) \int_S p(x, B)\nu(dx) \\ &= a\pi(B) + (1 - a)\pi(B) = \pi(B).\end{aligned}$$

□

3.4. **Recurrence and transience.** In this section, all Markov chains are assumed to have countable state spaces. Let $T_y^{(0)} = 0$ and set, for $k \geq 1$,

$$T_y^{(k)} = \min\{n > T_y^{(k-1)} : X_n = y\}.$$

T_y^k is the time of the k-th return to state y. Briefly, we let $T_y = T_y^{(1)}$. Set $\rho_{xy} = \mathbb{P}_x(T_y < \infty)$ for $x, y \in S$.

Definition 3.6. A state y is said to be *recurrent* if $\rho_{yy} = 1$ and *transient* if $\rho_{yy} < 1$.

Theorem 3.16. *For all* $x, y \in S$ *and* $k \geq 1$, $\mathbb{P}_x(T_y^{(k)} < \infty) = \rho_{xy}\rho_{yy}^{(k-1)}$, *where* $0^0 := 1$.

Proof. The case $k = 1$ is clear. Let $k > 1$ and set $N = T_y^{(k-1)}$,

$$f(\omega) = \begin{cases} 1 & \text{if } \omega_n = y \text{ for some } n \geq 1 \\ 0 & \text{o.w.} \end{cases}$$

Note that, on $\{N < \infty\}$, $f(X_N, X_{N+1}, ...) = 1$ if and only if $T_y^{(k)} < \infty$. Since N is a stopping time for X_n, by the strong Markov property, one has $X_N = y$ and

$$\mathbb{E}_x(f(X_N, X_{N+1}, ...)|\mathcal{F}_N) = \varphi(X_N) = \varphi(y) = \rho_{yy} \quad \text{on } \{N < \infty\},$$

where $\varphi(z) = \mathbb{E}_z(f(X_0, X_1, ...))$. Putting all above together, we have

$$\mathbb{P}_x(T_y^{(k)} < \infty) = \mathbb{E}_x(f(X_N, X_{N+1}, ...); N < \infty) = \rho_{yy}\mathbb{P}_x(T_y^{(k-1)} < \infty).$$

The desired identity is then proved by induction. □

Corollary 3.17. y *is recurrent if and only if* $\mathbb{P}_y(X_n = y \text{ i.o.}) = 1$; y *is transient if and only if* $\mathbb{P}_y(X_n = y \text{ i.o.}) = 0$.

Proof. Note that $\{X_n = y \text{ i.o.}\} = \{T_y^{(k)} < \infty, \forall k > 0\}$. By Theorem 3.16, one has $\mathbb{P}_y(T_y^{(k)} < \infty) = \rho_{yy}^k$ for all $k \geq 0$. The desired property is obvious from these observations. □

Theorem 3.18. *y is recurrent if and only if $\sum_{n=1}^{\infty} p^n(y,y) = \infty$.*

Proof. Let $N_y = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n = y\}}$. By definition, N_y is the number of visits to y at positive times and

$$\{N_y \geq k\} = \{T_y^{(k)} < \infty\}.$$

As a result of this equation and Theorem 3.16, one has

$$\mathbb{E}_x N_y = \sum_{k=1}^{\infty} \mathbb{P}_x(N_y \geq k) = \sum_{k=1}^{\infty} \mathbb{P}_x(T_y^{(k)} < \infty) = \sum_{k=1}^{\infty} \rho_{xy}\rho_{yy}^{k-1}.$$

Note that $\mathbb{E}_y N_y = \sum_{n=1}^{\infty} p^n(y,y)$. If y is transient, that is, $\rho_{yy} < 1$, then $\mathbb{E}_x N_y = \rho_{xy}/(1 - \rho_{yy}) < \infty$ for all $x \in S$. If y is recurrent, then $\mathbb{E}_y N_y = \infty$. □

Remark 3.8. If y is transient, then

$$\mathbb{E}_x N_y = \frac{\rho_{xy}}{1 - \rho_{yy}} \quad \forall x \in S.$$

Remark 3.9. We summarize the above discussion as follows. The following are equivalences of recurrence,

(1) y is recurrent;
(2) $\mathbb{P}_y(T_y < \infty) = \mathbb{P}_y(X_n = y \text{ for some } n > 0) = 1$;
(3) $\mathbb{P}_y(X_n = y \text{ i.o.}) = \mathbb{P}_y(T_y^{(k)} < \infty, \forall k > 0) = 1$;
(4) $\sum_n \mathbb{P}_y(X_n = y) = \mathbb{E}_y N_y = \infty$.

and the following are equivalences of transience,

(5) y is transient;
(6) $\mathbb{P}_y(T_y < \infty) = \mathbb{P}_y(X_n = y \text{ for some } n > 0) < 1$;
(7) $\mathbb{P}_y(X_n = y \text{ i.o.}) = \mathbb{P}_y(T_y^{(k)} < \infty, \forall k > 0) = 0$;
(8) $\sum_n \mathbb{P}_y(X_n = y) = \mathbb{E}_y N_y < \infty$.

From the view point of generating functions, let

$$u_x(s) = \sum_{n=0}^{\infty} p^n(x,x)s^n, \quad f_x(s) = \sum_{n=1}^{\infty} \mathbb{P}_x(T_x = n)s^n.$$

Exercise 3.8. Show that $u_x(s) = 1/(1 - f_x(s))$ for all $s \in [0,1]$ and $x \in S$, where $1/0 := \infty$.

Exercise 3.9. Prove that if x is transient, then

$$\mathbb{P}_x(T_x = \infty) = \left(\sum_{n=0}^{\infty} p^n(x,x)\right)^{-1}.$$

Exercise 3.10. Let X_n be a simple random walk on $\mathbb{Z}$ with transition probability

$$p(i, i+1) = p, \quad p(i, i-1) = q = 1 - p \quad \forall i \in \mathbb{Z}.$$

Use Taylor's expansion for $(1-x)^{-1/2}$ to show that, for all $i \in \mathbb{Z}$,

$$u_i(s) = (1 - 4pqs^2)^{-1/2}, \quad f_i(s) = 1 - (1 - 4pqs^2)^{1/2}, \quad \forall s \in [0,1],$$

and determine ρ_{00}.

Assume that y is recurrent and set $R_k = T_y^{(k)}$ for $k \geq 0$ and $r_k = R_k - R_{k-1}$. Here, the sequence R_k is called the renewal time of state y.

Theorem 3.19. *If y is recurrent, then, under $\mathbb{P}_y$, the sequence $(r_k, X_{R_{k-1}}, ..., X_{R_k-1})$ with $k \ge 1$ are i.i.d. and, for $n \ge 1$, $x_0, ..., x_n \in S$,*

$$\begin{aligned}\mathbb{P}_y((r_k, X_{R_{k-1}}, ..., X_{R_k-1}) = (n, x_0, ..., x_{n-1})) \\ = \delta_y(x_0)\left(\prod_{i=0}^{n-2} p(x_i, x_{i+1})(1-\delta_y(x_{i+1}))\right) p(x_{n-1}, y).\end{aligned}$$

Proof. Let $f_1(x) = \delta_y(x)$ and, for $n > 1$, let f_n be a function on S^n defined by

$$f_n(y_0, y_1, ..., y_{n-1}) = \delta_y(y_0)\left(\prod_{i=0}^{n-2} p(y_i, y_{i+1})(1-\delta_y(y_{i+1}))\right) p(y_{n-1}, y), \quad \forall y_0, ..., y_{n-1} \in S^n.$$

Since the range of $(r_k, X_{R_{k-1}}, ..., X_{R_k-1})$ is countable, it suffices to show that for $m \ge 1$, $a_1, ..., a_m \in \mathbb{N}$ and $x_{i,j} \in S$ with $0 \le j < a_i$ and $1 \le i \le m$,

$$\mathbb{P}_y(X_{b_i+j} = x_{i,j}, r_i = a_i, \forall 0 \le j < a_i, \forall 1 \le i \le m) = \prod_{i=1}^m f_{a_i}(x_{i,0}, ..., x_{i,a_i-1})$$

where $b_1 = 0$ and $b_i = a_1 + \cdots + a_{i-1}$. It is clear that the above identity holds if $x_{i,0} \ne y$ for some $1 \le i \le m$ or $x_{i,j} = y$ for some $1 \le i \le m$ and some $0 < j < a_i$. Assuming $x_{i,0} = y$ for all $1 \le i \le m$ and $x_{i,j} \ne y$ for all $1 \le i \le m$ and all $0 \le j < a_i$, one has

$$\begin{aligned}&\{X_{b_i+j} = x_{i,j}, r_i = a_i, \forall 0 \le j < a_i, \forall 1 \le i \le m\} \\ =&\{X_{b_i+j} = x_{i,j}, \forall 0 \le j < a_i, \forall 1 \le i \le m, X_{b_{m+1}} = y\}.\end{aligned}$$

By the Markov property, we obtain

$$\begin{aligned}&\mathbb{P}_y(X_{b_i+j} = x_{i,j}, \forall 0 \le j < a_i, \forall 1 \le i \le m, X_{b_{m+1}} = y) \\ =&\mathbb{P}_y(X_{b_m+j} = x_{m,j}, \forall 0 < j < a_i, X_{b_{m+1}} = y | X_{b_i+j} = x_{i,j}, \forall 0 \le j < a_i, \forall 1 \le i < m, X_{b_m} = y) \\ &\quad \times \mathbb{P}_y(X_{b_i+j} = x_{i,j}, \forall 0 \le j < a_i, \forall 1 \le i < m, X_{b_m} = y) \\ =&f_{a_m}(x_{m,0}, ..., x_{m,a_m-1})\mathbb{P}_y(X_{b_i+j} = x_{i,j}, \forall 0 \le j < a_i, \forall 1 \le i < m, X_{b_m} = y).\end{aligned}$$

The desired identity is then given by induction. ☐

3.5. **Group property of states.**

Theorem 3.20. *If x is recurrent and $\rho_{xy} > 0$, then y is recurrent and $\rho_{xy} = \rho_{yx} = 1$.*

Proof. Define $K = \inf\{k : p^k(x,y) > 0\}$. Since $\rho_{xy} > 0$, $K < \infty$. Let $y_1, ..., y_{K-1}$ be states in S such that

$$\prod_{i=1}^K p(y_{i-1}, y_i) > 0,$$

where $y_0 = x$ and $y_K = y$. It is easy to see from the definition of K that $y_i \notin \{x, y\}$ for all $1 \le i < K$. By the Markov property, we obtain

$$\mathbb{P}_x(T_x = \infty) \ge \mathbb{P}_x(X_i = y_i, \forall 1 \le i \le K, X_i \ne x, \forall i > K) = \prod_{i=1}^K p(y_{i-1}, y_i)(1-\rho_{yx}).$$

Thus, the recurrence of x implies $\rho_{yx} = 1$. Since $\rho_{yx} > 0$, one may choose $L > 0$ such that $p^L(y,x) > 0$. Putting all above together and then applying Theorem 3.18 gives

$$\sum_{n=1}^\infty p^n(y,y) \ge \sum_{n=1}^\infty p^{n+K+L}(y,y) \ge \sum_{n=1}^\infty p^L(y,x)p^n(x,x)p^K(x,y) = \infty.$$

This means that y is recurrent. The fact $\rho_{xy} = 1$ is immediate from this theorem with the exchange of x and y. □

Remark 3.10. If $\rho_{xy} > 0$ but $\rho_{yx} = 0$, then x must be transient.

Exercise 3.11. Prove by using the strong Markov property that $\rho_{xz} \geq \rho_{xy}\rho_{yz}$.

Corollary 3.21. *If x is transient and $\rho_{yx} > 0$, then y is transient.*

Definition 3.7. A set $C \subset S$ is said to be *closed* if $x \in C$ and $\rho_{xy} > 0$ implies $y \in C$. A set D is said to be *irreducible* if $\rho_{xy} > 0$ for all $x, y \in D$.

Remark 3.11. $\rho_{xy} > 0$ if and only if there exists $K > 0$ such that $p^K(x, y) > 0$.

Remark 3.12. C is closed if and only if $\mathbb{P}_x(X_n \in C) = 1$ for all $n \geq 1$ and $x \in C$.

Corollary 3.22. *Let R be the set of all recurrent states. Then, R is closed. Moreover, $R = \bigcup_i R_i$, where R_i's are closed and irreducible.*

Proof. The closedness of R is obvious from Theorem 3.20. To see a decomposition of R, let R_x be a subset of R defined by

$$R_x = \{y \in R : \rho_{xy} > 0\}.$$

As a result of Theorem 3.20 and Exercise 3.11, R_x is closed and it remains to show that $R_x = R_y$ for all $y \in R_x$. Let $z \in R_y$. Then, $\rho_{yz} = 1$. By Exercise 3.11, this implies $\rho_{xz} \geq \rho_{xy}\rho_{yz} = 1$. Hence, we have $R_y \subset R_x$. Note that $\{x, y\} \subset R_x \cap R_y$. Since $x \in R_y$, we have $R_x \subset R_y$. □

Proposition 3.23. *Let C be finite and closed. Then, C must contain a recurrent state. Moreover, if C is irreducible, then all states in C are recurrent.*

Proof. For the first part, recall the notation $N_y = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n = y\}}$. By Fubini's theorem, we have, for $x \in C$,

$$\sum_{y \in C} \mathbb{E}_x N_y = \sum_{y \in C} \sum_{n=1}^{\infty} p^n(x, y) = \sum_{n=1}^{\infty} \sum_{y \in C} p^n(x, y) = \sum_{n=1}^{\infty} 1 = \infty.$$

Since C is finite, $\mathbb{E}_x N_y = \infty$ for some $y \in C$ and $x \in C$. This implies that y must be recurrent, otherwise, by Remark 3.8,

$$\mathbb{E}_x N_y = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty,$$

which is a contradiction. The second part is clear from the irreducibility. □

Definition 3.8. A recurrent state x is called *positive recurrent* if $\mathbb{E}_x T_x < \infty$ and *null recurrent* if $\mathbb{E}_x T_x = \infty$.

Theorem 3.24. *If x is positive recurrent and $\rho_{xy} > 0$, then y is positive recurrent and $\mathbb{E}_x T_y < \infty$.*

Proof. The case $y = x$ is obvious and we assume in the following that $y \neq x$. Recall that, in Theorem 3.19, if x is recurrent with recurrent time R_k and $r_k = R_k - R_{k-1}$, then, under $\mathbb{P}_x$, $(r_k, X_{R_{k-1}}, ..., X_{R_k - 1})$ with $k \geq 1$ are i.i.d.

For $k \geq 1$, set

$$V_k = \begin{cases} 1 & \text{if } X_m = y \text{ for some } R_{k-1} < m < R_k \\ 0 & \text{o.w.} \end{cases}$$

and

$$U_k = \sum_{i=1}^{k} V_i, \quad V = \inf\{k \geq 1 : U_k = 2\}.$$

Note that $\mathbb{P}_x(V < \infty) = 1$ and, on $\{V = k\}$, $T_y^{(2)} \leq R_k$. This implies

$$\begin{aligned}\mathbb{E}_x T_y^{(2)} &= \sum_{k=1}^{\infty} \mathbb{E}(T_y^{(2)}; V = k) \leq \sum_{k=1}^{\infty} \mathbb{E}(R_k; V = k) \\ &= \sum_{k=1}^{\infty} \sum_{l=1}^{k} \mathbb{E}(r_l; V = k) = \sum_{l=1}^{\infty} \mathbb{E}(r_l; V \geq l)\end{aligned}$$

By Theorem 3.19, since $\{V \geq l\} = \{V \leq l-1\}^c \in \mathcal{F}_{R_{l-1}-1}$, $\{V \geq l\}$ and r_l are independent. This implies

$$\mathbb{E}_x T_y^{(2)} \leq \sum_{l=1}^{\infty} \mathbb{E}_x r_l \mathbb{P}_x(V \geq l) = \mathbb{E}_x T_x \mathbb{E}_x V.$$

As $V_1, V_2, \ldots$ are i.i.d., we have, for $l \geq 1$

$$\begin{aligned}\mathbb{P}_x(V > l) &= \sum_{i=1}^{l} \mathbb{P}_x(V_i = 1, V_j = 0, \forall j \neq i, j \leq l) + \mathbb{P}_x(V_i = 0, \forall 1 \leq i \leq l) \\ &= l\mathbb{P}_x(V_1 = 1)\mathbb{P}_x(V_1 = 0)^{l-1} + \mathbb{P}_x(V_1 = 0)^l,\end{aligned}$$

which yields

$$\mathbb{E}_x V \leq \frac{2}{\mathbb{P}_x(V_1 = 1)}.$$

By Theorem 3.18, as y is recurrent, one may choose $K > 0$ such that

$$\mathbb{P}_x(X_i \neq x, \forall 1 \leq i < K, X_K = y) > 0.$$

A a consequence, $\mathbb{P}_x(V_1 = 1) > 0$ and then $\mathbb{E}_x V < \infty$. Since x is positive recurrent, $\mathbb{E}_x T_y^{(2)} < \infty$, which leads to $\mathbb{E}_x T_y < \infty$ and

$$\infty > \mathbb{E}_x(T_y^{(2)} - T_y) = \mathbb{E}_x(\mathbb{E}_x(T_y^{(2)} - T_y | \mathcal{F}_{T_y})) = \mathbb{E}_y T_y.$$

□

Corollary 3.25. *Let R be the set of recurrent states. Then, for any closed and irreducible subset of R, either all states are positive recurrent or all states are null recurrent.*

Remark 3.13. We will prove in the next subsection that all states in a finite and closed set $C \subset S$ are positive recurrent.

3.6. **Stationary distributions.**

Definition 3.9. A measure π on S is said to be a *stationary measure* if

$$\sum_{y \in S} \pi(y) p(y, x) = \pi(x), \quad \forall x \in S.$$

If π is a probability, then we call it a *stationary distribution.*

Example 3.8 (Random walks on $\mathbb{Z}$). Let p be a transition probability on $S = \mathbb{Z}$ given by

$$p(i, i+1) = p, \quad p(i, i-1) = q = 1 - p, \quad \forall i \in \mathbb{Z},$$

with $p \in (0,1)$. Set $\pi(i) = (p/q)^i$. Then, π is a stationary distribution since

$$\sum_{i\in\mathbb{Z}} \pi(i)p(i,j) = \pi(j-1)p(j-1,j) + \pi(j+1)p(j+1,j)$$
$$= (p/q)^{j-1}p + (p/q)^{j+1}q = (p/q)^j(q+p) = (p/q)^j.$$

Example 3.9. Consider the Ehrenfest chain on $\{0,1,...,r\}$, that is, the transition probability p is defined by

$$p(i,i+1) = 1 - \frac{i}{r}, \quad \forall 0 \le i < r; \quad p(i,i-1) = \frac{i}{r}, \quad \forall 0 < i \le r.$$

Set $\pi(i) = \binom{r}{i}2^{-r}$. Then, π is a stationary distribution.

Example 3.10 (Birth and Death chains). A birth and death chain is a Markov chain on $S = \{0,1,2,...\}$ with transition probability

$$p(i,i+1) = p_i, \quad p(i,i) = r_i, \quad p(i,i-1) = q_i,$$

where $p_i + q_i + r_i = 1$ and $q_0 = 0$. Assume that $q_i > 0$ for all $i > 0$ and set

$$\pi(i) = \prod_{j=1}^{i} \frac{p_{j-1}}{q_j}.$$

Then, π is a stationary measure.

Exercise 3.12. Let π be a positive stationary distribution for a Markov chain X_n with transition probability p and set

$$q(x,y) = \frac{\pi(y)p(y,x)}{\pi(x)}, \quad \forall x,y \in S.$$

Let $Y_m = X_{n-m}$. Show that, if X_0 has distribution π, then $Y_0,...,Y_n$ forms a Markov chain with transition probability q. Here, Y_m is called the *reverse* or *dual* Markov chain for X_n and q is called the *dual transition probability.*

Exercise 3.13. Let π and q be as in Exercise 3.12. If $q = p$, then π is called a *reversible distribution* for p. Show that if π is reversible for p, then π is stationary.

Theorem 3.26. *Let x be a recurrent state. Then, the following map*

$$y \mapsto \pi_x(y) = \mathbb{E}_x\left(\sum_{i=0}^{T_x-1} \mathbf{1}_{\{X_i=y\}}\right) = \sum_{i=0}^{\infty} \mathbb{P}_x(X_i = y, T_x > i)$$

defines a stationary measure for p.

Remark 3.14. Note that $\pi_x(x) = 1$ and $\pi_x(y) < \infty$. Let $y \in S$ and assume that $\rho_{xy} > 0$. By Theorem 3.26, we have

$$1 = \pi_x(x) = \sum_{z\in S} \pi_x(z)p^n(z,x) \ge \pi_x(y)p^n(y,x), \quad \forall n > 0.$$

By Theorem 3.20, since x is recurrent and $\rho_{xy} > 0$, one has $\rho_{yx} = 1$. This implies $\pi_x(y) < \infty$. As a result, if p is irreducible, then $\pi_x(y) < \infty$ for all $y \in S$.

Proof of Theorem 3.26. We prove this theorem using *cycle trick.* Note that, for $y \neq x$,

$$\sum_{i=0}^{T_x-1} \mathbf{1}_{\{X_i=y\}} = \sum_{i=1}^{T_x} \mathbf{1}_{\{X_i=y\}} = \sum_{i=1}^{T_x-1} \mathbf{1}_{\{X_i=y\}}.$$

By Fubini's theorem, it is easy to see that

$$\sum_{y\in S}\pi_x(y)p(y,z)=\sum_{i=0}^{\infty}\sum_{y\in S}\mathbb{P}_x(X_i=y,T_x>i)p(y,z).$$

By the Markov property, if $z\neq x$, then

$$\begin{aligned}\sum_{y\in S}\mathbb{P}_x(X_i=y,T_x>i)p(y,z)&=\sum_{y\in S}\mathbb{P}_x(X_i=y,T_x>i,X_{i+1}=z)\\=&\sum_{y\in S}\mathbb{P}_x(X_i=y,X_{i+1}=z,T_x>i+1)=\mathbb{P}_x(X_{i+1}=z,T_x>i+1).\end{aligned}$$

This implies

$$\sum_{y\in S}\pi_x(y)p(y,z)=\sum_{i=0}^{\infty}\mathbb{P}_x(X_{i+1}=z,T_x>i+1)=\pi_x(z).$$

If $z=x$, we have

$$\begin{aligned}\sum_{y\in S}\mathbb{P}_x(X_i=y,T_x>i)p(y,x)&=\sum_{y\in S}\mathbb{P}_x(X_i=y,T_x>i,X_{i+1}=x)\\=&\sum_{y\in S}\mathbb{P}_x(X_i=y,T_x=i+1)=\mathbb{P}_x(T_x=i+1).\end{aligned}$$

Since x is recurrent, this implies

$$\sum_{y\in S}\pi_x(y)p(y,x)=\sum_{i=0}^{\infty}\mathbb{P}_x(T_x=i+1)=1=\pi_x(x).$$

□

Remark 3.15. If x is transient, then

$$\sum_{y\in S}\pi_x(y)p(y,z)=\pi_x(z),\quad \forall z\neq x.$$

But, for $z=x$, we have

$$\sum_{y\in S}\pi_x(y)p(y,x)<1=\pi_x(x).$$

Exercise 3.14. Recall the renewal chain with transition probability

$$p(0,j)=f_{j+1},\quad p(j+1,j)=1,\quad \forall j\geq 0.$$

Use Theorem 3.26 to show that $\pi(j)=\sum_{k\geq j}f_{k+1}$ is a stationary measure for p.

Theorem 3.27. *If p is irreducible and all states are recurrent, then the stationary measure is unique up to a multiple constant.*

Proof. Let π be a stationary measure. Since p is irreducible, it is clear that $\pi(y)<\infty$ for all $y\in S$. Fix $a\in S$. Note that

$$\pi(z)=\sum_{y\in S}\pi(y)p(y,z)=\pi(a)p(a,z)+\sum_{y\neq a}\pi(y)p(y,z).$$

Applying the second equation to the last summation yields

$$\pi(z)=\pi(a)p(a,z)+\pi(a)\sum_{y\neq a}p(a,y)p(y,z)+\sum_{y\neq a,x\neq a}\pi(x)p(x,y)p(y,z).$$

Inductively, one has

$$\begin{aligned}\pi(z) &= \sum_{m=1}^{n} \pi(a) \sum_{x_1,\dots,x_{m-1}\neq a, x_m=z} p(a,x_1)p(x_1,x_2)\cdots p(x_{m-1},x_m) \\ &\quad + \sum_{x_0,\dots,x_{n-1}\neq a, x_n=z} \pi(x_0)p(x_0,x_1)\cdots p(x_{n-1},x_n) \\ &\geq \pi(a)\sum_{m=1}^{n} \mathbb{P}_a(X_i \neq a, 1\leq i<m, X_m=z), \quad \forall n\geq 1.\end{aligned}$$

Since z is recurrent, this implies

$$\begin{aligned}\pi(z) &\geq \pi(a)\sum_{m=1}^{\infty} \mathbb{P}_a(X_i \neq a, 1\leq i<m, X_m=z) \\ &= \pi(a)\sum_{m=1}^{\infty} \mathbb{P}_a(T_a \geq m, X_m=z) = \pi(a)\pi_a(z),\end{aligned}$$

where π_a is the measure in Theorem 3.26. As a result, we obtain

$$\pi(a) = \sum_{z\in S}\pi(z)p^n(z,a) \geq \pi(a)\sum_{z\in S}\pi_a(z)p^n(z,a) = \pi(a)\pi_a(a) = \pi(a).$$

This implies that if $p^n(z,a)>0$, then $\pi(z)=\pi(a)\pi_a(z)$. Since p is irreducible, this must be true for all $z\in S$. □

As a consequence of Theorem 3.26 and 3.27, we have

Theorem 3.28. *If p has a stationary distribution π, then all states $y\in S$ satisfying $\pi(y)>0$ are positive recurrent. In particular, if p is assumed further irreducible, then $\pi(x)=1/\mathbb{E}_xT_x$ for all $x\in S$.*

Proof. Note that if $\pi(y)>0$, then

$$\sum_{x\in S}\pi(x)\sum_{n=1}^{\infty}p^n(x,y) = \sum_{n=1}^{\infty}\pi(y) = \infty.$$

This implies y has to be recurrent otherwise,

$$\sum_{x\in S}\pi(x)\sum_{n=1}^{\infty}p^n(x,y) = \sum_{x\in S}\pi(x)\frac{\rho_{xy}}{1-\rho_{yy}} \leq \frac{1}{1-\rho_{yy}} < \infty.$$

Next, we turn to prove the second part and assume that p is irreducible. By the irreducibility of p, $\pi(x)>0$ for all $x\in S$. By Theorems 3.26 and 3.27, x is recurrent and one may select a constant c_x such that $\pi=c_x\pi_x$. This implies

$$1 = c_x\sum_{y\in S}\pi_x(y) = c_x\sum_{y\in S}\sum_{i=0}^{\infty}\mathbb{P}_x(X_i=y, T_x>i) = c_x\sum_{i=0}^{\infty}\mathbb{P}_x(T_x>i) = c_x\mathbb{E}_xT_x,$$

which leads to $\pi(x)=c_x\pi_x(x)=c_x=1/\mathbb{E}_xT_x$ for all $x\in S$. Hence, $\mathbb{E}_xT_x<\infty$ or equivalently x is positive recurrent.

Back to the first part, let $\pi(y)>0$ and let C be the closed and irreducible set containing y. Let p_C be the submatrix of p indexed by C. It is an easily exercise to show that p_C is an irreducible transition probability on C and has $\pi|_C/\pi(C)$ as a (in fact, the) stationary distribution. Note that if X_n, Y_n be the Markov chain on S, C with transition probabilities p, p_C and T_x, T_x^C be the first return times of x in X_n, Y_n, then, given $X_0=Y_0=x$ with

$x \in C$, T_x and T_x^C share the same distribution and, thus, $\mathbb{E}_x T_x = \mathbb{E}_x^C T_x^C$ for all $x \in C$. As an immediate result of the second part, x is positive recurrent for all $x \in C$. □

Corollary 3.29. *Assume that p is irreducible and all states are recurrent. Then, all states are positive recurrent if and only if there is a stationary distribution π. In particular, $\pi > 0$.*

Corollary 3.30. *If S is finite and p is irreducible, then all states are positive recurrent and p has exactly one stationary distribution. Conversely, if p has null recurrent states, then $|S| = \infty$.*

Remark 3.16. If p is irreducible, then

$$\frac{1}{\mathbb{E}_x T_x} = \sum_{y \in S} \frac{p(y,x)}{\mathbb{E}_y T_y}, \quad \forall y \in S$$

and

$$\frac{\pi_x(y)}{\mathbb{E}_x T_x} = \frac{1}{\mathbb{E}_y T_y}, \quad \forall x, y \in S.$$

3.7. **Asymptotic behavior.** In this section, we will consider the long-term behavior of Markov chains on countable state spaces. Note that if y is transient, then

$$\sum_{n=1}^{\infty} p^n(x,y) = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty, \quad \forall x \in S.$$

This implies $p^n(x,y) \to 0$ as $n \to \infty$. When y is recurrent, we set

$$N_{n,y} = \sum_{m=1}^{n} \mathbf{1}_{\{X_n = y\}}.$$

Clearly, $N_{n,y} \to N_y := \sum_{m=1}^{\infty} \mathbf{1}_{\{X_m = y\}}$ as $n \to \infty$.

Theorem 3.31. *Assume that y is recurrent. Then, for any $x \in S$,*

$$\frac{N_{n,y}}{n} \to \frac{1}{\mathbb{E}_y T_y} \mathbf{1}_{\{T_y < \infty\}}, \quad \mathbb{P}_x\text{-a.s.}$$

Proof. Recall the following notations: Let $R_0 = 0$ and

$$R_k = T_y^{(k)}, \quad r_k = R_k - R_{k-1}, \quad \forall k \geq 1.$$

By Theorem 3.19, $r_1, r_2, ..$ are i.i.d. under $\mathbb{P}_y$. By the strong law of large numbers, one has

$$\frac{R_k}{k} = \frac{r_1 + r_2 + \cdots r_k}{k} \to \mathbb{E}_y T_y, \quad \mathbb{P}_y\text{-a.s.}$$

Since y is recurrent, $\mathbb{P}_y(N_y = \infty) = 1$. This implies $N_{n,y} \to \infty$ $\mathbb{P}_y$-a.s.. Note that $R_{N_{n,y}} \leq n < R_{N_{n,y}+1}$ and write

$$\frac{R_{N_{n,y}}}{N_{n,y}} \leq \frac{n}{N_{n,y}} < \frac{R_{N_{n,y}+1}}{N_{n,y}+1} \times \frac{N_{n,y}+1}{N_{n,y}}.$$

Passing n to the infinity yields that $N_{n,y}/n \to 1/\mathbb{E}_y T_y$ $\mathbb{P}_y$-a.s..

Next, assume $x \neq y$. Clearly, $N_{n,y} = 0$ on $\{T_y = \infty\}$ for all $n \geq 1$. This implies $N_{n,y}/n \to 0$. On $\{T_y < \infty\}$, one may use the strong Markov property to conclude that $r_1, r_2, ...$ are independent and $r_2, r_3, ...$ are identically distributed. As a result, one has

$$\frac{R_k}{k} = \frac{r_1}{k} + \frac{r_2 + \cdots + r_k}{k-1} \times \frac{k-1}{k} \to \mathbb{E}_y T_y, \quad \text{on } \{T_y < \infty\},$$

$\mathbb{P}_x$-a.s.. □

Note that $N_{n,y}/n$ is uniform bounded by 1. By the Lebesgue dominated convergence theorem, if y is recurrent, then

$$\frac{1}{n}\sum_{m=1}^{n} p^m(x,y) = \frac{\mathbb{E}_x N_{n,y}}{n} \to \frac{\rho_{xy}}{\mathbb{E}_y T_y} \quad \text{as } n\to\infty. \tag{3.11}$$

The above convergence also holds for transient states. It is worthwhile to note that such a convergence does not implies the convergence of $p^n(x,y)$. For a counterexample, let $S=\{0,1\}$ and p is a transition probability defined by

$$p(0,1)=p(1,0)=1,\quad p(0,0)=p(1,1)=0.$$

In this case, $p_{2n}=I$ and $p^{2n+1}=p$ and, thus, $p^n(x,y)$ never converges. The reason here is due to the periodicity of the appearance of states.

Definition 3.10. The *period* d of a state x is defined to be the greatest common divisor of $\{n\ge 1 : p^n(x,x)>0\}$. If $d=1$, x is also called *aperiodic.*

Lemma 3.32. *If x,y are states satisfying $\rho_{xy}\rho_{yx}>0$, then the periods of x and y are the same.*

Proof. Let $K>0$ and $L>0$ be such that $p^K(x,y)p^L(y,x)>0$ and d_x,d_y be the periods of x,y. Clearly, $d_x|(K+L)$. Note that if $p^n(y,y)>0$, then $p^{K+n+L}(x,x)>0$. This implies $d_x|n$ and, hence, $d_x|d_y$. Exchanging x and y in the above discussion yields $d_x=d_y$. ☐

Lemma 3.33. *If x is of period 1, then there is $n_0>0$ such that $p^n(x,x)>0$ for all $n\ge n_0$.*

Proof. Set $I=\{n\ge 1 : p^n(x,x)>0\}$ and $M=\min\{n-m | m,n\in I, m<n\}$. Our first step is to show that $M=1$ or equivalently there exists N such that $N,N+1\in I$. Assume the inverse, that is $M>1$. Let $n_1\in I$ be such that $n_1+M\in I$. Since the greatest common divisor of I is 1, we may choose $n_2\in I$ such that $M\nmid n_2$. Write $n_2=mM+r$ with $m\ge 0$ and $0<r<M$. Since I is closed under addition, and thus closed under multiplication,

$$(m+1)(n_1+M)\in I,\quad (m+1)n_1+n_2\in I.$$

Clearly, these two terms are not equal and subtracting one from the other yields

$$M\le (m+1)M-n_2=M-r<M,$$

a contradiction. Thus, $M=1$.

Now assume that $N\in I$ be such that $N+1\in I$ and let $n_0=N^2$. Obviously, $n_0\in I$ and for $n>n_0$, we may write $n-n_0=mN+r$ with $0<r<N$. This implies

$$n=n_0+mN+r=N^2+mN+r=N(N+m-r)+(N+1)r\in I.$$

☐

Theorem 3.34. *Suppose p is irreducible with stationary distribution π. If all states are aperiodic, then $p^n(x,y)\to\pi(y)$ as $n\to\infty$ for all $x,y\in S$.*

Proof. We prove this theorem by the method of *coupling.* Let q be a transition probability on S^2 defined by

$$q((x_1,y_1),(x_2,y_2))=p(x_1,x_2)p(y_1,y_2).$$

Step 1: q is irreducible and aperiodic. By the irreducibility of p, let K,L be such that $p^K(x_1,x_2)p^L(y_1,y_2)>0$. Since p is aperiodic, by Lemma 3.33, one may choose n_0 such that $p^M(x_2,x_2)p^M(y_2,y_2)>0$ for all $M\ge n_0$. This implies

$$q^{K+L+M}((x_1,y_1),(x_2,y_2))\ge p^K(x_1,x_2)p^{L+M}(x_2,x_2)p^L(y_1,y_2)p^{K+M}(y_2,y_2)>0.,$$

for $M>n_0$. This finishes the first step.

Step 2: Let X_n, Y_n are independent Markov chains on S with transition probability p and initial distributions μ_X, μ_Y. Then, (X_n, Y_n) is a Markov chain on S^2 with transition probability q and initial distribution $\mu_X \times \mu_Y$. Clearly, $(x, y) \in S^2 \mapsto \pi(x)\pi(y)$ defines a stationary distribution for q. This means that all states in S^2 are positive recurrent. Set $T_{(x,x)} = \inf\{n \geq 1 | (X_n, Y_n) = (x, x)\}$ and

$$T = \inf_{x \in S} T_{(x,x)} = \inf\{n \geq 1 | X_n = Y_n\}.$$

Since q is irreducible and all states in S are recurrent, $\mathbb{P}_\mu(T_{(x,x)} < \infty) = 1$ for any initial distribution μ and $x \in S$, which implies $\mathbb{P}_\mu(T < \infty) = 1$.
Step 3: By the Markov property, one has

$$\begin{aligned}
\mathbb{P}_\mu(X_n = y, T \leq n) &= \sum_{m=1}^{n} \sum_{x \in S} \mathbb{P}_\mu(T = m, X_m = x, X_n = y) \\
&= \sum_{m=1}^{n} \sum_{x \in S} \mathbb{P}_\mu(T = m, X_m = x)\mathbb{P}_\mu(X_n = y | T = m, X_m = x) \\
&= \sum_{m=1}^{n} \sum_{x \in S} \mathbb{P}_\mu(T = m, Y_m = x)\mathbb{P}_\mu(Y_n = y | T = m, Y_m = x) \\
&= \mathbb{P}_\mu(Y_n = y, T \leq n)
\end{aligned}$$

This implies

$$\mathbb{P}_\mu(X_n = y) - \mathbb{P}_\mu(Y_n = y) \leq \mathbb{P}_\mu(X_n = y, T > n).$$

Exchanging X_n, Y_n and summing up y gives

$$\sum_{y \in S} |\mathbb{P}_\mu(X_n = y) - \mathbb{P}_\mu(Y_n = y)| \leq 2\mathbb{P}_\mu(T > n).$$

Let $\mu(s, t) = \delta_x(s)\pi(t)$ for $(s, t) \in S^2$. This implies

$$\sum_{y \in S} |p^n(x, y) - \pi(y)| \leq 2\mathbb{P}_\mu(T > n) \to 0$$

as $n \to \infty$. □

In the following, we consider the periodic cases.

Lemma 3.35. *Let p be irreducible and recurrent with period $d > 1$. Fix $x \in S$ and set, for each $y \in S$, $K_y = \{n \geq 1 : p^n(x, y) > 0\}$.*

(1) *There is a unique $r_y \in \{0, 1, ..., d - 1\}$ such that $K_y \subset r_y + d\mathbb{Z}$.*
(2) *For $r \in \{0, 1, ..., d - 1\}$, let $S_r = \{y \in S : r_y = r\}$. Then, for any $y_i \in S_i$ and $y_j \in S_j$,*

$$\{n \geq 1 : p^n(y_i, y_j) > 0\} \subset (j - i) + d\mathbb{Z}.$$

Such a partition $S_0, ..., S_{d-1}$ of S is independent of the choice of $x \in S$.
(3) *For $0 \leq i < d$, p^d is an irreducible and aperiodic transition probability on S_i.*

Proof. For (1), let m be such that $p^m(y, x) > 0$. Then, for $n \in K_y$, $p^{n+m}(x, x) > 0$. This implies that $d | (n + m)$. By letting $r_y = (d - m) \bmod d$, we have $K_y \subset r_y + d\mathbb{Z}$. To see the uniqueness, let r'_y be another integer in $\{0, 1, ..., d - 1\}$ such that $K_y \subset r'_y + d\mathbb{Z}$. Then, $d | (r_y - r'_y)$ and this can be true only if $r_y = r'_y$.

For (2), let $n, m > 0$ be such that $p^n(y_i, y_j)p^m(x, y_i) > 0$. Then, $d | (m - i)$ and $d | (m + n - j)$. This implies $d | (n - (j - i))$. For (3), it follows immediately from (2) that $(p^d)|_{S_i \times S_i}$ is an irreducible transition probability on S_i for all $0 \leq i < d$. Note that, for $x \in S_i$, if x has period c under p^d, then x has period cd under p. Hence, p^d is aperiodic. □

Remark 3.17. The sets $S_0, S_1, ..., S_{d-1}$ are called the cyclic decomposition of S.

Theorem 3.36. *Let p be irreducible with stationary distribution π. Assume that all states in S are of period d and $S_0, S_1, ..., S_{d-1}$ be the cyclic decomposition of S in Lemma 3.35 corresponding to x. Then, for $y \in S_r$,*

$$\lim_{n\to\infty} p^{r+nd}(x,y) = d\pi(y).$$

Proof. Set $\widetilde{p} = p^d|_{S_0\times S_0}$. By Lemma 3.35, $\widetilde{p}$ is irreducible and aperiodic on S_0. Note that, for $z \in S_0$,

$$\sum_{y\in S_0} \pi(y)\widetilde{p}(y,z) = \sum_{y\in S} \pi(y)p^d(y,z) = \pi(z).$$

This implies that $\pi|_{S_0}$ is a stationary measure for $\widetilde{p}$. By Theorem 3.34, as $x \in S_0$,

$$\lim_{n\to\infty} \widetilde{p}^n(x,y) = \frac{\pi(y)}{\pi(S_0)}, \quad \forall y \in S_0. \tag{3.12}$$

Let $(X_n)_{n=0}^\infty$ be a Markov chain on S with transition probability p and set $Y_n = X_{nd}$. Clearly, $(Y_n)_{n=0}^\infty$ is a Markov chain on S_0. Set $\widetilde{T}_y = \inf\{n \ge 1 | Y_n = y\}$ and $T_y = \inf\{n \ge 1 | X_n = y\}$. It is easy to check that, under $\mathbb{P}_x$, $d\widetilde{T}_y = T_y$ and, by (3.11), we have

$$\frac{1}{n}\sum_{k=1}^n \widetilde{p}^k(x,y) \to \frac{1}{\mathbb{E}_y\widetilde{T}_y} = \frac{d}{\mathbb{E}_yT_y} = d\pi(y), \quad \forall y \in S_0.$$

In addition with (3.12), this implies $\pi(S_0) = 1/d$. As a result, it follows that, for $0 < r < d$ and $y \in S_r$,

$$p^{nd+r}(x,y) = \sum_{z\in S_0} p^{nd}(x,z)p^r(z,y) \to \sum_{z\in S_0} d\pi(z)p^r(z,y) = d\sum_{z\in S} \pi(z)p^r(z,y) = d\pi(y).$$

□

For null recurrent states, we have the following observation.

Theorem 3.37. *Suppose p is irreducible and all states are null recurrent. Then,*

$$\lim_{n\to\infty} p^n(x,y) = 0, \quad \forall x,y \in S.$$

Proof. We first consider the case that p is aperiodic. Let $x, y \in S$. Since y is null recurrent, $\mathbb{P}_y(T_y < \infty) = 1$ and $\mathbb{E}_yT_y = \infty$. Let $\epsilon > 0$ and choose $N > 0$ such that

$$\sum_{m=1}^N \mathbb{P}_y(T_y > m) \ge 2/\epsilon.$$

Note that, for $n \ge N$,

$$\begin{aligned} 1 &\ge \mathbb{P}_x(X_m = y, \text{for some } n - N \le m \le n) \\ &= \sum_{k=n-N}^n \mathbb{P}_x(X_k = y, X_{k+1} \ne y, ..., X_n \ne y) \\ &= \sum_{k=n-N}^n p^k(x,y)\mathbb{P}_y(T_y > n-k) = \sum_{m=0}^N p^{n-m}(x,y)\mathbb{P}_y(T_y > m). \end{aligned}$$

This implies that there is $0 \le m \le N$ such that $p^{n-m}(x,y) \le \epsilon/2$ or equivalently

$$\min_{0\le m\le N} p^{n+m}(x,y) \le \epsilon/2, \quad \forall n \ge 0.$$

Recall the coupling in the proof of Theorem 3.34 and let q be the corresponding transition probability. As before, q is irreducible and aperiodic. Note that if q is transient, then

$$0 = \lim_{n\to\infty} q^n((x,x)(y,y)) = \lim_{n\to\infty} p^n(x,y)^2, \quad \forall n \geq N.$$

If q is recurrent, then the coupling time T satisfies $\mathbb{P}_{\mu\times\nu}(T < \infty) = 1$ for any probabilities μ, ν on S. By setting $\mu = \delta_x$ and $\nu = p^m(x, \cdot)$ with $m = 1, 2, ..., N$, we have

$$|p^n(x,y) - p^{n+m}(x,y)| \leq \mathbb{P}_{\mu\times\nu}(T > n) \to 0, \quad \text{as } n \to \infty.$$

As a consequence, we may select $M > 0$ such that

$$\max_{0\leq m\leq N} |p^n(x,y) - p^{n+m}(x,y)| \leq \epsilon/2, \quad \forall n \geq M,$$

which leads to

$$p^n(x,y) \leq \max_{0\leq m\leq N} |p^n(x,y) - p^{n+m}(x,y)| + \min_{0\leq m\leq N} p^{n+m} \leq \epsilon, \quad \forall n \geq M.$$

The proof for periodic p is similar to that in Theorem 3.34 and is omitted. □