Lecture-16: Invariant Distribution

1 Transient and recurrent states

1.1 Hitting and return times

Definition 1.1. For a homogeneous Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, we can define **first hitting time** to state $x \in \mathcal{X}$, as

$$\tau_x^+ \triangleq \inf\{n \in \mathbb{N} : X_n = x\}.$$

If $X_0 = x$, then τ_x^+ is called the **first return time** to state x.

Lemma 1.2. *For an irreducible Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ on finite state space $\mathcal{X}$, we have $\mathbb{E}_x \tau_y^+ < \infty$ for all states $x, y \in \mathcal{X}$.*

Proof. From the definition of irreducibility, for each pair of states $z, w \in \mathcal{X}$, we have a positive integer $n_{zw} \in \mathbb{N}$ such that $p_{zw}^{n_{zw}} > \epsilon_{zw} > 0$. Since the state space $\mathcal{X}$ is finite, We define

$$\epsilon \triangleq \inf_{z,w \in \mathcal{X}} \epsilon_{zw} > 0, \qquad r \triangleq \sup_{z,w \in \mathcal{X}} n_{zw} \in \mathbb{N}.$$

Hence, there exists a positive integer $r \in \mathbb{N}$ and a real $\epsilon > 0$ such that $p_{zw}^{(n)} > \epsilon$ for some $n \leqslant r$ and all states $z, w \in \mathcal{X}$. It follows that $P(\cup_{n \in [r]} \{X_n = y\}) > \epsilon$ or $P_z \left\{\tau_y^+ > r\right\} \leqslant 1 - \epsilon$ for any initial condition $X_0 = z \in \mathcal{X}$. Therefore, we can write for $k \in \mathbb{N}$

$$P_x \left\{\tau_y^+ > kr\right\} = P_x \left\{\tau_y^+ > (k-1)r\right\} P(\left\{\tau_y^+ > kr\right\} \mid \left\{\tau_y^+ > (k-1)r, X_0 = x\right\}) \leqslant (1-\epsilon) P_x \left\{\tau_y^+ > (k-1)r\right\}.$$

By induction, we have $P_x \left\{\tau_y^+ > kr\right\} \leqslant (1-\epsilon)^k$. Since $P_x \left\{\tau_y^+ > n\right\}$ is decreasing in n, we can write

$$\mathbb{E}_x \tau_y^+ = \sum_{k \in \mathbb{Z}_+} \sum_{i=0}^{r-1} P_x\{\tau_y^+ > kr + i\} \leqslant \sum_{k \in \mathbb{Z}_+} r P_x\{\tau_y^+ > kr\} \leqslant \frac{r}{\epsilon} < \infty.$$

□

Corollary 1.3. *For an irreducible Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ on finite state space $\mathcal{X}$, we have $P_x \left\{\tau_y^+ < \infty\right\} = 1$ for all states $x, y \in \mathcal{X}$.*

Proof. This follows from the fact that τ_y^+ is a positive random variable with finite mean for all states $y \in \mathcal{X}$ and any initial state $x \in \mathcal{X}$. □

1.2 Recurrence and transience

Definition 1.4. Let $f_{xy}^{(n)}$ denote the probability that starting from state x, the first transition into state y happens at time n. Then, $f_{xy}^{(n)} = P_x \left\{\tau_y^+ = n\right\}$. Then we denote the probability of eventually entering state y given that we start at state x, as $f_{xy} = \sum_{n=1}^{\infty} f_{xy}^{(n)} = P_x \left\{\tau_y^+ < \infty\right\}$. The state y is said to be **transient** if $f_{yy} < 1$ and **recurrent** if $f_{yy} = 1$.

Definition 1.5. For a discrete time process $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, the total number of visits to a state $y \in \mathcal{X}$ in first n steps is denoted by $N_y(n) \triangleq \sum_{i=1}^{n} \mathbb{1}_{\{X_i = y\}}$. Total number of visits to state $y \in \mathcal{X}$ is denoted by $N_y \triangleq N_y(\infty)$.

Remark 1. From the linearity of expectations and monotone convergence theorem, we get $\mathbb{E}_y N_y = \sum_{n \in \mathbb{N}} p_{yy}^{(n)}$.

Lemma 1.6. *Consider a homogeneous Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$. For each $m \in \mathbb{Z}_+$ and state $x, y \in \mathcal{X}$, we have*

$$P_x \{N_y = m\} = \begin{cases} 1 - f_{xy} & m = 0, \\ f_{xy} f_{yy}^{m-1}(1 - f_{yy}) & m \in \mathbb{N}. \end{cases}$$

Proof. For each $k \in \mathbb{N}$, the time $\tau_y^+(k)$ of the kth visit to the state y is a stopping time. From strong Markov property, the next return to state y is independent of the past. That is, $(\tau_y^+(k+1) - \tau_y^+(k) : k \in \mathbb{N})$ is an *i.i.d.* sequence, distributed identically to τ_y^+ starting from an initial state $X_0 = y$. When $X_0 = x \neq y$, then τ_y^+ is independent of sequence $(\tau_y^+(k+1) - \tau_y^+(k) : k \in \mathbb{N})$ and distributed differently. We observe that

$$\{N_y = m\} = \left\{\tau_y^+(m) < \infty, \tau_y^+(m+1) = \infty\right\} = \cap_{k=1}^m \left\{\tau_y^+(k) - \tau_y^+(k-1) < \infty\right\} \cap \left\{\tau_y^+(m+1) - \tau_y^+(m) = \infty\right\}.$$

It follows from the strong Markov property for process X, that

$$P_x \{N_y = m\} = P_x \left\{\tau_y^+(1) < \infty\right\} \prod_{k=2}^m P_y \left\{\tau_y^+(k) - \tau_y^+(k-1) < \infty\right\} P_y \left\{\tau_y^+(m+1) = \infty\right\}.$$

□

Corollary 1.7. *For a homogeneous Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, we have $P_y \{N_y < \infty\} = \mathbb{1}_{\{f_{yy} < 1\}}$.*

Proof. We can write the event $\{N_y < \infty\}$ as the disjoint union of events $\{N_y = m\}$ for $m \in \mathbb{Z}_+$, and the result follows from additivity of probability over disjoint events, and the expression for the conditional probability mass function $P_y \{N_y = m\}$ in Lemma 1.6. □

Remark 2. In particular, this corollary implies the following.

1. A transient state is visited a finite amount of times almost surely.
2. A recurrent state is visited infinitely often almost surely.
3. Since $\sum_{y \in \mathcal{X}} N_y = \infty$, it follows that all states can be transient in a finite state Markov chain.

Proposition 1.8. *A state $y \in \mathcal{X}$ is recurrent iff $\sum_{k \in \mathbb{N}} p_{yy}^{(k)} = \infty$.*

Proof. For any state $y \in \mathcal{X}$, we can write

$$p_{yy}^{(k)} = P_x \{X_k = y\} = \mathbb{E}_x \mathbb{1}_{\{X_k = y\}}.$$

Using monotone convergence theorem to exchange expectation and summation, we obtain

$$\sum_{k \in \mathbb{N}} p_{yy}^{(k)} = \mathbb{E}_y \sum_{k \in \mathbb{N}} \mathbb{1}_{\{X_k = y\}} = \mathbb{E}_y N_y.$$

Thus, $\sum_{k \in \mathbb{N}} p_{yy}^{(k)}$ represents the expected number of returns $\mathbb{E}_y N_y$ to a state y starting from state y, which we know to be finite if the state is transient and infinite if the state is recurrent. □

Proposition 1.9. *Transience and recurrence are class properties.*

Proof. Let us start with proving recurrence is a class property. Let x be a recurrent state and let $x \leftrightarrow y$. Then, we will show that y is a recurrent state. From the reachability, there exist some $m, n > 0$, such that $p_{xy}^{(m)} > 0$ and $p_{yx}^{(n)} > 0$. As a consequence of the recurrence, $\sum_{s \in \mathbb{Z}_+} p_{xx}^{(s)} = \infty$. It follows that y is recurrent by observing

$$\sum_{k \in \mathbb{Z}_+} p_{yy}^{(k)} \geqslant \sum_{s \in \mathbb{Z}_+} p_{yy}^{(m+n+s)} \geqslant \sum_{s \in \mathbb{Z}_+} p_{yx}^{(n)} p_{xx}^{(s)} p_{xy}^{(m)} = \infty.$$

Now, if x were transient instead, we conclude that y is also transient by the following observation

$$\sum_{s \in \mathbb{Z}_+} p_{yy}^{(s)} \leqslant \frac{\sum_{s \in \mathbb{Z}_+} p_{xx}^{(m+n+s)}}{p_{yx}^{(n)} p_{xy}^{(m)}} < \infty.$$

□

Corollary 1.10. *If y is recurrent, then for any state x such that $x \leftrightarrow y$, $f_{xy} = 1$.*

2 Invariant distribution

Definition 2.1. For a time-homogeneous Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ with transition matrix P, a distribution $\pi \in \mathcal{M}(\mathcal{X})$ is called **invariant** if it is a left eigenvector of the probability transition matrix P with eigenvalue unity, or

$$\pi = \pi P.$$

Remark 3. Recall that $\nu(n) \in \mathcal{M}(\mathcal{X})$ where $\nu_x(n) = P\{X_n = x\}$ for all $x \in \mathcal{X}$, denotes the probability distribution of the Markov chain X being in one of the states at step $n \in \mathbb{N}$. Then, if $\nu(0) = \pi$, then $\nu(n) = \nu(0)P^n = \pi$ for all time-steps $n \in \mathbb{N}$.

Definition 2.2. For a time-homogeneous Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ with transition matrix P, the stationary distribution is defined as $\nu(\infty) \triangleq \lim_{n\to\infty} \nu(n)$.

Remark 4. For a Markov chain with initial distribution being invariant, the stationary distribution is invariant distribution.

Example 2.3 (Simple random walk on a directed graph). Let $G = (V, E)$ be a finite directed graph. We define a simple random walk on this graph as a Markov chain with state space V and transition matrix $P : V \times V \to [0,1]$ where $p_{xy} \triangleq \frac{1}{\deg_{\text{out}}(x)} \mathbb{1}_{\{(x,y)\in E\}}$. We observe that vector $(\deg_{\text{out}}(x) : x \in \mathcal{X})$ is a left eigenvector of the transition matrix P with unit eigenvalue. Indeed we can very that

$$\sum_{x\in\mathcal{X}} \deg_{\text{out}}(x) p_{xy} = \sum_{x\in\mathcal{X}} \mathbb{1}_{\{(x,y)\in E\}} = \deg_{\text{out}}(y).$$

Since $\sum_{x\in\mathcal{X}} \deg_{\text{out}}(x) = 2|E|$, it follows that $\pi : \mathcal{X} \to [0,1]$ defined by $\pi_x \triangleq \frac{\deg_{\text{out}}(x)}{2|E|}$ for each $x \in V$, is the equilibrium distribution of this simple random walk.

2.1 Existence of an invariant distribution

Proposition 2.4. *Consider an irreducible and aperiodic homogeneous DTMC $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ with transition matrix P and starting from initial state $X_0 = x$. Let the positive vector $\tilde{\pi}_x : \mathcal{X} \to \mathbb{R}_+$ defined as*

$$\tilde{\pi}_x(y) \triangleq \mathbb{E}_x \sum_{n=1}^{\tau_x^+} \mathbb{1}_{\{X_n=y\}} = \mathbb{E}_x \sum_{n\in\mathbb{N}} \mathbb{1}_{\{n\leqslant\tau_x^+\}} \mathbb{1}_{\{X_n=y\}}, \quad y \in \mathcal{X}.$$

Then $\tilde{\pi}_x = \tilde{\pi}_x P$ if $P_x\{\tau_x^+ < \infty\} = 1$, and $\pi \triangleq \frac{\tilde{\pi}_x}{\mathbb{E}_x\tau_x^+}$ is a stationary distribution if $\mathbb{E}_x\tau_x^+ < \infty$.

Proof. We will first show that π is a distribution on state space $\mathcal{X}$. We first observe that

$$\sum_{y\in\mathcal{X}} \tilde{\pi}_x(y) = \sum_{y\in\mathcal{X}} \sum_{n=1}^{\tau_x^+} \mathbb{1}_{\{X_n=y\}} = \sum_{n=1}^{\tau_x^+} \mathbb{1}_{\{X_n\in\mathcal{X}\}} = \mathbb{E}_x\tau_x^+.$$

Thus $\tilde{\pi}_x(y) = \mathbb{E}_x \sum_{n=1}^{\tau_x^+} \mathbb{1}_{\{X_n=y\}} \leqslant \mathbb{E}_x\tau_x^+$ for all states $y \in \mathcal{X}$. If $\mathbb{E}_x\tau_x^+ < \infty$, then $\tilde{\pi}_x(y) < \infty$ for each $y \in \mathcal{X}$. Further, we have $\tilde{\pi}_x(x) = 1$. Since $\tilde{\pi}_x(y) \geqslant 0$, it follows that $\frac{\tilde{\pi}_x}{\mathbb{E}_x\tau_x^+}$ is a distribution on the state space $\mathcal{X}$.

We next show that $\tilde{\pi}_x$ is an invariant distribution of DTMC X. Using the monotone convergence theorem, we can write

$$\sum_{w\in\mathcal{X}} \tilde{\pi}_x(w) p_{wz} = \sum_{n\in\mathbb{N}} \sum_{w\in\mathcal{X}} P_x\{\tau_x^+ \geqslant n, X_n = w\} P(\{X_{n+1} = z\} \mid \{X_n = w\}).$$

We first focus on the term $w = x$. We see that $\{X_n = x, \tau_x^+ \geqslant n\} = \{\tau_x^+ = n\}$. Hence, from the strong Markov property, we have $P_x\{X_n = x, X_{n+1} = z, \tau_x^+ \geqslant n\} = P_x\{\tau_x^+ = n\} p_{xz}$. Summing over all $n \in \mathbb{N}$, we get

$$\tilde{\pi}_x(x) p_{xz} = \sum_{n\in\mathbb{N}} P_x\{X_n = x, X_{n+1} = z, \tau_x^+ \geqslant n\} = p_{xz} \sum_{n\in\mathbb{N}} P_x\{\tau_x^+ = n\} = p_{xz}.$$

We next focus on the terms $w \neq x$, such that $\{X_n = w, \tau_x^+ \geqslant n\} = \{X_n = w, \tau_x^+ \geqslant n+1\} \in \mathcal{F}_n$. Hence, from the Markov property of X, we can write

$$\begin{aligned} P_x\left\{\tau_x^+ \geqslant n+1, X_n = w, X_{n+1} = z\right\} &= P_x\left\{\tau_x^+ \geqslant n, X_n = w\right\} P(\{X_{n+1} = z\} \mid \{X_n = w, \tau_x^+ \geqslant n, X_0 = x\}) \\ &= P_x\left\{\tau_x^+ \geqslant n, X_n = w\right\} p_{wz}. \end{aligned}$$

Summing both sides over $n \in \mathbb{N}$ and $w \neq x$, we get

$$\begin{aligned} \sum_{w \neq x} \tilde{\pi}_x(w) p_{wz} &= \sum_{n \in \mathbb{N}} \sum_{w \neq x} P_x\left\{\tau_x^+ \geqslant n+1, X_n = w, X_{n+1} = z\right\} = \sum_{n \geqslant 2} P_x\left\{\tau_x^+ \geqslant n, X_n = z\right\} \\ &= \tilde{\pi}_x(z) - P_x\left\{\tau_x^+ \geqslant 1, X_1 = z\right\} = \tilde{\pi}_x(z) - p_{xz}. \end{aligned}$$

The result follows from summing both the cases. □

2.2 Uniqueness of stationary distribution

Recall that distributions π on state space $\mathcal{X}$ such that $\pi P = \pi$ is called a stationary distribution. Similarly, a function $h : \mathcal{X} \to \mathbb{R}$ is called **harmonic at** x if

$$h(x) = \sum_{y \in \mathcal{X}} p_{xy} h(y).$$

A function is **harmonic on a subset** $D \subset \mathcal{X}$ if it is harmonic at every state $x \in D$. That is, $Ph = h$ for a function harmonic on the entire state space $\mathcal{X}$.

Lemma 2.5. *For a finite irreducible Markov chain, a function f that is harmonic on all states in $\mathcal{X}$ is a constant.*

Proof. Suppose h is not a constant, then there exists a state $x_0 \in \mathcal{X}$, such that $h(x_0) \geqslant h(y)$ for all states $y \in \mathcal{X}$. Since the Markov chain is irreducible, there exists a state $z \in \mathcal{X}$ such that $p_{x_0,z} > 0$. Let's assume $h(z) < h(x_0)$, then

$$h(x_0) = p_{x_0,z} h(z) + \sum_{y \neq z} p_{x_0,y} h(y) < h(x_0).$$

This implies that $h(x_0) = h(z)$ for all states z such that $p_{x_0,z} > 0$. By induction, this implies that any $h(x_0) = h(y)$ for any states y reachable from state x_0. Since all states are reachable from state x_0 by irreducibility, this implies h is a constant on the state space $\mathcal{X}$. □

Corollary 2.6. *For any irreducible and aperiodic finite Markov chain, there exists a unique stationary distribution π.*

Proof. For an aperiodic and irreducible DTMC $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ with finite state space $\mathcal{X}$, we have $P_x\left\{\tau_y^+ < \infty\right\} = 1$ and $\mathbb{E}_x \tau_y^+ < \infty$ for all states $x, y \in \mathcal{X}$. Therefore, we have seen the existence of a positive stationary distribution π for an irreducible and aperiodic finite Markov chain. Further, from previous Lemma we have that the dimension of null-space of $(P - I)$ is unity. Hence, the rank of $P - I$ is $|\mathcal{X}| - 1$. Therefore, all vectors satisfying $\nu = \nu P$ are scalar multiples of π. □

2.3 Stationary distribution for irreducible and aperiodic finite DTMC

For a finite state irreducible and aperiodic DTMC $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, we have $\mathbb{E}_x \tau_y^+ < \infty$ and $P_x\left\{\tau_y < \infty\right\} = 1$ for all $x, y \in \mathcal{X}$. That is, the return times are finite almost surely, and hence we can apply strong Markov property at these stopping times to obtain that DTMC X is a regenerative process with delayed renewal sequence $\tau_y^+ : \Omega \to \mathbb{N}^{\mathbb{N}}$, where $\tau_y^+(0) \triangleq 0$, and $\tau_y^+(n) \triangleq \inf\left\{m > \tau_y^+(n-1) : X_m = y\right\}$.

Theorem 2.7. *For a finite state irreducible and aperiodic Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, its invariant distribution is same as its stationary distribution.*

Proof. We can create an on-off alternating renewal function on this DTMC X, which is ON when in state y. Then, from the limiting ON probability of alternating renewal function, we know that

$$\pi(y) \triangleq \lim_{k \to \infty} P_x\left\{X_k = y\right\} = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_k = y\}} = \frac{1}{\mathbb{E}_y \tau_y^+}.$$

We observe that $\pi(y) = \frac{\tilde{\pi}_y(y)}{\mathbb{E}_y \tau_y^+}$ for each state $y \in \mathcal{X}$. From the uniqueness of invariant distribution, it follows that π is the unique invariant distribution of the DTMC X. We observe that $\pi(x)$ is the long-term average of the amount of time spent in state x and from renewal reward theorem $\pi(x) = \frac{1}{\mathbb{E}_x \tau_x^+}$. □