AI2613 随机过程 2020–2021 春季学期

Lecture 3 – Discrete Markov Chain (II)

2021 年 3 月 8 日

Lecturer: 张驰豪 *Scribe*: 杨宽

1 Notations and Conventions

To simplify our notations and statements, we will use (I) (A) (R) (S) (U) (C) to denote the following properties of Markov chains:

(I) irreducible,

(A) aperiodic,

(R) recurrent,

(S) ∃ a stationary distribution,

(U) ∃ a uniqueness stationary distribution,

(C) convergence.

Moreover, we will use the notations (NR) and (PR) to denote "null recurrent" and "postive recurrent" respectively, while the definition of "null recurrent" and "positive recurrent" will be given later.

2 An Example of Infinite Markov Chains

Review of the last lecture: for finite Markov chains, the fundamental theorem of Markov chains tells us

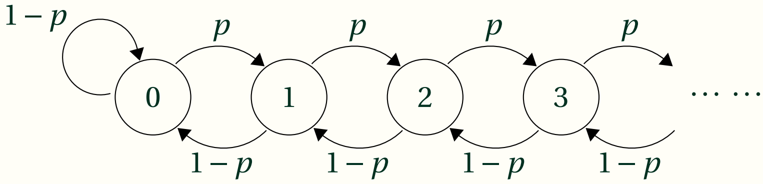
$$(\mathrm{I}) + (\mathrm{A}) \implies (\mathrm{U}) + (\mathrm{C}).$$

But how about infinite space?

Let's first consider the following example.

Example 1 (One-dimensional random walk with an absorbing barrier). The Markov chain is a one-dimensional random walk on $\mathbb{N}$ with an absorbing barrier at 0.

Question. What's the stationary distribution of the above Markov chain?



Clearly, we have some necessary conditions as follows: assume that there exists a stationary distribution π, then

$$\begin{aligned} & \pi(0) = (1-p)\cdot\pi(0) + (1-p)\cdot\pi(1) \\ \Longrightarrow\ & \pi(0) = \frac{1-p}{p}\cdot\pi(1)\,, \\ & \pi(1) = p\cdot\pi(0) + (1-p)\cdot\pi(2) \\ & \qquad = (1-p)\cdot\pi(1) + (1-p)\cdot\pi(2) \\ \Longrightarrow\ & \pi(1) = \frac{1-p}{p}\cdot\pi(2)\,, \\ & \qquad \text{and so on...} \end{aligned}$$

It implies that

$$\pi(i) = \frac{1-p}{p}\cdot\pi(i+1) \text{ for } i \geq 0 \quad \text{and} \quad \sum_{n=0}^{\infty}\pi(n) = 1\,.$$

So $\{\pi(n)\}$ should be a geometric progression, and there are three cases:

1. $p < 1/2$: we have

   $$\pi(i) = \left(\frac{p}{1-p}\right)^i\cdot\pi(0)\,,$$

   so

   $$1 = \sum_n\left(\frac{p}{1-p}\right)^n\cdot\pi(0) = \pi(0)\cdot\frac{1}{1-\frac{p}{1-p}} = \pi(0)\cdot\frac{1-p}{1-2p}\,,$$

   which implies that

   $$\pi(i) = \left(\frac{p}{1-p}\right)^i\cdot\frac{1-2p}{1-p}\,.$$

   (In fact, there exists a stationary distribution indeed.)
2. $p > 1/2$: $\pi(0) < \pi(1) < \cdots \Longrightarrow$ no stationary distribution.
3. $p = 1/2$: $\pi(0) = \pi(1) = \cdots \Longrightarrow$ no stationary distribution.

Although neither case 2 nor case 3 has a stationary distribution, they are still different. As we saw in the last lecture, case 3 is recurrent but case 2 is not.

Case 1 is also recurrent. However, there is a fundamental difference between the recurrence in case 1 and case 3 — case 1 is "positive recurrent" while case 3 is "null recurrent". We will introduce it later, but now let us focus on the core question:

Question. When does an infinite chain have a stationary distribution?

3 Law of Large Numbers

In order to answer the question above, we first review the law of larger numbers.

Definition 2 (Convergence). Let $X_0, X_1, X_2, \ldots$ be a sequence of random variables defined on an underlying sample space Ω and an underlying σ-algebra $\mathcal{F}$. We start by defining different modes of convergence.

- **Convergence in probability**. X_t is said to *converge to X in probability* (written $X_t \xrightarrow{P} X$) if

$$\forall \varepsilon > 0, \qquad \mathbf{Pr}[|X_n - X| > \varepsilon] \to 0 .$$

- **Almost sure convergence**. We say that the sequence X_t *converges almost surely* to X (written $X_t \xrightarrow{a.s.} X$), if $\exists M \in \mathcal{F}$, s.t.
  1. $\mathbf{Pr}[M] = 1$;
  2. $\forall \omega \in M$, $X_n(\omega) \to X(\omega)$ as $n \to \infty$.

  Namely,

$$\mathbf{Pr}\left[\lim_{n\to\infty} X_n = X\right] = 1 .$$

Fact 1. *Almost sure convergence $\implies$ convergence in probability.*

Example 3.

$$X_1, \ldots, X_n, \ldots \text{ where } X_n = \begin{cases} 1 & \text{with probability } \frac{1}{n}; \\ 0 & \text{with probability } 1 - \frac{1}{n}. \end{cases}$$

Then $X_n \xrightarrow{P} 0$ but $X_n \not\xrightarrow{a.s.} 0$.

Now we review the law of large numbers. Suppose that $X_1, \ldots, X_n, \ldots$ are i.i.d. random variables s.t. $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] < \infty$. Let $S_n \triangleq \frac{1}{n}\sum_{i=1}^n X_i$.

Theorem 2 (Weak Law of Large Numbers (WLLN)).

$$S_n \xrightarrow{P} \mu .$$

Namely, $\forall \varepsilon > 0$, $\mathbf{Pr}\left[\left|S_n - \mu\right| > \varepsilon\right] \to 0$.

Theorem 3 (Strong Law of Large Numbers (SLLN)).

$$S_n \xrightarrow{a.s.} \mu\,.$$

Namely, $\mathbf{Pr}\left[\lim_{n\to\infty} S_n = \mu\right] = 1$.

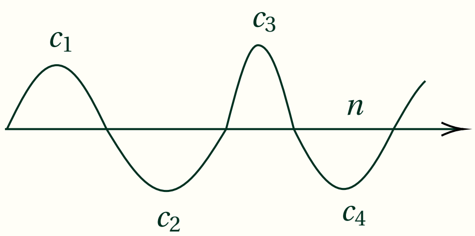
Recall that T_j is the first hitting time of j. We are now going to show the strong law of large numbers for Markov chains.

Theorem 4 (Strong Law of Large Numbers for Markov Chains). *Let $X_0, X_1, \ldots$ be a Markov chain starting at $X_0 = i$. Suppose that state i communicates with another state j. Then,*

$$P_i\left[\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}_{[X_t=j]} = \frac{1}{\mathbb{E}_j\left[T_j\right]}\right] = 1\,.$$

Proof. Consider the following three cases:

1. Case 1: j is transient. If j is transient, then $\mathbb{E}_j\left[T_j\right] = \infty$ since $P_j[T_j = \infty] > 0$. On the other hand, applying Proposition 10 in the last lecture we have $\mathbb{E}_j\left[N_j\right] < \infty$. Namely, $N_j = \lim_{n\to\infty}\sum_{t=1}^{n} \mathbb{1}_{[X_t=j]} < \infty$ with probability 1. Thus, $\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbb{1}_{[X_t=j]} = 0 = 1/\mathbb{E}_j\left[T_j\right]$ with probability 1.
2. Case 2: $i = j$ is recurrent. Let c_i be the length of the i-th cycle starting from j and then returning back to j. Then $c_1, c_2, \ldots$ are i.i.d. random variables with $\mathbb{E}[c_i] = \mathbb{E}_j\left[T_j\right]$. Let $S_k = c_1 + c_2 + \cdots + c_k$.



Let $v_n \triangleq \max\{k\colon S_k \le n\}$. So $S_{v_n} \le n \le S_{v_n+1}$, which yields that

$$\frac{S_{v_n}}{v_n} \le \frac{n}{v_n} \le \frac{S_{v_n+1}}{v_n}\,.$$

Since $v_n \to \infty$ as $n \to \infty$, applying the strong law of large numbers, we have

$$\frac{S_{v_n}}{v_n} \xrightarrow{a.s.} \mathbb{E}_j\left[T_j\right] \quad \text{and} \quad \frac{S_{v_n+1}}{v_n} = \frac{S_{v_n+1}}{v_n+1}\cdot\frac{v_n+1}{v_n} \xrightarrow{a.s.} \mathbb{E}_j\left[T_j\right]\,.$$

Thus $v_n/n \xrightarrow{a.s.} 1/\mathbb{E}_j\left[T_j\right]$.

3. Case 3: $i \neq j$ and j is recurrent. The finite path $i \to j$ is negligible. □

Next we need the following bounded convergence theorem:

Theorem 5 (Bounded Convergence Theorem). *Let $\{f_n\}$ be a sequence of bounded measurable functions on a set of finite measure E. If $f_n \to f$ uniformly on E, then*

$$\lim_{n\to\infty} \int_E f_n = \int_E f .$$

Namely, if X_n are bounded and $X_n \to c$ with probability 1, then $\mathbb{E}[X_n] \to c$ as $n \to \infty$.

Combining with the bounded convergence theorem, the strong law of large numbers for Markov chains yields the following two corollaries for finite chains.

Corollary 6. *For any irreducible Markov chain (if* (I)) and any two states i, j,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbf{P}^t(i,j) = \frac{1}{\mathbb{E}_j\left[T_j\right]} .$$

(Since $\mathbb{E}_i\left[\mathbb{1}_{[X_t=j]}\right] = \mathbf{P}^t(i,j)$.)

Corollary 7. *By the fundamental theorem of Markov chains,*

$$(\mathrm{I})+(\mathrm{A})+(\mathrm{S}) \implies \lim_{n\to\infty} \mathbf{P}^n(i,j) = \pi(j) ,$$

thus we have

$$(\mathrm{I})+(\mathrm{A})+(\mathrm{S}) \implies \frac{1}{\mathbb{E}_j\left[T_j\right]} = \pi(j) ,$$

where we use the Cesàro summation:

Proposition 8 (Cesàro summation). *Suppose $a_1, a_2, \ldots, a_n, \ldots$ is a sequence and $a_n \to a$. Then we have*

$$\frac{1}{n} \sum_{i=1}^{n} a_i \to a .$$

In fact, assuming (I) and (S) , we can obtain $\pi(j) = \frac{1}{\mathbb{E}_j[T_j]}$ directly, which we will show in the next section.

4 Existence of Stationary Distribution for Infinite Chains

Now we are ready to answer the following question.

Question. When does an infinite chain have a stationary distribution?

Definition 4 (Null Recurrence and Positive Recurrence). Recall that "recurrence" means $P_i[T_i < \infty] = 1$. There are two types of recurrence:

- **positive recurrence**: A state i is *positive recurrent* if $\mathbb{E}_i[T_i] < \infty$.
- **null recurrence**: A state i is *null recurrent* if $\mathbb{E}_i[T_i] = \infty$.

Theorem 9. *Assuming* (I), (PR) $\iff$ (S) + (U).

Proof. We first show the "$\Longleftarrow$" direction. Let $N_i(n)$ be the number of visits of state i in the first n steps. Assuming (I), the strong law of large numbers for Markov chains shows that

$$P_j\left[\lim_{n\to\infty} \frac{N_i(n)}{n} = \frac{1}{\mathbb{E}_i[T_i]}\right] = 1\,.$$

Suppose π is a stationary distribution. Let $X_0 \sim \pi$, i.e. $X_0 = i$ with probability $\pi(i)$. Then applying the bounded convergence theorem it gives that

$$\mathbb{E}_{X_0\sim\pi}\left[\frac{N_i(n)}{n}\right] = \sum_{t=1}^{n} \frac{\mathbb{E}_{X_0\sim\pi}\left[\mathbb{1}_{[X_t=i]}\right]}{n} = n\cdot\frac{\pi(i)}{n} = \frac{1}{\mathbb{E}_i[T_i]}\,.$$

Since the chain is irreducible, $\pi(i) > 0$ holds for every state i, and thus $\mathbb{E}_i[T_i] < \infty$ for all i.
Next we prove the "$\Longrightarrow$" direction. For the uniqueness part, let π be a stationary distribution. Applying the proof of the "$\Longleftarrow$" direction, the stationary distribution π satisfies $\pi(i) = 1/\mathbb{E}_i[T_i]$, thus the stationary distribution is unique.
For the existence part, we begin the proof by assuming that the state space $\mathcal{S}$ is finite. By Corollary 6,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbf{P}^t(i,j) = \frac{1}{\mathbb{E}_j[T_j]}\,.$$

Since $\sum_{j\in\mathcal{S}} \mathbf{P}^t(i,j) = 1$, we have

$$\sum_{j\in\mathcal{S}} \frac{1}{\mathbb{E}_j[T_j]} = 1\,,$$

which yields that π is a probability distribution where $\pi(i) = 1/\mathbb{E}_i[T_i]$. We claim that π is a stationary distribution for the chain.
Now we come to prove our claim. We write out the matrix equation $\mathbf{P}^t \cdot \mathbf{P} = \mathbf{P}^{t+1}$ as follows:

$$\sum_k \mathbf{P}^t(i,k)\cdot\mathbf{P}(k,j) = \mathbf{P}^{t+1}(i,j)\,.$$

Summing over $t = 1, 2, \ldots, n$, it gives that

$$\sum_k \left(\frac{1}{n} \sum_{t=1}^{n} \mathbf{P}^t(i,k) \right) \cdot \mathbf{P}(k,j) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{P}^{t+1}(i,j) .$$

Taking the limit as $n \to \infty$ for the both sides, it yields that

$$\sum_k \frac{1}{\mathbb{E}_k[T_k]} \cdot \mathbf{P}(k,j) = \frac{1}{\mathbb{E}_j[T_j]} .$$

Thus, π is indeed a stationary distribution of the chain.

Finally, we are going to handle the infinite case. Let $A \subseteq \mathcal{S}$ be a finite subset of $\mathcal{S}$. Note that when $\mathcal{S}$ is infinite, there exists a technical issue — $\mathbf{P}$ is no longer a matrix. But we can still define $\mathbf{P}^t(i,j)$ as the probability that starting from i the chain accesses j after exact t steps. Then Corollary 6 still holds, that is,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \mathbf{P}^t(i,j) = \frac{1}{\mathbb{E}_j[T_j]} .$$

(Corollary 6 is implied by Theorem 4 and 5, which does not necessarily require the finite space.)

So we have

$$\sum_{j \in A} \frac{1}{\mathbb{E}_j[T_j]} = \sum_{j \in A} \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \mathbf{P}^t(i,j) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{j \in A} \mathbf{P}^t(i,j) \le 1 .$$

Therefore, $c \triangleq \sum_{j \in \mathcal{S}} \frac{1}{\mathbb{E}_j[T_j]} \le 1$. Given (I), $c > 0$. Actually, we will see that c must be 1 later. Now, similarly to the finite space case, we also have

$$\sum_{k \in A} \mathbf{P}^t(i,k) \cdot \mathbf{P}(k,j) \le \mathbf{P}^{t+1}(i,j) ,$$

and thus,

$$\sum_{k \in A} \frac{1}{\mathbb{E}_k[T_k]} \cdot \mathbf{P}(k,j) \le \frac{1}{\mathbb{E}_j[T_j]} .$$

Taking the supremum over all finite subsets A of $\mathcal{S}$, it implies that

$$\sum_{k \in \mathcal{S}} \frac{1}{\mathbb{E}_k[T_k]} \cdot \mathbf{P}(k,j) \le \frac{1}{\mathbb{E}_j[T_j]} .$$

Taking the summation over all state j, it gives that

$$\sum_j \sum_k \frac{1}{\mathbb{E}_k[T_k]} \cdot \mathbf{P}(k,j) \le \sum_j \frac{1}{\mathbb{E}_j[T_j]} = c .$$

However, the left side of the above inequality is

$$\sum_j \sum_k \frac{1}{\mathbb{E}_k[T_k]} \cdot \mathbf{P}(k,j) = \sum_j c \cdot \mathbf{P}(k,j) = c ,$$

which is exactly the right side of the inequality. Hence for all $j \in \mathcal{S}$ we have

$$\sum_k \frac{1}{\mathbb{E}_k[T_k]} \cdot \mathbf{P}(k, j) = \frac{1}{\mathbb{E}_j[T_j]},$$

and thus $\tilde{\pi}(i) = \frac{1}{c} \cdot \frac{1}{\mathbb{E}_i[T_i]}$ is a stationary distribution. By the proof of the uniqueness part, if the chain does have a stationary distribution, then it has the unique stationary distribution π where $\pi(i) = 1/\mathbb{E}_i[T_i]$. So $c = 1$ and we complete our proof. □