Lecture 4: Markov Chains - Basics & Classification of States

1 Preliminaries

Definition 1.1. A discrete-time *Markov chain* is a discrete-time stochastic process $\{X_0, X_1, \dots\}$ whose state space I is countable (i.e. finite or enumerable), and for all $i, j \geq 0$ there is a given probability p_{ij} such that

$$P\big[X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots X_0 = i_0\big] = p_{ij},$$

for all $i_0, \dots, i_{n-1}, n \geq 0$.

The state space I is normally assumed to be a subset of $\mathbb{N}$ or $\mathbb{Z}$. Basically, if the current state at time n is i, then there is a fixed probability that the next state is j, no matter what states the process was in before time n. Hence,

$$p_{ij} \geq 0, \forall i, j \in I, \text{ and } \sum_{j \in I} p_{ij} = 1, \forall i \geq 0.$$

The matrix $\mathbf{P} = (P_{ij})$ is called the (one-step) *transition probability matrix*.

Definition 1.2 (Measures and distributions). A vector $\lambda = (\lambda_i : i \in I)$ consisting of coordinates $\lambda_i \geq 0, i \in I$, is called a *measure* on I. If the sum of all coordinates is 1, then it is called a *distribution*. Thus, each row of the transition probability matrix $\mathbf{P}$ of a Markov chain is a distribution.

It is easy to see that $\mathbf{P}^n$ is the n-step transition probability matrix; namely, $(\mathbf{P}^n)_{ij}$ is the probability $p_{ij}^{(n)}$ that the process from state i will be in state j after n transitions. To see that $p_{ij}^{(n)}$ is the same as $(\mathbf{P}^n)_{ij}$, we only need to observe that they both satisfy the *Chapman-Kolmogorov Equations*:

$$p_{ij}^{(m+n)} = \sum_{k=0}^{\infty} p_{ik}^{(n)} p_{kj}^{(m)}, \forall n, m \geq 0, i, j \in I.$$

Remark 1.3. If $\mathbf{P}$ is diagonalizable by a matrix S, i.e. $\mathbf{P} = S\Lambda S^{-1}$ for some diagonal matrix Λ consisting of $\mathbf{P}$'s eigenvalues on the diagonal, then the entries $(\mathbf{P}^n)_{ij}$ can easily be calculated since $\mathbf{P} = S\Lambda^n S^{-1}$, and the $(\Lambda^n)_{ii} = \lambda_{ii}^n$.

Given the distribution λ of X_0, we can easily calculate the distribution at time n:

$$\begin{aligned} P[X_n = j] &= \sum_{i \in I} P[X_n = j \mid X_0 = i] P[X_0 = i] \\ &= \sum_{i \in I} \lambda_i p_{ij}^{(n)}. \end{aligned}$$

In other words, X_n has distribution $\lambda \mathbf{P}^n$. Here, we think of λ as a row vector.

We shall often use the notation $P_i[F]$ to denote the probability of event F conditional on $X_0 = i$, i.e. $P_i[F] = P[F \mid X_0 = i]$.

2 Classification of states

Definition 2.1. Given a Markov chain $\mathbf{P}$, state j is said to be *reachable* from state i if $p_{ij}^{(n)} > 0$ for some $n \geq 0$. We also write $i \rightsquigarrow j$ to denote the fact that j is reachable from i. If $i \rightsquigarrow j$ and $j \rightsquigarrow i$, then states i and j are said to *communicate*. We use $i \leftrightarrow j$ to denote communication.

Proposition 2.2. *Communication is an equivalent relation.*

Proof. A property is equivalent if it is reflexive, symmetric and transitive. By default, $p_{ii}^{(0)} = 1 > 0$, hence $i \rightsquigarrow i$: reflexivity holds. If $i \rightsquigarrow j$, and $j \rightsquigarrow k$, then there are n and m such that $p_{ij}^{(n)} > 0$ and $p_{jk}^{(m)} > 0$. Thus,

$$p_{ik}^{(n+m)} = \sum_{j' \in I} p_{ij'}^{(n)} p_{j'k}^{(m)} \geq p_{ij}^{(n)} p_{jk}^{(m)} > 0,$$

implying $i \rightsquigarrow k$. Similarly, $k \rightsquigarrow j$ and $j \rightsquigarrow i$ imply $k \rightsquigarrow i$. Thus, the communication relation is transitive. The fact that it is also symmetric is trivial. $\square$

An equivalent relation partition the state space into *communication classes*, where two states are in the same class if and only if they communicate. We now investigate some class properties.

Definition 2.3. A (communication) class C is *closed* if for all $i \in C$, the fact that $i \rightsquigarrow j$ implies $j \in C$. A state i is *absorbing* if $\{i\}$ is a closed class by itself.

Example 2.4 (Gambler's Ruin Problem). A gambler goes to a casino where he bets 1 dollar every hand. He wins with probability p, and loses with probability $1-p$. He stops playing when he gets broke, or when he gets to N dollars. Figure 1 illustrates the process. In Figure 1, $\{0\}$, $\{N\}$, and $\{1, 2, ...N-1\}$

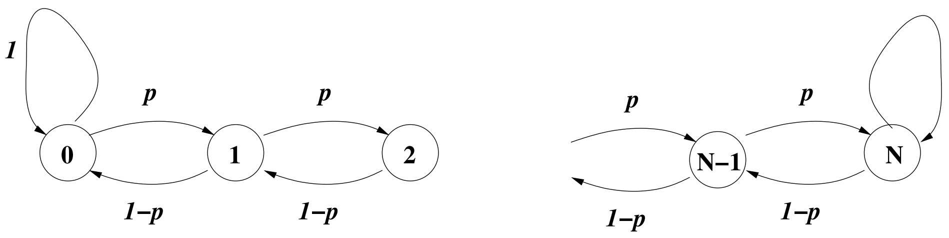

Figure 1: A Markov chain for the Gambler's ruin problem

are communication classes. States 0 and N are absorbing since the classes $\{0\}$ and $\{N\}$ are closed.

Denote by $f_{ij}^{(n)}$ the probability that the process visits state j from state i for the first time in n (≥ 1) steps, namely

$$f_{ij}^{(n)} = P[X_n = j \mid X_0 = i, X_k \neq j, k \in I]. \tag{1}$$

Then, the probability of ever getting from i to j is

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}, \tag{2}$$

and the mean number of steps required to reach j from i is

$$\mu_{ij} = \sum_{n=1}^{\infty} n f_{ij}^{(n)}. \tag{3}$$

Definition 2.5. A state is said to be *recurrent* if $f_{ii} = 1$ and *transient* if $f_{ii} < 1$. Furthermore, a recurrent state i is said to be *positive recurrent* if $\mu_{ii} < \infty$ and *null recurrent* if $\mu_{ii} = \infty$. The reader might also see *persistent* used instead of recurrent.

Example 2.6 (Positive Recurrence). For the example in Figure 2, let's calculate f_{11} and μ_{11}.

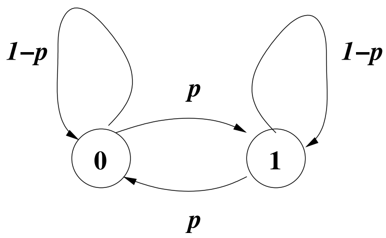

Figure 2: A Markov chain with two states

It is easy to see that

$$f_{11}^{(1)} = 1 - p, \quad f_{11}^{(2)} = p^2, \text{ and } f_{11}^{(3)} = p^2(1-p).$$

In general,

$$\begin{aligned} f_{11}^{(n)} &= p^2(1-p)^{(n-2)}, \\ f_{11} &= \sum_{n=1}^{\infty} f_{11}^{(n)} \\ &= (1-p) + p^2 \sum_{n=2}^{\infty} (1-p)^{n-2} \\ &= 1. \end{aligned}$$

Thus, state 1 is recurrent. Moreover,

$$\begin{aligned} \mu_{11} &= \sum_{n=1}^{\infty} n f_{11}^{(n)} \\ &= (1-p) + p^2 \sum_{n=2}^{\infty} n(1-p)^{n-2} \\ &= (1-p) + \frac{p^2}{(1-p)^2} \left(\sum_{n=1}^{\infty} n(1-p)^n - (1-p) \right) \\ &= 2. \end{aligned}$$

Thus state 1 is positive recurrent.

Example 2.7 (Null Recurrence). For the example in Figure 3, let's calculate f_{00} and μ_{00}.

$$f_{00}^{(n)} = \begin{cases} 0 & n = 1 \\ \frac{1}{n(n-1)} & n \geq 2 \end{cases}$$

Hence,

$$f_{00} = \sum_{n=1}^{\infty} f_{00}^{(n)} = \sum_{n=2}^{\infty} \frac{1}{n(n-1)} = \sum_{n=2}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n} \right) = 1.$$

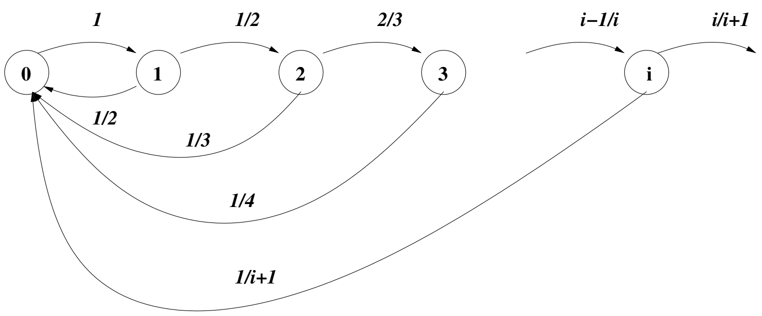

Figure 3: A null recurrent example

Thus state 0 is recurrent. Moreover, since

$$\mu_{00} = \sum_{n=1}^{\infty} n f_{00}^{(n)} = \sum_{n=2}^{\infty} \frac{1}{n-1} = \infty$$

state 0 is null recurrent.

Proposition 2.8. *Given a DTMC* $\mathbf{P}$*,*

(i) a state i is recurrent if and only if $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$,

(ii) a state i is transient if and only if $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$.

Proof. It is sufficient to show either (i) or (ii), and the other follows. We shall prove the later. Let V_i denote the number of visits to i starting from $X_0 = i$. Note that starting from i, the probability of visiting i exactly n times is $f_{ii}^{n-1}(1 - f_{ii})$. (Here, we make implicit use of the so-called *strong Markov property* to be introduced later.) Thus, V_i is a geometric random variable with parameter $(1 - f_{ii})$. The expected number of visits to a state i starting from i is then $E_i[V_i] = \frac{1}{1-f_{ii}} < \infty$.

Now, we calculate $E_i[V_i]$ in another way. Denote by I_n the random variable that indicates if the state of the process after n steps is i, namely

$$I_n = \begin{cases} 1 & X_n = i \\ 0 & \text{otherwise} \end{cases}$$

Then, we have

$$E_i[V_i] = E_i[\sum_{n=1}^{\infty} I_n] = \sum_{n=1}^{\infty} E_i[I_n] = \sum_{n=1}^{\infty} P_i[I_n = 1] = \sum_{n=1}^{\infty} P_i[X_n = i] = \sum_{n=1}^{\infty} p_{ii}^{(n)},$$

which completes the proof. ☐

Proposition 2.9. *Recurrence and transience are class properties. In other words, given a class C of a DTMC* $\mathbf{P}$*, if $i, j \in C$, and $i \leftrightarrow j$, then i and j have to be both transient or recurrent.*

Proof. Given communicating states i, j, we shall prove that j is transient if i is transient by showing that $\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$. Since i is transient, we know that

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty. \tag{6}$$

Since i and j communicate, we know that $\exists n_0, m_0$ such that $p_{ij}^{(n_0)} > 0$, and $p_{ji}^{(m_0)} > 0$. It is easy to see that

$$p_{ii}^{(n_0+n+m_0)} \geq p_{ij}^{(n_0)} p_{jj}^{(n)} p_{ji}^{(m_0)},$$

which implies

$$p_{jj}^{(n)} \leq \frac{p_{ii}^{(n_0+n+m_0)}}{p_{ij}^{(n_0)} p_{ji}^{(m_0)}}.$$

Summing the inequality above over all n, we get

$$\sum_{n=1}^{\infty} p_{jj}^{(n)} \leq \frac{1}{p_{ij}^{(n_0)} p_{ji}^{(m_0)}} \sum_{n=1}^{\infty} p_{ii}^{(n_0+n+m_0)} \leq \frac{1}{p_{ij}^{(n_0)} p_{ji}^{(m_0)}} \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty.$$

□

The next two results essentially say that recurrent classes and closed classes are the same for finite DTMCs.

Proposition 2.10. *Every recurrent class is closed.*

Proof. We will show this by contradiction. Given a recurrent class C, and state $i \in C$, we wish to show that if there is a state j and an integer n such that $p_{ij}^{(n)} > 0$ and $j \notin C$, then i cannot be recurrent.

The probability that a recurrent state i is visited exactly n times, starting from i, is $f_{ii}^{n-1}(1-f_{ii}) = 0$. Hence, a recurrent state is visited infinitely many times, or the probability that a recurrent state is visited infinite number of times is 1.

It is thus sufficient to show that there is a positive probability that i is visited only a finite number of times. Denote by $i\infty$ the event that i is visited an infinite number of times. We have

$$\begin{aligned} P_i[i\infty] &= \sum_{k \in I} P_i[i\infty \mid X_n = k] P_i[X_n = k] \\ &= P_i[i\infty \mid X_n = j] P_i[X_n = j] + \sum_{k \in I - \{j\}} P_i[i\infty \mid X_n = k] P_i[X_n = k] \end{aligned}$$

The first term in the right hand side is zero, since the process cannot return to i once it reaches j. Thus

$$\begin{aligned} P_i[i\infty] &= \sum_{k \in I - \{j\}} P_i[i\infty \mid X_n = k] P_i[X_n = k] \\ &\leq \sum_{k \neq j} p_{ik}^{(n)} \\ &= 1 - p_{ij}^{(n)} \\ &< 1 \end{aligned}$$

which is a contradiction, since we assumed that i is recurrent. □

Exercise 2.11. Show the previous proposition by showing that given a class C, and a state $i \in C$, if $i \rightsquigarrow j$ and $j \notin C$, then

$$f_{ii} < 1,$$

namely C is not recurrent.

Exercise 2.12. Show the previous proposition by showing that given a class C, and a state $i \in C$, if $i \rightsquigarrow j$ and $j \notin C$, then

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty,$$

namely C is not recurrent.

Proposition 2.13. *Every finite closed class is recurrent.*

Proof. If a finite closed class C is transient, then each state i in C can only be visited finite number t_i of times. But then at time $1 + \sum_{i \in C} t_i$ the process has to be somewhere. Contradiction! □

Infinite closed classes could be recurrent or transient. We shall give an example on this later.

3 Hitting time and mean hitting time

Definition 3.1. Given a set $A \subseteq I$, hitting time is defined as the minimum number of steps required for the process to be in any of the states in A:

$$H_A = inf\{n \geq 0 : X_n \in A\}.$$

Define h_i^A to be the probability that the process hits A from i:

$$h_i^A = P_i[H_A < \infty]$$

and μ_i^A the average number of steps before A is hit

$$\begin{aligned} \mu_i^A &= E[H_A] \\ &= \sum_{n=0}^{\infty} n P_i[H_A = n]. \end{aligned}$$

Lemma 3.2. *The h_i^A are the minimal non negative solution to the equations*

$$\begin{aligned} h_i^A &= 1, & \forall i \in A \\ h_i^A &= \textstyle\sum_{j \in I} p_{ij} h_j^A, & \forall i \notin A. \end{aligned}$$

Lemma 3.3. *The μ_i^A are the minimal non negative solution to the equations*

$$\begin{aligned} \mu_i^A &= 0, & \forall i \in A \\ \mu_i^A &= 1 + \textstyle\sum_{j \in I} p_{ij} \mu_j^A, & \forall i \notin A. \end{aligned}$$