SMSTC (2007/08)

Probability

www.smstc.ac.uk

Contents

SMSTC (2007/08)

Probability

Lecture 10: Markov chains in discrete time

Xuerong Mao, Strathclyde University[a]

www.smstc.ac.uk

Contents

[a] xuerong@stams.strath.ac.uk

10.1 Definition and basic properties

We now consider stochastic processes that move around on a countable (usually finite) state space S in discrete time. This means that at each time point an object moves from one position in the state space to another (or it may stay at the same position). Denote the position at time n by X_n for $n = 0, 1, \dots$ and $X_n \in S$. We shall assume the *Markov property* which is defined as follows.

Definition 10.1. The Markov property states that

$$\mathbf{P}(X_{n+m} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbf{P}(X_{n+m} = j \mid X_n = i) \qquad (10.1)$$

for all $n \geq 0$, $m \geq 1$ and $i, j \in S$. The stochastic process X_n with the Markov property is called a *Markov chain*.

Roughly the Markov property states that, given the state at time n, the behaviour after time n is independent of the behaviour before time n. In order to predict the future behaviour you need to know the current position, but information about how the process reached the current position (i.e. previous history) is of no further help.
Setting $m = 1$ in (10.1) gives

$$\mathbf{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbf{P}(X_{n+1} = j \mid X_n = i). \qquad (10.2)$$

The right-hand-side term is the probability that the Markov chain transits to state j at time $n+1$ given it is in state i at time n. Such a probability is called the *(one-step) transition probability* and is denoted by

$$p_{ij}(n) = \mathbf{P}(X_{n+1} = j \mid X_n = i).$$

If all the transition probabilities are independent of n, the Markov chain is said to be *stationary*. In this case, we can drop the n from $p_{ij}(n)$ and, clearly, we have

$$p_{ij} = \mathbf{P}(X_{n+1} = j \mid X_n = i) \quad \text{for all } i, j \in S,\ n \geq 0,$$

We shall consider only stationary Markov chains in this course.
The transition probabilities can be put into a square matrix, the rows and columns of which are indexed by the elements of S. Such matrices are called the *(one-step) transition matrices*.

Example 10.1. A very simple model for the weather from day to day. If it is raining today then the probability that it will rain tomorrow is 0.8 . If it is dry today, then the probability that it will rain tomorrow is 0.4 . The state space $S = \{\text{rain}, \text{dry}\}$. The transition matrix is

$$P = \begin{array}{c} \\ \text{rain} \\ \text{dry} \end{array} \begin{array}{c} \begin{array}{cc} \text{rain} & \text{dry} \end{array} \\ \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix} \end{array}.$$

Example 10.2. A maze used for training rats. There are 5 compartments labelled $1, \ldots, 6$ and connecting one or two-way doors. A rat moves in a Markov chain according to the transition matrix

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

(Because the labels are $1, \ldots, 5$ it is not necessary to label the rows and columns of the matrix.)

Example 10.3. Simple *random walk* with *absorbing barriers* at ± 2 (, also known as drunkard's ruin). Here the state space is $S = \{-2, -1, 0, 1, 2\}$. States ± 2 being absorbing barriers means that once the object reaches ± 2 it remains there. While the object is at one of the intermediate states, the probabilities of moving one step to the right/left/not moving are p, q, r, given non-negative numbers with sum 1. Here the transition matrix is

$$P = \begin{array}{c} \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccccc} -2 & -1 & 0 & 1 & 2 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ q & r & p & 0 & 0 \\ 0 & q & r & p & 0 \\ 0 & 0 & q & r & p \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

What do you notice about the transition matrices we have examined?
In fact all the entries are non-negative and each row sum is 1. Such a (square) matrix is called a *stochastic matrix*. Note also that, if we denote by $\mathbf{1}$ the column-vector with entries labelled by the elements of S and each equal to 1, then $P\mathbf{1} = \mathbf{1}$. This can be interpreted as $\mathbf{1}$ is a right eigenvector of P corresponding to the eigenvalue 1. (Recall that in general λ is an *eigenvalue* of the square matrix P if the equation $P\mathbf{x} = \lambda\mathbf{x}$ has a solution $\mathbf{x} \neq \mathbf{0}$. Then $\mathbf{x}$ is a corresponding right eigenvector.) So stochastic matrices always have the eigenvalue 1.

Now consider what happens in two steps starting from state i at time n.

$$\begin{aligned} p_{ij}^{(2)} &= \mathbf{P}(X_{n+2}=j \mid X_n=i) = \sum_{k\in S}\mathbf{P}(X_{n+2}=j \quad \& \quad X_{n+1}=k \mid X_n=i) \\ &= \sum_{k\in S}\mathbf{P}(X_{n+1}=k \mid X_n=i)\times\mathbf{P}(X_{n+2}=j \mid X_{n+1}=k \quad \& \quad X_n=i) \\ &= \sum_{k\in S}\mathbf{P}(X_{n+1}=k \mid X_n=i)\times\mathbf{P}(X_{n+2}=j \mid X_{n+1}=k) \\ &= \sum_{k\in S} p_{ik}p_{kj} = (P^2)_{ij}. \end{aligned}$$

So P^2 is the transition matrix that describes movements over two time units. Similarly P^n is the matrix of n-step transition probabilities.

One of the important objects in the study of Markov chains is to study the limiting behaviour of P^n as $n\to\infty$. Before we discuss this property, let us discuss how to compute various probabilities in terms of transition probabilities.

The transition probabilities describe movements of the Markov chain from one state to another. However, this is not enough to specify the probabilistic behaviour (or law) of the process $\{X_n\}_{n\geq 0}$. For this purpose, let us define the *initial distribution*

$$\lambda_i = \mathbf{P}(X_0=i), \quad i\in S.$$

We form these as a row vector $\lambda=(\lambda_i)_{i\in S}$. For example, if $S=\{0,1,2,\dots,N\}$, then

$$\lambda = (\lambda_0,\lambda_1,\lambda_2,\dots,\lambda_N).$$

Let us now explain that the transition matrix P and the initial distribution λ enable us to find, at least in principle, any probability connected with the process, such such $\mathbf{P}(X_n = i)$ or $\mathbf{P}(X_0 = i_0, \dots, X_n = i_n)$. Indeed,

$$\begin{aligned}\mathbf{P}(X_n = i) &= \sum_{k\in S} \mathbf{P}(X_0 = k \quad \& \quad X_n = i) \\ &= \sum_{k\in S} \mathbf{P}(X_0 = k)\mathbf{P}(X_n = i \mid X_0 = k) \\ &= \sum_{k\in S} \lambda_k p_{ki}^{(n)}.\end{aligned}$$

That is, in the matrix form, the probability distribution of X_n is given

$$(\mathbf{P}(X_n = i))_{i\in S} = \lambda P^n. \tag{10.3}$$

Moreover, compute the joint probability

$$\begin{aligned} & \mathbf{P}(X_0 = i_0,\ X_1 = i_1,\ \dots, X_{n-1} = i_{n-1},\ X_n = i_n) \\ = \ & \mathbf{P}(X_0 = i_0,\ X_1 = i_1,\ \dots, X_{n-1} = i_{n-1}) \\ & \times \mathbf{P}(X_n = i_n \mid X_0 = i_0,\ X_1 = i_1,\ \dots, X_{n-1} = i_{n-1}) \\ = \ & \mathbf{P}(X_0 = i_0,\ X_1 = i_1,\ \dots, X_{n-1} = i_{n-1})\mathbf{P}(X_n = i_n \mid X_{n-1} = i_{n-1}) \\ = \ & \mathbf{P}(X_0 = i_0,\ X_1 = i_1,\ \dots, X_{n-1} = i_{n-1}) p_{i_{n-1} i_n}.\end{aligned}$$

Repeating this procedure gives

$$\begin{aligned} & \mathbf{P}(X_0 = i_0,\ X_1 = i_1,\ \ldots, X_{n-1} = i_{n-1},\ X_n = i_n) \\ = \ & \mathbf{P}(X_0 = i_0) p_{i_0 i_1} \cdots p_{i_{n-1} i_n} \\ = \ & \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}. \end{aligned}$$

Similarly, the conditional probability

$$\begin{aligned} & \mathbf{P}(X_1 = i_1,\ \ldots, X_{n-1} = i_{n-1},\ X_n = i_n \mid X_0 = i_0) \\ & = p_{i_0 i_1} \cdots p_{i_{n-1} i_n}. \end{aligned}$$

These show that once the initial distribution λ and the transition matrix P are given, the probability distributions of the Markov chain $\{X_n\}_{n\geq 0}$ are determined. From now on, we will say that $\{X_n\}_{n\geq 0}$ is Markov (λ, P).

A particular interesting case is that the chain starts from a state i with probability 1, namely $P(X_0 = i) = 1$ but $P(X_0 = j) = 0$ if $j \neq i$. We denote this initial distribution by $\delta_i = (\delta_{ij})_{j\in S}$, where $\delta_{ij} = 1$ if $j = i$ and 0 otherwise. In this case, $\{X_n\}_{n\geq 0}$ is Markov (δ_i, P).

Let $\{X_n\}_{n\geq 0}$ be Markov (λ, P). In the case where $\lambda_i > 0$ we shall write $\mathbf{P}_i(A)$ for the conditional probability $\mathbf{P}(A \mid X_0 = i)$. By the Markov property at time 0, under $\mathbf{P}_i$, $\{X_n\}_{n\geq 0}$ is Markov (δ_i, P). So the behaviour of $\{X_n\}_{n\geq 0}$ under $\mathbf{P}_i$ does not depend on λ. More generally, we have the following theorem.

Theorem 10.1. *Let $\{X_n\}_{n\geq 0}$ be Markov (λ, P). Then, conditional on $X_m = i$, $\{X_{m+n}\}_{n\geq 0}$ is Markov (δ_i, P) and is independent of the random variables $X_0, \ldots, X_m$.*

10.2 Strong Markov property

Theorem 10.1 says that for each time m, conditional on $X_m = i$, the process after time m begins afresh from i. Suppose, instead of conditioning on $X_m = i$, we simply wait for the process to hit state i at some random time T_i. What can one say about the process after time T_i. What if we replaced T_i by a more general random time? In this section we shall identify a class of random times at which a version of the Markov property does hold.

A random variable $\tau : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ is called a *stopping time* or *Markov time* if the event $\{\tau = n\}$ depends only on $X_0, X_1, \ldots, X_n$ for $n = 0, 1, 2, \ldots$.

Let A be a subset of S. The *hitting time* H_A of A is

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

where we set $\inf \emptyset = \infty$ as usual. This is a stopping time because

$$\{H_A = n\} = \{X_0 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\}.$$

Moreover, the *first passage time*

$$T_j = \inf\{n \geq 1 : X_n = j\}$$

is a stopping time because

$$\{T_j = n\} = \{X_1 \neq j, \ldots, X_{n-1} \neq j, X_n = j\}.$$

The following theorem shows the *strong Markov property*—the Markov property holds at stopping times.

Theorem 10.2. ***(Strong Markov property)*** *Let $\{X_n\}_{n\geq 0}$ be Markov (λ, P) and let τ be a stopping time of $\{X_n\}_{n\geq 0}$. Then, conditional on $\tau < \infty$ and $X_\tau = i$, $\{X_{\tau+n}\}_{n\geq 0}$ is Markov (δ_i, P) and is independent of $X_0, \ldots, X_\tau$.*

Proof Let $B \subset \Omega$ be an event determined by $X_0, \ldots, X_\tau$. Clearly, $B \cap \{\tau = m\}$ is determined by $X_0, \ldots, X_m$. So, by the Markov property at time m,

$$\begin{aligned} &\mathbf{P}\big(\{X_{\tau+1} = j_1, \ldots, X_{\tau+n} = j_n\} \cap B \cap \{\tau = m\} \cap \{X_\tau = i\}\big) \\ =\ &\mathbf{P}\big(\{X_{m+1} = j_1, \ldots, X_{m+n} = j_n\} \mid B \cap \{\tau = m\} \cap \{X_m = i\}\big)\mathbf{P}\big(B \cap \{\tau = m\} \cap \{X_\tau = i\}\big) \\ =\ &\mathbf{P}_i(X_1 = j_1, \ldots, X_n = j_n)\mathbf{P}\big(B \cap \{\tau = m\} \cap \{X_\tau = i\}\big). \end{aligned}$$

Now sum over $m = 0, 1, \ldots$ and divide by $\mathbf{P}(\tau < \infty, X_\tau = i)$ to obtain

$$\begin{aligned} &\mathbf{P}\big(\{X_{\tau+1} = j_1, \ldots, X_{\tau+n} = j_n\} \cap B \mid \tau < \infty, X_\tau = i\}\big) \\ &= \mathbf{P}_i(X_1 = j_1, \ldots, X_n = j_n)\mathbf{P}\big(B \mid \tau < \infty, X_\tau = i\big) \end{aligned}$$

as required. □

10.3 Classification of states

It is sometimes possible to break a Markov chain into smaller pieces, each of which is relatively easy to understand, and which together give an understanding of the whole. This is done by identifying the *irreducible closed classes*.

Definition 10.2.
- If $p_{ii} = 1$, then i is an *absorbing state*.
- A non-empty subset C of the state space is called a *closed class* if it is not possible to leave C starting from a state in C, i.e. if $p_{ij} = 0$ for all states $i \in C$ and $j \notin C$.
- An *irreducible closed class* C is a closed class such that no proper subset of C is itself closed.
- A Markov chain is *irreducible* if S is an irreducible closed class, i.e. if there is no closed class other than S itself.

To identify irreducible closed classes, we introduce the concept of *communicating states*. We say that i reaches j if there is an n such that $p_{ij}^{(n)} > 0$. If i and j reach each other, we say that i and j communicate. Clearly, all states in an irreducible closed class communicate. To determine the closed classes, we just need to know which elements of P are positive and which 0.

Example 10.1 is irreducible; in Example 10.2, $\{4\}, \{5\}$ form two irreducible closed classes; in Example 10.3, $\{-2\}, \{2\}$ are irreducible closed classes.

Example 10.4. Stock control. Suppose that ordering of new stock takes place at the end of each week with a policy of not placing an order if there are one or more item remaining in stock and, when no items remain in stock, of ordering sufficient to meet any unfilled orders and make the stock up to three. The weekly demand has a known probability distribution; 0, 1, 2, or 3 with probability 0.2, 0.4, 0.3, 0.1 respectively. The demands in different weeks are independent. We shall define the state as the number of items in stock at the end of the week, counting unfilled orders as negative. The the state space is $S = \{-2, -1, 0, 1, 2, 3\}$. For example State -2 occurs if there was one item in stock at the end of the previous week and three items have been demanded. The transition matrix is

$$\begin{array}{c} \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\!\!\begin{array}{c} \begin{array}{cccccc} -2 & -1 & 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 0 & .1 & .3 & .4 & .2 \\ 0 & 0 & .1 & .3 & .4 & .2 \\ 0 & 0 & .1 & .3 & .4 & .2 \\ .1 & .3 & .4 & .2 & 0 & 0 \\ 0 & .1 & .3 & .4 & .2 & 0 \\ 0 & 0 & .1 & .3 & .4 & .2 \end{pmatrix} \end{array}.$$

This Markov chain is irreducible. We might want to answer questions like: how often is an order placed, how often are you unable to supply an item from stock, what is the average amount of stock held?

Example 10.5. Success runs. A coin is tossed independently until five heads in succession have been obtained. We denote the state here as the number of consecutive heads (up to 5) that have been obtained on the most recent tosses. S is $\{0, \dots, 5\}$. Suppose that the probability of a head is p at each toss and denote $1-p$ by q. The transition matrix is

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\!\!\begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} q & p & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 & 0 \\ q & 0 & 0 & p & 0 & 0 \\ q & 0 & 0 & 0 & p & 0 \\ q & 0 & 0 & 0 & 0 & p \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Here the state 5 is an absorbing state which forms the unique irreducible closed class $\{5\}$. We might want to answer question: How long does it take on average in order to obtain five heads in succession?

Example 10.6. Ehrenfest urn model for the diffusion of molecules of a gas through a membrane. There are N particles in a container which has a permeable partition. At each time point one of the particles chosen at random passes through the partition. Here we record the state as the number of particles to one side of the partition, so $S = \{0, \ldots, N\}$. When the state is i, there is probability i/N

that the next state is $i-1$, and probability $(N-i)/N$ that it is $i+1$. The transition matrix is

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ N \end{array} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & \dots & N-1 & N \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ \frac{1}{N} & 0 & \frac{N-1}{N} & 0 & \dots & 0 & 0 \\ 0 & \frac{2}{N} & 0 & \frac{N-2}{N} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix} \end{array}.$$

This is an irreducible Markov chain.

This example illustrates one further phenomenon which occurs — *periodicity*. In Example 10.6 the states that occur are alternately even and odd. If the process starts for example in state 0 then after an even number of steps it cannot avoid being in an even-numbered state, while after an odd number of steps it must be in an odd-numbered state. (We will see that this complicates the description of the limiting behaviour of P^n. In fact there are two limit matrices, one for the even powers and the other for the odd powers of P.)

We say that a state i has *period* d if the highest common factor (hcf) of $\{n : p_{ii}^{(n)} > 0\}$ is d. This means that if the chain is in state i at time n it can only return there at times of the form $n + kd$ for some integer k. If $p_{ii}^{(n)} = 0$ for all n, we say that state i has infinite period. A state with period 1 is called *aperiodic* (i.e. it does not have a period). It can be shown that all states in an irreducible closed class have the same period. Hence, for an irreducible Markov chain, all states have the same period and we say the chain has the period. If the period is 1 we call the Markov chain *aperiodic*. Once again to recognise periodicity it is necessary only to know which entries of P are positive and which 0. It is possible to invent Markov chains with any period (Try it.); however any period greater than 2 will usually be obvious. Where the period is 2, the states can be put into two classes (even/odd, white/black, etc.) and all steps of the process are from one class into the other, so that the process alternates between the two classes. Sometimes you will have to think a bit to see whether this is possible.

Let us now define two more important concepts.

Definition 10.3. A state i is *transient* if the probability starting from i of never returning to i is positive. A state i is *recurrent* if the probability of sooner or later returning to i starting from i is 1.

Obviously, a state i is transient if there is a state j which can be reached from i in one or more steps from which it is not possible to get back to i.

To establish more criteria on recurrence and transience, we let

$$f_{ij}^{(n)} = \mathbf{P}_i(T_j = n)$$

be the *first passage distribution* from state i to state j. We have

$$f_{ij}^{(n)} = \mathbf{P}_i(X_n = j,\ X_k \neq j,\ k = 1, \ldots, n-1).$$

Define

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} = \mathbf{P}_i(T_j < \infty).$$

From the definition, we see that the state i is recurrent if and only if

$$f_{ii} = \mathbf{P}_i(T_i < \infty) = 1,$$

and it is transient if and only if $f_{ii} < 1$. In particular, every state is either transient or recurrent. Let us also define inductively the *rth passage time* $T_i^{(r)}$ to state i by

$$T_i^{(0)} = 0, \quad T_i^{(1)} = T_i$$

and, for $r = 1, 2, \ldots$,

$$T_i^{(r+1)} = \inf\{n \geq T_i^{(r)} + 1 : X_n = i\}.$$

The length of the rth excursion to i is then

$$\tau_i^{(r)} = \begin{cases} T_i^{(r)} - T_i^{(r-1)} & \text{if } T_i^{(r-1)} < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 10.1. *For $r = 2, 3, \ldots$, conditional on $T_i^{(r-1)}$, $\tau_i^{(r)}$ is independent of $\{X_m : m \leq T_i^{(r-1)}\}$ and*

$$\mathbf{P}(\tau_i^{(r)} = n \mid T_i^{(r-1)} < \infty) = \mathbf{P}_i(T_i = n).$$

Proof This is a simple application of the strong Markov property at the stopping time $\tau = T_i^{(r-1)}$. Clearly, $X_\tau = i$ if $\tau < \infty$. So, conditional on $\tau < \infty$, $\{X_{\tau+n}\}_{n\geq 0}$ is Markov (δ_i, P) and independent of $X_0, X_1, \ldots, X_\tau$. But

$$\tau_i^{(r)} = \inf\{n \geq 1 : X_{\tau+n} = i\},$$

so $\tau_i^{(r)}$ is the first passage time of $\{X_{\tau+n}\}_{n\geq 0}$ to state i. □

Let us introduce the *number of visits* to state i:

$$V_i = \sum_{n=0}^{\infty} 1_{\{X_n = i\}}.$$

Note that

$$\mathbf{E}_i(V_i) = \mathbf{E}_i \sum_{n=0}^{\infty} 1_{\{X_n=i\}} = \sum_{n=0}^{\infty} \mathbf{E}_i(1_{\{X_n=i\}}) = \sum_{n=0}^{\infty} \mathbf{P}_i(X_n = i) = \sum_{n=0}^{\infty} p_{ii}^{(n)}.$$

Lemma 10.2. *For $r = 0, 1, \ldots$, we have that $\mathbf{P}_i(V_i > r) = (f_{ii})^r$.*

Proof We prove it by induction. When $r = 0$ the assertion holds clearly. Suppose inductively that the assertion holds for some $r \geq 0$. Observing that if $X_0 = i$ then $\{V_i > r+1\} = \{T_i^{(r+1)} < \infty\}$, we compute

$$\begin{aligned} \mathbf{P}_i(V_i > r+1) &= \mathbf{P}_i(T_i^{(r+1)} < \infty) \\ &= \mathbf{P}_i(T_i^{(r)} < \infty \text{ and } \tau_i^{(r+1)} < \infty) \\ &= \mathbf{P}_i(\tau_i^{(r+1)} < \infty \mid T_i^{(r)} < \infty)\mathbf{P}_i(T_i^{(r)} < \infty) \\ &= f_{ii}(f_{ii})^r = (f_{ii})^{r+1} \end{aligned}$$

by Lemma 10.1. So by induction the assertion holds for all r. □

The following theorem give criteria for recurrence and transience in terms of transition probabilities.

Theorem 10.3. *We have :*

(i) *A state i is recurrent if and only if $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$.*

(ii) *A state i is transient if and only if $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$.*

Proof If i is recurrent, then $f_{ii} = 1$ and, by Lemma 10.2,

$$\mathbf{P}_i(V_i = \infty) = \lim_{r\to\infty} \mathbf{P}_i(V_i > r) = 1$$

which implies

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \mathbf{E}_i(V_i) = \infty.$$

If i is transient, then $f_{ii} < 1$. Recall that one can compute the expectation of a non-negative integer-valued random variable as follows:

$$\sum_{r=0}^{\infty} \mathbf{P}(\xi > r) = \sum_{r=0}^{\infty} \sum_{v=r+1}^{\infty} \mathbf{P}(\xi = v)$$
$$= \sum_{v=1}^{\infty} \sum_{r=0}^{v-1} \mathbf{P}(\xi = v) = \sum_{v=1}^{\infty} v\mathbf{P}(\xi = v) = \mathbf{E}(\xi).$$

We then compute, by Lemma 10.2, that

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \mathbf{E}_i(V_i) = \sum_{r=0}^{\infty} \mathbf{P}_i(V_i) = \sum_{r=0}^{\infty} (f_{ii})^r = \frac{1}{1 - f_{ii}} < \infty$$

as required. □

In the proof above, we also see clearly that a state i is transient if and only if

$$\mathbf{P}_i(X_n = i \text{ for infinitely many } n) = 0$$

and that a state i is recurrent if and only if

$$\mathbf{P}_i(X_n = i \text{ for infinitely many } n) = 1.$$

Thus a recurrent state is one to which the chain keeps coming back and a transient states is one which the chain eventually leaves for ever. The following theorem shows that recurrence and transience are *class properties*.

Theorem 10.4. *Let C be an irreducible closed class. Then either all states in C are transient or all are recurrent.*

Proof Take any pair of states $, j \in C$ and suppose that i is transient. There exist $n, m > 0$ with $p_{ij}^{(n)} > 0$ and $p_{ji}^{(m)} > 0$. Noting that for all $r \geq 0$,

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

we have

$$\sum_{r=0}^{\infty} p_{jj}^{(r)} \leq \frac{1}{p_{ij}^{(n)} p_{ji}^{(m)}} \sum_{r=0}^{\infty} p_{ii}^{(r)} < \infty$$

by Theorem 10.3. Hence j is also transient by Theorem 10.3 again. □

In the light of this theorem it is natural to speak of a recurrent or transient class.

Theorem 10.5. *Every finite irreducible closed class is recurrent.*

Proof Suppose C is a finite irreducible closed class and the chain starts in C. Then there must exist some $i \in C$ for which we have

$$\begin{aligned} 0 &< \mathbf{P}(X_n = i \text{ for infinitely many } n) \\ &= \mathbf{P}(X_n = i \text{ for some } n)\mathbf{P}_i(X_n = i \text{ for infinitely many } n) \end{aligned}$$

by the strong Markov property. This shows that i is recurrent by Theorem 10.3 and hence C is recurrent by Theorem 10.4. □

In particular, if the state space is finite, then we have:

- the states which are not in any irreducible closed class are all transient;
- all states in an irreducible closed class are recurrent.

Example 10.1 is irreducible and both states are recurrent; in Example 10.2, $\{5\}$, $\{6\}$ are irreducible closed classes, states $4, 5$ are recurrent and states $1, 2, 3$ are transient; in Example 10.3, $\{-2\}, \{2\}$ are irreducible closed classes, states $-2, 2$ are recurrent and states $-1, 0, 1$ are transient; in Example 10.4, all states are recurrent; in Example 10.5, the state 5 is an absorbing state and the other states are transient; in Example 10.6, all states are recurrent.

For a recurrent state $f_{ii}^{(n)}$ is a probability distribution of the first passage time T_i with mean $\mu_i = \mathbf{E}_i(T_i) = \sum_{n=1}^{\infty} n f_{ii}^{(n)}$, the *mean recurrence time*. If $\mu_i = \infty$, state i is called *null*, otherwise it is called *positive*.

10.4 Stationary distribution

Definition 10.4. A *stationary distribution* is a probability distribution π on S, thought of as a row-vector, such that $\pi P = \pi$. It can also be called an *equilibrium* or *steady state* distribution.

Note that π is a left eigenvector corresponding to the eigenvalue 1. It follows from matrix theory that a Markov chain with finitely many states must have a left eigenvector for the eigenvalue 1. (It is not so elementary to show that this eigenvector can be take to have its entries all non-negative, so that division by a suitable constant gives a probability distribution.)
When we use the stationary distribution as initial distribution, namely $\{X_n\}_{n\geq 0}$ is Markov (π, P), we see from (10.3) that the probability distribution of X_n

$$(\mathbf{P}(X_n = i))_{i\in S} = \pi P^n = \pi \quad \text{for all } n.$$

We then say that the distributions of $\{X_n\}_{n\geq 0}$ are *time invariant* (another name for this is stationary). Therefore π is also known as the *stationary initial distribution*.
Which Markov chains have a stationary distribution? We will restrict attention to irreducible chains, since any other chain can be decomposed into irreducible subclasses.

Theorem 10.6. *An irreducible Markov chain has a stationary distribution if and only if it is positive recurrent. The stationary distribution is unique and given by* $\pi_i = \mu_i^{-1}$.

The proof of this is not given here but can be found in e.g. [2]. When the state space is finite, the stationary distribution can be obtained by solving linear equations. We look at some of the examples.

In Example 10.1, the Markov chain has a stationary distribution $\pi = (\pi_r, \pi_d)$ which obeys

$$\pi = \pi P \quad \text{subject to} \quad \pi_r + \pi_d = 1,$$

namely

$$\pi_r = 0.8\pi_r + 0.4\pi_d, \quad \pi_d = 0.2\pi_r + 0.6\pi_d, \quad \pi_r + \pi_d = 1.$$

Solving these equations gives $\pi_r = 2/3$ and $\pi_d = 1/3$. In other words, we obtain the stationary distribution $\pi = (2/3, 1/3)$.

In Example 10.4, let $\pi = (\pi_{-2}, \pi_{-1}, \pi_0, \pi_1, \pi_2, \pi_3)$ be the stationary distribution. Then

$$\pi = \pi P \quad \text{subject to} \quad \pi_{-2} + \pi_{-1} + \pi_0 + \pi_1 + \pi_2 + \pi_3 = 1.$$

Solving these equations gives $\pi = (5,\ 19,\ 40,\ 50,\ 40,\ 16)/170$.

In the case when a Markov chain has only one irreducible closed class C, first identify the periodicity for the states in the closed class. Then, solve the equations $\mathbf{y} = \mathbf{y}P_C$ with the sum of coefficients equal to 1, where P_C is the transition matrix on the class C. Finally, form the stationary distribution $\pi = (\pi_i)_{i \in S}$ by setting $\pi_i = y_i$ if $i \in C$ or otherwise $\pi_i = 0$.

Example 10.5 has one absorbing state 5, and the other states are transient. State 5 forms the only irreducible closed class $C = \{5\}$. The stationary distribution on this closed class is obviously $y_5 = 1$. Thus, the stationary distribution for the Markov chain is $\pi = (0,\ 0,\ 0,\ 0,\ 0,\ 1)$.

In Example 10.2, the states 5,6 form the only irreducible closed class $C = \{5, 6\}$, and the other states are transient. The corresponding

$$P_C = \begin{pmatrix} .5 & .5 \\ 1 & 0 \end{pmatrix}.$$

Let $\mathbf{y} = (y_5, y_6)$. Solving

$$\mathbf{y} = \mathbf{y}P_C \quad \text{subject to} \quad y_5 + y_6 = 1$$

gives $\mathbf{y} = (2/3, 1/3)$. Hence the stationary distribution for the Markov chain is $(0,\ 0,\ 0,\ 0,\ 2,\ 1)/3$. and each row of P^n tends to $(0,\ 0,\ 0,\ 0,\ 2,\ 1)/3$.

Example 10.6 is irreducible but with period 2. In the case when $n = 5$, it can be shown that the stationary distribution is the Binomial distribution $B(5,\ \frac{1}{2}) = (1, 5, 10, 10, 5, 1)/32$.
Equation $\pi = \pi P$ for the stationary distribution can be written

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}. \tag{10.4}$$

But, obviously,

$$\pi_j = \pi_j \sum_{i \in S} p_{ji} = \sum_{i \in S} \pi_j p_{ji}.$$

So

$$\sum_{i \in S} \pi_j p_{ji} = \sum_{i \in S} \pi_i p_{ij}. \tag{10.5}$$

We can interpret $\sum_{i \in S} \pi_j p_{ji}$ as the probability flux out of state j, and $\sum_{i \in S} \pi_i p_{ij}$ as the probability flux into state j. In this interpretation, it is natural to think of (10.5) as an equation of *full balance*. We observe that (10.5) holds if

$$\pi_j p_{ji} = \pi_i p_{ij}. \tag{10.6}$$

This is called the law of *detailed balance*, stating that the probability flux from i to j in equilibrium is the same as that from j to i.

10.5 Long term behaviour

Many physical systems tend to settle down to an *equilibrium state* regardless of its initial state. Let $\{X_n\}_{n\geq 0}$ be Markov (λ, P). Under suitable conditions

$$P^n \to \begin{pmatrix} \pi \\ \pi \\ \vdots \\ pi \end{pmatrix}. \tag{10.7}$$

We say that the chain has a *limiting distribution.* What this means is that if the chain is left running for a long time, it reaches an equilibrium situation regardless of its initial distribution. In this equilibrium situation the state occupancy probabilities are equal to the stationary distribution. Note namely that

$$(\mathbf{P}(X_n = i))_{i\in S} = \lambda P^n \to \lambda \begin{pmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{pmatrix} = \pi$$

regardless of λ. As the next example shows, there may be a stationary distribution without the chain having a limiting distribution.

Example 10.7. Let

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Then $\pi = (1, 1, 1)/3$, but P^n does not converge. Rather, it cycles through three different matrices. Notice that the period of this chain is 3.

Here is the main result of this section.

Theorem 10.7. *Let $\{X_n\}_{n \geq 0}$ be irreducible Markov (λ, P) and suppose that it has a stationary distribution π. If the chain is aperiodic, then each row of P^n tends to π. If the chain is periodic, then P^n does not tend to a limit but π still represents the proportion of time that is spent in the various states in the long run.*

The proof can be found in e.g. [3]. Let us discuss some examples.
In Example 10.1, we have

$$\lim_{n \to \infty} P^n = \begin{pmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{pmatrix}$$

In the long run it rains on 2/3 of the days.
In Example 10.4, each row of P^n tends to $(5,\ 19,\ 40,\ 50,\ 40,\ 16)/170$. So the states $-2, -1$, which correspond to having unfilled orders, occur in the long run in 24/170 of the weeks.
In Example 10.5, each row of P^n tends to $(0,\ 0,\ 0,\ 0,\ 0,\ 1)$.
In Example 10.2, and each row of P^n tends to $(0,\ 0,\ 0,\ 0,\ 2,\ 1)/3$.

In Example 3, there are two absorbing states. We are not yet in a position to give the limit of P^n. However we can say that the three central columns of P^n all tend to 0 because the corresponding states are all transient. We shall find the probabilities of absorption at 2 starting from the intermediate states later.

Example 6 is irreducible but with period 2. The case of $n = 5$ will be illustrated. It can be shown that the stationary distribution is the Binomial distribution $B(5,\ \frac{1}{2}) = (1, 5, 10, 10, 5, 1)/32$. Recall that this indicates the proportion of time that is spent in the various states. Suppose first that the process starts in state 0. The after an even number of steps it must be in an even-numbered state, while after an odd number of steps it cannot be in an even-numbered state. So the probability is *twice* the probability in the equilibrium distribution, but for the even-numbered states only. This gives

$$16P^{2n} \to \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 1 & 0 & 10 & 0 & 5 & 0 \\ 0 & 5 & 0 & 10 & 0 & 1 \\ 1 & 0 & 10 & 0 & 5 & 0 \\ 0 & 5 & 0 & 10 & 0 & 1 \\ 1 & 0 & 10 & 0 & 5 & 0 \\ 0 & 5 & 0 & 10 & 0 & 1 \end{pmatrix} \end{array}.$$

The limit of the odd powers of P is similar except that the two kinds of row are interchanged.

Note that we can reorder the rows and columns to get the following forms:

$$16P^{2n} \to \begin{array}{c} \\ 0 \\ 2 \\ 4 \\ 1 \\ 3 \\ 5 \end{array}\!\!\begin{array}{c} \begin{array}{cccccc} 0 & 2 & 4 & 1 & 3 & 5 \end{array} \\ \begin{pmatrix} 1 & 10 & 5 & 0 & 0 & 0 \\ 1 & 10 & 5 & 0 & 0 & 0 \\ 1 & 10 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 10 & 1 \\ 0 & 0 & 0 & 5 & 10 & 1 \\ 0 & 0 & 0 & 5 & 10 & 1 \end{pmatrix} \end{array} \qquad 16P^{2n+1} \to \begin{array}{c} \\ 0 \\ 2 \\ 4 \\ 1 \\ 3 \\ 5 \end{array}\!\!\begin{array}{c} \begin{array}{cccccc} 0 & 2 & 4 & 1 & 3 & 5 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 5 & 10 & 1 \\ 0 & 0 & 0 & 5 & 10 & 1 \\ 0 & 0 & 0 & 5 & 10 & 1 \\ 1 & 10 & 5 & 0 & 0 & 0 \\ 1 & 10 & 5 & 0 & 0 & 0 \\ 1 & 10 & 5 & 0 & 0 & 0 \end{pmatrix} \end{array}.$$

10.6 Ergodic theorems

Ergodic theorems concern the limiting behaviour of averages over time. Denote by $V_i(n)$ the number of visits to i before n:

$$V_i(n) = \sum_{k=0}^{n-1} 1_{\{X_k=i\}}.$$

Then $V_i(n)/n$ is the proportion of time before n spent in state i. The following ergodic theorem gives the long-run proportion of time spent by a Markov chain in each state.

Theorem 10.8. *If* $\{X_n\}_{n\geq 0}$ *is irreducible Markov* (λ, P)*, then*

$$\mathbf{P}\Big(\frac{V_i(n)}{n} \to \frac{1}{\mu_i} \text{ as } n \to \infty\Big) = 1,$$

where $\mu_i = E_i(T_i)$ *is the expected return time to state* i*. Moreover, in the positive recurrent case, for any bounded function* $f : S \to \mathbb{R}$ *we have*

$$\mathbf{P}\Big(\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \bar{f} \text{ as } n \to \infty\Big) = 1,$$

where

$$\bar{f} = \sum_{i\in S} \pi_i f(i)$$

and where $(\pi_i)_{i\in S}$ *is the unique stationary distribution.*

Proof If the chain is transient, then for any state i, $\mu_i = \infty$ while the total number V_i of visits to i is finite with probability 1, so

$$0 \le \frac{V_i(n)}{n} \le \frac{V_i}{n} \to 0 = \frac{1}{\mu_i}.$$

Suppose that the chain is recurrent and fix any state i. Without loss of generality we may assume that $\lambda = \delta_i$; otherwise consider $\{X_{T_i+n}\}_{n\ge 0}$ which is Markov (δ_i, P) by the strong Markov property. Recall $\tau_i^{(r)}$, the length of the rth excursion to i, as defined before. By Lemma 10.1, the non-negative random variables $\tau_i^{(1)}, \tau_i^{(2)}, \ldots$ are i.i.d. with $\mathbf{E}_i(\tau_i^{(r)}) = \mu_i$. It is easy to see that

$$\tau_i^{(1)} + \cdots + \tau_i^{(V_i(n)-1)} \le n - 1,$$

the left-hand side being the time of the last visit to i before n, while

$$\tau_i^{(1)} + \cdots + \tau_i^{(V_i(n))} \ge n,$$

the left-hand side being the time of the first visit to i after $n-1$. Hence

$$\frac{\tau_i^{(1)} + \cdots + \tau_i^{(V_i(n)-1)}}{V_i(n)} \le \frac{n}{V_i(n)} \le \frac{\tau_i^{(1)} + \cdots + \tau_i^{(V_i(n))}}{V_i(n)}. \tag{10.8}$$

By the strong law of large numbers

$$\mathbf{P}\Big(\frac{\tau_i^{(1)} + \cdots + \tau_i^{(n)}}{n} \to \mu_i \text{ as } n \to \infty\Big) = 1$$

and, since the chain is recurrent,

$$\mathbf{P}(V_i(n) \to \infty \text{ as } n \to \infty) = 1.$$

So letting $n \to \infty$ in (10.8), we get

$$\mathbf{P}\Big(\frac{n}{V_i(n)} \to \mu_i \text{ as } n \to \infty\Big) = 1,$$

which implies

$$\mathbf{P}\Big(\frac{V_i(n)}{n} \to \frac{1}{\mu_i} \text{ as } n \to \infty\Big) = 1.$$

In the positive recurrent case, the chain has the unique stationary distribution $(\pi_i)_{i \in S}$. Assume without loss of generality that $|f| \le 1$. For any $J \subset S$, we have

$$
\begin{aligned}
\Big|\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \bar{f}\Big| &= \Big|\sum_{i\in S}\Big(\frac{V_i(n)}{n} - \pi_i\Big) f(i)\Big| \\
&\le \sum_{i\in J}\Big|\frac{V_i(n)}{n} - \pi_i\Big| + \sum_{i\notin J}\Big|\frac{V_i(n)}{n} - \pi_i\Big| \\
&\le \sum_{i\in J}\Big|\frac{V_i(n)}{n} - \pi_i\Big| + \sum_{i\notin J}\Big(\frac{V_i(n)}{n} + \pi_i\Big) \\
&= \sum_{i\in J}\Big|\frac{V_i(n)}{n} - \pi_i\Big| + 1 - \sum_{i\in J}\frac{V_i(n)}{n} + \sum_{i\notin J}\pi_i \\
&= \sum_{i\in J}\Big|\frac{V_i(n)}{n} - \pi_i\Big| + \sum_{i\in J}\Big(\pi - \frac{V_i(n)}{n}\Big) + 2\sum_{i\notin J}\pi_i \\
&\le 2\sum_{i\in J}\Big|\frac{V_i(n)}{n} - \pi_i\Big| + 2\sum_{i\notin J}\pi_i.
\end{aligned}
$$

We proved above that

$$\mathbf{P}\Big(\frac{V_i(n)}{n} \to \frac{1}{\mu_i} \text{ as } n \to \infty \text{ for all } i\Big) = 1.$$

Given $\epsilon > 0$, choose J finite so that

$$\sum_{i\notin J}\pi_i < \frac{\epsilon}{4}.$$

and then choose $N = N(\omega)$ so that, for $n \geq N$,

$$\sum_{i \in J} \left| \frac{V_i(n)}{n} - \pi_i \right| < \frac{\epsilon}{4}.$$

Then, for $n \geq N$, we have

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \bar{f} \right| < \epsilon$$

as desired. □

10.7 Exercises

10–1. Let $\{X_n\}_{n\geq 0}$ be a Markov chain on $S = \{1, 2, 3\}$ with transition matrix

$$P = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 2/3 & 1/3 \\ 1/3 & 2/3 & 0 \end{pmatrix}.$$

Compute $\mathbf{P}(X_n = 1|X_0 = 1)$ for $n = 0, 1, 2, 3, 4, 5, 6$.

10–2. In Exercise 10-1, if the initial distribution is

$$\mathbf{P}(X_0 = 1) = \mathbf{P}(X_0 = 2) = \mathbf{P}(X_0 = 3) = 1/3,$$

Waht is $\mathbf{P}(Z_3 = 2 \text{ or } 3)$?

10–3. Prove that the Markov property (10.1) is equivalient to (10.2).

10–4. Let $\{X_n\}_{n\geq 0}$ be a Markov chain on $S = \{1, 2, 3, 4, 5\}$ with transition matrix

$$P = \begin{pmatrix} 0.5 & 0 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0.2 & 0.3 & 0.2 & 0.3 \\ 0.5 & 0 & 0 & 0 & 0.5 \end{pmatrix}.$$

Identify all closed classes. Which are irreducible? Which states are recurrent and which are transient?

10–5. Find the stationary distribution of the Markov chain in Exercise 10-1 and describe the limiting behaviour of the n-step transition probabilities $p_{ij}^{(n)}$ as $n \to \infty$.

10–6. Find the period of an irreducible Markov chain who has the transition matrix

$$(a) \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad (b) \begin{pmatrix} 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0 \end{pmatrix} \quad (c) \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

10–7. Let S be finite. Suppose that for some $i \in S$

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j \quad \text{for all } j \in S.$$

Show that $\pi = (\pi_j)_{j\in S}$ is a stationary distribution.

10–8. For disctinct states i and j, show that i reaches j if and only if there are some states $i_0, i_1, \ldots, i_n$ with $i_0 = i$ and $i_n = j$ such that $p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} > 0$.

References

[1] K.L. Chung, *Markov Chains with Stationary Transition Probabilities*, Springer, 2nd edition, 1967.

[2] P. Guttorp, *Stochastic Modelling of Scientific Data*, Chapman and Hall, 1995.

[3] J.R. Norris, *Markov Chains*, C.U.P., 1997.

[4] D. Revus, *Markov Chains*, North-Holland, 1984.

[5] D. Williams, *Proability with Martingales*, C.U.P., 1991.