1 Classification of states

In this section, we show several properties of DTMCs and the corresponding terminology.

1.1 Irreducibility

Informally, we say that the state j is accessible from state i if it is possible to get to state j from state i. Similarly, states i and j communicate if it is possible to go from state i to state j and vice versa.

Definition 1.1. A state j is called **accessible** from state i, denoted as $i \to j$, if there exists $n \in \{0, 1, \dots\}$ such that

$$\mathbb{P}(X_n = j \mid X_0 = i) = p_{ij}^{(n)} > 0.$$

We say states i and j **communicate**, denoted as $i \leftrightarrow j$, if there exist $m, n \in \{0, 1, \dots\}$ such that

$$\mathbb{P}(X_n = j \mid X_0 = i) = p_{ij}^{(n)} > 0 \quad \text{and} \quad \mathbb{P}(X_n = i \mid X_0 = j) = p_{ji}^{(m)} > 0.$$

Or equivalently, we have $i \leftrightarrow j$ means that $i \to j$ and $j \to i$.

Remark 1.2. In the above definition, we allow for $n = 0$ or $m = 0$. Since

$$\mathbb{P}(X_0 = j \mid X_0 = i) = p_{ij}^{(0)} = \delta_{ij} := \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases}$$

this implies that $i \leftrightarrow i$ for all $i \in S$. The symbol δ_{ij} is called the **Kronecker delta**.

The relation of communication divides the state space S into a partition of different **equivalence classes**. That is, all the states in one class communicate with each other, but not with any states from any other class.

Proposition 1.3 (*Equivalence Relation of Communication*)

The relation of communication is an **equivalence relation**. *That is, the following three conditions are satisfied:*

(1) Reflexivity: $i \leftrightarrow i$

(2) Symmetry: If $i \leftrightarrow j$, *then* $j \leftrightarrow i$

(3) Transitivity: If $i \leftrightarrow j$ *and* $j \leftrightarrow k$, *then* $i \leftrightarrow k$

Let

$$S_i = \{j \in S : j \leftrightarrow i\}$$

denote the set of states communicating with state i. If $i \leftrightarrow j$, then we have $S_i = S_j$ (see Problem 1.2). This implies that the communicating states can be divided into disjoint clusters or **communicating classes**.

Definition 1.4. A DTMC that has only one communication class is called **irreducible**.

In an irreducible DTMC, each state can be reached from any other state.

Definition 1.5. A state i is called **absorbing** if $p_{ii} = 1$

Since $p_{ij} \geq 0$ and $\sum_j p_{ij} = 1$, we must have $p_{ij} = 0$ for $j \neq i$ if state i is absorbing. Thus, once the DTMC enters state i, it will be trapped there forever.

Definition 1.6. A set of states C is called **closed**, if $p_{ij} = 0$ for all $i \in C$ and $j \in S \backslash C$.

If the DTMC is in a closed set C, then it will never leave this set again.

1.2 Periodicity

We now define a notion that measures how regularly a DTMC can return to a particular state.

Definition 1.7. The **period** of state i is defined as

$$d(i) = \gcd\left\{n \in \mathbb{N} : p_{ii}^{(n)} > 0\right\},$$

where gcd is the *greatest common divisor* of times at which return to state i is possible. If $d(i) = k$ for all $i \in S$, then we say that the DTMC has period k.

Remark 1.8. It is possible that the DTMC will never return to its current state, i.e. $p_{ii}^{(n)} = 0$ for all $n \in \mathbb{N}$. In this case, we let $d(i) = \infty$.

The period can be interpreted as the regular time interval in which is it possible to return to its current state.

Example 1.9. The simple random walk has period 2 because

$$\begin{cases} p_{ii}^{(n)} = 0, \text{ if } n \text{ is odd} \\ p_{ii}^{(n)} > 0, \text{ if } n \text{ is even} \end{cases}$$

However, the period does not mean that it is always possible to return to the state every period.

Remark 1.10. Note that $d(i) = k$ does **not** imply that $p_{ii}^{(k)} > 0$. For example, consider the DTMC with state space $S = \{1, 2, 3\}$ and transition matrix

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

with $p_{11}^{(2)} = p_{11}^{(3)} = 0.5$ and, hence, $d(1) = 1$. However, $p_{11}^{(1)} = p_{11} = 0$.

Definition 1.11. We call state i is **aperiodic** if $d(i) = 1$. The DTMC is called aperiodic if all states are aperiodic.

An aperiodic state has no pattern for return times. Naturally, if two states communicate with each other, then they share the same period.

Proposition 1.12 (*Class Property I*)

If $i \leftrightarrow j$, then $d(i) = d(j)$. In particular, if a DTMC is irreducible, then all its states have the same period.

Remark 1.13. There will be several results called **class properties** of a DTMC. These are properties shared by all states in the same communication class.

1.3 Recurrence and transience

For $n \in \mathbb{N}$, we define

$$f_{ij}^{(n)} = \mathbb{P}\left(X_n = j, X_{n-1} \neq j, \cdots, X_1 \neq j \middle| X_0 = i \right).$$

Then $f_{ij}^{(n)}$ is the probability that X visits state j for the first time at time n, given that $X_0 = i$. Clearly, if $X_0 = i$, then probability that X visits j for the first time is smaller than the probability X visits j starting at i

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} = f_{ij}^{(n)} + \sum_{k=1}^{n-1} f_{ij}^{(k)} p_{jj}^{(n-k)} \geq f_{ij}^{(n)}.$$

We also define

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

which is the probability that X ever visits state j, given that $X_0 = i$.

Definition 1.14.

1. A state i is called **recurrent** if $f_{ii} = 1$, i.e., if started at i, the DTMC will return to i at some time in the future with probability one.
2. A state i is called **transient** if $f_{ii} < 1$, i.e., there is a positive probability that the DTMC will never return to state i.

Another way to characterize recurrence and transience is through the total number of visits. Define

$$M_i = \sum_{n=1}^{\infty} \mathbb{1}\{X_n = i\}$$

which denotes the total number of times X visits state i after time 0. There is a direct relationship between recurrence, the number of total visits, and the transition probabilities.

Proposition 1.15 (*Equivalent Notions of Recurrence and Transience*)

$$\textit{state } i \textit{ is recurrent} \iff \mathbb{E}\left[\, M_i \middle|\, X_0 = i\right] = \infty \iff \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$$

$$\textit{state } i \textit{ is transient} \iff \mathbb{E}\left[\, M_i \middle|\, X_0 = i\right] < \infty \iff \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$$

The next proposition shows that recurrence and transience are class properties.

Proposition 1.16 (*Class Property II*)

Suppose that $i \leftrightarrow j$.

(i) If i is recurrent, then j is recurrent.

(ii) If i is transient, then j is transient.

As a result, if a DTMC is irreducible, then either all states are recurrent, or they are all transient. The latter is not possible if the state space is finite (see Proposition 2.8).

1.3.1 Mean Recurrence

We define the first time at which state i is visited after time 0 by:

$$T_i = \min\left\{n \in \mathbb{N} : X_n = i\right\}.$$

Note that T_i is a random stopping time. Clearly,

$$\mathbb{P}(T_i = k \mid X_0 = i) = \mathbb{P}\left(X_k = i, X_{k-1} \neq i, \cdots, X_1 \neq i \middle|\, X_0 = i\right) = f_{ii}^{(k)}. \tag{1}$$

Define the **mean recurrence time** of state i by

$$\mu_i = \mathbb{E}\left[T_i \mid X_0 = i\right] = \sum_{k=1}^{\infty} k\mathbb{P}\left(T_i = k \mid X_0 = i\right) = \sum_{k=1}^{\infty} k f_{ii}^{(k)}. \tag{2}$$

which denotes the average time to returns to state i. If a state i is recurrent, then the DTMC returns to i with probability one. However, the time T_i until the first return can have such a heavy tail that $\mu_i = \mathbb{E}[T_i|X_0 = i]$ is infinite. In this case, the state i is called null recurrent.

Definition 1.17. A recurrent state i is called **positive recurrent** if $\mu_i < \infty$, and **null recurrent** if $\mu_i = \infty$. We call the DTMC is positive (null, resp.) recurrent if all states are positive (null, resp.) recurrent.

Remark 1.18. If state i is recurrent, we have $f_{ii} = \sum_{k=1}^{\infty} f_{ii}^{(k)} = 1$, which implies

$$\lim_{k\uparrow\infty} f_{ii}^{(k)} = \mathbb{P}\left(T_i = \infty \mid X_0 = i\right) = 0.$$

However, this does not imply the finiteness of μ_i since we need $(kf_{ii}^{(k)})_{k\geq 1}$ so be summable.

The follow proposition shows that positive recurrence and null recurrence are also class properties.

Proposition 1.19 (*Class Property III*)

Suppose $i \leftrightarrow j$.

(i) If i is positive recurrent, then j is positive recurrent.

(ii) If i is null recurrent, then j is null recurrent.

To summarize, every state $i \in S$ can either be transient or recurrent. If it is recurrent, it must either be positive recurrent or null recurrent. Any of these properties is shared by all other states in the same communication class. This is called the **classification of states**.

Theorem 1.20 (*Decomposition theorem*)

The state space S of a DTMC can be partitioned uniquely as follows:

$$S = T \cup C_1 \cup C_2 \cup \cdots$$

where T is the set of all transient states, and C_i are irreducible closed sets of recurrent states.

These sets of states have the following interpretations:

- If the process is in any C_i, it never leaves it.
- If the process is in T, it may stay there forever but it can also move to one of the C_i, where it subsequently remains.

As a consequence, we can **reorder** the states such that the transition matrix can be rewritten as

$$\boldsymbol{P} = \begin{array}{c c} & \begin{array}{ccccc} C_1 & C_2 & C_3 & \cdots & T \end{array} \\ \begin{array}{c} C_1 \\ C_2 \\ C_3 \\ \vdots \\ T \end{array} & \begin{pmatrix} \boldsymbol{P}_1 & & & & \\ & \boldsymbol{P}_2 & & & \\ & & \boldsymbol{P}_3 & & \\ & & & \ddots & \\ \boldsymbol{Q}_1 & \boldsymbol{Q}_2 & \boldsymbol{Q}_3 & \cdots & \cdots \end{pmatrix} \end{array} \tag{3}$$

where $\boldsymbol{P}_i$ is the transition matrix of the DTMC *restriced to the closed set C_i*.

1.4 Example Problems

1.4.1 Proofs of Statements

Problem 1.1. Prove that the communication is an equivalence relation (Proposition 1.3).

Solution 1.1. Both reflexivity and symmetry are clear from the definition. Thus, we only need to show the transitivity. Suppose that $i \leftrightarrow j$ and $j \leftrightarrow k$, i.e. there exists $m, n \in \{0, 1, \dots\}$ such that $p_{ij}^{(m)} > 0$ and $p_{jk}^{(n)} > 0$. Then, by the Chapman–Kolmogorov equations,

$$p_{ik}^{(m+n)} = \sum_{l \in S} p_{il}^{(m)} p_{lk}^{(n)} \geq p_{ij}^{(m)} p_{jk}^{(n)} > 0,$$

which implies $i \to k$. An indentical argument shows that $k \to i$.

Problem 1.2. Let

$$C_i := \{j \in S : j \leftrightarrow i\}.$$

Show that

1. $i \in C_i$
2. $j \in C_i$ if and only if $i \in C_j$
3. if $i \leftrightarrow j$ then $C_i = C_j$

Furthermore, show that all distinct communication classes create a partition of S.

Solution 1.2. The first three properties follow from the fact that communication is an equivalence class:

1. Since $i \leftrightarrow i$ by reflexivity, we have that $i \in C_i$.
2. Since $i \leftrightarrow j$ and $j \leftrightarrow i$ by symmetry, we have $j \in C_i$ if and only if $i \in C_j$.
3. Suppose that $i \leftrightarrow j$. If $k \in C_i$ then $k \to i$, so $k \to j$ by transitivity. In particular, we have

$$C_i \subseteq C_j.$$

   An identical argument shows that then $C_j \subseteq C_i$ so $C_i = C_j$.

For the sake of contradiction, suppose that $i \in S$ belongs to distinct equivalence classes. That is, suppose that $i \in C_j$ and $i \in C_k$ but $C_i \neq C_j$. However, the second property implies that $j \in C_i$ and $k \in C_i$, so $j \leftrightarrow k$ by transitivity. The third property implies that $C_i = C_j$, which is a contradiction. We conclude that every $i \in S$ belongs to exactly one equivalence class, so all distinct communcation classes form a partition of S.

Problem 1.3. Prove the Class Property I (Proposition 1.12).

Solution 1.3. We only need to consider the case that $i \neq j$, since the case that $i = j$ is trivial. It suffices to show that $d(i)$ divides $d(j)$. Since symmetry will imply that $d(j)$ also divides $d(i)$, which implies that $d(i) = d(j)$.

We we define $N(i) = \{n : p_{ii}^{(n)} > 0\}$ then $d(i) = \gcd(N(i))$. We can define $N(j)$ in a similar way. Since $i \leftrightarrow j$, there exists m, n such that

$$p_{ij}^{(m)} > 0 \text{ and } p_{ji}^{(n)} > 0.$$

This implies by the Chapman–Kolmogorov equations that

$$p_{ii}^{(m+n)} \geq p_{ij}^{(m)} p_{ji}^{(n)} > 0$$

and in turn $m + n \in N(i)$. Next, for any $l \in N(j)$, we have $p_{jj}^{l} > 0$, so

$$p_{ii}^{(l+m+n)} \geq p_{ij}^{(m)} p_{jj}^{(l)} p_{ji}^{(n)} > 0.$$

Therefore, $m + n + l \in N(i)$ for all $l \in N(j)$. We conclude that $d(i)$ divides both $m + n$ and $m + n + l$ for any $l \in N(j)$. Therefore, $d(i)$ is a divisor of $N(j)$ and, hence, in particular divides $d(j)$.

Problem 1.4. Prove the equivalence of the different notions of recurrence and transcience (Proposition 1.15).

Solution 1.4. We first show the recurrence and transicience in terms of the expected visits. Note that if $f_{ii} < 1$,

$$\mathbb{P}\left(M_i = k \middle| X_0 = i\right) = \underbrace{f_{ii} \times f_{ii} \times \cdots \times f_{ii}}_{\text{return to state } i \text{ for exactly } k \text{ visits}} \times \underbrace{(1 - f_{ii})}_{\text{never visit } i \text{ ever again}}$$

In other words, conditionally on $X_0 = i$, M_i follows a geometric distribution with parameter $1 - f_{ii}$, so by the properties of the geometric distribution

$$\mathbb{E}\left[M_i \middle| X_0 = i\right] = \frac{f_{ii}}{1 - f_{ii}} < \infty.$$

By taking limits,

$$\lim_{f_{ii} \uparrow 1} \mathbb{E}\left[M_i \middle| X_0 = i\right] = \lim_{f_{ii} \uparrow 1} \frac{f_{ii}}{1 - f_{ii}} = \infty$$

and so if i is recurrent, we have

$$\mathbb{E}\left[M_i \middle| X_0 = i\right] = \infty.$$

To show that last equivalence in terms of the transition probabilities, notice that by linearity

$$\mathbb{E}\left[M_i \middle| X_0 = i\right] = \sum_{n=1}^{\infty} \mathbb{E}\left[\mathbb{1}\{X_n = i\} \middle| X_0 = i\right] = \sum_{n=1}^{\infty} \mathbb{P}\left[X_n = i \,\middle|\, X_0 = i\right] = \sum_{n=1}^{\infty} p_{ii}^{(n)}.$$

1.4.2 Applications

Problem 1.5. Is the random walk irreducible?

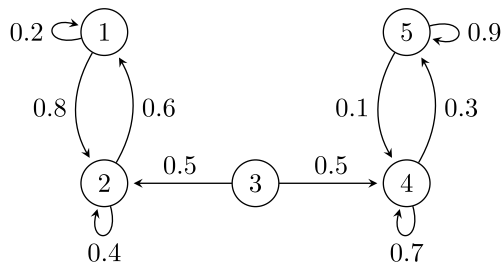
Solution 1.5. The random walk is irreducable because given any state $i \in \mathbb{Z}$, we have $i + 1 \leftrightarrow i$ and $i - 1 \leftrightarrow j$. Therefore, by repetitive applications of the transitivity property, we can conclude that $i \leftrightarrow j$ for any $i, j \in \mathbb{Z}$, so the random walk is irreducible.

Problem 1.6. Consider a DTMC with states $\{1,2,3,4,5\}$ and transition matrix

$$\boldsymbol{P} = \begin{pmatrix} 0.2 & 0.8 & & & \\ 0.6 & 0.4 & & & \\ & 0.5 & 0 & 0.5 & \\ & & & 0.7 & 0.3 \\ & & & 0.1 & 0.9 \end{pmatrix}$$

What are the communication classes of this DTMC.

Solution 1.6. We draw the corresponding state transition diagram associated with this DTMC:



Clearly, the communication classes are $\{1,2\}$, $\{3\}$, and $\{4,5\}$.

Remark 1.21. Notice that the values of the transition is not important to figure out the communication classes.

Problem 1.7. Consider a DTMC with transition matrix

$$\boldsymbol{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} p & 0 & 1-p & 0 \\ 0 & r & 0 & 1-r \\ q & 0 & 1-q & 0 \\ 0 & s & 0 & 1-s \end{pmatrix} \end{array}$$

where $p, q, r, s \in (0,1)$. What are the communication classes? Which ones are recurrent and which are transient? Show that we can $\boldsymbol{P}$ in the form of (3).

Solution 1.7. This DTMC is reducible with two closed recurrent classes $\{1,3\}$ and $\{2,4\}$. We can reorder the states to $\{1,3,2,4\}$ so that the transition matrix becomes

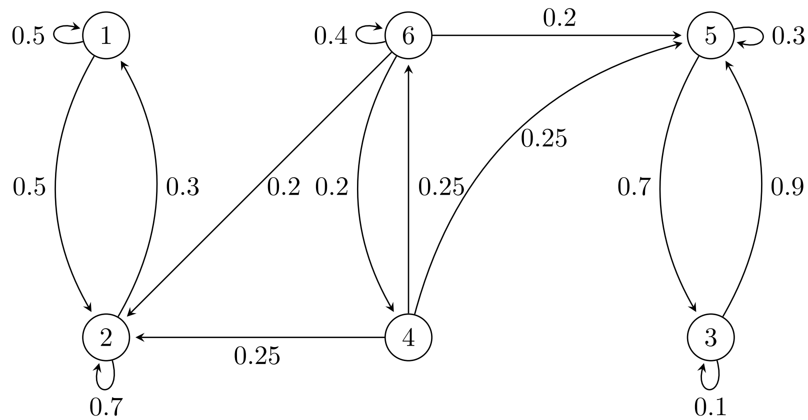
$$\boldsymbol{P} = \begin{array}{c} \\ 1 \\ 3 \\ 2 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 3 & 2 & 4 \end{array} \\ \begin{pmatrix} p & 1-p & 0 & 0 \\ q & 1-q & 0 & 0 \\ 0 & 0 & r & 1-r \\ 0 & 0 & s & 1-s \end{pmatrix} \end{array}$$

Problem 1.8. Consider the Markov chain with state space $\{1,2,3,4,5,6\}$ and transition matrix

$$\boldsymbol{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 3/10 & 7/10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/10 & 0 & 9/10 & 0 \\ 1/4 & 1/4 & 0 & 0 & 1/4 & 1/4 \\ 0 & 0 & 7/10 & 0 & 3/10 & 0 \\ 0 & 1/5 & 0 & 1/5 & 1/5 & 2/5 \end{pmatrix}.$$

What are the communication classes? Which ones are recurrent and which are transient? Show that we can $\boldsymbol{P}$ in the form of (3).

Solution 1.8. We draw the corresponding state transition diagram associated with this DTMC:



The communication classes are $\{1,2\}$, $\{3,5\}$, and $\{4,6\}$. The classes $\{1,2\}$ and $\{3,5\}$ are recurrent and the class $\{4,6\}$ is transient. After relabeling the states to $\{1,2,3,5,4,6\}$, the transition matrix becomes

$$\boldsymbol{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 3/10 & 7/10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/10 & 9/10 & 0 & 0 \\ 0 & 0 & 7/10 & 3/10 & 0 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 0 & 1/4 \\ 0 & 1/5 & 0 & 1/5 & 1/5 & 2/5 \end{pmatrix}.$$

To carefully show that the states are recurrent, notice that for example

$$f_{22} = \lim_{n\to\infty}(1 - 0.3 \cdot 0.5^n) = 1$$

since the only way we never visit 2 starting at 2 is if we move to state 1 and stay there forever. Since 2 is recurrent, but the Class property, we must also have that 1 is recurrent. The reccurence of 5 and 3 is similar.

To show that 4 is not recurrent, notice that with probability 0.25, we go to state 2, and state 2 it is impossible to return. Therefore,

$$1 - f_{44} \geq 1 - 0.25 = 0.75 \implies f_{44} \leq 0.75$$

since the event to go from 4 to 2 is contained in the event that you never return.

2 Stationary and limiting distributions

2.1 Stationary distributions

A stationary distribution is a special distribution of the states, such that if the Markov chain starts at a stationary distribution, all of its marginal states will remain at a stationary distribution.

Definition 2.1. A row vector $\boldsymbol{\pi} \in [0,1]^{|S|}$ is called a **stationary distribution** for a DTMC X with transition matrix $\boldsymbol{P}$ if

1. $\pi_i \geq 0$ for all $i \in S$ and $\sum_{i \in S} \pi_i = 1$,
2. $\boldsymbol{\pi} = \boldsymbol{\pi P}$, or equivalently, $\pi_j = \sum_{i \in S} \pi_i p_{ij}$ for all $j \in S$.

Remark 2.2. Obviously, by the properties of matrix multiplication,

$$\boldsymbol{\pi P}^n = \boldsymbol{\pi P} \cdot \boldsymbol{P}^{n-1} = \boldsymbol{\pi P}^{n-1} = \cdots = \boldsymbol{\pi P} = \boldsymbol{\pi},$$

Therefore, if we start the Markov chain with the stationary distribution so that $\boldsymbol{\nu}^{(0)} = \boldsymbol{\pi}$, then

$$\boldsymbol{\nu}^{(n)} = \boldsymbol{\nu}^{(0)} \boldsymbol{P}^n = \boldsymbol{\pi P}^n = \boldsymbol{\pi},$$

which means that X_n has the same distribution $\boldsymbol{\pi}$ for all $n \in \mathbb{N}$.

To find a stationary distribution, we can look for a nonnegative row vector π solving the linear equation

$$\boldsymbol{\pi} = \boldsymbol{\pi P}$$

under the constraints that

$$\begin{cases} \pi_i \geq 0 \text{ for all } i, \\ \sum_{i \in S} \pi_i = \boldsymbol{\pi} \cdot \mathbf{1} = 1, \end{cases}$$

where $\mathbf{1}$ is the row vector with all entries equal to 1. The solutions to such systems might not exist, nor are they necessarily unique.

Remark 2.3. A stationary distribution might not always exist (see Problem 2.5), and the stationary distribution might not be unique (see Problem 2.4).

We have the following two theorems establishing existence, and in the second case also uniqueness, of stationary distributions. Existence follows from:

Theorem 2.4 (*Existence*)

A DTMC with a finite state space S admits at least one stationary distribution $\boldsymbol{\pi}$.

Intuitively, the stationary distribution should be inversely proportional to average time it takes to return to that state. This fact and uniqueness follows from:

Theorem 2.5 (*Uniqueness*)

*An **irreducible** DTMC admits a stationary distribution $\boldsymbol{\pi}$ if and only if it is positive recurrent. In this case, $\boldsymbol{\pi}$ is unique and given by*

$$\pi_i = \frac{1}{\mu_i},$$

where μ_i is the mean recurrence time defined in (2).

By the definition of positive recurrence, we have that

$$\pi_i = \frac{1}{\mu_i} > 0 \qquad \text{for all } i \in S.$$

2.2 Limiting distribution

We are interested in the long run behavior of a DTMC. In particular, does there exist a possible limit,

$$\nu_i^{(\infty)} := \lim_{n\uparrow\infty} \nu^{(n)},$$

of the distributions $\nu^{(n)}$ of a DTMC at time n? The following result says that if such a limit exists, it must be stationary.

Proposition 2.6 (*Limit is Stationary*)

Suppose that the state space S is finite and that $\nu_i^{(\infty)} := \lim_n \nu_i^{(n)}$ exists for $i \in S$. Then $\nu^{(\infty)}$ is a stationary distribution.

Since $\boldsymbol{\nu}^{(n)} = \boldsymbol{\nu}^{(0)}\boldsymbol{P}^n$, the limit of $\boldsymbol{\nu}^{(n)}$ will exist if both S is finite and $\lim_{n\to\infty} \boldsymbol{P}^n$ exists. We will see that periodicity and recurrence/transience play an important role in the limiting behavior of a DTMC (see Problems 2.7, 2.8, 2.9). Periodicity is problematic because if states oscillate, then its limiting distribution may not exist. Next, if the state is transient then eventually the Markov chain will move away from that state:

1. the probability that the Markov chain is in the transient state j at time n tends to 0.
2. when starting in a null recurrent state, the Markov chain will spread out further and further, so that the probability of finding it in a particular state j at time n tends to 0.

Proposition 2.7

1. *If state j is transient, then*

$$\lim_{n\to\infty} p_{ij}^{(n)} = \mathbb{P}(X_n = j | X_0 = i) = 0, \text{ for any state } i.$$

2. *If state i is null recurrent, then*

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0, \text{ for any state } j.$$

This intuition suggests that if there are finitely many states, then it should be impossible for all states to be transient or spread out indefinitely. This is made precise below.

Proposition 2.8

*Suppose S is **finite**. Then:*

1. *at least one state is recurrent;*
2. *all recurrent states are positive recurrent.*

The preceding results can be summarized with the following fundamental result.

Theorem 2.9 (*Fundamental limit theorem*)

Suppose that the DTMC is irreducible, aperiodic, and positive recurrent. Then $\lim_{n\to\infty} p_{ij}^{(n)}$ *exists and is independent of state* i*. Furthermore,*

$$\lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} p_{jj}^{(n)} = \frac{1}{\mu_j} = \pi_j,$$

where μ_j *is the mean of the first time to revisit state* j *and* $\boldsymbol{\pi} = (\pi_j)_{j\in S}$ *is the unique stationary distribution from Theorem 2.5.*

Remark 2.10. In the case that the DTMC is irreducible, aperiodic, and null recurrent, the relation

$$\lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} p_{jj}^{(n)} = \frac{1}{\mu_j} = 0$$

holds due to Proposition 2.7 and the definition of null recurrence.

We now state an easy to check sufficient condition for the fundamental limit theorem to hold.

Proposition 2.11

Suppose that S *is finite and there exists* $n \in \mathbb{N}$ *such that such that* $p_{ij}^{(n)} > 0$ *for all* $i, j \in S$*. Then the DTMC is irreducible, aperiodic, and positive recurrent. In particular, the conclusion of Theorem 2.9 holds.*

Remark 2.12. Actually, the following partial converse to Proposition 2.11 is true: Suppose that S is finite and admits a probability distribution $\boldsymbol{\pi}$ such that $\pi_j > 0$ and $\pi_j = \lim_n p_{ij}^{(n)}$ for all $j \in S$. Then there exists $n \in \mathbb{N}$ such that such that $p_{ij}^{(n)} > 0$ for all $i, j \in S$.

2.3 Example Problems

2.3.1 Proofs of Statements

Problem 2.1. Prove that the limiting distribution must be stationary if it exists (Proposition 2.6).

Solution 2.1. Clearly, $\nu_i^{(\infty)}$ is nonnegative as the limit of the nonnegative numbers $\nu_i^{(n)}$. Next,

$$\sum_{i\in S} \nu_i^{(\infty)} = \sum_{i\in S} \lim_{n\uparrow\infty} \nu_i^{(n)} = \lim_{n\uparrow\infty} \sum_{i\in S} \nu_i^{(n)} = \lim_{n\uparrow\infty} 1 = 1,$$

where the interchange of summation and limits was possible, because S is finite. Hence, $\boldsymbol{\nu}^{(\infty)}$ is indeed a probability distribution on S. Next,

$$\begin{aligned}\nu_i^{(\infty)} &= \lim_{n\uparrow\infty} \nu_i^{(n)} = \lim_{n\uparrow\infty} \nu_i^{(n+1)} = \lim_{n\uparrow\infty} (\nu^{(n)} P)_i = \lim_{n\uparrow\infty} \sum_{j\in S} \nu_j^{(n)} p_{ji} \\ &= \sum_{j\in S} \lim_{n\uparrow\infty} \nu_j^{(n)} p_{ji} = \sum_{j\in S} \nu_j^{(\infty)} p_{ji},\end{aligned}$$

where we have used again the assumption that S is finite when interchanging limits and summation. Thus, $\boldsymbol{\nu}^{(\infty)} = \boldsymbol{\nu}^{(\infty)} \boldsymbol{P}$.

Problem 2.2. Prove Proposition 2.8.

Solution 2.2.

Part 1: We assume by way of contradiction that all states are transient. Then, by Proposition 2.7, for any fixed state i,

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0 \text{ for all } j \in S.$$

However, since S is finite, we can interchange the limit and sumamtion to see that

$$1 = \lim_{n\to\infty} \sum_{j\in S} p_{ij}^{(n)} = \sum_{j\in S} \lim_{n\to\infty} p_{ij}^{(n)} = \sum_{j\in S} 0 = 0,$$

which is a contradiction.

Part 2: We assume by way of contradiction that there exists a null recurrent state i. By Proposition 2.7

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0, \text{ for all } j \in S.$$

Thus, by the same argument as part (a), we reach a contradiction.

Problem 2.3. Prove Proposition 2.11.

Solution 2.3. We check each of the three properties:

Irreducible: It follows immediately that the DTMC has just one communication class and hence is irreducible.

Aperiodic: Therefore it will be aperiodic if we can show that one of its states, say i, has period $d(i) = 1$. Let $N(i) = \{k \in \mathbb{N} | p_{ii}^{(k)} > 0\}$. We clearly have $n \in N(i)$. Next, take some $j \in S$ with $p_{ji} > 0$. Such a j must exist since $p_{ii}^{(n)} > 0$. Then $p_{ii}^{(n+1)} \geq p_{ij}^{(n)} p_{ji} > 0$ and so $n+1 \in N(i)$. It follows that $d(i)$ divides both n and $n+1$ and hence must be equal to 1.

Reducible: Finally. the fact that the DTMC is positive recurrent follows from Proposition 2.8.

2.3.2 Applications

Problem 2.4. Consider the transition matrix

$$\boldsymbol{P} = \begin{pmatrix} & & 1 \\ & 1 & \\ 1 & & \end{pmatrix}$$

on $S = \{0, 1, 2\}$. Find all possible stationary distributions.

Solution 2.4. Any stationary distribution π must satisfy

$$\begin{cases} \pi_0 = \pi_2 \\ \pi_0 + \pi_1 + \pi_2 = 1 \\ \pi_i \geq 0, \text{ for } i = 0, 1, 2 \end{cases}$$

In other words,

$$\pi = \Big(\frac{1}{2}a, 1-a, \frac{1}{2}a)$$

is a stationary distribution for any $a \in [0, 1]$.

Problem 2.5. Consider the simple random walk with state space $S = \mathbb{Z}$. Show that the random walk does not have a stationary distribution.

Solution 2.5. Since

$$p_{i,i+1} = p_{i,i-1} = \frac{1}{2} \qquad \text{and} \qquad p_{i,j} = 0 \text{ otherwise},$$

and stationary distribution π would have to satisfy

$$\pi_i = (\boldsymbol{\pi} \boldsymbol{P})_i = \frac{1}{2}\pi_{i-1} + \frac{1}{2}\pi_{i+1} \qquad \text{for all } i \in \mathbb{Z}.$$

That is, for $i > 0$,

$$\pi_{i+1} - \pi_i = \pi_i - \pi_{i-1} = \cdots = \pi_1 - \pi_0,$$

and hence

$$\pi_{i+1} = \pi_0 + (i+1)(\pi_1 - \pi_0).$$

A similar reasoning applies in the case $i < 0$, and we get that

$$\pi_i = a + bi,$$

for $a = \pi_0$ and $b = \pi_1 - \pi_0$. But this form is incompatible with the requrements

$$\pi_i \geq 0 \qquad \text{and} \qquad \sum_{i \in \mathbb{Z}} \pi_i = 1$$

unless $\pi_i = 0$ for all i. Hence, a stationary distribution π cannot exist.

Remark 2.13. We have shown that a stationary distribution need not exist if S is infinite.

Problem 2.6. Show that the simple random walk is not positive recurrent.

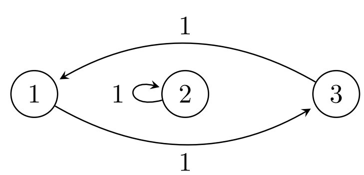
Solution 2.6. Clearly the simple random walk is irreducible. By Problem 2.5 we know that no stationary distribution exists. However, Theorem 2.5 states that if no stationary distribution exists, then the irreducible DTMC is not positive recurrent.

Problem 2.7. Consider the transition matrix

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

Classify the states of this DTMC. Does $\lim_{n \to \infty} \boldsymbol{P}^n$ exist?

Solution 2.7. This DTMC is reducible and has two recurrent classes $\{2\}$ and $\{1, 3\}$.



It is straightforward to verify that

$$\boldsymbol{P}^{2n} = I \text{ and } \boldsymbol{P}^{2n+1} = P, \text{ for any } n \in \mathbb{N}.$$

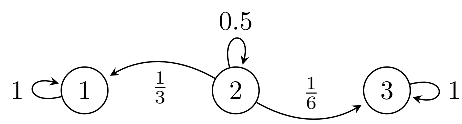
In other words, $\lim_{n\to\infty} \boldsymbol{P}^n$ does not exist.

Problem 2.8. Consider the transition matrix

$$\boldsymbol{P} = \begin{pmatrix} 1 & & \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \\ & & 1 \end{pmatrix}.$$

Classify the states of this DTMC. Does $\lim_{n\to\infty} \boldsymbol{P}^n$ exist?

Solution 2.8. This DTMC is reducible, has two recurrent classes $\{1\}$, $\{3\}$ and one transient class $\{2\}$.



We have

$$\lim_{n\to\infty} \boldsymbol{P}^n = \begin{pmatrix} 1 & & \\ \frac{2}{3} & 0 & \frac{1}{3} \\ & & 1 \end{pmatrix}.$$

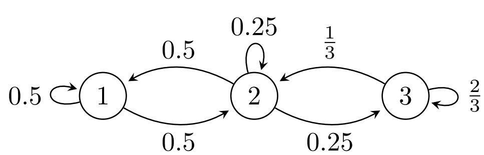
Note that $\lim_{n\to\infty} \boldsymbol{P}^n$ exists, but the rows are not identical. Also, $\lim_{n\to\infty} p_{i2}^{(n)} = 0$ for any $i = 1, 2, 3$.

Problem 2.9. Consider the transition matrix

$$\boldsymbol{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

Classify the states of this DTMC. Does $\lim_{n\to\infty} \boldsymbol{P}^n$ exist?

Solution 2.9. This DTMC is irreducible, aperiodic, and recurrent.



We can show numerically that

$$\lim_{n\to\infty} \boldsymbol{P}^n = \begin{pmatrix} \frac{4}{11} & \frac{4}{11} & \frac{3}{11} \\ \frac{4}{11} & \frac{4}{11} & \frac{3}{11} \\ \frac{4}{11} & \frac{4}{11} & \frac{3}{11} \end{pmatrix}.$$

3 Markov Chain Monte Carlo (MCMC)

3.1 Reversible Distributions

Definition 3.1. A probability distribution π on S is called a **reversible distribution** for our Markov chain if it satisfies the condition of **detailed balance**:

$$\pi_i p_{ij} = \pi_j p_{ji} \qquad \text{for all } i, j \in S. \tag{4}$$

By the chain rule for conditional expectations, we see that (4) can be written as

$$\mathbb{P}[X_0 = i, X_1 = j] = \mathbb{P}(X_0 = i)\,\mathbb{P}(X_1 = j \,|\, X_0 = i) = \mathbb{P}(X_0 = j)\,\mathbb{P}(X_1 = i \,|\, X_0 = j) = \mathbb{P}[X_0 = j, X_1 = i]$$

provided that $\boldsymbol{\pi}$ is the initial distribution. This explains why $\boldsymbol{\pi}$ is called "reversible". In fact, the next proposition shows that we can reverse multiple steps.

Proposition 3.2

If π is a reversible distribution and $\boldsymbol{\nu}^{(0)} = \boldsymbol{\pi}$, then the distribution of the Markov chain is invariant under time reversal. That is, for $n \in \mathbb{N}$ and $x_0, x_1, \ldots, x_n \in S$,

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_n = x_0, X_{n-1} = x_1, \ldots, X_0 = x_n)$$

The following fact is essential in the Metroplis–Hastings algorithm.

Proposition 3.3

Every reversible distribution $\boldsymbol{\pi}$ is also a stationary distribution.

3.2 Application: The Metropolis–Hastings algorithm

Suppose that S is a large (possibly infinite) state space and $\boldsymbol{\pi}$ is a given probability distribution on S. The goal is to sample from $\boldsymbol{\pi}$, via **Markov chain Monte Carlo (MCMC)**. More precisely, we want to construct a Markov chain on S so that its stationary distribution is $\boldsymbol{\pi}$. If we generate a series of values $(x_0, \ldots, x_n)$ from the Markov chain, then for x_n large, it will be close to $\boldsymbol{\pi}$.

Technically, the idea is to construct a DTMC X that satisfies the assumptions of Proposition 2.11 and has reversible distribution $\boldsymbol{\pi}$. Then we will have by Proposition 2.11 and Theorem 2.9:

$$\mathrm{Law}(X_n) = \boldsymbol{\nu}^{(n)} \approx \boldsymbol{\pi}.$$

Theorem 3.4 (*Metropolis–Hastings Algorithm*)

Given $\boldsymbol{\pi}$ and any proposal transition matrix $\boldsymbol{Q} = \{q_{ij}\}_{i,j\in S}$, we define

$$\alpha_{ij} := \min\left\{1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right\}$$

and

$$\boldsymbol{P} = (p_{ij})_{i,j\in S} \qquad \text{where} \qquad p_{ij} := \begin{cases} \alpha_{ij} q_{ij} & \text{if } i \neq j, \\ 1 - \sum_{k\neq i} \alpha_{ik} q_{ik} & \text{if } i = j. \end{cases}$$

Then:

1. *$\boldsymbol{P}$ is a transition matrix that defines a DTMC.*
2. *$\boldsymbol{\pi}$ is a stationary distribution for $\boldsymbol{P}$.*

A key consequence of this result is that it will allow us to compute averages with respect to $\boldsymbol{\pi}$. If S is finite and $\boldsymbol{P}$ satisfies, for example, the conditions of Proposition 2.11, then $p_{ij}^{(n)} \to \pi_j$ as $n \uparrow \infty$ and for any function $f : S \to \mathbb{R}$, then by the law of large numbers

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \longrightarrow \sum_{i\in S} f(i)\pi_i = \mathbb{E}_{X\sim\pi} f(X)$$

Whether the conditions of Proposition 2.11 is satisfied or not depends critically on the choice of the proposal $\boldsymbol{Q}$.

Remark 3.5. Given $\boldsymbol{Q}$ and $\boldsymbol{\pi}$, we can generate a Markov chain X with transition matrix $\boldsymbol{P}$ as follows:

- Given X_n, sample Y_{n+1} according to the distribution $q_{X_n,j}$, $j \in S$.
- Generate an independent Bernoulli random variable B with

$$\mathbb{P}[B = 1] = \alpha_{X_n,Y_{n+1}} \quad \text{and} \quad \mathbb{P}[B = 0] = 1 - \alpha_{X_n,Y_{n+1}}.$$

- If $B = 1$ set $X_{n+1} := Y_{n+1}$, otherwise set $X_{n+1} := X_n$.

The proof of this result is given in Problem 3.4.

3.3 Example Problems

3.3.1 Proofs of Statements

Problem 3.1. Prove Proposition 3.2.

Solution 3.1. Recall the formula for the joint distribution of a Markov chain,

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \nu_{x_0}^{(0)} p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{n-1}x_n}.$$

By the chain rule, we can flip indices to conclude that

$$\begin{aligned}
\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) &= \nu_{x_0}^{(0)} p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{n-1}x_n}\\
&= \pi_{x_0}p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{n-1}x_n}\\
&= p_{x_1x_0}\pi_{x_1}p_{x_1x_2}\cdots p_{x_{n-1}x_n}\\
&= p_{x_1x_0}p_{x_2,x_1}\pi_{x_2}\cdots p_{x_{n-1}x_n}\\
&\;\;\vdots\\
&= p_{x_1,x_0}p_{x_2,x_1}\cdots p_{x_n,x_{n-1}}\pi_{x_n}\\
&= \pi_{x_n}p_{x_n,x_{n-1}}\cdots p_{x_2,x_1}p_{x_1,x_0}\\
&= \mathbb{P}(X_0 = x_n, X_1 = x_{n-1}, \dots, X_n = x_0).
\end{aligned}$$

Problem 3.2. Prove Proposition 3.3.

Solution 3.2. We have

$$(\pi P)_j = \sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \underbrace{\sum_i p_{ji}}_{=1} = \pi_j$$

Problem 3.3. Prove the validity of the Metropolis algorithm in Proposition 3.4.

Solution 3.3.

Part 1: We clearly have $p_{ij} \geq 0$ for all $i \neq j$. Moreover,

$$\sum_{k \neq i} \alpha_{ik} q_{ik} \leq \sum_{k \neq i} q_{ik} = 1 - q_{ii} \leq 1.$$

Therefore, also $p_{ii} \geq 0$. The fact that $\sum_{j \in S} \pi_{ij} = 1$ is obvious. Hence P is indeed the transition matrix of a DTMC.

Part 2: To check the detailed balance equation, let $i \neq j$ be given. We may assume without loss of generality that $\pi_j q_{ji} \leq \pi_i q_{ij}$, for otherwise we can swap i and j. Then

$$\alpha_{ij} = \min\left\{1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right\} = \frac{\pi_j q_{ji}}{\pi_i q_{ij}}$$

and so

$$\pi_i p_{ij} = \pi_i \alpha_{ij} q_{ij} = \pi_i \frac{\pi_j q_{ji}}{\pi_i q_{ij}} q_{ij} = \pi_j q_{ji} = \pi_j \alpha_{ji} q_{ji} = \pi_j p_{ji},$$

because $\alpha_{ji} = 1$. Therefore, by Proposition 3.3, $\boldsymbol{\pi}$ is stationary.

Problem 3.4. Show that the algorithm in Remark 3.5 generates a DTMC with transition matrix $\boldsymbol{P}$.

Solution 3.4. Suppose that $X_n = i \in S$. Then if we generate according to the algorithm, then for $i \neq Y_{n+1}$,

$$\begin{aligned}
&\mathbb{P}(X_{n+1} = Y_{n+1} \mid X_n = i) \\
&= \mathbb{P}(X_{n+1} = j \mid X_n = i, Y_{n+1} = j, B = 1)\, \mathbb{P}(B = 1 \mid Y_{n+1} = j, X_n = i)\, \mathbb{P}(Y_{n+1} = j \mid X_n = i) \\
&+ \mathbb{P}(X_{n+1} = j \mid X_n = i, Y_{n+1} = j, B = 0)\, \mathbb{P}(B = 0 \mid Y_{n+1} = j, X_n = i)\, \mathbb{P}(Y_{n+1} = j \mid X_n = i) \\
&= \mathbb{P}(X_{n+1} = j \mid X_n = i, Y_{n+1} = j, B = 1)\alpha_{ij} q_{ij} + \mathbb{P}(X_{n+1} = j \mid X_n = i, Y_{n+1} = j, B = 0)(1 - \alpha_{ij}) q_{ij} \\
&= \begin{cases} q_{i,j}\alpha_{i,j} & \text{if } Y_{n+1} = j \neq i \\ q_{i,i} & \text{if } Y_{n+1} = j = i. \end{cases}
\end{aligned}$$

We used the fact that by the algorithm, if $j \neq i$, then

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, Y_{n+1} = j, B = 1) = 1$$

and if $j = i$, then both

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, Y_{n+1} = j, B = 1) = 1 \text{ and } \mathbb{P}(X_{n+1} = j \mid X_n = i, Y_{n+1} = j, B = 0) = 1.$$

We conclude by using the fact that since $\sum_{k \in S} \alpha_{ik} q_{ik} = 1$,

$$1 - \sum_{k \neq i} \alpha_{ik} q_{ik} = \alpha_{ii} q_{ii} = q_{ii}$$

since $\alpha_{i,i} = 1$.