Chapter 1

Discrete time Markov chains

In this course we consider a class of stochastic processes called Markov chains. The course is roughly equally divided between discrete-time and continuous-time Markov chains. We shall study various methods to help understand the behaviour of Markov chains, in particular over the long term.

1.1 What is a Markov chain?

1.1.1 Examples

A stochastic process is a mathematical model for a random development in time. Formally, a *discrete-time stochastic process* is a sequence $\{X_n : n = 0, 1, 2, \ldots\}$ of random variables. The value of the random variable X_n is interpreted as the state of the random development after n time steps.

Example 1.1. Suppose a virus can exist in two different strains α, β and in each generation either stays the same, or with probability $p \ll 1/2$ mutates to the other strain. Suppose the virus is in strain α initially, what is the probability that it is in the same strain after n generations?

We let X_n be the strain of the virus in the nth generation, which is a random variable with values in $\{\alpha, \beta\}$. The crucial point here is that the random variable X_n and X_{n+1} are *not independent* and things you have learnt about i.i.d. sequences of random variables do not apply here! To check this out note that

$$\mathbb{P}\{X_n = \alpha, X_{n+1} = \alpha\} = \mathbb{P}\{X_n = \alpha \text{ and no mutation occurs in step } n+1\} = (1-p)\mathbb{P}\{X_n = \alpha\},$$

$$\mathbb{P}\{X_n = \beta, X_{n+1} = \alpha\} = \mathbb{P}\{X_n = \beta \text{ and mutation occurs in step } n+1\} = p\mathbb{P}\{X_n = \beta\},$$

and hence

$$
\begin{aligned}
\mathbb{P}\{X_{n+1} = \alpha\} &= \mathbb{P}\{X_n = \alpha, X_{n+1} = \alpha\} + \mathbb{P}\{X_n = \beta, X_{n+1} = \alpha\} \\
&= (1-p)\mathbb{P}\{X_n = \alpha\} + p(1 - \mathbb{P}\{X_n = \alpha\}) \\
&= (1-2p)\mathbb{P}\{X_n = \alpha\} + p < (1-p).
\end{aligned}
$$

This gives,

$$\mathbb{P}(\{X_n = \alpha\} \cap \{X_{n+1} = \alpha\}) = (1-p)\mathbb{P}\{X_n = \alpha\} > \mathbb{P}\{X_{n+1} = \alpha\}\mathbb{P}\{X_n = \alpha\},$$

contradicting independence. Recall that another way to express the last formula is that

$$\mathbb{P}\{X_{n+1} = \alpha \mid X_n = \alpha\} > \mathbb{P}\{X_{n+1} = \alpha\}.$$

To study this process we therefore need a new theory, which is the theory of *discrete-time Markov chains.* The possible values of X_n are making up the statespace I of the chain, here $I = \{\alpha, \beta\}$. The characteristic feature of a Markov chain is that *the past influences the future only via the present.* For example you should check yourself that

$$\mathbb{P}\{X_{n+1} = \alpha \mid X_n = \alpha, X_{n-1} = \alpha\} = \mathbb{P}\{X_{n+1} = \alpha \mid X_n = \alpha\}.$$

Here the state of the virus at time $n + 1$ (future) does not depend on the state of the virus at time $n - 1$ (past) if the state at time n (present) is already known. The question raised above can be anserwed using this theory, we will give an intuitive argument below. ∎

Example 1.2 The simple random walk

A particle jumps about at random on the set $\mathbb{Z}$ of integers. At time 0 the particle is in fixed position $x \in \mathbb{Z}$. At each time $n \in \mathbb{N}$ a coin with probability p of heads and $q = 1 - p$ of tails is tossed. If the coin falls heads, then the particle jumps one position to the right, if the coin falls tails the particle jumps one position to the left. For $n \in \mathbb{N}$ the position X_n of the particle at time n is therefore

$$X_n = x + Y_1 + \cdots + Y_n = X_{n-1} + Y_n$$

where

$$Y_k := \begin{cases} 1 & \text{if } k\text{th toss is heads,} \\ -1 & \text{if } k\text{th toss is tails.} \end{cases}$$

The Y_ks are independent random variables with

$$\mathbb{P}\{Y_k = 1\} = p \text{ and } \mathbb{P}\{Y_k = -1\} = q.$$

The stochastic process $\{X_n : n \in \mathbb{N}\}$ has again the *time-parameter set* $\mathbb{N}$ and the *statespace* is the discrete (but infinite) set $\mathbb{Z}$ of integers. The rules of evaluation are given by the laws of Y_k and we note that

$$\begin{aligned}
&\mathbb{P}\{X_{n+1} = k \mid X_n = j, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1\} \\
&= \mathbb{P}\{Y_{n+1} = k - j \mid Y_n = j - x_{n-1}, Y_{n-1} = x_{n-1} - x_{n-2}, \ldots, Y_2 = x_2 - x_1, Y_1 = x_1 - x\}\} \\
&= \mathbb{P}\{Y_{n+1} = k - j\} = \begin{cases} 0 & \text{if } |k - j| \neq 1, \\ p & \text{if } k - j = 1, \\ q & \text{if } k - j = -1, \end{cases}
\end{aligned}$$

is again independent of $x_1, \ldots, x_{n-1}$. Observe that here the independence of the random variables Y_k plays an important role, though the random variables X_n are *not* independent. ∎

1.1.2 Intuition

Markov chain theory offers many important models for application and presents systematic methods to study certain questions. The existence of these systematic methods does not stop us from using intuition and common sense to guess the behaviour of the models.

For example, in Example 1.1 one can use the equation

$$\mathbb{P}\{X_{n+1} = \alpha \,|\, X_0 = \alpha\} = p\mathbb{P}\{X_n = \beta \,|\, X_0 = \alpha\} + (1-p)\mathbb{P}\{X_n = \alpha \,|\, X_0 = \alpha\}$$

together with

$$\mathbb{P}\{X_n = \beta \,|\, X_0 = \alpha\} = 1 - \mathbb{P}\{X_n = \alpha \,|\, X_0 = \alpha\},$$

to obtain for the desired quantity $p_n = \mathbb{P}\{X_n = \alpha \,|\, X_0 = \alpha\}$ the recursive relation

$$p_{n+1} = p(1 - p_n) + (1-p)p_n = p_n(1 - 2p) + p, \qquad p_0 = 1.$$

This has a unique solution given by

$$p_n = \frac{1}{2} + \frac{1}{2}(1 - 2p)^n.$$

As $n \to \infty$ this converges to the long term probability that the virus is in strain α, which is $1/2$ and therefore independent of the mutation probability p. The theory of Markov chains provides a systematic approach to this and similar questions.

1.1.3 Definition of discrete-time Markov chains

Suppose I is a *discrete*, i.e. finite or countably infinite, set. A *stochastic process* with statespace I and discrete time parameter set $\mathbb{N} = \{0, 1, 2, \ldots\}$ is a collection $\{X_n : n \in \mathbb{N}\}$ of random variables (on the same probability space) with values in I. The stochastic process $\{X_n : n \in \mathbb{N}\}$ is called a *Markov chain* with statespace I and discrete time parameter set $\mathbb{N}$ if its *law of evolution* is specified by the following:

(i) An *initial distribution* on the state space I given by a probability mass function $(w_i : i \in I)$, with $w_i \geq 0$ and $\sum_{i \in I} w_i = 1$.

(ii) A *one-step transition matrix* $P = (p_{ij} : i, j \in I)$ with $p_{ij} \geq 0$ for all $i, j \in I$ and

$$\sum_{j \in I} p_{ij} = 1 \qquad \text{for all } i \in I.$$

The law of evolution is given by

$$\boxed{\mathbb{P}\{X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\} = w_{x_0} p_{x_0 x_1} \cdots p_{x_{n-1} x_n}, \text{ for all } x_0, \ldots, x_n \in I.}$$

1.1.4 Discussion of the Markov property

Interpretation of the one-step transition matrix

All jumps from i to j in *one step* occur with probability p_{ij}, so if we fix a state i and observe where we are going in the next time step, then the probability distribution of the next location in I has the probability mass function $(p_{ij} : j \in I)$.

Given the present, the future is independent of the past

We have

$$\mathbb{P}\{X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\} = w_{x_0} p_{x_0 x_1} \cdots p_{x_{n-1} x_n},$$

and

$$\mathbb{P}\{X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}\} = w_{x_0} p_{x_0 x_1} \cdots p_{x_{n-2} x_{n-1}},$$

for all $x_0, \dots, x_n \in I$. Recall the definition of conditional probabilities

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Dividing the two equations gives

$$\mathbb{P}\{X_n = x_n \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}\} = p_{x_{n-1} x_n}. \tag{1.1.1}$$

In order to find $\mathbb{P}\{X_n = x_n \mid X_{n-1} = x_{n-1}\}$ we use the properties of the one-step transition matrix P to see that

$$
\begin{aligned}
\mathbb{P}\{X_n = x_n, \, X_{n-1} = x_{n-1}\} &= \sum_{x_0 \in I} \cdots \sum_{x_{n-2} \in I} \mathbb{P}\{X_n = x_n, \, \dots, X_0 = x_0\} \\
&= \sum_{x_0 \in I} \cdots \sum_{x_{n-2} \in I} w_{x_0} p_{x_0 x_1} \cdots p_{x_{n-2} x_{n-1}} p_{x_{n-1} x_n} \\
&= \sum_{x_0 \in I} \cdots \sum_{x_{n-2} \in I} \mathbb{P}\{X_0 = x_0, \dots, X_{n-1} = x_{n-1}\} p_{x_{n-1} x_n} \\
&= \mathbb{P}\{X_{n-1} = x_{n-1}\} \, p_{x_{n-1} x_n}.
\end{aligned}
$$

Hence,

$$\mathbb{P}\{X_n = x_n \mid X_{n-1} = x_{n-1}\} = p_{x_{n-1} x_n}. \tag{1.1.2}$$

Now compare (1.1.1) and (1.1.2). We get the formula known as the *Markov property*:

$$\boxed{\mathbb{P}\{X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_1 = x_1\} = \mathbb{P}\{X_n = x_n \mid X_{n-1} = x_{n-1}\} = p_{x_{n-1} x_n},}$$

for all $x_0, \dots, x_n \in I$ and $n \geq 1$.

1.1.5 Examples

Example 1.1 Recall the virus example. Initially the virus is in strain α, hence $w = (w_\alpha, w_\beta) = (1, 0)$. The P-matrix is

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

Example 1.2 For the simple random walk, the particle starts in a fixed point x, hence the initial distribution is given by

$$w_i = \mathbb{P}\{X_0 = i\} = \begin{cases} 1 & \text{if } i = x. \\ 0 & \text{otherwise.} \end{cases}$$

The one-step transition matrix is given by

$$p_{ij} = \mathbb{P}\{X_{n+1} = j \mid X_n = i\} = \begin{cases} 0 & \text{if } |j - i| \neq 1, \\ p & \text{if } j - i = 1, \\ q & \text{if } j - i = -1. \end{cases}$$

4

Example 1.3 Suppose that $X_0, X_1, X_2, \ldots$ is a sequence of independent and identically distributed random variables with

$$\mathbb{P}\{X_n = i\} = \mu(i) \text{ for all } n \in \mathbb{N}, i \in I,$$

for some finite statespace I and probability mass function $\mu : I \to [0,1]$. Then the initial distribution is $w = (w_i : i \in I)$ with $w_i = \mu(i)$ and the one-step transition matrix is

$$P = \begin{pmatrix} \mu(i_1) & \mu(i_2) & \cdots & \mu(i_n) \\ \mu(i_1) & \mu(i_2) & \cdots & \mu(i_n) \\ \cdots & \cdots & \cdots & \cdots \\ \mu(i_1) & \mu(i_2) & \cdots & \mu(i_n) \end{pmatrix}$$

where we enumerated the statespace $I = \{i_1, \ldots, i_n\}$.

Example 1.4   Random walk on a finite graph

A particle is moving on the graph below by starting on the top left vertex and at each time step moving along one of the adjacent edges to a neighbouring vertex, choosing the edge with equal probability and independently of all previous movements.
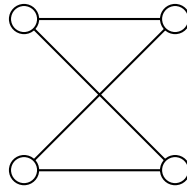
There are four vertices, which we enumerate from left to right by $\{1, \ldots, 4\}$. At time $n = 0$ we are in vertex 1. Hence $w = (1, 0, 0, 0)$ is the initial distribution. Each vertex has exactly two neighbours, so that it jumps to each neighbour with probability $1/2$. This gives

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

1.1.6   Fundamental questions about the long term behaviour

- Will the Markov chain converge to some "equilibrium regime"? And, what does this mean precisely?

- How much time does the Markov chain spend in the different states? What is the chain's favourite state? Does the answer depend on the starting position?

- How long does it take, on average, to get from some given state to another one?

1.2 n-step transition probabilities

Unless otherwise stated $X = \{X_n : n \in \mathbb{N}\}$ is a Markov chain with (discrete) statespace I and one-step transition matrix P. The Markov property shows that, whatever the initial distribution of the Markov chain is, we have

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

Let us consider a two-step transition,

$$
\begin{aligned}
\mathbb{P}\{X_{n+2} = j \mid X_n = i\} &= \sum_{k \in I} \mathbb{P}\{X_{n+2} = j, X_{n+1} = k \mid X_n = i\} \\
&= \sum_{k \in I} \mathbb{P}\{X_{n+1} = k \mid X_n = i\} \, \mathbb{P}\{X_{n+2} = j \mid X_{n+1} = k, X_n = i\} \\
&= \sum_{k \in I} p_{ik} p_{kj} = (P^2)_{ij},
\end{aligned}
$$

where P^2 is the product of the matrix P with itself. More generally,

$$\boxed{\mathbb{P}\{X_{n+k} = j \mid X_k = i\} = (P^n)_{ij}.}$$

Moreover, if the vector $(w_i : i \in I)$ is the initial distribution, we get

$$
\begin{aligned}
\mathbb{P}\{X_n = j\} &= \sum_{k \in I} \mathbb{P}\{X_0 = k\} \mathbb{P}\{X_n = j \mid X_0 = k\} \\
&= \sum_{k \in I} w_k \, (P^n)_{kj}.
\end{aligned}
$$

Hence we get

$$\boxed{\mathbb{P}\{X_n = j\} = (wP^n)_j.}$$

Hence, if we can calculate the matrix power P^n, we can find the distribution of X_n and the n-step transition probabilities, i.e. the probabilities of being in state j at time $n+k$ if we are in state i at time n.

If n is large (recall that we are particularly interested in the long term behaviour of the process!) it is not advisable to calculate P^n directly, but there are some more efficient methods, which we shall discuss now.

1.2.1 Calculation of matrix powers by diagonalization

We present the diagonalization method by an example. Suppose a village has three pubs along the main road. A customer decides after every pint whether he moves to the next pub on the left or on the right, and he chooses each option with the same probability. If there is no pub in this direction, he stays where he is for another pint. Here is a graph indicating the situation

The statespace of this Markov chain is $I = \{1, 2, 3\}$, where the three numbers indicate the three pubs and the one-step transition matrix is

$$
P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}
$$
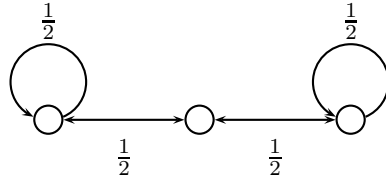
To find P^k for large k, we *diagonalize* P, if possible. This is happening in two steps.

Step 1 Find $\det(\lambda I - P)$, the characteristic polynomial, of P. Here

$$\lambda I - P = \begin{pmatrix} \lambda - \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \lambda & -\frac{1}{2} \\ 0 & -\frac{1}{2} & \lambda - \frac{1}{2} \end{pmatrix},$$

and hence

$$\begin{aligned}
\det(\lambda I - P) &= (\lambda - \tfrac{1}{2}) \det \begin{pmatrix} \lambda & -\frac{1}{2} \\ -\frac{1}{2} & \lambda - \frac{1}{2} \end{pmatrix} + \tfrac{1}{2} \det \begin{pmatrix} -\frac{1}{2} & 0 \\ -\frac{1}{2} & \lambda - \frac{1}{2} \end{pmatrix} \\
&= (\lambda - \tfrac{1}{2})(\lambda^2 - \lambda/2 - \tfrac{1}{4}) - \tfrac{1}{4}(\lambda - \tfrac{1}{2}) \\
&= (\lambda - \tfrac{1}{2})(\lambda - 1)(\lambda + \tfrac{1}{2}).
\end{aligned}$$

Thus P has three distinct eigenvalues

$$\lambda_1 = 1, \ \lambda_2 = \frac{1}{2}, \ \lambda_3 = -\frac{1}{2},$$

so it must be diagonalizable (which in this particular case was clear from the fact that P is symmetric).

Important remark: 1 is *always* an eigenvalue of p, because the vector $v = (1,\ldots,1)^T$ satisfies $Pv = v$. This is the fact that the row sums of P-matrices are 1!

Step 2 Now we find the corresponding eigenvectors v_1, v_2, v_3 with $Pv_i = \lambda_i v_i$. You can either solve the simultaneous equations $(P - \lambda_i I)v_i = 0$ or guess the answer.

Let S be the matrix with the eigenvectors as *columns*. In our case,

$$S = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & -1 & 1 \end{pmatrix},$$

then

$$S^{-1}PS = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} =: \Lambda.$$

Hence $P = S\Lambda S^{-1}$, and

$$P^2 = S\Lambda S^{-1} S\Lambda S^{-1} = S\Lambda^2 S^{-1}.$$

Continuing, we get

$$P^n = S \begin{pmatrix} \lambda_1^n & 0 & 0 \\ 0 & \lambda_2^n & 0 \\ 0 & 0 & \lambda_3^n \end{pmatrix} S^{-1}. \tag{1.2.1}$$

7

Hence we have P^n. Looking back at our example, in our case

$$S^{-1} = \frac{1}{6} \begin{pmatrix} 2 & 2 & 2 \\ 3 & 0 & -3 \\ 1 & -2 & 1 \end{pmatrix}.$$

If our friend starts at Pub 1, what is the probability that he is in Pub 2 after 3 pints, i.e. what is $\mathbb{P}\{X_3 = 2\}$? We find

$$P^3 = \frac{1}{6} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/8 & 0 \\ 0 & 0 & -1/8 \end{pmatrix} \begin{pmatrix} 2 & 2 & 2 \\ 3 & 0 & -3 \\ 1 & -2 & 1 \end{pmatrix} = \begin{pmatrix} 3/8 & 3/8 & 1/4 \\ 3/8 & 1/4 & 3/8 \\ 1/4 & 3/8 & 3/8 \end{pmatrix}.$$

(Check: row sums are still one, all entries positive!) Now we need the initial distribution, which is given by the probability mass function $w = (1, 0, 0)$ (start is in Pub 1) and we get

$$\mathbb{P}\{X_3 = 1\} = (wP^3)_1 = 3/8, \quad \mathbb{P}\{X_3 = 2\} = (wP^3)_2 = 3/8, \quad \mathbb{P}\{X_3 = 3\} = (wP^3)_3 = 1/4.$$

1.2.2 Tricks for Diagonalization

We continue with the same example and give some useful tricks to obtain a quick diagonalization of (small) matrices. In *step 1*, to find the eigenvalues of the matrix P, one has three automatic equations,

(1) 1 is always an eigenvalue for a P-matrix,

(2) the trace of P (=sum of the diagonal entries) equals the sum of the eigenvalues,

(3) the determinant of P is the product of the eigenvalues.

In the 2×2 or 3×3 case one may get the eigenvalues by solving the system of equations resulting from these three facts.

In our example we have $\lambda_1 = 1$ by Fact (1), Fact (2) gives $1 + \lambda_2 + \lambda_3 = 1$, which means $\lambda_2 = -\lambda_3$. Fact (3) gives $\lambda_2 \lambda_3 = -1/4$, hence $\lambda_2 = 1/2$, and $\lambda_3 = -1/2$ (or vice versa).

If we have distinct eigenvalues, the matrix is diagonalizable. Then from the representation of P^n in (1.2.1) we see that (in the 3×3 case) there exist matrices U_1, U_2 and U_3 with

$$P^n = \lambda_1^n U_1 + \lambda_2^n U_2 + \lambda_3^n U_3.$$

We know this without having to find the eigenvectors! Now taking the values $n = 0, 1, 2$ in the equation gives

(1) $U_1 + U_2 + U_3 = P^0 = I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$,

(2) $U_1 + \dfrac{1}{2}U_2 - \dfrac{1}{2}U_3 = P$,

(3) $U_1 + \dfrac{1}{4}U_2 + \dfrac{1}{4}U_3 = P^2$.

8

All one has to do is to solve the (matrix) simultaneous equations.

In the example $(1) - 4 \times (3)$ gives $-3U_1 = I - 4P^2$ and $(1) + 2 \times (2)$ gives $3U_1 + 2U_2 = I + 2P$. Now,

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

Hence

$$U_1 = \frac{1}{3}\left(4P^2 - I\right) = \frac{1}{3}\left[\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right] = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

And,

$$U_2 = \frac{1}{2}\left(I + 2P - 3U_1\right) = \frac{1}{2}\left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \right] = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix},$$

$$U_3 = I - U_1 - U_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \end{pmatrix}.$$

Hence, for all $n = 0, 1, 2, \ldots,$

$$P^n = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} + \left(\frac{1}{2}\right)^n \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} + \left(-\frac{1}{2}\right)^n \begin{pmatrix} \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \end{pmatrix}.$$

And recall that the n-step are just the entries of this matrix, i.e.

$$p_{ij}^{(n)} := \mathbb{P}\{X_n = j \mid X_0 = i\} = (P^n)_{ij},$$

for example $\mathbb{P}\{X_n = 2 \mid X_0 = 0\} = (P^n)_{02} = \frac{1}{3} - \frac{1}{2}\left(\frac{1}{2}\right)^n + \frac{1}{6}\left(-\frac{1}{2}\right)^n.$

As a further example look at

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

Then the eigenvalues are $\lambda_1 = 1$ (always) and λ_2, λ_3 satisfy $\lambda_1\lambda_2\lambda_3 = \det P = 0$, and $\lambda_1 + \lambda_2 + \lambda_3 = \text{trace } P = 1/2$. Hence $\lambda_2 = -1/2$ and $\lambda_3 = 0$. As the eigenvalues are distinct, the matrix P is diagonalizable and there exist matrices U_1, U_2, U_3 with

$$P^n = U_1 + \left(-\frac{1}{2}\right)^n U_2 + 0^n U_3, \text{ for all } n = 0, 1, 2, \ldots$$

Either take $n = 0, 1, 2$ and solve *or* find U_1 directly as

$$U_1 = \begin{pmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{pmatrix},$$

9

by solving $\pi P = \pi$ [see tutorials for this trick]. In the second case one only has to use the equations for $n = 0, 1$ and gets $I = U_1 + U_2 + U_3$ and $P = U_1 - \frac{1}{2}U_2$. Hence

$$U_2 = 2(U_1 - P) = \begin{pmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \\ -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

For $n \geq 1$ we do not need to find U_3 and get

$$P^n = \begin{pmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{pmatrix} + \left(-\frac{1}{2}\right)^n \begin{pmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \\ -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

1.2.3 The method of generating functions

This method if particularly efficient when we want to find certain entries of the matrix P^n without having to determine the complete matrix. The starting point is the geometric series,

$$1 + z + z^2 + \cdots = \frac{1}{1-z}, \text{ for all } |z| < 1.$$

For a finite square matrix A the same proof gives that

$$I + A + A^2 + \cdots = (I - A)^{-1},$$

if $\lim_{n \to \infty} A^n = 0$ and $I - A$ is invertible. This is the case if all the eigenvalues of A have modulus < 1.

Recall that the one-step transition matrix of a Markov chain does *not* fulfill this condition, as 1 is always an eigenvalue, however we have the following useful fact.

Lemma 1.1 *Suppose P is the one-step transition matrix of a finite state Markov chain, then all eigenvalues have modulus less or equal to one.*

We can thus use this series for the matrices θP for all $|\theta| < 1$. The idea is to expand $(I - \theta P)^{-1}$ as a power series in θ around $\theta = 0$ and to find P^n by comparing the coefficients of θ^n.

Let us look at an **example**. Let X be a Markov chain with statespace $I = \{0, 1, 2\}$ and one-step transition matrix

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

Then

$$I - \theta P = \begin{pmatrix} 1 & -\theta & 0 \\ -\frac{1}{4}\theta & 1 - \frac{1}{2}\theta & -\frac{1}{4}\theta \\ 0 & -\theta & 1 \end{pmatrix}.$$

To invert this matrix, recall that $A^{-1} = C^T / \det A$ where $C = \{c_{ij} : 0 \leq i, j \leq 2\}$ is the matrix of cofactors of A, i.e. c_{ij} is the determinant of the matrix A with the ith row and jth column removed,

multiplied with the factor $(-1)^{i+j}$. For example,

$$(I - \theta P)_{01}^{-1} = \frac{c_{10}}{\det(I - \theta P)} = \frac{-\det \begin{pmatrix} -\theta & 0 \\ -\theta & 1 \end{pmatrix}}{\det \begin{pmatrix} 1 - \frac{1}{2}\theta & -\frac{1}{4}\theta \\ -\theta & 1 \end{pmatrix} - (-\theta)\det \begin{pmatrix} -\frac{1}{4}\theta & -\frac{1}{4}\theta \\ 0 & 1 \end{pmatrix}}$$

$$= \frac{\theta}{1 - \frac{1}{2}\theta - \frac{1}{4}\theta^2 - \frac{1}{4}\theta^2} = \frac{\theta}{(1 - \theta)(1 + \frac{1}{2}\theta)}.$$

To expand this as a power series in θ we need to use partial fractions, write

$$\frac{\theta}{(1 - \theta)(1 + \frac{1}{2}\theta)} = \frac{a}{1 - \theta} + \frac{b}{1 + \frac{1}{2}\theta}.$$

then $\theta = a(1 + \frac{1}{2}\theta) + b(1 - \theta)$. Using this for $\theta = 1$ gives $a = 2/3$, and for $\theta = 0$ we get $b = -2/3$. We can thus continue with

$$(I - \theta P)_{01}^{-1} = \frac{2}{3}(1 - \theta)^{-1} - \frac{2}{3}(1 - (-\theta/2))^{-1} = \frac{2}{3}(1 + \theta + \theta^2 + \cdots) - \frac{2}{3}\left(1 - \frac{\theta}{2} + \frac{\theta^2}{4} - \frac{\theta^3}{8} + \cdots\right).$$

For the last expansion we have used the geometric series. Now we can compare the coefficients here with those coming from the matrix valued geometric series,

$$(I - \theta P)_{01}^{-1} = \sum_{n=0}^{\infty} \theta^n (P^n)_{01} = \sum_{n=0}^{\infty} \theta^n p_{01}^{(n)},$$

and get, for all $n \in \mathbb{N}$,

$$\mathbb{P}\{X_n = 1 \,|\, X_0 = 0\} = p_{01}^{(n)} = \frac{2}{3} - \frac{2}{3}\left(-\frac{1}{2}\right)^n.$$

You may wish to check that $p_{01}^{(0)} = 0$.

As another example we ask for the probability that, starting in state 0, we are again in state 0 after n time steps. Then,

$$(I - \theta P)_{00}^{-1} = \frac{c_{00}}{\det(I - \theta P)} = \frac{\det \begin{pmatrix} 1 - \frac{\theta}{2} & -\frac{\theta}{4} \\ -\theta & 1 \end{pmatrix}}{(1 - \theta)(1 + \frac{\theta}{2})} = \frac{1 - \frac{\theta}{2} - \frac{\theta^2}{4}}{(1 - \theta)(1 + \frac{\theta}{2})}.$$

Now we use partial fractions again (note that we need the α-term, as numerator and denominator have the same order),

$$\frac{1 - \frac{\theta}{2} - \frac{\theta^2}{4}}{(1 - \theta)(1 + \frac{\theta}{2})} = \alpha + \frac{\beta}{1 - \theta} + \frac{\gamma}{1 + \frac{\theta}{2}}.$$

To get α, β, γ, we look at

$$\alpha(1 - \theta)(1 + \frac{\theta}{2}) + \beta(1 + \frac{\theta}{2}) + \gamma(1 - \theta) = 1 - \frac{\theta}{2} - \frac{\theta^2}{4}.$$

Then $\theta = 1$ gives $\beta = 1/6$, $\theta = -2$ gives $\gamma = 1/3$ and comparing the θ^2 coefficient yields $\alpha = 1/2$. Hence,

$$
\begin{aligned}
(I - \theta P)_{00}^{-1} &= \frac{1}{2} + \frac{1}{6}(1-\theta)^{-1} + \frac{1}{3}\left(1 - \left(-\frac{\theta}{2}\right)\right)^{-1} \\
&= \frac{1}{2} + \frac{1}{6}(1 + \theta + \theta^2 + \theta^3 + \cdots) + \frac{1}{3}\left(1 - \frac{\theta}{2} + \frac{\theta^2}{4} - \cdots\right).
\end{aligned}
$$

Equating coefficients of θ^n gives,

$$
\mathbb{P}\{X_n = 0 \,|\, X_0 = 0\} =
\begin{cases}
\frac{1}{6} + \frac{1}{3}\left(-\frac{1}{2}\right)^n & \text{for } n \geq 1, \\
1 & \text{for } n = 0.
\end{cases}
$$

1.3   Hitting probabilities and expected waiting times

As always, X is a Markov chain with discrete statespace I and one-step transition matrix P. We now introduce a notation which allows us to look at the Markov chain and vary the initial distributions, this point of view turns out to be very useful later.

We denote by $\mathbb{P}_i$ the law of the chain X when X starts in the fixed state $i \in I$. In other words, for all $x_1, \ldots, x_n \in I$,
$$
\mathbb{P}_i\{X_1 = x_1, \ldots, X_n = x_n\} = p_{ix_1} p_{x_1 x_2} \cdots p_{x_{n-1} x_n}.
$$

We can also think of $\mathbb{P}_i$ as the law of the Markov chain conditioned on $X_0 = i$, i.e.

$$
\mathbb{P}_i(A) = \mathbb{P}\{A \,|\, X_0 = i\}.
$$

The law $\mathbb{P}_w$ of X with initial distribution w is then given as the *mixture* or weighted average of the laws $\mathbb{P}_i$, i.e. for the initial distribution $w = (w_i : i \in I)$ of the Markov chain and any event A,

$$
\boxed{\mathbb{P}_w(A) = \sum_{i \in I} \mathbb{P}_w(A \cap \{X_0 = i\}) = \sum_{i \in I} \mathbb{P}_w\{X_0 = i\}\mathbb{P}_i(A) = \sum_{i \in I} w_i \mathbb{P}_i(A).}
$$

We also use this notation for expected values, i.e. $\mathbb{E}_i$ refers to the expectation with respect to $\mathbb{P}_i$ and $\mathbb{E}_w$ refers to expectation with respect to $\mathbb{P}_w$.

Remember that in expressions such as $\mathbb{P}\{X_{n+1} = x_{n+1} \,|\, X_n = x_n\} = p_{x_n x_{n+1}}$ we do not have to write the index w because if we are told the state at time n we can forget about the initial distribution.

1.3.1   Hitting probabilities

We define the first hitting time T_j of a state $j \in I$ by

$$
T_j := \min\{n > 0 : X_n = j\}
$$

and understand that $T_j := \infty$ if the set is empty. Note the strict inequality here: we always have $T_j > 0$. Also observe the following equality of events,

$$
\{T_j = \infty\} = \{X \text{ never hits } j\}.
$$

Using this we can now define the *hitting probabilities* $(F_{ij} : i, j \in I)$, where F_{ij} is the probability that the chain started in i hits the state j (in positive, finite time). We let

$$F_{ij} := \mathbb{P}_i\{T_j < \infty\}.$$

Note that F_{ii} is in fact the probability that the chain started in i returns to i in finite time and need not be 1!

Example: We suppose a particle is moving on the integers $\mathbb{Z}$ as follows: it is started at the origin and in the first step moves to the left with probability $1/2$ and to the right with probability $1/2$. In each further step it moves one step away from the origin. (Exercise: write down the P-matrix!)
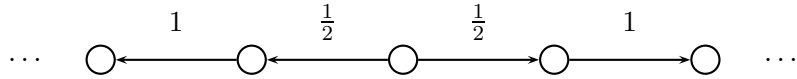
Here we have

$$F_{ij} = \begin{cases} 1 & \text{if } j > i > 0 \text{ or } j < i < 0, \\ 0 & \text{if } i < 0 < j \text{ or } i > 0 > j, \\ 1/2 & \text{if } i = 0, j \neq 0, \\ 0 & \text{if } i = j. \end{cases}$$

Theorem 1.2 *For a fixed state $b \in I$ the probabilities*

$$x_i := F_{i,b} := \mathbb{P}_i\{T_b < \infty\}, \text{ for } i \in I,$$

form the least non-negative solution of the system of equations,

$$x_i = \left(\sum_{j \neq b} p_{ij} x_j \right) + p_{ib} \quad \text{for } i \in I. \tag{1.3.1}$$

In particular, if $y_i \geq 0$ for all $i \in I$ form a solution of

$$y_i = \left(\sum_{j \neq b} p_{ij} y_j \right) + p_{ib}, \tag{1.3.2}$$

then $y_i \geq x_i$ for every $i \in I$.

Note: $y_i = 1$ for all $i \in I$ always solves (1.3.2).

Proof: First step: We show that $x_i = \mathbb{P}_i\{T_b < \infty\}$ is a solution of (1.3.1). Let $E = \{T_b < \infty\}$ be the event that the chain hits b. Then

$$\begin{aligned} x_i &= \mathbb{P}_i(E) = \sum_{j \in I} \mathbb{P}_i(E \cap \{X_1 = j\}) \\ &= \sum_{j \in I} \mathbb{P}_i\{X_1 = j\} \mathbb{P}_i\{E \mid X_1 = j\} \\ &= \sum_{j \in I} p_{ij} \mathbb{P}\{E \mid X_1 = j\}. \end{aligned}$$

Looking at the last probability we see that

$$\mathbb{P}\{E \mid X_1 = j\} = \begin{cases} 1 & \text{if } j = b, \\ \mathbb{P}_j\{T_b < \infty\} & \text{if } j \neq b. \end{cases}$$

This gives

$$x_i = \sum_{j \neq b} p_{ij} x_j + p_{ib}.$$

Second step: Suppose $y_i \geq 0$ forms a solution of (1.3.2). To finish the proof we have to show that $y_i \geq x_i$. Observe that,

$$
\begin{aligned}
y_i &= \left(\sum_{j \neq b} p_{ij} y_j \right) + p_{ib} \\
&= p_{ib} + \sum_{j \neq b} p_{ij} \left(\sum_{k \neq b} p_{jk} y_k + p_{jb} \right) \\
&= p_{ib} + \sum_{j \neq b} p_{ij} p_{jb} + \sum_{j \neq b} \sum_{k \neq b} p_{ij} p_{jk} y_k \\
&= \mathbb{P}_i\{X_1 = b\} + \mathbb{P}_i\{X_1 \neq b, X_2 = b\} + \sum_{j \neq b} \sum_{k \neq b} p_{ij} p_{jk} y_k.
\end{aligned}
$$

Repeated substitution yields,

$$
\begin{aligned}
y_i &= \mathbb{P}_i\{X_1 = b\} + \mathbb{P}_i\{X_1 \neq b, X_2 = b\} + \ldots + \mathbb{P}_i\{X_1 \neq b, \ldots, X_{n-1} \neq b, X_n = b\} \\
&\quad + \sum_{j_1 \neq b} \cdots \sum_{j_n \neq b} p_{ij_1} p_{j_1 j_2} \cdots p_{j_{n-1} j_n} y_{j_n}.
\end{aligned}
$$

Because $y_j \geq 0$ the last term is positive and we get, for all $n \in \mathbb{N}$,

$$
\begin{aligned}
y_i &\geq \mathbb{P}_i\{X_1 = b\} + \mathbb{P}_i\{X_1 \neq b, X_2 = b\} + \ldots + \mathbb{P}_i\{X_1 \neq b, \ldots, X_{n-1} \neq b, X_n = b\} \\
&= \mathbb{P}_i\{T_b = 1\} + \cdots + \mathbb{P}_i\{T_b = n\} \\
&= \mathbb{P}_i\{T_b \leq n\}.
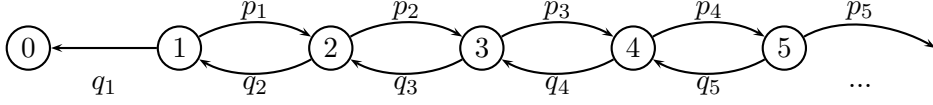\end{aligned}
$$

Hence, letting $n \to \infty$,

$$y_i \geq \lim_{n \to \infty} \mathbb{P}_i\{T_b \leq n\} = \mathbb{P}_i \left(\bigcup_n \{T_b \leq n\} \right) = \mathbb{P}_i\{T_b < \infty\} = x_i.$$

∎

Example Consider a Markov chain with state space $I = \{0, 1, 2, \ldots\}$ given by the diagram below, where, for $i = 1, 2, \ldots$, we have $0 < p_i = 1 - q_i < 1$. Here 0 is an absorbing state and

$$x_i := \mathbb{P}_i\{T_0 < \infty\}$$

is the hitting probability of state 0 if the chain is started in state $i \in I$.

For
$$\gamma_0 := 1, \gamma_i := \frac{q_i q_{i-1} \cdots q_1}{p_i p_{i-1} \cdots p_1} \qquad \text{for } i \geq 1,$$

- If $\sum_{i=0}^{\infty} \gamma_i = \infty$ we have $x_i = 1$ for all $i \in I$.

- If $\sum_{i=0}^{\infty} \gamma_i < \infty$, we have
$$x_i = \frac{\sum_{j=i}^{\infty} \gamma_j}{\sum_{j=0}^{\infty} \gamma_j} \text{ for } i \in I.$$

To prove this, we first use the 'one-step method' to find the equations $x_0 = 1$ and

$$x_i = q_i x_{i-1} + p_i x_{i+1}, \qquad \text{for } i \geq 1.$$

The solutions of this system are given by

$$x_i = 1 - A(\gamma_0 + \cdots \gamma_{i-1}), \qquad \text{for } i \geq 1,$$

for any choice of a fixed $A \in \mathbb{R}$. To prove this we have to

- check that the given form solves the equation,

- show that any solution is of this form.

For the first part we just plug in:

$$\begin{aligned} q_i x_{i-1} + p_i x_{i+1} &= q_i(1 - A(\gamma_0 + \cdots + \gamma_{i-2})) + p_i(1 - A(\gamma_0 + \cdots + \gamma_i)) \\ &= 1 - A(q_i(\gamma_0 + \cdots + \gamma_{i-2}) + p_i(\gamma_0 + \cdots + \gamma_i)), \end{aligned}$$

and

$$q_i(\gamma_0 + \cdots + \gamma_{i-2}) + p_i(\gamma_0 + \cdots + \gamma_i) = \gamma_0 + \cdots + \gamma_{i-2} + \frac{p_i q_{i-1} \cdots q_1 + q_i \cdots q_1}{p_{i-1} \cdots p_1} = \gamma_0 + \cdots + \gamma_{i-1}.$$

For the second part assume that $(x_i : i = 1, 2, \cdots)$ is *any* solution. Let $y_i = x_{i-1} - x_i$. Then $p_i y_{i+1} = q_i y_i$, which implies inductively that

$$y_{i+1} = \left(\frac{q_i}{p_i}\right) y_i = \gamma_i y_1.$$

Hence
$$x_0 - x_i = y_1 + \cdots + y_i = y_1(\gamma_0 + \cdots + \gamma_{i-1}).$$

Therefore, for $A := y_1$, we have
$$x_i = 1 - A(\gamma_0 + \cdots + \gamma_{i-1}).$$

15

To complete the proof we have to find the *smallest* nonnegative solution, which means we want to make A as large a spossible without making x_i negative. First suppose that $\sum_{i=0}^{\infty} \gamma_i = \infty$, then for any $A > 0$ the solution gets negative eventually, so that the largest nonnegative solution corresponds to the case $A = 0$, i.e. $x_i = 1$.

Next suppose that $\sum_{i=0}^{\infty} \gamma_i := M < \infty$. Then the solution remains nonnegative iff $1 - AM \geq 0$, i.e. if $A \leq 1/M$, so that the choice $A = 1/M$ gives the smallest solution. This solution is the one giving the right value for the hitting probabilities. ∎

Now we look at a simple random walk $\{X_n : n = 0, 1, 2\}$ with parameter $p \in [1/2, 1]$. If $X_0 = i > 0$, we are essentially in the situation of the previous example with $p_i = p$ and $q_i = 1 - p$. Then $\gamma_i = ((1-p)/p)^i$ and $\sum_{i=0}^{\infty} \gamma_i$ dicverges if $p = \frac{1}{2}$ and converges otherwise. Hence $F_{i0} = 1$ if $p = \frac{1}{2}$ and otherwise, for $i > 0$,

$$F_{i0} = \frac{\sum_{j=i}^{\infty} \left(\frac{1-p}{p}\right)^j}{\sum_{j=0}^{\infty} \left(\frac{1-p}{p}\right)^j} = \left(\frac{1-p}{p}\right)^i.$$

Note that, by the one-step method,

$$F_{00} = p\, F_{10} + q\, F_{-10} = p\,\frac{1-p}{p} + q = 1 - p + q = 2 - 2p.$$

Hence we can summarize our results for the random walk case as follows.

Theorem 1.3 *The hitting probabilities $F_{ij} = \mathbb{P}_i\{T_j < \infty\}$ for a simple random walk with parameter $p \in [1/2, 1]$ are given in the symmetric case $p = 1/2$ by*

$$F_{ij} = 1 \quad \text{for all } i, j \in \mathbb{Z},$$

and in the asymmetric case $p > 1/2$ by

$$F_{ij} = \begin{cases} 1 & \text{if } i < j, \\ 2 - 2p & \text{if } i = j, \\ \left(\frac{1-p}{p}\right)^{i-j} & \text{if } i > j. \end{cases}$$

By looking at $-X$ the theorem also gives full information about the case $p < 1/2$. As a special case note that, for all $p \in [0, 1]$,

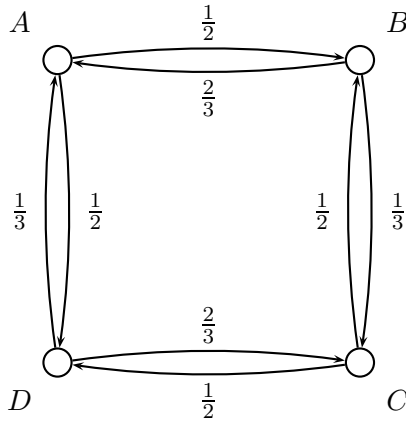$$\mathbb{P}_0\{X \text{ does not return to } 0\} = |p - q|.$$

1.3.2 Expected waiting times

In a manner similar to the problem of hitting probabilities, Theorem 1.2, one can prove the following theorem for the expected waiting time until a state b is hit for the first time.

Theorem 1.4 *Fix a state $b \in I$ and let $y_i := \mathbb{E}_i\{T_b\}$. Then*

$$y_i = 1 + \sum_{j \neq b} p_{ij} y_j \tag{1.3.3}$$

and y is the least nonnegative solution of this equation.

Note that $y_i = \infty$ is possible for some, or even all, i, and the convention $0 \times \infty = 0$ is in place.

Example 1.2 We again look at the simple random walk and calculate $y_i = \mathbb{E}\{T_0\}$, the average time taken to reach the origin if we start from the state i. To save some work we use the intuitively obvious extra equation $y_n = ny_1$ for all $n \geq 1$. Combining this with (1.3.3) for $i = 1$ gives

$$y_1 = 1 + py_2 = 1 + 2py_1.$$

If $p < 1/2$ (the case of a downward drift) we get the solution

$$\boxed{\mathbb{E}_n\{T_0\} = \frac{n}{1 - 2p} \text{ for all } n \geq 1.}$$

If $p \geq 1/2$ this gives $y_1 \geq 1 + y_1$, which implies $y_1 = \infty$. In particular, for $p = 1/2$ we get,

$$\boxed{\begin{array}{c} \text{The average waiting time until a symmetric, simple} \\ \text{random walk travels from state } i \text{ to state } j \text{ is infinite for all } i \neq j. \end{array}}$$

However, we know that the waiting time is finite almost surely! ∎

Example 1.6 Consider a Markov chain with statespace $I = \{A, B, C, D\}$ and jump probabilities given by the diagram.

Problem 1: Find the expected time until the chain started in C reaches A.

Solution: Let $x_i = \mathbb{E}_i\{T_A\}$. By considering the first step and using the Markov property (or just using (1.3.3)) we get

$$\begin{aligned} x_C &= 1 + \frac{1}{2}x_B + \frac{1}{2}x_D \\ x_B &= 1 + \frac{2}{3}0 + \frac{1}{3}x_C \\ x_D &= 1 + \frac{1}{3}0 + \frac{2}{3}x_C. \end{aligned}$$

Hence,

$$x_C = 1 + \tfrac{1}{2}(1 + \tfrac{1}{3}x_C) + \tfrac{1}{2}(1 + \tfrac{2}{3}x_C) = 2 + \tfrac{1}{2}x_C,$$

which implies $\mathbb{E}_C\{T_A\} = x_C = 4$. The expected time until the chain started in C reaches A is 4.

Problem 2: What is the probability that the chain started in A reaches the state C before B?

Solution: Let $x_i = \mathbb{P}_i\{T_C < T_B\}$. By considering the first step and using the Markov property we get

$$
\begin{aligned}
x_A &= \frac{1}{2}x_D + \frac{1}{2}0, \\
x_D &= \frac{1}{3}x_A + \frac{2}{3}1.
\end{aligned}
$$

Hence $2x_A = (1/3)x_A + (2/3)$, which gives $x_A = 2/5$. Hence the probability of hitting C before B when we start in A is 2/5. ∎

1.4 Classification of states and the renewal theorem

We now begin our study of the long-term behaviour of the Markov chain. Recall that $T_i = \inf\{n > 0 : X_n = i\}$ and

$$
F_{ii} = \mathbb{P}_i\{T_i < \infty\} = \mathbb{P}_i\{X \text{ returns to } i \text{ at some positive time}\}.
$$

Definition The state i is called

- *transient* if $F_{ii} < 1$, i.e. if there is a positive probability of escape from i,

- *recurrent* or *persistent* if $F_{ii} = 1$, i.e. if the chain returns to state i almost surely, and hence infinitely often.

Let us look at the *total number of visits* to a state i given by the random variable

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

where $1_{\{X_n = i\}}$ takes the value 1 if $X_n = i$ and 0 otherwise. Note that we include time $n = 0$.

If i is recurrent we have $\mathbb{P}_i\{V_i = \infty\} = 1$, in particular the expected number of visits to state i is $\mathbb{E}_i\{V_i\} = \infty$. If i is transient, then, for $n \geq 1$,

$$
\mathbb{P}_i\{V_i = n\} = F_{ii}^{n-1}(1 - F_{ii}),
$$

hence

$$
\mathbb{E}_i\{V_i\} = \sum_{n=1}^{\infty} n\mathbb{P}_i\{V_i = n\} = (1 - F_{ii}) \sum_{n=1}^{\infty} nF_{ii}^{n-1} = \frac{1 - F_{ii}}{(1 - F_{ii})^2} = \frac{1}{1 - F_{ii}} < \infty,
$$

recalling $\sum_{n=1}^{\infty} nx^{n-1} = (1 - x)^{-2}$ for $|x| < 1$.

Example: We look at the simple random walk again, focusing on the case $p > q$ of an upward drift. We know that $F_{ii} = 2q = 1 - (p - q)$, so $\mathbb{E}_i V_i = (p - q)^{-1}$.

Intuitively, this can be explained by recalling that $X_n/n \to p - q$. For large n, the walk must visit $\approx n(p - q)$ states in its first n steps, so it can spend a time of roughly $1/(p - q)$ in each state.

1.4.1 The renewal theorem

The aim of this section is to study the long term asymptotics of $\mathbb{P}_i\{X_n = i\} = (P^n)_{ii}$. We start by deriving a second formula for $\mathbb{E}_i\{V_i\}$. Directly from the definition of V_i we get

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

Recall that $\mathbb{P}_i\{X_n = i\} = (P^n)_{ii}$. Hence,

$$\boxed{\mathbb{E}_i\{V_i\} = \sum_{n=0}^{\infty}(P^n)_{ii}.}$$

We thus get the renewal theorem in the transient case.

Theorem 1.5 (Renewal Theorem in the transient case) *If i is a transient state of the Markov chain X, then*

$$\sum_{n=0}^{\infty}(P^n)_{ii} = \mathbb{E}_i\{V_i\} = \frac{1}{1 - F_{ii}} < \infty.$$

In particular, we have $\lim_{n\to\infty}(P^n)_{ii} = 0$.

The more interesting case of the renewal theorem refers to the recurrent case. In this case, $\mathbb{E}_i\{V_i\} = \sum_{n=0}^{\infty}(P^n)_{ii} = \infty$, leaving *open* whether $\lim_{n\to\infty}(P^n)_{ii} = 0$. In fact, as we shall see below, both cases can occur.

Definition A recurrent state i is called

- *positive recurrent* if $\mathbb{E}_i\{T_i\} < \infty$, i.e. if the mean time until return to i is finite,

- *null recurrent* if $\mathbb{E}_i\{T_i\} = \infty$.

For **example** we have seen before that in the symmetric simple random walk $\mathbb{E}_0\{T_0\} = \infty$, so 0 (and all other states) is null recurrent.

If we want study the limiting behaviour of $(P^n)_{ii}$ we first have to deal with the problem of *periodicity*. For example, for simple random walk we have

$$(P^n)_{ii} \begin{cases} = 0 & \text{if } n \text{ odd },\\ > 0 & \text{otherwise.} \end{cases}$$

We can only expect interesting behaviour for the limit of $(P^{2n})_{ii}$.

Generally, we define the *period* of a state $i \in I$ as the greatest common divisor of the set $\{n > 0 : (P^n)_{ii} > 0\}$. We write $d(i)$ for the period of the state i. For example, the period of every state in the simple random walk is 2. For another example let

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

Although one cannot return to the first state imediately, the period is one.

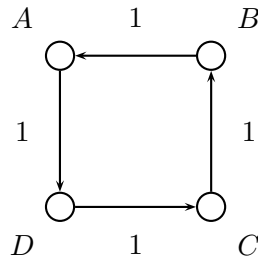**Theorem 1.6 (Renewal Theorem (main part))**

(a) If i is a null recurrent state, then $\lim_{n\to\infty}(P^n)_{ii} = 0$.

(b) If i is positive recurrent, then $(P^n)_{ii} = 0$ if n is not a multiple of $d(i)$. Otherwise,

$$\lim_{n\to\infty}(P^{nd(i)})_{ii} = \frac{d(i)}{\mathbb{E}_i\{T_i\}}.$$

We omit the proof and look at examples instead.

Example 1.7 This is a trivial example without randomness, a particle moves always one step counterclockwise through the graph.

Here the period of every state is 4 and the average return time is also 4, the transition probabilities satisfy $(P^{4n})_{ii} = 1$. The theorem holds (even without the limit!)

Example 1.8 Consider the example given by the following graph.
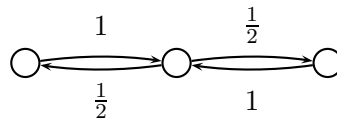
The one-step transition matrix is given by

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

All states have period 2. We can find

(a) P^n by diagonalization with tricks,

(b) $\mathbb{E}_i\{T_i\}$ by the "one-step method".

Then we can verify the statement of the theorem in our example.

(a) We have trace $P = 0$, $\det P = 0$ and hence eigenvalues $1, 0, -1$. Solving $\pi P = \pi$ gives

$$\pi = (1/4, 1/2, 1/4).$$

Hence

$$P^n = U_1 + (-1)^n U_2 + 0^n U_3 \text{ for all } n \geq 1,$$

and

$$U_1 = \begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/2 & 1/4 \end{pmatrix},$$

$$U_2 = U_1 - P = \begin{pmatrix} 1/4 & -1/2 & 1/4 \\ -1/4 & 1/2 & -1/4 \\ 1/4 & -1/2 & 1/4 \end{pmatrix},$$

and U_3 is irrelevant. For example, we get

$$(P^n)_{33} = \frac{1}{4} + (-1)^n \frac{1}{4} \text{ for all } n \geq 1.$$

(b) For $y_i = \mathbb{E}_i\{T_3\}$ we get the equations

$$
\begin{aligned}
y_1 &= 1 + y_2 \\
y_2 &= 1 + \tfrac{1}{2}y_1 \\
y_3 &= 1 + y_2.
\end{aligned}
$$

Solving gives, for example, $\mathbb{E}_3\{T_3\} = y_3 = 4$, $\mathbb{E}_1\{T_1\} = 4$ by symmetry and trivially $\mathbb{E}_2\{T_2\} = 2$. Checking the renewal theorem we observe that $(P^n)_{33} = 0$ if n is odd and

$$\lim_{n \to \infty} (P^{2n})_{33} = \lim_{n \to \infty} \frac{1}{4} + (-1)^{2n} \frac{1}{4} = \frac{1}{2} = \frac{d(2)}{\mathbb{E}_3\{T_3\}}.$$

1.4.2 Class properties and irreducibility

We say that the state i *communicates* with the state j, and write $i \leftrightarrow j$ if there exist $n, m \geq 0$ with $(P^n)_{ij} > 0$ and $(P^m)_{ji} > 0$. The relation $\leftrightarrow$ is an equivalence relation on the state space I, because

- $i \leftrightarrow i$,

- $i \leftrightarrow j$ implies $j \leftrightarrow i$,

- $i \leftrightarrow j$ and $j \leftrightarrow k$ implies $i \leftrightarrow k$.

Only the last statement is nontrivial. To prove it assume $i \leftrightarrow j$ and $j \leftrightarrow k$. Then there exist $n_1, n_2 \geq 0$ with $(P^{n_1})_{ij} > 0$ and $(P^{n_2})_{jk} > 0$. Then,

$$
\begin{aligned}
(P^{n_1+n_2})_{ik} &= \mathbb{P}_i\{X(n_1 + n_2) = k\} \\
&\geq \mathbb{P}_i\{X(n_1) = j, \, X(n_1 + n_2) = k\} \\
&\geq \mathbb{P}_i\{X(n_1) = j\}\mathbb{P}_i\{X(n_1 + n_2) = k \mid X(n_1) = j\} \\
&= (P^{n_1})_{ij}(P^{n_2})_{jk} > 0.
\end{aligned}
$$

Similarly, there exist m_1, m_2 with $(P^{m_1+m_2})_{ki} > 0$ and hence $i \leftrightarrow k$.

Since $\leftrightarrow$ is an equivalence relation, we can define the corresponding equivalence classes, which are called *communicating classes*. The class of i consists of all $j \in I$ with $i \leftrightarrow j$. A property of a state is a *class property* if whenever it holds for one state, it holds for all states in the same class. The following properties are class properties,

- i is transient,

- i is positive recurrent,

- i is null recurrent,

- i has period d.

Now we can attribute the property to the class, saying for example that a class has period d, etc. We can decompose the statespace I as a disjoint union

$$I = T \cup R_1 \cup R_2 \cup \dots,$$

where T is the *set* of transient states and $R_1, R_2, \dots$ are the recurrent classes.

If X starts in T, it can either stay in T forever (somehow drifting off to infinity) or get trapped (and stay forever) in one of the recurrent classes.
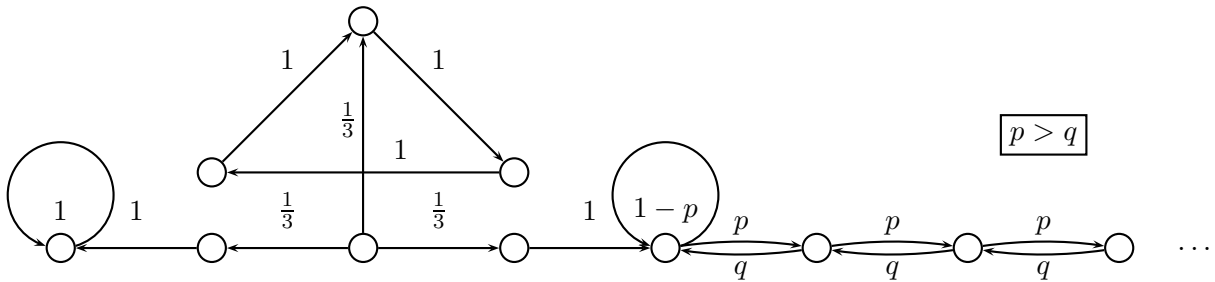

Figure 1.1: 6 classes, 2 recurrent, 4 transient

A Markov chain is called *irreducible* if all states communicate with each other, i.e. if I is the only communicating class. Thanks to the decomposition the study of general chains is frequently reduced to the study of irreducible chains.

1.5 The Big Theorem

1.5.1 The invariant distribution

Let π be a probability mass function on I, i.e. $\pi : I \to [0,1]$ with $\sum_{i \in I} \pi_i = 1$. π is called an *invariant distribution* or *equilibrium distribution* if $\pi P = \pi$, that is

$$\sum_{i \in I} \pi_i p_{ij} = \pi_j \text{ for all } j.$$

If such a π exists, then we can use is at as initial distribution to the following effect,

$$\mathbb{P}_\pi\{X_1 = j\} = \sum_{i \in I} \mathbb{P}_\pi\{X_0 = i, X_1 = j\} = \sum_{i \in I} \pi_i p_{ij} = \pi_j = \mathbb{P}_\pi\{X_0 = j\},$$

and, more generally,

$$\boxed{\mathbb{P}_\pi\{X_n = j\} = \pi_j \text{ for all } j \in I.}$$

In other words the law of X_n under $\mathbb{P}_\pi$ is π at all times n, we say that the system is in *equilibrium*.

We now assume that the chain X is irreducible. The all states have the same period d. The chain is called *aperiodic* if $d = 1$.

Theorem 1.7 (The Big Theorem) *Let X be an irreducible chain. Then the following statements are equivalent:*

- *X has a positive recurrent state,*

- *all states of X are positive recurrent,*

- *X has an invariant distribution π.*

If this holds, the invariant distribution is given by

$$\pi_i = \frac{1}{\mathbb{E}_i\{T_i\}} > 0.$$

Moreover,

(a) For all initial distributions w, with probability 1,

$$\frac{1}{n}\Big(\#visits\ to\ state\ i\ by\ time\ n\Big) \longrightarrow \pi_i.$$

(b)

$$\mathbb{E}_j\{\#visits\ to\ state\ i\ before\ T_j\} = \frac{\pi_i}{\pi_j}\ for\ all\ i \neq j.$$

(c) In the aperiodic case $d = 1$, we have for all initial distributions w,

$$\lim_{n\to\infty} \mathbb{P}_w\{X_n = j\} = \pi_j\ for\ all\ j \in I.$$

Note how (b) tallies with the following fact deduced from (a),

$$\frac{\#\text{visits to state } i \text{ by time } n}{\#\text{visits to state } j \text{ by time } n} \longrightarrow \frac{\pi_i}{\pi_j}.$$

We do not give the full proof here, but sketch the **proof of (c)**, because it is a nice example of a *coupling argument*. We let X be the Markov chain with initial distribution w and Y an independent Markov chain with the same P-matrix and initial distribution π. The proof comes in two steps.

Step 1. Fix any state $b \in I$ and let $T = \inf\{n > 0 : X_n = b \text{ and } Y_n = b\}$.
We show that $\mathbb{P}\{T < \infty\} = 1$.

The process $W = \{W_n = (X_n, Y_n) : n \in \mathbb{N}\}$ is a Markov chain with statespace $I \times I$ and n-step transition probabilities

$$(\tilde{P}^n)_{(i,k)(j,l)} = (P^n)_{ij}(P^n)_{kl},$$

which is positive for sufficiently large n since P is aperiodic. Hence W is irreducible. W has an invariant distribution given by $\tilde{\pi}_{(i,k)} = \pi_i \pi_k$ and hence — by the first part of the big theorem— it must be positive recurrent. Positive recurrence implies that the expected first hitting time of every state is finite with probability one, see Q1 on Sheet 6. Now observe that

$$T = \inf\{n > 0 \,:\, X_n = b \text{ and } Y_n = b\} = \inf\{n > 0 \,:\, W_n = (b,b)\},$$

and note that we have shown that $T < \infty$ almost surely.

Step 2. The trick is to use the finite time T to switch from the chain X to the chain Y. Let Z be the Markov chain given by

$$Z_n = \begin{cases} X_n & \text{if } n \leq T, \\ Y_n & \text{if } n \geq T. \end{cases}$$

It is intuitively obvious and not hard to show that Z is a Markov chain with the same P-matrix as X and initial distribution w. We have

$$\mathbb{P}\{Z_n = j\} = \mathbb{P}\{X_n = j \text{ and } n < T\} + \mathbb{P}\{Y_n = j \text{ and } n \geq T\}.$$

Hence,

$$\begin{aligned}
|\mathbb{P}\{X_n = j\} - \pi_j| &= |\mathbb{P}\{Z_n = j\} - \mathbb{P}\{Y_n = j\}| \\
&= |\mathbb{P}\{X_n = j \text{ and } n < T\} - \mathbb{P}\{Y_n = j \text{ and } n < T\}| \\
&\leq \mathbb{P}\{n < T\}.
\end{aligned}$$

As $\mathbb{P}\{T < \infty\} = 1$ the last term converges to 0 and we are done. ∎

1.5.2 Time-reversible Markov chains

We discuss the case of time-reversible or symmetrizable Markov chains. Suppose $(m_i \,:\, i \in I)$ is a collection of nonnegative numbers, not all zero. We call a Markov chain m-symmetrizable if we have

$$m_i p_{ij} = m_j p_{ji} \text{ for all } i,j \in I.$$

These equations are called *detailed balance equations*. They imply that

$$\sum_{i \in I} m_i p_{ij} = m_j \sum_{i \in I} p_{ji} = m_j \text{ for all } j \in I.$$

If, moreover, $M = \sum_{i \in I} m_i < \infty$, then

- $\pi_i = \frac{m_i}{M}$ is an invariant distribution and also solves the detailed balance equations,

- $\mathbb{P}_\pi\{X_0 = i_0, \ldots, X_n = i_n\} = \mathbb{P}_\pi\{X_0 = i_n, \ldots, X_n = i_0\}$.

In other words, under the law $\mathbb{P}_\pi$ the sequence $X_0, \ldots, X_n$ has the same law as the time-reversed sequence $X_n, \ldots, X_0$. Note that both statements are very easy to check!

Remarks:

- If X is π-symmetrizable, then π is an invariant distribution of X. But conversely, if π is an invariant distribution of X this does *not* imply that X is π-symmetrizable!

- It is sometimes much easier to solve the detailed balance equations and thus find an invariant distribution, rather than solving $\pi P = \pi$, see Example 1.11 below.

- If the invariant distribution does not solve the detailed balance equations, then they have no solution.

Example 1.9 Let X be a Markov chain with state space $I = \{0, 1, 2\}$ and transition matrix

$$P = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/6 & 1/3 \\ 1/4 & 1/6 & 7/12 \end{pmatrix}$$

- Find the equilibrium distribution π.

- Is P π-symmetrizable?

π has to satisfy $\pi_0 + \pi_1 + \pi_2 = 1$ and, from $\pi P = \pi$,

$$\begin{aligned} \tfrac{1}{2}\pi_0 + \tfrac{1}{2}\pi_1 + \tfrac{1}{4}\pi_2 &= \pi_0, \\ \tfrac{1}{4}\pi_0 + \tfrac{1}{6}\pi_1 + \tfrac{1}{6}\pi_2 &= \pi_1, \\ \tfrac{1}{4}\pi_0 + \tfrac{1}{3}\pi_1 + \tfrac{7}{12}\pi_2 &= \pi_2. \end{aligned}$$

This can be solved to give $\pi = (2/5, 1/5, 2/5)$. To check symmetrizability we have to verify $\pi_i p_{ij} = \pi_j p_{ji}$ for all $i, j \in \{0, 1, 2\}$. There are three non-trivial equations to be checked,

$$\begin{aligned} \pi_0 p_{01} = \pi_1 p_{10} &\Leftrightarrow \pi_0 = 2\pi_1 \\ \pi_0 p_{02} = \pi_2 p_{20} &\Leftrightarrow \pi_2 = \pi_0 \\ \pi_1 p_{12} = \pi_2 p_{21} &\Leftrightarrow \pi_2 = 2\pi_1. \end{aligned}$$

This is satisfied for our π. In fact one could have started with these equations and $\pi_0 + \pi_1 + \pi_2 = 1$ and the only solution is the invariant distribution π.

Example 1.10 Let X be a Markov chain with state space $I = \{0, 1, 2\}$ and transition matrix

$$P = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/6 & 1/3 \\ 1/2 & 1/6 & 1/3 \end{pmatrix}.$$

This matrix is not symmetrizable. Trying to find a suitable m leads to the equations

$$\begin{aligned} m_0 p_{01} = m_1 p_{10} &\Leftrightarrow m_0 = 2m_1 \\ m_0 p_{02} = m_2 p_{20} &\Leftrightarrow m_0 = 2m_2 \\ m_1 p_{12} = m_2 p_{21} &\Leftrightarrow m_2 = 2m_1. \end{aligned}$$

These equations have only the trivial solution $m_0 = m_1 = m_2 = 0$, which is not permitted in the definition of symmetrizability! Still, one can find an invariant distribution $\pi = (1/2, 5/24, 7/24)$.

1.6 Finding the invariant distribution using generating functions

Let X be an irreducible chain. The Big Theorem tells us that it is worth trying to find the invariant distribution, because it can tell us lots about the long term behaviour of the Markov chain, for example the *ergodic principle*

$$\frac{1}{n}\Big(\#\text{visits to state } j \text{ by time } n\Big) \longrightarrow \pi_j$$

and, in the aperiodic case,

$$\lim_{n\to\infty} \mathbb{P}_w\{X_n = j\} = \pi_j \text{ for all } j \in I.$$

If the state space $I = \mathbb{N}$, then we cannot find π by solving a finite system of linear equations, as before. Instead the powerful method of generating functions is available.

The *generating function* $\hat{\pi}$ of an invariant distribution is given by

$$\hat{\pi}(s) = \sum_{n=0}^{\infty} \pi_n s^n.$$

One of our requirements for an invariant distribution is that $\hat{\pi}(1) = 1$, which is equivalent to $\sum_{n=0}^{\infty} \pi_n = 1$. We study the method by looking at an example.

Example 1.11 Let X be a Markov chain with statespace $I = \mathbb{N}$ and one-step transition matrix

$$P = \begin{pmatrix} 6/7 & 0 & 1/7 & 0 & \cdots & \cdots \\ 6/7 & 0 & 0 & 1/7 & 0 & \cdots \\ 0 & 6/7 & 0 & 0 & 1/7 & \cdots \\ \cdots & \cdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

The system $\pi P = \pi$ gives

$$
\begin{aligned}
\frac{6}{7}\pi_0 + \frac{6}{7}\pi_1 &= \pi_0 \\
\frac{6}{7}\pi_2 &= \pi_1 \\
\frac{1}{7}\pi_0 + \frac{6}{7}\pi_3 &= \pi_2 \\
\frac{1}{7}\pi_1 + \frac{6}{7}\pi_4 &= \pi_3,
\end{aligned}
$$

and so on. Multiplying the equations above by $s, s^2, s^3, \dots$ gives for the generating function

$$\frac{1}{7}s^3\hat{\pi}(s) + \frac{6}{7}\Big(\hat{\pi}(s) - \pi_0\Big) + \frac{6}{7}s\pi_0 = s\hat{\pi}(s).$$

Rearranging gives,

$$\hat{\pi}(s)\Big(\frac{1}{7}s^3 - s + \frac{6}{7}\Big) = \pi_0\Big(\frac{6}{7} - \frac{6}{7}s\Big),$$

hence

$$\hat{\pi}(s) = \pi_0 \frac{6(1-s)}{s^3 - 7s + 6}.$$

To find π_0 recall $\hat{\pi}(1) = 1$. Here are two ways to use this.

First Method: Use L'Hôpital's rule to get

$$\hat{\pi}(1) = \lim_{s \to 1} \pi_0 \frac{-6}{3s^2 - 7} = \frac{3}{2}\pi_0,$$

hence $\pi_0 = \frac{2}{3}$.

Better Method: Factorize and cancel the common factor $1 - s$ if possible. Here

$$\hat{\pi}(s) = \pi_0 \frac{-6(1-s)}{(1-s)(s^2 + s - 6)}.$$

Hence $\hat{\pi}(1) = 6\pi_0/4$ and we arrive at the same conclusion.

Altogether

$$\hat{\pi}(s) = \frac{-4}{s^2 + s - 6} = \frac{-4}{(s+3)(s-2)}.$$

Once we have found $\hat{\pi}$ we can find π_n by writing $\hat{\pi}$ as a power series and equating coefficients. For this purpose first use *partial fractions*

$$\frac{-4}{(s+3)(s-2)} = \frac{a}{s+3} + \frac{b}{s-2},$$

which gives $-4 = a(s-2) + b(s+3)$ and $s = 2$ gives $b = -4/5$ and $s = -3$ gives $a = 4/5$, hence

$$\hat{\pi}(s) = \frac{4}{5}\frac{1}{s+3} - \frac{4}{5}\frac{1}{s-2}.$$

Now we use $(1-s)^{-1} = 1 + s + s^2 + s^3 + \cdots$ and obtain

$$\hat{\pi}(s) = \frac{4}{5}\left(\frac{1}{3}\left(1 - \frac{s}{3} + \frac{s^2}{3^2} - \frac{s^3}{3^3} + \cdots\right) + \frac{1}{2}\left(1 + \frac{s}{2} + \frac{s^2}{2^2} + \frac{s^3}{2^3} + \cdots\right)\right).$$

Hence the coefficient at s^n is, for all $n \geq 0$,

$$\pi_n = \frac{4}{5}\left(\frac{1}{2^{n+1}} + \frac{(-1)^n}{3^{n+1}}\right).$$

We have thus found the invariant distribution π.

To find the long term average state of the system, recall that we just have to find the mean of the distribution π, which is

$$\sum_{n=0}^{\infty} n\pi_n = \hat{\pi}'(1),$$

because $\hat{\pi}'(s) = \sum_{n=0}^{\infty} n\pi_n s^{n-1}$ by taking termwise derivatives of a power series. In our example

$$\hat{\pi}'(s) = \frac{4(1+2s)}{(s^2 + s - 6)^2},$$

then $\hat{\pi}'(1) = 3/4$ is the long time average state.