Markov Chains

These notes contain material prepared by colleagues who have also presented this course at Cambridge, especially James Norris. The material mainly comes from books of Norris, Grimmett & Stirzaker, Ross, Aldous & Fill, and Grinstead & Snell. Many of the examples are classic and ought to occur in any sensible course on Markov chains.

Contents

Table of Contents **i**

Schedules **iv**

1 Definitions, basic properties, the transition matrix **1**
- 1.1 An example and some interesting questions . . . 1
- 1.2 Definitions . . . 2
- 1.3 Where do Markov chains come from? . . . 3
- 1.4 How can we simulate them? . . . 3
- 1.5 The n-step transition matrix . . . 3
- 1.6 $P^{(n)}$ for a two-state Markov chain . . . 4

2 Calculation of n-step transition probabilities, class structure, absorption, and irreducibility **5**
- 2.1 Example: a three-state Markov chain . . . 5
- 2.2 Example: use of symmetry . . . 6
- 2.3 Markov property . . . 6
- 2.4 Class structure . . . 7
- 2.5 Closed classes . . . 8
- 2.6 Irreducibility . . . 8

3 Hitting probabilities and mean hitting times **9**
- 3.1 Absorption probabilities and mean hitting times . . . 9
- 3.2 Calculation of hitting probabilities and mean hitting times . . . 9
- 3.3 Absorption probabilities are minimal solutions to RHEs . . . 11
- 3.4 Gambler's ruin . . . 12

4 Survival probability for birth and death chains, stopping times and strong Markov property **13**
- 4.1 Survival probability for birth death chains . . . 13
- 4.2 Mean hitting times are minimal solutions to RHEs . . . 14
- 4.3 Stopping times . . . 15
- 4.4 Strong Markov property . . . 15

5 Recurrence and transience **17**
5.1 Recurrence and transience 17
5.2 Equivalence of recurrence and certainty of return 17
5.3 Equivalence of transience and summability of n-step transition probabilities 18
5.4 Recurrence as a class property 18
5.5 Relation with closed classes 19

6 Random walks in dimensions one, two and three **21**
6.1 Simple random walk on $\mathbb{Z}$ 21
6.2 Simple symmetric random walk on $\mathbb{Z}^2$ 22
6.3 Simple symmetric random walk on $\mathbb{Z}^3$ 23
6.4 *A continuized analysis of random walk on $\mathbb{Z}^3$* 24
6.5 *Feasibility of wind instruments* 24

7 Invariant distributions **25**
7.1 Examples of invariant distributions 25
7.2 Notation 25
7.3 What does an invariant measure or distribution tell us? 26
7.4 Invariant distribution is the solution to LHEs 26
7.5 Stationary distributions 28
7.6 Equilibrium distributions 28

8 Existence and uniqueness of invariant distribution, mean return time, positive and null recurrence **29**
8.1 Existence and uniqueness up to constant multiples 29
8.2 Mean return time, positive and null recurrence 31
8.3 Random surfer 32

9 Convergence to equilibrium of ergodic chains **33**
9.1 Equivalence of positive recurrence and the existence of an invariant distribution 33
9.2 Aperiodic chains 34
9.3 Convergence to equilibrium *and proof by coupling* 35

10 Long-run proportion of time spent in given state **37**
10.1 Ergodic theorem 37
10.2 *Kemeny's constant and the random target lemma* 39

11 Time reversal, detailed balance, reversibility, random walk on a graph **41**
11.1 Time reversal 41
11.2 Detailed balance 42
11.3 Reversibility 43
11.4 Random walk on a graph 43

12 Concluding problems and recommendations for further study **45**
12.1 Reversibility and Ehrenfest's urn model 45
12.2 Reversibility and reflected walk 46
12.3 *The probabilistic abacus* 47
12.4 *Random walk and electrical networks* 47
12.5 Probability courses in Part II 50

Appendix **51**

A Probability spaces **51**

B Historical notes **52**

C The probabilistic abacus for absorbing chains **53**

Index **56**

Topics marked * * are non-examinable. Furthermore, only Appendix A is examinable.

Richard Weber, October 2011

Schedules

Definition and basic properties, the transition matrix. Calculation of n-step transition probabilities. Communicating classes, closed classes, absorption, irreducibility. Calculation of hitting probabilities and mean hitting times; survival probability for birth and death chains. Stopping times and statement of the strong Markov property. [5]

Recurrence and transience; equivalence of transience and summability of n-step transition probabilities; equivalence of recurrence and certainty of return. Recurrence as a class property, relation with closed classes. Simple random walks in dimensions one, two and three. [3]

Invariant distributions, statement of existence and uniqueness up to constant multiples. Mean return time, positive recurrence; equivalence of positive recurrence and the existence of an invariant distribution. Convergence to equilibrium for irreducible, positive recurrent, aperiodic chains *and proof by coupling*. Long-run proportion of time spent in given state. [3]

Time reversal, detailed balance, reversibility; random walk on a graph. [1]

Learning outcomes

A Markov process is a random process for which the future (the next step) depends only on the present state; it has no memory of how the present state was reached. A typical example is a random walk (in two dimensions, the drunkards walk). The course is concerned with Markov chains in discrete time, including periodicity and recurrence. For example, a random walk on a lattice of integers returns to the initial position with probability one in one or two dimensions, but in three or more dimensions the probability of recurrence in zero. Some Markov chains settle down to an equilibrium state and these are the next topic in the course. The material in this course will be essential if you plan to take any of the applicable courses in Part II. Learning outcomes By the end of this course, you should:

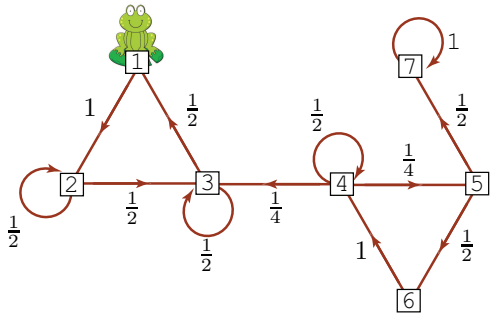
- understand the notion of a discrete-time Markov chain and be familiar with both the finite state-space case and some simple infinite state-space cases, such as random walks and birth-and-death chains;
- know how to compute for simple examples the n-step transition probabilities, hitting probabilities, expected hitting times and invariant distribution;
- understand the notions of recurrence and transience, and the stronger notion of positive recurrence;
- understand the notion of time-reversibility and the role of the detailed balance equations;
- know under what conditions a Markov chain will converge to equilibrium in long time;
- be able to calculate the long-run proportion of time spent in a given state.

1 Definitions, basic properties, the transition matrix

Markov chains were introduced in 1906 by Andrei Andreevich Markov (1856–1922) and were named in his honor.

1.1 An example and some interesting questions

Example 1.1. A frog hops about on 7 lily pads. The numbers next to arrows show the probabilities with which, at the next jump, he jumps to a neighbouring lily pad (and when out-going probabilities sum to less than 1 he stays where he is with the remaining probability).



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

There are 7 'states' (lily pads). In matrix P the element p_{57} $(= 1/2)$ is the probability that, when starting in state 5, the next jump takes the frog to state 7. We would like to know where do we go, how long does it take to get there, and what happens in the long run? Specifically:

(a) Starting in state 1, what is the probability that we are still in state 1 after 3 steps? $(p_{11}^{(3)} = 1/4)$ after 5 steps? $(p_{11}^{(5)} = 3/16)$ or after 1000 steps? $(\approx 1/5$ as $\lim_{n\to\infty} p_{11}^{(n)} \to 1/5)$

(b) Starting in state 4, what is the probability that we ever reach state 7? (1/3)

(c) Starting in state 4, how long on average does it take to reach either 3 or 7? (11/3)

(d) Starting in state 2, what is the long-run proportion of time spent in state 3? (2/5)

Markov chains models/methods are useful in answering questions such as: How long does it take to shuffle deck of cards? How likely is a queue to overflow its buffer? How long does it take for a knight making random moves on a chessboard to return to his initial square (answer 168, if starting in a corner, 42 if starting near the centre). What do the hyperlinks between web pages say about their relative popularities?

1.2 Definitions

Let I be a countable set, $\{i, j, k, \dots\}$. Each $i \in I$ is called a **state** and I is called the **state-space**.

We work in a **probability space** $(\Omega, \mathscr{F}, P)$. Here Ω is a set of outcomes, $\mathscr{F}$ is a set of subsets of Ω, and for $A \in \mathscr{F}$, $P(A)$ is the probability of A (see Appendix A).

The object of our study is a sequence of random variables $X_0, X_1, \dots$ (taking values in I) whose joint distribution is determined by simple rules. Recall that a random variable X with values in I is a function $X : \Omega \to I$.

A row vector $\lambda = (\lambda_i : i \in I)$ is called a **measure** if $\lambda_i \geq 0$ for all i.

If $\sum_i \lambda_i = 1$ then it is a **distribution** (or probability measure). We start with an **initial distribution** over I, specified by $\{\lambda_i : i \in I\}$ such that $0 \leq \lambda_i \leq 1$ for all i and $\sum_{i \in I} \lambda_i = 1$.

The special case that with probability 1 we start in state i is denoted $\lambda = \delta_i = (0, \dots, 1, \dots, 0)$.

We also have a **transition matrix** $P = (p_{ij} : i, j \in I)$ with $p_{ij} \geq 0$ for all i, j.

It is a **stochastic matrix**, meaning that $p_{ij} \geq 0$ for all $i, j \in I$ and $\sum_{j \in I} p_{ij} = 1$ (i.e. each row of P is a distribution over I).

Definition 1.2. We say that $(X_n)_{n \geq 0}$ is a **Markov chain** with initial distribution λ and transition matrix P if for all $n \geq 0$ and $i_0, \dots, i_{n+1} \in I$,

(i) $P(X_0 = i_0) = \lambda_{i_0}$;

(ii) $P(X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} \mid X_n = i_n) = p_{i_n i_{n+1}}$.

For short, we say $(X_n)_{n \geq 0}$ is ***Markov*$(\boldsymbol{\lambda}, \boldsymbol{P})$**. Checking conditions (i) and (ii) is usually the most helpful way to determine whether or not a given random process $(X_n)_{n \geq 0}$ is a Markov chain. However, it can also be helpful to have the alternative description which is provided by the following theorem.

Theorem 1.3. *$(X_n)_{n \geq 0}$ is Markov(λ, P) if and only if for all $n \geq 0$ and $i_0, \dots, i_n \in I$,*

$$P(X_0 = i_0, \dots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}. \tag{1.1}$$

Proof. Suppose $(X_n)_{n \geq 0}$ is *Markov*(λ, P). Then

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

On the other hand if (1.1) holds we can sum it over all $i_1, \dots, i_n$, to give $P(X_0 = i_0) = \lambda_0$, i.e (i). Then summing (1.1) on i_n we get $P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}) = \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}}$. Hence

$$P(X_n \mid X_0 = i_1, \dots, X_{n-1} = i_{n-1}) = \frac{P(X_0 = i_0, \dots, X_n = i_n)}{P(X_0 = i_0, \dots, X_{n-1} = i_{n-1})} = p_{i_{n-1} i_n},$$

which establishes (ii). □

1.3 Where do Markov chains come from?

At each time we apply some new 'randomness' to determine the next step, in a way that is a function only of the current state. We might take $U_1, U_2, \ldots$ as some i.i.d. random variables taking values in a set E and a function $F : I \times E \to I$.

Take $i \in I$ Set $X_0 = i$ and define recursively $X_{n+1} = F(X_n, U_{n+1})$, $n \geq 0$. Then $(X_n)_{n\geq 0}$ is *Markov*(δ_i, P) where $p_{ij} = P(F(i, U) = j)$.

1.4 How can we simulate them?

Use a computer to simulate $U_1, U_2, \ldots$ as i.i.d. $U[0, 1]$. Define $F(U, i) = J$ by the rules

$$\begin{aligned} U \in [0, p_{i1}) &\qquad\qquad &\implies j = 1 \\ U \in \left[\textstyle\sum_{k=1}^{j-1} p_{ik}, \sum_{k=1}^{j} p_{ik}\right), \quad j \geq 2 &\qquad\qquad &\implies J = j. \end{aligned}$$

1.5 The n-step transition matrix

Let A be an event. A convenient notation is $P_i(A) = P(A \mid X_0 = i)$. For example

$$P_i(X_1 = j) = p_{ij}.$$

Given the initial distribution λ, let us treat it as a row vector. Then

$$P(X_1 = j) = \sum_{i\in I} \lambda_i P_i(X_1 = j) = \sum_{i\in I} \lambda_i p_{ij}.$$

Similarly,

$$\begin{aligned} P_i(X_2 = j) &= \sum_k P_i(X_1 = k, X_2 = j) = \sum_k p_{ik}p_{kj} = (P^2)_{ij} \\ P(X_2 = j) &= \sum_{i,k} \lambda_i P_i(X_1 = k, X_2 = j) = \sum_{i,k} \lambda_i p_{ik}p_{kj} = (\lambda P^2)_j. \end{aligned}$$

Continuing in this way,

$$\begin{aligned} P_i(X_n = j) &= (\delta_i P^n)_j = (P^n)_{ij} = p_{ij}^{(n)} \\ P(X_n = j) &= \sum_{i_0,\ldots,i_{n-1}} \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} j} = (\lambda P^n)_j. \end{aligned}$$

Thus $P^{(n)} = (p_{ij}^{(n)})$, the n-step transition matrix, is simply P^n (P raised to power n).

Also, for all i, j and $n, m \geq 0$, the (obvious) **Chapman-Kolomogorov equations** hold:

$$p_{ij}^{(n+m)} = \sum_{k\in I} p_{ik}^{(n)} p_{kj}^{(m)}$$

named for their independent formulation by Chapman (a Trinity College graduate, 1880-1970), and Kolomogorov (1903-1987).

In Example 1.1

$$P^{(n)} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}^n \to \begin{pmatrix} \frac{1}{5} & \frac{2}{5} & \frac{2}{5} & 0 & 0 & 0 & 0 \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5} & 0 & 0 & 0 & 0 \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5} & 0 & 0 & 0 & 0 \\ \frac{2}{15} & \frac{4}{15} & \frac{4}{15} & 0 & 0 & 0 & \frac{1}{3} \\ \frac{1}{15} & \frac{2}{15} & \frac{2}{15} & 0 & 0 & 0 & \frac{2}{3} \\ \frac{2}{15} & \frac{4}{15} & \frac{4}{15} & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

1.6 $P^{(n)}$ for a two-state Markov chain

Example 1.4 (A two-state Markov chain)**.**



$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

The eigenvalues are 1 and $1 - \alpha - \beta$. So we can write

$$P = U \begin{pmatrix} 1 & 0 \\ 0 & 1-\alpha-\beta \end{pmatrix} U^{-1} \quad \Longrightarrow P^n = U \begin{pmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{pmatrix} U^{-1}$$

So $p_{11}^{(n)} = A + B(1-\alpha-\beta)^n$ for some A and B.

But $p_{11}^{(0)} = 1 = A+B$ and $p_{11}^{(1)} = 1-\alpha = A+B(1-\alpha-\beta)$. So $(A, B) = (\beta, \alpha)/(\alpha+\beta)$, i.e.

$$P_1(X_n = 1) = p_{11}^{(n)} = \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n.$$

Note that this tends exponentially fast to a limit of $\beta/(\alpha+\beta)$.

Other components of P^n can be compute similarly, and we have

$$P^n = \begin{pmatrix} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n & \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n \\ \frac{\beta}{\alpha+\beta} - \frac{\beta}{\alpha+\beta}(1-\alpha-\beta)^n & \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}(1-\alpha-\beta)^n \end{pmatrix}$$

Note. We might have reached the same answer by arguing:

$$p_{11}^{(n)} = p_{11}^{(n-1)} p_{11} + p_{12}^{(n-1)} p_{21} = p_{11}^{(n-1)}(1-\alpha) + \left(1 - p_{11}^{(n-1)}\right)\beta.$$

This gives the linear recurrence relation

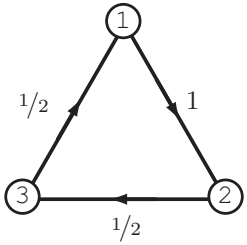
$$p_{11}^{(n)} = \beta + (1-\alpha-\beta) p_{11}^{(n-1)}$$

to which the solution is of the form $p_{11}^{(n)} = A + B(1-\alpha-\beta)^n$.

2 Calculation of n-step transition probabilities, class structure, absorption, and irreducibility

2.1 Example: a three-state Markov chain

Example 2.1.



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

Now $0 = \det(xI - P) = x(x - 1/2)^2 - 1/4 = (1/4)(x - 1)(4x^2 + 1)$.
Eigenvalues are 1, $\pm i/2$. This means that

$$p_{11}^{(n)} = A + B(i/2)^n + C(-i/2)^n.$$

We make the substitution:

$$(\pm i/2)^n = (1/2)^n e^{\pm in\pi/2} = (1/2)^n \Big(\cos(n\pi/2) \pm i \sin(n\pi/2)\Big).$$

Thus for some B' and C',

$$p_{11}^{(n)} = A + (1/2)^n \Big(B' \cos(n\pi/2) + C' \sin(n\pi/2)\Big).$$

We then use facts that $p_{11}^{(0)} = 1$, $p_{11}^{(1)} = 0$, $p_{11}^{(2)} = 0$ to fix A, B' and C' and get

$$p_{11}^{(n)} = \tfrac{1}{5} + \left(\tfrac{1}{2}\right)^n \left[\tfrac{4}{5} \cos \tfrac{n\pi}{2} - \tfrac{2}{5} \sin \tfrac{n\pi}{2}\right].$$

Note that the second term is exponentially decaying.
We have $p_{11}^{(5)} = \frac{3}{16}$, $\quad p_{11}^{(10)} = \frac{51}{256}$, $\quad |p_{11}^{(n)} - 1/5| < 2^{-n}$.

More generally, for chain with m states, and states i and j

(i) Compute the eigenvalues $\mu_1, \ldots, \mu_m$ of the $m \times m$ matrix P.

(ii) If the eigenvalues are distinct then $p_{ij}^{(n)}$ has the form (remembering that $\mu_1 = 1$),

$$p_{ij}^{(n)} = a_1 + a_2\mu_2^n + \cdots a_m\mu_m^n$$

for some constants $a_1, \ldots, a_M$. If an eigenvalue μ is repeated k times then there will be a term of $(b_0 + b_1 n + \cdots + b_{k-1}n^{k-1})\mu^n$.

(iii) Complex eigenvalues come in conjugate pairs and these can be written using sins and cosines, as in the example.

However, sometimes there is a more clever way.

2.2 Example: use of symmetry

Example 2.2 (Random walk on vertices of a complete graph)**.** Consider random walk on the complete graph K_4 (vertices of a tetrahedron), with

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

Eigenvalues are $\{1, -1/3, -1/3, -1/3\}$ so general solution is $p_{11}^{(n)} = A + (-1/3)^n(a + bn + cn^2)$. However, we may use symmetry in a helpful way. Notice that for $i \neq j$, $p_{ij}^{(n)} = (1/3)(1 - p_{ii}^{(n)})$. So

$$p_{11}^{(n)} = \sum_{j \neq 1} p_{1j} p_{j1}^{(n-1)} = (1/3)\Big(1 - p_{11}^{(n-1)}\Big).$$

Thus we have just a first order recurrence equation, to which the general solution is of the form $p_{11}^{(n)} = A + B(-1/3)^n$. Since $A + B = 1$ and $A + (-1/3)B = 0$ we have $p_{11}^{(n)} = 1/4 + (3/4)(-1/3)^n$. Obviously, $p_{11}^{(n)} \to 1/4$ as $n \to \infty$.

Do we expect the same sort of thing for random walk on corners of a cube? (No, $p_{11}^{(n)}$ does not tend to a limit since $p_{11}^{(n)} = 0$ if n is odd.)

2.3 Markov property

Theorem 2.3 (Markov property)**.** *Let $(X_n)_{n\geq 0}$ be Markov(λ, P). Then conditional on $X_m = i$, $(X_{m+n})_{n\geq 0}$ is Markov(δ_i, P) and is independent of the random variables $X_0, X_1, \ldots, X_m$.*

Proof (non-examinable). We have to show that for any event A determined by $X_0, \ldots, X_m$ we have

$$\begin{aligned} &P(\{X_m = i_m, \ldots, X_{m+n} = i_{m+n}\} \cap A \mid X_m = i) \\ &\quad = \delta_{ii_m} p_{i_m i_{m+1}} \cdots p_{i_{m+n-1} i_{m+n}} P(A \mid X_m = i) \end{aligned} \tag{2.1}$$

then the result follows from Theorem 1.3. First consider the case of elementary events like $A = \{X_0 = i_0, \ldots, X_m = i_m\}$. In that case we have to show that

$$\begin{aligned} &P(X_0 = i_0, \ldots, X_{m+n} = i_{m+n} \text{ and } i = i_m)/P(X_m = i) \\ &\quad = \delta_{ii_m} p_{i_m i_{m+1}} \cdots p_{i_{m+n-1} i_{m+n}} P(X_0 = i_0, \ldots, X_m = i_m \text{ and } i = i_m)/P(X_m = i) \end{aligned}$$

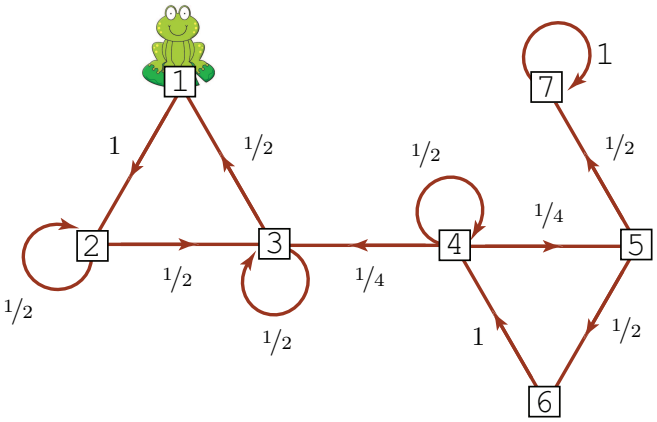
which is true by Theorem 1.3. In general, any event A determined by $X_0, \ldots, X_m$ may be written as a countable disjoint union of elementary events

$$A = \bigcup_{k=1}^{\infty} A_k$$

The desired identify (2.1) follows by summing the corresponding identities for the A_k. □

2.4 Class structure

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



We say that ***i* leads to *j*** and write $i \to j$ if

$$P_i(X_n = j \text{ for some } n \geq 0) > 0.$$

We say ***i* communicates with *j*** and write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$.

Theorem 2.4. *For distinct states i and j the following are equivalent.*

(i) $i \to j$;

(ii) $p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} > 0$ *for some states* $i_0, i_1, \ldots, i_n$ *and* $i_0 = i$ *and* $i_n = j$;

(iii) $p_{ij}^{(n)} > 0$ *for some* $n \geq 0$.

Proof. Observe that

$$p_{ij}^{(n)} \leq P_i(X_n = j \text{ for some } n \geq 0) \leq \sum_{n=0}^{\infty} p_{ij}^{(n)}$$

which proves the equivalence of (i) and (iii). Also,

$$p_{ij}^{(n)} = \sum_{i_1, \ldots, i_{n-1}} p_{i_0 i_1} p_{i i_2} \cdots p_{i_{n-1} j}$$

so that (ii) and (iii) are equivalent. □

2.5 Closed classes

It is clear from (ii) that $i \leftrightarrow j$ and $j \leftrightarrow k$ imply $i \leftrightarrow k$, and also that $i \leftrightarrow i$. So $\leftrightarrow$ satisfies the conditions for an equivalence relation on I and so partitions I into **communicating classes**. We say a C is a **closed class** if

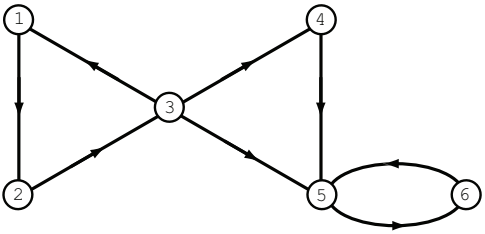
$$i \in C,\ i \to j \implies j \in C.$$

A closed class is one from which there is no escape. A state i is **absorbing** if $\{i\}$ is a closed class.

If C is not closed then it is **open** and there exist $i \in C$ and $j \notin C$ with $i \to j$ (you can escape).

Example 2.5. Find the classes in P and say whether they are open or closed.

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



The solution is obvious from the diagram. The classes are $\{1, 2, 3\}$, $\{4\}$ and $\{5, 6\}$, with only $\{5, 6\}$ being closed.

2.6 Irreducibility

A chain or transition matrix P in which I is a single class is called **irreducible**. It is easy to detect. We just have to check that $i \to j$ for every i, j.

3 Hitting probabilities and mean hitting times

3.1 Absorption probabilities and mean hitting times

Example 3.1.

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

Let us calculate probability of absorption in state 2.

$$P_1(\text{hit } 2) = \sum_{n=1}^{\infty} P(\text{hit 2 at time } n) = \sum_{n=1}^{\infty} (1-p)^{n-1} p = 1.$$

Similarly can find mean time to hit state 2:

$$\begin{aligned} E_1(\text{time to hit } 2) &= \sum_{n=1}^{\infty} nP(\text{hit 2 at time } n) \\ &= \sum_{n=1}^{\infty} n(1-p)^{n-1} p = -p\frac{d}{dp}\sum_{n=0}^{\infty}(1-p)^n = \frac{1}{p}. \end{aligned}$$

Alternatively, set $h = P_1(\text{hit } 2)$, $k = E_1(\text{time to hit } 2)$. Conditional on the first step,

$$h = (1-p)P_1(\text{hit } 2 \mid X_1 = 1) + pP_1(\text{hit } 2 \mid X_1 = 2) = (1-p)h + p \implies h = 1$$

$$k = 1 + (1-p)k + p.0 \implies k = 1/p.$$

3.2 Calculation of hitting probabilities and mean hitting times

Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix P. The first **hitting time** of a subset A of I is the random variable $H^A : \Omega \to \{0, 1, \dots\} \cup \{\infty\}$ given by

$$H^A(\omega) = \inf\{n \geq 0 : X_n(\omega) \in A\}$$

where we agree that the infimum of the empty set is ∞. The probability starting from i that $(X_n)_{n\geq 0}$ ever hits A is

$$h_i^A = P_i(H^A < \infty).$$

When A is a closed class, h_i^A is called an **absorption probability**. The **mean hitting time** for $(X_n)_{n\geq 0}$ reaching A is given by

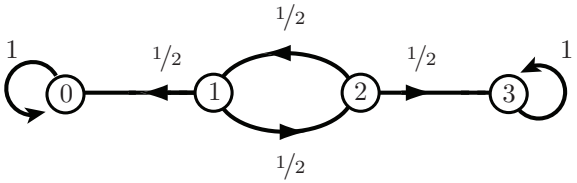
$$k_i^A = E_i(H^A) = \sum_{n<\infty} nP_i(H^A = n) + \infty P(H^A = \infty).$$

Informally,

$$h_i^A = P_i(\text{hit } A), \quad k_i^A = E_i(\text{time to hit } A).$$

Remarkably, these quantities can be calculated from certain simple linear equations. Let us consider an example.

Example 3.2. Symmetric random walk on the integers 0,1,2,3, with absorption at 0 and 3.



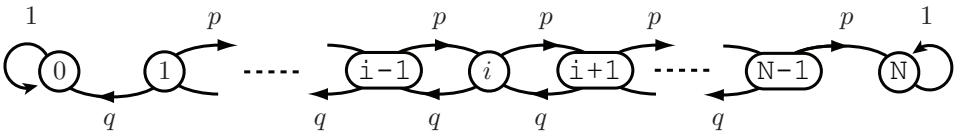
Starting from 2, what is the probability of absorption at 3? How long does it take until the chain is absorbed in 0 or 3?

Let $h_i = P_i(\text{hit } 3)$, $k_i = E_i(\text{time to hit } \{0,3\})$. Clearly,

$$\begin{aligned} h_0 &= 0 & k_0 &= 0 \\ h_1 &= \tfrac{1}{2}h_0 + \tfrac{1}{2}h_2 & k_1 &= 1 + \tfrac{1}{2}k_0 + \tfrac{1}{2}k_2 \\ h_2 &= \tfrac{1}{2}h_1 + \tfrac{1}{2}h_3 & k_2 &= 1 + \tfrac{1}{2}k_1 + \tfrac{1}{2}k_3 \\ h_3 &= 1 & k_3 &= 0 \end{aligned}$$

These are easily solved to give $h_1 = 1/3$, $h_2 = 2/3$, and $k_1 = k_2 = 2$.

Example 3.3 (Gambler's ruin on $0, \ldots, N$)**.** Asymmetric random walk on the integers $0, 1, \ldots, N$, with absorption at 0 and N. $0 < p = 1 - q < 1$.



Let $h_i = P_i(\text{hit } 0)$. So $h_0 = 1$, $h_N = 0$ and

$$h_i = qh_{i-1} + ph_{i+1}, \quad 1 \le i \le N-1.$$

Characteristic equation is $px^2 - x + q = (x-1)(px-q)$ so roots are $\{1, q/p\}$ and if $p \neq q$ general solution is $h_i = A + B(q/p)^i$. Using boundary conditions we find

$$h_i = \frac{(q/p)^i - (q/p)^N}{1 - (q/p)^N}.$$

If $p = q$ general solution is $h_i = A + Bi$ and using boundary conditions we get $h_i = 1 - i/N$.

Similarly, you should be able to show that if $p = q$, $k_i = E_i(\text{time to hit } \{0, N\}) = i(N-i)$.

3.3 Absorption probabilities are minimal solutions to RHEs

For a finite state space, as examples thus far, the equations have a unique solution. But if I is infinite there may be many solutions. The absorption probabilities are given by the minimal solution.

Theorem 3.4. *The vector of hitting probabilities $h^A = (h_i^A : i \in I)$ is the minimal non-negative solution to the system of linear equations*

$$\begin{cases} h_i^A = 1 & \text{for } i \in A \\ h_i^A = \sum_j p_{ij} h_j^A & \text{for } i \notin A \end{cases} \tag{3.1}$$

(Minimality means that if $x = (x_i : i \in I)$ is another solution with $x_i \geq 0$ then $x_i \geq h_i$ for all i.)

We call (3.1) a set of right hand equations (RHEs) because h^A occurs on the r.h.s. of P in $h^A = \tilde{P} h^A$ (where $\tilde{P}$ is the matrix obtained by deleting rows for which $i \in A$).

Proof. First we show that h^A satisfies (3.1). If $X_0 = i \in A$, then $H^A = 0$, so $h_i^A = 1$. If $X_0 \notin A$, then $H^A \geq 1$, so by the Markov property

$$h_i^A = P_i(H^A < \infty) = \sum_{j \in I} P_i(H^A < \infty, X_1 = j) = \sum_{j \in I} P_i(H^A < \infty \mid X_1 = j) P_i(X_1 = j)$$
$$= \sum_{j \in I} p_{ij} h_j^A \quad \left(\text{as } P_i(H^A < \infty \mid X_1 = j) = P_j(H^A < \infty) = h_j^A\right).$$

Suppose now that $x = (x_i : i \in I)$ is any solution to (3.1). Then $h_i^A = x_i = 1$ for $i \in A$. Suppose $i \notin A$, then

$$x_i = \sum_j p_{ij} x_j = \sum_{j \in A} p_{ij} x_j + \sum_{j \notin A} p_{ij} x_j.$$

Substitute for x_j to obtain

$$x_i = \sum_{j \in A} p_{ij} + \sum_{j \notin A} p_{ij} \left(\sum_{k \in A} p_{jk} + \sum_{k \notin A} p_{jk} x_k \right)$$
$$= P_i(X_1 \in A) + P_i(X_1 \notin A, X_2 \in A) + \sum_{j,k \notin A} p_{ij} p_{jk} x_k$$

By repeated substitution for x in the final terms we obtain

$$x_i = P_i(X_1 \in A) + P_i(X_1 \notin A, X_2 \in A) + \cdots + P_i(X_1 \notin A, X_2, \ldots, X_{n-1} \notin A, X_n \in A)$$
$$+ \sum_{j_1, \ldots, j_n \notin A} p_{ij_1} p_{j_1 j_2} \cdots p_{j_{n-1} j_n} x_{j_n}.$$

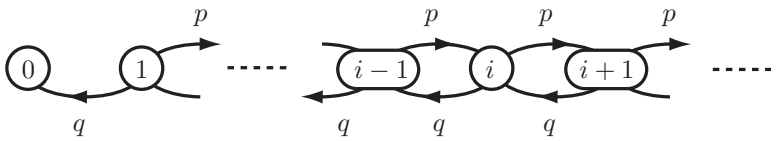
So since x_{j_n} is non-negative, $x_i \geq P_i(H^A \leq n)$. This implies

$$x_i \geq \lim_{n \to \infty} P_i(H^A \leq n) = P_i(H^A < \infty) = h_i. \qquad \square$$

Notice that if we try to use this theorem to solve Example 3.2 then (3.1) does not provide the information that $h_0 = 0$ since the second part of (3.1) only says $h_0 = h_0$. However, $h_0 = 0$ follows from the minimality condition.

3.4 Gambler's ruin

Example 3.5 (Gambler's ruin on $0, 1, \dots$)**.** $0 < p = 1 - q < 1$.



The transition probabilities are

$$\begin{aligned} &p_{00} = 1, \\ &p_{i,i-1} = q, \quad p_{i,i+1} = p \quad \text{for } i = 1, 2, \dots \end{aligned}$$

Set $h_i = P_i(\text{hit } 0)$, then it is the minimal non-negative solution to

$$\begin{aligned} &h_0 = 1, \\ &h_i = ph_{i+1} + qh_{i-1} \quad \text{for } i = 1, 2, \dots \end{aligned}$$

If $p \neq q$ this recurrence has general solution

$$h_i = A + B\,(q/p)^i\,.$$

If $p < q$ then the restriction to $0 \leq h_i \leq 1$ forces $B = 0$, so $h_i = 1$ for all i.
If $p > q$, then since $h_0 = 1$ we get a family of solutions

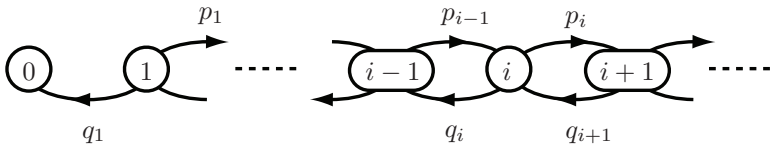
$$h_i = (q/p)^i + A\left(1 - (q/p)^i\right).$$

So the minimal non-negative solution is (with $A = 0$) $h_i = (q/p)^i$.
Finally, if $p = q = 1/2$ the recurrence relation has a general solution $h_i = A + Bi$ then again the restriction $0 \leq h_i \leq 1$ forces $B = 0$, so $h_i = 1$ for all i. Thus, even if you find a fair casino you are certain to end up broke. This apparent paradox is called **gambler's ruin**.

4 Survival probability for birth and death chains, stopping times and strong Markov property

4.1 Survival probability for birth death chains

Example 4.1. Birth and death chains. Consider the Markov chain with diagram



where, for $i = 1, 2, \ldots$, we have $0 < p_i = 1 - q_i < 1$. State i is that in which a population is i.

As in previous examples, 0 is an absorbing state and we wish to calculate the absorption probability starting from state i. So now $h_i = P_i(\text{hit } 0)$ is the extinction probability starting from state i. We write down the usual system of r.h. equations

$$\begin{aligned} h_0 &= 1, \\ h_i &= p_i h_{i+1} + q_i h_{i-1} \quad \text{for } i = 1, 2, \ldots \end{aligned}$$

This recurrence relation has variable coefficient so the usual technique fails. But consider $u_i = h_{i-1} - h_i$. Then $p_i u_{i+1} = q_i u_i$, so

$$u_{i+1} = \left(\frac{q_i}{p_i}\right) u_i = \left(\frac{q_i q_{i-1} \cdots q_1}{p_i p_{i-1} \cdots p_1}\right) u_1 = \gamma_i u_1$$

where the final equality defines γ_i. Then

$$u_1 + \cdots + u_i = h_0 - h_i$$

so

$$h_i = 1 - A(\gamma_0 + \cdots + \gamma_{i-1})$$

where $A = u_1$ and $\gamma_0 = 1$. At this point A remains to be determined. In the case $\sum_{i=0}^{\infty} \gamma_i = \infty$, the restriction $0 \le h_i \le 1$ forces $A = 0$ and $h_i = 1$ for all i. But if $\sum_{i=0}^{\infty} \gamma_i < \infty$ then we can take $A > 0$ so long as

$$1 - A(\gamma_0 + \cdots + \gamma_{i-1}) \ge 0 \quad \text{for all } i.$$

Thus the minimal non-negative solution occurs when $A = (\sum_{i=0}^{\infty} \gamma_i)^{-1}$ and then

$$h_i = \sum_{j=i}^{\infty} \gamma_j \Big/ \sum_{j=0}^{\infty} \gamma_j.$$

In this case, for $i = 1, 2, \ldots$, we have $h_i < 1$, so the population survives with positive probability.

4.2 Mean hitting times are minimal solutions to RHEs

A similar result to Theorem 3.4 can be proved for mean hitting times.

Theorem 4.2. *The vector of mean hitting times $k^A = (k_i^A : i \in I)$ is the minimal non-negative solution to the system of linear equations*

$$\begin{cases} k_i^A = 0 & \text{for } i \in A \\ k_i^A = 1 + \sum_j p_{ij} k_j^A & \text{for } i \notin A \end{cases} \tag{4.1}$$

Proof. First we show that k^A satisfies (4.1). If $X_0 = i \in A$, then $H^A = 0$, so $k_i^A = 0$. If $X_0 \notin A$, then $H^A \geq 1$, so by the Markov property

$$E_i(H^A \mid X_1 = j) = 1 + E_j(H^A) = 1 + k_j^A$$

and

$$\begin{aligned} k_i^A = E_i(H^A) &= E_i \sum_{j \in I} (H^A 1_{\{X_1 = j\}}) = \sum_{j \in I} E_i(H^A 1_{\{X_1 = j\}}) \\ &= \sum_{j \in I} E_i(H^A \mid X_1 = j) P_i(X_1 = j) = 1 + \sum_{j \notin A} p_{ij} k_j^A. \end{aligned}$$

In the first line above we use the fact that we can swap E_i and $\sum_{j \in I}$ for countable sums (Fubini's theorem).

Suppose now that $y = (y_i : i \in I)$ is any solution to (4.1). Then $k_i^A = y_i = 0$ for $i \in A$. Suppose $i \notin A$, then

$$\begin{aligned} y_i &= 1 + \sum_{j \notin A} p_{ij} y_j \\ &= 1 + \sum_{j \notin A} p_{ij} \left(1 + \sum_{k \notin A} p_{jk} y_k \right) \\ &= P_i(H^A \geq 1) + P_i(H^A \geq 2) + \sum_{j,k \notin A} p_{ij} p_{jk} y_k. \end{aligned}$$

By repeated substitution we obtain

$$y_i = P_i(H^A \geq 1) + \cdots + P_i(H^A \geq n) + \sum_{j_1, \ldots, j_n \notin A} p_{i j_1} p_{j_1 j_2} \cdots p_{j_{n-1} j_n} y_{j_n}.$$

So since y_{j_n} is non-negative

$$y_i \geq \lim_{n \to \infty} [P_i(H^A \geq 1) + \cdots + P_i(H^A \geq n)] = E_i(H^A) = k_i^A. \qquad \square$$

4.3 Stopping times

The Markov property says that for each time m, conditional on $X_m = i$, the process after time m is a Markov chain that begins afresh from i. What if we simply waited for the process to hit state i at some random time H?

A random variable $T : \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$ is called a **stopping time** if the event $\{T = n\}$ depends only on $X_0, \dots, X_n$ for $n = 0, 1, 2, \dots$. Alternatively, T is a stopping time if $\{T = n\} \in \mathscr{F}_n$ for all n, where this means $\{T = n\} \equiv \{(X_0, \dots, X_n) \in B_n\}$ for some $B_n \in I^{n+1}$.

Intuitively, this means that by watching the process you can tell when T occurs. If asked to stop at T you know when to stop.

Examples 4.3.

1. The hitting time: $H_i = \inf\{n \geq 0 : X_n = i\}$ is a stopping time.
2. $T = H_i + 1$ is a stopping time.
3. $T = H_i - 1$ is not a stopping time.
4. $T = L_i = \sup\{n \geq 0 : X_n = i\}$ is not a stopping time.

4.4 Strong Markov property

Theorem 4.4 (Strong Markov property)**.** *Let $(X_n)_{n\geq 0}$ be Markov(λ, P) and let T be a stopping time of $(X_n)_{n\geq 0}$. Then conditional on $T < \infty$ and $X_T = i$, $(X_{T+n})_{n\geq 0}$ is Markov(δ_i, P) and independent of the random variables $X_0, X_i, \dots, X_T$.*

Proof (not examinable). If B is an event determined by $X_0, X_1, \dots, X_T$ then $B \cap \{T = m\}$ is an event determined by $X_0, X_1, \dots, X_m$, so, by the Markov property at time m,

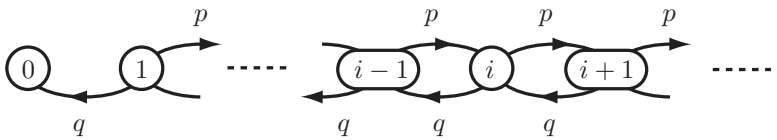
$$\begin{aligned} &P(\{X_T = j_0, X_{T+1} = j_1, \dots, X_{T+n} = j_n\} \cap B \cap \{T = m\} \cap \{X_T = i\}) \\ &\quad = P_i(\{X_0 = j_0, X_1 = j_1, \dots, X_n = j_n\}) P(B \cap \{T = m\} \cap \{X_T = i\}) \end{aligned}$$

where we have used the condition $T = m$ to replace m by T. Now sum over $m = 0, 1, 2, \dots$ and divide by $P(T < \infty, X_T = i)$ to obtain

$$\begin{aligned} &P(\{X_T = j_0, X_{T+1} = j_1, \dots, X_{T+n} = j_n\} \cap B \mid T < \infty, X_T = i) \\ &\quad P_i(\{X_0 = j_0, X_1 = j_1, \dots, X_n = j_n\}) P(B \mid T < \infty, X_T = i). \end{aligned}$$

□

Example 4.5 (Gambler's ruin again)**.** $0 < p = 1 - q < 1$.



Let $h_i = P_i(\text{hit } 0)$. Then

$$h_1 = ph_2 + q$$
$$h_2 = h_1^2$$

where the second expression is obtained by using strong Markov property to see that $P_2(\text{hit } 0) = P_2(\text{hit } 1)P_1(\text{hit } 0) = P_1(\text{hit } 0)^2$. So $h_1 = ph_1^2 + q \implies h_1 = 1 \text{ or } = q/p$. We take the minimal solution.

Similarly, let $k_i = E_i(\text{time to hit } 0)$. Assuming $q > p$,

$$k_1 = 1 + pk_2$$
$$k_2 = 2k_1 \quad \text{(by strong Markov property)}$$

$\implies k_1 = 1/(1-2p) = 1/(q-p)$.

5 Recurrence and transience

5.1 Recurrence and transience

Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix P. Define

$$H_i = \inf\{n \geq 0 : X_n = i\} = \textbf{hitting time} \text{ on } i$$
$$T_i = \inf\{n \geq 1 : X_n = i\} = \textbf{first passage time} \text{ to } i.$$

Note that H_i and T_i differ only if $X_0 = i$. Let

$$V_i = \sum_{n=0}^{\infty} 1_{\{X_n = i\}} = \text{number of visits to } i$$
$$f_i = P_i(T_i < \infty) = \textbf{return probability} \text{ to } i$$
$$m_i = E_i(T_i) = \textbf{mean return time} \text{ to } i.$$

We say that i is **recurrent** if $P_i(V_i = \infty) = 1$. Otherwise i is **transient**. A recurrent state is one that one that you keep coming back to. A transient state is one that you eventually leave forever. In fact, as we shortly see, $P_i(V_i = \infty)$ can only take the values 0 and 1 (not, say, 1/2).

5.2 Equivalence of recurrence and certainty of return

Lemma 5.1. *For all $k \geq 0$, $P_i(V_i \geq k+1) = (f_i)^k$.*

Proof. This is true for $k = 0$. Assume it is true for $k-1$. The kth visit to i is a stopping time, so by the strong Markov property

$$\begin{aligned} P_i(V_i \geq k+1) &= P_i(V_i \geq k+1 \mid V_i \geq k) P_i(V_i \geq k) \\ &= P_i(T_i < \infty)(f_i)^{k-1} \\ &= (f_i)^k. \end{aligned}$$

Hence the lemma holds by induction. □

Theorem 5.2. *We have the dichotomy*

$$i \text{ is recurrent} \iff f_i = 1$$
$$i \text{ is transient} \iff f_i < 1.$$

Proof. Observe that

$$P_i(V_i < \infty) = P_i\Big(\bigcup_{k\geq 1}\{V_i = k\}\Big) = \sum_{k=1}^{\infty} P_i(V_i = k) = \sum_{k=1}^{\infty}(1-f_i)f_i^{k-1} = \begin{cases} 0, & f_i = 1, \\ 1, & f_i < 1. \end{cases}$$

So i is recurrent iff $f_i = 1$, and transient iff $f_i < 1$. □

5.3 Equivalence of transience and summability of n-step transition probabilities

Theorem 5.3. *We have the dichotomy*

$$i \text{ is recurrent} \iff \sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty \iff f_i = 1$$

$$i \text{ is transient} \iff \sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty \iff f_i < 1.$$

Proof. If i is recurrent, this means $P_i(V_i = \infty) = 1$, and so

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \sum_{n=0}^{\infty} E_i\left(1_{\{X_n=i\}}\right) = E_i\left(\sum_{n=0}^{\infty} 1_{\{X_n=i\}}\right) = E_i(V_i) = \infty.$$

If i is transient then (by Theorem 5.2) $f_i < 1$ and

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = E_i(V_i) = \sum_{r=0}^{\infty} P_i(V_i > r) = \sum_{r=0}^{\infty} f_i^r = \frac{1}{1-f_i} < \infty.$$

□

5.4 Recurrence as a class property

Now we can show that recurrence and transience are class properties.

Theorem 5.4. *Let C be a communicating class. Then either all states in C are transient or all are recurrent.*

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

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

so

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

and so j is transient by Theorem 5.3 □

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

The following blue text gives an alternative way to reach the conclusion that recurrence and transience are class properties. In lectures we will probably omit this. But it is nice to see the contrast with matrix-method used in Theorem 5.4.

Theorem 5.5. *Suppose i is recurrent and $i \to j$. Then (a) $P_j(H_i < \infty) = 1$, (b) $P_i(H_j < \infty) = 1$, (c) j is recurrent.*

Note that (c) implies that recurrence and transience are class properties (see also Theorem 5.4). Also, (a) implies that every recurrent class is closed (see also Theorem 5.6).

Proof. (a) By the strong Markov property

$$0 = P_i(H_i = \infty) \geq P_i(H_j < \infty) P_j(H_i = \infty).$$

But $P_i(H_j < \infty) > 0$ so $P_j(H_i = \infty) = 0 \implies P_j(H_i < \infty) = 1$, which is (a).

(b) Let $T_i^{(0)} = 0$ and $T_i^{(k)}$ be the time of kth return to i. Consider the events $A_k = \left[X_n = j \text{ for some } T_i^{(k-1)} < n < T_i^{(k)}\right]$. Let $\rho = P_i(A_k)$. The A_k are independent, so

$$P_i(H_j < \infty) = P_i\big(\bigcup_k A_k\big) = 1 - P_i\big(\bigcap_k A_k^c\big) = \begin{cases} 0, & \rho = 0 \\ 1, & \rho > 0 \end{cases}$$

Now $i \to j$ implies $\rho > 0$, which forces $P_i(H_j < \infty) = 1$, which is (b).

Finally, by (a), (b) and the strong Markov property,

$$P_j(T_j < \infty) \geq P_j(H_i < \infty) P_i(H_j < \infty) = 1.$$

□

5.5 Relation with closed classes

Theorem 5.6. *Every recurrent class is closed.*

Proof. Let C be a class which is not closed. Then there exists $i \in C$ and $j \notin C$ and $m \geq 1$ with

$$P_i(X_m = j) > 0.$$

Since we have

$$P_i(V_i = \infty \mid X_m = j) = 0$$

this implies that

$$P_i(V_i = \infty) = \sum_k P_i(V_i = \infty \mid X_m = k\}) P_i(X_m = k) < 1$$

so i is not recurrent, and so neither is C. □

Theorem 5.7. *Every finite closed class is recurrent.*

Proof. Let C be such a class. Pick any initial distribution on C. Then $\sum_{i\in C} V_i = \infty$. Since C is finite some state must be visited infinitely often. So

$$1 = P\Big(\bigcup_{i\in C}\{V_i = \infty\}\Big) \le \sum_{i\in C} P(V_i = \infty)$$

So for some i, $0 < P(V_i = \infty) = P(H_i < \infty)P_i(V_i = \infty)$. But since $P_i(V_i = \infty)$ can only take values 0 or 1, we must have $P_i(V_i = \infty) = 1$. Thus i is recurrent, and so also C is recurrent. □

We will need the following result for Theorem 9.8.

Theorem 5.8. *Suppose P is irreducible and recurrent. Then for all $j \in I$ we have $P(T_j < \infty) = 1$.*

Proof. By the Markov property we have

$$P(T_j < \infty) = \sum_{i\in I} P(X_0 = i)P_i(T_j < \infty)$$

so it suffices to show that $P_i(T_j < \infty) = 1$ for all $i \in I$. Choose m with $p_{ji}^{(m)} > 0$. By Theorem 5.3 we have

$$\begin{aligned}
1 &= P_j(X_n = j \text{ for infinitely many } n)\\
&\le P_j(X_n = j \text{ for some } n \ge m+1)\\
&= \sum_{k\in I} P_j(X_n = j \text{ for some } n \ge m+1 \mid X_m = k)P_j(X_m = k)\\
&= \sum_{k\in I} P_k(T_j < \infty)p_{jk}^{(m)}
\end{aligned}$$

But $\sum_{k\in I} p_{jk}^{(m)} = 1$ so we must have $P_i(T_j < \infty) = 1$. □

We conclude with a criterion by which to tell whether an irreducible chain is transient.

Theorem 5.9. *An irreducible Markov chain is transient iff for some state i there exists a nonzero vector y such that $y_j = \sum_{k\neq i} p_{jk}y_k$, for all $j \neq i$, and $|y_j| \le 1$ for all j.*

**Proof*.* If the chain is transient then we may take $y_j = P_j(T_i < \infty)$, with $0 < y_j < 1$. On the other hand, if there exists a y as described in the theorem statement, then $|y_j| \le \sum_{k\neq i} p_{jk}|y_k|$. By repeatedly substitution of this into itself (in the manner of the proofs of Theorems 3.4 and 4.2) we can show

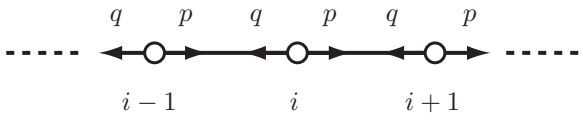
$$0 < |y_j| \le \sum_{k_1 k_2 \ldots k_m \neq i} p_{jk_1}p_{k_1k_2}\ldots p_{k_{m-1}k_m} = P_j(T_i > m).$$

Taking a limit as $m \to \infty$ we have $P_j(T_i = \infty) > 0$ and so the chain is transient. □

6 Random walks in dimensions one, two and three

6.1 Simple random walk on $\mathbb{Z}$

Example 6.1 (Simple random walk on $\mathbb{Z}$)**.** The simple random walk on $\mathbb{Z}$ has diagram



where $0 < p = 1 - q < 1$.

Consider $f_0 = P_0(\text{return to } 0)$. Let $h_i = P_i(\text{hit } 0)$. Then

$$f_0 = qh_{-1} + ph_1.$$

Using the strong Markov property $h_1 = q + ph_1^2$. The smallest solution is $h_1 = \min\{1, q/p\}$. So if $q = p = 1/2$ then $h_{-1} = h_1 = 1 \implies f_0 = 1$ and the random walk is recurrent. But if $q < p$ then $h_1 < 1$, which implies $f_0 < 1$ and transience. Similarly, if $q > p$ then $h_{-1} < 1$ and we have transience.

We may also analyse this random walk using Theorem 5.3 (testing to see whether $\sum_n p_{00}^{(n)}$ is finite or infinite. This is more complicated method, but we now do it as a warm-up for our subsequent analysis of random walk on $\mathbb{Z}^2$.

Suppose we start at 0. Obviously, we cannot return to 0 after an odd number of steps, so $p_{00}^{(2n+1)} = 0$ for all n. Any given sequence of steps of length $2n$ from 0 to 0 occurs with probability $p^n q^n$, there being n steps right and n steps left, and the number of such sequences is the number of ways of choosing the n right steps from $2n$. Thus

$$p_{00}^{(2n)} = \binom{2n}{n} p^n q^n.$$

Stirling's approximation provides an approximation to $n!$ for large n: it is known that

$$n! \sim A\sqrt{n}(n/e)^n \quad \text{as } n \to \infty$$

for some $A \in [1, \infty)$ (we do not need here that $A = \sqrt{2\pi}$). Thus

$$p_{00}^{(2n)} = \frac{(2n)!}{(n!)^2}(pq)^n \sim \frac{(4pq)^n}{A\sqrt{n/2}} \quad \text{as } n \to \infty.$$

In the symmetric case $p = q = 1/2$ we have that $4pq = 1$ and so for some N and all $n \geq N$,

$$p_{00}^{(2n)} \geq \frac{1}{2A\sqrt{n}}$$

so

$$\sum_{n=N}^{\infty} p_{00}^{(2n)} \geq \frac{1}{2A} \sum_{n=N}^{\infty} \frac{1}{\sqrt{n}} = \infty$$

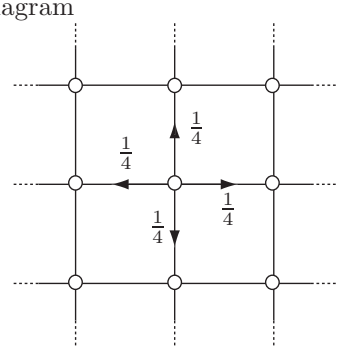
which show that the random walk is recurrent. On the other hand, if $p \neq q$ then $4pq = r < 1$, so by a similar argument, for some N,

$$\sum_{n=N}^{\infty} p_{00}^{(2n)} \leq \frac{1}{A} \sum_{n=N}^{\infty} r^n < \infty$$

and so the random walk is transient.

6.2 Simple symmetric random walk on $\mathbb{Z}^2$

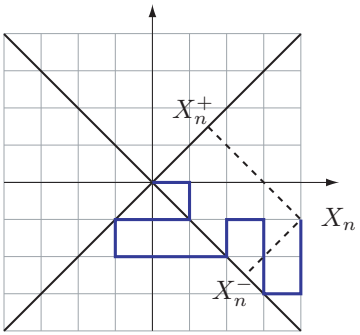
Example 6.2 (Simple symmetric random walk on $\mathbb{Z}^2$)**.** The simple random walk on $\mathbb{Z}^2$ has diagram



and transition probabilities

$$p_{ij} = \begin{cases} 1/4 & \text{if } |i-j| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Suppose we start at 0. Let us call the walk X_n and write X_n^+ and X_n^- for the projection of the walk on the diagonal lines $y = \pm x$.



Then X_n^+ and X_n^- are independent symmetric random walks on $2^{-1/2}\mathbb{Z}$, and $X_n = 0$ if and only if $X_n^+ = 0 = X_n^-$. This makes clear that for X_n we have

$$p_{00}^{(n)} = \left(\binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \right)^2 \sim \frac{2}{A^2 n} \quad \text{as } n \to \infty$$

by Stirling's formula. Then $\sum_{n=0}^{\infty} p_{00}^{(n)} = \infty$ by comparison with $\sum_{n=0}^{\infty} 1/n$ and the walk is recurrent.

6.3 Simple symmetric random walk on $\mathbb{Z}^3$

Example 6.3 (Simple symmetric random walk on $\mathbb{Z}^3$)**.** The transition probabilities are given by

$$p_{ij} = \begin{cases} 1/6 & \text{if } |i-j| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Thus the chain jumps to each of its nearest neighbours with equal probability. There is no easy way to map this into three independent walks on $\mathbb{Z}$ as we have done above for the random walk on $\mathbb{Z}^2$ (although see §6.4 below).

Suppose we start at 0. We can only return to zero after an even number of steps, say $2n$. Of these $2n$ steps there must be i up, i down, j north, j south, k east, k west for some $i, j, k \geq 0$, with $i + j + k = n$. By counting ways this can be done we get

$$p_{00}^{(2n)} = \sum_{\substack{i,j,k \geq 0 \\ i+j+k=n}} \frac{(2n)!}{(i!j!k!)^2} \left(\frac{1}{6}\right)^{2n} = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \sum_{\substack{i,j,k \geq 0 \\ i+j+k=n}} \binom{n}{i\ j\ k}^2 \left(\frac{1}{3}\right)^{2n}.$$

Now

$$\sum_{\substack{i,j,k \geq 0 \\ i+j+k=n}} \binom{n}{i\ j\ k} \left(\frac{1}{3}\right)^n = 1$$

the left hand side being the total probability of all the ways of placing n balls in 3 boxes. For the case where $n = 3m$ we have

$$\binom{n}{i\ j\ k} = \frac{n!}{i!j!k!} \leq \frac{n!}{m!m!m!} = \binom{n}{m\ m\ m}$$

for all i, j, k, so

$$p_{00}^{(2n)} \leq \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \binom{n}{m\ m\ m} \left(\frac{1}{3}\right)^n \sim \frac{1}{2A^3} \left(\frac{6}{n}\right)^{3/2} \quad \text{as } n \to \infty.$$

Hence $\sum_{m=0}^{\infty} p_{00}^{(6m)} < \infty$ by comparison with $\sum_{n=0}^{\infty} n^{-3/2}$. But $p_{00}^{(6m)} \geq (1/6)^2 p_{00}^{(6m-2)}$ and $p_{00}^{(6m)} \geq (1/6)^4 p_{00}^{(6m-4)}$ so for all m we must have

$$\sum_{n=0}^{\infty} p_{00}^{(n)} < \infty$$

and the walk is transient. In fact, it can be shown that the probability of return to the origin is about $0.340537329544\ldots$.

The above results were first proved in 1921 by the Hungarian mathematician George Pólya (1887–1985).

6.4 *A continuized analysis of random walk on $\mathbb{Z}^3$*

There is a nice way to analyse the random walk on $\mathbb{Z}^3$ that avoids the need for the intricacies above. Consider six independent processes $X^+(t)$, $X^-(t)$, $Y^+(t)$, $Y^-(t)$, $Z^+(t)$, $Z^-(t)$. Each is nondecreasing in continuous time on the states $0, 1, 2, \ldots$, making a transition $+1$ at times that are separated as i.i.d. exponential(1/2) random variables. Consequently, successive jumps (in one of the 6 directions) are separated as i.i.d. $\sim$exponential(3) (i.e. as events in a Poisson process of rate 3). Also, $P(X^+(t) = i) = (t/2)^i e^{-t/2}/i!$, and similarly for the others. At time t,

$$P(X^+(t) = X^-(t)) = \sum_{i=0}^{\infty} \Big((t/2)^i e^{-t/2}/i! \Big)^2 = I_0(t) e^{-t} \sim (2\pi t)^{-1/2},$$

where I_0 is a modified Bessel function of the first kind of order 0. Thus, for this continuous-time process $p_{00}(t) \sim (2\pi t)^{-3/2}$. Notice that this process observed at its jump times is equivalent to discrete-time random walk on $\mathbb{Z}^3$. Since the mean time between the jumps of the continuous process is 1/3, we find $\sum_{n=0}^{\infty} p_{00}^{(n)} = 3 \int_0^{\infty} p_{00}(t) dt < \infty$.

6.5 *Feasibility of wind instruments*

Example 6.4. Lord Rayleigh in "On the theory of resonance" (1899) proposed a model for wind instruments in which the creation of resonance through a vibrating column of air requires repeated expansion and contraction of a mass of air at the mouth of the instrument, air being modelled as an incompressible fluid.

Think instead about an infinite rectangular lattice of cities. City (0,0) wishes to expand its tax base and does this by inducing a business from a neighboring city to relocate to it. The impoverished city does the same (choosing to "beggar-its-neighbour" randomly amongst its 4 neighbours since "beggars can't be choosers"), and this continues, just like a 2-D random walk. Unfortunately, this means that with probability 1 the walk returns to the origin city who eventually finds that one of its own businesses is induced away by one of its neighbours, leaving it no better off than at the start. We might say that it is "infinitely-hard to expand the tax base by a beggar-your-neighbour policy". However, in 3-D there is a positive probability (about 0.66) that the city $(0, 0)$ will never be beggared by one of its 6 neighbours.

By analogy, we see that in 2-D it will be "infinitely hard" to expand the air at the mouth of the wind instrument, but in 3-D the energy required is finite. That is why Doyle and Snell say wind instruments are possible in our 3-dimensional world, but not in Flatland.

We will learn in Lecture 12 something more about the method that Rayleigh used to show that the energy required to create a vibrating column of air in 3-D is finite.

7 Invariant distributions

7.1 Examples of invariant distributions

Many of the long-time properties of Markov chains are connected with the notion of an invariant distribution or measure. Remember that a measure λ is any row vector $(\lambda_i : i \in I))$ with non-negative entries. A measure λ is an **invariant measure** if $\lambda P = \lambda$. A invariant measure λ is an **invariant distribution** if $\sum_i \lambda_i = 1$. The terms **equilibrium**, **stationary**, and **steady state** are also used to mean the same thing.

Example 7.1. Here are some invariant distributions for some 2-state Markov chains.

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \lambda = (\tfrac{1}{2}, \tfrac{1}{2}), \quad (\tfrac{1}{2}, \tfrac{1}{2}) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = (\tfrac{1}{2}, \tfrac{1}{2})$$

$$P = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}, \quad \lambda = (\tfrac{1}{2}, \tfrac{1}{2}), \quad (\tfrac{1}{2}, \tfrac{1}{2}) \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} = (\tfrac{1}{2}, \tfrac{1}{2})$$

$$P = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \lambda = (\tfrac{2}{3}, \tfrac{1}{3}), \quad (\tfrac{2}{3}, \tfrac{1}{3}) \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = (\tfrac{2}{3}, \tfrac{1}{3})$$

(A) Does an invariant measure always exist? Can there be more than one?

(B) Does an invariant distribution always exist? Can there be more than one?

(C) What does an invariant measure or distribution tell me about the chain?

(D) How do I calculate π? (Left hand equations, detailed balance, symmetry).

7.2 Notation

We repeat and add to the notation definitions in §5.1.

$$T_i = \inf\{n \geq 1 : X_n = i\} = \textbf{first passage time} \text{ to } i$$
$$m_i = E_i(T_i) = \textbf{mean return time} \text{ to } i$$

$$V_i(n) = \sum_{k=0}^{n-1} 1_{\{X_k = i\}} = \text{number of visits to } j \text{ before time } n$$

$$V_i^k = V_i(T_k) = \text{number of visits to } i \text{ before first return to } k$$

$$\gamma_i^k = E_k(V_i^k) = \text{mean number of visits to } i \text{ between successive visits to } k$$

Notice that $V_k^k = 1$ and $\gamma_k^k = 1$.

7.3 What does an invariant measure or distribution tell us?

Suppose π is an invariant distribution and λ is an invariant measure. Under suitable conditions,

$$m_i = \frac{1}{\pi_i}, \quad \gamma_j^i = \frac{\pi_j}{\pi_i} \ \left(= \frac{\lambda_j}{\lambda_i} \right)$$

and

$$P\left(\frac{V_j(n)}{n} \to \pi_j \text{ as } n \to \infty \right) = 1$$

or, as we say, $V_j(n)/n \to \pi_j$ **almost surely**.

Example 7.2. Consider the two state chain with

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

Then

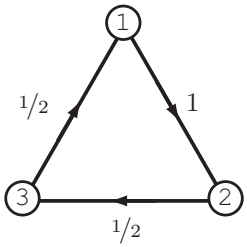
$$P^n \to \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{pmatrix}$$

$\pi = (\pi_1, \pi_2) = (\beta/(\alpha+\beta), \alpha/(\alpha+\beta)$ is an invariant distribution.

7.4 Invariant distribution is the solution to LHEs

How do we calculate the π_i?

Example 7.3. Consider again Example 2.1.



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

To find an invariant distribution we write down the components of the vector equation $\pi = \pi P$. (We call these **left hand equations** (LHEs) as π appears on the left of P.)

$$\begin{aligned} \pi_1 &= \pi_3 \tfrac{1}{2} \\ \pi_2 &= \pi_1 \tfrac{1}{2} + \pi_3 \tfrac{1}{2} \\ \pi_3 &= \pi_1 \tfrac{1}{2} + \pi_3 \tfrac{1}{2} \end{aligned}$$

The right-hand sides give the probabilities for X_1, when X_0 has distribution π, and the equations require X_1 also to have distribution π,. The equations are homogeneous

so one of them is redundant, and another equation is required to fix π uniquely. That equation is

$$\pi_1 + \pi_2 + \pi_3 = 1,$$

and so we find that $\pi = (1/5, 2/5, 1/5)$. Recall that in Example 2.1

$$p_{11}^{(n)} \to 1/5 \text{ as } n \to \infty$$

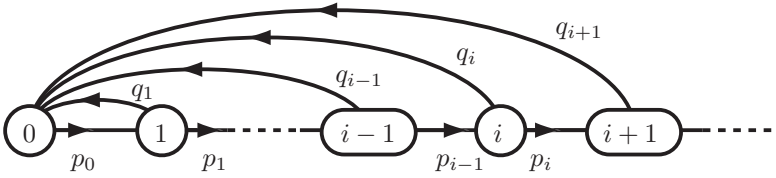
so this confirms Theorem 7.6, below. Alternatively, knowing that $p_{11}^{(n)}$ has the form

$$p_{11}^{(n)} = a + \left(\frac{1}{2}\right)^n \left(b \cos \frac{n\pi}{2} + c \sin \frac{n\pi}{2}\right)$$

we can use Theorem 7.6 and knowledge of π_1 to identify $a = 1/5$, instead of working out $p_{11}^{(2)}$ in Example 2.1.

Alternatively, we may sometimes find π from the **detailed balance equations** $\pi_i p_{ij} = \pi_j p_{ji}$ for all i, j. If these equations have a solution then it is an invariant distribution for P (since by summing on j we get $\pi_j = \sum_j \pi_j p_{ji}$). For example, in Example 7.2 the detailed balance equation is $\pi_1 \alpha = \pi_2 \beta$. Together with $\pi_1 + \pi_2 = 1$ this gives $\pi = (\beta, \alpha)/(\alpha + \beta)$. We will say more about this in Lecture 11.

Example 7.4. Consider a **success-run chain** on $\{0, 1, \ldots\}$, whose transition probabilities are given by $p_{i,i+1} = p_i$, $p_{i0} = q_i = 1 - p_i$.



Then the left-hand (invariant measure equations) read

$$\pi_0 = \sum_{i=0}^{\infty} \pi_i q_i \quad \Longrightarrow \; p_0 \pi_0 = \sum_{i=1}^{\infty} \pi_i q_i$$
$$\pi_i = \pi_{i-1} p_{i-1}, \quad \text{for } i \geq 1.$$

Let $r_0 = 1$, $r_i = p_0 p_1 \cdots p_{i-1}$. So $\pi_i = r_i \pi_0$.

We choose p_i converging sufficiently rapidly to 1 so that

$$r := \prod_{i=0}^{\infty} p_i > 0 \quad \text{(which is equivalent to } \sum_{i=0}^{\infty} q_i < \infty.)$$

So $q_i\pi_i = (1-p_i)r_i\pi_0 = (r_i - r_{i+1})\pi_0$. Also

$$\begin{aligned}\pi_0 &= \sum_{i=0}^{\infty}\pi_i q_i = \lim_{n\to\infty}\sum_{i=0}^{n}\pi_i q_i = \lim_{n\to\infty}\sum_{i=0}^{n}(r_i - r_{i+1})\pi_0 \\ &= \lim_{n\to\infty}(r_0 - r_{n+1})\pi_0 = (1-r)\pi_0\end{aligned}$$

so $\pi_0 = 0$ and there is no invariant measure.

7.5 Stationary distributions

The following result explains why invariant distributions are also called **stationary** distributions.

Theorem 7.5. *Let $(X_n)_{n\geq 0}$ be Markov(λ, P) and suppose that λ is invariant for P. Then $(X_{m+n})_{n\geq 0}$ is also Markov(λ, P).*

Proof. By the Markov property we have $P(X_m = i) = (\lambda P^m)_i = \lambda_i$ for all i, and clearly, conditional on $X_{m+n} = i$, X_{m+n+1} is independent of $X_m, X_{m+1}, \ldots, X_{m+n}$ and has distribution $(p_{ij} : j \in I)$ □

7.6 Equilibrium distributions

The next result explains why invariant distributions are also called **equilibrium** distributions.

Theorem 7.6. *Let I be finite. Suppose for some $i \in I$ that*

$$p_{ij}^{(n)} \to \pi_j \quad \text{for all } j \in I.$$

Then $\pi = (\pi_j : j \in I)$ is an invariant distribution.

Proof. We have

$$\sum_{j\in I}\pi_j = \sum_{j\in I}\lim_{n\to\infty}p_{ij}^{(n)} = \lim_{n\to\infty}\sum_{j\in I}p_{ij}^{(n)} = 1$$

and

$$\pi_j = \lim_{n\to\infty}p_{ij}^{(n)} = \lim_{n\to\infty}\sum_{k\in I}p_{ik}^{(n)}p_{kj} = \sum_{k\in I}\lim_{n\to\infty}p_{ik}^{(n)}p_{kj} = \sum_{k\in I}\pi_k p_{kj}$$

where we have used finiteness of I to justify interchange of summation and limit operations. Hence π is an invariant distribution. □

Notice that for any of the symmetric random walks discussed previously we have $p_{ij}^{(n)} \to 0$ as $n \to \infty$ for all $i, j \in I$. The limit is certainly invariant, but it is not a distribution. Theorem 7.6 is not a very useful result but it serves to indicate a relationship between invariant distributions n-step transition probabilities. In Theorem 9.8 we shall prove a sort of converse, which is much more useful.

8 Existence and uniqueness of invariant distribution, mean return time, positive and null recurrence

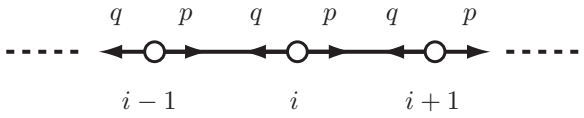
8.1 Existence and uniqueness up to constant multiples

Example 8.1 (Non-uniqueness of invariant distribution)**.**

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

All distributions are invariant. (Example Sheet 2 has a less trivial example.)

Example 8.2 (Non-uniqueness with irreducibility)**.** Suppose $p \neq q$.



$$\pi_i = \pi_{i-1}p + \pi_{i+1}q$$

Set $\lambda_i = 1$, $\mu_i = (p/q)^i$, $i \in \mathbb{Z}$. Then both λ and μ are invariant measures. However, this Markov chain is transient.

The next theorem shows that an invariant measure such that $\lambda_i = 1$ for some i must have $0 < \lambda_j < \infty$ for all j.

Theorem 8.3. *Suppose P is irreducible and $\lambda \geq 0$ and $\lambda = \lambda P$. Then $\lambda \equiv 0$ or $(0 < \lambda_i < \infty$ for all $i)$ or $\lambda \equiv \infty$.*

Proof. $\lambda = \lambda P = \lambda P^n$. Given i, j there exists n with $p_{ij}^{(n)} > 0$ (since irreducible) so

$$\lambda_j = \sum_k \lambda_k p_{kj}^{(n)} \geq \lambda_i p_{ij}^{(n)}$$

So $\lambda_j < \infty \implies \lambda_i < \infty$ and $\lambda_i > 0 \implies \lambda_j > 0$. □

In the next two results we shall show that every irreducible and recurrent stochastic matrix P has an essentially unique positive invariant measure. The proofs rely heavily on the probabilistic interpretation so it is worth noting at the outset that, for a finite state-space I, the existence of an invariant row vector is a simple piece of linear algebra: the row sums of P are all 1, so the column vector of ones is an eigenvector with eigenvalue 1, so P must also have a row eigenvector with eigenvalue 1.

Recall that for a fixed state k, we define for each i the expected time spent in i between visits to k:

$$\gamma_i^k = E_k \sum_{n=0}^{T_k - 1} 1_{\{X_n = i\}}.$$

Here the sum of indicator functions counts the number of times at which $X_n = i$ before the first passage to k at time T_k.

Theorem 8.4 (Existence of an invariant measure)**.** *Let P be irreducible and recurrent. Then*

(i) $\gamma_k^k = 1$.

(ii) $\gamma^k = (\gamma_i^k : i \in I)$ *satisfies* $\gamma^k P = \gamma^k$.

(iii) $0 < \gamma_i^k < 1$ *for all* i.

Proof. (i) is obvious. (ii) For $n = 1, 2, \ldots$ the event $\{n \leq T_k\}$ depends only on $X_0, \ldots, X_{n-1}$ so, by the Markov property at $n-1$,

$$P_k(X_{n-1} = i, X_n = j \text{ and } n \leq T_k) = P_k(X_{n-1} = i \text{ and } n \leq T_k)p_{ij}$$

Since P is recurrent, $P_k(T_k < \infty) = 1$ and $X_0 = X_{T_k} = k$ with probability 1. So

$$\begin{aligned}
\gamma_j^k &= E_k \sum_{n=1}^{T_k} 1_{\{X_n = j\}} = E_k \sum_{n=1}^{\infty} 1_{\{X_n = j \text{ and } n \leq T_k\}} \\
&= \sum_{i \in I} \sum_{n=1}^{\infty} P_k(X_{n-1} = i, X_n = j \text{ and } n \leq T_k) \\
&= \sum_{i \in I} p_{ij} \sum_{n=1}^{\infty} P_k(X_{n-1} = i \text{ and } n \leq T_k) \\
&= \sum_{i \in I} p_{ij} E_k \sum_{n=1}^{\infty} 1_{\{X_{n-1} = i \text{ and } n \leq T_k\}} \\
&= \sum_{i \in I} p_{ij} E_k \sum_{m=0}^{\infty} 1_{\{X_m = i \text{ and } m \leq T_k - 1\}} \\
&= \sum_{i \in I} p_{ij} E_k \sum_{m=0}^{T_k - 1} 1_{\{X_m = i\}} = \sum_{i \in I} \gamma_i^k p_{ij}
\end{aligned}$$

(iii) Since P is irreducible, for each i there exists $n, m \geq 0$ with $p_{ik}^{(n)}, p_{ki}^{(m)} > 0$. Then $\gamma_i^k \geq \gamma_k^k p_{ki}^{(m)} > 0$ and $\gamma_i^k p_{ik}^{(n)} \leq \gamma_k^k = 1$ by (i) and (ii). □

Theorem 8.5 (Uniqueness of an invariant measure)**.** *Let P be irreducible and let λ be an invariant measure for P with $\lambda_k = 1$. Then $\lambda \geq \gamma^k$. If in addition P is recurrent, then $\lambda = \gamma^k$.*

Proof. For each $j \in I$ we have

$$\begin{aligned}
\lambda_j &= \sum_{i_0 \in I} \lambda_{i_0} p_{i_0 j} = p_{kj} + \sum_{i_0 \neq k} \lambda_{i_0} p_{i_0 j} \\
&= p_{kj} + \sum_{i_0 \neq k} p_{k i_0} p_{i_0 j} + \sum_{i_0, i_1 \neq k} \lambda_{i_1} p_{i_1 i_0} p_{i_0 j} \\
&\vdots \\
&= p_{kj} + \sum_{i_0 \neq k} p_{k i_0} p_{i_0, j} + \sum_{i_0, i_1 \neq k} p_{k i_1} p_{i_1 i_0} p_{i_0, j} + \sum_{i_0, i_1 \neq k} \lambda_{i_0} p_{i_0 i_1} p_{i_1, j} \\
&\quad + \cdots + \sum_{i_0, \ldots, i_{n-1} \neq k} p_{k i_{n-1}} \cdots p_{i_1 i_0} p_{i_0 j} + \sum_{i_0, \ldots, i_n \neq k} \lambda_{i_n} p_{i_n i_{n-1}} \cdots p_{i_1 i_0} p_{i_0 j} \\
&\geq P_k(X_1 = j \text{ and } T_k \geq 1) + P_k(X_2 = j \text{ and } T_k \geq 2) + \cdots + P_k(X_n = j \text{ and } T_k \geq n) \\
&\to \gamma_j^k \text{ as } n \to \infty
\end{aligned}$$

So $\lambda \geq \gamma^k$. If P is recurrent, then γ^k is invariant by Theorem 8.4, so $\mu = \lambda - \gamma^k$ is also invariant and $\mu \geq 0$. Since P is irreducible, given $i \in I$, we have $p_{ik}^{(n)} > 0$ for some n, and $0 = \mu_k = \sum_j \mu_j p_{jk}^{(n)} \geq \mu_i p_{ik}^{(n)}$, so $\mu_i = 0$. □

8.2 Mean return time, positive and null recurrence

Recall that a state is recurrent if $P_i(V_i = \infty) = 1$, i.e.

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

We showed in Theorem 5.3 that this is equivalent to $f_i = 1$, i.e.

$$P_i(T_i < \infty) = 1.$$

If in addition, the **mean return time** (or **expected return time**)

$$m_i = E(T_i)$$

is finite, then we say that i is **positive recurrent**. A recurrent state that fails to have this property is called **null recurrent**, i.e. $m_i = \infty$.

We show in the next lecture that the following trichotomy holds for an irreducible Markov chain:

- P is transient;
- P is null-recurrent with $m_i = \infty$ for all i;
- P is positive-recurrent with an invariant distribution π, and $m_i \ (= 1/\pi_i) < \infty$ for all i.

8.3 Random surfer

Example 8.6 (Random surfer and PageRank)**.** The designer of a search engine trys to please three constituencies: searchers (who want useful search results), paying advertisers (who want to see 'click throughs'), and the search provider itself (who wants to maximize advertising revenue). The order in which search results are listed is a very important part of the design.

Google PageRank lists search results according to the probabilities with which a person randomly clicking on links will arrive at any particular page.

Let us model the web as a graph $G = (V, E)$ in which the vertices of the graph correspond to web pages. Suppose $|V| = n$. There is a directed edge from i to j if and only if page i contains a link to page j. Imagine that a random surfer jumps from his current page i by randomly choosing an outgoing link from amongst the $L(i)$ links if $L(i) > 0$, or if $L(i) = 0$ chooses a random page. Let

$$\hat{p}_{ij} = \begin{cases} 1/L(i) & \text{if } L(i) > 0 \text{ and } (i,j) \in E, \\ 1/n & \text{if } L(i) = 0. \end{cases}$$

To avoid the possibility that $\hat{P} = (\hat{p}_{ij})$ might not be irreducible or aperiodic (a complicating property that we discuss in the next lecture) we make a small adjustment, choosing $\alpha \in [0, 1)$, and set

$$p_{ij} = \alpha \hat{p}_{ij} + (1 - \alpha)(1/n).$$

In other words, we imagine that at each click the surfer chooses with probability α a random page amongst those that are linked from his current page (or a random page if there are no such links), and with probability $1 - \alpha$ chooses a completely random page. The invariant distribution satisfying $\pi P = \pi$ tells us the proportions of time that a random surfer will spend on the various pages. So if $\pi_i > \pi_j$ then i is more important than page j and should be ranked higher.

The solution of $\pi P = \pi$ was at the heart of Google's (original) page-ranking algorithm. In fact, it is not too difficult to estimate π because P^n converges very quickly to its limit

$$\Pi = 1\pi^\top = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \\ \pi_1 & \pi_2 & \cdots & \pi_n \\ \vdots & & & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix}$$

These days the ranking algorithm is more complicated and a trade secret.

9 Convergence to equilibrium of ergodic chains

9.1 Equivalence of positive recurrence and the existence of an invariant distribution

Theorem 9.1. *Let P be irreducible. Then the following are equivalent:*

(i) every state is positive recurrent;

(ii) some state i is positive recurrent;

(iii) P has an invariant distribution, π say.

Moreover, when (iii) holds we have $m_i = 1/\pi_i$ for all i.

Proof. (i) $\Longrightarrow$ (ii) is obvious. (ii) $\Longrightarrow$ (iii) If i is positive recurrent, it is certainly recurrent, so P is recurrent. By Theorem 8.4, γ^i is then invariant. But

$$\sum_{j\in I} \gamma^i_j = m_i < \infty$$

so $\pi_j = \gamma^i_k/m_i$ defines an invariant distribution.

(iii) $\Longrightarrow$ (i) Take any state k. Since P is irreducible and $\sum_{i\in I}\pi_i = 1$ we have $\pi_k = \sum_{i\in I}\pi_i p^{(n)}_{ik} > 0$ for some n. Set $\lambda_i = \pi_i/\pi_k$. Then λ is an invariant measure with $\lambda_k = 1$ So by Theorem 8.5, $\lambda \geq \gamma^k$. Hence

$$m_k = \sum_{i\in I}\gamma^k_i \leq \sum_{i\in I}\frac{\pi_i}{\pi_k} = \frac{1}{\pi_k} < \infty \tag{9.1}$$

and so k is positive recurrent.

To complete the proof we return to the argument for (iii) $\Longrightarrow$ (i) armed with the knowledge that P is recurrent and so $\lambda = \gamma^k$ and the inequality (9.1) is in fact an equality. □

Example 9.2 (Simple symmetric random walk on $\mathbb{Z}$)**.** The simple symmetric random walk on $\mathbb{Z}$ is clearly irreducible, and by Example 6.1 it is recurrent. Consider the measure $\pi_i = 1$ for all i. Then $\pi_i = \pi_{i-1}(1/2) + \pi_{i+1}(1/2)$ and so π is invariant. Any invariant distribution must be a scalar multiple of π. Since $\sum_{i\in I}\pi_i = \infty$ there can be no invariant distribution and so the walk is null recurrent.

Example 9.3. The existence of an invariant measure, does not guarantee recurrence. Consider, for example, the simple symmetric random walk on $\mathbb{Z}^3$, which is transient, but has invariant measure π given by $\pi = 1$ for all i.

9.2 Aperiodic chains

We next turn to an investigation of the limiting behaviour of the n-step transition probabilities $p_{ij}^{(n)}$ as $n \to \infty$. As we saw in Theorem 7.6, if the state-space is finite and if for some i the limit $p_{ij}^{(n)} \to \pi_j$ exists for all j, then π must be an invariant distribution. But, as the following example shows, the limit does not always exist.

Example 9.4. Consider the two-state chain with transition matrix

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

Then $P^2 = I = P^{2n}$ and $P^{2n+1} = P$. Thus $p_{ij}^{(n)}$ fails to converge for all i, j.

Let us call a state i **aperiodic** if $p_{ii}^{(n)} > 0$ for all sufficiently large n. The behaviour of the chain in Example 9.4 is connected with its periodicity.

Lemma 9.5. *A state i is aperiodic if there exist $n_1, \ldots, n_k \geq 1$ with no common factor and such that $p_{ii}^{(n_j)} > 0$, for all $j = 1, \ldots, k$.*

Proof. For all sufficiently large n we can write $n = a_1 n_1 + \cdots + a_k n_k$ using some non-negative integers $a_1, \ldots, a_k$. So then

$$p_{ii}^{(n)} \geq \underbrace{p_{ii}^{(n_1)} \cdots p_{ii}^{(n_1)}}_{a_1 \text{ times}} \underbrace{p_{ii}^{(n_2)} \cdots p_{ii}^{(n_2)}}_{a_2 \text{ times}} \cdots \underbrace{p_{ii}^{(n_k)} \cdots p_{ii}^{(n_k)}}_{a_k \text{ times}} > 0. \qquad \square$$

This lemma this shows that a state which is not aperiodic is rightly called **periodic**: because the set of n for which $p_{ii}^{(n)} = 0$ must essentially be, for some $d \geq 2$, the set of n which are divisible by d. One can also show that if two states communicate then they must have the same period.

Lemma 9.6. *Suppose P is irreducible, and has an aperiodic state i. Then, for all states j and k, $p_{jk}^{(n)} > 0$ for all sufficiently large n. In particular, all states are aperiodic.*

Proof. There exist $r, s > 0$ with $p_{ji}^{(r)}, p_{ik}^{(s)} > 0$. Then

$$p_{jk}^{(r+n+s)} \geq p_{ji}^{(r)} p_{ii}^{(n)} p_{ik}^{(s)} > 0$$

for all sufficiently large n. $\square$

A Markov chain is called a **regular chain** if some power of the transition matrix has only positive elements. We have shown above that if a Markov chain has a finite state space, is aperiodic and irreducible then it is regular.

An aperiodic, irreducible Markov and positive recurrent chain is called an **ergodic chain**.

9.3 Convergence to equilibrium *and proof by coupling*

We begin this section with a **magic trick**.

Example 9.7. (from Grinstead and Snell's 'Introduction to Probability', 2nd edition, page 452, also in Grimmett and Stirzaker)

Here is a trick to try on your friends. Shuffle a deck of cards and deal them out one at a time. Count the face cards each as ten. Ask your friend to look at one of the first ten cards; if this card is a six, she is to look at the card that turns up six cards later; if this card is a three, she is to look at the card that turns up three cards later, and so forth. Eventually she will reach a point where she is to look at a card that turns up x cards later but there are not x cards left. You then tell her the last card that she looked at even though you did not know her starting point. You tell her you do this by watching her, and she cannot disguise the times that she looks at the cards. In fact you just do the same procedure and, even though you do not start at the same point as she does, you will most likely end at the same point. Why?

Using a computer, I have simulated this game 10,000 times. In each game I have computing whether or not two players finish with the same last card, doing this for each of the 100 possible pairs of their initial choices of numbers. Between games I have shuffled the pack of cards by applying a random number of in- and out-shuffles that are chosen by a simulations of 20–100 Bernoulli trials. I find that the probability that the two players' last cards match one another is about 0.85. This increases to 0.90 if using two decks of cards and 0.96 with four decks of cards.

The proof of Theorem 9.8 that now follows is much like this magic trick in that it works by coupling two Markov chains. It is beautiful and ingenious. I like what J. Michael Steele says:

Coupling is one of the most powerful of the "genuinely probabilistic" techniques. Here by "genuinely probabilistic" we mean something that works directly with random variables rather than with their analytical co-travelers (like distributions, densities, or characteristic functions).

If you understand the proof of Theorem 9.8 and have a good picture what is going on in your mind, then I think you are developing excellent intuitive probabilistic insight.

Theorem 9.8 (Convergence to equilibrium)**.** *Let P be the transition matrix of an ergodic Markov chain (i.e. irreducible, aperiodic and positive recurrent), with invariant distribution π. Then (for any initial distribution) $P(X_n = j) \to \pi_j$ as $n \to \infty$ for all j. In particular*

$$p_{ij}^{(n)} \to \pi_j \text{ as } n \to \infty \text{ for all } i, j.$$

Proof. We use a coupling argument that is due to French mathematician Vincent Doblin (1937). Let $(X_n)_{n\geq 0}$ be and $(Y_n)_{n\geq 0}$ be independent and *Markov*(λ, P) and *Markov*(π, P), respectively.

Fix a reference state b and set $T = \inf\{n \geq 1 : X_n = Y_n = b\}$.

Step 1. We show $P(T < \infty) = 1$. To see this, observe that process $W_n = (X_n, Y_n)$ is a Markov chain on $I \times I$ with transition probabilities

$$\tilde{p}_{(i,k)(j,\ell)} = p_{ij} p_{k\ell}$$

and initial distribution

$$\mu_{(i,k)} = \lambda_i \pi_k.$$

Since P is aperiodic, for all states i, j, k, ℓ we have

$$\tilde{p}^{(n)}_{(i,k)(j,\ell)} = p^{(n)}_{ij} p^{(n)}_{k\ell} > 0$$

for sufficiently large n, so $\tilde{P}$ is irreducible. Also, $\tilde{P}$ has an invariant distribution of

$$\tilde{\pi}_{(i,k)} = \pi_i \pi_k.$$

So by Theorem 9.1, $\tilde{P}$ is positive recurrent. But T is the first passage time of W_n to (b, b) and so $P(T < \infty) = 1$ by Theorem 5.8.

Step 2. Since the initial distribution of Y_0 is the invariant distribution, we know that

$$\pi_i = P(Y_n = i) = P(Y_n = i, n \geq T) + P(Y_n = i, n < T).$$

Now $P(Y_n = i, n < T) \leq P(n < T)$. Also, $P(n < T) \to P(T = \infty) = 0$ and hence $P(Y_n = i, n < T) \to 0$ as $n \to \infty$. So we must have that $P(Y_n = i, n \geq T) \to \pi_i$. Also,

$$\begin{aligned} P(X_n = i) &= P(X_n = i, n \geq T) + P(X_n = i, n < T) \\ &= P(Y_n = i, n \geq T) + P(X_n = i, n < T). \end{aligned}$$

Again, $P(X_n = i, n < T) \to 0$, and we have seen above that $P(Y_n = i, n \geq T) \to \pi_i$. This proves $P(X_n = i) \to \pi_i$. □

To understand this proof one should see what goes wrong when P is not aperiodic. Consider the two-state chain of Example 9.4 which has $(1/2, 1/2)$ as its unique invariant distribution. We start $(X_n)_{n \geq 0}$ from 0 and $(Y_n)_{n \geq 0}$ with equal probability from 0 or 1. However, if $Y_0 = 1$, then, because of periodicity, $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$ will never meet, and the proof fails.

10 Long-run proportion of time spent in given state

10.1 Ergodic theorem

Theorem 10.1 (Strong law of large numbers)**.** *Let $Y_1, Y_2 \dots$ be a sequence of independent and identically distributed non-negative random variables with $E(Y_i) = \mu$. Then*

$$P\left(\frac{Y_1 + \cdots + Y_n}{n} \to \mu \text{ as } n \to \infty\right) = 1.$$

We do not give a proof.

Recall the definition $V_i(n) = \sum_{k=0}^{n-1} 1_{\{X_k = i\}}$ =number of visits to i before time n.

Theorem 10.2 (Ergodic theorem)**.** *Let P be irreducible and let λ be any distribution. Suppose $(X_n)_{0 \le n \le N}$ is Markov(λ, P). Then*

$$P\left(\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ as } n \to \infty\right) = 1$$

where $m_i = E_i(T_i)$ is the expected return time to state i.

Moreover, in the positive recurrent case, for any bounded function $f : I \to \mathbb{R}$ we have

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

where

$$\bar{f} = \sum_i \pi_i f(i)$$

and where $(\pi_i : i \in I)$ is the unique invariant distribution.

Notice that for this result we do not need P to be aperiodic.

Proof. We prove only the first part. If P is transient, then, with probability 1, the total number V_i of visits to i is finite, so

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

Suppose that P is recurrent and fix a state i. For $H = T_i$ we have $P(T < \infty) = 1$ by Theorem 5.8 and $(X_T + n)_{n \ge 0}$ is Markov(δ_i, P) and independent of $X_0, X_1, \dots, X_T$ by the strong Markov property. The long-run proportion of time spent in i is the same for $(X_{T+n})_{n \ge 0}$ and $(X_n)_{n \ge 0}$ so it suffices to consider the case $\lambda = \delta_i$.

Recall that the **first passage time** to state i is a random variable T_i defined by

$$T_i(\omega) = \inf\{n \ge 1 : X_n(\omega) = i\}$$

where $\inf \emptyset = \infty$. We now define inductively the ***r*th passage time** $T_i^{(r)}$ to state i by

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

and for $r = 0, 1, 2, \ldots$

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

The length of the ***r*th excursion** is then

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

Write $S_i^{(r)}$ for the length of the rth excursion to i. One can see, using the strong Markov property, that the non-negative random variables $S_i^{(1)}, S_i^{(2)}, \ldots$ are independent and identically distributed with $E_i(S_i^{(r)}) = m_i$. Now

$$S_i^{(1)} + \cdots + S_i^{(V_i(n)-1)} \leq n - 1,$$

the left-hand side being the time of the last visit to i before n. Also

$$S_i^{(1)} + \cdots + S_i^{(V_i(n))} \geq n,$$

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

$$\frac{S_i^{(1)} + \cdots + S_i^{(V_i(n)-1)}}{V_i(n)} \leq \frac{n}{V_i(n)} \leq \frac{S_i^{(1)} + \cdots + S_i^{(V_i(n))}}{V_i(n)}. \tag{10.1}$$

By the strong law of large numbers

$$P\left(\frac{S_i^{(1)} + \cdots + S_i^{(n)}}{n} \to m_i \text{ as } n \to \infty\right) = 1$$

and, since P is recurrent

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

So letting $n \to \infty$ in (10.1), we get

$$P\left(\frac{n}{V_i(n)} \to m_i \text{ as } n \to \infty\right) = 1.$$

which implies

$$P\left(\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ as } n \to \infty\right) = 1. \qquad \square$$

10.2 *Kemeny's constant and the random target lemma*

In this section we explore a bit further the calculation of mean recurrence times. Define

$$m_{ij} = E_i(T_j) = E_i(\inf\{n \geq 1 : X_n = j\})$$

so $m_{ii} = m_i$. How can we compute these mean recurrence times?

Throughout this section we suppose that the Markov chain is irreducible and positive recurrent. Suppose $(X_n)_{n\geq 0}$ is *Markov*(δ_i, P) and $(Y_n)_{n\geq 0}$ is *Markov*(π, P), where π is the invariant distribution. Define

$$z_{ij} := E\left[\sum_{n=0}^{\infty}\left(1_{\{X_n=j\}} - 1_{\{Y_n=j\}}\right)\right] = \sum_{n=0}^{\infty}(p_{ij}^{(n)} - \pi_j). \tag{10.2}$$

An application of the coupling idea (that eventually $X_n = Y_n$) can be used to show that this sum is finite. Then

$$\begin{aligned} z_{ij} &= E\left[\sum_{n=0}^{T_j-1}(1_{\{X_n=j\}} - \pi_j)\right] + E\left[\sum_{n=T_j}^{\infty}(1_{\{X_n=j\}} - \pi_j)\right] \\ &= \delta_{ij} - \pi_j E_i(T_j) + z_{jj}. \end{aligned}$$

Hence by putting $i = j$ we get $m_i = E_i(T_i) = 1/\pi_i$, as we have seen before. For $i \neq j$,

$$m_{ij} = E_i(T_j) = \frac{z_{jj} - z_{ij}}{\pi_j}.$$

We will figure out how to compute the z_{ij} shortly. First, let us consider the question of how long it takes to reach equilibrium. Suppose j is a random state that is chosen according to the invariant distribution. Then starting from state i, the expected number of steps required to hit state j (possibly 0 if $i = j$) is

$$\sum_{j\neq i}\pi_j E_i(T_j) = \sum_{j\neq i}(z_{jj} - z_{ij}) = \sum_j(z_{jj} - z_{ij}) = \sum_j z_{jj} \tag{10.3}$$

since from (10.2) we see that $\sum_j z_{ij} = 0$. The right hand side of (10.3) is (surprisingly!) a constant K (called **Kemeny's constant**) which is independent of i. This is known as the **random target lemma**. There is a nice direct proof also.

Lemma 10.3 (random target lemma)**.** *Starting in state i, the expected time of hitting a state j that is chosen randomly according to the invariant distribution, is $\sum_{j\neq i}\pi_j E_i(T_j) = K$ (a constant independent of i).*

Proof. We wish to compute $k_i = \sum_j \pi_j m_{ij} - \pi_i m_i$. The first step takes us 1 closer to the target state, unless (with probability π_i) we start in the target. So

$$k_i = 1 + \sum_j p_{ij}k_j - \pi_i m_i = \sum_j p_{ij}k_j.$$

Writing $k = (k_1, \ldots, k_n)$, this says $k = Pk$. Since each k_i is some weighted-average of $k_1, \ldots, k_n$ we see that k_i must be a constant, say $k_i = K$. □

Now we show how to compute the matrix $Z = (z_{ij})$.

Let Π be the matrix in which each row is π. Suppose that the Markov chain is ergodic (aperiodic, irreducible and positive recurrent). Then we know that $P^n \to \Pi$. It is easy to see that $\Pi P = P\Pi = \Pi\Pi = \Pi$ and so $(P^k - \Pi)^k = (P^k - \Pi)$. Let us define $\bar{Z}$ by

$$\begin{aligned} \bar{Z} := Z + \Pi &= I + (P - \Pi) + (P^2 - \Pi) + \cdots \\ &= I + (P - \Pi) + (P - \Pi)^2 + \cdots \\ &= (I - (P - \Pi))^{-1}. \end{aligned} \tag{10.4}$$

The right hand side of (10.4) converges since each terms is growing closer to 0 geometrically fast. Equivalently, we see that $(I - P + \Pi)$ has an inverse. For if there were to exist x such that $(I - P + \Pi)x = 0$ then $\pi(I - P + \Pi)x = \pi - \pi - \pi x = 0 \implies \pi x = 0 \implies \Pi x = 0 \implies (I - P)x = 0$. The only way this can happen is if x is a constant vector. Since π has only positive components this implies $x = 0$.

We have seen above that $m_{ij} = (z_{ij} - z_{jj})/\pi_j$. and that $K = \text{trace}(Z) = \text{trace}(\bar{Z}) - 1$. The eigenvalues of $\bar{Z}$ are the reciprocals of the eigenvalues of $\bar{Z}^{-1} = I - P + \Pi$, and these are the same as $1, 1 - \lambda_2, \ldots, 1 - \lambda_N$, where $1, \lambda_2, \ldots, \lambda_N$ are the eigenvalues of P. Thus since the trace of a matrix is the sum of its eigenvalues,

$$K = \sum_{i=2}^{N} \frac{1}{1 - \lambda_i}.$$

Recall that in the two-state chain of Example 1.4 the eigenvalues are $(1, 1 - \alpha - \beta)$. So in this chain $K = 1/(\alpha + \beta)$.

11 Time reversal, detailed balance, reversibility, random walk on a graph

11.1 Time reversal

For Markov chains, the past and future are independent given the present. This property is symmetrical in time and suggests looking at Markov chains with time running backwards. On the other hand, convergence to equilibrium shows behaviour which is asymmetrical in time: a highly organised state such as a point mass decays to a disorganised one, the invariant distribution. This is an example of entropy increasing. So if we want complete time-symmetry we must begin in equilibrium. The next result shows that a Markov chain in equilibrium, run backwards, is again a Markov chain. The transition matrix may however be different.

Theorem 11.1. *Let P be irreducible and have an invariant distribution π. Suppose that $(X_n)_{0\le n\le N}$ is Markov(π, P) and set $Y_n = X_{N-n}$. Then $(Y_n)_{0\le n\le N}$ is Markov$(\pi, \hat{P})$ where $\hat{P}$ is given by*

$$\pi_j \hat{p}_{ji} = \pi_i p_{ij}$$

and P is also irreducible with invariant distribution π.

Proof. First we check that P is a stochastic matrix:

$$\sum_i \hat{p}_{ji} = \frac{1}{\pi_j} \sum_i \pi_i p_{ij} = 1$$

since π is invariant for P. Next we check that π is invariant for $\hat{P}$:

$$\sum_{j\in I} \pi_j \hat{p}_{ji} = \sum_{j\in I} \pi_i p_{ij} = \pi_i$$

since P is a stochastic matrix.

We have

$$\begin{aligned} P(Y_0 = i_0, Y_1 = i_1, \dots, Y_N = i_N) &= P(X_0 = i_N, X_1 = i_{N-1}, \dots, X_N = i_0) \\ &= \pi_{i_N} p_{i_N i_{N-1}} \cdots p_{i_1 i_0} \\ &= \pi_{i_0} \hat{p}_{i_0 i_1} \cdots \hat{p}_{i_{N-1} i_N} \end{aligned}$$

so by Theorem 1.3 $(Y_n)_{0\le n\le N}$ is *Markov*$(\pi, \hat{P})$.

Finally, since P is irreducible, for each pair of states i, j, there is a chain of states $i_0 = i, i_i, \dots, i_{n-1}, i_n = j$ with $p_{i_0 i_1} \cdots p_{i_{n-1} i_n} > 0$. So

$$\hat{p}_{i_n i_{n-1}} \cdots \hat{p}_{i_{n-1} i_n} = \pi_{i_0} p_{i_0 i_1} \cdots \hat{p}_{i_{n-1} i_n} / \pi_{i_n} > 0$$

and thus $\hat{P}$ is also irreducible. □

11.2 Detailed balance

A stochastic matrix P and a measure λ are said to be in **detailed balance** if

$$\lambda_i p_{ij} = \lambda_j p_{ji} \text{ for all } i, j.$$

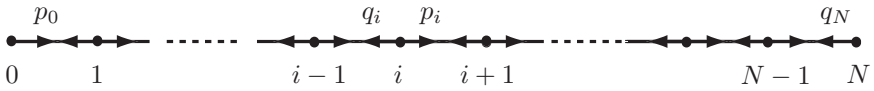
Though obvious, the following result is worth remembering because when a solution λ to the detailed balance equations exists, it is often easier to find by the detailed balance equations than by the equation $\lambda = \lambda P$.

Lemma 11.2. *If P and λ are in detailed balance, then λ is invariant for P.*

Proof. We have $(\lambda P)_i = \sum_j \lambda_j p_{ji} = \sum_j \lambda_i p_{ij} = \lambda_i$. □

Examples 11.3. (a) A two-state chain like $P = \begin{pmatrix} 1/2 & 1/2 \\ 2/3 & 1/3 \end{pmatrix}$.

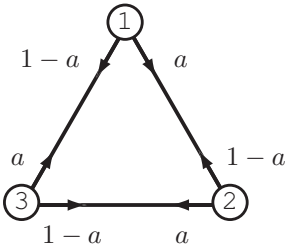
(b) Any walk on $0, \ldots, N$ with $(p_{i,i-1}, p_{i,i+1}) = (q_i, p_i)$:



We have $\pi_i p_i = \pi_{i+1} q_{i+1} \implies \pi_i = \dfrac{p_{i-1} \cdots p_0}{q_i \cdots q_1} \pi_0$.

(c) Consider walk on a triangle with

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



Detailed balance equations are

$$\pi_1 a = \pi_2 (1-a)$$
$$\pi_2 a = \pi_3 (1-a)$$
$$\pi_3 a = \pi_1 (1-a)$$

So there is a solution if $a = 1/2$. But not if $a \neq 1/2$.

11.3 Reversibility

Let $(X_n)_{n\geq 0}$ be *Markov*(λ, P), with P irreducible. We say that $(X_n)_{n\geq 0}$ is **reversible** if, for all $N \geq 1$, $(X_{N-n})_{0\leq n\leq N}$ s also *Markov*(λ, P).

Theorem 11.4. *Let P be an irreducible stochastic matrix and let λ be a distribution. Suppose that $(X_n)_{n\geq 0}$ is Markov(λ, P). Then the following are equivalent:*

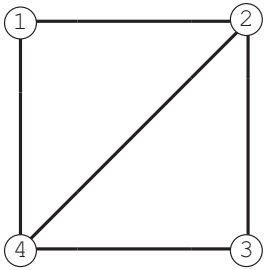
(a) $(X_n)_{n\geq 0}$ is reversible;

(b) P and λ are in detailed balance.

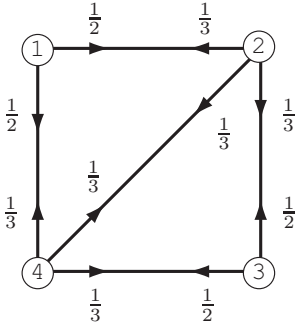
Proof. Both (a) and (b) imply that λ is invariant for P. Then both (a) and (b) are equivalent to the statement that $\hat{P} = P$ in Theorem 11.1. □

11.4 Random walk on a graph

Example 11.5. Random walk on a graph. A graph G is a countable collection of states, usually called vertices., some of which are joined by edges, for example:



Thus a graph is a partially drawn Markov chain diagram. There is a natural way to complete the diagram which gives rise to the random walk on G.



The valence v_i of vertex i is the number of edges at i. We have to assume that every vertex has finite valence. The random walk on G picks edges with equal probability:

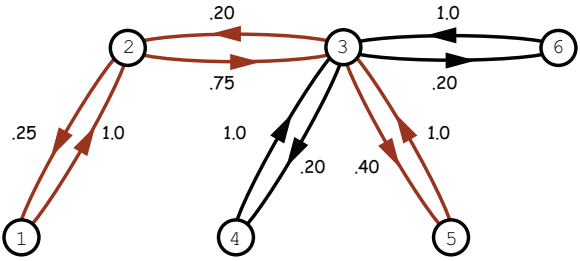
Thus the transition probabilities are given by

$$p_{ij} = \begin{cases} 1/v_i & \text{if } (i,j) \text{ is an edge} \\ 0 & \text{otherwise} \end{cases}$$

We assume G is connected, so that P is irreducible. It is easy to see that P is in detailed balance with $v = (v_i : i \in G)$. So, if the total valence $\sigma = \sum_i v_i$ is finite. then $\pi_i = v_i/\sigma$ is invariant and P is reversible.

The invariant distribution in this example is $\pi = (1/10)(2, 3, 3, 2)$. So, for example, $m_1 = 1/\pi_i = 5$.

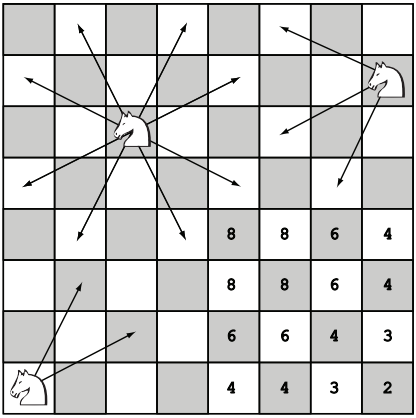
Example 11.6. Random walk on a tree with arbitrary transition probabilities.



We can clearly solve detailed balance equations if there are no cycles. E.g.

$$\pi_1 \times 1.0 \times 0.75 \times 0.40 = \pi_5 \times 1.0 \times 0.20 \times 0.25$$

Example 11.7 (Random chess board knight)**.** A random knight makes each permissible move with equal probability. If it starts in a corner, how long on average will it take to return?



This is an example of a random walk on a graph: the vertices are the squares of the chess board and the edges are the moves that the knight can take. The diagram shows valences of the 64 vertices of the graph. We know by Theorem 9.1 and the preceding example that

$$E_c(T_c) = 1/\pi_c = \frac{\sum_i v_i}{v_c}$$

so all we have to do is, identify valences. The four corner squares have valence 2; and the eight squares adjacent to the corners have valence 3. There are 20 squares of valence 4; 16 of valence 6; and the 16 central squares have valence 8. Hence,

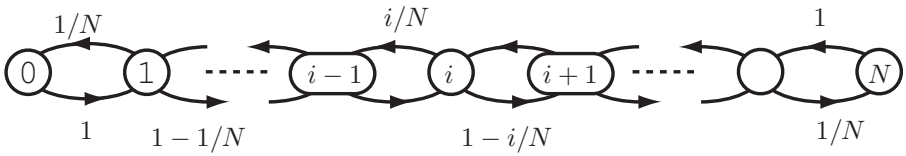
$$E_c(T_c) = \frac{8 + 24 + 80 + 96 + 128}{2} = 168.$$

Obviously this is much easier than solving sets of 64 simultaneous linear equations to find π from $\pi P = \pi$, or calculating $E_c(T_c)$ using Theorem 3.4.

12 Concluding problems and recommendations for further study

12.1 Reversibility and Ehrenfest's urn model

Example 12.1 (Ehrenfest's urn model)**.** In Ehrenfest's model for gas diffusion N balls are placed in 2 urns. Every second, a ball is chosen at random and moved to the other urn (Paul Ehrenfest, 1880–1933). At each second we let the number of balls in the first urn, i, be the state. From state i we can pass only to state $i-1$ and $i+1$, and the transition probabilities are given by



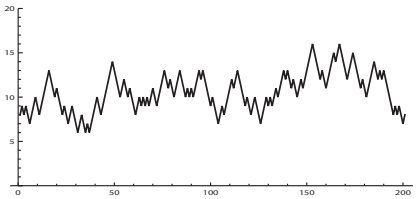
This defines the transition matrix of an irreducible Markov chain. Since each ball moves independently of the others and is ultimately equally likely to be in either urn, we can see that the invariant distribution π is the binomial distribution with parameters N and $1/2$. It is easy to check that this is correct (from detailed balance equations). So,

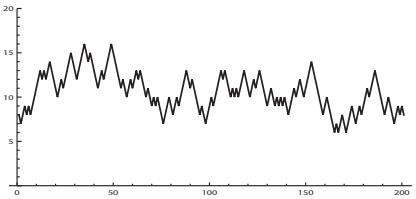
$$\pi_i = \frac{1}{2^N}\binom{N}{i}, \quad \text{and} \quad m_i = 2^N \Big/ \binom{N}{i}.$$

Consider in particular the central term $i = N/2$. We have seen that this term is approximately $1/\sqrt{\pi N/2}$. Thus we may approximate $m_{N/2}$ by $\sqrt{\pi N/2}$.

Assume that we let our system run until it is in equilibrium. At this point, a video is made, shown to you, and you are asked to say if it is shown in the forward or the reverse direction. It would seem that there should always be a tendency to move toward an equal proportion of balls so that the correct order of time should be the one with the most transitions from i to $i-1$ if $i > N/2$ and i to $i+1$ if $i < N/2$. However, the reversibility of the process means this is not the case.

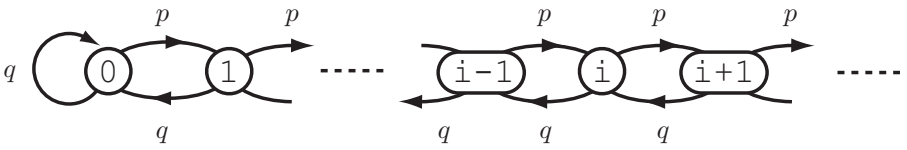
The following chart shows $\{X_n\}_{n=300}^{500}$ in a simulation of an urn with 20 balls, started at $X_0 = 10$. Next to it the data is shown reversed. There is no apparent difference.





12.2 Reversibility and reflected walk

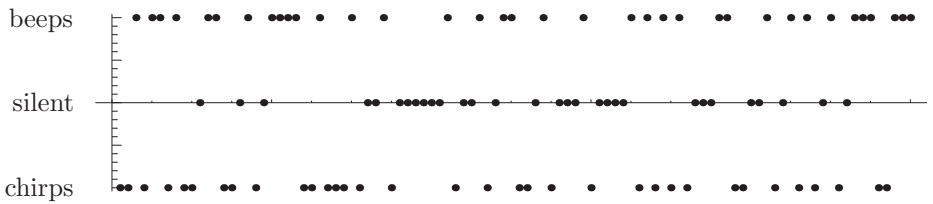
Example 12.2. Consider the following random walk with $0 < p = 1 - q < 1$.



Suppose we do not observe X_n, but whenever there is transition to the right we hear a beep. Whenever there is a transition to the left we hear a chirp. There is silence whenever the transition $0 \to 0$ occurs. What if the distribution of time between successive beeps? between successive chirps?

The distance between two successive rightward steps is distributed as a geometric random variable Y with parameter p, such that $P(Y = k) = q^{k-1}p$, $k = 1, 2, \ldots$. So this is the distribution of the time between beeps. Clearly $EY = 1/p$. As for the distance between leftward steps, these are distributed as a geometric random variable, Z, with $P(Z = k) = p^{k-1}q$, and $EZ = 1/q$, but only if $X_n > 0$. If $X_n = 0$ then the time until the next chirp has the distribution of $Y + Z$. This suggests that we might be able to hear some difference in the processes of beeps and chirps.

However, $(X_n)_{n\geq 0}$ is reversible. So once we have reached equilibrium the reversed process looks the same as forward process. Leftward and rightward steps are interchanged when we run the process in reverse. This means that the statistics of the chirps process must be the same as that of beeps process, and so the distribution between successive chirps is also a geometric random variable with parameter p. The following simulations is for $p = 0.45$.



Burke's Theorem (1956) is a version of this result in continuous time. It states that in a $M/M/1$ queue (in which arrivals form a Poisson process of rate λ, there is a single server, and service times are independent and exponentially distributed with mean $1/\mu$ (and $\mu > \lambda$)) the steady state departure process looks like a Poisson process with rate parameter λ. This means that someone who watches only the departure process cannot tell that there is a queue located between the input and output!

12.3 *The probabilistic abacus*

Peter Winkler (*Mathematical Mind-Benders*, 2007) poses the following puzzle. "In a circle are n children, each holding some gumdrops. The teacher gives an additional gumdrop to each child who has an odd number, then each child passes half of his or her gumdrops to the child on his left or on his right. These two steps are repeated until they no longer have any effect. Prove that this process will indeed eventually terminate, with all kids have the same (even) number of gumdrops."

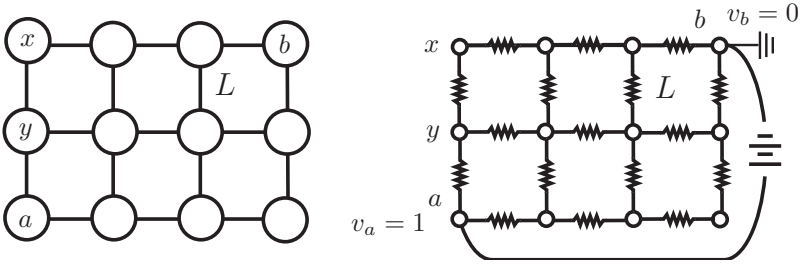
To solve this puzzle, let $m_i(0)$ be the number of gumdrops held by the child i just before they pass for the first time. Let $M = \max_i m_i(0)$. Notice that $m_i(1) \leq M$ for all i, and inductively $m_i(t) \leq M$ for all i and t. But until the game ends the teacher gives out at least 1 gumdrop per round. Hence the game must terminate within a finite number of rounds.

Grinstead and Snell describe a similar algorithm which works for more general Markov chains as follows (pages 445–447). Let $P = (p_{ij})$ be the $n \times n$ transition matrix of an irreducible and aperiodic Markov chain of n states, with all entries rational. A class of n children plays a game of successive rounds. Initially, child i has $m_i(0)$ gumdrops, and at end of a round t she has $m_i(t)$ gumdrops. The teacher then hands out gumdrops so that each child has $n_i(t)$ gumdrops, where $n_i(t)$ is the least number not less than $m_i(t)$ such that $p_{ij}n_i(t)$ is an integer, for every j. Finally, for each i and j, child i passes $p_{ij}n_i(t)$ gumdrops to child j. That completes the round, child i now has $m_i(t+1)$ gumdrops, and the game continues to the round $t+1$. Again, we can see that in a finite number of rounds we are guaranteed to reach a situation in which the teacher does not have to give out any more gumdrops, and that the holdings of gumdrops correspond to an invariant measure.

(This is the algorithm for computing an invariant distribution that is described by Arthur Engel, Wahrscheinlichkeitsrechnung und Statistik, vol. 2 (Stuttgart: Klett Verlag, 1976). He also describes an algorithm for an absorbing Markov chain that calculates the absorption probabilities. See Appendix C.)

12.4 *Random walk and electrical networks*

Recall the problem of finding absorption probabilities. Consider random walk on a graph and let $h_x = P_x(H_a < H_b)$, the probability when starting from x of hitting a before b.



This satisfies the right hand equations $h = Ph$, with boundary conditions, i.e.

$$h_x = \sum_y p_{xy} h_y = \sum_y \frac{1}{d_x} h_y, \ x \neq a, b, \quad \text{with } h_a = 1, \ h_b = 0. \tag{12.1}$$

where d_x is the degree of node x.

Now consider the electrical network in which a 1 ohm resistor replaces each link and a battery holds the voltages at a and b to 1 and 0 respectively. By Ohm's Law the current from x to y is

$$i_{xy} = \frac{\text{voltage drop from } x \text{ to } y}{\text{resistence between } x \text{ and } y} = \frac{v_x - v_y}{1}.$$

By Kirchhoff's Laws, the current into x must be equal to the current out. So

$$0 = \sum_y i_{xy} = \sum_y (v_x - v_y) \implies v_x = \sum_y \frac{1}{d_x} v_y.$$

Thus v satisfies the same equations (12.1) as does h and so $h = v$. A function h which satisfies $h = Ph$ is said to be a **harmonic function**.

We can now consider the following questions: (i) what is the probability, starting at a that we return to a before hitting b? (ii) how does this probability change if we remove a link, such as L?

We have seen that to answer (i) we just put a battery between a and b, to fix $v_a = 1$ and $v_b = 0$, and unit resistors at each link. Then v_y is the probability that a walker who starts at y hits a before b. Also, $\sum_y p_{ay} v(y)$ is the probability that a walker who starts at a returns to a before reaching b.

The current flowing out of a (and in total around the whole network) is

$$\sum_y i_{ay} = \sum_y p_{ay}(v_a - v_y) = 1 - \sum_y p_{ay} v_y = 1 - P_a(H_a < H_b).$$

Rayleigh's Monotonicity Law says that increasing the resistance of any resistor in the network cannot decrease the resistance between any two points. So if we remove a link, such as L, (which is equivalent to increasing the resistance of the corresponding resistor to infinity), then the resistance between points a and b increases, and so that a voltage difference of 1 drives less current around the network and thus the quantity $\sum_y p_{ay} v(y)$ increases, i.e. $P_a(H_a < H_b)$ increases. It is perhaps a bit surprising that the removal of a link always has this effect, no matter where that link is located in the graph.

We may let the size of the network increase towards that of the infinite lattice of resistors and ask "What happens to the effective resistence between the node at the centre and those nodes around the outer boundary as the distance between them tends to infinity?". By what we know of recurrence properties of random walk, it must be that in 1 and 2 dimensions the effective resistence tends to infinity, whereas in 3 dimensions it tends to something finite.

(*This next bit will probably not be lectured.*) We can also give a nice interpretation of current. Suppose the walker begins at a and walks until he reaches b; note that if he returns to a before reaching b, he keeps on going. Let $u_x = E_a[V_x(T_b)]$ be the expected number of visits to state x before reaching b. Clearly, for $x \neq a, b$,

$$u_x = \sum_y u_y p_{yx}, \quad \text{which is equivalent to } \frac{u_x}{d_x} = \sum_y p_{xy} \frac{u_y}{d_y}.$$

Thus u_x/d_x is harmonic, with $u_b = 0$. So $u = v$ for a network in which we fix $v_b = 0$. The current from x to adjacent node y is

$$i_{xy} = v_x - v_y = \frac{u_x}{d_x} - \frac{u_y}{d_y} = u_x p_{xy} - u_y p_{yx}.$$

Now $u_x p_{xy}$ is just the expected value of the number of times that the walker will pass along the edge from x to y. Similarly, $u_y p_{yx}$ is just the expected value of the number of times that the walker will pass along the edge from y to x. So the current i_{xy} has an interpretation as the expected value of the net number of times that the walker will pass along the edge from x to y, when starting at a and continuing until hitting b. We should fix v_a so that $\sum_y i_{ay} = 1$ since, given a start at a, this is necessarily 1.

12.5 Probability courses in Part II

I will say a few sentences about how things are developed in Part II courses. All of these build on Probability IA, and your knowledge of Markov chains will also be useful (though that is not essential).

Probability and Measure: definitions and properties of notions such as expectation and the strong law of large numbers are developed more fully and rigorously than in Probability IA. Limits are a key theme. For example, one can prove that that certain 'tail events' can only have probability of 0 or 1.

An example of this we have seen: the event $\{T_i < \infty \mid X_0 = i\}$ has probability 0 or 1 (never 1/2).

Applied Probability: Markov processes in continuous time, with applications in queueing, telecommunication networks, insurance ruin, and epidemics.

Imagine the frog in Example 1.1 waits for an exponentially distributed time before hopping to a new lily pad. What now is $p_{57}(t)$, $t \geq 0$?

Optimization and Control: Markov chains combined with costs, rewards, and optimization. A controller can pick the transition matrix $P_{ij}(a)$, to be a function of his action a (e.g. how should a gambler bet who wants to maximize his change of reaching 10 pounds? How many flowers should the theatre owner send to a disguntled opera singer?).

Perhaps we would like to steer the frog to arrive at some particular lily pad in the least possible time, or at last cost.

Suppose three frogs are placed at different nodes of $\mathbb{Z}^2$. At each step we can choose one of the frogs to make a random hop to one of its neighbouring nodes. We wish to make our choices so as to minimize the expected time until we first have a frog at the origin.

Stochastic Financial Models: more on random walks, Brownian motion, stochastic models used to price financial products.

We might like to design optimal bets on where the frog will end up, or hedge the risk that he ends up in an unfavourable place.

We might known that the frog will eventually reach state 1 or 2, and then win a prize worth $£r_i$, $i = 1, 2$. How much would we be willing to pay for the right (option) to buy his final prize for $£s$?

I hope some of you will want to do all these courses in Part II!

Appendix

A Probability spaces

The models in this course take place in a **probability space** $(\Omega, \mathscr{F}, P)$. Let us briefly review what this means.

- Ω is a set of outcomes ($\omega \in \Omega$)
- $\mathscr{F}$ is a nonempty set of subsets of Ω (corresponding to possible events), and such that

  (i) $\Omega \in \mathscr{F}$

  (ii) $A \in \mathscr{F} \implies A^c \in \mathscr{F}$

  (iii) $A_n \in \mathscr{F}$ for $n \in \mathbb{N} \implies \bigcup_{n\in\mathbb{N}} A_n \in \mathscr{F}$ (a union of countably many members of $\mathscr{F}$)

  Note that these imply $\bigcap_{n\in\mathbb{N}} A_n \in \mathscr{F}$, because $\bigcap_n A_n = (\bigcup_n A_n^c)^c$.

  A collection of subsets with these properties is called a **$\boldsymbol{\sigma}$-algebra** or σ-field. ('sigma' refers to the fact we are dealing with countable unions.)

- $P : \mathscr{F} \to [0,1]$ satisfies
  - $P(\Omega) = 1$
  - $A_n \in \mathscr{F}$ for $n \in \mathbb{N}$, A_n disjoint $\implies P(\bigcup_n A_n) = \sum_{n=1}^{\infty} P(A_n)$.

Note that this sum is well-defined, independently of how we enumerate the A_n. For if $(B_n : n \in \mathbb{N})$ is an enumeration of the same sets, but in a different order, then given any n there exists an m such that

$$\bigcup_{i=1}^{n} A_i \subseteq \bigcup_{i=1}^{m} B_i$$

and it follows that $\sum_{i=1}^{\infty} P(A_i) \leq \sum_{i=1}^{\infty} P(B_i)$. Clearly, $\geq$ is also true, so the sums are equal. Another important consequence of these axioms is that we can take limits.

Proposition A.1. *If $(A_n)_{n\in\mathbb{N}}$ is a sequence of events and $\lim_{n\to\infty} A_n = A$, then $\lim_{n\to\infty} P(A_n) = P(A)$.*

We use this repeatedly. For example, if V_i is the number of visits to i, then $P_i(V_i = \infty) = P_i(\lim_{k\to\infty}[V_i \geq k]) = \lim_{k\to\infty} P_i(V_i \geq k) = \lim_{k\to\infty} f_i^k$, which is 0 or 1, since $0 \leq f_i \leq 1$ (we cannot have $P_i(V_i = \infty) = 1/2$).

- $X : \Omega \to I$ is a **random variable** if

$$\{X = i\} = \{\omega \in \Omega : X(\omega) = i\} \in \mathscr{F} \text{ for all } i \in I.$$

Then $P(X = i) = P(\{X = i\})$. This is why we require $\{X = i\} \in \mathscr{F}$.

B Historical notes

Grinstead and Snell (Introduction to Probability, 1997) say the following (pages 464–465):

"Markov chains were introduced by Andrei Andreevich Markov (1856–1922) and were named in his honor. He was a talented undergraduate who received a gold medal for his undergraduate thesis at St. Petersburg University. Besides being an active research mathematician and teacher, he was also active in politics and participated in the liberal movement in Russia at the beginning of the twentieth century. In 1913, when the government celebrated the 300th anniversary of the House of Romanov family, Markov organized a counter-celebration of the 200th anniversary of Bernoulli's discovery of the Law of Large Numbers.

Markov was led to develop Markov chains as a natural extension of sequences of independent random variables. In his first paper, in 1906, he proved that for a Markov chain with positive transition probabilities and numerical states the average of the outcomes converges to the expected value of the limiting distribution (the fixed vector). In a later paper he proved the central limit theorem for such chains. Writing about Markov, A. P. Youschkevitch remarks:

> Markov arrived at his chains starting from the internal needs of probability theory, and he never wrote about their applications to physical science. For him the only real examples of the chains were literary texts, where the two states denoted the vowels and consonants.[1] ed. C. C. Gillespie (New York: Scribner's Sons, 1970), pp. 124–130.)

In a paper written in 1913,[2] Markov chose a sequence of 20,000 letters from Pushkin's *Eugene Onegin* to see if this sequence can be approximately considered a simple chain. He obtained the Markov chain with transition matrix

$$\begin{pmatrix} .128 & .872 \\ .663 & .337 \end{pmatrix}$$

The fixed vector for this chain is (.432, .568), indicating that we should expect about 43.2 percent vowels and 56.8 percent consonants in the novel, which was borne out by the actual count."

[1] See 'Dictionary of Scientific Biography,

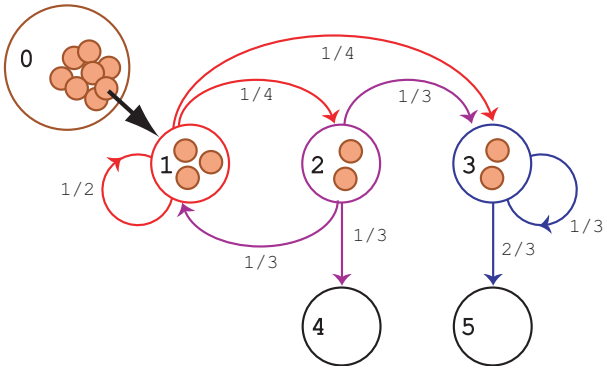
[2] A. A. Markov, "An Example of Statistical Analysis of the Text of Eugene Onegin Illustrating the Association of Trials into a Chain," *Bulletin de l'Acadamie Imperiale des Sciences de St. Petersburg,* ser. 6, vol. 7 (1913), pp. 153–162.

C The probabilistic abacus for absorbing chains

This section is about Engel's 'probabilistic abacus', an algorithmic method of computing the absorption probabilities of an absorbing Markov chain. The version of the abacus that calculates the invariant distribution of a recurrent Markov chain has already been described in §12.3. These algorithms are of no practical value! However, they are interesting because it is challenging to understand why they work, and thinking about that will help to build your intuition for Markov processes.

Consider an absorbing Markov chain with rational transition probabilities, as in

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



Clearly $\{1, 2, 3\}$ are transient and $\{4, 5\}$ are absorbing.

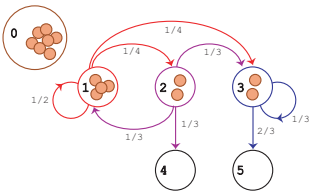
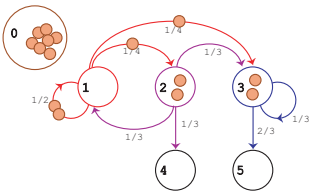
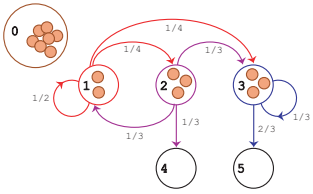
Suppose we want to find α_{14}, the probability that starting in state 1 absorption takes place in state 4. We can do this by Engel's algorithm, which is equivalent to playing a so-called **chip firing game**. It works like this. We create one node for each state and put some chips (or tokens) at the nodes corresponding to the non-absorbing states, $\{1, 2, 3\}$. Suppose that there are integers $r_i, r_{i1}, \ldots, r_{in}$ such that $p_{ij} = r_{ij}/r_i$ for all j. If there were r_i chips at node i we could 'fire' or 'make a move' in node i. This means taking r_i chips from node i and moving r_{ij} of them to node j, for each j.

The *critical loading* is one in which each node has one less chip that it needs to fire, i.e. $c_i = r_i - 1$. So $c_1 = 3$ and $c_2 = c_3 = 2$. We start with a critical loading by placing tokens at nodes 1, 2, 3, 4, 5 in numbers:

$$(3, 2, 2, 0, 0),$$

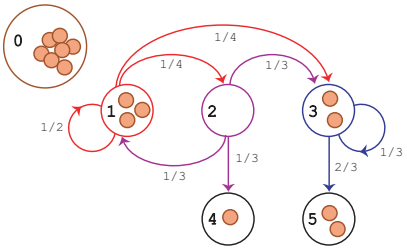
as shown above, and then add a large number of tokens to another node 0. 'Firing' node 0 means moving one token from node 0 to node 1. Engel's algorithm also imposes the rule that node 0 may be fired only if no other node can fire. Starting from the critical loading we fire node 0 and then node 1:

$$(3, 2, 2, 0, 0) \xrightarrow{0} (4, 2, 2, 0, 0) \xrightarrow{1} (2, 3, 3, 0, 0)$$

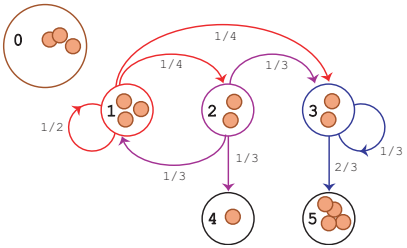
Now nodes 2 or 3 could fire. Suppose we fire 3 and then 2:

$$(2,3,3,0,0) \xrightarrow{3} (2,3,1,0,2) \xrightarrow{2} (3,0,2,1,2)$$



Notice that the same loading would be reached if we had fired 2 and then 3. This fact will be important later on. Now we fire the sequence 0, 1, 3, 0, 0, 1, 0:

$$(3,0,2,1,2) \xrightarrow{0} (4,0,2,1,2) \xrightarrow{1} (2,1,3,1,2) \xrightarrow{3} (2,1,1,1,4)$$
$$\xrightarrow{0} (3,1,1,1,4) \xrightarrow{0} (4,1,1,1,4) \xrightarrow{1} (2,2,2,1,4) \xrightarrow{0} (3,2,2,1,4)$$



At this point we stop, because nodes 1, 2 and 3 now have exactly the same loading as at the start. We are at $(3,2,2,0,0) + (0,0,0,1,4)$. We have fired 0 five times and ended up back at the critical loading, but with 1 token in node 4 and 4 tokens in node 5. Thus $\alpha_{14} = 1/5$ and $\alpha_{15} = 4/5$.

Why does this algorithm work? It is fairly obvious that if the initially critically loaded configuration of the transient states reoccurs then the numbers of tokens that have appeared in the nodes that correspond to the absorbing states must be in quantities that are in proportion to the absorptions probabilities, α_{1j}. But why is the initial critically loaded configuration guaranteed to eventually reappear? This puzzled Engel

in 1976, and was proved circa 1979 by Peter Doyle, whose proof we now present. It is much more difficult than to prove that the abacus described in §12.3 terminates. (The proof is interesting. There is, in fact, a rich literature on the properties of chip firing games and this proof generalises to show that many chip firing games have the property that the termination state does not depend on the order in which moves are made.)

Proof that the critical loading reoccurs.

1. At times that we fire node 0 the loadings of the nodes corresponding to the non-absorbing states are no greater than their critical loadings. There are only finitely many such possible loadings. So some loading must reoccur. Call this loading L.

2. Start by loading L in white tokens. By hypothesis, if we continue from this initially loading then we will eventually return to it after some sequence of firings, say $f = (f_0, f_{i_2}, \ldots, f_{i_N})$, of which m of them are firings of node 0. Let us add red tokens to each node so as to take the loading up to the critical loading. Put just m white tokens at node 0. We now compare the result of two different strategies of play.

Strategy I is to follow Engel's algorithm until we run out of tokens in node 0.

Strategy II is to make the firing sequence f, using only white tokens. Here we no longer impose the rule that node 0 can fire only if no other node can. When we run out of token in node 0, L reoccurs and so we are back at the critical loading.

The proof is finished by proving that from any initial loading, the loading obtained at the end, when the m tokens in node 0 have run out, is independent of any choices that we make along the way about which of several available nodes we might fire. Assuming this is true, then by starting in the critical loading and following Engel's algorithm until tokens in node 0 run out, we end in the same place as Strategy II, which we know ends in the critical loading. So the critical loading reoccurs under Engel's algorithm also.

3. To finish, we must prove our claim. Suppose that up to the point that no more firings become possible Strategy I fires sequentially the nodes $e_1, \ldots, e_N$, and Strategy II fires sequentially the nodes $f_1, \ldots, f_M$. We now show that these are simply the same set of firing arranged in different orders, and hence that $N = M$ and the terminal loadings are identical.

Consider firing e_1. If none of $f_1, \ldots, f_M$ are node e_1 then this sequence f cannot have terminated, because there must be enough tokens in node e_1 to fire it. So some $f_{i_1} = e_1$. Suppose we have proved that $e_1, \ldots, e_k$ are in 1–1 correspondence with some $f_{i_1}, \ldots, f_{i_k}$. Consider e_{k+1}. If there is a f_{k+1} amongst the f_i which occurs in addition to $f_{i_1}, \ldots, f_{i_k}$, and before these are all carried out, then we can choose to match this to e_{k+1}. Suppose there was no such f_i. The firing e_{k+1} was made possible by $e_1, \ldots, e_k$. It must also have been made possible by $f_{i_1}, \ldots, f_{i_k}$. Any other firings that have been made can only help by adding tokens to node e_{k+1}. So there must be at least one more firing subsequent to all of $f_{i_1}, \ldots, f_{i_k}$ that is of node e_{k+1}. So again, we can make a match. Thus, inductively, the e_i's are contained within the f_i's. By that same argument the reverse is true and we conclude that the two sets of firings are the same. □

Index

absorbing, 8
absorption probability, 9
almost surely, 26
aperiodic, 34

birth-death chain, 13

Chapman-Kolomogorov equations, 3
chip firing game, 53
class property, 18
closed class, 8, 19
communicating classes, 8
convergence to equilibrium, *see* ergodic theorem
coupling, 35

detailed balance, 27, 42
distribution, 2
Doblin, V., 35

Ehrenfest's urn model, 45
electrical networks, 47
Engel's probabilistic abacus, 47, 53
equilibrium distribution, 25, 28
equivalence relation, 8
ergodic chain, 34
ergodic theorem, 37
evening out the gumdrops, 46
expected return time, *see* mean return time

first passage time, 17, 25, 37

gambler's ruin, 12, 15

harmonic function, 48
hitting time, 9, 17

initial distribution, 2
invariant distribution, 25–33
    existence and uniqueness, 29
invariant measure, 25
irreducible, 8

Kemeny's constant, *see* random target lemma

left hand equations, 26

magic trick, 35
Markov chain, 2
    simulation of, 3
Markov property, 6
Markov, A. A., 1, 52
mean hitting time, 9
mean recurrence times, 39
mean return time, 17, 25, 31
measure, 2

n-step transition matrix, 3
null recurrent, 31

open class, 8

periodic, 34
Pólya, G., 23
positive recurrent, 31
probabilistic abacus, 47, 53
probability space, 2, 51

random surfer, 32
random target lemma, 39
random variable, 51
random walk, 10, 33
    and electrical networks, 47–49
    of a knight on a chessboard, 44
    on a complete graph, 6
    on a graph, 43
    on a triangle, 42
    on $\{0, 1, \ldots, N\}$, 10
    on $\{0, 1, \ldots\}$, 12
    on $\mathbb{Z}^d$, 21–24
recurrent, 17
regular chain, 34
return probability, 17
reversibility, 42–46

reversible, 43
right hand equations, 11
rth excursion, 38
rth passage time, 38

σ-algebra, 51
σ-field, 51
state, 2
state-space, 2
stationary distribution, 25, 28
steady state distribution, 25
stochastic matrix, 2
stopping time, 15
strong law of large numbers, 37
strong Markov property, 15
success-run chain, 27

transient, 17
transition matrix, 2
two-state Markov chain, 4